ANALYSIS II

Classroom Notes

with an Appendix: German Translation of Section 7

H.-D. Alber

Contents

1 Sequences of functions, uniform convergence, power series

1.1 Pointwise convergence

In section 4 of the lecture notes to the Analysis I course we introduced the exponential function

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

For every $n \in \mathbb{N}$ we define the polynomial function $f_n : \mathbb{R} \to \mathbb{R}$ by

$$f_n(x) := \sum_{k=0}^{n} \frac{x^k}{k!} \,.$$

Then $\{f_n\}_{n=1}^{\infty}$ is a sequence of functions with the property that

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

for every $x \in \mathbb{R}$. We say that the sequence $\{f_n\}_{n=1}^{\infty}$ converges pointwise to the exponential function.

Definition 1.1 *Let D be a set (not necessarily a set of real numbers), and let $\{f_n\}_{n=1}^{\infty}$ be a sequence of functions $f_n : D \to \mathbb{R}$. This sequence is said to converge pointwise, if a function $f : D \to \mathbb{R}$ exists such that*

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

for all $x \in D$. We call f the pointwise limit function of $\{f_n\}_{n=1}^{\infty}$.

The sequence $\{f_n\}_{n=1}^{\infty}$ of functions converges pointwise if and only if the numerical sequence $\{f_n(x)\}_{n=1}^{\infty}$ converges for every $x \in D$. For, if $\{f_n\}_{n=1}^{\infty}$ converges pointwise, then $\{f_n(x)\}_{n=1}^{\infty}$ converges by definition. On the other hand, if $\{f_n(x)\}_{n=1}^{\infty}$ converges for every $x \in D$, then a function $f : D \to \mathbb{R}$ is defined by

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

and so $\{f_n\}_{n=1}^{\infty}$ converges pointwise.

Clearly, this shows that the limit function of a pointwise convergent function sequence is uniquely determined. Moreover, together with the Cauchy convergence criterion for numerical sequences it immediately yields the following

Theorem 1.2 *A sequence* $\{f_n\}_{n=1}^{\infty}$ *of functions* $f_n : D \to \mathbb{R}$ *converges pointwise, if and only if to every* $x \in D$ *and to every* $\varepsilon > 0$ *there is a number* $n_0 \in \mathbb{N}$ *such that*

$$|f_n(x) - f_m(x)| < \varepsilon$$

for all $n, m \geq n_0$ *.*

With quantifiers this can be written as

$$\underset{x>0}{\forall} \; \underset{\varepsilon>0}{\forall} \; \underset{n_0\in\mathbb{N}}{\exists} \; \underset{n,m\geq n_0}{\forall} : |f_n(x) - f_m(x)| < \varepsilon \, .$$

Examples

1. Let $D = [0,1]$ and $x \mapsto f_n(x) := x^n$. Since for $x \in [0,1)$ we have $\lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} x^n = 0$, and since $\lim_{n\to\infty} f_n(1) = \lim_{n\to\infty} 1^n = 1$, the function sequence $\{f_n\}_{n=1}^{\infty}$ converges pointwise to the limit function $f : [0,1] \to \mathbb{R}$,

$$f(x) = \begin{cases} 0\,, & 0 \leq x < 1 \\ 1\,, & x = 1\,. \end{cases}$$

2. Above we considered the sequence of polynomial functions $\{f_n\}_{n=1}^{\infty}$ with $f_n(x) = \sum_{k=0}^{n} \frac{x^k}{k!}$, which converges pointwise to the exponential function. This sequence $\left\{\sum_{k=0}^{n} \frac{x^k}{k!}\right\}_{n=1}^{\infty}$ can also be called a *function series*.

3. Let $D = [0,2]$ and

$$f_n(x) = \begin{cases} nx\,, & 0 \leq x \leq \frac{1}{n} \\ 2 - nx\,, & \frac{1}{n} < x < \frac{2}{n} \\ 0\,, & \frac{2}{n} \leq x \leq 2\,. \end{cases}$$

This function sequence $\{f_n\}_{n=1}^{\infty}$ converges pointwise to the null function in $[0,2]$.

Proof: It must be shown that for all $x \in D$

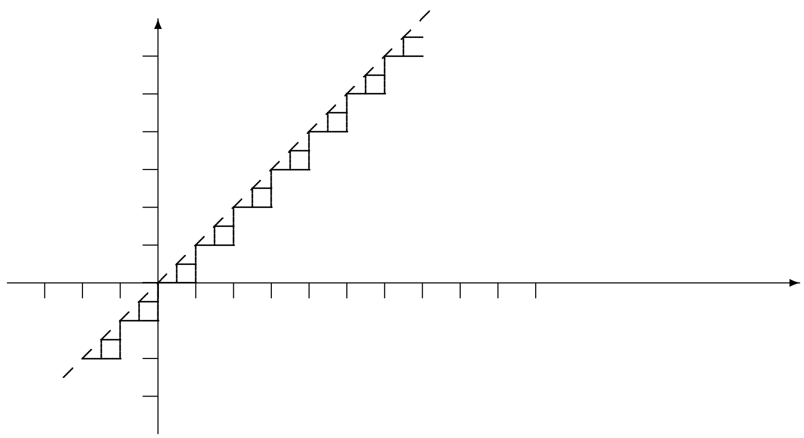
$$\lim_{n\to\infty} f_n(x) = 0.$$

For $x = 0$ we obviously have $\lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} 0 = 0$. Thus, let $x > 0$. Then there is $n_0 \in \mathbb{N}$ such that $\frac{2}{n_0} \leq x$. Since $\frac{2}{n} \leq \frac{2}{n_0} \leq x$ for $n \geq n_0$, the definition of f_n yields $f_n(x) = 0$ for all these n, whence

$$\lim_{n\to\infty} f_n(x) = 0.$$

■

4. Let $D = \mathbb{R}$ and $x \mapsto f_n(x) = \frac{1}{n}[nx]$. Here $[nx]$ denotes the greatest integer less or equal to nx.



$\{f_n\}_{n=1}^{\infty}$ converges pointwise to the identity mapping $x \mapsto f(x) := x$.

Proof: Let $x \in \mathbb{R}$ and $n \in \mathbb{N}$. Then there is $k \in \mathbb{Z}$ with $x \in [\frac{k}{n}, \frac{k+1}{n})$, hence $nx \in [k, k+1)$, and therefore

$$f_n(x) = \frac{1}{n}[nx] = \frac{k}{n}.$$

From $\frac{k}{n} \leq x < \frac{k+1}{n}$ it follows that

$$0 \leq x - \frac{k}{n} < \frac{1}{n},$$

which yields $|x - f_n(x)| = |x - \frac{k}{n}| < \frac{1}{n}$. This implies

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

■

1.2 Uniform convergence, continuity of the limit function

Suppose that $D \subseteq \mathbb{R}$ and that $\{f_n\}_{n=1}^{\infty}$ is a sequence of continuous functions $f_n = D \to \mathbb{R}$, which converges pointwise. It is natural to ask whether the limit function $f : D \to \mathbb{R}$ is continuous. However, the first example considered above shows that this need not be the case, since

$$x \mapsto f_n(x) = x^n : [0,1] \to \mathbb{R}$$

is continuous, but the limit function

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

is discontinuous. To be able to conclude that the limit function is continuous, a stronger type of convergence must be introduced:

Definition 1.3 *Let D be a set (not necessarily a set of real numbers), and let $\{f_n\}_{n=1}^{\infty}$ be a sequence of functions $f_n : D \to \mathbb{R}$. This sequence is said to be uniformly convergent, if a function $f : D \to \mathbb{R}$ exists such that to every $\varepsilon > 0$ there is a number $n_0 \in \mathbb{N}$ with*

$$|f_n(x) - f(x)| < \varepsilon$$

for all $n \geq n_0$ and all $x \in D$. The function f is called limit function.

With quantifiers, this can be written as

$$\underset{\varepsilon>0}{\forall} \quad \underset{n_0\in\mathbb{N}}{\exists} \quad \underset{x\in D}{\forall} \quad \underset{n\geq n_0}{\forall} \;:\; |f_n(x) - f(x)| < \varepsilon\,.$$

Note that for pointwise convergence the number n_0 may depend on $x \in D$, but for uniform convergence it must be possible to choose the number n_0 independently of $x \in D$. It is obvious that if $\{f_n\}_{n=1}^{\infty}$ converges uniformly, then it also converges pointwise, and the limit functions of uniform convergence and pointwise convergence coincide.

Examples

1. Let $D = [0,1]$ and $x \mapsto f_n(x) := x^n : D \to \mathbb{R}$. We have shown above that the sequence $\{f_n\}_{n=1}^{\infty}$ converges pointwise. However, this sequence is not uniformly convergent.

Proof: If this sequence would converge uniformly, the limit function had to be

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

since this is the pointwise limit function. We show that for this function the negation of the statement in the definition of uniform convergence is true:

$$\underset{\varepsilon>0}{\exists} \quad \underset{n_0\in\mathbb{N}}{\forall} \quad \underset{x\in D}{\exists} \quad \underset{n\geq n_0}{\exists} \;:\; |f_n(x) - f(x)| \geq \varepsilon\,.$$

Choose $\varepsilon = \frac{1}{2}$ and n_0 arbitrarily. The negation is true if $x \in (0,1)$ can be found with

$$|f_{n_0}(x) - f(x)| = |f_{n_0}(x)| = x^{n_0} = \frac{1}{2} = \varepsilon\,.$$

This is equivalent to

$$x = \Big(\frac{1}{2}\Big)^{\frac{1}{n_0}} = 2^{-\frac{1}{n_0}} = e^{-\frac{\log 2}{n_0}}\,.$$

$\frac{\log 2}{n_0} > 0$ and the strict monotonicity of the exponential function imply $0 < e^{-\frac{\log 2}{n_0}} < e^0 = 1$, whence $0 < \left(\frac{1}{2}\right)^{\frac{1}{n_0}} < 1$, whence $x = \left(\frac{1}{2}\right)^{\frac{1}{n_0}}$ has the sought properties. ∎

2. Let $\{f_n\}_{n=1}^{\infty}$ be the sequence of functions defined in example 3 of section 1.1. This sequence converges pointwise to the function $f = 0$, but it does not converge uniformly. Otherwise it had to converge uniformly to $f = 0$. However, choose $\varepsilon = 1$, let $n_0 \in \mathbb{N}$ be arbitrary and set $x = \frac{1}{n_0}$. Then

$$|f_n\Big(\frac{1}{n_0}\Big) - f\Big(\frac{1}{n_0}\Big)| = |f_n\Big(\frac{1}{n_0}\Big)| = 1 \geq \varepsilon\,,$$

which negates the statement in the definition of uniform convergence.

3. Let $D = \mathbb{R}$ and $x \mapsto f_n(x) = \frac{1}{n}[nx]$. The sequence $\{f_n\}_{n=1}^{\infty}$ converges uniformly to $x \mapsto f(x) = x$. To verify this, let $\varepsilon > 0$ and remember that in example 4 of section 1.1 we showed that

$$|f_n(x) - f(x)| = |f_n(x) - x| < \frac{1}{n}$$

for all $x \in \mathbb{R}$ and all $n \in \mathbb{N}$. Hence, if we choose $n_0 \in \mathbb{N}$ such that $\frac{1}{n_0} < \varepsilon$, we obtain for all $n \geq n_0$ and all $x \in \mathbb{R}$

$$|f_n(x) - f(x)| < \frac{1}{n} \leq \frac{1}{n_0} < \varepsilon\,.$$

Uniform convergence is important because of the following

Theorem 1.4 *Let $D \subseteq \mathbb{R}$, let $a \in D$ and let all the functions $f_n : D \to \mathbb{R}$ be continuous at a. Suppose that the sequence of functions $\{f_n\}_{n=1}^{\infty}$ converges uniformly to the limit function $f : D \to \mathbb{R}$. Then f is continuous at a.*

Proof: Let $\varepsilon > 0$. We have to find $\delta > 0$ such that for all $x \in D$ with $|x - a| < \delta$

$$|f(x) - f(a)| < \varepsilon$$

holds. To determine such a number δ, note that for all $x \in D$ and all $n \in \mathbb{N}$

$$\begin{aligned} |f(x) - f(a)| &= |f(x) - f_n(x) + f_n(x) - f_n(a) + f_n(a) - f(a)| \\ &\leq |f(x) - f_n(x)| + |f_n(x) - f_n(a)| + |f_n(a) - f(a)|\,. \end{aligned}$$

Since $\{f_n\}_{n=1}^{\infty}$ converges uniformly to f, there is $n_0 \in \mathbb{N}$ with $|f_n(y) - f(y)| < \frac{\varepsilon}{3}$ for all $n \geq n_0$ and all $y \in D$, whence

$$|f(x) - f(a)| \leq \frac{2}{3}\varepsilon + |f_{n_0}(x) - f_{n_0}(a)|\,.$$

Since f_{n_0} is continuous, there is $\delta > 0$ such that $|f_{n_0}(x) - f_{n_0}(a)| < \frac{\varepsilon}{3}$ for all $x \in D$ with $|x - a| < \delta$. Thus, if $|x - a| < \delta$

$$|f(x) - f(a)| < \frac{2}{3}\varepsilon + \frac{1}{3}\varepsilon = \varepsilon\,,$$

which proves that f is continuous at a. ∎

This theorem shows that

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

Hence, for a uniformly convergent sequence of functions the limits $\lim_{x\to a}$ and $\lim_{n\to\infty}$ can be interchanged.

Corollary 1.5 *The limit function of a uniformly convergent sequence of continuous functions is continuous.*

Example 2 considered above shows that the limit function can be continuous even if the sequence $\{f_n\}_{n=1}^{\infty}$ does not converge uniformly. However, we have

Theorem 1.6 (of Dini) *Let $D \subseteq \mathbb{R}$ be compact, let $f_n : D \to \mathbb{R}$ and $f : D \to \mathbb{R}$ be continuous, and assume that the sequence of functions $\{f_n\}_{n=1}^{\infty}$ converges pointwise and monotonically to f, i.e. the sequence $\{|f_n(x) - f(x)|\}_{n=1}^{\infty}$ is a decreasing null sequence for every $x \in D$. Then $\{f_n\}_{n=1}^{\infty}$ converges uniformly to f. (Ulisse Dini, 1845-1918).*

Proof: Let $\varepsilon > 0$. To every $x \in D$ a neighborhood $U(x)$ is associated as follows: $\lim_{n\to\infty} f_n(x) = f(x)$ implies that a number $n_0 = n_0(x, \varepsilon)$ exists such that $|f_{n_0}(x) - f(x)| < \varepsilon$. Since f and f_{n_0} are continuous, also $|f_{n_0} - f|$ is continuous, hence there is an open neighborhood $U(x)$ of x such that $|f_{n_0}(y) - f(y)| < \varepsilon$ holds for all $y \in U(x) \cap D$. The system $\mathcal{U} = \{U(x) \mid x \in D\}$ of these neighborhoods is an open covering of the compact set D, hence finitely many of these neighborhoods $U(x_1), \ldots, U(x_m)$ suffice to cover D. Let

$$\tilde{n} = \max\{n_0(x_i, \varepsilon) \mid i = 1, \ldots, m\}\ .$$

To every $x \in D$ there is a number $i \in \{1, \ldots, m\}$ with $x \in U(x_i)$. Then, by construction of $U(x_i)$,

$$|f_{n_0(x_i,\varepsilon)}(x) - f(x)| < \varepsilon\,,$$

whence, since $\{f_n(x)\}_{n=1}^{\infty}$ converges monotonically to $f(x)$,

$$|f_n(x) - f(x)| < \varepsilon$$

for all $n \geq n_0(x_i, \varepsilon)$. In particular, this inequality holds for all $n \geq \tilde{n}$. Since $\tilde{n}$ is independent of x, this proves that $\{f_n\}_{n=1}^{\infty}$ converges uniformly to f. ∎

1.3 Supremum norm

For the definition of convergence and limits of numerical sequences the absolute value, a tool to measure distance for numbers, was of crucial importance. Up to now we have not introduced a tool to measure distance of functions, but we were nevertheless able to define two different types of convergence of sequences of functions, the pointwise convergence and the uniform convergence. Since functions with domain D and target set $\mathbb{R}$ are elements of the algebra $F(D, \mathbb{R})$, it is natural to ask whether a tool can be introduced, which allows to measure the distance of two elements from $F(D, \mathbb{R})$, and which can be used to define convergence on the set $F(D, \mathbb{R})$ just as the absolute value could be used to define convergence on the set $\mathbb{R}$. Here we shall show that this is indeed possible on the smaller algebra $B(D, \mathbb{R})$ of bounded real valued functions. The resulting type of convergence of sequences of functions from $B(D, \mathbb{R})$ is the uniform convergence.

Definition 1.7 *Let D be a set (not necessarily a set of real numbers), and let $f : D \to \mathbb{R}$ be a bounded function. The nonnegative number*

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

is called the supremum norm of f.

The norm has properties similar to the properties of the absolute value on $\mathbb{R}$. This is shown by the following

Theorem 1.8 *Let $f, g : D \to \mathbb{R}$ be bounded functions and c be a real number. Then*

(i) $\|f\| = 0 \iff f = 0$

(ii) $\|cf\| = |c| \, \|f\|$

(iii) $\|f + g\| \leq \|f\| + \|g\|$

(iv) $\|fg\| \leq \|f\| \, \|g\|$.

Proof: (i) and (ii) are obvious. To prove (iii), note that for $x \in D$

$$\begin{aligned} |(f+g)(x)| &= |f(x)+g(x)| \le |f(x)| + |g(x)| \\ &\le \sup_{y\in D} |f(y)| + \sup_{y \in D} |g(y)| = \|f\| + \|g\| \,. \end{aligned}$$

Thus, $\|f\| + \|g\|$ is an upper bound for the set $\{|(f+g)(x)| \bigm| x \in D\}$, whence for the least upper bound

$$\|f+g\| = \sup_{x\in D} |(f+g)(x)| \le \|f\| + \|g\| \,.$$

To prove (iv), we use that for $x \in D$

$$|(fg)(x)| = |f(x)g(x)| = |f(x)|\,|g(x)| \le \|f\|\,\|g\| \,,$$

whence

$$\|fg\| = \sup_{x\in D} |(fg)(x)| \le \|f\|\,\|g\| \,.$$

■

Definition 1.9 *Let V be a vector space. A mapping $\|\cdot\| : V \to [0,\infty)$ which has the properties*

(i) $\|v\| = 0 \iff v = 0$

(ii) $\|cv\| = |c|\,\|v\|$ *(positive homogeneity)*

(iii) $\|v+u\| \le \|v\| + \|u\|$ *(triangle inequality)*

is called a norm on V. If V is an algebra, then $\|\cdot\| : V \to [0,\infty)$ is called an algebra norm, provided that (i) - (iii) and

(iv) $\|uv\| \le \|u\|\,\|v\|$

are satisfied. A vector space or an algebra with norm is called a normed vector space or a normed algebra.

Clearly, the absolute value $|\cdot| : \mathbb{R} \to [0,\infty)$ has the properties (i) - (iv) of the preceding definition, hence $|\cdot|$ is an algebra norm on $\mathbb{R}$ and $\mathbb{R}$ is a normed algebra. The preceding theorem shows that the supremum norm $\|\cdot\| : B(D,\mathbb{R}) \to [0,\infty)$ is an algebra norm on the set $B(D,\mathbb{R})$ of bounded real valued functions, and $B(D,\mathbb{R})$ is a normed algebra.

Definition 1.10 *A sequence of functions* $\{f_n\}_{n=1}^{\infty}$ *from* $B(D, \mathbb{R})$ *is said to converge with respect to the supremum norm to a function* $f \in B(D, \mathbb{R})$ *, if to every* $\varepsilon > 0$ *there is a number* $n_0 \in \mathbb{N}$ *such that*

$$\|f_n - f\| < \varepsilon$$

for all $n \geq n_0$ *, or, equivalently, if*

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

Theorem 1.11 *A sequence* $\{f_n\}_{n=1}^{\infty}$ *from* $B(D, \mathbb{R})$ *converges to* $f \in B(D, \mathbb{R})$ *with respect to the supremum norm, if and only if* $\{f_n\}_{n=1}^{\infty}$ *converges uniformly to* f *.*

Proof: $\{f_n\}_{n=1}^{\infty}$ converges uniformly to f , if and only if to every $\varepsilon > 0$ there is $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ and all $x \in D$

$$|f_n(x) - f(x)| \leq \varepsilon \,.$$

This holds if and only if for all $n \geq n_0$

$$\|f_n - f\| = \sup_{x \in D} |f_n(x) - f(x)| \leq \varepsilon \,,$$

hence if and only if $\{f_n\}_{n=1}^{\infty}$ converges to f with respect to the supremum norm. ∎

Definition 1.12 *A sequence* $\{f_n\}_{n=1}^{\infty}$ *of functions from* $B(D, \mathbb{R})$ *is said to be a Cauchy sequence, if to every* $\varepsilon > 0$ *there is* $n_0 \in \mathbb{N}$ *such that*

$$\|f_n - f_m\| < \varepsilon$$

for all $n, m \geq n_0$ *.*

Theorem 1.13 *A sequence* $\{f_n\}_{n=1}^{\infty}$ *of functions from* $B(D, \mathbb{R})$ *converges uniformly, if and only if it is a Cauchy sequence.*

Proof: If $\{f_n\}_{n=1}^{\infty}$ converges uniformly, then there is a function $f \in B(D, \mathbb{R})$, the limit function, such that $\{\|f_n - f\|\}_{n=1}^{\infty}$ is a null sequence. Hence to $\varepsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that for $n, m \geq n_0$

$$\|f_n - f_m\| = \|f_n - f + f - f_m\| \leq \|f_n - f\| + \|f - f_m\| < 2\varepsilon \,.$$

This shows that $\{f_n\}_{n=1}^{\infty}$ is a Cauchy sequence.

Conversely, assume that $\{f_n\}_{n=1}^\infty$ is a Cauchy sequence. To prove that this sequence converges, we first must identify the limit function. To this end we show that $\{f_n(x)\}_{n=1}^\infty$ is a Cauchy sequence of real numbers for every $x \in D$. For, since $\{f_n\}_{n=1}^\infty$ is a Cauchy sequence, to $\varepsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that for all $n, m \geq n_0$

$$|f_n(x) - f_m(x)| \leq \|f_n - f_m\| < \varepsilon\,,$$

and so $\{f_n(x)\}_{n=1}^\infty$ is indeed a Cauchy sequence of real numbers. Since every Cauchy sequence of real numbers converges, we obtain that $\{f_n\}_{n=1}^\infty$ converges pointwise with limit function $f : D \to \mathbb{R}$ defined by

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

We show that $\{f_n\}_{n=1}^\infty$ even converges uniformly to f. For, using again that $\{f_n\}_{n=1}^\infty$ is a Cauchy sequence, to $\varepsilon > 0$ there is $n_0 \in \mathbb{N}$ with $\|f_n - f_m\| < \varepsilon$ for $n, m \geq n_0$. Therefore we obtain for $x \in D$ and $n \geq n_0$

$$|f_n(x) - f(x)| = |f_n(x) - \lim_{m\to\infty} f_m(x)| = \lim_{m\to\infty} |f_n(x) - f_m(x)| \leq \varepsilon\,,$$

whence

$$\|f_n - f\| = \sup_{x\in D} |f_n(x) - f(x)| \leq \varepsilon$$

for $n \geq n_0$, since ε is independent of x. ∎

1.4 Uniformly converging series of functions

Let D be a set and let $f_n : D \to \mathbb{R}$ be functions. The series of functions $\sum_{n=1}^\infty f_n$ is said to be uniformly convergent, if the sequence $\{\sum_{n=1}^m f_n\}_{m=1}^\infty$ is uniformly convergent.

Theorem 1.14 (Criterion of Weierstraß) *Let $f_n : D \to \mathbb{R}$ be bounded functions satisfying $\|f_n\| \leq c_n$, and let $\sum_{n=1}^\infty c_n$ be convergent. Then the series of functions $\sum_{n=1}^\infty f_n$ converges uniformly.*

Proof: It suffices to show that $\{\sum_{n=1}^m f_n\}_{m=1}^\infty$ is a Cauchy sequence. Let $\varepsilon > 0$. Since $\sum_{k=1}^\infty c_k$ converges, there is $n_0 \in \mathbb{N}$ such that $\left|\sum_{k=n}^m c_k\right| = \sum_{k=n}^m c_k < \varepsilon$ for all $m \geq n \geq n_0$, whence

$$\|\sum_{k=n}^m f_k\| \leq \sum_{k=n}^m \|f_k\| \leq \sum_{k=n}^m c_k < \varepsilon\,,$$

for all $m \geq n \geq n_0$. ∎

1.5 Differentiability of the limit function

Let D be a subset of $\mathbb{R}$. We showed that a uniformly convergent sequence $\{f_n\}_{n=1}^{\infty}$ of continuous functions has a continuous limit function $f : D \to \mathbb{R}$. One can ask the question what type of convergence is needed to ensure that a sequence of differentiable functions has a differentiable limit function? Simple examples show that uniform convergence is not sufficient to ensure this. The following is a slightly different question: Assume that $\{f_n\}_{n=1}^{\infty}$ is a uniformly convergent sequence of differentiable functions with limit function f. If f is differentiable, does this imply that the sequence of derivatives $\{f_n'\}_{n=1}^{\infty}$ converges pointwise to f'? Also this need not be true, as is shown by the following example: Let $D = [0,1]$ and let $x \mapsto f_n(x) = \frac{1}{n}x^n : [0,1] \to \mathbb{R}$. The sequence $\{f_n\}_{n=1}^{\infty}$ of differentiable functions converges uniformly to the differentiable limit function $f = 0$. The sequence of derivatives $\{f_n'\}_{n=1}^{\infty} = \{x^{n-1}\}_{n=1}^{\infty}$ does not converge uniformly on $[0,1]$, but it converges pointwise to the limit function

$$g(x) = \begin{cases} 0\,, & 0 \leq x < 1 \\ 1\,, & x = 1\,. \end{cases}$$

However, $g \neq f' = 0$.

Our original question is answered by the following

Theorem 1.15 *Let $-\infty < a < b < \infty$ and let $f_n : [a,b] \to \mathbb{R}$ be differentiable functions. If the sequence $\{f_n'\}_{n=1}^{\infty}$ of derivatives converges uniformly and the sequence $\{f_n\}_{n=1}^{\infty}$ converges at least in one point $x_0 \in [a,b]$, then the sequence $\{f_n\}_{n=1}^{\infty}$ converges uniformly to a differentiable limit function $f : [a,b] \to \mathbb{R}$ and*

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

for all $x \in [a,b]$.

This means that under the convergence condition given in this theorem, derivation (which is a limit process) can be interchanged with the limit with respect to n:

$$\big(\lim_{n\to\infty} f_n\big)' = \lim_{n\to\infty} f_n'\,.$$

Proof: First we show that $\{f_n\}_{n=1}^{\infty}$ converges uniformly. Let $\varepsilon > 0$. For $x \in [a,b]$

$$|f_m(x) - f_n(x)| \quad \leq | \quad \big(f_m(x) - f_n(x)\big) - \big(f_m(x_0) - f_n(x_0)\big)| + |f_m(x_0) - f_n(x_0)|\,. \quad (*)$$

Since $f_m - f_n$ is differentiable, the mean value theorem yields for a suitable z between x_0 and x

$$|\big(f_m(x) - f_n(x)\big) - \big(f_m(x_0) - f_n(x_0)\big)| = |f'_m(z) - f'_n(z)|\,|x - x_0|\,.$$

The sequence of derivatives converges uniformly. Therefore there is $n_0 \in \mathbb{N}$ such that for all $m, n \geq n_0$

$$|f'_m(z) - f'_n(z)| < \frac{\varepsilon}{2(b-a)}\,,$$

hence

$$|\big(f_m(x) - f_n(x)\big) - \big(f_m(x_0) - f_n(x_0)\big)| \leq \frac{\varepsilon}{2}\,,$$

for all $m, n \geq n_0$ and all $x \in [a, b]$. By assumption the numerical sequence $\{f_n(x_0)\}_{n=1}^\infty$ converges, hence there is $n_1 \in \mathbb{N}$ such that for all $m, n \geq n_1$

$$|f_m(x_0) - f_n(x_0)| \leq \frac{\varepsilon}{2}\,.$$

The last two estimates and $(*)$ together yield

$$|f_m(x) - f_n(x)| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

for all $m, n \geq n_2 = \max\{n_0, n_1\}$ and all $x \in [a, b]$. This implies that $\{f_n\}_{n=1}^\infty$ converges uniformly. The limit function is denoted by f.

Let $c \in [a, b]$ and for $x \in [a, b]$ set

$$F(x) = \begin{cases} \dfrac{f(x) - f(c)}{x - c} - m\,, & x \neq c \\ 0\,, & x = c\,, \end{cases}$$

with $m = \lim_{n\to\infty} f'_n(c)$. The statement of the theorem follows if F is continuous at the point $x = c$, since continuity of F implies that f is differentiable at c with derivative $f'(c) = m = \lim_{n\to\infty} f'_n(c)$. For the proof that F is continuous at c, set

$$F_n(x) = \begin{cases} \dfrac{f_n(x) - f_n(c)}{x - c} - f'_n(c)\,, & x \neq c \\ 0\,, & x = c\,. \end{cases}$$

Obviously $F(x) = \lim_{n\to\infty} F_n(x)$, and since F_n is continuous due to the differentiability of f_n, the continuity of F follows if it can be shown that $\{F_n\}_{n=1}^\infty$ converges uniformly. This follows by application of the mean value theorem to the differentiable function $f_m - f_n$:

$$\begin{aligned} F_m(x) - F_n(x) &= \begin{cases} \dfrac{\big(f_m(x) - f_n(x)\big) - \big(f_m(c) - f_n(c)\big)}{x - c} - \big(f'_m(c) - f'_n(c)\big)\,, & x \neq c \\ 0\,, & x = c \end{cases} \\ &= \big(f'_m(z) - f'_n(z)\big) - \big(f'_m(c) - f'_n(c)\big)\,, \end{aligned}$$

for a suitable z between x and c if $x \neq c$, and for $z = c$ if $x = c$. By assumption $\{f_n'\}_{n=1}^{\infty}$ converges uniformly, consequently there is $n_0 \in \mathbb{N}$ such that for all $m, n \geq n_0$ and all $y \in [a, b]$

$$|f_m'(y) - f_n'(y)| < \varepsilon\,,$$

whence

$$\begin{aligned} |F_m(x) - F_n(x)| &\leq |f_m'(z) - f_n'(z)| + |f_m'(c) - f_n'(c)| \\ &< \varepsilon + \varepsilon = 2\varepsilon\,, \end{aligned}$$

for all $m, n \geq n_0$ and all $x \in [a, b]$. This shows that $\{F_n\}_{n=1}^{\infty}$ converges uniformly and completes the proof. ■

1.6 Power series

Let a numerical sequence $\{a_n\}_{n=1}^{\infty}$ and a real number x_0 be given. For arbitrary $x \in \mathbb{R}$ consider the series

$$\sum_{n=0}^{\infty} a_n (x - x_0)^n\,.$$

This series is called a power series. a_n is called the n-th coefficient, x_0 is the center of expansion of the power series. The Taylor series and the series for exp, sin and cos are power series. These examples show that power series are interesting mainly as function series

$$x \mapsto \sum_{n=0}^{\infty} f_n(x)$$

with $f_n(x) = a_n(x - x_0)^n$. First the convergence of power series must be investigated:

Theorem 1.16 *Let*

$$\sum_{n=0}^{\infty} a_n (x - x_0)^n$$

be a power series.

(i) *Suppose first that*

$$a = \overline{\lim_{n\to\infty}} \sqrt[n]{|a_n|} < \infty\,.$$

Then the power series is in case

$$
\begin{array}{ll}
a = 0: & \textit{absolutely convergent for all } x \in \mathbb{R} \\
a > 0: & \begin{cases} \textit{absolutely convergent for } |x - x_0| < \frac{1}{a} \\ \textit{convergent or divergent for } |x - x_0| = \frac{1}{a} \\ \textit{divergent for } |x - x_0| > \frac{1}{a}. \end{cases}
\end{array}
$$

(ii) *If* $\left\{ \sqrt[n]{|a_n|} \right\}_{n=1}^{\infty}$ *is unbounded, then the power series converges only for* $x = x_0$.

Proof: By the root test, the series $\sum_{n=0}^{\infty} a_n (x - x_0)^n$ converges absolutely if

$$
\overline{\lim_{n \to \infty}} \sqrt[n]{|a_n|\, |x - x_0|^n} = |x - x_0| \, \overline{\lim_{n \to \infty}} \sqrt[n]{|a_n|} = |x - x_0| a < 1 ,
$$

and diverges if

$$
\overline{\lim_{n \to \infty}} \sqrt[n]{|a_n|\, |x - x_0|^n} = |x - x_0| a > 1 .
$$

This proves (i). If $\left\{ \sqrt[n]{|a_n|} \right\}_{n=1}^{\infty}$ is unbounded, then for $x \neq x_0$ also $\left\{ |x - x_0| \sqrt[n]{|a_n|} \right\}_{n=1}^{\infty} = \left\{ \sqrt[n]{|a_n (x - x_0)^n|} \right\}_{n=1}^{\infty}$ is unbounded, hence $\{a_n (x - x_0)^n\}_{n=1}^{\infty}$ is not a null sequence, and consequently $\sum_{n=0}^{\infty} a_n (x - x_0)^n$ diverges. This proves (ii) ■

Definition 1.17 *Let* $a = \overline{\lim}_{n \to \infty} \sqrt[n]{|a_n|}$. *The number*

$$
r = \begin{cases} \frac{1}{a}, & \textit{if } a \neq 0 \\ \infty, & \textit{if } a = 0 \\ 0, & \textit{if } \left\{ \sqrt[n]{|a_n|} \right\}_{n=1}^{\infty} \textit{ is unbounded} \end{cases}
$$

is called radius of convergence and the open interval

$$
(x_0 - r \,, \, x_0 + r) = \left\{ x \in \mathbb{R} \;\middle|\; |x - x_0| < r \right\}
$$

is called interval of convergence of the power series

$$
\sum_{n=0}^{\infty} a_n (x - x_0)^n .
$$

Examples

1. The power series

$$
\sum_{n=0}^{\infty} x^n , \quad \sum_{n=1}^{\infty} \frac{1}{n} x^n
$$

both have radius of convergence equal to 1. This is evident for the first series. To prove it for the second series, note that

$$
\lim_{n \to \infty} \sqrt[n]{n} = \lim_{n \to \infty} e^{\frac{1}{n} \log n} = e^{\lim_{n \to \infty} (\frac{1}{n} \log n)} = e^0 = 1 ,
$$

since $\lim_{x\to\infty} \frac{\log x}{x} = 0$, by the rule of de l'Hospital. Thus, the radius of convergence of the second series is given by

$$r = \frac{1}{\overline{\lim}_{n\to\infty} \sqrt[n]{\frac{1}{n}}} = \frac{1}{\lim_{n\to\infty} \sqrt[n]{\frac{1}{n}}} = \lim_{n\to\infty} \sqrt[n]{n} = 1 .$$

For $x = 1$ both power series diverge, for $x = -1$ the first one diverges, the second one converges.

2. In Analysis I it was proved that the exponential series

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

converges absolutely for all $x \in \mathbb{R}$. (To verify this use the ratio test, for example.) Consequently, the radius of convergence r must be infinite. For, if r would be finite, the exponential series had to diverge for all x with $|x| > r$, which is excluded. (This implies $\frac{1}{r} = \overline{\lim}_{n\to\infty} \sqrt[n]{\frac{1}{n!}} = 0$, by the way.)

Theorem 1.18 *Let $\sum_{n=0}^{\infty} a_n(x-x_0)^n$ and $\sum_{n=0}^{\infty} b_n(x-x_0)^n$ be power series with radii of convergence r_1 and r_2, respectively. Then for all x with $|x-x_0| < r = \min(r_1, r_2)$*

$$\begin{aligned} \sum_{n=0}^{\infty} a_n(x-x_0)^n + \sum_{n=0}^{\infty} b_n(x-x_0)^n &= \sum_{n=0}^{\infty} (a_n + b_n)(x-x_0)^n \\ \Big[\sum_{n=0}^{\infty} a_n(x-x_0)^n\Big] \Big[\sum_{n=0}^{\infty} b_n(x-x_0)^n\Big] &= \sum_{n=0}^{\infty} \Big(\sum_{k=0}^{n} a_k b_{n-k}\Big)(x-x_0)^n . \end{aligned}$$

Proof: The statements follow immediately from the theorems about computing with series and about the Cauchy product of two series. (We note that the radii of convergence of both series on the right are at least equal to r, but can be larger.) ∎

Theorem 1.19 *Let $\sum_{n=0}^{\infty} a_n(x-x_0)^n$ be a power series with radius of convergence r. Then this series converges uniformly in every compact interval $[x_0 - r_1, x_0 + r_1]$ with $0 \le r_1 < r$.*

Proof: Let $c_n = |a_n|\, r_1^n$. Then

$$\overline{\lim_{n\to\infty}} \sqrt[n]{c_n} = \overline{\lim_{n\to\infty}} \sqrt[n]{|a_n| r_1^n} = r_1 \frac{1}{r} < 1 ,$$

whence the root test implies that the series

$$\sum_{n=0}^{\infty} c_n$$

converges. Because of $|a_n(x-x_0)^n| \le |a_n| r_1^n = c_n$ for all x with $|x-x_0| \le r_1$, the Weierstraß criterion (Theorem 1.14) yields that the power series $\sum_{n=0}^{\infty} a_n(x-x_0)^n$ converges uniformly for $x \in [x_0 - r_1\,,\, x_0 + r_1] = \{y \mid |y-x_0| \le r_1\}$. ∎

Corollary 1.20 *Let $\sum_{n=0}^{\infty} a_n(x-x_0)^n$ be a power series with radius of convergence $r > 0$. Then the function $f : (x_0 - r\,,\, x_0 + r) \to \mathbb{R}$ defined by*

$$f(x) = \sum_{n=0}^{\infty} a_n(x-x_0)^n$$

is continuous.

Proof: Since $\{x \mapsto \sum_{n=0}^{m} a_n(x-x_0)^n\}_{m=0}^{\infty}$ is a sequence of continuous functions, which converges uniformly in every compact interval $[x_0 - r_1\,,\, x_0 + r_1]$ with $r_1 < r$, the limit function f is continuous in each of these intervals. Hence f is continuous in the union

$$(x_0 - r\,,\, x_0 + r) = \bigcup_{0<r_1<r} [x_0 - r_1\,,\, x_0 + r_1]\,. \qquad ∎$$

Let

$$f(x) = \sum_{n=0}^{\infty} a_n(x-x_0)^n$$

be a power series with radius of convergence $r > 0$. Each of the polynomials $f_m(x) = \sum_{n=0}^{m} a_n(x-x_0)^n$ is differentiable with derivative

$$f_m'(x) = \sum_{n=1}^{m} n a_n (x-x_0)^{n-1}\,.$$

$\sum_{n=1}^{\infty} n a_n (x-x_0)^{n-1}$ is a power series, whose radius of convergence r_1 is equal to r. To verify this, note that

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

and that

$$\overline{\lim_{n\to\infty}} \sqrt[n]{|n a_n|} = \lim_{n\to\infty} \sqrt[n]{n}\; \overline{\lim_{n\to\infty}} \sqrt[n]{|a_n|} = \overline{\lim_{n\to\infty}} \sqrt[n]{|a_n|} = \frac{1}{r}\,,$$

which implies that the series $\sum_{n=1}^{\infty} n a_n (x-x_0)^{n-1}$ converges for all x with $|x-x_0| < r$ and diverges for all x with $|x-x_0| > r$. By Theorem 1.16 this can only be true if $r_1 = r$.

Thus, Theorem 1.19 implies that the sequence $\{f_m'\}_{m=1}^{\infty}$ of derivatives converges uniformly in every compact subinterval of the interval of convergence $(x_0 - r, x_0 + r)$.

Consequently, we can use Theorem 1.15 to conclude that the limit function $f(x) = \sum_{n=0}^{\infty} a_n(x-x_0)^n$ is differentiable with derivative

$$f'(x) = \lim_{m\to\infty} f'_m(x) = \sum_{n=1}^{\infty} n a_n (x-x_0)^{n-1}$$

in all these subintervals. Hence f is differentiable with derivative given by this formula in the interval of convergence $(x_0 - r\,,\, x_0 + r)$, which is the union of these subintervals.

Repeating these arguments we obtain

Theorem 1.21 *Let $f(x) = \sum_{n=0}^{\infty} a_n(x-x_0)^n$ be a power series with radius of convergence $r > 0$. Then f is infinitely differentiable in the interval of convergence. All the derivatives can be computed termwise:*

$$f^{(k)}(x) = \sum_{n=k}^{\infty} n(n-1)\ldots(n-k+1)a_n(x-x_0)^{n-k}\,.$$

Example: In the interval $(0,2]$ the logarithm can be expanded into the power series

$$\log x = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n}(x-1)^n\,.$$

In section 7.4 of the lecture notes to Analysis I we proved that this equation holds true for $\frac{1}{2} \leq x \leq 2$. To verify that it also holds for $0 < x < \frac{1}{2}$, note that the radius of convergence of the power series on the right is

$$r = \frac{1}{\overline{\lim}_{n\to\infty} \sqrt[n]{\left|\frac{(-1)^{n-1}}{n}\right|}} = \lim_{n\to\infty} \sqrt[n]{n} = 1\,.$$

Hence, this power series converges in the interval of convergence $\big\{x \;\big|\; |x-1| < 1\big\} = (0,2)$ and represents there an infinitely differentiable function. The derivative of this function is

$$\begin{aligned}\Big[\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n}\,(x-1)^n\Big]' &= \sum_{n=1}^{\infty}(-1)^{n-1}(x-1)^{n-1} = \sum_{n=0}^{\infty}(1-x)^n \\ &= \frac{1}{1-(1-x)} = \frac{1}{x} = (\log x)'\,.\end{aligned}$$

Consequently $\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n}(x-1)^n$ and $\log x$ both are antiderivatives of $\frac{1}{x}$ in the interval $(0,2)$, and therefore differ at most by a constant:

$$\log x = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n}(x-1)^n + C\,.$$

To determine C, set $x = 1$. From $\log(1) = 0$ we obtain $C = 0$.

Theorem 1.22 (Identity theorem for power series) *Let the radii of convergence r_1 and r_2 of the power series $\sum_{n=0}^{\infty} a_n(x-x_0)^n$ and $\sum_{n=0}^{\infty} b_n(x-x_0)^n$ be greater than zero. Assume that these power series coincide in a neighborhood $U_r(x_0) = \{x \in \mathbb{R} \mid |x - x_0| < r\}$ of x_0 with $r \leq \min(r_1, r_2)$:*

$$\sum_{n=0}^{\infty} a_n(x - x_0)^n = \sum_{n=0}^{\infty} b_n(x - x_0)^n$$

for all $x \in U_r(x_0)$. Then $a_n = b_n$ for all $n = 0, 1, 2, \ldots$.

Proof: First choose $x = x_0$, which immediately yields

$$a_0 = b_0 \,.$$

Next let $n \in \mathbb{N} \cup \{0\}$ and assume that $a_k = b_k$ for $0 \leq k \leq n$. It must be shown that $a_{n+1} = b_{n+1}$ holds. From the assumptions of the theorem and from the assumption of the induction it follows that

$$\sum_{k=n+1}^{\infty} a_k(x - x_0)^k = \sum_{k=n+1}^{\infty} b_k(x - x_0)^k \,,$$

hence

$$(x - x_0)^{n+1} \sum_{k=n+1}^{\infty} a_k(x - x_0)^{k-n-1} = (x - x_0)^{n+1} \sum_{k=n+1}^{\infty} b_k(x - x_0)^{k-n-1}$$

for all $x \in U_r(x_0)$. For x from this neighborhood with $x \neq x_0$ this implies

$$\sum_{k=n+1}^{\infty} a_k(x - x_0)^{k-n-1} = \sum_{k=n+1}^{\infty} b_k(x - x_0)^{k-n-1} \,.$$

The continuity of power series thus implies

$$\begin{aligned} a_{n+1} &= \sum_{k=n+1}^{\infty} a_k(x_0 - x_0)^{k-n-1} = \lim_{x \to x_0} \sum_{k=n+1}^{\infty} a_k(x - x_0)^{k-n-1} \\ &= \lim_{x \to x_0} \sum_{k=n+1}^{\infty} b_k(x - x_0)^{k-n-1} = \sum_{k=n+1}^{\infty} b_k(x_0 - x_0)^{k-n-1} = b_{n+1} \,. \end{aligned}$$

■

Every power series defines a continuous function in the interval of convergence. Information about continuity of the power series on the boundary of the interval of convergence is provided by the following

Theorem 1.23 *Let $\sum_{n=0}^{\infty} a_n(x-x_0)^n$ be a power series with positive radius of convergence, let $z \in \mathbb{R}$ be a boundary point of the interval of convergence and assume that $\sum_{n=0}^{\infty} a_n(z-x_0)^n$ converges. Then the power series converges uniformly in the interval $[z, x_0]$ (if $z < x_0$), or in the interval $[x_0, z]$ (if $x_0 < z$), respectively.*

A **proof** of this theorem can be found in the book: M. Barner, F. Flohr: Analysis I, p. 317, 318 (in German).

Corollary 1.24 (Abel's limit theorem) *If a power series converges at a point on the boundary of the interval of convergence, then it is continuous at this point. (Niels Hendrick Abel, 1802-1829).*

1.7 Trigonometric functions continued

Since sine is defined by a power series with interval of convergence equal to $\mathbb{R}$,

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

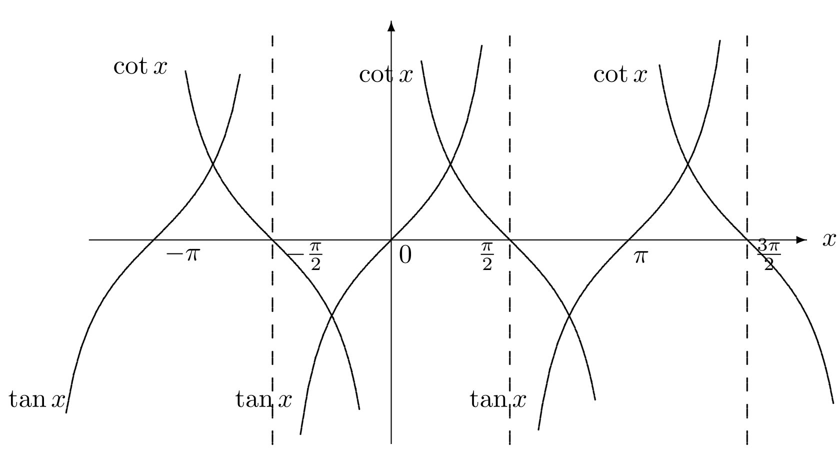
the derivative of sin can be computed by termwise differentiation of the power series, hence

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

This result has been proved in Analysis I using the addition theorem for sine.

Tangent and cotangent. One defines

$$\tan x = \frac{\sin x}{\cos x}, \quad \cot x = \frac{\cos x}{\sin x} = \frac{1}{\tan x}.$$



From the addition theorems for sine and cosine addition theorems for tangent and cotangent can be derived:

$$\begin{aligned} \tan(x+y) &= \frac{\tan x + \tan y}{1 - \tan x\ \tan y} \\ \cot(x+y) &= \frac{\cot x\ \cot y - 1}{\cot x + \cot y}\,. \end{aligned}$$

The derivatives are

$$\begin{aligned} \tan' x &= \Big(\frac{\sin x}{\cos x}\Big)' = \frac{\cos^2 x + \sin^2 x}{\cos^2 x} = \frac{1}{\cos^2 x} \\ \cot' x &= \Big(\frac{\cos x}{\sin x}\Big)' = \frac{-\sin^2 x - \cos^2 x}{\sin^2 x} = \frac{-1}{\sin^2 x}\,. \end{aligned}$$

Inverse trigonometric functions. sine and cosine are periodic, hence not injective, and consequently do not have inverse functions. However, if sine and cosine are restricted to suitable intervals, inverse functions do exist.

By definition of π, we have $\cos x > 0$ for $x \in (-\frac{\pi}{2}, \frac{\pi}{2})$, hence because of $\sin' x = \cos x$, the sine function is strictly increasing in the interval $[-\frac{\pi}{2}, \frac{\pi}{2}]$. Consequently, $\sin : [-\frac{\pi}{2}, \frac{\pi}{2}] \to [-1,1]$ has an inverse function. Moreover, inverse functions also exist to other restrictions of sine:

$$\sin : [\pi(n+\frac{1}{2}), \pi(n+\frac{3}{2})] \to [-1,1]\,, \quad n \in \mathbb{Z}\,.$$

If one speaks of the inverse function of sine, one has to specify which one of these infinitely many inverses are meant. If no specification is given, the inverse function

$$\arcsin :\ [-1,1] \to [-\frac{\pi}{2}, \frac{\pi}{2}]$$

of $\sin : [-\frac{\pi}{2}, \frac{\pi}{2}] \to [-1,1]$ is meant. Because of reasons, which have their origin in the theory of functions of a complex variable, the infinitely many inverse functions

$$x \mapsto (\arcsin x) + 2n\pi\,, \quad n \in \mathbb{Z}$$

and

$$x \mapsto -(\arcsin x) + (2n+1)\pi, \quad n \in \mathbb{Z}$$

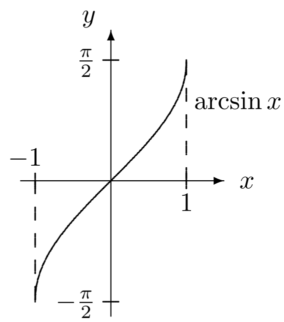
are called *branches of the inverse function of sine* or *branches of arc sine* ("Zweige des Arcussinus"). The function $\arcsin : [-1,1] \to [-\frac{\pi}{2}, \frac{\pi}{2}]$ is called *principle branch* of the inverse function ("Hauptwert der Umkehrfunktion").

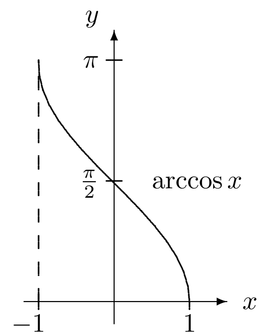
Correspondingly, the inverse function

$$\arccos : [-1,1] \to [0,\pi]$$

to the function cos : $[0, \pi] \to [-1, 1]$ is called principle branch of the inverse function of cosine, but there exist the infinitely many other inverse functions

$$x \to \pm(\arccos x) + 2n\pi, \quad n \in \mathbb{Z}.$$



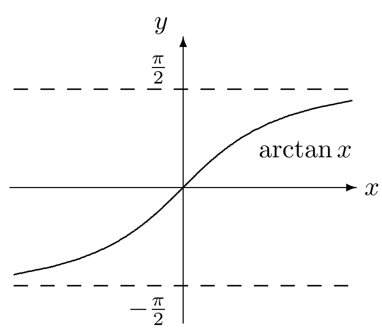


A similar situation arises with tangent and cotangent. The principle branch of the inverse function of tangent is the function

$$\arctan : [-\infty, \infty] \to \left[-\frac{\pi}{2}, \frac{\pi}{2} \right].$$

One calls this function arc tangent ("Arcustangens"), but there are infinitely many other branches of the inverse function

$$x \mapsto \quad \arctan x + n\pi, \quad n \in \mathbb{Z}$$



In the following we consider the principle branches of the inverse functions. For the derivatives one obtains

$$\begin{aligned}
(\arcsin x)' &= \frac{1}{\sin'(\arcsin x)} = \frac{1}{\cos(\arcsin x)} \\
&= \frac{1}{\sqrt{1-(\sin(\arcsin x))^2}} = \frac{1}{\sqrt{1-x^2}} \\
(\arccos x)' &= \frac{1}{\cos'(\arccos x)} = \frac{-1}{\sin(\arccos x)} \\
&= \frac{-1}{\sqrt{1-(\cos(\arccos x))^2}} = \frac{-1}{\sqrt{1-x^2}} \\
(\arctan x)' &= \frac{1}{\tan'(\arctan x)} = \big(\cos(\arctan x)\big)^2 \\
&= \frac{1}{1+(\tan(\arctan x))^2} = \frac{1}{1+x^2}.
\end{aligned}$$

The functions arcsin, arccos and arctan can be expanded into power series. For example,

$$\frac{d}{dt}(\arctan x) = \frac{1}{1+x^2} = \sum_{n=0}^{\infty} (-1)^n x^{2n},$$

if $|x| < 1$. Also the power series

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1}\, x^{2n+1}$$

has radius of convergence equal to 1, and it is an antiderivative of $\sum_{n=0}^{\infty} (-1)^n x^{2n}$, hence

$$\arctan x = \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1}\, x^{2n+1} + C$$

for $|x| < 1$, with a suitable constant C. From $\arctan 0 = 0$ we obtain $C = 0$, thus

$$\arctan x = \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1}\, x^{2n+1}$$

for all $x \in \mathbb{R}$ with $|x| < 1$. The convergence criterion of Leibniz shows that the power series on the right converges for $x = 1$, hence Abel's limit theorem implies that the function given by the power series is continuous at 1. Since arctan is continuous, the power series and the function arctan define two continuous extensions of the function arctan from the interval $(-1, 1)$ to $(-1, 1]$. Since the continuous extension is unique, we must have

$$\arctan 1 = \sum_{n=1}^{\infty} \frac{(-1)^n}{2n+1}\,.$$

Because of

$$\cos(2x) = (\cos x)^2 - (\sin x)^2 = 2(\cos x)^2 - 1,$$

it follows

$$0 = 2\Big(\cos\frac{\pi}{4}\Big)^2 - 1,$$

hence

$$\cos\frac{\pi}{4} = \sqrt{\frac{1}{2}}$$

and

$$\sin\frac{\pi}{4} = \sqrt{1 - (\cos\frac{\pi}{4})^2} = \sqrt{\frac{1}{2}}\,,$$

thus

$$\tan\frac{\pi}{4} = \frac{\sin\frac{\pi}{4}}{\cos\frac{\pi}{4}} = 1\,.$$

This yields

$$\arctan 1 = \frac{\pi}{4}\,,$$

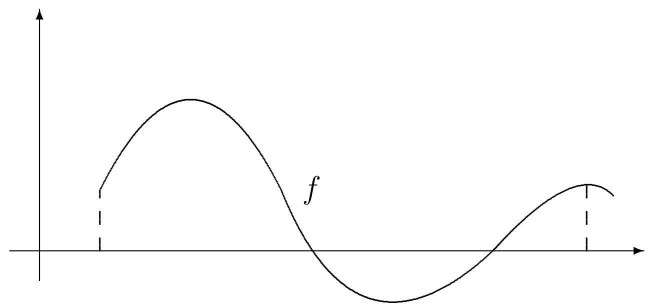
whence

$$\frac{\pi}{4} = \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1}\,.$$

Theoretically this series allows to compute π, but the convergence is slow.

2 The Riemann integral

For a class of real functions as large as possible one wants to determine the area of the surface bounded by the graph of the function and the abscissa. This area is called the integral of the function.



To determine this area might be a difficult task for functions as complicated as the Dirichlet function

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

and in fact, the Riemann integral, which we are going to discuss in this section, is not able to assign a surface area to this function. The Riemann integral was historically the first rigorous notion of an integral. It was introduced by Riemann in his Habilitation thesis 1854. Today mathematicians use a more general and advanced integral, the Lebesgue integral, which can assign an area to the Dirichlet function. The value of the Lebesgue integral of the Dirichlet function is 0. (Bernhard Riemann 1826 – 1866, Henri Lebesgue 1875 – 1941)

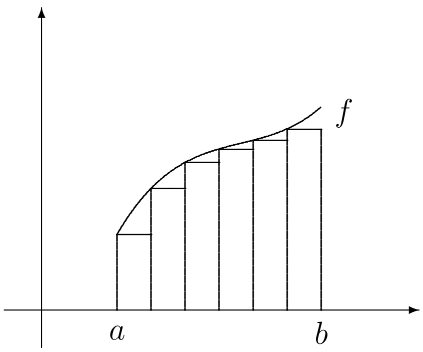
2.1 Definition of the Riemann integral

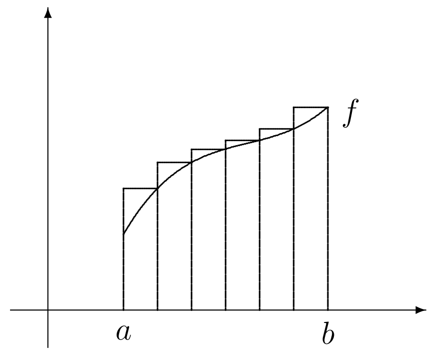
Let $-\infty < a < b < \infty$ and let $f : [a, b] \to \mathbb{R}$ be a given function. It suggests itself to compute the area below the graph of f by inscribing rectangles into this surface. If we refine the subdivision, the total area of these rectangles will converge to the area of the surface below the graph of f. It is also possible to cover the area below the graph of f by rectangles. Again, if the subdivision is refined, the total area of these rectangles will converge to the area of the surface below the graph of f.

Therefore one expects that in both approximating processes the total areas of the rectangles will converge to the same number. The area of the surface below the graph of f is defined to be this number.

Of course, the total areas of the inscribed rectangles and of the covering rectangles will not converge to the same number for all functions f. An example for this is the Dirichlet function.

Those functions f, for which these areas converge to the same number, are called Riemann integrable, and the number is called Riemann integral of f over the interval $[a, b]$.





This program will now be carried through rigorously.

Definition 2.1 *Let* $-\infty < a < b < \infty$ *. A partition* P *of the interval* $[a, b]$ *is a finite set* $\{x_0, \dots x_n\} \subseteq \mathbb{R}$ *with*

$$a = x_0 < x_1 < \dots < x_{n-1} < x_n = b\,.$$

For brevity we set $\Delta x_i = x_i - x_{x-1}$ $(i = 1, \dots, n)$ *.*

Let $f : [a, b] \to \mathbb{R}$ be a bounded real function and $P = \{x_0, \dots, x_n\}$ a partition of $[a, b]$. For $i = 1, \dots, n$ set

$$\begin{aligned} M_i &= \sup \{f(x) \mid x_{i-1} \leq x \leq x_i\}\ , \\ m_i &= \inf \{f(x) \mid x_{i-1} \leq x \leq x_i\}\ , \end{aligned}$$

and define

$$\begin{aligned} U(P, f) &= \sum_{i=1}^{n} M_i\,\Delta x_i \\ L(P, f) &= \sum_{i=1}^{n} m_i\,\Delta x_i\,. \end{aligned}$$

Since f is bounded, there exist numbers m, M such that

$$m \leq f(x) \leq M$$

for all $x \in [a,b]$. This implies $m \leq m_i \leq M_i \leq M$ for all $i = 1, \ldots, n$, hence

$$m(b-a) = \sum_{i=1}^{n} m\,\Delta x_i \leq \sum_{i=1}^{n} m_i\,\Delta x_i = L(P,f) \qquad (*)$$

$$\leq \sum_{i=1}^{n} M_i\,\Delta x_i = U(P,f) \leq \sum_{i=1}^{n} M\,\Delta x_i = M(b-a)\,.$$

Consequently, the infimum and the supremum

$$\overline{\int_a^b} f\,dx = \inf\{U(P,f) \mid P \text{ is a partition of } [a,b]\}$$

$$\underline{\int_a^b} f\,dx = \sup\{L(P,f) \mid P \text{ is a partition of } [a,b]\}$$

exist. The numbers $\overline{\int_a^b} f\,dx$ and $\underline{\int_a^b} f\,dx$ are called upper and lower Riemann integral of f.

Definition 2.2 *A bounded function $f : [a,b] \to \mathbb{R}$ is called Riemann integrable, if the upper Riemann integral $\overline{\int_a^b} f\,dx$ and the lower Riemann integral $\underline{\int_a^b} f\,dx$ coincide. The common value or the upper and lower Riemann integral is denoted by*

$$\int_a^b f\,dx \quad \textit{or} \quad \int_a^b f(x)\,dx$$

and called the Riemann integral of f. The set of Riemann integrable functions defined on the interval $[a,b]$ is denoted by $\mathcal{R}([a,b])$.

2.2 Criteria for Riemann integrable functions

To work with Riemann integrable functions, one needs simple criteria for a function to be Riemann integrable. In this section we derive such criteria.

Definition 2.3 *Let P, P_1, P_2 and P^* be partitions of $[a,b]$. The partition P^* is called a refinement of P if $P \subseteq P^*$ holds. P^* is called common refinement of P_1 and P_2 if $P^* = P_1 \cup P_2$.*

Theorem 2.4 *Let $f : [a,b] \to \mathbb{R}$ and let P^* be a refinement of the partition P of $[a,b]$. Then*

$$L(P,f) \leq L(P^*,f)$$

$$U(P^*,f) \leq U(P,f).$$

Proof: Let $P = \{x_0, \ldots, x_n\}$ and assume first that P^* contains exactly one point x^* more than P. Then there are $x_{j-1}, x_j \in P$ with $x_{j-1} < x^* < x_j$. Let

$$\begin{aligned} w_1 &= \inf\{f(x) \,\big|\, x_{j-1} \le x \le x^*\}, \\ w_2 &= \inf\{f(x) \,\big|\, x^* \le x \le x_j\}, \end{aligned}$$

and for $i = 1, \ldots, n$

$$m_i = \inf\{f(x) \,\big|\, x_{i-1} \le x \le x_i\}.$$

Then $w_1, w_2 \ge m_j$, hence

$$\begin{aligned} L(P,f) &= \sum_{i=1}^{n} m_i \Delta x_i = \sum_{i=1}^{j-1} m_i \Delta x_i \\ &\quad + m_j(x^* - x_{j-1} + x_j - x^*) + \sum_{i=j+1}^{n} m_i \Delta x_i \\ &\le \sum_{i=1}^{j-1} m_i \Delta x_i + w_1(x^* - x_{j-1}) + w_2(x_j - x^*) + \sum_{i=j+1}^{n} m_i \Delta x_i \\ &= L(P^*, f). \end{aligned}$$

By induction we conclude that $L(P,f) \le L(P^*, f)$ holds if P^* contains k points more than P for any k. The second inequality stated in the theorem is proved analogously. ∎

Theorem 2.5 *Let $f : [a,b] \to \mathbb{R}$ be bounded. Then*

$$\underline{\int_a^b} f\, dx \le \overline{\int_a^b} f\, dx.$$

Proof: Let P_1 and P_2 be partitions and let P^* be the common refinement. Inequality $(*)$ proved above shows that

$$L(P^*, f) \le U(P^*, f).$$

Combination of this inequality with the preceding theorem yields

$$L(P_1, f) \le L(P^*, f) \le U(P^*, f) \le U(P_2, f),$$

whence

$$L(P_1, f) \le U(P_2, f)$$

for all partitions P_1 and P_2 of $[a,b]$. Therefore $U(P_2, f)$ is an upper bound of the set

$$\big\{L(P,f) \,\big|\, P \text{ is a partition of } [a,b]\big\},$$

hence the least upper bound $\underline{\int_a^b} f\,dx$ of this set satisfies

$$\underline{\int_a^b} f\,dx \le U(P_2, f).$$

Since this inequality holds for every partition P_2 of $[a, b]$, it follows that $\underline{\int_a^b} f\,dx$ is a lower bound of the set

$$\{U(P, f) \,|\, P \text{ is a partition of } [a, b]\},$$

hence the greatest lower bound of this set satisfies

$$\underline{\int_a^b} f\,dx \le \overline{\int_a^b} f\,dx.$$

■

Theorem 2.6 *Let $f : [a, b] \to \mathbb{R}$ be bounded. The function f belongs to $\mathcal{R}([a, b])$ if and only if to every $\varepsilon > 0$ there is a partition P of $[a, b]$ such that*

$$U(P, f) - L(P, f) < \varepsilon.$$

Proof: First assume that to every $\varepsilon > 0$ there is a partition P with $U(P, f)-L(P, f) < \varepsilon$. Since

$$L(P, f) \le \underline{\int_a^b} f\,dx \le \overline{\int_a^b} f\,dx \le U(P, f),$$

it follows that

$$0 \le \overline{\int_a^b} f\,dx - \underline{\int_a^b} f\,dx \le U(P, f) - L(P, f) < \varepsilon,$$

hence

$$0 \le \overline{\int_a^b} f\,dx - \underline{\int_a^b} f\,dx < \varepsilon$$

for every $\varepsilon > 0$. This implies

$$\overline{\int_a^b} f\,dx = \underline{\int_a^b} f\,dx,$$

thus $f \in \mathcal{R}([a, b])$.

Conversely, let $f \in \mathcal{R}([a, b])$. By definition of the infimum and the supremum to every $\varepsilon > 0$ there are partitions P_1 and P_2 with

$$\begin{aligned}
\int_a^b f\,dx &= \overline{\int_a^b} f\,dx \le U(P_1, f) < \overline{\int_a^b} f\,dx + \frac{\varepsilon}{2} \\
\int_a^b f\,dx &= \underline{\int_a^b} f\,dx \ge L(P_2, f) > \underline{\int_a^b} f\,dx - \frac{\varepsilon}{2}.
\end{aligned}$$

Let P be the common refinement of P_1 and P_2. Then

$$\int_a^b f\,dx - \frac{\varepsilon}{2} < L(P,f) \le \int_a^b f\,dx \le U(P,f) < \int_a^b f\,dx + \frac{\varepsilon}{2},$$

hence

$$U(P,f) - L(P,f) < \varepsilon.$$

■

From this theorem we can conclude that $C([a,b]) \subseteq \mathcal{R}([a,b])$:

Theorem 2.7 *Let $f : [a,b] \to \mathbb{R}$ be continuous. Then f is Riemann integrable. Furthermore, to every $\varepsilon > 0$ there is $\delta > 0$ such that*

$$\Big|\sum_{i=1}^{n} f(t_i)\Delta x_i - \int_a^b f\,dx\Big| < \varepsilon$$

for every partition $P = \{x_0, \ldots, x_n\}$ of $[a,b]$ with

$$\max\{\Delta x_1, \ldots, \Delta x_n\} < \delta$$

and for every choice of points $t_1, \ldots, t_n$ with $t_i \in [x_{i-1}, x_i]$.

Note that if $\{P_j\}_{j=1}^{\infty}$ is a sequence of partitions $P_j = \{x_0^{(j)} = a,\ x_1^{(j)}, \ldots, x_{n_j}^{(j)} = b\}$ of $[a,b]$ with

$$\lim_{j\to\infty} \max\{\Delta x_1^{(j)}, \ldots, \Delta x_{n_j}^{(j)}\} = 0$$

and if $t_i^{(j)} \in [x_{i-1}^{(j)}, x_i^{(j)}]$, then this theorem implies

$$\int_a^b f\,dx = \lim_{j\to\infty} \sum_{i=1}^{n_j} f(t_i^{(j)})\Delta x_i^{(j)}.$$

The integral is the limit of the *Riemann sums* $\sum_{i=1}^{n} f(t_i)\Delta x_i$.

Proof: Let $\varepsilon > 0$. We set

$$\eta = \frac{\varepsilon}{b-a}.$$

As a continuous function on the compact interval $[a,b]$, the function f is bounded and uniformly continuous (cf. Theorem 6.43 of the lecture notes to the Analysis I course). Therefore there exists $\delta > 0$ such that for all $x, t \in [a,b]$ with $|x - t| < \delta$

$$|f(x) - f(t)| < \eta. \qquad (*)$$

We choose a partition $P = \{x_0, \ldots, x_n\}$ of $[a,b]$ with $\max\{\Delta x_1, \ldots, \Delta x_n\} < \delta$. Then $(*)$ implies for all $x, t \in [x_{i-1}, x_i]$

$$f(x) - f(t) < \eta,$$

hence

$$\begin{aligned} M_i - m_i &= \sup_{x_{i-1} \le x \le x_i} f(x) - \inf_{x_{i-1} \le t \le x_i} f(t) \\ &= \max_{x_{i-1} \le x \le x_i} f(x) - \min_{x_{i-1} \le t \le x_i} f(t) \\ &= f(x_0) - f(t_0) < \eta, \end{aligned}$$

for suitable $x_0, t_0 \in [x_{i-1}, x_i]$. This yields

$$\begin{aligned} U(P, f) - L(P, f) &= \sum_{i=1}^{n} (M_i - m_i) \Delta x_i < \eta \sum_{i=1}^{n} \Delta x_i \\ &= \eta(b - a) = \varepsilon. \end{aligned} \tag{$**$}$$

Since $\varepsilon > 0$ was arbitrary, the preceding theorem implies $f \in \mathcal{R}([a, b])$. From $(**)$ and from the inequalities

$$\begin{aligned} L(P, f) &= \sum_{i=1}^{n} m_i \Delta x_i \le \sum_{i=1}^{n} f(t_i) \Delta x_i \le \sum_{i=1}^{n} M_i \Delta x_i \le U(P, f) \\ L(P, f) &\le \int_a^b f \, dx \le U(P, f) \end{aligned}$$

we infer that

$$\Big| \int_a^b f \, dx - \sum_{i=1}^{n} f(t_i) \Delta x_i \Big| < \varepsilon. \qquad \blacksquare$$

Also the class of monotone functions is a subset of $\mathcal{R}([a, b])$:

Theorem 2.8 *Let $f : [a, b] \to \mathbb{R}$ be monotone. Then f is Riemann integrable.*

Proof: Assume that f is increasing. f is bounded because of $f(a) \le f(x) \le f(b)$ for all $x \in [a, b]$. Let $\varepsilon > 0$. To arbitrary $n \in \mathbb{N}$ set

$$x_i = a + \frac{b - a}{n} i,$$

for $i = 0, 1, \ldots, n$. Then $P = \{x_0, \ldots, x_n\}$ is a partition of $[a, b]$, and since f is increasing we obtain

$$\begin{aligned} m_i &= \inf_{x_{i-1} \le x \le x_i} f(x) &= f(x_{i-1}) \\ M_i &= \sup_{x_{i-1} \le x \le x_i} f(x) &= f(x_i), \end{aligned}$$

thence

$$U(P, f) - L(P, f) = \sum_{i=1}^{n} (M_i - m_i) \Delta x_i$$

$$= \sum_{i=1}^{n} \Big(f(x_i) - f(x_{i-1}) \Big) \frac{b - a}{n} = \Big(f(b) - f(a) \Big) \frac{b - a}{n} < \varepsilon,$$

where the last inequality sign holds if $n \in \mathbb{N}$ is chosen sufficiently large. By Theorem 2.6, this inequality shows that $f \in \mathcal{R}([a,b])$.

For decreasing f the proof is analogous. ■

Example: Let $-\infty < a < b < \infty$. The function $\exp : [a,b] \to \mathbb{R}$ is continuous and therefore Riemann integrable. The value of the integral is

$$\int_a^b e^x dx = e^b - e^a.$$

To verify this equation we use Theorem 2.7. For every $n \in \mathbb{N}$ and all $i = 0, 1, \dots, n$ we set $x_i^{(n)} = a + \frac{i}{n}(b-a)$. Then $\{P_n\}_{n=1}^\infty$ with $P_n = \{x_0^{(n)}, \dots, x_n^{(n)}\}$ is a sequence of partitions of $[a,b]$ satisfying

$$\lim_{n\to\infty} \max\{\Delta x_1^{(n)}, \dots, \Delta x_n^{(n)}\} = \lim_{n\to\infty} \frac{b-a}{n} = 0.$$

Thus, with $t_i^{(n)} = x_{i-1}^{(n)}$ we obtain

$$\begin{aligned}
\int_a^b e^x dx &= \lim_{n\to\infty} \sum_{i=1}^n \exp(t_i^{(n)}) \Delta x_i^{(n)} \\
&= \lim_{n\to\infty} \sum_{i=1}^n \exp\Big(a + \frac{i-1}{n}(b-a)\Big) \frac{b-a}{n} \\
&= \lim_{n\to\infty} e^a \frac{b-a}{n} \sum_{i=1}^n \left[e^{(b-a)/n}\right]^{i-1} \\
&= e^a \lim_{n\to\infty} \frac{b-a}{n} \frac{\left[e^{(b-a)/n}\right]^n - 1}{e^{(b-a)/n} - 1} \\
&= \frac{e^a(e^{b-a} - 1)}{\lim_{n\to\infty} \frac{e^{(b-a)/n} - 1}{(b-a)/n}} = e^b - e^a,
\end{aligned}$$

since $\lim_{x\to 0} \frac{e^x - 1}{x} = 1$, by the rule of de l'Hospital.

2.3 Simple properties of the integral

Theorem 2.9 (i) *If $f_1, f_2 \in \mathcal{R}([a,b])$, then $f_1 + f_2 \in \mathcal{R}([a,b])$, and*

$$\int_a^b (f_1 + f_2) dx = \int_a^b f_1\, dx + \int_a^b f_2\, dx.$$

If $g \in \mathcal{R}([a,b])$ and $c \in \mathbb{R}$, then $cg \in \mathcal{R}([a,b])$ and

$$\int_a^b cg\, dx = c \int_a^b g\, dx.$$

Hence $\mathcal{R}([a,b])$ *is a vector space.*

(ii) *If* $f_1, f_2 \in \mathcal{R}([a,b])$ *and* $f_1(x) \le f_2(x)$ *for all* $x \in [a,b]$, *then*

$$\int_a^b f_1\,dx \le \int_a^b f_2\,dx.$$

(iii) *If* $f \in \mathcal{R}([a,b])$ *and if* $a < c < b$, *then*

$$f\big|_{[a,c]} \in \mathcal{R}([a,b]), \quad f\big|_{[c,b]} \in \mathcal{R}([a,b])$$

and

$$\int_a^c f\,dx + \int_c^b f\,dx = \int_a^b f\,dx.$$

(iv) *If* $f \in \mathcal{R}([a,b])$ *and* $|f(x)| \le M$ *for all* $x \in [a,b]$, *then*

$$\Big|\int_a^b f\,dx\Big| \le M(b-a).$$

Proof: (i) Let $f = f_1 + f_2$ and let P be a partition of $[a,b]$. Then

$$\begin{aligned}
\inf_{x_{i-1}\le x\le x_i} f(x) &= \inf_{x_{i-1}\le x\le x_i} \Big(f_1(x) + f_2(x)\Big) \\
&\ge \inf_{x_{i-1}\le x\le x_i} f_1(x) + \inf_{x_{i-1}\le x\le x_i} f_2(x) \\
\sup_{x_{-1}\le x\le x_i} f(x) &= \sup_{x_{i-1}\le x\le x_i} \Big(f_1(x) + f_2(x)\Big) \\
&\le \sup_{x_{i-1}\le x\le x_i} f_1(x) + \sup_{x_{i-1}\le x\le x_i} f_2(x),
\end{aligned}$$

hence

$$\begin{aligned}
L(P,f_1) + L(P,f_2) &\le L(P,f) \\
U(P,f) &\le U(P,f_1) + U(P,f_2).
\end{aligned} \tag{$*$}$$

Let $\varepsilon > 0$. Since f_1 and f_2 are Riemann integrable, there exist partitions P_1 and P_2 such that for $j = 1,2$

$$U(P_j, f_j) - L(P_j, f_j) < \varepsilon.$$

For the common refinement P of P_1 and P_2 we have $L(P_j,f_j) \le L(P,f_j)$ and $U(P,f_j) \le U(P_j,f_j)$, hence, for $j = 1,2$,

$$U(P,f_j) - L(P,f_j) < \varepsilon. \tag{$**$}$$

From this inequality and from $(*)$ we obtain

$$U(P,f) - L(P,f) \le U(P,f_1) + U(P,f_2) - L(P,f_1) - L(P,f_2) < 2\varepsilon.$$

Since $\varepsilon > 0$ was chosen arbitrarily, this inequality and Theorem 2.6 imply $f = f_1 + f_2 \in \mathcal{R}([a,b])$.

From $(**)$ we also obtain

$$U(P, f_j) < L(P, f_j) + \varepsilon \leq \int_a^b f_j \, dx + \varepsilon \,,$$

whence, observing $(*)$

$$\begin{aligned} \int_a^b f \, dx &\leq U(P, f) \leq U(P, f_1) + U(P, f_2) \\ &\leq \int_a^b f_1 \, dx + \int_a^b f_2 \, dx + 2\varepsilon \,. \end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, this yields

$$\int_a^b f \, dx \leq \int_a^b f_1 \, dx + \int_a^b f_2 \, dx \,. \qquad (***)$$

Similarly, $(**)$ yields

$$L(P, f_j) > U(P, f_j) - \varepsilon \geq \int_a^b f \, dx - \varepsilon \,,$$

which together with $(*)$ results in

$$\begin{aligned} \int_a^b f \, dx &\geq L(P, f) \geq L(P, f_1) + L(P, f_2) \\ &\geq \int_a^b f_1 \, dx + \int_a^b f_2 \, dx - 2\varepsilon \,, \end{aligned}$$

from which we conclude that

$$\int_a^b f \, dx \geq \int_a^b f_1 \, dx + \int_a^b f_2 \, dx \,.$$

This inequality and $(***)$ yield

$$\int_a^b f \, dx = \int_a^b f_1 \, dx + \int_a^b f_2 \, dx \,.$$

To prove that $cg \in \mathcal{R}([a,b])$ we note that the definition of $L(P, cg)$ immediately yields for every partition P of $[a,b]$

$$L(P, cg) = \begin{cases} cL(P, g) \,, & \text{if } c \geq 0 \\ cU(P, g) \,, & \text{if } c < 0 \,. \end{cases}$$

Thus, for $c \geq 0$

$$\begin{aligned}
\underline{\int_a^b} cg\,dx &= \sup\big\{cL(P,g)\ \big|\ P \text{ is a partition of } [a,b]\big\} \\
&= c\,\sup\big\{L(P,g)\ \big|\ P \text{ is a partition of } [a,b]\big\} = c\underline{\int_a^b} g\,dx = c\int_a^b g\,dx\,,
\end{aligned}$$

and for $c < 0$

$$\begin{aligned}
\underline{\int_a^b} cg\,dx &= \sup\big\{cU(P,g)\ \big|\ P \text{ is a partition of } [a,b]\big\} \\
&= c\inf\big\{U(P,g)\ \big|\ P \text{ is a partition of } [a,b]\big\} = c\overline{\int_a^b} g\,dx = c\int_a^b g\,dx\,.
\end{aligned}$$

In the same manner

$$\overline{\int_a^b} cg\,dx = c\int_a^b g\,dx\,.$$

Therefore

$$\underline{\int_a^b} cg\,dx = c\int_a^b g\,dx = \overline{\int_a^b} cg\,dx\,,$$

which implies $cg \in \mathcal{R}([a,b])$ and $\int_a^b cg\,dx = c\int_a^b g\,dx$.

This completes the proof of (i). The proof of (ii) is left as an exercise. To prove (iii), note first that to any partition P of $[a,b]$ we can define a refinement P^* by

$$P^* = P \cup \{c\}\,.$$

Theorem 2.4 implies

$$L(P,f) \leq L(P^*,f), \quad U(P^*,f) \leq U(P,f)\,. \tag{$*$}$$

From P^* we obtain partitions P^*_- of $[a,c]$ and P^*_+ of $[c,b]$ by setting $P^*_- = P^* \cap [a,c]$ and $P^*_+ = P^* \cap [c,b]$, and if $P^* = \{x_0,\ldots,x_n\}$ with $x_j = c$, then

$$L(P^*,f) = \sum_{i=1}^n m_i\,\Delta x_i = \sum_{i=1}^j m_i\,\Delta x_i + \sum_{i=j+1}^n m_i\,\Delta x_i = L(P^*_-,f) + L(P^*_+,f)\,.$$

Here for simplicity we wrote $L(P^*_-,f)$ instead of $L(P^*_-,f|_{[a,c]})$ and $U(P^*_-,f)$ instead of $U(P^*_-,f|_{[c,b]})$. Similarly

$$U(P^*,f) = U(P^*_-,f) + U(P^*_+,f)\,.$$

From $(*)$ and from these equations we conclude

$$\begin{aligned} L(P,f) &\leq L(P_-^*,f)+L(P_+^*,f) \leq \underline{\int_a^c} f\,dx + \underline{\int_c^b} f\,dx \\ U(P,f) &\geq U(P_-^*,f)+U(P_+^*,f) \geq \overline{\int_a^c} f\,dx + \overline{\int_c^b} f\,dx\,. \end{aligned}$$

These estimates hold for any partition P of $[a,b]$, whence

$$\begin{aligned} \int_a^b f\,dx &= \underline{\int_a^b} f\,dx \leq \underline{\int_a^c} f\,dx + \underline{\int_c^b} f\,dx \\ \int_a^b f\,dx &= \overline{\int_a^b} f\,dx \geq \overline{\int_a^c} f\,dx + \overline{\int_c^b} f\,dx\,. \end{aligned}$$

Since $\underline{\int_a^c} f\,dx \leq \overline{\int_a^c} f\,dx$ and $\underline{\int_c^b} f\,dx \leq \overline{\int_c^b} f\,dx$, these inequalities can only hold if

$$\underline{\int_a^c} f\,dx = \overline{\int_a^c} f\,dx\,, \quad \underline{\int_c^b} f\,dx = \overline{\int_c^b} f\,dx\,,$$

hence $f|_{[a,c]} \in \mathcal{R}([a,c])$, $f|_{[c,b]} \in \mathcal{R}([c,b])$, and

$$\int_a^c f\,dx + \int_c^b f\,dx = \int_a^b f\,dx\,.$$

This proves (iii). The obvious proof of (iv) is left as an exercise. ■

Theorem 2.10 *Let $-\infty < m < M < \infty$ and $f \in \mathcal{R}([a,b])$ with $f : [a,b] \to [m,M]$. Let $\Phi : [\,m,M] \to \mathbb{R}$ be continuous and let $h = \Phi \circ f$. Then $h \in \mathcal{R}([a,b])$.*

Proof: Let $\varepsilon > 0$. Since Φ is uniformly continuous on $[m,M]$, there is a number $\delta > 0$ such that for all $s,t \in [m,M]$ with $|s-t| \leq \delta$

$$|\Phi(s) - \Phi(t)| < \varepsilon\,.$$

Moreover, since $f \in \mathcal{R}([a,b])$ there is a partition $P = \{x_0, \ldots, x_n\}$ of $[a,b]$ such that

$$U(P,f) - L(P,f) < \varepsilon\delta\,. \qquad (*)$$

Let

$$\begin{aligned} M_i &= \sup_{x_{i-1}\leq x\leq x_i} f(x)\,, & m_i &= \inf_{x_{i-1}\leq x\leq x_i} f(x) \\ M_i^* &= \sup_{x_{i-1}\leq x\leq x_i} h(x)\,, & m_i^* &= \inf_{x_{i-1}\leq x\leq x_i} h(x) \end{aligned}$$

and

$$\begin{aligned} A &= \big\{ i \,\big|\, i \in \mathbb{N}, \quad 1 \le i \le n, \quad M_i - m_i < \delta \big\} \\ B &= \{1, \dots, n\} \setminus A \,. \end{aligned}$$

If $i \in A$, then for all x, y with $x_{i-1} \le x, y \le x_i$

$$|h(x) - h(y)| = |\Phi\big(f(x)\big) - \Phi\big(f(y)\big)| < \varepsilon \,,$$

since $|f(x) - f(y)| \le M_i - m_i < \delta$. This yields for $i \in A$

$$M_i^* - m_i^* \le \varepsilon \,.$$

If $i \in B$, then

$$M_i^* - m_i^* \le 2\|\Phi\| \,,$$

with the supremum norm $\|\Phi\| = \sup_{m \le t \le M} |\Phi(t)|$. Furthermore, $(*)$ yields

$$\delta \sum_{i \in B} \Delta x_i \le \sum_{i \in B} (M_i - m_i) \Delta x_i \le \sum_{i=1}^{n} (M_i - m_i) \Delta x_i = U(P, f) - L(P, f) < \varepsilon\delta \,,$$

whence

$$\sum_{i \in B} \Delta x_i \le \varepsilon \,.$$

Together we obtain

$$\begin{aligned} U(P, h) - L(P, h) &= \sum_{i \in A} (M_i^* - m_i^*) \Delta x_i + \sum_{i \in B} (M_i^* - m_i^*) \Delta x_i \\ &\le \varepsilon \sum_{i \in A} \Delta x_i + 2\|\Phi\| \sum_{i \in B} \Delta x_i \\ &\le \varepsilon(b - a) + 2\|\Phi\|\varepsilon = \varepsilon(b - a + 2\|\Phi\|). \end{aligned}$$

Since ε was chosen arbitrarily, we conclude from this inequality that $h \in \mathcal{R}([a, b])$, using Theorem 2.6. ■

Corollary 2.11 *Let $f, g \in \mathcal{R}([a, b])$. Then*

(i) $fg \in \mathcal{R}([a, b])$

(ii) $|f| \in \mathcal{R}([a, b])$ *and* $\Big| \int_a^b f \, dx \Big| \le \int_a^b |f| \, dx.$

Proof: (i) Setting $\Phi(t) = t^2$ in the preceding theorem yields $f^2 = \Phi \circ f \in \mathcal{R}([a, b])$. From

$$fg = \frac{1}{4} \left[(f + g)^2 - (f - g)^2 \right]$$

we conclude with this result that also $fg \in \mathcal{R}([a,b])$.
(ii) Setting $\Phi(t) = |t|$ in the preceding theorem yields $|f| = \Phi \circ f \in \mathcal{R}([a,b])$. Choose $c = \pm 1$ such that

$$c \int_a^b f\,dx \geq 0.$$

Then

$$\Big| \int_a^b f\,dx \Big| = c \int_a^b f\,dx = \int_a^b cf\,dx \leq \int_a^b |f| dx,$$

since $cf(x) \leq |f(x)|$ for all $x \in [a,b]$. ■

2.4 Fundamental theorem of calculus

Let $-\infty < a < b < \infty$ and $f \in \mathcal{R}\big([a,b]\big)$. One defines

$$\int_b^a f\,dx = -\int_a^b f\,dx\,.$$

Then

$$\int_u^v f\,dx + \int_v^w f\,dx = \int_u^w f\,dx\,,$$

if u, v, w are arbitrary points of $[a,b]$.

Theorem 2.12 (Mean value theorem of integration) *Let $f : [a,b] \to \mathbb{R}$ be continuous. Then there is a point c with $a \leq c \leq b$ such that*

$$\int_a^b f\,dx = f(c)(b-a)\,.$$

Proof: f is Riemann integrable, since f is continuous. Since the integral is monotone, we obtain

$$\begin{aligned}(b-a)\min_{x\in[a,b]} f(x) &= \int_a^b \min_{y\in[a,b]} f(y)\,dx \leq \int_a^b f(x)\,dx \\ &\leq \int_a^b \max_{y\in[a,b]} f(y)\,dx = \max_{x\in[a,b]} f(x)(b-a)\,.\end{aligned}$$

Since f attains the minimum and the maximum on $[a,b]$, by the intermediate value theorem there exists a number $c \in [a,b]$ such that

$$f(c) = \frac{1}{b-a}\int_a^b f\,dx\,.$$

■

Theorem 2.13 *Let* $f \in \mathcal{R}\big([a,b]\big)$. *Then*

$$F(x) = \int_a^x f(t)\, dt$$

defines a continuous function $F : [a,b] \to \mathbb{R}$.

Proof: There is M with $|f(x)| \le M$ for all $x \in [a,b]$. Thus, for $x, x_0 \in [a,b]$ with $x_0 < x$

$$\Big|F(x) - F(x_0)\Big| = \Big|\int_a^x f(t)\, dt - \int_a^{x_0} f(t)\, dt\Big| = \Big|\int_{x_0}^x f(t)\, dt\Big| \le M(x - x_0)\,.$$

This estimate implies that F is continuous on $[a,b]$. ■

Theorem 2.14 *Let* $f \in \mathcal{R}\big([a,b]\big)$ *be continuous. Then the function* $F : [a,b] \to \mathbb{R}$ *defined by*

$$F(x) = \int_a^x f(t)\, dt$$

is continuously differentiable with

$$F' = f\,.$$

Therefore F *is an antiderivative of* f .

Proof: Let $x_0 \in [a,b]$. The mean value theorem of integration implies

$$\begin{aligned}
\lim_{x \to x_0} \frac{F(x) - F(x_0)}{x - x_0} &= \lim_{x \to x_0} \frac{1}{x - x_0} \Big[\int_a^x f(t)\, dt - \int_a^{x_0} f(t)\, dt\Big] \\
&= \lim_{x \to x_0} \frac{1}{x - x_0} \int_{x_0}^x f(t)\, dt = \lim_{x \to x_0} \frac{1}{x - x_0} f(y)(x - x_0) \\
&= \lim_{x \to x_0} f(y) = f(x_0)\,,
\end{aligned}$$

for suitable y between x_0 and x . Therefore F is differentiable with $F' = f$. Since f is continuous by assumption, F is continuously differentiable. ■

Theorem 2.15 (Fundamental theorem of calculus) *Let* F *be an antiderivative of the continuous function* $f : [a,b] \to \mathbb{R}$. *Then*

$$\int_a^b f(t)\, dt = F(b) - F(a) = F(x)\Big|_a^b\,.$$

Proof: The functions $x \mapsto \int_a^x f(t)\, dt$ and F both are antiderivatives of f . Since two antiderivatives differ at most by a constant c , we obtain

$$F(x) = \int_a^x f(t)\, dt + c$$

for all $x \in [a,b]$. This implies $c = F(a)$, whence $F(b) - F(a) = \int_a^b f(t)\,dt$. ■

This theorem is so important because it simplifies the otherwise so tedious computation of integrals.

Examples. 1.) Let $0 < a < b$ and $c \in \mathbb{R}$, $c \neq -1$. Then

$$\int_a^b x^c\,dx = \frac{1}{c+1}\,x^{c+1}\Big|_a^b .$$

For $c < -1$ one obtains

$$\lim_{m\to\infty} \int_a^m x^c\,dx = \lim_{m\to\infty} \frac{1}{c+1}\,m^{c+1} - \frac{1}{c+1}\,a^{c+1} = -\frac{1}{c+1}\,a^{c+1} .$$

Therefore one defines for $a > 0$ and $c < -1$

$$\int_a^\infty x^c\,dx := \lim_{m\to\infty} \int_a^m x^c\,dx = -\frac{1}{c+1}\,a^{c+1} .$$

The integral $\int_a^\infty x^c\,dx$ is called improper Riemann integral and one says that for $c < -1$ the function $x \mapsto x^c$ is improperly Riemann integrable over the interval $[a,\infty)$ with $a > 0$. In particular, one obtains

$$\int_1^\infty x^{-2}\,dx = 1 .$$

For $c < 0$ the function $x \mapsto x^c$ is not defined at $x = 0$ and unbounded on every interval $(0,b]$ with $b > 0$. Therefore the Riemann integral $\int_0^b x^c\,dx$ is not defined. However, for $-1 < c < 0$ one obtains

$$\lim_{\substack{\varepsilon\to 0\\ \varepsilon>0}} \int_\varepsilon^b x^c\,dx = \frac{1}{c+1}\,b^{c+1} - \lim_{\substack{\varepsilon\to 0\\ \varepsilon>0}} \frac{1}{c+1}\,\varepsilon^{c+1} = \frac{1}{c+1}\,b^{c+1} .$$

Therefore the improper Riemann integral

$$\int_0^b x^c\,dx := \lim_{\substack{\varepsilon\to 0\\ \varepsilon>0}} \int_\varepsilon^b x^c\,dx = \frac{1}{c+1}\,b^{c+1}$$

is defined, x^c is improperly Riemann integrable over $(0,b]$ for $-1 < c < 0$ and $b > 0$. In particular, one obtains

$$\int_0^1 x^{-\frac{1}{2}}\,dx = 2 .$$

2.) For $0 < a < b < \infty$

$$\int_a^b \frac{1}{x}\,dx = \log b - \log a .$$

Neither of the limits $\lim_{b\to\infty}\int_a^b \frac{1}{x}\,dx$, $\lim_{a\to 0}\int_a^b \frac{1}{x}\,dx$ exists, so x^{-1} is not improperly Riemann integrable over $[a,\infty)$ or $(0,b]$.

3.) Let $-1<a<b<1$. Then

$$\int_a^b \frac{1}{\sqrt{1-x^2}}\,dx = \arcsin b - \arcsin a\,.$$

One defines

$$\int_{-1}^{1} \frac{1}{\sqrt{1-x^2}}\,dx = \lim_{\substack{b\to 1\\ b<1}} \lim_{\substack{a\to -1\\ a>-1}} \int_a^b \frac{1}{\sqrt{1-x^2}}\,dx$$

$$= \lim_{\substack{b\to 1\\ b<1}} \arcsin b - \lim_{\substack{a\to -1\\ a>-1}} \arcsin a = \frac{\pi}{2} - \Big(-\frac{\pi}{2}\Big) = \pi\,.$$

$\frac{1}{\sqrt{1-x^2}}$ is improperly Riemann integrable over the interval $(-1,1)$.

Theorem 2.16 (Substitution) *Let f be continuous, let $g:[a,b]\to\mathbb{R}$ be continuously differentiable and let the composition $f\circ g$ be defined. Then*

$$\int_a^b f\big(g(t)\big)\,g'(t)\,dt = \int_{g(a)}^{g(b)} f(x)\,dx\,.$$

Proof: Since g is a continuous function defined on a compact interval, the range of g is a compact interval $[c,d]$. Therefore we can restrict f to this interval. As a continuous function, $f:[c,d]\to\mathbb{R}$ is Riemann integrable, hence has an antiderivative $F:[c,d]\to\mathbb{R}$. The chain rule implies

$$(F\circ g)' = (F'\circ g)\cdot g' = (f\circ g)\cdot g'\,,$$

whence

$$F\big(g(b)\big) - F\big(g(a)\big) = \int_a^b f\big(g(t)\big)\,g'(t)\,dt\,.$$

Combination of this equation with

$$F\big(g(b)\big) - F\big(g(a)\big) = \int_{g(a)}^{g(b)} f(x)\,dx$$

yields the statement. ∎

Remark: If g^{-1} exists, the rule of substitution can be written in the form

$$\int_a^b f(x)\,dx = \int_{g^{-1}(a)}^{g^{-1}(b)} f\big(g(t)\big)\,g'(t)\,dt\,.$$

Example. We want to compute $\int_0^1 \sqrt{1-x^2}\,dx$. With the substitution $x = x(t) = \cos t$ it follows because of the invertibility of cosine on the interval $[0, \frac{\pi}{2}]$ that

$$\begin{aligned}\int_0^1 \sqrt{1-x^2}\,dx &= \int_{x^{-1}(0)}^{x^{-1}(1)} \sqrt{1-x(t)^2}\,\frac{dx(t)}{dt}\,dt \\ &= \int_{\frac{\pi}{2}}^{0} \sqrt{1-(\cos t)^2}\,(-\sin t)\,dt = \int_0^{\pi/2} (\sin t)^2\,dt \\ &= \int_0^{\pi/2} \big(\frac{1}{2} - \frac{1}{2}\cos(2t)\big)\,dt = \frac{\pi}{4} - \frac{1}{4}\sin(2t)\Big|_0^{\pi/2} = \frac{\pi}{4}\,,\end{aligned}$$

where we used the addition theorem for cosine:

$$\cos(2t) = \cos(t+t) = (\cos t)^2 - (\sin t)^2 = 1 - (\sin t)^2 - (\sin t)^2 = 1 - 2(\sin t)^2\,.$$

Theorem 2.17 (Product integration) *Let $f : [a,b] \to \mathbb{R}$ be continuous, let F be an antiderivative of f and let $g : [a,b] \to \mathbb{R}$ be continuously differentiable. Then*

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

Proof: The product rule gives $(F \cdot g)' = F' \cdot g + F \cdot g' = f \cdot g + F \cdot g'$, thus

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

■

Example. With $f(x) = g(x) = \sin x$ and $F(x) = -\cos x$ we obtain

$$\begin{aligned}\int_0^\pi (\sin x)^2\,dx &= -\cos x\,\sin x\Big|_0^\pi + \int_0^\pi (\cos x)^2\,dx \\ &= -\cos x\,\sin x\Big|_0^\pi + \int_0^\pi \big(1-(\sin x)^2\big)\,dx = \pi - \int_0^\pi (\sin x)^2\,dx\,,\end{aligned}$$

hence

$$\int_0^\pi (\sin x)^2\,dx = \frac{\pi}{2}\,.$$

3 Continuous mappings on $\mathbb{R}^n$

3.1 Norms on $\mathbb{R}^n$

Let $n \in \mathbb{N}$. On the set of all n-tupels of real numbers

$$\{x = (x_1, x_2, \dots, x_n) \mid x_i \in \mathbb{R},\ i = 1, \dots, n\}$$

the operations of addition and multiplication by real numbers are defined by

$$\begin{aligned} x + y &:= (x_1 + y_1, \dots, x_n + y_n) \\ cx &:= (cx_1, \dots, cx_n). \end{aligned}$$

The set of n-tupels together with these operations is a vector space denoted by $\mathbb{R}^n$. A basis of this vector space is for example given by

$$e_1 = (1, 0, \dots, 0),\ e_2 = (0, 1, 0, \dots, 0), \dots,\ e_n = (0, \dots, 0, 1).$$

On $\mathbb{R}^n$, norms can be defined in different ways. I consider three examples of norms:

1.) **The maximum norm:**

$$\|x\|_\infty := \max\{|x_1|, \dots, |x_n|\}.$$

To prove that this is a norm, the properties

(i) $\|x\|_\infty = 0 \iff x = 0$

(ii) $\|cx\|_\infty = |c|\,\|x\|_\infty$ (positive homogeneity)

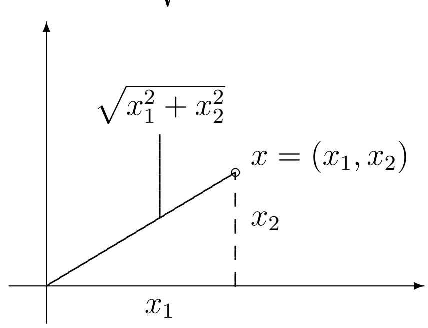
(iii) $\|x + y\|_\infty \le \|x\|_\infty + \|y\|_\infty$ (triangle inequality)

must be verified. (i) and (ii) are obviously satisfied. To prove (iii) note that there exists $i \in \{1, \dots, n\}$ such that $\|x + y\|_\infty = |x_i + y_i|$. Then

$$\|x + y\|_\infty = |x_i + y_i| \le |x_i| + |y_i| \le \|x\|_\infty + \|y\|_\infty.$$

2.) **The Euclidean norm:**

$$|x| := \sqrt{x_1^2 + \dots + x_n^2}.$$



Using the **scalar product**

$$x \cdot y := x_1 y_1 + x_2 y_2 + \ldots + x_n y_n \in \mathbb{R}$$

this can also be written as

$$|x| = \sqrt{x \cdot x}\,.$$

It is obvious that $|x| = 0 \iff x = 0$ and $|cx| = |c|\,|x|$ hold. To verify that $|\cdot|$ is a norm on $\mathbb{R}^n$, it thus remains to verify the triangle inequality. To this end one first proves the

Cauchy-Schwarz inequality

$$|x \cdot y| \leq |x|\,|y|\,.$$

Proof: The quadratic polynomial in t

$$|x|^2 t^2 + 2\,x \cdot y\,t + |y|^2 = |tx + y|^2 \geq 0$$

cannot have two different zeros, whence the discriminant must satisfy

$$(x \cdot y)^2 - |x|^2\,|y|^2 \leq 0\,.$$

■

Now the triangle inequality is obtained as follows:

$$\begin{aligned} |x+y|^2 &= (x+y) \cdot (x+y) = |x|^2 + 2x \cdot y + |y|^2 \\ &\leq |x|^2 + 2|x \cdot y| + |y|^2 \\ &\leq |x|^2 + 2|x|\,|y| + |y|^2 = (|x| + |y|)^2\,, \end{aligned}$$

whence

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

3.) **The p-norm:**

Let p be a real number with $p \geq 1$. Then the p-norm is defined by

$$\|x\|_p := \big(|x_1|^p + \ldots + |x_n|^p\big)^{\frac{1}{p}}\,.$$

Note that the 2-norm is the Euclidean norm:

$$\|x\|_2 = |x|\,.$$

Here we only verify that $\|\cdot\|_1$ is a norm. Since $\|x\|_1 = 0 \iff x = 0$ and $\|cx\|_1 = |c|\,\|x\|_1$ are evident, we have to show that the triangle inequality is satisfied:

$$\|x + y\|_1 = \sum_{i=1}^{n} |x_i + y_i| \leq \sum_{i=1}^{n} \big(|x_i| + |y_i|\big) = \|x\|_1 + \|y\|_1\,.$$

Definition 3.1 *Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$. A sequence $\{x_k\}_{k=1}^{\infty}$ with $x_k \in \mathbb{R}^n$ is said to converge, if $a \in \mathbb{R}^n$ exists such that*

$$\lim_{k\to\infty} \|x_k - a\| = 0\,.$$

a is called limit or limit element of the sequence $\{x_k\}_{k=1}^{\infty}$.

Just as in $\mathbb{R} = \mathbb{R}^1$ one proves that a sequence cannot converge to two different limit elements. Hence the limit of a sequence is unique. This limit is denoted by

$$a = \lim_{k\to\infty} x_k\,.$$

In this definition of convergence on $\mathbb{R}^n$ a norm is used. Hence, it seems that convergence of a sequence depends on the norm chosen. The following results show that this is not the case.

Lemma 3.2 *A sequence $\{x_k\}_{k=1}^{\infty}$ with $x_k = \big(x_k^{(1)}, \ldots, x_k^{(n)}\big) \in \mathbb{R}^n$ converges to $a = (a^{(1)}, \ldots, a^{(n)})$ with respect to the maximum norm, if and only if every sequence of components $\big\{x_k^{(i)}\big\}_{k=1}^{\infty}$ converges to $a^{(i)}$, $i = 1, \ldots, n$.*

Proof: The statement follows immediately from the inequalities

$$|x_k^{(i)} - a^{(i)}| \leq \|x_k - a\|_\infty \leq |x_k^{(1)} - a^{(1)}| + \ldots + |x^{(n)} - a^{(n)}|\,.$$

■

Theorem 3.3 *Let $\{x_k\}_{k=1}^{\infty}$ with $x_k \in \mathbb{R}^n$ be a sequence bounded with respect to the maximum norm, i.e. there is a constant $c > 0$ with $\|x_k\|_\infty \leq c$ for all $k \in \mathbb{N}$. Then the sequence $\{x_k\}_{k=1}^{\infty}$ possesses a subsequence, which converges with respect to the maximum norm.*

Proof: Since $|x_k^{(i)}| \leq\| x_k\|_\infty$ for $i = 1, \ldots, n$, all the component sequences are bounded. Therefore by the Bolzano-Weierstraß Theorem for sequences in $\mathbb{R}$, the sequence $\big\{x_k^{(1)}\big\}_{k=1}^{\infty}$ possesses a convergent subsequence $\big\{x_{k(j)}^{(1)}\big\}_{j=1}^{\infty}$. Then $\big\{x_{k(j)}^{(2)}\big\}_{j=1}^{\infty}$ is a bounded subsequence of $\big\{x_k^{(2)}\big\}_{k=1}^{\infty}$, hence it has a convergent subsequence $\big\{x_{k(j(\ell))}^{(2)}\big\}_{\ell=1}^{\infty}$. Also $\big\{x_{k(j(\ell))}^{(1)}\big\}_{\ell=1}^{\infty}$ converges as a subsequence of the converging sequence $\big\{x_{k(j)}^{(1)}\big\}_{j=1}^{\infty}$. Thus, for the subsequence $\big\{x_{k(j(\ell))}\big\}_{\ell=1}^{\infty}$ of $\big\{x_k\big\}_{k=1}^{\infty}$ the first two component sequences converge. We proceed in the same way and obtain after n steps a subsequence $\big\{x_{k_s}\big\}_{s=1}^{\infty}$ of $\big\{x_k\big\}_{k=1}^{\infty}$, for which all component sequences converge. By the preceding lemma this implies that $\big\{x_{k_s}\big\}_{s=1}^{\infty}$ converges with respect to the maximum norm. ■

Theorem 3.4 *Let $\|\cdot\|$ and $|\!|\!|\cdot|\!|\!|$ be norms on $\mathbb{R}^n$. Then there exist constants $a, b > 0$ such that for all $x \in \mathbb{R}^n$*

$$a\|x\| \leq |\!|\!|x|\!|\!| \leq b\|x\|\,.$$

Proof: Obviously it suffices to show that for any norm $\|\cdot\|$ on $\mathbb{R}^n$ there exist constants $a, b > 0$ such that for the maximum norm $\|\cdot\|_\infty$

$$\|x\| \leq a\|x\|_\infty\,, \quad \|x\|_\infty \leq b\|x\|\,,$$

for all $x \in \mathbb{R}^n$. The first one of these estimates is obtained as follows:

$$\begin{aligned} \|x\| &= |x_1e_1 + x_2e_2 + \ldots + x_ne_n| \\ &\leq \|x_1e_1\| + \ldots + \|x_ne_n\| = |x_1|\,\|e_1\| + \ldots + |x_n|\,\|e_n\| \\ &\leq \big(\|e_1\| + \ldots + \|e_n\|\big)\,\|x\|_\infty = a\|x\|_\infty\,, \end{aligned}$$

where $a = \|e_1\| + \ldots + \|e_n\|$.

The second one of these estimates is proved by contradiction: Suppose that such a constant $b > 0$ would not exist. Then for every $k \in \mathbb{N}$ we can choose an element $x_k \in \mathbb{R}^n$ such that

$$\|x_k\|_\infty > k\,\|x_k\|\,.$$

Set $y_k = \frac{x_k}{\|x_k\|_\infty}$. The sequence $\{y_k\}_{k=1}^\infty$ satisfies

$$\|y_k\| = \Big\|\frac{x_k}{\|x_k\|_\infty}\Big\| = \frac{1}{\|x_k\|_\infty}\,\|x_k\| < \frac{1}{k}$$

and

$$\|y_k\|_\infty = \Big\|\frac{x_k}{\|x_k\|_\infty}\Big\|_\infty = \frac{1}{\|x_k\|_\infty}\,\|x_k\|_\infty = 1\,.$$

Therefore by Theorem 3.3 the sequence $\{y_k\}_{k=1}^\infty$ has a subsequence $\big\{y_{k_j}\big\}_{j=1}^\infty$, which converges with respect to the maximum norm. For brevity we set $z_j = y_{k_j}$. Let z be the limit of $\{z_j\}_{j=1}^\infty$. Then

$$\lim_{j\to\infty} \|z_j - z\|_\infty = 0\,,$$

hence, since $\|z_j\|_\infty = \|y_{k_j}\|_\infty = 1$,

$$1 = \lim_{j\to\infty} \|z_j\|_\infty = \lim_{j\to\infty} \|z_j - z + z\|_\infty \leq \|z\|_\infty + \lim_{j\to\infty} \|z_j - z\|_\infty = \|z\|_\infty\,,$$

whence $z \neq 0$. On the other hand, $\|z_j\| = \|y_{k_j}\| < \frac{1}{k_j} \leq \frac{1}{j}$ together with the estimate $\|x\| \leq a\|x\|_\infty$ proved above implies

$$\begin{aligned} \|z\| &= \|z - z_j + z_j\| = \lim_{j\to\infty} \|z - z_j + z_j\| \\ &\leq \lim_{j\to\infty} \|z - z_j\| + \lim_{j\to\infty} \|z_j\| \leq a \lim_{j\to\infty} \|z - z_j\|_\infty + \lim_{j\to\infty} \frac{1}{j} = 0\,, \end{aligned}$$

hence $z = 0$. This is a contradiction, hence a constant b must exist such that $\|x\|_\infty \leq b\|x\|$ for all $x \in \mathbb{R}$. ∎

Definition 3.5 *Let $\|\cdot\|$ and $|\!|\!|\cdot|\!|\!|$ be norms on a vector space V. If constant $a, b > 0$ exist such that*

$$a\|v\| \leq |\!|\!|v|\!|\!| \leq b\|v\|$$

for all $v \in V$, then these norms are said to be equivalent.

The above theorem thus shows that on $\mathbb{R}^n$ all norms are equivalent. From the definition of convergence it immediately follows that a sequence converging with respect to a norm also converges with respect to an equivalent norm. Therefore **on $\mathbb{R}^n$** the definition of convergence does not depend on the norm.

Moreover, since all norms on $\mathbb{R}^n$ are equivalent to the maximum norm, from Lemma 3.2 and Theorem 3.3 we immediately obtain

Lemma 3.6 *A sequence in $\mathbb{R}^n$ converges to $a \in \mathbb{R}^n$ if and only if the component sequences all converge to the components of a.*

Theorem 3.7 (Theorem of Bolzano-Weierstraß for $\mathbb{R}^n$) *Every bounded sequence in $\mathbb{R}^n$ possesses a convergent subsequence.*

Lemma 3.8 (Cauchy convergence criterion) *Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$. A sequence $\{x_k\}_{k=1}^\infty$ in $\mathbb{R}^n$ converges if and only if to every $\varepsilon > 0$ there is a $k_0 \in \mathbb{N}$ such that for all $k, \ell \geq k_0$*

$$\|x_k - x_\ell\| < \varepsilon\,.$$

Proof: $\{x_k\}_{k=1}^\infty$ is a Cauchy sequence on $\mathbb{R}^n$ if and only if every component sequence $\{x_k^{(i)}\}_{k=1}^\infty$ for $i = 1, \ldots, n$ is a Cauchy sequence in $\mathbb{R}$. For, there are constants, $a, b > 0$ such that for all $i = 1, \ldots, n$

$$\begin{aligned} a|x_k^{(i)} - x_\ell^{(i)}| &\leq a\|x_k - x_\ell\|_\infty \leq \|x_k - x_\ell\| \\ &\leq b\|x_k - x_\ell\|_\infty \leq b\big(|x_k^{(1)} - x_\ell^{(1)}| + \ldots + |x_k^{(n)} - x_\ell^{(n)}|\big)\,. \end{aligned}$$

The statement of the lemma follows from this observation, from the fact that the component sequences converge in $\mathbb{R}$ if and only if they are Cauchy sequences, and from the fact that a sequence converges in $\mathbb{R}^n$ if and only if all the component sequences converge. ∎

Infinite series: Let $\{x_k\}_{k=1}^{\infty}$ be a sequence in $\mathbb{R}^n$. By the infinite series $\sum_{k=1}^{\infty} x_k$ one means the sequence $\{s_\ell\}_{\ell=1}^{\infty}$ of partial sums $s_\ell = \sum_{k=1}^{\ell} x_k$. If $\{s_\ell\}_{\ell=1}^{\infty}$ converges, then $s = \lim_{\ell\to\infty} s_\ell$ is called the sum of the series $\sum_{k=1}^{\infty} x_k$. One writes

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

A series is said to converge absolutely, if

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

converges, where $\|\cdot\|$ is a norm on $\mathbb{R}^n$. From

$$\|\sum_{k=\ell}^{m} x_k\| \leq \sum_{k=\ell}^{m} \|x_k\|$$

and from the Cauchy convergence criterion it follows that an absolutely convergent series converges. The converse is in general not true.

A series converges absolutely if and only if every component series converges absolutely. This implies that every rearrangement of an absolutely convergent series in $\mathbb{R}^n$ converges to the same sum, since this holds for the component series.

3.2 Topology of $\mathbb{R}^n$

In the following we denote by $\|\cdot\|$ a norm on $\mathbb{R}^n$.

Definition 3.9 *Let $a \in \mathbb{R}^n$ and $\varepsilon > 0$. The set*

$$U_\varepsilon(a) = \{x \in \mathbb{R}^n \mid \|x - a\| < \varepsilon\}$$

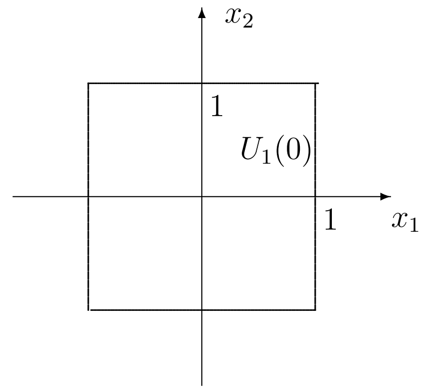
is called open ε-neighborhood of a with respect the the norm $\|\cdot\|$, or ball with center a and radius ε.

A subset U of $\mathbb{R}^n$ is called neighborhood of a if U contains an ε-neighborhood of a.

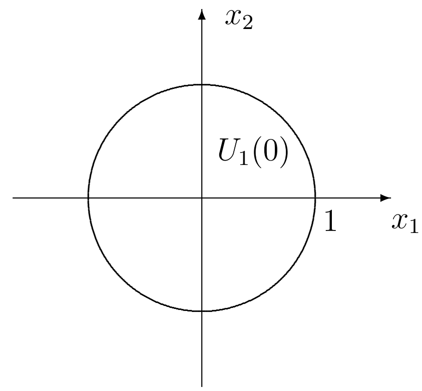
The set $U_1(0) = \{x \in \mathbb{R}^n \mid \|x\| < 1\}$ is called open unit ball with respect to $\|\cdot\|$.

In $\mathbb{R}^2$ the unit ball can be pictured for the different norms:

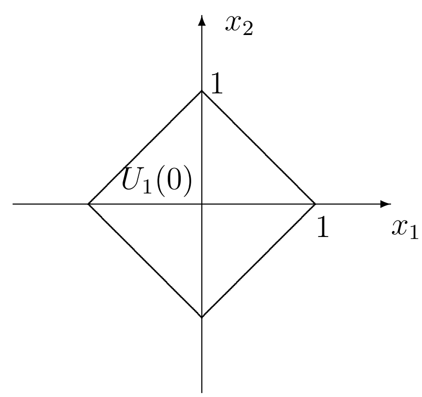
Maximum norm $\|\cdot\|_\infty$:



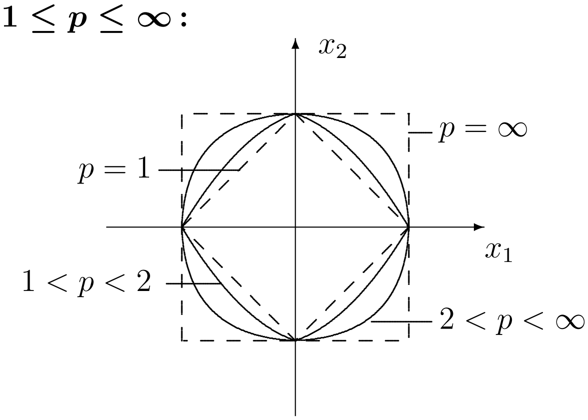
Euclidean norm $|\cdot|$:



1-norm $\|\cdot\|_1$:



***p*-norm $\|\cdot\|_p$ with $1 \le p \le \infty$:**



Whereas the ε-neighborhoods of a point a differ for different norms, the notion of a neighborhood is independent of the norm. For, let $\|\cdot\|$ and $\boldsymbol{|}\cdot\boldsymbol{|}$ be norms on $\mathbb{R}^n$. We show that every ε-neighborhood with respect to $\|\cdot\|$ of $a \in \mathbb{R}^n$ contains a δ-neighborhood with respect to $\boldsymbol{|}\cdot\boldsymbol{|}$.

To this end let

$$\begin{aligned} U_\varepsilon(a) &= \{x \in \mathbb{R}^n \mid \|x-a\| < \varepsilon\}, \\ V_\varepsilon(a) &= \{x \in \mathbb{R}^n \mid \boldsymbol{|}x-a\boldsymbol{|} < \varepsilon\}. \end{aligned}$$

Since all norms on $\mathbb{R}^n$ are equivalent, there is a constant $c > 0$ such that

$$c\|x-a\| \le \boldsymbol{|}x-a\boldsymbol{|}$$

for all $x \in \mathbb{R}^n$. Therefore, if $x \in V_{c\varepsilon}(a)$ then $\boldsymbol{|}x-a\boldsymbol{|} < c\varepsilon$, which implies $\|x-y\| \le \frac{1}{c}\boldsymbol{|}x-a\boldsymbol{|} < \varepsilon$, and this means $x \in U_\varepsilon(a)$. Consequently, with $\delta = c\varepsilon$,

$$V_\delta(a) \subseteq U_\varepsilon(a).$$

This result implies that if U is a neighborhood of a with respect to $\|\cdot\|$, then it contains a neighborhood $U_\varepsilon(a)$, and then also the neighborhood $V_{c\varepsilon}(a)$, hence U is a neighborhood of a with respect to the norm $\boldsymbol{|}\cdot\boldsymbol{|}$ as well. Consequently, a neighborhood of a with respect to one norm is a neighborhood of a with respect to every other norm on $\mathbb{R}^n$. Therefore the definition of a neighborhood is independent of the norm.

Definition 3.10 *Let M be a subset of $\mathbb{R}^n$. A point $x \in \mathbb{R}^n$ is called interior point of M, if M contains an ε-neighborhood of x, hence if M is a neighborhood of x.*

$x \in \mathbb{R}^n$ is called accumulation point of M, if every neighborhood of x contains a point of M different from x.

$x \in \mathbb{R}$ is called boundary point of M, if every neighborhood of x contains a point of M and a point of the complement $\mathbb{R}^n \setminus M$.

M is called open, if it only consists of its interior points. M is called closed, if it contains all its accumulation points.

The following statements are proved exactly as in $\mathbb{R}^1$:

The complement of an open set is closed, the complement of a closed set is open. The union of an arbitrary system of open sets is open, the intersection of finitely many open sets is open. The intersection of an arbitrary system of closed sets is closed, the union of finitely many closed sets is closed.

A subset M of $\mathbb{R}^n$ is called bounded, if there exists a positive constant C such that

$$\|x\| \le C$$

for all $x \in M$. The number

$$\operatorname{diam}(M) := \sup_{y,x \in M} \|y - x\|$$

is called diameter of the bounded set M.

Theorem 3.11 *Let $\{A_k\}_{k=1}^{\infty}$ be a sequence of bounded, closed, nonempty subsets A_k of $\mathbb{R}^n$ with $A_{k+1} \subseteq A_k$ and with*

$$\lim_{k\to\infty} \operatorname{diam}(A_k) = 0 .$$

Then there is $x \in \mathbb{R}^n$ such that

$$\bigcap_{k=1}^{\infty} A_k = \{x\} .$$

Proof: For every $k \in \mathbb{N}$ choose $x_k \in A_k$. Then the sequence $\{x_k\}_{k=1}^{\infty}$ is a Cauchy sequence, sind $\lim_{k\to\infty} \operatorname{diam}(A_k) = 0$ implies that to $\varepsilon > 0$ there is k_0 such that diam $A_k < \varepsilon$ for all $k \ge k_0$. Thus, $A_{k+\ell} \subseteq A_k$ implies for all $k \ge k_0$ that

$$\|x_{k+\ell} - x_k\| \le \operatorname{diam}\,(A_k) < \varepsilon .$$

The limit x of $\{x_k\}_{k=1}^{\infty}$ satisfies $x \in \bigcap_{k=1}^{\infty} A_k$. For, if $j \in \mathbb{N}$ would exist with $x \notin A_j$, then, since $\mathbb{R}^n \backslash A_j$ is open, a neighborhood $U_\varepsilon(x)$ could be chosen such that $U_\varepsilon(x) \cap A_j = \emptyset$. Thus, $U_\varepsilon(x) \cap A_{j+\ell} = \emptyset$, since $A_{j+\ell} \subseteq A_j$, which implies $\|x - x_{j+\ell}\| \ge \varepsilon$ for all ℓ. This contradicts the property that x is the limit of $\{x_k\}_{k=1}^{\infty}$, and therefore x belongs to the intersection of all sets A_k.

This intersection does not contain any other point. For if $y \in \bigcap_{k=1}^{\infty} A_k$, then $\|x-y\| \le \operatorname{diam}\,(A_k)$ for all k, whence

$$\|x - y\| = \lim_{k\to\infty} \|x - y\| \le \lim_{k\to\infty} \operatorname{diam}\,(A_k) = 0 .$$

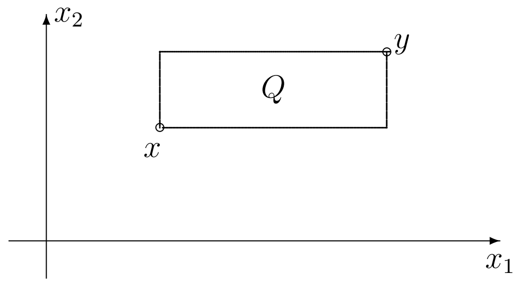
Consequently $y = x$, which proves $\bigcap_{k=1}^{\infty} A_k = \{x\}$. ■

Definition 3.12 *Let $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$. The set*

$$Q = \{z = (z_1, \ldots, z_n) \in \mathbb{R}^n \mid x_i \le z_i \le y_i\,,\ i = 1, \ldots, n\}$$

is called closed interval in $\mathbb{R}^n$. If $y_1 - x_1 = y_2 - x_2 = \ldots = y_n - x_n = a \ge 0$, then this set is called a cube with edge length a.

Let M be a subset of $\mathbb{R}^n$. A system $\mathcal{U}$ of open subsets of $\mathbb{R}^n$ such that $M \subseteq \bigcup_{U\in\mathcal{U}} U$ is called an open covering of M.



Theorem 3.13 *Let $M \subseteq \mathbb{R}^n$. The following three statements are equivalent:*

(i) *M is bounded and closed.*

(ii) *Let $\mathcal{U}$ be an open covering of M. Then there are finitely many $U_1, \ldots, U_m \in \mathcal{U}$ such that $M \subseteq \bigcup_{i=1}^m U_i$.*

(iii) *Every infinite subset of M possesses an accumulation point in M.*

Proof: (i) $\Rightarrow$ (ii): Assume that M is bounded and closed, but that there is an open covering $\mathcal{U}$ of M for which (ii) is not satisfied. As a bounded set M is contained in a sufficiently large closed cube W. Subdivide this cube into 2^n closed cubes with edge length halved. By assumption, there is at least one of the smaller cubes, denoted by W_1, such that $W_1 \cap M$ cannot be covered by finitely many sets from $\mathcal{U}$. Now subdivide W_1 and select W_2 analogously. The sequence $\{M \cap W_k\}_{k=1}^\infty$ of closed sets thus constructed, has the following properties:

1.) $M \cap W \supseteq M \cap W_1 \supseteq M \cap W_2 \supseteq \ldots$

2.) $\lim\limits_{k\to\infty} \text{diam}\,(M \cap W_k) = 0$

3.) $M \cap W_k$ cannot be covered by finitely many sets from $\mathcal{U}$.

3.) implies $M \cap W_k \neq \emptyset$. Therefore, by 1.) and 2.) the sequence $\{M \cap W_k\}_{k=1}^\infty$ satisfies the assumptions of Theorem 3.11, hence there is $x \in \mathbb{R}^n$ such that

$$x \in \bigcap_{k=1}^\infty (M \cap W_k)\,.$$

Since $x \in M$, there is $U \in \mathcal{U}$ with $x \in U$. The set U is open, and therefore contains an ε-neighborhood of x, and then also a δ-neighborhood of x with respect to the maximum norm. Because $\lim_{k\to\infty}$ diam $(W_k) \to 0$ and because $x \in W_k$ for all k, this δ-neighborhood contains the cubes W_k for all sufficiently large k. Hence U contains $M \cap W_k$ for all sufficiently large k. Thus, $M \cap W_k$ can be covered by one set from $\mathcal{U}$, contradicting 3.). We thus conclude that if (i) holds, then also (ii) must be satisfied.

(ii) $\Rightarrow$ (iii): Assume that (ii) holds and let A be a subset of M which does not have accumulation points in M. Then no one of the points of M is an accumulation point of A, consequently to every $x \in M$ there is an open neighborhood, which does not contain a point from A different from x. The system of all these neighborhoods is an open covering of M, hence finitely many of these neighborhoods cover M. Since everyone of these neighborhoods contains at most one point from A, we conclude that A must be finite. An infinite subset of M must thus have an accumulation point in M.

(iii) $\Rightarrow$ (i). Assume that (iii) is satisfied. If M would not be bounded, to every $k \in \mathbb{N}$ there would exist $x_k \in M$ such that

$$\|x_k\| \geq k\,.$$

Let A denote the set of these points. A is an infinite subset of M, but it does not have an accumulation point. For, to an accumulation point y of A there must exist infinitely many $x \in A$ satisfying $\|x - y\| < 1$, which implies

$$\|x\| = \|x - y + y\| \leq \|x - y\| + \|y\| < 1 + \|y\|\,.$$

This is not possible, since A only contains finitely many points with norm smaller than $1 + \|y\|$ Thus, the infinite subset A of M does not have an accumulation point. Since this contradicts (iii), M must be bounded.

Let x be an accumulation point of M. For every $k \in \mathbb{N}$ we can select $x_k \in M$ with $0 < \|x_k - x\| < \frac{1}{k}$. The sequence $\{x_k\}_{k=1}^{\infty}$ converges to x, hence x is the only accumulation point of this sequence. Therefore x must belong to M by (iii), thus M contains all its accumulation points, whence M is closed. ■

Definition 3.14 *A subset of $\mathbb{R}^n$ is called compact, if it has one (and therefore all) of the three properties stated in the preceding theorem.*

Theorem 3.15 *A subset M of $\mathbb{R}^n$ is compact, if and only if every sequence in M possesses a convergent subsequence with limit contained in M.*

This theorem is proved as in $\mathbb{R}^1$ (cf. Theorem 6.15 in the classroom notes to Analysis I.)

A set M with the property that every sequence in M has a subsequence converging in M, is called **sequentially compact**. Therefore, in $\mathbb{R}^n$ a set is compact if and only if it is sequentially compact. Finally, just as in $\mathbb{R}^1$, from the Theorem of Bolzano-Weierstraß for sequences (Theorem 3.7) we obtain

Theorem 3.16 (Theorem of Bolzano-Weierstraß for sets in $\mathbb{R}^n$) *Every bounded infinite subset of $\mathbb{R}^n$ has an accumulation point.*

The **proof** is the same as the proof of Theorem 6.11 in the classroom notes to Analysis I.

3.3 Continuous mappings from $\mathbb{R}^n$ to $\mathbb{R}^m$

Let D be a subset of $\mathbb{R}^n$. We consider mappings $f : D \to \mathbb{R}^m$. Such mappings are called *functions of n variables*.

For $x \in D$ let $f_1(x), \ldots, f_m(x)$ denote the components of the element $f(x) \in \mathbb{R}^m$. This defines mappings

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

Conversely, let m mappings $f_1, \ldots, f_m : D \to \mathbb{R}$ be given. Then a mapping

$$f : D \to \mathbb{R}^m$$

is defined by

$$f(x) := \big(f_1(x), \ldots, f_m(x)\big)\,.$$

Thus, every mapping $f : D \to \mathbb{R}^m$ with $D \subseteq \mathbb{R}^n$ is specified by m equations

$$\begin{aligned} y_1 &= f_1(x_1, \ldots, x_n) \\ &\vdots \\ y_m &= f_m(x_1, \ldots, x_n)\,. \end{aligned}$$

Examples
1.) Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a mapping, which satisfies for all $x, y \in \mathbb{R}^n$ and all $c \in \mathbb{R}$

$$f(x + y) = f(x) + f(y)$$
$$f(cx) = cf(x)$$

Then f is called a linear mapping. The study of linear mappings from $\mathbb{R}^n$ to $\mathbb{R}^m$ is the topic of linear algebra. From linear algebra one knows that $f : \mathbb{R}^n \to \mathbb{R}^m$ is a linear mapping if and only if there exists a matrix

$$A = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \\ a_{m1} & \ldots & a_{mn} \end{pmatrix}$$

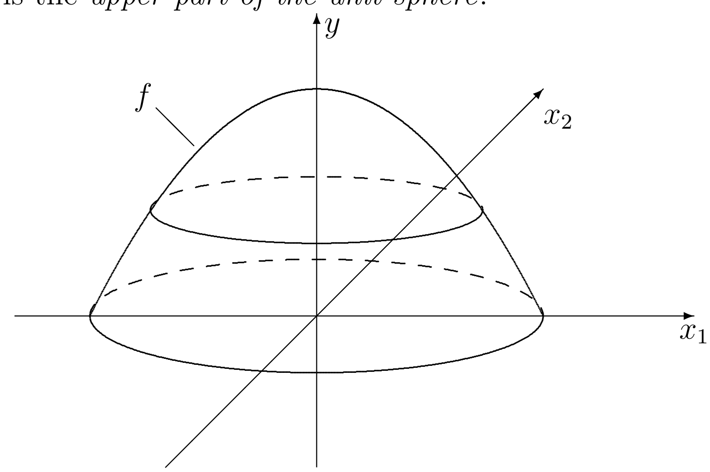
with $a_{ij} \in \mathbb{R}$ such that

$$f(x) = Ax = \begin{pmatrix} a_{11}x_1 & + \ldots + & a_{1n}x_n \\ \vdots & & \\ a_{m1}x_1 & + \ldots + & a_{mn}x_n \end{pmatrix}.$$

2.) Let $n = 2$, $m = 1$ and $D = \left\{ x \in \mathbb{R}^2 \;\middle|\; |x| < 1 \right\}$. A mapping $f : D \to \mathbb{R}$ is defined by

$$f(x) = f(x_1, x_2) = \sqrt{1 - x_1^2 - x_2^2}\,.$$

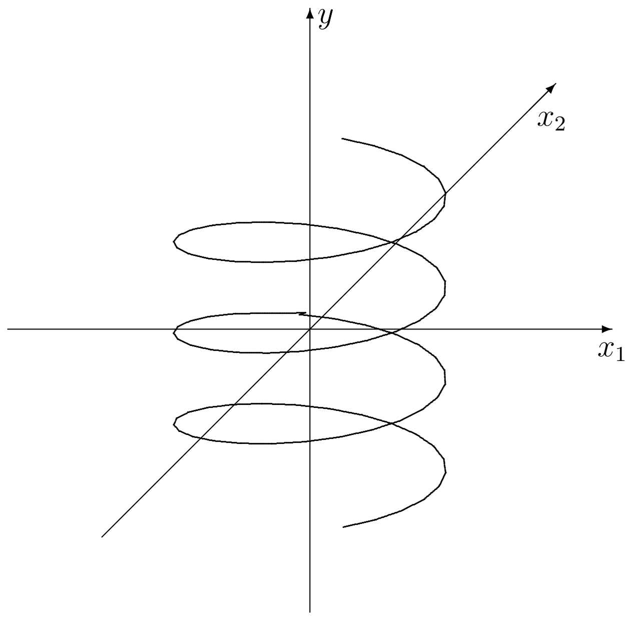
The graph of a mapping from a subset D of $\mathbb{R}^2$ to $\mathbb{R}$ is a surface in $\mathbb{R}^3$. In the present example graph f is the *upper part of the unit sphere*:



3.) Every mapping $f : \mathbb{R} \to \mathbb{R}^m$ is called a path in $\mathbb{R}^m$. For example, let for $t \in \mathbb{R}$

$$f(t) = \begin{pmatrix} f_1(t) \\ f_2(t) \\ f_3(t) \end{pmatrix} = \begin{pmatrix} \cos t \\ \sin t \\ t \end{pmatrix}$$

The range of f is a *helix*.



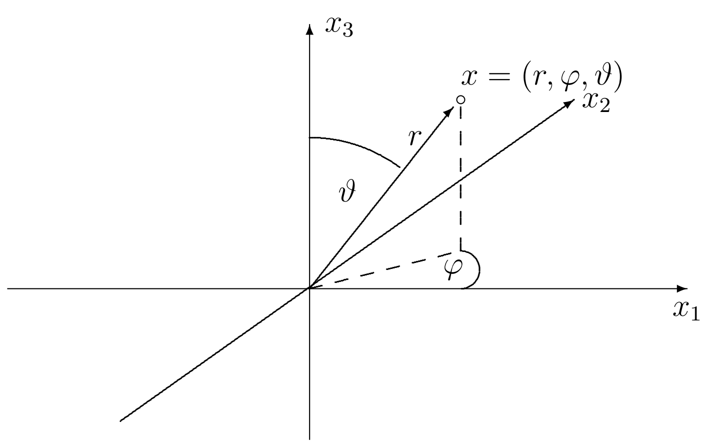
4.) Polar coordinates: Let

$$D = \left\{(r,\varphi,\psi) \in \mathbb{R}^3 \mid 0 < r,\ 0 \le \varphi < 2\pi,\ 0 < \vartheta < \pi\right\} \subseteq \mathbb{R}^3\,,$$

and let $f : D \to \mathbb{R}^3$,

$$f(r,\varphi,\psi) = \begin{pmatrix} r\cos\varphi\sin\vartheta \\ r\sin\varphi\sin\vartheta \\ r\cos\vartheta \end{pmatrix}.$$

The range of this mapping is $\mathbb{R}^3$ without the x_3-axis:



Definition 3.17 *Let D be a subset of $\mathbb{R}^n$. A mapping $f : D \to \mathbb{R}^m$ is said to be continuous at $a \in D$, if to every neighborhood V of $f(a)$ there is a neighborhood U of a such that $f(U \cap D) \subseteq V$.*

Since every neighborhood of a point contains an ε-neighborhood of this point, irrespective of the norm we use to define ε-neighborhoods, we obtain an equivalent formulation if in this definition we replace V by $V_\varepsilon\big(f(a)\big)$ and U by $U_\delta(a)$. Thus, using the definition of ε-neighborhoods, we immediately get the following

Theorem 3.18 *Let $D \subseteq \mathbb{R}^n$. A mapping $f : D \to \mathbb{R}^m$ is continuous at $a \in D$ if and only if to every $\varepsilon > 0$ there is $\delta > 0$ such that*

$$\|f(x) - f(a)\| < \varepsilon$$

for all $x \in D$ with $\|x - a\| < \delta$.

Note that in this theorem we denoted the norms in $\mathbb{R}^n$ and $\mathbb{R}^m$ with the same symbol $\|\cdot\|$.

Almost all results for continuous real functions transfer to continuous functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ with the same proofs. An example is the following

Theorem 3.19 *Let $D \subseteq \mathbb{R}^n$. A function $f : D \to \mathbb{R}^m$ is continuous at $a \in D$, if and only if for every sequence $\{x_k\}_{k=1}^\infty$ with $x_k \in D$ and $\lim_{k\to\infty} x_k = a$*

$$\lim_{k\to\infty} f(x_k) = f(a)$$

holds.

Proof: Cf. the proof of Theorem 6.21 of the classroom notes to Analysis I.

Definition 3.20 *Let $f : D \to \mathbb{R}^m$ and let $a \in \mathbb{R}^n$ be an accumulation point of D. Let $b \in \mathbb{R}^m$. One says that f has the limit b at a and writes*

$$\lim_{x\to a} f(x) = b$$

if to every $\varepsilon > 0$ there is $\delta > 0$ such that

$$\|f(x) - b\| < \varepsilon$$

for all $x \in D\setminus\{a\}$ with $\|x - y\| < \delta$.

Theorem 3.21 *Let $f : D \to \mathbb{R}^m$ and let a be an accumulation point. $\lim_{x\to a} f(x) = b$ holds if and only if for every sequence $\{x_k\}_{k=1}^\infty$ with $x_k \in D\setminus\{a\}$ and $\lim_{k\to\infty} x_k = a$*

$$\lim_{k\to\infty} f(x_k) = b$$

holds.

Proof: Cf. the proof of Theorem 6.39 of the classroom notes to Analysis I.

Example: Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by

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

This function is continuous at every point $(x,y) \in \mathbb{R}^2$ with $(x,y) \neq 0$, but it is not continuous at $(x,y) = 0$. For

$$f(x,0) = f(0,y) = 0\,,$$

whence f vanishes identically on the lines $y = 0$ and $x = 0$. However, on the diagonal $x = y$

$$f(x,y) = f(x,x) = \frac{2x^2}{2x^2} = 1\,.$$

For the two sequences $\{z_k\}_{k=1}^{\infty}$ with $z_k = (\frac{1}{k}, 0)$ and $\{\tilde{z}_k\}_{k=1}^{\infty}$ with $\tilde{z}_\ell = \left(\frac{1}{k}, \frac{1}{k}\right)$ we therefore have $\lim_{k\to\infty} z_k = \lim_{k\to\infty} \tilde{z}_k = 0$, but

$$\lim_{k\to\infty} f(z_k) = 0 = f(0) \neq 1 = \lim_{k\to\infty} f(\tilde{z}_k)\,.$$

Therefore, by Theorem 3.19 f is not continuous at $(0,0)$, and by Theorem 3.21 does not have a limit at $(0,0)$. Hence f cannot be made into a function continuous at $(0,0)$ by modifying the value $f(0,0)$.

Observe however, that the function

$$x \mapsto f(x,y) : \mathbb{R} \to \mathbb{R}$$

is continuous for every $y \in \mathbb{R}$, and

$$y \mapsto f(x,y) : \mathbb{R} \to \mathbb{R}$$

is continuous for every $x \in \mathbb{R}$. Therefore f is continuous in every variable, but as a function $f : \mathbb{R}^2 \to \mathbb{R}$ it is not continuous at $(0,0)$.

Theorem 3.22 *Let $D \subseteq \mathbb{R}^n$ and let $f : D \to \mathbb{R}^m$. The function f is continuous at a point $a \in D$, if and only if all the component functions $f_1, \ldots, f_m : D \to \mathbb{R}$ are continuous at a.*

Proof: f is continuous at a, if and only if for every sequence $\{x_k\}_{k=1}^{\infty}$ with $x_k \in D$ and $\lim_{k\to\infty} x_k = a$ the sequence $\{f(x_k)\}_{k=1}^{\infty}$ converges to $f(a)$. This holds if and only if every component sequence $\{f_i(x_k)\}_{k=1}^{\infty}$ converges to $f_i(a)$ for $i = 1, \ldots, m$, and this is equivalent to the continuity of f_i at a for $i = 1, \ldots, m$. ■

Definition 3.23 *Let $D \subseteq \mathbb{R}^n$. A function $f : D \to \mathbb{R}^m$ is said to be continuous if it is continuous at every point of D.*

Definition 3.24 *Let D be a subset of $\mathbb{R}^n$. A subset D' of D is said to be relatively open with respect to D, if there exists an open subset O of $\mathbb{R}^n$ such that $D' = O \cap D$.*

Thus, for example, every subset D of $\mathbb{R}^n$ is relatively open with respect to itself, since $D = D \cap \mathbb{R}^n$ and $\mathbb{R}^n$ is open.

Lemma 3.25 *A subset D' of D is relatively open with respect to D, if and only if for every $x \in D$ there is a neighborhood U of x such that $U \cap D \subseteq D'$.*

Proof: If D' is relatively open, there is an open subset O of $\mathbb{R}^n$ such that $D' = O' \cap D$. For every $x \in D'$ the set O is the sought neighborhood.

Conversely, assume that to every $x \in D'$ there is a neighborhood $U(x)$ with $U(x) \cap D \subseteq D'$. Since every neighborhood contains an open neighborhood, we can assume that $U(x)$ is open. Then

$$D' \subseteq D \cap \bigcup_{x \in D'} U(x) = \bigcup_{x \in D'} \big(D \cap U(x)\big) \subseteq D',$$

whence $D' = D \cap O$ with the open set $O = \bigcup_{x \in D'} U(x)$. Consequently D' is relatively open with respect to D. ■

Theorem 3.26 *Let $D \subseteq \mathbb{R}^n$. A function $f : D \to \mathbb{R}^m$ is continuous, if and only if for each open set O of $\mathbb{R}^m$ the inverse image $f^{-1}(O)$ is relatively open with respect to D.*

Proof: Let f be continuous and $x \in f^{-1}(O)$. Then $f(x)$ belongs to the open set O, whence O is a neighborhood of $f(x)$. Therefore, by definition of continuity, there is a neighborhood V of x such that $f(V \cap D) \subseteq O$, which implies $V \cap D \subseteq f^{-1}(O)$. Thus, $f^{-1}(O)$ is relatively open with respect to D.

Assume conversely that the inverse image of every open set is relatively open in D. Let $x \in D$ and let U be an open neighborhood of $f(x)$. Then $f^{-1}(U)$ is relatively open, whence there is an open set $O \subseteq \mathbb{R}^n$ such that $f^{-1}(U) = O \cap D$. This implies $x \in f^{-1}(U) \subseteq O$, whence O is a neighborhood of x. For this neighborhood of x we have

$$f(O \cap D) = f\big(f^{-1}(U)\big) \subseteq U,$$

hence f is continuous. ■

The following theorems and the corollary are proved as the corresponding theorems in $\mathbb{R}$.

Theorem 3.27 (i) *Let $D \subseteq \mathbb{R}^n$ and let $f : D \to \mathbb{R}^m$, $g : D \to \mathbb{R}^m$ be continuous. Then also the mappings $f + g : D \to \mathbb{R}^m$ and $cf : D \to \mathbb{R}^m$ are continuous for every $c \in \mathbb{R}$.*

(ii) *Let $f : D \to \mathbb{R}$ and $g : D \to \mathbb{R}$ be continuous. Then also $f \cdot g : D \to \mathbb{R}$ and*

$$\frac{f}{g} : \big\{x \in D \mid g(x) \neq 0\big\} \to \mathbb{R}$$

are continuous.

(iii) *Let $f : D \to \mathbb{R}^m$ and $\varphi : D \to \mathbb{R}$ be continuous. Then also φf is continuous.*

Theorem 3.28 *Let $D_1 \subseteq \mathbb{R}^n$ and $D_2 \subseteq \mathbb{R}^p$. Assume that $f : D_1 \to D_2$ and $g : D_2 \to \mathbb{R}^m$ are continuous. Then $g \circ f : D_1 \to \mathbb{R}^m$ is continuous.*

This theorem is proved just as Theorem 6.25 in the classroom notes of Analysis I.

Definition 3.29 *Let D be a subset of $\mathbb{R}^n$. A mapping $f : D \to \mathbb{R}^m$ is said to be uniformly continuous, if to every $\varepsilon > 0$ there is $\delta > 0$ such that*

$$\|f(x) - f(y)\| < \varepsilon$$

for all $x, y \in D$ satisfying $\|x - y\| < \delta$.

Theorem 3.30 *Let $D \subseteq \mathbb{R}^n$ be compact and $f : D \to \mathbb{R}^m$ be continuous. Then f is uniformly continuous and $f(D) \subseteq \mathbb{R}^m$ is compact.*

Corollary 3.31 *Let $D \subseteq \mathbb{R}^n$ be compact and $f : D \to \mathbb{R}$ be continuous. Then f attains the maximum and minimum.*

Definition 3.32 *A subset M of $\mathbb{R}^n$ is said to be connected, if it has the following property: Let U_1, U_2 be relatively open subsets of M such that $U_1 \cap U_2 = \emptyset$ and $U_1 \cup U_2 = M$. Then $M = U_1$ and $U_2 = \emptyset$ or $M = U_2$ and $U_1 = \emptyset$.*

Example Every interval in $\mathbb{R}$ is connected.

Theorem 3.33 *Let D be a connected subset of $\mathbb{R}^n$ and $f : D \to \mathbb{R}^m$ be continuous. Then $f(D)$ is a connected subset of $\mathbb{R}^m$.*

Proof: Let U_1 and U_2 be relatively open subsets of $f(D)$ with $U_1 \cap U_2 = \emptyset$ and $U_1 \cup U_2 = f(D)$. With suitable open subsets O_1, O_2 of $\mathbb{R}^m$ we thus have $U_1 = O_1 \cap f(D)$ and $U_2 = O_2 \cap f(D)$, whence the continuity of f implies that $f^{-1}(U_1) = f^{-1}(O_1)$ and $f^{-1}(U_2) = f^{-1}(O_2)$ are relatively open subsets of D satisfying $f^{-1}(U_1) \cap f^{-1}(U_2) = \emptyset$ and $f^{-1}(U_1) \cup f^{-1}(U_2) = D$. Thus, since D is connected, it follows that $f^{-1}(U_1) = \emptyset$ or $f^{-1}(U_2) = \emptyset$, hence $U_1 = \emptyset$ or $U_2 = \emptyset$. Consequently, $f(D)$ is connected. ■

Definition 3.34 *Let $[a,b]$ be an interval in $\mathbb{R}$ and let $\gamma : [a,b] \to \mathbb{R}^m$ be continuous. Then γ is called a path in $\mathbb{R}^m$.*

Definition 3.35 *A subset M of $\mathbb{R}^n$ is said to be pathwise connected, if any two points in M can be connected by a path in M, i.e. if to $x, y \in M$ there is an interval $[a,b]$ and a continuous mapping $\gamma : [a,b] \to M$ such that $\gamma(a) = x$ and $\gamma(b) = y$.*

$\gamma(a)$ is called starting point, $\gamma(b)$ end point of γ.

Theorem 3.36 *Let $D \subseteq \mathbb{R}^n$ be pathwise connected and let $f : D \to \mathbb{R}^m$ be continuous. Then $f(D)$ is pathwise connected.*

Proof: Let $u, v \in f(D)$ and let $x \in f^{-1}(u)$ and $y \in f^{-1}(v)$. Then there is a path γ, which connects x with y in D. Thus, $f \circ \gamma$ is a path which connects u with v in $f(D)$. ■

Theorem 3.37 *Let $M \subseteq \mathbb{R}^m$ be pathwise connected. Then M is connected.*

Proof: Suppose that M is not connected. Then there are relatively open subsets $U_1 \neq \emptyset$ and $U_2 \neq \emptyset$ such that $U_1 \cap U_2 = \emptyset$ and $U_1 \cup U_2 = M$. Select $x \in U_1$ and $y \in U_2$ and let $\gamma : [a,b] \to M$ be a path connecting x with y. Since M is not connected, it follows that the set $\gamma([a,b])$ is not connected. To see this, set

$$\begin{aligned} V_1 &= \gamma([a,b]) \cap U_1 , \\ V_2 &= \gamma([a,b]) \cap U_2 . \end{aligned}$$

Then V_1 and V_2 are relatively open subsets of $\gamma([a,b])$ satisfying $V_1 \cap V_2 = \emptyset$ and $V_1 \cap V_2 = \gamma([a,b])$. Therefore, since $x \in V_1$, $y \in V_2$ implies $V_1 \neq \emptyset$, $V_2 \neq \emptyset$, it follows that $\gamma([a,b])$ is not connected.

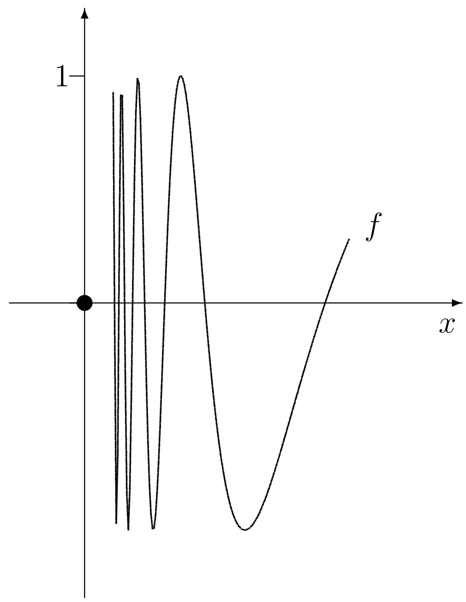
On the other hand, since $[a,b]$ is connected and since γ is continuous, the set $\gamma([a,b])$ must be connected. Our assumption has thus led to a contradiction, hence M is connected.

■

Example. Consider the mapping $f : [0, \infty) \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} \sin \frac{1}{x} , & x > 0 \\ 0 , & x = 0 . \end{cases}$$

Then $M = \text{graph}(f) = \big\{ \big(x, f(x)\big) \mid x \in [0, \infty) \big\}$ is a subset of $\mathbb{R}^2$, which is connected, but not pathwise connected.



To prove that M is not pathwise connected, assume the contrary. Then, since $(0,0) \in M$ and $(x_0, 1) \in M$ with $x_0 = 1/\frac{\pi}{2}$, a path $\gamma : [a,b] \to M$ exists such that $\gamma(a) = (0,0)$ and $\gamma(b) = (x_0, 1)$. The component functions γ_1 and γ_2 are continuous. Since to every $x \geq 0$ a unique $y \in \mathbb{R}$ exists such that $(x,y) \in M$, namely $y = f(x)$, these component functions satisfy for all $c \in [a,b]$

$$\gamma(c) = \big(\gamma_1(c)\,,\, \gamma_2(c)\big) = \Big(\gamma_1(c)\,,\, f\big(\gamma_1(c)\big)\Big)\,,$$

hence

$$\gamma_2 = f \circ \gamma_1\,.$$

However, this is a contradiction, since $f \circ \gamma_1$ is not continuous.

To see this, set

$$x_n = \frac{1}{\frac{\pi}{2} + 2n\pi}\,.$$

Then $\{x_n\}_{n=1}^{\infty}$ is a null sequence with

$$\gamma_1(a) = 0 < x_n < x_0 = \gamma_1(b)\,.$$

Therefore the intermediate value theorem implies that a sequence $\{c_n\}_{n=1}^{\infty}$ exists with $a \leq c_n \leq b$ such that

$$\gamma_1(c_n) = x_n\,.$$

The bounded sequence $\{c_n\}_{n=1}^{\infty}$ has a convergent subsequence $\{c_{n_j}\}_{j=1}^{\infty}$ with limit

$$c = \lim_{j\to\infty} c_{n_j} \in [a,b]\,.$$

From the continuity of γ_1 it follows that

$$\gamma_1(c) = \lim_{j\to\infty} \gamma_1(c_{n_j}) = \lim_{j\to\infty} x_{n_j} = \lim_{n\to\infty} x_n = 0\,,$$

hence

$$(f\circ\gamma_1)(c) = f\big(\gamma_1(c)\big) = f(0) = 0\,,$$

but

$$\begin{aligned}\lim_{j\to\infty}(f\circ\gamma_1)(c_{n_j}) &= \lim_{j\to\infty} f\big(\gamma_1(c_{n_j})\big) = \lim_{j\to\infty} f(x_{n_j})\\ &= \lim_{j\to\infty}\sin\Big(\frac{\pi}{2}+2n_j\pi\Big) = \lim_{j\to\infty} 1 = 1 \neq (f\circ\gamma_1)(c)\,,\end{aligned}$$

which proves that $f\circ\gamma_1$ is not continuous at c.

To prove that M is connected, assume the contrary. Then there are relatively open subsets U_1, U_2 of M satisfying $U_1 \neq \emptyset$, $U_2 \neq \emptyset$, $U_1\cap U_2 = \emptyset$, and $U_1\cup U_2 = M$. the set

$$M' = \big\{\big(x, f(x)\big) \mid x > 0\big\} \subseteq M$$

is connected as the image of the connected set $(0,\infty)$ under the continuous map

$$x \mapsto \big(x, f(x)\big) : (0,\infty) \to \mathbb{R}^2.$$

Consequently, $U_1\cap M' = \emptyset$ or $U_2\cap M' = \emptyset$. Without restriction of equality we assume that $U_1\cap M' = \emptyset$. Then $U_2 = M'$ and $U_1 = \big\{(0,0)\big\}$. However, this is a contradiction, since $\big\{(0,0)\big\}$ is not relatively open with respect to M. Otherwise an open set $O\subseteq\mathbb{R}^2$ would exist such that $\big\{(0,0)\big\} = M\cap O$, hence $(0,0)\in O$, and therefore O would contain an ε-neighborhood of $(0,0)$. Since $\sin\big(\frac{1}{x}\big)$ has infinitely many zeros in every neighborhood of $x=0$, the ε-neighborhood of $(0,0)$ would contain besides $(0,0)$ infinitely many points of M on the positive real axis, hence $M\cap O \neq \big\{(0,0)\big\}$. Consequently, M is connected. ■

This example shows that the statement of the preceding theorem cannot be inverted.

Theorem 3.38 *Let D be a compact subset of $\mathbb{R}^n$ and $f: D\to\mathbb{R}^m$ be continuous and injective. Then the inverse $f^{-1}: f(D)\to D$ is continuous.*

The proof of this theorem is obtained by a slight modification of the proof of Theorem 6.28 in the classroom notes of Analysis I.

Definition 3.39 *let $D\subseteq\mathbb{R}^n$ and $W\subseteq\mathbb{R}^m$. A mapping $f: D\to W$ is called homeomorphism, if f is bijective, continuous and has a continuous inverse.*

3.4 Uniform convergence, the normed spaces of continuous and linear mappings

Definition 3.40 *Let D be a nonempty set and let $f : D \to \mathbb{R}^m$ be bounded. Then*

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

is called the supremum norm of f. Here $\|\cdot\|$ denotes a norm on $\mathbb{R}^m$.

As for real valued mappings it follows that $\|\cdot\|_\infty$ is a norm on the vector space $B(D,\mathbb{R}^m)$ of bounded mappings from D to $\mathbb{R}^m$, cf. the proof of Theorem 1.8. Therefore, with this norm $B(D,\mathbb{R}^m)$ is a normed space. Of course, the supremum norm on $B(D,\mathbb{R}^m)$ depends on the norm on $\mathbb{R}^m$ used to define the supremum norm. However, from the equivalence of all norms on $\mathbb{R}^m$ it immediately follows that the supremum norms on $B(D,\mathbb{R}^m)$ obtained from different norms on $\mathbb{R}^m$ are equivalent. Therefore the following definition does not depend on the supremum norm chosen:

Definition 3.41 *Let D be a nonempty set and let $\{f_k\}_{k=1}^\infty$ be a sequence of functions $f_k \in B(D,\mathbb{R}^m)$. The sequence $\{f_k\}_{k=1}^\infty$ is said to converge uniformly, if $f \in B(D,\mathbb{R}^m)$ exists such that*

$$\lim_{k\to\infty} \|f_k - f\|_\infty = 0.$$

Theorem 3.42 *A sequence $\{f_k\}_{k=1}^\infty$ with $f_k \in B(D,\mathbb{R}^m)$ converges uniformly if and only if to every $\varepsilon > 0$ there is $k_0 \in \mathbb{N}$ such that for all $k, \ell \geq k_0$*

$$\|f_k - f_\ell\|_\infty < \varepsilon.$$

(Cauchy convergence criterion.)

This theorem is proved as Corollary 1.5.

Definition 3.43 *A normed vector space with the property that every Cauchy sequence converges, is called a complete normed space or a Banach space (Stefan Banach, 1892 – 1945).*

Corollary 3.44 *The space $B(D,\mathbb{R}^m)$ with the supremum norm is a Banach space.*

Theorem 3.45 *Let $D \subseteq \mathbb{R}^n$ and let $\{f_k\}_{k=1}^\infty$ be a sequence of continuous functions $f_k \in B(D,\mathbb{R}^m)$, which converges uniformly to $f \in B(D,\mathbb{R}^m)$. Then f is continuous.*

This theorem is proved as Corollary 1.5. For a subset D of $\mathbb{R}^n$ we denote by $C(D, \mathbb{R}^m)$ the set of all continuous functions from D to $\mathbb{R}^m$. This is a linear subspace of the vector space of all functions from D to $\mathbb{R}^m$. Also the set of all bounded continuous functions $C(D, \mathbb{R}^m) \cap B(D, \mathbb{R}^m)$ is a vector space. As a subspace of $B(D, \mathbb{R}^m)$ it is a normed space with the supremum norm. From the preceding theorem we obtain the following important result:

Corollary 3.46 *For $D \subseteq \mathbb{R}^n$ the normed space $C(D, \mathbb{R}^m) \cap B(D, \mathbb{R}^m)$ is complete, hence it is a Banach space.*

Proof: Let $\{f_k\}_{k=1}^{\infty}$ be a Cauchy sequence in $C(D, \mathbb{R}^m) \cap B(D, \mathbb{R}^m)$. Then this sequence converges with respect to the supremum norm to a function $f \in B(D, \mathbb{R}^m)$. The preceding theorem implies that $f \in C(D, \mathbb{R}^m)$, since $f_k \in C(D, \mathbb{R}^m)$ for all k. Thus, $f \in C(D, \mathbb{R}^m) \cap B(D, \mathbb{R}^m)$, and $\{f_k\}_{k=1}^{\infty}$ converges with respect to the supremum norm to f. Therefore every Cauchy sequence converges in $C(D, \mathbb{R}^m) \cap B(D, \mathbb{R}^m)$, hence this space is complete.

■

By $L(\mathbb{R}^n, \mathbb{R}^m)$ we denote the set of all linear mappings $f : \mathbb{R}^n \to \mathbb{R}^m$. Since for linear mappings f, g and for a real number c the mappings $f + g$ and cf are linear, $L(\mathbb{R}^n, \mathbb{R}^m)$ is a vector space.

Theorem 3.47 *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be linear. Then f is continuous. If f differs from zero, then f is unbounded.*

Proof: To f there exists a unique $m \times n$–Matrix $(a_{ij})_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}}$ such that

$$\begin{aligned} f_1(x_1, \ldots, x_n) &= a_{11}x_1 + \ldots + a_{1n}x_n \\ &\vdots \\ f_m(x_1, \ldots, x_n &= a_{m1}x_1 + \ldots + a_{mn}x_n. \end{aligned}$$

Since everyone of the expressions on the right depends continuously on $x = (x_1, \ldots, x_n)$, it follows that all component functions of f are continuous, hence f is continuous.

If f differs from 0, there is $x \in \mathbb{R}^n$ with $f(x) \neq 0$. From the linearity we then obtain for $\lambda \in \mathbb{R}$

$$|f(\lambda x)| = |\lambda f(x)| = |\lambda| \, |f(x)|,$$

which can be made larger than any constant by choosing $|\lambda|$ sufficiently large. Hence f is not bounded. ■

We want to define a norm on the linear space $L(\mathbb{R}^n, \mathbb{R}^m)$. It is not possible to use the supremum norm, since every linear mapping $f \neq 0$ is unbounded, hence, the supremum of the set

$$\{\|f(x)\| \,\big|\, x \in \mathbb{R}^n\}$$

does not exist. Instead, on $L(\mathbb{R}^n, \mathbb{R}^m)$ a norm can be defined as follows: Let $B = \{x \in \mathbb{R}^n \,\big|\, \|x\| \le 1\}$ be the closed unit ball in $\mathbb{R}^n$. The set B is bounded and closed, hence compact. Thus, since $f \in L(\mathbb{R}^n, \mathbb{R}^m)$ is continuous and since every continuous map is bounded on compact sets, the supremum

$$\|f\| := \sup_{x \in B} \|f(x)\|$$

exists. The following lemma shows that the mapping $\|\cdot\| : L(\mathbb{R}^n, \mathbb{R}^m) \to [0, \infty)$ thus defined is a norm:

Lemma 3.48 *Let $f, g : \mathbb{R}^n \to \mathbb{R}^m$ be linear, let $c \in \mathbb{R}$ and $x \in \mathbb{R}^n$. Then*

(i) $f = 0 \Longleftrightarrow \|f\| = 0$,

(ii) $\|cf\| = |c|\,\|f\|$

(iii) $\|f + g\| \le \|f\| + \|g\|$

(iv) $\|f(x)\| \le \|f\|\,\|x\|$.

Proof: We first prove (iv). For $x = 0$ the linearity of f implies $f(x) = 0$, whence $\|f(x)\| = 0 \le \|f\|\,\|x\|$. For $x \neq 0$ we have $\|\frac{x}{\|x\|}\| = 1$, hence $\frac{x}{\|x\|} \in B$. Therefore the linearity of f yields

$$\begin{aligned} \|f(x)\| &= \big\|f\big(\|x\|\frac{x}{\|x\|}\big)\big\| = \big\|\,\|x\| f\big(\frac{x}{\|x\|}\big)\big\| \\ &= \|x\|\,\big\|f\big(\frac{x}{\|x\|}\big)\big\| \le \|x\| \sup_{y \in B} \|f(y)\| = \|x\|\,\|f\|. \end{aligned}$$

To prove (i), let $f = 0$. Then $\|f\| = \sup_{x \in B} \|f(x)\| = 0$. On the other hand, if $\|f\| = 0$, we conclude from (iv) for all $x \in \mathbb{R}^n$ that

$$\|f(x)\| \le \|f\|\,\|x\| = 0,$$

hence $f(x) = 0$, and therefore $f = 0$. (ii) and (iii) are proved just as the corresponding properties for the supremum norm in Theorem 1.8. ∎

Definition 3.49 *For $f \in L(\mathbb{R}^n, \mathbb{R}^m)$*

$$\|f\| = \sup_{\|x\| \le 1} \|f(x)\|$$

is called the operator norm of f.

With this norm $L(\mathbb{R}^n, \mathbb{R}^m)$ is a normed vector space. To every linear mapping $A : \mathbb{R}^n \to \mathbb{R}^m$ there is associated a unique $m \times n$–matrix, which we also denote by A, such that $A(x) = Ax$. Here Ax denotes the matrix multiplication. The question arises, whether the operator norm $\|A\|$ can be computed from the elements of the matrix A. To give a partial answer, we define for $A = (a_{ij})$,

$$\|A\|_\infty = \max_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}} |a_{ij}|.$$

Theorem 3.50 *There exist constants $c, C > 0$ such that for every $A \in L(\mathbb{R}^n, \mathbb{R}^m)$*

$$c\|A\|_\infty \le \|A\| \le C\|A\|_\infty.$$

Proof: Note first that there exist constants $c_1, \ldots, c_3 > 0$ such that for all $x \in \mathbb{R}^m$ and $y \in \mathbb{R}^n$

$$c_1\|x\|_\infty \le \|x\| \le c_2\|x\|_\infty, \quad \|y\|_1 \le c_3\|y\|,$$

because all norms on $\mathbb{R}^n$ are equivalent. For $1 \le j \le n$ let e_j denote the j–th unit vector of $\mathbb{R}^n$ and let

$$a^{(j)} = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix} \in \mathbb{R}^m$$

be the j–th column vector of the matrix $A = (a_{ij})$. Then for $x \in \mathbb{R}^n$

$$\|A(x)\| = \|Ax\| = \|\sum_{j=1}^{n} a^{(j)} x_j\|. \qquad (*)$$

Setting $x = e_j$ in this equation yields

$$\|a^{(j)}\| = \|A(e_j)\| \le \|A\|\,\|e_j\|,$$

hence, with $c_4 = \max\limits_{1 \le j \le n} \|e_j\|$,

$$\|A_\infty\| = \max_{1 \le j \le n} \|a^{(j)}\|_\infty \le \frac{1}{c_1} \max_{1 \le j \le n} \|a^{(j)}\| \le \frac{c_4}{c_1}\|A\|.$$

On the other hand, for $\|x\| \leq 1$ equation $(*)$ yields

$$\begin{aligned} \|A(x)\| &\leq \sum_{j=1}^{n} \|a^{(j)}\| \, |x_j| \leq c_2 \|A\|_\infty \sum_{j=1}^{n} |x_j| \\ &= c_2 \|A\|_\infty \|x\|_1 \leq c_2 \|A\|_\infty c_3 \|x\| \leq c_2 c_3 \|A\|_\infty \end{aligned}$$

whence

$$\|A\| = \sup_{\|x\| \leq 1} \|A(x)\| \leq c_2 c_3 \|A\|_\infty .$$

■

4 Differentiable mappings on $\mathbb{R}^n$

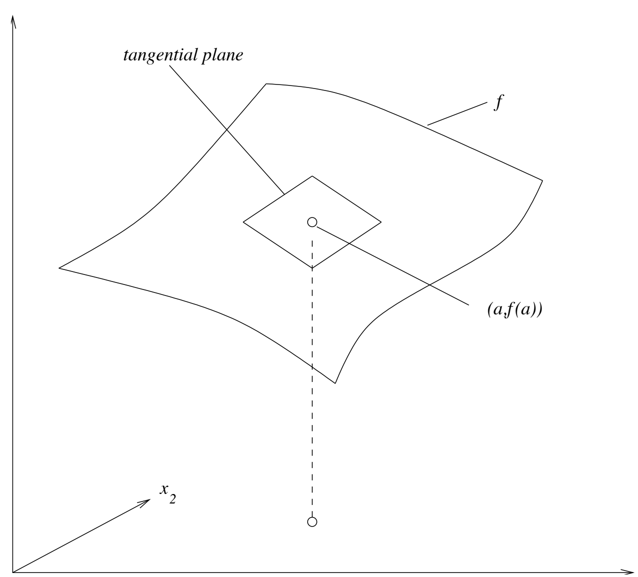
4.1 Definition of the derivative

The derivative of a real function f at a satisfies the equation

$$f(x) = f(a) + f'(a)(x-a) + r(x)(x-a)\,,$$

where the function r is continuous at a and satisfies $r(a) = 0$. Since $x \mapsto f'(a)x$ is a linear map from $\mathbb{R}$ to $\mathbb{R}$, the interpretation of this equation is that under all affine maps $x \mapsto f(a) + T(x-a)$, where $T : \mathbb{R} \to \mathbb{R}$ is linear, the one obtained by choosing $T(x) = f'(a)x$ is the best approximation of the function f in a neighborhood of a.

Viewed in this way, the notion of the derivative can be generalized immediately to mappings $f : D \to \mathbb{R}^m$ with $D \subseteq \mathbb{R}^n$. Thus, the derivative of f at $a \in D$ is the linear map $T : \mathbb{R}^n \to \mathbb{R}^m$ such that under all affine functions the mapping $x \mapsto f(a) + T(x-a)$ approximates f best in a neighborhood of a.



For a mapping $f : \mathbb{R}^2 \to \mathbb{R}$ this means that the linear mapping $T : \mathbb{R}^2 \to \mathbb{R}$, the derivative of f at a, must be chosen such that the graph of the mapping $x \mapsto f(a) + T(x-a)$ is equal to the tangential plane of the graph of f at $\big(a, f(a)\big)$.

This idea leads to the following rigorous definition of a differentiable function:

Definition 4.1 *Let U be an open subset of $\mathbb{R}^n$. A function $f : U \to \mathbb{R}^m$ is said to be differentiable at the point $a \in U$, if there is a linear mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ and a function $r : U \to \mathbb{R}^m$, which is continuous at a and satisfies $r(a) = 0$, such that for all $x \in U$*

$$f(x) = f(a) + T(x-a) + r(x)\,\|x-a\|\,.$$

Therefore to verify that f is differentiable at $a \in D$ a linear mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ must be found such that the function r defined by

$$r(x) := \frac{f(x) - f(a) - T(x-a)}{\|x - a\|}$$

satisfies

$$\lim_{r \to a} r(x) = 0\,.$$

Later we show how T can be found. However, there is at most one such T:

Lemma 4.2 *The linear mapping T is uniquely determined.*

Proof: Let $T_1, T_2 : \mathbb{R}^n \to \mathbb{R}^m$ be linear mappings and $r_1, r_2 : U \to \mathbb{R}^m$ be functions with $\lim_{x \to a} r_1(x) = \lim_{x \to a} r_2(x) = 0$, such that for $x \in U$

$$\begin{aligned} f(x) &= f(a) + T_1(x-a) + r_1(x)\,\|x-a\| \\ f(x) &= f(a) + T_2(x-a) + r_2(x)\,\|x-a\|\,. \end{aligned}$$

Then

$$(T_1 - T_2)(x-a) = \big(r_2(x) - r_1(x)\big)\,\|x-a\|\,.$$

Let $h \in \mathbb{R}^n$. Then, $x = a + th \in U$ for all sufficiently small $t > 0$ since U is open, whence

$$(T_1 - T_2)(th) = t(T_1 - T_2)(h) = \big(r_2(a+th) - r_1(a+th)\big)\,\|th\|\,,$$

thus

$$(T_1 - T_2)(h) = \lim_{t \to 0}(T_1 - T_2)(h) = \lim_{t \to 0}\big(r_2(a+th) - r_1(a+th)\big)\,\|h\| = 0\,.$$

This implies $T_1 = T_2$, since $h \in \mathbb{R}^n$ was chosen arbitrarily. ∎

Definition 4.3 *Let $U \subseteq \mathbb{R}^n$ be open and let $f : U \to \mathbb{R}^m$ be differentiable at $a \in U$. Then the unique linear mapping $T : \mathbb{R}^n \to \mathbb{R}^m$, for which a function $r : U \to \mathbb{R}^m$ satisfying $\lim_{x \to a} r(x) = 0$ exists, such that*

$$f(x) = f(a) + T(x-a) + r(x)\,\|x-a\|$$

holds for all $x \in U$, is called derivative of f at a. This linear mapping is denoted by $f'(a) = T$.

Mostly we drop the brackets around the argument and write $T(h) = Th = f'(a)h$.

For a real valued function f the derivative is a linear mapping $f'(a) : \mathbb{R}^n \to \mathbb{R}$. Such linear mappings are also called linear forms. In this case $f'(a)$ can be represented by a $1\times n$-matrix, and we normally identify $f'(a)$ with this matrix. The transpose $[f'(a)]^T$ of this $1\times n$-matrix is a $n\times 1$-matrix, a column vector. For this transpose one uses the notation

$$\operatorname{grad} f(a) = [f'(a)]^T .$$

$\operatorname{grad} f(a)$ is called the gradient of f at a. With the scalar product on $\mathbb{R}^n$ the gradient can be used to represent the derivative of f: For $h \in \mathbb{R}^n$ we have

$$f'(a)h = \big(\operatorname{grad} f(a)\big) \cdot h .$$

If $h \in \mathbb{R}^n$ is a unit vector and if t runs through $\mathbb{R}$, then the point th moves along the straight line through the origin with direction h. A differentiable real function is defined by

$$t \mapsto \big(\operatorname{grad} f(a)\big) \cdot th = t\big(\operatorname{grad} f(a) \cdot h\big) .$$

The derivative is $\operatorname{grad} f(a) \cdot h$, and this derivative attains the maximum value

$$\operatorname{grad} f(a) \cdot h = |\operatorname{grad} f(a)|$$

if h has the direction of $\operatorname{grad} f(a)$. Since $f(a) + \operatorname{grad} f(a) \cdot (th) = f(a) + f'(a)th$ approximates the value $f(a+th)$, it follows that the vector $\operatorname{grad} f(a)$ points into the direction of steepest ascent of the function f at a, and the length of $\operatorname{grad} f(a)$ determines the slope of f in this direction.

Lemma 4.4 *Let $U \subseteq \mathbb{R}^n$ be an open set. The function $f : U \to \mathbb{R}^m$ is differentiable at $a \in U$, if and only if all component functions $f_1, \ldots, f_m : U \to \mathbb{R}$ are differentiable in a. The derivatives satisfy*

$$(f_j)'(a) = \big(f'(a)\big)_j , \quad j = 1, \ldots, m .$$

Proof: If the derivatives $f'(a)$ exist, then the components satisfy

$$\lim_{h\to 0} \frac{f_j(a+h) - f_j(a) - \big(f'(a)\big)_j h}{\|h\|} = 0 .$$

Since $\big(f'(a)\big)_j : \mathbb{R}^n \to \mathbb{R}$ is linear, it follows that f_j is differentiable at a with derivative $(f_j)'(a) = \big(f'(a)\big)_j$. Conversely, if the derivative $(f_j)'(a)$ of f_j exists at a for all $j =$

$1, \dots, m$, then a linear mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is defined by

$$Th = \begin{pmatrix} (f_1)'(a)h \\ \vdots \\ (f_m)'(a)h \end{pmatrix},$$

for which

$$\lim_{h \to 0} \frac{f(a+h) - f(a) - Th}{\|h\|} = 0 .$$

Thus, f is differentiable at a with derivative $f'(a) = T$. ■

4.2 Directional derivatives and partial derivatives

Let $U \subseteq \mathbb{R}^n$ be an open set, let $a \in U$ and let $f : U \to \mathbb{R}^m$. Let $v \in \mathbb{R}^n$ be a given vector. Since U is open, there is $\delta > 0$ such that $a + tv \in U$ for all $t \in \mathbb{R}$ with $|t| < \delta$; hence $f(a + tv)$ is defined for all such t. If t runs through the interval $(-\delta, \delta)$, then $a + tv$ runs through a line segment passing through a, which has the direction of the vector v.

Definition 4.5 *We call the limit*

$$D_v f(a) = \lim_{t \to 0} \frac{f(a + tv) - f(a)}{t}$$

derivative of f at a in the direction of the vector v, if this limit exists.

It is possible that the directional derivative $D_v f(a)$ exists, even if f is not differentiable at a. Also, it can happen that the derivative of f at a exists in the direction of some vectors, and does not exist in the direction of other vectors. In any case, the directional derivative contains useful information about the function f. However, if f is differentiable at a, then all directional derivatives of f exist at a:

Lemma 4.6 *Let $U \subseteq \mathbb{R}^n$ be open, let $a \in U$ and let $f : U \to \mathbb{R}^m$ be differentiable at a. Then the directional derivative $D_v f(a)$ exists for every $v \in \mathbb{R}^n$ and satisfies*

$$D_v f(a) = f'(a)v .$$

Proof: Set $x = a + tv$ with $t \in \mathbb{R}$, $t \neq 0$. Then by definition of the derivative $f'(a)$

$$f(a + tv) = f(a) + f'(a)(tv) + r(tv + a)\, |t|\, \|v\| ,$$

hence

$$\frac{f(a + tv) - f(a)}{t} = f'(a)v + r(tv + a)\, \frac{|t|}{t}\, \|v\| .$$

Since $\frac{|t|}{t} = \pm 1$ and since $\lim_{t\to 0} r(tv + a) = r(a) = 0$, it follows that $\lim_{t\to 0} r(tv + a)\frac{|t|}{t}\,\|v\| = 0$, hence

$$\lim_{t\to 0} \frac{f(a+tv) - f(a)}{t} = f'(a)v\,.$$

■

This result can be used to compute $f'(a)$: If $v_1, \ldots, v_n$ is a basis of $\mathbb{R}^n$, then every vector $v \in \mathbb{R}$ can be represented as a linear combination $v = \sum_{i=1}^n \alpha_i v_i$ of the basis vectors with uniquely determined numbers $\alpha_i \in \mathbb{R}$. The linearity of $f'(a)$ thus yields

$$f'(a)v = f'(a)\Big(\sum_{i=1}^n \alpha_i v_i\Big) = \sum_{i=1}^n \alpha_i f'(a)v_i = \sum_{i=1}^n \alpha_i D_{v_i} f(a)\,.$$

Therefore $f'(a)$ is known if the directional derivatives $D_{v_i} f(a)$ for the basis vectors are known. It suggests itself to use the standard basis $e_1, \ldots, e_n$. The directional derivative $D_{e_i} f(a)$ is called *i-th partial derivative of f at a*. For the i-th partial derivative one uses the notations

$$D_i f,\ \frac{\partial f}{\partial x_i}\,,\ f_{x_i},\ f'_{x_i},\ f_{|i}\,.$$

For $i = 1, \ldots, n$ and $j = 1, \ldots, m$ we have

$$\begin{aligned}
\frac{\partial f}{\partial x_i}(a) &= \lim_{t\to 0} \frac{f(a+te_i) - f(a)}{t} = \lim_{x_i \to a_i} \frac{f(a_1, \ldots, x_i, \ldots, a_n) - f(a_1, \ldots, a_i, \ldots, a_n)}{x_i - a_i}\,, \\
\frac{\partial f_j}{\partial x_i}(a) &= \lim_{x_i \to a_i} \frac{f_j(a_1, \ldots, x_i, \ldots, a_n) - f_j(a_1, \ldots, a_i, \ldots, a_n)}{x_i - a_i}\,.
\end{aligned}$$

Consequently, to compute partial derivatives the differential calculus for functions of one real variable suffices.

To construct $f'(a)$ from the partial derivatives one proceeds as follows: If $f'(a)$ exists, then all the partial derivatives $D_i f(a) = \frac{\partial f}{\partial x_i}(a)$ exist. For arbitrary $h \in \mathbb{R}^n$ we have $h = \sum_{i=1}^n h_i e_i$, where $h_i \in \mathbb{R}$ are the components of h, hence

$$f'(a)h = f'(a)\Big(\sum_{i=1}^n h_i e_i\Big) = \sum_{i=1}^n \big(f'(a)e_i\big)h_i = \sum_{i=1}^n D_i f(a) h_i\,,$$

or, in matrix notation,

$$f'(a)h = \begin{pmatrix} \big(f'(a)h\big)_1 \\ \vdots \\ \big(f'(a)h\big)_m \end{pmatrix} = \begin{pmatrix} D_1 f_1(a) & \ldots & D_n f_1(a) \\ \vdots & & \\ D_1 f_m(a) & \ldots & D_n f_m(a) \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}$$

Thus,

$$f'(a) = \begin{pmatrix} D_1 f_1(a) & \dots & D_n f_1(a) \\ \vdots & & \\ D_1 f_m(a) & \dots & D_n f_m(a) \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \dots & \frac{\partial f_1}{\partial x_n}(a) \\ \vdots & & \\ \frac{\partial f_m}{\partial x_1}(a) & \dots & \frac{\partial f_m}{\partial x_n}(a) \end{pmatrix}$$

is the representation of $f'(a)$ as $m \times n$-matrix belonging to the standard bases $e_1, \dots, e_n$ of $\mathbb{R}^n$ and $e_1, \dots, e_m$ of $\mathbb{R}^m$. This matrix is called Jacobi-matrix of f at a. (Carl Gustav Jacob Jacobi 1804–1851).

It is possible that all partial derivatives exist at a without f being differentiable at a. Then the Jacobi-matrix can be formed, but it does not represent the derivative $f'(a)$, which does not exist.

Therefore, to check whether f is differentiable at a, one first verifies that all partial derivatives exist at a. This is a necessary condition for the existence of $f'(a)$. Then one forms the Jacobi-matrix

$$T = \Big(\frac{\partial f_i}{\partial x_j}(a)\Big)_{\substack{i=1,\dots,m \\ j=1,\dots,n}},$$

and tests whether for this matrix

$$\lim_{h \to 0} \frac{f(a+h) - f(a) - Th}{\|h\|} = 0$$

holds. If this holds, then f is differentiable at a with derivative $f'(a) = T$.

Examples

1.) Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$f(x_1, x_2) = \begin{pmatrix} f_1(x_2, x_2) \\ f_2(x_1, x_2) \end{pmatrix} = \begin{pmatrix} x_1^2 - x_2^2 \\ 2x_1 x_2 \end{pmatrix}.$$

At $a = (a_1, a_2) \in \mathbb{R}^2$ the Jacobi-matrix is

$$T = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \frac{\partial f_1}{\partial x_2}(a) \\ \frac{\partial f_2}{\partial x_1}(a) & \frac{\partial f_2}{\partial x_2}(a) \end{pmatrix} = \begin{pmatrix} 2a_1 & -2a_2 \\ 2a_2 & 2a_1 \end{pmatrix}.$$

To test the differentiability of f at a, set for $h = (h_1, h_2) \in \mathbb{R}^2$ and $i = 1, 2$

$$r_i(h) = \frac{f_i(a+h) - f_i(a) - T_i(h)}{\|h\|},$$

hence

$$\begin{aligned} r_1(h) &= \frac{(a_1+h_1)^2 - (a_2+h_2)^2 - a_1^2 + a_2^2 - 2a_1h_1 + 2a_2h_2}{\|h\|} = \frac{h_1^2 - h_2^2}{\|h\|}, \\ r_2(h) &= \frac{2(a_1+h_1)(a_2+h_2) - 2a_1a_2 - 2a_2h_1 - 2a_1h_2}{\|h\|} = \frac{2h_1h_2}{\|h\|}. \end{aligned}$$

Using the maximum norm $\|\cdot\| = \|\cdot\|_\infty$, we obtain

$$\begin{aligned} |r_1(h)| &\leq 2\|h\|_\infty \\ |r_2(h)| &\leq 2\|h\|_\infty , \end{aligned}$$

thus

$$\lim_{h\to 0} \|r(h)\|_\infty = \lim_{h\to 0} \|\big(r_1(h), r_2(h)\big)\|_\infty \leq \lim_{h\to 0} 2\|h\|_\infty = 0 .$$

Therefore f is differentiable at a. Since a was arbitrary, f is everywhere differentiable, i.e. f is differentiable.

2.) Let the affine map $f : \mathbb{R}^n \to \mathbb{R}^m$ be defined by

$$f(x) = Ax + c ,$$

where $c \in \mathbb{R}^m$ and $A : \mathbb{R}^n \to \mathbb{R}^m$ is linear. Then f is differentiable with derivative $f'(a) = A$ for all $a \in \mathbb{R}^n$. For,

$$\frac{f(a+h) - f(a) - Ah}{\|h\|} = \frac{A(a+h) + c - Aa - c - Ah}{\|h\|} = 0 .$$

3.) Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$f(x_1, x_2) = \begin{cases} 0 , & \text{for } (x_1, x_2) = 0 , \\ \dfrac{|x_1|\, x_2}{\sqrt{x_1^2 + x_2^2}} , & \text{for } (x_1, x_2) \neq 0 . \end{cases}$$

This function is not differentiable at $a = 0$, but it has all the directional derivatives at 0. To see that all directional derivatives exist, let $v = (v_1, v_2)$ be a vector from $\mathbb{R}^2$ different from zero. Then

$$D_v f(0) = \lim_{t\to 0} \frac{f(tv) - f(0)}{t} = \lim_{t\to 0} \frac{t|t|\, |v_1| v_2}{t|t|\, \sqrt{v_1^2 + v_2^2}} = \frac{|v_1|\, v_2}{\sqrt{v_1^2 + v_2^2}} .$$

To see that f is not differentiable at 0, note that the partial derivatives satisfy

$$\frac{\partial f}{\partial x_1}(0) = 0 , \quad \frac{\partial f}{\partial x_2}(0) = 0 .$$

Therefore, if f would be differentiable at 0, the derivative had to be

$$f'(0) = \Big(\frac{\partial f}{\partial x_1}(0) \quad \frac{\partial f}{\partial x_2}(0) \Big) = (0 \quad 0) .$$

Consequently, all directional derivatives would satisfy

$$D_v f(0) = f'(0) v = 0 .$$

Yet, the preceding calculation yields for the derivative in the direction of the diagonal vector $v = (1,1)$ that

$$D_v\, f(0) = \frac{1}{\sqrt{2}}\,.$$

Therefore $f'(0)$ cannot exist.

We note that $|f(x_1,x_2)| = \frac{|x_1\,x_2|}{|x|} \leq |x|$, which implies that f is continuous at 0.

4.3 Elementary properties of differentiable mappings

In the preceding example f was not differentiable at 0, but had all the directional derivatives and was continuous at 0. Here is an example of a function $f : \mathbb{R}^2 \to \mathbb{R}$, which has all the directional derivatives at 0, yet is not continuous at 0: f is defined by

$$f(x_1,x_2) = \begin{cases} 0, & \text{for } (x_1,x_2) = 0 \\ \dfrac{x_1 x_2^2}{x_1^2 + x_2^6}\,, & \text{for } (x_1,x_2) \neq 0\,. \end{cases}$$

To see that all directional derivatives exist at 0, let $v = (v_1, v_2) \in \mathbb{R}^2$ with $v \neq 0$. Then

$$D_v\, f(0) = \lim_{t\to 0} \frac{f(tv) - f(0)}{t} = \begin{cases} \displaystyle\lim_{t\to 0} \frac{v_1 v_2^2}{v_1^2 + t^4 v_2^6} = \frac{v_2^2}{v_1}\,, & \text{if } v_1 \neq 0 \\ 0\,, & \text{if } v_1 = 0\,. \end{cases}$$

Yet, for $h = (h_1, \sqrt{h_1})$ with $h_1 > 0$ we have

$$\lim_{h_1\to 0} f(h) = \lim_{h_1\to 0} \frac{h_1^2}{h_1^2 + h_1^3} = \lim_{h_1\to 0} \frac{1}{1 + h_1} = 1 \neq f(0)\,.$$

Therefore f is not continuous at 0. Together with the next result we obtain as a consequence that f is not differentiable at 0:

Theorem 4.7 *Let U be an open subset of $\mathbb{R}^n$, let $a \in U$ and let $f : U \to \mathbb{R}^m$ be differentiable at a. Then there is a constant $c > 0$ such that for all x from a neighborhood of a*

$$\|f(x) - f(a)\| \leq c\|x - a\|\,.$$

In particular, f is continuous at a.

Proof: We have

$$f(x) = f(a) + f'(a)(x-a) + r(x)\,\|x-a\|\,,$$

whence, with the operator norm $\|f'(a)\|$ of the linear mapping $f'(a) : \mathbb{R}^n \to \mathbb{R}^m$,

$$\|f(x) - f(a)\| \leq \|f'(a)\|\,\|x-a\| + \|r(x)\|\,\|x-a\|\,.$$

Since $\lim_{x\to a} r(x) = 0$, there is $\delta > 0$ such that

$$\|r(x)\| \le 1$$

for all $x \in D$ with $\|x - a\| < \delta$, whence for these x

$$\|f(x) - f(a)\| \le \big(\|f'(a)\| + 1\big)\,\|x - a\| = c\,\|x - a\|\,,$$

with $c = \|f'(a)\| + 1$. In particular, this implies

$$\lim_{x\to a} \|f(x) - f(a)\| \le \lim_{x\to a} c\|x - a\| = 0\,,$$

whence f is continuous at a. ∎

Theorem 4.8 *Let $U \subseteq \mathbb{R}^n$ be open and $a \in U$. If $f : U \to \mathbb{R}^m$ and $g : U \to \mathbb{R}^m$ are differentiable at a, then also $f + g$ and cf are differentiable at a for all $c \in \mathbb{R}$, and*

$$\begin{aligned}(f+g)'(a) &= f'(a) + g'(a)\\ (cf)'(a) &= cf'(a)\,.\end{aligned}$$

Proof: We have for $h \in \mathbb{R}^n$ with $a + h \in U$

$$\begin{aligned}f(a+h) &= f(a) + f'(a)h + r_1(a+h)\,\|h\|\,, \quad \lim_{h\to 0} r_1(a+h) = 0\\ g(a+h) &= g(a) + g'(a)h + r_2(a+h)\,\|h\|\,, \quad \lim_{h\to 0} r_2(a+h) = 0\,.\end{aligned}$$

Thus

$$(f+g)(a+h) = (f+g)(a) + \big(f'(a) + g'(a)\big)h + (r_1 + r_2)(a+h)\,\|h\|$$

with $\lim_{h\to 0}(r_1 + r_2)(a+h) = 0$. Consequently $f + g$ is differentiable at a with derivative $(f+g)'(a) = f'(a) + g'(a)$. The statement for cf follows in the same way. ∎

Theorem 4.9 (Product rule) *Let $U \subseteq \mathbb{R}^n$ be open and let $f, g : U \to \mathbb{R}$ be differentiable at $a \in U$. Then $f \cdot g : U \to \mathbb{R}$ is differentiable at a with derivative*

$$(f \cdot g)'(a) = f(a)\,g'(a) + g(a)\,f'(a)\,.$$

Proof: We have for $a + h \in U$

$$\begin{aligned}(f \cdot g)(a+h) &= \big(f(a) + f'(a)h + r_1(a+h)\,\|h\|\big) \cdot \big(g(a) + g'(a)h + r_2(a+h)\,\|h\|\big)\\ &= (f \cdot g)(a) + f(a)\,g'(a)h + g(a)\,f'(a)h + r(a+h)\,\|h\|\,,\end{aligned}$$

where

$$
\begin{aligned}
r(a+h)\,\|h\| &= \big(f'(a)h\, g'(a)\,\frac{h}{\|h\|}\big)\,\|h\| + \big(g(a)+g'(a)h\big)\, r_1(a+h)\,\|h\| \\
&\quad + \big(f(a)+f'(a)h\big) r_2(a+h)\,\|h\| + r_1(a+h)\, r_2(a+h)\,\|h\|^2 .
\end{aligned}
$$

The absolute value is a norm on $\mathbb{R}$. Since $r(a+h) \in \mathbb{R}$, we thus obtain with the operator norms $\|f'(a)\|$, $\|g'(a)\|$,

$$
\begin{aligned}
\lim_{h\to 0} |r(a+h)| &\le \lim_{h\to 0} \Big[\big(\|f'(a)\|\,\|h\|\,\|g'(a)\|\big) \\
&\quad + \big(|g(a)| + \|g'(a)\|\,\|h\|\big)\,|r_1(a+h)| \\
&\quad + \big(|f(a)| + \|f'(a)\|\,\|h\|\big)\,|r_2(a+h)| \\
&\quad + |r_1(a+h)|\,|r_2(a+h)|\,\|h\| \Big] = 0 .
\end{aligned}
$$

Since $f(a)\, g'(a)h + g(a)\, f'(a)h = \big(f(a)\, g'(a) + g(a)\, f'(a)\big)h$, it follows that $f \cdot g$ is differentiable at a with derivative given by this linear mapping. ■

Theorem 4.10 (Chain rule) *Let $U \subseteq \mathbb{R}^p$ and $V \subseteq \mathbb{R}^n$ be open, let $f : U \to V$ and $g : V \to \mathbb{R}^m$. Suppose that $a \in U$, that f is differentiable at a and that g is differentiable at $b = f(a)$. Then $g \circ f : U \to \mathbb{R}^n$ is differentiable at a with derivative*

$$(g \circ f)'(a) = g'\big(f(a)\big) \circ f'(a) .$$

Remark: Since $g'(b)$ and $f'(a)$ can be represented by matrices, $g'(b) \circ f'(a)$ can also be written as $g'(b)\, f'(a)$, employing matrix multiplication.

Proof: For brevity we set

$$T_2 = g'(b)\,, \quad T_1 = f'(a)\,,$$

and for $h \in \mathbb{R}^p$ with $a + h \in U$

$$R(h) = (g \circ f)(a+h) - (g \circ f)(a) - T_2 T_1 h\,.$$

The statement of the theorem follows if it can be shown that

$$\lim_{h\to 0} \frac{\|R(h)\|}{\|h\|} = 0.$$

We have for $x \in U$ and $y \in V$

$$
\begin{aligned}
f(x) - f(a) - T_1(x-a) &= r_1(x-a)\|x-a\|, \qquad \lim_{h\to 0} r_1(h) = 0 \\
g(y) - g(b) - T_2(y-b) &= r_2(y-b)\|y-b\|, \qquad \lim_{k\to 0} r_2(k) = 0.
\end{aligned}
$$

Since T_2 is linear, we thus obtain for $x = a + h$ and $y = f(a+h)$

$$\begin{aligned} R(h) &= g\big(f(a+h)\big) - g\big(f(a)\big) - T_2\big(f(a+h) - f(a)\big) \\ &\quad + T_2\big(f(a+h) - f(a) - T_1 h\big) \\ &= r_2\big(f(a+h) - f(a)\big)\|f(a+h) - f(a)\| + T_2\big(r_1(h)\|h\|\big), \end{aligned}$$

which yields

$$\begin{aligned} \lim_{h\to 0} \frac{\|R(h)\|}{\|h\|} &\le \lim_{h\to 0} \Big[\frac{1}{\|h\|}\, \|r_2\big(f(a+h) - f(a)\big)\|\, \|f(a+h) - f(a)\| \\ &\quad + \big\|T_2\big(r_1(h)\|h\|\big)\big\|\Big]. \end{aligned}$$

Since f is differentiable at a, for $\|h\|$ sufficiently small the estimate $\|f(a+h) - f(a)\| \le c\|h\|$ holds, cf. Theorem 4.7. Therefore, with the operator norm $\|T_2\|$ we conclude that

$$\lim_{h\to 0} \frac{\|R(h)\|}{\|h\|} \le \lim_{h\to 0} \Big[\|r_2\big(f(a+h) - f(a)\big)\|c + \|T_2\|\, \|r_1(h)\|\Big] = 0.$$

■

For the Jacobi–matrices of $f : U \to \mathbb{R}^n$, $g : V \to \mathbb{R}^m$ and $h : U \to \mathbb{R}^m$ we thus obtain

$$\begin{pmatrix} \frac{\partial h_1}{\partial x_1}(a) & \dots & \frac{\partial h_1}{\partial x_p}(a) \\ \vdots & & \\ \frac{\partial h_m}{\partial x_1}(a) & \dots & \frac{\partial h_m}{\partial x_p}(a) \end{pmatrix} = \begin{pmatrix} \frac{\partial g_1}{\partial y_1}(b) & \dots & \frac{\partial g_1}{\partial y_n}(b) \\ \vdots & & \\ \frac{\partial g_m}{\partial y_1}(b) & \dots & \frac{\partial g_m}{\partial y_n}(b) \end{pmatrix} \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \dots & \frac{\partial f_1}{\partial x_p}(a) \\ \vdots & & \\ \frac{\partial f_n}{\partial x_1}(a) & \dots & \frac{\partial f_n}{\partial x_p}(a) \end{pmatrix}.$$

Thus,

$$\frac{\partial h_j}{\partial x_i}(a) = \sum_{k=1}^{n} \frac{\partial g_j}{\partial y_k}(b) \frac{\partial f_k}{\partial x_i}(a), \quad i = 1, \dots, p, \;\; j = 1, \dots, m.$$

Corollary 4.11 *Let U be an open subset of $\mathbb{R}^n$, let $a \in U$ and let $f : U \to \mathbb{R}$ be differentiable at a and satisfy $f(a) \neq 0$. Then $\frac{1}{f}$ is differentiable at a with derivative*

$$\Big(\frac{1}{f}\Big)'(a) = -\frac{1}{f(a)^2}\, f'(a).$$

Proof: Consider the differentiable function $g : \mathbb{R}\backslash\{0\} \to \mathbb{R}$ defined by $g(x) = \frac{1}{x}$. Then

$$\frac{1}{f} = g \circ f : \{x \in U \,\big|\, f(x) \neq 0\} \to \mathbb{R}$$

is differentiable at a with derivative

$$\left(\frac{1}{f}\right)'(a) = g'\big(f(a)\big)f'(a) = -\frac{1}{f(a)^2}\,f'(a). \qquad \blacksquare$$

Assume that U and V are open subsets of $\mathbb{R}^n$ and that $f : U \to V$ is an invertible map with inverse $f^{-1} : V \to U$. If $a \in U$, if f is differentiable at a and if f^{-1} is differentiable at $b = f(a) \in V$, then the derivative $(f^{-1})'(b)$ can be computed from $f'(a)$ using the chain rule. To see this, note that

$$f^{-1} \circ f = \mathrm{id}_U.$$

The identity mapping id_U is obtained as the restriction of the identity mapping $id_{\mathbb{R}^n}$ to U. Since $id_{\mathbb{R}^n}$ is linear, it follows that id_U is differentiable at every $c \in U$ with derivative $(id_U)'(x) = id_{\mathbb{R}^n}$. Consequently

$$id_{\mathbb{R}^n} = (id_U)'(a) = (f^{-1} \circ f)'(a) = (f^{-1})'(b)\,f'(a)\,.$$

From linear algebra we know that this equation implies that $(f^{-1})'(b)$ is the inverse of $f'(a)$. Consequently, $\big(f'(a)\big)^{-1}$ exists and

$$(f^{-1})'(b) = \big(f'(a)\big)^{-1},$$

or

$$(f^{-1})'(b) = \big[f'\big(f^{-1}(b)\big)\big]^{-1}\,.$$

Thus, if one assumes that $f'(a)$ exists and that the inverse mapping is differentiable at $f(a)$, one can conclude that the linear mapping $f'(a)$ is invertible. On the other hand, if one assumes that $f'(a)$ exists and is invertible and that the inverse mapping is continuous at $f(a)$, one can conclude that the inverse mapping is differentiable at $f(a)$. This is shown in the following theorem. We remark that the linear mapping $f'(a)$ is invertible if and only if the determinant $\det f'(a)$ differs from zero, where $f'(a)$ is identified with the $n\times n$-matrix representing the linear mapping $f'(a)$.

Theorem 4.12 *Let $U \subseteq \mathbb{R}^n$ be an open subset, let $a \in U$ and let $f : U \to \mathbb{R}^n$ be one-to-one. If f is differentiable at a with invertible derivative $f'(a)$, if the range $f(U)$ contains a neighborhood of $b = f(a)$, and if the inverse mapping $f^{-1} : f(U) \to U$ of f is continuous at b, then f^{-1} is differentiable at b with derivative*

$$(f^{-1})'(b) = \big(f'(a)\big)^{-1} = \Big(f'\big(f^{-1}(b)\big)\Big)^{-1}.$$

Proof: For brevity we set $g = f^{-1}$. First it is shown that there is a neighborhood $V \subseteq f(U)$ of b and a constant $c > 0$ such that

$$\frac{\|g(y) - g(b)\|}{\|y - b\|} \le c \qquad (*)$$

for all $y \in V$.

Since f is differentiable at a, we have for $x \in U$

$$f(x) - f(a) = f'(a)(x-a) + r(x)\,\|x-a\|\,, \tag{$**$}$$

where r is continuous at a and satisfies $r(a) = 0$. Let $y \in f(U)$. Employing $(**)$ with $x = g(y)$ and noting that $b = f(a)$, we obtain from the inverse triangle inequality that

$$\begin{aligned}
\frac{\|g(y)-g(b)\|}{\|y-b\|} &= \frac{\|g(y)-a\|}{\|f\big(g(y)\big)-f(a)\|} \\
&= \frac{\|g(y)-a\|}{\big\|f'(a)\big(g(y)-a\big)+r\big(g(y)\big)\,\|g(y)-a\|\,\big\|} \\
&\le \frac{\|\big(f'(a)\big)^{-1}f'(a)\big(g(y)-a\big)\,\|}{\|f'(a)\big(g(y)-a\big)\|-\|r\big(g(y)\big)\|\,\|\big(f'(a)\big)^{-1}f'(a)\big(g(y)-a\big)\|} \\
&\le \frac{\|\big(\big(f'(a)\big)^{-1}\|\,\|f'(a)\big(g(y)-a\big)\|}{\|f'(a)\big(g(y)-a\big)\|\,\Big(1-\|r\big(g(y)\big)\|\,\|\big(f'(a)\big)^{-1}\|\Big)} \\
&= \frac{\|\big(f'(a)\big)^{-1}\|}{1-\|r\big(g(y)\big)\|\,\|\big(f'(a)\big)^{-1}\|}
\end{aligned}$$

The inequality $(*)$ is obtained from this estimate. To see this, note that by assumption g is continuous at b and that r is continuous at $a = g(b)$, hence $r \circ g$ is continuous at b. Thus,

$$\lim_{y\to b} r\big(g(y)\big) = r\big(g(b)\big) = r(a) = 0.$$

Using $(*)$ the theorem can be proved as follows: we have to show that

$$\lim_{y\to b} \frac{g(y)-g(b)-\big(f'(a)\big)^{-1}(y-b)}{\|y-b\|} = 0\,.$$

Employing $(**)$ again,

$$\begin{aligned}
&\frac{g(y)-a-\big(f'(a)\big)^{-1}(y-b)}{\|y-b\|} \\
&\quad= \frac{g(y)-a-\big(f'(a)\big)^{-1}\Big(f\big(g(y)\big)-f(a)\Big)}{\|y-b\|} \\
&\quad= \frac{g(y)-a-\big(f'(a)\big)^{-1}\Big(f'(a)\big(g(y)-a\big)+r\big(g(y)\big)\,\|g(y)-a\|\Big)}{\|y-b\|} \\
&\quad= -f'(a)\Big(r\big(g(y)\big)\Big)\,\frac{\|g(y)-a\|}{\|y-b\|}\,.
\end{aligned}$$

With $a = g(b)$ we thus obtain from $(*)$

$$\lim_{y\to b} \Big\| \frac{g(y) - g(b) - \big(f'(a)\big)^{-1}(y-b)}{\|y-b\|} \Big\|$$
$$\leq \lim_{y\to b} \|f'(a)\| \, \|r\big(g(y)\big)\| \, c = c\,\|f'(a)\| \lim_{y\to b} \|r\big(g(y)\big)\| = 0\,.$$

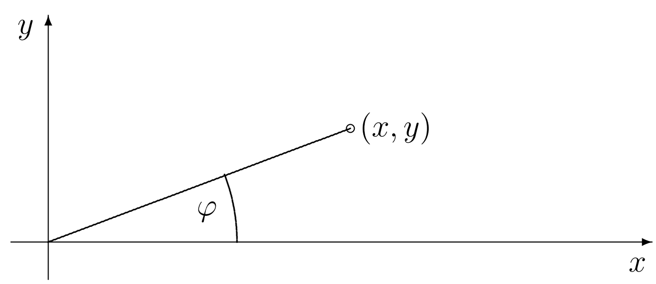
■

Example (Polar coordinates) Let

$$U = \big\{(r,\varphi) \;\big|\; r > 0\,, \quad 0 < \varphi < 2\pi\big\} \subseteq \mathbb{R}^2,$$

and let $f = (f_1, f_2) : U \to \mathbb{R}^2$ be defined by

$$\begin{aligned} x &= f_1(r,\varphi) = r\cos\varphi \\ y &= f_2(r,\varphi) = r\sin\varphi\,. \end{aligned}$$



This mapping is one-to-one with range

$$f(U) = \mathbb{R}^2 \backslash \big\{(x,0) \;\big|\; x \geq 0\big\}\,,$$

and has a continuous inverse. From a theorem proved in the next section it follows that f is differentiable. Thus,

$$f'(r,\varphi) = \begin{pmatrix} \frac{\partial f_1}{\partial r}(r,\varphi) & \frac{\partial f_1}{\partial \varphi}(r,\varphi) \\ \frac{\partial f_2}{\partial r}(r,\varphi) & \frac{\partial f_2}{\partial \varphi}(r,\varphi) \end{pmatrix} = \begin{pmatrix} \cos\varphi & -r\,\sin\varphi \\ \sin\varphi & r\,\cos\varphi \end{pmatrix}.$$

This matrix is invertible for $(r,\varphi) \in U$, hence the derivative $(f^{-1})'(x,y)$ exists for every $(x,y) = f(r,\varphi) = (r\,\cos\varphi\,,\, r\sin\varphi)$ and can be computed without having to determine the inverse function f^{-1}:

$$\begin{aligned} (f^{-1})'(x,y) &= \big(f'(r,\varphi)\big)^{-1} = \begin{pmatrix} \cos\varphi & -r\,\sin\varphi \\ \sin\varphi & r\,\cos\varphi \end{pmatrix}^{-1} \\ &= \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\frac{1}{r}\sin\varphi & \frac{1}{r}\cos\varphi \end{pmatrix} = \begin{pmatrix} \frac{x}{\sqrt{x^2+y^2}} & \frac{y}{\sqrt{x^2+y^2}} \\ \frac{-y}{x^2+y^2} & \frac{x}{x^2+y^2} \end{pmatrix} \end{aligned}$$

4.4 Mean value theorem

The mean value theorem for real functions can be generalized to *real valued* functions:

Theorem 4.13 (Mean value theorem) *Let U be an open subset of $\mathbb{R}^n$, let $f : U \to \mathbb{R}$ be differentiable, and let $a, b \in U$ be points such that the line segment connecting these points is contained in U. Then there is a point c from this line segment with*

$$f(b) - f(a) = f'(c)(b - a)\,.$$

Proof: Define a function $\gamma : [0,1] \to U$ by $t \mapsto \gamma(t) := a + t(b - a)$. This function maps the interval $[0,1]$ onto the line segment connecting a and b. The affine function γ is differentiable with derivative

$$\gamma'(t) = b - a\,.$$

Let $F = f \circ \gamma$ be the composition. Since f and γ are differentiable, $F : [0,1] \to \mathbb{R}$ is differentiable. Thus, the mean value theorem for real functions implies that there is $\vartheta \in (0,1)$ such that

$$f(b) - f(a) = F(1) - F(0) = F'(\vartheta) = f'\big(\gamma(\vartheta)\big)\,\gamma'(\vartheta) = f'(c)(b - a)\,,$$

where we have set $c = \gamma(\vartheta)$. ∎

Of course, the mean value theorem can also be formulated as follows: If U contains together with the points x and $x + h$ also the line segment connecting these points, then there is a number ϑ with $0 < \vartheta < 1$ such that

$$f(x + h) - f(x) = f'(x + \vartheta h)h\,.$$

The mean value theorem does not hold for functions $f : U \to \mathbb{R}^m$ with $m > 1$, but the following weaker result can often be used as a replacement for the mean value theorem:

Corollary 4.14 *Let $U \subseteq \mathbb{R}^n$ be open and let $f : U \to \mathbb{R}^m$ be differentiable. Assume that x and $x + h$ are points from U such that the line segment $\ell = \{x + th \mid 0 \le t \le 1\}$ connecting x and $x + h$ is contained in U. If the derivative of f is bounded on ℓ by a constant $S \ge 0$, i.e. if for all $0 \le t \le 1$ the operator norm of the derivative satisfies*

$$\|f'(x + th)\| \le S\,,$$

then

$$\|f(x + h) - f(x)\| \le S\|h\|\,.$$

To prove this corollary we need the following lemma, which we do not prove:

Lemma 4.15 *Let $\|\cdot\|$ be a norm on $\mathbb{R}^m$. Then to every $u \in \mathbb{R}^m$ there is a linear mapping $A_u : \mathbb{R}^m \to \mathbb{R}$ such that $\|A_u\| = 1$ and $A_u(u) = \|u\|$.*

Example: For the Euclidean norm $\|\cdot\| = |\cdot|$ define A_u by

$$A_u(v) = \frac{u}{|u|} \cdot v\,, \quad v \in \mathbb{R}^m\,.$$

Then $A_u(u) = \frac{u}{|u|} \cdot u = |u|$ and

$$\begin{aligned} 1 = \Big|\frac{u}{|u|}\Big| &= \frac{1}{|u|} A_u(u) \leq \frac{1}{|u|} \|A_u\|\,|u| = \|A_u\| \\ &= \sup_{|v|\leq 1} |A_u(v)| = \sup_{|v|\leq 1} \Big|\frac{u}{|u|} \cdot v\Big| \leq \sup_{|v|\leq 1} \frac{|u| \cdot |v|}{|u|} = 1\,, \end{aligned}$$

Hence $\|A_u\| = 1$.

Proof of the corollary: To $f(x+h)-f(x) \in \mathbb{R}^m$ choose the linear mapping $A : \mathbb{R}^m \to \mathbb{R}$ such that $\|A\| = 1$ and $A\big(f(x+h) - f(x)\big) = \|f(x+h) - f(x)\|$. As a linear mapping, A is differentiable with derivative $A'(y) = A$ for all $y \in \mathbb{R}^m$. Thus, from the mean value theorem applied to the differentiable function $F = A \circ f : U \to \mathbb{R}$ we conclude that a number ϑ with $0 < \vartheta < 1$ exists such that

$$\begin{aligned} &\|f(x+h) - f(x)\| = A\big(f(x+h) - f(x)\big) \\ &\quad = A\big(f(x+h)\big) - A\big(f(x)\big) = F(x+h) - F(x) = F'(x+\vartheta h)h \\ &\quad = Af'(x+\vartheta h)h \leq \|A\|\,\|f'(x+\vartheta h)\|\,\|h\| \leq S\,\|h\|\,. \end{aligned}$$

∎

Theorem 4.16 *Let U be an open and pathwise connected subset of $\mathbb{R}^n$, and let $f : U \to \mathbb{R}^m$ be differentiable. Then f is constant if and only if $f'(x) = 0$ for all $x \in U$.*

To prove this theorem, the following lemma is needed:

Lemma 4.17 *Let $U \subseteq \mathbb{R}^n$ be open and pathwise connected. Then all points $a, b \in U$ can be connected by a polygon in U, i.e. by a curve consisting of finitely many straight line segments.*

A **proof** of this lemma can be found in the book of Barner-Flohr, Analysis II, p. 56.

Proof of the theorem: If f is constant, then evidently $f'(x) = 0$ for all $x \in U$. To prove the converse, assume that $f'(x) = 0$ for all $x \in U$. Let a, b be two arbitrary points in U. These points can be connected in U by a polygon with the corner points

$$a_0 = a\,,\; a_1, \ldots, a_{k-1}\,,\; a_k = b\,.$$

We apply Corollary 4.14 to the line segment connecting a_j and a_{j+1} for $j = 0, 1, \ldots, k-1$. Since $f'(x) = 0$ for all $x \in U$, the operator norm $\|f'(x)\|$ is bounded on this line segment by 0. Therefore Corollary 4.14 yields $\|f(a_{j+1}) - f(a_j)\| \leq 0$, hence $f(a_{j+1}) = f(a_j)$ for all $j = 0, 1, \ldots, k-1$, which implies

$$f(b) = f(a)\,.$$

■

From the existence of all the partial derivatives $\frac{\partial f}{\partial x_1}(a), \ldots, \frac{\partial f}{\partial x_n}(a)$ at a, one cannot conclude that f is differentiable at a. However, we have the following useful criterion for differentiability of f at a:

Theorem 4.18 *Let U be an open subset of $\mathbb{R}^n$ with $a \in U$ and let $f : U \to \mathbb{R}^m$. If all partial derivatives $\frac{\partial f_j}{\partial x_i}$ exist in U for $i = 1, \ldots, n$ and $j = 1, \ldots, m$, and if all the functions $x \mapsto \frac{\partial f_j}{\partial x_i}(x) : U \to \mathbb{R}$ are continuous at a, then f is differentiable at a.*

Proof: It suffices to prove that all the component functions $f_1, \ldots, f_m$ are differentiable at a. Thus, we can asssume that $f : U \to \mathbb{R}$ is real valued. We have to show that

$$\lim_{h \to 0} \frac{f(a+h) - f(a) - Th}{\|h\|_\infty} = 0$$

for the linear mapping T with the matrix representation

$$T = \Big(\frac{\partial f}{\partial x_1}(a), \ldots, \frac{\partial f}{\partial x_n}(a)\Big)\,.$$

For $h = (h_1, \ldots, h_n) \in \mathbb{R}^n$ define

$$\begin{aligned} a_0 &:= a \\ a_1 &:= a_0 + h_1 e_1 \\ a_2 &:= a_1 + h_2 e_2 \\ &\vdots \\ a + h = a_n &:= a_{n-1} + h_n e_n\,, \end{aligned}$$

where $e_1, \dots, e_n$ is the canonical basis of $\mathbb{R}^n$. Then

$$f(a+h) - f(a) = \big(f(a+h) - f(a_{n-1})\big) + \big(f(a_{n-1}) - f(a_{n-2})\big) + \ldots + \big(f(a_1) - f(a)\big)\,. \quad (*)$$

If x runs through the line segment connecting a_{j-1} to a_j, then only the component x_j of x is varying. Since by assumption the mapping $x_j \to f(x_1, \dots, x_j, \dots, x_n)$ is differentiable, the mean value theorem can be applied to every term on the right hand side of $(*)$. Let c_j be the intermediate point on the line segment connecting a_{j-1} to a_j. Then

$$f(a+h) - f(a) = \sum_{j=1}^{n} \big(f(a_j) - f(a_{j-1})\big) = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(c_j) h_j\,,$$

whence

$$\begin{aligned}
|f(a+h) - f(a) - Th| &= \Big|\sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(c_j) h_j - \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(a) h_j\Big| \\
&= \Big|\sum_{j=1}^{n} \Big(\frac{\partial f}{\partial x_j}(c_j) - \frac{\partial f}{\partial x_j}(a)\Big) h_j\Big| \\
&\leq \|h\|_\infty \sum_{j=1}^{n} \Big|\frac{\partial f}{\partial x_j}(c_j) - \frac{\partial f}{\partial x_j}(a)\Big|\,.
\end{aligned}$$

Because the intermediate points satisfy $\|c_j - a\|_\infty \leq \|h\|_\infty$ for all $j = 1, \dots, n$, it follows that $\lim_{h \to 0} c_j = a$ for all intermediate points. The continuity of the partial derivatives at a thus implies

$$\lim_{h \to 0} \frac{|f(a+h) - f(a) - Th|}{\|h\|_\infty} \leq \lim_{h \to 0} \sum_{j=1}^{n} \Big|\frac{\partial f}{\partial x_j}(c_j) - \frac{\partial f}{\partial x_j}(a)\Big| = 0\,.$$

■

Example: Let $s \in \mathbb{R}$ and let $f : \mathbb{R}^n \backslash \{0\} \to \mathbb{R}$ be defined by

$$f(x) = (x_1^2 + \ldots + x_n^2)^s.$$

This mapping is differentiable, since the partial derivatives

$$\frac{\partial f}{\partial x_j}(x) = s(x_1^2 + \ldots + x_n^2)^{s-1} 2x_j$$

are continuous in $\mathbb{R}^n \backslash \{0\}$.

4.5 Continuously differentiable mappings, second derivative

Let $U \subseteq \mathbb{R}^n$ be open and let $f : U \to \mathbb{R}^m$ be differentiable at every $x \in U$. Then

$$x \mapsto f'(x) : U \to L(\mathbb{R}^n, \mathbb{R}^m)$$

defines a mapping from U into the set of linear mappings from $\mathbb{R}^n$ to $\mathbb{R}^m$. If one applies the linear mapping $f'(x)$ to a vector $h \in \mathbb{R}^n$, a vector of $\mathbb{R}^m$ is obtained. Thus, f' can also be considered to be a mapping from $U \times \mathbb{R}^n$ to $\mathbb{R}^m$:

$$(x, h) \mapsto f'(x)h : U \times \mathbb{R}^n \to \mathbb{R}^m.$$

This mapping is linear with respect to the second argument. What view one takes depends on the situation.

Since $L(\mathbb{R}^n, \mathbb{R}^m)$ is a normed space, one can define continuity of the function f' as follows:

Definition 4.19 *Let $U \subseteq \mathbb{R}^n$ be an open set and let $f : U \to \mathbb{R}^m$ be differentiable.*

(i) *$f' : U \to L(\mathbb{R}^n, \mathbb{R}^m)$ is said to be continuous at $a \in U$ if to every $\varepsilon > 0$ there is $\delta > 0$ such that for all $x \in U$ with $\|x - a\| < \delta$*

$$\|f'(x) - f'(a)\| < \varepsilon\,.$$

(ii) *f is said to be continuously differentiable if $f' : U \to L(\mathbb{R}^n, \mathbb{R}^m)$ is continuous.*

(iii) *Let $U, V \subseteq \mathbb{R}^n$ be open and let $f : U \to V$ be continuously differentiable and invertible. If the inverse $f^{-1} : V \to U$ is also continuously differentiable and invertible. If the inverse $f^{-1} : V \to U$ is also continuously differentiable, then f is called a diffeomorphism.*

Here $\|f'(x) - f'(a)\|$ denotes the operator norm of the linear mapping $\big(f'(x) - f'(a)\big) : \mathbb{R}^n \to \mathbb{R}^m$. The following result makes this definition less abstract:

Theorem 4.20 *Let $U \subseteq \mathbb{R}^n$ be open and let $f : U \to \mathbb{R}^m$. Then the following statements are equivalent:*

(i) *f is continuously differentiable.*

(ii) *All partial derivatives $\frac{\partial}{\partial x_i} f_j$ with $1 \leq i \leq n$, $1 \leq j \leq m$ exist in U and are continuous functions*

$$x \mapsto \frac{\partial}{\partial x_i} f_j(x) : U \to \mathbb{R}\,.$$

(iii) *f is differentiable and the mapping $x \mapsto f'(x)h : U \to \mathbb{R}^m$ is continuous for every $h \in \mathbb{R}^n$.*

Proof: First we show that (i) and (ii) are equivalent. If f is differentiable, then all partial derivatives exist in U. Conversely, if all partial derivatives exist in U and are continuous, then by Theorem 4.18 the function f is differentiable. Hence, it remains to show that f' is continuous if and only if all partial derivatives are continuous.

For $a, x \in U$ let

$$\|f'(x) - f'(a)\|_\infty = \max_{\substack{i=1,\ldots,n \\ j=1,\ldots,m}} \Big|\frac{\partial f_j}{\partial x_i}(x) - \frac{\partial f_j}{\partial x_i}(a)\Big|\,. \tag{$*$}$$

By Theorem 3.50 there exist constants $c, C > 0$, which are independent of x and a, such that $c\|f'(x) - f'(a)\|_\infty \leq \|f'(x) - f'(a)\| \leq C\|f'(x) - f'(a)\|_\infty$. From this estimate and from $(*)$ we see that

$$\lim_{x\to 0} \|f'(x) - f'(a)\| = 0$$

holds if and only if

$$\lim_{x\to a} \frac{\partial f_j}{\partial x_i}(x) = \frac{\partial f_j}{\partial x_i}(a)$$

for all $1 \leq i \leq n\,,\ 1 \leq j \leq m$. By Definition 4.19 this means that f' is continuous at a if and only if all partial derivatives are continuous at a.

To prove that (iii) is equivalent to the first two statements of the theorem it suffices to remark that if f is differentiable, then

$$x \mapsto f'(x)h = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x)h_i : U \to \mathbb{R}^m.$$

By choosing for h vectors from the standard basis $e_1, \ldots, e_n$ of $\mathbb{R}^n$, we immediately see from this equation that $x \mapsto f'(x)h$ is continuous for every $h \in \mathbb{R}^n$, if and only if all partial derivatives are continuous. ∎

The derivative $f : U \to \mathbb{R}^m$ is a mapping $f' : U \to L(\mathbb{R}^n, \mathbb{R}^m)$. Since $L(\mathbb{R}^n, \mathbb{R}^m)$ is a normed space, it is possible to define the derivative of f' at x, which is a linear mapping from $\mathbb{R}^n$ to $L(\mathbb{R}^n, \mathbb{R}^m)$. One denotes this derivative by $f''(x)$ and calls it the second derivative of f at x. Thus, if f is two times differentiable, then

$$f'' : U \to L\big(\mathbb{R}^n, L(\mathbb{R}^n, \mathbb{R}^m)\big)\,.$$

Less abstractly, I define the second derivative in the following equivalent way:

Definition 4.21 (i) *Let $U \subseteq \mathbb{R}^n$ be open and let $f : U \to \mathbb{R}^m$ be differentiable. f is said to be two times differentiable at a point $x \in U$, if to every fixed $h \in \mathbb{R}^n$ the mapping $g_h : U \to \mathbb{R}^m$ defined by*

$$g_h(x) = f'(x)h$$

is differentiable at x.

(ii) *The function $f''(x) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$ defined by*

$$f''(x)(h,k) = g_h'(x)(k)$$

is called the second derivative of f at x. If $f : U \to \mathbb{R}^m$ is two times differentiable (i.e., two times differentiable at every $x \in U$), then

$$f'' : U \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m.$$

Theorem 4.22 *Let $U \subseteq \mathbb{R}^n$ be open with $x \in U$ and let $f : U \to \mathbb{R}^m$ be differentiable.*

(i) *If f is two times differentiable at x, then all second partial derivatives of f at x exist, and for $h = (h_1, \dots h_n) \in \mathbb{R}^n$ and $k = (k_1, \dots, k_n) \in \mathbb{R}^n$*

$$f''(x)(h,k) = \sum_{j=1}^{n} \sum_{i=1}^{n} \frac{\partial}{\partial x_j} \frac{\partial}{\partial x_i} f(x) h_i k_j \,.$$

(ii) *$f''(x)$ is bilinear, i.e. $(h,k) \to f''(x)(h,k)$ is linear in both arguments.*

Proof: If f is two times differentiable at x, then by definition the function

$$y \mapsto g_k(y) = f'(y)h = \sum_{i=1}^{n} \frac{\partial}{\partial x_i} f(y) h_i$$

is differentiable at $y = x$, hence

$$f''(x)(h,k) = g_h'(x)k = \sum_{j=1}^{n} \frac{\partial}{\partial x_j} g_h(x) k_j = \sum_{j=1}^{n} \frac{\partial}{\partial x_j} \Big(\sum_{i=1}^{n} \frac{\partial}{\partial x_i} f(x) h_i \Big) k_j \,.$$

With $h = e_i$ and $k = e_j$, where e_i and e_j are vectors from the standard basis of $\mathbb{R}^n$, this formula implies that the second partial derivative $\frac{\partial}{\partial x_j} \frac{\partial}{\partial x_i} f(x)$ exists. Thus, in this formula the partial derivative and the summation can be interchanged, hence the stated representation formula for $f''(x)(h,k)$ results. The bilinearity of $f''(x)$ follows immediately from this representation formula. ∎

For the *second partial derivatives* $\frac{\partial}{\partial x_j}\frac{\partial}{\partial x_i}f(x)$ of f one also uses the notation

$$\frac{\partial^2 f}{\partial x_j \partial x_i} = \frac{\partial}{\partial x_j}\frac{\partial}{\partial x_i}f, \qquad \frac{\partial^2 f}{\partial x_i^2} = \frac{\partial}{\partial x_i}\frac{\partial}{\partial x_i}f.$$

Note that

$$\frac{\partial^2 f}{\partial x_j \partial x_i}(x) = \begin{pmatrix} \dfrac{\partial^2}{\partial x_j \partial x_i} f_1(x) \\ \vdots \\ \dfrac{\partial^2}{\partial x_j \partial x_i} f_m(x) \end{pmatrix} \in \mathbb{R}^m.$$

For $m = 1$, the second partial derivatives $\frac{\partial^2}{\partial x_j \partial x_i}f(x)$ are real numbers. Thus, for $f : U \to \mathbb{R}$ we obtain a matrix representation for $f''(x)$:

$$\begin{aligned} f''(x)(h,k) &= \sum_{j=1}^{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial x_j \partial x_i}f(x)h_i k_j \\ &= (h_1,\ldots,h_n)\begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1^2}(x) & \ldots & \dfrac{\partial^2 f}{\partial x_n \partial x_1}(x) \\ \vdots & & \\ \dfrac{\partial^2 f}{\partial x_1 \partial x_n}(x) & \ldots & \dfrac{\partial^2 f}{\partial x_n^2}(x) \end{pmatrix}\begin{pmatrix} k_1 \\ \vdots \\ k_n \end{pmatrix} = hHk, \end{aligned}$$

with the Hessian matrix

$$H = \Big(\frac{\partial^2 f}{\partial x_j \partial x_i}\Big)_{j,i=1,\ldots,n}.$$

(Ludwig Otto Hesse 1811 – 1874). For $f : U \to \mathbb{R}^m$ with $m > 1$ one obtains

$$\big(f''(x)\big)_\ell(h,k) = hH_\ell k,$$

where H_ℓ is the Hessian matrix for the component function f_ℓ of f. In particular, this yields

$$\big(f''(x)\big)_\ell(h,k) = (f_\ell)''(x)(h,k),$$

i.e. the ℓ–th component of $f''(x)$ is the second derivative of the component function f_ℓ.

It is possible, that all second partial derivatives of f at x exist, even if f is not two times differentiable at x. In this case the Hessian matrices H_ℓ can be formed, but they do not represent the second derivative at f at x, which does not exist. If f is two times differentiable at x, then the Hessian matrices H_ℓ are symmetric, i.e.

$$\frac{\partial^2}{\partial x_j \partial x_i}f_\ell(x) = \frac{\partial^2}{\partial x_i \partial x_j}f_\ell(x)$$

for all $1 \le i, j \le n$, hence the order of differentiation does not matter. This follows from the following.

Theorem 4.23 (of H.A. Schwarz) *Let $U \subseteq \mathbb{R}^n$ be open, let $x \in U$ and let f be two times differentiable at x. Then for all $h, k \in \mathbb{R}^n$*

$$f''(x)(h,k) = f''(x)(k,h).$$

(Hermann Amandus Schwartz, 1843 – 1921)

Proof: Obviously the bilinear mapping $f''(x)$ is symmetric, if and only if every component function $((f''(x))_\ell$ is symmetric. Therefore it suffices to show that every component is symmetric. Since $(f''(x))_\ell = (f_\ell)''(x)$ and since $f_\ell : U \to \mathbb{R}$ is real valued, it is sufficient to prove that for every real valued function $f : U \to \mathbb{R}$ the second derivative $f''(x)$ is symmetric. We thus assume that f is real valued.

To prove symmetry, we show that for all $h, k \in \mathbb{R}^n$

$$\lim_{\substack{s\to 0\\ s>0}} \frac{f(x+sh+sk) - f(x+sh) - f(x+sk) + f(x)}{s^2} = f''(x)(h,k). \qquad (*)$$

The statement of the theorem ist a consequence of this formula, since the left hand side remains unchanged if h and k are interchanged.

By definition, $f''(x)(h,k)$ is the derivative of the function $x \mapsto f'(x)h$. Thus, for all $h, k \in \mathbb{R}^n$,

$$f'(x+k)h - f'(x)h = f''(x)(h,k) + R_x(h,k)\|k\| \qquad (**)$$

with

$$\lim_{k\to 0} R_x(h,k) = 0.$$

$R_x(h,k)$ is linear with respect to h, since $f'(x+k)h$, $f'(x)h$ and $f''(x)(h,k)$ are linear with respect to h. We show that a number ϑ with $0 < \vartheta < 1$ exists, which depends on h and k, such that

$$\begin{aligned} & f(x+h+k) - f(x+h) - f(x+k) + f(x) \qquad (+)\\ = \ & f''(x)(h,k) + R_x(h, \vartheta h + k)\|\vartheta h + k\| - R_x(h, \vartheta h)\|\vartheta h\|. \end{aligned}$$

For, let $F : [0,1] \to \mathbb{R}$ be defined by

$$F(t) = f(x+th+k) - f(x+th).$$

F is differentiable, whence the mean value theorem implies that $0 < \vartheta < 1$ exists with

$$F(1) - F(0) = F'(\vartheta).$$

Therefore, with the definition of F and with $(**)$,

$$\begin{aligned} & f(x+h+k)-f(x+h)-f(x+k)+f(x)=F(1)-F(0) \\ = \; & F'(\vartheta)=f'(x+\vartheta h+k)h-f'(x+\vartheta h)h \\ = \; & \big(f'(x+\vartheta h+k)h-f'(x)h\big)-\big(f'(x+\vartheta h)h-f'(x)h\big) \\ = \; & \big(f''(x)(h,\vartheta h+k)+R_x(h,\vartheta h+k)\|\vartheta h+k\|\big) \\ & -\big(f''(x)(h,\vartheta h)+R_x(h,\vartheta h)\|\vartheta h\|\big) \\ = \; & f''(x)(h,k)+R_x(h,\vartheta h+k)\|\vartheta h+k\|-R_x(h,\vartheta h)\|\vartheta h\|, \end{aligned}$$

which is $(+)$. In the last step we used the linearity of $f''(x)$ in the second argument.

Let $s>0$. If one replaces in $(+)$ the vector k by sk and the vector h by sh, then on the right hand side the factor s^2 can be extracted, because of the bilinearity or linearity or the positive homogeneity of all the terms. The result is

$$\begin{aligned} & f(x+sh+sk)-f(x+sh)-f(x+sk)+f(x) \\ = \; & s^2\Big[f''(x)(h,k)+R_x\big(h,s(\vartheta h+k)\big)\|\vartheta h+k\|-R_x(h,s\vartheta h)\|\vartheta h\|\Big]. \end{aligned}$$

Since

$$\lim_{s\to 0} R_x\big(h,s(\vartheta h+k)\big)=0, \quad \lim_{s\to 0} R_x(h,s\vartheta h)=0,$$

this equation yields $(*)$. ■

Example: Let $f:\mathbb{R}^2\to\mathbb{R}$

$$f(x_1,x_2)=x_1^2x_2+x_1+x_2^3.$$

The partial derivatives of every order exist and are continuous. This implies that f is continuously differentiable. We have

$$\operatorname{grad} f(x)=\begin{pmatrix} \dfrac{\partial f}{\partial x_1}(x) \\ \dfrac{\partial f}{\partial x_2}(x) \end{pmatrix}=\begin{pmatrix} 2x_1x_2+1 \\ x_1^2+3x_2^2 \end{pmatrix}.$$

For $h\in\mathbb{R}^2$ the partial derivatives of

$$x\mapsto f'(x)h=\operatorname{grad} f(x)\cdot h=\frac{\partial f}{\partial x_1}(x)h_1+\frac{\partial f}{\partial x_2}(x)h_2$$

are

$$\frac{\partial}{\partial x_i}\big(f'(x)h\big)=\frac{\partial^2 f}{\partial x_i\partial x_1}(x)h_1+\frac{\partial^2 f}{\partial x_i\partial x_2}(x)h_2\,, \quad i=1,2\,,$$

hence these partial derivatives are continuous, and so $x \mapsto f'(x)h$ is differentiable. Thus, by definition f is two times differentiable with the Hessian matrix

$$f''(x) = H = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1^2}(x) & \dfrac{\partial^2 f}{\partial x_2 \partial x_1}(x) \\ \dfrac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \dfrac{\partial^2 f}{\partial x_2^2}(x) \end{pmatrix} = \begin{pmatrix} 2x_2 & 2x_1 \\ 2x_1 & 6x_2 \end{pmatrix}.$$

4.6 Higher derivatives, Taylor formula

Higher derivatives are defined by induction: Let $U \subseteq \mathbb{R}^n$ be open. The p-th derivative of $f : U \to \mathbb{R}^m$ at x is a mapping

$$f^{(p)}(x) : \underbrace{\mathbb{R}^n \times \ldots \times \mathbb{R}^n}_{p\text{-factors}} \to \mathbb{R}^m$$

obtained as follows: If f is $(p-1)$–times differentiable and if for all $h_1, \ldots, h_{p-1} \in \mathbb{R}^n$ the mapping

$$x \mapsto f^{(p-1)}(x)(h_1, \ldots, h_{p-1}) : U \to \mathbb{R}^m$$

is differentiable at x, then f is said to be p-times continuously differentiable at x with p-th derivative $f^{(p)}(x)$ defined by

$$f^{(p)}(x)(h_1, \ldots, h_p) = \big[f^{p-1}(\cdot)(h_1, \ldots, h_{p-1})\big]'(x)h_p,$$

for $h_1, \ldots, h_p \in \mathbb{R}^n$.

The function $(h_1, \ldots, h_p) \to f^{(p)}(x)(h_1, \ldots, h_p)$ is linear in all its arguments, and from the theorem of H.A. Schwartz one obtaines by induction that it is totally symmetric: For $1 \leq i \leq j \leq p$

$$f^{(p)}(x)(h_1, \ldots, h_i, \ldots, h_j, \ldots, h_p) = f^{(p)}(x)(h_1, \ldots, h_j, \ldots, h_i, \ldots, h_p)\,.$$

From the representation formula for the second derivatives one immediately obtains by induction for $h^{(j)} = (h_1^{(j)}, \ldots, h_n^{(j)}) \in \mathbb{R}^n$

$$f^{(p)}(x)(h^{(1)}, \ldots, h^{(p)}) = \sum_{i_1=1}^{n} \ldots \sum_{i_p=1}^{n} \frac{\partial^p f}{\partial x_{i_1} \ldots \partial x_{i_p}}(x) h_{i_1}^{(1)} \ldots h_{i_p}^{(p)}\,.$$

In accordance with Theorem 4.20, one says that f is p-times continuously differentiable, if f is p-times differentiable and the mapping $x \mapsto f^{(p)}(x)(h^{(1)}, \ldots, h^{(p)}) : U \to \mathbb{R}^m$ is continuous for all $h^{(1)}, \ldots, h^{(p)} \in \mathbb{R}^n$. By choosing in the above representation formula of

$f^{(p)}$ for $h^{(1)}, \ldots, h^{(p)}$ vectors from the standard basis $e_1, \ldots, e_n$ of $\mathbb{R}^n$, it is immediately seen that f is p-times continuously differentiable, if and only if all partial derivatives of f up to the order p exist and are continuous.

If $f^{(p)}$ exists for all $p \in \mathbb{N}$, then f is said to be infinitely differentiable. This happens if and only if all partial derivatives of any order exist in U.

Theorem 4.24 (Taylor formula) *Let U be an open subset of $\mathbb{R}^n$, let $f : U \to \mathbb{R}$ be $(p+1)$-times differentiable, and assume that the points x and $x+h$ together with the line segment connecting these points belong to U. Then there is a number ϑ with $0 < \vartheta < 1$ such that*

$$f(x+h) = f(x) + f'(x)h + \frac{1}{2!} f''(x)(h,h) + \ldots + \frac{1}{p!} f^{(p)}(x)(\underbrace{h, \ldots, h}_{p\text{-}times}) + R_p(x,h)\,,$$

where

$$R_p(x,h) = \frac{1}{(p+1)!}\, f^{(p+1)}(x+\vartheta h)(\underbrace{h, \ldots, h}_{p+1\text{-}times})\,.$$

Proof: Let $\gamma : [0,1] \to U$ be defined by $\gamma(t) = x + th$. To $F = f \circ \gamma : [0,1] \to \mathbb{R}$ apply the Taylor formula for real functions:

$$F(1) = \sum_{j=0}^{p} \frac{F^{(j)}(0)}{j!} + \frac{1}{(p+1)!}\, F^{(p+1)}(\vartheta)\,.$$

Insertion of the derivatives

$$\begin{aligned} F'(t) &= f'\big(\gamma(t)\big)\,\gamma'(t) = f'\big(\gamma(t)\big)h\,, \\ F''(t) &= f''\big(\gamma(t)\big)\,\big(h, \gamma'(t)\big) = f''\big(\gamma(t)\big)(h,h)\,, \\ &\vdots \\ F^{p+1}(t) &= f^{p+1}\big(\gamma(t)\big)\big(h, \ldots, \gamma'(t)\big) = f^{p+1}\big(\gamma(t)\big)(h, \ldots, h), \end{aligned}$$

into this formula yields the statement. ∎

Using the representation of $f^{(k)}$ by partial derivatives the Taylor formula can also be written as

$$\begin{aligned} f(x+h) &= \sum_{j=0}^{p} \frac{1}{j!}\Big[\sum_{i_1,\ldots,i_j=1}^{n} \frac{\partial^j f(x)}{\partial x_{i_1} \ldots \partial x_{i_j}}\, h_{i_1} \ldots h_{i_j}\Big] \\ &\quad + \frac{1}{(p+1)!} \sum_{i_1,\ldots,i_{p+1}=1}^{n} \frac{\partial^{p+1} f(x+\vartheta h)}{\partial x_{i_1} \ldots \partial x_{i_{p+1}}}\, h_{i_1} \ldots h_{i_{p+1}}\,. \end{aligned}$$

In this formula the notation can be simplified using multi-indices. For a multi-index $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}_0^n$ and for $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ set

$$\begin{aligned}
|\alpha| &:= \alpha_1 + \ldots + \alpha_n \qquad \text{(length of } \alpha\text{)} \\
\alpha! &:= \alpha_1! \ldots \alpha_n! \\
x^\alpha &:= x_1^{\alpha_1} \ldots x_n^{\alpha_n}\,, \\
D^\alpha f(x) &:= \frac{\partial^{|\alpha|} f(x)}{\partial^{\alpha_1} x_1 \ldots \partial^{\alpha_n} x_n}\,.
\end{aligned}$$

If α is a fixed multi-index with length $|\alpha| = j$, then the sum

$$\sum_{i_1,\ldots,i_j=1}^{n} \frac{\partial^j f(x)}{\partial x_{i_1} \ldots \partial x_{i_j}}\, h_{i_1} \ldots h_{i_j}$$

contains $\frac{j!}{\alpha_1! \ldots \alpha_n!}$ terms, which are obtained from $D^\alpha f(x) h^\alpha$ by interchanging the order, in which the derivatives are taken. Using this, the Taylor formula can be written in the compact form

$$\begin{aligned}
f(x+h) &= \sum_{j=0}^{p} \sum_{|\alpha|=j} \frac{1}{\alpha!} D^\alpha f(x) h^\alpha + \sum_{|\alpha|=p+1} \frac{1}{\alpha!} D^\alpha f(x+\vartheta h) h^\alpha \\
&= \sum_{|\alpha|\le p} \frac{1}{\alpha!} D^\alpha f(x) h^\alpha + \sum_{|\alpha|=p+1} \frac{1}{\alpha!} D^\alpha f(x+\vartheta h) h^\alpha\,.
\end{aligned}$$

5 Local extreme values, inverse function and implicit function

5.1 Local extreme values

Definition 5.1 *Let $U \subseteq \mathbb{R}^n$ be open, let $f : U \to \mathbb{R}$ be differentiable and let $a \in U$. If $f'(a) = 0$, then a is called critical point of f.*

Theorem 5.2 *Let $U \subseteq \mathbb{R}^n$ be open and let $f : U \to \mathbb{R}$ be differentiable. If f has a local extreme value at a, then a is a critical point of f.*

Proof: Without restriction of generality we assume that f has a local maximum at a. Then there is a neighborhood V of a such that $f(x) \le f(a)$ for all $x \in V$. Let $h \in \mathbb{R}^n$ and choose $\delta > 0$ small enough such that $a + th \in V$ for all $t \in \mathbb{R}$ with $|t| \le \delta$. Let $F : [-\delta, \delta] \to \mathbb{R}$ be defined by

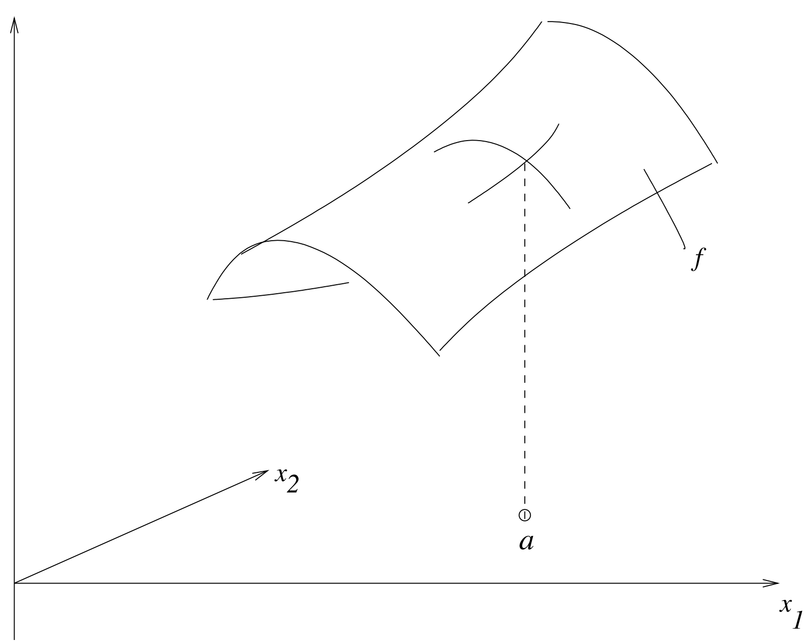
$$F(t) = f(a + th)\,.$$

Then F has a local maximum at $t = 0$, hence

$$0 = F'(0) = f'(a)h\,.$$

Since this holds for every $h \in \mathbb{R}^n$, it follows that $f'(a) = 0$. ■

Thus, if f has a local extreme value at a, then a is necessarily a critical point. For example, the saddle point a in the following picture is a critical point, but f has not an extreme value there.



This example shows that for functions of several variables the situation is more complicated than for functions of one variable. Still, also for functions of several variables the second derivative can be used to formulate a sufficient criterion for an extreme value. To this end some definitions and results for quadratic forms are needed, which we state without proof:

Definition 5.3 *Let $Q : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be a bilinear mapping. Then the mapping $h \to Q(h,h) : \mathbb{R}^n \to \mathbb{R}$ is called a quadratic form. A quadratic form is called*

(i) *positive definite, if $Q(h,h) > 0$ for all $h \neq 0$,*

(ii) *positive semi-definite, if $Q(h,h) \geq 0$ for all h,*

(iii) *negative definite, if $Q(h,h) < 0$ for all $h \neq 0$,*

(iv) *negative semi definite, if $Q(h,h) \leq 0$ for all h,*

(v) *indefinite, if $Q(h,h)$ has positive and negative values.*

To a quadratic form one can always find a symmetric coefficient matrix

$$C = \begin{pmatrix} c_{11} & \dots & c_{1n} \\ \vdots & & \\ c_{n1} & \dots & c_{nn} \end{pmatrix}$$

such that

$$Q(h,h) = \sum_{i,j=1}^{n} c_{ij} h_i h_j = h \cdot Ch \,.$$

From this representation it follows that for a quadratic form the mapping $h \mapsto Q(h,h) : \mathbb{R}^n \to \mathbb{R}$ is continuous. The quadratic form $Q(h,h)$ is positive definite, if

$$c_{11} > 0\,, \quad \det \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} > 0\,, \quad \det \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & x_{32} & c_{33} \end{pmatrix} > 0\,, \dots, \det(c_{ij})_{i,j=1,\dots,n} > 0\,.$$

If $f : U \to \mathbb{R}$ is two times differentiable at $x \in U$, then $(h,k) \mapsto f''(x)(h,k)$ is bilinear, hence $h \mapsto f''(x)(h,h)$ is a quadratic form. Since

$$f''(x)(h,h) = \sum_{i,j=1}^{n} \frac{\partial^2 f(x)}{\partial x_i \partial x_j} h_i h_j \,,$$

the coefficient matrix to this quadratic form is the Hessian matrix

$$H = \Big(\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\Big)_{i,j=1,\ldots,n}.$$

By the theorem of H.A. Schwarz, this matrix is symmetric.

Now we can formulate a sufficient criterion for extreme values:

Theorem 5.4 *Let $U \subseteq \mathbb{R}^n$ be open, let $f : U \to \mathbb{R}$ be two times continuously differentiable, and let $a \in U$ be a critical point of f. If the quadratic form $f''(a)(h,h)$*

(i) *is positive definite, then f has a local minimum at a,*

(ii) *is negative definite, then f has a local maximum at a,*

(iii) *is indefinite, then f does not have an extreme value at a.*

Proof: The Taylor formula yields

$$f(x) = f(a) + f'(a)(x-a) + \frac{1}{2} f''\big(a + \vartheta(x-a)\big)(x-a, x-a),$$

with a suitable $0 < \vartheta < 1$. Thus, since $f'(a) = 0$,

$$\begin{aligned} f(x) &= f(a) + \frac{1}{2} f''\big(a + \vartheta(x-a)\big)(x-a, x-a) \qquad (*) \\ &= f(a) + \frac{1}{2} f''(a)(x-a, x-a) + R(x)(x-a, x-a), \end{aligned}$$

with

$$\begin{aligned} R(x)(h,k) &= \frac{1}{2} f''\big(a + \vartheta(x-a)\big)(h,k) - \frac{1}{2} f''(a)(h,k) \\ &= \frac{1}{2} \sum_{i,j=1}^{n} \Big(\frac{\partial^2 f\big(a + \vartheta(x-a)\big)}{\partial x_i \partial x_j} - \frac{\partial^2 f(a)}{\partial x_i \partial x_j}\Big) h_j k_i. \end{aligned}$$

Since by assumption f is two times continuously differentiable, the second partial derivatives are continuous. Hence to every $\varepsilon > 0$ there is $\delta > 0$ such that for all $x \in U$ with $\|x-a\| < \delta$ and for all $1 \le i,\, j \le n$

$$\Big|\frac{\partial^2 f\big(a + \vartheta(x-a)\big)}{\partial x_i \partial x_j} - \frac{\partial^2 f(a)}{\partial x_i \partial x_j}\Big| < \frac{2}{n^2}\varepsilon .$$

Consequently, for $x \in U$ with $\|x-a\| < \delta$

$$|R(x)(h,h)| \le \frac{1}{2} \sum_{i,j=1}^{n} \frac{2}{n^2}\varepsilon \, \|h\|_\infty \, \|h\|_\infty \le \varepsilon c^2 \|h\|^2, \qquad (+)$$

where in the last step we used that there is a constant $c > 0$ with $\|h\|_\infty \leq c\,\|h\|$ for all $h \in \mathbb{R}^n$.

Assume now that $f''(a)(h,h) > 0$ is a positive definite quadratic form. Then $f''(a)(h,h) > 0$ for all $h \in \mathbb{R}^n$ with $h \neq 0$, and since the continuous mapping $h \mapsto f''(a)(h,h) : \mathbb{R}^n \to \mathbb{R}$ attains the minimum on the closed and bounded, hence compact set $\{h \in \mathbb{R}^n \bigm| \|h\| = 1\}$ at a point h_0 from this set, it follows for all $h \in \mathbb{R}^n$ with $h \neq 0$

$$f''(a)(h,h) = \|h\|^2 f''(a)\Big(\frac{h}{\|h\|}\,,\,\frac{h}{\|h\|}\Big) \geq \|h\|^2 \min_{\|\eta\|=1} f''(a)(\eta,\eta) = \kappa\,\|h\|^2$$

with

$$\kappa = f''(a)(h_0,h_0) > 0\,.$$

Now choose $\varepsilon = \frac{\kappa}{4c^2}$. Then this estimate and $(*), (+)$ yield that there is $\delta > 0$ such that for all $x \in U$ with $\|x-a\| < \delta$

$$\begin{aligned} f(x) - f(a) &= \frac{1}{2} f''(a)(x-a,x-a) + R(x)(x-a,x-a) \\ &\geq \frac{\kappa}{2}\,\|x-a\|^2 - \frac{\kappa}{4}\,\|x-a\|^2 = \frac{\kappa}{4}\|x-a\|^2 \geq 0. \end{aligned}$$

This means that f attains a local minimum at a.

In the same way one proves that a local maximum is attained at a if $f''(a)(h,h)$ is negative definite. If $f''(a)(h,h)$ is indefinite, there is $h_0 \in \mathbb{R}^n$, $k_0 \in \mathbb{R}^n$ with $\|h_0\| = \|k_0\| = 1$ and with

$$\kappa_1 := f''(a)(h_0,h_0) > 0\,, \quad \kappa_2 := f''(a)(k_0,k_0) < 1\,.$$

From these relations we conclude as above that for all points x on the straight line through a with direction vector h_0 sufficiently close to a the difference $f(x) - f(a)$ is positive, and for x on the straight line through a with direction vector k_0 sufficiently close to a the difference $f(x) - f(a)$ is negative. Thus, f does not attain an extreme value at a. ■

Example: Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x,y) = 6xy - 3y^2 - 2x^3$. All partial derivatives of all orders exist, hence f is infinitely differentiable. Therefore the assumptions of the Theorems 5.2 and 5.4 are satisfied. Thus, if (x,y) is a critical point, then

$$\operatorname{grad} f(x,y) = \begin{pmatrix} \dfrac{\partial f}{\partial x}(x,y) \\ \dfrac{\partial f}{\partial y}(x,y) \end{pmatrix} = \begin{pmatrix} 6y - 6x^2 \\ 6x - 6y \end{pmatrix} = 0\,,$$

which yields for the critical points $(x, y) = (0, 0)$ and $(x, y) = (1, 1)$.

To determine, whether these critical points are extremal points, the Hessian matrix must be computed at these points. The Hessian is

$$H(x,y) = \begin{pmatrix} \frac{\partial^2}{\partial x^2} f(x,y) & \frac{\partial^2}{\partial y \partial x} f(x,y) \\ \frac{\partial^2}{\partial x \partial y} f(x,y) & \frac{\partial^2}{\partial y^2} f(x,y) \end{pmatrix} = \begin{pmatrix} -12x & 6 \\ 6 & -6 \end{pmatrix}.$$

The quadratic form $f''(0,0)(h,h)$ defined by the Hessian matrix

$$H(0,0) = \begin{pmatrix} 0 & 6 \\ 6 & -6 \end{pmatrix}$$

is indefinite. For, if $h = (1, 1)$ then

$$f''(0,0)(h,h) = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 0 & 6 \\ 6 & -6 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 6 \\ 0 \end{pmatrix} = 6\,,$$

and if $h = (0, 1)$ then

$$f''(0,0)(h,h) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 0 & 6 \\ 6 & -6 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 6 \\ -6 \end{pmatrix} = -6\,,$$

Therefore $(0, 0)$ is not an extremal point of f. On the other hand, the quadratic form $f''(1,1)(h,h)$ defined by the matrix

$$H(1,1) = \begin{pmatrix} -12 & 6 \\ 6 & -6 \end{pmatrix}$$

is negative definite. For, by the criterion given above the matrix $-H(1,1)$ is positive definite since $12 > 0$ and

$$\det \begin{pmatrix} 12 & -6 \\ -6 & 6 \end{pmatrix} = 72 - 36 > 0\,.$$

Consequently $H(1, 1)$ is negative definite and $(1, 1)$ a local maximum of f.

5.2 Banach's fixed point theorem

In this section we state and prove the Banach fixed point theorem, a tool which we need in the later investigations and which has many important applications in mathematics.

Definition 5.5 *Let X be a set and let $d : X \times X \to \mathbb{R}$ be a mapping with the properties*

(i) $d(x,y) \geq 0\,, \quad d(x,y) = 0 \Leftrightarrow x = y$

(ii) $d(x,y) = d(y,x)$ *(symmetry)*

(iii) $d(x,y) \leq d(x,z) + d(z,y)$ *(triangle inequality)*

Then d is called a metric on X, and (X,d) is called a metric space. $d(x,y)$ is called the distance of x and y.

Examples 1.) Let X be a normed vector space. We denote the norm by $\|\cdot\|$. Then a metric is defined by $d(x,y) := \|x-y\|$. With this definition of the norm, every normed space becomes a metric space. In particular, $\mathbb{R}^n$ is a metric space.

2.) Let X be a nonempty set. We define a metric on X by

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

This metric is called degenerate.

3.) On $\mathbb{R}$ a metric is defined by

$$d(x,y) = \frac{|x-y|}{1+|x-y|}\,.$$

To see that this is a metric, note that the properties (i) and (ii) of Definition 5.5 are obviously satisfied. It remains to show that the triangle inequality holds. To this end note that $t \mapsto \frac{t}{1+t} : [0,\infty) \to [0,\infty)$ is strictly increasing, since $\frac{d}{dt}\frac{t}{1+t} = \frac{1}{1+t}(1-\frac{t}{1+t}) > 0$. Thus, for $x,y,z \in \mathbb{R}$

$$\begin{aligned} d(x,y) &= \frac{|x-y|}{1+|x-y|} \leq \frac{|x-z|+|z-y|}{1+|x-z|+|z-y|} \\ &\leq \frac{|x-z|}{1+|x-z|} + \frac{|z-y|}{1+|z-y|} = d(x,z) + d(z,y)\,. \end{aligned}$$

On a metric space X, a topology can be defined. For example, an ε-neighborhood $B_\varepsilon(x)$ of the point $x \in X$ is defined by

$$B_\varepsilon(x) = \left\{ y \in X \;\middle|\; d(x,y) < \varepsilon \right\}\,.$$

Based on this definition, open and closed sets and continuous functions between metric spaces can be defined. A subset of a metric space is called compact, if it has the Heine-Borel covering property.

Definition 5.6 *Let (X, d) be a metric space.*
(i) A sequence $\{x_n\}_{n=1}^{\infty}$ with $x_n \in X$ is said to converge, if $x \in X$ exists such that to every $\varepsilon > 0$ there is $n_0 \in \mathbb{N}$ with

$$d(x_n, x) < \varepsilon$$

for all $n \geq n_0$. The element x is called the limit of $\{x_n\}_{n=1}^{\infty}$.
(ii) A sequence $\{x_n\}_{n=1}^{\infty}$ with $x_n \in X$ is said to be a Cauchy sequence, if to every $\varepsilon > 0$ there is n_0 such that for all $n, k \geq n_0$

$$d(x_n, x_k) < \varepsilon \,.$$

Every converging sequence is a Cauchy sequence, but the converse is not necessarily true.

Definition 5.7 *A metric space (X, d) with the property that every Cauchy sequence converges, is called a complete metric space.*

Definition 5.8 *Let (X, d) be a metric space. A mapping $T : X \to X$ is said to be a contraction, if there is a number ϑ with $0 \leq \vartheta < 1$ such that for all $x, y \in X$*

$$d(Tx, Ty) \leq \vartheta d(x, y) \,.$$

Theorem 5.9 (Banach fixed point theorem) *Let (X, d) be a complete metric space and let $T : X \to X$ be a contraction. Then T possesses exactly one fixed point x, i.e. there is exactly one $x \in X$ such that*

$$Tx = x \,.$$

For arbitrary $x_0 \in X$ define the sequence $\{x_k\}_{k=1}^{\infty}$ by

$$\begin{aligned} x_1 &= Tx_0 \\ x_{k+1} &= Tx_k \,. \end{aligned}$$

Then

$$d(x, x_k) \leq \frac{\vartheta^k}{1 - \vartheta} \, d(x_1, x_0) \,,$$

hence

$$\lim_{k \to \infty} x_k = x \,.$$

Proof: First we show that T can have at most one fixed point. Let $x, y \in X$ be fixed points, hence $Tx = x$, $Ty = y$. Then

$$d(x,y) = d(Tx,Ty) \leq \vartheta d(x,y)\,,$$

which implies $(1-\vartheta)\,d(x,y) = 0$, whence $d(x,y) = 0$, and so $x = y$.

Next we show that a fixed point exists. Let $\{x_k\}_{k=1}^{\infty}$ be the sequence defined above. Then for $k \geq 1$

$$d(x_{k+1},x_k) = d(Tx_k,Tx_{k-1}) \leq \vartheta d(x_k,x_{k-1})\,.$$

The triangle inequality yields

$$d(x_{k+\ell},x_k) \leq d(x_{k+\ell},x_{k+\ell-1}) + d(x_{k+\ell-1},x_{k+\ell-2}) + \ldots + d(x_{k+1},x_k)\,,$$

thus

$$\begin{aligned} d(x_{k+\ell},x_k) &\leq (\vartheta^{\ell-1} + \vartheta^{\ell-2} + \ldots + \vartheta + 1)\,d(x_{k+1},x_k) \\ &\leq \frac{1-\vartheta^{\ell}}{1-\vartheta}\,\vartheta^k\,d(x_1,x_0) \leq \frac{\vartheta^k}{1-\vartheta}\,d(x_1,x_0)\,. \qquad (*) \end{aligned}$$

Since $\lim_{k\to\infty} \vartheta^k = 0$, if follows from this estimate that $\{x_k\}_{k=1}^{\infty}$ is a Cauchy sequence. Since the space X is complete, it has a limit x . For this limit we obtain

$$\begin{aligned} d(Tx,x) &= \lim_{k\to\infty} d(Tx,x) \\ &\leq \lim_{k\to\infty} \big[d(Tx,Tx_k) + d(Tx_k,x_{k+1}) + d(x_{k+1},x)\big] \\ &\leq \lim_{k\to\infty} \big[\vartheta d(x,x_k) + d(x_{k+1},x_{k+1}) + d(x_{k+1},x)\big] = 0\,, \end{aligned}$$

hence $Tx = x$, which shows that x is the uniquely determined fixed point. Moreover, $(*)$ yields

$$\begin{aligned} d(x,x_k) &= \lim_{\ell\to\infty} d(x,x_k) \\ &\leq \lim_{\ell\to\infty} \big[d(x,x_{k+\ell}) + d(x_{k+\ell},x_k)\big] \leq \frac{\vartheta^k}{1-\vartheta}\,d(x_1,x_0)\,. \end{aligned}$$

■

5.3 Local invertibility

Since $f'(a)$ is an approximation to f in a neighborhood of a, one can ask whether invertibility of $f'(a)$ (i.e. $\det f'(a) \neq 0$) already suffices to conclude that f is one–to–one in a neighborhood of a. The following example shows that in general this is not true:

Example: Let $f : (-1, 1) \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} x + 3x^2 \sin \dfrac{1}{x}, & x \neq 0 \\ 0, & x = 0. \end{cases}$$

f is differentiable for all $|x| < 1$ with derivative

$$f'(x) = \begin{cases} 1 + 6x \sin \dfrac{1}{x} - 3 \cos \dfrac{1}{x}, & x \neq 0 \\ 1, & x = 0. \end{cases}$$

In every neighborhood of 0 there are infinitely many intervals, which belong to $(0, \infty)$, and in which f' is continuous and has negative values. Thus, in such an interval one can find $0 < x_1 < x_2$ with $f(x_1) > f(x_2) > 0$. On the other hand, since f is continuous and satisfies $f(0) = 0$, the intermediate value theorem implies that the interval $(0, x_1)$ contains a point x_3 with $f(x_2) = f(x_3)$. Hence in no neighborhood of 0 the function f is one–to–one.

Since $f'(0) = 1$ and since in every neighborhood of 0 there are points x with $f'(x) < 0$, it follows that f' is not continuous at 0. Requiring that f' is continuous, changes the situation:

Theorem 5.10 *Let $U \subseteq \mathbb{R}^n$ be open, let $a \in U$, let $f : U \to \mathbb{R}^n$ be continuously differentiable, and assume that the derivative $f'(a)$ is invertible. Let $b = f(a)$. Then there is a neighborhood V of a and a neighborhood W of b, such that $f|_V : V \to W$ is bijective with a continuously differentiable inverse $g : W \to V$. (Clearly, $g'(y) = [f'(g(y))]^{-1}$.)*

Proof: We first assume that $a = 0$, $f(0) = 0$, hence $b = 0$, and $f'(0) = I$, where $I : \mathbb{R}^n \to \mathbb{R}^n$ is the identity mapping. It suffices to show that there is an open neighborhood W of 0 and a neighborhood W' of 0, such that every $y \in W$ has a unique inverse image under f in W'. Since f is continuous, it follows that $f^{-1}(W)$ is open, hence $V = f^{-1}(W) \cap W'$ is a neighborhood of 0, and $f : V \to W$ is invertible.

To construct W, we define for $y \in \mathbb{R}^n$ the mapping $\Phi_y : U \to \mathbb{R}^n$ by

$$\Phi_y(x) = x - f(x) + y.$$

Every fixed point x of this mapping is an inverse image of y under f. We choose $W = U_r(0)$ and show that if $r > 0$ is chosen sufficiently small, then for every $y \in U_r(0)$ the mapping Φ_y has a unique fixed point in the closed ball $W' = \overline{U_{2r}(0)}$.

This is guaranteed by the Banach fixed point theorem, if we can show that Φ_y maps $\overline{U_{2r}(0)}$ into itself and is a contraction on $\overline{U_{2r}(0)}$.

Note first that the continuity of f' implies that there is $r > 0$ such that for all $x \in \overline{U_{2r}(0)}$ with the operator norm

$$\|I - f'(x)\| = \|f'(0) - f'(x)\| \le \frac{1}{2},$$

whence

$$\|\Phi_y'(x)\| = \|I - f'(x)\| \le \frac{1}{2}.$$

For $x \in \overline{U_{2r}(0)}$ the line segment connecting this point to 0 is contained in $\overline{U_{2r}(0)}$, hence Corollary 4.14 yields for such x

$$\|x - f(x)\| = \|\Phi_y(x) - \Phi_y(0)\| \le \frac{1}{2}\|x\| \le r.$$

Thus, for $y \in U_r(0)$ and $x \in \overline{U_{2r}(0)}$,

$$\|\Phi_y(x)\| = \|x - f(x) + y\| \le \|x - f(x)\| + \|y\| \le 2r.$$

consequently, Φ_y maps $\overline{U_{2r}(0)}$ into itself. To prove that $\Phi_y : \overline{U_{2r}(0)} \to \overline{U_{2r}(0)}$ is a contraction for every $y \in U_r(0)$, we use again Corollary 4.14. Since for $x, z \in \overline{U_{2r}(0)}$ also the line segment connecting these points is contained in $\overline{U_{2r}(0)}$, it follows that

$$\|\Phi_y(x) - \Phi(z)\| \le \frac{1}{2}\|x - z\|.$$

Consequently, for every $y \in U_r(0)$ the mapping Φ_y is a contraction on the complete metric space $\overline{U_{2r}(0)}$, whence has a unique fixed point $x \in \overline{U_{2r}(0)}$. Since x is an inverse image of y under f, a local inverse $g : W \to V$ of f is defined by

$$g(y) = x.$$

We must show that g is continuously differentiable. Note first that if x_1 is a fixed point of Φ_{y_1} and x_2 is a fixed point of Φ_{y_2}, then

$$\begin{aligned} \|x_1 - x_2\| &= \|\Phi_{y_1}(x_1) - \Phi_{y_2}(x_2)\| \le \|\Phi_0(x_1) - \Phi_0(x_2)\| + \|y_1 - y_2\| \\ &\le \frac{1}{2}\|x_1 - x_2\| + \|y_1 - y_2\|, \end{aligned}$$

which implies

$$\|g(y_1) - g(y_2)\| = \|x_1 - x_2\| \le 2\|y_1 - y_2\|.$$

Hence, g is continuous. To verify that g is differentiable, note that $\det f'(x) \neq 0$ for all x in a neighborhood of 0, hence $f'(x)$ is invertible for all x from this neighborhood. To see this, remember that f' is continuous. By Theorem 4.20 this implies that the partial derivatives of f, which form the coefficients of the matrix $f'(x)$, depend continuously on x. Because $\det f'(x)$ consists of sums of products of these coefficients, it is a continuous function of x, hence it differs from zero in a neighborhood of 0, since $\det f'(0) = 1$. In this neighborhood $f'(x)$ is invertible. Therefore, since the inverse g is continuous, Theorem 4.12 implies that g is differentiable. Finally, from the formula

$$g'(y) = \Big[f'\big(g(y)\big) \Big]^{-1}$$

it follows that g' is continuous. Here we use that the coefficients of the inverse $(f'(x))^{-1}$ are determined via determinants (Cramer's rule), and thus depend continuously on the coefficients of $f'(x)$.

To prove the theorem for a function f with the properties stated in the theorem, consider the two affine invertible mappings $A, B : \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$\begin{aligned} Ax &= x + a \\ By &= \big(f'(a)\big)^{-1}(y - b). \end{aligned}$$

Then $H = B \circ f \circ A$ is defined in the open set $U - a = \{x - a \,\big|\, x \in U\}$ containing 0, $H(0) = (f'(a))^{-1}(f(a) - b) = 0$, and

$$H'(0) = B' f'(a) A' = \big(f'(a)\big)^{-1} f'(a) = I.$$

The preceding considerations show that neighborhoods V', W' of 0 exist such that $H : V' \to W'$ is invertible. Since $f = B^{-1} \circ H \circ A^{-1}$, it thus follows that f has the local inverse

$$g = A \circ H^{-1} \circ B : W \to V$$

with the neighborhoods $W = B^{-1}(W')$ of b and $V = A(V')$ of a. The local inverse H^{-1} is continuously differentiable, hence also g is continuously differentiable. ∎.

Example: Let $f : \mathbb{R}^3 \to \mathbb{R}^3$ be defined by

$$\begin{aligned} f_1(x_1, x_2, x_3) &= x_1 + x_2 + x_3\,, \\ f_2(x_1, x_2, x_3) &= x_2 x_3 + x_3 x_1 + x_1 x_2\,, \\ f_3(x_1, x_2, x_3) &= x_1 x_2 x_3. \end{aligned}$$

Since all partial derivatives exist and are continuous, it follows that f is continuously differentiable with

$$f'(x) = \begin{pmatrix} 1 & 1 & 1 \\ x_3 + x_2 & x_3 + x_1 & x_2 + x_1 \\ x_2 x_3 & x_1 x_3 & x_1 x_2 \end{pmatrix},$$

hence

$$\begin{aligned} \det f'(x) &= \begin{vmatrix} 1 & 0 & 0 \\ x_3 + x_2 & x_1 - x_2 & x_1 - x_3 \\ x_2 x_3 & (x_1 - x_2)x_3 & (x_1 - x_3)x_2 \end{vmatrix} \\ &= (x_1 - x_2)(x_1 - x_3)x_2 - (x_1 - x_2)(x_1 - x_3)x_3 \\ &= (x_1 - x_2)(x_1 - x_3)(x_2 - x_3)\,. \end{aligned}$$

Thus, let $b = f(a)$ with $(a_1 - a_2)(a_1 - a_3)(a_2 - a_3) \neq 0$. Then there are neighborhoods V of a and W of b, such that the system of equations

$$\begin{aligned} y_1 &= x_1 + x_2 + x_3 \\ y_2 &= x_2 x_3 + x_3 x_1 + x_1 x_2 \\ y_3 &= x_1 x_2 x_3 \end{aligned}$$

has a unique solution $x \in V$ to every $y \in W$.

We remark that the local invertibility does not imply global invertibility. One can see this at the following example: Let $f : \{(x, y) \in \mathbb{R}^2 \mid y > 0\} \to \mathbb{R}^2$ be defined by

$$\begin{aligned} f_1(x, y) &= y \cos x \\ f_2(x, y) &= y \sin x\,. \end{aligned}$$

f is continuously differentiable with

$$\det f'(x, y) = \begin{vmatrix} -y \sin x & \cos x \\ y \cos x & \sin x \end{vmatrix} = -y \sin^2 x - y \cos^2 x = -y \neq 0$$

for all (x, y) from the domain of definition. Consequently f is locally invertible at every point. Yet, f is not globally invertible, since f is 2π-periodic with respect to the x variable.

5.4 Implicit functions

Let a function $f : \mathbb{R}^{n+m} \to \mathbb{R}^n$ be given with the components $f_1, \ldots, f_n$, and let $y = (y_1, \ldots, y_m)$ be given. Can one determine $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ such that the equations

$$\begin{aligned} f_1(x_1, \ldots, x_n, y_1, \ldots, y_m) &= 0 \\ \vdots & \\ f_n(x_1, \ldots, x_n, y_1, \ldots, y_m) &= 0 \end{aligned}$$

hold? These are n equations for n unknowns $x_1, \ldots, x_n$. First we study the situation for a linear function $f = A : \mathbb{R}^{n+m} \to \mathbb{R}^n$,

$$A(x,y) = \begin{pmatrix} A_1(x,y) \\ \vdots \\ A_n(x,a) \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + \ldots + a_{1n}x_n + b_{11}y_1 + b_{1m}y_m \\ \vdots \\ a_{n1}x_1 + \ldots + a_{nn}x_n + b_{n1}y_1 + b_{nm}y_m \end{pmatrix}.$$

Suppose that A has the property

$$A(h,0) = 0 \Rightarrow h = 0\,.$$

A has this property, if and only if the matrix

$$\begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \\ a_{n1} & \ldots & a_{nn} \end{pmatrix} = \begin{pmatrix} \frac{\partial A_1}{\partial x_1} & \ldots & \frac{\partial A_1}{\partial x_n} \\ \vdots & & \\ \frac{\partial A_n}{\partial x_1} & \ldots & \frac{\partial A_n}{\partial x_n} \end{pmatrix}$$

is invertible, hence if and only if

$$\det\Big(\frac{\partial A_j}{\partial x_i}\Big)_{i,j=1,\ldots,n} \neq 0\,.$$

Under this condition the mapping

$$h \mapsto Ch := A(h,0) : \mathbb{R}^n \to \mathbb{R}^n$$

is invertible, consequently the system of equations

$$A(h,k) = A(h,0) + A(0,k) = Ch + A(0,k) = 0$$

has for every $k \in \mathbb{R}^m$ the unique solution

$$h = \varphi(k) := -C^{-1}A(0,k)\,.$$

For $\varphi : \mathbb{R}^m \to \mathbb{R}^n$ one has

$$A\big(\varphi(k), k\big) = 0\,,$$

for all $k \in \mathbb{R}^m$. One says that the function φ is implicitly given by this equation.

The theorem about implicit functions concerns the same situation for continuously differentiable functions f, which are not necessarily linear:

Theorem 5.11 (about implicit functions) *Let $D \subseteq \mathbb{R}^{n+m}$ be open and let $f : D \to \mathbb{R}^n$ be continuously differentiable. Suppose that there is $a \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ with $(a,b) \in D$, such that $f(a,b) = 0$ and*

$$\det \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a,b) & \dots & \frac{\partial f_1}{\partial x_n}(a,b) \\ \vdots & & \\ \frac{\partial f_n}{\partial x_1}(a,b) & \dots & \frac{\partial f_n}{\partial x_n}(a,b) \end{pmatrix} \neq 0 \,. \tag{$*$}$$

Then there is a neighborhood $U \subseteq \mathbb{R}^m$ of b and a uniquely determined continuously differentiable function $\varphi : U \to \mathbb{R}^n$ such that $\varphi(b) = a$ and for all $y \in U$

$$f\big(\varphi(y), y\big) = 0 \,.$$

Proof: Consider the mapping $F : D \to \mathbb{R}^{n+m}$,

$$F(x,y) = \big(f(x,y), y\big) \in \mathbb{R}^{n+m} \,.$$

Then

$$F(a,b) = \big(f(a,b), b\big) = (0,b) \,.$$

Since f is continuously differentiable, all the partial derivatives of F exist and are continuous in D, hence F is continuously differentiable in D. The derivative $F'(a,b)$ is given by

$$F'(a,b) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_1}{\partial x_n} & \frac{\partial f_1}{\partial y_1} & \dots & \frac{\partial f_1}{\partial y_m} \\ \vdots & & & & & \\ \frac{\partial f_n}{\partial x_1} & \dots & \frac{\partial f_n}{\partial x_n} & \frac{\partial f_n}{\partial y_1} & \dots & \frac{\partial f_n}{\partial y_m} \\ 0 & \dots & 0 & 1 & \dots & 0 \\ \vdots & & & \vdots & & \\ 0 & \dots & 0 & 0 & & 1 \end{pmatrix} ,$$

where the partial derivatives are computed at (a,b). Thus, for $h \in \mathbb{R}^n$, $k \in \mathbb{R}^m$

$$F'(a,b)(h,k) = \big(f'(a,b)(h,k), k\big) \,.$$

This linear mapping is invertible. For, if

$$F'(a,b)(h,k) = \big(f'(a,b)(h,k), k\big) = 0 \,,$$

then $k = 0$, therefore $f'(a,b)(h,0) = 0$, which together with $(*)$ yields $h = 0$. Consequently the null space of the linear mapping $F'(a,b)$ consists only of the set $\{0\}$, hence it is invertible.

Therefore the assumptions of Theorem 5.10 are satisfied, and it follows that there are neighborhoods V of (a,b) and W of $(0,b)$ in $\mathbb{R}^{n+m}$ such that

$$F\big|_V : U \to W$$

is invertible. The inverse $F^{-1} : W \to V$ is of the form

$$F^{-1}(z,w) = \big(\phi(z,w), w\big),$$

with a continuously differentiable function $\phi : W \to \mathbb{R}^n$. Now set

$$U = \{w \in \mathbb{R}^m \mid (0,w) \in W\} \subseteq \mathbb{R}^m$$

and define $\varphi : U \to \mathbb{R}^n$ by

$$\varphi(w) = \phi(0,w)\,.$$

U is a neighborhood of b since W is a neighborhood of $(0,b)$, and for all $w \in U$

$$(0,w) = F\big(F^{-1}(0,w)\big) = F\big(\phi(0,w), w\big) = F\big(\varphi(w), w\big) = \big(f\big(\varphi(w), w\big), w\big)\,,$$

whence

$$f\big(\varphi(w), w\big) = 0\,.$$

∎

The derivative of the function φ can be computed using the chain rule: For the derivative $\frac{d}{dy} f\big(\varphi(y), y\big)$ of the function $y \mapsto f\big(\varphi(y), \varphi\big)$ we obtain

$$\begin{aligned} 0 &= \frac{d}{dy} f\big(\varphi(y), y\big) = \big(\frac{\partial}{\partial x} f\,,\ \frac{\partial}{\partial y} f\big)\big(\varphi(y), y\big) \begin{pmatrix} \varphi'(y) \\ id_{\mathbb{R}^m} \end{pmatrix} \\ &= \big(\frac{\partial}{\partial x} f\big)\big(\varphi(y), y\big) \circ \varphi'(y) + \big(\frac{\partial}{\partial y} f\big)\big(\varphi(y), y\big)\,. \end{aligned}$$

Thus,

$$\varphi'(y) = -\Big[\big(\frac{\partial}{\partial x} f\big)\big(\varphi(y), y\big)\Big]^{-1} \circ \big(\frac{\partial}{\partial y} f\big)\big(\varphi(y), y\big)\,.$$

Here we have set

$$\begin{aligned} \frac{\partial}{\partial x} f(x,y) &= \big(\frac{\partial f_j}{\partial x_i}(x,y)\big)_{i,j=1,\ldots,n} \\ \frac{\partial}{\partial y} f(x,y) &= \big(\frac{\partial f_j}{\partial y_i}(x,y)\big)_{j=1,\ldots,n\,,\, i=1,\ldots,m}\,. \end{aligned}$$

Examples:

1.) Let an equation

$$f(x_1,\dots,x_n)=0$$

be given with continuously differentiable $f:\mathbb{R}^n\to\mathbb{R}$. To given $x_1,\dots,x_{n-1}$ we seek x_n such that this equation is satisfied, i.e. we want to solve this equation for x_n. Assume that $a=(a_1,\dots,a_n)\in\mathbb{R}^n$ is given such that

$$f(a_1,\dots,a_n)=0$$

and

$$\frac{\partial f}{\partial x_n}(a_1,\dots,a_n)\neq 0\,.$$

Then the implicit function theorem implies that there is a neighborhood $U\subseteq\mathbb{R}^{n-1}$ of $(a_1,\dots,a_{n-1})$, such that to every $(x_1,\dots,x_{n-1})\in U$ a unique $x_n=\varphi(x_1,\dots,x_{n-1})$ can be found, which solves the equation

$$f(x_1,\dots,x_{n-1},x_n)=0\,,$$

and which is a continuously differentiable function of $(x_1,\dots,x_{n-1})$ and satisfies $x_n=a_n$ for $(x_1,\dots,x_{n-1})=(a_1,\dots,a_{n-1})$. For the derivative of the function φ one obtains

$$\operatorname{grad}\varphi(x_1,\dots,x_{n-1})=\frac{-1}{\frac{\partial}{\partial x_n}f(x_1,\dots,x_n)}\operatorname{grad}_{n-1}f(x_1,\dots,x_n)=\frac{-1}{\frac{\partial f}{\partial x_n}}\begin{pmatrix}\frac{\partial}{\partial x_1}f\\ \vdots\\ \frac{\partial}{\partial x_{n-1}}f\end{pmatrix},$$

where $x_n=\varphi(x_1,\dots x_{n-1})$.

2.) Let $f:\mathbb{R}^3\to\mathbb{R}^2$ be defined by

$$\begin{aligned}f_1(x,y,z)&=3x^2+xy-z-3\\ f_2(x,y,z)&=2xz+y^3+xy\,.\end{aligned}$$

We have $f(1,0,0)=0$. To given $z\in\mathbb{R}$ from a neighborhood of 0 we seek $(x,y)\in\mathbb{R}^2$ such that $f(x,y,z)=0$. To this end we must test, whether the matrix

$$\begin{pmatrix}\frac{\partial f_1}{\partial x}(x,y,z) & \frac{\partial f_1}{\partial y}(x,y,z)\\ \frac{\partial f_2}{\partial x}(x,y,z) & \frac{\partial f_2}{\partial y}(x,y,z)\end{pmatrix}=\begin{pmatrix}6x+y & x\\ 2z+y & 3y^2+x\end{pmatrix}$$

is invertible at $(x,y,z)=(1,0,0)$. At this point, the determinant of this matrix is

$$\begin{vmatrix}6 & 1\\ 0 & 1\end{vmatrix}=6\neq 0\,,$$

hence the matrix is invertible. Consequently, a sufficiently small number $\delta > 0$ and a continuously differentiable function $\varphi : (-\delta, \delta) \to \mathbb{R}^2$ with $\varphi(0) = (1,0)$ can be found such that $f\big(\varphi_1(z), \varphi_2(z), z\big) = 0$ for all z with $|z| < \delta$. For the derivative of φ we obtain with $(x,y) = \varphi(z)$

$$\begin{aligned}
\varphi'(z) &= -\begin{pmatrix} 6x+y & x \\ 2z+y & 3y^2+x \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial f_1}{\partial z}(x,y,z) \\ \frac{\partial f_2}{\partial z}(x,y,z) \end{pmatrix} \\
&= \frac{-1}{(6x+y)(3y^2+x) - x(2z+y)} \begin{pmatrix} 3y^2+x & -x \\ -(2z+y) & 6x+y \end{pmatrix} \begin{pmatrix} -1 \\ 2x \end{pmatrix} \\
&= \frac{-1}{(6x+y)(3y^2+x) - x(2z+y)} \begin{pmatrix} -3y^2 - x - 2x^2 \\ 2z + y + 12x^2 + 2xy \end{pmatrix} .
\end{aligned}$$

Since $\varphi(0) = (1,0)$, we obtain in particular

$$\varphi'(0) = -\frac{1}{6} \begin{pmatrix} -3 \\ 12 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ -2 \end{pmatrix} .$$

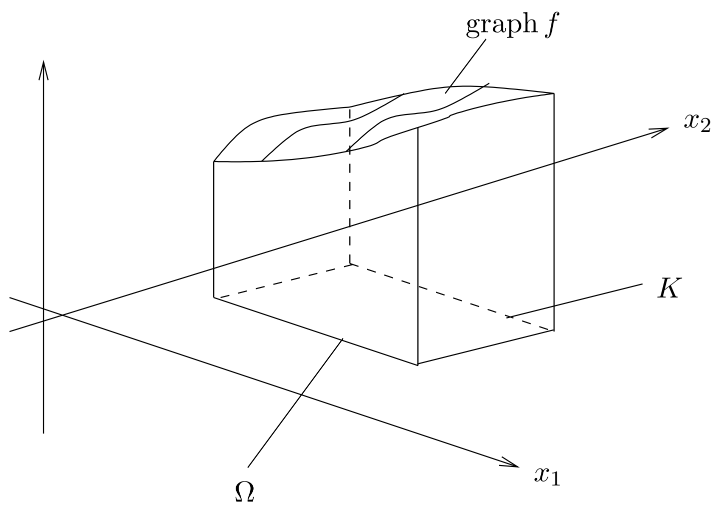
6 Integration of functions of several variables

6.1 Definition of the integral

Let Ω be a bounded subset of $\mathbb{R}^2$ and let $f : \Omega \to \mathbb{R}$ be a real valued function. If f is continuous, then graph f is a surface in $\mathbb{R}^3$. We want to define the integral

$$\int_\Omega f(x)dx$$

such that its value is equal to the volume of the subset K of $\mathbb{R}^3$, which lies between the graph of f and the x_1, x_2-plane. More generally, we want to define integrals for functions defined on $\mathbb{R}^n$, such that for $n = 2$ the integral has this property.



Definition 6.1 Let

$$Q = \{x \in \mathbb{R}^n \mid a_i \leq x_i < b_i,\ i = 1, \ldots, n\}$$

be a bounded, half open interval in $\mathbb{R}^n$. A partition P of Q is a cartesian product

$$P = P_1 \times \ldots \times P_n,$$

where $P_i = \{x_0^{(i)}, \ldots, x_{k_i}^{(i)}\}$ is a partition of $[a_i, b_i]$, for every $i = 1, \ldots, n$.

Q is partitioned into $k = k_1 \cdot k_2 \ldots k_n$ half open subintervals $Q_1, \ldots Q_k$ of the form

$$Q_j = [x_{p_1}^{(1)}, x_{p_1+1}^{(1)}) \times \ldots \times [x_{p_n}^{(n)}, x_{p_n+1}^{(n)}).$$

The number

$$|Q_j| = (x_{p_1+1}^{(1)} - x_{p_1}^{(1)}) \ldots (x_{p_n+1}^{(n)} - x_{p_n}^{(n)})$$

is called measure of Q_j. For a bounded function $f : Q \to \mathbb{R}$ define

$$\begin{aligned} M_j &= \sup f(Q_j), & m_j &= \inf f(Q_j), \\ U(p, f) &= \sum_{j=1}^{k} M_j |Q_j|, & L(P, f) &= \sum_{j=1}^{k} m_j |Q_j|. \end{aligned}$$

The upper and lower Darboux integrals are

$$\begin{aligned} \overline{\int_Q} f\,dx &= \inf\{U(P, f) \mid P \text{ is a partition of Q}\}, \\ \underline{\int_Q} f\,dx &= \sup\{L(P, f) \mid P \text{ is a partition of } Q\}. \end{aligned}$$

Definition 6.2 A bounded function $f : Q \to \mathbb{R}$ is called Riemann integrable, if the upper and lower Darboux integrals coincide. The common value is denoted by

$$\int_Q f\,dx \quad \text{or} \quad \int_Q f(x)dx$$

and is called the Riemann integral of f.

To define the integral on more general domains, let $\Omega \subseteq \mathbb{R}^n$ be a bounded subset and let $f : \Omega \to \mathbb{R}$. Choose a bounded interval Q such that $\Omega \subseteq Q$ and extend f to a function $f_Q : Q \to \mathbb{R}$ by

$$f_Q(x) = \begin{cases} f(x), & x \in \Omega, \\ 0, & x \in Q \backslash \Omega. \end{cases}$$

Definition 6.3 A bounded function $f : \Omega \to \mathbb{R}$ is called Riemann integrable over Ω if the extension f_Q is integrable over Q. We set

$$\int_\Omega f(x)dx = \int_Q f_Q(x)\,dx.$$

The multi-dimensional integral shares most of the properties with the one-dimensional integral. We do not repeat the proofs, since they are almost the same. Differences arise mainly from the more complicated structure of the domain of integration. Whether a function is integrable over a domain Ω depends not only on the properties of the function but also on the properties of Ω.

Definition 6.4 A bounded set $\Omega \subseteq \mathbb{R}^n$ is called Jordan-measurable, if the characteristic function $\chi_\Omega : \mathbb{R}^n \to \mathbb{R}$ defined by

$$\chi_\Omega(x) = \begin{cases} 1, & x \in \Omega \\ 0, & x \in \mathbb{R}^n \backslash \Omega \end{cases}$$

is integrable. In this case $|\Omega| = \int_\Omega 1\,dx$ is called the Jordan measure of Ω.

Of course, a bounded interval $Q \subseteq \mathbb{R}^n$ is measurable, and the previously given definition of $|Q|$ coincides with the new definition.

Theorem 6.5 *If the compact domain $\Omega \subseteq \mathbb{R}^n$ is Jordan measurable and if $f : \Omega \to \mathbb{R}$ is continuous, then f is integrable.*

A **proof** of this theorem can be found in the book "Lehrbuch der Analysis, Teil 2" of H. Heuser, p. 455.

6.2 Limits of integrals, parameter dependent integrals

Theorem 6.6 *Let $\Omega \subseteq \mathbb{R}^n$ be a bounded set and let $\{f_k\}_{k=1}^{\infty}$ be a sequence of Riemann integrable functions $f_k : \Omega \to \mathbb{R}$, which converges uniformly to a Riemann integrable function $f : \Omega \to \mathbb{R}$. Then*

$$\lim_{k\to\infty} \int_{\Omega} f_k(x)dx = \int_{\Omega} f(x)dx.$$

Remark It can be shown that the uniform limit f of a sequence of integrable functions is automatically integrable.

Proof Let $\varepsilon > 0$. Then there is $k_0 \in \mathbb{N}$ such that for all $k \geq k_0$ and all $x \in \Omega$ we have

$$|f_k(x) - f(x)| < \varepsilon,$$

hence

$$\Big|\int_{\Omega} \big(f_k(x) - f(x)\big)dx\Big| \leq \int_{Q} |f_k(x) - f(x)|dx \leq \int_{Q} \varepsilon\, dx \leq \varepsilon|Q|.$$

By definition, this means that $\lim_{k\to\infty} \int_{\Omega} f_k(x)dx = \int_{\Omega} f(x)dx$. ■

Corollary 6.7 *Let $D \subseteq \mathbb{R}^k$ and let $Q \subseteq \mathbb{R}^m$ be a bounded interval. If $f : D \times \overline{Q} \to \mathbb{R}$ is continuous, then the function $F : D \to \mathbb{R}$ defined by the parameter dependent integral*

$$F(x) = \int_{Q} f(x,t)dt$$

is continuous.

Proof Let $x_0 \in D$ and let $\{x_k\}_{k=1}^{\infty}$ be a sequence with $x_k \in D$ and $\lim_{k\to\infty} x_k = x_0$. Then x_0 is the only accumulation point of the set $M = \{x_k \mid k \in \mathbb{N}\} \cup \{x_0\}$, from which it is immediately seen that $M \times \overline{Q}$ is closed and bounded, hence it is a compact subset of $D \times \overline{Q}$. Therefore the continuous function f is uniformly continuous on $M \times \overline{Q}$. This implies that to every $\varepsilon > 0$ there is $\delta > 0$ such that for all $y \in M$ with $|y - x_0| < \delta$ and all $t \in Q$ we have

$$|f(y,t) - f(x_0,t)| < \varepsilon.$$

Choose $k_0 \in \mathbb{N}$ such that $|x_k - x_0| < \delta$ for all $k \geq k_0$. This implies for $k \geq k_0$ and for all $t \in Q$ that

$$|f(x_k,t) - f(x_0,t)| < \varepsilon,$$

which shows that the sequence $\{f_k\}_{k=1}^{\infty}$ of continuous functions $f_k : Q \to \mathbb{R}$ defined by $f_k(t) = f(x_k,t)$ converges uniformly to the continuous function $f_\infty(t) = f(x_0,t)$. Theorem 6.6 implies

$$\lim_{k\to\infty} F(x_k) = \lim_{k\to\infty} \int_Q f(x_k,t)dt = \int_Q f(x,t)dx = F(x).$$

Therefore F is continuous. ∎

6.3 The Theorem of Fubini

The computation of integrals by approximation of the integrand by step functions is impractible. For onedimensional integrals the computation is strongly simplified by the Fundamental Theorem of Calculus. In this section we show that multidimensional integrals can be computed as iterated onedimensional integrals, which makes also the computation of these integrals practicle.

We first consider integrals of step functions. Let

$$\begin{aligned} Q &= \{x \in \mathbb{R}^n \mid a_i \leq x_i < b_i,\ i = 1,\ldots,n\} \\ Q' &= \{x' \in \mathbb{R}^{n-1} \mid a_i \leq x_i < b_i,\ i = 1,\ldots,n-1\} \end{aligned}$$

be half open intervals. If

$$P = P_1 \times P_2 \times \ldots \times P_n$$

is a partition of Q, then $P' = P_1 \times \ldots \times P_{n-1}$ is a partition of Q'. Let $Q'_1, \ldots, Q'_k$ be the subintervals of Q' generated by P' and let $I_1, \ldots, I_{k'} \subseteq [a_n, b_n)$ be the half open subintervals generated by P_n. Then all the subintervals of Q generated by P are given by

$$Q'_j \times I_\ell, \quad 1 \leq j \leq k,\ 1 \leq \ell \leq k'.$$

For the characteristic functions $\chi_{Q'_j \times I_\ell}$ and the measures $|Q'_j \times I_\ell|$ we have

$$\chi_{Q'_j \times I_\ell}(x) = \chi_{Q'_j}(x')\chi_{I_\ell}(x_n) \quad \text{and} \quad |Q'_j \times I_\ell| = |Q'_j|\,|I_\ell|.$$

Let $s : Q \to \mathbb{R}$ be a step function of the form

$$s(x) = \sum_{\substack{j=1,\ldots,k\\ \ell=1,\ldots,k'}} r_{j\ell}\, \chi_{Q'_j \times I_\ell}(x) = \sum_{\ell=1}^{k'} \Big(\sum_{j=1}^{k} r_{j\ell}\, \chi_{Q'_j}(x') \Big) \chi_{I_\ell}(x_n),$$

with given numbers $r_{j\ell} \in \mathbb{R}$. The last equality shows that for every fixed $x_n \in [a_n, b_n)$ the function $x' \mapsto s(x', x_n)$ is a step function on Q' with integral

$$\int_{Q'} s(x', x_n)\, dx' = \sum_{\ell=1}^{k'} \Big(\sum_{j=1}^{k} r_{j\ell}\, |Q'_j| \Big) \chi_{I_\ell}(x_n),$$

and this formula shows that $x_n \mapsto \int_{Q'} s(x', x_n)\, dx'$ is a step function on $[a_n, b_n)$. For the integral of this step function over the interval $[a_n, b_n)$ we thus find

$$\begin{aligned}
\int_{a_n}^{b_n} \int_{Q'} s(x', x_n)\, dx'\, dx_n &= \sum_{\ell=1,\ldots,k'} \Big(\sum_{j=1,\ldots,k} r_{j\ell}\, |Q'_j| \Big) |I_\ell| \\
&= \sum_{\substack{j=1,\ldots,k\\ \ell=1,\ldots,k'}} r_{j\ell}\, |Q'_j \times I_\ell| = \int_Q s(x)\, dx.
\end{aligned}$$

For step functions the n-dimensional integral $\int_Q s(x)\, dx$ can thus be computed as an iterated integral. This is also true for continuous functions:

Theorem 6.8 (Guido Fubini, 1879 – 1943) *Let*

$$\begin{aligned}
Q &= \{x \in \mathbb{R}^n \mid a_i \le x_i < b_i,\ i = 1, \ldots, n\} \\
Q' &= \{x' \in \mathbb{R}^{n-1} \mid a_i \le x_i < b_i,\ i = 1, \ldots, n-1\}.
\end{aligned}$$

Then for every continuous function $f : \overline{Q} \to \mathbb{R}$ the function $F : [a_n, b_n] \to \mathbb{R}$ defined by

$$F(x_n) = \int_{Q'} f(x', x_n) dx'$$

is integrable and

$$\int_Q f(x) dx = \int_{a_n}^{b_n} F(x_n) dx_n = \int_{a_n}^{b_n} \int_{Q'} f(x', x_n) dx' dx_n. \tag{6.1}$$

Proof By Corollary 6.7 the function F is continuous, whence it is integrable. To verify (6.1) we approximate f by step functions. Choose a sequence of partitions $\{P^{(\ell)}\}_{\ell=1}^{\infty}$ of Q such that

$$\sup_{j=1,\ldots,j_\ell} \operatorname{diam}(Q_j^{(\ell)}) \le \frac{1}{\ell}\,, \tag{6.2}$$

where $Q_1^{(\ell)}, \dots Q_{j_\ell}^{(\ell)}$ are the subintervals of Q generated by the partition $P^{(\ell)}$. Choose $x_j^{(\ell)} \in Q_j^{(\ell)}$ and define step functions $s_\ell : Q \to \mathbb{R}$ by

$$s_\ell(x) = \sum_{j=1}^{j_\ell} f(x_j^{(\ell)}) \chi_{Q_j^{(\ell)}}(x).$$

The sequence $\{s_\ell\}_{\ell=1}^\infty$ converges uniformly to f. To verify this, note that the continuous function f is uniformly continuous on the compact set $\overline{Q}$. It thus follows that to given $\varepsilon > 0$ there is $\delta > 0$ such that $|f(x) - f(y)| < \varepsilon$ for all $x, y \in Q$ satisfying $|x - y| < \delta$. Choose ℓ_0 with $1/\ell_0 < \delta$. For every $\ell \geq \ell_0$ and every $x \in Q$ there is exactly one number j such that $x \in Q_j^\ell$. From (6.2) we thus conclude that $|x - x_j^\ell| \leq \operatorname{diam}(Q_j^{(\ell)}) \leq 1/\ell \leq 1/\ell_0 < \delta$, hence

$$|f(x) - s_\ell(x)| = |f(x) - f(x_j^\ell)| < \varepsilon.$$

This inequality shows that indeed $\{s_\ell\}_{\ell=1}^\infty$ converges uniformly to f, since ℓ_0 is independent of $x \in Q$. Therefore Theorem 6.6 can be applied. We find that

$$\lim_{\ell \to \infty} \int_Q s_\ell(x) dx = \int_Q f(x) dx. \tag{6.3}$$

Moreover, for the step function $S_\ell : [a_n, b_n] \to \mathbb{R}$ defined by

$$S_\ell(x_n) = \int_{Q'} s_\ell(x', x_n) dx'$$

it follows that

$$|F(x_n) - S_\ell(x_n)| \leq \int_{Q'} |f(x', x_n) - s_\ell(x', x_n)| \, dx' \leq \sup_{y \in Q} |f(y) - s_\ell(y)| \, |Q'|.$$

The right hand side is independent of x_n and converges to zero for $\ell \to \infty$, hence $\{S_\ell\}_{\ell=1}^\infty$ converges to F uniformly on $[a_n, b_n]$. Consequently, Theorem 6.6 implies

$$\lim_{\ell \to \infty} \int_{a_n}^{b_n} S_\ell(x_n) dx_n = \int_{a_n}^{b_n} F(x_n) dx_n.$$

Since (6.1) holds for step functions, it follows from this equation and from (6.3) that

$$\begin{aligned} \int_{a_n}^{b_n} F(x_n) dx_n &= \lim_{\ell \to \infty} \int_{a_n}^{b_n} S_\ell(x_n) dx_n \\ &= \lim_{\ell \to \infty} \int_{a_n}^{b_n} \int_{Q'} s_\ell(x', x_n) dx' dx_n = \lim_{\ell \to \infty} \int_Q s_\ell(x) dx = \int_Q f(x) dx. \end{aligned}$$

■

Remarks By repeated application of this theorem we obtain that

$$\int_Q f(x) dx = \int_{a_n}^{b_n} \dots \int_{a_1}^{b_1} f(x_1, \dots, x_n) dx_1 \dots dx_n.$$

It is obvious from the proof that in the Theorem of Fubini the coordinate x_n can be replaced by any other coordinate. Therefore the order of integration in the iterated integral can be replaced by any other order.

The Theorem of Fubini holds not only for continuous functions, but for any integrable function. In the general case both the formulation of the theorem and the proof are more complicated.

6.4 The transformation formula

The transformation formula generalizes the rule of substitution for one-dimensional integrals. We start with some preparations.

Definition 6.9 (i) Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuous. The support of f is defined by

$$\operatorname{supp} f = \overline{\{x \in \mathbb{R}^n \mid f(x) \neq 0\}}.$$

(ii) Let $D \subseteq \mathbb{R}^n$ and let $\{U_i\}_{i=1}^{\infty}$ be an open covering of D. For every $i \in \mathbb{N}$ let $\varphi_i : \mathbb{R}^n \to \mathbb{R}$ be a continuous function with compact support contained in U_i, such that

$$\sum_{i=1}^{\infty} \varphi_i(x) = 1, \quad \text{for all } x \in D.$$

Then $\{\varphi_i\}_{i=1}^{\infty}$ is called partition of unity on D subordinate to the covering $\{U_i\}_{i=1}^{\infty}$.

Theorem 6.10 *Let $D \subseteq \mathbb{R}^n$ be a compact set and let $B_{r_1}(z_1), \ldots, B_{r_m}(z_m)$ be open balls in $\mathbb{R}^n$ with $D \subseteq B_{r_1}(z_1) \cup \ldots \cup B_{r_m}(z_m)$. Then there is a partition of unity $\{\varphi_i\}_{i=1}^{\infty}$ on D subordinate to the covering $\{B_{r_i}(z_i)\}_{i=1}^{m}$.*

Proof: Let $C = \mathbb{R}^n \setminus \bigcup_{i=1}^{m} B_{r_i}(z_i)$. The distance $\operatorname{dist}(D, C) = \inf\{|x-y| \mid x \in D,\ y \in C\}$ is positive. Otherwise there would be sequences $\{x_j\}_{j=1}^{\infty}$, $\{y_j\}_{j=1}^{\infty}$, $x_j \in D$, $y_j \in C$ such that $\lim_{j\to\infty} |x_j - y_j| = 0$. Since D is compact, $\{x_j\}_{j=1}^{\infty}$ would have an accumulation point $x_0 \in D$. Since x_0 would also be an accumulation point of $\{y_j\}_{j=1}^{\infty}$ and since C is closed, it would follow that $x_0 \in C$, hence $D \cap C \neq \emptyset$, which contradicts the assumptions.

Therefore we can choose balls $B_i' = B_{r_i'}(z_i)$, $i = 1, \ldots m$, with $r_i' < r_i$, such that $D \subseteq \bigcup_{i=1}^{m} B_i'$. For $1 \leq i \leq m$, let β_i be a continuous function on $\mathbb{R}^n$ with support in $B_{r_i}(z_i)$, such that $\beta_i(z) = 1$ for $z \in B_i'$. Put $\varphi_1 = \beta_1$ and set

$$\varphi_j = (1 - \beta_1)(1 - \beta_2) \cdots (1 - \beta_{j-1})\beta_j\,, \quad \text{for } 2 \leq j \leq m.$$

Every φ_j is a continuous function. By induction one obtains that for $1 \leq l \leq m$,

$$\varphi_1 + \cdots + \varphi_l = 1 - (1 - \beta_l)(1 - \beta_{l-1}) \cdots (1 - \beta_1).$$

Every $x \in D$ belongs to at least one B'_i, hence $1 - \beta_i(x) = 0$. For $l = m$ the product on the right hand side thus vanishes on D, so that $\sum_{i=1}^m \varphi_i(x) = 1$ for all $x \in D$. ■

Theorem 6.11 *Let $U \subseteq \mathbb{R}^n$ be an open set with $0 \in U$ and let $T : U \to \mathbb{R}^n$ be continuously differentiable such that $T(0) = 0$ with invertible derivative $T'(0) : \mathbb{R}^n \to \mathbb{R}^n$. Then there is a number $j \in \{1, \ldots, n\}$ and there are neighborhoods V, W of 0 in $\mathbb{R}^n$, such that the decomposition*

$$T(x) = h\big(g(Bx)\big)$$

is valid for all $x \in V$, where the linear operator $B : \mathbb{R}^n \to \mathbb{R}^n$ merely interchanges the x_j- and x_n-coordinate, and where the functions $h : W \to \mathbb{R}^n$, $g : B(V) \to W$ are of the form

$$g(x) = \begin{pmatrix} x_1 \\ \vdots \\ x_{n-1} \\ g_n(x) \end{pmatrix}, \quad h(x) = \begin{pmatrix} h_1(x) \\ \vdots \\ h_{n-1}(x) \\ x_n \end{pmatrix}, \tag{6.4}$$

and are continuously differentiable with $\det h' \neq 0$ in W, $\det g' \neq 0$ in $B(V)$.

Proof The last row of the Jacobi matrix $T'(0) = \big(\frac{\partial T_i}{\partial x_j}(0)\big)_{i,j=1,\ldots,n}$ contains at least one non-zero element, since otherwise $T'(0)$ would not be invertible. Let this be $\frac{\partial T_n}{\partial x_j}(0)$. Now define

$$\tilde{g}(x) = \begin{pmatrix} x_1 \\ \vdots \\ x_{n-1} \\ T_n(x_1, \ldots, x_{j-1}, x_n, x_{j+1}, \ldots, x_{n-1}, x_j) \end{pmatrix}. \tag{6.5}$$

Then $\tilde{g} : U \to \mathbb{R}^n$ is continuously differentiable with $\tilde{g}(0) = 0$ and

$$\tilde{g}'(x) = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ \frac{\partial T_n}{\partial x_1} & & \cdots & \frac{\partial T_n}{\partial x_j} \end{pmatrix},$$

whence $\det \tilde{g}'(0) = \frac{\partial T_n}{\partial x_j}(0) \neq 0$. Consequently the Inverse Function Theorem 5.10 implies that there are neighborhoods $\tilde{V} \subseteq U$ and W of 0 such the restriction $g : \tilde{V} \to W$ of $\tilde{g}$ to

$\tilde{V}$ is one-to-one and such that the inverse $g^{-1} : W \to \tilde{V}$ is continuously differentiable with nonvanishing determinants $\det g'$ and $\det(g^{-1})'$. Of course, we have $g^{-1}(0) = 0$. Now set $h = T \circ B \circ g^{-1}$. Then h is defined on W and is continuously differentiable with $h(0) = 0$. Also, for $y = g(x)$ we obtain from the definition of g that

$$h_n(y) = T_n\big(Bg^{-1}\big(g(x)\big)\big) = T_n(Bx) = T_n(x_1, \ldots, x_n, \ldots, x_j) = g_n(x) = y_n.$$

This equation and (6.5) show that h and g have the form required in (6.4). Set $V = B^{-1}(\tilde{V})$. Then $h \circ g \circ B : V \to \mathbb{R}^n$, and we have

$$h \circ g \circ B = T \circ B \circ g^{-1} \circ g \circ B = T,$$

which is the decomposition of T required in the theorem. The chain rule yields

$$h' = (T \circ B \circ g^{-1})' = (T' \circ B \circ g^{-1})(B \circ g^{-1})(g^{-1})',$$

whence $\det h' = \big(\det T'(B \circ g^{-1})\big) \det B \, \det(g^{-1})'$. We have $\det B = \pm 1$. Moreover, $\det(g^{-1})'$ does not vanish by construction. Thus, because $\det T'(0) \neq 0$ and because $\det T'$ is continuous, we can reduce the sizes of V and W, if necessary, such that $\det h'(x) \neq 0$ for all $x \in W$. ■

With this theorem we can prove the transformation rule, which generalizes the rule of substitution:

Theorem 6.12 (Transformation rule) *Let $U \subseteq \mathbb{R}^n$ be open and let $T : U \to \mathbb{R}^n$ be a continuously differentiable transformation such that $|\det T'(x)| > 0$ for all $x \in U$. Suppose that Ω is a compact Jordan-measurable subset of U and that $f : T(\Omega) \to \mathbb{R}$ is continuous. Then $T(\Omega)$ is a Jordan measurable subset of $\mathbb{R}^n$, the function f is integrable over $T(\Omega)$ and*

$$\int_{T(\Omega)} f(y)\, dy = \int_\Omega f\big(T(x)\big) |\det T'(x)|\, dx. \tag{6.6}$$

Proof For simplicity we prove this theorem only in the special case when Ω is connected and when $f : \mathbb{R}^n \to \mathbb{R}$ is a continuous function with $\operatorname{supp} f \subseteq T(\Omega)$. In this case f is defined outside of $T(\Omega)$ and vanishes there. Moreover, $f \circ T$ is defined in U with support contained in Ω. We can therefore extend $f \circ T$ by 0 to a continuous function on $\mathbb{R}^n$. Hence, we can extend the domain of integration on both sides of (6.6) to $\mathbb{R}^n$.

Consider first the case $n = 1$. By assumption Ω is compact and connected, hence Ω is an interval $[a, b]$. Since $\det T'(x) = T'(x)$ vanishes nowhere, $T'(x)$ is either everywhere positive in $[a, b]$ or everywhere negative. In the first case we have $T(a) < T(b)$, in the second case $T(b) < T(a)$. If we take the plus sign in the first case and the minus sign in the second case we obtain from the rule of substitution

$$\begin{aligned}\int_{T([a,b])} f(y)\,dy &= \pm\int_{T(a)}^{T(b)} f(y)\,dy = \pm\int_a^b f\big(T(x)\big)T'(x)\,dx\\ &= \int_a^b f\big(T(x)\big)|T'(x)|\,dx = \int_a^b f\big(T(x)\big)|\det T'(x)|\,dx.\end{aligned}$$

Therefore (6.6) holds for $n = 1$. Assume next that $n \geq 2$ and that (6.6) holds for $n-1$. We shall prove that this implies that (6.6) holds for n, from which the statement of the theorem follows by induction.

Assume first that the transformation is of the special form $T(x) = (x', x_n) = \big(x', T_n(x', x_n)\big)$. Then the Theorem of Fubini yields

$$\begin{aligned}&\int_{\mathbb{R}^n} f(y)\,dy = \int_{\mathbb{R}^{n-1}}\int_{\mathbb{R}} f(y', y_n)\,dy_n\,dy'\\ &= \int_{\mathbb{R}^{n-1}}\int_{\mathbb{R}} f(\big(x', T_n(x', x_n)\big)|\frac{\partial}{\partial x_n}T_n(x', x_n)|\,dx_n\,dx'\\ &= \int_{\mathbb{R}^{n-1}}\int_{\mathbb{R}} f\big(T(x)\big)|\det T'(x)|\,dx_n\,dx' = \int_{\mathbb{R}^n} f\big(T(x)\big)|\det T'(x)|\,dx,\end{aligned}$$

since $\det T'(x) = \frac{\partial}{\partial x_n}T_n(x)$. The transformation formula thus holds in this case. Next, assume that the transformation is of the special form $T(x) = \big(\tilde{T}(x', x_n), x_n\big)$ with $\tilde{T}(x', x_n) \in \mathbb{R}^{n-1}$. With the Jacobi matrix $\partial_{x'}\tilde{T}(x) = \Big(\frac{\partial \tilde{T}_i}{\partial x_j}(x)\Big)_{i,j=1,\ldots,n-1}$ we have

$$\det T'(x) = \det\begin{pmatrix}\partial_{x'}\tilde{T}(x) & 0\\ 0 & 1\end{pmatrix} = \det\big(\partial_{x'}\tilde{T}(x)\big).$$

Since by assumption the transformation rule holds for $n-1$, we thus have

$$\begin{aligned}&\int_{\mathbb{R}^n} f(y)\,dy = \int_{\mathbb{R}}\int_{\mathbb{R}^{n-1}} f(y', y_n)\,dy'\,dy_n\\ &= \int_{\mathbb{R}}\int_{\mathbb{R}^{n-1}} f\big(\tilde{T}(x', x_n), x_n\big)|\det\big(\partial_{x'}\tilde{T}(x', x_n)\big)|dx'\,dx_n\\ &= \int_{R^n} f\big(T(x)\big)|\det T'(x)|\,dx.\end{aligned}$$

The transformation formula (6.6) therefore holds also in this case. It also holds when the transformation T is a linear operator B, which merely interchanges coordinates, since this amounts to a change of the order of integration when the integral is computed iteratively, and by the Theorem of Fubini the order of integration does not matter.

If (6.6) holds for the transformations R and S, then it also holds for the transformation $T = R \circ S$. For,

$$\begin{aligned}\int_{\mathbb{R}^n} f(z)\,dz &= \int_{\mathbb{R}^n} f\big(R(y)\big)|\det R'(y)|\,dy \\ &= \int_{\mathbb{R}^n} f\Big(R\big(S(x)\big)\Big)|\det R'\big(S(x)\big)|\,|\det S'(x)|\,dx \\ &= \int_{\mathbb{R}^n} f\big(T(x)\big)|\det\big(R'\big(S(x)\big)S'(x)\big)|\,dx = \int_{\mathbb{R}^n} f\big(T(x)\big)|\det T'(x)|\,dx,\end{aligned}$$

since by the determinant multiplication theorem for $n \times n$-matrices M_1 and M_2 we have $\det M_1 \det M_2 = \det(M_1 M_2)$.

If T has the properties stated in the theorem and if $y \in U$, then the transformation $\hat{T}(x-y) = T(x) - T(y)$ satisfies all assumptions of Theorem 6.11, since $\hat{T}(0) = 0$. It follows by this theorem that there is a neighborhood V of y such that the decomposition

$$T(x) = T(y) + h\Big(g\big(B(x-y)\big)\Big)$$

holds for $x \in V$ with elementary transformations h, g and B, for which we showed above that (6.6) holds; since (6.6) also holds for the transformations which merely consist in subtraction of y or addition of $T(y)$, it also holds for the composition T of these elementary transformations. We thus proved that each point $y \in U$ has a neighborhood $V(y)$ such that (6.6) holds for all continuous f, for which $\operatorname{supp}(f \circ T) \subseteq V(y)$.

Since $\det T'(y) \neq 0$, the inverse function theorem implies that T is locally a diffeomorphismus. Therefore $T(V(y))$ contains an open neighborhood of $T(y)$. If $\operatorname{supp} f$ is a subset of this neighborhood, we have $\operatorname{supp}(f \circ T) \subseteq V(y)$, whence (6.6) holds for all such f. We conclude that each point $z \in T(\Omega)$ has a neighborhood $W(z)$, which we can choose to be an open ball, such that (6.6) holds for all continuous f whose support lies in $W(z)$.

Since $T(\Omega)$ is compact, there are points $z_1, \ldots, z_p$ in $T(\Omega)$ such that the union of the open balls $W(z_i)$ covers $T(\Omega)$. By Theorem 6.10 there is a partition of unity $\{\varphi_i\}_{i=1}^p$ on $T(\Omega)$ subordinate to the covering $\{W(z_i)\}_{i=1}^p$. If f is a continuous function with $\operatorname{supp} f \subseteq T(\Omega)$, we thus have for every $x \in \mathbb{R}^n$

$$f(x) = f(x)\sum_{i=1}^p \varphi_i(x) = \sum_{i=1}^p \big(\varphi_i(x)f(x)\big).$$

Since $\operatorname{supp}(\varphi_i f) \subseteq \operatorname{supp}\varphi_i \subseteq W(z_i)$, the transformation equation (6.6) holds for every $\varphi_i f$, whence it holds for the sum of these functions, which is f. ∎

7 p-dimensional surfaces in $\mathbb{R}^{\mathrm{m}}$, curve- and surface integrals, Theorems of Gauß and Stokes

7.1 p-dimensional patches of a surface, submanifolds

Let $L(\mathbb{R}^n, \mathbb{R}^m)$ be the vector space of all linear mappings from $\mathbb{R}^n$ to $\mathbb{R}^m$. For $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ the range $A(\mathbb{R}^n)$ is a linear subspace of $\mathbb{R}^m$.

Definition 7.1 Let $A \in L(\mathbb{R}^n, \mathbb{R}^m)$. The dimension of this subspace $A(\mathbb{R}^n)$ is called rank of A.

From linear algebra we know that a linear mapping $A : \mathbb{R}^p \to \mathbb{R}^n$ with rank p is injective.

Definition 7.2 Let $U \subseteq \mathbb{R}^p$ be an open set and $p < n$. Let the transformation $\gamma : U \to \mathbb{R}^n$ be continuously differentiable and assume that the derivative

$$\gamma'(u) \in L(\mathbb{R}^p, \mathbb{R}^n)$$

has the rank p for all $u \in U$. Then γ is called a parametric representation or simply a parametrization of a p-dimensional surface patch in $\mathbb{R}^n$. If $p = 1$, then γ is called a parametric representation of a curve in $\mathbb{R}^n$.

Note that γ need not to be injective. The surface may have double points.

Example 1: Let $U = \{(u, v) \in \mathbb{R}^2 \,|\, u^2 + v^2 < 1\}$ and let $\gamma : U \to \mathbb{R}^3$ be defined by

$$\gamma(u,v) = \begin{pmatrix} \gamma_1(u,v) \\ \gamma_2(u,v) \\ \gamma_3(u,v) \end{pmatrix} = \begin{pmatrix} u \\ v \\ \sqrt{1-(u^2+v^2)} \end{pmatrix},$$

then γ is the parametric representation of the upper half of the unit sphere in $\mathbb{R}^3$. To see this, observe that the two columns of the matrix

$$\gamma'(u,v) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -\frac{u}{\sqrt{1-(u^2+v^2)}} & -\frac{v}{\sqrt{1-(u^2+v^2)}} \end{pmatrix}.$$

are linearly independent for all $(u, v) \in U$, whence the rank is 2.

Example 2: In the preceding example the patch of a surface is given by the graph of a function. More general, let $U \subseteq \mathbb{R}^p$ be an open set and $f : U \to \mathbb{R}^{n-p}$ be continuously differentiable. Then the graph of f is a p-dimensional patch of a surface which is embedded in $\mathbb{R}^n$. The mapping $\gamma : U \to \mathbb{R}^n$,

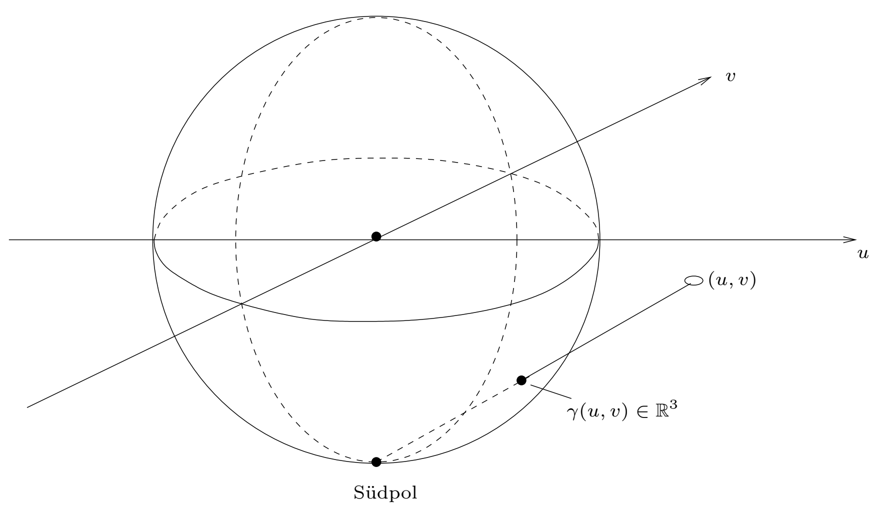
$$\begin{aligned}
\gamma_1(u) &:= u_1 \\
\gamma_2(u) &:= u_2 \\
&\vdots \\
\gamma_p(u) &:= u_p \\
\gamma_{p+1}(u) &:= f_1(u_1 \ldots, u_p) \\
&\vdots \\
\gamma_n(u) &:= f_{n-p}(u_1 \ldots, u_p)
\end{aligned}$$

is a parametric representation of this surface, since the column vectors of the matrix

$$\gamma'(u) = \begin{pmatrix} 1 & \ldots & 0 \\ \vdots & \ddots & \\ 0 & \ldots & 1 \\ \partial_{x_1} f_1(u) & \ldots & \partial_{x_p} f_1(u) \\ \vdots & & \vdots \\ \partial_{x_1} f_{n-p}(u) & \ldots & \partial_{x_p} f_{n-p}(u) \end{pmatrix},$$

are linearly independent. Therefore the rank is p.

Example 3: By stereographic projection, the sphere with center in the origin, which is punched at the south pole, can be mapped one-to-one onto the plane. The inverse γ of this projection maps the plane onto the punched sphere:



From the figure we see that the components $\gamma_1, \ldots, \gamma_3$ of the mapping $\gamma : \mathbb{R}^2 \to \mathbb{R}^3$ satisfy

$$\frac{\gamma_1}{\gamma_2} = \frac{u}{v}, \qquad \frac{\sqrt{u^2+v^2} - \sqrt{\gamma_1^2+\gamma_2^2}}{\gamma_3} = \frac{\sqrt{u^2+v^2}}{-1}, \qquad \gamma_1^2 + \gamma_2^2 + \gamma_3^2 = 1.$$

Solution of these equations for $\gamma_1, \ldots, \gamma_3$ yields

$$\begin{aligned} \gamma_1(u,v) &= \frac{2u}{1+u^2+v^2} \\ \gamma_2(u,v) &= \frac{2v}{1+u^2+v^2} \\ \gamma_3(u,v) &= \frac{1-u^2-v^2}{1+u^2+v^2} \,. \end{aligned}$$

The derivation is

$$\gamma'(u,v) = \frac{2}{(1+u^2+v^2)^2} \begin{pmatrix} 1-u^2+v^2 & -2uv \\ -2uv & 1+u^2-v^2 \\ -2u & -2v \end{pmatrix} .$$

For $u^2+v^2 \neq 1$ we have

$$\begin{aligned} \begin{vmatrix} \partial_{x_1}\gamma_1(u,v) & \partial_{x_2}\gamma_1(u,v) \\ \partial_{x_1}\gamma_2(u,v) & \partial_{x_2}\gamma_2(u,v) \end{vmatrix} &= (1+(v^2-u^2))(1-(v^2-u^2)) - 4u^2v^2 \\ &= 1-(v^2-u^2)^2 - 4u^2v^2 = 1-(v^2+u^2)^2 \neq 0\,, \end{aligned}$$

and for $u \neq 0$

$$\begin{aligned} \begin{vmatrix} \partial_{x_1}\gamma_2(u,v) & \partial_{x_2}\gamma_2(u,v) \\ \partial_{x_1}\gamma_3(u,v) & \partial_{x_2}\gamma_3(u,v) \end{vmatrix} &= 4uv^2 + 2u(1+u^2-v^2) \\ &= 2u(1+u^2+v^2) \neq 0\,. \end{aligned}$$

Correspondingly, for $v \neq 0$ we get

$$\begin{vmatrix} \partial_{x_1}\gamma_1(u,v) & \partial_{x_2}\gamma_1(u,v) \\ \partial_{x_1}\gamma_3(u,v) & \partial_{x_2}\gamma_3(u,v) \end{vmatrix} = -2v(1+u^2+v^2) \neq 0\,.$$

These relations show that $\gamma'(u,v)$ has rank 2 for all $(u,v) \in \mathbb{R}^2$, which shows that γ is a parametrization of the unit sphere with the south pole removed.

Example 4: Let $\tilde{\gamma}$ be the restriction of the parametrization γ from Example 3 to the unit disk $U = \{(u,v) \in \mathbb{R}^2 \mid u^2+v^2 < 1\}$. This restriction is a parametrization of the upper half of the unit sphere, which differs from the parametrization of Example 1.

Definition 7.3 Let $U, V \subseteq \mathbb{R}^p$ be open sets and $\gamma : U \to \mathbb{R}^n$, $\tilde{\gamma} : V \to \mathbb{R}^n$ be parametrizations of p-dimensional surface patches. γ and $\tilde{\gamma}$ are called equivalent, if there exists a diffeomorphism $\varphi : V \to U$ with

$$\tilde{\gamma} = \gamma \circ \varphi\,.$$

This is an equivalence relation on the set of the parametric representation of surface patches.

Example 5: Let $\gamma : U \to \mathbb{R}^3$ be the parametrization of the upper half of the unit sphere in Example 1 and let $\tilde{\gamma} : \tilde{U} \to \mathbb{R}^3$ be the corresponding parametrization in Example 4. These parametrizations are equivalent. For, a diffeomorphism $\gamma : U \to U$ is given by

$$\varphi(u, v) = \begin{pmatrix} \frac{2u}{1+u^2+v^2} \\ \frac{2v}{1+u^2+v^2} \end{pmatrix}.$$

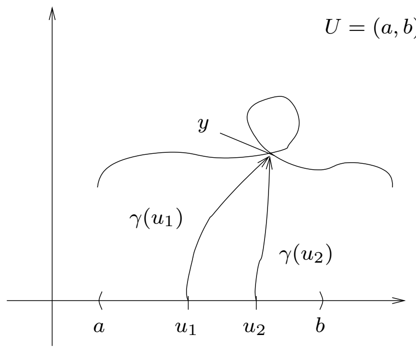
We have

$$(\gamma \circ \varphi)(u, v) = \begin{pmatrix} \frac{2u}{1+u^2+v^2} \\ \frac{2v}{1+u^2+v^2} \\ \sqrt{1 - \frac{4u^2+4v^2}{(1+u^2+v^2)^2}} \end{pmatrix} = \frac{1}{1+u^2+v^2} \begin{pmatrix} 2u \\ 2v \\ 1-u^2-v^2 \end{pmatrix} = \tilde{\gamma}(u, v).$$

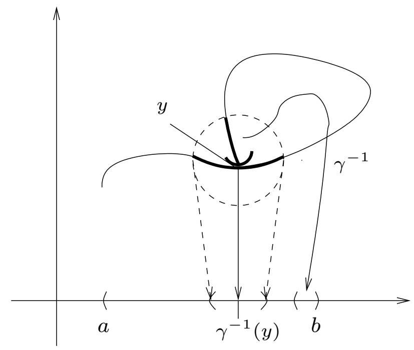
In Example 3 a parametric representation for the punched sphere is given. However, for topological reasons there exists no parametric representation $\gamma : U \to \mathbb{R}^3$ of the entire sphere. To parametrize the entire sphere we have to split it into at least two parts, which can be parametrized separately. Therefore we define:

Definition 7.4 Let $U \subseteq \mathbb{R}^p$ be an open set. A parametrization $\gamma : U \to \mathbb{R}^n$ of a p–dimensional surface patch is called simple if γ is injective with continuous inverse γ^{-1}. In this case the range $F = \gamma(U)$ is called a simple p-dimensional surface patch.

The figure following below explains this definition at the example of a curve in $\mathbb{R}^2$.



γ is not injective: The two different parameter values u_1 and u_2 are mapped to the same double point of the curve.



γ^{-1} is not continuous: The image of every sphere around y contains points, whose distance to $\gamma^{-1}(y)$ is greater than $\varepsilon = \frac{1}{2}\big(b - \gamma^{-1}(u)\big)$.

Examples of parametrizations $\gamma : (a, b) \to \mathbb{R}^2$, which are not simple.

Definition 7.5 A subset $M \subseteq \mathbb{R}^n$ is called p-dimensional submanifold of $\mathbb{R}^n$ if there exists for each $x \in M$ an open n-dimensional neighborhood $V(x)$ of x and a mapping γ_x with the properties:

(i) $V(x) \cap M$ is a simple p-dimensional surface patch parametrized by γ_x.

(ii) If x and y are two points in M with

$$N = \big(V(x) \cap M\big) \cap \big(V(y) \cap M\big) \neq \emptyset,$$

then $\gamma_x : \gamma_x^{-1}(N) \to M$ and $\gamma_y : \gamma_y^{-1}(N) \to M$ are equivalent parametrizations of N.

The inverse mapping $\kappa_x = \gamma_x^{-1} : V(x) \cap M \to U \subseteq \mathbb{R}^p$ is called a coordinate mapping or a chart. The set $\{\kappa_x \mid x \in M\}$ of charts is called an atlas of M.

Observe that two charts κ_x and κ_y of the atlas of M need not be different. For, if $y \in M$ belongs to the domain of definition $V(x) \cap M$ of the chart κ_x, then $\kappa_y = \kappa_x$ is allowed by this definition.

Example 6: Let $S = \{x \in \mathbb{R}^3 \mid |x| = 1\}$ be the unit sphere in $\mathbb{R}^3$. The stereographic projection of Example 3, which maps $S \backslash \{(0, 0, -1)\}$ onto $U = \mathbb{R}^2$, is a chart of S, a second chart is given by the stereographic projection of $S \backslash \{(0, 0, 1)\}$ from the north pole onto $\mathbb{R}^2$. Therefore the unit sphere is a two-dimensional submanifold of $\mathbb{R}^3$ with an atlas consisting of two charts only.

Definition 7.6 Let M be a p-dimensional submanifold of $\mathbb{R}^n$ and x a point in M. If γ is a parametrization of M in a neighborhood of x with $x = \gamma(u)$, then the range of the linear mapping $\gamma'(u)$ is a p-dimensional subspace of $\mathbb{R}^n$. This subspace is called tangent space of M in x, written $T_x(M)$ or simply T_xM.

The definition of $T_x(M)$ is independent of the chosen parametrization. To see this, assume that $\tilde{\gamma}$ is a parametrization equivalent to γ with $x = \tilde{\gamma}(\tilde{u})$ and that φ is a diffeomorphism with $\gamma = \tilde{\gamma} \circ \varphi$ and $\tilde{u} = \varphi(u)$. Then the chain rule gives

$$\gamma'(u) = \tilde{\gamma}'(\tilde{u})\varphi'(u).$$

Since $\varphi'(u)$ is an invertible linear mapping, this equation implies that $\gamma'(u)$ and $\tilde{\gamma}'(\tilde{u})$ have the same ranges.

7.2 Integration on patches of a surface

Let $M \subseteq \mathbb{R}^n$ be a simple p-dimensional surface patch parametrized by $\gamma : U \to M$. For $1 \leq i,\, j \leq p$ let the continuous functions $g_{ij} : U \to \mathbb{R}$ be defined by

$$g_{ij}(u) = \frac{\partial \gamma}{\partial u_i}(u) \cdot \frac{\partial \gamma}{\partial u_j}(u) = \begin{pmatrix} \frac{\partial \gamma_1}{\partial u_i}(u) \\ \vdots \\ \frac{\partial \gamma_n}{\partial u_i}(u) \end{pmatrix} \cdot \begin{pmatrix} \frac{\partial \gamma_1}{\partial u_j}(u) \\ \vdots \\ \frac{\partial \gamma_n}{\partial u_j}(u) \end{pmatrix} = \sum_{k=1}^{n} \frac{\partial \gamma_k}{\partial u_i}(u) \frac{\partial \gamma_k(u)}{\partial u_j} \quad .$$

Definition 7.7 For $u \in U$ let

$$G(u) = \begin{pmatrix} g_{11}(u) & \dots & g_{1p}(u) \\ \vdots & & \\ g_{p1}(u) & \dots & g_{pp}(u) \end{pmatrix}.$$

The function $g : U \to \mathbb{R}$ defined by $g(u) := \det(G(u))$ is called Gram's determinant to the parametrization γ.

To motivate this definition fix $u \in U$. Then

$$h \to \gamma(u) + \gamma'(u)h : \mathbb{R}^p \to \mathbb{R}^n$$

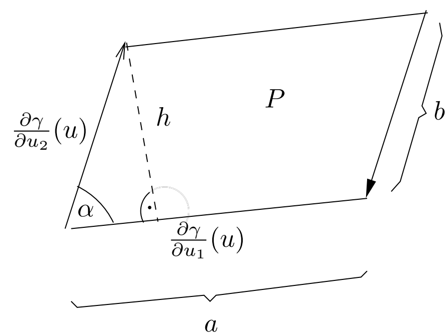
is the parametrization of a planar surface which is tangential to the surface patch M at the point $x = \gamma(u)$. The partial derivatives $\frac{\partial \gamma}{\partial u_1}(u), \dots, \frac{\partial \gamma}{\partial u_p}(u)$ are vectors lying in the tangent space T_xM of M at the point x, a p-dimensional linear subspace of $\mathbb{R}^p$, and even generate this vector space because by assumption the matrix $\gamma'(u)$ has rank p. The column vectors $\frac{\partial\gamma}{\partial u_1}(u), \ldots, \frac{\partial\gamma}{\partial u_p}(u)$ of this matrix are called tangent vectors of M in $\gamma(u)$. The set

$$P = \Big\{ \sum_{i=1}^{p} r_i \frac{\partial\gamma}{\partial u_i}(u) \;\Big|\; r_i \in \mathbb{R}\,,\, 0 \le r_i \le 1 \Big\}$$

is a subset of the tangent space, a parallelotope.

Theorem 7.8 *We have $g(u) > 0$ and $\sqrt{g(u)}$ is equal to the p-dimensional volume of the parallelotope P.*

For simplicity we prove this theorem for $n = 2$ only. In this case P is the parallelogram shown in the figure.



With $a = \left|\frac{\partial\gamma}{\partial u_1}(u)\right|$ and $b = \left|\frac{\partial\gamma}{\partial u_2}(u)\right|$ it follows that

$$\begin{aligned}
\sqrt{g(u)} &= \sqrt{\det(G(u))} \\
&= \sqrt{\begin{vmatrix} \frac{\partial\gamma}{\partial u_1}(u)\cdot\frac{\partial\gamma}{\partial u_1}(u) & \frac{\partial\gamma}{\partial u_1}(u)\cdot\frac{\partial\gamma}{\partial u_2}(u) \\ \frac{\partial\gamma}{\partial u_2}(u)\cdot\frac{\partial\gamma}{\partial u_1}(u) & \frac{\partial\gamma}{\partial u_2}(u)\cdot\frac{\partial\gamma}{\partial u_2}(u) \end{vmatrix}} = \sqrt{\begin{vmatrix} a^2 & ab\cos\alpha \\ ab\cos\alpha & b^2 \end{vmatrix}} \\
&= \sqrt{a^2b^2 - a^2b^2\cos^2\alpha} = ab\sqrt{1-\cos^2\alpha} = ab\sin\alpha = b\cdot h = \text{ area of } P\,. \qquad \blacksquare
\end{aligned}$$

Definition 7.9 Let $f : M \to \mathbb{R}$ be a function. f is called integrable over the p-dimensional surface patch M if the function

$$u \to f(\gamma(u))\sqrt{g(u)}$$

is integrable over U. The integral of f over M is defined by

$$\int_M f(x)dS(x) := \int_U f(\gamma(u))\sqrt{g(u)}du.$$

$dS(x)$ is called the p-dimensional surface element of M in x. Symbolically one writes

$$dS(x) = \sqrt{g(u)}du\,, \quad x = \gamma(u)\,.$$

Next we show that the integral is well defined by verifying that the value of the integral $\int_U f(\gamma(u))\sqrt{g(u)}du$ does not change if the parametrization γ is replaced with an equivalent one.

Theorem 7.10 *Let $U, \tilde{U} \subseteq \mathbb{R}^p$ be open sets, let $\gamma : U \to M$ and $\tilde{\gamma} : \tilde{U} \to M$ be equivalent parametrizations of the surface patch M and let $\varphi : \tilde{U} \to U$ be a diffeomorphism with $\tilde{\gamma} = \gamma \circ \varphi$. The Gram determinants for the parametrizations γ and $\tilde{\gamma}$ are denoted by $g : U \to \mathbb{R}$ and $\tilde{g} : \tilde{U} \to \mathbb{R}$, respectively. Then we have:*

(i) For all $u \in \tilde{U}$

$$\tilde{g}(u) = g(\varphi(u))|\det \varphi'(u)|^2.$$

(ii) If $(f \circ \gamma)\sqrt{g}$ is integrable over U, then $(f \circ \tilde{\gamma})\sqrt{\tilde{g}}$ is integrable over $\tilde{U}$ with

$$\int_U f(\gamma(u))\sqrt{g(u)}du = \int_{\tilde{U}} f(\tilde{\gamma}(v))\sqrt{\tilde{g}(v)}dv\,.$$

Proof: (i) From

$$g_{ij}(u) = \sum_{k=1}^{n} \frac{\partial \gamma_k(u)}{\partial u_i}\,\frac{\partial \gamma_k(u)}{\partial u_j}\,,$$

we obtain that

$$G(u) = [\gamma'(u)]^T \gamma'(u)\,.$$

From the chain rule and the rule of multiplication of determinants we thus conclude

$$\begin{aligned}
\tilde{g} &= \det \tilde{G} = \det([\tilde{\gamma}']^T \tilde{\gamma}') \\
&= \det([(\gamma' \circ \varphi)\varphi']^T (\gamma' \circ \varphi)\varphi') = \det(\varphi'^T [\gamma' \circ \varphi]^T [\gamma' \circ \varphi]\varphi') \\
&= (\det \varphi')\det([\gamma' \circ \varphi]^T [\gamma' \circ \varphi])(\det \varphi') = (\det \varphi')^2 (g \circ \varphi)\,.
\end{aligned}$$

(ii) Using part (i) of the proposition we obtain from the transformation rule (Theorem 6.10) that

$$\int_U f(\gamma(u))\sqrt{g(u)}du = \int_{\tilde{U}} f((\gamma \circ \varphi)(v))\sqrt{g(\varphi(v))}|\det \varphi'(v)|dv = \int_{\tilde{U}} f(\tilde{\gamma}(v))\sqrt{\tilde{g}(v)}dv\,.$$

■

7.3 Integration on submanifolds

Now the definition of integrals on patches of surface on submanifolds has to be generalized. I restrict myself to p-dimensional submanifolds M of $\mathbb{R}^n$, which can be covered by finitely many simple surface patches $V_1, \ldots, V_m$. Thus, assume that $M = \bigcup_{j=1}^{m} V_j$. To every $1 \leq j \leq m$ let $U_j \subseteq \mathbb{R}^p$ be an open set and $\kappa_j : V_j \subseteq M \to U_j$ a chart. The inverse mappings $\gamma_j = \kappa_j^{-1} : U_j \to V_j$ are simple parametrizations.

Definition 7.11 A family $\{\alpha_j\}_{j=1}^m$ of functions $\alpha_j : M \to \mathbb{R}$ is called a partition of unity of locally integrable functions, which is subordinate to the covering $\{V_j\}_{j=1}^m$, if

(i) $0 \leq \alpha_j \leq 1, \qquad \alpha_j|_{M \setminus V_j} = 0,$

(ii) $\sum_{j=1}^{m} \alpha_j(x) = 1, \quad \text{for all } x \in M,$

(iii) The function $\alpha_j \circ \gamma_j : U_j \to \mathbb{R}$ is locally integrable, i.e. for all $R > 0$ there exists the integral
$$\int\limits_{U_j \cap \{|u|<R\}} \alpha_j(\gamma_j(u))du \,.$$

Definition 7.12 Let M be a p-dimensional submanifold of $\mathbb{R}^n$, which can be covered by finitely many simple surface patches $V_1, \ldots, V_M$. A function $f : M \to \mathbb{R}$ is called integrable over M, if $f_{|V_j}$ is integrable for all j. In this case one sets

$$\int\limits_M f(x)dS(x) = \sum_{j=1}^{m} \int\limits_{V_j} \alpha_j(x)f(x)dS(x)$$

with a partition of unity $\{\alpha_j\}_{j=1}^m$ of locally integrable functions subordinate to the covering $\{V_j\}_{j=1}^m$.

The function $\alpha_j(x)f(x)$ is integrable over V_j, since by assumption $(f \circ \gamma_j)\sqrt{g_j}$ is integrable over U_j with the Gram determinant g_j to the parametrization γ_j. Thus, since $0 \leq \alpha_j(x) \leq 1$, the function $(\alpha_j \circ \gamma_j)(f \circ \gamma_j)\sqrt{g_j}$ is also integrable over U_j as a product of an integrable and a bounded locally integrable function.

It must be shown that the definition of the integral is independent of the choice of the covering of M by simple surface patches and of the choice of the partition of unity:

Theorem 7.13 *Let M be a p-dimensional submanifold in $\mathbb{R}^n$ and let*

$$\gamma_k : U_k \to V_k\,, \quad k = 1, \ldots, m$$
$$\tilde{\gamma}_j : \tilde{U}_j \to \tilde{V}_j\,, \quad j = 1, \ldots, l$$

be simple parametrizations with $\bigcup_{k=1}^{m} V_k = \bigcup_{j=1}^{l} \tilde{V}_j = M$. Assume that if

$$D_{jk} = \tilde{V}_j \cap V_k \neq \emptyset$$

holds, then

$$U_{kj} = \gamma_k^{-1}(D_{jk})\,, \quad \tilde{U}_{kj} = \tilde{\gamma}_j^{-1}(D_{jk})$$

are Jordan-measurable subsets of $\mathbb{R}^n$ and

$$\gamma_k : U_{kj} \to D_{jk}, \quad \tilde{\gamma}_j : \tilde{U}_{kj} \to D_{jk}$$

are equivalent parametrizations.

Assume that the partitions of unity $\{\alpha_k\}_{k=1}^{m}$ and $\{\beta_j\}_{j=1}^{\ell}$ are subordinate to the coverings $\{V_k\}_{k=1}^{m}$ and $\{\tilde{V}_j\}_{j=1}^{\ell}$, respectively. Then

$$\sum_{k=1}^{m} \int_{V_k} \alpha_k(x) f(x) dS(x) = \sum_{j=1}^{l} \int_{\tilde{V}_j} \beta_j(x) f(x) dS(x)\,. \tag{7.1}$$

Proof: First I show that $\beta_j \alpha_k f$ is integrable over V_k and over $\tilde{V}_j$ with

$$\int_{\tilde{V}_j} \beta_j(x)\alpha_k(x) f(x) dS(x) = \int_{V_k} \beta_j(x)\alpha_k(x) f(x) dS(x)\,. \tag{7.2}$$

To see this, assume that g_k and $\tilde{g}_j$ are the Gram determinants to γ_k and $\tilde{\gamma}_j$, respectively. If the function $[(\alpha_k f) \circ \gamma_k]\sqrt{g_k}$ is integrable over U_k, then this function is also integrable over U_{kj}, since U_{kj} is a Jordan measurable subset of U_k. According to theorem 7.10 $[(\alpha_k f) \circ \tilde{\gamma}_j]\sqrt{\tilde{g}_j}$ is then integrable over $\tilde{U}_{jk}$. By assumption, $\beta_j \circ \tilde{\gamma}_j$ is locally integrable over $\tilde{U}_j$, therefore this function is also locally integrable over $\tilde{U}_{jk}$ because $\tilde{U}_{jk}$ is a Jordan measurable subset of $\tilde{U}_j$. From $0 \leq \beta_j \circ \tilde{\gamma}_j \leq 1$ we thus conclude that the product

$$[(\beta_j \alpha_k f) \circ \tilde{\gamma}_j]\sqrt{\tilde{g}_j} = (\beta_j \circ \tilde{\gamma}_j)[(\alpha_k f) \circ \tilde{\gamma}_j]\sqrt{\tilde{g}_j}$$

is integrable over $\tilde{U}_{jk}$, and by the equivalence of the parametrizations $\gamma_k : U_{kj} \to D_{jk}$ and $\tilde{\gamma}_j : \tilde{U}_{jk} \to D_{jk}$ it thus follows that

$$\int_{\tilde{U}_{jk}} [(\beta_j \alpha_k f) \circ \tilde{\gamma}_j]\sqrt{\tilde{g}_j} du = \int_{U_{kj}} [(\beta_j \alpha_k f) \circ \gamma_k]\sqrt{g_k} du\,. \tag{7.3}$$

From $(\beta_j\alpha_k)(x) = 0$ for all $x \in M \setminus D_{jk}$ we get $[(\beta_j\alpha_k f) \circ \tilde{\gamma}_j](u) = 0$ for all $u \in \tilde{U}_j \setminus \tilde{U}_{jk}$ and $[(\beta_j\alpha_k f) \circ \gamma_k](u) = 0$ for all $u \in U_k \setminus U_{kj}$. Therefore the domains of integration in (7.3) can be extended without modification of the values of the integrals. It follows that

$$\int_{\tilde{U}_j} [(\beta_j\alpha_k f) \circ \tilde{\gamma}_j]\sqrt{\tilde{g}_j}du = \int_{U_k} [(\beta_j\alpha_k f) \circ \gamma_k]\sqrt{g_k}du \,.$$

Since $\tilde{\gamma}_j : \tilde{U}_j \to \tilde{V}_j$ and $\gamma_k : U_k \to V_k$ are parametrizations, this means that (7.2) is satisfied. Together with $\sum\limits_{j=1}^{\ell} \beta_j(x) = 1$ and $\sum\limits_{k=1}^{m} \alpha_k(x) = 1$ it follows from (7.2)

$$\sum_{k=1}^{m} \int_{V_k} \alpha_k(x)f(x)dS(x) = \sum_{k=1}^{m} \int_{V_k} \sum_{j=1}^{\ell} \beta_j(x)\alpha_k(x)f(x)dS(x)$$

$$= \sum_{j=1}^{\ell}\sum_{k=1}^{m} \int_{V_k} \beta_j(x)\alpha_k(x)f(x)dS(x) = \sum_{j=1}^{\ell}\sum_{k=1}^{m} \int_{\tilde{V}_j} \beta_j(x)\alpha_k(x)f(x)dS(x)$$

$$= \sum_{j=1}^{\ell} \int_{\tilde{V}_j} \sum_{k=1}^{m} \alpha_k(x)\beta_j(x)f(x)dS(x) = \sum_{j=1}^{\ell} \int_{\tilde{V}_j} \beta_j(x)f(x)dS(x) \,,$$

and this is (7.1) ■

7.4 The Integral Theorem of Gauß

To formulate the Gauß Theorem I need two definitions:

Definition 7.14 (Normal vector)

(i) Let $A \subseteq \mathbb{R}^n$ be a compact set. We say that A has a smooth boundary, if ∂A is an $(n-1)$-dimensional submanifold of $\mathbb{R}^n$.

(ii) Let $x \in A$. If the nonzero vector $\nu \in \mathbb{R}^n$ is orthogonal to all vectors in the tangent space $T_x(\partial A)$ of ∂A at x, then ν is called normal vector of ∂A at x. If $|\nu| = 1$ holds, then ν is a unit normal vector. If ν points to the exterior of A, then ν is called exterior normal vector.

Definition 7.15 (Divergence) Let $U \subseteq \mathbb{R}^n$ be an open set and let $f : U \to \mathbb{R}^n$ be differentiable. Then the function $\operatorname{div} f : U \to \mathbb{R}$ is defined by

$$\operatorname{div} f(x) := \sum_{i=1}^{n} \frac{\partial}{\partial x_i} f_i(x) \,.$$

$\operatorname{div} f$ is called the divergence of f.

Theorem 7.16 (Theorem of Gauß) *Let $A \subseteq \mathbb{R}^n$ be a compact set with smooth boundary; let $U \subseteq \mathbb{R}^n$ be an open set with $A \subseteq U$ and let $f : U \to \mathbb{R}^n$ be continuously differentiable. $\nu(x)$ denotes the exterior normal vector to ∂A at x. Then*

$$\int\limits_{\partial A} \nu(x) \cdot f(x) dS(x) = \int\limits_{A} \operatorname{div} f(x) dx \, .$$

For $n = 1$ the theorem says: Let $a, b \in \mathbb{R}$, $a < b$. Then

$$f(b) - f(a) = \int\limits_a^b \frac{d}{dx} f(x) dx \, ,$$

and we see that the Theorem of Gauß is the generalization of the fundamental theorem of calculus to $\mathbb{R}^n$.

Example for an application: A body A is submerged in a liquid with specific weight c. The surface of the liquid is given by the plane $x_3 = 0$. Then the pressure at a point $x = (x_1, x_2, x_3) \in \mathbb{R}^3$ with $x_3 < 0$ is

$$-cx_3 \, .$$

If $x \in \partial A$, then this pressure acts on the body with the force per unit area

$$-cx_3(-\nu(x)) = cx_3\nu(x)$$

in direction of the external unit normal vector $\nu(x)$ to ∂A at x. The total force on the body is thus equal to

$$K = \begin{pmatrix} K_1 \\ K_2 \\ K_3 \end{pmatrix} = \int\limits_{\partial A} cx_3\nu(x) dS(x) \, .$$

Application of the Gaussian Theorem to the functions $f_1, f_2, f_3 : A \to \mathbb{R}^3$ defined by

$$f_1(x_1, x_2, x_3) = (x_3, 0, 0), \quad f_2(x_1, x_2, x_3) = (0, x_3, 0), \quad f_3(x_1, x_2, x_3) = (0, 0, x_3)$$

yields for $i = 1, 2$

$$K_i = \int_{\partial A} cx_3\nu_i(x) dS(x) = c \int \nu(x) \cdot f_i(x) dS(x) = c \int_A \frac{\partial}{\partial x_i} x_3 \, dx = 0,$$

and for $i = 3$

$$K_3 = \int_{\partial A} cx_3\nu_3(x) dS(x) = c \int \nu(x) \cdot f_3(x) dS(x) = c \int_A \frac{\partial}{\partial x_3} x_3 dx = c \int_A dx = c \operatorname{Vol}(A).$$

K has the direction of the positive x_3-axis. Therefore K is a buoyant force acting on A with the value $c \operatorname{Vol}(A)$. This is equal to the weight of the displaced liquid.

7.5 Green's formulae

Let $U \subseteq \mathbb{R}^n$ be an open set, let $A \subseteq U$ be a compact set with smooth boundary, and for $x \in \partial A$ let $\nu(x)$ be the exterior unit normal to ∂A at x.

In the following we write $\nabla f(x)$ for differentiable $f : U \to \mathbb{R}$ to denote the gradient $\operatorname{grad} f(x) \in \mathbb{R}^n$.

Definition 7.17 Let the function $f : U \to \mathbb{R}$ be continuously differentiable. Then the normal derivative of f at the point $x \in \partial A$ is defined by

$$\frac{\partial f}{\partial \nu}(x) := f'(x)\nu(x) = \nu(x) \cdot \nabla f(x) = \sum_{i=1}^{n} \frac{\partial f(x)}{\partial x_i} \nu_i(x) .$$

The normal derivative of f is the directional derivative of f in the direction of ν. For twice differentiable $f : U \to \mathbb{R}$ set

$$\Delta f(x) := \sum_{i=1}^{n} \frac{\partial^2}{\partial x_i^2} f(x) .$$

Δ is called Laplace operator.

Theorem 7.18 *For $f, g \in C^2(U, \mathbb{R})$ we have*

(i) **Green's first identity:**

$$\int\limits_{\partial A} f(x) \frac{\partial g}{\partial \nu}(x) dS(x) \quad = \quad \int\limits_{A} \big(\nabla f(x) \cdot \nabla g(x) + f(x) \Delta g(x)\big) dx.$$

(ii) **Green's second identity:**

$$\int\limits_{\partial A} \Big(f(x) \frac{\partial g}{\partial \nu}(x) - g(x) \frac{\partial f}{\partial \nu}(x)\Big) dS(x) = \int\limits_{A} \big(f(x) \Delta g(x) - g(x) \Delta f(x)\big) dx.$$

Proof: To prove Green's first identity apply the Gaußian Theorem to the continuously differentiable function

$$f \, \nabla g : U \to \mathbb{R}^n.$$

Hence follows

$$\int\limits_{\partial A} f(x) \frac{\partial g}{\partial \nu}(x) dS(x) = \int\limits_{\partial A} \nu(x) \cdot (f \, \nabla g)(x) dS(x)$$
$$= \int\limits_{A} \operatorname{div} (f \, \nabla g)(x) dx = \int\limits_{A} (\nabla f(x) \cdot \nabla g(x) + f(x) \Delta g(x)) dx .$$

To prove Green's second identity use Green's first identity. We obtain

$$\begin{aligned}
&\int\limits_{\partial A} \Big(f(x)\frac{\partial g}{\partial\nu}(x) - g(x)\frac{\partial f}{\partial\nu}(x)\Big)dS(x)\\
&= \int\limits_A (\nabla f(x)\cdot\nabla g(x) + f(x)\Delta g(x))dx - \int\limits_A (\nabla f(x)\cdot\nabla g(x) + g(x)\Delta f(x))dx\\
&= \int\limits_A (f(x)\Delta g(x) - g(x)\Delta f(x))dx\,. \qquad \blacksquare
\end{aligned}$$

7.6 The Integral Theorem of Stokes

Let $U \subseteq \mathbb{R}^2$ be an open set and let $A \subseteq U$ be a compact set with smooth boundary. Then the boundary ∂A is a continuously differentiable curve. If $g : U \to \mathbb{R}^2$ is continuously differentiable, the Theorem of Gauß becomes

$$\int\limits_A \left(\frac{\partial g_1}{\partial x_1}(x) + \frac{\partial g_2}{\partial x_2}(x)\right)dx = \int\limits_{\partial A} (\nu_1(x)g_1(x) + \nu_2(x)g_2(x))ds(x), \tag{7.4}$$

with the exterior unit normal vector $\nu(x) = (\nu_1(x), \nu_2(x))$. If $f : U \to \mathbb{R}^2$ is another continuously differentiable function and if we choose for g in (7.4) the function

$$g(x) := \begin{pmatrix} f_2(x)\\ -f_1(x)\end{pmatrix},$$

then we obtain

$$\begin{aligned}
\int\limits_A \left(\frac{\partial f_2}{\partial x_1}(x) - \frac{\partial f_1}{\partial x_2}(x)\right)dx &= \int\limits_{\partial A} (\nu_1(x)f_2(x) - \nu_2(x)f_1(x))ds(x)\\
&= \int\limits_{\partial A} \tau(x)\cdot f(x)ds(x)\,, \qquad (7.5)
\end{aligned}$$

where

$$\tau(x) = \begin{pmatrix} -\nu_2(x)\\ \nu_1(x)\end{pmatrix}.$$

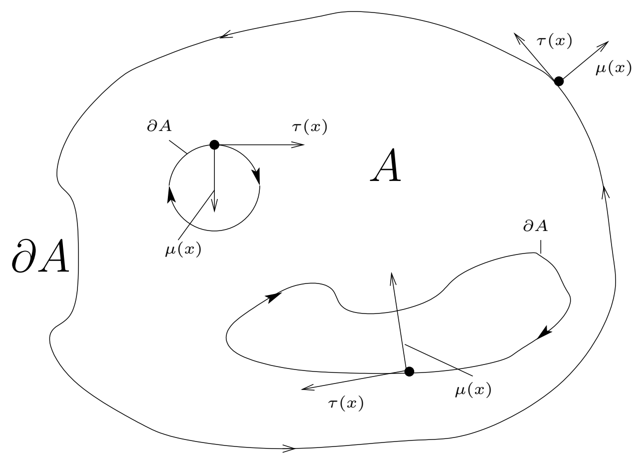
$\tau(x)$ is a unit vector perpendicular to the normal vector $\nu(x)$ and is obtained by rotating $\nu(x)$ by 90^o in the mathematically positive sense (counterclockwise). Therefore $\tau(x)$ is a unit tangent vector to ∂A in $x \in \partial A$. If we define for differentiable $f : U \to \mathbb{R}^2$ the **rotation** of f by

$$\operatorname{rot} f(x) := \frac{\partial f_2}{\partial x_1}(x) - \frac{\partial f_1}{\partial x_2}(x)\,,$$

then (7.5) can be written in the form

$$\int\limits_A \operatorname{rot} f(x)dx = \int\limits_{\partial A} \tau(x)\cdot f(x)ds(x)\,.$$

This formula is called **Stokes theorem in the plane**. Note that A is not assumed to be "simply connected". This means that A can have "holes":



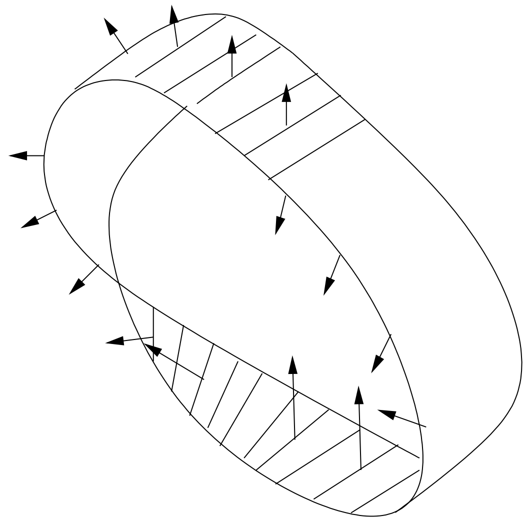
We can identify the subset $A \subseteq \mathbb{R}^2$ with the planar submanifold $Ax\{0\}$ of $\mathbb{R}^3$ and the integral over A in the Stokes formula with the surface integral over this submanifold. This interpretation suggests that this formula can be generalized and that Stokes formula is not only valid for planar submanifolds but also for more general 2-dimensional submanifolds of $\mathbb{R}^3$. As a matter of fact Stokes formula is valid for orientable submanifolds of $\mathbb{R}^3$ with boundary. To define these, we need some preparations.

Definition 7.19 Let $M \subseteq \mathbb{R}^3$ be a 2-dimensional submanifold. A unit normal vector field ν of M is a continuous mapping $\nu : M \to \mathbb{R}^3$, such that every $a \in M$ is mapped to a unit normal vector $\nu(a)$ to M at a.

A 2-dimensional submanifold M of $\mathbb{R}^3$ is called orientable, if there exists a unit normal field on M.

Example: The unit sphere $M = \{x \in \mathbb{R}^3 \mid |x| = 1\}$ is orientable. A unit normal field is $\nu(a) = \frac{a}{|a|}$, $a \in M$.

In contrast, the Möbius strip is not orientable:



Möbiusband

Definition 7.20 Let $V \subseteq \mathbb{R}^p$ be a neighborhood of 0 and $U = V \cap \big(R^{p-1} \times [0,\infty)\big)$. A function $\gamma : U \to \mathbb{R}^n$, which is continuously differentiable up to the boundary and for which $\gamma'(u)$ has rank p for all $u \in U$, is called a parametrization of a surface patch with boundary. If γ is injective and has a continuous inverse, then γ is called a simple parametrization and $F = \gamma(U)$ is called a simple p-dimensional surface patch with boundary. The set $\partial F = \gamma\big(V \cap (\mathbb{R}^{p-1} \times \{0\})\big) \subseteq \mathbb{R}^n$ is called the boundary of F.

Note that ∂F is a simple $(p-1)$-dimensional surface patch with parametrization given by $u' \mapsto \gamma(u', 0)$. We generalize Definition 7.5 of a submanifold and call a set $M \subset \mathbb{R}^n$ a p-dimensional submanifold with boundary, if the sets $M \cap V(x)$ in this definition are simple p-dimensional surface patches with or without boundary $\partial\big(M \cap V(x)\big)$, and if the boundary of M defined by

$$\partial M = \bigcup_{x \in M} \partial\big(M \cap V(x)\big)$$

is not empty. ∂M is a $(p-1)$-dimensional submanifold of $\mathbb{R}^n$.

For all the points x of a p-dimensional submanifold M with boundary including the boundary points the tangential space $T_x M$ is given by Definition 7.6.

Let M be a two-dimensional orientable submanifold in $\mathbb{R}^3$ with boundary. Then ∂M is a one-dimensional submanifold of $\mathbb{R}^3$, a curve. At $x \in \partial M$ the tangent space $T_x(\partial M)$ is one-dimensional, the tangent space $T_x M$ is two-dimensional. Therefore $T_x M$ contains exactly one unit vector $\mu(x)$, which is normal to $T_x(\partial M)$ and points out of M. With a unit normal vector field ν on M we define a unit tangent vector field $\tau : \partial M \to \mathbb{R}^3$ by setting

$$\tau(x) = \nu(x) \times \mu(x), \quad x \in \partial M.$$

We say that the vector field τ orients ∂M positively with respect to ν.

Definition 7.21 Let $U \subseteq \mathbb{R}^3$ be an open set and $f : U \to \mathbb{R}^3$ be a differentiable function. The rotation of f

$$\operatorname{rot} f : U \to \mathbb{R}^3$$

is defined by

$$\operatorname{rot} f(x) := \begin{pmatrix} \frac{\partial f_3}{\partial x_2} - \frac{\partial f_2}{\partial x_3} \\ \frac{\partial f_1}{\partial x_3} - \frac{\partial f_3}{\partial x_1} \\ \frac{\partial f_2}{\partial x_1} - \frac{\partial f_1}{\partial x_2} \end{pmatrix}.$$

Theorem 7.22 (Integral Theorem of Stokes) *Let M be a compact two-dimensional orientable submanifold of $\mathbb{R}^3$ with boundary, let $\nu : M \to \mathbb{R}^3$ be a unit normal vector field and let $\tau : \partial M \to \mathbb{R}^3$ be a unit tangent vector field, which orients ∂M positively with respect to ν. Assume that $U \subseteq \mathbb{R}^3$ is an open set with $M \subseteq U$ and that $f : U \to \mathbb{R}^3$ is continuously differentiable. Then*

$$\int_B \nu(x) \cdot rot\, f(x) dS(x) = \int_{\partial B} \tau(x) \cdot f(x) ds(x).$$

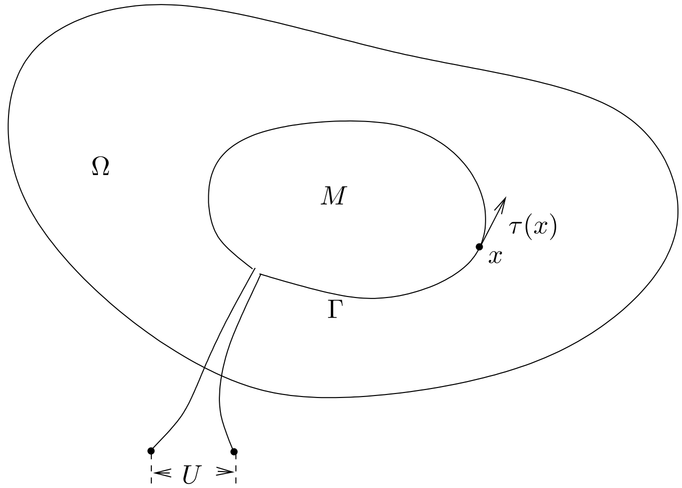
Example: Let $\Omega \subseteq \mathbb{R}^3$ be a domain in $\mathbb{R}^3$. In Ω there exists an electric field E, which depends on the location $x \in \Omega$ and the time $t \in \mathbb{R}$. Thus, E is a vector field

$$E : \Omega \times \mathbb{R} \to \mathbb{R}^3 .$$

The corresponding magnetic induction is a vector field

$$B : \Omega \times \mathbb{R} \to \mathbb{R}^3.$$

We place a wire loop Γ in Ω. This wire loop is the boundary of a surface $M \subseteq \Omega$:



If B varies in time, then an electric voltage is induced in Γ. We can calculate this voltage as follows: For all $(x,t) \in \Omega \times \mathbb{R}$ we have

$$\mathrm{rot}_x E(x,t) = -\frac{\partial}{\partial t} B(x,t)\,.$$

This is one of the Maxwell equations, which expresses Faraday's law of induction. Therefore it follows from Stokes' Theorem with a unit normal vector field $\nu : M \to \mathbb{R}^3$

$$\begin{aligned} U(t) &= \int\limits_{\Gamma} \tau(x) \cdot E(x,t) ds(x) = \int\limits_{M} \nu(x) \cdot \mathrm{rot}_x E(x,t) dS(x) \\ &= -\int\limits_{M} \nu(x) \cdot \frac{\partial}{\partial t} B(x,t) dS(x) = -\frac{\partial}{\partial t} \int\limits_{M} \nu(x) \cdot B(x,t) dS(x)\,. \end{aligned}$$

The integral $\int\limits_M \nu(x) \cdot B(x,t) dS(x)$ is called *flux of the magnetic induction through* M. Therefore $U(t)$ is equal to the negative time variation of the flux of B through M.

Appendix

German translation of Section 7

A p-dimensionale Flächen im $\mathbb{R}^m$, Flächenintegrale, Gaußscher und Stokescher Satz

A.1 p-dimensionale Flächenstücke, Untermannigfaltigkeiten

Wie früher bezeichne $L(\mathbb{R}^n, \mathbb{R}^m)$ den Vektorraum aller linearen Abbildungen von $\mathbb{R}^n$ nach $\mathbb{R}^m$. Für $A \in L(\mathbb{R}^n, \mathbb{R}^m)$ ist die Bildmenge $A(\mathbb{R}^n)$ ein linearer Unterraum von $\mathbb{R}^m$.

Definition A.1 Sei $A \in L(\mathbb{R}^n, \mathbb{R}^m)$. Als Rang von A bezeichnet man die Dimension des Unterraumes $A(\mathbb{R}^n)$.

Aus der Theorie der linearen Abbildungen ist bekannt, dass eine lineare Abbildung $A : \mathbb{R}^p \to \mathbb{R}^n$ mit Rang p injektiv ist.

Definition A.2 Sei $U \subseteq \mathbb{R}^p$ eine offene Menge und sei $p < n$. Die Abbildung $\gamma : U \to \mathbb{R}^n$ sei stetig differenzierbar und die Ableitung

$$\gamma'(u) \in L(\mathbb{R}^p, \mathbb{R}^n)$$

habe für alle $u \in U$ den Rang p. Dann heißt γ Parameterdarstellung eines p-dimensionalen Flächenstückes im $\mathbb{R}^n$. Ist $p = 1$, dann heißt γ Parameterdarstellung einer Kurve im $\mathbb{R}^n$.

Man beachte, daß γ nicht injektiv zu sein braucht. Die Fläche kann "Doppelpunkte" haben.

Beispiel 1: Sei $U = \{(u, v) \in \mathbb{R}^2 \,|\, u^2 + v^2 < 1\}$ und sei $\gamma : U \to \mathbb{R}^3$ definiert durch

$$\gamma(u, v) = \begin{pmatrix} \gamma_1(u, v) \\ \gamma_2(u, v) \\ \gamma_3(u, v) \end{pmatrix} = \begin{pmatrix} u \\ v \\ \sqrt{1 - (u^2 + v^2)} \end{pmatrix}.$$

Dann ist γ die Parameterdarstellung der oberen Hälfte der Einheitssphäre im $\mathbb{R}^3$. Denn es gilt

$$\gamma'(u, v) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -\frac{u}{\sqrt{1-(u^2+v^2)}} & -\frac{v}{\sqrt{1-(u^2+v^2)}} \end{pmatrix}.$$

Die beiden Spalten in dieser Matrix sind für alle $(u, v) \in U$ linear unabhängig, also ist der Rang 2.

Beispiel 2: Im vorangehenden Beispiel ist das Flächenstück durch den Graphen einer Funktion gegeben. Allgemeiner sei $U \subseteq \mathbb{R}^p$ eine offene Menge und sei $f : U \to \mathbb{R}^{n-p}$ stetig differenzierbar. Dann ist der Graph von f ein in den $\mathbb{R}^n$ eingebettetes p-dimensionales Flächenstück. Die Abbildung $\gamma : U \to \mathbb{R}^n$,

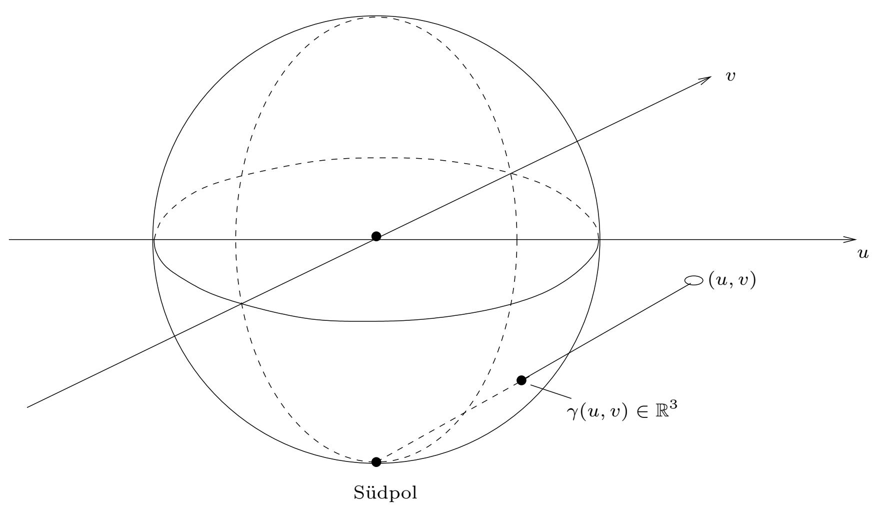
$$\begin{aligned}
\gamma_1(u) &:= u_1\\
\gamma_2(u) &:= u_2\\
&\vdots\\
\gamma_p(u) &:= u_p\\
\gamma_{p+1}(u) &:= f_1(u_1 \ldots, u_p)\\
&\vdots\\
\gamma_n(u) &:= f_{n-p}(u_1 \ldots, u_p)
\end{aligned}$$

ist eine Parameterdarstellung dieser Fläche . Denn es gilt

$$\gamma'(u) = \begin{pmatrix} 1 & \ldots & 0 \\ \vdots & \ddots & \\ 0 & \ldots & 1 \\ \partial_{x_1} f_1(u) & \ldots & \partial_{x_p} f_1(u) \\ \vdots & & \vdots \\ \partial_{x_1} f_{n-p}(u) & \ldots & \partial_{x_p} f_{n-p}(u) \end{pmatrix},$$

und alle Spalten dieser Matrix sind linear unabhängig, also ist der Rang p.

Beispiel 3: Durch stereographische Projektion kann die am Südpol gelochte Sphäre mit Mittelpunkt im Ursprung eineindeutig auf die Ebene abgebildet werden, also umgekehrt auch die Ebene auf die gelochte Sphäre:



$$\frac{\gamma_1}{\gamma_2} = \frac{u}{v}, \quad \frac{\sqrt{u^2+v^2} - \sqrt{\gamma_1^2+\gamma_2^2}}{\gamma_3} = \frac{\sqrt{u^2+v^2}}{-1}, \quad \gamma_1^2+\gamma_2^2+\gamma_3^2 = 1.$$

Aus den in der Abbildung angegebenen, aus den geometrischen Verhältnissen abgeleiteten Gleichungen erhält man für die Abbildung $\gamma : \mathbb{R}^2 \to \mathbb{R}^3$ der stereographischen Projektion, daß

$$\begin{aligned}\gamma_1(u,v) &= \frac{2u}{1+u^2+v^2}\\ \gamma_2(u,v) &= \frac{2v}{1+u^2+v^2}\\ \gamma_3(u,v) &= \frac{1-u^2-v^2}{1+u^2+v^2}\,.\end{aligned}$$

Die Ableitung ist

$$\gamma'(u,v) = \frac{2}{(1+u^2+v^2)^2}\begin{pmatrix} 1-u^2+v^2 & -2uv \\ -2uv & 1+u^2-v^2 \\ -2u & -2v \end{pmatrix}.$$

Für $u^2+v^2 \neq 1$ ist

$$\begin{aligned}\begin{vmatrix} \partial_u\gamma_1(u,v) & \partial_v\gamma_1(u,v) \\ \partial_u\gamma_2(u,v) & \partial_v\gamma_2(u,v) \end{vmatrix} &= (1+(v^2-u^2))(1-(v^2-u^2)) - 4u^2v^2 \\ &= 1-(v^2-u^2)^2 - 4u^2v^2 = 1-(v^2+u^2)^2 \neq 0\,.\end{aligned}$$

Für $u \neq 0$ gilt

$$\begin{aligned}\begin{vmatrix} \partial_u\gamma_2(u,v) & \partial_v\gamma_2(u,v) \\ \partial_u\gamma_3(u,v) & \partial_v\gamma_3(u,v) \end{vmatrix} &= 4uv^2 + 2u(1+u^2-v^2) \\ &= 2u(1+u^2+v^2) \neq 0\,,\end{aligned}$$

und für $v \neq 0$ entsprechend

$$\begin{vmatrix} \partial_u\gamma_1(u,v) & \partial_v\gamma_1(u,v) \\ \partial_u\gamma_3(u,v) & \partial_v\gamma_3(u,v) \end{vmatrix} = -2v(1+u^2+v^2) \neq 0\,,$$

also hat γ' immer den Rang 2, und somit ist γ eine Parameterdarstellung der Einheitssphäre bei herausgenommenem Südpol.

Beispiel 4: Es sei $\tilde{\gamma}$ die Einschränkung der Parametrisierung γ aus Beispiel 3 auf die Einheitskreisscheibe $U = \{(u,v) \in \mathbb{R}^2 \mid u^2+v^2 < 1\}$. Dies liefert eine Parametrisierung der oberen Hälfte der Einheitssphäre, die sich von der Parametrisierung aus Beispiel 1 unterscheidet.

Definition A.3 Seien $U, V \subseteq \mathbb{R}^p$ offene Mengen, $\gamma : U \to \mathbb{R}^n$, $\tilde{\gamma} : V \to \mathbb{R}^n$ seien Parameterdarstellungen von p-dimensionalen Flächenstücken. γ und $\tilde{\gamma}$ heißen äquivalent, wenn ein Diffeomorphismus $\varphi : V \to U$ existiert mit

$$\tilde{\gamma} = \gamma \circ \varphi\,.$$

Dies ist eine Äquivalenzrelation unter den Parameterdarstellungen von Flächenstücken.

Beispiel 5: Sei $\gamma : U \to \mathbb{R}^3$ die Parametrisierung der oberen Hälfte der Einheitssphäre aus Beispiel 1 und sei $\tilde{\gamma} : U \to \mathbb{R}^3$ die entsprechende Parametrisierung aus Beispiel 4. Diese Parametrisierungen sind äquivalent. Denn ein Diffeomorphismus $\varphi : U \to U$ ist gegeben durch

$$\varphi(u,v) = \begin{pmatrix} \frac{2u}{1+u^2+v^2} \\ \frac{2v}{1+u^2+v^2} \end{pmatrix}.$$

Für diesen Diffeomorphismus gilt

$$(\gamma \circ \varphi)(u,v) = \begin{pmatrix} \frac{2u}{1+u^2+v^2} \\ \frac{2v}{1+u^2+v^2} \\ \sqrt{1 - \frac{4u^2+4v^2}{(1+u^2+v^2)^2}} \end{pmatrix} = \frac{1}{1+u^2+v^2} \begin{pmatrix} 2u \\ 2v \\ 1-u^2-v^2 \end{pmatrix} = \tilde{\gamma}(u,v).$$

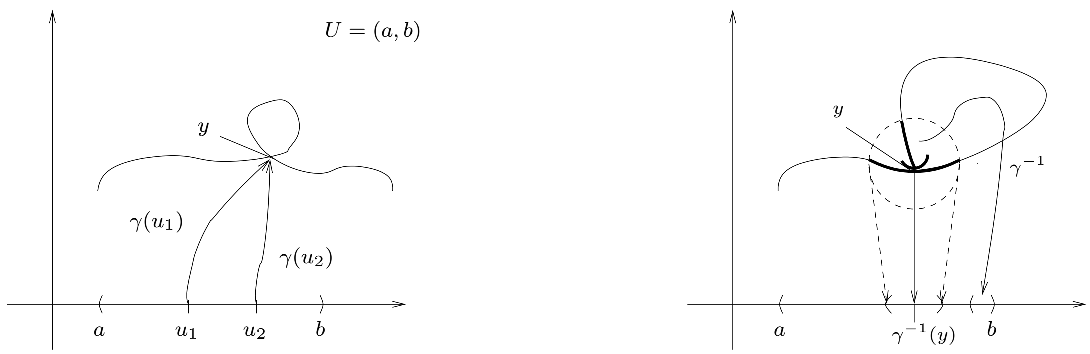
In Beispiel 3 ist eine Parameterdarstellung für die gelochte Sphäre angegeben. Für die gesamte Sphäre gibt es jedoch aus topologischen Gründen keine Parameterdarstellung $\gamma : U \to \mathbb{R}^3$. Zur Parametrisierung muss sie daher in mindestens zwei Teile aufgeteilt werden, die einzeln parametrisiert werden können. Deswegen definiert man:

Definition A.4 Sei $U \subseteq \mathbb{R}^p$ eine offene Menge. Eine Parametrisierung $\gamma : U \to \mathbb{R}^n$ eines p–dimensionalen Flächenstückes heißt einfach, wenn γ injektiv und die Umkehrabbildung γ^{-1} stetig ist. In diesem Fall bezeichnet man die Bildmenge $F = \gamma(U)$ als einfaches p–dimensionales Flächenstück.

Die unten folgende Abbildung erläutert diese Definition am Beispiel einer Kurve im $\mathbb{R}^2$.

Definition A.5 Eine Teilmenge $M \subseteq \mathbb{R}^n$ heißt p-dimensionale Untermannigfaltigkeit des $\mathbb{R}^n$, wenn es zu jedem $x \in M$ eine offene n-dimensionale Umgebung $V(x)$ und eine Abbildung γ_x gibt mit folgenden Eigenschaften:

(i) $V(x) \cap M$ ist ein einfaches p–dimensionales Flächenstück, das durch γ_x parametrisiert wird.



γ ist nicht injektiv: Die beiden verschiedenen Parameterwerte u_1 und u_2 werden auf denselben Doppelpunkt y der Kurve abgebildet.

γ^{-1} ist nicht stetig: Das Bild jeder Kugel um y enthält Punkte, deren Abstand von $\gamma^{-1}(y)$ größer als $\varepsilon = \frac{1}{2}(b - \gamma^{-1}(y))$ ist.

Figure 1: Beispiele für nicht einfache Parametrisierungen $\gamma : (a, b) \to \mathbb{R}^2$.

(ii) Sind x und y zwei Punkte aus M mit

$$N = \big(V(x) \cap M\big) \cap \big(V(y) \cap M\big) \neq \emptyset,$$

dann sind $\gamma_x : \gamma_x^{-1}(N) \to M$, $\gamma_y : \gamma_y^{-1}(N) \to M$ äquivalente Parametrisierungen von N.

Die Umkehrabbildung $\kappa_x = \gamma_x^{-1} : V(x) \cap M \to U \subseteq \mathbb{R}^p$ heißt Karte der Untermannigfaltigkeit M. Die Menge $\{\kappa_x \mid x \in M\}$ der Karten heißt Atlas von M.

Man beachte, dass zwei Karten κ_x und κ_y aus dem Atlas von M nicht notwendigerweise verschieden sein müssen. Denn gehört $y \in M$ zum Definitionsbereich $V(x) \cap M$ der Karte κ_x, dann ist nach dieser Defintion $\kappa_y = \kappa_x$ erlaubt.

Beispiel 6: Es sei $S = \{x \in \mathbb{R}^3 \mid |x| = 1\}$ die Einheitssphäre im $\mathbb{R}^3$. Die stereographische Projektion aus Beispiel 3, die die Menge $S \setminus \{(0, 0, -1)\}$ auf $U = \mathbb{R}^2$ abbildet, ist eine Karte von S, eine zweite Karte erhält man, wenn man die Menge $S \setminus \{(0, 0, 1)\}$ stereographisch vom Nordpol aus auf den $\mathbb{R}^2$ abbildet. Die Einheitsspäre ist daher eine zweidimensionale Untermannigfaltigkeit des $\mathbb{R}^3$ mit einem Atlas, der nur aus zwei Karten besteht.

Definition A.6 Es sei M eine p-dimensionale Untermannigfaltigkeit des $\mathbb{R}^n$ und x ein Punkt von M. Ist γ eine Parametrisierung von M in einer Umgebung von x mit $x = \gamma(u)$, dann ist der Wertebereich der linearen Abbildung $\gamma'(u)$ ein p-dimensionaler Unterraum von $\mathbb{R}^n$. Dieser Wertebereich heißt Tangentialraum von M im Punkt x. Man schreibt dafür $T_x(M)$ oder auch einfach T_xM.

Die Definition von $T_x(M)$ hängt nicht von der gewählten Parametrisierung ab. Denn ist $\tilde{\gamma}$ eine zu γ äquivalente Parametrisierung mit $x = \tilde{\gamma}(\tilde{u})$ und ist φ ein Diffeomorphismus mit $\gamma = \tilde{\gamma} \circ \varphi$ und mit $\tilde{u} = \varphi(u)$, dann liefert die Kettenregel

$$\gamma'(u) = \tilde{\gamma}'(\tilde{u})\varphi'(u).$$

Weil $\varphi'(u)$ eine invertierbare lineare Abbildung ist folgt, dass $\gamma'(u)$ und $\tilde{\gamma}'(\tilde{u})$ denselben Wertebereich haben.

A.2 Integration auf Flächenstücken

Sei $M \subseteq \mathbb{R}^n$ ein einfaches p–dimensionales Flächenstück, das durch $\gamma : U \to M$ parametrisiert wird. Für $1 \leq i, j \leq p$ seien die stetigen Funktionen $g_{ij} : U \to \mathbb{R}$ definiert durch

$$g_{ij}(u) = \frac{\partial \gamma}{\partial u_i}(u) \cdot \frac{\partial \gamma}{\partial u_j}(u) = \begin{pmatrix} \frac{\partial \gamma_1}{\partial u_i}(u) \\ \vdots \\ \frac{\partial \gamma_n}{\partial u_i}(u) \end{pmatrix} \cdot \begin{pmatrix} \frac{\partial \gamma_1}{\partial u_j}(u) \\ \vdots \\ \frac{\partial \gamma_n}{\partial u_j}(u) \end{pmatrix} = \sum_{k=1}^{n} \frac{\partial \gamma_k}{\partial u_i}(u) \frac{\partial \gamma_k(u)}{\partial u_j} \quad .$$

Definition A.7 Für $u \in U$ sei

$$G(u) = \begin{pmatrix} g_{11}(u) & \dots & g_{1p}(u) \\ \vdots & & \\ g_{p1}(u) & \dots & g_{pp}(u) \end{pmatrix} .$$

Die durch $g(u) := \det(G(u))$ definierte Funktion $g : U \to \mathbb{R}$ heißt Gramsche Determinante zur Parameterdarstellung γ.

Zur **Motivation** dieser Definition sei $u \in U$ fest gewählt. Dann ist

$$h \to \gamma(u) + \gamma'(u)h : \mathbb{R}^p \to \mathbb{R}^n$$

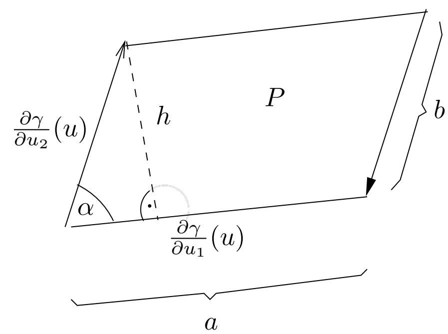
Parameterdarstellung eines ebenen Flächenstückes, das im Punkt $x = \gamma(u)$ tangential ist an das Flächenstück M. Die partiellen Ableitungen $\frac{\partial \gamma}{\partial u_1}(u), \dots, \frac{\partial \gamma}{\partial u_p}(u)$ sind Vektoren, die im Tangentialraum $T_x M$ von M im Punkt x liegen, einem p-dimensionalen linearen Unterraum von $\mathbb{R}^p$, und diesen Unterraum sogar aufspannen, weil die Matrix $\gamma'(u)$ nach Voraussetzung den Rang p hat. $\frac{\partial \gamma}{\partial u_1}(u_0), \dots, \frac{\partial \gamma}{\partial u_p}(u)$ heißen Tangentialvektoren von M im Punkt x. Die Menge

$$P = \left\{ \sum_{i=1}^{p} r_i \, \frac{\partial \gamma}{\partial u_i}\,(u) \,\middle|\, r_i \in \mathbb{R}\,,\, 0 \leq r_i \leq 1 \right\}$$

ist eine Teilmenge des Tangentialraumes, ein Parallelotop.

Satz A.8 *Es gilt $g(u) > 0$ und $\sqrt{g(u)}$ ist gleich dem p-dimensionalen Volumen des Parallelotops P.*

Der Einfachheit halber beweisen wir diesen Satz nur für $n = 2$. Im diesem Fall ist P das im Bild dargestellte Parallelogramm.



Mit $a = \left|\frac{\partial\gamma}{\partial u_1}(u)\right|$ und $b = \left|\frac{\partial\gamma}{\partial u_2}(u)\right|$ gilt

$$\begin{aligned}\sqrt{g(u)} &= \sqrt{\det(G(u))}\\ &= \sqrt{\left|\begin{matrix} \frac{\partial\gamma}{\partial u_1}(u)\cdot\frac{\partial\gamma}{\partial u_1}(u) & \frac{\partial\gamma}{\partial u_1}(u)\cdot\frac{\partial\gamma}{\partial u_2}(u) \\ \frac{\partial\gamma}{\partial u_2}(u)\cdot\frac{\partial\gamma}{\partial u_1}(u) & \frac{\partial\gamma}{\partial u_2}(u)\cdot\frac{\partial\gamma}{\partial u_2}(u)\end{matrix}\right|} = \sqrt{\left|\begin{matrix} a^2 & ab\cos\alpha \\ ab\cos\alpha & b^2\end{matrix}\right|}\\ &= \sqrt{a^2b^2 - a^2b^2\cos^2\alpha} = ab\sqrt{1-\cos^2\alpha} = ab\sin\alpha = b\cdot h = \text{ Fläche von } P\,. \qquad \blacksquare\end{aligned}$$

Definition A.9 Sei M ein p–dimensionales Flächenstück und $f : M \to \mathbb{R}$ eine Funktion. f heißt integrierbar über M, wenn die Funktion

$$u \to f(\gamma(u))\sqrt{g(u)}$$

über U integrierbar ist. Man definiert dann das Integral von f über M durch

$$\int\limits_M f(x)dS(x) := \int\limits_U f(\gamma(u))\sqrt{g(u)}du.$$

Man nennt $dS(x)$ das p-dimensionale Flächenelement von M an der Stelle x. Symbolisch gilt

$$dS(x) = \sqrt{g(u)}du\,, \quad x = \gamma(u)\,.$$

Als nächstes zeigen wir, dass diese Definition sinnvoll ist, d. h. dass der Wert des Integrals $\int\limits_U f(\gamma(u))\sqrt{g(u)}du$ sich nicht ändert wenn die Parametrisierung γ durch eine äquivalente ersetzt wird.

Satz A.10 *Seien* $U, \tilde{U} \subseteq \mathbb{R}^p$ *offene Mengen, seien* $\gamma : U \to M$ *sowie* $\tilde{\gamma} : \tilde{U} \to M$ *äquivalente Parameterdarstellungen des Flächenstückes* M *und sei* $\varphi : \tilde{U} \to U$ *ein Diffeomorphismus mit* $\tilde{\gamma} = \gamma \circ \varphi$*. Die Gramschen Determinanten zu den Parameterdarstellungen* γ *und* $\tilde{\gamma}$ *werden mit* $g : U \to \mathbb{R}$ *beziehungsweise* $\tilde{g} : \tilde{U} \to \mathbb{R}$ *bezeichnet.*

(i) Dann gilt

$$\tilde{g}(u) = g(\varphi(u))|\det \varphi'(u)|^2$$

für alle $u \in \tilde{U}$*.*

(ii) Ist $(f \circ \gamma)\sqrt{g}$ *über* U *integrierbar, dann auch* $(f \circ \tilde{\gamma})\sqrt{\tilde{g}}$ *über* $\tilde{U}$ *und es gilt*

$$\int\limits_U f(\gamma(u))\sqrt{g(u)}du = \int\limits_{\tilde{U}} f(\tilde{\gamma}(v))\sqrt{\tilde{g}(v)}dv \,.$$

Beweis: (i) Es gilt

$$g_{ij}(u) = \sum_{k=1}^{n} \frac{\partial \gamma_k(u)}{\partial u_i} \frac{\partial \gamma_k(u)}{\partial u_j} \,,$$

also ist

$$G(u) = [\gamma'(u)]^T \gamma'(u) \,.$$

Nach der Kettenregel und dem Determinantenmultiplikationssatz gilt also

$$\begin{aligned}
\tilde{g} &= \det \tilde{G} = \det([\tilde{\gamma}']^T \tilde{\gamma}') \\
&= \det([(\gamma' \circ \varphi)\varphi']^T (\gamma' \circ \varphi)\varphi') = \det(\varphi'^T [\gamma' \circ \varphi]^T [\gamma' \circ \varphi]\varphi') \\
&= (\det \varphi') \det([\gamma' \circ \varphi]^T [\gamma' \circ \varphi])(\det \varphi') = (\det \varphi')^2 (g \circ \varphi) \,.
\end{aligned}$$

(ii) Nach dem Transformationssatz ist $(f \circ \gamma)\sqrt{g}$ über U integrierbar, genau dann wenn $(f \circ \gamma \circ \varphi)\sqrt{g \circ \varphi}|\det \varphi'| = (f \circ \tilde{\gamma})\sqrt{\tilde{g}}$ über $\tilde{U}$ integrierbar ist. Außerdem ergeben Teil (i) der Behauptung und der Transformationssatz, daß

$$\int\limits_U f(\gamma(u))\sqrt{g(u)}du = \int\limits_{\tilde{U}} f((\gamma \circ \varphi)(v))\sqrt{g(\varphi(v))}|\det \varphi'(v)|dv = \int\limits_{\tilde{U}} f(\tilde{\gamma}(v))\sqrt{\tilde{g}(v)}dv \,.$$

■

A.3 Integration auf Untermannigfaltigkeiten

Nun soll die Definition des Integrals von Flächenstücken auf Untermannigfaltigkeiten verallgemeinert werden. Ich beschränke mich dabei auf p-dimensionale Untermannigfaltigkeiten M des $\mathbb{R}^n$, die durch endlich viele einfache Flächenstücke $V_1, \ldots, V_m$ überdeckt werden können. Es gelte also $M = \bigcup_{j=1}^{m} V_j$. Zu jedem $1 \leq j \leq m$ sei $U_j \subseteq \mathbb{R}^p$ eine offene Menge und $\kappa_j : V_j \subseteq M \to U_j$ eine Karte. Die Umkehrabbildungen $\gamma_j = \kappa_j^{-1} : U_j \to V_j$ sind einfache Parametrisierungen.

Definition A.11 Eine Familie $\{\alpha_j\}_{j=1}^m$ von Funktionen $\alpha_j = M \to \mathbb{R}$ heißt eine der Überdeckung $\{V_j\}_{j=1}^m$ von M untergeordnete Zerlegung der Eins aus lokal integrierbaren Funktionen, wenn gilt

(i) $0 \leq \alpha_j \leq 1, \quad \alpha_j|_{M\setminus V_j} = 0$,

(ii) $\sum_{j=1}^{m} \alpha_j(x) = 1, \quad$ für alle $x \in M$,

(iii) $\alpha_j \circ \gamma_j : U_j \to \mathbb{R}$ ist lokal integrierbar, d. h. für alle $R > 0$ existiere das Integral

$$\int_{U_j \cap \{|u|<R\}} \alpha_j(\gamma_j(u))du\,.$$

Definition A.12 Es sei M eine p-dimensionale Untermannigfaltigkeit des $\mathbb{R}^n$, zu der eine endliche Überdeckung $\{V_j\}_{j=1}^m$ aus einfachen Flächenstücken existiere. Eine Funktion $f : M \to \mathbb{R}$ heißt integrierbar über M, falls $f_{|V_j}$ für alle j integrierbar ist. Man setzt dann

$$\int_M f(x)dS(x) = \sum_{j=1}^{m} \int_{V_j} \alpha_j(x)f(x)dS(x)$$

mit einer der Überdeckung $\{V_j\}_{j=1}^m$ von M untergeordneten Partition der Eins $\{\alpha_j\}_{j=1}^m$ aus lokal integrierbaren Funktionen.

Die Funktion $\alpha_j(x)f(x)$ ist über V_j integrierbar, weil nach Voraussetzung $(f \circ \gamma_j)\sqrt{g_j}$ über U_j integrierbar ist mit der Gramschen Determinanten g_j zur Parametrisierung γ_j. Wegen $0 \leq \alpha_j(x) \leq 1$ ist also auch $(\alpha_j \circ \gamma_j)(f \circ \gamma_j)\sqrt{g_j}$ über U_j integrierbar als Produkt einer integrierbaren und einer beschränkten, lokal integrierbaren Funktion.

Es muß noch gezeigt werden, daß die Definition des Integrals unabhängig von der Wahl der Überdeckung von M durch einfache Flächenstücke und von der Wahl der Partition der Eins ist:

Satz A.13 *Sei M eine p-dimensionale Untermannigfaltigkeit im $\mathbb{R}^n$ und seien*

$$\gamma_k : U_k \to V_k\,, \quad k = 1, \ldots, m$$
$$\tilde{\gamma}_j : \tilde{U}_j \to \tilde{V}_j\,, \quad j = 1, \ldots, l$$

einfache Parametrisierungen mit $\bigcup_{k=1}^{m} V_k = \bigcup_{j=1}^{l} \tilde{V}_j = M$. *Gilt*

$$D_{jk} = \tilde{V}_j \cap V_k \neq \emptyset,$$

dann seien

$$U_{kj} = \gamma_k^{-1}(D_{jk})\,, \quad \tilde{U}_{kj} = \tilde{\gamma}_j^{-1}(D_{jk})$$

Jordan-messbare Teilmengen von $\mathbb{R}^p$ *und*

$$\gamma_k : U_{kj} \to D_{jk}, \quad \tilde{\gamma}_j : \tilde{U}_{kj} \to D_{jk}$$

seien äquivalente Parametrisierungen.

Das Funktionensystem $\{\alpha_k\}_{k=1}^m$ *sei eine der Überdeckung* $\{V_k\}_{k=1}^m$ *und das System* $\{\beta_j\}_{j=1}^l$ *eine der Überdeckung* $\left\{\tilde{V}_j\right\}_{j=1}^l$ *untergeordnete Zerlegung der Eins. Dann gilt*

$$\sum_{k=1}^{m} \int\limits_{V_k} \alpha_k(x) f(x) dS(x) = \sum_{j=1}^{l} \int\limits_{\tilde{V}_j} \beta_j(x) f(x) dS(x)\,. \tag{A.1}$$

Beweis: Zunächst zeige ich, daß $\beta_j \alpha_k f$ sowohl über V_k als auch über $\tilde{V}_j$ integrierbar ist mit

$$\int\limits_{\tilde{V}_j} \beta_j(x) \alpha_k(x) f(x) dS(x) = \int\limits_{V_k} \beta_j(x) \alpha_k(x) f(x) dS(x)\,. \tag{A.2}$$

Um dies einzusehen, seien g_k beziehungsweise $\tilde{g}_j$ die Gramschen Determinanten zu γ_k und $\tilde{\gamma}_j$. Wenn die Funktion $[(\alpha_k f) \circ \gamma_k]\sqrt{g_k}$ über U_k integrierbar ist, dann ist diese Funktion auch über U_{kj} integrierbar, weil U_{kj} eine Jordan-messbare Teilmenge von U_k ist. Nach Satz A.10 ist dann $[(\alpha_k f) \circ \tilde{\gamma}_j]\sqrt{\tilde{g}_j}$ über $\tilde{U}_{jk}$ integrierbar. Nach Voraussetzung ist $\beta_j \circ \tilde{\gamma}_j$ über $\tilde{U}_j$ lokal integrierbar, also ist diese Funktion auch über $\tilde{U}_{jk}$ lokal integrierbar, weil $\tilde{U}_{jk}$ eine Jordan-messbare Teilmenge von $\tilde{U}_j$ ist. Wegen $0 \leq \beta_j \circ \tilde{\gamma}_j \leq 1$ folgt, daß das Produkt

$$[(\beta_j \alpha_k f) \circ \tilde{\gamma}_j]\sqrt{\tilde{g}_j} = (\beta_j \circ \tilde{\gamma}_j)[(\alpha_k f) \circ \tilde{\gamma}_j]\sqrt{\tilde{g}_j}$$

über $\tilde{U}_{jk}$ integrierbar ist, und wegen der Äquivalenz der Parametrisierungen $\gamma_k : U_{kj} \to D_{jk}$ und $\tilde{\gamma}_j : \tilde{U}_{jk} \to D_{jk}$ gilt folglich

$$\int\limits_{\tilde{U}_{jk}} [(\beta_j \alpha_k f) \circ \tilde{\gamma}_j]\sqrt{\tilde{g}_j} du = \int\limits_{U_{kj}} [(\beta_j \alpha_k f) \circ \gamma_k]\sqrt{g_k} du\,. \tag{A.3}$$

Da $(\beta_j\alpha_k)(x) = 0$ für alle $x \in M \setminus D_{jk}$, ist $[(\beta_j\alpha_k f) \circ \tilde{\gamma}_j](u) = 0$ für alle $u \in \tilde{U}_j \setminus \tilde{U}_{jk}$ und $[(\beta_j\alpha_k f) \circ \gamma_k](u) = 0$ für alle $u \in U_k \setminus U_{kj}$, also können in (A.3) die Integrationsbereiche ausgedehnt werden ohne Änderung der Integrale. Dies bedeutet, dass

$$\int\limits_{\tilde{U}_j} [(\beta_j\alpha_k f) \circ \tilde{\gamma}_j]\sqrt{\tilde{g}_j}du = \int\limits_{U_k} [(\beta_j\alpha_k f) \circ \gamma_k]\sqrt{g_k}du\,,$$

gilt, und diese Gleichung ist äquivalent (A.2), weil $\tilde{\gamma}_j : \tilde{U}_j \to \tilde{V}_j$ und $\gamma_k : U_k \to V_k$ Parametrisierungen sind.

Zusammen mit $\sum_{j=1}^{\ell} \beta_j(x) = 1$ und $\sum_{k=1}^{m} \alpha_k(x) = 1$ folgt aus (A.2), dass

$$\sum_{k=1}^{m} \int\limits_{V_k} \alpha_k(x)f(x)dS(x) = \sum_{k=1}^{m} \int\limits_{V_k} \sum_{j=1}^{\ell} \beta_j(x)\alpha_k(x)f(x)dS(x)$$

$$= \sum_{j=1}^{\ell} \sum_{k=1}^{m} \int\limits_{V_k} \beta_j(x)\alpha_k(x)f(x)dS(x) = \sum_{j=1}^{\ell} \sum_{k=1}^{m} \int\limits_{\tilde{V}_j} \beta_j(x)\alpha_k(x)f(x)dS(x)$$

$$= \sum_{j=1}^{\ell} \int\limits_{\tilde{V}_j} \sum_{k=1}^{m} \alpha_k(x)\beta_j(x)f(x)dS(x) = \sum_{j=1}^{\ell} \int\limits_{\tilde{V}_j} \beta_j(x)f(x)dS(x)\,,$$

und dies ist die Gleichung (A.1). ∎

A.4 Der Gaußsche Integralsatz

Zur Formulierung des Gaußschen Satzes benötige ich zwei Definitionen:

Definition A.14

(i) Sei $A \subseteq \mathbb{R}^n$ eine kompakte Menge. Man sagt, A habe glatten Rand, wenn ∂A eine $(n-1)$-dimensionale Untermannigfaltigkeit von $\mathbb{R}^n$ ist.

(ii) Sei $x \in A$. Ist der von Null verschiedene Vektor $\nu \in \mathbb{R}^n$ orthogonal zu allen Vektoren im Tangentialraum $T_x(\partial A)$ von ∂A im Punkt x, dann heißt ν Normalenvektor zu ∂A im Punkt x. Gilt $|\nu| = 1$, dann heißt ν Einheitsnormalenvektor. Zeigt ν ins Äußere von A, dann heißt ν äußerer Normalenvektor.

Definition A.15 (Divergenz) Sei $U \subseteq \mathbb{R}^n$ eine offene Menge und $f : U \to \mathbb{R}^n$ sei differenzierbar. Dann ist die Funktion $\operatorname{div} f : U \to \mathbb{R}$ definiert durch

$$\operatorname{div} f(x) := \sum_{i=1}^{n} \frac{\partial}{\partial x_i} f_i(x)\,.$$

Man nennt $\operatorname{div} f$ die Divergenz von f.

Satz A.16 (Gaußscher Integralsatz) *Sei $A \subseteq \mathbb{R}^n$ eine kompakte Menge mit glattem Rand, $U \subseteq \mathbb{R}^n$ sei eine offene Menge mit $A \subseteq U$ und $f : U \to \mathbb{R}^n$ sei stetig differenzierbar. $\nu(x)$ bezeichne den äußeren Einheitsnormalenvektor an ∂A im Punkt x. Dann gilt*

$$\int_{\partial A} \nu(x) \cdot f(x) dS(x) = \int_A \text{div } f(x) dx \,.$$

Für $n = 1$ lautet der Satz: Seien $a, b \in \mathbb{R}$, $a < b$. Dann ist

$$f(b) - f(a) = \int_a^b \frac{d}{dx} f(x) dx \,,$$

und man sieht, daß der Gaußsche Satz die Verallgemeinerung des Hauptsatzes der Differential- und Integralrechnung auf den $\mathbb{R}^n$ ist.

Anwendungsbeispiel: Ein Körper A befinde sich in einer Flüssigkeit mit dem spezifischen Gewicht c, deren Oberfläche mit der Ebene $x_3 = 0$ zusammenfalle. Der Druck im Punkt $x = (x_1, x_2, x_3) \in \mathbb{R}^3$ mit $x_3 < 0$ ist dann

$$-cx_3 \,.$$

Ist $x \in \partial A$, dann resultiert aus diesem Druck die Kraft pro Flächeneinheit

$$-cx_3(-\nu(x)) = cx_3\nu(x)$$

auf den Körper in Richtung des äußeren Normaleneinheitsvektors $\nu(x)$ an ∂A im Punkt x. Für die gesamte Oberflächenkraft ergibt sich

$$K = \begin{pmatrix} K_1 \\ K_2 \\ K_3 \end{pmatrix} = \int_{\partial A} cx_3\nu(x) dS(x) \,.$$

Anwendung des Gaußschen Satzes auf die Funktionen $f_1, \ldots, f_3 : A \to \mathbb{R}^3$ mit

$$f_1(x_1, x_2, x_3) = (x_3, 0, 0), \quad f_2(x_1, x_2, x_3) = (0, x_3, 0), \quad f_3(x_1, x_2, x_3) = (0, 0, x_3)$$

liefert für $i = 1, 2$

$$K_i = \int_{\partial A} cx_3\nu_i(x) dS(x) = c \int \nu(x) \cdot f_i(x) dS(x) = c \int_A \frac{\partial}{\partial x_i} x_3 dx = 0,$$

und für $i = 3$

$$K_3 = \int_{\partial A} cx_3\nu_3(x) dS(x) = c \int \nu(x) \cdot f_3(x) dS(x) = c \int_A \frac{\partial}{\partial x_3} x_3 dx = c \int_A dx = c \,\text{Vol}\,(A) \,.$$

K ist somit in Richtung der positiven x_3-Achse gerichtet, also erfährt A einen Auftrieb der Größe $c \,\text{Vol}\,(A)$. Dies ist gleich dem Gewicht der verdrängten Flüssigkeit.

A.5 Greensche Formeln

Es sei $U \subseteq \mathbb{R}^n$ eine offene Menge, $A \subseteq U$ sei eine kompakte Menge mit glattem Rand, und für $x \in \partial A$ sei $\nu(x)$ die äußere Einheitsnormale an ∂A im Punkt x. Für differenzierbares $f : U \to \mathbb{R}$ bezeichne $\nabla f(x) \in \mathbb{R}^n$ im Folgenden den Gradienten grad $f(x)$. Man nennt ∇ den Nablaoperator.

Definition A.17 Die Funktion $f : U \to \mathbb{R}$ sei stetig differenzierbar. Dann definiert man die Normalableitung von f im Punkt $x \in \partial A$ durch

$$\frac{\partial f}{\partial \nu}(x) := f'(x)\nu(x) = \nu(x) \cdot \nabla f(x) = \sum_{i=1}^{n} \frac{\partial f(x)}{\partial x_i}\nu_i(x)\,.$$

Die Normalableitung von f ist die Richtungsableitung von f in Richtung von ν. Für zweimal differenzierbares $f : U \to \mathbb{R}$ sei

$$\Delta f(x) := \sum_{i=1}^{n} \frac{\partial^2}{\partial x_i^2} f(x).$$

Δ heißt Laplace-Operator.

Satz A.18 *Für $f, g \in C^2(U, \mathbb{R})$ gelten*

(i) **Erste Greensche Formel:**

$$\int_{\partial A} f(x)\frac{\partial g}{\partial \nu}(x)dS(x) = \int_A \big(\nabla f(x) \cdot \nabla g(x) + f(x)\Delta g(x)\big)dx$$

(ii) **Zweite Greensche Formel:**

$$\int_{\partial A} \Big(f(x)\frac{\partial g}{\partial \nu}(x) - g(x)\frac{\partial f}{\partial \nu}(x)\Big)dS(x) = \int_A \big(f(x)\Delta g(x) - g(x)\Delta f(x)\big)dx\,.$$

Beweis: Zum Beweis der ersten Greenschen Formel wende den Gaußschen Integralsatz auf die stetig differenzierbare Funktion

$$f\,\nabla g : U \to \mathbb{R}^n$$

an. Es folgt

$$\int_{\partial A} f(x)\frac{\partial g}{\partial \nu}(x)dS(x) = \int_{\partial A} \nu(x) \cdot (f\,\nabla g)(x)dS(x)$$
$$= \int_A \operatorname{div}(f\,\nabla g)(x)dx = \int_A (\nabla f(x) \cdot \nabla g(x) + f(x)\Delta g(x))dx\,.$$

Für den Beweis der zweiten Greenschen Formel benützt man die erste Greensche Formel. Danach gilt

$$\begin{aligned}
&\int\limits_{\partial A} \Big(f(x)\frac{\partial g}{\partial \nu}(x) - g(x)\frac{\partial f}{\partial \nu}(x)\Big) dS(x) \\
&= \int\limits_A (\nabla f(x)\cdot\nabla g(x) + f(x)\Delta g(x))dx - \int\limits_A (\nabla f(x)\cdot\nabla g(x) + g(x)\Delta f(x))dx \\
&= \int\limits_A (f(x)\Delta g(x) - g(x)\Delta f(x))dx\,.
\end{aligned}$$

■

A.6 Der Stokesche Integralsatz

Sei $U \subseteq \mathbb{R}^2$ eine offene Menge und sei $A \subseteq U$ eine kompakte Menge mit glattem Rand. Dann ist der Rand ∂A eine stetig differenzierbare Kurve. Für stetig differenzierbares $g : U \to \mathbb{R}^2$ nimmt der Gaußsche Satz die Form

$$\int\limits_A \left(\frac{\partial g_1}{\partial x_1}(x) + \frac{\partial g_2}{\partial x_2}(x)\right) dx = \int\limits_{\partial A} (\nu_1(x)g_1(x) + \nu_2(x)g_2(x))ds(x) \tag{A.4}$$

an mit dem äußeren Normaleneinheitsvektor $\nu(x) = (\nu_1(x), \nu_2(x))$. Ist $f : U \to \mathbb{R}^2$ eine andere stetig differenzierbare Funktion und wählt man für g in (A.4) die Funktion

$$g(x) := \begin{pmatrix} f_2(x) \\ -f_1(x) \end{pmatrix},$$

dann erhält man

$$\begin{aligned}
\int\limits_A \left(\frac{\partial f_2}{\partial x_1}(x) - \frac{\partial f_1}{\partial x_2}(x)\right) dx &= \int\limits_{\partial A} (\nu_1(x)f_2(x) - \nu_2(x)f_1(x))ds(x) \\
&= \int\limits_{\partial A} \tau(x)\cdot f(x)ds(x)\,, \qquad \text{(A.5)}
\end{aligned}$$

mit

$$\tau(x) = \begin{pmatrix} -\nu_2(x) \\ \nu_1(x) \end{pmatrix}.$$

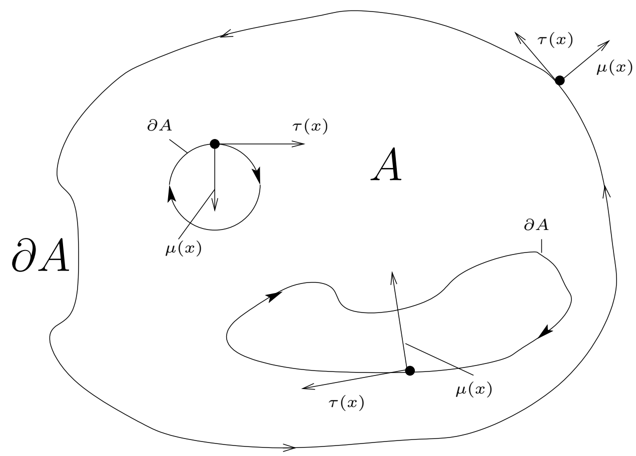
$\tau(x)$ ist ein Einheitsvektor, der senkrecht auf dem Normalenvektor $\nu(x)$ steht, also ist $\tau(x)$ ein Einheitstangentenvektor an ∂A im Punkt $x \in \partial A$, und zwar derjenige, den man aus $\nu(x)$ durch Drehung um 90^o im mathematisch positiven Sinn erhält. Definiert man für differenzierbares $f : U \to \mathbb{R}^2$ die **Rotation** von f durch

$$\operatorname{rot} f(x) := \frac{\partial f_2}{\partial x_1}(x) - \frac{\partial f_1}{\partial x_2}(x)\,,$$

dann kann (A.5) in der Form

$$\int_A \operatorname{rot} f(x)dx = \int_{\partial A} \tau(x) \cdot f(x)ds(x)$$

geschrieben werden. Diese Formel heißt **Stokescher Satz in der Ebene**. Man beachte, dass A nicht als "einfach zusammenhängend" vorausgesetzt wurde. Das heißt, dass A "Löcher" haben kann:



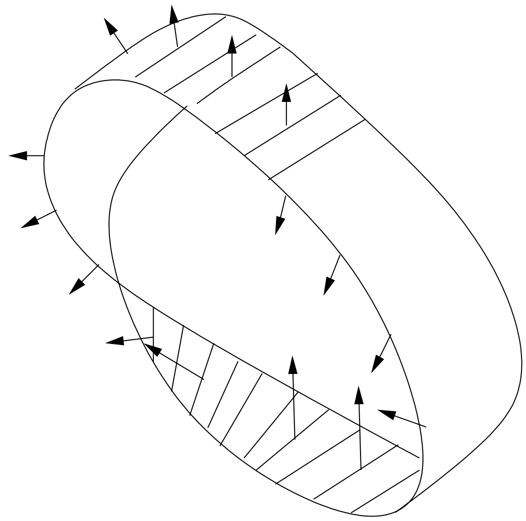
Man kann die Teilmenge $A \subseteq \mathbb{R}^2$ mit der ebenen Untermannigfaltigkeit $A \times \{0\}$ im $\mathbb{R}^3$ identifizieren und das Integral über A im Stokeschen Satz mit dem Flächenintegral über diese Untermannigfaltigkeit. Diese Interpretation legt die Vermutung nahe, dass diese Formel verallgemeinert werden kann und der Stokesche Satz nicht nur für ebene Untermannigfaltigkeiten, sondern für allgemeinere 2-dimensionale Untermannigfaltigkeiten des $\mathbb{R}^3$ gilt. In der Tat gilt der Stokesche Satz für orientierbare Untermannigfaltigkeiten des $\mathbb{R}^3$, die folgendermaßen definiert sind:

Definition A.19 Sei $M \subseteq \mathbb{R}^3$ eine 2-dimensionale Untermannigfaltigkeit. Unter einem Einheitsnormalenfeld ν von M versteht man eine stetige Abbildung $\nu : M \to \mathbb{R}^3$ mit der Eigenschaft, dass für jedes $a \in M$ der Vektor $\nu(a)$ ein Einheitsnormalenvektor von M in a ist.

Eine 2-dimensionale Untermannigfaltigkeit M des $\mathbb{R}^3$ heißt orientierbar, wenn ein Einheitsnormalenfeld auf M existiert.

Beispiel: Die Einheitssphäre $M = \{x \in \mathbb{R}^3 \mid |x| = 1\}$ ist orientierbar. Ein Einheitsnormalenfeld ist $\nu(a) = \frac{a}{|a|}$, $a \in M$.

Dagegen ist das Möbiusband nicht orientierbar:



Möbiusband

Definition A.20 Sei $U \subseteq \mathbb{R}^3$ eine offene Menge und $f : U \to \mathbb{R}^3$ differenzierbar. Die Rotation von f

$$\operatorname{rot} f : U \to \mathbb{R}^3$$

ist definiert durch

$$\operatorname{rot} f(x) := \begin{pmatrix} \dfrac{\partial f_3}{\partial x_2}(x) - \dfrac{\partial f_2}{\partial x_3}(x) \\ \dfrac{\partial f_1}{\partial x_3}(x) - \dfrac{\partial f_3}{\partial x_1}(x) \\ \dfrac{\partial f_2}{\partial x_1}(x) - \dfrac{\partial f_1}{\partial x_2}(x) \end{pmatrix} .$$

Satz A.21 (Stokesscher Integralsatz) *Sei M eine 2-dimensionale orientierbare Untermannigfaltigkeit des $\mathbb{R}^3$, und sei $\nu : M \to \mathbb{R}^3$ ein Einheitsnormalenfeld. Sei $B \subseteq M$ eine kompakte Menge mit glattem Rand (d. h. ∂B sei eine differenzierbare Kurve.) Für $x \in \partial B$ sei $\mu(x) \in T_x M$ der aus B hinausweisende Einheitsnormalenvektor. Außerdem sei*

$$\tau(x) = \nu(x) \times \mu(x) \quad x \in \partial B .$$

$\tau(x)$ ist ein Einheitstangentenvektor an ∂B. Schließlich seien $U \subseteq \mathbb{R}^3$ eine offene Menge mit $B \subseteq U$ und $f : U \to \mathbb{R}^3$ eine stetig differenzierbare Funktion. Dann gilt:

$$\int_B \nu(x) \cdot \operatorname{rot} f(x) dS(x) = \int_{\partial B} \tau(x) \cdot f(x) ds(x) .$$

Beispiel: Sei $\Omega \subseteq \mathbb{R}^3$ ein Gebiet im $\mathbb{R}^3$. In Ω existiere ein elektrisches Feld E, das vom Ort $x \in \Omega$ und der Zeit $t \in \mathbb{R}$ abhängt. Also gilt

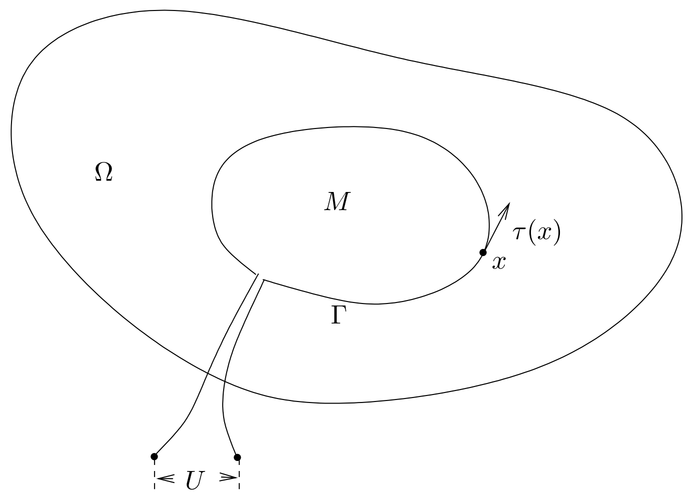
$$E : \Omega \times \mathbb{R} \to \mathbb{R}^3 \, .$$

Ebenso sei

$$B : \Omega \times \mathbb{R} \to \mathbb{R}^3$$

die magnetische Induktion.

Sei $\Gamma \subseteq \Omega$ eine Drahtschleife. Diese Drahtschleife berande eine Fläche $M \subseteq \Omega$:



In Γ wird durch die Änderung von B eine elektrische Spannung U induziert. Diese Spannung kann folgendermaßen berechnet werden: Es gilt für alle $(x,t) \in \Omega \times \mathbb{R}$

$$\mathrm{rot}_x E(x,t) = -\frac{\partial}{\partial t} B(x,t) \, .$$

Dies ist eine der Maxwellschen Gleichungen. Also folgt aus dem Stokeschen Satz mit einem Einheitsnormalenfeld $\nu : M \to \mathbb{R}^3$

$$\begin{aligned} U(t) &= \int_\Gamma \tau(x) \cdot E(x,t) ds(x) = \int_M \nu(x) \cdot \mathrm{rot}_x E(x,t) dS(x) \\ &= -\int_M \nu(x) \cdot \frac{\partial}{\partial t} B(x,t) dS(x) = -\frac{\partial}{\partial t} \int_M \nu(x) \cdot B(x,t) dS(x) \, . \end{aligned}$$

Das Integral $\int_M \nu(x) \cdot B(x,t) dS(x)$ heißt **Fluß der magnetischen Induktion durch** M. Somit ist $U(t)$ gleich der negativen zeitlichen Änderung des Flusses von B durch M.