Smoothness of real functions

Petr Lachout

Version: February 26, 2017

Contents

Used symbols

$\mathbb{R}$ ...... real line, Euclidean space
$\mathbb{R}^*$ ...... extended real line
$\mathbb{N}$ ...... natural numbers

$\mathsf{e}_{i:n} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \leftarrow i^{th}\text{row}$ ...... i^{th} unit vector of dimension n

$\operatorname{int}(A)$ ...... interior of a set A
$\operatorname{rint}(A)$ ...... relative interior of a set A
$\operatorname{clo}(A)$ ...... closure of a set A
$\partial(A)$ ...... boundary of a set A
$\mathsf{Dom}(f)$ ...... domain of a function f
$\mathsf{graph}(f)$ ...... graph of a function f
$\mathsf{epi}(f)$ ...... epigraph of a function f
$\mathsf{hypo}(f)$ ...... hypograph of a function f
$f_{x,s}$ ...... a function f restricted to a line going through x with direction s
$\mathsf{D}_{x,s}$ ...... definition region of $f_{x,s}$

Chapter 1

General notions

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

Definition 1.1 *For a function $f : \mathbb{R}^n \to \mathbb{R}^*$, we define its epigraph (cz. epigraf) and hypograph (cz. hypograf)*

$$\begin{aligned} \mathsf{epi}\,(f) &= \left\{ \begin{pmatrix} x \\ \eta \end{pmatrix} : f(x) \le \eta,\ x \in \mathbb{R}^n,\ \eta \in \mathbb{R} \right\}, && (1.1) \\ \mathsf{hypo}\,(f) &= \left\{ \begin{pmatrix} x \\ \eta \end{pmatrix} : f(x) \ge \eta,\ x \in \mathbb{R}^n,\ \eta \in \mathbb{R} \right\} && (1.2) \end{aligned}$$

and its domain (cz. doména)

$$\mathsf{Dom}\,(f) = \{x : f(x) < +\infty,\ x \in \mathbb{R}^n\}. \tag{1.3}$$

Definition 1.2 *We say, function $f : \mathbb{R}^n \to \mathbb{R}^*$ is proper (cz. vlastní), if $\mathsf{Dom}\,(f) \ne \emptyset$ 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 $\inf\left\{\widetilde{f}(x) : x \in \mathbb{R}^n\right\}$, where

$$\begin{aligned} \widetilde{f}(x) &= f(x) \quad \text{if } x \in D, && (1.4) \\ &= +\infty \quad \text{otherwise.} && (1.5) \end{aligned}$$

Epigraph of a function is a particular set, and, mapping between a function and its epigraph is a bijection.

Lemma 1.3 *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*

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

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.

Infimum and supremum of functions is related to union and intersection of epigraphs.

Lemma 1.4 *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,*

$$\mathsf{epi}\left(\sup_{i\in I} f_i\right) = \bigcap_{i\in I} \mathsf{epi}\left(f_i\right), \ \mathsf{hypo}\left(\inf_{i\in I} f_i\right) = \bigcap_{i\in I} \mathsf{hypo}\left(f_i\right), \tag{1.6}$$

$$\mathsf{epi}\left(\inf_{i\in I} f_i\right) \supset \bigcup_{i\in I} \mathsf{epi}\left(f_i\right), \ \mathsf{hypo}\left(\sup_{i\in I} f_i\right) \supset \bigcup_{i\in I} \mathsf{hypo}\left(f_i\right). \tag{1.7}$$

If I is finite, we receive equalities

$$\mathsf{epi}\left(\min_{i\in I} f_i\right) = \bigcup_{i\in I} \mathsf{epi}\left(f_i\right), \ \mathsf{hypo}\left(\max_{i\in I} f_i\right) = \bigcup_{i\in I} \mathsf{hypo}\left(f_i\right). \tag{1.8}$$

Proof: Statement is a direct consequence of Lemma 1.3. 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 a finite number 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.

Important role is played by restrictions of functions to lines.

Definition 1.5 *Let $\mathsf{D} \subset \mathbb{R}^n$, $\mathsf{D} \neq \emptyset$ and $f : \mathsf{D} \to \mathbb{R}$. For all $x \in \mathsf{D}$, $s \in \mathbb{R}^n$, we define a restriction of f to a line going through x with direction s (cz. restrikce funkce f na přímku) as $f_{x,s} : \mathsf{D}_{x,s} \to \mathbb{R} : t \in \mathsf{D}_{x,s} \mapsto f\left(x + ts\right)$, where $\mathsf{D}_{x,s} = \{t \,:\, x + ts \in \mathsf{D}, t \in \mathbb{R}\}$.*

To abbreviate notation, we will employ shifts of a set.

Definition 1.6 *Let $\mathsf{D} \subset \mathbb{R}^n$, $\mathsf{D} \neq \emptyset$ and $x \in \mathbb{R}^n$. We define D shifted to x (cz. posun množiny D) as $\tilde{\mathsf{D}}_x = \mathsf{D} - x = \{y - x \,:\, y \in \mathsf{D}\}$.*

Chapter 2

Differentiability of a function

2.1 On the real line

Definition 2.1 *Let $\mathsf{D} \subset \mathbb{R}$, $x \in \operatorname{int}(\mathsf{D})$ and $f : \mathsf{D} \to \mathbb{R}$. 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*

$$f(y) = f(x) + f'(x)(y-x) + |y-x| R_1(y-x; f, x), \tag{2.1}$$

where $R_1(\cdot; f, x) : \tilde{\mathsf{D}}_x \to \mathbb{R}$ and $\lim_{\substack{y \to x \\ y \in \mathsf{D}}} R_1(y-x; f, x) = 0$; for $\tilde{\mathsf{D}}_x$ see Definition 1.6.

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

Lemma 2.2 *If $\mathsf{D} \subset \mathbb{R}$, $x \in \operatorname{int}(\mathsf{D})$ and $f : \mathsf{D} \to \mathbb{R}$ is differentiable at x then f is continuous at x.*

Proof: Continuity of f at x follows immediately (2.1).

Q.E.D.

Lemma 2.3 *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,*

$$\int_a^b f'(s)\, \mathsf{d}s = f(b) - f(a). \tag{2.2}$$

2.2 Several arguments

Definition 2.4 *Let $\mathsf{D} \subset \mathbb{R}^n$, $x \in \operatorname{int}(\mathsf{D})$, $f : \mathsf{D} \to \mathbb{R}$ 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 $t \in \mathsf{D}_{x,h}$ we have*

$$f(x+th) = f(x) + f'(x;h)\, t + |t| R_1(t; f, x, h), \tag{2.3}$$

where $R_1(\cdot; f, x) : \mathsf{D}_{x,h} \to \mathbb{R}$ and $\lim_{\substack{t \to 0 \\ t \in \mathsf{D}_{x,h}}} R_1(t; f, x, h) = 0$; for $\mathsf{D}_{x,h}$ see Definition 1.5.

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 2.5 *Let $\mathsf{D} \subset \mathbb{R}^n$, $x \in \operatorname{int}(\mathsf{D})$, $f : \mathsf{D} \to \mathbb{R}$ and $i \in \{1, 2, \ldots, n\}$. 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 denote*

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

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

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

In the text, we are using differentiability of a function convenient for optimization, see e.g. [1], [5]. We will introduce necessary terminology and basic properties of differentiable functions.

Definition 2.6 *Let $\mathsf{D} \subset \mathbb{R}^n$, $x \in \operatorname{int}(\mathsf{D})$ and $f : \mathsf{D} \to \mathbb{R}$. We say, f is differentiable at x (or, possesses total differential at x, is 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*

$$f(y) = f(x) + \langle \nabla f(x), y - x \rangle + \|y - x\| R_1(y - x; f, x) , \tag{2.4}$$

where $R_1(\cdot; f, x) : \tilde{\mathsf{D}}_x \to \mathbb{R}$ and $\lim\limits_{\substack{y \to x \\ y \in \mathsf{D}}} R_1(y - x; f, x) = 0$.

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

Definition 2.7 *Let $\mathsf{D} \subset \mathbb{R}^n$, $x \in \operatorname{int}(\mathsf{D})$ and $f : \mathsf{D} \to \mathbb{R}$.*

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 gradient ∇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 gradient ∇f is continuous at $\mathcal{U}(x, \delta)$.

Gradient is necessary for expansion (2.4).

Lemma 2.8 *Let $\mathsf{D} \subset \mathbb{R}^n$, $x \in \operatorname{int}(\mathsf{D})$ and $f : \mathsf{D} \to \mathbb{R}$. Let f fulfill an expansion for all $y \in \mathsf{D}$*

$$f(y) = f(x) + \langle \xi, y - x \rangle + \|y - x\| \varphi(y - x), \tag{2.5}$$

where $\xi \in \mathbb{R}^n$, $\varphi : \tilde{\mathsf{D}}_x \to \mathbb{R}$ and $\lim\limits_{\substack{y \to x \\ y \in \mathsf{D}}} \varphi(y - x) = 0$.

Then f is differentiable at x, $\xi = \nabla f(x)$, $\varphi = R_1(\cdot; f, x)$ and $f'(x; h) = \langle \nabla f(x), h \rangle$ for all directions $h \in \mathbb{R}^n$.

Proof: Using (2.5) for a direction $h \in \mathbb{R}^n$ and $t \in \mathbb{R}$ small enough, we have

$$f(x + th) = f(x) + \langle \xi, th \rangle + \|th\| \varphi(th).$$

Consider derivative ratio and let $t \to 0$:

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

Setting $h = \mathsf{e}_{i:n}$, we receive $\xi_i = \frac{\partial f}{\partial x_i}(x)$.
We have verified ξ is the gradient of f at x, f is differentiable at x, $\varphi = R_1(\cdot; f, x)$ and directional derivatives possess announced form.

Q.E.D.

Lemma 2.9 *Let* $\mathsf{D} \subset \mathbb{R}^n$, $x \in \operatorname{int}(\mathsf{D})$ *and* $f : \mathsf{D} \to \mathbb{R}$.
If f *is differentiable at* x *then* f *is continuous at* x.

Proof: Continuity of f at x follows immediately (2.4).

Q.E.D.

There are nice consequences for restrictions to lines.

Lemma 2.10 *Let* $\mathsf{D} \subset \mathbb{R}^n$, $x \in \mathsf{D}$, $h \in \mathbb{R}^n$ *and* $f : \mathsf{D} \to \mathbb{R}$.

(i) *If* $t \in \mathbb{R}$, $x+th \in \operatorname{int}(\mathsf{D})$ *and* f *is differentiable at* $x+th$ *then directional derivative* $f'(x+th;h)$ *exists and* $f_{x,h}$ *is differentiable at* t *with*

$$f'_{x,h}(t) = f'(x+th;h) = \langle \nabla f(x+th), h \rangle . \tag{2.6}$$

(ii) *Let* $\alpha, \beta \in \mathbb{R}$, $\alpha < \beta$ *and* $x+th \in \operatorname{int}(\mathsf{D})$ *for all* $t \in (\alpha, \beta)$. *If* f *is differentiable at* $x+th$ *for all* $t \in (\alpha,\beta)$, $f_{x,h}$ *is right-continuous at* α *and left-continuous at* β *then*

$$\begin{aligned} f(x+\beta h) - f(x+\alpha h) &= f_{x,h}(\beta) - f_{x,h}(\alpha) = \int_\alpha^\beta f'_{x,h}(t)\, \mathsf{d}t \qquad (2.7) \\ &= \int_\alpha^\beta \langle \nabla f(x+th), h \rangle\, \mathsf{d}t. \end{aligned}$$

Proof: (i) follows Lemma 2.8 and (ii) is a consequence of Lemma 2.3.

Q.E.D.

2.3 Vector valued functions

Start with a curve.

Definition 2.11 *Let* $\mathsf{D} \subset \mathbb{R}$, $t \in \operatorname{int}(\mathsf{D})$ *and* $f : \mathsf{D} \to \mathbb{R}^m$. *Consider the function expressed as a vector of functions* $f = (f_1, f_2, \ldots, f_m)^\top$. *We say,*

- f *is* <u>*differentiable at* t</u> *if* f_j *is differentiable at* t *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* <u>*differentiable at* S</u> *if* f_j *is differentiable at* S *for each* $j \in \{1,2,\ldots,m\}$.

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

Definition 2.12 *Let* $n, m \in \mathbb{N}$, $x \in \mathbb{R}^n$ *and* $A \in \mathbb{R}^{n\times m}$. *We define denotation*

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

Using matrix notation, we have $\lessdot A, x \gtrdot = A^\top x$.

Definition 2.13 *Let $\mathsf{D} \subset \mathbb{R}^n$, $n \geq 2$, $x \in \operatorname{int}(\mathsf{D})$ and $f : \mathsf{D} \to \mathbb{R}^m$. Consider the function expressed 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 S if f_j is differentiable at S for each $j \in \{1, 2, \ldots, m\}$.*

Lemma 2.14 *Let $\mathsf{D} \subset \mathbb{R}^n$, $x \in \operatorname{int}(\mathsf{D})$ and $f : \mathsf{D} \to \mathbb{R}$. 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*

$$f(y) = f(x) + \lessdot \nabla f(x), y - x \gtrdot + \|y - x\| R_1(y - x; f, x), \tag{2.8}$$

where $R_1(\cdot; f, x) : \tilde{\mathsf{D}}_x \to \mathbb{R}^m$ and $\lim\limits_{\substack{y \to x \\ y \in \mathsf{D}}} R_1(y - x; f, x) = 0$.

The expression becomes more simple for a curve. Let $\mathsf{D} \subset \mathbb{R}$, $t \in \operatorname{int}(\mathsf{D})$ and $f : \mathsf{D} \to \mathbb{R}^m$. 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

$$f(s) = f(t) + (s - t) f'(t) + |s - t| R_1(s - t; f, t), \tag{2.9}$$

where $R_1(\cdot; f, t) : \tilde{\mathsf{D}}_t \to \mathbb{R}^m$ and $\lim\limits_{\substack{s \to t \\ s \in \mathsf{D}}} R_1(s - t; f, t) = 0$.

Proof: It is a straightforward rewriting of definition.

Q.E.D.

2.4 Chain rule

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

Lemma 2.15 *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*

$$(f \circ g)'(t) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(g(t))\, g_i'(t) = \langle \nabla f(g(t)), g'(t) \rangle. \tag{2.10}$$

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

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

where

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

Thus, $f \circ g$ is differentiable at t and (2.10) is shown.

Q.E.D.

2.5 Second derivative

Also, second derivative will be employed.

Definition 2.16 *Let $\mathsf{D} \subset \mathbb{R}^n$, $x \in \operatorname{int}(\mathsf{D})$ and $f : \mathsf{D} \to \mathbb{R}$. 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 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 by $\nabla^2 f(x) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x)\right)_{i=1,j=1}^{n,n}$ and called Hessian matrix.

Definition 2.17 *Let $\mathsf{D} \subset \mathbb{R}^n$, $x \in \operatorname{int}(\mathsf{D})$ and $f : \mathsf{D} \to \mathbb{R}$. We say, f is twice differentiable at x (or, possessing 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*

$$\begin{aligned} f(y) &= f(x) + \langle \nabla f(x), y - x \rangle + \frac{1}{2} \langle y - x, H_f(x)(y - x) \rangle \\ &\quad + \|y - x\|^2 R_2(y - x; f, x), \end{aligned} \tag{2.11}$$

where $R_2(\cdot; f, x) : \tilde{\mathsf{D}}_x \to \mathbb{R}$ and $\lim_{\substack{y \to x \\ y \in \mathsf{D}}} R_2(y - x; f, x) = 0$.

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

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

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

Lemma 2.18 *Let $\mathsf{D} \subset \mathbb{R}^n$, $x \in \operatorname{int}(\mathsf{D})$ and $f : \mathsf{D} \to \mathbb{R}$. If f is twice differentiable at x then matrix $H_f(x)$ is uniquely determined.*

Proof: Assume two symmetric matrices A, B such that for all $y \in \mathsf{D}$ we have

$$
\begin{aligned}
f(y) &= f(x) + \langle \nabla f(x), y-x \rangle + \frac{1}{2} \langle y-x, A(y-x) \rangle + \|y-x\|^2 \rho(y-x), \\
f(y) &= f(x) + \langle \nabla f(x), y-x \rangle + \frac{1}{2} \langle y-x, B(y-x) \rangle + \|y-x\|^2 \chi(y-x),
\end{aligned}
$$

where $\rho : \tilde{\mathsf{D}}_x \to \mathbb{R}$, $\chi : \tilde{\mathsf{D}}_x \to \mathbb{R}$ with $\lim_{\substack{y \to x \\ y \in \mathsf{D}}} \rho(y-x) = 0$, $\lim_{\substack{y \to x \\ y \in \mathsf{D}}} \chi(y-x) = 0$. Then,

$$
0 = \frac{1}{2} \langle y-x, (A-B)(y-x) \rangle + \|y-x\|^2 (\rho(y-x) - \chi(y-x)).
$$

Fix $h \in \mathbb{R}^n$. Hence, $x + \alpha h \in \mathsf{D}$ for $\alpha > 0$ sufficiently small, since $x \in \operatorname{int}(\mathsf{D})$. Then,

$$
0 = \frac{1}{2} \langle \alpha h, (A-B)\alpha h \rangle + \|\alpha h\|^2 (\rho(\alpha h) - \chi(\alpha h)).
$$

After multiplication with $\frac{2}{\alpha^2}$, we receive

$$
0 = \langle h, (A-B)h \rangle + 2 \|h\|^2 (\rho(\alpha h) - \chi(\alpha h)).
$$

Letting α vanish we receive

$$
0 = \langle h, (A-B)h \rangle \quad \text{for all} \quad h \in \mathbb{R}^n.
$$

That indicates $A = B$, because $A - B$ is symmetric.

Q.E.D.

Lemma 2.19 *Let $\mathsf{D} \subset \mathbb{R}^n$, $x \in \operatorname{int}(\mathsf{D})$ and $f : \mathsf{D} \to \mathbb{R}$. If f is differentiable at a neighborhood of x and ∇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

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

According to Lemma 2.10

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

Using expansion of gradient, we are receiving

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

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

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

$$
\lim_{t \to 0} \frac{f(x+th) - f(x) - t \langle \nabla f(x), h \rangle}{t^2} = \frac{1}{2} \langle h, H_f(x) h \rangle . \tag{2.12}
$$

(ii) *If* f *is differentiable at a neighborhood of* x *and* ∇f *is differentiable at* x*, then,* $\nabla^2 f(x)$ *exists and restriction* $f_{x,h}$ *possesses derivatives*

$$
f'_{x,h}(t) = \langle \nabla f(x+th), h \rangle \quad \textit{for all } t \textit{ small enough,} \tag{2.13}
$$

$$
f''_{x,h}(0) = \left\langle h, \nabla^2 f(x) h \right\rangle . \tag{2.14}
$$

Proof:

1. (i) follows (2.11), 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 2.19 and (2.4), (2.8), since for $s \neq 0$

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

Q.E.D.

2.6 Arguments for differentiability

Existence and continuity of gradient, resp. of Hessian, are sufficient conditions for differentiability in the sense of Definitions 2.6 and 2.17.

Lemma 2.21 *Let $I \subset \mathbb{R}$, $\mathrm{int}\,(I) \neq \emptyset$, $\mathsf{D} \subset \mathbb{R}^n$, $\mathrm{int}\,(\mathsf{D}) \neq \emptyset$, $g : I \to \mathsf{D}$, $f : \mathsf{D} \to \mathbb{R}$ and $t \in \mathrm{int}\,(I)$ such that $g(t) \in \mathrm{int}\,(\mathsf{D})$. If 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*

$$(f \circ g)'(t) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(g(t))\, g_i'(t) = \langle \nabla f(g(t)), g'(t) \rangle . \tag{2.15}$$

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

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

Then,

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

Divide formula by $s - t$ and let $s \to t$.
We receive formula (2.15), since gradient of f is continuous at $g(t)$.

Q.E.D.

Using Lemma 2.21, we derive differentiability of a function.

Lemma 2.22 *Let $\mathsf{D} \subset \mathbb{R}^n$, $x \in \mathrm{int}\,(\mathsf{D})$ and $f : \mathsf{D} \to \mathbb{R}$. If gradient of f exists on a neighborhood of $g(t)$ and is continuous at x, then f is differentiable at x with*

$$\begin{aligned} &f(x + h) = f(x) + \langle \nabla f(x), h \rangle + \|h\|\, R_1(h; f, x), \\ &|R_1(h; f, x)| \leq \sqrt{n} \max\{\|\nabla f(x + \chi) - \nabla f(x)\| : \chi \in \mathbb{R}^n, \|\chi\| \leq \|h\|\} \end{aligned} \tag{2.16}$$

if $h \in \mathbb{R}^n$ is sufficiently small.

Proof: For $h \in \mathbb{R}^n$, $i \in \{1, 2, \ldots, n\}$, $u \in [0, 1]$, we denote

$$\xi(u, x, h, i) = (x_1, \ldots, x_{i-1}, x_i + uh_i, x_{i+1} + h_{i+1}, \ldots, x_n + h_n,) .$$

For $h \in \mathbb{R}^n$ sufficiently small, we receive an expansion

$$\begin{aligned} f(x + h) - f(x) &= \sum_{i=1}^{n} [f(\xi(1, x, h, i)) - f(\xi(0, x, h, i))] \\ &= \sum_{i=1}^{n} h_i \int_0^1 \frac{\partial f}{\partial x_i}(\xi(u, x, h, i))\, \mathsf{d}u \\ &= \langle \nabla f(x), h \rangle + \sum_{i=1}^{n} h_i \int_0^1 \frac{\partial f}{\partial x_i}(\xi(u, x, h, i)) - \frac{\partial f}{\partial x_i}(x)\, \mathsf{d}u \\ &= \langle \nabla f(x), h \rangle + \|h\|\, R_1(h; f, x), \\ |R_1(h; f, x)| &\leq \sqrt{n} \max\{\|\nabla f(x + \chi) - \nabla f(x)\| : \chi \in \mathbb{R}^n, \|\chi\| \leq \|h\|\} . \end{aligned}$$

Q.E.D.

Lemma 2.23 *Let* $\mathsf{D} \subset \mathbb{R}^n$, $x \in \operatorname{int}(\mathsf{D})$ *and* $f : \mathsf{D} \to \mathbb{R}$. *Then,* f *is continuously differentiable at a neighborhood of* x *if and only if there is* $\delta > 0$ *such that* ∇f *exists on* $\mathcal{U}(x, \delta)$ *and is continuous at* $\mathcal{U}(x, \delta)$.

Proof: A consequence of Lemma 2.22.

Q.E.D.

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

$$\begin{aligned} f(x+h) &= f(x) + \langle \nabla f(x), h \rangle + \frac{1}{2} \langle h, \nabla^2 f(x) h \rangle + \|h\|^2 R_2(h; f, x), \qquad (2.17) \\ |R_2(h; f, x)| &\leq \frac{n}{2} \max \left\{ \left\| \nabla^2 f(x+\chi) - \nabla^2 f(x) \right\| : \chi \in \mathbb{R}^n, \|\chi\| \leq \|h\| \right\} \end{aligned}$$

if $h \in \mathbb{R}^n$ *is sufficiently small. Moreover,* $H_f(x) = \nabla^2 f(x)$.

Proof:

1. Symmetry of Hessian

   Take two coordinates $i, j \in \{1, 2, \ldots, n\}$, $i \neq j$ and for $\alpha, \beta \in \mathbb{R}$ consider:

$$\begin{aligned} & f(x + \alpha \mathsf{e}_{i:n} + \beta \mathsf{e}_{j:n}) - f(x + \alpha \mathsf{e}_{i:n}) - f(x + \beta \mathsf{e}_{j:n}) + f(x) \\ =\ & \beta \int_0^1 \frac{\partial f}{\partial x_j}(x + \alpha \mathsf{e}_{i:n} + u\beta \mathsf{e}_{j:n}) - \frac{\partial f}{\partial x_j}(x + u\beta \mathsf{e}_{j:n}) \, \mathsf{d}u \\ =\ & \alpha\beta \int_0^1 \left(\int_0^1 \frac{\partial^2 f}{\partial x_j \partial x_i}(x + v\alpha \mathsf{e}_{i:n} + u\beta \mathsf{e}_{j:n}) \, \mathsf{d}v \right) \mathsf{d}u. \end{aligned}$$

   Since second partial derivatives are continuous at x, we observe

$$\lim_{\substack{\alpha \to 0 \\ \beta \to 0}} \frac{f(x + \alpha \mathsf{e}_{i:n} + \beta \mathsf{e}_{j:n}) - f(x + \alpha \mathsf{e}_{i:n}) - f(x + \beta \mathsf{e}_{j:n}) + f(x)}{\alpha\beta} = \frac{1}{2} \frac{\partial^2 f}{\partial x_j \partial x_i}(x).$$

   By definitions of partial derivatives we have

$$\begin{aligned} \lim_{\alpha \to 0} \lim_{\beta \to 0} \frac{f(x + \alpha \mathsf{e}_{i:n} + \beta \mathsf{e}_{j:n}) - f(x + \alpha \mathsf{e}_{i:n}) - f(x + \beta \mathsf{e}_{j:n}) + f(x)}{\alpha\beta} &= \frac{1}{2} \frac{\partial^2 f}{\partial x_j \partial x_i}(x), \\ \lim_{\beta \to 0} \lim_{\alpha \to 0} \frac{f(x + \alpha \mathsf{e}_{i:n} + \beta \mathsf{e}_{j:n}) - f(x + \alpha \mathsf{e}_{i:n}) - f(x + \beta \mathsf{e}_{j:n}) + f(x)}{\alpha\beta} &= \frac{1}{2} \frac{\partial^2 f}{\partial x_i \partial x_j}(x). \end{aligned}$$

   Finally, $\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x)$.

2. For $h \in \mathbb{R}^n$, $i \in \{1, 2, \ldots, n\}$, $u \in [0, 1]$, we denote

$$\xi(u, x, h, i) \quad = \quad (x_1, \ldots, x_{i-1}, x_i + uh_i, x_{i+1} + h_{i+1}, \ldots, x_n + h_n,) \, .$$

For $h \in \mathbb{R}^n$ sufficiently small, we receive an expansion

$$\begin{aligned}
f\left(x+h\right) - f\left(x\right) &= \sum_{i=1}^{n} \left[f\left(\xi(1, x, h, i)\right) - f\left(\xi(0, x, h, i)\right) \right] \\
&= \sum_{i=1}^{n} h_i \int_0^1 \frac{\partial f}{\partial x_i} \left(\xi(u, x, h, i)\right) \mathsf{d}u \\
&= \left\langle \nabla f\left(x\right), h \right\rangle + \sum_{i=1}^{n} h_i \int_0^1 \frac{\partial f}{\partial x_i} \left(\xi(u, x, h, i)\right) - \frac{\partial f}{\partial x_i} \left(x\right) \mathsf{d}u \\
&= \left\langle \nabla f\left(x\right), h \right\rangle + \sum_{i=1}^{n} h_i \int_0^1 \frac{\partial f}{\partial x_i} \left(\xi(u, x, h, i)\right) - \frac{\partial f}{\partial x_i} \left(\xi(0, x, h, i)\right) \mathsf{d}u \\
&\quad + \sum_{i=1}^{n} h_i \left(\sum_{k=i+1}^{n} \int_0^1 \frac{\partial f}{\partial x_i} \left(\xi(1, x, h, k)\right) - \frac{\partial f}{\partial x_i} \left(\xi(0, x, h, k)\right) \mathsf{d}u \right) \\
&= \left\langle \nabla f\left(x\right), h \right\rangle + \sum_{i=1}^{n} h_i^2 \int_0^1 u \int_0^1 \frac{\partial^2 f}{\partial x_i \partial x_i} \left(\xi(uv, x, h, i)\right) \mathsf{d}v \, \mathsf{d}u \\
&\quad + \sum_{i=1}^{n} h_i \left(\sum_{k=i+1}^{n} h_k \int_0^1 \int_0^1 \frac{\partial^2 f}{\partial x_i \partial x_k} \left(\xi(v, x, h, k)\right) \mathsf{d}v \, \mathsf{d}u \right) \\
&= \left\langle \nabla f\left(x\right), h \right\rangle + \frac{1}{2} \left\langle h, \nabla^2 f\left(x\right) h \right\rangle \\
&\quad + \sum_{i=1}^{n} h_i^2 \int_0^1 u \int_0^1 \frac{\partial^2 f}{\partial x_i \partial x_i} \left(\xi(uv, x, h, i)\right) - \frac{\partial^2 f}{\partial x_i \partial x_i} \left(x\right) \mathsf{d}v \, \mathsf{d}u \\
&\quad + \sum_{i=1}^{n} \sum_{k=i+1}^{n} h_i h_k \int_0^1 \int_0^1 \frac{\partial^2 f}{\partial x_i \partial x_k} \left(\xi(v, x, h, k)\right) - \frac{\partial^2 f}{\partial x_i \partial x_k} \left(x\right) \mathsf{d}v \, \mathsf{d}u \\
&= \left\langle \nabla f\left(x\right), h \right\rangle + \frac{1}{2} \left\langle h, \nabla^2 f\left(x\right) h \right\rangle + \|h\|^2 R_2\left(h; f, x\right),
\end{aligned}$$

where

$$\left| R_2\left(h; f, x\right) \right| \quad \leq \quad \frac{n}{2} \max \left\{ \left\| \nabla^2 f\left(x + \chi\right) - \nabla^2 f\left(x\right) \right\| \, : \, \chi \in \mathbb{R}^n, \|\chi\| \leq \|h\| \right\} .$$

Hence, function f is twice differentiable, because Hessian is continuous at x.

Q.E.D.

Chapter 3

Convex functions

3.1 Definition of a convex function

Definition 3.1 *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 3.2 *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 3.3: *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*

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y)\,. \tag{3.1}$$

Proof:

1. Let f is convex.

   Then accordingly to Lemma 3.2, $\mathsf{Dom}\,(f)$ is a convex set.

   Let $x, y \in \mathsf{Dom}\,(f)$ and $0 < \lambda < 1$.

   Then for all $\eta, \xi \in \mathbb{R}$ fulfilling $f(x) \le \eta$ and $f(y) \le \xi$,

   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, $f(\lambda x + (1-\lambda)y) \le \lambda\eta + (1-\lambda)\xi < +\infty$.

   Minimum over all possible η, ξ is giving (3.1).

2. Let property (3.1) 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 3.3 shows, that new definition 3.1 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 convex function attaining value $-\infty$ is degenerated.

Lemma 3.4 *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 (3.1), we receive

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

Q.E.D.

Theorem 3.5: *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 $i \in \{1, 2, \ldots, n\}$.
$\mathsf{Dom}\,(f)$ is convex, therefore,

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

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

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

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

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

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

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

Finally, we are receiving

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

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

$$\mathsf{epi}(f) = \operatorname{clo}(\mathsf{epi}(f)) \cap (\mathsf{Dom}(f) \times \mathbb{R}) . \tag{3.2}$$

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

Q.E.D.

Theorem possesses a nice consequence.

Consiquence: 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 3.6, since $\mathsf{Dom}(f)$ is a closed set, and hence,

$$\mathsf{epi}(f) = \operatorname{clo}(\mathsf{epi}(f)) \cap (\mathsf{Dom}(f) \times \mathbb{R}) \quad \text{is a closed set.}$$

Q.E.D.

Theorem 3.7: *Let I be a nonempty index set and a convex function $f_i : \mathbb{R}^n \to \mathbb{R}^*$ is given for all $i \in I$. Then $\sup\limits_{i \in I} f_i : \mathbb{R}^n \to \mathbb{R}^*$ is also convex.*

Proof: According to Lemma 1.4 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, thus, $\sup_{i\in I} f_i$ is a convex function and a proof is done.

Q.E.D.

Definition 3.8 *Let $f : \mathbb{R}^n \to \mathbb{R}^*$ be a function. We define (lower) level sets of f (cz. (dolní) úrovňové množiny) $\mathsf{lev}_{\leq\Delta} f = \{x \in \mathsf{D} : f(x) \leq \Delta\}$, $\mathsf{lev}_{<\Delta} f = \{x \in \mathsf{D} : f(x) < \Delta\}$ for all $\Delta \in \mathbb{R}$, and, (upper) level sets of f (cz. (horní) úrovňové množiny) $\mathsf{lev}_{\geq\Delta} f = \{x \in \mathsf{D} : f(x) \geq \Delta\}$, $\mathsf{lev}_{>\Delta} f = \{x \in \mathsf{D} : f(x) > \Delta\}$ for all $\Delta \in \mathbb{R}$.*

Theorem 3.9: *If $f : \mathbb{R}^n \to \mathbb{R}^*$ is a convex function, then $\mathsf{lev}_{\leq\alpha} f$, $\mathsf{lev}_{<\alpha} f$ are convex sets for all $\alpha \in \mathbb{R}$.*

Proof: It is sufficient to verify $\mathsf{lev}_{<\alpha} f$ is convex, since $\mathsf{lev}_{\leq\alpha} f = \bigcap_{\beta>\alpha} \mathsf{lev}_{<\beta} f$.
Take $y, z \in \mathsf{lev}_{<\alpha} f$ and $0 < \lambda < 1$. Then, $y, z \in \mathsf{Dom}\,(f)$ and we have

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

It means $\lambda y + (1-\lambda)z \in \mathsf{lev}_{<\alpha} f$. Thus, $\mathsf{lev}_{<\alpha} f$ is a convex set.

Q.E.D.

As a consequence of Theorem 3.9, we are receiving that the set of all feasible solutions (cz. množina přípustných řešení) of a convex program is convex, i.e.

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

is a convex set, having functions $g_1, g_2, \ldots, g_k$ convex.
Convexity of level sets $\{x : f(x) \leq \alpha\}$, $\{x : f(x) < \alpha\}$ is not implying convexity of the function f.

Example 3.10: 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 $\{x : f(x) \leq \alpha\} = (0, \mathsf{e}^\alpha]$, $\{x : f(x) < \alpha\} = (0, \mathsf{e}^\alpha)$ are convex for all $\alpha \in \mathbb{R}$.

△

Definition 3.11 *We say, function $f : \mathbb{R}^n \to \mathbb{R}^*$ is*

i) strictly convex (cz. ryze konvexní), if $\mathsf{Dom}\,(f)$ is a nonempty convex set and 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.

Lemma 3.12 *If $f : \mathbb{R}^n \to \mathbb{R}^*$ is a strictly convex function with $\mathsf{Dom}(f)$ containing two different points then f is proper.*

Proof: Assume $x \in \mathsf{Dom}(f)$ with $f(x) = -\infty$.
There is $y \in \mathsf{Dom}(f)$, $y \neq x$.
Then, for all $\alpha \in (0,1)$ we have $f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha) f(y) = -\infty$.
That is impossible.
Therefore, f must be proper.

Q.E.D.

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

Theorem 3.13: *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 $\mathsf{Dom}(f)$.*
The set of all global minimizers f on $\mathsf{Dom}(f)$ is convex.

Proof: Let $\hat{x} \in \mathsf{Dom}(f)$ be a local minimum of f on $\mathsf{Dom}(f)$, but, it is not a 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 a contradiction, since $\hat{x}$ is a local minimum of f on $\mathsf{Dom}(f)$.
Hence, $\hat{x}$ is a global minimum of f on $\mathsf{Dom}(f)$.
Denote $\Delta = \inf\{f(y) : y \in \mathbb{R}^n\}$. The set of all global minimizers f on $\mathsf{Dom}(f)$.

$$\{x \in \mathsf{Dom}(f) : f(x) = \Delta\} \quad = \quad \{x \in \mathbb{R}^n : f(x) \leq \Delta\}$$

is convex, according to Theorem 3.9.

Q.E.D.

Theorem 3.14: *Let $f : \mathbb{R}^n \to \mathbb{R}^*$ be a strictly convex function with $\mathsf{Dom}(f)$ containing two different points. If there is a local minimum of f on $\mathsf{Dom}(f)$, then, it is uniquely determined and it is the unique global minimum of f on $\mathsf{Dom}(f)$.*

Proof: According to Theorem 3.13, 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.

3.2 Basic properties of convex functions

Let us recapitulate basic properties of convex functions.

3.2.1 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 on 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 a point or an interval.) Consider smoothness of convex functions.

Theorem 3.15: *Let $J \subset \mathbb{R}$ be an interval and $f : J \to \mathbb{R}$ be a convex function.*

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

ii) Derivative from left and from right exist at each point $t \in \operatorname{int}(J)$; 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 $t, s \in \operatorname{int}(J)$, $t < s$.

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

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 3.16: *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 possesses second derivative on J, then*
  f is convex $\Leftrightarrow$ $f'' \geq 0$ on J.

3.2.2 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 3.17 *Let $\mathsf{D} \subset \mathbb{R}^n$, $\mathsf{D} \neq \emptyset$ be a convex set, $f_1 : \mathsf{D} \to \mathbb{R}$, $f_2 : \mathsf{D} \to \mathbb{R}$, $\dots$, $f_k : \mathsf{D} \to \mathbb{R}$ be convex functions, $a_1 \geq 0$, $a_2 \geq 0$, $\dots$, $a_k \geq 0$. Then, $\sum_{i=1}^{k} a_i f_i : \mathsf{D} \to \mathbb{R}$ is a convex function.*

Proof: A proof is straightforward.

Q.E.D.

Theorem 3.18 (Jensen's 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 $\mathsf{P}(X \in \mathsf{D}) = 1$ 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 [2], Theorem 5.9, p.26.

Q.E.D.

A consequence of Theorem 3.18 is a generalization of inequality (3.1) ("deterministic Jensen's inequality").

Theorem 3.19: *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*

$$f\left(\sum_{i=1}^{k} p_i x_i\right) \leq \sum_{i=1}^{k} p_i f(x_i) \tag{3.3}$$

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 3.18. 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 restrictions to lines.

Theorem 3.20: *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 restriction $f_{x,s} : \mathsf{D}_{x,s} \to \mathbb{R}$ is convex for all $x \in \mathsf{D}$, $s \in \mathbb{R}^n$.*

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 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

$$\begin{aligned}
&f_{x,s}\left(\lambda t_1 + (1-\lambda)t_2\right) = \\
&= f\left(x + (\lambda t_1 + (1-\lambda)t_2)s\right) = f\left(\lambda(x + t_1 s) + (1-\lambda)(x + t_2 s)\right) \\
&\leq \lambda f\left(x + t_1 s\right) + (1-\lambda) f\left(x + t_2 s\right) = \lambda f_{x,s}\left(t_1\right) + (1-\lambda) f_{x,s}\left(t_2\right).
\end{aligned}$$

   We have verified $f_{x,s}$ is a convex function on an interval $\mathsf{D}_{x,s}$.

3. Let function $f_{x,s}$ 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

$$\begin{aligned}
&f\left(\lambda x + (1-\lambda)y\right) = f\left(y + \lambda s\right) = f_{y,s}\left(\lambda\right) \\
&\leq \lambda f_{y,s}\left(1\right) + (1-\lambda) f_{y,s}\left(0\right) = \lambda f\left(x\right) + (1-\lambda) f\left(y\right).
\end{aligned}$$

   We have verified f is a convex function.

Q.E.D.

This property enables us generalize criteria for convex function identification. Properties of the first and the second derivative of restrictions to lines were prepared in Lemmas 2.10, 2.20.

Theorem 3.21: *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 D. Then,*

$$f \text{ is convex} \quad \Leftrightarrow \quad \begin{array}{l} t \in \mathsf{D}_{x,s} \mapsto \langle \nabla f\left(x + ts\right), s \rangle \text{ is nondecreasing on } \mathsf{D}_{x,s} \\ \text{for all } x \in \mathsf{D},\ s \in \mathbb{R}^n. \end{array} \tag{3.4}$$

Proof: According to Theorem 3.20 we have to verify convexity of all restrictions to lines.
Take $x \in \mathsf{D}$, $s \in \mathbb{R}^n$ and consider function $f_{x,s}$.
Function f is differentiable at D, therefore according to Lemma 2.10, we have

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

Hence, according to Theorem 3.16

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

The statement is proved.

Q.E.D.

Theorem 3.22: *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 D and ∇f is differentiable at D, then, $\nabla^2 f$ exists on D, f is twice differentiable at D with*

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

and

$$f \text{ is convex} \quad \Leftrightarrow \quad H_f\left(x\right) \text{ is positively semidefinite for all } x \in \mathsf{D}. \tag{3.5}$$

Proof: According to Theorem 3.20 we have to verify convexity of all restrictions to lines.
Take $x \in \mathsf{D}$, $s \in \mathbb{R}^n$ and consider function $f_{x,s}$.
According to Lemma 2.20, we have

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

Hence,

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

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 3.23 *For a 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.*

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

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

Lemma 3.24 *For a 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 corner principle minors of matrix A are positive, i.e.*

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

- *There are a regular matrix Q and a diagonal matrix Λ with positive members on 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 3.25: *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 3.5.

Q.E.D.

Theorem 3.26: *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* D*, then*

$$f \text{ is convex} \iff \forall\, x, y \in \mathsf{D} \text{ we have } f(x) - f(y) \geq \langle \nabla f(y), x - y \rangle . \tag{3.6}$$

Proof:

1. Let f be convex.
   Take $x, y \in \mathsf{D}$ and denote $h = x - y$.
   According to Lemma 2.10, $f_{y,h}$ is a differentiable convex function on $\mathsf{D}_{y,h}$. Then, its derivative
   $$f'_{y,h}(\mu) \quad = \quad \langle \nabla f(y + \mu h), h \rangle \quad \text{is nondecreasing.}$$
   According to "Theorem on mean value" there is $\theta \in (0,1)$ such that
   $$\begin{aligned} f(x) - f(y) &= f_{y,h}(1) - f_{y,h}(0) = f'_{y,h}(\theta) \\ &\geq f'_{y,h}(0) = \langle \nabla f(y), h \rangle = \langle \nabla f(y), x - y \rangle , \end{aligned}$$
   since derivative 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 $v, w \in \mathsf{D}$, $\quad \lambda \in (0,1)$ and denote $z = \lambda v + (1-\lambda) w$.
   According to assumption we have:
   $$\begin{aligned} f(v) - f(z) \geq \langle \nabla f(z), v - z \rangle , \\ f(w) - f(z) \geq \langle \nabla f(z), w - z \rangle . \end{aligned}$$
   Hence,
   $$\begin{aligned} \lambda f(v) + (1-\lambda) f(w) &\geq f(z) + \langle \nabla f(z), \lambda(v - z) + (1-\lambda)(w - z) \rangle \\ &= f(z) = f(\lambda v + (1-\lambda) w) . \end{aligned}$$
   According to Theorem 3.3, f is convex.

Q.E.D.

This property is generalized by notion of subgradient.

Definition 3.27 *Let* $\mathsf{D} \subset \mathbb{R}^n$, $\mathsf{D} \neq \emptyset$ *be a set,* $f : \mathsf{D} \to \mathbb{R}$ *be a function,* $x \in \mathsf{D}$ *and* $a \in \mathbb{R}^n$*. We say,* a *is* *a subgradient of* f *at* $x \in \mathsf{D}$ *(cz. subgradient), if we have*

$$f(y) - f(x) \geq \langle a, y - x \rangle \text{ for all } y \in \mathsf{D}. \tag{3.7}$$

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

Using subdiferential we can equivalently rewrite definition of global minimum.

Theorem 3.28: *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* 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 subdiferencial are helpful tools for describing convex functions.

Theorem 3.29: *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* $x \in \operatorname{rint}(\mathsf{D})$*.*

Proof: Without any loss of generality we can assume $\operatorname{int}(\mathsf{D}) \neq \emptyset$.
Take $x \in \operatorname{int}(\mathsf{D})$.
Then, $(x, f(x)) \in \partial(\mathsf{epi}(f))$ and there is a supporting hyperplane determined by proper $\alpha \in \mathbb{R}^n$ 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 η can be arbitrary large, therefore, $\beta \geq \mathbf{0}$.

1) Assume $\beta = \mathbf{0}$.
Since $x \in \operatorname{int}(\mathsf{D})$, there is $\delta > 0$ such that $\mathcal{U}_\delta(x) \subset \mathsf{D}$.
Therefore, for all $y \in \mathcal{U}_\delta(x)$ we have $\langle\, \alpha, y \,\rangle \geq \langle\, \alpha, x \,\rangle$.
Then $\alpha = \mathbf{0}$.
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 subdiferentials is in power if function is defined on an open set.

Theorem 3.30: *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 3.29, $\partial f(x) \neq \emptyset$ 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 D is a convex set.
   Take $\alpha \in \partial f(z)$, which exists according to our assumption.
   Definition of subgradient is giving
   $$f(x) - f(z) \geq \langle \alpha, x - z \rangle,$$
   $$f(y) - f(z) \geq \langle \alpha, y - z \rangle.$$
   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 3.3.

Q.E.D.

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

Theorem 3.31: *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 $x \in \operatorname{rint}(\mathsf{D})$.*

Proof: Accordingly to Theorem 3.29, the condition is fulfilled for a convex function. We have to show the opposite implication, only.

1. Accordingly to Theorem 3.30, $f : \operatorname{rint}(\mathsf{D}) \to \mathbb{R}$ is convex.
2. Take $x, y \in \mathsf{D}$ and $0 < \lambda < 1$.
   Since D is convex, we have $\mathsf{D} \subset \operatorname{clo}(\operatorname{rint}(\mathsf{D}))$.
   Then, there are sequences $x_k, y_k \in \operatorname{rint}(\mathsf{D})$ 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 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.

Lemma 3.32 *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 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 λ 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 3.33 *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 single-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 3.34 *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 single-point set.*

3. *f possesses a gradient at y.*

3.3 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. $f : \mathbb{R}^n \to (\mathbb{R}^*)^m$. We also understand such a function as a vector of functions $f = (f_1, f_2, \dots, f_m)^\top$.

Definition 3.35 *For a function $f : \mathbb{R}^n \to (\mathbb{R}^*)^m$, we define its epigraph (cz. epigraf)*

$$\mathsf{epi}\,(f) \;=\; \left\{ \begin{pmatrix} x \\ \eta \end{pmatrix} : f\,(x) \leq \eta,\ x \in \mathbb{R}^n,\ \eta \in \mathbb{R}^m \right\}, \tag{3.8}$$

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

$$\mathsf{Dom}\,(f) \;=\; \{x : f_i\,(x) < +\infty \text{ for all } i \in \{1, 2, \dots, m\},\ x \in \mathbb{R}^n\}, \tag{3.9}$$

$$\mathsf{WDom}\,(f) \;=\; \{x : f_i\,(x) < +\infty \text{ for some } i \in \{1, 2, \dots, m\},\ x \in \mathbb{R}^n\}. \tag{3.10}$$

Definition 3.36 *A function $f : \mathbb{R}^n \to (\mathbb{R}^*)^m$ is called monotone (cz. monotónní), if $f\,(x) \leq f\,(y)$ whenever $x \leq y$.*

Definition 3.37 *A function $f : \mathbb{R}^n \to (\mathbb{R}^*)^m$ is convex (cz. konvexní), if $\mathsf{WDom}\,(f) = \mathsf{Dom}\,(f)$ and $\mathsf{epi}\,(f)$ is a convex set.*

Convexity of a function can be equivalently explained.

Lemma 3.38 *A function $f : \mathbb{R}^n \to (\mathbb{R}^*)^m$ is convex if and only if $\mathsf{Dom}\,(f) = \mathsf{Dom}\,(f_1) = \mathsf{Dom}\,(f_2) = \dots = \mathsf{Dom}\,(f_m)$, $\mathsf{Dom}\,(f)$ is a convex set and f_i is a convex function for each $i \in \{1, 2, \dots, m\}$.*

Theorem 3.39: *A function $f : \mathbb{R}^n \to (\mathbb{R}^*)^m$ is convex if and only if $\mathsf{Dom}\,(f) = \mathsf{Dom}\,(f_1) = \mathsf{Dom}\,(f_2) = \dots = \mathsf{Dom}\,(f_m)$, $\mathsf{Dom}\,(f)$ is a convex set and for all $x, y \in \mathsf{Dom}\,(f)$ and $0 < \lambda < 1$ we have*

$$f\,(\lambda x + (1 - \lambda) y) \leq \lambda f\,(x) + (1 - \lambda) f\,(y). \tag{3.11}$$

Proof: Statement is a consequence of Theorem 3.3 and Lemma 3.38.

Q.E.D.

Composition of functions preserves convexity of functions under some circumstances.

Lemma 3.40 *If $\mathcal{X} \subset \mathbb{R}^n$ is a convex set, $h : \mathbb{R}^n \to \mathbb{R}^m$ is an affine linear 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 linear function.
For $x, y \in \mathcal{X}$ and $\lambda \in (0, 1)$ we have

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

Q.E.D.

Lemma 3.41 *If $\mathcal{X} \subset \mathbb{R}^n$ is a convex set, $h : \mathcal{X} \to \mathbb{R}^m$ is a convex function and $g : \operatorname{conv}(h(\mathcal{X})) \to \mathbb{R}$ 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

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

Q.E.D.

3.4 Generalization of convex functions

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

Definition 3.42 *Let $\mathsf{S} \subset \mathsf{D} \subset \mathbb{R}^n$, S be an 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 S (cz. pseudokonvexní v bodě x vzhledem k S), if*

$$\forall y \in \mathsf{S}, \ \langle \nabla f(x), y - x \rangle \geq 0 \text{ implies } f(y) \geq f(x).$$

2. *We say, f is strictly pseudoconvex at x with respect to S (cz. striktně pseudokonvexní v bodě x vzhledem k S), if*

$$\forall y \in \mathsf{S}, \ y \neq x, \ \langle \nabla f(x), y - x \rangle \geq 0 \text{ implies } f(y) > f(x).$$

3. *We say, f is pseudoconcave at x with respect to S (cz. pseudokonkávní v bodě x vzhledem k S), if $-f$ is pseudoconvex at x with respect to S.*

4. *We say, f is strictly pseudoconcave at x with respect to S (cz. striktně pseudokonkávní v bodě x vzhledem k S), if $-f$ is strictly pseudoconvex at x with respect to S.*

Definition 3.43 *Let $\mathsf{S} \subset \mathsf{D} \subset \mathbb{R}^n$, S be a nonempty open set and $f : \mathsf{D} \to \mathbb{R}$ be differentiable at S.*

1. *We say, f is pseudoconvex on S (cz. pseudokonvexní na S), if*

$$\forall x, y \in \mathsf{S}, \ \langle \nabla f(x), y - x \rangle \geq 0 \text{ implies } f(y) \geq f(x).$$

2. *We say, f is strictly pseudoconvex on S (cz. striktně pseudokonvexní na S), if*

$$\forall x, y \in \mathsf{S}, \ x \neq y, \ \langle \nabla f(x), y - x \rangle \geq 0 \text{ implies } f(y) > f(x).$$

3. *We say, f is pseudoconcave on S (cz. pseudokonkávní na S), if $-f$ is pseudoconvex on S.*

4. *We say, f is strictly pseudoconcave on S (cz. striktně pseudokonkávní na S), if $-f$ is strictly pseudoconvex on S.*

Definition 3.44 *Let $\mathsf{S} \subset \mathsf{D} \subset \mathbb{R}^n$, S be a convex set, $x \in \mathsf{S}$ and $f : \mathsf{D} \to \mathbb{R}$.*

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

Lemma 3.45 *Let $\mathsf{S} \subset \mathsf{D} \subset \mathbb{R}^n$, S be a nonempty convex set and $f : \mathsf{D} \to \mathbb{R}$ be quasiconvex on S. Then, level sets of f fulfill $\mathsf{S} \cap \mathsf{lev}_{\le\Delta} f = \{x \in \mathsf{S} : f\,(x) \le \Delta\}$, $\mathsf{S} \cap \mathsf{lev}_{<\Delta} f = \{x \in \mathsf{S} : f\,(x) < \Delta\}$ are convex for all $\Delta \in \mathbb{R}$.*

Proof: It is sufficient to verify $\mathsf{S} \cap \mathsf{lev}_{<\Delta} f$ is convex, since $\mathsf{S} \cap \mathsf{lev}_{\le\Delta} f = \bigcap_{\beta>\Delta} \mathsf{S} \cap \mathsf{lev}_{<\beta} f$.
Take $y, z \in \mathsf{S} \cap \mathsf{lev}_{<\Delta} f$ and $0 < \lambda < 1$. Because f is quasiconvex on S, we have
$$f\,(\lambda y + (1-\lambda)z) \le \max\{f\,(y)\,, f\,(z)\} < \Delta.$$
It means $\lambda y + (1-\lambda)z \in \mathsf{S} \cap \mathsf{lev}_{<\Delta} f$. Thus, $\mathsf{S} \cap \mathsf{lev}_{<\Delta} f$ is a convex set.

Q.E.D.

Lemma 3.46 *Let $\mathsf{S} \subset \mathsf{D} \subset \mathbb{R}^n$, S be a nonempty open convex set and $h : \mathsf{D} \to \mathbb{R}$ be differentiable at S and pseudoconvex on S. Then, h is quasiconvex on S.*

Proof: Take $x, y \in \mathsf{S}$, $0 < \lambda < 1$ and denote $z = \lambda x + (1-\lambda)y$. Hence, $z \in \mathsf{S}$, since S is a convex set.
Assume $h\,(z) > h\,(x)$.
Function h is pseudoconvex on S, therefore, $\langle\, \nabla h\,(z)\,, x - z\,\rangle < 0$.
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,
$$\begin{aligned} y - z &= -\frac{\lambda}{(1-\lambda)}(x-z), \\ \langle\, \nabla h\,(z)\,, y - z\,\rangle &= -\frac{\lambda}{(1-\lambda)}\,\langle\, \nabla h\,(z)\,, x - z\,\rangle > 0\,. \end{aligned}$$
Function h is pseudoconvex on S, therefore, $h\,(z) \le h\,(y)$.
Finally, $h\,(\lambda x + (1-\lambda)y) \le \max\{h\,(x)\,, h\,(y)\}$ and h is quasiconvex on S.

Q.E.D.

Lemma 3.47 *Let $\mathsf{S} \subset \mathsf{D} \subset \mathbb{R}^n$, S be a nonempty open convex set, $f : \mathsf{D} \to \mathbb{R}$ be differentiable at S and pseudoconvex on S. Then, level sets of f fulfill $\mathsf{S} \cap \mathsf{lev}_{\le\Delta} f = \{x \in \mathsf{S} : f\,(x) \le \Delta\}$, $\mathsf{S} \cap \mathsf{lev}_{<\Delta} f = \{x \in \mathsf{S} : f\,(x) < \Delta\}$ are convex for all $\Delta \in \mathbb{R}$.*

Proof: The statement is a direct consequence of Lemma 3.46 and Lemma 3.45.

Q.E.D.

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

Example 3.48: Consider $\mathsf{D} = \mathbb{R}$, $\mathsf{S} = (-2, 1)$ and $f\,(x) = -x^2$. Function f is pseudoconvex at -1 with respect to S, but it is not quasiconvex at -1 with respect to S.

△

Bibliography

[1] Bazara, M.S.; Sherali, H.D.; Shetty, C.M.: *Nonlinear Programming. Theory and Algorithms.* 2nd edition, Wiley, New York, 1993.

[2] Lachout, P.: *Teorie pravděpodobnosti.* Skripta MFF UK, Praha, 1998.

[3] Lachout, P.: *Konvexita v konečné dimenzi.* Text na mé webové stránce, verze 16.10.2011.

[4] Rockafellar, T.: *Convex Analysis.*, Springer-Verlag, Berlin, 1975.

[5] Rockafellar, T.; Wets, R. J.-B.: *Variational Analysis.* Springer-Verlag, Berlin, 1998.

Index