# Optimization Theory - direct approach
# Petr Lachout

- working text for the lecture
"NMSA403 Optimization Theory (Teorie optimalizace)"

Version - October 14, 2018

# Contents

# Used symbols

$\mathrm{int}\,(A)$ ...... interior of $A$
$\mathrm{clo}\,(A)$ ...... closure of $A$
$\partial\,(A)$ ...... boundary of $A$
$\mathsf{Fin}\,(A)$ ...... set of all finite subsets of $A$

# Chapter 1

# Geometry in $\mathbb{R}^n$

## 1.1 Euclidean space $\mathbb{R}^n$

We will consider a finite dimensional Euclidean space $\mathbb{R}^n$ containing column vectors of reals. The space is equipped with scalar product (or, inner product) (cz. skalární součin)

$$\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i. \tag{1.1}$$

We accept this notation to recall general situation based on duality between spaces $\mathcal{X}$ and

$$\mathcal{X}^* = \{f : f : \mathcal{X} \to \mathbb{R} \quad \text{is a continuous linear function}\},$$

where scalar product $\langle x, x^* \rangle$, $x \in \mathcal{X}$, $x^* \in \mathcal{X}^*$ means evaluation of a function at a given point, i.e. $\langle x, x^* \rangle = x^*(x)$.

Duality of a finite dimensional Euclidean space is simple. The space is finite dimensional Hilbert space and Hilbert spaces are self-dual, i.e. dual space can be identified with the original space.

Topology of the space is given by the norm $\|x\| = \sqrt{\langle x, x \rangle}$.

In some places we employ transposition of vectors or matrices, denoted by $x^\top$, $A^\top$.

We distinguish two types of "linear" functions:

- A linear function (cz. lineární funkce) is a function $f : \mathbb{R}^n \to \mathbb{R}$ fulfilling $f(ax + by) = af(x) + bf(y)$ for all $x, y \in \mathbb{R}^n$, $a, b \in \mathbb{R}$.
- An affine function (cz. afinní funkce) is a function $f : \mathbb{R}^n \to \mathbb{R}$ fulfilling $f(ax + (1-a)y) = af(x) + (1-a)f(y)$ for all $x, y \in \mathbb{R}^n$, $a \in \mathbb{R}$.

These two notions are closely related, each linear function is of a form $f(x) = \langle \gamma, x \rangle$ and each affine function can be written in a form $f(x) = \langle \gamma, x \rangle + \alpha$.

## 1.2 Convex sets

We recall definition of convex sets and their basic properties. Properties are stated without proofs; interested reader can find proofs and additional information at [5].

**Definition 1.1** *We call a set $A \subset \mathbb{R}^n$ to be convex (cz. konvexní) if for any couple of points $x, y \in A$ and $0 < \lambda < 1$ we have $\lambda x + (1-\lambda)y \in A$.*

Note, that empty set fulfills the definition and, therefore, is a convex set.

Definition of convex sets can be expressed verbally: "Each convex linear combination of two points from $A$ belongs again to $A$." or geometrically: "Whole segment between two points from $A$ belongs to $A$."

Immediately, we observe convex set contains each convex linear combination of finite number of points of it.

**Lemma 1.2** *Let $A \subset \mathbb{R}^n$ be a convex set. Then, for each $k \in \mathbb{N}$, $a_1, a_2, \ldots, a_k \in A$, $\lambda_1, \lambda_2, \ldots, \lambda_k \in [0,1]$, $\sum_{i=1}^k \lambda_i = 1$ we have $\sum_{i=1}^k \lambda_i a_i \in A$.*

Some of basic operations preserves set convexity. Consider arithmetic operations.

**Lemma 1.3** *If $A, B \subset \mathbb{R}^n$ are convex sets and $\alpha, \beta \in \mathbb{R}$, then*

$$\alpha A + \beta B = \{\alpha a + \beta b \,:\, a \in A,\ b \in B\}$$

*is a convex set. Particularly, $\alpha A$, $A + B$, $A - B$ (! Different to set difference $A \setminus B$!) are convex sets.*

Consider set operations. Intersection preserves convexity.

**Lemma 1.4** *Let $I \neq \emptyset$ be an index set and for each $i \in I$ a convex set $A_i \subset \mathbb{R}^n$ be given. Then $\bigcap_{i \in I} A_i$ is a convex set.*

Unfortunately, results of the other set operations, i.e. union, complement and set difference, are typically nonconvex.

Consider topological operations.

**Lemma 1.5** *If $A \subset \mathbb{R}^n$ is a convex set, then $\operatorname{clo}(A)$, $\operatorname{int}(A)$, $\operatorname{rint}(A)$ are convex sets.*

Unfortunately, boundary $\partial(A)$ is typically nonconvex.

### 1.2.1 Particular convex sets

Remember other helpful notions.

**Definition 1.6** *If $A \subset \mathbb{R}^n$, then symbol $\operatorname{conv}(A)$ denotes the smallest convex set containing set $A$. Set $\operatorname{conv}(A)$ is called convex hull of $A$ (cz. konvexní obal).*

Convex hull can be constructed.

**Lemma 1.7** *If $A \subset \mathbb{R}^n$ then*

$$\operatorname{conv}(A) = \left\{ \sum_{s \in I} \lambda(s) s \,:\, \lambda(s) \geq 0 \ \forall\, s \in I\,,\ \sum_{s \in I} \lambda(s) = 1\,,\ I \in \mathsf{Fin}(A) \right\}.$$

In $n$-dimensional Euclidean space convex linear combinations of at most $n+1$ points from the set are sufficient.

**Theorem 1.8** (Caratheodory): *Pro $A \subset \mathbb{R}^n$ we have*

$$\operatorname{conv}(A) = \left\{ \sum_{s \in I} \lambda(s) s \,:\, \lambda(s) \geq 0 \ \forall\, s \in I\,,\ \sum_{s \in I} \lambda(s) = 1\,,\ I \subset A\,,\ \operatorname{card}(I) \leq n+1 \right\}.$$

Important role is played by polyhedral sets.

**Definition 1.9** *A set $A \subset \mathbb{R}^n$ is called*

*i)* *a convex polyhedral set (cz. konvexní polyedrická množina) if there is a finite number of closed half-spaces $H_1, H_2, \ldots, H_k$ such that $A = \bigcap_{i=1}^k H_i$.*

*ii)* *convex polyhedron (cz. konvexní polyedr), if there is a finite set $S \subset \mathbb{R}^n$ such that $A = \mathrm{conv}\,(S)$.*

Let us note, each closed half-space can be expressed as $\{x \in \mathbb{R}^n : \langle \gamma, x \rangle \geq b\}$ or $\{x \in \mathbb{R}^n : \langle \gamma, x \rangle \leq b\}$, where $\gamma \in \mathbb{R}^n$, $\gamma \neq \mathbf{0}$, $b \in \mathbb{R}$ are properly chosen.

Let us note, there is a family of locally simplicial sets, e.g. union of two convex polyhedral sets is a locally simplicial set. Such sets can be considered as "nonconvex polyhedral sets".

**Theorem 1.10**: *Convex polyhedron is a compact.*

**Theorem 1.11**: *Let $A \subset \mathbb{R}^n$, then $A$ is a convex polyhedron if and only if $A$ is a bounded convex polyhedral set.*

Recall definition and properties of cones.

**Definition 1.12** *A set $A \subset \mathbb{R}^n$ is called*

*i)* *a cone (with vertex at origin) (cz. kužel), if $\mathbf{0} \in A$ and for each point $s \in A$ and $\alpha > 0$ we have $\alpha s \in A$.*

*ii)* *a cone with vertex at point $p \in \mathbb{R}^n$, whenever $A - p$ is a cone.*

*iii)* *a convex cone (cz. konvexní kužel), whenever $A$ is a cone and a convex set.*

**Definition 1.13** *If $A \subset \mathbb{R}^n$, $\mathrm{pos}\,(A)$ denotes the smallest convex cone containing set $A$.*
*Set $\mathrm{pos}\,(A)$ is called a nonnegative hull of the set $A$ (cz. nezáporný obal).*

Nonnegative hull of a set can be constructed.

**Lemma 1.14** *If $A \subset \mathbb{R}^n$ then*

$$\mathrm{pos}\,(A) \quad = \quad \left\{ \sum_{s \in I} \lambda(s) s \ : \ \lambda(s) \geq 0 \ \forall\, s \in I\,,\ I \subset A \text{ is finite} \right\}.$$

Working in a finite dimension, construction can be simplified.

**Theorem 1.15** (Caratheodory): *If $A \subset \mathbb{R}^n$ then*

$$\mathrm{pos}\,(A) \quad = \quad \left\{ \sum_{s \in I} \lambda(s) s \ : \ \lambda(s) \geq 0 \ \forall\, s \in I\,,\ I \subset A\,,\ \mathrm{card}\,(I) \leq n \right\}.$$

Finally, we have to introduced a polyhedral cone.

**Definition 1.16** *A set $A \subset \mathbb{R}^n$ is called convex polyhedral cone (cz. konvexní polyedrický kužel), if there is a finite set $S \subset \mathbb{R}^n$ such that $A = \mathrm{pos}\,(S)$.*

**Theorem 1.17**: *Let $A \subset \mathbb{R}^n$ contains no line. Then, $A$ is a convex polyhedral cone if and only if $A$ is a cone and a convex polyhedral set.*

### 1.2.2 Directions of convex sets

Directions in which a convex set is unbounded are important for optimization.

**Definition 1.18** *Let $A \subset \mathbb{R}^n$ be a nonempty set. We call $s \in \mathbb{R}^n$ a direction of $A$ (cz. směr), whenever there is a point $a \in A$ such that for each $\alpha > 0$ we have $a + \alpha s \in A$.*

*Set of all directions of $A$ will be denoted by $\mathsf{direct}(A)$.*

*We understand $\mathsf{direct}(\emptyset) = \{0\}$.*

Directions of sets possess following properties.

**Lemma 1.19** *If $A \subset \mathbb{R}^n$ then $\mathsf{direct}(A)$ is a cone.*

**Lemma 1.20** *If $A \subset \mathbb{R}^n$ is a convex set then $\mathsf{direct}(A)$ is a convex cone.*

**Lemma 1.21** *If $A \subset \mathbb{R}^n$ is a closed convex set, then $\mathsf{direct}(A)$ is a closed convex cone.*

**Lemma 1.22** *If $A \subset \mathbb{R}^n$ is a convex set, $s \in \mathsf{direct}(A)$ and $a \in \operatorname{rint}(A)$, then for each $\alpha > 0$ we have $a + \alpha s \in A$.*

**Lemma 1.23** *If $A \subset \mathbb{R}^n$ is a closed convex set, $s \in \mathsf{direct}(A)$ and $a \in A$, then for each $\alpha > 0$ is a $a + \alpha s \in A$.*

Closed convex set can be decomposed.

**Theorem 1.24**: *A closed convex set $A \subset \mathbb{R}^n$ can be decomposed as algebraic sums of sets*

$$A = \mathsf{L}(A) + \mathsf{K}(A) + \mathsf{btt}(A), \tag{1.2}$$

$$A = \mathsf{L}(A) + \mathsf{K}(A) + \operatorname{conv}(\mathsf{btt}(A)), \tag{1.3}$$

*where*

$$\begin{aligned}
\mathsf{L}(A) &= \mathsf{direct}(A) \cap -\mathsf{direct}(A) \quad \text{is a subspace of } \mathbb{R}^n, \\
\mathsf{L}(A)^{\perp} & \quad \text{is a subspace of } \mathbb{R}^n \text{ perpendicular at } \mathsf{L}(A) \text{ and } \mathbb{R}^n = \mathsf{L}(A) \oplus \mathsf{L}(A)^{\perp}, \\
D &= A \cap \mathsf{L}(A)^{\perp} \quad \text{is a projection } A \text{ at } \mathsf{L}(A)^{\perp}, \\
\mathsf{K}(A) &= \mathsf{direct}(D) = \mathsf{direct}(A) \cap \mathsf{L}(A)^{\perp} \quad \text{is a projection } \mathsf{direct}(A) \text{ at } \mathsf{L}(A)^{\perp}, \\
\mathsf{btt}(A) &= \{x \in D : \forall s \in \mathsf{K}(A),\ s \neq 0 \text{ we have } x - s \notin D\}.
\end{aligned}$$

## 1.3 Extreme points and extreme directions

Convex polyhedral sets can be fully characterized by their extreme points and extreme directions.

**Definition 1.25** *Let $A \subset \mathbb{R}^n$ be a nonempty convex set. We call $a \in A$ to be an extreme point of $A$ (cz. krajní bod), whenever there are no $x, y \in A$, $x \neq y$ and no $0 < \lambda < 1$ such that $a = \lambda x + (1 - \lambda) y$.*

*The set of all extreme points of $A$ will be denoted by $\operatorname{ext}(A)$.*

**Definition 1.26** *Let $A \subset \mathbb{R}^n$ be a nonempty convex set. We call $s \in \mathsf{direct}(A)$ to be an extreme direction of $A$ (cz. krajní směr), whenever $s \neq 0$ and there are no $x, y \in \mathsf{direct}(A)$, $x, y \notin \operatorname{pos}(\{s\})$ and no $\lambda > 0$, $\varphi > 0$ such that $s = \lambda x + \varphi y$.*

*We denote*

$$\operatorname{extd}(A) = \{s \in \mathsf{direct}(A) : \|s\| = 1,\ s \text{ is an extreme direction of } A\}.$$

Hence, we have a characterization.

**Lemma 1.27** *Let $A \subset \mathbb{R}^n$ be a nonempty convex set. Then, $A$ is a convex polyhedron if and only if* $\mathrm{ext}\,(A)$ *is a finite set and* $A = \mathrm{conv}\,(\mathrm{ext}\,(A))$.

**Lemma 1.28** *Let $A \subset \mathbb{R}^n$ be a nonempty convex set containing no line. Then, $A$ is a convex polyhedral cone if and only if* $\mathrm{extd}\,(A)$ *is a finite set and* $A = \mathrm{pos}\,(\mathrm{extd}\,(A))$.

**Lemma 1.29** *Let $A \subset \mathbb{R}^n$ be a nonempty convex set containing no line. Then, $A$ is a convex polyhedral set if and only if* $\mathrm{ext}\,(A)$, $\mathrm{extd}\,(A)$ *are finite sets and*
$A = \mathrm{conv}\,(\mathrm{ext}\,(A)) + \mathrm{pos}\,(\mathrm{extd}\,(A))$.

## 1.4 Separation of sets

Developing of mathematical programing theory, description, formulation and justification of methodology convenient for solving of optimization programs need Hahn-Banach theorem, see [3] chapter 2. Actually, we do not need this theorem in full general setting. For our purposes, theorem on separation of sets in finite dimension is sufficient.

Let us begin with a geometrical law for parallelogram.

**Lemma 1.30** *If $x, y \in \mathbb{R}^n$, then we have $\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2$.*

**Proof:** Formula is straightforward since

$$\begin{aligned} \|x+y\|^2 + \|x-y\|^2 &= \|x\|^2 + 2\langle x, y\rangle + \|y\|^2 + \|x\|^2 - 2\langle x, y\rangle + \|y\|^2 \\ &= 2\|x\|^2 + 2\|y\|^2 . \end{aligned}$$

Q.E.D.

Theory is based on theorem on projection of a point on a convex closed set.

**Theorem 1.31** (on projection): *Let $K \subset \mathbb{R}^n$ be a convex closed set, $K \neq \emptyset$ and $x \in \mathbb{R}^n$ be a given point. Then, there is an uniquely determined point $\hat{x} \in K$ such that*
$\|x - \hat{x}\| = \min\{\|x-y\| : y \in K\}$.

*The point $\hat{x} \in K$ is also uniquely determined by a condition*

$$\langle x - \hat{x}, y - \hat{x}\rangle \leq 0 \text{ for each } y \in K. \tag{1.4}$$

*The point $\hat{x}$ is called the projection of point $x$ on set $K$ (cz. projekce na množinu).*

**Proof:** Begin our proof with denotation $\Delta = \inf\{\|x-y\| : y \in K\}$.
Evidently, $0 \leq \Delta < +\infty$, since $K$ is non-empty.

1. Existence

   Take a sequence $y_i \in K$, $i \in \mathbb{N}$ such that $\lim_{i\to+\infty} \|x - y_i\| = \Delta$.

   Using lemma 1.30, we are receiving an equality for each couple of indexes $i, j \in \mathbb{N}$

$$\begin{aligned} \|y_i - y_j\|^2 &= \|(x-y_j) - (x-y_i)\|^2 \\ &= 2\|x-y_i\|^2 + 2\|x-y_j\|^2 - \|2x - (y_i+y_j)\|^2 \\ &= 2\|x-y_i\|^2 + 2\|x-y_j\|^2 - 4\left\|x - \frac{1}{2}(y_i+y_j)\right\|^2 . \end{aligned}$$

We know $\frac{1}{2}(y_i + y_j) \in K$ because $K$ is convex. Hence,

$$\|y_i - y_j\|^2 \leq 2\|x - y_i\|^2 + 2\|x - y_j\|^2 - 4\Delta^2 \longrightarrow 0 \text{ whenever } i, j \to +\infty.$$

We have shown the sequence $y_i$, $i \in \mathbb{N}$ is Cauchy in $\mathbb{R}^n$, therefore, there is $\hat{x} \in \mathbb{R}^n$ such that $y_i \longrightarrow \hat{x}$. Since $K$ is a closed set, we have $\hat{x} \in K$. Moreover, $\|x - \hat{x}\| = \Delta$ because norm is continuous.

2. Uniqueness.
   Let $x^*, x^{**} \in K$ and $\|x - x^*\| = \|x - x^{**}\| = \Delta$.
   Using similar arguments as in the first part of the proof, we are receiving:

$$\|x^* - x^{**}\|^2 = 2\|x - x^*\|^2 + 2\|x - x^{**}\|^2 - 4\left\|x - \frac{1}{2}(x^* + x^{**})\right\|^2 \leq$$

$$\leq 2\|x - x^*\|^2 + 2\|x - x^{**}\|^2 - 4\Delta^2 = 0.$$

   Consequently, $x^* = x^{**}$.

3. Equivalent characterization using condition (1.4).
   (a) Let $\hat{x} \in K$ be with $\|x - \hat{x}\| = \Delta$, and, assume the condition (1.4) is not valid.
   Then, there is $y \in K$ such that $\langle x - \hat{x}, y - \hat{x} \rangle > 0$ (the angle is acute).
   For $\alpha \in \mathbb{R}$, we denote $z(\alpha) = (1 - \alpha)\hat{x} + \alpha y$. Then, a function

$$f(\alpha) = \|x - z(\alpha)\|^2 = \|(1 - \alpha)(x - \hat{x}) + \alpha(x - y)\|^2 =$$
$$= (1 - \alpha)^2\|x - \hat{x}\|^2 + 2\alpha(1 - \alpha)\langle x - \hat{x}, x - y \rangle + \alpha^2\|x - y\|^2$$

   possesses a first derivative

$$f'(\alpha) =$$
$$= -2(1 - \alpha)\|x - \hat{x}\|^2 + 2(1 - 2\alpha)\langle x - \hat{x}, x - y \rangle + 2\alpha\|x - y\|^2.$$

   Plugging $\alpha = 0$, we obtain

$$\begin{aligned} f(0) &= \Delta^2, \\ f'(0) &= -2\|x - \hat{x}\|^2 + 2\langle x - \hat{x}, x - y \rangle \\ &= -2\langle x - \hat{x}, x - \hat{x} \rangle + 2\langle x - \hat{x}, x - y \rangle \\ &= 2\langle x - \hat{x}, \hat{x} - y \rangle = -2\langle x - \hat{x}, y - \hat{x} \rangle < 0. \end{aligned}$$

   Consequently, $f(\alpha) < \Delta^2$ in a right neighborhood of zero. That is a contradiction with the definition of constant $\Delta$.

   (b) Let the condition (1.4) be fulfilled for $\hat{x} \in K$ and all $y \in K$. Then,

$$\|x - y\|^2 = \|(x - \hat{x}) + (\hat{x} - y)\|^2$$
$$= \|x - \hat{x}\|^2 + 2\langle x - \hat{x}, \hat{x} - y \rangle + \|\hat{x} - y\|^2 \geq \|x - \hat{x}\|^2.$$

   Consequently, $\|x - \hat{x}\| = \Delta$.

Q.E.D.

Geometrical meaning of the condition (1.4) is that the angle between $x - \hat{x}$ and $y - \hat{x}$ is not acute for any $y \in K$ (the angle is right or obtuse).

**Theorem 1.32** (separation of a point and a set): *Let $K \subset \mathbb{R}^n$, $K \neq \emptyset$ be a convex set and $x \notin \mathrm{clo}(K)$. Then, there is $\gamma \in \mathbb{R}^n$, $\gamma \neq \mathbf{0}$ such that*

$$\inf \{\langle \gamma, y \rangle \, : \, y \in K\} > \langle \gamma, x \rangle .$$

*For example one can choose $\gamma = \hat{x} - x$, where $\hat{x}$ is the projection of $x$ on the set $\mathrm{clo}(K)$. Hence,*

$$\inf \{\langle \gamma, y \rangle \, : \, y \in K\} = \min \{\langle \gamma, y \rangle \, : \, y \in \mathrm{clo}(K)\} = \langle \gamma, \hat{x} \rangle > \langle \gamma, x \rangle .$$

**Proof:** Accordingly to Theorem 1.31, there is a point $\hat{x} \in \mathrm{clo}(K)$ uniquely determined by property (1.4), i.e.

$$\langle x - \hat{x}, y - \hat{x} \rangle \leq 0 \quad \text{for all} \quad y \in \mathrm{clo}(K) .$$

Set $\gamma = \hat{x} - x$. Then,

$$\begin{aligned} \langle \gamma, y - \hat{x} \rangle &= -\langle x - \hat{x}, y - \hat{x} \rangle \geq 0 \text{ for each } \; y \in K, \\ \langle \gamma, \hat{x} - x \rangle &= \|\gamma\|^2 = \langle \hat{x} - x, \hat{x} - x \rangle > 0 \text{ since } \gamma \neq \mathbf{0}. \end{aligned}$$

Consequently,

$$\langle \gamma, y \rangle \geq \langle \gamma, \hat{x} \rangle > \langle \gamma, x \rangle \text{ for each } \; y \in K.$$

Q.E.D.

Theorem 1.32 is saying there is a closed half-space $H = \{\xi \in \mathbb{R}^n \, : \, \langle \gamma, \xi \rangle \geq c\}$ such that $K \subset H$ and distance of $x$ from $H$ is positive.

**Theorem 1.33** (supporting hyperplane): *Let $K \subset \mathbb{R}^n$, $K \neq \emptyset$ be a convex set and $x \in \partial(K)$. Then, there is $\gamma \in \mathbb{R}^n$, $\gamma \neq \mathbf{0}$ such that $\inf \{\langle \gamma, \xi \rangle \, : \, \xi \in K\} = \langle \gamma, x \rangle$.*

**Proof:** Let $x \in \partial(K)$.

1. There is a sequence of points $y_k \notin \mathrm{clo}(K)$, $k \in \mathbb{N}$ such that $y_k \to x$.
   According to Theorem 1.32, for each $k \in \mathbb{N}$ there is $\gamma_k \in \mathbb{R}^n$, $\gamma_k \neq \mathbf{0}$ such that

   $$\inf \{\langle \gamma_k, \xi \rangle \, : \, \xi \in K\} > \langle \gamma_k, y_k \rangle .$$

   Normalized vectors $\beta_k = \frac{\gamma_k}{\|\gamma_k\|}$, $k \in \mathbb{N}$ belong in a compact sphere
   $S = \{x \in \mathbb{R}^n \, : \, \|x\| = 1\}$.
   Hence, we can select a convergent subsequence $\beta_{k_m} \to \gamma \in S$ letting $m \to +\infty$.
   Then, for each $\xi \in K$ we receive:

   $$\langle \gamma, \xi \rangle = \lim_{m \to +\infty} \langle \beta_{k_m}, \xi \rangle \geq \lim_{m \to +\infty} \langle \beta_{k_m}, y_{k_m} \rangle = \langle \gamma, x \rangle .$$

   We have shown $\inf \{\langle \gamma, \xi \rangle \, : \, \xi \in K\} \geq \langle \gamma, x \rangle$.

2. There is a sequence of points $z_k \in \operatorname{rint}(K)$, $k \in \mathbb{N}$ such that $z_k \to x$.
   Then, $\langle \gamma, z_k \rangle \to \langle \gamma, x \rangle$.

We derived $\inf \{\langle \gamma, \xi \rangle : \xi \in K\} = \langle \gamma, x \rangle$, therefore, required statement is proved.

Q.E.D.

Theorem 1.33 is saying that there is a hyperplane $L = \{\xi \in \mathbb{R}^n : \langle \gamma, \xi \rangle = c\}$ and its closed half-space $H = \{\xi \in \mathbb{R}^n : \langle \gamma, \xi \rangle \geq c\}$ such that $x \in L \cap \operatorname{clo}(K)$ and $K \subset H$. Hyperplane $L$ is called a supporting hyperplane of $K$ at $x$ (cz. opěrná nadrovina) and half-space $H$ is called a supporting half-space of $K$ at $x$ (cz. opěrný poloprostor).

Theorem 1.32 is giving a nice characterization of closed convex sets.

**Theorem 1.34** (characterization of closed convex sets): *Let $K \subset \mathbb{R}^n$, $K \neq \emptyset$. Then, $K$ is a closed convex set if and only if $K = \bigcap_{H \in \mathcal{H}_K} H$, where*

$$\mathcal{H}_K = \{H \subset \mathbb{R}^n \,:\, H \text{ is a closed half-space containing } K\}\,.$$

**Proof:** Closed convex half-spaces are closed convex sets. Hence, each their intersection is a closed convex set. We need to show opposite implication, only.

Assume $K$ is a closed convex set.
Denote $C = \bigcap_{H \in \mathcal{H}_K} H$.

1. Evidently $C \supset K$.
2. Take $y \notin K$.
   Accordingly to Theorem 1.32 there is a vector $\gamma \in \mathbb{R}^n$, $\gamma \neq \mathbf{0}$ such that $\inf \{\langle \gamma, x \rangle : x \in K\} > \langle \gamma, y \rangle$.
   Denoting $\Delta := \inf \{\langle \gamma, x \rangle : x \in K\}$, we obtain a closed half-space $H := \{x \in \mathbb{R}^n : \langle \gamma, x \rangle \geq \Delta\}$ with properties $K \subset H$ and $y \notin H$.
   Consequently, $y \notin C$ and, therefore, $C \subset K$.

Q.E.D.

The characterization can be improved.

**Theorem 1.35** (characterization of closed convex sets 2): *Let $K \subset \mathbb{R}^n$ be a non-empty set. Then, $K$ is a closed convex set if and only if $K = \bigcap_{H \in \mathcal{S}_K} H$, where*

$$\mathcal{S}_K = \{H \subset \mathbb{R}^n \,:\, H \text{ is a supporting half-space of } K\}\,.$$

**Proof:** We have to show implication from left-hand side to right-hand side, only. Assume for that $K$ is a closed convex set and denote $C = \bigcap_{H \in \mathcal{S}_K} H$.

1. Evidently $C \supset K$.
2. Take $y \notin K$.
   Set $\gamma = \hat{y} - y$, where $\hat{y}$ is the projection of $y$ to $K$.
   Accordingly to Theorem 1.32 or Theorem 1.33, we have

$$\min \{\langle \gamma, x \rangle \,:\, x \in K\} = \langle \gamma, \hat{y} \rangle > \langle \gamma, y \rangle\,.$$

   Then, $H := \{x \in \mathbb{R}^n : \langle \gamma, x \rangle \geq \langle \gamma, \hat{y} \rangle\}$ is a supporting half-space of $K$ and $y \notin H$.
   Consequently, $y \notin C$ and, therefore, $C \subset K$.

We have derived $K = C$.

Q.E.D.

Theorems 1.32 and 1.33 solve fully separation of a point and a convex set. Now, we start to consider separation of two sets. Two cases will be distinguished.

**Definition 1.36** *Let $A, B \subset \mathbb{R}^n$ be non-empty sets.*

*i) Sets $A$, $B$ are properly separable (cz. neostře oddělitelné), if there is $\gamma \in \mathbb{R}^n$, $\gamma \neq \mathbf{0}$ such that*

$$\langle \gamma, a \rangle \geq \langle \gamma, b \rangle \text{ for all } a \in A\,,\ b \in B. \tag{1.5}$$

*ii) Sets $A$, $B$ are strictly separable (cz. ostře oddělitelné), if there is $\gamma \in \mathbb{R}^n$, $\gamma \neq \mathbf{0}$ and $c, d \in \mathbb{R}$ such that*

$$\langle \gamma, a \rangle \geq c > d \geq \langle \gamma, b \rangle \text{ for all } a \in A\,,\ b \in B. \tag{1.6}$$

Introduced separations are simply related.

**Lemma 1.37** *If two sets are strictly separable, then they are properly separable.*

**Proof:** Implication is evident.

Q.E.D.

Separation of two sets can be expressed in an equivalent way.

**Lemma 1.38** *Non-empty sets $A, B \subset \mathbb{R}^n$ are properly separable if and only if there is $\gamma \in \mathbb{R}^n$, $\gamma \neq \mathbf{0}$ such that*

$$\inf \{\langle \gamma, a \rangle : a \in A\} \geq \sup \{\langle \gamma, b \rangle : b \in B\}\,. \tag{1.7}$$

**Proof:** Property (1.7) is an equivalent rewriting of (1.5).

Q.E.D.

**Lemma 1.39** *Non-empty sets $A, B \subset \mathbb{R}^n$ are strictly separable if and only if there is $\gamma \in \mathbb{R}^n$, $\gamma \neq \mathbf{0}$ such that*

$$\inf \{\langle \gamma, a \rangle : a \in A\} > \sup \{\langle \gamma, b \rangle : b \in B\}\,. \tag{1.8}$$

**Proof:** Property (1.6) implies (1.8).
If (1.8) is in power, it is sufficient to set

$$c := \inf \{\langle \gamma, a \rangle : a \in A\}\,,\ d := \sup \{\langle \gamma, b \rangle : b \in B\}$$

and we are receiving (1.6).

Q.E.D.

Separation of sets is equivalent with separation of their convex hulls.

**Lemma 1.40** *Let $A, B \subset \mathbb{R}^n$ be non-empty sets. Then, we have:*

- *Sets $A$, $B$ are properly separable if and only if $\operatorname{conv}(A)$, $\operatorname{conv}(B)$ are properly separable.*
- *Sets $A$, $B$ are strictly separable if and only if $\operatorname{conv}(A)$, $\operatorname{conv}(B)$ are strictly separable.*

**Proof:** Equivalences are evident, since, for each $\gamma \in \mathbb{R}^n$ and $C \subset \mathbb{R}^n$ we have

$$\inf\{\langle \gamma, c \rangle \,:\, c \in C\} = \inf\{\langle \gamma, c \rangle \,:\, c \in \operatorname{conv}(C)\}\,.$$

Q.E.D.

In the following case, we are able strictly separate. given sets.

**Theorem 1.41** (strict separation of convex sets): *Let $A \subset \mathbb{R}^n$ be a non-empty closed convex set and $B \subset \mathbb{R}^n$ be a non-empty compact convex set. If $A \cap B = \emptyset$, then $A$, $B$ can be strictly separated.*

**Proof:** Set

$$K = A - B = \{a - b \,:\, a \in A, b \in B\}\,.$$

1. We know, the set $K$ is non-empty and convex.
2. Also, $\mathbf{0} \notin K$, since $A \cap B = \emptyset$.
3. We need to show closedness of $K$.
   Let $y_i \in K$, $i \in \mathbb{N}$ be such that $y_i \to \hat{y} \in \mathbb{R}^n$ if $i \to +\infty$.
   According to definition of $K$, for each $i \in \mathbb{N}$ there are $a_i \in A$, $b_i \in B$ such that $y_i = a_i - b_i$.
   Since $B$ is a compact, we can select a subsequence $b_{i_k}$, $k \in \mathbb{N}$ such that $b_{i_k} \to \hat{b} \in B$ if $k \to +\infty$.
   Then, $a_{i_k} = y_{i_k} + b_{i_k} \to \hat{a} = \hat{y} + \hat{b} \in \mathbb{R}^n$ if $k \to +\infty$.
   Set $A$ is closed, therefore, $\hat{a} \in A$.
   We have found $\hat{y} = \hat{a} - \hat{b} \in K$ and, consequently, $K$ is a closed set.

The set $K$ and the point $\mathbf{0}$ fulfill assumptions of Theorem 1.32. Then, there is $\gamma \in \mathbb{R}^n$, $\gamma \neq \mathbf{0}$ such that $\inf\{\langle \gamma, x \rangle \,:\, x \in K\} > \langle \gamma, \mathbf{0} \rangle = 0$.
According to the definition of $K$, we have

$$\begin{aligned} 0 &< \inf\{\langle \gamma, x \rangle \,:\, x \in K\} = \inf\{\langle \gamma, a - b \rangle \,:\, a \in A\,,\ b \in B\} = \\ &= \inf\{\langle \gamma, a \rangle \,:\, a \in A\} - \sup\{\langle \gamma, b \rangle \,:\, b \in B\}\,. \end{aligned}$$

Hence, $\inf\{\langle \gamma, a \rangle \,:\, a \in A\} > \sup\{\langle \gamma, b \rangle \,:\, b \in B\}$, which means strict separability of sets $A$ and $B$, according to Lemma 1.39.

Q.E.D.

Also, there is a nice case in which given sets can be properly separated.

**Theorem 1.42** (proper separation of convex sets): *Let $A, B \subset \mathbb{R}^n$ be non-empty convex sets. If $\operatorname{rint}(A) \cap \operatorname{rint}(B) = \emptyset$, then $A$, $B$ can be properly separated.*

**Proof:** Set

$$K = A - B = \{a - b \,:\, a \in A, b \in B\}\,.$$

1. Set $K$ is non-empty and convex.
2. We need to show $\mathbf{0} \notin \mathrm{rint}\,(K)$.

   Without any loss of generality we will work in the space $\mathsf{Aff}\,(K)$, where $\mathrm{rint}\,(K) = \mathrm{int}\,(K)$.

   Assume $\mathbf{0} \in \mathrm{int}\,(K)$.

   Then, there is $\bar{x} \in A \cap B$ and $\varepsilon > 0$ such that $\mathcal{U}_\varepsilon\,(\mathbf{0}) \subset K$.

   Sets are convex, therefore, there are points $a^1 \in \mathrm{rint}\,(A)$ and $b^1 \in \mathrm{rint}\,(B)$ such that $\|a^1 - \bar{x}\| < \frac{\varepsilon}{2}$, $\|b^1 - \bar{x}\| < \frac{\varepsilon}{2}$.

   Then, $a^1 - b^1, b^1 - a^1 \in \mathcal{U}_\varepsilon\,(\mathbf{0}) \subset K$.

   Then, there are $a^2 \in A$ and $b^2 \in B$ such that $a^2 - b^2 = b^1 - a^1$.

   After this, the point $\frac{a^1+a^2}{2} = \frac{b^1+b^2}{2} \in \mathrm{rint}\,(A) \cap \mathrm{rint}\,(B)$, which is a contradiction with the assumption that the intersection of relative interiors of considered sets is empty.

Consequently, the set $K$ and the point $\mathbf{0}$ can be properly separated; either, according to Theorem 1.32, if $\mathbf{0} \notin \mathrm{clo}\,(K)$, or, according to Theorem 1.33, if $\mathbf{0} \in \partial\,(K)$.
Thus, there is $\gamma \in \mathbb{R}^n$, $\gamma \neq \mathbf{0}$ such that $\inf\{\langle \gamma, x \rangle \,:\, x \in K\} \geq \langle \gamma, \mathbf{0} \rangle = 0$.
Applying the definition of set $K$, we are receiving

$$\begin{aligned} 0 &\leq \inf\{\langle \gamma, x \rangle \,:\, x \in K\} = \inf\{\langle \gamma, a - b \rangle \,:\, a \in A\,,\; b \in B\} \\ &= \inf\{\langle \gamma, a \rangle \,:\, a \in A\} - \sup\{\langle \gamma, b \rangle \,:\, b \in B\}\,. \end{aligned}$$

Therefore, $\inf\{\langle \gamma, a \rangle \,:\, a \in A\} \geq \sup\{\langle \gamma, b \rangle \,:\, b \in B\}$, which means proper separability of $A$, $B$, according to Lemma 1.38.

Q.E.D.

## 1.5 Results from linear algebra

**Lemma 1.43** *Let $\mathcal{G} \subset \mathbb{R}^n$ be open, $\mathbf{0} \in \mathcal{G}$ and $x \in \mathbb{R}^n$.*
*If $\langle y, x \rangle \geq 0$ for each $y \in \mathcal{G}$ then $x = \mathbf{0}$.*

**Proof:** Since $\mathcal{G} \subset \mathbb{R}^n$ is open and $\mathbf{0} \in \mathcal{G}$, there is $\delta > 0$ such that $-\delta x \in \mathcal{G}$.
Hence, $0 \leq \langle x, -\delta x \rangle = -\delta \langle x, x \rangle = -\delta \|x\|^2$.
Consequently, $x = \mathbf{0}$.

Q.E.D.

**Lemma 1.44** *Let $\mathcal{G} \subset \mathbb{R}^n$ be open, $\mathbf{0} \in \mathcal{G}$ and $x \in \mathbb{R}^n$.*
*If $\langle y, x \rangle \geq 0$ for each $y \geq \mathbf{0}$, $y \in \mathcal{G}$ then $x \geq \mathbf{0}$.*

**Proof:** Define $\xi \in \mathbb{R}^n$ such that for $i \in \{1, 2, \dots, n\}$ we set $\xi_i = \max\{0, -x_i\}$.
Since $\mathcal{G} \subset \mathbb{R}^n$ is open and $\mathbf{0} \in \mathcal{G}$, there is $\delta > 0$ such that $\delta\xi \in \mathcal{G}$.
Moreover, $\delta\xi \geq \mathbf{0}$, therefore,
$0 \leq \langle x, \delta\xi \rangle = -\delta \sum_{x_i<0} x_i^2 = -\delta \|\xi\|^2$.
Thus, $\xi = \mathbf{0}$ and, consequently, $x \geq \mathbf{0}$.

Q.E.D.

**Lemma 1.45** *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix.
If $\langle y, Ay \rangle = 0$ for each $y \geq \mathbf{0}$, $y \in \mathbb{R}^n$ then $A = \mathbf{0}$.*

**Proof:**

1. Take $i \in \{1, 2, \dots, n\}$, then

$$0 = \langle \mathsf{e}_{i:n}, A\mathsf{e}_{i:n} \rangle = A_{i,i}.$$

2. Take $i, j \in \{1, 2, \dots, n\}$, $i \neq j$, then

$$\begin{aligned} 0 &= \langle \mathsf{e}_{i:n} + \mathsf{e}_{j:n}, A(\mathsf{e}_{i:n} + \mathsf{e}_{j:n}) \rangle \\ &= \langle \mathsf{e}_{i:n}, A\mathsf{e}_{i:n} \rangle + \langle \mathsf{e}_{i:n}, A\mathsf{e}_{j:n} \rangle + \langle \mathsf{e}_{j:n}, A\mathsf{e}_{i:n} \rangle + \langle \mathsf{e}_{j:n}, A\mathsf{e}_{j:n} \rangle \\ &= A_{i,i} + A_{i,j} + A_{j,i} + A_{j,j} = 2A_{i,j}, \end{aligned}$$

   because of symmetry.

We have proved $A = \mathbf{0}$.

Q.E.D.

**Theorem 1.46** (Farkas): *Let $A \in \mathbb{R}^{m\times n}$ be a matrix and $b \in \mathbb{R}^m$ be a vector. Then, an equation $Ax = b$ possesses a non-negative solution if and only if for all $u \in \mathbb{R}^m$ fulfilling $A^\top u \geq \mathbf{0}$ we have $\langle b, u \rangle \geq \mathbf{0}$.*

**Proof:**

1. If $x \geq \mathbf{0}$, $Ax = b$, $u \in \mathbb{R}^m$, $A^\top u \geq \mathbf{0}$, then we have

$$\langle b, u \rangle = \langle Ax, u \rangle = \langle x, A^\top u \rangle \geq 0,$$

   since scalar product of non-negative vectors is non-negative.

2. Assume $Ax = b$ possesses no non-negative solution.

   Denote $K := \{Ax : x \geq \mathbf{0}\}$.

   Then, $b \notin K$ and $K$ is a non-empty convex closed cone. Hence, assumptions of Theorem 1.32 are fulfilled and there is $u \in \mathbb{R}^m$, $u \neq \mathbf{0}$ such that
   $\langle u, b \rangle < \inf \{\langle u, y \rangle : y \in K\}$.

   (a) Choosing $x = \mathbf{0}$ we receive $\mathbf{0} = A\mathbf{0} \in K$. Consequently, $\langle u, b \rangle < 0$.

(b) Take $y \in K$ and $\xi > 0$. Then, $\xi y \in K$ and we have

$$\langle u, b \rangle < \langle u, \xi y \rangle = \xi \langle u, y \rangle .$$

Therefore,

$$\langle u, y \rangle > \frac{1}{\xi} \langle u, b \rangle \xrightarrow[\xi \to +\infty]{} 0.$$

We have found $\langle u, y \rangle \geq 0$.
Hence, for all $x \geq \mathbf{0}$ we have $\langle u, Ax \rangle = \left\langle A^\top u, x \right\rangle \geq 0$.
According to Lemma 1.44, $A^\top u \geq \mathbf{0}$.

We have found $u$ such that $\langle b, u \rangle < 0$ and $A^\top u \geq \mathbf{0}$.

Therefore, right-hand side is also violated.

Q.E.D.

Later another theorem based on set separation will be employed.

**Theorem 1.47** (Gordan): *Let $A \in \mathbb{R}^{m \times n}$. Then, just one of the following problems possesses a solution:*

1. *There is $p \in \mathbb{R}^m$ such that $p^\top A < \mathbf{0}$.*
2. *There is $x \in \mathbb{R}^n$, $x \geq \mathbf{0}$, $x \neq \mathbf{0}$ such that $Ax = \mathbf{0}$.*

**Proof:**

1. Let $p \in \mathbb{R}^m$ be such that $p^\top A < \mathbf{0}$.

   Then for each $x \in \mathbb{R}^n$, $x \geq \mathbf{0}$, $x \neq \mathbf{0}$, we have $p^\top Ax < 0$.

   Consequently, $Ax \neq \mathbf{0}$ and the second problem possesses no solution.

2. Let there is no $p \in \mathbb{R}^m$ such that $p^\top A < \mathbf{0}$.

   Consider two sets

   $$\mathcal{M} = \left\{ A^\top p \,:\, p \in \mathbb{R}^m \right\}, \; \mathcal{N} = \left\{ z \in \mathbb{R}^n \,:\, z < \mathbf{0} \right\}.$$

   Sets $\mathcal{M}$, $\mathcal{N}$ are non-empty convex and $\mathcal{M} \cap \mathcal{N} = \emptyset$.

   Accordingly to Theorem 1.42, they can be properly separated. Then there is a $q \in \mathbb{R}^n$, $q \neq \mathbf{0}$ such that $\forall\, u \in \mathcal{M}$, $\forall\, v \in \mathcal{N}$ we have $\langle q, u \rangle \geq \langle q, v \rangle$.

   (a) Consider a sequence $v_k = -\frac{1}{k}\mathbf{1}$, $k \in \mathbb{N}$. Hence, $v_k \in \mathcal{N}$, $k \in \mathbb{N}$ and $\forall\, u \in \mathcal{M}$, $\forall\, k \in \mathbb{N}$ we have

   $$\langle q, u \rangle \geq \langle q, v_k \rangle = -\frac{1}{k} \sum_{j=1}^{n} q_j .$$

   Letting $k$ go to $+\infty$, $\forall\, u \in \mathcal{M}$ we have $\langle q, u \rangle \geq 0$.

(b) Consider for index $i \in \{1, 2, \dots, n\}$ a sequence $w_k = -\mathbb{1} - k\mathsf{e}_{i::n}$, $k \in \mathbb{N}$. Hence, $w_k \in \mathcal{N}$, $k \in \mathbb{N}$ and for a fixed $u \in \mathcal{M}$, $\forall\, k \in \mathbb{N}$ we have

$$\langle\, q, u \,\rangle \geq \langle\, q, w_k \,\rangle = -\sum_{j=1}^{n} q_j - kq_i.$$

Consequently, $\forall\, k \in \mathbb{N}$ we have

$$q_i \geq -\frac{1}{k}\left(\sum_{j=1}^{n} q_j + \langle\, q, u \,\rangle\right).$$

Letting $k$ go to $+\infty$, we receive $q \geq \mathbf{0}$.

(c) For every $p \in \mathbb{R}^m$, we have $A^\top p \in \mathcal{M}$.
Consequently, $\left\langle\, q, A^\top p \,\right\rangle \geq 0$.
Therefore for each $p \in \mathbb{R}^m$, we have $\left\langle\, q, A^\top p \,\right\rangle = \langle\, Aq, p \,\rangle \geq 0$.
According to Lemma 1.43, $Aq = \mathbf{0}$.

We have found $q \in \mathbb{R}^n$ a solution of the second problem.

Q.E.D.

# Chapter 2

# Functions

## 2.1 General notions

We consider functions defined on a finite dimensional Euclidean space with values in an extended real line, i.e. real values extended with $+\infty$ and $-\infty$. Extended real line is denoted by $\mathbb{R}^*$.

**Definition 2.1** *For a function $f : \mathbb{R}^n \to \mathbb{R}^*$, we define its epigraph (cz. epigraf) and hypograph (cz. hypograf)*

$$\mathsf{epi}\,(f) = \left\{ \begin{pmatrix} x \\ \eta \end{pmatrix} : f(x) \leq \eta,\ x \in \mathbb{R}^n,\ \eta \in \mathbb{R} \right\}, \tag{2.1}$$

$$\mathsf{hypo}\,(f) = \left\{ \begin{pmatrix} x \\ \eta \end{pmatrix} : f(x) \geq \eta,\ x \in \mathbb{R}^n,\ \eta \in \mathbb{R} \right\} \tag{2.2}$$

*and its domain (cz. doména)*

$$\mathsf{Dom}\,(f) = \{x : f(x) < +\infty,\ x \in \mathbb{R}^n\}. \tag{2.3}$$

**Definition 2.2** *We say, function $f : \mathbb{R}^n \to \mathbb{R}^*$ is proper (cz. vlastní), if $\mathsf{Dom}\,(f) \neq \emptyset$ and $f(x) > -\infty$ for all $x \in \mathbb{R}^n$.*

Acceptance of value $+\infty$ is important for optimization, particularly for its theory. It allows more simple and readable description of an optimization program.

For example optimization program $\inf\{f(x) : x \in D\}$ can be rewritten as an unconstrained problem $\inf\left\{\tilde{f}(x) : x \in \mathbb{R}^n\right\}$, where

$$\tilde{f}(x) = f(x) \quad \text{if} \quad x \in D, \tag{2.4}$$

$$= +\infty \quad \text{otherwise.} \tag{2.5}$$

Epigraph of a function is a particular set.

**Lemma 2.3** *Set $E \subset \mathbb{R}^{n+1}$ is an epigraph of a function $f : \mathbb{R}^n \to \mathbb{R}^*$ if and only if for all $x \in \mathbb{R}^n$ we have*

$$\left\{\eta : \begin{pmatrix} x \\ \eta \end{pmatrix} \in E\right\} \quad \text{is either} \quad \emptyset \quad \text{or} \quad \mathbb{R} \quad \text{or} \quad [\hat{\eta}, +\infty) \text{ for a proper } \hat{\eta} \in \mathbb{R}.$$

*If $E$ is an epigraph of a function $f : \mathbb{R}^n \to \mathbb{R}^*$, then $f(x) = \min\left\{\eta : \begin{pmatrix} x \\ \eta \end{pmatrix} \in E\right\}$.*

**Proof:** Property is evident.

Q.E.D.

Mapping between a function and its epigraph is a 1-1 bijection.

**Lemma 2.4** *Let $I$ be an index set and for each $i \in I$ a function $f_i : \mathbb{R}^n \to \mathbb{R}^*$ be given. Then,*

$$\mathsf{epi}\left(\sup_{i\in I} f_i\right) = \bigcap_{i\in I} \mathsf{epi}\,(f_i)\,,\ \ \mathsf{hypo}\left(\inf_{i\in I} f_i\right) = \bigcap_{i\in I} \mathsf{hypo}\,(f_i)\,, \tag{2.6}$$

$$\mathsf{epi}\left(\inf_{i\in I} f_i\right) \supset \bigcup_{i\in I} \mathsf{epi}\,(f_i)\,,\ \ \mathsf{hypo}\left(\sup_{i\in I} f_i\right) \supset \bigcup_{i\in I} \mathsf{hypo}\,(f_i)\,. \tag{2.7}$$

*If $I$ is finite, we receive equalities*

$$\mathsf{epi}\left(\inf_{i\in I} f_i\right) = \bigcup_{i\in I} \mathsf{epi}\,(f_i)\,,\ \ \mathsf{hypo}\left(\sup_{i\in I} f_i\right) = \bigcup_{i\in I} \mathsf{hypo}\,(f_i)\,. \tag{2.8}$$

**Proof:** Statement is a direct consequence of Lemma 2.3. Intersection or union of a finite number of intervals of type $[\xi, +\infty)$ is giving again an interval of the same type. Union of infinite number of such intervals can violate this property. Similarly, Intersection or union of intervals of type $(-\infty, \xi]$ is again a interval of the same type. Union of infinite number of such intervals can violate this property.

Q.E.D.

## 2.2 Differentiability of a function

### 2.2.1 On the real line

**Definition 2.5** *Let $\mathsf{D} \subset \mathbb{R}$, $\mathsf{D} \neq \emptyset$, $f : \mathsf{D} \to \mathbb{R}$ and $x \in \mathrm{int}\,(\mathsf{D})$. We say, $f$ is differentiable at $x$ (cz. diferencovatelná v bodě $x$) if there is $f'(x) \in \mathbb{R}$ such that for all $y \in \mathsf{D}$ we have*

$$f(y) = f(x) + f'(x)(y-x) + |y-x|\, R_1(y-x; f, x)\,, \tag{2.9}$$

*where $\lim_{h\to\mathbf{0}} R_1(h; f, x) = 0$.*

*Equivalently, $f$ is differentiable at $x$ if and only if $\lim_{h\to 0} \frac{f(x+h)-f(x)}{h} = f'(x) \in \mathbb{R}$.*

*If $\mathsf{S} \subset \mathrm{int}\,(\mathsf{D})$, then we say $f$ is differentiable at $\mathsf{S}$ (cz. diferencovatelná v množině $\mathsf{S}$), if it is differentiable at each point $x \in \mathsf{S}$.*

**Lemma 2.6** *If $\mathsf{D} \subset \mathbb{R}$, $\mathsf{D} \neq \emptyset$ and $f : \mathsf{D} \to \mathbb{R}$ is differentiable at $x \in \mathrm{int}\,(\mathsf{D})$ then $f$ is continuous at $x$.*

**Proof:** Continuity of $f$ at $x$ follows immediately (2.9).

Q.E.D.

**Lemma 2.7** *Let $a, b \in \mathbb{R}$, $a < b$, $f : [a, b] \to \mathbb{R}$ be differentiable at $(a, b)$, right-continuous at $a$ and left-continuous at $b$. Then,*

$$\int_a^b f'(s)\, \mathsf{d}s \ = \ f(b) - f(a)\,. \tag{2.10}$$

### 2.2.2 Several arguments

**Definition 2.8** *Let* $\mathsf{D} \subset \mathbb{R}^n$, $\mathsf{D} \neq \emptyset$, $f : \mathsf{D} \to \mathbb{R}$, $x \in \operatorname{int}(\mathsf{D})$ *and* $h \in \mathbb{R}^n$. *We say,* $f$ *is differentiable at* $x$ *in direction* $h$ *(cz. diferencovatelná v bodě* $x$ *ve směru* $h$*) if there is* $f'(x;h) \in \mathbb{R}$ *such that for all* $t \in \mathbb{R}$, $x + th \in \mathsf{D}$ *we have*

$$f(x+th) = f(x) + f'(x;h)\,t + |t|\, R_1(t; f, x, h), \tag{2.11}$$

*where* $\lim_{s \to 0} R_1(s; f, x, h) = 0$.

*Equivalently,* $f$ *is differentiable at* $x$ *in direction* $h$ *if and only if* $\lim_{t \to 0} \frac{f(x+th)-f(x)}{t} = f'(x;h) \in \mathbb{R}$.

**Definition 2.9** *Let* $\mathsf{D} \subset \mathbb{R}^n$, $\mathsf{D} \neq \emptyset$, $f : \mathsf{D} \to \mathbb{R}$, $x \in \operatorname{int}(\mathsf{D})$. *For* $i \in \{1, 2, \ldots, n\}$ *we say,* $f$ *possesses a partial derivative at* $x$ *w.r.t.* $x_i$ *(cz. parciální derivace v bodě* $x$ *vzhledem k* $x_i$*) if* $f$ *is differentiable at* $x$ *in direction* $\mathsf{e}_{i:n}$ *and we set*

$$\frac{\partial f}{\partial x_i}(x) = f'(x; \mathsf{e}_{i:n}).$$

*If* $f$ *possesses a partial derivative at* $x$ *w.r.t.* $x_i$ *for all* $i \in \{1, 2, \ldots, n\}$ *we say* $f$ *possesses a gradient at* $x$ *(cz. gradient v bodě* $x$*) and we denote*

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_i}(x)\right)_{i=1}^{n}.$$

In the text, we are using differentiability of a function convenient for optimization, see e.g. [1], [6]. We will introduce necessary terminology and basic properties of differentiable functions.

**Definition 2.10** *Let* $\mathsf{D} \subset \mathbb{R}^n$, $\mathsf{D} \neq \emptyset$, $f : \mathsf{D} \to \mathbb{R}$ *and* $x \in \operatorname{int}(\mathsf{D})$. *We say,* $f$ *is differentiable at* $x$ *(or, possesses total differential at* $x$*, Fréchet differentiable at* $x$*) (cz. diferencovatelná v bodě* $x$*) if* $f$ *possesses a gradient* $\nabla f(x) \in \mathbb{R}^n$ *and for all* $y \in \mathsf{D}$ *we have*

$$f(y) = f(x) + \langle \nabla f(x), y - x \rangle + \|y - x\|\, R_1(y - x; f, x), \tag{2.12}$$

*where* $\lim_{h \to \mathbf{0}} R_1(h; f, x) = 0$.

*If* $\mathsf{S} \subset \operatorname{int}(\mathsf{D})$, *then we say* $f$ *is differentiable at* $\mathsf{S}$ *(cz. diferencovatelná v množině* $\mathsf{S}$*), if it is differentiable at each point* $x \in \mathsf{S}$.

**Definition 2.11** *Let* $\mathsf{D} \subset \mathbb{R}^n$, $\mathsf{D} \neq \emptyset$, $f : \mathsf{D} \to \mathbb{R}$ *and* $x \in \operatorname{int}(\mathsf{D})$.

*We say,* $f$ *is continuously differentiable at* $x$ *(cz. spojitě diferencovatelná v bodě* $x$*), if there is* $\delta > 0$ *such that* $\mathcal{U}(x, \delta) \subset \mathsf{D}$, $f$ *is differentiable at* $\mathcal{U}(x, \delta)$ *and the gradient* $\nabla f$ *is continuous at* $x$.

*We say,* $f$ *is continuously differentiable at a neighborhood of* $x$ *(cz. spojitě diferencovatelná v okolí bodu* $x$*), if there is* $\delta > 0$ *such that* $\mathcal{U}(x, \delta) \subset \mathsf{D}$, $f$ *is differentiable at* $\mathcal{U}(x, \delta)$ *and the gradient* $\nabla f$ *is continuous at* $\mathcal{U}(x, \delta)$.

Gradient is necessary for expansion (2.12).

**Lemma 2.12** *Let* $\mathsf{D} \subset \mathbb{R}^n$, $\mathsf{D} \neq \emptyset$, $f : \mathsf{D} \to \mathbb{R}$ *and* $x \in \operatorname{int}(\mathsf{D})$. *Let* $f$ *fulfills an expansion for all* $y \in \mathsf{D}$

$$f(y) = f(x) + \langle \xi, y - x \rangle + \|y - x\|\, R_1(y - x; f, x), \tag{2.13}$$

*where* $\xi \in \mathbb{R}^n$ *and* $\lim_{h \to \mathbf{0}} R_1(h; f, x) = 0$.

*Then* $f$ *is differentiable at* $x$, $\xi = \nabla f(x)$ *and* $f'(x;h) = \langle \nabla f(x), h \rangle$ *for all directions* $h \in \mathbb{R}^n$.

**Proof:** Using (2.13) for a direction $h \in \mathbb{R}^n$ and $t \in \mathbb{R}$ small enough, we have

$$f(x+th) = f(x) + \langle \xi, th \rangle + \|th\| R_1(th; f, x),$$

where $\lim_{h \to \mathbf{0}} R_1(h; f, x) = 0$ .
Consider derivative ratio and let $t \to 0$:

$$\lim_{t \to 0} \frac{f(x+th) - f(x)}{t} = \langle \xi, h \rangle + \|h\| \lim_{t \to 0} \frac{|t|}{t} R_1(th; f, x) = \langle \xi, h \rangle .$$

Setting $h = \mathsf{e}_{i:n}$, we receive $\xi_i = \frac{\partial f}{\partial x_i}(x)$.
We have verified $\xi$ is the gradient of $f$ at $x$, $f$ is differentiable at $x$ and directional derivatives possess announced form.

Q.E.D.

**Lemma 2.13** *If $\mathsf{D} \subset \mathbb{R}^n$, $\mathsf{D} \neq \emptyset$ and $f : \mathsf{D} \to \mathbb{R}$ is differentiable at $x \in \operatorname{int}(\mathsf{D})$ then $f$ is continuous at $x$.*

**Proof:** Continuity of $f$ at $x$ follows immediately (2.12).

Q.E.D.

**Lemma 2.14** *Let $\mathsf{D} \subset \mathbb{R}^n$, $\mathsf{D} \neq \emptyset$ and $f : \mathsf{D} \to \mathbb{R}$. Consider $x \in \mathsf{D}$ and $h \in \mathbb{R}^n$ such that $x + th \in \mathsf{D}$ for all $0 \leq t \leq 1$. Define function $\varphi : [0,1] \to \mathbb{R} : t \in [0,1] \to f(x+th)$.*

- ***(i)*** *If $0 < t < 1$, $x + th \in \operatorname{int}(\mathsf{D})$ and $f$ is differentiable at $x + th$ then $\varphi$ is differentiable at $t$ and $\varphi'(t) = \langle \nabla f(x+th), h \rangle$.*
- ***(ii)*** *If $x + th \in \operatorname{int}(\mathsf{D})$ and $f$ is differentiable at $x + th$ for all $0 < t < 1$, $\varphi$ is continuous at $0$ from right and $\varphi$ is continuous at $1$ from left then*

$$f(x+h) - f(x) = \varphi(1) - \varphi(0) = \int_0^1 \langle \nabla f(x+th), h \rangle \, \mathsf{d}t. \tag{2.14}$$

**Proof:** (i) follows Lemma 2.12 and (ii) is a consequence of Lemma 2.7.

Q.E.D.

### 2.2.3 Vector valued functions

Start with a curve.

**Definition 2.15** *Let $\mathsf{D} \subset \mathbb{R}$, $\mathsf{D} \neq \emptyset$, $f : \mathsf{D} \to \mathbb{R}^m$ and $t \in \operatorname{int}(\mathsf{D})$. Express the function as a vector of functions $f = (f_1, f_2, \ldots, f_m)^\top$. We say,*

- *$f$ is differentiable at $x$ if $f_j$ is differentiable at $x$ for each $j \in \{1, 2, \ldots, m\}$. We denote the derivative by $f'(t) = (f_1'(t), f_2'(t), \ldots, f_m'(t))^\top$.*
- *If $\mathsf{S} \subset \operatorname{int}(\mathsf{D})$, $f$ is differentiable at $\mathsf{S}$ if $f_j$ is differentiable at $\mathsf{S}$ for each $j \in \{1, 2, \ldots, m\}$.*

And now a general case. We start with a notion of multidimensional scalar product.

**Definition 2.16** *Let $n, m \in \mathbb{N}$, $x \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n\times m}$. We define denotation*

$$\lessdot A, x \gtrdot \quad = \quad \left(\langle A_{\cdot,1}, x\rangle, \langle A_{\cdot,2}, x\rangle, \ldots, \langle A_{\cdot,m}, x\rangle\right)^\top .$$

*Using matrix notation, we can write $\lessdot A, x\gtrdot = A^\top x$.*

**Definition 2.17** *Let $\mathsf{D} \subset \mathbb{R}^n$, $\mathsf{D} \neq \emptyset$, $n \geq 2$, $f : \mathsf{D} \to \mathbb{R}^m$ and $x \in \operatorname{int}(\mathsf{D})$. Express the function as a vector of functions $f = (f_1, f_2, \ldots, f_m)^\top$. We say,*

- *$f$ possesses a gradient at $x$ if $f_j$ possesses a gradient at $x$ for each $j \in \{1, 2, \ldots, m\}$. We denote $\nabla f(x) = (\nabla f_1(x), \nabla f_2(x), \ldots, \nabla f_m(x))$.*
- *$f$ is differentiable at $x$ if $f_j$ is differentiable at $x$ for each $j \in \{1, 2, \ldots, m\}$.*
- *If $\mathsf{S} \subset \operatorname{int}(\mathsf{D})$, $f$ is differentiable at $\mathsf{S}$ if $f_j$ is differentiable at $\mathsf{S}$ for each $j \in \{1, 2, \ldots, m\}$.*

**Lemma 2.18** *Let $\mathsf{D} \subset \mathbb{R}^n$, $\mathsf{D} \neq \emptyset$, $f : \mathsf{D} \to \mathbb{R}^m$ and $x \in \operatorname{int}(\mathsf{D})$. Then, $f$ is differentiable at $x$ if and only if $f$ possesses a gradient $\nabla f(x) \in \mathbb{R}^{n\times m}$ and for all $y \in \mathsf{D}$ we have*

$$f(y) = f(x) + \lessdot \nabla f(x), y - x \gtrdot + \|y - x\| R_1(y - x; f, x), \tag{2.15}$$

*where $R_1(\cdot; f, x) : \mathbb{R}^n \to \mathbb{R}^m$ and $\lim_{h\to \mathbf{0}} R_1(h; f, x) = 0$.*

*The expression is more simple for $n = 1$. Let $\mathsf{D} \subset \mathbb{R}$, $\mathsf{D} \neq \emptyset$, $f : \mathsf{D} \to \mathbb{R}^m$ and $t \in \operatorname{int}(\mathsf{D})$. Then, $f$ is differentiable at $t$ if and only if $f$ possesses a derivative $f'(t) \in \mathbb{R}^m$ and for all $s \in \mathsf{D}$ we have*

$$f(s) = f(t) + (s - t)f'(t) + |s - t| R_1(s - t; f, t), \tag{2.16}$$

*where $R_1(\cdot; f, x) : \mathbb{R} \to \mathbb{R}^m$ and $\lim_{h\to \mathbf{0}} R_1(h; f, x) = 0$.*

**Proof:** It is a straightforward rewriting of definition.

Q.E.D.

### 2.2.4 Chain rule

Differentiability directly implies chain rule (cz. řetízkové pravidlo).

**Lemma 2.19** *Let $I \subset \mathbb{R}$, $\operatorname{int}(I) \neq \emptyset$, $\mathsf{D} \subset \mathbb{R}^n$, $\operatorname{int}(\mathsf{D}) \neq \emptyset$, $g : I \to \mathsf{D}$, $f : \mathsf{D} \to \mathbb{R}$ and $t \in \operatorname{int}(I)$ such that $g(t) \in \operatorname{int}(\mathsf{D})$. If $f$ is differentiable at $g(t)$ and $g$ is differentiable at $t$, then $f \circ g$ is differentiable at $t$ and*

$$(f \circ g)'(t) \quad = \quad \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(g(t))\, g_i'(t) = \langle \nabla f(g(t)), g'(t)\rangle . \tag{2.17}$$

**Proof:** Take $s \in I$, $s \neq t$. Accordingly to differentiability of $f$ at $g(t)$ and differentiability of $g$ at $t$, we have

$$\begin{aligned}
&f(g(s)) - f(g(t)) \\
&\quad = \langle \nabla f(g(t)), g(s) - g(t)\rangle + \|g(s) - g(t)\| R_1(g(s) - g(t); f, g(t)) \\
&\quad = \langle \nabla f(g(t)), (s - t)\, g'(t) + |s - t| R_1(s - t; g, t)\rangle \\
&\qquad + \|(s - t)\, g'(t) + |s - t| R_1(s - t; g, t)\| R_1(g(s) - g(t); f, g(t)) \\
&\quad = \langle \nabla f(g(t)), g'(t)\rangle (s - t) + |s - t| \langle \nabla f(g(t)), R_1(s - t; g, t)\rangle \\
&\qquad + |s - t| \left\| \frac{s - t}{|s - t|} g'(t) + R_1(s - t; g, t)\right\| R_1(g(s) - g(t); f, g(t)) \\
&\quad = \langle \nabla f(g(t)), g'(t)\rangle (s - t) + |s - t| R_1(s - t; f \circ g, t),
\end{aligned}$$

where

$$
\begin{aligned}
R_1(w; f\circ g, t) &= \langle \nabla f(g(t)), R_1(w; g, t) \rangle \\
&\quad + \|g'(t) + R_1(w; g, t)\| \, R_1(g(t+w) - g(t); f, g(t)) \\
&\quad \text{ if } w > 0, \\
&= \langle \nabla f(g(t)), R_1(w; g, t) \rangle \\
&\quad + \|-g'(t) + R_1(w; g, t)\| \, R_1(g(t+w) - g(t); f, g(t)) \\
&\quad \text{ if } w < 0, \\
&= 0 \text{ if } w = 0,
\end{aligned}
$$

Thus, $f \circ g$ is differentiable at $t$ and (2.17) is shown.

Q.E.D.

### 2.2.5 Second derivative

Also, second derivative will be employed.

**Definition 2.20** *Let $\mathsf{D} \subset \mathbb{R}^n$, $\mathsf{D} \neq \emptyset$, $f : \mathsf{D} \to \mathbb{R}$ and $x \in \operatorname{int}(\mathsf{D})$. We say, $f$ possesses second partial derivatives at $x$ (cz. má druhé parciální derivace v $x$), if $f$ possesses a gradient on a neighborhood of $x$ and all partial derivatives of the gradient at $x$ exists; i.e. $\frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right)(x)$ exists for all indexes $i, j \in \{1, 2, \ldots, n\}$.*

*Then, we denote $\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right)(x)$ for all $i, j \in \{1, 2, \ldots, n\}$. Matrix of second partial derivatives is denoted $\nabla^2 f(x) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x)\right)_{i=1,j=1}^{n,n}$ and called Hessian matrix.*

**Definition 2.21** *Let $\mathsf{D} \subset \mathbb{R}^n$, $\mathsf{D} \neq \emptyset$, $f : \mathsf{D} \to \mathbb{R}$ and $x \in \operatorname{int}(\mathsf{D})$. We say, $f$ is twice differentiable at $x$ (or, possesses second Peano derivative) (cz. dvakrát diferencovatelná v $x$), if there is a gradient $\nabla f(x) \in \mathbb{R}^n$ and a symmetric matrix $H_f(x) \in \mathbb{R}^{n\times n}$ such that for all $y \in \mathsf{D}$ we have*

$$
\begin{aligned}
f(y) &= f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2} \langle y - x, H_f(x)(y - x) \rangle \\
&\quad + \|y - x\|^2 R_2(y - x; f, x),
\end{aligned}
\tag{2.18}
$$

*where $\lim_{h\to\mathbf{0}} R_2(h; f, x) = 0$ .*

*If $\mathsf{S} \subset \operatorname{int}(\mathsf{D})$, then we say, $f$ is twice differentiable at $\mathsf{S}$ (cz. dvakrát diferencovatelná v množině $\mathsf{S}$), if it is twice differentiable at each $x \in \mathsf{S}$.*

Matrix $H_f(x)$ can differ from Hessian matrix. The reasons are

- $\nabla f$ does not exist in any neighborhood of $x$,
- $\nabla f$ exists in a neighborhood of $x$ and $\nabla^2 f(x)$ does not exist.
- $\nabla f$ exists in a neighborhood of $x$, $\nabla^2 f(x)$ exist, but, asymmetric.

**Lemma 2.22** *Let $\mathsf{D} \subset \mathbb{R}^n$, $\mathsf{D} \neq \emptyset$, $f : \mathsf{D} \to \mathbb{R}$ and $x \in \operatorname{int}(\mathsf{D})$. If $f$ is twice differentiable at $x$ then matrix $H_f(x)$ is unequally determined.*

**Proof:** Since $H_f(x)$ is symmetric, its uniqueness follows from Lemma 1.45.

Q.E.D.

**Lemma 2.23** *Let* $\mathsf{D} \subset \mathbb{R}^n$, $\mathsf{D} \neq \emptyset$, $f : \mathsf{D} \to \mathbb{R}$ *and* $x \in \operatorname{int}(\mathsf{D})$. *If* $f$ *is differentiable at a neighborhood of* $x$ *and* $\nabla f$ *is differentiable at* $x$, *then,* $\nabla^2 f(x)$ *exists and* $f$ *is twice differentiable at* $x$ *with*

$$H_f(x) = \frac{1}{2}\nabla^2 f(x) + \frac{1}{2}\left(\nabla^2 f(x)\right)^\top.$$

*If, moreover, Hessian matrix is symmetric, i.e.* $\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x)$ *for all* $i, j \in \{1, 2, \ldots, n\}$, *then*

$$H_f(x) = \nabla^2 f(x).$$

**Proof:** According to our assumptions, there is $\delta > 0$ such that $\mathcal{U}(x, \delta) \subset \mathsf{D}$ and for all $y \in \mathcal{U}(x, \delta)$, $h \in \mathbb{R}^n$, $\|h\| < \delta - \|y - x\|$ we have

$$f(y+h) - f(y) = \langle \nabla f(y), h \rangle + \|h\| R_1(h; f, y),$$
$$\nabla f(y) - \nabla f(x) = \lessdot \left(\nabla^2 f(x)\right)^\top, y - x \gtrdot + \|y - x\| R_1(y - x; \nabla f, x).$$

According to Lemma 2.14

$$f(x+h) - f(x) = \int_0^1 \langle \nabla f(x + th), h \rangle \, \mathrm{d}t.$$

Plugging in expansion of the gradient, we are receiving

$$\begin{aligned}
& f(x+h) - f(x) - \nabla f(x) = \int_0^1 \langle \nabla f(x+th) - \nabla f(x), h \rangle \, \mathrm{d}t \\
&= \int_0^1 \left\langle \lessdot \left(\nabla^2 f(x)\right)^\top, th \gtrdot + \|th\| R_1(th; \nabla f, x), h \right\rangle \mathrm{d}t \\
&= \int_0^1 t \left\langle \lessdot \left(\nabla^2 f(x)\right)^\top, h \gtrdot, h \right\rangle \mathrm{d}t + \int_0^1 |t| \langle \|h\| R_1(th; \nabla f, x), h \rangle \, \mathrm{d}t \\
&= \frac{1}{2}\left\langle h, \left(\nabla^2 f(x)\right)^\top h \right\rangle + \|h\|^2 \int_0^1 |t| \left\langle R_1(th; \nabla f, x), \frac{h}{\|h\|} \right\rangle \mathrm{d}t \\
&= \frac{1}{2}\left\langle h, \frac{1}{2}\left(\nabla^2 f(x) + \left(\nabla^2 f(x)\right)^\top\right) h \right\rangle + \|h\|^2 \int_0^1 |t| \left\langle R_1(th; \nabla f, x), \frac{h}{\|h\|} \right\rangle \mathrm{d}t,
\end{aligned}$$

where

$$\lim_{h \to \mathbf{0}} \int_0^1 |t| \left\langle R_1(th; \nabla f, x), \frac{h}{\|h\|} \right\rangle \mathrm{d}t = 0 \quad \text{since} \quad \lim_{s \to \mathbf{0}} R_1(s; \nabla f, x) = 0.$$

We have proved $f$ is twice differentiable at $x$ with $H_f(x) = \frac{1}{2}\left(\nabla^2 f(x) + \left(\nabla^2 f(x)\right)^\top\right)$.

Q.E.D.

**Lemma 2.24** *Let* $\mathsf{D} \subset \mathbb{R}^n$, $\mathsf{D} \neq \emptyset$, $f : \mathsf{D} \to \mathbb{R}$, $x \in \operatorname{int}(\mathsf{D})$ *and* $h \in \mathbb{R}^n$.

***(i)*** *If $f$ is twice differentiable at $x$, then*

$$\lim_{t\to 0}\frac{f(x+th)-f(x)-t\langle \nabla f(x),h\rangle}{t^2} = \frac{1}{2}\langle h, H_f(x)\,h\rangle. \tag{2.19}$$

***(ii)*** *Let us denote $\mathsf{D}_h = \{t\in\mathbb{R} : x+th\in\mathsf{D}\}$. If $f$ is differentiable at a neighborhood of $x$ and $\nabla f$ is differentiable at $x$, then, $\nabla^2 f(x)$ exists and function $\varphi : \mathsf{D}_h \to \mathbb{R} : t\in\mathsf{D}_h \to f(x+th)$ possesses derivatives*

$$\varphi'(t) = \langle \nabla f(x+th), h\rangle \quad \textit{for all } t \textit{ small enough}, \tag{2.20}$$

$$\varphi''(0) = \left\langle h, \nabla^2 f(x)\,h\right\rangle. \tag{2.21}$$

**Proof:**

1. (i) follows (2.18), since for $t\neq 0$

$$\frac{f(x+th)-f(x)-t\langle \nabla f(x),h\rangle}{t^2} = \frac{1}{2}\langle h, H_f(x)\,h\rangle + \|h\|^2 R_2(th; f, x).$$

2. (ii) follows Lemma 2.23 and (2.12), (2.15), since for $s\neq 0$

$$\begin{aligned}\frac{\varphi(t+s)-\varphi(t)}{s} &= \frac{f(x+(t+s)h)-f(x+th)}{s}\\ &= \langle \nabla f(x+th), h\rangle + \|h\|\,R_1(sh; f, x+th),\\ \frac{\varphi'(s)-\varphi'(0)}{s} &= \frac{\langle \nabla f(x+sh),h\rangle - \langle \nabla f(x),h\rangle}{s}\\ &= \left\langle h, \nabla^2 f(x)\,h\right\rangle + \|h\|\,R_1(sh;\nabla f, x).\end{aligned}$$

Q.E.D.

### 2.2.6 Arguments for differentiability

Existence and continuity of the gradient, resp. of Hessian, are sufficient conditions for differentiability in the sense of Definitions 2.10 and 2.21.

**Lemma 2.25** *Let $I\subset\mathbb{R}$, $\operatorname{int}(I)\neq\emptyset$, $\mathsf{D}\subset\mathbb{R}^n$, $\mathsf{D}\neq\emptyset$, $g : I\to\mathsf{D}$, $f : \mathsf{D}\to\mathbb{R}$ and $t\in\operatorname{int}(I)$ such that $g(t)\in\operatorname{int}(\mathsf{D})$. If a gradient of $f$ exists on a neighborhood of $g(t)$ and is continuous at $g(t)$ and $g$ is differentiable at $t$, then $f\circ g$ is differentiable at $t$ with*

$$(f\circ g)'(t) = \sum_{i=1}^{n}\frac{\partial f}{\partial x_i}(g(t))\,g_i'(t) = \langle \nabla f(g(t)), g'(t)\rangle. \tag{2.22}$$

**Proof:** For $s\in I$, $s\neq t$, $i\in\{1,2,\ldots,n\}$, $u\in[0,1]$, we denote

$$\xi(u,s,i) = (g_1(t),\ldots,g_{i-1}(t), g_i(t)+u(g_i(s)-g_i(t)), g_{i+1}(s),\ldots,g_n(s)).$$

Then,

$$\begin{aligned} f\circ g(s) - f\circ g(t) &= \sum_{i=1}^{n}\left[f(\xi(1,s,i)) - f(\xi(0,s,i))\right]\\ &= \sum_{i=1}^{n}\int_0^1 \frac{\partial f}{\partial x_i}(\xi(u,s,i))\,(g_i(s)-g_i(t))\,\mathsf{d}u.\end{aligned}$$

Divide formula by $s-t$ and let $s \to t$.
We receive formula (2.22), since the gradient of $f$ is continuous at $g(t)$.

Q.E.D.

Using Lemma 2.25, we derive differentiability of a function.

**Lemma 2.26** *Let* $\mathsf{D} \subset \mathbb{R}^n$, $\mathsf{D} \neq \emptyset$, $f : \mathsf{D} \to \mathbb{R}$ *and* $x \in \mathrm{int}(\mathsf{D})$. *If a gradient of* $f$ *exists on a neighborhood of* $g(t)$ *and is continuous at* $x$, *then* $f$ *is differentiable at* $x$ *with*

$$\begin{aligned} & f(x+h) = f(x) + \langle \nabla f(x), h \rangle + \|h\| R_1(h; f, x), \\ & |R_1(h; f, x)| \leq \max\{\|\nabla f(x+uh) - \nabla f(x)\| : 0 \leq u \leq 1\} \end{aligned} \tag{2.23}$$

*if* $h$ *is sufficiently small.*

**Proof:** Using Lemma 2.25 for $h \in \mathbb{R}^n$ sufficiently small, we receive an expansion

$$\begin{aligned} & f(x+h) - f(x) = \int_0^1 \langle \nabla f(x+uh), h \rangle \,\mathrm{d}u \\ = \quad & \langle \nabla f(x), h \rangle + \int_0^1 \langle \nabla f(x+uh) - \nabla f(x), h \rangle \,\mathrm{d}u \\ = \quad & \langle \nabla f(x), h \rangle + \|h\| R_1(h; f, x), \\ & |R_1(h; f, x)| \leq \max\{\|\nabla f(x+uh) - \nabla f(x)\| : 0 \leq u \leq 1\}. \end{aligned}$$

Q.E.D.

**Lemma 2.27** *Let* $\mathsf{D} \subset \mathbb{R}^n$, $\mathsf{D} \neq \emptyset$, $f : \mathsf{D} \to \mathbb{R}$ *and* $x \in \mathrm{int}(\mathsf{D})$. *Then,*
$f$ *is continuously differentiable at a neighborhood of* $x$ *if and only if there is* $\delta > 0$ *such that* $\nabla f$ *exists at* $\mathcal{U}(x, \delta)$ *and is continuous at* $\mathcal{U}(x, \delta)$.

**Proof:** A consequence of Lemma 2.26.

Q.E.D.

**Lemma 2.28** *Let* $\mathsf{D} \subset \mathbb{R}^n$, $\mathsf{D} \neq \emptyset$, $f : \mathsf{D} \to \mathbb{R}$ *and* $x \in \mathrm{int}(\mathsf{D})$. *If* $\nabla f$, $\nabla^2 f$ *exist on a neighborhood of* $x$ *and* $\nabla^2 f$ *is continuous at* $x$, *then Hessian* $\nabla^2 f(x)$ *is a symmetric matrix and* $f$ *is twice differentiable at* $x$ *with*

$$\begin{aligned} f(x+h) \quad = \quad & f(x) + \langle \nabla f(x), h \rangle + \frac{1}{2} \langle h, \nabla^2 f(x) h \rangle + \\ & + \frac{1}{2} \|h\|^2 R_2(h; f, x), \\ & |R_2(h; f, x)| \leq \max\left\{\left\|\nabla^2 f(x+uh) - \nabla^2 f(x)\right\| : 0 \leq u \leq 1\right\} \end{aligned} \tag{2.24}$$

*if* $h$ *sufficiently small. Moreover,* $H_f(x) = \nabla^2 f(x)$.

**Proof:** Take $x, h \in \mathbb{R}^n$. Using twice Lemma 2.25, we derive

$$
\begin{aligned}
& f(x+h) - f(x) = \int_0^1 \langle \nabla f(x+uh), h \rangle \,\mathrm{d}u \\
= \quad & \langle \nabla f(x), h \rangle + \int_0^1 \langle \nabla f(x+uh) - \nabla f(x), h \rangle \,\mathrm{d}u \\
= \quad & \langle \nabla f(x), h \rangle + \int_0^1 \int_0^u \left\langle h, \nabla^2 f(x+vh)\, h \right\rangle \mathrm{d}v\,\mathrm{d}u \\
= \quad & \langle \nabla f(x), h \rangle + \frac{1}{2} \left\langle h, \nabla^2 f(x)\, h \right\rangle + \\
& + \int_0^1 \int_0^u \left\langle h, \left(\nabla^2 f(x+vh) - \nabla^2 f(x)\right) h \right\rangle \mathrm{d}v\,\mathrm{d}u \\
= \quad & \langle \nabla f(x), h \rangle + \frac{1}{2} \left\langle h, \nabla^2 f(x)\, h \right\rangle + \frac{1}{2} \|h\|^2 R_2(h; f, x), \\
& |R_2(h; f, x)| \le \max\left\{ \left\| \nabla^2 f(x+uh) - \nabla^2 f(x) \right\| \,:\, 0 \le u \le 1 \right\}.
\end{aligned}
$$

Q.E.D.

## 2.3 Convex functions

### 2.3.1 Definition of a convex function

**Definition 2.29** *A function $f : \mathbb{R}^n \to \mathbb{R}^*$ is convex (cz. konvexní), if $\mathsf{epi}(f)$ is a convex set.*

Convexity of a function can be equivalently explained.

**Lemma 2.30** *If a function $f : \mathbb{R}^n \to \mathbb{R}^*$ is convex, then $\mathsf{Dom}(f)$ is a convex set.*

**Proof:** Let $x, y \in \mathsf{Dom}(f)$ and $0 < \lambda < 1$.
Then, there is $\eta, \xi \in \mathbb{R}$ such that $f(x) \le \eta$ and $f(y) \le \xi$.
Hence, $\begin{pmatrix} x \\ \eta \end{pmatrix}, \begin{pmatrix} y \\ \xi \end{pmatrix} \in \mathsf{epi}(f)$.
Since $\mathsf{epi}(f)$ is convex, $(\lambda x + (1-\lambda)y, \lambda\eta + (1-\lambda)\xi) \in \mathsf{epi}(f)$.
Hence, $f(\lambda x + (1-\lambda)y) \le \lambda\eta + (1-\lambda)\xi < +\infty$.
Therefore, $\lambda x + (1-\lambda)y \in \mathsf{Dom}(f)$ and convexity of $\mathsf{Dom}(f)$ is shown.

Q.E.D.

**Theorem 2.31**: *Function $f : \mathbb{R}^n \to \mathbb{R}^*$ is convex if and only if $\mathsf{Dom}(f)$ is a convex set and for all $x, y \in \mathsf{Dom}(f)$ and $0 < \lambda < 1$ we have*

$$
f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y). \tag{2.25}
$$

**Proof:**

1. Let $f$ is convex.
   Then accordingly to Lemma 2.30, $\mathsf{Dom}\,(f)$ is a convex set.
   Let $x, y \in \mathsf{Dom}\,(f)$ and $0 < \lambda < 1$.
   Then for all $\eta, \xi \in \mathbb{R}$ fulfilling $f\,(x) \leq \eta$ and $f\,(y) \leq \xi$,
   one has $\begin{pmatrix} x \\ \eta \end{pmatrix}, \begin{pmatrix} y \\ \xi \end{pmatrix} \in \mathsf{epi}\,(f)$.
   $\mathsf{epi}\,(f)$ is convex, then, $(\lambda x + (1-\lambda)y, \lambda\eta + (1-\lambda)\xi) \in \mathsf{epi}\,(f)$.
   Hence, $f\,(\lambda x + (1-\lambda)y) \leq \lambda\eta + (1-\lambda)\xi < +\infty$.
   Minimum over all possible $\eta, \xi$ is giving (2.25).

2. Let property (2.25) be fulfilled.
   Take $\begin{pmatrix} x \\ \eta \end{pmatrix}, \begin{pmatrix} y \\ \xi \end{pmatrix} \in \mathsf{epi}\,(f)$ and $0 < \lambda < 1$. Then,

$$\lambda\eta + (1-\lambda)\xi \geq \lambda f\,(x) + (1-\lambda) f\,(y) \geq f\,(\lambda x + (1-\lambda)y)\,.$$

   Hence, $(\lambda x + (1-\lambda)y, \lambda\eta + (1-\lambda)\xi) \in \mathsf{epi}\,(f)$.
   We found $\mathsf{epi}\,(f)$ is a convex set, therefore, $f$ is a convex function.

Q.E.D.

Theorem 2.31 shows, that new definition 2.29 coincides with classical definition of a convex function, if function is proper and a restriction $f : \mathsf{Dom}\,(f) \to \mathbb{R}$ is considered.
A note, convex function attaining value $-\infty$ is degenerated.

**Lemma 2.32** *Let $f : \mathbb{R}^n \to \mathbb{R}^*$ be a convex function. Then, either $f\,(x) \in \mathbb{R}$ for all $x \in \mathsf{Dom}\,(f)$ or $f\,(x) = -\infty$ for all $x \in \mathrm{rint}\,(\mathsf{Dom}\,(f))$.*

**Proof:** Let $x \in \mathsf{Dom}\,(f)$ and $f\,(x) = -\infty$.
If $y \in \mathrm{rint}\,(\mathsf{Dom}\,(f))$, then there is $z \in \mathsf{Dom}\,(f)$ and $0 < \lambda \leq 1$ such that $y = \lambda x + (1-\lambda)z$.
Using property (2.25), we receive

$$f\,(y) \;=\; f\,(\lambda x + (1-\lambda)z) \leq \lambda f\,(x) + (1-\lambda) f\,(z) = -\infty.$$

Q.E.D.

**Theorem 2.33**: *If function $f : \mathbb{R}^n \to \mathbb{R}^*$ is convex and proper, then it is continuous on* $\mathrm{rint}\,(\mathsf{Dom}\,(f))$.

**Proof:** Without any loss of generality we can assume, $\mathrm{int}\,(\mathsf{Dom}\,(f)) \neq \emptyset$. Otherwise, we will consider the problem in coordinate system of the smallest lineal containing $\mathsf{Dom}\,(f)$.

Let $x \in \mathrm{int}\,(\mathsf{Dom}\,(f))$.
Then, there is $\Delta > 0$ such that $x + \Delta \mathsf{e}_{i:n}, x - \Delta \mathsf{e}_{i:n} \in \mathsf{Dom}\,(f)$ for all $i = 1, 2, \ldots, n$.
$\mathsf{Dom}\,(f)$ is convex, therefore,

$$\mathcal{M} = \mathrm{conv}\left(\{x + \Delta \mathsf{e}_{i:n}, x - \Delta \mathsf{e}_{i:n} \;:\; i = 1, 2, \ldots, n\}\right) \subset \mathsf{Dom}\,(f)\,.$$

Each point $y \in \mathcal{M}$ can be written as

$$\begin{aligned} y &= \sum_{i=1}^{n} \lambda_{i,+}(x + \Delta \mathsf{e}_{i:n}) + \sum_{i=1}^{n} \lambda_{i,-}(x - \Delta \mathsf{e}_{i:n}), \\ \text{where} & \sum_{i=1}^{n} \lambda_{i,+} + \sum_{i=1}^{n} \lambda_{i,-} = 1,\ \lambda_{i,+}, \lambda_{i,-} \geq 0. \end{aligned}$$

Hence for $y \in \mathcal{M}$ we receive a bound

$$\begin{aligned} f(y) &\leq \sum_{i=1}^{n} \lambda_{i,+} f(x + \Delta \mathsf{e}_{i:n}) + \sum_{i=1}^{n} \lambda_{i,-} f(x - \Delta \mathsf{e}_{i:n}) \leq \Xi < +\infty, \\ \text{where} &\ \Xi := \max \{ f(x + \Delta \mathsf{e}_{i:n}), f(x - \Delta \mathsf{e}_{i:n}) \,:\, i = 1, 2, \ldots, n \}. \end{aligned}$$

Point $y \in \mathcal{M}$ can be also represented as $y = x + \delta s$, where $\sum_{i=1}^{n} |s_i| = \Delta$ and $0 \leq \delta \leq 1$. Then,

$$\begin{aligned} f(y) &= f(x + \delta s) = f((1-\delta)x + \delta(x+s)) \leq (1-\delta) f(x) + \delta f(x+s) \\ &\leq (1-\delta) f(x) + \delta \Xi, \\ f(x) &= f\left( \frac{1}{1+\delta}(x + \delta s) + \frac{\delta}{1+\delta}(x - s) \right) \\ &\leq \frac{1}{1+\delta} f(x + \delta s) + \frac{\delta}{1+\delta} f(x - s) \\ &\leq \frac{1}{1+\delta} f(y) + \frac{\delta}{1+\delta} \Xi . \end{aligned}$$

Finally, we are receiving

$$(1+\delta) f(x) - \delta \Xi \leq f(y) \leq (1-\delta) f(x) + \delta \Xi .$$

Thus, $f$ is continuous at each $x \in \operatorname{int}(\mathsf{Dom}(f))$.

Q.E.D.

Continuity of a convex function at boundary of its domain is not an easy task. A necessary condition is valid for a general proper function.

**Theorem 2.34**: *Let a function $f : \mathbb{R}^n \to \mathbb{R}^*$ be proper and continuous on $\mathsf{Dom}(f)$. Then,*

$$\mathsf{epi}(f) = \operatorname{clo}(\mathsf{epi}(f)) \cap (\mathsf{Dom}(f) \times \mathbb{R}) . \tag{2.26}$$

**Proof:** Let $x \in \mathsf{Dom}(f)$ and $\begin{pmatrix} x \\ \eta \end{pmatrix} \in \operatorname{clo}(\mathsf{epi}(f))$.

Then, there is a sequence $(x_k, \eta_k) \in \mathsf{epi}(f)$ converging to $\begin{pmatrix} x \\ \eta \end{pmatrix}$.

Hence, we have $f(x_k) \leq \eta_k$.
Function is continuous on $\mathsf{Dom}(f)$, after a limit passage we receive $f(x) \leq \eta$.
Thus, we have shown $\begin{pmatrix} x \\ \eta \end{pmatrix} \in \mathsf{epi}(f)$.

Q.E.D.

Theorem possesses a nice consequence.

**Consequence:** Let $f : \mathbb{R}^n \to \mathbb{R}^*$ be a proper function continuous on $\mathsf{Dom}\,(f)$ and $\mathsf{Dom}\,(f)$ be a closed set. Then, $\mathsf{epi}\,(f)$ is also a closed set. ♣

**Proof:** Statement is a direct consequence of Theorem 2.34, since $\mathsf{Dom}\,(f)$ is a closed set, and hence,

$$\mathrm{clo}\,(\mathsf{epi}\,(f)) \cap (\mathsf{Dom}\,(f) \times \mathbb{R}) = \mathrm{clo}\,(\mathsf{epi}\,(f))\,.$$

Q.E.D.

**Theorem 2.35**: *If functions $f_i : \mathbb{R}^n \to \mathbb{R}^*$ are convex for all $i \in I$, then $\sup_{i \in I} f_i : \mathbb{R}^n \to \mathbb{R}^*$ is also convex.*

**Proof:** According to Lemma 2.4 we have $\mathsf{epi}\left(\sup_{i\in I} f_i\right) = \bigcap_{i\in I} \mathsf{epi}\,(f_i)$.
Intersection of convex sets is a convex set; see Lemma 1.4.
The statement is proved.

Q.E.D.

**Theorem 2.36**: *If $f : \mathbb{R}^n \to \mathbb{R}^*$ is a convex function, then, sets $\{x : f(x) \le \alpha\}$, $\{x : f(x) < \alpha\}$ are convex for all $\alpha \in \mathbb{R}$.*

*These sets are called level sets of $f$ (cz. úrovňové množiny funkce $f$).*

**Proof:** It is sufficient to verify that the set $\{x : f(x) < \alpha\}$ is convex, since

$$\{x : f(x) \le \alpha\} = \bigcap_{\beta > \alpha} \{x : f(x) < \beta\}\,.$$

Take $y, z \in \{x : f(x) < \alpha\}$ and $0 < \lambda < 1$. Then, $y, z \in \mathsf{Dom}\,(f)$ and we have

$$f(\lambda y + (1-\lambda)z) \le \lambda f(y) + (1-\lambda) f(z) < \alpha.$$

Q.E.D.

As a consequence of Theorem 2.36, we are receiving that the set of all feasible solutions (cz. množina přípustných řešení) of a convex program is convex, i.e.

$$\{x \in \mathbb{R}^n : g_1(x) \le \alpha_1, g_2(x) \le \alpha_2, \ldots, g_k(x) \le \alpha_k\}$$

is a convex set, having functions $g_1, g_2, \ldots, g_k$ convex.

Convexity of a sets $\{x : f(x) \le \alpha\}$ and $\{x : f(x) < \alpha\}$ is not implying that function $f$ is convex.

**Example 2.37:** Function

$$\begin{aligned} f(x) &= \log(x) \quad \text{if } x > 0, \\ &= +\infty \quad \text{otherwise} \end{aligned}$$

is not convex, but, its level sets $\{x : f(x) \le \alpha\} = (0, \mathsf{e}^\alpha]$, $\{x : f(x) < \alpha\} = (0, \mathsf{e}^\alpha)$ are convex for all $\alpha \in \mathbb{R}$.

△

**Definition 2.38** *We say, function $f : \mathbb{R}^n \to \mathbb{R}^*$ is*

*i)* *strictly convex (cz. ryze konvexní), if for all couple of points $x, y \in \mathsf{Dom}(f)$, $x \neq y$ and $0 < \lambda < 1$ we have inequality*

$$f(\lambda x + (1-\lambda)y) \quad < \quad \lambda f(x) + (1-\lambda) f(y).$$

*ii)* *concave (cz. konkávní), if the function $-f$ is convex.*

*iii)* *strictly concave, (cz. ryze konkávní), if the function $-f$ is strictly convex.*

Concave function can be equivalently defined as a function with convex hypograph.

Consider, strictly convex function is always proper. The only exception is if $\mathsf{Dom}(f)$ is a one-point set.

Convex functions are very important at optimization theory, since its local minima are immediately global minima.

**Theorem 2.39**: *Let $f : \mathbb{R}^n \to \mathbb{R}^*$ be a proper convex function. Then, each local minimum of $f$ on $\mathsf{Dom}(f)$ is a global minimum of $f$ on $\mathsf{Dom}(f)$.*

*Set of all global minimizers is convex.*

**Proof:** Let $\hat{x} \in \mathsf{Dom}(f)$ be a local minimum of $f$ on $\mathsf{Dom}(f)$, but, it is no global minimum.
Then, there is $y \in \mathsf{Dom}(f)$ such that $f(y) < f(\hat{x})$.
Then, for all $\alpha \in (0,1)$ we have $f(\alpha\hat{x} + (1-\alpha)y) \leq \alpha f(\hat{x}) + (1-\alpha) f(y) < f(\hat{x})$.
That is contradiction, since $\hat{x}$ is a local minimum of $f$ on $\mathsf{Dom}(f)$.
Hence, $\hat{x}$ is a global minimum of $f$ on $\mathsf{Dom}(f)$.
Set of all global minimizers $\{x \in \mathbb{R}^n : f(x) \leq \inf\{f(y) : y \in \mathbb{R}^n\}\}$ is convex, according to Theorem 2.36.

Q.E.D.

**Theorem 2.40**: *Let $f : \mathbb{R}^n \to \mathbb{R}^*$ be a proper strictly convex function which possesses a local minimum on $\mathsf{Dom}(f)$, then, it possesses unique global minimum on $\mathsf{Dom}(f)$. This global minimum is unique local minimum of $f$ on $\mathsf{Dom}(f)$.*

**Proof:** According to Theorem 2.39, each local minimum of $f$ on $\mathsf{Dom}(f)$ is also its global minimum.
It is sufficient to show uniqueness of global minimum of $f$ on $\mathsf{Dom}(f)$.
Take $\hat{x}$ a global minimum of $f$ on $\mathsf{Dom}(f)$ and $y \in \mathsf{Dom}(f)$, $y \neq \hat{x}$. Applying strict convexity of $f$, we receive

$$f\left(\frac{1}{2}\hat{x} + \frac{1}{2}y\right) < \frac{1}{2} f(\hat{x}) + \frac{1}{2} f(y).$$

Hence,

$$f(y) > 2f\left(\frac{1}{2}\hat{x} + \frac{1}{2}y\right) - f(\hat{x}) \geq 2f(\hat{x}) - f(\hat{x}) = f(\hat{x}).$$

We have shown $\hat{x}$ is unique global minimum of $f$ on $\mathsf{Dom}(f)$.

Q.E.D.

Let us recapitulate basic properties of convex functions.

### 2.3.2 Convex functions of one variable

This section sums up basic properties of convex functions of one variable. Presented results are listed without proofs. Interested readers can consult basic textbooks on mathematical analysis and probability theory.

We will consider a function $f : J \to \mathbb{R}$ defined on a convex set $J \subset \mathbb{R}$ (Recall simple structure of convex sets on real line. They are either empty set or point or interval.) Consider smoothness of convex functions.

**Theorem 2.41**: *Let $J \subset \mathbb{R}$ be an interval and $f : J \to \mathbb{R}$ be a convex function.*

*i) Function $f$ is continuous on $\mathrm{int}\,(J)$ and it can jump in extreme points of $J$. Jumps must keep bounds: $f(a) \geq f(a+)$ if $a$ is a left extreme point of $J$, $f(a) \geq f(a-)$ if $a$ is a right extreme point of $J$.*

*ii) Derivative from left and from right exist at each point $t \in \mathrm{int}\,(J)$; i.e.*

$$\begin{aligned} f'_+(t) &= \lim_{h\to 0+} \frac{f(t+h)-f(t)}{h} \in \mathbb{R}, \\ f'_-(t) &= \lim_{h\to 0-} \frac{f(t+h)-f(t)}{h} \in \mathbb{R}. \end{aligned}$$

*We have $f'_-(t) \leq f'_+(t) \leq f'_-(s) \leq f'_+(s)$, whenever $t, s \in \mathrm{int}\,(J)$, $t < s$.*

*iii) $f'$ exists on $J$ except at most countably many point.*

*iv) $f''$ exists on $J$ except a set of Lebesgue measure zero.*

*v) $f$ fulfills an inequality $f\left(\sum_{i=1}^k p_i x_i\right) \leq \sum_{i=1}^k p_i f(x_i)$ for all $x_1, x_2, \ldots, x_k \in J$, $p_1 \geq 0$, $p_2 \geq 0$, $\ldots$, $p_k \geq 0$, $\sum_{i=1}^k p_i = 1$.*

*vi) $f$ fulfills Jensen inequality, i.e. $f(\mathsf{E}[X]) \leq \mathsf{E}[f(X)]$ for each real random variable $X$ with a finite mean and with $P(X \in J) = 1$.*

Recall, v) is a particular case of vi) . To see that, consider a random variable $X$ attaining values $x_1, x_2, \ldots, x_k$ with probabilities $p_1, p_2, \ldots, p_k$.

Now, we recall some basic criteria indicating convex functions.

**Theorem 2.42**: *Let $J \subset \mathbb{R}$ be an open interval and $f : J \to \mathbb{R}$ be a function, then we have:*

- *Function $f$ is convex $\Leftrightarrow$ $f'_+$ exists nondecreasing on $J$ $\Leftrightarrow$*

  *$\Leftrightarrow$ $f'_-$ exists nondecreasing on $J$.*

- *If $f$ is differentiable on $J$, then*

  *$f$ is convex $\Leftrightarrow$ $f'$ is nondecreasing on $J$.*

- *If $f$ is twice differentiable on $J$, then*

  *$f$ is convex $\Leftrightarrow$ $f'' \geq 0$ on $J$.*

### 2.3.3 Convex function of several variables

This section sums up basic properties of convex functions of several variables. Presented results are listed without proofs. Interested readers can consult basic textbooks on mathematical analysis, linear algebra and probability theory.

Consider a function $f : \mathsf{D} \to \mathbb{R}$ defined on a convex set $\mathsf{D} \subset \mathbb{R}^n$.

**Lemma 2.43** *Let $\mathsf{D} \subset \mathbb{R}^n$, $\mathsf{D} \neq \emptyset$ be a convex set and $f_1 : \mathsf{D} \to \mathbb{R}$, $f_2 : \mathsf{D} \to \mathbb{R}$, $\dots$, $f_k : \mathsf{D} \to \mathbb{R}$ be convex functions. Then, $\sum_{i=1}^k a_i f_i : \mathsf{D} \to \mathbb{R}$ is a convex function for all $a_1 \geq 0$, $a_2 \geq 0$, $\dots$, $a_k \geq 0$.*

**Proof:** A proof is straightforward.

Q.E.D.

**Theorem 2.44** (Jensen inequality): *Let $\mathsf{D} \subset \mathbb{R}^n$ be a nonempty convex set and $f : \mathsf{D} \to \mathbb{R}$ be a convex function. If a real random vector $X = (X_1, X_2, \dots, X_k)^\top$ possesses finite mean and $P(X \in \mathsf{D}) = 1$ then we have*
*$\mathsf{E}[X] \in \mathsf{D}$ and $f(\mathsf{E}[X]) \leq \mathsf{E}[f(X)]$.*

**Proof:** A proof can be found for example in [4], Theorem 5.9, p.26.

Q.E.D.

A consequence of Theorem 2.44 is a generalization of inequality (2.25) ("deterministic Jensen inequality").

**Theorem 2.45**: *Let $\mathsf{D} \subset \mathbb{R}^n$, $\mathsf{D} \neq \emptyset$ be a convex set and $f : \mathsf{D} \to \mathbb{R}$ be a convex function. Then, an inequality*

$$f\left(\sum_{i=1}^k p_i x_i\right) \leq \sum_{i=1}^k p_i f(x_i) \tag{2.27}$$

*holds for all $x_1, x_2, \dots, x_k \in \mathsf{D}$, $p_1 \geq 0$, $p_2 \geq 0$, $\dots$, $p_k \geq 0$, $\sum_{i=1}^k p_i = 1$.*

**Proof:** The statement is a particular case of Theorem 2.44. To see that, consider a random variable $X$ attaining values $x_1, x_2, \dots, x_k$ with probabilities $p_1, p_2, \dots, p_k$.

Q.E.D.

Convexity of a function can be verified by means of functions of one variable.

**Theorem 2.46**: *Let $\mathsf{D} \subset \mathbb{R}^n$, $\mathsf{D} \neq \emptyset$ be a convex set and $f : \mathsf{D} \to \mathbb{R}$. Then, function $f$ is convex if and only if functions $\varphi_{x,s} : \mathsf{D}_{x,s} \to \mathbb{R}$ are convex for all $x \in \mathsf{D}$ and all $s \in \mathbb{R}^n$, where $\varphi_{x,s}(t) = f(x + ts)$ and $\mathsf{D}_{x,s} = \{t : x + ts \in \mathsf{D}, t \in \mathbb{R}\}$. (Let us recall set $\mathsf{D}_{x,s}$ is always an interval.)*

**Proof:**

1. Take $x \in \mathsf{D}$ and $s \in \mathbb{R}^n$.
   For $t_1, t_2 \in \mathsf{D}_{x,s}$ and $0 < \lambda < 1$ we have
   $$x + (\lambda t_1 + (1-\lambda)t_2)s = \lambda(x + t_1 s) + (1-\lambda)(x + t_2 s) \in \mathsf{D},$$
   since $x + t_1 s, x + t_2 s \in \mathsf{D}$ and $\mathsf{D}$ is a convex set.
   We have proved $\mathsf{D}_{x,s}$ is a convex subset of $\mathbb{R}$, therefore, it is an interval.
2. Let $f$ be a convex function and $x \in \mathsf{D}$, $s \in \mathbb{R}^n$.
   For $t_1, t_2 \in \mathsf{D}_{x,s}$ and $0 < \lambda < 1$ we have
   $$\begin{aligned} &\varphi_{x,s}(\lambda t_1 + (1-\lambda)t_2) = \\ &= f(x + (\lambda t_1 + (1-\lambda)t_2)s) = f(\lambda(x + t_1 s) + (1-\lambda)(x + t_2 s)) \\ &\le \lambda f(x + t_1 s) + (1-\lambda) f(x + t_2 s) = \lambda \varphi_{x,s}(t_1) + (1-\lambda)\varphi_{x,s}(t_2). \end{aligned}$$
   We have verified $\varphi_{x,s}$ is a convex function on an interval $\mathsf{D}_{x,s}$.
3. Let function $\varphi_{x,s}$ be convex on $\mathsf{D}_{x,s}$ for all $x \in \mathsf{D}$ and $s \in \mathbb{R}^n$.
   Take $x, y \in \mathsf{D}$, $0 < \lambda < 1$ and set $s = x - y$. Then, we have
   $$\begin{aligned} &f(\lambda x + (1-\lambda)y) = f(y + \lambda s) = \varphi_{y,s}(\lambda) \\ &\le \lambda \varphi_{y,s}(1) + (1-\lambda)\varphi_{y,s}(0) = \lambda f(x) + (1-\lambda) f(y). \end{aligned}$$
   We have verified $f$ is a convex function.

Q.E.D.

This property enables us generalize criteria for convex function identification.
We will need first and second derivative of projections.

**Lemma 2.47** *Let $\mathsf{D} \subset \mathbb{R}^n$, $\mathsf{D} \neq \emptyset$ be a convex open set and $f : \mathsf{D} \to \mathbb{R}$.*

- *If $f$ is differentiable at $\mathsf{D}$ and $x \in \mathsf{D}$, $s \in \mathbb{R}^n$, $t \in \mathsf{D}_{x,s}$, we have*
  $$\varphi'_{x,s}(t) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x + ts)\, s_i = \langle \nabla f(x + ts), s \rangle.$$
- *If $f$ is twice differentiable at $\mathsf{D}$ and $x \in \mathsf{D}$, $s \in \mathbb{R}^n$, $t \in \mathsf{D}_{x,s}$, we have*
  $$\begin{aligned} &\lim_{u \to 0} \frac{\varphi_{x,s}(t+u) - \varphi_{x,s}(t) - u \langle \nabla f(x + ts), s \rangle}{u^2} = \\ &= \frac{1}{2} \langle s, H_f(x + ts)\, s \rangle. \end{aligned}$$
- *If $f$ is differentiable at $\mathsf{D}$ and $\nabla f$ is differentiable at $\mathsf{D}$, then, $\nabla^2 f$ exists on $\mathsf{D}$ and for $x \in \mathsf{D}$, $s \in \mathbb{R}^n$, $t \in \mathsf{D}_{x,s}$, we have*
  $$\varphi'_{x,s}(t) = \langle \nabla f(x + ts), s \rangle, \qquad (2.28)$$
  $$\varphi''_{x,s}(t) = \langle s, \nabla^2 f(x + ts)\, s \rangle. \qquad (2.29)$$

**Proof:** Statement is a consequence of Lemmas 2.14, 2.24.

Q.E.D.

**Theorem 2.48**: *Let* $\mathsf{D} \subset \mathbb{R}^n$, $\mathsf{D} \neq \emptyset$ *be a convex open set and* $f : \mathsf{D} \to \mathbb{R}$ *be differentiable at* $\mathsf{D}$. *Then,*

$$f \text{ is convex} \quad \Leftrightarrow \quad \begin{array}{l} t \in \mathsf{D}_{x,s} \mapsto \langle \nabla f(x+ts), s \rangle \text{ is nondecreasing on } \mathsf{D}_{x,s} \text{ for all} \\ x \in \mathsf{D},\ s \in \mathbb{R}^n. \end{array} \tag{2.30}$$

**Proof:** According to Theorem 2.46 we have to verify convexity of all one-dimensional projections of $f$.
Take $x \in \mathsf{D}$, $s \in \mathbb{R}^n$ and consider function $\varphi_{x,s}$.
Function $f$ is differentiable at $\mathsf{D}$, therefore, according to Lemma 2.47, we have

$$\varphi'_{x,s}(t) = \langle \nabla f(x+ts), s \rangle .$$

Hence, according to Theorem 2.42

$$\varphi_{x,s} \text{ is convex} \quad \Leftrightarrow \quad t \in \mathsf{D}_{x,s} \mapsto \langle \nabla f(x+ts), s \rangle \quad \text{is a nondecreasing function.}$$

The statement is proved.

Q.E.D.

**Theorem 2.49**: *Let* $\mathsf{D} \subset \mathbb{R}^n$, $\mathsf{D} \neq \emptyset$ *be a convex open set and* $f : \mathsf{D} \to \mathbb{R}$. *If* $f$ *is differentiable at* $\mathsf{D}$ *and* $\nabla f$ *is differentiable at* $\mathsf{D}$, *then,* $\nabla^2 f$ *exists on* $\mathsf{D}$, $f$ *is twice differentiable at* $\mathsf{D}$ *with*

$$H_f(x) \quad = \quad \frac{1}{2}\nabla^2 f(x) + \frac{1}{2}\left(\nabla^2 f(x)\right)^\top$$

*and*

$$f \text{ is convex} \quad \Leftrightarrow \quad H_f(x) \text{ is positively semidefinite for all } x \in \mathsf{D}. \tag{2.31}$$

**Proof:** According to Theorem 2.46 we have to verify convexity of all one-dimensional projections of $f$.
Take $x \in \mathsf{D}$, $s \in \mathbb{R}^n$ and consider function $\varphi_{x,s}$.
According to Lemma 2.24, we have

$$\varphi''_{x,s}(t) \quad = \quad \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 f}{\partial x_i \partial x_j}(x+ts) s_i s_j = s^\top \nabla^2 f(x+ts)\ s.$$

Hence,

$$\begin{aligned} \varphi_{x,s} \text{ is convex} \quad &\Longleftrightarrow \quad \forall\, t \in \mathsf{D}_{x,s} \quad \text{we have } s^\top \nabla^2 f(x+ts)\ s \geq 0 \\ &\Longleftrightarrow \quad \forall\, t \in \mathsf{D}_{x,s} \quad \text{we have } s^\top H_f(x+ts)\ s \geq 0. \end{aligned}$$

Finally, function $f$ is convex if and only if $H_f(x)$ is positively semidefinite for all $x \in \mathsf{D}$.

Q.E.D.

Let us recall notion of positively semidefinite matrix and its equivalent definitions.

**Lemma 2.50** *For symmetric matrix $A \in \mathbb{R}^{n\times n}$ the following is equivalent:*

- *$A$ is positively semidefinite.*
- *For all $x \in \mathbb{R}^n$ we have $x^\top Ax \geq 0$.*
- *All eigenvalues of matrix $A$ are nonnegative.*
- *Determinants of all principle minors of matrix $A$ are nonnegative, i.e.*

$$\forall\, I \subset \{1, 2, \ldots, n\},\ I \neq \emptyset \quad \text{we have} \quad \det\left(A_{i,j}, i, j \in I\right) \geq 0.$$

- *There are a regular matrix $Q$ and a diagonal matrix $\Lambda$ with nonnegative members on its diagonal such that $A = Q^\top \Lambda Q$.*

**Lemma 2.51** *For symmetric matrix $A \in \mathbb{R}^{n\times n}$ the following is equivalent:*

- *$A$ is positively definite.*
- *For all $x \in \mathbb{R}^n$, $x \neq \mathbf{0}$ we have $x^\top Ax > 0$.*
- *All eigenvalues of matrix $A$ are positive.*
- *Determinants of all principle corner minors of matrix $A$ are positive, i.e.*

$$\forall\, k \in \{1, 2, \ldots, n\} \quad \text{we have} \quad \det\left(A_{i,j}, i, j \in \{1, 2, \ldots, k\}\right) > 0.$$

- *There are a regular matrix $Q$ and a diagonal matrix $\Lambda$ with positive members on its diagonal such that $A = Q^\top \Lambda Q$.*

Let us recall expression of form $A = Q^\top \Lambda Q$ means transformation of a quadratic form to its polar base. For that there is an effective algorithm known as Gauss-Jordan elimination. In fact, it is Gauss elimination applied to rows and columns at ones, i.e. each elementary transformation applied to rows must be applied to columns, too.

Let us recall smoothness of convex functions.

**Theorem 2.52**: *Let $\mathsf{D} \subset \mathbb{R}^n$, $\mathsf{D} \neq \emptyset$ be a convex set and $f : \mathsf{D} \to \mathbb{R}$ be a convex function. Then, $f$ is continuous on* $\operatorname{rint}(\mathsf{D})$.

**Proof:** Theorem is a reformulation of Theorem 2.33.

Q.E.D.

**Theorem 2.53**: *Let $\mathsf{D} \subset \mathbb{R}^n$, $\mathsf{D} \neq \emptyset$ be an open convex set and $f : \mathsf{D} \to \mathbb{R}$ be a function. If $f$ is differentiable at $\mathsf{D}$, then*

$$f \text{ is convex} \iff \forall\, x, y \in \mathsf{D} \text{ we have } f(x) - f(y) \geq \langle \nabla f(y), x - y \rangle. \tag{2.32}$$

**Proof:**

1. Let $f$ be convex.
   Take $x, y \in \mathsf{D}$ and denote $h = x - y$, $\varphi(\mu) = f(y + \mu h)$.
   Then, $\varphi$ is a differentiable convex function on $\mathsf{D}_{y,h}$ and its derivatives is

$$\varphi'(\mu) \quad = \quad \langle \nabla f(y + \mu h), h \rangle .$$

   According to "Theorem on mean value" there is $\theta \in (0,1)$ such that

$$\begin{aligned} f(x) - f(y) &= \varphi(1) - \varphi(0) = \varphi'(\theta) \\ &\geq \varphi'(0) = \langle \nabla f(y), h \rangle = \langle \nabla f(y), x - y \rangle , \end{aligned}$$

   since derivatives of a convex differentiable function is nondecreasing.

2. Let $\forall\, x, y \in \mathsf{D}$ we have $f(x) - f(y) \geq \langle \nabla f(y), x - y \rangle$
   Take $y, z \in \mathsf{D}$, $\lambda \in (0,1)$ and denote $x = \lambda y + (1-\lambda) z$.
   According to assumption we have:

$$\begin{aligned} f(y) - f(x) \geq \langle \nabla f(x), y - x \rangle , \\ f(z) - f(x) \geq \langle \nabla f(x), z - x \rangle . \end{aligned}$$

   Hence,

$$\begin{aligned} \lambda f(y) + (1-\lambda) f(z) &\geq f(x) + \langle \nabla f(x), \lambda(y - x) + (1-\lambda)(z - x) \rangle \\ &= f(x) = f(\lambda y + (1-\lambda) z) . \end{aligned}$$

   According to Theorem 2.31, $f$ is convex.

Q.E.D.

This property is generalized by a subgradient and a subdifferential.

**Definition 2.54** *Let $\mathsf{D} \subset \mathbb{R}^n$, $\mathsf{D} \neq \emptyset$ be a set and $f : \mathsf{D} \to \mathbb{R}$ be a function. We say, $f$ possesses at $x \in \mathsf{D}$ a subgradient $a \in \mathbb{R}^n$ (cz. subgradient), if we have*

$$f(y) - f(x) \geq \langle a, y - x \rangle \text{ for all } y \in \mathsf{D}. \tag{2.33}$$

*Set of all subgradients at $x$ will be called a subdifferential of $f$ at $x$ (cz. subdiferenciál) and will be denoted by $\partial f(x)$.*

Using subdifferential we can equivalently rewrite definition of global minimum.

**Theorem 2.55**: *Let $\mathsf{D} \subset \mathbb{R}^n$, $x^* \in \mathsf{D}$ and $f : \mathsf{D} \to \mathbb{R}$ be a function. Then, $x^*$ is a global minimum of $f$ on $\mathsf{D}$ if and only if $\mathbf{0} \in \partial f(x^*)$.*

**Proof:** Statement is a trivial consequence of subgradient definition, since

$$\mathbf{0} \in \partial f(x^*) \quad \Longleftrightarrow \quad \forall\, x \in \mathsf{D}\ f(x) \geq f(x^*) .$$

Q.E.D.

The rewriting can be understand as a generalization of methodology to determine local minima seeking for zero derivative. Unfortunately, it is a rewriting having no practical importance. Nothing new is received by this idea.

Subgradient and subdifferential are helpful tools for describing local minima of a convex function; as we will see at Chapter 3.

**Lemma 2.56** *If $\mathcal{G} \subset \mathbb{R}^n$ is nonempty open convex set, $f : \mathcal{G} \to \mathbb{R}$ is a convex function and $y \in \mathcal{G}$. If $f$ possesses a gradient at $y$, then $\partial f(y) = \{\nabla f(y)\}$.*

**Proof:** Take $\eta \in \partial f(y)$ and $i \in \{1, 2, \ldots, n\}$.
For sufficiently small $\lambda > 0$, we have $y + \lambda \mathsf{e}_{i:n}, y - \lambda \mathsf{e}_{i:n} \in \mathcal{G}$. Therefore using convexity of $f$, we are receiving bounds

$$\begin{aligned} f(y + \lambda \mathsf{e}_{i:n}) - f(y) &\geq \langle \eta, \lambda \mathsf{e}_{i:n} \rangle = \lambda \eta_i, \\ f(y - \lambda \mathsf{e}_{i:n}) - f(y) &\geq \langle \eta, -\lambda \mathsf{e}_{i:n} \rangle = -\lambda \eta_i. \end{aligned}$$

Dividing by $\lambda$ and letting $\lambda \to 0+$, we find

$$\begin{aligned} \frac{\partial f}{\partial x_i}(y) &\geq \eta_i, \\ -\frac{\partial f}{\partial x_i}(y) &\geq -\eta_i. \end{aligned}$$

Consequently, $\eta = \nabla f(y)$ for each $\eta \in \partial f(y)$.
That is $\partial f(y) = \{\nabla f(y)\}$.

Q.E.D.

**Lemma 2.57** *Let $\mathcal{G} \subset \mathbb{R}^n$ be a nonempty open convex set, $f : \mathcal{G} \to \mathbb{R}$ be a convex function and $y \in \mathcal{G}$. If $\partial f(y)$ is a one-point set, then $f$ is differentiable at $y$ and $\partial f(y) = \{\nabla f(y)\}$.*

**Proof:** Theorem from mathematical analysis.

Q.E.D.

Previous observations can be summed up in a lemma.

**Lemma 2.58** *Let $\mathcal{G} \subset \mathbb{R}^n$ be a nonempty open convex set, $f : \mathcal{G} \to \mathbb{R}$ be a convex function and $y \in \mathcal{G}$. Hence, the following is equivalent:*

1. *$f$ is differentiable at $y$ and $\partial f(y) = \{\nabla f(y)\}$.*
2. *$\partial f(y)$ is a one-point set.*
3. *$f$ possesses a gradient at $y$.*

Results on separation of convex bodies have consequences for convex function.

**Theorem 2.59**: *Let $\mathsf{D} \subset \mathbb{R}^n$ be a nonempty convex set and $f : \mathsf{D} \to \mathbb{R}$ be a convex function. Then, $\partial f(x) \neq \emptyset$ for each $x \in \mathrm{rint}(\mathsf{D})$.*

**Proof:** Without any loss of generality we can assume $\text{int}(\mathsf{D}) \neq \emptyset$.
Take $x \in \text{int}(\mathsf{D})$.
Then, $(x, f(x)) \in \partial(\mathsf{epi}(f))$ and according to Theorem 1.33 there are $\alpha \in \mathbb{R}^n$ and $\beta \in \mathbb{R}$ such that $\begin{pmatrix} \alpha \\ \beta \end{pmatrix} \neq \mathbf{0}$ and for all $(y, \eta) \in \mathsf{epi}(f)$ we have

$$\langle \alpha, y \rangle + \beta\eta \geq \langle \alpha, x \rangle + \beta f(x).$$

Number $\eta$ can be arbitrary large, therefore, $\beta \geq \mathbf{0}$.

1) Assume $\beta = \mathbf{0}$.
   Since $x \in \text{int}(\mathsf{D})$, there is $\delta > 0$ such that $\mathcal{U}_\delta(x) \subset \mathsf{D}$.
   Therefore, for all $y \in \mathcal{U}_\delta(x)$ we have $\langle \alpha, y \rangle \geq \langle \alpha, x \rangle$.
   According to Lemma 1.43, $\alpha = \mathbf{0}$.
   Hence, $\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \mathbf{0}$ which is a contradiction, because the vector must not be the origin.

2) Assume $\beta > \mathbf{0}$.
   Consequently, for all $(y, \eta) \in \mathsf{epi}(f)$ we have

$$\left\langle \frac{1}{\beta}\,\alpha, y \right\rangle + \eta \geq \left\langle \frac{1}{\beta}\,\alpha, x \right\rangle + f(x).$$

   Therefore, for all $y \in \mathsf{Dom}(f)$ we have

$$f(y) - f(x) \geq \left\langle \frac{1}{\beta}\,\alpha, x - y \right\rangle = \left\langle -\frac{1}{\beta}\,\alpha, y - x \right\rangle.$$

We have found $\beta > \mathbf{0}$ and $-\frac{1}{\beta}\,\alpha \in \partial f(x)$. Theorem is proved.

Q.E.D.

Equivalent description of a convex function using non-emptiness of subdifferentials is in power if function definition region is an open set.

**Theorem 2.60**: *Let $\mathsf{D} \subset \mathbb{R}^n$ be an open convex set and $f : \mathsf{D} \to \mathbb{R}$. Then, $f$ is a convex function if and only if $\partial f(x) \neq \emptyset$ for each $x \in \mathsf{D}$.*

**Proof:**

1. According to Theorem 2.59, $\partial f(x) \neq \emptyset$ for each $x \in \mathsf{D}$.
2. Assume $\partial f(x) \neq \emptyset$ for each $x \in \mathsf{D}$.
   Take $x, y \in \mathsf{D}$ and $0 < \lambda < 1$.
   Then $z = \lambda x + (1 - \lambda)y \in \mathsf{D}$, since $\mathsf{D}$ is a convex set.
   Take $\alpha \in \partial f(z)$, which exists according to our assumption.
   Definition of subgradient is giving to us

$$f(x) - f(z) \geq \langle \alpha, x - z \rangle,$$
$$f(y) - f(z) \geq \langle \alpha, y - z \rangle.$$

Therefore,

$$\lambda(f(x) - f(z)) + (1-\lambda)(f(y) - f(z)) \geq \lambda \langle \alpha, x - z \rangle + (1-\lambda) \langle \alpha, y - z \rangle .$$

Hence,

$$\lambda f(x) + (1-\lambda)f(y) - f(z) \geq \langle \alpha, \lambda x + (1-\lambda)y - z \rangle = 0.$$

We have shown

$$\lambda f(x) + (1-\lambda)f(y) \geq f(\lambda x + (1-\lambda)y).$$

Thus, $f$ is convex according to Theorem 2.31.

Q.E.D.

For a continuous function, the characterization is also in power.

**Theorem 2.61**: *Let $\mathsf{D} \subset \mathbb{R}^n$ be a convex set and $f : \mathsf{D} \to \mathbb{R}$ be a continuous function. Then, $f$ is a convex function if and only if $\partial f(x) \neq \emptyset$ for each $x \in \mathrm{rint}(\mathsf{D})$.*

**Proof:** Accordingly to Theorem 2.59, the condition is fulfilled for a convex function. We have to show the opposite implication, only.

1. Accordingly to Theorem 2.60, $f : \mathrm{rint}(\mathsf{D}) \to \mathbb{R}$ is convex.
2. Take $x, y \in \mathsf{D}$ and $0 < \lambda < 1$.
   Since $\mathsf{D}$ is convex, we have $\mathsf{D} \subset \mathrm{clo}(\mathrm{rint}(\mathsf{D}))$.
   Then, there are sequences $x_k, y_k \in \mathrm{rint}(\mathsf{D})$ such that $x_k \to x$ and $y_k \to y$.
   For each $k \in \mathbb{N}$, we have

   $$\lambda f(x_k) + (1-\lambda)f(y_k) \geq f(\lambda x_k + (1-\lambda)y_k).$$

   After limit passage $k \to +\infty$ and using continuity of $f$ on $\mathsf{D}$, we receive

   $$\lambda f(x) + (1-\lambda)f(y) \geq f(\lambda x + (1-\lambda)y).$$

We have proved $f$ is convex.

Q.E.D.

### 2.3.4 Vector valued convex functions

In this section we consider functions defined on a finite dimensional Euclidean space with values in a Cartesian product of finite number of extended real lines, i.e. $f : \mathbb{R}^n \to (\mathbb{R}^*)^m$.

**Definition 2.62** *For a function $f : \mathbb{R}^n \to (\mathbb{R}^*)^m$, we define its epigraph (cz. epigraf)*

$$\mathsf{epi}(f) = \left\{ \begin{pmatrix} x \\ \eta \end{pmatrix} : f(x) \leq \eta,\ x \in \mathbb{R}^n,\ \eta \in \mathbb{R}^m \right\}, \tag{2.34}$$

*domain (cz. doména) and weak domain (cz. slabá doména)*

$$\mathsf{Dom}(f) = \{ x : f(x) < +\infty,\ x \in \mathbb{R}^n \}, \tag{2.35}$$

$$\mathsf{WDom}(f) = \{ x : f_i(x) < +\infty \text{ for some } i \in \{1, 2, \ldots, m\},\ x \in \mathbb{R}^n \}. \tag{2.36}$$

**Definition 2.63** *A function $f : \mathbb{R}^n \to (\mathbb{R}^*)^m$ is called* monotone *(cz. monotónní), if $f(x) \le f(y)$ whenever $x \le y$.*

**Definition 2.64** *A function $f : \mathbb{R}^n \to (\mathbb{R}^*)^m$ is* convex *(cz. konvexní), if $\mathsf{epi}(f)$ is a convex set and $\mathsf{WDom}(f) = \mathsf{Dom}(f)$.*

Convexity of a function can be equivalently explained.

**Lemma 2.65** *A function $f : \mathbb{R}^n \to (\mathbb{R}^*)^m$ is convex if and only if $\mathsf{Dom}(f_1) = \mathsf{Dom}(f_2) = \cdots = \mathsf{Dom}(f_m) = \mathsf{Dom}(f)$, $\mathsf{Dom}(f)$ is a convex set and $f_i$ is a convex function for each $i \in \{1, 2, \ldots, m\}$.*

**Theorem 2.66**: *Function $f : \mathbb{R}^n \to (\mathbb{R}^*)^m$ is convex if and only if $\mathsf{Dom}(f_1) = \mathsf{Dom}(f_2) = \cdots = \mathsf{Dom}(f_m) = \mathsf{Dom}(f)$, $\mathsf{Dom}(f)$ is a convex set and for all $x, y \in \mathsf{Dom}(f)$ and $0 < \lambda < 1$ we have*

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y). \tag{2.37}$$

**Proof:** Statement is a consequence of Theorem 2.31 and Lemma 2.65.

Q.E.D.

Composition of functions preserves convexity of functions under some circumstances.

**Lemma 2.67** *If $\mathcal{X} \subset \mathbb{R}^n$ is a convex set, $h : \mathbb{R}^n \to \mathbb{R}^m$ is an affine function and $g : h(\mathcal{X}) \to \mathbb{R}$ is a convex function, then, $f : \mathcal{X} \to \mathbb{R} : x \mapsto g(h(x))$ is a convex function.*

**Proof:** Assumptions are corectly formulated, since $h(\mathcal{X})$ is a convex set, because $\mathcal{X}$ is a convex set and $h : \mathbb{R}^n \to \mathbb{R}^m$ is an affine function.
For $x, y \in \mathcal{X}$ and $\lambda \in (0, 1)$ we have

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &= g(h(\lambda x + (1-\lambda)y)) \\ &= g(\lambda h(x) + (1-\lambda) h(y)) \\ &\le \lambda g(h(x)) + (1-\lambda) g(h(y)) \\ &= \lambda f(x) + (1-\lambda) f(y). \end{aligned}$$

Q.E.D.

**Lemma 2.68** *If $\mathcal{X} \subset \mathbb{R}^n$ is a convex set, $h : \mathcal{X} \to \mathbb{R}^m$ is a convex function and $g : \mathrm{conv}(h(\mathcal{X})) \to \mathbb{R}$ is a monotone convex function, then, $f : \mathcal{X} \to \mathbb{R} : x \mapsto g(h(x))$ is a convex function.*

**Proof:** For $x, y \in \mathcal{X}$ and $\lambda \in (0, 1)$ we can estimate

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &= g(h(\lambda x + (1-\lambda)y)) \\ &\le g(\lambda h(x) + (1-\lambda) h(y)) \\ &\le \lambda g(h(x)) + (1-\lambda) g(h(y)) \\ &= \lambda f(x) + (1-\lambda) f(y). \end{aligned}$$

Q.E.D.

### 2.3.5 Generalization of convex functions

Now, we introduce generalizations of convex functions which are useful for optimization.

**Definition 2.69** *Let* $\mathsf{S} \subset \mathsf{D} \subset \mathbb{R}^n$*,* $\mathsf{S}$ *be a nonempty open set,* $x \in \mathsf{S}$ *and* $f : \mathsf{D} \to \mathbb{R}$ *be differentiable at* $x$*.*

1. *We say, $f$ is pseudoconvex at $x$ with respect to $\mathsf{S}$ (cz. pseudokonvexní v bodě $x$ vzhledem k $\mathsf{S}$), if*
$$\forall\, y \in \mathsf{S},\ \langle \nabla f(x), y - x \rangle \geq 0 \text{ we have } f(y) \geq f(x)\,.$$
2. *We say, $f$ is strictly pseudoconvex at $x$ with respect to $\mathsf{S}$ (cz. striktně pseudokonvexní v bodě $x$ vzhledem k $\mathsf{S}$), if*
$$\forall\, y \in \mathsf{S},\ y \neq x,\ \langle \nabla f(x), y - x \rangle \geq 0 \text{ we have } f(y) > f(x)\,.$$
3. *We say, $f$ is pseudoconcave at $x$ with respect to $\mathsf{S}$ (cz. pseudokonkávní v bodě $x$ vzhledem k $\mathsf{S}$), if $-f$ is pseudoconvex at $x$ with respect to $\mathsf{S}$.*
4. *We say, $f$ is strictly pseudoconcave at $x$ with respect to $\mathsf{S}$ (cz. striktně pseudokonkávní v bodě $x$ vzhledem k $\mathsf{S}$), if $-f$ is strictly pseudoconvex at $x$ with respect to $\mathsf{S}$.*

**Definition 2.70** *Let* $\mathsf{S} \subset \mathsf{D} \subset \mathbb{R}^n$*,* $\mathsf{S}$ *be a nonempty open set and* $f : \mathsf{D} \to \mathbb{R}$ *be differentiable at* $\mathsf{S}$*.*

1. *We say, $f$ is pseudoconvex on $\mathsf{S}$ (cz. pseudokonvexní na $\mathsf{S}$), if*
$$\forall\, x, y \in \mathsf{S},\ \langle \nabla f(x), y - x \rangle \geq 0 \text{ we have } f(y) \geq f(x)\,.$$
2. *We say, $f$ is strictly pseudoconvex on $\mathsf{S}$ (cz. striktně pseudokonvexní na $\mathsf{S}$), if*
$$\forall\, x, y \in \mathsf{S},\ x \neq y,\ \langle \nabla f(x), y - x \rangle \geq 0 \text{ we have } f(y) > f(x)\,.$$
3. *We say, $f$ is pseudoconcave on $\mathsf{S}$ (cz. pseudokonkávní na $\mathsf{S}$), if $-f$ is pseudoconvex on $\mathsf{S}$.*
4. *We say, $f$ is strictly pseudoconcave on $\mathsf{S}$ (cz. striktně pseudokonkávní na $\mathsf{S}$), if $-f$ is strictly pseudoconvex on $\mathsf{S}$.*

**Definition 2.71** *Let* $\mathsf{S} \subset \mathsf{D} \subset \mathbb{R}^n$*,* $\mathsf{S}$ *be a nonempty convex set,* $f : \mathsf{D} \to \mathbb{R}$ *and* $x \in \mathsf{S}$*.*

1. *We say, $f$ is quasiconvex at $x$ with respect to $\mathsf{S}$ (cz. quasikonvexní v bodě $x$ vzhledem k $\mathsf{S}$) if*
$$\forall\, y \in \mathsf{S},\ 0 < \lambda < 1 \text{ we have } f(\lambda x + (1-\lambda)y) \leq \max\{f(x), f(y)\}.$$
2. *We say, $f$ is quasiconvex on $\mathsf{S}$ (cz. quasikonvexní na $\mathsf{S}$), if*
$$\forall\, y, z \in \mathsf{S},\ 0 < \lambda < 1 \text{ we have } f(\lambda y + (1-\lambda)z) \leq \max\{f(y), f(z)\}.$$

**Lemma 2.72** *Let* $\mathsf{D} \subset \mathbb{R}^n$*,* $\mathsf{D} \neq \emptyset$ *be an open convex set and* $h : \mathsf{D} \to \mathbb{R}$ *be quasiconvex on* $\mathsf{D}$*. Then, (lower) level sets of $h$ (cz. (dolní) úrovňové množiny) are convex , i.e.* $\forall\, \Delta \in \mathbb{R}$ *sets* $\mathsf{lev}_{\leq \Delta} h = \{x \in \mathsf{D} : f(x) \leq \Delta\}$ *and* $\mathsf{lev}_{< \Delta} h = \{x \in \mathsf{D} : f(x) < \Delta\}$ *are convex.*

**Proof:** The statement is a direct consequence of Definition 2.71.

Q.E.D.

**Lemma 2.73** *Let $\mathsf{S} \subset \mathsf{D} \subset \mathbb{R}^n$, $\mathsf{S}$ be a nonempty open convex set and $h : \mathsf{D} \to \mathbb{R}$ be differentiable at $\mathsf{S}$ and pseudoconvex on $\mathsf{S}$. Then, $h$ is quasiconvex on $\mathsf{S}$.*

**Proof:** Take $x, y \in \mathsf{S}$, $0 < \lambda < 1$ and denote $z = \lambda x + (1-\lambda)y$.
Assume $h(z) > h(x)$.
Function $h$ is pseudoconvex, therefore, $\langle \nabla_x h(z), x - z \rangle < 0$.
Consider,

$$\begin{aligned} x - z &= x - (\lambda x + (1-\lambda)y) = (1-\lambda)(x-y), \\ y - z &= y - (\lambda x + (1-\lambda)y) = \lambda(y-x) = -\lambda(x-y). \end{aligned}$$

Hence,

$$\begin{aligned} y - z &= -\frac{\lambda}{(1-\lambda)}(x-z), \\ \langle \nabla_x h(z), y - z \rangle &= -\frac{\lambda}{(1-\lambda)} \langle \nabla_x h(z), x - z \rangle > 0\,. \end{aligned}$$

Function $h$ is pseudoconvex, therefore, $h(z) \le h(y)$.
Finally, $h(z) \le \max\{h(x), h(y)\}$ and $h$ is quasiconvex on $\mathsf{S}$.

Q.E.D.

**Lemma 2.74** *Let $\mathsf{D} \subset \mathbb{R}^n$, $\mathsf{D} \neq \emptyset$ be an open convex set and be differentiable at $\mathsf{S}$ and pseudoconvex on $\mathsf{S}$. Then, (lower) level sets of $h$ (cz. (dolní) úrovňové množiny) are convex , i.e. $\forall \Delta \in \mathbb{R}$ sets $\mathsf{lev}_{\le\Delta} h = \{x \in \mathsf{D} : f(x) \le \Delta\}$ and $\mathsf{lev}_{<\Delta} h = \{x \in \mathsf{D} : f(x) < \Delta\}$ are convex.*

**Proof:** The statement is a direct consequence of Lemma 2.73 and Lemma 2.72.

Q.E.D.

Function pseudoconvex at a point does not have to be quasiconvex at the point; see an example.

**Example 2.75:** Consider $\mathsf{D} = (-2, 1)$ and $f(x) = -x^2$. Function $f$ is pseudoconvex at $-1$, but it is not quasiconvex at $-1$.

△

# Chapter 3

# Mathematical programming

In this chapter we introduce a proper course how to seek for local minimizers of a mathematical program. The procedure ideas are taken from the book [1].

Mathematical program means a task

$$\min_{x\in\mathbf{M}} f(x), \tag{3.1}$$

where $\mathbf{M} \subset \mathcal{G} \subset \mathbb{R}^n$, $\mathbf{M} \neq \emptyset$, $\mathcal{G}$ is open and $f : \mathcal{G} \to \mathbb{R}$.

Our aim is to introduce necessary and sufficient condition for a given point to be a local minimizer. Step by step we will consider mathematical programs with different description.

We divide conditions to conditions of the first order, i.e. conditions based on gradient or on its generalization, and conditions of the second order, i.e. conditions based on second partial derivatives or its generalization.

We will denote:

- FONC (First Order Necessary Condition) (cz. nutná podmínka optimality prvého řádu),
- FOSC (First Order Sufficient Condition) (cz. postačující podmínka optimality prvého řádu),
- SONC (Second Order Necessary Condition) (cz. nutná podmínka optimality druhého řádu),
- SOSC (Second Order Sufficient Condition) (cz. postačující podmínka optimality druhého řádu).

## 3.1 Convex objective function

We start with convex programming (CP), i.e. objective function $f$ is a convex function and $\mathbf{M}$ is a convex set. For this case we possess a full description by a first order condition.

**Theorem 3.1** (FONC+FOSC for CP): *Let $\mathbf{M} \subset \mathcal{G} \subset \mathbb{R}^n$, $\mathbf{M}$ be a convex set, $\mathcal{G}$ be an open convex set, $\widetilde{x} \in \mathbf{M}$ and $f : \mathcal{G} \to \mathbb{R}$ be a convex function. Then, $\widetilde{x}$ is a global minimum of $f$ on $\mathbf{M}$ if and only if there is $\eta \in \partial f(\widetilde{x})$ such that $\forall\, x \in \mathbf{M}$ we have $\langle \eta, x - \widetilde{x} \rangle \geq 0$.*

**Proof:** We have to show both implications.

1. Let $\eta \in \partial f(\widetilde{x})$ and $\forall\, x \in \mathbf{M}$ we have $\langle \eta, x - \widetilde{x} \rangle \geq 0$.

   According to definition of subgradient, Definition 2.54, we have

$$\forall\, x \in \mathcal{G} \qquad f(x) \geq f(\widetilde{x}) + \langle \eta, x - \widetilde{x} \rangle .$$

Using the condition for $x \in \mathbf{M}$, we receive

$$f(x) \geq f(\widetilde{x}) + \langle \eta, x - \widetilde{x} \rangle \geq f(\widetilde{x}).$$

We have verified $\widetilde{x}$ is a global minimum of $f$ on $\mathbf{M}$.

2. Let $\widetilde{x}$ be a global minimum of $f$ on $\mathbf{M}$.
   Consider two sets

$$\begin{aligned} \Psi &= \{(x,y) \,:\, x + \widetilde{x} \in \mathcal{G},\ y > f(x+\widetilde{x}) - f(\widetilde{x})\}, \\ \Gamma &= \{(x,y) \,:\, x + \widetilde{x} \in \mathbf{M},\ y \leq 0\}. \end{aligned}$$

The sets can be also expressed as

$$\begin{aligned} \Psi &= (\mathsf{epi}(f) \setminus \mathsf{graph}(f)) - (\widetilde{x}, f(\widetilde{x})), \\ \Gamma &= (\mathbf{M} - \widetilde{x}) \times (-\infty, 0]. \end{aligned}$$

Sets are convex and $\Psi \cap \Gamma = \emptyset$.
According to Theorem 1.42, $\Psi$, $\Gamma$ can be properly separated.
Hence, there are $\alpha \in \mathbb{R}^n$, $\beta \in \mathbb{R}$, $(\alpha, \beta) \neq 0$ such that

$$\inf\{\langle \alpha, x \rangle + \beta y \,:\, (x,y) \in \Psi\} \geq \sup\{\langle \alpha, x \rangle + \beta y \,:\, (x,y) \in \Gamma\}.$$

We know $(\mathbf{0}, 0) \in \Gamma$ and $(\mathbf{0}, 2^{-k}) \in \Psi$ for each $k \in \mathbb{N}$.
Consequently,

$$\inf\{\langle \alpha, x \rangle + \beta y \,:\, (x,y) \in \Psi\} = 0 = \sup\{\langle \alpha, x \rangle + \beta y \,:\, (x,y) \in \Gamma\}.$$

Realize about a sign of $\beta$.

(a) Assume $\beta < 0$.
Then, we receive a contradiction

$$\inf\{\langle \alpha, x \rangle + \beta y \,:\, (x,y) \in \Psi\} = -\infty,$$

since $(\mathbf{0}, y) \in \Psi$ for any $y > 0$.

(b) Assume $\beta = 0$.
Hence, $\langle \alpha, x \rangle \geq 0$ for each $x \in \mathcal{G} - \widetilde{x}$,
since $(x, f(x+\widetilde{x}) - f(\widetilde{x}) + 1) \in \Psi$.
We know $\widetilde{x} \in \mathbf{M} \subset \mathcal{G}$ and $\mathcal{G} \subset \mathbb{R}^n$ is open, therefore, $\mathcal{G} - \widetilde{x}$ is open and $\mathbf{0} \in \mathcal{G} - \widetilde{x}$.
According to Lemma 1.43, $\alpha = \mathbf{0}$.
That is a contradiction with property $(\alpha, \beta) \neq 0$.

We found $\beta > 0$. Then,

$$\begin{aligned} &\langle \alpha, x \rangle + \beta y \geq 0 \quad \text{for each } x \in \mathcal{G} - \widetilde{x},\ y > f(x+\widetilde{x}) - f(\widetilde{x}), \\ &\langle \alpha, \xi \rangle + \beta \eta \leq 0 \quad \text{for each } \xi \in \mathbf{M} - \widetilde{x},\ \eta \leq \mathbf{0}. \end{aligned}$$

Hence, letting $y \to f(x+\widetilde{x}) - f(\widetilde{x})$ and setting $\eta = \mathbf{0}$ we receive

$$\begin{aligned} &\langle \alpha, x \rangle + \beta (f(x+\widetilde{x}) - f(\widetilde{x})) \geq 0 \quad \text{for each } x \in \mathcal{G} - \widetilde{x}, \\ &\langle \alpha, \xi \rangle \leq 0 \quad \text{for each } \xi \in \mathbf{M} - \widetilde{x}. \end{aligned}$$

Set $\gamma = -\frac{1}{\beta}\,\alpha$ and shift points back; i.e. $x + \widetilde{x} \rightsquigarrow x$,

$$-\left\langle\, \gamma, x - \widetilde{x} \,\right\rangle + f\left(x\right) - f\left(\widetilde{x}\right) \geq 0 \quad \text{for each } x \in \mathcal{G},$$
$$-\left\langle\, \gamma, x - \widetilde{x} \,\right\rangle \leq 0 \quad \text{for each } x \in \mathbf{M}.$$

Consequently,

$$f\left(x\right) \geq f\left(\widetilde{x}\right) + \left\langle\, \gamma, x - \widetilde{x} \,\right\rangle \quad \text{for each } x \in \mathcal{G},$$
$$\left\langle\, \gamma, x - \widetilde{x} \,\right\rangle \geq 0 \quad \text{for each } x \in \mathbf{M}.$$

We have found $\gamma \in \partial f\left(\widetilde{x}\right)$ with required property.

Q.E.D.

The condition becomes simpler for points from interior of $\mathbf{M}$.

**Lemma 3.2** *If $\mathbf{M} \subset \mathcal{G} \subset \mathbb{R}^n$, $\mathbf{M}$ is a convex set, $\mathcal{G}$ is open convex, $\widetilde{x} \in \operatorname{int}\left(\mathbf{M}\right)$ and $f : \mathcal{G} \to \mathbb{R}$ is a convex function, then*

$$\widetilde{x} \textit{ is a global minimum of } f \textit{ on } \mathbf{M} \iff \mathbf{0} \in \partial f\left(\widetilde{x}\right).$$

**Proof:** According to Theorem 3.1, $\widetilde{x}$ is a global minimum of $f$ on $\mathbf{M}$ if and only if there is $\eta \in \partial f\left(\widetilde{x}\right)$ such that for each $x \in \mathbf{M}$ we have $\left\langle\, \eta, x - \widetilde{x} \,\right\rangle \geq 0$.
Since $\widetilde{x} \in \operatorname{int}\left(\mathbf{M}\right)$, there is $\varepsilon > 0$ such that $\mathcal{U}\left(\widetilde{x}, \varepsilon\right) \subset \mathbf{M}$.
Then for each $x \in \mathcal{U}\left(\widetilde{x}, \varepsilon\right)$, we have $\left\langle\, \eta, x - \widetilde{x} \,\right\rangle \geq 0$.
Hence accordingly to Lemma 1.43, $\eta = \mathbf{0}$.

Q.E.D.

**Lemma 3.3** *Let $\mathbf{M} \subset \mathcal{G} \subset \mathbb{R}^n$, $\mathbf{M}$ be a convex set, $\mathcal{G}$ be an open convex set, $\widetilde{x} \in \mathbf{M}$ and $f : \mathcal{G} \to \mathbb{R}$ be a convex function having a gradient at $\widetilde{x}$. Then*

$$\widetilde{x} \textit{ is a global minimum of } f \textit{ on } \mathbf{M} \iff \forall\, x \in \mathbf{M} \textit{ we have } \left\langle\, \nabla f\left(\widetilde{x}\right), x - \widetilde{x} \,\right\rangle \geq 0.$$

**Proof:** According to lemma 2.58, $\partial f\left(\widetilde{x}\right) = \left\{\nabla f\left(\widetilde{x}\right)\right\}$, since $f$ possesses a gradient at $\widetilde{x}$.
Consequently, the statement follows directly Theorem 3.1.

Q.E.D.

The characterization coincides with classical one for interior points of $\mathbf{M}$.

**Lemma 3.4** *Let $\mathbf{M} \subset \mathcal{G} \subset \mathbb{R}^n$, $\mathbf{M}$ be a convex set, $\mathcal{G}$ be an open convex set, $\widetilde{x} \in \operatorname{int}\left(\mathbf{M}\right)$ and $f : \mathcal{G} \to \mathbb{R}$ be a convex function having a gradient at $\widetilde{x}$. Then,*

$$\widetilde{x} \textit{ is a global minimum of } f \textit{ on } \mathbf{M} \iff \nabla f\left(\widetilde{x}\right) = \mathbf{0}.$$

**Proof:** The statement follows Lemma 3.3.

Q.E.D.

## 3.2 Concave objective function

In this chapter we consider local minima of a concave function on a convex set. Let us formulate a sufficient condition of the first order.

**Theorem 3.5** (FONC for a concave function): *Let* $\mathbf{M} \subset \mathcal{G} \subset \mathbb{R}^n$*,* $\mathbf{M}$ *be a non-empty convex set,* $\mathcal{G}$ *be an open convex set,* $f : \mathcal{G} \to \mathbb{R}$ *be a concave function and* $\widetilde{x} \in \mathbf{M}$*. If* $\widetilde{x}$ *is a local minimum of* $f$ *on* $\mathbf{M}$*, then for each* $\eta \in \mathbb{R}^n$*,* $-\eta \in \partial(-f)(\widetilde{x})$ *we have* $\langle \eta, x - \widetilde{x} \rangle \geq 0 \ \forall x \in \mathbf{M}$*.*

**Proof:** Let $\widetilde{x}$ be a local minimum of $f$ on $\mathbf{M}$.
Hence, there is $\delta > 0$ such that $\widetilde{x}$ is a global minimum of $f$ on $\mathbf{M} \cap \mathcal{U}(\widetilde{x}, \delta)$.
If $x \in \mathbf{M}$, then there is $\lambda > 0$ such that $\widetilde{x} + \lambda(x - \widetilde{x}) \in \mathbf{M} \cap \mathcal{U}(\widetilde{x}, \delta)$.
Therefore, $f(\widetilde{x} + \lambda(x - \widetilde{x})) \geq f(\widetilde{x})$.
Taking $\eta \in \mathbb{R}^n$, $-\eta \in \partial(-f)(\widetilde{x})$, we estimate

$$f(\widetilde{x}) \leq f(\widetilde{x} + \lambda(x - \widetilde{x})) \leq f(\widetilde{x}) + \lambda \langle \eta, x - \widetilde{x} \rangle .$$

Consequently, for each $\eta \in \mathbb{R}^n$, $-\eta \in \partial(-f)(\widetilde{x})$ and $x \in \mathbf{M}$ we have $\langle \eta, x - \widetilde{x} \rangle \geq 0$.

Q.E.D.

Unfortunately, the condition is not sufficient, see the following example.

**Example 3.6:** Consider function $f : \mathbb{R} \to \mathbb{R} : x \mapsto -x^2$ and $\mathbf{M} = [-1, 2]$.
Condition from Theorem 3.5 is fulfilled in points $-1$, $0$, $2$. Points $-1$, $2$ are local minima of $f$ on $\mathbf{M}$. The point $0$ is a global maximum of $f$, but, no local minimum.

△

If a concave function possesses a local minimum in an interior point of the set of feasible solutions, then the function must be flat in its neighborhood.

**Lemma 3.7** *Let* $\mathbf{M} \subset \mathcal{G} \subset \mathbb{R}^n$*,* $\mathbf{M}$ *be a non-empty convex set,* $\mathcal{G}$ *be an open convex set,* $f : \mathcal{G} \to \mathbb{R}$ *be a concave function and* $\widetilde{x} \in \operatorname{int}(\mathbf{M})$*. If* $\widetilde{x}$ *is a local minimum of* $f$ *on* $\mathbf{M}$*, then* $\partial(-f)(\widetilde{x}) = \{\mathbf{0}\}$ *and there is* $\delta > 0$ *such that* $\mathcal{U}(\widetilde{x}, \delta) \subset \mathbf{M}$ *and for all* $x \in \mathcal{U}(\widetilde{x}, \delta)$ *we have* $f(x) = f(\widetilde{x})$*.*

**Proof:** Let $\widetilde{x} \in \operatorname{int}(\mathbf{M})$ be a local minimum of $f$ on $\mathbf{M}$.

1. Take $\eta \in \mathbb{R}^n$, $-\eta \in \partial(-f)(\widetilde{x})$.
   Hence, there is $\delta > 0$ such that $\widetilde{x} - \delta\eta \in \mathbf{M}$.
   According to Theorem 3.5 we have $\langle \eta, -\delta\eta \rangle = -\delta \|\eta\|^2 \geq 0$.
   Finally, $\eta = \mathbf{0}$ and $\partial(-f)(\widetilde{x}) = \{\mathbf{0}\}$.
2. There is $\delta > 0$ such that $\mathcal{U}(\widetilde{x}, \delta) \subset \mathbf{M}$ and for all $x \in \mathcal{U}(\widetilde{x}, \delta)$ we have $f(x) \geq f(\widetilde{x})$.
   $f$ is a concave, therefore, for all $x \in \mathcal{U}(\widetilde{x}, \delta)$ and $\eta \in \partial(-f)(\widetilde{x}) = \{\mathbf{0}\}$ we have $f(\widetilde{x}) \leq f(x) \leq f(\widetilde{x}) + \langle \eta, x - \widetilde{x} \rangle = f(\widetilde{x})$.
   Finally, $f(x) = f(\widetilde{x})$ for all $x \in \mathcal{U}(\widetilde{x}, \delta)$.

Q.E.D.

**Lemma 3.8** *Let $\mathbf{M} \subset \mathcal{G} \subset \mathbb{R}^n$, $\mathbf{M}$ be a non-empty convex set, $\mathcal{G}$ be an open convex set, $f : \mathcal{G} \to \mathbb{R}$ be a concave function, $\widetilde{x} \in \mathbf{M}$ and $f$ possess a gradient at $\widetilde{x}$. If $\widetilde{x}$ is a local minimum of $f$ on $\mathbf{M}$, then*

$$\forall\, x \in \mathbf{M} \text{ we have } \left\langle \nabla f\left(\widetilde{x}\right), x - \widetilde{x} \right\rangle \geq 0.$$

**Proof:** We know $-f$ is a convex function which possesses a gradient at $\widetilde{x}$. According to lemma 2.58, subdifferential of $-f$ at $\widetilde{x}$ contains the gradient of $-f$, only. Consequently, the statement follows directly from Theorem 3.5.

Q.E.D.

**Lemma 3.9** *Let $\mathbf{M} \subset \mathcal{G} \subset \mathbb{R}^n$, $\mathbf{M}$ be a non-empty convex set, $\mathcal{G}$ be an open convex set, $f : \mathcal{G} \to \mathbb{R}$ be a concave function, $\widetilde{x} \in \operatorname{int}(\mathbf{M})$ and $f$ possess a gradient at $\widetilde{x}$. If $\widetilde{x}$ is a local minimum of $f$ on $\mathbf{M}$, then $\nabla f\left(\widetilde{x}\right) = \mathbf{0}$ and there is $\delta > 0$ such that $\mathcal{U}\left(\widetilde{x}, \delta\right) \subset \mathbf{M}$ and for all $x \in \mathcal{U}\left(\widetilde{x}, \delta\right)$ we have $f\left(x\right) = f\left(\widetilde{x}\right)$.*

**Proof:** We know, $-f$ is a convex function differentiable at $\widetilde{x}$. According to lemma 2.58, its subdifferential contains only the gradient. Consequently, the statement follows directly from Theorem 3.5.

Q.E.D.

## 3.3 Open set of feasible solutions

Previous chapter considered convex and concave objective functions for which optimal solutions are well characterized by means of subgradients and subdifferentials.

In this chapter, we will consider free minimum (cz. volné minimum); i.e. a minimum of a function on an open set. Unconstrained minimum is a particullar case of such a program.

For a general function, subdifferentials are typically empty. Therefore, we are dealing with functions differentiable (or twice differentiable) at a given point; for definitions see Section 2.2. Nevertheless, presented sufficient conditions require some kind of convexity for objective function; employed properties are introduced in Section 2.3.5.

Now, we introduce optimality conditions of the first and of the second order for problems with an open set of feasible solutions.

**Theorem 3.10** (FONC for FM): *Let $\mathbf{M} \subset \mathbb{R}^n$ be an open set, $f : \mathbf{M} \to \mathbb{R}$ and $x^* \in \mathbf{M}$. If $x^*$ is a local minimum of $f$ on $\mathbf{M}$ and $f$ possesses a gradient at $x^*$, then $\nabla f\left(x^*\right) = \mathbf{0}$.*

**Proof:** Assume, $\nabla f\left(x^*\right) \neq \mathbf{0}$.
Hence, there is an index $i \in \{1, 2, \ldots, n\}$ such that $\frac{\partial f}{\partial x_i}\left(x^*\right) \neq 0$. Since $\mathbf{M} \subset \mathbb{R}^n$ is an open set, we have:

- If $\frac{\partial f}{\partial x_i}\left(x^*\right) > 0$, then there is $\alpha > 0$ such that $x^* - \alpha \mathsf{e}_{i:n} \in \mathbf{M}$ and $f\left(x^* - \alpha \mathsf{e}_{i:n}\right) < f\left(x^*\right)$.
- If $\frac{\partial f}{\partial x_i}\left(x^*\right) < 0$, then there is $\alpha > 0$ such that $x^* + \alpha \mathsf{e}_{i:n} \in \mathbf{M}$ and $f\left(x^* + \alpha \mathsf{e}_{i:n}\right) < f\left(x^*\right)$.

We found out, $x^*$ cannot be a local minimum of $f$ on $\mathbf{M}$.

Q.E.D.

**Theorem 3.11** (FOSC for FM): *Let $\mathbf{M} \subset \mathbb{R}^n$ be an open set, $f : \mathbf{M} \to \mathbb{R}$, $x^* \in \mathbf{M}$ and $f$ possesses a gradient at $x^*$. If $\nabla f(x^*) = \mathbf{0}$ and there is $\delta > 0$ such that $\mathcal{U}_\delta(x^*) \subset \mathbf{M}$ and $f$ is convex on $\mathcal{U}_\delta(x^*)$, then $x^*$ is a local minimum of $f$ on $\mathbf{M}$.*

**Proof:** The statement is a consequence of Lemma 3.4.

Q.E.D.

**Theorem 3.12** (SONC for FM): *Let $\mathbf{M} \subset \mathbb{R}^n$ be an open set, $f : \mathbf{M} \to \mathbb{R}$ and $x^* \in \mathbf{M}$. If $x^*$ is a local minimum of $f$ on $\mathbf{M}$ and $f$ is twice differentiable at $x^*$, then $\nabla f(x^*) = \mathbf{0}$ and $H_f(x^*)$ is a positively semidefinite matrix.*

**Proof:** Since function $f$ is twice differentiable at $x^*$ (see Definition 2.21), we have an expansion

$$f(x^* + h) = f(x^*) + \langle \nabla f(x^*), h \rangle + \frac{1}{2} \langle h, H_f(x^*) h \rangle + \|h\|^2 R_2(h; f, x^*) \quad \forall h \in \mathbb{R}^n,$$

where $\lim_{h \to \mathbf{0}} R_2(h; f, x^*) = 0$ and $H_f(x^*)$ is a symmetric matrix.
According to Theorem 3.10, $\nabla f(x^*) = \mathbf{0}$. The expansion becomes to be more simple

$$f(x^* + h) = f(x^*) + \frac{1}{2} \langle h, H_f(x^*) h \rangle + \|h\|^2 R_2(h; f, x^*).$$

Multiplying $h$ by $\alpha > 0$, we receive

$$f(x^* + \alpha h) = f(x^*) + \frac{\alpha^2}{2} \langle h, H_f(x^*) h \rangle + \alpha^2 \|h\|^2 R_2(\alpha h; f, x^*).$$

Point $x^*$ is a local minimum of $f$ on $\mathbb{R}^n$, therefore, for $\alpha > 0$ small enough

$$0 \le \frac{f(x^* + \alpha h) - f(x^*)}{\alpha^2} = \frac{1}{2} \langle h, H_f(x^*) h \rangle + \|h\|^2 R_2(\alpha h; f, x^*).$$

We know, $\lim_{\alpha \to 0+} R_2(\alpha h; f, x^*) = 0$.
Letting $\alpha \to 0+$, we receive

$$\forall h \in \mathbb{R}^n, \ h \ne \mathbf{0} : \ 0 \le \langle h, H_f(x^*) h \rangle.$$

According to Definition 2.21, $H_f(x^*)$ is a symmetric matrix.
We have derived, $H_f(x^*)$ is a positively semidefinite matrix,

Q.E.D.

**Theorem 3.13** (SOSC for FM): *Let $\mathbf{M} \subset \mathbb{R}^n$ be an open set, $f : \mathbf{M} \to \mathbb{R}$ and $x^* \in \mathbf{M}$. If $f$ is twice differentiable at $x^*$, $\nabla f(x^*) = \mathbf{0}$ and $H_f(x^*)$ is a positively definite matrix, then $x^*$ is a strict local minimum of $f$ on $\mathbf{M}$.*

**Proof:** Since function $f$ is twice differentiable at $x^*$ (see Definition 2.21), we have an expansion

$$f(x^* + h) = f(x^*) + \langle \nabla f(x^*), h \rangle + \frac{1}{2} \langle h, H_f(x^*) h \rangle + \|h\|^2 R_2(h; f, x^*) \quad \forall h \in \mathbb{R}^n,$$

where $\lim_{h \to \mathbf{0}} R_2(h; f, x^*) = 0$ .

Matrix $H_f(x^*)$ is positively definite. Hence, its eigen values are positive. Let us denote its smallest eigen value by symbol $\lambda$.
Assumption $\nabla f(x^*) = \mathbf{0}$, simplifies the expansion.
For $h$ small enough, we have

$$\begin{aligned}
& f(x^* + h) - f(x^*) = \frac{1}{2}\left\langle h, H_f(x^*)\, h \right\rangle + \|h\|^2 R_2(h; f, x^*) \\
&= \frac{1}{2}\|h\|^2 \left( \left\langle \frac{h}{\|h\|}, H_f(x^*) \frac{h}{\|h\|} \right\rangle + 2R_2(h; f, x^*) \right) \\
&\geq \frac{1}{2}\|h\|^2 \left(\lambda + 2R_2(h; f, x^*)\right) \\
&\geq \frac{1}{4}\|h\|^2 \lambda > 0.
\end{aligned}$$

We detect, $x^*$ is a strict local minimum of $f$ on $\mathbf{M}$.

Q.E.D.

# Chapter 4

# Nonlinear programming

In this chapter, we deal with a nonlinear program (NLP) (cz. úloha nelineárního programování)

$$\min\left\{f(x) : g_j(x) \le 0\,,\ j = 1,\ldots,m\,,\ h_k(x) = 0,\ k = 1,\ldots,q\,,\ x \in \mathbb{R}^n\right\}, \tag{4.1}$$

where functions $f,\ g_j,\ j = 1,\ldots,m,\ h_k,\ k = 1,\ldots,q$ are defined on $\mathbb{R}^n$ and with values in $\mathbb{R}^*$. For simplicity, we denote $\mathsf{J} = \{1,\ldots,m\}$, $\mathsf{K} = \{1,\ldots,q\}$, $g = \begin{pmatrix} g_1 \\ \vdots \\ g_m \end{pmatrix}$, $h = \begin{pmatrix} h_1 \\ \vdots \\ h_q \end{pmatrix}$, and for $L \subset \mathsf{J}$, $\widetilde{L} \subset \mathsf{K}$ $g_L = (g_j, j \in L)$, $h_{\widetilde{L}} = \left(h_k, k \in \widetilde{L}\right)$.

We will introduce properties helping us to identify possible solutions of (4.1). We have to restrict our interest to a convenient set on which all functions are real-valued, eventually, smooth or differentiable.

Set of all feasible solutions of (4.1) is

$$\mathbf{M} = \{x \in \mathbb{R}^n : g_j(x) \le 0\ \forall j \in \mathsf{J}\,,\ h_k(x) = 0\ \forall k \in \mathsf{K}\}\,. \tag{4.2}$$

Main idea for solving nonlinear programs is a relaxation. We divide constraints in two groups $(\mathsf{J}_1, \mathsf{K}_1)$ and $(\mathsf{J}_2, \mathsf{K}_2)$, where $\mathsf{J}_1 \subset \mathsf{J}$, $\mathsf{K}_1 \subset \mathsf{K}$, $\mathsf{J}_2 \subset \mathsf{J}$, $\mathsf{K}_2 \subset \mathsf{K}$, $\mathsf{J}_1 \cap \mathsf{J}_2 = \emptyset$, $\mathsf{J}_1 \cup \mathsf{J}_2 = \mathsf{J}$, $\mathsf{K}_1 \cap \mathsf{K}_2 = \emptyset$, $\mathsf{K}_1 \cup \mathsf{K}_2 = \mathsf{K}$. Violation of constraints with indexes $(\mathsf{J}_1, \mathsf{K}_1)$ will be penalized as a part of our objective function and received objective function will be optimized under constraints with indexes $(\mathsf{J}_2, \mathsf{K}_2)$.

We select a convenient open set $\mathcal{G} \supset \mathbf{M}$ such that all functions $f$, $g_j$, $j \in \mathsf{J}$, $h_k$, $k \in \mathsf{K}$ are real-valued on $\mathcal{G}$ and consider relaxed set of feasible solutions

$$\mathbf{M}_2 = \{x \in \mathbb{R}^n : g_j(x) \le 0\,,\ j \in \mathsf{J}_2\,,\ h_k(x) = 0,\ k \in \mathsf{K}_2\}\,. \tag{4.3}$$

Set $\mathcal{G}$ represents a convenient definition region for all functions appearing in the nonlinear program.

We define Lagrange function (cz. Lagrangeova funkce)

$$\begin{aligned} L(x;u,v) &= \left\langle \begin{pmatrix} 1 \\ u \\ v \end{pmatrix}, \begin{pmatrix} f(x) \\ g_{\mathsf{J}_1}(x) \\ h_{\mathsf{K}_1}(x) \end{pmatrix} \right\rangle = f(x) + \langle u, g_{\mathsf{J}_1}(x) \rangle + \langle v, h_{\mathsf{K}_1}(x) \rangle \\ &= f(x) + \sum_{j \in \mathsf{J}_1} u_j g_j(x) + \sum_{k \in \mathsf{K}_1} v_k h_k(x) \end{aligned} \tag{4.4}$$

for $x \in \mathcal{G} \cap \mathbf{M}_2$, $u \in \mathbb{R}^{\mathsf{J}_1}_{+,0}$, $v \in \mathbb{R}^{\mathsf{K}_1}$.

Our aim is to suggest penalization $u \in \mathbb{R}^{\mathsf{J}_1}_{+,0}$, $v \in \mathbb{R}^{\mathsf{K}_1}$ such that, a solution of relaxed program

$$\min \{L(x; u, v) \, : \, x \in (\mathcal{G} \cap \mathbf{M}_2)\} \tag{4.5}$$

would be a solution of the original problem (4.1).

Part of our lecture will follow this general division of constraints. Rest of this chapter will consider $\mathsf{J}_1 = \mathsf{J}$, $\mathsf{K}_1 = \mathsf{K}$ and later all non-negativity constraints will be included in $\mathsf{J}_2$.

## 4.1 Constraint extrema

At first, let us recall constraint extrema problem (cz. úloha o vázaném extrému) introduced in lectures on mathematical analysis.

The task is to find all local extrema of a given real function $f : \mathbb{R}^n \to \mathbb{R}$ on a given set

$$\mathbf{M} = \{x \in \mathbb{R}^n \, : \, h_k(x) = 0\,, \; k = 1, \ldots, q\}\,, \tag{4.6}$$

provided all functions $f$, $h_k$, $k = 1, \ldots, q$ are differentiable on $\mathbb{R}^n$.

The task is treated by means of a Lagrange function

$$L(x; \lambda) = \left\langle \begin{pmatrix} 1 \\ \lambda \end{pmatrix}, \begin{pmatrix} f(x) \\ h(x) \end{pmatrix} \right\rangle = f(x) + \langle \lambda, h(x) \rangle = f(x) + \sum_{k=1}^{q} \lambda_k h_k(x)\,. \tag{4.7}$$

Set $\mathbf{M}$ can be geometrically interpreted as a manifold in $\mathbb{R}^n$; particularly, as a hyperplane if $h_k, \; k = 1, \ldots, q$ are affine functions.

**Definition 4.1** *Point $x^0 \in \mathbf{M} = \{x \in \mathbb{R}^n \, : \, h_k(x) = 0\,, \; k = 1, \ldots, q\}$ is called a regular point of $\mathbf{M}$ (cz. regulární bod), whenever functions $h_k$, $k = 1, \ldots, q$ are differentiable at $x^0$ and it is fulfilled:*

*i) $h_k(x^0) = 0$ for each $k = 1, \ldots, q$.*

*ii) Gradients $\nabla h_k(x^0)$, $k = 1, \ldots, q$ are linearly independent, i.e.*

$$\textit{matrix} \quad \left( \frac{\partial h_k(x^0)}{\partial x_i} \right)_{i=1,k=1}^{n,q} \quad \textit{is of full rank } q. \tag{4.8}$$

Regularity of a point depends on description of set $\mathbf{M}$.

**Example 4.2:** Consider a set $\mathbf{M} = \left\{ \begin{pmatrix} 0 \\ x_2 \end{pmatrix} \in \mathbb{R}^2 \, : \, x_2 \in \mathbb{R} \right\}$, which is the axis $x_2$.

- Consider description $\mathbf{M} = \{x \in \mathbb{R}^2 \, : \, h(x) = 0\}$, where $h(x_1, x_2) = x_1$.

  Then, each point of $\mathbf{M}$ is regular, since $\nabla h(x) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \neq \mathbf{0}$ for each $x \in \mathbb{R}^2$.

- Consider description $\mathbf{M} = \{x \in \mathbb{R}^2 \, : \, g(x) = 0\}$, where $g(x_1, x_2) = x_1^2$.

  Then, no point of $\mathbf{M}$ is regular, since $\nabla g(x) = \begin{pmatrix} 2x_1 \\ 0 \end{pmatrix}$ for each $x \in \mathbb{R}^2$, and consequently, $\nabla g(x) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ for each $x \in \mathbf{M}$.

△

**Theorem 4.3** (Constraint Extreme): *Let $x^*$ be a local extreme of a function $f$ on a set $\mathbf{M} = \{x \in \mathbb{R}^n \,:\, h_k(x) = 0,\ k = 1, \dots, q\}$, where $f$, $h_k$, $k = 1, \dots, q$ are differentiable at $x^*$.*

*If $x^*$ is a regular point of $\mathbf{M}$, then there is a vector $\lambda^* \in \mathbb{R}^m$ such that*

$$\nabla L(x^*, \lambda^*) = \nabla f(x^*) + \sum_{k=1}^{q} \lambda_k^* \nabla h_k(x^*) = 0.$$

## 4.2 Global minimum

In this section we consider a general partition of constraints into two groups $(\mathsf{J}_1, \mathsf{K}_1)$ and $(\mathsf{J}_2, \mathsf{K}_2)$, where $\mathsf{J}_1 \cup \mathsf{K}_1 \neq \emptyset$. For $\mathsf{J}_1 \cup \mathsf{K}_1 = \emptyset$, the relaxed program would coincide with the original task.

Let us introduce conditions giving global minimum of of (4.1).

**Definition 4.4** *Let $x^* \in \mathbb{R}^n$, $u^* \in \mathbb{R}^{\mathsf{J}_1}$, $v^* \in \mathbb{R}^{\mathsf{K}_1}$ and $\mathcal{G} \supset \mathbf{M}$. We say, that the triplet $(x^*, u^*, v^*)$ fulfills saddle point condition (SPC) (cz. sedlový bod, globální podmínky optimality) for (4.1) on the region $(\mathcal{G} \cap \mathbf{M}_2) \times \mathbb{R}_{+,0}^{\mathsf{J}_1} \times \mathbb{R}^{\mathsf{K}_1}$, whenever*

$$x^* \in \mathcal{G} \cap \mathbf{M}_2\,,\ \ u^* \geq 0$$

*and for all $x \in \mathcal{G} \cap \mathbf{M}_2$, $u \in \mathbb{R}^{\mathsf{J}_1}$, $u \geq \mathbf{0}$, $v \in \mathbb{R}^{\mathsf{K}_1}$ we have*

$$L(x; u^*, v^*) \geq L(x^*; u^*, v^*) \geq L(x^*; u, v)\,. \tag{4.9}$$

*Another often used terminology is that the triplet $(x^*, u^*, v^*)$ is a saddle point of the Lagrange function (4.4) on the region $(\mathcal{G} \cap \mathbf{M}_2) \times \mathbb{R}_{+,0}^{\mathsf{J}_1} \times \mathbb{R}^{\mathsf{K}_1}$.*

**Theorem 4.5** (SPC ⇒ NLP): *Let $x^* \in \mathbb{R}^n$ and $\mathcal{G} \supset \mathbf{M}$ be such that all functions considered in (4.1) attain real values on $\mathcal{G}$. If there are $u^* \in \mathbb{R}^{\mathsf{J}_1}$, $v^* \in \mathbb{R}^{\mathsf{K}_1}$ such that the triplet $(x^*, u^*, v^*)$ fulfills (SPC) for (4.1) on the region $(\mathcal{G} \cap \mathbf{M}_2) \times \mathbb{R}_{+,0}^{\mathsf{J}_1} \times \mathbb{R}^{\mathsf{K}_1}$, then $x^*$ is a global minimum of (4.1).*

*Moreover, a complementarity condition (cz. podmínka komplementarity) for (4.1) is fulfilled:*

$$\langle u^*, g_{\mathsf{J}_1}(x^*) \rangle = \sum_{j \in \mathsf{J}_1} u_j^* g_j(x^*) = 0. \tag{4.10}$$

**Proof:** According to (SPC), we have $x^* \in \mathcal{G} \cap \mathbf{M}_2$, $u^* \in \mathbb{R}^{\mathsf{J}_1}$, $u^* \geq 0$, $v^* \in \mathbb{R}^{\mathsf{K}_1}$.
For (4.1), the particular form of (4.9) is

$$\begin{aligned} f(x) + \langle u^*, g_{\mathsf{J}_1}(x) \rangle + \langle v^*, h_{\mathsf{K}_1}(x) \rangle \ &\geq\ f(x^*) + \langle u^*, g_{\mathsf{J}_1}(x^*) \rangle + \langle v^*, h_{\mathsf{K}_1}(x^*) \rangle \\ &\geq\ f(x^*) + \langle u, g_{\mathsf{J}_1}(x^*) \rangle + \langle v, h_{\mathsf{K}_1}(x^*) \rangle \end{aligned} \tag{4.11}$$

for all $x \in \mathcal{G} \cap \mathbf{M}_2$, $u \in \mathbb{R}^{\mathsf{J}_1}$, $u \geq \mathbf{0}$, $v \in \mathbb{R}^{\mathsf{K}_1}$.
Consider right-hand side of (4.11), i.e. inequality

$$\langle u^*, g_{\mathsf{J}_1}(x^*) \rangle + \langle v^*, h_{\mathsf{K}_1}(x^*) \rangle \geq \langle u, g_{\mathsf{J}_1}(x^*) \rangle + \langle v, h_{\mathsf{K}_1}(x^*) \rangle \quad \forall u \in \mathbb{R}^{\mathsf{J}_1}, u \geq \mathbf{0}, v \in \mathbb{R}^{\mathsf{K}_1}. \tag{4.12}$$

- Take $j \in \mathsf{J}_1$. Fixing $u_\iota = 0$ for all $\iota \in \mathsf{J}_1 \setminus \{j\}$, $v = \mathbf{0}$ and plugging in (4.12), we are receiving

  $$\langle u^*, g_{\mathsf{J}_1}(x^*) \rangle + \langle v^*, h_{\mathsf{K}_1}(x^*) \rangle \geq u_j g_j(x^*) \quad \forall u_j \geq 0.$$

  Hence,

  $$g_j(x^*) \leq \lim_{u_j \to +\infty} \frac{\langle u^*, g_{\mathsf{J}_1}(x^*) \rangle + \langle v^*, h_{\mathsf{K}_1}(x^*) \rangle}{u_j} = 0.$$

- Take $k \in \mathsf{K}_1$. Fixing $u = \mathbf{0}$, $v_\iota = 0$ for all $\iota \in \mathsf{K}_1 \setminus \{k\}$ and plugging in (4.12), we are receiving
$$\langle u^*, g_{\mathsf{J}_1}(x^*) \rangle + \langle v^*, h_{\mathsf{K}_1}(x^*) \rangle \geq v_k h_k(x^*) \quad \forall v_k \in \mathbb{R}.$$
Hence,
$$\begin{aligned} h_k(x^*) &\leq \lim_{v_k \to +\infty} \frac{\langle u^*, g_{\mathsf{J}_1}(x^*) \rangle + \langle v^*, h_{\mathsf{K}_1}(x^*) \rangle}{v_k} = 0, \\ h_k(x^*) &\geq \lim_{v_k \to -\infty} \frac{\langle u^*, g_{\mathsf{J}_1}(x^*) \rangle + \langle v^*, h_{\mathsf{K}_1}(x^*) \rangle}{v_k} = 0. \end{aligned}$$
- From (4.12), we derived $g_j(x^*) \leq 0$ for all $j \in \mathsf{J}_1$, $h_k(x^*) = 0$ for all $k \in \mathsf{K}_1$.
Since $x^* \in \mathbf{M}_2$, we have checked $x^* \in \mathbf{M}$.

Consequently, $\langle v^*, h_{\mathsf{K}_1}(x^*) \rangle = \langle v, h_{\mathsf{K}_1}(x^*) \rangle = 0$.
Plugging $u = \mathbf{0}$, $v = \mathbf{0}$ in (4.12), we observe $\langle u^*, g_{\mathsf{J}_1}(x^*) \rangle \geq 0$.
Since $u^* \geq \mathbf{0}$ and $g_j(x^*) \leq 0$ for all $j \in \mathsf{J}_1$, we have $\langle u^*, g_{\mathsf{J}_1}(x^*) \rangle \leq 0$ and, therefore, the complementarity condition $\langle u^*, g_{\mathsf{J}_1}(x^*) \rangle = 0$ is verified.
Thus, $x^*$ is a feasible solutions of (4.1) and the complementarity condition is shown. The left-hand side of (4.11) is reduced to
$$f(x) + \langle u^*, g_{\mathsf{J}_1}(x) \rangle + \langle v^*, h_{\mathsf{K}_1}(x) \rangle \geq f(x^*) \quad \forall\, x \in \mathcal{G} \cap \mathbf{M}_2. \tag{4.13}$$
Take $x \in \mathbf{M}$, i.e. $g_j(x) \leq 0$ for all $j \in \mathsf{J}$, $h_k(x) = 0$ for all $k \in \mathsf{K}$ and plug it in (4.13). Hence, $\langle u^*, g_{\mathsf{J}_1}(x) \rangle \leq 0$, $\langle v^*, h_{\mathsf{K}_1}(x) \rangle = 0$ and we have $f(x) \geq f(x^*)$ for all $x \in \mathbf{M}$.
It means $x^*$ is a global minimum of (4.1).

Q.E.D.

Reverse implication is not valid in general. It needs some regularity condition. Having a convex program, we can employ Slater's regularity condition.

**Definition 4.6** *Let $\mathcal{G} \supset \mathbf{M}$ be an open set, functions $f$, $g_j$, $j \in \mathsf{J}$, $h_k$, $k \in \mathsf{K}$ be real-valued on $\mathcal{G}$ and functions $h_k, k \in \mathsf{K}$ possess gradients on $\mathcal{G}$. We say, Slater's regularity condition (Slater) (cz. Slaterova podmínka regularity) is fulfilled for (4.1) whenever there is $\widetilde{x} \in \mathbf{M}$ such that $g_j(\widetilde{x}) < 0$ for all $j \in \mathsf{J}$ and gradients $\nabla h_k(\widetilde{x})$, $k \in \mathsf{K}$ are linearly independent, i.e.*
$$\text{matrix} \quad \left( \frac{\partial h_k(\widetilde{x})}{\partial x_i} \right)_{i=1,k=1}^{n,q} \quad \text{is of full rank } q.$$

**Theorem 4.7** (NLP ⇒ SPC): *Let $\mathcal{G} \supset \mathbf{M}$ be an open convex set, functions $f$, $g_j$, $j \in \mathsf{J}$, $h_k$, $k \in \mathsf{K}$ be real-valued on $\mathcal{G}$, functions $f$, $g_j$, $j \in \mathsf{J}$ be convex on $\mathcal{G}$, functions $h_k$, $k \in \mathsf{K}$ be affine on $\mathcal{G}$ (hence, defined on whole $\mathbb{R}^n$), and, (4.1) fulfill (Slater).*

*If $x^*$ is a global minimum of (4.1), then there are $u^* \in \mathbb{R}^{\mathsf{J}_1}$, $v^* \in \mathbb{R}^{\mathsf{K}_1}$ such that $(x^*, u^*, v^*)$ fulfills (SPC) for (4.1) on the region $(\mathcal{G} \cap \mathbf{M}_2) \times \mathbb{R}^{\mathsf{J}_1}_{+,0} \times \mathbb{R}^{\mathsf{K}_1}$. Moreover, the triplet fulfills the complementarity condition for (4.1), i.e. $\langle u^*, g_{\mathsf{J}_1}(x^*) \rangle = 0$.*

**Proof:** We will show existence of $u^*$, $v^*$ by means of a separation theorem. Consider two sets:
$$\mathcal{A} = \left\{ \begin{pmatrix} \eta \\ \xi \\ \zeta \end{pmatrix} : \begin{array}{l} \eta \in \mathbb{R},\ \xi \in \mathbb{R}^{\mathsf{J}_1},\ \zeta \in \mathbb{R}^{\mathsf{K}_1} \text{ fulfill} \\ \eta \geq f(x),\ \xi_j \geq g_j(x),\ j \in \mathsf{J}_1,\ \zeta_k = h_k(x),\ k \in \mathsf{K}_1 \text{ for some } x \in \mathcal{G} \cap \mathbf{M}_2 \end{array} \right\},$$
$$\mathcal{B} = \left\{ \begin{pmatrix} \eta \\ \xi \\ \zeta \end{pmatrix} : \eta \in \mathbb{R},\ \xi \in \mathbb{R}^{\mathsf{J}_1},\ \zeta \in \mathbb{R}^{\mathsf{K}_1},\ \eta < f(x^*),\ \xi < 0,\ \zeta = 0 \right\}.$$

1. Evidently $\mathcal{B}$ is a convex set. We have to prove $\mathcal{A}$ is a convex set.

   Let $\begin{pmatrix} \eta^1 \\ \xi^1 \\ \zeta^1 \end{pmatrix}, \begin{pmatrix} \eta^2 \\ \xi^2 \\ \zeta^2 \end{pmatrix} \in \mathcal{A}$, $\lambda \in (0,1)$ and $\begin{pmatrix} \eta \\ \xi \\ \zeta \end{pmatrix} = \lambda \begin{pmatrix} \eta^1 \\ \xi^1 \\ \zeta^1 \end{pmatrix} + (1-\lambda) \begin{pmatrix} \eta^2 \\ \xi^2 \\ \zeta^2 \end{pmatrix}$.

   Hence, there are $x^1, x^2 \in \mathcal{G} \cap \mathbf{M}_2$ fulfilling

$$\eta^1 \geq f\left(x^1\right), \; \xi^1_j \geq g_j(x^1), \; j \in \mathsf{J}_1, \; \zeta^1_k = h_k(x^1), \; k \in \mathsf{K}_1,$$
$$\eta^2 \geq f\left(x^2\right), \; \xi^2_j \geq g_j(x^2), \; j \in \mathsf{J}_1, \; \zeta^2_k = h_k(x^2), \; k \in \mathsf{K}_1.$$

   Set $\mathcal{G}$ is assumed convex and set $\mathbf{M}_2$ is convex, since functions $g_j$, $j \in \mathsf{J}_2$ are convex on $\mathcal{G}$ and functions $h_k$, $k \in \mathsf{K}_2$ are affine. Therefore, $x = \lambda x^1 + (1-\lambda)x^2 \in \mathcal{G} \cap \mathbf{M}_2$.

   Since functions $f$, $g_j$, $j \in \mathsf{J}_1$ are convex and functions $h_k$, $k \in \mathsf{K}_1$ are affine, we have

$$\eta = \lambda\eta^1 + (1-\lambda)\eta^2 \geq \lambda f\left(x^1\right) + (1-\lambda) f\left(x^2\right) \geq f(x),$$
$$\xi_j = \lambda\xi^1_j + (1-\lambda)\xi^2_j \geq \lambda g_j(x^1) + (1-\lambda) g_j(x^2) \geq g_j(x) \quad \forall j \in \mathsf{J}_1,$$
$$\zeta_k = \lambda\zeta^1_k + (1-\lambda)\zeta^2_k = \lambda h_k(x^1) + (1-\lambda) h_k(x^2) = h_k(x) \quad \forall k \in \mathsf{K}_1.$$

   Finally, $\begin{pmatrix} \eta \\ \xi \\ \zeta \end{pmatrix} \in \mathcal{A}$ and, therefore, $\mathcal{A}$ is a convex set.

2. We will show, $\mathcal{A} \cap \mathcal{B} = \emptyset$.

   Assume $\begin{pmatrix} \eta \\ \xi \\ \zeta \end{pmatrix} \in \mathcal{B} \cap \mathcal{A}$.

   Hence, there is $x \in \mathcal{G} \cap \mathbf{M}_2$ such that $f(x^*) > \eta \geq f(x)$, $g_j(x) \leq \xi_j < 0$ for all $j \in \mathsf{J}_1$ and $h_k(x) = 0$ for all $k \in \mathsf{K}_1$.

   Since $x \in \mathbf{M}_2$, we have $g_j(x) \leq 0$ for all $j \in \mathsf{J}_2$ and $h_k(x) = 0$ for all $k \in \mathsf{K}_2$.

   Consequently, $x \in \mathbf{M}$ and $f(x^*) > f(x)$, which contradicts our assumption that $x^*$ is a global minimum of (4.1).

Assumptions of Theorem 1.42 are verified, therefore, there are $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}^{\mathsf{J}_1}$, $\gamma \in \mathbb{R}^{\mathsf{K}_1}$, $\begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} \neq \mathbf{0}$ such that

$$\left\langle \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}, \begin{pmatrix} \eta \\ \xi \\ \zeta \end{pmatrix} \right\rangle \geq \left\langle \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}, \begin{pmatrix} a \\ b \\ \mathbf{0} \end{pmatrix} \right\rangle \quad \forall \begin{pmatrix} \eta \\ \xi \\ \zeta \end{pmatrix} \in \mathcal{A}, \begin{pmatrix} a \\ b \\ \mathbf{0} \end{pmatrix} \in \mathcal{B}. \tag{4.14}$$

Members of $a$ and $b$ can be arbitrary negative, therefore, $\alpha \geq 0$, $\beta \geq \mathbf{0}$. Otherwise right-hand side of the inequality could be arbitrary large positive, consequently, the inequality would fail.

For $x \in \mathcal{G} \cap \mathbf{M}_2$ we have $\begin{pmatrix} f(x) \\ g_{\mathsf{J}_1}(x) \\ h_{\mathsf{K}_1}(x) \end{pmatrix} \in \mathcal{A}$. Therefore for all $\begin{pmatrix} a \\ b \\ \mathbf{0} \end{pmatrix} \in \mathcal{B}$

$$\left\langle \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}, \begin{pmatrix} f(x) \\ g_{\mathsf{J}_1}(x) \\ h_{\mathsf{K}_1}(x) \end{pmatrix} \right\rangle \geq \left\langle \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}, \begin{pmatrix} a \\ b \\ \mathbf{0} \end{pmatrix} \right\rangle.$$

Letting $a \to f(x^*)$, $b \to \mathbf{0}$, we receive

$$\left\langle \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}, \begin{pmatrix} f(x) \\ g_{\mathsf{J}_1}(x) \\ h_{\mathsf{K}_1}(x) \end{pmatrix} \right\rangle \geq \left\langle \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}, \begin{pmatrix} f(x^*) \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix} \right\rangle . \tag{4.15}$$

1. Assume $\alpha = 0$.

   Hence,

   $$\beta \geq \mathbf{0}, \begin{pmatrix} \beta \\ \gamma \end{pmatrix} \neq \mathbf{0} \quad \text{and} \quad \left\langle \begin{pmatrix} \beta \\ \gamma \end{pmatrix}, \begin{pmatrix} g_{\mathsf{J}_1}(x) \\ h_{\mathsf{K}_1}(x) \end{pmatrix} \right\rangle \geq \left\langle \begin{pmatrix} \beta \\ \gamma \end{pmatrix}, \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix} \right\rangle = 0 \quad \forall\, x \in \mathcal{G} \cap \mathbf{M}_2.$$

   According to (Slater), there is $\widetilde{x} \in \mathbf{M}$ such that $g_j(\widetilde{x}) < 0$ for all $j \in \mathsf{J}$ and $h_k(\widetilde{x}) = 0$ for all $k \in \mathsf{K}$. Therefore, $\beta = \mathbf{0}$. Hence, the condition is reduced to

   $$\gamma \neq \mathbf{0} \quad \text{and} \quad \langle \gamma, h_{\mathsf{K}_1}(x) \rangle \geq 0 \quad \forall\, x \in \mathcal{G} \cap \mathbf{M}_2.$$

   Functions $g_j$, $j \in \mathsf{J}$ are convex on the open set $\mathcal{G}$, therefore, they are continuous on $\mathcal{G}$.

   Since $g_j(\widetilde{x}) < 0$ for all $j \in \mathsf{J}$, there is $\delta > 0$ such that $\mathcal{U}_\delta(\widetilde{x}) \subset \mathcal{G}$ and

   $$\forall x \in \mathcal{U}_\delta(\widetilde{x}) \quad \text{we have} \quad g_j(x) < 0 \text{ for all } j \in \mathsf{J}.$$

   Hence,

   $$\forall\, x \in \mathcal{U}_\delta(\widetilde{x}) \text{ fulfilling } h_k(x) = 0 \text{ for all } k \in \mathsf{K}_2 \text{ we have } x \in \mathcal{G} \cap \mathbf{M}_2 \text{ and } \langle \gamma, h_{\mathsf{K}_1}(x) \rangle \geq 0.$$

   Consider $\Delta \in \mathcal{U}_\delta(\mathbf{0})$ fulfilling $h_k(\widetilde{x} + \Delta) = 0$ for all $k \in \mathsf{K}_2$.

   Hence, $\widetilde{x} + \Delta \in \mathcal{G} \cap \mathbf{M}_2$ and $\langle \gamma, h_{\mathsf{K}_1}(\widetilde{x} + \Delta) \rangle \geq 0$.

   Functions $h_k$, $k \in \mathsf{K}$ are affine, therefore,

   $$\begin{aligned} &\langle \nabla h_k(\widetilde{x}), \Delta \rangle = h_k(\widetilde{x} + \Delta) - h_k(\widetilde{x}) = 0 \text{ for all } k \in \mathsf{K}_2, \\ &\left\langle \sum_{k \in \mathsf{K}_1} \gamma_k \nabla h_k(\widetilde{x}), \Delta \right\rangle = \sum_{k \in \mathsf{K}_1} \gamma_k \langle \nabla h_k(\widetilde{x}), \Delta \rangle = \sum_{k \in \mathsf{K}_1} \gamma_k \left[ h_k(\widetilde{x} + \Delta) - h_k(\widetilde{x}) \right] \\ &= \langle \gamma, h_{\mathsf{K}_1}(\widetilde{x} + \Delta) \rangle \geq 0. \end{aligned}$$

   There is an unique rewriting $\sum_{k \in \mathsf{K}_1} \gamma_k \nabla h_k(\widetilde{x}) = \omega + \kappa$, where $\langle \nabla h_k(\widetilde{x}), \omega \rangle = 0$ for all $k \in \mathsf{K}_2$ and $\kappa \in \mathcal{L}(\nabla h_k(\widetilde{x}), k \in \mathsf{K}_2)$.

   There is $\rho > 0$ such that $\rho\omega \in \mathcal{U}_\delta(\mathbf{0})$, hence, plugging $\Delta = -\rho\omega$, we are observing

   $$\left\langle \sum_{k \in \mathsf{K}_1} \gamma_k \nabla h_k(\widetilde{x}), -\rho\omega \right\rangle = \langle \omega + \kappa, -\rho\omega \rangle = \langle \omega, -\rho\omega \rangle + \langle \kappa, -\rho\omega \rangle = -\rho \|\omega\|^2 \geq 0.$$

   Hence, $\omega = 0$ and $\sum_{k \in \mathsf{K}_1} \gamma_k \nabla h_k(\widetilde{x}) = \kappa \in \mathcal{L}(\nabla h_k(\widetilde{x}), k \in \mathsf{K}_2)$, which violates linear independence of gradients.

   Finally, $\alpha > 0$.

2. We need to show complementarity.

   Since $x^*$ is a feasible solutions of (4.1) and $\beta \geq 0$, we have $\langle \beta, g_{\mathsf{J}_1}(x^*) \rangle \leq 0$. In the same time according to (4.15), we also have $\langle \beta, g_{\mathsf{J}_1}(x^*) \rangle \geq 0$.

   Therefore, $\langle \beta, g_{\mathsf{J}_1}(x^*) \rangle = 0$.

Thus, we can divide inequality (4.15) by $\alpha$:

$$\left\langle \begin{pmatrix} 1 \\ \frac{1}{\alpha}\beta \\ \frac{1}{\alpha}\gamma \end{pmatrix}, \begin{pmatrix} f(x) \\ g_{\mathsf{J}_1}(x) \\ h_{\mathsf{K}_1}(x) \end{pmatrix} \right\rangle \geq f(x^*) \quad \forall\, x \in \mathcal{G} \cap \mathbf{M}_2.$$

Denote $u^* = \frac{1}{\alpha}\beta$, $v^* = \frac{1}{\alpha}\gamma$. Hence, $u^* \geq 0$ and

$$\left\langle \begin{pmatrix} 1 \\ u^* \\ v^* \end{pmatrix}, \begin{pmatrix} f(x) \\ g_{\mathsf{J}_1}(x) \\ h_{\mathsf{K}_1}(x) \end{pmatrix} \right\rangle \geq f(x^*) \quad \forall\, x \in \mathcal{G} \cap \mathbf{M}_2. \tag{4.16}$$

Since $\langle u^*, g_{\mathsf{J}_1}(x^*) \rangle = 0$, $h_k(x^*) = 0$ for all $k \in \mathsf{K}$, $\langle u, g_{\mathsf{J}_1}(x^*) \rangle \leq 0$ for all $u \geq 0$, we can continue

$$\left\langle \begin{pmatrix} 1 \\ u^* \\ v^* \end{pmatrix}, \begin{pmatrix} f(x) \\ g_{\mathsf{J}_1}(x) \\ h_{\mathsf{K}_1}(x) \end{pmatrix} \right\rangle \geq \left\langle \begin{pmatrix} 1 \\ u^* \\ v^* \end{pmatrix}, \begin{pmatrix} f(x^*) \\ g_{\mathsf{J}_1}(x^*) \\ h_{\mathsf{K}_1}(x^*) \end{pmatrix} \right\rangle \geq \left\langle \begin{pmatrix} 1 \\ u \\ v \end{pmatrix}, \begin{pmatrix} f(x^*) \\ g_{\mathsf{J}_1}(x^*) \\ h_{\mathsf{K}_1}(x^*) \end{pmatrix} \right\rangle. \tag{4.17}$$

We have proved chain of inequalities (4.17) which is (SPC) for (4.1).

Q.E.D.

## 4.3 Local minimum

In this section, local minima will discussed. To do that conditions developed in previous section will be restricted to a neighborhood of the given point.

**Definition 4.8** *Let $\mathcal{G} \supset \mathbf{M}$ be an open set, functions $f$, $g_j$, $j \in \mathsf{J}$, $h_k$, $k \in \mathsf{K}$ be real-valued on $\mathcal{G}$, $x^* \in \mathbb{R}^n$, $u^* \in \mathbb{R}^{\mathsf{J}_1}$, $v^* \in \mathbb{R}^{\mathsf{K}_1}$. We say, the triplet $(x^*, u^*, v^*)$ fulfills local saddle point condition (locSPC) (cz. lokálně podmínku sedlového bodu) for (4.1) on the region $(\mathcal{G} \cap \mathbf{M}_2) \times \mathbb{R}^{\mathsf{J}_1}_{+,0} \times \mathbb{R}^{\mathsf{K}_1}$, whenever there is $\delta > 0$ such that the triplet $(x^*, u^*, v^*)$ fulfills (SPC) for (4.1) on the region $(\mathcal{U}_\delta(x^*) \cap \mathcal{G} \cap \mathbf{M}_2) \times \mathbb{R}^{\mathsf{J}_1}_{+,0} \times \mathbb{R}^{\mathsf{K}_1}$.*

**Theorem 4.9** (locSPC ⇒ locNLP): *Let $\mathcal{G} \supset \mathbf{M}$ be an open set, functions $f$, $g_j$, $j \in \mathsf{J}$, $h_k$, $k \in \mathsf{K}$ be real-valued on $\mathcal{G}$ and $x^* \in \mathbb{R}^n$. If there are $u^* \in \mathbb{R}^{\mathsf{J}_1}$, $v^* \in \mathbb{R}^{\mathsf{K}_1}$ such that the triplet $(x^*, u^*, v^*)$ fulfills (locSPC) for (4.1) on the region $(\mathcal{G} \cap \mathbf{M}_2) \times \mathbb{R}^{\mathsf{J}_1}_{+,0} \times \mathbb{R}^{\mathsf{K}_1}$, then $x^*$ is a local minimum of (4.1).*

*Moreover, the complementarity condition for (4.1) is fulfilled, i.e. $\langle u^*, g_{\mathsf{J}_1}(x^*) \rangle = 0$.*

**Proof:** According to our assumptions, we possess $\delta > 0$ such that the triplet $(x^*, u^*, v^*)$ fulfills (SPC) for (4.1) on the region $(\mathcal{U}_\delta(x^*) \cap \mathcal{G} \cap \mathbf{M}_2) \times \mathbb{R}^{\mathsf{J}_1}_{+,0} \times \mathbb{R}^{\mathsf{K}_1}$.
Take $0 < \varepsilon < \delta$ and include a new constraint in (4.1)

$$\min\left\{ f(x) \,:\, g_j(x) \leq 0\,,\ j \in \mathsf{J}\,,\ \|x - x^*\| \leq \varepsilon,\ h_k(x) = 0,\ k \in \mathsf{K}\,,\ x \in \mathbb{R}^n \right\}. \tag{4.18}$$

We subsume the constraint $\|x - x^*\| \leq \varepsilon$ in the second group of constraints, i.e. it remains a constraint in the relaxed program. Hence, the triplet $(x^*, u^*, v^*)$ fulfills (SPC) for the program (4.18) on the region $(\mathcal{G} \cap (\mathbf{M}_2 \cap \mathcal{V}_\varepsilon(x^*))) \times \mathbb{R}^{\mathsf{J}_1}_{+,0} \times \mathbb{R}^{\mathsf{K}_1}$.
According to Theorem 4.5, $x^*$ is a global minimum of (4.18).
It means, $x^*$ is a local minimum of (4.1).

Q.E.D.

Reverse implication is not valid in general. For example, it is in power for a program, which is locally convex and fulfills local Slater's regularity condition.

**Definition 4.10** *Let $\mathcal{G} \supset \mathbf{M}$ be an open set, functions $f$, $g_j$, $j \in \mathsf{J}$, $h_k$, $k \in \mathsf{K}$ be real-valued on $\mathcal{G}$, functions $h_k, k \in \mathsf{K}$ possess gradients on $\mathcal{G}$ and $\mathcal{U} \subset \mathbb{R}^n$ be a non-empty open set.*

*We say, local Slater's regularity condition (locSlater) (cz. lokální Slaterova podmínka regularity) is fulfilled on $\mathcal{U}$ for (4.1) whenever there is $\widetilde{x} \in \mathbf{M} \cap \mathcal{U}$ such that $g_j(\widetilde{x}) < 0$ for all $j \in \mathsf{J}$ and gradients $\nabla h_k(\widetilde{x})$, $k \in \mathsf{K}$ are linearly independent.*

**Theorem 4.11** (locNLP ⇒ locSPC): *Let $\mathcal{G} \supset \mathbf{M}$ be an open set, functions $f$, $g_j$, $j \in \mathsf{J}$, $h_k$, $k \in \mathsf{K}$ be real-valued on $\mathcal{G}$ and $x^* \in \mathbf{M}$. Let $\delta > 0$ be such that $\mathcal{U}_\delta(x^*) \subset \mathcal{G}$, functions $f, g_j, j \in \mathsf{J}$ be convex on $\mathcal{U}_\delta(x^*)$, functions $h_k, k \in \mathsf{K}$ be affine on $\mathcal{U}_\delta(x^*)$, and, (4.1) fulfill (locSlater) on $\mathcal{U}_\delta(x^*)$.*

*If $x^*$ is a local minimum of (4.1), then there are $u^* \in \mathbb{R}^{\mathsf{J}_1}$, $v^* \in \mathbb{R}^{\mathsf{K}_1}$ such that $(x^*, u^*, v^*)$ fulfills (locSPC) for (4.1) on the region $(\mathcal{G} \cap \mathbf{M}_2) \times \mathbb{R}^{\mathsf{J}_1}_{+,0} \times \mathbb{R}^{\mathsf{K}_1}$. Moreover, the complementarity condition for (4.1) is fulfilled, i.e. $\langle u^*, g_{\mathsf{J}_1}(x^*) \rangle = 0$.*

**Proof:** According to our assumptions, $x^*$ is a local minimum of (4.1).
Therefore, there is $0 < \varepsilon < \delta$ such that $x^*$ is a global minimum of

$$\min \{ f(x) \,:\, g_j(x) \le 0\,,\ j \in \mathsf{J}\,,\ \|x - x^*\| \le \varepsilon,\ h_k(x) = 0,\ k \in \mathsf{K}\,,\ x \in \mathbb{R}^n \}\ . \tag{4.19}$$

We subsume the constraint $\|x - x^*\| \le \varepsilon$ in the second group of constraints, i.e. it remains a constraint in the relaxed program.
We assume $\mathcal{U}_\delta(x^*) \subset \mathcal{G}$, functions $f, g_j, j \in \mathsf{J}$ are convex on $\mathcal{U}_\delta(x^*)$, functions $h_k, k \in \mathsf{K}$ are affine on $\mathcal{U}_\delta(x^*)$, and, (4.1) fulfill (locSlater) on $\mathcal{U}_\delta(x^*)$. Thus, there is $\widetilde{x} \in \mathbf{M} \cap \mathcal{U}_\delta(x^*)$ with $g_j(\widetilde{x}) < 0$ for all $j \in \mathsf{J}$ and gradients $\nabla h_k(\widetilde{x})$, $k \in \mathsf{K}$ are linearly independent.
Hence for all $0 < \lambda < 1$, $g_j(\lambda\widetilde{x} + (1-\lambda)x^*) < 0$ for all $j \in \mathsf{J}$ and gradients $\nabla h_k(\lambda\widetilde{x} + (1-\lambda)x^*) = \nabla h_k(\widetilde{x})$, $k \in \mathsf{K}$ are linearly independent. Therefore, (4.19) fulfills (Slater).
According to Theorem 4.7, there are $u^* \in \mathbb{R}^{\mathsf{J}_1}$, $v^* \in \mathbb{R}^{\mathsf{K}_1}$ such that the triplet $(x^*, u^*, v^*)$ fulfills (SPC) for the program (4.19) on the region $(\mathcal{G} \cap (\mathbf{M}_2 \cap \mathcal{V}_\varepsilon(x^*))) \times \mathbb{R}^{\mathsf{J}_1}_{+,0} \times \mathbb{R}^{\mathsf{K}_1}$ and the complementarity condition is fulfilled.
Finally, $(x^*, u^*, v^*)$ fulfills (locSPC) for (4.1) on the region $(\mathcal{G} \cap \mathbf{M}_2) \times \mathbb{R}^{\mathsf{J}_1}_{+,0} \times \mathbb{R}^{\mathsf{K}_1}$ and the complementarity condition is fulfilled.

Q.E.D.

Let us recall basic relation between global and local minimum.

**Theorem 4.12** (locNLP ⇔ NLP):

*i) If $x^*$ is a global minimum of (4.1), then it is a local minimum of (4.1).*

*ii) Let $\mathbf{M}$ be a convex set, $f$ be a convex function. If $x^*$ is a local minimum of (4.1), then it is a global minimum of (4.1).*

**Proof:** The first statement follows directly definitions of global and local minimum. The second statement was shown as Theorem 2.39.

Q.E.D.

**Theorem 4.13** (locSPC $\Leftrightarrow$ SPC): *Let $\mathcal{G} \supset \mathbf{M}$ be an open set, functions $f$, $g_j$, $j \in \mathsf{J}$, $h_k$, $k \in \mathsf{K}$ be real-valued on $\mathcal{G}$ and $x^* \in \mathbf{M}$, $u^* \in \mathbb{R}^{\mathsf{J}_1}$, $v^* \in \mathbb{R}^{\mathsf{K}_1}$.*

*i) If $(x^*, u^*, v^*)$ fulfills (SPC) for (4.1) on the region $(\mathcal{G} \cap \mathbf{M}_2) \times \mathbb{R}^{\mathsf{J}_1}_{+,0} \times \mathbb{R}^{\mathsf{K}_1}$, then it fulfills (locSPC) for (4.1) on the region $(\mathcal{G} \cap \mathbf{M}_2) \times \mathbb{R}^{\mathsf{J}_1}_{+,0} \times \mathbb{R}^{\mathsf{K}_1}$.*

*ii) Let $\mathcal{G}$ be an open convex set, functions $f$, $g_j$, $j \in \mathsf{J}$ be convex on $\mathcal{G}$, functions $h_k$, $k \in \mathsf{K}$ be affine on $\mathcal{G}$. If $(x^*, u^*, v^*)$ fulfills (locSPC) for (4.1) on the region $(\mathcal{G} \cap \mathbf{M}_2) \times \mathbb{R}^{\mathsf{J}_1}_{+,0} \times \mathbb{R}^{\mathsf{K}_1}$, then it fulfills (SPC) for (4.1) on the region $(\mathcal{G} \cap \mathbf{M}_2) \times \mathbb{R}^{\mathsf{J}_1}_{+,0} \times \mathbb{R}^{\mathsf{K}_1}$.*

**Proof:** The first statement follows directly from definitions of (SPC), (locSPC). The second statement needs a proof.

If (locSPC) is fulfilled, there is $\delta > 0$ such that the triplet $(x^*, u^*, v^*)$ fulfills (SPC) for (4.1) on the region $(\mathcal{U}_\delta(x^*) \cap \mathcal{G} \cap \mathbf{M}_2) \times \mathbb{R}^{\mathsf{J}_1}_{+,0} \times \mathbb{R}^{\mathsf{K}_1}$.
Take $x \in \mathcal{G} \cap \mathbf{M}_2$.
Set $\mathcal{G} \cap \mathbf{M}_2$ is convex, therefore, whole segment connecting points $x$ and $x^*$ lays in $\mathcal{G} \cap \mathbf{M}_2$.
Moreover, there is $0 < \lambda < 1$ such that $y = \lambda x^* + (1-\lambda)x \in \mathcal{U}_\delta(x^*)$.
Under assumptions of the second statement, Lagrange function $L$ is convex in $x$, thus, we have

$$L(x; u^*, v^*) \geq L(y; u^*, v^*) \geq L(x^*; u^*, v^*).$$

Condition (SPC) is proved, since the right-hand side of the inequality coincides with (locSPC).

Q.E.D.

## 4.4 Karush-Kuhn-Tucker optimality condition

In this section, we will consider program (4.1) and all constraints will be subsumed in Lagrange function as its penalization; i.e. $\mathsf{J}_1 = \mathsf{J}$, $\mathsf{K}_1 = \mathsf{K}$, $\mathsf{J}_2 = \emptyset$, $\mathsf{K}_2 = \emptyset$, therefore, $\mathbf{M}_2 = \mathbb{R}^n$.

Substantially influenced persons:

We choose convenient open set $\mathcal{G} \supset \mathbf{M}$ on which all functions appearing in (4.1) are real-valued and we consider Lagrange function on $\mathcal{G} \times \mathbb{R}^m_{+,0} \times \mathbb{R}^q$, since $\mathbf{M}_2 = \mathbb{R}^n$, defined for $x \in \mathcal{G}$, $u \in \mathbb{R}^m_{+,0}$, $v \in \mathbb{R}^q$, as

$$\begin{aligned} L(x; u, v) &= \left\langle \begin{pmatrix} 1 \\ u \\ v \end{pmatrix}, \begin{pmatrix} f(x) \\ g(x) \\ h(x) \end{pmatrix} \right\rangle = f(x) + \langle u, g(x) \rangle + \langle v, h(x) \rangle \qquad (4.20) \\ &= f(x) + \sum_{j=1}^{m} u_j g_j(x) + \sum_{k=1}^{q} v_k h_k(x). \end{aligned}$$

Let us recall differentiability of Lagrange function.

- If functions $f$, $g_j$, $j \in \mathsf{J}$, $h_k$, $k \in \mathsf{K}$ possess gradients at $x$, then $L(\bullet; u, v)$ possesses a gradient at $x$ and

$$\nabla_x L(x; u, v) = \nabla f(x) + \sum_{j \in \mathsf{J}} u_j \nabla g_j(x) + \sum_{k \in \mathsf{K}} v_k \nabla h_k(x). \qquad (4.21)$$

- If functions $f,\ g_j,\ j\in\mathsf{J},\ h_k,\ k\in\mathsf{K}$ are differentiable at $x$, then $L(\bullet;u,v)$ is differentiable at $x$ and (4.21) is valid.
- If functions $f,\ g_j,\ j\in\mathsf{J},\ h_k,\ k\in\mathsf{K}$ are twice differentiable at $x$, then $L(\bullet;u,v)$ is twice differentiable at $x$ and

$$H_L(x;u,v) \quad = \quad H_f(x) + \sum_{j\in\mathsf{J}} u_j H_{g_j}(x) + \sum_{k\in\mathsf{K}} v_k H_{h_k}(x)\,. \tag{4.22}$$

Let us introduce Karush-Kuhn-Tucker optimality conditions.

**Definition 4.14** *Let $\mathcal{G}\supset\mathbf{M}$ be an open set, functions $f,\ g_j,\ j\in\mathsf{J},\ h_k,\ k\in\mathsf{K}$ be real-valued on $\mathcal{G}$, $x^*\in\mathbf{M}$, $u^*\in\mathbb{R}^m$, $v^*\in\mathbb{R}^q$, functions $f,\ g_j,\ j\in\mathsf{J},\ h_k,\ k\in\mathsf{K}$ possess gradients at $x^*$. We say, $(x^*,u^*,v^*)$ fulfills Karush-Kuhn-Tucker optimality conditions (KKT) (cz. Karushovy-Kuhnovy-Tuckerovy podmínky optimality, lokální podmínky optimality) for (4.1), whenever following conditions are fulfilled:*

*i) Primal Feasibility condition (PFC) (cz. přípustnost): $g_j(x^*)\le 0,\ j\in\mathsf{J},\ h_k(x^*)=0,\ k\in\mathsf{K}$.*

*ii) Dual Feasibility condition (DFC) (cz. optimalita): $\nabla_x L(x^*;u^*,v^*)=0$.*

*iii) Complementarity Slackness condition (CSC) (cz. komplementarita): $u^*\ge 0$, $\langle u^*, g(x^*)\rangle = 0$.*

**Theorem 4.15** (locSPC ⇒ KKT): *Let $\mathcal{G}\supset\mathbf{M}$ be an open set, functions $f,\ g_j,\ j\in\mathsf{J},\ h_k,\ k\in\mathsf{K}$ be real-valued on $\mathcal{G}$, $x^*\in\mathbf{M}$, $u^*\in\mathbb{R}^m$, $v^*\in\mathbb{R}^q$, functions $f,\ g_j,\ j\in\mathsf{J},\ h_k,\ k\in\mathsf{K}$ possess gradients at $x^*$. If $(x^*,u^*,v^*)$ fulfills (locSPC) for (4.1) on the region $\mathcal{G}\times\mathbb{R}^m\times\mathbb{R}^q$, then it fulfills (KKT) for (4.1).*

**Proof:** According to Theorem 4.9, $x^*$ is a local minimum of (4.1) and the complementarity condition is fulfilled. Therefore, conditions $(PFC)$ and $(CSC)$ are fulfilled.
Since $x^*$ is a local minimum of the function $L(\bullet;u^*,v^*)$ on $\mathcal{G}$, the function possesses a gradient at $x^*$ and $\mathcal{G}$ is an open set, the condition $(DFC)$ must be fulfilled, according to Theorem 3.10.

Q.E.D.

Let us define auxiliary notions.

**Definition 4.16** *For $x\in\mathbf{M}$, we define a set of indexes of active constraints (cz. množina indexů aktivních omezení)*

$$\mathsf{J}_g(x) \quad = \quad \{j\in\mathsf{J} : g_j(x)=0\}\,. \tag{4.23}$$

**Definition 4.17** *Let $\mathcal{G}\supset\mathbf{M}$ be an open set, functions $f,\ g_j,\ j\in\mathsf{J},\ h_k,\ k\in\mathsf{K}$ be real-valued on $\mathcal{G}$, $x\in\mathbf{M}$, $u\in\mathbb{R}^m$, $u\ge 0$, functions $f,\ g_j,\ j\in\mathsf{J},\ h_k,\ k\in\mathsf{K}$ possess gradients at $x$.*
*We define approximative set of feasible directions (cz. aproximující množina přípustných směrů)*

$$\mathsf{U}(x) = \left\{ z\in\mathbb{R}^n : \begin{array}{l} z\ne 0, \\ \langle\nabla g_j(x), z\rangle \le 0 \quad \mathit{for} \quad j\in\mathsf{J}_g(x), \\ \langle\nabla h_k(x), z\rangle = 0 \quad \mathit{for} \quad k\in\mathsf{K}, \end{array} \right\}$$

*second approximative set of feasible directions (cz. druhá aproximující množina přípustných směrů)*

$$\mathsf{V}(x,u) = \left\{ z\in\mathbb{R}^n : \begin{array}{l} z\ne 0, \\ \langle\nabla g_j(x), z\rangle = 0 \quad \mathit{for} \quad j\in\mathsf{J}_g(x),\ u_j>0, \\ \langle\nabla g_j(x), z\rangle \le 0 \quad \mathit{for} \quad j\in\mathsf{J}_g(x),\ u_j=0, \\ \langle\nabla h_k(x), z\rangle = 0 \quad \mathit{for} \quad k\in\mathsf{K}, \end{array} \right\}$$

*approximative set of improving directions* (*cz. aproximující množina zlepšujících směrů*)

$$\mathsf{Z}(x) = \left\{ z \in \mathbb{R}^n : \begin{array}{c} \langle \nabla g_j(x), z \rangle \le 0, \ j \in \mathsf{J}_g(x), \\ \langle \nabla h_k(x), z \rangle = 0, \ k \in \mathsf{K}, \\ \langle \nabla f(x), z \rangle < 0. \end{array} \right\} \tag{4.24}$$

**Theorem 4.18** (Basic theorem on KKT): *Let $\mathcal{G} \supset \mathbf{M}$ be an open set, functions $f$, $g_j$, $j \in \mathsf{J}$, $h_k$, $k \in \mathsf{K}$ be real-valued on $\mathcal{G}$, $x^* \in \mathbf{M}$, $u^* \in \mathbb{R}^m$, $v^* \in \mathbb{R}^q$, functions $f$, $g_j$, $j \in \mathsf{J}$, $h_k$, $k \in \mathsf{K}$ possess gradients at $x^*$. Hence, $\mathsf{Z}(x^*) = \emptyset$ if and only if there are $u^* \in \mathbb{R}^m$, $v^* \in \mathbb{R}^q$ such that $(x^*, u^*, v^*)$ fulfills $(DFC)$ and $(CSC)$.*

**Proof:** Since $x^* \in \mathbf{M}$, it is sufficient to consider following chain of equivalences.

There are $u^* \in \mathbb{R}^m$, $v^* \in \mathbb{R}^q$ such that $(x^*, u^*, v^*)$ fulfills $(DFC)$ and $(CSC)$.

$$\Updownarrow$$

There are $u^* \in \mathbb{R}^m$, $u^* \ge 0$, $v^* \in \mathbb{R}^q$ such that $(x^*, u^*, v^*)$ fulfills

$$\begin{aligned} \left\langle \begin{pmatrix} 1 \\ u^* \\ v^* \end{pmatrix}, \begin{pmatrix} \nabla f(x^*) \\ \nabla g(x^*) \\ \nabla h(x^*) \end{pmatrix} \right\rangle &= 0, \\ \langle u^*, g(x^*) \rangle &= 0. \end{aligned}$$

$$\Updownarrow$$

There are $u^* \in \mathbb{R}^m$, $u^* \ge 0$, $v^* \in \mathbb{R}^q$ fulfilling

$$\begin{aligned} \left\langle \begin{pmatrix} u^* \\ v^* \end{pmatrix}, \begin{pmatrix} \nabla g(x^*) \\ \nabla h(x^*) \end{pmatrix} \right\rangle &= -\nabla f(x^*), \\ \langle u^*, g(x^*) \rangle &= 0. \end{aligned}$$

$$\Updownarrow$$

There are $u^* \in \mathbb{R}^m$, $v^* \in \mathbb{R}^q$, $u^* \ge 0$, $u_j^* = 0$ for all $j \notin \mathsf{J}_g(x^*)$ fulfilling

$$\sum_{j \in \mathsf{J}_g(x^*)} u_j^* \nabla g_j(x^*) + \sum_{k=1}^{q} v_k^* \nabla h_k(x^*) = -\nabla f(x^*).$$

$\Updownarrow$ according to Theorem 1.46 (Farkas Theorem)

We have for each $z \in \mathbb{R}^n$

$$\begin{bmatrix} z^\top \nabla g_j(x^*) \le 0 & \forall j \in \mathsf{J}_g(x^*), \\ z^\top \nabla h_k(x^*) = 0 & \forall k \in \mathsf{K} \end{bmatrix} \Longrightarrow -z^\top \nabla f(x^*) \le 0.$$

$$\Updownarrow$$

We have for each $z \in \mathbb{R}^n$

$$\begin{bmatrix} \langle \nabla g_j(x^*), z \rangle \le 0 & \forall j \in \mathsf{J}_g(x^*), \\ \langle \nabla h_k(x^*), z \rangle = 0 & \forall k \in \mathsf{K} \end{bmatrix} \Longrightarrow \langle \nabla f(x^*), z \rangle \ge 0.$$

$$\Updownarrow$$

$$\mathsf{Z}(x^*) = \emptyset.$$

Q.E.D.

The approximative set of improving directions $\mathsf{Z}(x)$ is an attempt indicate directions leading from the point $x \in \mathbf{M}$ in the set of all feasible solutions $\mathbf{M}$ while objective function is decreasing. Thus, $x$ could be a minimum of (4.1) only if $\mathsf{Z}(x) = \emptyset$. Unfortunately, this scheme is not true in general; see example 4.29. The program (4.1) must fulfill some additional property; so called a regularity condition.

**Theorem 4.19**: *Let $\mathcal{G} \supset \mathbf{M}$ be an open set, functions $f$, $g_j$, $j \in \mathsf{J}$, $h_k$, $k \in \mathsf{K}$ be real-valued on $\mathcal{G}$, $\widetilde{x} \in \mathbf{M}$, functions $g_j$, $j \in \mathsf{J}$, $h_k$, $k \in \mathsf{K}$ possess gradients at $\widetilde{x}$, and, $f$ be differentiable at $\widetilde{x}$.*

*If $\widetilde{x}$ is a local minimum of (4.1) and for each $z \in \mathsf{Z}(\widetilde{x})$ there are $\lambda > 0$, $\underline{\theta} < 0 < \overline{\theta}$ and a vector function $\Psi : (\underline{\theta}, \overline{\theta}) \to \mathcal{G}$ fulfilling:*

$$\Psi \text{ is differentiable at } 0, \ \ \Psi(0) = \widetilde{x}\,, \ \ \Psi'(0) = \lambda z, \ \ \forall\, 0 < \theta < \overline{\theta} : \ \Psi(\theta) \in \mathbf{M}, \tag{4.25}$$

*then $\mathsf{Z}(\widetilde{x}) = \emptyset$.*

**Proof:** Let $\widetilde{x}$ be a local minimum of (4.1), $z \in \mathsf{Z}(\widetilde{x})$ and $\Psi$ be a vector function fulfilling (4.25). Consider the function $F(\theta) = f(\Psi(\theta))$ defined for all $\theta \in (\underline{\theta}, \overline{\theta})$. The function $F$ is differentiable at 0 with derivative

$$F'(0) \ = \ \langle \nabla f(\Psi(0)), \Psi'(0) \rangle = \langle \nabla f(\widetilde{x}), \lambda z \rangle = \lambda \langle \nabla f(\widetilde{x}), z \rangle \,.$$

Since $\widetilde{x}$ is a local minimum of (4.1), there is $\varepsilon > 0$ such that $\mathcal{U}_\varepsilon(\widetilde{x}) \subset \mathcal{G}$ and $\widetilde{x}$ is a global minimum of $f$ on $\mathcal{U}_\varepsilon(\widetilde{x}) \cap \mathbf{M}$.
Since $\Psi$ is differentiable at 0, there is $\theta_0 \in (0, \overline{\theta})$ such that $\Psi(\theta) \in \mathcal{U}_\varepsilon(\widetilde{x}) \cap \mathbf{M}$ for all $\theta \in (0, \theta_0)$.
Hence, we observe $F(0) \le F(\theta)$ for all $\theta \in (0, \theta_0)$.
After that,

$$0 \le F'(0) \ = \ \lambda \langle \nabla f(\widetilde{x}), z \rangle \,.$$

That is in contradiction with $\langle \nabla f(\widetilde{x}), z \rangle < 0$.

Thus, $\mathsf{Z}(\widetilde{x}) = \emptyset$.

Q.E.D.

A condition implying $\mathsf{Z}(\widetilde{x}) = \emptyset$ for a local minimum $\widetilde{x}$ is called constraint qualification (cz. podmínka regularity). The most general constraint qualification we will meet in the lecture is Kuhn-Tucker constraint qualification, which was published in 1956.

**Definition 4.20** *Let $\mathcal{G} \supset \mathbf{M}$ be an open set, functions $g_j$, $j \in \mathsf{J}$, $h_k$, $k \in \mathsf{K}$ be real-valued on $\mathcal{G}$, $\widetilde{x} \in \mathbf{M}$, functions $g_j$, $j \in \mathsf{J}$, $h_k$, $k \in \mathsf{K}$ possess gradients at $\widetilde{x}$.*

*We say, Kuhn-Tucker constraint qualification (K-T) (cz. Kuhnova-Tuckerova podmínka regularity) is fulfilled for (4.1) at $\widetilde{x}$ if for each $z \in \mathsf{U}(\widetilde{x})$ there are $\lambda > 0$, $\underline{\theta} < 0 < \overline{\theta}$ and a vector function $\Psi : (\underline{\theta}, \overline{\theta}) \to \mathbb{R}^n$ fulfilling:*

$$\Psi \text{ is differentiable at } 0, \ \ \Psi(0) = \widetilde{x}\,, \ \ \Psi'(0) = \lambda z, \ \ \forall\, 0 < \theta < \overline{\theta} : \ \Psi(\theta) \in \mathbf{M}. \tag{4.26}$$

**Remark 4.21:** Let us note that (K-T) condition is independent on $f$. Therefore, it is the same for minimization and maximization.

♠

**Theorem 4.22** (locNLP $\Rightarrow$ KKT): *Let $\mathcal{G} \supset \mathbf{M}$ be an open set, functions $f$, $g_j$, $j \in \mathsf{J}$, $h_k$, $k \in \mathsf{K}$ be real-valued on $\mathcal{G}$, $x^* \in \mathbf{M}$, functions $f$, $g_j$, $j \in \mathsf{J}$, $h_k$, $k \in \mathsf{K}$ possess gradients at $x^*$.*

*If $x^*$ is a local minimum of (4.1) and constraint qualification (K-T) is fulfilled for (4.1) at $x^*$, then, there are $u^* \in \mathbb{R}^m$, $v^* \in \mathbb{R}^q$ such that $(x^*, u^*, v^*)$ fulfills (KKT) for (4.1).*

**Proof:** Condition (K-T) evidently implies condition (4.25).
Therefore according to Theorem 4.19, $\mathsf{Z}(x^*) = \emptyset$.
That is equivalent with (KKT), according to Theorem 4.18.

Q.E.D.

Kuhn-Tucker constraint qualification is often in power.

**Definition 4.23** *We say, (4.1) fulfills affine constraint qualification (Affine) if all functions $g_j$, $j \in \mathsf{J}$, $h_k$, $k \in \mathsf{K}$ are affine.*

**Theorem 4.24**: *If (4.1) fulfills (Affine), hence, (K-T) is fulfilled for (4.1) at each point of* $\mathbf{M}$.

**Proof:** Denote $g_j(x) = \langle \alpha_j, x \rangle + \beta_j$, $h_k(x) = \langle a_k, x \rangle + b_k$.
Take $\widetilde{x} \in \mathbf{M}$.
Let $z \in \mathbb{R}^n$, $z \neq 0$ fulfills

$$\langle \nabla g_j(\widetilde{x}), z \rangle = \langle \alpha_j, z \rangle \le 0 \quad \forall\, j \in \mathsf{J}_g(\widetilde{x}), \ \langle \nabla h_k(\widetilde{x}), z \rangle = \langle a_k, z \rangle = 0 \quad \forall\, k \in \mathsf{K}.$$

Define a vector function $\Psi(\theta) = \widetilde{x} + \theta z$, $\theta \in \mathbb{R}$.

Evidently, $\Psi(0) = \widetilde{x}$ and $\Psi'(0) = z$. We have to show, $\Psi(\theta) \in \mathbf{M}$ for all $\theta \ge 0$ small enough.

1. For each $j \in \mathsf{J}_g(\widetilde{x})$ and each $\theta \ge 0$ we have

$$g_j(\Psi(\theta)) = \langle \alpha_j, \Psi(\theta) \rangle + \beta_j = \langle \alpha_j, \widetilde{x} \rangle + \beta_j + \theta \langle \alpha_j, z \rangle = g_j(\widetilde{x}) + \theta \langle \alpha_j, z \rangle \le 0,$$

   since $g_j(\widetilde{x}) = 0$ and $\langle \alpha_j, z \rangle \le 0$.

2. For each $j \notin \mathsf{J}_g(\widetilde{x})$ and each $\theta \ge 0$ we have

$$g_j(\Psi(\theta)) = \langle \alpha_j, \Psi(\theta) \rangle + \beta_j = \langle \alpha_j, \widetilde{x} \rangle + \beta_j + \theta \langle \alpha_j, z \rangle = g_j(\widetilde{x}) + \theta \langle \alpha_j, z \rangle.$$

   We define

$$\begin{aligned} \theta_j &= +\infty && \text{if } \langle \alpha_j, z \rangle \le 0, \\ &= -\frac{g_j(\widetilde{x})}{\langle \alpha_j, z \rangle} && \text{if } \langle \alpha_j, z \rangle > 0. \end{aligned}$$

   Since $g_j(\widetilde{x}) < 0$, $\theta_j > 0$ and $g_j(\Psi(\theta)) \le 0$ for all $\theta \in [0, \theta_j]$.

3. For each $k = 1, 2, \ldots, q$ and each $\theta \ge 0$ we have

$$h_k(\Psi(\theta)) = \langle a_k, \Psi(\theta) \rangle + b_k = \langle a_k, \widetilde{x} \rangle + b_k + \theta \langle a_k, z \rangle = h_k(\widetilde{x}) + \theta \langle a_k, z \rangle = 0,$$

   since $h_k(\widetilde{x}) = 0$ and $\langle a_k, z \rangle = 0$.

Consequently, $\Psi(\theta) \in \mathbf{M}$ for all $\theta \in [0, \theta_0]$, where $\theta_0 = \min\{\theta_j, \ j \notin \mathsf{J}_g(\widetilde{x})\}$.

The function $\Psi$ fulfills (4.26), thus, the condition (K-T) is in power for (4.1) at each point of $\mathbf{M}$.

Q.E.D.

Recall, the set of all feasible solutions of (4.1) is a convex polyhedral set in (Affine) case.

**Theorem 4.25**: *Let $\mathcal{G} \supset \mathbf{M}$ be an open set, functions $g_j$, $j \in \mathsf{J}$, $h_k$, $k \in \mathsf{K}$ be real-valued on $\mathcal{G}$, $g_j$, $j \in \mathsf{J}$ be differentiable convex functions and $h_k$, $k \in \mathsf{K}$ be affine functions. If (Slater) is fulfilled then (K-T) is in power for (4.1) at each point of $\mathbf{M}$.*

**Proof:** See [1].

Q.E.D.

**Definition 4.26** *Let $\mathcal{G} \supset \mathbf{M}$ be an open set, functions $g_j$, $j \in \mathsf{J}$, $h_k$, $k \in \mathsf{K}$ be real-valued on $\mathcal{G}$, $\widetilde{x} \in \mathbf{M}$, functions $g_j$, $j \in \mathsf{J}$, $h_k$, $k \in \mathsf{K}$ possess gradients at $\widetilde{x}$. We say, condition of linear independence (LI) (cz. lineární nezávislost) is fulfilled for (4.1) at $\widetilde{x}$, whenever gradients*

$$\left(\nabla g_j\left(\widetilde{x}\right),\ j \in \mathsf{J}_g\left(\widetilde{x}\right), \nabla h_k\left(\widetilde{x}\right),\ k \in \mathsf{K}\right)$$

*are linearly independent.*

**Theorem 4.27**: *Let $\mathcal{G} \supset \mathbf{M}$ be an open set, functions $g_j$, $j \in \mathsf{J}$, $h_k$, $k \in \mathsf{K}$ be real-valued on $\mathcal{G}$, $\widetilde{x} \in \mathbf{M}$ and functions $g_j$, $j \in \mathsf{J}$, $h_k$, $k \in \mathsf{K}$ possess gradients at $\widetilde{x}$. If condition (LI) is fulfilled for (4.1) at $\widetilde{x}$ then (K-T) is fulfilled for (4.1) at $\widetilde{x}$.*

**Proof:** See [1].

Q.E.D.

**Remark 4.28:** Condition (4.8) of point regularity from Definition 4.1 is a particular case of (LI), especially if $\mathsf{J} = \emptyset$.

♠

Let us introduce an example fulfilling neither (LI) nor (K-T), nevertheless, $\mathsf{Z}\left(\hat{x}\right) = \emptyset$.

**Example 4.29:** Consider program

$$\min\left\{x_2 \,:\, (1-x_1)^3 - x_2 \geq 0\,,\ x_1 \geq 0, x_2 \geq 0\right\}. \tag{4.27}$$

Its set of all optimal solutions is $\mathbf{M}^\star = \{(t,0) \,:\, 0 \leq t \leq 1\}$.
Consider the point $x^* = (1,0) \in \mathbf{M}^\star$.

Denote functions

$$f\left(x\right) = x_2,\ g_1\left(x\right) = -(1-x_1)^3 + x_2,\ g_2\left(x\right) = -x_1,\ g_3\left(x\right) = -x_2.$$

Lagrange function is of shape

$$L\left(x;u\right) = f\left(x\right) + u_1 g_1\left(x\right) + u_2 g_2\left(x\right) + u_3 g_3\left(x\right) = x_2 - u_1[(1-x_1)^3 - x_2] - u_2 x_1 - u_1 x_2.$$

All functions possess gradients at $x^*$

$$\nabla f\left(x^*\right) = \begin{pmatrix} 0 \\ 1 \end{pmatrix},\ \nabla g_1\left(x^*\right) = \begin{pmatrix} 0 \\ 1 \end{pmatrix},\ \nabla g_2\left(x^*\right) = \begin{pmatrix} -1 \\ 0 \end{pmatrix},\ \nabla g_3\left(x^*\right) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}.$$

- Condition (LI) is not fulfilled for (4.27) at $x^*$, because $\mathsf{J}_g(x^*) = \{1,3\}$ and gradients $\nabla g_1(x^*)$, $\nabla g_3(x^*)$ are linearly dependent, precisely, $\nabla g_1(x^*) = -\nabla g_3(x^*)$.

- Condition (K-T) is not fulfilled for (4.27) at $x^*$. For example direction $z = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ fulfills

$$\langle \nabla g_1(x^*), z \rangle = \left\langle \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\rangle = 0, \ \langle \nabla g_3(x^*), z \rangle = \left\langle \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix} \right\rangle = 0.$$

  Assume a vector function $\Psi : (\underline{\theta}, \overline{\theta}) \to \mathcal{G}$ and $\lambda > 0$ such that $\Psi(0) = x^*$, $\Psi'(0) = \lambda z = \begin{pmatrix} \lambda \\ 0 \end{pmatrix}$.
  Hence, the first coordinate of $\Psi$ must fulfill $\Psi(\theta)_1 > 1$ for all $\theta > 0$ sufficiently small.
  Thus, $\Psi(\theta) \notin \mathbf{M}$ for all $\theta > 0$ sufficiently small.
  Finally, direction $z = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ violates (K-T).

- The approximative set of improving directions is

$$\begin{aligned} \mathsf{Z}(x^*) &= \left\{ z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} : \langle \nabla f(x^*), z \rangle < 0 , \ \langle \nabla g_1(x^*), z \rangle \le 0 , \ \langle \nabla g_3(x^*), z \rangle \le 0 \right\} \\ &= \left\{ z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} : \left\langle \begin{pmatrix} 0 \\ 1 \end{pmatrix}, z \right\rangle = z_2 < 0 , \ \left\langle \begin{pmatrix} 0 \\ 1 \end{pmatrix}, z \right\rangle = z_2 \le 0 , \right. \\ &\qquad \left. \left\langle \begin{pmatrix} 0 \\ -1 \end{pmatrix}, z \right\rangle = -z_2 \le 0 \right\} \\ &= \emptyset. \end{aligned}$$

  According to Theorem 4.18, there is $u^* \in \mathbb{R}^3$ such that $(x^*, u^*)$ fulfills (KKT).
  For example, such a vector is $u^* = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$.

△

Condition (KKT) needs some additional property to indicate a local minimum of (4.1).

**Definition 4.30** *Let $\mathcal{G} \supset \mathbf{M}$ be an open set, functions $f$, $g_j$, $j \in \mathsf{J}$, $h_k$, $k \in \mathsf{K}$ be real-valued on $\mathcal{G}$, $x \in \mathbf{M}$, $u \in \mathbb{R}^m$, $v \in \mathbb{R}^q$, functions $f$, $g_j$, $j \in \mathsf{J}$, $h_k$, $k \in \mathsf{K}$ be twice differentiable at $x$.*

*We say, $(x, u, v)$ fulfills Second Order Sufficient Condition (SOSC) (cz. postačující podmínka 2.řádu), if $u \ge 0$ and for all directions $z \in \mathsf{V}(x, u)$ we have $\langle H_L(x; u, v) z, z \rangle > 0$.*

**Theorem 4.31** (KKT ⇒ locNLP): *Let $\mathcal{G} \supset \mathbf{M}$ be an open set, functions $f$, $g_j$, $j \in \mathsf{J}$, $h_k$, $k \in \mathsf{K}$ be real-valued on $\mathcal{G}$, $x^* \in \mathbf{M}$, $u^* \in \mathbb{R}^m$, $v^* \in \mathbb{R}^q$, functions $f$, $g_j$, $j \in \mathsf{J}$, $h_k$, $k \in \mathsf{K}$ be twice differentiable at $x^*$.*

*If $(x^*, u^*, v^*)$ fulfills (KKT) and (SOSC) for (4.1), then $x^*$ is a strict local minimum of (4.1).*

**Proof:** Statement will be proved by a contradiction.

Assume, $x^*$ is not a strict local minimum of (4.1).
Hence, there is a sequence $x_\beta \in \mathbf{M}$, $x_\beta \neq x^*$, $\beta \in \mathbb{N}$ converging to $x^*$ with $f(x_\beta) \le f(x^*)$ for all $\beta \in \mathbb{N}$.
Select a subsequence $x_{\beta_\gamma}$, $\gamma \in \mathbb{N}$ such that

$$\frac{x_{\beta_\gamma} - x^*}{\left\| x_{\beta_\gamma} - x^* \right\|} \xrightarrow[\gamma \to +\infty]{} \eta.$$

All functions are (even twice) differentiable at $x^*$, therefore, we are receiving expansions

$$\begin{aligned}
&0 \geq f\left(x_{\beta_\gamma}\right) - f\left(x^*\right) = \left\langle \nabla f\left(x^*\right), x_{\beta_\gamma} - x^* \right\rangle + \left\| x_{\beta_\gamma} - x^* \right\| R_1\left(x_{\beta_\gamma} - x^*; f, x^*\right) , \\
&0 \geq g_j\left(x_{\beta_\gamma}\right) = g_j\left(x_{\beta_\gamma}\right) - g_j\left(x^*\right) = \left\langle \nabla g_j\left(x^*\right), x_{\beta_\gamma} - x^* \right\rangle + \left\| x_{\beta_\gamma} - x^* \right\| R_1\left(x_{\beta_\gamma} - x^*; g_j, x^*\right) \\
&\qquad \text{for all} \quad j \in \mathsf{J}_g\left(x^*\right) , \\
&0 = h_k\left(x_{\beta_\gamma}\right) = h_k\left(x_{\beta_\gamma}\right) - h_k(x^*) = \left\langle \nabla h_k\left(x^*\right), x_{\beta_\gamma} - x^* \right\rangle + \left\| x_{\beta_\gamma} - x^* \right\| R_1\left(x_{\beta_\gamma} - x^*; h_k, x^*\right) \\
&\qquad \text{for all} \quad k \in \mathsf{K} .
\end{aligned}$$

After division by $\left\| x_{\beta_\gamma} - x^* \right\|$ and letting $\gamma \to +\infty$ we are receiving

$$\begin{aligned}
&0 \geq \langle \nabla f\left(x^*\right), \eta \rangle , \\
&0 \geq \langle \nabla g_j\left(x^*\right), \eta \rangle \quad \forall\, j \in \mathsf{J}_g\left(x^*\right) , \\
&0 = \langle \nabla h_k\left(x^*\right), \eta \rangle \quad \forall\, k \in \mathsf{K} .
\end{aligned}$$

Condition (KKT) is assumed for (4.1) at $x^*$, thus, condition $(DFC)$ is in power, which means

$$0 = \nabla_x L\left(x^*; u^*, v^*\right) = \nabla f\left(x^*\right) + \sum_{j=1}^{m} u_j^* \nabla g_j\left(x^*\right) + \sum_{k=1}^{q} v_k^* \nabla h_k\left(x^*\right) .$$

After that,

$$\begin{aligned}
&0 = \langle \nabla_x L\left(x^*; u^*, v^*\right), \eta \rangle = \langle \nabla f\left(x^*\right), \eta \rangle + \sum_{j=1}^{m} u_j^* \langle \nabla g_j\left(x^*\right), \eta \rangle + \sum_{k=1}^{q} v_k^* \langle \nabla h_k\left(x^*\right), \eta \rangle \\
&= \langle \nabla f\left(x^*\right), \eta \rangle + \sum_{j=1}^{m} u_j^* \langle \nabla g_j\left(x^*\right), \eta \rangle \leq \langle \nabla f\left(x^*\right), \eta \rangle .
\end{aligned}$$

Consequently,

$$\langle \nabla f\left(x^*\right), \eta \rangle = 0, \ \langle \nabla g_j\left(x^*\right), \eta \rangle = 0 \quad \forall\, j \in \mathsf{J}_g\left(x^*\right) , \ u_j^* > 0 .$$

We observe $\eta \in \mathsf{V}\left(x^*, u^*\right)$.
All functions are twice differentiable at $x^*$, therefore, Lagrange's function is also twice differentiable at $x^*$ and we have

$$\begin{aligned}
0 \quad &\geq \quad f\left(x_{\beta_\gamma}\right) - f\left(x^*\right) \\
&= \quad L\left(x_{\beta_\gamma}; u^*, v^*\right) - L\left(x^*; u^*, v^*\right) + \sum_{j=1}^{m} u_j^* \left(g_j\left(x^*\right) - g_j\left(x_{\beta_\gamma}\right)\right) + \sum_{k=1}^{q} v_k^* \left(h_k\left(x^*\right) - h_k\left(x_{\beta_\gamma}\right)\right) \\
&= \quad L\left(x_{\beta_\gamma}; u^*, v^*\right) - L\left(x^*; u^*, v^*\right) - \sum_{j=1}^{m} u_j^* g_j\left(x_{\beta_\gamma}\right) \\
&\geq \quad L\left(x_{\beta_\gamma}; u^*, v^*\right) - L\left(x^*; u^*, v^*\right) \\
&= \quad \left\langle \nabla_x L\left(x^*; u^*, v^*\right), x_{\beta_\gamma} - x^* \right\rangle + \frac{1}{2} \left\langle H_L\left(x^*; u^*, v^*\right)\left(x_{\beta_\gamma} - x^*\right), x_{\beta_\gamma} - x^* \right\rangle \\
&\qquad + \left\| x_{\beta_\gamma} - x^* \right\|^2 R_2\left(x_{\beta_\gamma} - x^*; L, x^*\right) \\
&= \quad \frac{1}{2} \left\langle H_L\left(x^*; u^*, v^*\right)\left(x_{\beta_\gamma} - x^*\right), x_{\beta_\gamma} - x^* \right\rangle + \left\| x_{\beta_\gamma} - x^* \right\|^2 R_2\left(x_{\beta_\gamma} - x^*; L, x^*\right) .
\end{aligned}$$

Dividing by $\left\|x_{\beta_\gamma} - x^*\right\|^2$ and letting $\gamma \to +\infty$, we are receiving

$$0 \geq \frac{1}{2} \left\langle H_L \left(x^*; u^*, v^*\right) \eta, \eta \right\rangle .$$

That is in contradiction with (SOSC).

Thus, we have proved $x^*$ is a strict local minimum of (4.1).

Q.E.D.

**Theorem 4.32** (KKT ⇒ locSPC): *Let $\mathcal{G} \supset \mathbf{M}$ be an open set, functions $f$, $g_j$, $j \in \mathsf{J}$, $h_k$, $k \in \mathsf{K}$ be real-valued on $\mathcal{G}$, $x^* \in \mathbf{M}$, functions $f$, $g_j$, $j \in \mathsf{J}$ possess gradients at $x^*$. Let there be $\delta > 0$ such that $\mathcal{U}_\delta(x^*) \subset \mathcal{G}$, functions $f, g_j, j \in \mathsf{J}$ be convex on $\mathcal{U}_\delta(x^*)$ and functions $h_k, k \in \mathsf{K}$ be affine on $\mathcal{U}_\delta(x^*)$.*

*If there are $u^* \in \mathbb{R}^m$, $v^* \in \mathbb{R}^q$ such that $(x^*, u^*, v^*)$ fulfills (KKT) for (4.1), then $(x^*, u^*, v^*)$ fulfills (locSPC) for (4.1) on the region $\mathcal{G} \times \mathbb{R}^m \times \mathbb{R}^q$. Moreover, $x^*$ is a local minimum of (4.1) and $(CSC)$ is in power.*

**Proof:** We know, that Lagrange function is convex in $x$ on $\mathcal{U}_\delta(x^*)$. (KKT) is assumed, therefore, $(CSC)$ is in power and for all $x \in \mathcal{U}_\delta(x^*)$ we have

$$L(x; u^*, v^*) \geq L(x^*; u^*, v^*) + \langle \nabla L(x^*; u^*, v^*), x - x^* \rangle = L(x^*; u^*, v^*) .$$

For $u \in \mathbb{R}^{\mathsf{J}}$, $u \geq \mathbf{0}$, $v \in \mathbb{R}^{\mathsf{K}}$ we have

$$\begin{aligned} L(x^*; u, v) &= L(x^*; u^*, v^*) + \langle u - u^*, g(x^*) \rangle + \langle v - v^*, h(x^*) \rangle \\ &= L(x^*; u^*, v^*) + \langle u, g(x^*) \rangle \leq L(x^*; u^*, v^*) . \end{aligned}$$

We derived, $(x^*, u^*, v^*)$ fulfills (locSPC) for (4.1) on the region $\mathcal{G} \times \mathbb{R}^m \times \mathbb{R}^q$.
According to Theorem 4.9, $x^*$ is a local minimum of (4.1) and the complementarity condition is fulfilled.

Q.E.D.

**Definition 4.33** *If for all $j \in \mathsf{J}_g(x^*)$ we have $u_j^* > 0$, then we are speaking on strict complementarity (cz. silná komplementarita).*

Consider particular cases.

**Example 4.34:** Let $f \in \mathsf{C}^2$, $\mathbf{M} = \{x \in \mathbb{R}^n : Ax \leq b\}$ and denote $g(x) = Ax - b$.
Lagrange function is

$$L(x; u) = \left\langle \begin{pmatrix} 1 \\ u \end{pmatrix}, \begin{pmatrix} f(x) \\ g(x) \end{pmatrix} \right\rangle = f(x) + \langle u, Ax - b \rangle$$

According to Theorem 4.24, condition (K-T) is fulfilled at each point of $\mathbf{M}$.

Condition (KKT) is:

i) $Ax^* - b \leq 0$.

ii) $\nabla f(x^*) + A^\top u^* = 0$.

iii) $\langle\, u^*, Ax^* - b\,\rangle = 0,$ .

Assume, $\mathbf{M} \neq 0$ and $\nabla^2_{x,x} f\,(x)$ positively definite for all $x \in \mathbb{R}^n$; thus $f$ is strictly convex. Hence, $\nabla^2_{x,x} L\,(x^*, u^*) = \nabla^2_{x,x} f\,(x^*)$ and

$$y^\top \nabla^2_{x,x} L\,(x^*, u^*)\, y = y^\top \nabla^2_{x,x} f\,(x^*)\, y > 0$$

for all $y \in \mathbb{R}^n, y \neq 0$. Thus, (SOSC) is fulfilled.

Finally for this program, $x^* \in \mathbb{R}^n$ is a local minimum of $f$ on $\mathbf{M}$ if and only if there is $u^* \geq 0$ such that $(x^*, u^*)$ fulfills (KKT).

△

**Example 4.35:** Consider a program $\min\{\langle\, c, x\,\rangle \,:\, Ax = b\,,\ x \geq 0\}$. Thus,

$$f\,(x) = \langle\, c, x\,\rangle\,,\ g\,(x) = x,\ h\,(x) = Ax - b,$$
$$L\,(x; u, v) = \langle\, c, x\,\rangle - \langle\, u, x\,\rangle + \langle\, v, Ax - b\,\rangle\,.$$

According to Theorem 4.24, condition (K-T) is fulfilled at each point of $\mathbf{M}$.

Condition (KKT) is:

i) $Ax^* = b,\ x^* \geq 0.$

ii) $c - u^* + A^\top v^* = 0.$

iii) $\langle\, u^*, x^*\,\rangle = 0,\ u^* \geq 0.$

From ii) follows $u^* = c + A^\top v^* \geq 0$, since $u^* \geq 0$. Plugging it in iii) we receive

$$0 = \langle\, u^*, x^*\,\rangle = \langle\, c, x^*\,\rangle + \left\langle\, A^\top v^*, x^*\,\right\rangle = \langle\, c, x^*\,\rangle + \langle\, v^*, Ax^*\,\rangle = \langle\, c, x^*\,\rangle + \langle\, v^*, b\,\rangle = \langle\, c, x^*\,\rangle + \langle\, b, v^*\,\rangle\,.$$

Denoting $y^* = -v^*$, (KKT) are transformed to

i) $Ax^* = b,\ x^* \geq 0.$

ii) $A^\top y^* \leq c.$

iii) $\langle\, c, x^*\,\rangle = \langle\, b, y^*\,\rangle.$

If $x^*$ is an optimal solution of (LP), then according to Theorem 4.22 it must fulfill (KKT). Which means, there is $y^*$ an optimal solution of a corresponding dual program.

Actually, we derived duality in Linear Programming. Deriving seems to be simple, based on Theorems 4.18 and 4.22, only. Unfortunately, it only seems to be simple. In proof of Theorem 4.22 we employed Farkas Theorem, which is equivalent to the duality in Linear Programming.

△

**Example 4.36:** Continue with Example 4.35 and recall denotation $\mathcal{P}(x^*) = \{i \in I \,:\, x_i^* > 0\}$.
In the Example, the set of indexes of all active constraints is $\mathsf{J}_g(x^*) = \{i \in I \,:\, x_i^* = 0\} = I \setminus \mathcal{P}(x^*)$.
Assume, strict complementarity for each $(x^*, u^*)$ fulfilling (KKT); i.e. for each $i \in I$ we have either $x_i^* > 0$, $u_i^* = 0$ or $x_i^* = 0$, $u_i^* > 0$. Problematic case $x_i^* = 0$, $u_i^* = 0$ is missing.
Hence, (KKT) with strict complementarity is saying, there is no $z \neq 0$ with properties $Az = \mathbf{0}$ and $z_j = 0$ for all $j \in \mathsf{J}_g(x^*)$.
It means, equation $A_{J\times\mathcal{P}(x^*)}\xi = \mathbf{0}$ possesses unique solution $\xi = \mathbf{0}$.
Hence, columns of the matrix $A_{J\times\mathcal{P}(x^*)}$ are linearly independent and, therefore, $x^*$ is a basic solution of the equation $Ax = b$; see a description of Simplex Method.

△

**Remark 4.37:** Conditions (KKT) for (4.1) can be written using gradients of Lagrange function (4.20):

i) $\nabla_u L(x^*, v^*, u^*) \leq 0$, $\nabla_v L(x^*, v^*, u^*) = 0$.

ii) $\nabla L(x^*, v^*, u^*) = 0$.

iii) $u^* \geq 0$, $\langle u^*, \nabla_u L(x^*, v^*, u^*) \rangle = 0$.

♠

## 4.5 Symmetric nonlinear programming

In this section, we introduce another class of optimization programs in which (KKT) condition can be easily formulated.

We will treat symmetric nonlinear program (cz. symetrická úloha NLP)

$$\min\{f(x) \,:\, g_j(x) \leq 0,\ j \in \mathsf{J},\ x_i \geq 0,\ i \in I\} \tag{4.28}$$

Such a program will be denoted by (SNLP).

Set of all feasible solutions is again denoted by

$$\mathbf{M} = \{x \in \mathbb{R}^n \,:\, g_j(x) \leq 0,\ j \in \mathsf{J},\ x_i \geq 0,\ i \in I\} \tag{4.29}$$

Consider, for a symmetric program the set of indexes of active constraints consists from two parts

$$\mathsf{J}_g(x) = \mathsf{J}_{g,g}(x) \cup \mathsf{J}_{g,x}(x)\ ,\ \text{kde } \mathsf{J}_{g,g}(x) = \{j \in \mathsf{J} \,:\, g_j(x) = 0\}\ ,\ \mathsf{J}_{g,x}(x) = \{i \in I \,:\, x_i = 0\}\,.$$

Approximative set of improving directions possesses a particular shape

$$\mathsf{Z}(x^*) = \{z \in \mathbb{R}^n \,:\, \langle \nabla g_j(x^*), z \rangle \leq 0,\ j \in \mathsf{J}_{g,g}(x^*)\,,\ z_i \geq 0,\ i \in \mathsf{J}_{g,x}(x^*)\ ,\ \langle \nabla f(x^*), z \rangle < 0\}\,.$$

For symmetric program (4.28), a trimmed shape of Lagrange function is useful. All constraints determined by functions $g_j$, $j \in \mathsf{J}$ are included in Lagrange function as penalization and all constraints given by nonnegativness of variables remain constraints in relaxed program. Trimmed Lagrange function is of shape

$$L^S(x; y) = \left\langle \begin{pmatrix} 1 \\ y \end{pmatrix}, \begin{pmatrix} f(x) \\ g(x) \end{pmatrix} \right\rangle = f(x) + \langle y, g(x) \rangle = f(x) + \sum_{j=1}^{m} y_j g_j(x) \tag{4.30}$$

defined for all $x \in \mathcal{G}$, $y \in \mathbb{R}^m$, $x \geq 0$, $y \geq 0$.

Global and local optimality conditions are as follows.

**Definition 4.38** *Let $\mathcal{G} \supset \mathbf{M}$ be an open set, functions $f$, $g_j$, $j \in \mathsf{J}$ be real-valued on $\mathcal{G}$, $x^* \in \mathbb{R}^n$, $y^* \in \mathbb{R}^m$. We say, $(x^*, y^*)$ fulfills saddle point condition (SSPC) for symmetric program (4.28) on the region $\left(\mathcal{G} \cap \mathbb{R}^n_{+,0}\right) \times \mathbb{R}^m_{+,0}$, whenever*

$$x^* \in \mathcal{G}\,,\ x^* \geq 0\,,\ y^* \geq 0$$

*and for all $x \in \mathcal{G}$, $y \in \mathbb{R}^m$, $x \geq 0$, $y \geq 0$ we have*

$$L^S(x; y^*) \geq L^S(x^*; y^*) \geq L^S(x^*; y)\,. \tag{4.31}$$

*Another often used terminology, $(x^*, y^*)$ is a saddle point of Lagrange function (4.30).*

**Definition 4.39** *Let $\mathcal{G} \supset \mathbf{M}$ be an open set, functions $f$, $g_j$, $j \in \mathsf{J}$ be real-valued on $\mathcal{G}$, $x^* \in \mathbb{R}^n$, $y^* \in \mathbb{R}^m$, functions $g_j$, $j \in \mathsf{J}$ possess gradients at $x^*$. We say, $(x^*, y^*)$ fulfills Karush-Kuhn-Tucker optimality conditions (SKKT) for symmetric program (4.28), whenever*

$$x^* \geq 0,\ \nabla_x L^S(x^*; y^*) \geq 0,\ \left\langle x^*, \nabla_x L^S(x^*; y^*) \right\rangle = 0, \tag{4.32}$$

$$y^* \geq 0,\ \nabla_y L^S(x^*; y^*) \leq 0,\ \left\langle y^*, \nabla_y L^S(x^*; y^*) \right\rangle = 0. \tag{4.33}$$

**Theorem 4.40** (SSPC ⇒ SNLP): *Let $\mathcal{G} \supset \mathbf{M}$ be an open set, functions $f$, $g_j$, $j \in \mathsf{J}$ be real-valued on $\mathcal{G}$ and $x^* \in \mathcal{G}$. If there is $y^* \in \mathbb{R}^m$ such that $(x^*, y^*)$ fulfills (SSPC) for the program (4.28), then $x^*$ is an optimal solution of (4.28).*

**Proof:** Statement is a particular case of Theorem 4.5.

Q.E.D.

**Theorem 4.41** (SSPC ⇒ SKKT): *Let $\mathcal{G} \supset \mathbf{M}$ be an open set, functions $f$, $g_j$, $j \in \mathsf{J}$ be real-valued on $\mathcal{G}$, $x^* \in \mathbb{R}^n$, $y^* \in \mathbb{R}^m$, functions $f$, $g_j$, $j \in \mathsf{J}$ possess gradients at $x^*$.*

*If $(x^*, y^*)$ fulfills (SSPC) on the region $\left(\mathcal{G} \cap \mathbb{R}^n_{+,0}\right) \times \mathbb{R}^q_{+,0}$, then the condition (SKKT) is also fulfilled.*

**Proof:** Assume $(x^*, y^*)$ fulfills (SSPC).

1. We know, $x^* \in \mathcal{G}$, $x^* \geq 0$ is a minimum of $L^S(\bullet; y^*)$ on $\mathcal{G}$ and $L^S(\bullet; y^*)$ possesses a gradient at $x^*$.

   For $i \in I$, partial derivative of Lagrange function with respect to $x_i$ fulfills

   $$\frac{\partial L^S}{\partial x_i}(x^*; y^*) = \lim_{t \to 0+} \frac{L^S(x^* + t\,\mathsf{e}_i; y^*) - L^S(x^*; y^*)}{t} \geq 0\,.$$

   If $x^*_i > 0$, we have another observation

   $$\frac{\partial L^S}{\partial x_i}(x^*; y^*) = \lim_{t \to 0-} \frac{L^S(x^* + t\,\mathsf{e}_i; y^*) - L^S(x^*; y^*)}{t} \leq 0\,.$$

   Thus, condition (4.32) is proved.

2. We know, $y^* \in \mathbb{R}^m$, $y^* \geq 0$ is a maximum of $L^S(x^*; \bullet)$ on $\mathbb{R}^m_{+,0}$ and $L^S(x^*; \bullet)$ is affine.

   For $j \in \mathsf{J}$, partial derivative of Lagrange function with respect to $y_j$ fulfills

   $$\frac{\partial L^S}{\partial y_k}(x^*; y^*) \lim_{t \to 0+} \frac{L^S(x^*; y^* + t\,\mathsf{e}_k) - L^S(x^*; y^*)}{t} \leq 0\,.$$

If $y_k^* > 0$, then we have another observation

$$\frac{\partial L^S}{\partial y_k}(x^*; y^*) = \lim_{t \to 0-} \frac{L^S(x^*; y^* + t\,\mathsf{e}_k) - L^S(x^*; y^*)}{t} \geq 0\,.$$

Thus, condition (4.33) is proved.

Thus, conditions (SKKT) are proved.

Q.E.D.

**Theorem 4.42** (SSPC $\Leftrightarrow$ SKKT): *Let $\mathcal{G} \supset \mathbf{M}$ be an open convex set, functions $f$, $g_j$, $j \in \mathsf{J}$ be real-valued and convex on $\mathcal{G}$, $x^* \in \mathbb{R}^n$, $y^* \in \mathbb{R}^m$, functions $f$, $g_j$, $j \in \mathsf{J}$ possess gradients at $x^*$.*

*Then, $(x^*, y^*)$ fulfills condition (SSPC) on the region $\left(\mathcal{G} \cap \mathbb{R}^n_{+,0}\right) \times \mathbb{R}^q_{+,0}$ if and only if $(x^*, y^*)$ fulfills condition (SKKT).*

**Proof:** According to Theorem 4.41, we know (SSPC) implies (SKKT). Thus, It is sufficient to show opposite implication. Assume for that, $(x^*, y^*)$ fulfills condition (SKKT).

1. Functions $L^S(\bullet; y^*)$ is convex on $\mathcal{G}$, therefore using (4.32), we have

$$\begin{aligned} L^S(x; y^*) &\geq L^S(x^*; y^*) + \left\langle \nabla_x L^S(x^*; y^*), x - x^* \right\rangle \\ &= L^S(x^*; y^*) + \left\langle \nabla_x L^S(x^*; y^*), x \right\rangle \geq L^S(x^*; y^*) \quad \forall\, x \in \mathcal{G}\,,\ x \geq 0. \end{aligned}$$

2. Functions $L^S(x^*; \bullet)$ are affine on whole $\mathbb{R}^m$, therefore using (4.33), we have

$$\begin{aligned} L^S(x^*; y) &= L^S(x^*; y^*) + \left\langle \nabla_y L^S(x^*; y^*), y - y^* \right\rangle \\ &= L^S(x^*; y^*) + \left\langle \nabla_y L^S(x^*; y^*), y \right\rangle \leq L^S(x^*; y^*) \quad \forall\, y \geq 0. \end{aligned}$$

We have proved $(x^*, y^*)$ fulfills condition (SSPC).

Q.E.D.

Consider an illustrative example.

**Example 4.43:** Consider program $\min\left\{-x : x^2 \leq 0, x \geq 0\right\}$.

The program possesses unique feasible solution $x^* = 0$. Thus $x^* = 0$ must be an optimal solution. Shape of trimmed Lagrange function is $L^S(x; y) = -x + yx^2$.

Assume, $(0, y^*)$ fulfills (SSPC), i.e.

$$y^* \geq 0,\ -x + y^* x^2 \geq 0 \quad \forall x \geq 0.$$

Hence,

$$y^* \geq \frac{1}{x} \quad \forall x > 0.$$

No such $y^*$ exists.

△

**Remark 4.44:** Let us specify meaning of (Slater) for (4.28); i.e. there is a point $\widetilde{x} \in \mathbb{R}^n$ with properties $\widetilde{x}_i > 0$ for each $i \in I$ and $g_j(\widetilde{x}) < 0$ for each $j \in \mathsf{J}$.

♠

**Theorem 4.45** (SNLP ⇒ SSPC): *Let $\mathcal{G} \supset \mathbf{M}$ be an open convex set, functions $f$, $g_j$, $j \in \mathsf{J}$ be real-valued and convex on $\mathcal{G}$ and (4.28) fulfill (Slater).*

*If $x^*$ is an optimal solution of (4.28), then there is $y^* \in \mathbb{R}^q$ such that $(x^*, y^*)$ fulfills condition (SSPC).*

**Proof:** Statement is a particular case of Theorem 4.7.

Q.E.D.

**Theorem 4.46** (SNLP ⇔ SKKT): *Let $\mathcal{G} \supset \mathbf{M}$ be an open convex set, functions $f$, $g_j$, $j \in \mathsf{J}$ be real-valued and convex on $\mathcal{G}$, $x^* \in \mathbf{M}$, functions $f$, $g_j$, $j \in \mathsf{J}$ possess gradients at $x^*$ and (4.28) fulfill (Slater).*

*Then, $x^*$ is minimum of (4.28) if and only if there is $y^* \geq 0$ such that $(x^*, y^*)$ fulfills (SKKT).*

**Proof:** Statement follows Theorems 4.42 and 4.45.

Q.E.D.

Conditions (SKKT) (hence also (KKT)) differ for maximum and minimum.

**Example 4.47:**

- Consider program $\min\left\{-x_1^2 - x_2^2 \,:\, x_1 + x_2 \leq 1, x_1 \geq 0, x_2 \geq 0\right\}$.

  Drawing the set of all feasible solutions (a triangle) and contour lines of objective function $\left\{x \in \mathbb{R}^2 \,:\, -x_1^2 - x_2^2 = r\right\}$, one can identify all minima points; i.e. $[1, 0]$ and $[0, 1]$.

  Corresponding trimmed Lagrange function is

  $$L^S(x; y) = -x_1^2 - x_2^2 + y(x_1 + x_2 - 1).$$

  Conditions (SKKT) for a couple $(x^*, y^*)$ are

  $$\nabla_x L^S(x^*; y^*) = \begin{pmatrix} -2x_1^* + y^* \\ -2x_2^* + y^* \end{pmatrix} \geq 0,\; x_1^* \geq 0,\; x_2^* \geq 0, \tag{4.34}$$
  $$\left\langle x^*, \nabla_x L^S(x^*; y^*) \right\rangle = -2(x_1^*)^2 + x_1^* y^* - 2(x_2^*)^2 + x_2^* y^* = 0,$$
  $$\nabla_y L^S(x^*; y^*) = x_1^* + x_2^* - 1 \leq 0,\; y^* \geq 0, \tag{4.35}$$
  $$\left\langle y^*, \nabla_y L^S(x^*; y^*) \right\rangle = y^*(x_1^* + x_2^* - 1) = 0.$$

  All solutions $(x_1^*, x_2^*; y^*)$ of (SKKT) are $[0, 0; 0]$, $[0, 1; 2]$, $[\frac{1}{2}, \frac{1}{2}; 1]$, $[1, 0; 2]$.

- Consider program $\max\left\{-x_1^2 - x_2^2 \,:\, x_1 + x_2 \leq 1, x_1 \geq 0, x_2 \geq 0\right\}$.

  We will investigate equivalent program $\min\left\{x_1^2 + x_2^2 \,:\, x_1 + x_2 \leq 1, x_1 \geq 0, x_2 \geq 0\right\}$.

  Objective function is strictly convex and set of all feasible solutions is convex (actually a triangle). Hence, the problem possesses uniquely determined global (also local) minimum $[0, 0]$.

  Corresponding trimmed Lagrange function is

  $$L^S(x; y) = x_1^2 + x_2^2 + y(x_1 + x_2 - 1).$$

Conditions (SKKT) for a couple $(x^*, y^*)$ are

$$\nabla_x L^S(x^*; y^*) = \begin{pmatrix} 2x_1^* + y^* \\ 2x_2^* + y^* \end{pmatrix} \geq 0,\ x_1^* \geq 0,\ x_2^* \geq 0, \tag{4.36}$$

$$\left\langle x^*, \nabla_x L^S(x^*; y^*) \right\rangle = 2(x_1^*)^2 + x_1^* y^* + 2(x_2^*)^2 + x_2^* y^* = 0,$$

$$\nabla_y L^S(x^*; y^*) = x_1^* + x_2^* - 1 \leq 0,\ y^* \geq 0, \tag{4.37}$$

$$\left\langle y^*, \nabla_y L^S(x^*; y^*) \right\rangle = y^*(x_1^* + x_2^* - 1) = 0.$$

Conditions (SKKT) possess unique solution $(x_1^*, x_2^*; y^*) = [0, 0, 0]$.

△

## 4.6 Importance of Lagrange coefficients

The last thing we should touch in this text is importance or interpretation of Lagrange coefficients. We will see, they allow us to estimate value of objective function after a certain change of considered program.

**Theorem 4.48** (Interpretation of Lagrange coefficients): *Let $f, g_j : \mathbb{R}^n \to \mathbb{R}$ be convex functions for all $j \in \mathsf{J}$ and $b^{(1)}, b^{(2)} \in \mathbb{R}^m$. Consider two symmetric programs*

$$\min \left\{ f(x)\ :\ g_j(x) \leq b_j^{(1)}, j \in \mathsf{J},\ , x \geq 0\,,\ x \in \mathbb{R}^n \right\}, \tag{4.38}$$

$$\min \left\{ f(x)\ :\ g_j(x) \leq b_j^{(2)},\ j \in \mathsf{J},\ , x \geq 0\,,\ x \in \mathbb{R}^n \right\}. \tag{4.39}$$

*Assume, (Slater) is in power for both programs and both programs possess an optimal solution.*

*Select and denote $x^{(1)}, x^{(2)}$ an optimal solution of (4.38) resp. of (4.39) and $y^{(1)}, y^{(2)}$ corresponding Lagrange coefficients, then we have*

$$\left\langle y^{(1)}, b^{(1)} - b^{(2)} \right\rangle \leq f\left(x^{(2)}\right) - f\left(x^{(1)}\right) \leq \left\langle y^{(2)}, b^{(1)} - b^{(2)} \right\rangle. \tag{4.40}$$

**Proof:** Considered programs are nice, their optimal solutions are fully characterized by (SSPC) and $\mathcal{G}$ can be taken as $\mathbb{R}^n$. Write Lagrange function for the program (4.38)

$$L^{(1)}(x; y) = f(x) + \left\langle y, g(x) - b^{(1)} \right\rangle.$$

The couple $(x^{(1)}, y^{(1)})$ fulfills (SSPC). The left-hand side of it is

$$f(x) + \left\langle y^{(1)}, g(x) - b^{(1)} \right\rangle \geq f\left(x^{(1)}\right) + \left\langle y^{(1)}, g\left(x^{(1)}\right) - b^{(1)} \right\rangle \quad \forall x \geq 0\,.$$

Using complementarity $\left\langle y^{(1)}, g\left(x^{(1)}\right) - b^{(1)} \right\rangle = 0$, we are receiving

$$f(x) + \left\langle y^{(1)}, g(x) - b^{(1)} \right\rangle \geq f\left(x^{(1)}\right) \quad \forall x \geq 0\,. \tag{4.41}$$

Particularly $x^{(2)} \geq 0$ can by plugged in (4.41)

$$f\left(x^{(2)}\right) + \left\langle y^{(1)}, g\left(x^{(2)}\right) - b^{(1)} \right\rangle \geq f\left(x^{(1)}\right).$$

After rewriting and using $g\left(x^{(2)}\right) \leq b^{(2)}$, we are receiving

$$f\left(x^{(2)}\right) - f\left(x^{(1)}\right) \geq \left\langle y^{(1)}, b^{(1)} - g\left(x^{(2)}\right)\right\rangle \geq \left\langle y^{(1)}, b^{(1)} - b^{(2)}\right\rangle .$$

That is the left-hand side of (4.40).
Reversing order of programs we are receiving

$$f\left(x^{(1)}\right) - f\left(x^{(2)}\right) \geq \left\langle y^{(2)}, b^{(2)} - b^{(1)}\right\rangle .$$

That is the right-hand side of (4.40).

Q.E.D.

Estimates presented in Theorem 4.48 possess economical interpretation, application. They allow us to estimate behavior of optimal value of objective function after a change of constraints; e.g. increasing or decreasing of demands; change of rules given by government; increasing or reduction of firm production; enlarging or decreasing of storage capacity; etc.

# Bibliography


[1] Bazara, M.S.; Sherali, H.D.; Shetty, C.M.: *Nonlinear Programming. Theory and Algorithms.* 2nd edition, Wiley, New York, 1993.

[2] Plesník, J.; Dupačová, J.; Vlach, M.: Lineárne programovanie, Alfa, Bratislava, 1990.

[3] Habala, P.; Hájek, P.; Zizler, V.: *Introduction to Banach Spaces I-II.* MatfyzPress, Praha, 1996.

[4] Lachout, P.: *Teorie pravděpodobnosti.* Skripta MFF UK, Praha, 1998.

[5] Rockafellar, T.: *Convex Analysis.*, Springer-Verlag, Berlin, 1975.

[6] Rockafellar, T.; Wets, R. J.-B.: *Variational Analysis.* Springer-Verlag, Berlin, 1998.

# Index