Mathematics for Computer Scientists III

Prof. Dr. Henryk Zähle

Winter Term 2024/25

Saarland University

Table of contents

1. Multidimensional analysis

1. Multidimensional analysis

1.1 The space $\mathbb{R}^n$

Forster, Analysis 2, Sections I.1, I.2

Recall

The n-dimensional Euclidean space is defined by

$$\mathbb{R}^n := \big\{(x_1, \ldots, x_n) : x_1, \ldots, x_n \in \mathbb{R}\big\}.$$

It is known from Part II that it is an n-dimensional vector space if it is equipped with

vector add.: $(x_1, \ldots, x_n) + (y_1, \ldots, y_n) := (x_1 + y_1, \ldots, x_n + y_n)$

scalar mult.: $\alpha(x_1, \ldots, x_n) := (\alpha x_1, \ldots, \alpha x_n)$ for $\alpha \in \mathbb{R}$

A vector $(x_1, \ldots, x_n)$ in $\mathbb{R}^n$ can be identified with both

the $1 \times n$ matrix $[x_1 \ \cdots \ x_n]$,

the $n \times 1$ matrix $[x_1 \ \cdots \ x_n]^{\mathsf{T}}$.

$\mathrm{e}_i := (0, \ldots, 0, 1, 0, \ldots, 0)$ i-th unit vector, $i = 1, \ldots, n$.

Recall

For vectors $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$ in $\mathbb{R}^n$ we define

$\|x\| := (\sum_{i=1}^n x_i^2)^{1/2}$ **Euclidean norm** of x,

$\langle x, y \rangle := \sum_{i=1}^n x_i y_i$ **Euclidean scalar product** of x and y.

Note that

$\|x\|$ is the length of the vector x (☞ slide 5),

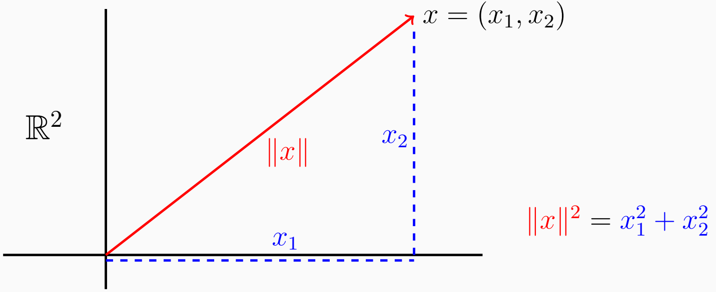
$\|x - y\|$ is the distance between x and y (☞ slide 6),

The mapping $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ is a norm on $\mathbb{R}^n$ (☞ slide 7),

$\langle x, y \rangle = [x_1 \; \cdots \; x_n][y_1 \; \cdots \; y_n]^\mathsf{T}$.

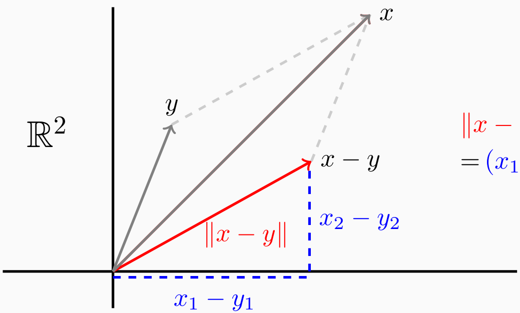
Recall

$\|x\|$ is the length of the vector x:



Recall

$\|x-y\|$ is the distance between x and y:



$$\|x-y\|^2 = (x_1-y_1)^2 + (x_2-y_2)^2$$

Recall

The mapping $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ is a norm on $\mathbb{R}^n$, i.e.

- $\|x\| \geq 0$ for all $x \in \mathbb{R}^n$.
  $\|x\| = 0$ if and only if $x = (0, \ldots, 0)$.
- $\|\lambda x\| = |\lambda| \cdot \|x\|$ for all $\lambda \in \mathbb{R}$ and $x \in \mathbb{R}^n$.
- $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in \mathbb{R}^n$.

Example

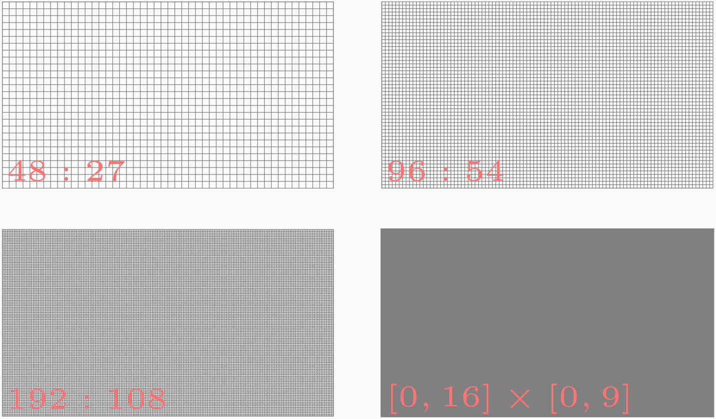
In image analysis, the computer screen is often regraded as a subset D of $\mathbb{R}^2$, e.g.

$$D := [0,16] \times [0,9] \subseteq \mathbb{R}^2.$$

In this case, the elements of D can be seen as pixels and $\|x-y\|$ is the distance between pixel x and pixel y.

- In point of fact, a $16:9$ screen is a "grid" consisting of, for instance, 1920×1080 pixels (note that $1920:1080 = 16:9$).
- Therefore, D is actually just a continuous approximation of a discrete set as $\{1,\dots,1920\} \times \{1,\dots,1080\}$ (☞ slide 9).

Example



Definition 1.1.1

Let $(x^N)_{N\in\mathbb{N}}$ be a sequence in $\mathbb{R}^n$ and $x^0 \in \mathbb{R}^n$.

$(x^N)_{N\in\mathbb{N}}$ is said to **converge to** x^0 if $\lim_{N\to\infty} \|x^N - x^0\| = 0$, i.e.

$$\forall\, \varepsilon > 0: \quad \exists\, N_0 \in \mathbb{N}: \quad \forall\, N \geq N_0: \quad \|x^N - x^0\| \leq \varepsilon.$$

In this case, we also write $\lim_{N\to\infty} x^N = x^0$ or $x^N \to x^0$.

Lemma 1.1.2

Let $(x^N)_{N\in\mathbb{N}} = (x_1^N, \ldots, x_n^N)_{N\in\mathbb{N}}$ *be a sequence in* $\mathbb{R}^n$*,*
and $x^0 = (x_1^0, \ldots, x_n^0) \in \mathbb{R}^n$*. Then:*

$$x^N \to x^0 \qquad \textit{(in } \mathbb{R}^n\textit{)}$$

if and only if

$$x_i^N \to x_i^0 \qquad \textit{(in } \mathbb{R}\textit{)} \qquad \textit{for all } i = 1, \ldots, n.$$

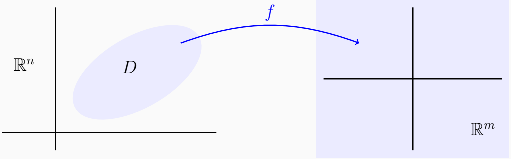
1. Multidimensional analysis

1.2 $\mathbb{R}^m$-valued functions on $\mathbb{R}^n$

Courant/John, Intro. to Calculus and Analysis II/1, Section 1.2

The basic object of Chapter 1 is a function defined on a subset D of $\mathbb{R}^n$ and taking values in $\mathbb{R}^m$:

$$f : \quad D \quad \longrightarrow \; \mathbb{R}^m$$
$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \longmapsto f(x_1, \dots, x_n) = \begin{bmatrix} f_1(x_1, \dots, x_n) \\ \vdots \\ f_m(x_1, \dots, x_n) \end{bmatrix}$$



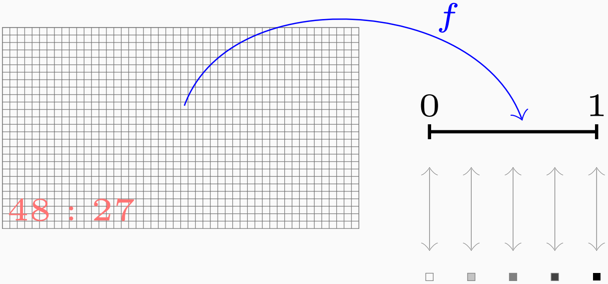
Definition 1.2.1

In the framework above, the set D is called the **domain of** f.

Example

In image analysis, monochrome images are obtained by assigning a relative luminance (a value in the unit interval $[0,1]$) to each pixel.

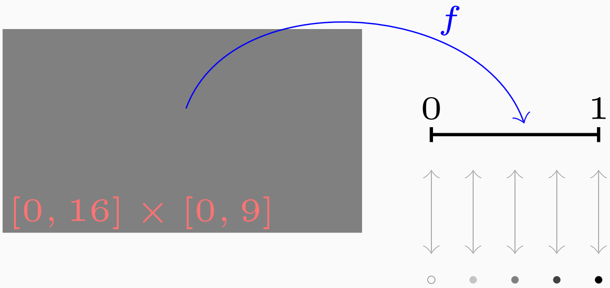
That is, a monochrome image on a 16:9 computer screen can be seen as a mapping from (e.g.) $\{1,\ldots,48\} \times \{1,\ldots,27\}$ to $[0,1]$ ($\subseteq \mathbb{R}$).



Example

For a realistic screen resolution (as 1920×1080) the discrete domain D $(= \{1, \ldots, 1920\} \times \{1, \ldots, 1080\})$ can be replaced by its continuous approximation $D := [0, 16] \times [0, 9]$ (☞ slide 9).

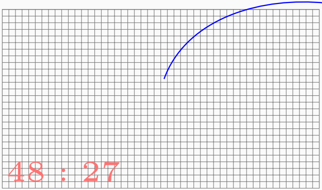
Thus, to some extent a monochrome image on the screen can also be seen as a mapping f from $[0, 16] \times [0, 9]$ $(\subseteq \mathbb{R}^2)$ to $([0, 1] \subseteq)$ $\mathbb{R}$.



Example

In image analysis, colour images are obtained by assigning to each pixel a triplet (R, G, B) in a RGB color space as, for instance, $\{0, \ldots, 255\}^3$.

That is, a colour image on a 16:9 computer screen can be seen as a mapping from (e.g.) $\{1, \ldots, 48\} \times \{1, \ldots, 27\}$ to $\{0, \ldots, 255\}^3$.



$$\xrightarrow{f} \quad \{0, \ldots, 255\} \times \{0, \ldots, 255\} \times \{0, \ldots, 255\}$$

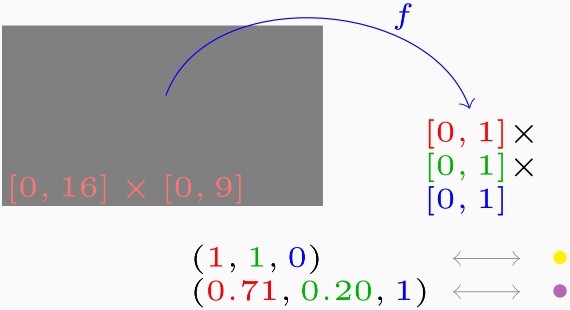
$$(255, 255, 0) \longleftrightarrow$$

$$(180, 50, 255) \longleftrightarrow$$

Example

For a realistic screen resolution (as 1920×1080) the discrete domain D $(= \{1, \ldots, 1920\} \times \{1, \ldots, 1080\})$ can be replaced by its continuous approximation $D := [0, 16] \times [0, 9]$ (☞ slide 9).

Thus, to some extent a colour image on the screen can also be seen as a mapping f from $[0, 16] \times [0, 9]$ $(\subseteq \mathbb{R}^2)$ to $([0, 1]^3 \subseteq)$ $\mathbb{R}^3$.



Definition 1.2.2

Let $D \subseteq \mathbb{R}^n$ and $f : D \to \mathbb{R}^m$ be any function.

The **graph of** f is the set defined by

$$G_f := \big\{(x_1, \ldots, x_n, y_1, \ldots, y_m) \in D \times \mathbb{R}^m \ : \\ f(x_1, \ldots, x_n) = (y_1, \ldots, y_m)\big\}.$$

The **level set of** f **at level** $c = (c_1, \ldots, c_m) \in \mathbb{R}^m$ is the set defined by

$$L_f(c) := \big\{(x_1, \ldots, x_n) \in D \,:\, f(x_1, \ldots, x_n) = (c_1, \ldots, c_m)\big\}.$$

Visualisation of f when $n = m = 1$:

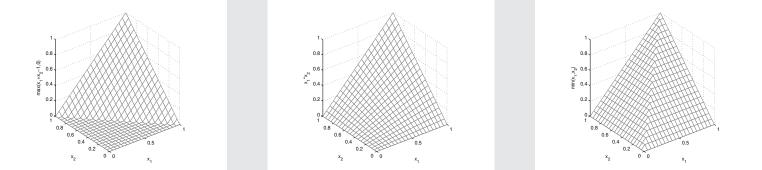
by means of G_f

Visualisation of f when $n = 2$, $m = 1$:

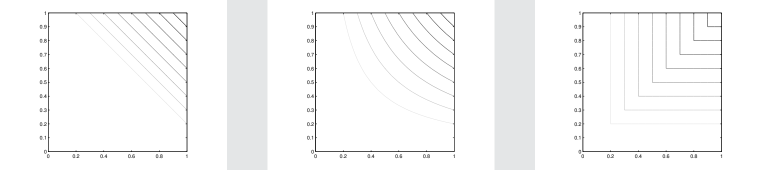
by means of G_f, or by means of $L_f(c)$ for $c \in C \subseteq \mathbb{R}$ (☞ 1.2.3)

Example 1.2.3

$f(x_1, x_2) :=$	$f(x_1, x_2) :=$	$f(x_1, x_2) :=$
$\max\{x_1 + x_2 - 1, 0\}$	$x_1 \cdot x_2$	$\min\{x_1, x_2\}$



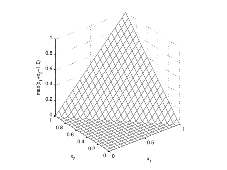
Graph G_f (for the domain $D := [0, 1]^2$).

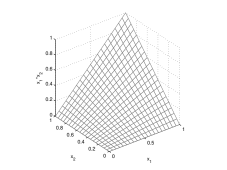


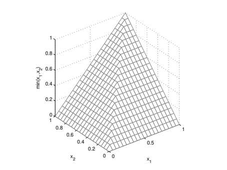
Level sets $L_f(c)$ for some levels c (and the domain $D := [0, 1]^2$).

Example 1.2.3

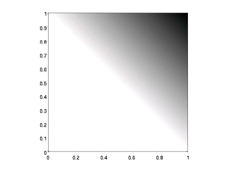
$f(x_1, x_2) :=$	$f(x_1, x_2) :=$	$f(x_1, x_2) :=$
$\max\{x_1 + x_2 - 1; 0\}$	$x_1 \cdot x_2$	$\min\{x_1; x_2\}$

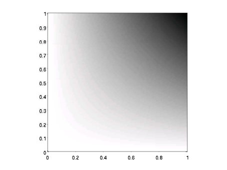


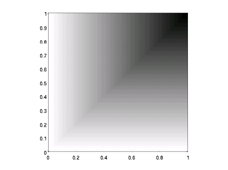




Graph G_f (for the domain $D := [0, 1]^2$).







Level sets $L_f(c)$ for "continuum" of levels c (and the domain $D := [0, 1]^2$).

1. Multidimensional analysis

1.3 Continuity

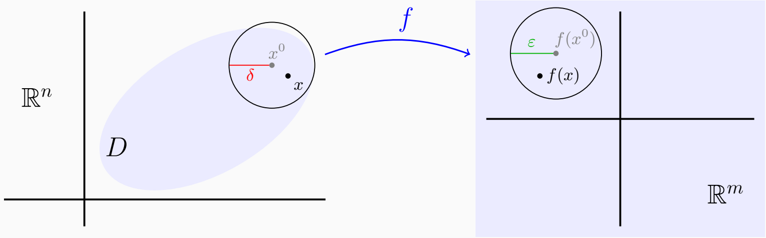
Forster, Analysis 2, Section I.2
Heuser, Lehrbuch der Analysis 2, Section XIV.113
Courant/John, Intro. to Calculus and Analysis II/1, Section 1.3

Definition 1.3.1

Let $D \subseteq \mathbb{R}^n$ and $x^0 \in D$. A function $f : D \to \mathbb{R}^m$ is said to be **continuous at** x^0 if

$$\forall\, \varepsilon > 0 : \; \exists\, \delta > 0 : \; \forall\, x \in D \text{ with } \|x - x^0\| < \delta : \; \|f(x) - f(x^0)\| < \varepsilon.$$

The function f is said to be **continuous** if it is continuous at every point of D, and in this case it is also referred to as a C^0 **function**.



Definition 1.3.1

Let $D \subseteq \mathbb{R}^n$ and $x^0 \in D$. A function $f : D \to \mathbb{R}^m$ is said to be **continuous at** x^0 if

$$\forall\, \varepsilon > 0 : \; \exists\, \delta > 0 : \; \forall\, x \in D \text{ with } \|x - x^0\| < \delta : \; \|f(x) - f(x^0)\| < \varepsilon.$$

The function f is said to be **continuous** if it is continuous at every point of D, and in this case it is also referred to as a C^0 **function**.

Remark 1.3.2 (Local property)

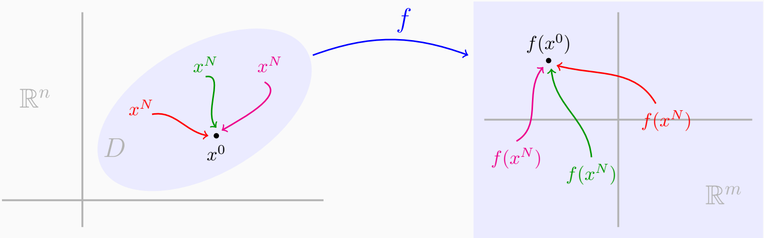
Let $D \subseteq \mathbb{R}^n$, $x^0 \in D$, $r > 0$, and $f : D \to \mathbb{R}^m$ be any function.

Then f is continuous at x^0 if and only if its restriction $f|_{D_{x_0,r}}$ to the set $D_{x_0,r} := D \cap \{x \in \mathbb{R}^n : \|x - x^0\| < r\}$ is continuous at x^0.

Theorem 1.3.3 (Sequential criterion of continuity)

Let $D \subseteq \mathbb{R}^n$ and $x^0 \in D$. A function $f : D \to \mathbb{R}^m$ is continuous at x^0 if and only if for each sequence $(x^N)_{N\in\mathbb{N}}$ in D we have:

$$x^N \to x^0 \qquad \Longrightarrow \qquad f(x^N) \to f(x^0).$$



Theorem 1.3.3 (Sequential criterion of continuity)

Let $D \subseteq \mathbb{R}^n$ and $x^0 \in D$. A function $f : D \to \mathbb{R}^m$ is continuous at x^0 if and only if for each sequence $(x^N)_{N\in\mathbb{N}}$ in D we have:

$$x^N \to x^0 \qquad \Longrightarrow \qquad f(x^N) \to f(x^0).$$

The equivalent condition in 1.3.3 is comparatively strong. In fact, it is stricter than the condition that for each $i = 1, \dots, n$:

$$f(x^0 + h_N \mathsf{e}_i) \to f(x^0)$$

for any sequence $(h_N)_{N\in\mathbb{N}}$ in $\mathbb{R}$ with $h_N \to 0$ (☞ Exercise 3).

In particular, to show that f is continuous at $x^0 = (x_1^0, \dots, x_n^0)$, it is **not** sufficient to show that for each $i = 1, \dots, n$ the mapping $x_i \mapsto f(x_1^0, \dots, x_{i-1}^0, x_i, x_{i+1}^0, \dots, x_n^0)$ is continuous at x_i^0.

Theorem 1.3.3 (Sequential criterion of continuity)

Let $D \subseteq \mathbb{R}^n$ and $x^0 \in D$. A function $f : D \to \mathbb{R}^m$ is continuous at x^0 if and only if for each sequence $(x^N)_{N\in\mathbb{N}}$ in D we have:

$$x^N \to x^0 \qquad \Longrightarrow \qquad f(x^N) \to f(x^0).$$

The equivalent condition in 1.3.3 is comparatively strong. In fact, it is stricter than the condition that for each $i = 1, \ldots, n$:

$$f(x^0 + h_N \mathsf{e}_i) \to f(x^0)$$

for any sequence $(h_N)_{N\in\mathbb{N}}$ in $\mathbb{R}$ with $h_N \to 0$ (☞ Exercise 3).

It can even happen that f is **not** continuous at x^0, but for **each** vector $v \in \mathbb{R}^n$ we have:

$$f(x^0 + h_N v) \to f(x^0)$$

for any sequence $(h_N)_{N\in\mathbb{N}}$ in $\mathbb{R}$ with $h_N \to 0$ (☞ Exercise 4*).

Theorem 1.3.3 (Sequential criterion of continuity)

Let $D \subseteq \mathbb{R}^n$ and $x^0 \in D$. A function $f : D \to \mathbb{R}^m$ is continuous at x^0 if and only if for each sequence $(x^N)_{N\in\mathbb{N}}$ in D we have:

$$x^N \to x^0 \qquad \Longrightarrow \qquad f(x^N) \to f(x^0).$$

Remark 1.3.4

Let $D \subseteq \mathbb{R}^n$, $x^0 \in D$ and $f : D \to \mathbb{R}^m$. If we can find a sequence $(x^N)_{N\in\mathbb{N}}$ in D such that $x^N \to x^0$, but $f(x^N) \not\to f(x^0)$, then 1.3.3 implies that f is **not** continuous at x^0.

Example 1.3.5

Show that the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x_1, x_2) := \begin{cases} \frac{x_1^2 x_2}{x_1^2 + x_2^2} & , \quad (x_1, x_2) \neq (0,0) \\ 0 & , \quad (x_1, x_2) = (0,0) \end{cases}$$

is continuous at $(0,0)$.

Example 1.3.6

Show that the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x_1, x_2) := \begin{cases} \frac{x_1 x_2}{x_1^2 + x_2^2} & , \quad (x_1, x_2) \neq (0,0) \\ 0 & , \quad (x_1, x_2) = (0,0) \end{cases}$$

is **not** continuous at $(0,0)$.

Remark 1.3.7

Let $D \subseteq \mathbb{R}^n$ and $x^0 \in D$. A function $f = (f_1, \ldots, f_m) : D \to \mathbb{R}^m$ is continuous at $x^0 = (x_1^0, \ldots, x_n^0)$ if and only if for every $i = 1, \ldots, m$ the i-th coordinate $f_i : D \to \mathbb{R}$, i.e. the mapping

$$D \longrightarrow \mathbb{R}, \qquad (x_1, \ldots, x_n) \longmapsto f_i(x_1, \ldots, x_n)$$

is continuous at x^0. This follows easily from 1.3.3 and 1.1.2.

Example 1.3.8

As a direct consequence of 1.3.6 and 1.3.7 we obtain that the function $f = (f_1, f_2) : \mathbb{R}^2 \to \mathbb{R}^2$ is **not** continuous at $(0, 0)$, if $f_1 : \mathbb{R}^2 \to \mathbb{R}$ and $f_2 : \mathbb{R}^2 \to \mathbb{R}$ are the functions defined in 1.3.5 and 1.3.6, respectively.

Proposition 1.3.9

Let $D \subseteq \mathbb{R}^n$, $E \subseteq \mathbb{R}^m$ *and* $x^0 \in D$. *Let* $f : D \to \mathbb{R}^m$ *and* $g : E \to \mathbb{R}^k$ *be any functions such that* $\{f(x) : x \in D\} \subseteq E$.

If f *is continuous at* x^0 *and* g *is continuous at* $y^0 := f(x^0)$, *then the composition* $g \circ f : D \to \mathbb{R}^k$ *is continuous at* x^0.

Lemma 1.3.10

The following functions are continuous:

- $\mathrm{add} : \mathbb{R}^2 \to \mathbb{R}$ *defined by* $\mathrm{add}(x_1, x_2) := x_1 + x_2$.
- $\mathrm{mult} : \mathbb{R}^2 \to \mathbb{R}$ *defined by* $\mathrm{mult}(x_1, x_2) := x_1 \cdot x_2$.
- $\mathrm{divi} : \mathbb{R} \times \mathbb{R}_0 \to \mathbb{R}$ *defined by* $\mathrm{divi}(x_1, x_2) := x_1 / x_2$.

where $\mathbb{R}_0 := \mathbb{R} \setminus \{0\}$.

Proposition 1.3.11

Let $D \subseteq \mathbb{R}^n$, $x^0 \in D$ and $f, g : D \to \mathbb{R}$. If f and g are continuous at x^0, then the function

(i) $f + g : D \to \mathbb{R}$ *is continuous at* x^0.

(ii) $f \cdot g : D \to \mathbb{R}$ *is continuous at* x^0.

(iii) $f/g : D \to \mathbb{R}$ *is continuous at* x^0 *(provided that* $g \neq 0$ *on* D*).*

Remark 1.3.12

Let $D_* \subseteq \mathbb{R}$ and $x_*^0 \in D_*$, and $\varphi : D_* \to \mathbb{R}$ be continuous at x_*^0.

Then, for any $n \in \mathbb{N}$ and $D := \mathbb{R} \times \cdots \times \mathbb{R} \times D_* \times \mathbb{R} \cdots \times \mathbb{R}$ ($\subseteq \mathbb{R}^n$), the function $f : D \to \mathbb{R}$ defined by

$$f(x_1, \ldots, x_*, \ldots, x_n) := \varphi(x_*)$$

is continuous at $x^0 = (x_1^0, \ldots, x_*^0, \ldots, x_n^0)$ (for any choice of the x_j^0).

Remark 1.3.13

Let $D_i \subseteq \mathbb{R}$, $x_i^0 \in D_i$, and $\varphi_{i,1}, \ldots, \varphi_{i,k} : D_i \to \mathbb{R}$ be continuous at x_i^0, $i = 1, \ldots, n$. Then, for $D := D_1 \times \cdots \times D_n$, the function $f : D \to \mathbb{R}$ defined by

$$f(x_1, \ldots, x_n) := \sum_{j=1}^{k} \varphi_{1,j}(x_1) \cdots \varphi_{n,j}(x_n)$$

is continuous at $x^0 = (x_1^0, \ldots, x_n^0)$. This follows from 1.3.12 and 1.3.11.

Example 1.3.14

Show that the function $f : (0, \infty) \times (0, \infty) \times \mathbb{R} \to \mathbb{R}$ defined by

$$f(x_1, x_2, x_3) := 1 + x_1 \cos(x_3) + \ln(x_1) e^{x_3} / x_2$$

is continuous.

Example 1.3.15

Show that the function $f : \mathbb{R}^2 \to \mathbb{R}$ in Example 1.3.5 is continuous.

1. Multidimensional analysis

1.4 Open and closed sets in $\mathbb{R}^n$

Forster, Analysis 2, Sections I.1, I.3
Courant/John, Intro. to Calculus and Analysis II/1, Section 1.1

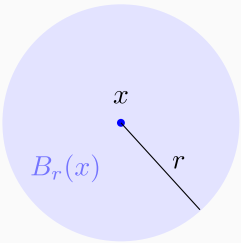
Definition 1.4.1

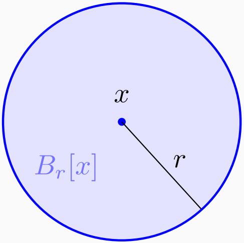
The **open ball around** $x \in \mathbb{R}^n$ **with radius** $r > 0$ is defined by

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

The **closed ball around** $x \in \mathbb{R}^n$ **with radius** $r > 0$ is defined by

$$B_r[x] := \{y \in \mathbb{R}^n : \|x - y\| \le r\}.$$



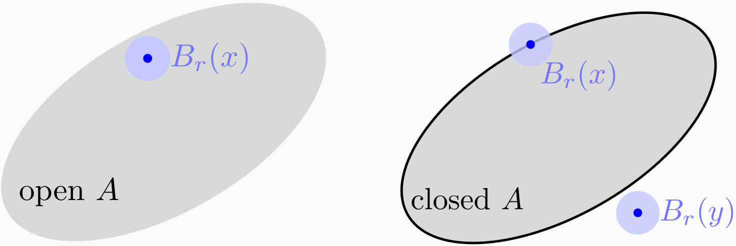


Definition 1.4.2

A nonempty subset A of $\mathbb{R}^n$ is said to be

- **open** if for each $x \in A$ there exists an $r > 0$ such that $B_r(x) \subseteq A$.
- **closed** if $\mathbb{R}^n \setminus A$ is open.

By convention, the empty set $\varnothing$ is both open and closed.



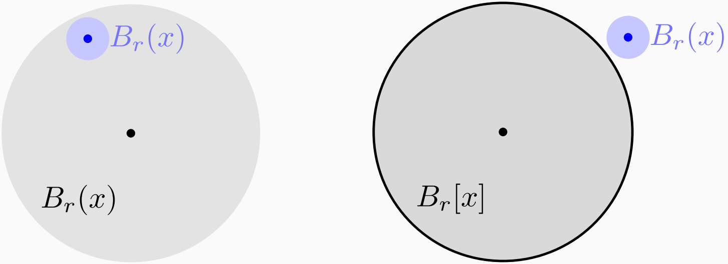
Example 1.4.3

Let $x \in \mathbb{R}^n$ and $r > 0$.

(i) $\mathbb{R}^n$ is open and closed.

(ii) $B_r(x)$ is open.

(iii) $B_r[x]$ is closed.



Definition 1.4.4

A subset A of $\mathbb{R}^n$ is said to be

- **bounded** if there exists an $r > 0$ such that $A \subseteq B_r(0)$.
- **compact** if it is closed and bounded.

Example 1.4.5

Let $a = (a_1, a_2)$ and $b = (b_1, b_2)$ be such that $a_1 < b_1$ and $a_2 < b_2$.

- $(a_1, b_1) \times (a_2, b_2)$ is bounded but not compact.
- $(a_1, b_1] \times (a_2, b_2]$ is bounded but not compact.
- $[a_1, b_1] \times [a_2, b_2]$ is compact (in particular bounded).
- $(-\infty, b_1] \times (a_2, b_2]$ is not bounded (in particular not compact).
- $(-\infty, b_1] \times (-\infty, b_2]$ is not bounded (in particular not compact).
- ...

Definition 1.4.6

Let A be a subset of $\mathbb{R}^n$. Then

- $A^\circ := \{x \in \mathbb{R}^n | \, \exists \, r > 0 : \; B_r(x) \subseteq A\}$ is the **interior of** A.
- $\overline{A} := \{x \in \mathbb{R}^n | \, \forall \, r > 0 : \; B_r(x) \cap A \neq \varnothing\}$ is the **closure of** A.
- $\partial A := \overline{A} \setminus A^\circ$ is the **boundary of** A.

Example 1.4.7

Let $x \in \mathbb{R}^n$ and $r > 0$.

(i) $B_r(x)^\circ = B_r(x)$ and $B_r[x]^\circ = B_r(x)$.

(ii) $\overline{B_r(x)} = B_r[x]$ and $\overline{B_r[x]} = B_r[x]$.

(iii) $\partial B_r(x) = S_r(x)$ and $\partial B_r[x] = S_r(x)$.

Here $S_r(x) := \{y \in \mathbb{R}^n : \|x - y\| = r\}$ is the surface of the ball.

Note that $B_r[x] = B_r(x) \uplus S_r(x)$.

Proposition 1.4.8

Let A be a subset of $\mathbb{R}^n$. Then:

(i) A° *is open.*

(ii) $\overline{A}$ *is closed.*

(iii) ∂A *is closed.*

1. Multidimensional analysis

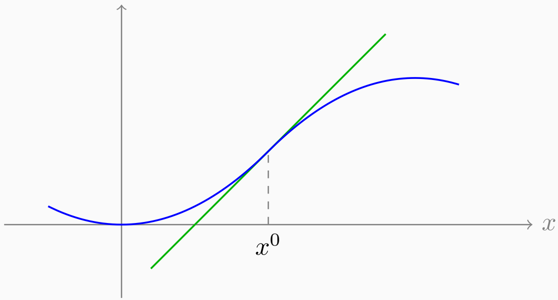
1.5 Partial differentiability of $\mathbb{R}$-valued functions in n variables

Forster, Analysis 2, Section I.5
Heuser, Lehrbuch der Analysis 2, Section XX.162
Courant/John, Intro. to Calculus and Analysis II/1, Section 1.4

Recall

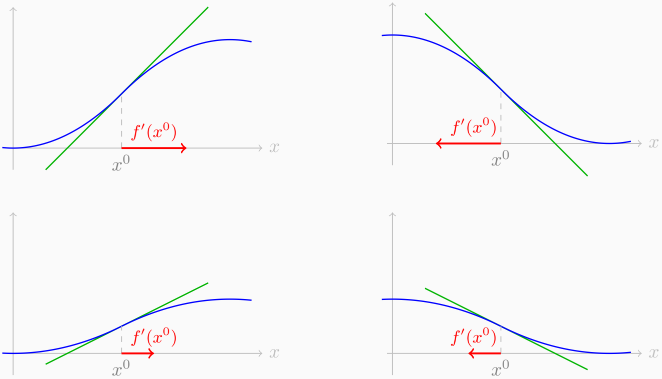
Let $D \subseteq \mathbb{R}$ be open (i.e. $n = 1$ as in Part I).

For a function $f : D \to \mathbb{R}$, the derivative $f'(x^0)$ at $x^0 \in D$ (if it exists) specifies the slope of the tangent of f's graph at point $(x^0, f(x^0))$.



Recall

If we regard $f'(x^0)$ as a vector, it can be seen as the direction in which f is increasing, and its absolute value can be seen as the "rate" of increase.



Objective

Finding an analogous vector that can be seen as the direction in which a function $f : D \to \mathbb{R}$ (with $D \subseteq \mathbb{R}^n$ open) increases fastest at a fixed point $x^0 \in D$, and whose norm can be seen as the "rate" of the (fastest) increase at x^0.

This vector will be introduced in 1.5.9. In fact, it will be shown in 1.8.3 that the vector introduced in 1.5.9 has the properties mentioned above.

For 1.5.9 we need the concept of partial derivatives (☞ 1.5.1).

Preliminary remark

Let $D \subseteq \mathbb{R}^n$, $x^0 \in D$ and $f : D \to \mathbb{R}$.

In what follows, we will frequently consider limits of the form

$$\lim_{h \to 0} \frac{f(x^0 + hv) - f(x^0)}{h}$$

for various vectors $v \in \mathbb{R}^n$. By def., this limit exists and is equal to d if

$$\forall\, \varepsilon > 0 : \ \exists\, \delta > 0 : \ \forall\, h \in (-\delta, \delta) \setminus \{0\} :$$

$$\left| \frac{f(x^0 + hv) - f(x^0)}{h} - d \right| < \varepsilon.$$

For this to make sense, we need to make sure that $x^0 + hv$ lies in the domain D of f for sufficiently small h.

Preliminary remark

The latter condition (on the previous slide) can be ensured by assuming that D is an open set:

Let $D \subseteq \mathbb{R}^n$ be an *open* set and $x^0 \in D$.

Then by Definition 1.4.2 we can find an $r > 0$ such that $B_r(x^0) \subseteq D$.

In particular, for any $v \in \mathbb{R}^n$ we have that

$$x^0 + hv \in B_r(x^0) \subseteq D \qquad \text{for all } h \in (-h_0, h_0),$$

where $h_0 := r/\|v\|$ or $:= \infty$ according to whether $\|v\| > 0$ or $\|v\| = 0$.

First order partial derivatives

Definition 1.5.1

Let $D \subseteq \mathbb{R}^n$ be open, $x^0 \in D$ and $f : D \to \mathbb{R}$.

f is called **partially differentiable at x^0 w.r.t. x_i** if the following limit exists:

$$\frac{\partial f}{\partial x_i}(x^0) := \lim_{h \to 0} \frac{f(x^0 + h\mathsf{e}_i) - f(x^0)}{h}.$$

Then $\frac{\partial f}{\partial x_i}(x^0)$ is called **partial derivative of f w.r.t. x_i at point x^0**.

$x^0 + h\mathsf{e}_i = (x_1^0, \ldots, x_{i-1}^0, x_i^0 + h, x_{i+1}^0, \ldots, x_n^0)$

Partial differentiability of f at x^0 w.r.t. x_i means that the univariate mapping $x_i \mapsto f(x_1^0, \ldots, x_{i-1}^0, x_i, x_{i+1}^0, \ldots, x_n^0)$ is differentiable at x_i^0.

In this case, the mapping $x_i \mapsto f(x_1^0, \ldots, x_{i-1}^0, x_i, x_{i+1}^0, \ldots, x_n^0)$ must be continuous at x_i^0.

Definition 1.5.1

Let $D \subseteq \mathbb{R}^n$ be open, $x^0 \in D$ and $f : D \to \mathbb{R}$.

f is called **partially differentiable at x^0 w.r.t. x_i** if the following limit exists:

$$\frac{\partial f}{\partial x_i}(x^0) := \lim_{h \to 0} \frac{f(x^0 + h\mathsf{e}_i) - f(x^0)}{h}.$$

Then $\frac{\partial f}{\partial x_i}(x^0)$ is called **partial derivative of f w.r.t. x_i at point x^0**.

$$x^0 + h\mathsf{e}_i = (x_1^0, \ldots, x_{i-1}^0, x_i^0 + h, x_{i+1}^0, \ldots, x_n^0)$$

$\frac{\partial f}{\partial x_i}(x^0)$ specifies the change of f when taking an infinitesimally small step from x^0 in direction e_i.

Sometimes it is more convenient to write $\frac{\partial}{\partial x_i} f$ instead of $\frac{\partial f}{\partial x_i}$ (☞ s. 52).

Definition 1.5.1

Let $D \subseteq \mathbb{R}^n$ be open, $x^0 \in D$ and $f : D \to \mathbb{R}$.

f is called **partially differentiable at x^0 w.r.t. x_i** if the following limit exists:

$$\frac{\partial f}{\partial x_i}(x^0) := \lim_{h \to 0} \frac{f(x^0 + h\mathsf{e}_i) - f(x^0)}{h}\,.$$

Then $\frac{\partial f}{\partial x_i}(x^0)$ is called **partial derivative of f w.r.t. x_i at point x^0**.

f is called **partially differentiable w.r.t. x_i** if it is partially differentiable w.r.t. x_i at every point of D.

f is called **partially differentiable (at x^0)** if it is partially differentiable (at x^0) w.r.t. x_i for each $i = 1, \ldots, n$.

Definition 1.5.1

Let $D \subseteq \mathbb{R}^n$ be open, $x^0 \in D$ and $f : D \to \mathbb{R}$.

f is called **partially differentiable at x^0 w.r.t. x_i** if the following limit exists:

$$\frac{\partial f}{\partial x_i}(x^0) := \lim_{h \to 0} \frac{f(x^0 + h\mathsf{e}_i) - f(x^0)}{h}\,.$$

Then $\frac{\partial f}{\partial x_i}(x^0)$ is called **partial derivative of f w.r.t. x_i at point x^0**.

Remark 1.5.2 (Local property)

Let $D \subseteq \mathbb{R}^n$ be open, $x^0 \in D$, $r > 0$ and $f : D \to \mathbb{R}^m$. Then:

f is partially differentiable at x^0 if and only if its restriction $f|_{D_{x_0,r}}$ to the set $D_{x_0,r} := D \cap B_r(x^0)$ is partially differentiable at x^0.

Remark 1.5.3

Let $D \subseteq \mathbb{R}^n$ be open and $x^0 \in D$. Let $f, g : D \to \mathbb{R}$ be two functions that are partially differentiable at x^0 w.r.t. x_i. Moreover, let $\alpha, \beta \in \mathbb{R}$.

Then the functions $\alpha f + \beta g$, $f \cdot g$ and f/g (if $g \neq 0$ on D) are partially differentiable at x^0 w.r.t. x_i and we have:

(i) $\frac{\partial(\alpha f+\beta g)}{\partial x_i}(x^0) = \alpha \frac{\partial f}{\partial x_i}(x^0) + \beta \frac{\partial g}{\partial x_i}(x^0)$ **linearity**,

(ii) $\frac{\partial(f \cdot g)}{\partial x_i}(x^0) = \frac{\partial f}{\partial x_i}(x^0) g(x^0) + f(x^0) \frac{\partial g}{\partial x_i}(x^0)$ **product rule**,

(iii) $\frac{\partial(f/g)}{\partial x_i}(x^0) = \frac{\frac{\partial f}{\partial x_i}(x^0) g(x^0) - f(x^0) \frac{\partial g}{\partial x_i}(x^0)}{g(x^0)^2}$ **quotient rule**.

This follows directly form the analogous result for $n = 1$ (☞ Part I).

Partial differentiability of f at x^0 does **not** generally imply that f is continuous at x^0.

It is true that in this case the mapping

$$x_i \mapsto f(x_1^0, \ldots, x_{i-1}^0, x_i, x_{i+1}^0, \ldots, x_n^0)$$

is continuous at x_i^0 for every $i = 1, \ldots, n$ (☞ slide 42), but this does **not** generally imply that f is continuous at x^0 (☞ Slide 22).

Example 1.5.4

Show that the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined in 1.3.6 is partially differentiable at $(0, 0)$, and recall that f is **not** continuous at $(0, 0)$.

Note, however, 1.5.6.

Definition 1.5.5

Let $D \subseteq \mathbb{R}^n$ be open, $x^0 \in D$ and $f : D \to \mathbb{R}$.

f is called **continuously partially differentiable at x^0 w.r.t.** x_i if it is partially differentiable w.r.t. x_i at each point of a ball $B_r(x^0)$ around x^0 and the mapping $B_r(x^0) \to \mathbb{R}$, $x \mapsto \frac{\partial f}{\partial x_i}(x)$ is continuous at x^0.

f is called **continuously partially differentiable w.r.t.** x_i if it is partially differentiable w.r.t. x_i and the mapping $D \to \mathbb{R}$, $x \mapsto \frac{\partial f}{\partial x_i}(x)$ is continuous.

f is called **continuously partially differentiable (at x^0)** if it is continuously partially differentiable (at x^0) w.r.t. x_i for all $i = 1, \ldots, n$.

f is called a C^1 **function** if it is continuously partially differentiable.

Remark 1.5.6

Let $D \subseteq \mathbb{R}^n$ be open, $x^0 \in D$, $f : D \to \mathbb{R}$. Then (☞ 1.7.5 & 1.7.4):

If f is continuously partially diff.ble at x^0, then it is continuous at x^0. In particular, any C^1 function is continuous.

Example 1.5.7

Show that the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x_1, x_2) := x_1^2 e^{x_1} x_2$ is continuously partially differentiable.

Example 1.5.8

At which points is the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x_1, x_2) := \begin{cases} \frac{x_1\, x_2^2}{x_1^2 + x_2^4} & , \quad (x_1, x_2) \neq (0,0) \\ 0 & , \quad (x_1, x_2) = (0,0) \end{cases}$$

partially diff.ble? At which points is it continuously partially diff.ble?

Definition 1.5.9

Let $D \subseteq \mathbb{R}^n$ be open, $x^0 \in D$ and $f : D \to \mathbb{R}$.

If f is partially differentiable at x^0, then the vector

$$\nabla f(x^0) := \big(\tfrac{\partial f}{\partial x_1}(x^0), \ldots, \tfrac{\partial f}{\partial x_n}(x^0)\big)$$

is called the **gradient of f at x^0**.

The gradient is typically understood as a column vector ($n \times 1$ matrix), i.e.

$$\nabla f(x^0) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x^0) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x^0) \end{bmatrix}.$$

If f is even *continuously* partially differentiable, then $\nabla f(x^0)$ is the direction in which f increases fastest at x^0 and its norm $\|\nabla f(x^0)\|$ can be seen as the "rate" of the (fastest) increase at x^0 (☞ 1.8.3).

Higher order partial derivatives

Let $f : D \to \mathbb{R}$ be a function defined on an open set $D \subseteq \mathbb{R}^n$.
For fixed $x_0 \in D$, $k \geq 2$ and $i_1, \ldots, i_k \in \{1, \ldots, n\}$, we will say that

f satisfies condition $(\mathrm{PD})_{x^0; i_1, \ldots, i_k}$ if we can find an $r > 0$ such that the following assertions hold true, where $D_{x^0,r} := D \cap B_r(x^0)$:

(1) f is partially diff.ble at every point of $D_{x^0,r}$ w.r.t. x_{i_1}.

(2) $\frac{\partial}{\partial x_{i_1}} f$ is partially diff.ble at every point of $D_{x^0,r}$ w.r.t. x_{i_2}.

(3) $\frac{\partial}{\partial x_{i_2}} (\frac{\partial}{\partial x_{i_1}} f)$ is partially diff.ble at every point of $D_{x^0,r}$ w.r.t. x_{i_3}.

$\vdots$

(k-1) $\frac{\partial}{\partial x_{i_{k-2}}} (\cdots (\frac{\partial}{\partial x_{i_2}} (\frac{\partial}{\partial x_{i_1}} f)) \cdots)$ is partially diff.ble at every point of $D_{x^0,r}$ w.r.t. $x_{i_{k-1}}$.

(k) $\frac{\partial}{\partial x_{i_{k-1}}} (\frac{\partial}{\partial x_{i_{k-2}}} (\cdots (\frac{\partial}{\partial x_{i_2}} (\frac{\partial}{\partial x_{i_1}} f)) \cdots))$ is partially diff.ble at x^0 w.r.t. x_{i_k}.

Definition 1.5.10

Let $D \subseteq \mathbb{R}^n$ be open, $x^0 \in D$, and $f : D \to \mathbb{R}$. Moreover, let $k \geq 2$.

f is called **k-times partially differentiable at x^0 w.r.t. $x_{i_1}, \dots, x_{i_k}$** if f satisfies condition (PD)$_{x^0;i_1,\dots,i_k}$. In this case,

$$\frac{\partial^k f}{\partial x_{i_k} \partial x_{i_{k-1}} \cdots \partial x_{i_1}}(x^0) := \frac{\partial}{\partial x_{i_k}}\Big(\frac{\partial}{\partial x_{i_{k-1}}}\Big(\cdots\Big(\frac{\partial}{\partial x_{i_2}}\Big(\frac{\partial}{\partial x_{i_1}} f\Big)\Big)\cdots\Big)\Big)(x^0)$$

is called **k-th order partial derivative of f at x^0 w.r.t. $x_{i_1}, \dots, x_{i_k}$**.

f is called **k-times partially differentiable at x^0** if f satisfies condition (PD)$_{x^0;i_1,\dots,i_k}$ for every choice of $i_1, \dots, i_k \in \{1, \dots, n\}$.

f is called **k-times partially differentiable (w.r.t. $x_{i_1}, \dots, x_{i_k}$)** if f is k-times partially differentiable (w.r.t. $x_{i_1}, \dots, x_{i_k}$) at every point of D.

Definition 1.5.10

Let $D \subseteq \mathbb{R}^n$ be open, $x^0 \in D$, and $f : D \to \mathbb{R}$. Moreover, let $k \geq 2$.

f is called k-**times partially differentiable at** x^0 **w.r.t.** $x_{i_1}, \dots, x_{i_k}$ if f satisfies condition (PD)$_{x^0;i_1,\dots,i_k}$. In this case

$$\frac{\partial^k f}{\partial x_{i_k} \partial x_{i_{k-1}} \cdots \partial x_{i_1}}(x^0) := \frac{\partial}{\partial x_{i_k}}\Big(\frac{\partial}{\partial x_{i_{k-1}}}\Big(\cdots\Big(\frac{\partial}{\partial x_{i_2}}\Big(\frac{\partial}{\partial x_{i_1}} f\Big)\Big)\cdots\Big)\Big)(x^0)$$

is called k-**th order partial derivative of** f **at** x^0 **w.r.t.** $x_{i_1}, \dots, x_{i_k}$.

In general we do **not** have $\frac{\partial^2 f}{\partial x_{i_1} \partial x_{i_2}}(x^0) = \frac{\partial^2 f}{\partial x_{i_2} \partial x_{i_1}}(x^0)$.

Note, however, 1.5.12.

Let $f : D \to \mathbb{R}$ be a function defined on an open set $D \subseteq \mathbb{R}^n$.
For fixed $x_0 \in D$, $k \geq 2$ and $i_1, \ldots, i_k \in \{1, \ldots, n\}$, we will say that

f satisfies condition $(\mathrm{PD})'_{x^0;i_1,\ldots,i_k}$ if we can find an $r > 0$ such that the following assertions hold true, where $D_{x^0,r} := D \cap B_r(x^0)$:

(1) f is partially diff.ble at every point of $D_{x^0,r}$ w.r.t. x_{i_1}.

(2) $\frac{\partial}{\partial x_{i_1}} f$ is partially diff.ble at every point of $D_{x^0,r}$ w.r.t. x_{i_2}.

(3) $\frac{\partial}{\partial x_{i_2}}(\frac{\partial}{\partial x_{i_1}} f)$ is partially diff.ble at every point of $D_{x^0,r}$ w.r.t. x_{i_3}.

$\vdots$

(k-1) $\frac{\partial}{\partial x_{i_{k-2}}}(\cdots(\frac{\partial}{\partial x_{i_2}}(\frac{\partial}{\partial x_{i_1}} f))\cdots)$ is partially diff.ble at every point of $D_{x^0,r}$ w.r.t. $x_{i_{k-1}}$.

(k) $\frac{\partial}{\partial x_{i_{k-1}}}(\frac{\partial}{\partial x_{i_{k-2}}}(\cdots(\frac{\partial}{\partial x_{i_2}}(\frac{\partial}{\partial x_{i_1}} f))\cdots))$ is partially diff.ble at every point of $D_{x^0,r}$ w.r.t. x_{i_k}.

Definition 1.5.11

Let $D \subseteq \mathbb{R}^n$ be open, $x^0 \in D$, and $f : D \to \mathbb{R}$. Moreover, let $k \geq 2$.

f is called k-**times continuously partially differentiable at** x^0 if it satisfies condition $(\mathrm{PD})'_{x^0;i_1,\ldots,i_k}$ for every choice of $i_1, \ldots, i_k \in \{1, \ldots, n\}$ and all k-th order partial derivatives are continuous at x^0.

f is called k-**times continuously partially differentiable** if f is k-times continuously partially differentiable at every point of D. In this case, f is called a C^k **function**.

Theorem 1.5.12 (Schwarz)

Let $D \subseteq \mathbb{R}^n$ be open, $x^0 \in D$, $f : D \to \mathbb{R}$. If f is twice continuously partially diff.ble at x^0, then for any $i_1, i_2 \in \{1, \ldots, n\}$ we have

$$\frac{\partial^2 f}{\partial x_{i_1} \partial x_{i_2}}(x^0) = \frac{\partial^2 f}{\partial x_{i_2} \partial x_{i_1}}(x^0).$$

By an induction on k (noting that a permutation can be obtained by successively interchanging 'neighbours') we can conclude:

Corollary 1.5.13

Let $D \subseteq \mathbb{R}^n$ be open, $x^0 \in D$, $f : D \to \mathbb{R}$. If f is k-times continuously partially diff.ble at x^0, then for any $i_1, \ldots, i_k \in \{1, \ldots, n\}$ we have

$$\frac{\partial^k f}{\partial x_{i_1} \cdots \partial x_{i_k}}(x^0) = \frac{\partial^k f}{\partial x_{i_{\pi(1)}} \cdots \partial x_{i_{\pi(k)}}}(x^0)$$

for each permutation $\pi(1), \ldots, \pi(k)$ of the numbers $1, \ldots, k$.

Definition 1.5.14

Let $D \subseteq \mathbb{R}^n$ be open, $x^0 \in D$ and $f : D \to \mathbb{R}$. If f is twice partially differentiable at x^0, then

$$\operatorname{Hess} f(x^0) := \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(x^0) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x^0) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x^0) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x^0) & \frac{\partial^2 f}{\partial x_2 \partial x_2}(x^0) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x^0) \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x^0) & \frac{\partial^2 f}{\partial x_n \partial x_2}(x^0) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n}(x^0) \end{bmatrix}$$

is called the **Hessian matrix of f at x^0**.

The Hessian matrix appears in the Taylor expansion (of order ≥ 2) of a multidimensional function (☞ 1.9.5) and plays a crucial role for the characterisation of (local) minimisers/maximisers of a multidimensional function (☞ 1.11.8).

It is symm. if f is twice *continuously* partially diff.ble at x^0 (☞ 1.5.12).

Definition 1.5.15

Let $D \subseteq \mathbb{R}^n$ be open, $x^0 \in D$ and $f : D \to \mathbb{R}$. If f is twice partially differentiable at x^0, then

$$\Delta f(x^0) := \sum_{i=1}^{n} \frac{\partial^2 f}{\partial^2 x_i^2}(x^0)$$

is called the **Laplacian (or Laplace operator) of f at x^0**.

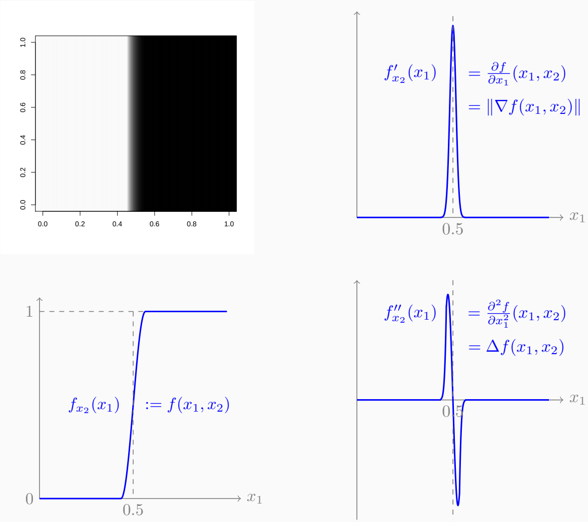
Example 1.5.16

In image analysis, the gradient $\nabla f(\cdot)$ and the Laplacian $\Delta f(\cdot)$ are used to detect edges in continuous monochrome "images" f (☞ slide 13). More precisely,

an edge in "image" f

... consists of points x where $\Delta f(x) = 0$ and $\|\nabla f(x)\|$ is large

(see slide 60 for an illustration). But there may be other such points!



Note that here all functions are indep. of x_2, in part. $\frac{\partial^j f}{\partial x_2^j} \equiv 0$, $j = 1, 2$.

1. Multidimensional analysis

1.6 Partial differentiability of $\mathbb{R}^m$-valued functions in n variables

Definition 1.6.1

Let $D \subseteq \mathbb{R}^n$ be open, $x^0 \in D$ and $f = (f_1, \ldots, f_m) : D \to \mathbb{R}^m$.

f is called **partially differentiable at x^0 w.r.t. x_i** if each of the coordinates $f_1, \ldots, f_m$ is partially diff.ble at x^0 w.r.t. x_i. In this case,

$$\frac{\partial f}{\partial x_i}(x^0) := \begin{bmatrix} \frac{\partial f_1}{\partial x_i}(x^0) \\ \vdots \\ \frac{\partial f_m}{\partial x_i}(x^0) \end{bmatrix}$$

is called the **partial derivative of f at x^0 w.r.t. x_i**.

f is called **partially differentiable w.r.t. x_i** if it is partially differentiable w.r.t. x_i at every point of D.

f is called **partially differentiable** (at x^0) if it is partially differentiable (at x^0) w.r.t. x_i for all $i = 1, \ldots, n$.

Remark 1.6.2

$f = (f_1, \dots, f_m)$ is partially differentiable at x^0 w.r.t. x_i if and only if there exists a $(d_1, \dots, d_m) \in \mathbb{R}^m$ such that

$$\lim_{h\to 0} \begin{bmatrix} \frac{f_1(x^0+h\mathrm{e}_i)-f_1(x^0)}{h} \\ \vdots \\ \frac{f_m(x^0+h\mathrm{e}_i)-f_m(x^0)}{h} \end{bmatrix} = \begin{bmatrix} d_1 \\ \vdots \\ d_m \end{bmatrix}.$$

In this case, $(d_1, \dots, d_m) = (\frac{\partial f_1}{\partial x_i}(x^0), \dots, \frac{\partial f_m}{\partial x_i}(x^0))$.
This is an immediate consequence of 1.5.1 and 1.1.2.

Remark 1.6.3

The definitions of **continuously partially differentiable**, k-**times (continuously) partially diff.ble** and C^k **function** are analogous to the case $m = 1$ (☞ 1.5.5). In particular, $f = (f_1, \dots, f_m)$ is a C^k function if and only if each of the coordinates $f_1, \dots, f_m$ is a C^k function.

Remark 1.6.4

Let $D \subseteq \mathbb{R}^n$ be open and $x^0 \in D$. Let $f, g : D \to \mathbb{R}^m$ and $\varphi : D \to \mathbb{R}$. Moreover, let $\alpha, \beta \in \mathbb{R}$.

If f, g and φ are partially differentiable at x^0 w.r.t. x_i, then the functions $\alpha f + \beta g$ and $\varphi \cdot f$ are partially differentiable at x^0 w.r.t. x_i and

(i) $\frac{\partial(\alpha f+\beta g)}{\partial x_i}(x^0) = \alpha \frac{\partial f}{\partial x_i}(x^0) + \beta \frac{\partial g}{\partial x_i}(x^0)$ **linearity**,

(ii) $\frac{\partial(\varphi \cdot f)}{\partial x_i}(x^0) = \frac{\partial \varphi}{\partial x_i}(x^0) f(x^0) + \varphi(x^0) \frac{\partial f}{\partial x_i}(x^0)$ **product rule**.

This follows directly from 1.5.3.

Definition 1.6.5

Let $D \subseteq \mathbb{R}^n$ be open, $x^0 \in D$ and $f : D \to \mathbb{R}^m$. If $f = (f_1, \dots, f_m)$ is partially differentiable at x^0, then

$$Jf(x^0) := \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x^0) & \frac{\partial f_1}{\partial x_2}(x^0) & \cdots & \frac{\partial f_1}{\partial x_n}(x^0) \\ \frac{\partial f_2}{\partial x_1}(x^0) & \frac{\partial f_2}{\partial x_2}(x^0) & \cdots & \frac{\partial f_2}{\partial x_n}(x^0) \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(x^0) & \frac{\partial f_m}{\partial x_2}(x^0) & \cdots & \frac{\partial f_m}{\partial x_n}(x^0) \end{bmatrix}$$

is called the **Jacobian matrix of f at x^0**.

The Jacobian matrix appears in the Taylor expansion (of order ≥ 1) of a multidimensional function (☞ 1.7.3, 1.9.5).

In general, the Jacobian matrix is neither symmetric nor quadratic.

For $m = 1$, we have $Jf(x^0) = \nabla f(x^0)^\mathsf{T}$.

1. Multidimensional analysis

1.7 Total differentiability

Forster, Analysis 2, Section I.6
Heuser, Lehrbuch der Analysis 2, Sections XX.163–165
Courant/John, Intro. to Calculus and Analysis II/1, Sec. 1.5, 1.6

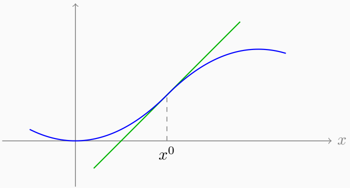
Recall

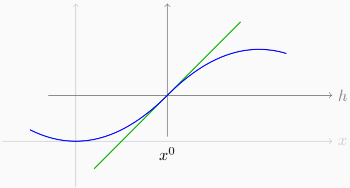
Let $D \subseteq \mathbb{R}$ be open (i.e. $n = 1$ as in Part I).

For a function $f : D \to \mathbb{R}$, the derivative $f'(x^0)$ at $x^0 \in D$ (if it exists) can be seen as a linear map that approxim. the change of f "around" x^0:

$$f(x^0 + h) - f(x^0) = \underbrace{f'(x^0) \cdot h}_{\text{linear in } h} + r(h),$$

where $r(h) := f(x^0 + h) - f(x^0) - f'(x^0) \cdot h$ satisfies $\lim_{h \to 0} \frac{r(h)}{h} = 0$.





Objective

Finding a linear map that can be seen as an approximation of the change of f "around" $x^0 \in D$ for a function $f : D \to \mathbb{R}^m$ (with $D \subseteq \mathbb{R}^n$ open) .

This map will be introduced in 1.7.1.

Note at this point that it is known (from Part II) that for any linear map $L : \mathbb{R}^n \to \mathbb{R}^m$ there exists a matrix $A \in \mathbb{R}^{m \times n}$ such that

$$L(x) = Ax \qquad \text{for all } x \in \mathbb{R}^n.$$

This is the matrix representation of the linear map L.
In this sense, the map L can be identified with the matrix A.

Definition 1.7.1

Let $D \subseteq \mathbb{R}^n$ be open, $x^0 \in D$ and $f = (f_1, \dots, f_m) : D \to \mathbb{R}^m$.

f is called **totally differentiable at** x^0 if one can find a $\delta > 0$, a matrix $A = A_{f,x^0} \in \mathbb{R}^{m \times n}$ and a map $r = r_{f,x^0} : B_\delta(\mathbf{0}) \to \mathbb{R}^m$ such that $B_\delta(x^0) \subseteq D$ and

$$f(x^0 + x) - f(x^0) = Ax + r(x) \quad \text{for all } x \in B_\delta(\mathbf{0}), \tag{1.1}$$

$$\lim_{x \to \mathbf{0}} r(x)/\|x\| = \mathbf{0}. \tag{1.2}$$

In this case, $A = A_{f,x^0}$ is called the **total derivative of f at x^0**.

f is called **totally differentiable** if it is totally differentiable at every point of D.

For simplicity, the word "totally" / "total" is often omitted.

Remark 1.7.2

Conditions (1.1)–(1.2) are equivalent to

$$\lim_{x \to \mathbf{0}} \frac{f(x^0 + x) - f(x^0) - Ax}{\|x\|} = \mathbf{0}\,, \tag{1.3}$$

since r is of the form $r(x) = f(x^0 + x) - f(x^0) - Ax$.

Theorem 1.7.3

Let $D \subseteq \mathbb{R}^n$ be open, $x^0 \in D$ and $f = (f_1, \ldots, f_m) : D \to \mathbb{R}^m$.

If f is totally differentiable at x^0, then f is partially differentiable at x^0 and the derivative $A = A_{f,x^0}$ of f at x^0 (☞ 1.7.1) coincides with the Jacobian matrix of f at x^0, i.e.

$$A_{f,x^0} = Jf(x^0).$$

That is, if the total derivative of f at x^0 exists, it is given by $Jf(x^0)$.

Theorem 1.7.4

Let $D \subseteq \mathbb{R}^n$ *be open,* $x^0 \in D$ *and* $f = (f_1, \ldots, f_m) : D \to \mathbb{R}^m$*.*

If f *is totally differentiable at* x^0*, then* f *is continuous at* x^0*.*

Theorem 1.7.5

Let $D \subseteq \mathbb{R}^n$ *be open,* $x^0 \in D$ *and* $f = (f_1, \ldots, f_m) : D \to \mathbb{R}^m$*.*

If f *is continuously partially differentiable at* x^0*, then* f *is totally differentiable at* x^0*.*

Remark 1.7.6

In view of 1.7.5, a continuously partially differentiable function (i.e. a C^1 function) is often simply said to be **continuously differentiable**.

A k-times continuously partially differentiable function (i.e. a C^k function) is simply said to be k**-times continuously differentiable**.

$$\begin{tikzcd}[column sep=small] & \text{continuously differentiable} \arrow[d, Rightarrow, "1.7.5"] & \\ & \text{totally differentiable} \arrow[dl, Rightarrow, "1.7.4"'] \arrow[dr, Rightarrow, "1.7.3"] & \\ \text{continuous} & & \text{partially differentiable} \end{tikzcd}$$

Other implications do **not** generally apply!

Other implications do **not** generally apply!

- partially differentiable but not continuous

  ☞ 1.5.4

- partially differentiable but not continuously (totally) differentiable

  ☞ 1.5.8 (and its proof; take 1.7.4 into account)

- totally differentiable but not continuously differentiable

  ☞ $n = m = 1$ ✓

- continuous but not continuously (totally, partially) differentiable

  ☞ $n = m = 1$ ✓

Remark 1.7.7

Let $D \subseteq \mathbb{R}^n$ be open, $x^0 \in D$ and $f, g : D \to \mathbb{R}^m$. Let $\alpha, \beta \in \mathbb{R}$.

If f and g are totally differentiable at x^0, then the function $\alpha f + \beta g$ is totally differentiable at x^0 and we have

$$J(\alpha f + \beta g)(x^0) = \alpha J f(x^0) + \beta J g(x^0) \qquad \textbf{linearity}. \tag{1.4}$$

The first assertion can be derived easily from (1.1)–(1.2) or from (1.3).

(1.4) follows directly from 1.6.4(i).

Theorem 1.7.8 (Chain rule)

Let $D \subseteq \mathbb{R}^n$, $E \subseteq \mathbb{R}^m$ *be open sets,* $x^0 \in D$, $f : D \to E$, $g : E \to \mathbb{R}^k$.

If f *is totally differentiable at* x^0 *and* g *is totally differentiable at* $y^0 := f(x^0)$, *then the composition* $g \circ f : D \to \mathbb{R}^k$ *is totally differentiable at* x^0 *and we have*

$$J(g \circ f)(x^0) = \underbrace{Jg(f(x^0))}_{\text{'outer derivative'}} \cdot \underbrace{Jf(x^0)}_{\text{'inner derivative'}} . \tag{1.5}$$

Remark 1.7.9

If $n = k = 1$ (i.e. $\mathbb{R} \ni x \mapsto g \circ f(x) = g(f_1(x), \ldots, f_m(x)) \in \mathbb{R}$), then (1.5) reads as

$$(g \circ f)'(x^0) = \nabla g(f(x^0))^{\mathsf{T}} \begin{bmatrix} f_1'(x^0) \\ \vdots \\ f_m'(x^0) \end{bmatrix} = \sum_{i=1}^m \frac{\partial g}{\partial x_i}(f(x^0)) \cdot f_i'(x^0).$$

1. Multidimensional analysis

1.8 Directional derivative

Forster, Analysis 2, Section I.6
Heuser, Lehrbuch der Analysis 2, Section XX.166
Courant/John, Intro. to Calculus and Analysis II/1, Section 1.5

Definition 1.8.1

Let $D \subseteq \mathbb{R}^n$ be open, $x^0 \in D$, $f : D \to \mathbb{R}$. Let $v \in \mathbb{R}^n$ with $\|v\| = 1$.

The **directional derivative of f at x^0 in direction** v is defined by

$$D_v f(x^0) := \lim_{h \to 0} \frac{f(x^0 + hv) - f(x^0)}{h},$$

provided the limit exists.

$D_v f(x^0)$ specifies the change of f when taking an infinitesimally small step from x^0 in direction v, i.e. $D_v f(x^0) =$ slope of f in direction v.

$D_{\mathsf{e}_i} f(x^0) = \frac{\partial f}{\partial x_i}(x^0),\ i = 1, \ldots, n.$

For $n = 1$: $\quad D_{+1} f(x^0) := \lim_{h \to 0} \frac{f(x^0+h)-f(x^0)}{h} = +f'(x^0),$

For $n = 1$: $\quad D_{-1} f(x^0) := \lim_{h \to 0} \frac{f(x^0-h)-f(x^0)}{h} = -f'(x^0).$

Theorem 1.8.2

Let $D \subseteq \mathbb{R}^n$ be open, $x^0 \in D$ and $f : D \to \mathbb{R}$.

If f is totally differentiable at x^0, then the directional derivative $D_v f(x^0)$ does exist for all $v \in \mathbb{R}^n$ with $\|v\| = 1$ and we have

$$D_v f(x^0) = \langle v, \nabla f(x^0)\rangle \quad \textit{for all } v \in \mathbb{R}^n \textit{ with } \|v\| = 1.$$

In particular, the slope in direction v is equal to 0 for all directional vectors v that are perpendicular to the gradient $\nabla f(x^0)$.

Obviously the slope is 0 when taking an infinitesimally small step from x^0 along the level set at level $c := f(x^0)$.

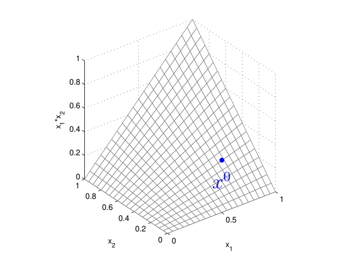
That is, the gradient $\nabla f(x^0)$ at x^0 is perpendicular to the tangent of the level set at level $c := f(x^0)$ at x^0 (☞ slide 76).

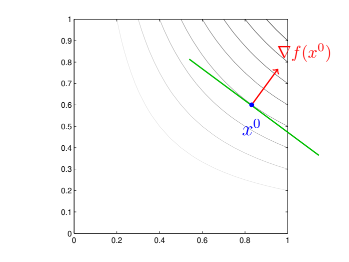
Theorem 1.8.2

Let $D \subseteq \mathbb{R}^n$ be open, $x^0 \in D$ and $f : D \to \mathbb{R}$.

If f is totally differentiable at x^0, then the directional derivative $D_v f(x^0)$ does exist for all $v \in \mathbb{R}^n$ with $\|v\| = 1$ and we have

$$D_v f(x^0) = \langle v, \nabla f(x^0) \rangle \qquad \text{for all } v \in \mathbb{R}^n \text{ with } \|v\| = 1.$$





Corollary 1.8.3

Let $D \subseteq \mathbb{R}^n$ be open, $x^0 \in D$ and $f : D \to \mathbb{R}$.

If f is totally differentiable at x^0 with $\nabla f(x^0) \neq [0, \ldots, 0]^\mathsf{T}$, then the directional derivative $D_v f(x^0)$ is maximal for the direction

$$v^* := \frac{\nabla f(x^0)}{\|\nabla f(x^0)\|}$$

and the maximal slope is given by

$$D_{v^*} f(x^0) = \|\nabla f(x^0)\|.$$

1. Multidimensional analysis

1.9 Mean value theorem and Taylor's theorem

Forster, Analysis 2, Section I.7
Heuser, Lehrbuch der Analysis 2, Section XX.167, XX.168
Courant/John, Intro. to Calculus and Analysis II/1, Section 1.7

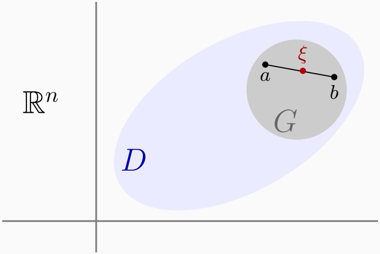
Theorem 1.9.1 (Mean value theorem, multidimensional version)

Let $D \subseteq \mathbb{R}^n$ and $f : D \to \mathbb{R}$.

Let $G \subseteq D$ be open, and $a, b \in G$ be such that the line segment between a and b is contained in G (i.e. $\{a + t(b-a) : t \in [0,1]\} \subseteq G$).

If f is totally differentiable at every point of G, then there exists a $\theta = \theta_{f,a,b} \in (0,1)$ such that, for $\xi = \xi_{f,a,b} := a + \theta(b-a)$, we have

$$f(b) - f(a) = \langle \nabla f(\xi), b - a \rangle.$$



Theorem 1.9.1 (Mean value theorem, multidimensional version)

Let $D \subseteq \mathbb{R}^n$ and $f : D \to \mathbb{R}$.

Let $G \subseteq D$ be open, and $a, b \in G$ be such that the line segment between a and b is contained in G (i.e. $\{a + t(b-a) : t \in [0,1]\} \subseteq G$).

If f is totally differentiable at every point of G, then there exists a $\theta = \theta_{f,a,b} \in (0,1)$ such that, for $\xi = \xi_{f,a,b} := a + \theta(b-a)$, we have

$$f(b) - f(a) = \langle \nabla f(\xi), b - a \rangle.$$

For $n = 1$, this is the mean value theorem known from Part I:

$$f(b) - f(a) = f'(\xi)(b-a) \quad \text{for some } \xi \in [a,b]$$

For general n, it is a special case of 1.9.3 (choose $k := 0$ there) but only under a stronger assumption (C^1 instead of total differentiability).

In some situations it is easy to specify θ explicitly (☞ 1.9.2).
In general, however, it is not necessarily easy to specify θ explicitly.

Example 1.9.2

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x_1, x_2) := \cos(x_1) + \sin(x_2)$ and choose $a := (0, 0)$ and $b := (\frac{\pi}{2}, \frac{\pi}{2})$.

It obviously holds that $\nabla f(x_1, x_2) = [-\sin(x_1), \cos(x_2)]^{\mathsf{T}}$.

By 1.9.1 there must exist a $\theta \in (0, 1)$ such that

$$\begin{aligned} 0 &= f(\tfrac{\pi}{2}, \tfrac{\pi}{2}) - f(0,0) \\ &= f(b) - f(a) \;=\; \langle \nabla f(a + \theta(b-a)), b - a \rangle \\ &= (b-a)^{\mathsf{T}} \nabla f(a + \theta(b-a)) \;=\; [\tfrac{\pi}{2}, \tfrac{\pi}{2}] \nabla f(\theta \tfrac{\pi}{2}, \theta \tfrac{\pi}{2}) \\ &= [\tfrac{\pi}{2}, \tfrac{\pi}{2}] \, [-\sin(\tfrac{\theta\pi}{2}), \cos(\tfrac{\theta\pi}{2})]^{\mathsf{T}} \;=\; \tfrac{\pi}{2} \left(\cos(\tfrac{\theta\pi}{2}) - \sin(\tfrac{\theta\pi}{2}) \right), \end{aligned}$$

and $\theta := \frac{1}{2}$ is easily seen to satisfy this equation (☞ Part I).

Let $D \subseteq \mathbb{R}^n$ be open, $x^0 \in D$ and $f : D \to \mathbb{R}$.

Recall

If f is a C^1 function, then the change of f "around" x^0 can be approximated by a linear function (i.e. by an order 1 polynomial):

$$f(x^0 + x) - f(x^0) \approx Jf(x^0)x = (\nabla f(x^0))^\mathsf{T} x = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(x^0) x_i$$

Objective

Approximation of the change of f "around" x^0 by an order k polynomial, provided that f is a C^k function:

$$f(x^0 + x) - f(x^0) \approx \sum_{1 \leq |\alpha| \leq k} \frac{1}{\alpha_1!} \cdots \frac{1}{\alpha_n!} \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}(x^0) x_1^{\alpha_1} \cdots x_n^{\alpha_n}$$

Here the sum ranges over all $\alpha = (\alpha_1, \ldots, \alpha_n) \in \{0, \ldots, k\}^n$ for which the sum of the coordinates $|\alpha| := \sum_{i=1}^n \alpha_i$ is contained in $\{1, \ldots, k\}$.

Let $D \subseteq \mathbb{R}^n$ be open, $x^0 \in D$ and $f : D \to \mathbb{R}$.

Recall

If f is a C^1 function, then the change of f "around" x^0 can be approximated by a linear function (i.e. by an order 1 polynomial):

$$f(x^0 + x) - f(x^0) \approx Jf(x^0)x = (\nabla f(x^0))^{\mathsf{T}} x = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x^0) x_i$$

Objective

Approximation of the change of f "around" x^0 by an order k polynomial, provided that f is a C^k function:

$$f(x^0 + x) - f(x^0) \approx \sum_{1 \le |\alpha| \le k} \frac{1}{\alpha_1!} \cdots \frac{1}{\alpha_n!} \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}(x^0) x_1^{\alpha_1} \cdots x_n^{\alpha_n}$$

For $n = 1$ this is the Taylor approximation known from Part I:

$$f(x^0 + x) - f(x^0) \approx \sum_{\alpha=1}^{k} \frac{1}{\alpha!} f^{(\alpha)}(x^0) x^{\alpha}$$

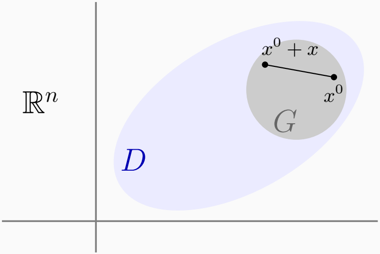
Theorem 1.9.3 (Taylor, multidimensional version)

Let $D \subseteq \mathbb{R}^n$, $x^0 \in D$ and $f : D \to \mathbb{R}$.

Let $G \subseteq D$ be open such that $x^0 \in G$. Moreover, let $x \in G$ such that the line segment between x^0 and $x^0 + x$ is contained in G.

Let $k \in \mathbb{N}_0$ and assume that f is a C^{k+1} function on G, i.e. that f is $(k+1)$-times continuously differentiable at every point of G.

$\mathbb{N}_0 := \{0, 1, \ldots\}$



Theorem 1.9.3 (Taylor, multidimensional version)

Let $D \subseteq \mathbb{R}^n$, $x^0 \in D$ *and* $f : D \to \mathbb{R}$.

Let $G \subseteq D$ *be open such that* $x^0 \in G$. *Moreover, let* $x \in G$ *such that the line segment between* x^0 *and* $x^0 + x$ *is contained in* G.

Let $k \in \mathbb{N}_0$ *and assume that* f *is a* C^{k+1} *function on* G, *i.e. that* f *is* $(k+1)$*-times continuously differentiable at every point of* G.

Then there exists a $\theta = \theta_{f,x^0,x} \in [0,1]$ *such that*

$$\begin{aligned} &f(x^0 + x) - f(x^0) \\ &= \textstyle\sum_{1 \leq |\alpha| \leq k} \frac{1}{\alpha_1!} \cdots \frac{1}{\alpha_n!} \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}(x^0)\, x_1^{\alpha_1} \cdots x_n^{\alpha_n} + r_k(x) \end{aligned}$$

with Lagrange remainder $r_k(x) = r_{k,f,x^0}(x)$ *given by*

$$r_k(x) := \textstyle\sum_{|\alpha| = k+1} \frac{1}{\alpha_1!} \cdots \frac{1}{\alpha_n!} \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}(x^0 + \theta x)\, x_1^{\alpha_1} \cdots x_n^{\alpha_n}.$$

Corollary 1.9.4

Let $D \subseteq \mathbb{R}^n$, $x^0 \in D$, $f : D \to \mathbb{R}$. Let $\delta > 0$ be s. t. $B_\delta(x^0) \subseteq D$.

Let $k \in \mathbb{N}$ and assume that f is a C^k function on $B_\delta(x^0)$, i.e. that f is k-times continuously differentiable at every point of $B_\delta(x^0)$. Then

$$f(x^0 + x) \tag{1.6}$$
$$= f(x^0) + \sum_{1 \le |\alpha| \le k} \frac{1}{\alpha_1!} \cdots \frac{1}{\alpha_n!} \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}(x^0)\, x_1^{\alpha_1} \cdots x_n^{\alpha_n} + o(\|x\|^k)$$

for all $x \in B_\delta(x^0)$.

In (1.6) the symbol $o(\|x\|^k)$ denotes a function $r = r_{f,x_0,\delta} : B_\delta(\mathbf{0}) \to \mathbb{R}$ that satisfies $\lim_{x \to \mathbf{0}} r(x)/\|x\|^k = 0$.

$o(\|\cdot\|^k)$ is the so-called Landau notation (or O notation) of $r(\cdot)$.

(1.6) is said to be the **Taylor formula for f at x^0 of order k.**

Remark 1.9.5

In the setting of 1.9.4 with $k = 2$, the Taylor formula (1.6) is

$$f(x^0 + x) \;=\; f(x^0) + \nabla f(x^0)^{\mathsf{T}} x + \tfrac{1}{2} x^{\mathsf{T}} \mathrm{Hess}\, f(x^0) x \;+\; o(\|x\|^2) \qquad (1.7)$$

and can be reformulated as

$$\begin{aligned} f(y) \;=\;& f(x^0) + \nabla f(x^0)^{\mathsf{T}} (y - x^0) \qquad (1.8) \\ &+ \tfrac{1}{2} (y - x^0)^{\mathsf{T}} \mathrm{Hess}\, f(x^0)(y - x^0) \;+\; o(\|y - x^0\|^2). \end{aligned}$$

Example 1.9.6

Show that $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x_1, x_2) := x_1 \sin(x_2)$ is a C^2 function and determine the second order Taylor formula of f at $x^0 = (0, 0)$ (i.e. (1.7) and/or (1.8) for $n = 2$ and $x^0 = (0, 0)$).

Definition 1.9.7

In the setting of 1.9.4, we can define a map $T_k(x^0;\,\cdot\,) : B_\delta(x^0) \to \mathbb{R}$ by

$$T_k^f(x^0;y) \;:=$$

$$f(x^0) + \sum_{1\le|\alpha|\le k} \tfrac{1}{\alpha_1!}\cdots\tfrac{1}{\alpha_n!}\,\tfrac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1}\cdots\partial x_n^{\alpha_n}}(x^0)\,(y_1 - x_1^0)^{\alpha_1}\cdots(y_n - x_n^0)^{\alpha_n}.$$

This map is called the **Taylor polynomial of order k of f at x^0**.

The Taylor polynomials of order 1 and 2 are

$$T_1^f(x^0;y) = f(x^0) + \nabla f(x^0)^\mathsf{T}(y - x^0),$$

$$T_2^f(x^0;y) = f(x^0) + \nabla f(x^0)^\mathsf{T}(y - x^0) + \tfrac{1}{2}(y - x^0)^\mathsf{T}\mathrm{Hess}\, f(x^0)(y - x^0).$$

Definition 1.9.7

In the setting of 1.9.4, we can define a map $T_k(x^0;\,\cdot\,) : B_\delta(x^0) \to \mathbb{R}$ by

$$T_k^f(x^0;y) \,:=$$

$$f(x^0) + \sum_{1\le|\alpha|\le k} \tfrac{1}{\alpha_1!}\cdots\tfrac{1}{\alpha_n!}\,\tfrac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1}\cdots\partial x_n^{\alpha_n}}(x^0)\,(y_1-x_1^0)^{\alpha_1}\cdots(y_n-x_n^0)^{\alpha_n}.$$

This map is called the **Taylor polynomial of order k of f at x^0**.

In the setting of 1.9.4, the Taylor formula (1.6) can be written as

$$f(x^0+x) \,=\, T_k^f(x^0;x^0+x) + o(\|x\|^k),$$

which is equivalent to

$$f(y) \,=\, T_k^f(x^0;y) + o(\|y-x^0\|^k).$$

$T_k^f(x^0;\,\cdot\,)$ is also called **k-th order Taylor approximation of f at x^0**.

1. Multidimensional analysis

1.10 Numerical differentiation

Will be skipped for now.

1. Multidimensional analysis

1.11 Extremum points of $\mathbb{R}$-valued functions on $\mathbb{R}^n$

Forster, Analysis 2, Section I.7
Heuser, Lehrbuch der Analysis 2, Section XX.173
Courant/John, Intro. to Calculus and Analysis II/1, Section 3.7

Definition 1.11.1

Let $D \subseteq \mathbb{R}^n$ and $f : D \to \mathbb{R}$. A point $x \in D$ is called a

minimum (maximum) point of f if

$$\forall\, y \in D\text{:}\quad f(x) \leq (\geq)\ f(y).$$

local minimum (maximum) point of f if

$$\exists\, \delta > 0\text{:}\quad \forall\, y \in B_\delta(x) \cap D\text{:}\quad f(x) \leq (\geq)\ f(y).$$

(local) extremum point of f if

it is a (local) minimum point or a (local) maximum point of f.

If D is open, then we can replace $B_\delta(x) \cap D$ by $B_\delta(x)$.

Definition 1.11.2

Let $D \subseteq \mathbb{R}^n$ and $f : D \to \mathbb{R}$. A point $x \in D$ is called a

isolated minimum (maximum) point of f if

$$\forall\, y \in D\colon \;\; f(x) \leq (\geq)\; f(y), \quad \text{and}$$

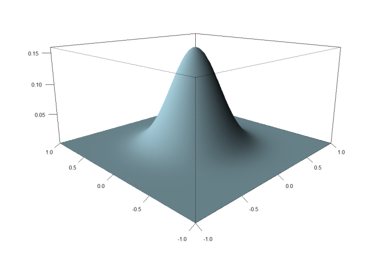
$$\exists\, \delta > 0\colon \;\; \forall\, y \in B_\delta(x) \cap D,\; y \neq x\colon \;\; f(x) < (>)\; f(y).$$

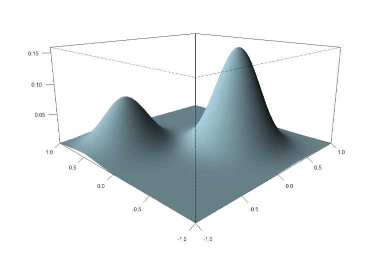
isolated local minimum (maximum) point of f if

$$\exists\, \delta > 0\colon \;\; \forall\, y \in B_\delta(x) \cap D,\; y \neq x\colon \;\; f(x) < (>)\; f(y).$$

isolated (local) extremum point of f if

it is an isolated (local) minimum point or an isolated (local) maximum point of f.





One isolated (local) maximum point.

One isolated maximum point, and one further isolated local maximum point.

Theorem 1.11.3 (☞ Satz I.3.7 in Forster 2)

Let $D \subseteq \mathbb{R}^n$ and $f : D \to \mathbb{R}$. If D is compact and f is continuous, then f possesses both a maximum point and a minimum point.

Recall

If $n = 1$, the following two assertions hold true (☞ Part I) for a function $f : D \to \mathbb{R}$ (with $D \subseteq \mathbb{R}$ open) and $x \in D$:

(i) If f is differentiable at x and x is a local extremum point of f, then $f'(x) = 0$.

(ii) If f is twice continuously differentiable at x, $f'(x) = 0$ and $f''(x) > 0$ (< 0), then x is an isolated local minimum (maximum) point of f.

Objective

Analogous statements about a function $f : D \to \mathbb{R}$ with $D \subseteq \mathbb{R}^n$.

(i) ☞ 1.11.4

(ii) ☞ 1.11.8

Theorem 1.11.4 (Necessary condition for local extremum point)

Let $D \subseteq \mathbb{R}^n$ *be open,* $x \in D$ *and* $f : D \to \mathbb{R}$*.*

If f *is totally differentiable at* x *and* x *is a local extremum point of* f*, then* x *is a stationary point of* f *(i.e.* $\nabla f(x) = \mathbf{0}$*).*

Definition 1.11.5

Let $D \subseteq \mathbb{R}^n$ be open and $f : D \to \mathbb{R}$ be partially differentiable.

Then a point $x \in D$ is called a **stationary point of** f if $\nabla f(x) = \mathbf{0}$.

The necessary condition $\nabla f(x) = \mathbf{0}$ in 1.11.4 typically leads to a system of n (possibly nonlinear) equations in n variables.

Note that not every stationary point is a local extremum point (☞ 1.11.7). This leads to the concept of saddle points (☞ 1.11.6).

Theorem 1.11.4 (Necessary condition for local extremum point)

Let $D \subseteq \mathbb{R}^n$ be open, $x \in D$ and $f : D \to \mathbb{R}$.

If f is totally differentiable at x and x is a local extremum point of f, then x is a stationary point of f (i.e. $\nabla f(x) = \mathbf{0}$).

Definition 1.11.5

Let $D \subseteq \mathbb{R}^n$ be open and $f : D \to \mathbb{R}$ be partially differentiable.

Then a point $x \in D$ is called a **stationary point of** f if $\nabla f(x) = \mathbf{0}$.

Definition 1.11.6

Let $D \subseteq \mathbb{R}^n$ be open, and $f : D \to \mathbb{R}$ be partially differentiable.

Then a stationary point $x \in D$ of f is called a **saddle point of** f if it is **not** a local extremum point of f.

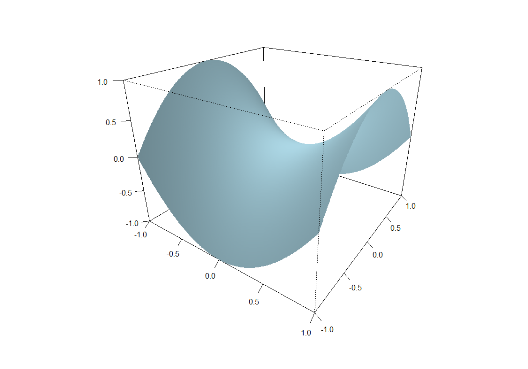
That is, a stationary point $x \in D$ of f is a saddle point of f if and only if

$$\forall\, \delta > 0\colon \ \ \exists\, x_1, x_2 \in B_\delta(x) \cap D\colon \ \ f(x_1) < f(x) < f(x_2).$$

Example 1.11.7

The function $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x_1, x_2) := x_1^2 - x_2^2$.

Show that f is a C^1 function and find all stationary points of f. Moreover, show that f does not possess any local extremum point.



Theorem 1.11.8 (Sufficient condition for local extremum point)

Let $D \subseteq \mathbb{R}^n$, $x \in D$, $f : D \to \mathbb{R}$ *and* $\delta > 0$ *be such that* $B_\delta(x) \subseteq D$.

Assume that f *is a* C^2 *function on* $B_\delta(x)$ *(i.e. that* f *is twice continuously differentiable at every point of* $B_\delta(x)$*) and that* x *is a stationary point of* f *(i.e.* $\nabla f(x) = \mathbf{0}$*). Then:*

(i) *If* $\mathrm{Hess}\, f(x)$ *is positive definite, then* x *is an isolated local minimum point of* f.

(ii) *If* $\mathrm{Hess}\, f(x)$ *is negative definite, then* x *is an isolated local maximum point of* f.

(iii) *If* $\mathrm{Hess}\, f(x)$ *is indefinite, then* x *is a saddle point of* f.

A sort of converse of 1.11.8 can be found in 1.11.11 below.

Remark 1.11.9

A matrix $H \in \mathbb{R}^{n\times n}$ is positive definite if and only if the leading principal minors $\delta_1(H), \ldots, \delta_n(H)$ are all strictly positive.

A matrix $H \in \mathbb{R}^{n\times n}$ is negative definite if and only if the matrix $-H$ is positive definite.

These assertions are known from Part II.

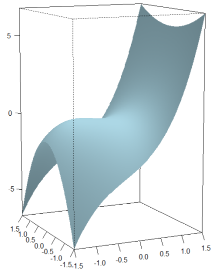
Recall that for $r = 1, \ldots, n$ the r-th leading principal minor is defined by

$$\delta_r(H) := \det(H_r) \quad \text{with} \quad H_r := \begin{bmatrix} h_{1,1} & \cdots & h_{1,r} \\ \vdots & & \vdots \\ h_{r,1} & \cdots & h_{r,r} \end{bmatrix}.$$

Example 1.11.10

The fct. $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x_1, x_2) := x_2^2(x_1 - 1) + x_1^2(x_1 + 1)$. Show that f is a C^2 function and find all stationary points of f.

Moreover, classify the stationary points, i.e. decide for any stationary point whether it is a local minimum point, a local maximum point or a saddle point of f.



Theorem 1.11.11

Let $D \subseteq \mathbb{R}^n$, $x \in D$, $f : D \to \mathbb{R}$ and $\delta > 0$ be such that $B_\delta(x) \subseteq D$.

Assume that f is a C^2 function on $B_\delta(x)$ (i.e. that f is twice continuously differentiable at every point of $B_\delta(x)$). Then:

(i) *If x is a local minimum point of f, then $\operatorname{Hess} f(x)$ is positive semi-definite.*

(ii) *If x is a local maximum point of f, then $\operatorname{Hess} f(x)$ is negative semi-definite.*

If $f : D \to \mathbb{R}$ (with $D \subseteq \mathbb{R}^n$ open) is a C^2 fct., then the following implications hold for any stationary point x of f. Other implications do **not** generally apply!

$$\begin{tikzcd}
x \text{ local minimum point} \arrow[r, Rightarrow, "1.11.11(\mathrm{i})"'] & \operatorname{Hess} f(x) \text{ positive semi-definite} \\
x \text{ isolated local minimum point} \arrow[u, Rightarrow] & \operatorname{Hess} f(x) \text{ positive definite} \arrow[l, Rightarrow, "1.11.8(\mathrm{i})"'] \arrow[u, Rightarrow] \\
x \text{ local maximum point} \arrow[r, Rightarrow, "1.11.11(\mathrm{ii})"'] & \operatorname{Hess} f(x) \text{ negative semi-definite} \\
x \text{ isolated local maximum point} \arrow[u, Rightarrow] & \operatorname{Hess} f(x) \text{ negative definite} \arrow[l, Rightarrow, "1.11.8(\mathrm{ii})"'] \arrow[u, Rightarrow] \\
x \text{ saddle point} & \operatorname{Hess} f(x) \text{ indefinite} \arrow[l, Rightarrow, "1.11.8(\mathrm{iii})"]
\end{tikzcd}$$

Remark 1.11.12 (Global extremum points)

Theorems 1.11.4 and 1.11.8 can be used to find and to classify stationary points of a C^2 function $f : D \to \mathbb{R}$ (with $D \subseteq \mathbb{R}^n$ open).

However, additional work is required to decide whether the local extremum points found are even "global" extremum points.

Here are two examples:

(1) Assume that $f : \mathbb{R}^n \to \mathbb{R}$ is a C^2 function. Then one can use 1.11.4 and 1.11.8 to specify the local extremum points of f:

One of these points, say x, is a minimum (maximum) point of f if

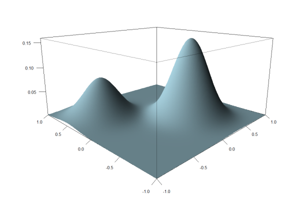
- it has the smallest (greatest) function value among all local minimum (maximum) points of f, and
- $\liminf_{N\to\infty} f(x^N) \geq f(x)$ ($\limsup_{N\to\infty} f(x^N) \leq f(x)$) holds for *every* sequence $(x^N)_{N\in\mathbb{N}}$ in $\mathbb{R}^n$ with $\|x^N\| \to \infty$.

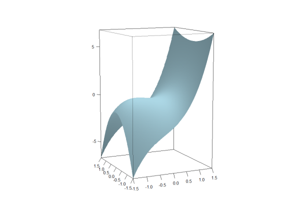
Remark 1.11.12 (Global extremum points, continued)

(2) Assume that $f : \overline{D} \to \mathbb{R}$ is a continuous function on the closed interval $\overline{D} := [a_1, b_1] \times \cdots \times [a_n, b_n]$ and that its restriction to the open interval $D := (a_1, b_1) \times \cdots \times (a_n, b_n)$ is a C^2 function. Then one can use 1.11.4 and 1.11.8 to specify the local extremum points of f on D.

One of these points, say x, is a minimum (maximum) point of f if

- it has the smallest (greatest) function value among all local minimum (maximum) points of f on D, and
- $f(x) \leq (\geq)\ f(y)$ holds for *every* point y on the boundary of $[a_1, b_1] \times \cdots \times [a_n, b_n]$.





In order to show that a local minimum (maximum) point x of a function $f : D \to \mathbb{R}$ is **not** a minimum (maximum) point of f, it is of course sufficient to find **one** point $y \in D$ with $f(y) < (>)\ f(x)$.

Example 1.11.13

Decide whether or not the local extremum points of the function f in Example 1.11.10 are even extremum points of f.

1. Multidimensional analysis

1.12 Newton's method

Forster, Analysis 1, Section 17
Heuser, Lehrbuch der Analysis 2, Section XXII.189

Motivation

In Section 1.11 it was mentioned (☞ slide 94) that the necessary condition $\nabla f(x) = \mathbf{0}$ for x to be a local extremum point of f typically leads to a system of n equations in n variables:

$$\begin{aligned} \frac{\partial f}{\partial x_1}(x_1, \ldots, x_n) &= 0 \\ &\vdots \\ \frac{\partial f}{\partial x_n}(x_1, \ldots, x_n) &= 0 \end{aligned}$$

If these n equations are not linear ones, so that methods such as the Gaussian elimination do **not** apply, then the determination of solutions $x = (x_1, \ldots, x_n)$ is not necessarily straightforward.

Motivation

The equation system on the previous slide is a special equation system of the general form

$$\begin{aligned} f_1(x_1,\dots,x_n) &= 0 \\ &\vdots \\ f_n(x_1,\dots,x_n) &= 0 \end{aligned}$$

Solving this equation system is equivalent to finding the zeros of the $\mathbb{R}^n$-valued function $f := (f_1,\dots,f_n)$.

If the explicit determination of the zeros is difficult/impossible, then numerical methods such as the Newton method can be used.

On the next few slides, Newton's method will be introduced, first for $n = 1$, then for arbitrary $n \in \mathbb{N}$.

Newton's method for $n = 1$

Let $f : D \to \mathbb{R}$ (with $D \subseteq \mathbb{R}$ open) be a C^1 function.

To obtain an approximation of a zero x^* of f, i.e. of an $x^* \in D$ that satisfies $f(x^*) = 0$, one can proceed as follows:

0. Choose an $x_0 \in D$ which is known to be "close" to a zero x^* of f.
1. Replace $f(x)$ by the order-1 Taylor polynomial of f at x_0 (☞ 1.9.7),

   $$T_1^f(x_0; x) = f(x_0) + f'(x_0)(x - x_0),$$

   and specify the zero, x_1, of $x \mapsto T_1^f(x_0; x)$ (if it exists):

   $$x_1 := x_0 - \frac{f(x_0)}{f'(x_0)} \qquad \big(\iff 0 = f(x_0) + f'(x_0)(x_1 - x_0) \big).$$

2. Replace $f(x)$ by the order-1 Taylor polynomial of f at x_1 (☞ 1.9.7),

   $$T_1^f(x_1; x) = f(x_1) + f'(x_1)(x - x_1),$$

   and specify the zero, x_2, of $x \mapsto T_1^f(x_1; x)$ (if it exists):

   $$x_2 := x_1 - \frac{f(x_1)}{f'(x_1)} \qquad \big(\iff 0 = f(x_1) + f'(x_1)(x_2 - x_1) \big).$$

$\vdots$

Let $f : D \to \mathbb{R}$ (with $D \subseteq \mathbb{R}$ open) be a C^1 function.

To obtain an approximation of a zero x^* of f, i.e. of an $x^* \in D$ that satisfies $f(x^*) = 0$, one can proceed as follows:

0. Choose an $x_0 \in D$ which is known to be "close" to a zero x^* of f.

   $\vdots$

k. Replace $f(x)$ by the order-1 Taylor polynomial of f at x_{k-1},

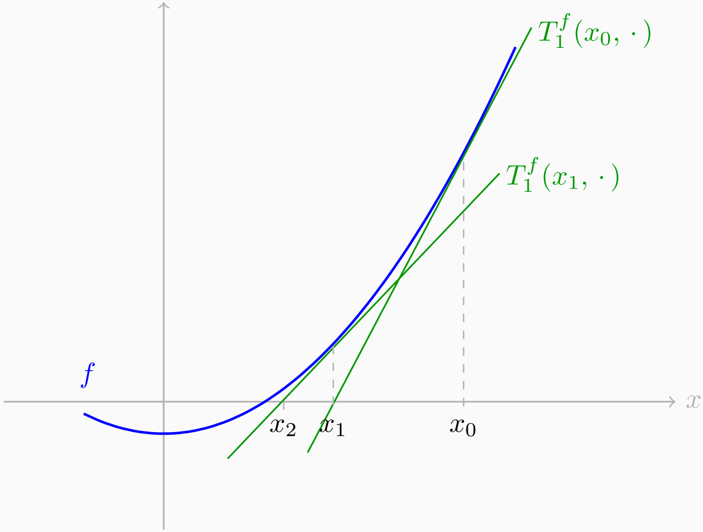
$$T_1^f(x_{k-1}; x) = f(x_{k-1}) + f'(x_{k-1})(x - x_{k-1}),$$

and specify the zero, x_k, of $x \mapsto T_1^f(x_{k-1}; x)$ (if it exists):

$$x_k := x_{k-1} - \frac{f(x_{k-1})}{f'(x_{k-1})}. \tag{1.9}$$

Stop if $f(x_k)$ is "sufficiently close" to 0.
Otherwise continue with Step $k + 1$.

This is the **Newton method**. See slide 110 for an illustration.



$T_1^f(x_0,\cdot)$
$T_1^f(x_1,\cdot)$
f
x
x_2
x_1
x_0

Let $(x_k)_{k\in\mathbb{N}}$ be the sequence obtained by performing the iterative steps on slide 109.

If $(x_k)_{k\in\mathbb{N}}$ is convergent, then the limit is a zero of f.

Indeed: If x^* is used to denote the limit of $(x_k)_{k\in\mathbb{N}}$, then by the continuity of f and f' we have that

$$x^* = \lim_{k\to\infty} x_k = \lim_{k\to\infty} \Big(x_{k-1} - \tfrac{f(x_{k-1})}{f'(x_{k-1})}\Big) = x^* - \tfrac{f(x^*)}{f'(x^*)}.$$

This implies $0 = \frac{f(x^*)}{f'(x^*)}$, i.e. $f(x^*) = 0$.

However, the sequence $(x_k)_{k\in\mathbb{N}}$ doesn't always converge.

The following Theorem 1.12.1 provides a sufficient condition for the convergence of $(x_k)_{k\in\mathbb{N}}$.

Theorem 1.12.1 (☞ Satz 17.2 in Forster 1)

Let $f : [a, b] \to \mathbb{R}$ be a continuous convex function that is twice differentiable on (a, b) and that satisfies $f(a) < 0 < f(b)$. Then:

(i) *There does exist exactly one $x^* \in (a, b)$ such that $f(x^*) = 0$.*

(ii) *If $x_0 \in [a, b]$ is an arbitrary point with $f(x_0) \geq 0$, then the sequence $(x_k)_{k\in\mathbb{N}}$ defined by*

$$x_k := x_{k-1} - \frac{f(x_{k-1})}{f'(x_{k-1})}, \qquad k \in \mathbb{N}$$

is well defined and converges from above to x^*.

(iii) *If $f'(x^*) \geq C$ and $f''(x) \leq K$ for all $x \in (x^*, b]$, for some fixed constants $C > 0$ and $K \geq 0$, then*

$$|x_{k+1} - x_k| \leq |x^* - x_k| \leq \frac{K}{2C}|x_k - x_{k-1}|^2 \quad \text{for all } k \in \mathbb{N}.$$

Analogous statements apply if f if concave or if $f(a) > 0 > f(b)$.

Remark 1.12.2

(i) Under the assumptions of Theorem 1.12.1(iii), we have

$$|x^* - x_k| \leq \tfrac{K}{2C}|x^* - x_{k-1}|^2 \quad \text{for all } k \in \mathbb{N}.$$

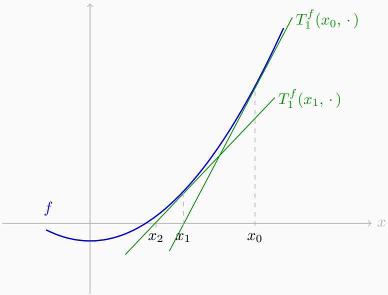
That is, in this case the convergence of x_k to x^* is "quadratic".

(ii) If the computation of $f'(\cdot)$ is costly, then one can replace (1.9) by

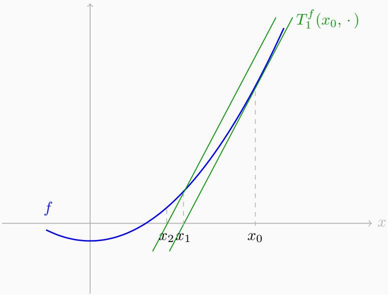
$$x_k := x_{k-1} - \tfrac{f(x_{k-1})}{f'(x_0)},$$

which leads to the **simplified Newton method**. However, for the simplified Newton method the convergence is only "linear".

For (ii), note that x_k is such that $0 = f(x_{k-1}) + f'(x_0)(x_k - x_{k-1})$.
See slide 114 for an illustration.



Newton method



Simplified Newton method

Example 1.12.3

Let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) := x^2 - 2$.

Specify both the Newton method and the simplified Newton method for finding the zero $x^* := \sqrt{2}$ of f and perform the first four steps of these methods. In either case, set $x_0 := 1$.

Newton's method for arbitrary $n \in \mathbb{N}$

Let $f = (f_1, \ldots, f_n) : D \to \mathbb{R}^n$ (with $D \subseteq \mathbb{R}^n$ open) be a C^1 function.

To obtain an approximation of a zero x^* of f, i.e. of an $x^* \in D$ that satisfies $f(x^*) = \mathbf{0}$, one can proceed as follows:

0. Choose an $x_{(0)} \in D$ which is known to be "close" to a zero x^* of f.

   $\vdots$

k. Replace $f(x)$ by the order-1 Taylor polynomial of f at $x_{(k-1)}$,

$$T_1^f(x_{(k-1)}; x) = f(x_{(k-1)}) + Jf(x_{(k-1)})(x - x_{(k-1)}),$$

and specify a zero $x_{(k)}$ of $x \mapsto T_1^f(x_{(k-1)}; x)$ (if it exists).
Note that you only need to solve a system of *linear* equations here
(☞ Slides 118, 119).

Stop if $f(x_{(k)})$ is "sufficiently close" to $\mathbf{0}$.
Otherwise continue with Step $k + 1$.

This is the multidimensional **Newton method**.

Note

By the Taylor polynomial

$$T_1^f(x_{(k-1)}; x) = f(x_{(k-1)}) + Jf(x_{(k-1)})(x - x_{(k-1)})$$

of $f = (f_1, \ldots, f_n)$ that appeared on the previous slide we mean

$$\begin{bmatrix} T_1^{f_1}(x_{(k-1)}, x) \\ \vdots \\ T_1^{f_n}(x_{(k-1)}, x) \end{bmatrix}$$

$$= \begin{bmatrix} f_1(x_{(k-1)}) \\ \vdots \\ f_n(x_{(k-1)}) \end{bmatrix} + \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x_{(k-1)}) & \cdots & \frac{\partial f_1}{\partial x_n}(x_{(k-1)}) \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1}(x_{(k-1)}) & \cdots & \frac{\partial f_n}{\partial x_n}(x_{(k-1)}) \end{bmatrix} \begin{bmatrix} x_1 - x_{(k-1),1} \\ \vdots \\ x_n - x_{(k-1),n} \end{bmatrix}$$

To find a zero of this polynomial, we need to solve the following **linear** equation system: ...

Note

...

$$
\begin{bmatrix} \frac{\partial f_1}{\partial x_1}\big(x_{(k-1)}\big) & \cdots & \frac{\partial f_1}{\partial x_n}\big(x_{(k-1)}\big) \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1}\big(x_{(k-1)}\big) & \cdots & \frac{\partial f_n}{\partial x_n}\big(x_{(k-1)}\big) \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}
$$

$$
= \begin{bmatrix} \frac{\partial f_1}{\partial x_1}\big(x_{(k-1)}\big) & \cdots & \frac{\partial f_1}{\partial x_n}\big(x_{(k-1)}\big) \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1}\big(x_{(k-1)}\big) & \cdots & \frac{\partial f_n}{\partial x_n}\big(x_{(k-1)}\big) \end{bmatrix} \begin{bmatrix} x_{(k-1),1} \\ \vdots \\ x_{(k-1),n} \end{bmatrix} - \begin{bmatrix} f_1\big(x_{(k-1)}\big) \\ \vdots \\ f_n\big(x_{(k-1)}\big) \end{bmatrix}
$$