V. Functions of several variables

V.1. $\mathbb{R}^n$ as a linear and metric space

Definition. The set $\mathbb{R}^n$, $n \in \mathbb{N}$, is the set of all ordered n-tuples of real numbers, i.e.

$$\mathbb{R}^n = \{[x_1, \dots, x_n] : x_1, \dots, x_n \in \mathbb{R}\}.$$

For $\boldsymbol{x} = [x_1, \dots, x_n] \in \mathbb{R}^n$, $\boldsymbol{y} = [y_1, \dots, y_n] \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ we set

$$\boldsymbol{x} + \boldsymbol{y} = [x_1 + y_1, \dots, x_n + y_n], \qquad \alpha\boldsymbol{x} = [\alpha x_1, \dots, \alpha x_n].$$

Further, we denote $\boldsymbol{o} = [0, \dots, 0]$ – the *origin*.

Definition. The *Euclidean metric (distance) on* $\mathbb{R}^n$ is the function $\rho\colon \mathbb{R}^n \times \mathbb{R}^n \to [0, +\infty)$ defined by

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

The number $\rho(\boldsymbol{x}, \boldsymbol{y})$ is called the *distance of the point $\boldsymbol{x}$ from the point $\boldsymbol{y}$*.

Theorem 1 (properties of the Euclidean metric)**.** *The Euclidean metric ρ has the following properties:*

(i) $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n\colon \rho(\boldsymbol{x}, \boldsymbol{y}) = 0 \Leftrightarrow \boldsymbol{x} = \boldsymbol{y}$,

(ii) $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n\colon \rho(\boldsymbol{x}, \boldsymbol{y}) = \rho(\boldsymbol{y}, \boldsymbol{x})$, *(symmetry)*

(iii) $\forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^n\colon \rho(\boldsymbol{x}, \boldsymbol{y}) \le \rho(\boldsymbol{x}, \boldsymbol{z}) + \rho(\boldsymbol{z}, \boldsymbol{y})$, *(triangle inequality)*

(iv) $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n, \forall \lambda \in \mathbb{R}\colon \rho(\lambda\boldsymbol{x}, \lambda\boldsymbol{y}) = |\lambda|\rho(\boldsymbol{x}, \boldsymbol{y})$, *(homogeneity)*

(v) $\forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^n\colon \rho(\boldsymbol{x} + \boldsymbol{z}, \boldsymbol{y} + \boldsymbol{z}) = \rho(\boldsymbol{x}, \boldsymbol{y})$. *(translation invariance)*

Definition. Let $\boldsymbol{x} \in \mathbb{R}^n,\ r \in \mathbb{R}, r > 0$. The set $B(\boldsymbol{x}, r)$ defined by

$$B(\boldsymbol{x}, r) = \{\boldsymbol{y} \in \mathbb{R}^n;\ \rho(\boldsymbol{x}, \boldsymbol{y}) < r\}$$

is called an *open ball with radius r centred at $\boldsymbol{x}$* or the *neighbourhood of $\boldsymbol{x}$*.

Definition. Let $M \subset \mathbb{R}^n$. We say that $\boldsymbol{x} \in \mathbb{R}^n$ is an *interior point of* M, if there exists $r > 0$ such that $B(\boldsymbol{x}, r) \subset M$.
The set of all interior points of M is called the *interior of* M and is denoted by $\operatorname{Int} M$.
The set $M \subset \mathbb{R}^n$ is *open in* $\mathbb{R}^n$, if each point of M is an interior point of M, i.e. if $M = \operatorname{Int} M$.

Theorem 2 (properties of open sets)**.**

(i) The empty set and $\mathbb{R}^n$ are open in $\mathbb{R}^n$.

(ii) Let $G_\alpha \subset \mathbb{R}^n$, $\alpha \in A \neq \emptyset$, be open in $\mathbb{R}^n$. Then $\bigcup_{\alpha \in A} G_\alpha$ is open in $\mathbb{R}^n$.

(iii) Let $G_i \subset \mathbb{R}^n$, $i = 1, \dots, m$, be open in $\mathbb{R}^n$. Then $\bigcap_{i=1}^{m} G_i$ is open in $\mathbb{R}^n$.

Remark.
(ii) *A union of an arbitrary system of open sets is an open set.*
(iii) *An intersection of a finitely many open sets is an open set.*

Definition. Let $M \subset \mathbb{R}^n$ and $\boldsymbol{x} \in \mathbb{R}^n$. We say that $\boldsymbol{x}$ is a *boundary point of* M if for each $r > 0$

$$B(\boldsymbol{x}, r) \cap M \neq \emptyset \quad \text{and} \quad B(\boldsymbol{x}, r) \cap (\mathbb{R}^n \setminus M) \neq \emptyset.$$

The *boundary of* M is the set of all boundary points of M (notation $\operatorname{bd} M$).
The *closure* of M is the set $M \cup \operatorname{bd} M$ (notation $\overline{M}$).
A set $M \subset \mathbb{R}^n$ is said to be *closed in* $\mathbb{R}^n$ if it contains all its boundary points, i.e. if $\operatorname{bd} M \subset M$, or in other words if $\overline{M} = M$.

Definition. Let $\boldsymbol{x}^j \in \mathbb{R}^n$ for each $j \in \mathbb{N}$ and $\boldsymbol{x} \in \mathbb{R}^n$. We say that a sequence $\{\boldsymbol{x}^j\}_{j=1}^{\infty}$ *converges to* $\boldsymbol{x}$, if

$$\lim_{j \to \infty} \rho(\boldsymbol{x}, \boldsymbol{x}^j) = 0.$$

The vector $\boldsymbol{x}$ is called the *limit of the sequence* $\{\boldsymbol{x}^j\}_{j=1}^{\infty}$.
The sequence $\{\boldsymbol{y}^j\}_{j=1}^{\infty}$ of points in $\mathbb{R}^n$ is called *convergent* if there exists $\boldsymbol{y} \in \mathbb{R}^n$ such that $\{\boldsymbol{y}^j\}_{j=1}^{\infty}$ converges to $\boldsymbol{y}$.

Remark. The sequence $\{\boldsymbol{x}^j\}_{j=1}^{\infty}$ converges to $\boldsymbol{x} \in \mathbb{R}^n$ if and only if

$$\forall \varepsilon \in \mathbb{R}, \varepsilon > 0 \,\exists j_0 \in \mathbb{N} \,\forall j \in \mathbb{N}, j \geq j_0 \colon \boldsymbol{x}^j \in B(\boldsymbol{x}, \varepsilon).$$

Theorem 3 (convergence is coordinatewise)**.** *Let $\boldsymbol{x}^j \in \mathbb{R}^n$ for each $j \in \mathbb{N}$ and let $\boldsymbol{x} \in \mathbb{R}^n$. The sequence $\{\boldsymbol{x}^j\}_{j=1}^{\infty}$ converges to $\boldsymbol{x}$ if and only if for each $i \in \{1, \dots, n\}$ the sequence of real numbers $\{x_i^j\}_{j=1}^{\infty}$ converges to the real number x_i.*

Remark. Theorem 3 says that the convergence in the space $\mathbb{R}^n$ is the same as the "coordinatewise" convergence. It follows that a sequence $\{\boldsymbol{x}^j\}_{j=1}^{\infty}$ has at most one limit. If it exists, then we denote it by $\lim_{j\to\infty} \boldsymbol{x}^j$. Sometimes we also write simply $\boldsymbol{x}^j \to \boldsymbol{x}$ instead of $\lim_{j\to\infty} \boldsymbol{x}^j = \boldsymbol{x}$.

Theorem 4 (characterisation of closed sets)**.** *Let $M \subset \mathbb{R}^n$. Then the following statements are equivalent:*

(i) M is closed in $\mathbb{R}^n$.

(ii) $\mathbb{R}^n \setminus M$ is open in $\mathbb{R}^n$.

(iii) Any $\boldsymbol{x} \in \mathbb{R}^n$ which is a limit of a sequence from M belongs to M.

Theorem 5 (properties of closed sets)**.**

(i) The empty set and the whole space $\mathbb{R}^n$ are closed in $\mathbb{R}^n$.

(ii) Let $F_\alpha \subset \mathbb{R}^n$, $\alpha \in A \neq \emptyset$, be closed in $\mathbb{R}^n$. Then $\bigcap_{\alpha\in A} F_\alpha$ is closed in $\mathbb{R}^n$.

(iii) Let $F_i \subset \mathbb{R}^n$, $i = 1, \dots, m$, be closed in $\mathbb{R}^n$. Then $\bigcup_{i=1}^{m} F_i$ is closed in $\mathbb{R}^n$.

Remark.
(ii) *An intersection of an arbitrary system of closed sets is closed.*
(iii) *A union of finitely many closed sets is closed.*

Theorem 6. *Let $M \subset \mathbb{R}^n$. Then the following holds:*

(i) The set $\overline{M}$ is closed in $\mathbb{R}^n$.

(ii) The set $\operatorname{Int} M$ is open in $\mathbb{R}^n$.

(iii) The set M is open in $\mathbb{R}^n$ if and only if $M = \operatorname{Int} M$.

Remark. The set $\operatorname{Int} M$ is the largest open set contained in M in the following sense: If G is a set open in $\mathbb{R}^n$ and satisfying $G \subset M$, then $G \subset \operatorname{Int} M$. Similarly $\overline{M}$ is the smallest closed set containing M.

Definition. We say that the *set* $M \subset \mathbb{R}^n$ is *bounded* if there exists $r > 0$ such that $M \subset B(\boldsymbol{o}, r)$. A *sequence* of points in $\mathbb{R}^n$ is *bounded* if the set of its members is bounded.

Theorem 7. *A set $M \subset \mathbb{R}^n$ is bounded if and only if its closure $\overline{M}$ is bounded.*

V.2. Continuous functions of several variables

Definition. Let $M \subset \mathbb{R}^n$, $\boldsymbol{x} \in M$, and $f \colon M \to \mathbb{R}$. We say that f is *continuous at $\boldsymbol{x}$ with respect to M*, if we

$$\forall \varepsilon \in \mathbb{R}, \varepsilon > 0 \,\exists \delta \in \mathbb{R}, \delta > 0 \,\forall \boldsymbol{y} \in B(\boldsymbol{x}, \delta) \cap M \colon f(\boldsymbol{y}) \in B(f(\boldsymbol{x}), \varepsilon).$$

We say that f is *continuous at the point $\boldsymbol{x}$* if it is continuous at $\boldsymbol{x}$ with respect to a neighbourhood of $\boldsymbol{x}$, i.e.

$$\forall \varepsilon \in \mathbb{R}, \varepsilon > 0 \,\exists \delta \in \mathbb{R}, \delta > 0 \,\forall \boldsymbol{y} \in B(\boldsymbol{x}, \delta) \colon f(\boldsymbol{y}) \in B(f(\boldsymbol{x}), \varepsilon).$$

Theorem 8. *Let $M \subset \mathbb{R}^n$, $\boldsymbol{x} \in M$, $f \colon M \to \mathbb{R}$, $g \colon M \to \mathbb{R}$, and $c \in \mathbb{R}$. If f and g are continuous at the point $\boldsymbol{x}$ with respect to M, then the functions cf, $f + g$ a fg are continuous at $\boldsymbol{x}$ with respect to M. If the function g is nonzero at $\boldsymbol{x}$, then also the function f/g is continuous at $\boldsymbol{x}$ with respect to M.*

Theorem 9. *Let $r, s \in \mathbb{N}$, $M \subset \mathbb{R}^s$, $L \subset \mathbb{R}^r$, and $\boldsymbol{y} \in M$. Let $\varphi_1, \dots, \varphi_r$ be functions defined on M, which are continuous at $\boldsymbol{y}$ with respect to M and $[\varphi_1(\boldsymbol{x}), \dots, \varphi_r(\boldsymbol{x})] \in L$ for each $\boldsymbol{x} \in M$. Let $f \colon L \to \mathbb{R}$ be continuous at the point $[\varphi_1(\boldsymbol{y}), \dots, \varphi_r(\boldsymbol{y})]$ with respect to L. Then the compound function $F \colon M \to \mathbb{R}$ defined by*

$$F(\boldsymbol{x}) = f\big(\varphi_1(\boldsymbol{x}), \dots, \varphi_r(\boldsymbol{x})\big), \quad \boldsymbol{x} \in M,$$

is continuous at $\boldsymbol{y}$ with respect to M.

Theorem 10 (Heine)**.** *Let $M \subset \mathbb{R}^n$, $\boldsymbol{x} \in M$, and $f\colon M \to \mathbb{R}$. Then the following are equivalent.*

(i) The function f is continuous at $\boldsymbol{x}$ with respect to M.

(ii) $\lim_{j\to\infty} f(\boldsymbol{x}^j) = f(\boldsymbol{x})$ for each sequence $\{\boldsymbol{x}^j\}_{j=1}^{\infty}$ such that $\boldsymbol{x}^j \in M$ for $j \in \mathbb{N}$ and $\lim_{j\to\infty} \boldsymbol{x}^j = \boldsymbol{x}$.

Definition. Let $M \subset \mathbb{R}^n$ and $f\colon M \to \mathbb{R}$. We say that f is *continuous on M* if it is continuous at each point $\boldsymbol{x} \in M$ with respect to M.

Remark. The functions $\pi_j\colon \mathbb{R}^n \to \mathbb{R}$, $\pi_j(\boldsymbol{x}) = x_j$, $1 \le j \le n$, are continuous on $\mathbb{R}^n$. They are called *coordinate projections*.

Theorem 11. *Let f be a continuous function on $\mathbb{R}^n$ and $c \in \mathbb{R}$. Then the following holds:*

(i) The set $\{\boldsymbol{x} \in \mathbb{R}^n;\ f(\boldsymbol{x}) < c\}$ is open in $\mathbb{R}^n$.

(ii) The set $\{\boldsymbol{x} \in \mathbb{R}^n;\ f(\boldsymbol{x}) > c\}$ is open in $\mathbb{R}^n$.

(iii) The set $\{\boldsymbol{x} \in \mathbb{R}^n;\ f(\boldsymbol{x}) \le c\}$ is closed in $\mathbb{R}^n$.

(iv) The set $\{\boldsymbol{x} \in \mathbb{R}^n;\ f(\boldsymbol{x}) \ge c\}$ is closed in $\mathbb{R}^n$.

(v) The set $\{\boldsymbol{x} \in \mathbb{R}^n;\ f(\boldsymbol{x}) = c\}$ is closed in $\mathbb{R}^n$.

Definition. We say that a set $M \subset \mathbb{R}^n$ is *compact* if for each sequence of elements of M there exists a convergent subsequence with a limit in M.

Theorem 12 (characterisation of compact subsets of $\mathbb{R}^n$)**.** *The set $M \subset \mathbb{R}^n$ is compact if and only if M is bounded and closed.*

Lemma 13. *Omitted.*

Definition. Let $M \subset \mathbb{R}^n$, $\boldsymbol{x} \in M$, and let f be a function defined at least on M (i.e. $M \subset D_f$). We say that f attains at the point $\boldsymbol{x}$ its

- *maximum on M* if $f(\boldsymbol{y}) \le f(\boldsymbol{x})$ for every $\boldsymbol{y} \in M$,
- *local maximum with respect to M* if there exists $\delta > 0$ such that $f(\boldsymbol{y}) \le f(\boldsymbol{x})$ for every $\boldsymbol{y} \in B(\boldsymbol{x}, \delta) \cap M$,
- *strict local maximum with respect to M* if there exists $\delta > 0$ such that $f(\boldsymbol{y}) < f(\boldsymbol{x})$ for every $\boldsymbol{y} \in \big(B(\boldsymbol{x}, \delta) \setminus \{\boldsymbol{x}\}\big) \cap M$.

The notions of a *minimum*, a *local minimum*, and a *strict local minimum* with respect to M are defined in analogous way.

Definition. We say that a function f attains a *local maximum* at a point $\boldsymbol{x} \in \mathbb{R}^n$ if $\boldsymbol{x}$ is a local maximum with respect to some neighbourhood of $\boldsymbol{x}$.

Similarly we define *local minimum*, *strict local maximum* and *strict local minimum*.

Theorem 14 (attaining extrema)**.** *Let $M \subset \mathbb{R}^n$ be a non-empty compact set and $f\colon M \to \mathbb{R}$ a function continuous on M. Then f attains its maximum and minimum on M.*

Corollary. *Let $M \subset \mathbb{R}^n$ be a non-empty compact set and $f\colon M \to \mathbb{R}$ a continuous function on M. Then f is bounded on M.*

Definition. We say that a function f of n variables has a limit at a point $\boldsymbol{a} \in \mathbb{R}^n$ equal to $A \in \mathbb{R}^*$ if

$$\forall \varepsilon \in \mathbb{R}, \varepsilon > 0\ \exists \delta \in \mathbb{R}, \delta > 0\ \forall \boldsymbol{x} \in B(\boldsymbol{a}, \delta) \setminus \{\boldsymbol{a}\}\colon f(\boldsymbol{x}) \in B(A, \varepsilon).$$

Remark.

- Each function has at a given point at most one limit. We write $\lim_{\boldsymbol{x}\to\boldsymbol{a}} f(\boldsymbol{x}) = A$.
- The function f is continuous at $\boldsymbol{a}$ if and only if $\lim_{\boldsymbol{x}\to\boldsymbol{a}} f(\boldsymbol{x}) = f(\boldsymbol{a})$.
- For limits of functions of several variables one can prove similar theorems as for limits of functions of one variable (arithmetics, the sandwich theorem, …).

Theorem 15. *Let $r, s \in \mathbb{N}$, $\boldsymbol{a} \in \mathbb{R}^s$, and let $\varphi_1, \dots, \varphi_r$ be functions of s variables such that $\lim_{\boldsymbol{x}\to\boldsymbol{a}} \varphi_j(\boldsymbol{x}) = b_j$, $j = 1, \dots, r$. Set $\mathbf{b} = [b_1, \dots, b_r]$. Let f be a function of r variables which is continuous at the point $\mathbf{b}$. If we define a compound function F of s variables by*

$$F(\boldsymbol{x}) = f(\varphi_1(\boldsymbol{x}), \varphi_2(\boldsymbol{x}), \dots, \varphi_r(\boldsymbol{x})),$$

then $\lim_{\boldsymbol{x}\to\boldsymbol{a}} F(\boldsymbol{x}) = f(\boldsymbol{b})$.

V.3. Partial derivatives and tangent hyperplane

Set $\boldsymbol{e}^j = [0,\dots,0,\underset{j\text{th coordinate}}{1},0,\dots,0]$.

Definition. Let f be a function of n variables, $j \in \{1,\dots,n\}$, $\boldsymbol{a} \in \mathbb{R}^n$. Then the number

$$\begin{aligned}\frac{\partial f}{\partial x_j}(\boldsymbol{a}) &= \lim_{t\to 0} \frac{f(\boldsymbol{a}+t\boldsymbol{e}^j) - f(\boldsymbol{a})}{t} \\ &= \lim_{t\to 0} \frac{f(a_1,\dots,a_{j-1},a_j+t,a_{j+1},\dots,a_n) - f(a_1,\dots,a_n)}{t}\end{aligned}$$

is called the *partial derivative (of first order) of function f according to jth variable at the point $\boldsymbol{a}$* (if the limit exists).

Theorem 16 (necessary condition of the existence of local extremum)**.** *Let $G \subset \mathbb{R}^n$ be an open set, $\boldsymbol{a} \in G$, and suppose that a function $f\colon G \to \mathbb{R}$ has a local extremum (i.e. a local maximum or a local minimum) at the point $\boldsymbol{a}$. Then for each $j \in \{1,\dots,n\}$ the following holds:*

The partial derivative $\dfrac{\partial f}{\partial x_j}(\boldsymbol{a})$ either does not exist or it is equal to zero.

Definition. Let $G \subset \mathbb{R}^n$ be a non-empty open set. If a function $f\colon G \to \mathbb{R}$ has all partial derivatives continuous at each point of the set G (i.e. the function $\boldsymbol{x} \mapsto \frac{\partial f}{\partial x_j}(\boldsymbol{x})$ is continuous on G for each $j \in \{1,\dots,n\}$), then we say that f is of the *class $\mathcal{C}^1$ on G*. The set of all of these functions is denoted by $C^1(G)$.

Remark. If $G \subset \mathbb{R}^n$ is a non-empty open set and and $f,g \in C^1(G)$, then $f+g \in C^1(G)$, $f-g \in C^1(G)$, and $fg \in C^1(G)$. If moreover $g(\boldsymbol{x}) \neq 0$ for each $\boldsymbol{x} \in G$, then $f/g \in C^1(G)$.

Proposition 17 (weak Lagrange theorem)**.** *Omitted.*

Definition. Let $G \subset \mathbb{R}^n$ be an open set, $\boldsymbol{a} \in G$, and $f \in C^1(G)$. Then the graph of the function

$$T\colon \boldsymbol{x} \mapsto f(\boldsymbol{a}) + \frac{\partial f}{\partial x_1}(\boldsymbol{a})(x_1-a_1) + \frac{\partial f}{\partial x_2}(\boldsymbol{a})(x_2-a_2) + \dots + \frac{\partial f}{\partial x_n}(\boldsymbol{a})(x_n-a_n), \quad \boldsymbol{x} \in \mathbb{R}^n,$$

is called the *tangent hyperplane* to the graph of the function f at the point $[\boldsymbol{a}, f(\boldsymbol{a})]$.

Theorem 18 (tangent hyperplane)**.** *Let $G \subset \mathbb{R}^n$ be an open set, $\boldsymbol{a} \in G$, $f \in C^1(G)$, and let T be a function whose graph is the tangent hyperplane of the function f at the point $[\boldsymbol{a}, f(\boldsymbol{a})]$. Then*

$$\lim_{\boldsymbol{x}\to\boldsymbol{a}} \frac{f(\boldsymbol{x}) - T(\boldsymbol{x})}{\rho(\boldsymbol{x},\boldsymbol{a})} = 0.$$

Theorem 19. *Let $G \subset \mathbb{R}^n$ be an open non-empty set and $f \in C^1(G)$. Then f is continuous on G.*

Remark. Existence of partial derivatives at a **does not** imply continuity at a.

Theorem 20 (derivative of a composite function; chain rule)**.** *Let $r,s \in \mathbb{N}$ and let $G \subset \mathbb{R}^s$, $H \subset \mathbb{R}^r$ be open sets. Let $\varphi_1,\dots,\varphi_r \in C^1(G)$, $f \in C^1(H)$ and $[\varphi_1(\boldsymbol{x}),\dots,\varphi_r(\boldsymbol{x})] \in H$ for each $\boldsymbol{x} \in G$. Then the compound function $F\colon G \to \mathbb{R}$ defined by*

$$F(\boldsymbol{x}) = f\big(\varphi_1(\boldsymbol{x}),\varphi_2(\boldsymbol{x}),\dots,\varphi_r(\boldsymbol{x})\big), \quad \boldsymbol{x} \in G,$$

is of the class $\mathcal{C}^1$ on G. Let $\boldsymbol{a} \in G$ and $\mathbf{b} = [\varphi_1(\boldsymbol{a}),\dots,\varphi_r(\boldsymbol{a})]$. Then for each $j \in \{1,\dots,s\}$ we have

$$\frac{\partial F}{\partial x_j}(\boldsymbol{a}) = \sum_{i=1}^{r} \frac{\partial f}{\partial y_i}(\boldsymbol{b}) \frac{\partial \varphi_i}{\partial x_j}(\boldsymbol{a}).$$

Definition. Let $G \subset \mathbb{R}^n$ be an open set, $\boldsymbol{a} \in G$, and $f \in C^1(G)$. The *gradient of f at the point $\boldsymbol{a}$* is the vector

$$\nabla f(\boldsymbol{a}) = \left[\frac{\partial f}{\partial x_1}(\boldsymbol{a}), \frac{\partial f}{\partial x_2}(\boldsymbol{a}), \dots, \frac{\partial f}{\partial x_n}(\boldsymbol{a})\right].$$

Remark. The gradient of f at a points in the direction of steepest growth of f at a. At every point, the gradient is perpendicular to the contour of f.

Definition. Let $G \subset \mathbb{R}^n$ be an open set, $\boldsymbol{a} \in G$, $f \in C^1(G)$, and $\nabla f(\boldsymbol{a}) = \boldsymbol{o}$. Then the point $\boldsymbol{a}$ is called a *stationary* (or *critical*) *point* of the function f.

Definition. Let $G \subset \mathbb{R}^n$ be an open set, $f\colon G \to \mathbb{R}$, $i, j \in \{1, \dots, n\}$, and suppose that $\frac{\partial f}{\partial x_i}(\boldsymbol{x})$ exists finite for each $\boldsymbol{x} \in G$. Then the *partial derivative of the second order* of the function f according to ith and jth variable at a point $\boldsymbol{a} \in G$ is defined by

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{a}) = \frac{\partial\left(\frac{\partial f}{\partial x_i}\right)}{\partial x_j}(\boldsymbol{a})$$

If $i = j$ then we use the notation $\frac{\partial^2 f}{\partial x_i^2}(\boldsymbol{a})$.

Similarly we define higher order partial derivatives.

Remark. In general it is not true that $\frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{a}) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\boldsymbol{a})$.

Theorem 21 (interchanging of partial derivatives)**.** *Let $i, j \in \{1, \dots, n\}$ and suppose that a function f has both partial derivatives $\frac{\partial^2 f}{\partial x_i \partial x_j}$ and $\frac{\partial^2 f}{\partial x_j \partial x_i}$ on a neighbourhood of a point $\boldsymbol{a} \in \mathbb{R}^n$ and that these functions are continuous at $\boldsymbol{a}$. Then*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{a}) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\boldsymbol{a}).$$

Definition. Let $G \subset \mathbb{R}^n$ be an open set and $k \in \mathbb{N}$. We say that a function f is of the *class $\mathcal{C}^k$ on G*, if all partial derivatives of f of all orders up to k are continuous on G. The set of all of these functions is denoted by $C^k(G)$.

We say that a function f is of the *class $\mathcal{C}^\infty$ on G*, if all partial derivatives of all orders of f are continuous on G. The set of all of these functions is denoted by $C^\infty(G)$.

V.4. Implicit function theorem and Lagrange multiplier theorem

Theorem 22 (implicit function)**.** *Let $G \subset \mathbb{R}^{n+1}$ be an open set, $F\colon G \to \mathbb{R}$, and $\tilde{\boldsymbol{x}} \in \mathbb{R}^n$, $\tilde{y} \in \mathbb{R}$ such that $[\tilde{\boldsymbol{x}}, \tilde{y}] \in G$. Suppose that*

(i) $F \in C^1(G)$,

(ii) $F(\tilde{\boldsymbol{x}}, \tilde{y}) = 0$,

(iii) $\dfrac{\partial F}{\partial y}(\tilde{\boldsymbol{x}}, \tilde{y}) \neq 0$.

Then there exist a neighbourhood $U \subset \mathbb{R}^n$ of the point $\tilde{\boldsymbol{x}}$ and a neighbourhood $V \subset \mathbb{R}$ of the point $\tilde{y}$ such that for each $\boldsymbol{x} \in U$ there exists a unique $y \in V$ satisfying $F(\boldsymbol{x}, y) = 0$. If we denote this y by $\varphi(\boldsymbol{x})$, then the resulting function φ is in $C^1(U)$ and

$$\frac{\partial \varphi}{\partial x_j}(\boldsymbol{x}) = -\frac{\frac{\partial F}{\partial x_j}(\boldsymbol{x}, \varphi(\boldsymbol{x}))}{\frac{\partial F}{\partial y}(\boldsymbol{x}, \varphi(\boldsymbol{x}))} \quad \text{for } \boldsymbol{x} \in U,\ j \in \{1, \dots, n\}.$$

Theorem 23 (Lagrange multiplier theorem)**.** *Let $G \subset \mathbb{R}^2$ be an open set, $f, g \in C^1(G)$, $M = \{[x, y] \in G;\ g(x, y) = 0\}$ and let $[\tilde{x}, \tilde{y}] \in M$ be a point of local extremum of f with respect to M. Then at least one of the following conditions holds:*

(I) $\nabla g(\tilde{x}, \tilde{y}) = \boldsymbol{o}$,

(II) there exists $\lambda \in \mathbb{R}$ satisfying

$$\begin{aligned}\frac{\partial f}{\partial x}(\tilde{x}, \tilde{y}) + \lambda \frac{\partial g}{\partial x}(\tilde{x}, \tilde{y}) = 0,\\ \frac{\partial f}{\partial y}(\tilde{x}, \tilde{y}) + \lambda \frac{\partial g}{\partial y}(\tilde{x}, \tilde{y}) = 0.\end{aligned}$$

Theorem 24 (implicit functions)**.** *Let $m, n \in \mathbb{N}$, $k \in \mathbb{N} \cup \{\infty\}$, $G \subset \mathbb{R}^{n+m}$ an open set, $F_j\colon G \to \mathbb{R}$ for $j = 1, \dots, m$, $\tilde{\boldsymbol{x}} \in \mathbb{R}^n$, $\tilde{\boldsymbol{y}} \in \mathbb{R}^m$, $[\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}] \in G$. Suppose that*

(i) $F_j \in C^k(G)$ *for all* $j \in \{1, \dots, m\}$,

(ii) $F_j(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) = 0$ *for all* $j \in \{1, \dots, m\}$,

(iii) $\begin{vmatrix} \frac{\partial F_1}{\partial y_1}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) & \dots & \frac{\partial F_1}{\partial y_m}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) \\ \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial y_1}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) & \dots & \frac{\partial F_m}{\partial y_m}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) \end{vmatrix} \neq 0.$

Then there are a neighbourhood $U \subset \mathbb{R}^n$ of $\tilde{\boldsymbol{x}}$ and a neighbourhood $V \subset \mathbb{R}^m$ of $\tilde{\boldsymbol{y}}$ such that for each $\boldsymbol{x} \in U$ there exists a unique $\boldsymbol{y} \in V$ satisfying $F_j(\boldsymbol{x}, \boldsymbol{y}) = 0$ for each $j \in \{1, \dots, m\}$. If we denote the coordinates of this $\boldsymbol{y}$ by $\varphi_j(\boldsymbol{x})$, then the resulting functions φ_j are in $C^k(U)$.

Remark. The symbol in the condition (iii) of Theorem 24 is called a *determinant*. The general definition will be given later.

For $m = 1$ we have $\begin{vmatrix} a \end{vmatrix} = a$, $a \in \mathbb{R}$. In particular, in this case the condition (iii) in Theorem 24 is the same as the condition (iii) in Theorem 22.

For $m = 2$ we have $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$, $a, b, c, d \in \mathbb{R}$.

Theorem 25 (Lagrange multipliers theorem)**.** *Let $m, n \in \mathbb{N}$, $m < n$, $G \subset \mathbb{R}^n$ an open set, $f, g_1, \dots, g_m \in C^1(G)$,*

$$M = \{\boldsymbol{z} \in G;\ g_1(\boldsymbol{z}) = 0, g_2(\boldsymbol{z}) = 0, \dots, g_m(\boldsymbol{z}) = 0\}$$

and let $\tilde{\boldsymbol{z}} \in M$ be a point of local extremum of f with respect to the set M. Then at least one of the following conditions holds:

(I) the vectors

$$\nabla g_1(\tilde{\boldsymbol{z}}), \nabla g_2(\tilde{\boldsymbol{z}}), \dots, \nabla g_m(\tilde{\boldsymbol{z}})$$

are linearly dependent,

(II) there exist numbers $\lambda_1, \lambda_2, \dots, \lambda_m \in \mathbb{R}$ satisfying

$$\nabla f(\tilde{\boldsymbol{z}}) + \lambda_1 \nabla g_1(\tilde{\boldsymbol{z}}) + \lambda_2 \nabla g_2(\tilde{\boldsymbol{z}}) + \cdots + \lambda_m \nabla g_m(\tilde{\boldsymbol{z}}) = \boldsymbol{o}.$$

Remark.

- The notion of *linearly dependent vectors* will be defined later.
  For $m = 1$: One vector is linearly dependent if it is the zero vector.
  For $m = 2$: Two vectors are linearly dependent if one of them is a multiple of the other one.
- The numbers $\lambda_1, \dots, \lambda_m$ are called the *Lagrange multipliers*.

V.5. Concave and quasiconcave functions

Definition. Let $M \subset \mathbb{R}^n$. We say that M is *convex* if

$$\forall \boldsymbol{x}, \boldsymbol{y} \in M\ \forall t \in [0,1]\colon t\boldsymbol{x} + (1-t)\boldsymbol{y} \in M.$$

Definition. Let $M \subset \mathbb{R}^n$ be a convex set and f a function defined on M. We say that f is

- *concave on M* if
$$\forall \boldsymbol{a}, \boldsymbol{b} \in M\ \forall t \in [0,1]\colon f(t\boldsymbol{a} + (1-t)\boldsymbol{b}) \geq tf(\boldsymbol{a}) + (1-t)f(\boldsymbol{b}),$$
- *strictly concave on M* if
$$\forall \boldsymbol{a}, \boldsymbol{b} \in M, \boldsymbol{a} \neq \boldsymbol{b}\ \forall t \in (0,1)\colon f(t\boldsymbol{a} + (1-t)\boldsymbol{b}) > tf(\boldsymbol{a}) + (1-t)f(\boldsymbol{b}).$$

Remark. By changing the inequalities to the opposite we obtain a definition of a *convex* and a *strictly convex* function.

Remark. A function f is convex (strictly convex) if and only if the function $-f$ is concave (strictly concave).

All the theorems in this section are formulated for concave and strictly concave functions. They have obvious analogies that hold for convex and strictly convex functions.

Remark.

- If a function f is strictly concave on M, then it is concave on M.
- Let f be a concave function on M. Then f is strictly concave on M if and only if the graph of f "does not contain a segment", i.e.
$$\neg\big(\exists \boldsymbol{a}, \boldsymbol{b} \in M, \boldsymbol{a} \neq \boldsymbol{b},\ \forall t \in [0,1]\colon f(t\boldsymbol{a} + (1-t)\boldsymbol{b}) = tf(\boldsymbol{a}) + (1-t)f(\boldsymbol{b})\big).$$

Theorem 26. *Let f be a function concave on an open convex set $G \subset \mathbb{R}^n$. Then f is continuous on G.*

Theorem 27 (characterisation of strictly concave functions of the class $\mathcal{C}^1$)**.** *Let $G \subset \mathbb{R}^n$ be a convex open set and $f \in C^1(G)$. Then the function f is strictly concave on G if and only if*

$$\forall \boldsymbol{x}, \boldsymbol{y} \in G, \boldsymbol{x} \neq \boldsymbol{y} \colon f(\boldsymbol{y}) < f(\boldsymbol{x}) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\boldsymbol{x})(y_i - x_i).$$

Theorem 28 (characterisation of concave functions of the class $\mathcal{C}^1$)**.** *Let $G \subset \mathbb{R}^n$ be a convex open set and $f \in C^1(G)$. Then the function f is concave on G if and only if*

$$\forall \boldsymbol{x}, \boldsymbol{y} \in G \colon f(\boldsymbol{y}) \leq f(\boldsymbol{x}) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\boldsymbol{x})(y_i - x_i).$$

Corollary 29. *Let $G \subset \mathbb{R}^n$ be a convex open set, $f \in C^1(G)$, and let $\boldsymbol{a} \in G$ be a critical point of f (i.e. $\nabla f(\boldsymbol{a}) = \boldsymbol{o}$). If f is concave on G, then $\boldsymbol{a}$ is a maximum point of f on G. If f is strictly concave on G, then $\boldsymbol{a}$ is a strict maximum point of f on G.*

Theorem 30 (level sets of concave functions)**.** *Let f be a function concave on a convex set $M \subset \mathbb{R}^n$. Then for each $\alpha \in \mathbb{R}$ the set $Q_\alpha = \{\boldsymbol{x} \in M;\ f(\boldsymbol{x}) \geq \alpha\}$ is convex.*

Definition. Let $M \subset \mathbb{R}^n$ be a convex set and let f be a function defined on M. We say that f is

- *quasiconcave on M* if
$$\forall \boldsymbol{a}, \boldsymbol{b} \in M\ \forall t \in [0,1] \colon f(t\boldsymbol{a} + (1-t)\boldsymbol{b}) \geq \min\{f(\boldsymbol{a}), f(\boldsymbol{b})\},$$

- *strictly quasiconcave on M* if
$$\forall \boldsymbol{a}, \boldsymbol{b} \in M, \boldsymbol{a} \neq \boldsymbol{b},\ \forall t \in (0,1) \colon f(t\boldsymbol{a} + (1-t)\boldsymbol{b}) > \min\{f(\boldsymbol{a}), f(\boldsymbol{b})\}.$$

Remark. By changing the inequalities to the opposite and changing the minimum to a maximum we obtain a definition of a *quasiconvex* and a *strictly quasiconvex* function.

Remark. A function f is quasiconvex (strictly quasiconvex) if and only if the function $-f$ is quasiconcave (strictly quasiconcave).

All the theorems in this section are formulated for quasiconcave and strictly quasiconcave functions. They have obvious analogies that hold for quasiconvex and strictly quasiconvex functions.

Remark.

- If a function f is strictly quasiconcave on M, then it is quasiconcave on M.
- Let f be a quasiconcave function on M. Then f is strictly quasiconcave on M if and only if the graph of f "does not contain a horizontal segment", i.e.
$$\neg\big(\exists \boldsymbol{a}, \boldsymbol{b} \in M, \boldsymbol{a} \neq \boldsymbol{b},\ \forall t \in [0,1] \colon f(t\boldsymbol{a} + (1-t)\boldsymbol{b}) = f(\boldsymbol{a})\big).$$

Remark. Let $M \subset \mathbb{R}^n$ be a convex set and f a function defined on M.

- If f is concave on M, then f is quasiconcave on M.
- If f is strictly concave on M, then f is strictly quasiconcave on M.

Theorem 31 (characterization of quasiconcave functions using level sets)**.** *Let $M \subset \mathbb{R}^n$ be a convex set and f a function defined on M. Then f is quasiconcave on M if and only if for each $\alpha \in \mathbb{R}$ the set $Q_\alpha = \{\boldsymbol{x} \in M;\ f(\boldsymbol{x}) \geq \alpha\}$ is convex.*

Theorem 32 (a uniqueness of an extremum)**.** *Let f be a strictly quasiconcave function on a convex set $M \subset \mathbb{R}^n$. Then there exists at most one point of maximum of f.*

Corollary. *Let $M \subset \mathbb{R}^n$ be a convex, closed, bounded and nonempty set and f a continuous and strictly quasiconcave function on M. Then f attains its maximum at exactly one point.*

Theorem 33 (sufficient condition for concave and convex functions in $\mathbb{R}^2$)**.** *Let $G \subset \mathbb{R}^2$ be convex and $f \in C^2(G)$.*
If $\frac{\partial^2 f}{\partial x^2} \leq 0$, $\frac{\partial^2 f}{\partial y^2} \leq 0$, and $\frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial y}\right)^2 \geq 0$ hold on G, then f is concave on G.
If $\frac{\partial^2 f}{\partial x^2} \geq 0$, $\frac{\partial^2 f}{\partial y^2} \geq 0$, and $\frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial y}\right)^2 \geq 0$ hold on G, then f is convex on G.

VI. Matrix calculus

VI.1. Basic operations with matrices

Definition. A table of numbers

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix},$$

where $a_{ij} \in \mathbb{R}$, $i = 1, \dots, m$, $j = 1, \dots, n$, is called a *matrix of type* $m \times n$ (shortly, an *m-by-n matrix*). We also write $(a_{ij})_{\substack{i=1..m \\ j=1..n}}$ for short.

An n-by-n matrix is called a *square matrix of order* n.

The set of all m-by-n matrices is denoted by $M(m \times n)$.

Definition. Let

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}.$$

The n-tuple $(a_{i1}, a_{i2}, \dots, a_{in})$, where $i \in \{1, 2, \dots, m\}$, is called the *ith row* of the matrix $\boldsymbol{A}$.

The m-tuple $\begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}$, where $j \in \{1, 2, \dots, n\}$, is called the *jth column* of the matrix $\boldsymbol{A}$.

Definition. We say that two matrices are equal, if they are of the same type and the corresponding elements are equal, i.e. if $\boldsymbol{A} = (a_{ij})_{\substack{i=1..m \\ j=1..n}}$ and $\boldsymbol{B} = (b_{uv})_{\substack{u=1..r \\ v=1..s}}$, then $\boldsymbol{A} = \boldsymbol{B}$ if and only if $m = r$, $n = s$ and $a_{ij} = b_{ij}\ \forall i \in \{1, \dots, m\}, \forall j \in \{1, \dots, n\}$.

Definition. Let $\boldsymbol{A}, \boldsymbol{B} \in M(m \times n)$, $\boldsymbol{A} = (a_{ij})_{\substack{i=1..m \\ j=1..n}}$, $\boldsymbol{B} = (b_{ij})_{\substack{i=1..m \\ j=1..n}}$, $\lambda \in \mathbb{R}$. The *sum of the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$* is the matrix defined by

$$\boldsymbol{A} + \boldsymbol{B} = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \dots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \dots & a_{2n} + b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m1} & \dots & a_{mn} + b_{mn} \end{pmatrix}.$$

The product of the real number λ and the matrix $\boldsymbol{A}$ (or the λ-multiple of the matrix $\boldsymbol{A}$) is the matrix defined by

$$\lambda \boldsymbol{A} = \begin{pmatrix} \lambda a_{11} & \lambda a_{12} & \dots & \lambda a_{1n} \\ \lambda a_{21} & \lambda a_{22} & \dots & \lambda a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda a_{m1} & \lambda a_{m2} & \dots & \lambda a_{mn} \end{pmatrix}.$$

Proposition 34 (basic properties of the sum of matrices and of a multiplication by a scalar)**.** *The following holds:*

- $\forall \boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C} \in M(m \times n)$: $\boldsymbol{A} + (\boldsymbol{B} + \boldsymbol{C}) = (\boldsymbol{A} + \boldsymbol{B}) + \boldsymbol{C}$, *(associativity)*
- $\forall \boldsymbol{A}, \boldsymbol{B} \in M(m \times n)$: $\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{B} + \boldsymbol{A}$, *(commutativity)*
- $\exists! \boldsymbol{O} \in M(m \times n)\ \forall \boldsymbol{A} \in M(m \times n)$: $\boldsymbol{A} + \boldsymbol{O} = \boldsymbol{A}$, *(existence of a zero element)*
- $\forall \boldsymbol{A} \in M(m \times n)\ \exists \boldsymbol{C}_{\boldsymbol{A}} \in M(m \times n)$: $\boldsymbol{A} + \boldsymbol{C}_{\boldsymbol{A}} = \boldsymbol{O}$, *(existence of an opposite element)*
- $\forall \boldsymbol{A} \in M(m \times n)\ \forall \lambda, \mu \in \mathbb{R}$: $(\lambda\mu)\boldsymbol{A} = \lambda(\mu \boldsymbol{A})$,
- $\forall \boldsymbol{A} \in M(m \times n)$: $1 \cdot \boldsymbol{A} = \boldsymbol{A}$,
- $\forall \boldsymbol{A} \in M(m \times n)\ \forall \lambda, \mu \in \mathbb{R}$: $(\lambda + \mu)\boldsymbol{A} = \lambda \boldsymbol{A} + \mu \boldsymbol{A}$,
- $\forall \boldsymbol{A}, \boldsymbol{B} \in M(m \times n)\ \forall \lambda \in \mathbb{R}$: $\lambda(\boldsymbol{A} + \boldsymbol{B}) = \lambda \boldsymbol{A} + \lambda \boldsymbol{B}$.

Remark.

- The matrix $\boldsymbol{O}$ from the previous proposition is called a *zero matrix* and all its elements are all zeros.

- The matrix $\boldsymbol{C_A}$ from the previous proposition is called a *matrix opposite to* $\boldsymbol{A}$. It is determined uniquely, it is denoted by $-\boldsymbol{A}$, and it satisfies $-\boldsymbol{A} = (-a_{ij})_{\substack{i=1..m \\ j=1..n}}$ and $-\boldsymbol{A} = -1 \cdot \boldsymbol{A}$.

Definition. Let $\boldsymbol{A} \in M(m \times n)$, $\boldsymbol{A} = (a_{is})_{\substack{i=1..m \\ s=1..n}}$, $\boldsymbol{B} \in M(n \times k)$, $\boldsymbol{B} = (b_{sj})_{\substack{s=1..n \\ j=1..k}}$. Then the *product of matrices* $\boldsymbol{A}$ and $\boldsymbol{B}$ is defined as a matrix $\boldsymbol{AB} \in M(m \times k)$, $\boldsymbol{AB} = (c_{ij})_{\substack{i=1..m \\ j=1..k}}$, where

$$c_{ij} = \sum_{s=1}^{n} a_{is} b_{sj}.$$

Theorem 35 (properties of the matrix multiplication)**.** *Let* $m, n, k, l \in \mathbb{N}$. *Then:*

(i) $\forall \boldsymbol{A} \in M(m \times n)\ \forall \boldsymbol{B} \in M(n \times k)\ \forall \boldsymbol{C} \in M(k \times l)\colon \boldsymbol{A}(\boldsymbol{BC}) = (\boldsymbol{AB})\boldsymbol{C}$, *(associativity of multiplication)*

(ii) $\forall \boldsymbol{A} \in M(m \times n)\ \forall \boldsymbol{B}, \boldsymbol{C} \in M(n \times k)\colon \boldsymbol{A}(\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{AB} + \boldsymbol{AC}$, *(distributivity from the left)*

(iii) $\forall \boldsymbol{A}, \boldsymbol{B} \in M(m \times n)\ \forall \boldsymbol{C} \in M(n \times k)\colon (\boldsymbol{A} + \boldsymbol{B})\boldsymbol{C} = \boldsymbol{AC} + \boldsymbol{BC}$, *(distributivity from the right)*

(iv) $\exists! \boldsymbol{I} \in M(n \times n)\ \forall \boldsymbol{A} \in M(n \times n)\colon \boldsymbol{IA} = \boldsymbol{AI} = \boldsymbol{A}$. *(existence and uniqueness of an identity matrix* $\boldsymbol{I}$*)*

Remark. Warning! The matrix multiplication is not commutative.

Definition. A *transpose* of a matrix

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{pmatrix}$$

is the matrix

$$\boldsymbol{A}^T = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ a_{13} & a_{23} & \dots & a_{m3} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{pmatrix},$$

i.e. if $\boldsymbol{A} = (a_{ij})_{\substack{i=1..m \\ j=1..n}}$, then $\boldsymbol{A}^T = (b_{uv})_{\substack{u=1..n \\ v=1..m}}$, where $b_{uv} = a_{vu}$ for each $u \in \{1, \dots, n\}$, $v \in \{1, 2, \dots, m\}$.

Theorem 36 (properties of the transpose of a matrix)**.** *Platí:*

(i) $\forall \boldsymbol{A} \in M(m \times n)\colon \left(\boldsymbol{A}^T\right)^T = \boldsymbol{A}$,

(ii) $\forall \boldsymbol{A}, \boldsymbol{B} \in M(m \times n)\colon (\boldsymbol{A} + \boldsymbol{B})^T = \boldsymbol{A}^T + \boldsymbol{B}^T$,

(iii) $\forall \boldsymbol{A} \in M(m \times n)\ \forall \boldsymbol{B} \in M(n \times k)\colon (\boldsymbol{AB})^T = \boldsymbol{B}^T \boldsymbol{A}^T$.

Definition. We say that the matrix $\boldsymbol{A} \in M(n \times n)$ is *symmetric* if $\boldsymbol{A} = \boldsymbol{A}^T$.

VI.2. Invertible matrices

Definition. Let $\boldsymbol{A} \in M(n \times n)$. We say that $\boldsymbol{A}$ is an *invertible* matrix if there exist $\boldsymbol{B} \in M(n \times n)$ such that

$$\boldsymbol{AB} = \boldsymbol{BA} = \boldsymbol{I}.$$

Definition. We say that the matrix $\boldsymbol{B} \in M(n \times n)$ is an *inverse* of a matrix $\boldsymbol{A} \in M(n \times n)$ if $\boldsymbol{AB} = \boldsymbol{BA} = \boldsymbol{I}$.

Remark. A matrix $\boldsymbol{A} \in M(n \times n)$ is invertible if and only if it has an inverse.

Remark.

- If $\boldsymbol{A} \in M(n \times n)$ is invertible, then it has exactly one inverse, which is denoted by $\boldsymbol{A}^{-1}$.
- If some matrices $\boldsymbol{A}, \boldsymbol{B} \in M(n \times n)$ satisfy $\boldsymbol{AB} = \boldsymbol{I}$, then also $\boldsymbol{BA} = \boldsymbol{I}$.

Theorem 37 (operations with invertible matrices)**.** *Let* $\boldsymbol{A}, \boldsymbol{B} \in M(n \times n)$ *be invertible matrices. Then*

(i) $\boldsymbol{A}^{-1}$ *is invertible and* $\left(\boldsymbol{A}^{-1}\right)^{-1} = \boldsymbol{A}$,

(ii) $\boldsymbol{A}^T$ *is invertible and* $\left(\boldsymbol{A}^T\right)^{-1} = \left(\boldsymbol{A}^{-1}\right)^T$,

(iii) $\boldsymbol{AB}$ *is invertible and* $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$.

Definition. Let $k, n \in \mathbb{N}$ and $\boldsymbol{v}^1, \dots, \boldsymbol{v}^k \in \mathbb{R}^n$. We say that a vector $\boldsymbol{u} \in \mathbb{R}^n$ is a *linear combination of the vectors* $\boldsymbol{v}^1, \dots, \boldsymbol{v}^k$ *with coefficients* $\lambda_1, \dots, \lambda_k \in \mathbb{R}$ if

$$\boldsymbol{u} = \lambda_1\boldsymbol{v}^1 + \dots + \lambda_k\boldsymbol{v}^k.$$

By a *trivial linear combination* of vectors $\boldsymbol{v}^1, \dots, \boldsymbol{v}^k$ we mean the linear combination $0 \cdot \boldsymbol{v}^1 + \dots + 0 \cdot \boldsymbol{v}^k$. Linear combination which is not trivial is called *non-trivial*.

Definition. We say that vectors $\boldsymbol{v}^1, \dots, \boldsymbol{v}^k \in \mathbb{R}^n$ are *linearly dependent* if there exists their non-trivial linear combination which is equal to the zero vector. We say that vectors $\boldsymbol{v}^1, \dots, \boldsymbol{v}^k \in \mathbb{R}^n$ are *linearly independent* if they are not linearly dependent, i.e. if whenever $\lambda_1, \dots, \lambda_k \in \mathbb{R}$ satisfy $\lambda_1\boldsymbol{v}^1 + \dots + \lambda_k\boldsymbol{v}^k = \boldsymbol{o}$, then $\lambda_1 = \lambda_2 = \dots = \lambda_k = 0$.

Remark. Vectors $\boldsymbol{v}^1, \dots, \boldsymbol{v}^k$ are linearly dependent if and only if one of them can be expressed as a linear combination of the others.

Definition. Let $\boldsymbol{A} \in M(m \times n)$. The *rank* of the matrix $\boldsymbol{A}$ is the maximal number of linearly independent row vectors of $\boldsymbol{A}$, i.e. the rank is equal to $k \in \mathbb{N}$ if

(i) there is k linearly independent row vectors of $\boldsymbol{A}$ and

(ii) each l-tuple of row vectors of $\boldsymbol{A}$, where $l > k$, is linearly dependent.

The rank of the zero matrix is zero. Rank of $\boldsymbol{A}$ is denoted by $\operatorname{rank}(\boldsymbol{A})$.

Definition. We say that a matrix $\boldsymbol{A} \in M(m \times n)$ is in a *row echelon form* if for each $i \in \{2, \dots, m\}$ the ith row of $\boldsymbol{A}$ is either a zero vector or it has more zeros at the beginning than the $(i-1)$th row.

Remark. The rank of a row echelon matrix is equal to the number of its non-zero rows.

Definition. The *elementary row operations* on the matrix $\boldsymbol{A}$ are:

(i) interchange of two rows,

(ii) multiplication of a row by a non-zero real number,

(iii) addition of a multiple of a row to another row.

Definition. A matrix *transformation* is a finite sequence of elementary row operations. If a matrix $\boldsymbol{B} \in M(m \times n)$ results from the matrix $\boldsymbol{A} \in M(m \times n)$ by applying a transformation T on the matrix $\boldsymbol{A}$, then this fact is denoted by $\boldsymbol{A} \overset{T}{\rightsquigarrow} \boldsymbol{B}$.

Theorem 38 (properties of matrix transformations)**.**

(i) Let $\boldsymbol{A} \in M(m \times n)$*. Then there exists a transformation transforming* $\boldsymbol{A}$ *to a row echelon matrix.*

(ii) Let T_1 *be a transformation applicable to* m*-by-*n *matrices. Then there exists a transformation* T_2 *applicable to* m*-by-*n *matrices such that for any two matrices* $\boldsymbol{A}, \boldsymbol{B} \in M(m \times n)$ *we have* $\boldsymbol{A} \overset{T_1}{\rightsquigarrow} \boldsymbol{B}$ *if and only if* $\boldsymbol{B} \overset{T_2}{\rightsquigarrow} \boldsymbol{A}$.

(iii) Let $\boldsymbol{A}, \boldsymbol{B} \in M(m \times n)$ *and there exist a transformation* T *such that* $\boldsymbol{A} \overset{T}{\rightsquigarrow} \boldsymbol{B}$*. Then* $\operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{B})$.

Remark. Similarly as the elementary row operations one can define also elementary column operations. It can be shown that the elementary column operations do not change the rank of the matrix.

Remark. It can be shown that $\operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{A}^T)$ for any $\boldsymbol{A} \in M(m \times n)$.

Theorem 39 (reprezentation of a transformation)**.** *Let* T *be a transformation on* $m \times n$ *matrices. Then there exists an invertible matrix* $\boldsymbol{C}_T \in M(m \times m)$ *satisfying:*

whenever we apply the transformation T *to a matrix* $\boldsymbol{A} \in M(m \times n)$*, we obtain the matrix* $\boldsymbol{C}_T\boldsymbol{A}$.

Remark. Also the converse is true: For every invertible matrix $\boldsymbol{C}$ the mapping $\boldsymbol{A} \mapsto \boldsymbol{CA}$ is a transformation.

Lemma 40. *Let* $\boldsymbol{A} \in M(n \times n)$ *and* $\operatorname{rank}(\boldsymbol{A}) = n$*. Then there exists a transformation transforming* $\boldsymbol{A}$ *to* $\boldsymbol{I}$.

Theorem 41. *Let* $\boldsymbol{A} \in M(n \times n)$*. Then* $\boldsymbol{A}$ *is invertible if and only if* $\operatorname{rank}(\boldsymbol{A}) = n$.

VI.3. Determinants

Definition. Let $\boldsymbol{A} \in M(n \times n)$. The symbol $\boldsymbol{A}_{ij}$ denotes the $(n-1)$-by-$(n-1)$ matrix which is created from $\boldsymbol{A}$ by omitting the ith row and the jth column.

Definition. Let $\boldsymbol{A} = (a_{ij})_{i,j=1..n}$. The *determinant* of the matrix $\boldsymbol{A}$ is defined by

$$\det \boldsymbol{A} = \begin{cases} a_{11} & \text{if } n = 1, \\ \sum_{i=1}^{n} (-1)^{i+1} a_{i1} \det \boldsymbol{A}_{i1} & \text{if } n > 1. \end{cases}$$

For $\det \boldsymbol{A}$ we will also use the symbol

$$\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \ddots & \vdots & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}.$$

Theorem 42 (cofactor expansion)**.** *Let $\boldsymbol{A} = (a_{ij})_{i,j=1..n}$, $k \in \{1, \dots, n\}$. Then*

$$\det \boldsymbol{A} = \sum_{i=1}^{n} (-1)^{i+k} a_{ik} \det \boldsymbol{A}_{ik} \quad \textit{(expansion along kth column)},$$

$$\det \boldsymbol{A} = \sum_{j=1}^{n} (-1)^{k+j} a_{kj} \det \boldsymbol{A}_{kj} \quad \textit{(expansion along kth row)}.$$

Lemma 43. *Let $j, n \in \mathbb{N}$, $j \leq n$, and the matrices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C} \in M(n \times n)$ coincide at each row except for the jth row. Let the jth row of $\boldsymbol{A}$ be equal to the sum of the jth rows of $\boldsymbol{B}$ and $\boldsymbol{C}$. Then* $\det \boldsymbol{A} = \det \boldsymbol{B} + \det \boldsymbol{C}$.

$$\begin{vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{j-1,1} & \dots & a_{j-1,n} \\ u_1+v_1 & \dots & u_n+v_n \\ a_{j+1,1} & \dots & a_{j+1,n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{j-1,1} & \dots & a_{j-1,n} \\ u_1 & \dots & u_n \\ a_{j+1,1} & \dots & a_{j+1,n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{vmatrix} + \begin{vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{j-1,1} & \dots & a_{j-1,n} \\ v_1 & \dots & v_n \\ a_{j+1,1} & \dots & a_{j+1,n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{vmatrix}$$

Theorem 44 (determinant and transformations)**.** *Let $\boldsymbol{A}, \boldsymbol{A}' \in M(n \times n)$.*

(i) If the matrix $\boldsymbol{A}'$ is created from the matrix $\boldsymbol{A}$ by multiplying one row in $\boldsymbol{A}$ by a real number μ, then $\det \boldsymbol{A}' = \mu \det \boldsymbol{A}$.

(ii) If the matrix $\boldsymbol{A}'$ is created from $\boldsymbol{A}$ by interchanging two rows in $\boldsymbol{A}$ (i.e. by applying the elementary row operation of the first type), then $\det \boldsymbol{A}' = -\det \boldsymbol{A}$.

(iii) If the matrix $\boldsymbol{A}'$ is created from $\boldsymbol{A}$ by adding a μ-multiple of a row in $\boldsymbol{A}$ to another row in $\boldsymbol{A}$ (i.e. by applying the elementary row operation of the third type), then $\det \boldsymbol{A}' = \det \boldsymbol{A}$.

(iv) If $\boldsymbol{A}'$ is created from $\boldsymbol{A}$ by applying a transformation, then $\det \boldsymbol{A} \neq 0$ *if and only if* $\det \boldsymbol{A}' \neq 0$.

Remark. The determinant of a matrix with a zero row is equal to zero. The determinant of a matrix with two identical rows is also equal to zero.

Definition. Let $\boldsymbol{A} = (a_{ij})_{i,j=1..n}$. We say that $\boldsymbol{A}$ is an *upper triangular matrix* if $a_{ij} = 0$ for $i > j$, $i, j \in \{1, \dots, n\}$. We say that $\boldsymbol{A}$ is a *lower triangular matrix* if $a_{ij} = 0$ for $i < j$, $i, j \in \{1, \dots, n\}$.

Theorem 45 (determinant of a triangular matrix)**.** *Let $\boldsymbol{A} = (a_{ij})_{i,j=1..n}$ be an upper or lower triangular matrix. Then*

$$\det \boldsymbol{A} = a_{11} \cdot a_{22} \cdot \dots \cdot a_{nn}.$$

Theorem 46 (determinant and invertibility)**.** *Let $\boldsymbol{A} \in M(n \times n)$. Then $\boldsymbol{A}$ is invertible if and only if* $\det \boldsymbol{A} \neq 0$.

Theorem 47 (determinant of a product)**.** *Let $\boldsymbol{A}, \boldsymbol{B} \in M(n \times n)$. Then* $\det \boldsymbol{AB} = \det \boldsymbol{A} \cdot \det \boldsymbol{B}$.

Theorem 48 (determinant of a transpose)**.** *Let $\boldsymbol{A} \in M(n \times n)$. Then* $\det \boldsymbol{A}^T = \det \boldsymbol{A}$.

VI.4. Systems of linear equations

A system of m *equations in* n *unknowns* $x_1, \dots, x_n$:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2, \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &= b_m, \end{aligned} \tag{S}$$

where $a_{ij} \in \mathbb{R}$, $b_i \in \mathbb{R}$, $i = 1, \dots, m$, $j = 1, \dots, n$. The matrix form is

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},$$

where $\boldsymbol{A} = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \in M(m \times n)$, is called the *coefficient matrix*, $\boldsymbol{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \in M(m \times 1)$ is called the *vector of the right-hand side* and $\boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in M(n \times 1)$ is the *vector of unknowns*.

Definition. The matrix

$$(\boldsymbol{A}|\boldsymbol{b}) = \left(\begin{array}{ccc|c} a_{11} & \dots & a_{1n} & b_1 \\ \vdots & \ddots & \vdots & \vdots \\ a_{m1} & \dots & a_{mn} & b_m \end{array}\right)$$

is called the *augmented matrix of the system* (S).

Proposition 49 (solutions of a transformed system)**.** *Let* $\boldsymbol{A} \in M(m \times n)$*,* $\boldsymbol{b} \in M(m \times 1)$ *and let* T *be a transformation of matrices with* m *rows. Denote* $\boldsymbol{A} \overset{T}{\rightsquigarrow} \boldsymbol{A}'$*,* $\boldsymbol{b} \overset{T}{\rightsquigarrow} \boldsymbol{b}'$*. Then for any* $\boldsymbol{y} \in M(n \times 1)$ *we have* $\boldsymbol{A}\boldsymbol{y} = \boldsymbol{b}$ *if and only if* $\boldsymbol{A}'\boldsymbol{y} = \boldsymbol{b}'$*, i.e. the systems* $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ *and* $\boldsymbol{A}'\boldsymbol{x} = \boldsymbol{b}'$ *have the same set of solutions.*

Theorem 50 (Rouché-Fontené)**.** *The system* (S) *has a solution if and only if its coefficient matrix has the same rank as its augmented matrix.*

Systems of n equations in n variables

Theorem 51 (solvability of an $n \times n$ system)**.** *Let* $\boldsymbol{A} \in M(n \times n)$*. Then the following statements are equivalent:*

(i) *the matrix* $\boldsymbol{A}$ *is invertible,*

(ii) *for each* $\boldsymbol{b} \in M(n \times 1)$ *the system* (S) *has a unique solution,*

(iii) *for each* $\boldsymbol{b} \in M(n \times 1)$ *the system* (S) *has at least one solution,*

(iv) $\det \boldsymbol{A} \neq 0$.

Theorem 52 (Cramer's rule)**.** *Let* $\boldsymbol{A} \in M(n \times n)$ *be an invertible matrix,* $\boldsymbol{b} \in M(n \times 1)$*,* $\boldsymbol{x} \in M(n \times 1)$*, and* $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$*. Then*

$$x_j = \frac{\begin{vmatrix} a_{11} & \dots & a_{1,j-1} & b_1 & a_{1,j+1} & \dots & a_{1n} \\ \vdots & & & \vdots & & & \vdots \\ a_{n1} & \dots & a_{n,j-1} & b_n & a_{n,j+1} & \dots & a_{nn} \end{vmatrix}}{\det \boldsymbol{A}}$$

for $j = 1, \dots, n$.

VI.5. Definiteness of matrices

Definition. We say that a symmetric matrix $\boldsymbol{A} \in M(n \times n)$ is

- *positive definite* (PD), if $\boldsymbol{u}^T\boldsymbol{A}\boldsymbol{u} > 0$ for all $\boldsymbol{u} \in \mathbb{R}^n$, $\boldsymbol{u} \neq \boldsymbol{o}$,
- *negative definite* (ND), if $\boldsymbol{u}^T\boldsymbol{A}\boldsymbol{u} < 0$ for all $\boldsymbol{u} \in \mathbb{R}^n$, $\boldsymbol{u} \neq \boldsymbol{o}$,
- *positive semidefinite* (PSD), if $\boldsymbol{u}^T\boldsymbol{A}\boldsymbol{u} \geq 0$ for all $\boldsymbol{u} \in \mathbb{R}^n$,
- *negative semidefinite* (NSD), if $\boldsymbol{u}^T\boldsymbol{A}\boldsymbol{u} \leq 0$ for all $\boldsymbol{u} \in \mathbb{R}^n$,
- *indefinite* (ID), if there exist $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ such that $\boldsymbol{u}^T\boldsymbol{A}\boldsymbol{u} > 0$ and $\boldsymbol{v}^T\boldsymbol{A}\boldsymbol{v} < 0$.

Proposition 53 (definiteness of diagonal matrices)**.** *Let $\boldsymbol{A} \in M(n \times n)$ be diagonal (i.e. $a_{ij} = 0$ whenever $i \neq j$). Then*

- *$\boldsymbol{A}$ is PD if and only if $a_{ii} > 0$ for all $i = 1, 2, \ldots, n$,*
- *$\boldsymbol{A}$ is ND if and only if $a_{ii} < 0$ for all $i = 1, 2, \ldots, n$,*
- *$\boldsymbol{A}$ is PSD if and only if $a_{ii} \geq 0$ for all $i = 1, 2, \ldots, n$,*
- *$\boldsymbol{A}$ is NSD if and only if $a_{ii} \leq 0$ for all $i = 1, 2, \ldots, n$,*
- *$\boldsymbol{A}$ is ID if and only if there exist i, $j \in \{1, 2, \ldots, n\}$ such that $a_{ii} > 0$ and $a_{jj} < 0$.*

Proposition 54 (necessary conditions for definiteness)**.** *Let $\boldsymbol{A} \in M(n \times n)$ be a symmetric matrix. Then*

- *If $\boldsymbol{A}$ is PD, then $a_{ii} > 0$ for all $i = 1, 2, \ldots, n$,*
- *If $\boldsymbol{A}$ is ND, then $a_{ii} < 0$ for all $i = 1, 2, \ldots, n$,*
- *If $\boldsymbol{A}$ is PSD, then $a_{ii} \geq 0$ for all $i = 1, 2, \ldots, n$,*
- *If $\boldsymbol{A}$ is NSD, then $a_{ii} \leq 0$ for all $i = 1, 2, \ldots, n$,*
- *If there exist i, $j \in \{1, 2, \ldots, n\}$ such that $a_{ii} > 0$ and $a_{jj} < 0$, then $\boldsymbol{A}$ is ID.*

Theorem 55 (Sylvester's criterion)**.** *Let $\boldsymbol{A} = (a_{ij}) \in M(n \times n)$ be a symmetric matrix. Then $\boldsymbol{A}$ is*

- *positive definite if and only if*
$$\begin{vmatrix} a_{11} & \ldots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \ldots & a_{kk} \end{vmatrix} > 0 \quad \textit{for all } k = 1, \ldots, n,$$
- *negative definite if and only if*
$$(-1)^k \begin{vmatrix} a_{11} & \ldots & a_{1k} \\ \vdots & & \vdots \\ a_{k1} & \ldots & a_{kk} \end{vmatrix} > 0 \quad \textit{for all } k = 1, \ldots, n,$$
- *positive semidefinite if and only if*
$$\begin{vmatrix} a_{i_1 i_1} & \ldots & a_{i_1 i_k} \\ \vdots & & \vdots \\ a_{i_k i_1} & \ldots & a_{i_k i_k} \end{vmatrix} \geq 0$$
for each k-tuple of integers $1 \leq i_1 < \cdots < i_k \leq n$, $k = 1, \ldots, n$,
- *negative semidefinite if and only if*
$$(-1)^k \begin{vmatrix} a_{i_1 i_1} & \ldots & a_{i_1 i_k} \\ \vdots & & \vdots \\ a_{i_k i_1} & \ldots & a_{i_k i_k} \end{vmatrix} \geq 0$$
for each k-tuple of integers $1 \leq i_1 < \cdots < i_k \leq n$, $k = 1, \ldots, n$.

Let $f \in C^2(G)$. Then the matrix

$$H_f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \ldots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) & \ldots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \frac{\partial^2 f}{\partial x_n \partial x_2}(x) & \ldots & \frac{\partial^2 f}{\partial x_n^2 x_2}(x) \end{pmatrix}$$

is called *Hessian matrix* of f.

Theorem 56. *Let $G \subset \mathbb{R}^n$ be convex and $f \in C^2(G)$. If the Hessian matrix of f is positive semidefinite for every $x \in G$, then f is convex on G. If the Hessian matrix of f is positive definite for every $x \in G$, then f is strictly convex on G.*

VII. Antiderivatives and Riemann integral

VII.1. Antiderivatives

Definition. Let f be a function defined on an open interval I. We say that a function $F\colon I \to \mathbb{R}$ is an *antiderivative of f on I* if for each $x \in I$ the derivative $F'(x)$ exists and $F'(x) = f(x)$.

Remark. An antiderivative of f is sometimes called a function primitive to f.
If F is an antiderivative of f on I, then F is continuous on I.

Theorem 57 (Uniqueness of an antiderivative)**.** *Let F and G be antiderivatives of f on an open interval I. Then there exists $c \in \mathbb{R}$ such that $F(x) = G(x) + c$ for each $x \in I$.*

Remark. The set of all antiderivatives of f on an open interval I is denoted by

$$\int f(x)\,\mathrm{d}x.$$

The fact that F is an antiderivative of f on I is expressed by

$$\int f(x)\,\mathrm{d}x \overset{c}{=} F(x), \quad x \in I.$$

Table of basic antiderivatives

- $\displaystyle\int x^n\,\mathrm{d}x \overset{c}{=} \frac{x^{n+1}}{n+1}$ on $\mathbb{R}$ for $n \in \mathbb{N} \cup \{0\}$; on $(-\infty, 0)$ and on $(0, \infty)$ for $n \in \mathbb{Z}$, $n < -1$,
- $\displaystyle\int x^\alpha\,\mathrm{d}x \overset{c}{=} \frac{x^{\alpha+1}}{\alpha+1}$ on $(0, +\infty)$ for $\alpha \in \mathbb{R} \setminus \{-1\}$,
- $\displaystyle\int \frac{1}{x}\,\mathrm{d}x \overset{c}{=} \log|x|$ on $(0, +\infty)$ and on $(-\infty, 0)$,
- $\displaystyle\int e^x\,\mathrm{d}x \overset{c}{=} e^x$ on $\mathbb{R}$,
- $\displaystyle\int \sin x\,\mathrm{d}x \overset{c}{=} -\cos x$ on $\mathbb{R}$,
- $\displaystyle\int \cos x\,\mathrm{d}x \overset{c}{=} \sin x$ on $\mathbb{R}$,
- $\displaystyle\int \frac{1}{\cos^2 x}\,\mathrm{d}x \overset{c}{=} \operatorname{tg} x$ on each of the intervals $(-\frac{\pi}{2} + k\pi, \frac{\pi}{2} + k\pi)$, $k \in \mathbb{Z}$,
- $\displaystyle\int \frac{1}{\sin^2 x}\,\mathrm{d}x \overset{c}{=} -\operatorname{cotg} x$ on each of the intervals $(k\pi, \pi + k\pi)$, $k \in \mathbb{Z}$,
- $\displaystyle\int \frac{1}{1+x^2}\,\mathrm{d}x \overset{c}{=} \operatorname{arctg} x$ on $\mathbb{R}$,
- $\displaystyle\int \frac{1}{\sqrt{1-x^2}}\,\mathrm{d}x \overset{c}{=} \arcsin x$ on $(-1, 1)$,
- $\displaystyle\int -\frac{1}{\sqrt{1-x^2}}\,\mathrm{d}x \overset{c}{=} \arccos x$ on $(-1, 1)$.

Theorem 58 (Existence of an antiderivative)**.** *Let f be a continuous function on an open interval I. Then f has an antiderivative on I.*

Theorem 59 (Linearity of antiderivatives)**.** *Suppose that f has an antiderivative F on an open interval I, g has an antiderivative G on I, and let $\alpha, \beta \in \mathbb{R}$. Then the function $\alpha F + \beta G$ is an antiderivative of $\alpha f + \beta g$ on I.*

Theorem 60 (substitution)**.**

(i) Let F be an antiderivative of f on (a, b). Let $\varphi\colon (\alpha, \beta) \to (a, b)$ have a finite derivative at each point of (α, β). Then

$$\int f\bigl(\varphi(x)\bigr)\varphi'(x)\,\mathrm{d}x \overset{c}{=} F\bigl(\varphi(x)\bigr) \quad \text{on } (\alpha, \beta).$$

(ii) Let φ be a function with a finite derivative in each point of (α, β) such that the derivative is either everywhere positive or everywhere negative, and such that $\varphi\big((\alpha,\beta)\big) = (a,b)$. Let f be a function defined on (a,b) and suppose that

$$\int f\big(\varphi(t)\big)\varphi'(t)\,\mathrm{d}t \overset{c}{=} G(t) \quad \text{on } (\alpha,\beta).$$

Then

$$\int f(x)\,\mathrm{d}x \overset{c}{=} G\big(\varphi^{-1}(x)\big) \quad \text{on } (a,b).$$

Theorem 61 (integration by parts). *Let I be an open interval and let the functions f and g be continuous on I. Let F be an antiderivative of f on I and G an antiderivative of g on I. Then*

$$\int f(x)G(x)\,\mathrm{d}x = F(x)G(x) - \int F(x)g(x)\,\mathrm{d}x \quad \text{on } I.$$

Example. Denote $I_n = \displaystyle\int \frac{1}{(1+x^2)^n}\,\mathrm{d}x$, $n \in \mathbb{N}$. Then

$$\begin{aligned} I_{n+1} &= \frac{x}{2n(1+x^2)^n} + \frac{2n-1}{2n} I_n, x \in \mathbb{R}, \quad n \in \mathbb{N}, \\ I_1 &\overset{c}{=} \operatorname{arctg} x, x \in \mathbb{R}. \end{aligned}$$

Definition. A *rational function* is a ratio of two polynomials, where the polynomial in the denominator is not a zero polynomial.

Theorem ("fundamental theorem of algebra"). *Let $n \in \mathbb{N}$, $a_0, \dots, a_n \in \mathbb{C}$, $a_n \neq 0$. Then the equation*

$$a_n z^n + a_{n-1}z^{n-1} + \dots + a_1 z + a_0 = 0$$

has at least one solution $z \in \mathbb{C}$.

Lemma 62 (polynomial division). *Let P and Q be polynomials (with complex coefficients) such that Q is not a zero polynomial. Then there are uniquely determined polynomials S and R satisfying:*

- $\deg R < \deg Q$,
- $P(x) = S(x)Q(x) + R(x)$ *for all* $x \in \mathbb{C}$.

If P and Q have real coefficients then so have S and R.

Corollary. *If P is a polynomials and $\lambda \in \mathbb{C}$ its root (i.e. $P(\lambda) = 0$), then there is a polynomial S satisfying $P(x) = (x-\lambda)S(x)$ for all $x \in \mathbb{C}$.*

Theorem 63 (factorisation into monomials). *Let $P(x) = a_n x^n + \dots + a_1 x + a_0$ be a polynomial of degree $n \in \mathbb{N}$. Then there are numbers $x_1, \dots, x_n \in \mathbb{C}$ such that*

$$P(x) = a_n(x-x_1)\cdots(x-x_n), \quad x \in \mathbb{C}.$$

Definition. Let P be a polynomial that is not zero, $\lambda \in \mathbb{C}$, and $k \in \mathbb{N}$. We say that λ is a *root of multiplicity* k of the polynomial P if there is a polynomial S satisfying $S(\lambda) \neq 0$ and $P(x) = (x-\lambda)^k S(x)$ for all $x \in \mathbb{C}$.

Theorem 64 (roots of a polynomial with real coefficients). *Let P be a polynomial with real coefficients and $\lambda \in \mathbb{C}$ a root of P of multiplicity $k \in \mathbb{N}$. Then the also the conjugate number $\overline{\lambda}$ is a root of P of multiplicity k.*

Theorem 65 (factorisation of a polynomial with real coefficients). *Let $P(x) = a_n x^n + \dots + a_1 x + a_0$ be a polynomial of degree n with real coefficients. Then there exist real numbers $x_1, \dots, x_k$, $\alpha_1, \dots, \alpha_l$, $\beta_1, \dots, \beta_l$ and natural numbers $p_1, \dots, p_k$, $q_1, \dots, q_l$ such that*

- $P(x) = a_n(x-x_1)^{p_1}\cdots(x-x_k)^{p_k}(x^2+\alpha_1 x+\beta_1)^{q_1}\cdots(x^2+\alpha_l x+\beta_l)^{q_l}$,
- *no two polynomials from $x-x_1, x-x_2, \dots, x-x_k$, $x^2+\alpha_1 x+\beta_1, \dots, x^2+\alpha_l x+\beta_l$ have a common root,*
- *the polynomials $x^2+\alpha_1 x+\beta_1, \dots, x^2+\alpha_l x+\beta_l$ have no real root.*

Theorem 66 (decomposition to partial fractions). *Let P, Q be polynomials with real coefficients such that $\deg P < \deg Q$ and let*

$$Q(x) = a_n(x-x_1)^{p_1}\cdots(x-x_k)^{p_k}(x^2+\alpha_1 x+\beta_1)^{q_1}\cdots(x^2+\alpha_l x+\beta_l)^{q_l}$$

be a factorisation of from Theorem 65. Then there exist unique real numbers $A_1^1, \dots, A_{p_1}^1, \dots, A_1^k, \dots, A_{p_k}^k$, $B_1^1, C_1^1, \dots, B_{q_1}^1, C_{q_1}^1, \dots, B_1^l, C_1^l, \dots, B_{q_l}^l, C_{q_l}^l$ such that

$$\begin{aligned} \frac{P(x)}{Q(x)} = \frac{A_1^1}{(x-x_1)} + \dots + \frac{A_{p_1}^1}{(x-x_1)^{p_1}} + \dots + \frac{A_1^k}{(x-x_k)} + \dots + \frac{A_{p_k}^k}{(x-x_k)^{p_k}} + \\ + \frac{B_1^1 x + C_1^1}{(x^2+\alpha_1 x+\beta_1)} + \dots + \frac{B_{q_1}^1 x + C_{q_1}^1}{(x^2+\alpha_1 x+\beta_1)^{q_1}} + \dots + \\ + \frac{B_1^l x + C_1^l}{(x^2+\alpha_l x+\beta_l)} + \dots + \frac{B_{q_l}^l x + C_{q_l}^l}{(x^2+\alpha_l x+\beta_l)^{q_l}}, x \in \mathbb{R} \setminus \{x_1, \dots, x_k\}. \end{aligned}$$

VII.2. Riemann integral

Definition. A finite sequence $\{x_j\}_{j=0}^n$ is called a *partition of the interval* $[a,b]$ if

$$a = x_0 < x_1 < \cdots < x_n = b.$$

The points $x_0, \ldots, x_n$ are called the *partition points*.

We say that a partition D' of an interval $[a,b]$ is a *refinement of the partition* D of $[a,b]$ if each partition point of D is also a partition point of D'.

Definition. Suppose that $a, b \in \mathbb{R}$, $a < b$, the function f is bounded on $[a,b]$, and $D = \{x_j\}_{j=0}^n$ is a partition of $[a,b]$. Denote

$$\overline{S}(f,D) = \sum_{j=1}^n M_j(x_j - x_{j-1}), \text{ where } M_j = \sup\{f(x);\ x \in [x_{j-1}, x_j]\},$$

$$\underline{S}(f,D) = \sum_{j=1}^n m_j(x_j - x_{j-1}), \text{ where } m_j = \inf\{f(x);\ x \in [x_{j-1}, x_j]\},$$

$$\overline{\int_a^b} f = \inf\bigl\{\overline{S}(f,D);\ D \text{ is a partition of } [a,b]\bigr\},$$

$$\underline{\int_a^b} f = \sup\bigl\{\underline{S}(f,D);\ D \text{ is a partition of } [a,b]\bigr\}.$$

Definition. We say that a function f has the *Riemann integral* over the interval $[a,b]$ if $\overline{\int_a^b} f = \underline{\int_a^b} f$. The value of the integral of f over $[a,b]$ is then equal to the common value of $\overline{\int_a^b} f = \underline{\int_a^b} f$. We denote it by $\int_a^b f$. If $a > b$, then we define $\int_a^b f = -\int_b^a f$, and in case that $a = b$ we put $\int_a^b f = 0$.

Remark. Let D, D' be partitions of $[a,b]$, D' refines D, and let f be a bounded function on $[a,b]$. Then

$$\underline{S}(f,D) \le \underline{S}(f,D') \le \overline{S}(f,D') \le \overline{S}(f,D).$$

Suppose that D_1, D_2 are partitions of $[a,b]$ and a partition D' refines both D_1 and D_2. Then

$$\underline{S}(f,D_1) \le \underline{S}(f,D') \le \overline{S}(f,D') \le \overline{S}(f,D_2).$$

It easily follows that $\underline{\int_a^b} f \le \overline{\int_a^b} f$.

Theorem 67. *(i) Suppose that f has the Riemann integral over $[a,b]$ and let $[c,d] \subset [a,b]$. Then f has the Riemann integral also over $[c,d]$.*

(ii) Suppose that $c \in (a,b)$ and f has the Riemann integral over the intervals $[a,c]$ and $[c,b]$. Then f has the Riemann integral over $[a,b]$ and

$$\int_a^b f = \int_a^c f + \int_c^b f. \tag{1}$$

Remark. The formula (1) holds for all $a, b, c \in \mathbb{R}$ if the integral of f exists over the interval $\bigl[\min\{a,b,c\}, \max\{a,b,c\}\bigr]$.

Theorem 68 (linearity of the Riemann integral)**.** *Let f and g be functions with Riemann integral over $[a,b]$ and let $\alpha \in \mathbb{R}$. Then*

(i) the function αf has the Riemann integral over $[a,b]$ and

$$\int_a^b \alpha f = \alpha \int_a^b f,$$

(ii) the function $f+g$ has the Riemann integral over $[a,b]$ and

$$\int_a^b f+g=\int_a^b f+\int_a^b g.$$

Theorem 69. *Let $a,b\in\mathbb{R}$, $a<b$, and let f and g be functions with Riemann integral over $[a,b]$. Then:*

(i) If $f(x)\le g(x)$ for each $x\in[a,b]$, then

$$\int_a^b f\le\int_a^b g.$$

(ii) The function $|f|$ has the Riemann integral over $[a,b]$ and

$$\left|\int_a^b f\right|\le\int_a^b |f|.$$

Theorem 70. *Let f be a function continuous on an interval $[a,b]$, $a,b\in\mathbb{R}$. Then f has the Riemann integral on $[a,b]$.*

Theorem 71. *Let f be a function continuous on an interval (a,b) and let $c\in(a,b)$. If we denote $F(x)=\int_c^x f(t)\,\mathrm{d}t$ for $x\in(a,b)$, then $F'(x)=f(x)$ for each $x\in(a,b)$. In other words, F is an antiderivative of f on (a,b).*

Theorem 72 (Newton-Leibniz formula)**.** *Let f be a function continuous on an interval $(a-\varepsilon,b+\varepsilon)$, $a,b\in\mathbb{R}$, $a<b$, $\varepsilon>0$ and let F be an antiderivative of f on $(a-\varepsilon,b+\varepsilon)$. Then*

$$\int_a^b f(x)\,\mathrm{d}x=F(b)-F(a). \tag{2}$$

Remark. The Newton-Leibniz formula (2) holds even if $b<a$ (if $F'=f$ on $(b-\varepsilon,a+\varepsilon)$). Let us denote

$$[F]_a^b=F(b)-F(a).$$

Theorem 73 (integration by parts)**.** *Suppose that the functions f, g, f' a g' are continuous on an interval $[a,b]$. Then*

$$\int_a^b f'g=[fg]_a^b-\int_a^b fg'.$$

Theorem 74 (substitution)**.** *Let the function f be continuous on an interval $[a,b]$. Suppose that the function φ has a continuous derivative on $[\alpha,\beta]$ and φ maps $[\alpha,\beta]$ into the interval $[a,b]$. Then*

$$\int_\alpha^\beta f\big(\varphi(x)\big)\varphi'(x)\,\mathrm{d}x=\int_{\varphi(\alpha)}^{\varphi(\beta)} f(t)\,\mathrm{d}t.$$