

5th Semester

- [M5170 Matematické programování](#)
 - [Konvexní množiny](#)
 - [Oddělování konvexních množin](#)
 - [Konvexní funkce](#)
 - [Subgradient a subdiferenciál a Fenchelova transformace](#)
 - [Numerické metody v R](#)
 - [Numerické metody v \$\mathbb{R}^n\$](#)
 - [Nutné a postačující podmínky optimality](#)
 - [Duální úloha](#)
 - [Analýza citlivosti](#)
- [M5120 Lineární statistické modely](#)
 - [2. cvičení](#)
 - [3. cvičení](#)
 - [4. cvičení](#)
 - [5. cvičení](#)
 - [7. cvičení](#)
 - [9. cvičení](#)
 - [10. cvičení](#)

M5170 Matematické programování

Konvexní množiny

```

 $\xdef\scal#1#2{\langle #1, #2 \rangle}$ 
 $\xdef\norm#1{\left|\right| #1 \right|}$ 
 $\xdef\dist{\rho}$ 
 $\xdef\and{\&}$ 
 $\xdef\brackets#1{\left\{ #1 \right\}}$ 
 $\xdef\parc#1#2{\frac {\partial #1}{\partial #2}}$ 
 $\xdef\mtr#1{\begin{pmatrix} #1 \end{pmatrix}}$ 
 $\xdef\bm#1{\boldsymbol{#1}}$ 
 $\xdef\mc#1{\mathcal{#1}}$ 
 $\xdef\vv#1{\mathbf{#1}}$ 
 $\xdef\vvp#1{\pmb{#1}}$ 
 $\xdef\ve{\varepsilon}$ 
 $\xdef\l{\lambda}$ 
 $\xdef\th{\vartheta}$ 
 $\xdef\alpha{\alpha}$ 
 $\xdef\tagged#1{(\text{#1})}$ 
 $\xdef\tagged*#1{\text{#1}}$ 
 $\xdef\tagEqHere#1#2{\href{#2\#eq-#1}{(\text{#1})}}$ 
 $\xdef\tagDeHere#1#2{\href{#2\#de-#1}{\text{#1}}}$ 
 $\xdef\tagEq#1{\href{\#eq-#1}{(\text{#1})}}$ 
 $\xdef\tagDe#1{\href{\#de-#1}{\text{#1}}}$ 
 $\xdef\T#1{\htmlId{eq-#1}{#1}}$ 
 $\xdef\D#1{\htmlId{de-#1}{\vv{#1}}}$ 
 $\xdef\conv#1{\mathrm{conv}\, , #1}$ 
 $\xdef\cone#1{\mathrm{cone}\, , #1}$ 
 $\xdef\aff#1{\mathrm{aff}\, , #1}$ 
 $\xdef\lin#1{\mathrm{Lin}\, , #1}$ 
 $\xdef\span#1{\mathrm{span}\, , #1}$ 
 $\xdef\O{\mathcal O}$ 
 $\xdef\ri#1{\mathrm{ri}\, , #1}$ 
 $\xdef\rd#1{\mathrm{r}\partial\, , #1}$ 
 $\xdef\interior#1{\mathrm{int}\, , #1}$ 
 $\xdef\proj{\Pi}$ 
 $\xdef\epi#1{\mathrm{epi}\, , #1}$ 
 $\xdef\grad#1{\mathrm{grad}\, , #1}$ 
 $\xdef\hess#1{\nabla^2\, , #1}$ 
 $\xdef\subdif#1{\partial #1}$ 
 $\xdef\co#1{\mathrm{co}\, , #1}$ 
 $\xdef\iter#1{\wedge^{[#1]}}$ 

```

Definice $\mathcal{D}\{1.1\}$ (Konvexní množina)

Nechť $X \subseteq \mathbb{R}^n$. Množina X se nazývá **konvexní**, jestliže pro všechna $x_1, x_2 \in X$ a pro každé $\lambda \in [0,1]$ platí $\lambda x_1 + (1 - \lambda) x_2 \in X$ \tag{KM}

“ Speciálně prázdnou množinu \emptyset považujeme za **konvexní**

Operace nad konvexními množinami

Mějme $X_i, i \in I$ konvexní množiny. Potom

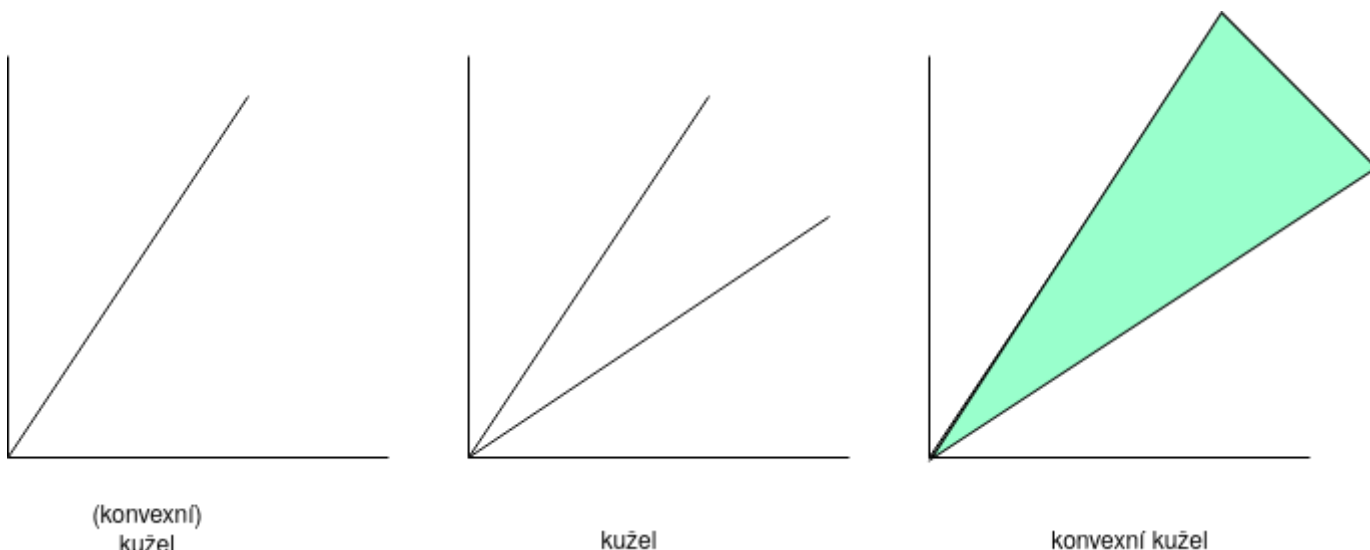
- jejich sjednocení $\bigcup_{i \in I} X_i$ je konvexní množina
- jejich **součet** $\lambda_1 X_1 + \dots + \lambda_m X_m = \{x \in \mathbb{R}^n \mid x = \sum_{i=1}^m \lambda_i x_i \text{ pro nějaká } x_i \in X_i\}$ je opět **konvexní**

Vlastnosti konvexních množin

Definice $\mathcal{D}\{2.1.3\}$ (Speciální množiny)

Množina $X \subseteq \mathbb{R}^n$ se nazývá

- **kužel**, jestliže pro každé $x \in X$ a pro každé $\lambda \in [0, \infty)$ je také $\lambda x \in X$
- **konvexní kužel**, jestliže je množina X konvexní a současně **kuzelem**
- **afinní**, jestliže pro každé $x_1, x_2 \in X$ a pro každé $\lambda \in \mathbb{R}$ platí $\lambda x_1 + (1 - \lambda)x_2 \in X$



“ **Polyedr** je **mnohostěn** v \mathbb{R}^n . Dále **ohraňčený polyedr** nazveme **polytop**

Dále si rozeberme různé kombinace bodů

Definice $\{2.1.4\}$ (Lineární kombinace)

Nechť $x_1, \dots, x_m \in \mathbb{R}^n$. Lineární kombinace $\lambda_1 x_1 + \dots + \lambda_m x_m$ se nazývá

- **konvexní**, jestliže $\lambda_1, \dots, \lambda_m \geq 0$ a $\sum_{i=1}^m \lambda_i = 1$
- **nezáporná**, jestliže $\lambda_1, \dots, \lambda_m \geq 0$
- **afinní**, jestliže $\sum_{i=1}^m \lambda_i = 1$.

Tedy jistě platí

- Množina obsahující všechny lineární kombinace libovolných dvou svých bodů (tj. s libovolnými dvěma body obsahuje i přímku procházející těmito body a počátek) je **vektorový (lineární) prostor**
- Množina obsahující všechny afinní kombinace libovolných dvou svých bodů (tj. s libovolnými dvěma body obsahuje i přímku procházející těmito body) je **afinní**
- Množina obsahující všechny nezáporné kombinace libovolných dvou svých bodů (tj. s libovolnými dvěma body obsahuje i celou výšeč určenou polopřímkami vycházejícími z počátku a procházejícími těmito body) je **konvexní kužel**

- Množina obsahující všechny konvexní kombinace dvou libovolných svých bodů (tj. s libovolnými dvěma body obsahuje i celou úsečku je spojující) je **konvexní**

Definice $\S{2.1.6}$ (Obaly)

Nechť $X \subseteq \mathbb{R}^n$

- průnik všech konvexních množin obsahujících množinu X se nazývá **konvexní obal** množiny X a značí se $\text{conv } X$.
- průnik všech konvexních *kuželů* obsahujících množinu X se nazývá **kónický obal** množiny X a značí se $\text{cone } X$.
- průnik všech *afinních* množin obsahujících množinu X se nazývá **afinní obal** množiny X a značí se $\text{aff } X$. Jeho *zaměření* se nazývá **lineární obal** množiny X a značí se $\text{lin } X$. **Dimenze** afinního obalu množiny X se značí $\dim X$ a klademe $\dim X := \dim \{\text{lin } X\}$.

“ Všimněme si, že $\text{span } X = \{ \sum_{i=1}^m \lambda_i x_i \mid x_i \in X, \lambda_i \in \mathbb{R} \}$, ale $\text{lin } X = \text{span} \{x_2 - x_1, x_3 - x_1, \dots, x_m - x_1\}$. (pro $X = \{x_1, \dots, x_m\}$) Viz obrázek z přednášky

“ Jinak řečeno, **konvexní obal** je nejmenší konvexní množina obsahující X ve smyslu množinové inkluze.
Kónický obal je nejmenší *konvexní kužel* obsahující X atd..

Jako **simplex** definujeme **konvexní obal** $n+1$ **afinně nezávislých** bodů $v_1, \dots, v_{n+1} \in \mathbb{R}^m$, kde $m \geq n$. Pod pojmem **afinně nezávislé** body rozumíme, že vektory $v_2 - v_1, v_3 - v_1, \dots, v_{n+1} - v_1$ jsou **lineárně nezávislé**.

Věta $\S{2.1.7}$

Nechť $X \subseteq \mathbb{R}^n$. Pak platí

- $\text{conv } X = \{x \mid x = \sum_{i=1}^m \lambda_i x_i, \text{ kde } m \in \mathbb{N} \text{ je libovolné}, x_1, \dots, x_m \in X, \lambda_1, \dots, \lambda_m \geq 0, \sum_{i=1}^m \lambda_i = 1\}$
- $\text{cone } X = \{x \mid x = \sum_{i=1}^m \lambda_i x_i, \text{ kde } m \in \mathbb{N} \text{ je libovolné}, x_1, \dots, x_m \in X, \lambda_1, \dots, \lambda_m \geq 0\}$
- $\text{aff } X = \{x \mid x = \sum_{i=1}^m \lambda_i x_i, \text{ kde } m \in \mathbb{N} \text{ je libovolné}, x_1, \dots, x_m \in X, \lambda_1, \dots, \lambda_m \in \mathbb{R}, \sum_{i=1}^m \lambda_i = 1\}$

Libovolný bod x v **konvexní kuželu** v \mathbb{R}^n lze vyjádřit pomocí **nezáporné kombinace** n bodů

Věta 2.1.9 (Caratheódoryho)

Nechť $X \subseteq \mathbb{R}^n$. Každý bod konvexního obalu $\text{conv } X$ může být vyjádřen jako konvexní kombinace nejvýše $n+1$ prvků množiny X , tj. pro $x \in \text{conv } X$ existují $x_1, \dots, x_{n+1} \in X$ a $\lambda_1, \dots, \lambda_{n+1} \geq 0$ splňující $\sum_{i=1}^{n+1} \lambda_i = 1$ taková, že $x = \lambda_1 x_1 + \dots + \lambda_{n+1} x_{n+1}$

“ **POZOR: Univerzální konvexní báze** (stejná pro všechny $x \in \text{conv } X$) konvexního obalu $\text{conv } X$ **nemusí existovat!**

Lze ukázat, že pokud $X \subseteq \mathbb{R}^n$ je **kompaktní** množina, pak $\text{conv } X$ je **také kompaktní**.

“ To stejné **neplatí** o uzavřenosti.

Zobecnění vnitřku množiny

Definice 2.1.11 (Relativně vnitřní bod)

Nechť $X \subseteq \mathbb{R}^n$. Bod $x^* \in X$ se nazývá **relativně vnitřním** bodem množiny X , jestliže existuje okolí $O(x^*)$ bodu x^* takové, že $O(x^*) \cap \text{aff } X \subseteq X$. Množinu všech *relativně vnitřních* bodů nazýváme **relativním vnitřkem** množiny X a značíme $\text{ri } X$.

Množina $\text{rd } X := \overline{X} \setminus \text{ri } X$ se nazývá **relativní hranice** množiny X .

“ Jistě platí $\text{interior } X \subseteq \text{ri } X$
a také $\text{ri } X \subseteq X \subseteq \overline{X} \subseteq \text{aff } X$

Platí $\overline{\text{ri } X} = \overline{X}$, tj. **relativní vnitřek** je dost velký na vygenerování **uzávěru**.

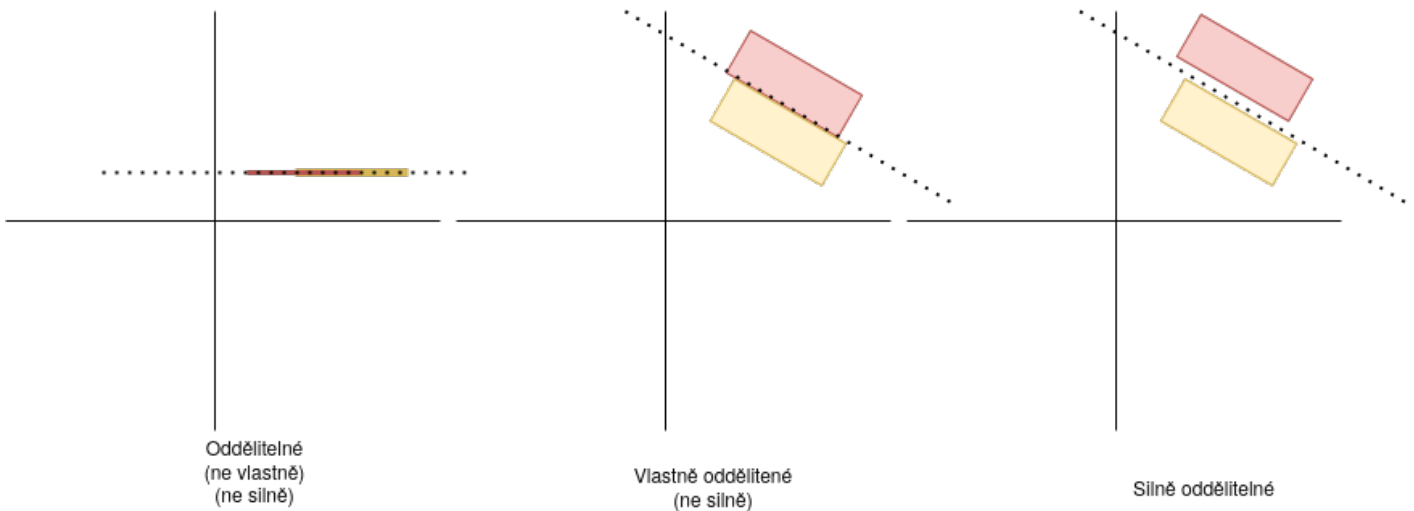
Oddělitelnost množin

Definice $\S 2.3.1$ (Oddělitelnost množin)

- **oddělitelné**, jestliže existuje $p \in \mathbb{R}^n \setminus \text{brackets}\{\mathbb{V} 0\}$ takové, že $\|\mathbf{x}_1\| \leq \|\mathbf{x}_2\|$ pro každé $\mathbf{x}_1 \in X_1, \mathbf{x}_2 \in X_2$.
- **vlastně oddělitelné**, jestliže jsou *oddělitelné* a zároveň existují body $\mathbf{x}_1^* \in X_1, \mathbf{x}_2^* \in X_2$ takové, že $\|\mathbf{x}_1^*\| > \|\mathbf{x}_2^*\|$
- **silně oddělitelné**, jestliže existuje $p \in \mathbb{R}^n \setminus \text{brackets}\{\mathbb{V} 0\}$ takové, že $\inf_{\mathbf{x}_1 \in X_1} \|\mathbf{x}_1\| > \sup_{\mathbf{x}_2 \in X_2} \|\mathbf{x}_2\|$, je-li navíc β

$\bigcup_{x_2 \in X_2} \text{scal } p \{x_2\}, \bigcap_{x_1 \in X_1} \text{scal } p \{x_1\}$, nadrovina $H_{p,\beta} := \{x \in \mathbb{R}^n \mid \text{scal } p x = \beta\}$ se nazývá **oddělující nadrovinou** množin X_1 a X_2 .

“ Ve vyjádření $\{x \in \mathbb{R}^n \mid \text{scal } p x = \beta\}$ značí p normálový vektor nadroviny a β její posunutí



Definice $D\{2.3.2\}$ (Projekce bodu)

Nechť $X \subseteq \mathbb{R}^n$ je **neprázdná** množina a $x \in \mathbb{R}^n$. Bod $x^* \in X$ nazveme **projekcí bodu x na množinu X** a označíme $\text{proj}_X(x)$, jestliže $\| \text{proj}_X(x) - x \| \leq \| y - x \|$ pro každé $y \in X$.

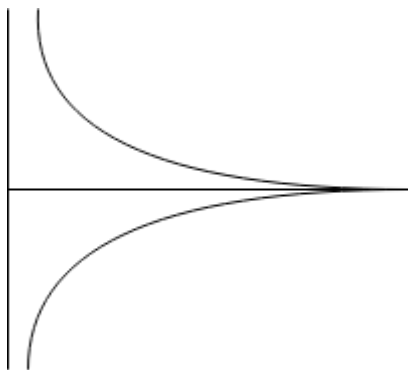
Věta $D\{2.3.4\}$

Neprázdné konvexní množiny $X_1, X_2 \subseteq \mathbb{R}^n$ jsou **silně oddělitelné** právě tehdy, když mají nenulovou vzdálenost, tj. $\text{dist}(X_1, X_2) := \inf_{x_1 \in X_1, x_2 \in X_2} \|x_1 - x_2\| > 0$, což je *ekvivalentní* s podmínkou $0 \notin \overline{X_1 - X_2}$.

Pod *kompaktní* množinou myslíme množinu, která je *ohraničená* (má konečný průměr) a *uzavřená* (obsahuje svou hranici).

Pokud jsou množiny $X_1, X_2 \subseteq \mathbb{R}^n$ *neprázdné*, *konvexní* a *disjunktní* a navíc BÚNO je X_1 **uzavřená** a X_2 **kompaktní**, tak jsou množiny **silně oddělitelné**.

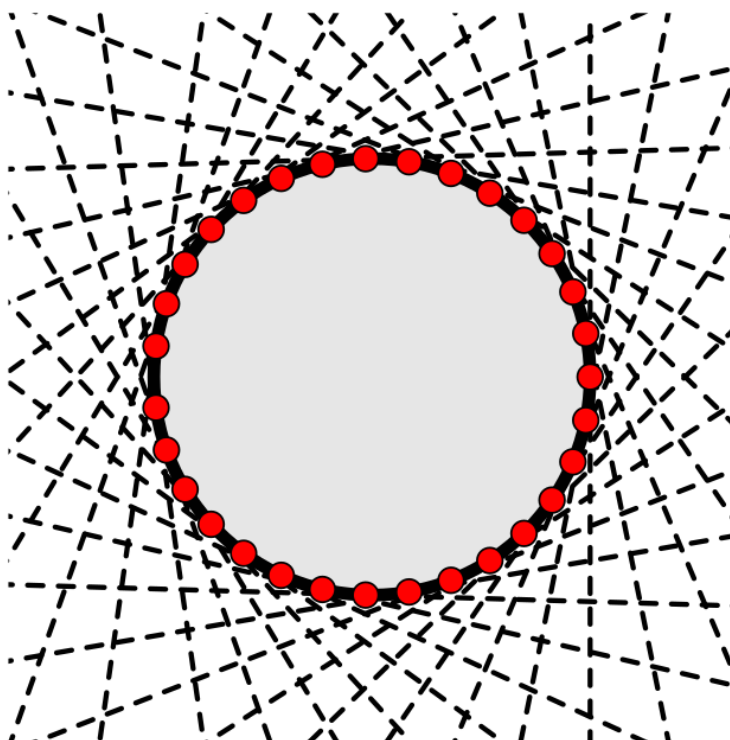
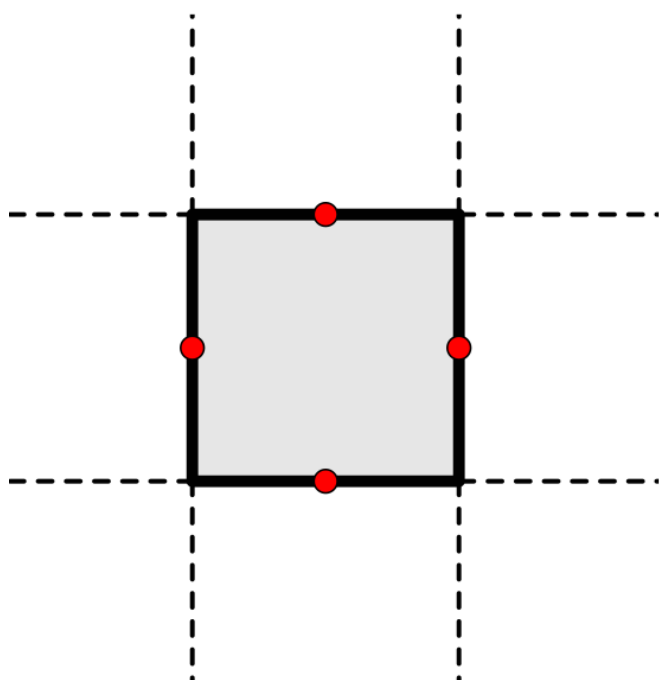
Požadavek **kompaktnosti** množiny X_2 vynechat nelze, viz protipříklad dvou hyperbol (obrázek je pouze ilustrativní)



Věta $\{2.3.5a\}$ (Geometrický popis konvexních množin)

Libovolná **uzavřená** konvexní množina $X \subseteq \mathbb{R}^n$ je řešením (*nekonečné*) soustavy **neostrých** lineárních rovnic.

“ **Geometricky**: každá uzavřená konvexní množina $X \subseteq \mathbb{R}^n$ je **průnikem uzavřených poloprostorů**, konkrétně **všech** uzavřených poloprostorů obsahujících X ”



Opěrné nadroviny

Definice $\{2.3.6\}$ (Opěrná nadrovina)

Nechť $X \subseteq \mathbb{R}^n$ je *neprázdná* množina a nechť $a \in \partial X := \overline{X} \setminus \text{interior } X$. Nadrovina $H_{\{p, \beta\}}$ se nazývá

- **opěrnou nadrovinou** množiny X v bodě a , jestliže $\forall x \in X, \exists \alpha \geq 0$ pro každé $x \in X$
- **vlastní opěrnou nadrovinou** množiny X , jestliže je **opěrnou nadrovinou** množiny X a zároveň existuje $x^* \in X$ takové, že $\exists \alpha \in \mathbb{R} \text{ s } \alpha > 0$

“ Jinak řečeno množina musí ležet pouze v **jednom** z poloprostorů určených opěrnou nadrovinou.

Věta 2.3.7 (Existenci opěrné nadroviny)

Nechť $X \subseteq \mathbb{R}^n$ je neprázdná konvexní množina a nechť $a \in \mathbb{R}^n$ $a \notin X$. Pak v bodě a existuje **vlastní** opěrná nadrovina množiny X .

“ Pro relativní vnitřek $\text{ri } X$ množiny X a její vnitřek platí $\text{interior } X \subseteq \text{ri } X$ a tedy jistě $\overline{X} \setminus \text{ri } X = \text{rd } X \subseteq \text{partial } X = \overline{X} \setminus \text{interior } X$

Podmínky oddělitelnosti

Věta 2.3.7a (Oddělitelnost množin)

Nechť $X_1, X_2 \subseteq \mathbb{R}^n$ jsou neprázdné, konvexní a disjunktní množiny. Pak pro tyto množiny existuje oddělující nadrovina.

Věta 2.3.8

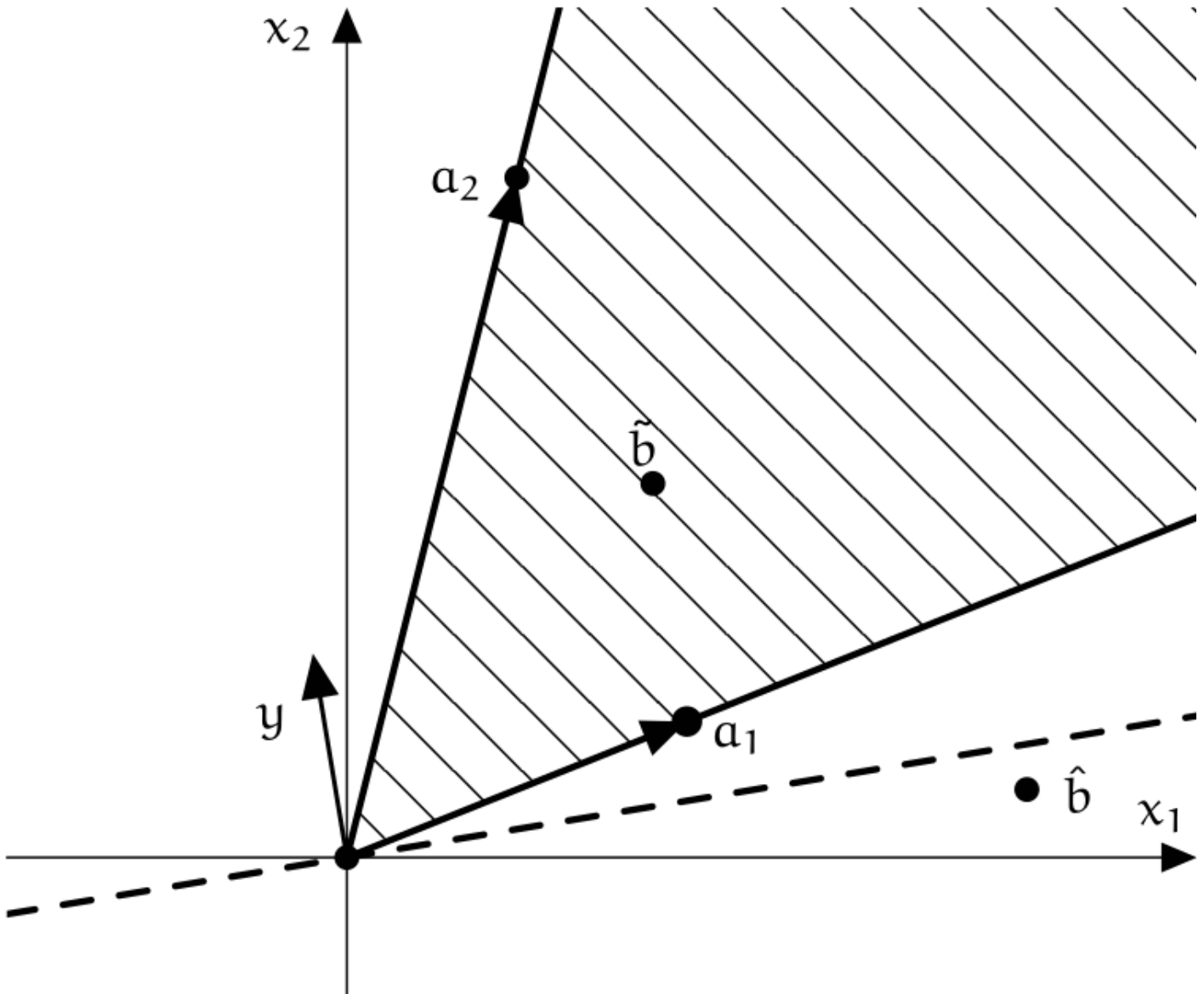
Neprázdné konvexní množiny $X_1, X_2 \subseteq \mathbb{R}^n$ jsou **vlastně** oddělitelné právě tehdy, když $X_1 \cap X_2 = \emptyset$.

Věta 2.3.9 (Farkas & Minkowski)

Nechť $A \in \mathbb{R}^{m \times n}$ a $b \in \mathbb{R}^m$. Potom je **právě jeden** z následujících systémů rovnic a nerovnic řešitelný: $Ax = b, x \geq 0$ (2.3.5) $A^T y \geq 0, y \leq b$ (2.3.6)

Jinak řečeno soustava $\text{tagEq}\{2.3.5\}$ má řešení právě tehdy, když pro všechna $y \in \mathbb{R}^m$ platí $A^T y \geq 0$ a zároveň $\text{scal } y b \geq 0$

Ještě jinak můžeme větu formulovat tak, že buď b leží v *konvexním kuželu* $\text{cone}\{a_i\}_{i=1}^n$ nebo jsou b a *konvexní kužel* **silně oddělitelné**.



Z této věty pak plynou tzv. **věty o alternativě**, které můžeme najít například v lineárním programování.

Tvrzení **Věty** $\text{tagDe}\{2.3.9\}$ můžeme také napsat jako:

Jestliže systém $f_0(x) := \text{scal } \{a_0\} x < 0$, $f_i(x) := \text{scal } \{a_i\} x \leq 0$, $i \in \{1, \dots, m\}$ nemá pro daná $a_0, \dots, a_m \in \mathbb{R}^n$ řešení na \mathbb{R}^n , pak existují čísla $y_1, \dots, y_m \geq 0$ taková, že $a_0 + \sum_{i=1}^m y_i a_i = 0$, tj. $f_0(x) + \sum_{i=1}^m y_i f_i(x) = 0$ pro každé $x \in \mathbb{R}^n$.

Toto plyne z toho, že pokud v $\text{tagDe}\{2.3.9\}$ položíme $b = a_0$ a $A = -(a_1, \dots, a_m)$ (a zaměníme-li x a y), pak podle předpokladu **nemá** systém $A^T x \geq 0$, $\text{scal } x \{a_0\}$

< 0 řešení. A tedy dostáváme, že systém $Ay = a_0, y \geq 0$ řešení **mít musí**.

Věta 2.3.12 (Fan & Glicksburg & Hoffman)

Nechť množina $X \subseteq \mathbb{R}^n$ je konvexní, funkce $f_1, \dots, f_k: X \rightarrow \mathbb{R}$ jsou **konvexní** a funkce f_{k+1}, \dots, f_m **afinní**, tj. pro $j \in \{k+1, \dots, m\}$ máme $f_j = \sum_{i=1}^n a_{ji} x_i + \beta_j$ pro vhodná $a_j \in \mathbb{R}^n$ a $\beta_j \in \mathbb{R}$. Jestliže systém nerovností a rovností
$$\begin{cases} f_i(x) < 0, & i \in \{1, \dots, k\} \\ f_j(x) = 0, & j \in \{k+1, \dots, m\} \end{cases} \tag{T{2.3.8}}$$
 nemá řešení na X , pak existují takové konstanty $y_1, \dots, y_k \geq 0$ a $y_{k+1}, \dots, y_m \in \mathbb{R}$ že pro alespoň jedno $l \in \{1, \dots, m\}$ je $y_l \neq 0$ a pro všechna $x \in X$ platí
$$\sum_{i=1}^m y_i f_i(x) \geq 0 \tag{T{2.3.9}}$$

Následující věta udává podmínky, které zajišťují kladnost jistého význačného y_i (BÚNO $y_i = 0$) ve vztahu 2.3.9. Po vydělení vztahu 2.3.9 tímto y_i dostaneme BÚNO $y_i = 1$.

Věta 2.3.13 (Podmínky regularity)

Nechť množina $X \subseteq \mathbb{R}^n$ je konvexní a funkce $f_0, \dots, f_m: X \rightarrow \mathbb{R}$ jsou konvexní. Jestliže systém nerovností
$$f_0(x) < 0, \tag{T{2.3.10}}$$

$$f_i(x) < 0, \quad i \in \{1, \dots, m\} \tag{T{2.3.11}}$$
 nemá řešení na X a podsystém 2.3.11 má řešení na X , pak existují $y_1, \dots, y_m \geq 0$ taková, že pro všechna $x \in X$ platí
$$f_0(x) + \sum_{i=1}^m y_i f_i(x) \geq 0 \tag{T{2.3.12}}$$

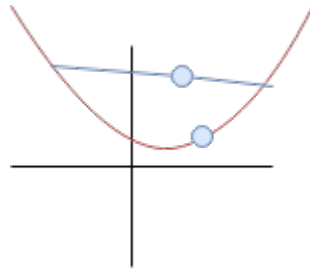
Definice $\S 2.2.1$ (Konvexní funkce)

- **konvexní** na X , jestliže pro všechna $x_1, x_2 \in X$ a každé $\lambda \in [0,1]$ platí
$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda) f(x_2) \tag{2.2.1}$$
- **ostře konvexní** na X , jestliže nerovnost $\tag{2.2.1}$ je **ostrá** pro všechna $x_1, x_2 \in X, x_1 \neq x_2$ a každé $\lambda \in (0,1)$.
- **silně konvexní** na X s **konstantou silné konvexnosti** $\theta > 0$, jestliže pro všechna $x_1, x_2 \in X$ a každé $\lambda \in [0,1]$ platí
$$\underbrace{f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda) f(x_2)}_{\tag{2.2.1}} - \theta \lambda(1 - \lambda) \|x_1 - x_2\|^2 \tag{2.2.2}$$

Věta 2.2.2 (Konvexnost nadgrafu)

Nechť $X \subseteq \mathbb{R}^n$ je *konvexní* množina a nechť $f : X \rightarrow \mathbb{R}$. Funkce f je **konvexní** na X právě tehdy, když její **nadgraf (epigraf)** $\text{epi } f := \{ (x, \beta) \in \mathbb{R}^{n+1} \mid x \in X, \beta \geq f(x) \}$ je **konvexní** množina.

Pro **ostře** konvexní funkci musí "tyto dva body" vždy ležet nad sebou (myšleny jejich souřadnice na ose y). Navíc pro **silnou** konvexnost mezi nimi musí vždy být alespoň daná mezera.



“ Tyto body **nemusí** ležet nad sebou (na svislé přímce). Navíc ještě

f **konvexní** $\iff -f$ **konkávní**

Kombinace konvexních funkcí

Věta $\{2.2.3\}$ (Nezáporná lineární kombinace konvexních funkcí)

Nechť $X \subseteq \mathbb{R}^n$ je konvexní množina, funkce $f_1, \dots, f_m: X \rightarrow \mathbb{R}$ jsou konvexní na X a $\alpha_1, \dots, \alpha_m \geq 0$ jsou daná čísla. Potom $F(x) = \alpha_1 f_1(x) + \dots + \alpha_m f_m(x)$ je **konvexní**.

Věta $\{2.2.4\}$ (*Sublevel set*)

Nechť $X \subseteq \mathbb{R}^n$ je konvexní množina a $f: X \rightarrow \mathbb{R}$ je *konvexní* funkce na X . Pak pro libovolné $K \in \mathbb{R}$ je odpovídající **dolní vrstevnicová množina** (**sublevel set**) $V_K := \{x \in X \mid f(x) \leq K\}$ také **konvexní**.

“ Platí pouze tato implikace: f **konvexní** \implies *sublevel set* **konvexní**
Například x^3 má konvexní *sublevel set*, ale sama konvexní **není**.

Přesněji říkáme, že pokud má funkce f **konvexní sublevel set**, pak f je **kvazikonvexní**.

Věta $\{2.2.5\}$ (Jensen)

Nechť $X \subseteq \mathbb{R}^n$ je konvexní množina a funkce $f: X \rightarrow \mathbb{R}$ je konvexní na X . Pak pro libovolné $m \in \mathbb{N}$, $x_1, \dots, x_m \in X$ a čísla $\lambda_1, \dots, \lambda_m \geq 0$ splňující $\sum_{i=1}^m \lambda_i = 1$ platí $f\left(\sum_{i=1}^m \lambda_i x_i\right) \leq \sum_{i=1}^m \lambda_i f(x_i)$.
 $\{2.2.3\}$ Je-li navíc funkce f **ostře konvexní** a $\lambda_1, \dots, \lambda_m \in (0,1)$, pak rovnost v $\{2.2.3\}$ nastane *právě tehdy*, když $x_1 = \dots = x_m$.

První část **Věty** $\tagDe{2.2.5}$ lze jistě podle **Definice** $\tagDe{2.2.1}$ nahradit ekvivalencí

“ Z Jensenovy nerovnosti $\tagEq{2.2.3}$ lze odvodit například **AG nerovnost**
$$\frac{x_1 + \dots + x_m}{m} \leq \sqrt[m]{x_1 \cdot \dots \cdot x_m}$$

Lokalizace minima konvexní funkce

Věta $\tagD{2.2.6}$

Nechť $X \subseteq \mathbb{R}^n$ je konvexní množina a funkce $f: X \rightarrow \mathbb{R}$ konvexní. Potom následující tvrzení jsou pravdivá

- Libovolné **lokální minimum** funkce f na X je současně **globálním minimem**.
- Množina bodů množiny X , v nichž funkce f nabývá svého **minima** na X , je **konvexní**. Je-li funkce dokonce **ostře konvexní**, pak je tato množina **nejvýše jednovýbová**.
- Je-li funkce f **diferencovatelná** na **otevřené** množině $U \subseteq X$ a $x^* \in U$ je jejím **stacionárním bodem**, tj. $\nabla f(x^*) = 0$, pak x^* je bodem **globálního minima** funkce f na množině X .

Z Věty $\tagD{2.2.6}$ mimo jiné plyne, že je-li $f: X \rightarrow \mathbb{R}$ (*ostře*) konvexní a **spojitá** funkce na konvexní a **kompaktní** množině $X \subseteq \mathbb{R}^n$, pak f má na X (*právě jedno*) **globální minimum**.

Věta $\tagD{2.2.7}$ (Základní věta konvexního programování)

Máme-li konvexní funkci $f: X \rightarrow \mathbb{R}$ na polytopu $X := \text{conv}\{x_1, \dots, x_m\} \subseteq \mathbb{R}^n$, pak je maximum funkce f na X dosaženo v některém z bodů x_1, \dots, x_m .

Obecněji: je-li X konvexní a **kompaktní** množina, pak maximum nastává v **extrémním bodě** (tj. v takovém bodě, který **není netriviální** konvexní kombinací dvou bodů z X)

“ Z Věty $\tagD{2.2.7}$ plyne **základní věta lineárního programování**:
Je-li funkce f **afinní** (taková funkce je konvexní i konkávní zároveň), pak **globální maximum** nastává v některém z bodů x_1, \dots, x_m , tj. v některém z "**vrcholů**" polytopu.

Vlastnosti konvexních funkcí

Věta 2.4.1 (Spojitost konvexní funkce)

Nechť $X \subseteq \mathbb{R}^n$ je konvexní množina a funkce $f: X \rightarrow \mathbb{R}$ je konvexní na X . Pak f je **spojitá** pro každé $x \in \text{int } X$.

Dále ještě známe několik podmínek zaručujících konvexnost funkce $f: \mathbb{R} \rightarrow \mathbb{R}$:

- má-li f vlastní derivaci v **otevřeném** intervalu I , pak f je (ostře) konvexní na I právě tehdy, když f' je **neklesající (rostoucí)** na I .
- má-li f vlastní derivaci v **otevřeném** intervalu I , pak f je (ostře) konvexní na I právě tehdy, když pro každé $x, x^* \in I$ platí $f(x) \geq f(x^*) + f'(x^*)(x - x^*)$, tj. graf funkce f leží **nad tečnou** sestrojenou v **libovolném** bodě.
- má-li f vlastní **druhou** derivaci v **otevřeném** intervalu I , pak f je konvexní na I právě tehdy, když funkce $f''(x) \geq 0$ (je-li $f''(x) > 0$ na I , pak **ostře konvexní**).

Tyto tvrzení si nyní rozšíříme pro $f: \mathbb{R}^n \rightarrow \mathbb{R}$ pro **silnou** konvexnost s konstantnou θ (volbou $\theta = 0$ dostáváme **ostrou** konvexnost)

Věta 2.4.2

Nechť $X \subseteq \mathbb{R}^n$ je konvexní množina a funkce f diferencovatelná na otevřené množině $\text{int } X$. Pak f je **silně** konvexní na X s konstantou **silné** konvexnosti $\theta \geq 0$ právě tehdy, když pro každé $x, x^* \in X$ platí $f(x) \geq f(x^*) + \langle \text{grad } f(x^*), x - x^* \rangle + \frac{\theta}{2} \|x - x^*\|^2$ (tag 2.4.1)

Ve výraz $\langle \text{grad } f(x^*), x - x^* \rangle$ hraje úlohu **tečné nadroviny** v bodě x^* s **normálovým vektorem** $\text{grad } f(x^*)$ (tečným jak na nadrovinu, tak na funkci f v bodě x^*)

Ještě jinými slovy je z Věty 2.4.2 plyne, že nadrovina daná $\langle \text{grad } f(x) - \text{grad } f(x^*), x - x^* \rangle \geq \theta \|x - x^*\|^2$ **opěrnou nadrovinou** pro f

Důsledek 2.4.5

Nechť $X \subseteq \mathbb{R}^n$ je konvexní množina splňující $\text{int } X \neq \emptyset$. Nechť funkce $f: X \rightarrow \mathbb{R}$ je **dvakrát** spojitě diferencovatelná na **otevřené** množině $\text{int } X$ s maticí druhých derivací $\text{Hess } f(x)$ (*Hessova matice*). Pak f je **silně** konvexní na X s konstantou silné konvexnosti $\theta \geq 0$ právě tehdy, když pro každé $x \in X$ a $h \in \mathbb{R}^n$

platí $\|\text{Hess } f(x)\|_h \geq 2 \|\text{Hess } f(x)\|_h^2$ $\tag{T{2.4.3}}$, jinými slovy $\text{Hess } f(x) \geq 2I$ pro všechna $x \in X$.

Z Důsledku $\tag{2.4.5}$ plynou následující tři implikace

- $\text{Hess } f(x) \geq 0$ pro všechna $x \in X$ $\implies f$ je konvexní na X
- $\text{Hess } f(x) > 0$ pro všechna $x \in X$ $\implies f$ je **ostře** konvexní na X
- $\text{int } X \neq \emptyset$ a f je konvexní na X $\implies \text{Hess } f(x) \geq 0$ pro všechna $x \in X$

Subgradient a subdiferenciál a Fenchelova transformace

```

 $\xdef\scal#1#2{\langle #1, #2 \rangle}$ 
 $\xdef\norm#1{\left\| \right. #1 \left. \right\|}$ 
 $\xdef\dist{\rho}$ 
 $\xdef\and{\&}$ 
 $\xdef\brackets#1{\left\{ #1 \right\}}$ 
 $\xdef\parc#1#2{\frac {\partial #1}{\partial #2}}$ 
 $\xdef\mtr#1{\begin{pmatrix} #1 \end{pmatrix}}$ 
 $\xdef\bm#1{\boldsymbol{#1}}$ 
 $\xdef\mc#1{\mathcal{#1}}$ 
 $\xdef\vv#1{\mathbf{#1}}$ 
 $\xdef\vvp#1{\pmb{#1}}$ 
 $\xdef\ve{\varepsilon}$ 
 $\xdef\l{\lambda}$ 
 $\xdef\th{\vartheta}$ 
 $\xdef\alpha{\alpha}$ 
 $\xdef\tagged#1{(\text{#1})}$ 
 $\xdef\tagged*#1{\text{#1}}$ 
 $\xdef\tagEqHere#1#2{\href{#2\#eq-#1}{(\text{#1})}}$ 
 $\xdef\tagDeHere#1#2{\href{#2\#de-#1}{\text{#1}}}$ 
 $\xdef\tagEq#1{\href{\#eq-#1}{(\text{#1})}}$ 
 $\xdef\tagDe#1{\href{\#de-#1}{\text{#1}}}$ 
 $\xdef\T#1{\htmlld{eq-#1}{#1}}$ 
 $\xdef\D#1{\htmlld{de-#1}{\vv{#1}}}$ 
 $\xdef\conv#1{\mathrm{conv}\,, #1}$ 
 $\xdef\cone#1{\mathrm{cone}\,, #1}$ 
 $\xdef\aff#1{\mathrm{aff}\,, #1}$ 
 $\xdef\lin#1{\mathrm{Lin}\,, #1}$ 
 $\xdef\span#1{\mathrm{span}\,, #1}$ 
 $\xdef\O{\mathcal O}$ 
 $\xdef\ri#1{\mathrm{ri}\,, #1}$ 
 $\xdef\rd#1{\mathrm{r}\partial\,, #1}$ 
 $\xdef\interior#1{\mathrm{int}\,, #1}$ 
 $\xdef\proj{\Pi}$ 
 $\xdef\epi#1{\mathrm{epi}\,, #1}$ 
 $\xdef\grad#1{\mathrm{grad}\,, #1}$ 
 $\xdef\hess#1{\nabla^2\,, #1}$ 
 $\xdef\subdif#1{\partial #1}$ 
 $\xdef\co#1{\mathrm{co}\,, #1}$ 

```

Subgradient a subdiferenciál

Definice $\mathcal{D}\{2.5.1\}$ (Subgradient a subdiferenciál)

Nechť $X \subseteq \mathbb{R}^n$ je konvexní množina. Vektor $a \in \mathbb{R}^n$ se nazývá **subgradient** funkce $f: X \rightarrow \mathbb{R}$ v bodě $x^* \in X$, jestliže $f(x) - f(x^*) \geq \langle a, x - x^* \rangle$ $\forall x \in X$. Množina všech **subgradientů** funkce f v bodě $x^* \in X$ se nazývá **subdiferenciál** funkce f v bodě x^* a značí se $\partial f(x^*)$. Funkce f se nazývá **subdiferencovatelná** v bodě $x^* \in X$, jestliže $\partial f(x^*) \neq \emptyset$.

„Jistě platí podle Věty $\mathcal{D}\{2.4.2\}$ $\forall f: X \rightarrow \mathbb{R}$ konvexní a $x^* \in \text{ri } X$, $\nabla f(x^*) \in \partial f(x^*)$ “

Speciálně, je-li $f: X \subseteq \mathbb{R} \rightarrow \mathbb{R}$ konvexní a $x^* \in \text{ri } X$, pak podle Věty $\mathcal{D}\{2.4.7\}$ existují **jednostranné** derivace $f'_-(x^*)$, $f'_+(x^*)$, přičemž platí $f'_-(x^*) \leq f'_+(x^*)$. V tomto případě pak máme $\partial f(x^*) = [f'_-(x^*), f'_+(x^*)]$.

$\rangle, f'_+(x^*)]$

Věta 2.5.4

Nechť $X \subseteq \mathbb{R}^n$ je konvexní množina a $f: X \rightarrow \mathbb{R}$

- Je-li funkce f **konvexní** a $x^* \in \text{ri } X$, pak $\text{subd} f(x^*)$ je **neprázdná, uzavřená a konvexní** množina
- Je-li $\text{subd} f(x)$ **neprázdná** pro každé $x \in X$, pak f je **konvexní** na X

Fenchelova transformace

Fenchelova transformace je transformace, která k funkci $f: X \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ přiřadí **konvexní** funkci $f^*: \mathbb{R}^n \rightarrow \mathbb{R}$. Této přidružené funkci f^* se v řeči *optimalizace/matematického programování* říká **duální úloha** (viz Definice 4.3.3).

Definice 2.6.1 (Fenchelova transformace)

Nechť $f: \mathbb{R}^n \rightarrow \mathbb{R}$. Funkce $f^*(y) := \sup_{x \in \mathbb{R}^n} [\langle x, y \rangle - f(x)]$ se nazývá **Fenchelova transformace** funkce f (nebo také **(konvexně) konjugovanou funkcí** funkce f).

“ Jistě $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ a proto definujeme ještě **efektivní definiční obor** $D^*(f) = \{x \in D(f) \mid f(x) < \infty\}$

Lemma 2.6.3

Nechť je dána funkce $f: X \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ a f^* je její Fenchelova transformace. Pak následující tvrzení jsou pravdivá:

- Funkce f^* je konvexní na množině $Y := \{y \in \mathbb{R}^n \mid f^*(y) < \infty\}$
- Pro každé $x \in X$ a $y \in \mathbb{R}^n$ platí tzv. *Fenchelova(-Youngova) nerovnost* $f(x) + f^*(y) \geq \langle x, y \rangle$ přičemž **rovnost nastane** právě tehdy, když $y \in \text{subd} f(x)$.
- Je-li $f(x) \leq g(x)$ na X , pak $f^*(y) \leq g^*(y)$ pro všechna $y \in \mathbb{R}^n$.

Věta 2.6.6 (Fenchel & Moreau)

Nechť $X \subseteq \mathbb{R}^n$ je konvexní množina a funkce $f: X \rightarrow \mathbb{R}$ **konvexní** na X . Pak v **každém bodě spojitosti** funkce f platí tzv. *Fenchelova rovnost* $f^{**} = f$

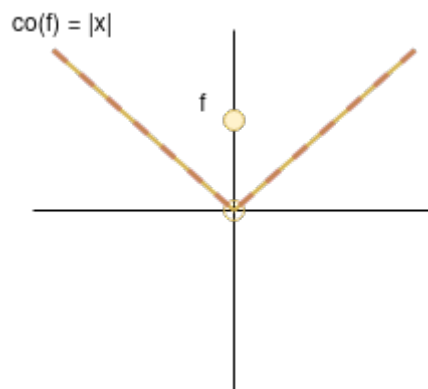
Jinak řečeno, **druhá** Fenchelova transformace ke **konvexní** funkci je s touto funkcí **totožná**, tj. $f^{**} \equiv f$ pro f **konvexní**. Navíc jelikož f^* je vždy konvexní, tak dostáváme, že počítat **třetí** Fenchelovu transformaci f^{**} nemá smysl, protože bude totožná s první Fenchelovou transformací f^* .

Definice 2.6.7 (Obálka funkce)

Nechť $X \subseteq \mathbb{R}^n$ je konvexní množina a $f: X \rightarrow \mathbb{R}$. Potom funkce $g(x) := \sup \{h(x) \mid h \text{ je konvexní a } h(x) \leq f(x) \text{ } \forall x \in X\}$ se nazývá **konvexní obálka (obal) funkce** f a značí se $\text{co } f$.

Jinak řečeno, $\text{co } f$ je **největší** konvexní funkce, která je **majorizována** funkcí f .

Jistě platí $D(\text{co } f) = \text{conv}(D(f))$, z čehož plyne $\text{conv}(\text{epi } f) \subseteq \text{epi}(\text{co } f)$



Zde $\text{conv}(\text{epi } f) = \text{epi} \{|x|\} \setminus \{0\}$, ale $0 \in \text{epi}(\text{co } f)$

Věta 2.6.9

Nechť $X \subseteq \mathbb{R}^n$ je konvexní množina a $f: X \rightarrow \mathbb{R}$. Potom pro každé $x \in X$ platí $\text{co } f(x) = f^{**}(x)$

Rychlost konvergence

Nechť jsou dány 2 posloupnosti $\{e_k\}_{k=0}^{\infty}$ a $\{h_k\}_{k=0}^{\infty}$ takové, že $e_k \in [0, \infty)$, $e_k \rightarrow 0$ a $h_k \in [0, \infty)$, $h_k \rightarrow 0$. Řekneme, že posloupnost $\{e_k\}$ konverguje **rychleji** (*pomaleji*) než $\{h_k\}$, pokud existuje index $\tilde{k} \in \mathbb{N}_0$ takový, že $e_k \leq (\geq) h_k$ $\forall k \in [\tilde{k}, \infty) \cap \mathbb{N}_0$.

Nechť je dána posloupnost $\{e_k\}_{k=0}^{\infty}$ splňující $e_k \in [0, \infty)$ a $e_k \rightarrow 0$. Řekneme, že posloupnost $\{e_k\}$ konverguje

- **alespoň lineárně s rychlostí** $\beta \in (0,1)$, pokud konverguje **rychleji** než **geometrická** posloupnost se členy tvaru $q \cdot \beta^k$, kde $q > 0$ a $\beta \in (\beta, 1)$.
 - čím **větší** β , tím **pomaleji** jde tato geometrická posloupnost k nule - tj. konverguje rychleji než geometrická posloupnost s β větší než β
- **nejvýše lineárně s rychlostí** $\beta \in (0,1)$, pokud konverguje **pomaleji** než **geometrická** posloupnost se členy tvaru $q \cdot \beta^k$, kde $q > 0$ a $\beta \in (\beta, 1)$

$(0, \beta)$.

- **lineárně s rychlostí** $\beta \in (0,1)$, pokud konverguje **nejvýše** a současně **alespoň** lineárně s rychlostí β .
- **superlineárně (sublineárně)**, pokud konverguje **rychleji** (*pomaleji*) než libovolná geometrická posloupnost se členy tvaru $q \beta^k$, kde $q > 0$ a $\beta \in (0,1)$.

Definice $\mathcal{D}\{3.4\}$

Nechť je dána posloupnost $\{e_k\}_{k=0}^{\infty}$ splňující $e_k \in [0, \infty)$ a $e_k \rightarrow 0$, přičemž $\{e_k\}$ konverguje **superlineárně**. Řekneme, že posloupnost $\{e_k\}$ konverguje

- **alespoň superlineárně s řádem** $p > 1$, pokud konverguje **rychleji** než všechny posloupnosti se členy tvaru $q \beta^{\lfloor p^k \rfloor}$, kde $q > 0$ a $\beta \in (0,1)$ a $\lfloor p^k \rfloor \in (1, p)$
 - čím **větší** $\lfloor p^k \rfloor$, tím **rychleji** posloupnost $q \beta^{\lfloor p^k \rfloor}$ konverguje - tj. $\{e_k\}$ konverguje rychleji než všechny posloupnosti s **menším** p
- **nejvýše superlineárně s řádem** $p > 1$, pokud konverguje **pomaleji** než všechny posloupnosti se členy tvaru $q \beta^{\lfloor p^k \rfloor}$, kde $q > 0$ a $\beta \in (0,1)$ a $\lfloor p^k \rfloor \in (p, \infty)$
- **superlineárně s řádem** $p > 1$, pokud konverguje **nejvýše** a současně **alespoň** superlineárně s řádem p .
- **superlineárně s řádem** $p = 1$, pokud konverguje **pomaleji** než všechny posloupnosti se členy tvaru $q \beta^{\lfloor p^k \rfloor}$, kde $q > 0$ a $\beta \in (0,1)$ a $\lfloor p^k \rfloor \in (1, \infty)$

Metody

Definice $\mathcal{D}\{3.1.1\}$ (Unimodální funkce)

Nechť je dán interval $I \subset \mathbb{R}$ a funkce $f: I \rightarrow \mathbb{R}$. Řekneme, že f je **unimodální** na I , jestliže existuje $x^* \in I$ takové, že

- $f(x_1) > f(x_2)$ pro libovolná $x_1, x_2 \in I$ splňující $x^* > x_1 > x_2$
- $f(x_1) < f(x_2)$ pro libovolná $x_1, x_2 \in I$ splňující $x^* < x_1 < x_2$

Jinými slovy, **unimodální** funkce je **klesající** na $(-\infty, x^*) \cap I$ (tj. *nalevo* od x^*) a **rostoucí** na $I \cap (x^*, \infty)$ (tj. *napravo* od x^*).

Unimodalita **neimplikuje** konvexnost (*ani se spojitostí*), pouze **kvazikonvexnost**

Naopak, *konvexní* funkce **nemusí** nutně být *unimodální* (ale **ostrá konvexnost** \implies **unimodalita**)

“ Konvexní funkce nemusí být např. jen rostoucí, ale i **neklesající**

V této části budeme řešit úlohu $f(x) \rightarrow \min, \text{ } x \in I := [a, b]$ $\tag{T{3.1.1}}$

Lemma $D{3.1.2}$

Nechť $f: I \rightarrow \mathbb{R}$ je **unimodální** na I a $x_1, x_2 \in I$ jsou takové, že $x_1 < x_2$.

- Je-li $f(x_1) \leq f(x_2)$, pak $x^* \leq x_2$
- Je-li $f(x_1) \geq f(x_2)$, pak $x^* \geq x_1$

Upozornění:

Dále uvažujme **POUZE UNIMODÁLNÍ** funkce.

Navíc nechť N značí **povolený počet vyčíslení** a **přesnost** těchto metod je dáno jako $\bar{x} - x^* |$, kde x^* je **přesné** řešení úlohy $\tag{Eq{3.1.1}}$ a \bar{x} jeho nalezená aproximace.

Metoda prostého dělení

“ Tato metoda **není** efektivní a je to *de facto* hrubá síla

Podle parity N určíme dělicí body intervalu I .

N liché	N sudé
$x_i := a + \{b - a \over N + 1\} i, \text{ } i=1, \dots, N = 2k - 1$	$x_{2i} := a + \{b - a \over k + 1\} i \text{ } \text{and } \text{ } x_{2i-1} := x_{2i} - \delta, \text{ } i = 1, \dots, k := N/2,$

kde δ je *vhodné malé číslo*.

Poté vyčíslíme $f(x_1), \dots, f(x_N)$ (což v případě N **sudého** a $\delta \in \{0, \{b-a \over k+1\}\}$ znamená **pouze** k vyčíslení) a nechť v x_j nastává nejmenší hodnota, tj. $f(x_j) = \min_{1 \leq i \leq N} f(x_i)$ Pak z **Lemma $D{3.1.2}$** plyne, že $x^* \in [x_{j-1}, x_{j+1}]$ a tento interval nazveme **interval lokalizace minima (ILM)** a za aproximaci x^* vezmeme *střed* ILM, tj. $\bar{x} := \{x_{j-1} + x_{j+1} \over 2\}$

Pro **délku** $|I_N|$ intervalu lokalizace minima platí $|I_N| := \max_{1 \leq i \leq N} (x_{i+1} - x_{i-1}) = \begin{cases} 2 \cdot \frac{b-a}{N+1}, & N = 2k-1 \\ \frac{b-a}{N/2} + \delta, & N = 2k \end{cases}$

“ Pro N **sudé** je **poslední** interval delší, proto dostáváme takový tvar $|I_N|$

“ Přesnost této metody je dána **polovinou ILM**, tj. $\frac{|I_N|}{2}$

Rychlost konvergence této metody je **sublineární**, navíc je tento algoritmus **pasivní**, tj. volba x_{m+1} **nezáleží** na x_1, \dots, x_m (závisí pouze na N , či na N a δ).

“ Při rovnosti funkčních hodnot **preferujeme konec** (teoreticky by to mělo být jedno)

Metoda půlení intervalu

Nechť nyní $N = 2k$. Položme $a_0 = a$, $b_0 = b$ a $x_i^- := \frac{a_{i-1} + b_{i-1}}{2} - \delta$ a $x_i^+ := \frac{a_{i-1} + b_{i-1}}{2} + \delta$, kde $\delta > 0$ je dostatečně malé a $i = 1, \dots, k$. Vyčíslíme funkci v x_i^- , x_i^+ , tj. dostaneme $f(x_i^-)$, $f(x_i^+)$. Pak

- jestliže $f(x_i^-) < f(x_i^+)$, pak podle **Lemma** 3.1.2 je **ILM** $[a_{i-1}, x_i^+]$ $\implies a_i = a_{i-1}, b_i = x_i^+$
- jestliže $f(x_i^-) > f(x_i^+)$, pak podle **Lemma** 3.1.2 je **ILM** $[x_i^-, b_{i-1}]$ $\implies a_i = x_i^-, b_i = b_{i-1}$

Takto můžeme tento proces opakovat (k -krát, jelikož máme $N = 2k$ povolených vyčíslení), kdy za a, b volíme krajní body **ILM** pro každý krok. Zřejmě, jako aproximaci x^* v k -tém kroku bereme **střed ILM** pro k -tý krok.

Délka **ILM** je v tomto případě $|I_k| = \frac{b-a}{2^k} + \frac{(2^k - 1)\delta}{2^{k-1}}$, přičemž $\delta \in [0, \frac{b-a}{2}]$ a navíc pro $k \rightarrow \infty$ je $|I_k| \rightarrow 2\delta$.

“ Z tohoto vyplývá, že čím menší δ , tím je metoda přesnější. Nicméně ve skutečnosti se můžeme dostat k zaokrouhlovacím chybám, které dokonce mohou způsobit, že špatně určíme velikosti $f(x_i^-)$, $f(x_i^+)$ (tím pádem bychom řekli, že x^* je v opačném intervalu, než je ve skutečnosti)

Tato metoda konverguje **lineárně** s rychlostí $1/2$.

“ Při rovnosti funkčních hodnot **zapomínáme konec** (teoreticky by to mělo být jedno)

Metoda zlatého řezu

Myšlenka metody zlatého řezu "vylepšuje" metodu půlení intervalu tak, že každá další iterace umožňuje pouze **jedno** další vyčíslení.

“ Zde τ je řešení rovnice $\tau^2 - \tau - 1 = 0$, tj. $\frac{1}{\tau} \approx 0.618$

Mějme funkci f , interval $[a, b]$, přesnost ϵ nebo počet vyčíslení $N \geq 2$:

- (Inicializace) Položíme $a_0 := a$, $b_0 := b$ a $k := 1$. Vypočteme $\mu_1 := a_0 + \frac{b_0 - a_0}{\tau^2}$ a $\lambda_1 := a_0 + \frac{b_0 - a_0}{\tau}$
- Je-li $k = N$, pokračujeme částí 5., jinak následuje krok 3.
- Vyčíslíme $f(\lambda_k)$ a $f(\mu_k)$. Jestliže $f(\lambda_k) \geq f(\mu_k)$:
 - Položíme $a_k := \lambda_k$, $b_k := b_{k-1}$, $\lambda_{k+1} := \mu_k$ a $\mu_{k+1} := f(\lambda_{k+1}) := f(\mu_k)$, $\mu_{k+1} := a_k + \frac{b_k - a_k}{\tau}$ a pokračujeme na krok 4.
 - Položíme $a_k := a_{k-1}$, $b_k := \mu_k$, $\lambda_{k+1} := \lambda_k$ a $\mu_{k+1} := f(\mu_{k+1}) := f(\lambda_k)$, $\lambda_{k+1} := a_k + \frac{b_k - a_k}{\tau^2}$ a pokračujeme na krok 4.
- Položíme $k := k+1$ a pokračujeme krokem 2.
- Stanovíme poslední ILM jako $[a_{k-1}, b_{k-1}]$ a vypočteme $\bar{x} := \frac{a_{k-1} + b_{k-1}}{2}$. **KONEC**

Tato metoda konverguje **lineárně** s rychlostí $\frac{1}{\tau} \approx 0.618$

“ Toto **neznámá**, že by **na stejný počet vyčíslení** byla tato metoda horší než [metoda půlení intervalu](#)

Průběh MZŘ s výchozím intervalem $I = [a_0, b_0]$ s délkou $\ell_0 := b_0 - a_0$.

i (i + 1 vyčíslení)	vzdálenost od a_{i-1} (b_{i-1})		délka ILM
	λ_i (μ_i)	μ_i (λ_i)	
1	ℓ_0/τ^2	ℓ_0/τ	ℓ_0/τ
2	ℓ_0/τ^3	ℓ_0/τ^2	ℓ_0/τ^2
\vdots	\vdots	\vdots	\vdots
$N - 1$	ℓ_0/τ^N	ℓ_0/τ^{N-1}	ℓ_0/τ^{N-1}

Fibonacciho metoda

V této poslední metodě uvažujme, že zkrácení δ může být **jiné** v každé kroku metody.

“ Necht F_n je n -té Fibonacciho číslo

Máme povoleno N vyčíslení, takže $M = N - 1$ a $\lambda_i = a_{i-1} + \{F_{N-i-1} \over F_{N-i+1}\} l_{i-1} = b_{i-1} - \{F_{N-1} \over F_{N-1+i}\} l_{i-1}$ $\mu_i = a_{i-1} + \{F_{N-1} \over F_{N-1+i}\} l_{i-1} = b_{i-1} - \{F_{N-i-1} \over F_{N-i+1}\} l_{i-1}$

Průběh FM s výchozím intervalem $I = [a_0, b_0]$ s délkou $\ell_0 := b_0 - a_0$.

i (i + 1 vyčíslení)	vzdálenost od a _{i-1} (b _{i-1})		délka ILM
	λ _i (μ _i)	μ _i (λ _i)	
1	$\frac{F_{N-2}}{F_N} \ell_0$	$\frac{F_{N-1}}{F_N} \ell_0$	$\frac{F_{N-1}}{F_N} \ell_0$
2	$\frac{F_{N-3}}{F_N} \ell_0$	$\frac{F_{N-2}}{F_N} \ell_0$	$\frac{F_{N-2}}{F_N} \ell_0$
⋮	⋮	⋮	⋮
i	$\frac{F_{N-i-1}}{F_N} \ell_0$	$\frac{F_{N-i}}{F_N} \ell_0$	$\frac{F_{N-i}}{F_N} \ell_0$
⋮	⋮	⋮	⋮
N − 2	$\frac{1}{F_N} \ell_0$	$\frac{2}{F_N} \ell_0$	$\frac{2}{F_N} \ell_0$
N − 1 & δ > 0	$\frac{1}{F_N} \ell_0$	$\frac{1}{F_N} \ell_0 + \delta$	$\frac{1}{F_N} \ell_0$ nebo $\frac{1}{F_N} \ell_0 + \delta$
N − 1 & δ < 0	$\frac{1}{F_N} \ell_0 + \delta$	$\frac{1}{F_N} \ell_0$	$\frac{1}{F_N} \ell_0$ nebo $\frac{1}{F_N} \ell_0 + \delta $

Tato metoda konverguje **lineárně** s rychlostí $\frac{1}{\tau} \approx 0.618$, tj. *stejně* jako [metoda zlatého řezu](#)

“ Fibonacciho metoda je (mírně) přesnější, než *metoda zlatého řezu* (která lze vnímat jako *limitní varianta* Fibonacciho metody). Nicméně u Fibonacciho metody je při změně N potřeba **všechny body přepočítat**, což u metody zlatého řezu **není**.

Numerické metody v \mathbb{R}^n

```

 $\langle \cdot, \cdot \rangle$  \xdef\scal#1#2{\langle #1, #2 \rangle} \xdef\norm#1{\left|\right| #1 \right|\right|}
\rho \xdef\dist{\rho} \xdef\and{\&} \xdef\brackets#1{\left\{ #1 \right\}} \xdef\parc#1#2{\frac {\partial
#1}{\partial #2}} \xdef\mtr#1{\begin{pmatrix} #1 \end{pmatrix}}
\bm#1{\boldsymbol{#1}} \xdef\mc#1{\mathcal{#1}}
\vv#1{\mathbf{#1}} \xdef\vp#1{\pmb{#1}} \xdef\ve{\varepsilon} \xdef\l{\lambda}
\th{\vartheta} \xdef\alpha{\alpha} \xdef\vf{\varphi} \xdef\Tagged#1{\text{#1}}
\tagged*#1{\text{#1}} \xdef\tagEqHere#1#2{\href{#2\#eq-#1}{\text{#1}}}
\tagDeHere#1#2{\href{#2\#de-#1}{\text{#1}}} \xdef\tagEq#1{\href{\#eq-
#1}{\text{#1}}} \xdef\tagDe#1{\href{\#de-#1}{\text{#1}}} \xdef\T#1{\htmlId{eq-
#1}{#1}} \xdef\D#1{\htmlId{de-#1}{\vv{#1}}} \xdef\conv#1{\mathrm{conv}\, , #1}
\cone#1{\mathrm{cone}\, , #1} \xdef\aff#1{\mathrm{aff}\, , #1} \xdef\lin#1{\mathrm{Lin}\, ,
#1} \xdef\span#1{\mathrm{span}\, , #1} \xdef\O{\mathcal O} \xdef\ri#1{\mathrm{ri}\, , #1}
\rd#1{\mathrm{r}\partial\, , #1} \xdef\interior#1{\mathrm{int}\, , #1} \xdef\proj{\Pi}
\epi#1{\mathrm{epi}\, , #1} \xdef\grad#1{\mathrm{grad}\, , #1}
\gradT#1{\mathrm{grad}^T #1} \xdef\gradx#1{\mathrm{grad}_x #1}
\hess#1{\nabla^2, #1} \xdef\hessx#1{\nabla^2_x #1} \xdef\jacobx#1{D_x #1}
\jacob#1{D #1} \xdef\subdif#1{\partial #1} \xdef\co#1{\mathrm{co}\, , #1}
\iter#1{\text{#1}} \xdef\str{\text{*}} \xdef\spv{\mc V} \xdef\civ{\mc U}
\knvxProg{\tagEqHere{4.1}{./nutne-a-postacujici-podminky-optimality} \, \text{and} \,
\tagEqHere{4.2}{nutne-a-postacujici-podminky-optimality}}

```

Budeme se věnovat úlohám (přesněji numerickým metodám jejich řešení) typu $f(x) \rightarrow \min$, $x \in \mathbb{R}^n$, $\tag{T{3.2.1}}$ kde $f: \mathbb{R}^n \rightarrow \mathbb{R}$ je (jednou/dvakrát/třikrát) **spojitě diferencovatelná** funkce. Obecně jsou metody numerické optimalizace založeny na **minimalizační posloupnosti** $\{x^k\}$ definované jako $x^{k+1} = x^k + a_k h_k$, kde $a_k \in \mathbb{R}$ se nazývá **délka k -tého kroku** a vektor $h_k \in \mathbb{R}^n$ je **směr k -tého vektoru**.

“ Budeme uvažovat tzv. **přesnou minimalizaci**, kdy dílčí minimalizace řešíme přesně (nikoliv numerickými metodami)

“ Všechny následující metody jsou **spádové**

Metoda největšího spádu

U této metody volíme $h_k = -\text{grad } f(x^{[k]})$, přičemž délku kroku volíme **přesným řešením** úlohy $f(x^{[k+1]}) = f(x^{[k]} - \alpha_k \text{grad } f(x^{[k]})) = \min_{\alpha \geq 0} f(x^{[k]} - \alpha \cdot \text{grad } f(x^{[k]}))$. Dále bude platit, že vektory určené body $x^{[k+1]}$, $x^{[k]}$ a $x^{[k+2]}$, $x^{[k+1]}$ jsou na sebe **ortogonální**. Z tohoto dostáváme, že pro daný směr hledáme **nejblížší** vrstevnici, která bude **tečná** k tomuto vektoru.

Tato metoda je **prvního řádu** (stačí nám pouze gradient).

“ V některých případech dochází k tzv. "cik-cak" efektu (klikatění), kdy se minimalizující posloupnost dostává k optimu velmi pomalu. Toto se děje například u **Rosenbrockovy (banánové) funkce** (jedna z *testovacích funkcí*)

Kvadratické funkce

Nechť f je kvadratická funkce tvaru $f(x) = \frac{1}{2} \text{scal} \{Qx\} x - x^T b$, kde $Q = Q^T > 0$ je **symetrická** $n \times n$ matice a $b \in \mathbb{R}^n$. Taková funkce f je **ostře** (i *silně*) konvexní. Z pozitivní definitnosti Q dostáváme, že vlastní čísla matice Q jsou **kladné** a můžeme je uspořádat následovně $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$. Díky tomu můžeme úlohu $\tag{Eq{3.2.1}}$ vyřešit "přímo" jako $x^* = Q^{-1} b$, nicméně počítání inverze může být **velice náročné** (výpočetně).

V tomto případě je gradient g funkce f dán jako $g(x) := \text{grad } f(x) = Qx - b$, tedy v jednotlivých iteracích dostáváme $g_k := Qx^{[k]} - b$ a α_k můžeme určit jako $\alpha_k = \frac{g_k^T g_k}{g_k^T Q g_k} \in [1/\lambda_n, 1/\lambda_1]$

Nyní se zaměříme na konvergenci metody, kterou můžeme zkoumat pomocí $E(x) := f(x) - f(x^*) = \frac{1}{2} (x - x^*)^T Q (x - x^*)$

Lemma $\tag{D{3.2.1}}$ (Konvergence metody největšího spádu)

Platí $E(x^{[k+1]}) = \left(1 - \frac{(g_k^T g_k)^2}{(g_k^T Q g_k)(g_k^T Q^{-1} g_k)}\right) E(x^{[k]})$

Z Lemmatu $\tag{De{3.1.2}}$ okamžitě plyne, že pokud pro nějaké $k \in \mathbb{N}$ nastane $1 = \frac{(g_k^T g_k)^2}{(g_k^T Q g_k)(g_k^T Q^{-1} g_k)}$, $\tag{T{3.2.1-a}}$ tak v $k+1$ metoda největšího spádu nalezne řešení **přesně**. V opačném případě je metoda **nekonečně-kroková**.

Rovnost $\tag{Eq{3.2.1-a}}$ nastane v případě, že g_k je **vlastním vektorem** matice Q , jinak řečeno gradient musí mířit **do středu** elipsy (*elipsoidu*).

V případě, že $\|x_1 - x_2\| = \|x_2 - x_3\| = \dots = \|x_{n-1} - x_n\|$ je konvergence **superlineární**. Naopak pokud $\|x_1 - x_2\| = \|x_2 - x_3\| = \dots = \|x_{n-1} - x_n\| \neq 0$, tak může být konvergence **velice pomalá**. Ve skutečnosti ještě rychlost konvergence závisí na počátečním $x^{[0]}$

Nekvadratické funkce

V případě nekvadratické funkce je metoda největšího spádu schopna nalézt **pouze stacionární body**.

Lemma 3.2.2iii (Lokální konvergence)

Nechť $f: \mathbb{R}^n \rightarrow \mathbb{R}$ je **spojitě diferencovatelná**. Jestliže bod $x^{[0]} \in \mathbb{R}^n$ je takový, že množina $\{x \in \mathbb{R}^n \mid f(x) \leq f(x^{[0]})\}$ je **ohraničená**, pak posloupnost $\{x^{[k]}\}$ generovaná metodou největšího spádu konverguje k bodu x^* , kde $\nabla f(x^*) = 0$.

Pokud konverguje $\{x^{[k]}\}$ k bodu x^* a funkce f je **dvakrát** spojitě diferencovatelná **na okolí** x^* a platí $\alpha \leq \lambda_{\min}(\nabla^2 f(x^*)) \leq A$, kde $\alpha, A > 0$ (tedy f je v okolí x^* **silně konvexní**), pak metoda konverguje (**alespoň**) s rychlostí $\left(\frac{A - \alpha}{A + \alpha}\right)^2$

“ Tedy i pro nekvadratické funkce hraje velkou roli podmíněnost matice $\nabla^2 f(x^*)$

“ Metoda největšího spádu se nejčastěji využívá v jiných metodách jako pomocné, když ony metody samotné v tu chvíli neposkytují dostatečné zlepšení

Celkem můžeme *metodu největšího spádu* shrnout:

- **globální** konvergence (pro nekvadratické metody za dalších předpokladů)
- **pomalá** konvergence
 - mnohdy numericky ani nekonverguje
- je základem pro další (lepší) metody

Newtonova metoda

Hlavní myšlenkou Newtonovy metody je, že v $(k+1)$ -kroku, kde $k \in \mathbb{N}_0$, funkci f aproximujeme **Taylorovým polynomem druhého řádu** se středem v bodě $x^{[k]}$ a jako $x^{[k+1]}$ volíme bod, ve kterém tento polynom nabývá svého minima.

Jinak řečeno místo **tečné nadroviny** k funkci konstruuujeme **tečnou n -rozměrnou parabolou**

Tedy místo funkce f uvažujeme v k -tém kroku $T_k(x) := f(x^{[k]}) + \nabla f(x^{[k]})^T (x - x^{[k]}) + \frac{1}{2} (x - x^{[k]})^T \nabla^2 f(x^{[k]}) (x - x^{[k]}) \approx f(x)$

Jelikož hledáme řešení $T_k(x) \rightarrow \min$, tak výraz výše první zderivujeme, z čehož dostaneme $\nabla T_k(x) = \nabla f(x^{[k]}) + \nabla^2 f(x^{[k]}) (x - x^{[k]})$, což v případě **regulární** matice $\nabla^2 f(x^{[k]})$ vede na $x^{[k+1]} = x^{[k]} - (\nabla^2 f(x^{[k]}))^{-1} \nabla f(x^{[k]})$

“ Pro $\nabla^2 f(x) > 0$ je funkce **konvexní** a nalezneme **minimum**

Nicméně je důležité podotknout, že výpočet $(\nabla^2 f(x^{[k]}))^{-1}$ je **velmi výpočetně náročný**. Avšak v případě **kvadratické funkce** Newtonova metoda nalezne řešení **v jednom kroku**, tedy její rychlost konvergence je **superlineární** s řádem ∞ .

Regularita matice $\nabla^2 f(x^{[k]})$ je velmi důležitá pro konvergenci, viz následující věta.

Věta $D\{3.2.5\}$

Nechť $f \in C^3$ v okolí bodu $x^* \in \mathbb{R}^n$, který je **nedegenerovaným minimem**, tj. $\nabla f(x^*) = 0$ a $\nabla^2 f(x^*) > 0$. Potom pro $x^{[0]} \in \mathbb{R}^n$ dostatečně blízko x^* konverguje $\{x^{[k]}\}$ generovaná Newtonovou metodou k bodu x^* **superlineárně** s řádem (alespoň) $p = 2$ (tj. *kvadraticky*).

“ Zde určit, co znamená "dostatečně blízko x^* " je obtížné

Celkem můžeme *Newtonovu metodu* shrnout jako:

- **velmi rychlá** konvergence
- nutnost **dostatečně blízké počáteční aproximace**
- velmi velká výpočetní náročnost při velkém n (počtu dimenzí)

Metoda sdružených gradientů

Uvažujme situaci v úloze $\text{tagEq}\{3.2.1\}$, kdy máme funkci $f(x) = \frac{1}{2} x^T Q x - b^T x$ $\text{tag}\{T\{MSG\}\}$, kde $Q = Q^T \in \mathbb{R}^{n \times n}$, $Q > 0$ je symetrická matice a $b \in \mathbb{R}^n$.

Pak nalezení úlohy $\text{tagEq}\{3.2.1\}$ a $\text{tagEq}\{MSG\}$ je ekvivalentní s řešením úlohy $Qx = b$ $\text{tag}\{T\{MSGa\}\}$, kterou umíme řešit například Gaussovou eliminací.

Metoda sdružených gradientů je v případě $Q > 0$ přímou metodou, která dojde k řešení $\text{tagEq}\{MSGa\}$ po n krocích. Nicméně tento fakt lze brát také jako, že je to iterační metoda s velmi rychlou konvergencí v případě pozitivně definitní matice.

Definice $\text{D}\{3.2.7\}$ (Q -sdružené vektory)

Nechť $Q = Q^T \in \mathbb{R}^{n \times n}$ je **pozitivně definitní**. Vektory $h_1, h_2 \in \mathbb{R}^n \setminus \{0\}$ se nazývají **Q -sdružené** (nebo také **Q -ortogonální**), jestliže $\text{scal}\{Qh_1\}{h_2} = h_1^T Q h_2 = 0$. Systém vektorů $\{h_0, \dots, h_{m-1}\} \in \mathbb{R}^n \setminus \{0\}$ pro $m \in \{2, \dots, n\}$ se nazývá **Q -sdružený**, jestliže $\text{scal}\{Q h_i\}{h_j} = 0$ pro $i \neq j$.

Věta $\text{D}\{3.2.8\}$

Nechť systém vektorů $\{h_0, \dots, h_{m-1}\} \in \mathbb{R}^n \setminus \{0\}$ s $m \in \{2, \dots, n\}$ je **Q -sdružený**. Potom jsou tyto vektory **lineárně nezávislé**.

Věta $\text{D}\{3.2.9\}$

Nechť $m \in \{2, \dots, n\}$ a mějme systém **Q -sdružených** vektorů $\{h_0, \dots, h_{m-1}\} \in \mathbb{R}^n$. Nechť $x_{\text{iter } 0}$ je **dáno** a body $x_{\text{iter } 1}, \dots, x_{\text{iter } m}$ jsou dány jako $x_{\text{iter } k+1} = x_{\text{iter } k} + \alpha_k h_k = x_{\text{iter } 0} + \sum_{i=0}^k \alpha_i h_i$, $\forall k \in \{0, \dots, m-1\}$, $\text{tag}\{T\{3.2.8\}\}$ kde α_k jsou volena tak, že $f(x_{\text{iter } k} + \alpha_k h_k) = \min_{\alpha \in \mathbb{R}} f(x_{\text{iter } k} + \alpha h_k)$ pro $k \in \{0, \dots, m-1\}$ (tj. jsou volena **přesnou minimalizací**). Pak pro kvadratickou funkci f definovanou v $\text{tagEq}\{MSG\}$ platí $f(x_{\text{iter } m}) = \min_{x \in X_m} f(x)$, kde $X_m := \text{lin}\{h_0, \dots, h_{m-1}\}$ (viz **Definice** $\text{tagDeHere}\{2.1.6\}$./konvexni-mnoziny) - lineární obal). Zejména pro $m = n$ dostáváme $f(x_{\text{iter } n}) = \min_{x \in \mathbb{R}^n} f(x)$, tj. $x_{\text{iter } n}$ je řešením úlohy $\text{tagEq}\{3.2.1\}$ a $\text{tagEq}\{MSG\}$.

“ Nalezení **Q -sdružených** vektorů lze provést zobecněným **Gram-Schmidtovým ortogonalizačním procesem** (ten je uveden v Lin. Alg. ve speciálním tvaru pro $Q = I$).

Explicitně můžeme odvodit délku α_k -tého kroku jako $\alpha_k = -\frac{h_k^T \text{grad } f(x_{\text{iter } k})}{h_k^T Q h_k}$ $\text{tag}\{T\{3.2.9\}\}$

Celkem můžeme metodu popsat následovně $h_0 := -\text{grad } f(x_{\text{iter } 0})$, $\forall k \quad h_k := -\text{grad } f(x_{\text{iter } k}) + \beta_{k-1} h_{k-1}$ $\text{tag}\{3.2.10\}$ $\beta_{k-1} := \frac{\text{grad } f(x_{\text{iter } k})^T h_{k-1}}{h_{k-1}^T Q h_{k-1}}$ $\text{tag}\{3.2.11\}$, přičemž body minimalizující posloupnosti jsou počítány podle **Věty** $\text{tagDe}\{3.2.9\}$.

“ Tento výpočet lze "zjednodušit", viz $\text{Tagged}\{3.2.12\}$ v přednášce.

Hlavní výhodou **metody sdružených gradientů** je její **snadná implementace**, naopak nevýhodou citlivost na podmíněnost matice Q . Také se daří říct, že metoda sdružených gradientů **konverguje nejrychleji** z metod založených pouze na *maticovém násobení*.

Pro nekvadratické funkce používáme stejný algoritmus jako doted' až na volbu β_k , ale metodu restartujeme po n krocích

Věta D{3.2.13} (Rychlost konvergence)

Nechť $f \in C^3$ na \mathbb{R}^n , $x^0 \in \mathbb{R}^n$ a x^* je **nedegenerované lokální minimum**, tj. $\text{grad } f(x^*) = 0$ a $\text{hess } f(x^*) > 0$. Nechť x^k je výsledek **metody sdružených gradientů** s cyklem délky n a výchozím bodem x^{k-1} a nechť $x^k \rightarrow x^*$ pro $k \rightarrow \infty$. Potom **minimalizující posloupnost** $\{x^k\}$ konverguje **superlineárně** s řádem **alespoň** $p = 2$.

“ MSG souvisí s metodami *Krylovových podprostorů*.

Obecný úvod

- m - počet omezení

- k - počet nerovností (tzn. g_1, \dots, g_k jsou **nerovnosti**, zbytek rovnosti)

Omezení zakomponovaná v P se nazývají **přímá**, naopak omezení ve formě g_i se nazývají **funkcionální**. Dále definujeme

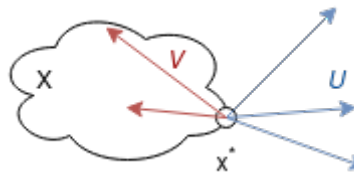
- **množinu přípustných vektorů**

$$\text{spv}(x^*, X) := \{h \in \mathbb{R}^n \mid \exists \lambda, \lambda_0 > 0 : x^* + th \in X \text{ pro } t \in (0, \lambda_0)\}$$

- je to **kužel**

- **množinu spádových vektorů (kužel zlepšujících vektorů)**

$$\text{civ}(x^*, f) := \{h \in \mathbb{R}^n \mid \exists \lambda, \lambda_0 > 0 : x^* + th \in D(f); \text{ a } f(x^* + th) < f(x^*) \text{ pro } t \in (0, \lambda_0)\}$$



Lemma 4.1.1

Nechť $X \subseteq \mathbb{R}^n$ a $f: X \rightarrow \mathbb{R}$ jsou dány. Je-li bod $x^* \in X$ lokálním řešením úlohy 4.1, potom $\text{spv}(x^*, X) \cap \text{civ}(x^*, f) = \emptyset$

Definice 4.1.2 (Stacionární bod)

Nechť množina $X \subseteq \mathbb{R}^n$ je **konvexní** a funkce $f: X \rightarrow \mathbb{R}$ je **diferencovatelná** na (nějaké otevřené množině obsahující) X . Řekneme, že bod $x^* \in X$ je **stacionárním bodem** úlohy 4.1 (nebo také **stacionárním bodem funkce** f na množině X), jestliže $\nabla f(x^*)^T (x - x^*) \geq 0$ pro každé $x \in X$.

“ Výraz v 4.1.2 je **směrová derivace** x^* do libovolného bodu v X - v těchto směrech musí být f **neklesající**

Pro $X = \mathbb{R}^n$ je podmínka 4.1.2 splněna **pouze** v případě $\nabla f(x^*) = 0$

Dále ukažme, že **stacionární bod** ve smyslu Definice 4.1.2 má vlastnosti, které od něj očekáváme.

Věta 4.1.3 (Vlastnosti stacionárního bodu)

Nechť $f: X \rightarrow \mathbb{R}$ je **diferencovatelná** na (nějaké otevřené množině obsahující **konvexní** množinu) $X \subseteq \mathbb{R}^n$.

- **Množinu indexů** všech funkcí, které se v bodě x realizují jako rovnost $S(x) := \bigcup_{k=1}^{\infty} \{x, x_{k+1}, \dots, x_m\}$, $\forall x \in X$

Věta 4.2.1 (Lagrangeův princip)

Nechť množina $P \subseteq \mathbb{R}^n$ je **konvexní**, funkce $f, g_1, \dots, g_k : P \rightarrow \mathbb{R}$ jsou **diferencovatelné** v bodě $x^* \in X$ a g_{k+1}, \dots, g_m jsou **spojitě diferencovatelné** na nějakém okolí bodu x^* . Je-li bod $x^* \in X$ **lokálním řešením** úlohy (4.1) & (4.2), pak existují **Lagrangeovy multiplikátory** $y_0 > 0$ a $y \in Q$ takové, že **ne všechna** y_0, \dots, y_m jsou **nulová** a platí $\nabla_x L(x, y_0, y) = 0$ a $x - x^* \geq 0$ $\forall x \in P$ (4.3) $y_i g_i(x) = 0$, $i = 1, \dots, m$ (4.4)

Podmínka (4.3) znamená, že x musí být **stacionárním bodem** funkce $L(x, y_0, y)$ (*podmínka stacionarity*). Dále podmínka (4.4) se nazývá **podmínka komplementarity** a požadavek $y_1, \dots, y_k > 0$ jako **podmínka duality**.

“Jistě $y_0, y_1, \dots, y_k \geq 0$, $y_{k+1}, \dots, y_m \in \mathbb{R}$ $\iff y_0 > 0$ a $y \in Q$

Jelikož situace s $y_0 = 0$ je problematická, existují podmínky na zaručení $y_0 \neq 0$, což je ekvivalentní s $y_0 = 1$. Tyto podmínky se nazývají **podmínky kvalifikovaného omezení**:

- **Regulárnost bodu** x , tj. bod x je **regulární**, pokud jsou $\nabla g_i(x)$ pro $i \in S(x)$ **lineárně nezávislé** (tj. *gradienty aktivních omezení jsou LNŽ*)
- **afinní omezení** - funkce g_1, \dots, g_m jsou **afinní**
- **Slaterova podmínka** - g_1, \dots, g_k jsou **konvexní**, g_{k+1}, \dots, g_m jsou **afinní**, **konstantní** vektory ∇g_i jsou **lineárně nezávislé** pro $i \in \{k+1, \dots, m\}$ a existuje $\bar{x} \in P$ takové, že $g_i(\bar{x}) < 0$ pro $i \in \{1, \dots, k\}$ a $g_j(\bar{x}) = 0$ pro $j \in \{k+1, \dots, m\}$

Důsledek 4.2.2

Nechť $P \subseteq \mathbb{R}^n$ je **konvexní** množina, funkce f, g_1, \dots, g_m jsou **diferencovatelné** na (nějaké otevřené množině obsahující) P a pro $x \in X$ existují multiplikátory $y \in Q$ takové, že platí (4.3) & (4.4) s $y_0 = 1$. Nechť je dále splněn (alespoň) jeden z následujících předpokladů:

1. funkce $L(x, y)$ je **konvexní** na množině P
2. úloha (4.1) & (4.2) je úlohou **konvexního programování**, tj. na **konvexní** množině P jsou funkce f, g_1, \dots, g_k **konvexní** a g_{k+1}, \dots, g_m **afinní**

Pak bod x je **globálním řešením** úlohy (4.1) & (4.2).

Věta D{4.2.3} (Karushova-Kuhnova-Tuckerova v diferenciálním tvaru)

Nechť $P \subseteq \mathbb{R}^n$ je **konvexní** množina, funkce f, g_1, \dots, g_k **konvexní** a **diferencovatelné** na (nějaké otevřené množině obsahující) P , funkce g_{k+1}, \dots, g_m **afinní** na P a nechť platí (alespoň) jedna z následujících podmínek:

1. **(LNZ)** množina P je **otevřená**, vektory $\nabla g_i(x)$, $i \in S(x)$ jsou **lineárně nezávislé** pro každé $x \in X$.
2. **(Slaterova)** funkcionální omezení jsou pouze tvaru **nerovností**, tj. $k = m$, a **existuje** bod $\bar{x} \in P$ takový, že $g_i(\bar{x}) < 0$ pro $i \in \{1, \dots, k\}$
3. **(lineární)** množina P je **polyedr** a funkce g_1, \dots, g_k jsou **afinní**

Pak x^* je řešením úlohy D{4.1} & D{4.2} právě tehdy, když existuje $y^* \in Q$ takové, že platí D{4.2.3} & D{4.2.4} s $y_0^* = 1$.

Věta D{4.2.4}

Nechť funkce f, g_1, \dots, g_m jsou **dvakrát spojitě** diferencovatelné v bodě x^* a $x^* \in \text{interior } P$ je takový, že **existují** multiplikátory $y^* \in Q$ splňující D{4.2.3} & D{4.2.4} s $y_0^* = 1$ a současně $y_i^* > 0$ pro $i \in I(x^*)$ (tzv. **podmínka ostré komplementarity**), tj. $\nabla_x L(x^*, y^*) = 0$, $g_i(x^*) \leq 0$ pro $i \in \{1, \dots, k\}$, $g_i(x^*) = 0$ pro $i \in \{k+1, \dots, m\}$, $y_i^* > 0$ pro $i \in I(x^*)$, $y_i^* = 0$ pro $i \in \{1, \dots, k\} \setminus I(x^*)$, $y_i^* \in \mathbb{R}$ pro $i \in \{k+1, \dots, m\}$. Jestliže $\nabla^2_x L(x^*, y^*) > 0$ na $\ker(\nabla^T g_i(x^*))_{i \in S(x^*)}$, tj. $h^T \nabla^2_x L(x^*, y^*) h > 0$ pro všechna $h \in \mathbb{R}^n \setminus \{0\}$ taková, že $\nabla g_i(x^*) h = 0$ pro $i \in S(x^*)$, pak bod x^* je **ostré lokální minimum** funkce f na množině X .

Duální úloha

```

 $\left\langle \cdot, \cdot \right\rangle$   $\left\| \cdot \right\|$   $\frac{\partial}{\partial \cdot}$   $\begin{pmatrix} \cdot \\ \cdot \end{pmatrix}$   $\boldsymbol{\cdot}$   $\mathcal{M}$   $\mathbf{v}$   $\mathbf{v}^\top$   $\varepsilon$   $\lambda$   $\theta$   $\alpha$   $\varphi$   $\text{\tagged{#1}}$   $\text{\tagged{#1}}$   $\text{\tagEqHere{#2}{#eq-#1}}{\text{\tagged{#1}}}$   $\text{\tagDeHere{#2}{#de-#1}}{\text{\tagged{#1}}}$   $\text{\tagEq{#1}}{\text{\tagged{#1}}}$   $\text{\tagDe{#1}}{\text{\tagged{#1}}}$   $\text{\tagT{#1}}{\text{\tagged{#1}}}$   $\text{\tagD{#1}}{\text{\tagged{#1}}}$   $\text{\tagConv{#1}}$   $\text{\tagCone{#1}}$   $\text{\tagAff{#1}}$   $\text{\tagLin{#1}}$   $\text{\tagSpan{#1}}$   $\text{\tagO{#1}}$   $\text{\tagRi{#1}}$   $\text{\tagRd{#1}}$   $\text{\tagInterior{#1}}$   $\text{\tagProj{#1}}$   $\text{\tagEpi{#1}}$   $\text{\tagGrad{#1}}$   $\text{\tagGradT{#1}}$   $\text{\tagGradX{#1}}$   $\text{\tagHess{#1}}$   $\text{\tagHessX{#1}}$   $\text{\tagSubdif{#1}}$   $\text{\tagCo{#1}}$   $\text{\tagIter{#1}}$   $\text{\tagStr{#1}}$   $\text{\tagSpv{#1}}$   $\text{\tagCiv{#1}}$   $\text{\tagKnvxProg{#1}}$   $\text{\tagEqHere{4.1}}$   $\text{\tagEqHere{4.2}}$ 

```

Definice $\text{\tagD{4.3.1}}$ (Kuhnovy-Tuckerovy vektory)

Vektor $\mathbf{y} \in Q$ (prvních k složek je nezáporných) se nazývá **Kuhnovým-Tuckerovým** vektorem (**K-T** vektorem) úlohy \tagKnvxProg , jestliže $f(\mathbf{x}) + \sum_{i=1}^m y_i g_i(\mathbf{x}) = L(\mathbf{x}, \mathbf{y}) \quad \forall \mathbf{x} \in P$, $\text{\tagT{4.3.0}}$ kde $f(\mathbf{x}) := \inf_{\mathbf{x} \in X} f(\mathbf{x})$ je hodnota úlohy \tagKnvxProg .

“ **K-T** vektor pro danou úlohu **nemusí** existovat

Věta $\text{\tagD{4.3.2}}$

Nechť úloha \tagKnvxProg je úlohou **konvexního programování**, tj. množina $P \subseteq \mathbb{R}^n$ je **konvexní**, funkce f, g_1, \dots, g_k **konvexní** a g_{k+1}, \dots, g_m jsou **afinní**, a nechť dále platí (alespoň) jedna z podmínek **regularity**:

1. **(Slaterova)** $k = m$ a existuje $\bar{\mathbf{x}} \in P$ takové, že $g_i(\bar{\mathbf{x}}) < 0$ pro $i = 1, \dots, m$

2. **(lineární)** množina P je **polyedr**, funkce f, g_1, \dots, g_k jsou **afinní** a $X \neq \emptyset$.

Pak **existuje K-T** vektor úlohy knvxProg .

“ Zde se podmínka 2. liší od podmínky 3. ve **Větě 4.2.3** - vyžaduje ještě **neprázdnot** přípustné množiny

Úloha konvexního programování splňující nějakou z podmínek z Věty 4.3.2 se nazývá **regulární**.

Definice 4.3.3 (Duální úloha)

Nechť $y \in Q$. Definujme funkci $\varphi(y) := \inf_{x \in P} L(x, y) = \inf_{x \in P} (f(x) + \sum_{i=1}^m y_i g_i(x))$ a množinu (tzv. **efektivní definiční obor**) $Y := \{y \in Q \mid \varphi(y) > -\infty\}$. Pak úloha $\varphi(y) \rightarrow \max, y \in Y$ se nazývá **duální úlohou** k úloze knvxProg . Číslo $\varphi^* := \sup_{y \in Y} \varphi(y)$ se nazývá **hodnotou duální úlohy**.

Úloha 4.3.1 je úlohou **konkávního** programování, tj. množina Y je **konvexní** a funkce φ je **konkávní** na Y .

Věta 4.3.5 (Slabá věta o dualitě)

Pro každé $x \in X$ a každé $y \in Q$ platí $f(x) \geq \varphi(y)$. Zejména, pokud $X \neq \emptyset$ a $Y \neq \emptyset$, pak $\varphi^* \leq f^*$.

“ V případě $X = \emptyset$ a/nebo $Y = \emptyset$ je nerovnost splněna *triviálně*, neboť $\inf \emptyset = -\infty$ a $\sup \emptyset = -\infty$.

Věta 4.3.5 říká, že pro **duální rozdíl** $g(x, y)$ s $x \in X \neq \emptyset$ a $y \in Y \neq \emptyset$ bude platit $g(x, y) := f(x) - \varphi(y) \geq 0$. Navíc číslo $g(x^*, y^*) := f^* - \varphi^*$ udává tzv. **optimální duální rozdíl** (*optimal duality gap*). Dá se také říct, že pro libovolné $y \in Q$ je hodnota $\varphi(y)$ **dolní hranicí minima účelové** funkce úlohy knvxProg .

Certifikát optimality

Jsou-li $x^* \in X$ a $y^* \in Q$ taková, že platí $f(x^*) = \varphi(y^*)$, pak x^* a y^* jsou **optimálními řešeními** svých příslušných úloh.

Duální rozdíl je úzce spjat s **existencí K-T vektorů**. Jestliže je duální rozdíl **nenulový**, tj. $f^* > v^*$, pak **množina K-T vektorů musí být prázdná**.

“ Jinak řečeno, **existence K-T** vektorů zaručuje $f^* = v^*$

Věta 4.3.6 (Silná věta o dualitě)

Nechť úloha knvxProg je **regulární úlohou** konvexního programování (viz. Věta 4.3.2). Pokud $f^* > -\infty$, pak platí tzv. **vztah duality** $f^* = v^*$, tj.
$$\inf_{x \in P} \sup_{y \in Q} L(x,y) = \sup_{y \in Q} \inf_{x \in P} L(x,y),$$
 přičemž množina řešení duální úlohy 4.3.1 je **neprázdná a shodná** s množinou všech **K-T vektorů** úlohy knvxProg .

Z Věty 4.3.6 vyplývá, že pokud knvxProg je **regulární úlohou konvexního programování** (viz Věta 4.3.2) a

- 1. jestliže $Y \neq \emptyset$, pak **duální úloha je řešitelná** a $f^* > -\infty$
- 2. jestliže $Y = \emptyset$, pak $f^* = -\infty$

Celkem z Vět 4.3.5, 4.3.6 a z bezprostředně výše uvedeného důsledku vyplývá, že v případě **regulární úlohy konvexního programování** mohou nastat **pouze 2** možnosti

Duální Ú. \ Primární Ú.	Nepřípustná ($f^* = -\infty$)	Přípustná a Omezená	Neomezená ($f^* = -\infty$)
Neomezená ($v^* = \infty$)	NE (Ano bez regularity)	NE	NE
Přípustná a Omezená	NE (Možná bez regularity)	ANO	NE
Nepřípustná ($v^* = -\infty$)	NE (Ano bez regularity)	NE	ANO

“ Z **regularity** plyne, že $X \neq \emptyset$

Věta 4.3.8 (Kuhnova-Tuckerova v nediferenciálním tvaru)

Nechť úloha knvxProg je **regulární úlohou konvexního programování** (viz Věta 4.3.2). Pak $x^* \in X$ je řešením této úlohy právě tehdy, když platí (*alespoň*) jedna z podmínek:

- 1. existuje $y^* \in Q$ takové, že $f(x^*) = v(y^*)$

2. existuje $y \in Q$ takové, že

$$L(x, y) = \min_{x \in P} L(x, y), \quad g_i(x) = 0,$$

$\forall i \in \{1, \dots, m\}$. Navíc množina takovýchto vektorů $y \in Q$ **splývá** s množinou **řešení duální** úlohy (a podle Věty 4.3.6 tedy i s množinou **K-T vektorů** úlohy knvxProg).

“ Jsou-li navíc funkce f, g_1, \dots, g_m **diferencovatelné** v bodě x , pak podmínka 4.3.2 je **ekvivalentní** s podmínkou 4.2.3 pro $y_0 = 1$, zatímco 4.3.3 **odpovídá** 4.2.4.

Tedy Věta 4.3.8 je skutečně **zobecnění KKT Věty** 4.2.3 pro případ **nediferencovatelných** funkcí

Také se dá říct, že koncept **K-T vektorů** je **zobecněním Lagrangeových multiplikátorů** s **kvalifikovanými omezeními** (kvůli $y_0 = 1$) - **K-T vektory splývají s multiplikátory** za podmínek Věty 4.2.3.

Definice 4.3.9 (Sedlový bod)

Bod $[x, y] \in P \times Q$ se nazývá **sedlovým bodem** Lagrangeovy funkce $L(x, y)$ úlohy knvxProg na $P \times Q$, jestliže $L(x, y) \leq L(x, y^*) \leq L(x^*, y) \quad \forall x \in P, y \in Q$, tj. platí $L(x, y) = \max_{y \in Q} L(x, y) = \min_{x \in P} L(x, y)$.

Věta 4.3.10 (Kuhnova-Tuckerova pro sedlový bod)

Nechť úloha knvxProg je **regulární úlohou** konvexní programování (viz Věta 4.3.2). Pak bod $x \in P$ je **řešením** úlohy knvxProg právě tehdy, když **existuje** $y \in Q$ takové, že $[x, y]$ je **sedlovým bodem** Lagrangeovy funkce $L(x, y)$ úlohy knvxProg na $P \times Q$.

Analýza citlivosti

```

 $\langle \cdot, \cdot \rangle$   $\| \cdot \|$   $\frac{\partial}{\partial \cdot}$   $\begin{pmatrix} \cdot \\ \cdot \end{pmatrix}$   $\mathbf{\cdot}$   $\mathcal{\cdot}$   $\mathbf{v}$   $\mathbf{v}^p$   $\mathbf{v}^\epsilon$   $\lambda$   $\theta$   $\alpha$   $\varphi$   $\text{\tagged{#1}}$   $\text{\tagged{#1}} \text{\tagEqHere{#2}{#eq-#1}} \text{\tagDeHere{#2}{#de-#1}} \text{\tagEq{#1}} \text{\tagDe{#1}} \text{\tagT{#1}} \text{\tagD{#1}} \text{\tagConv{#1}} \text{\tagCone{#1}} \text{\tagAff{#1}} \text{\tagLin{#1}} \text{\tagSpan{#1}} \text{\tagO{#1}} \text{\tagRi{#1}} \text{\tagRd{#1}} \text{\tagInterior{#1}} \text{\tagProj{#1}} \text{\tagEpi{#1}} \text{\tagGrad{#1}} \text{\tagGradT{#1}} \text{\tagGradX{#1}} \text{\tagHess{#1}} \text{\tagHessX{#1}} \text{\tagJacobX{#1}} \text{\tagJacob{#1}} \text{\tagSubdif{#1}} \text{\tagCo{#1}} \text{\tagIter{#1}} \text{\tagStr{#1}} \text{\tagSpv{#1}} \text{\tagCiv{#1}} \text{\tagKnvxProg{#1}} \text{\tagEqHere{4.1}} \text{\tagEqHere{4.2}}$ 
```

Nyní se budeme věnovat řešení úlohy matematické programování v **závislosti na parametru**.

První se podíváme na úlohu matematické programování **s omezeními pouze ve tvaru rovností**, ale s obecnou závislostí na parametrech.

A v druhém (a posledním) případě se podíváme na úlohu **s omezeními ve tvaru nerovností**, ale s parametry **pouze** v podobě **absolutních členů** ve funkcích zadávajících tyto omezení.

“ Zde je dobré podotknout, že druhý případ je mírně užitečnější a navíc si uvědomme, že **1 rovnost lze napsat jako 2 nerovnosti**

Úlohy s rovnostmi

Věta $\text{\tagD{4.4.1}}$ (O obálce)

Mějme úlohu $f(x,r) \rightarrow \min, \text{qquad} g_1(x,r) = 0, \dots, g_m(x,r) = 0 \tag{T{AC.1}}$ kde $x \in \mathbb{R}^n, r \in \mathbb{R}^k, f, g_1, \dots, g_m \in C^1$. Pripusťme, že pro **každou hodnotu parametru** r má úloha $\tag{AC.1}$ **jediné řešení**, které označíme $x^*(r)$. Potom **hodnota úlohy** $\tag{AC.1}$ je $f^*(r) = f(x^*(r), r)$. Je-li $x^*(r)$ **diferencovatelná** vzhledem k r a **Jacobiho matice** $G(x^*(r), r) \in \mathbb{R}^{m \times n}$ má **plnou hodnost** m , pak platí $\frac{\partial f^*(r)}{\partial r_i} = \frac{\partial f}{\partial r_i}(x^*(r), r) + \sum_{j=1}^m y_j(r) \frac{\partial g_j}{\partial r_i}(x^*(r), r)$

Obálka je křivka, která se **množiny křivek dotýká** (tj. se dotýká každé křivky) a má společnou tečnu s danou křivkou

Jinak řečeno je **tečná** k množině **křivek**



Požadavky Věty $\tag{4.4.1}$ jsou relativně silné, proto uveďme její "slabší verzi".

Věta $\tag{4.4.2}$

Nechť $f, g_1, \dots, g_m \in C^2$ a x^* je **lokálním řešením** úlohy $f(x) \rightarrow \min, \text{qquad} g_1(x) = 0, \dots, g_m(x) = 0$ s odpovídajícími Lagrangeovými multiplikátory y^* . Nechť dále tato dvojice splňuje **postačující podmínku druhého řádu**, tj. $\text{hess}_x L(x^*, y^*) > 0$ na $\ker \text{Jacobi } G(x^*)$, přičemž současně x^* je **regulárním bodem**, tj. $\text{Jacobi } G(x^*) \in \mathbb{R}^{m \times n}$ má **plnou hodnost** m . Uvažme úlohu **parametrického programování** $f(x) \rightarrow \min, \text{qquad} G(x) = u \tag{T{AC.2}}$, kde $u \in \mathbb{R}^m$. Pak **existuje otevřená koule** S se **středem v počátku** ($u = 0$) taková, že pro každé $u \in S$ existuje **lokální řešení** $x^*(u) \in \mathbb{R}^n$ úlohy $\tag{AC.2}$ a odpovídající $y^*(u) \in \mathbb{R}^m$. Navíc $x^*(\cdot)$ a $y^*(\cdot)$ jsou **spojitě diferencovatelné** funkce na S a platí $x^*(0) = x^*, y^*(0) = y^*$ a pro každé $u \in S$ máme $\text{grad } f^*(u) = -y^*(u)$, kde $f^*(u)$ značí **optimální hodnotu úlohy** $\tag{AC.2}$ vzhledem k u , tj. klademe $f^*(u) := f(x^*(u))$.

Jednoduše řečeno se optimální hodnota mění **podle Lagrangeových multiplikátorů** pro danou hodnotu u

Úlohy s nerovnostmi

Uvažujme úlohu závislou na m -tici parametrů $b = (b_1, \dots, b_m)^T \in \mathbb{R}^m$, tj. $f(x) \rightarrow \min, \quad x \in X(b) := \{x \in P \mid g_i(x) \leq b_i, \quad i \in \{1, \dots, m\}\}$ $\tag{T{4.4.2}}$ A dále zavedme značení $G(x) := (g_1(x), \dots, g_m(x))^T, \quad X(b) := \{x \in P \mid G(x) \leq b\}$ $B := \{b \in \mathbb{R}^m \mid X(b) \neq \emptyset\}, \quad F(b) := \inf_{x \in X(b)} f(x), \quad b \in B$ množinu **K-T vektorů** úlohy $\tag{4.4.2}$ označme $Y(b) := \{y \in \mathbb{R}^m \mid y \geq 0, \quad F(b) \leq f(x) + \sum_{i=1}^m y_i (G_i(x) - b_i) \quad \forall x \in P\}$ a **subdiferenciál** funkce $F(b)$ (viz Definice $\tag{2.5.1}$ [./subgradient-a-subdiferencial-a-fenchelova-transformace](#)) označme $\text{subdif } F(b) := \{a \in \mathbb{R}^m \mid F(b') - F(b) \leq \sum_{i=1}^m a_i (b'_i - b_i) \quad \forall b' \in B\}$

Věta $\tag{4.4.3}$

Nechť množina $P \subseteq \mathbb{R}^n$ je **konvexní**, funkce f, g_1, \dots, g_m jsou **konvexní** na P a platí $0 \in B, F(0) > -\infty$ a $Y(0) \neq \emptyset$. Potom

- množina B je **konvexní**
- funkce $F(b)$ je **konečná, konvexní a nerostoucí** na B
- platí $\text{subdif } F(b) = -Y(b)$ pro **všechna** $b \in B$

“ Z předpokladů Věty $\tag{4.4.3}$ plyne, že

- úloha $\tag{4.4.2}$ je **přípustná** pro $b = 0$
- úloha $\tag{4.4.2}$ **má řešení** pro $b = 0$
- množina **K-T vektorů** úlohy $\tag{4.4.2}$ je **neprázdná** (toto **není** splněno automaticky, viz Věta $\tag{4.3.2}$ [./dualni-uloha](#))

Navíc je-li F dokonce **diferencovatelná** v b , pak $\text{subdif } F(b)$ je **jednoprvková** množina, která obsahuje **pouze** $-\text{grad } F(b)$ a tedy tento vektor musí být roven $(-1) \cdot$ **jediný K-T vektor** této úlohy. (Toto je analogie Věty *O obálce* $\tag{4.4.1}$)

“ Podle Věty $\tag{4.3.6}$ [./dualni-uloha](#) jsme popsali **K-T vektory** úlohy knvxProg pomocí řešení **duální úlohy** $\tag{4.3.1}$ [./dualni-uloha](#). Část 3. této věty jim navíc dává ještě charakteristiku **subgradientu**

hodnoty úlohy parametrického programování §4.4.2§.

“ V případě **regulární** úlohy *konvexního programování* dostáváme zkombinováním těchto dvou výsledků §4.4.3§ $F(b) = -\inf_{y \in Y(b)} \sum_{i=1}^m g_i(y)$, kde $Y(b)$ je množina řešení **duální úlohy** (viz Věta §4.3.6§).

Pomocí Věty §4.4.3§ jsme schopni dostat zajímavé výsledky o původní úloze *matematického programování*, tj. §4.4.2§ s $b = 0$.

Důsledek §4.4.4§

Nechť množina $P \subseteq \mathbb{R}^n$ je **konvexní**, funkce f, g_1, \dots, g_m jsou **konvexní** na P , platí $F(0) > -\infty$ a **existuje** $\bar{x} \in P$ takové, že $G(\bar{x}) < 0$ (viz **Slaterova podmínka** ve Větě §4.3.2§). Potom $0 \in B$ a

1. funkce $F(\cdot)$ je **spojitá** v bodě $b = 0$
2. pro libovolné $h \in \mathbb{R}^m$ **existuje jednostranná směrová** derivace $F'_h(0) = \max_{y \in Y(0)} \sum_{i=1}^m h_i y_i$
3. funkce F je **diferencovatelná** v bodě $b = 0$ právě tehdy, když $Y(0)$ je **jednoprvková** množina, tj. $Y(0) = \{y\}$. Navíc platí $\nabla F(0) = -y$.

“ Z části 3. okamžitě plyne, že pokud **existuje více K-T vektorů** \iff funkce F **není diferencovatelná**

Fenchelova transformace a duální úloha

Lze ukázat, že pro $F(b)$ **hodnotu primární úlohy** je $F(y) = -\inf_{x \in \mathbb{R}^n} \sum_{i=1}^m g_i(x) + y^T x$ a tedy duální úlohu §4.3.1§ je možné psát jako $\max_{y \geq 0} -F(y)$

M5120 Lineární statistické modely

2. cvičení

```


$$\begin{aligned}
&\def\mc{#1}{\mathcal{#1}} \def\scal#1#2{\langle #1, #2 \rangle} \def\N{\mathbb{N}} \\
&\def\R{\mathbb{R}} \def\Q{\mathbb{Q}} \def\Z{\mathbb{Z}} \def\D{\mathbb{D}} \\
&\def\bm#1{\boldsymbol{#1}} \def\vv#1{\mathbf{#1}} \def\vp#1{\pmb{#1}} \\
&\def\floor#1{\lfloor #1 \rfloor} \def\ceil#1{\lceil #1 \rceil} \def\grad#1{\mathrm{grad} , #1} \\
&\def\ve{\varepsilon}
\end{aligned}$$


```

Lineární model

Obecně má tvar $Y_i = \beta_0 + \beta_1 x_{i,1} + \dots + \beta_k x_{i,k} + \varepsilon_i$, kde $i = 1, \dots, n$. Y_i je naše "cílová proměnná" (regresand). Proměnné $x_{i,1}, \dots, x_{i,k}$ jsou kovariáty (regresor, prediktor) a jsou pevně dané. Dále máme regresní koeficienty β_0, \dots, β_k a ε_i je náhodná proměnná chyby.

Také platí $\varepsilon_i \sim^{iid} (0, \sigma^2)$ $E(\varepsilon_i) = 0$ $var(\varepsilon_i) = \sigma^2$ $cov(\varepsilon_i) = 0$

Celkem máme $\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{1,1} & \dots & x_{1,k} \\ 1 & x_{2,1} & \dots & x_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & \dots & x_{n,k} \end{pmatrix} \cdot \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$ \tag{LSM} A vektorově $\mathbf{Y} = \underbrace{\mathbf{X}}_{\text{matice plánu}} \cdot \mathbf{\beta} + \mathbf{\varepsilon}$

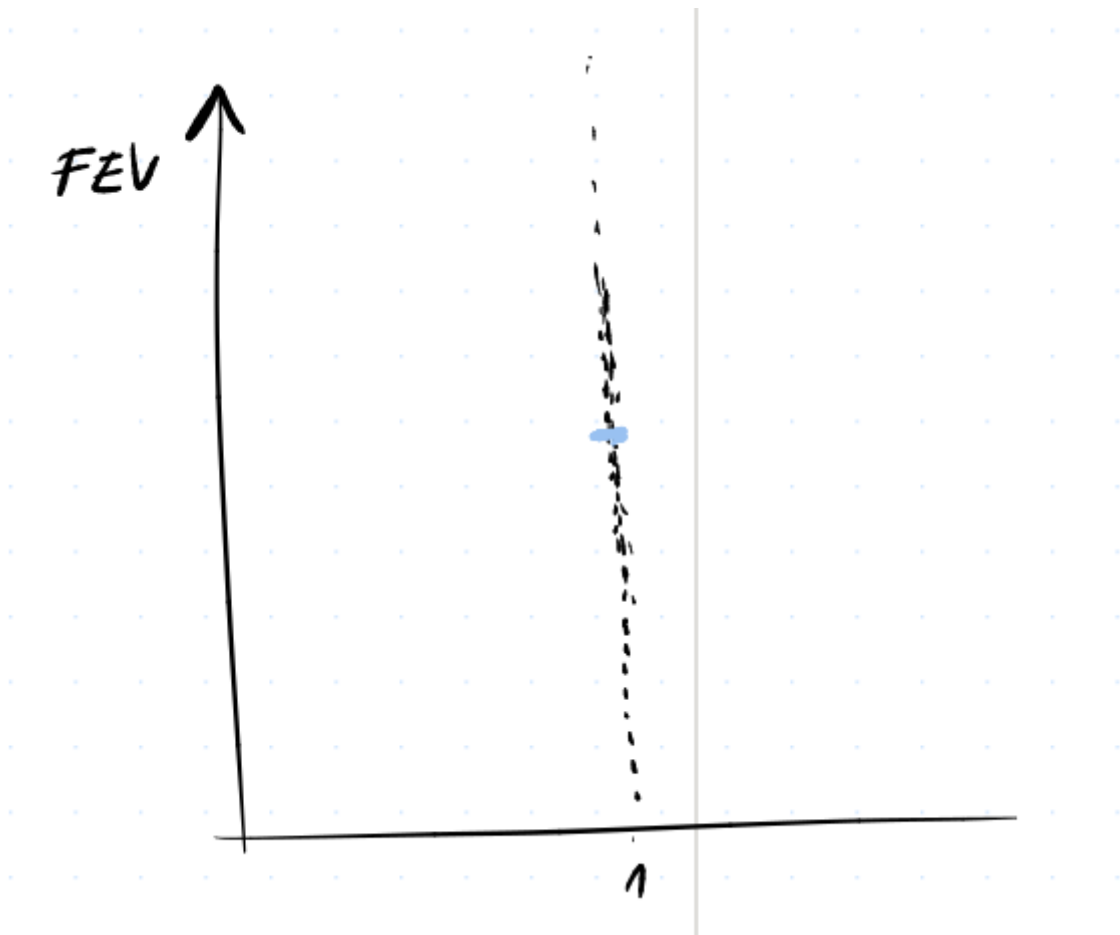
Tedy pro 2. cvičení

02 / a)

$$FEV_i = \beta_0 + \varepsilon_i$$

Můžeme si představit jako funkci $y = \beta_0$

A matice plánu bude $\begin{pmatrix} 1 & 1 & \dots & 1 \end{pmatrix}$ Tedy kapacita plic je podle tohoto modelu konstantní a vizuálně znázorněné jako

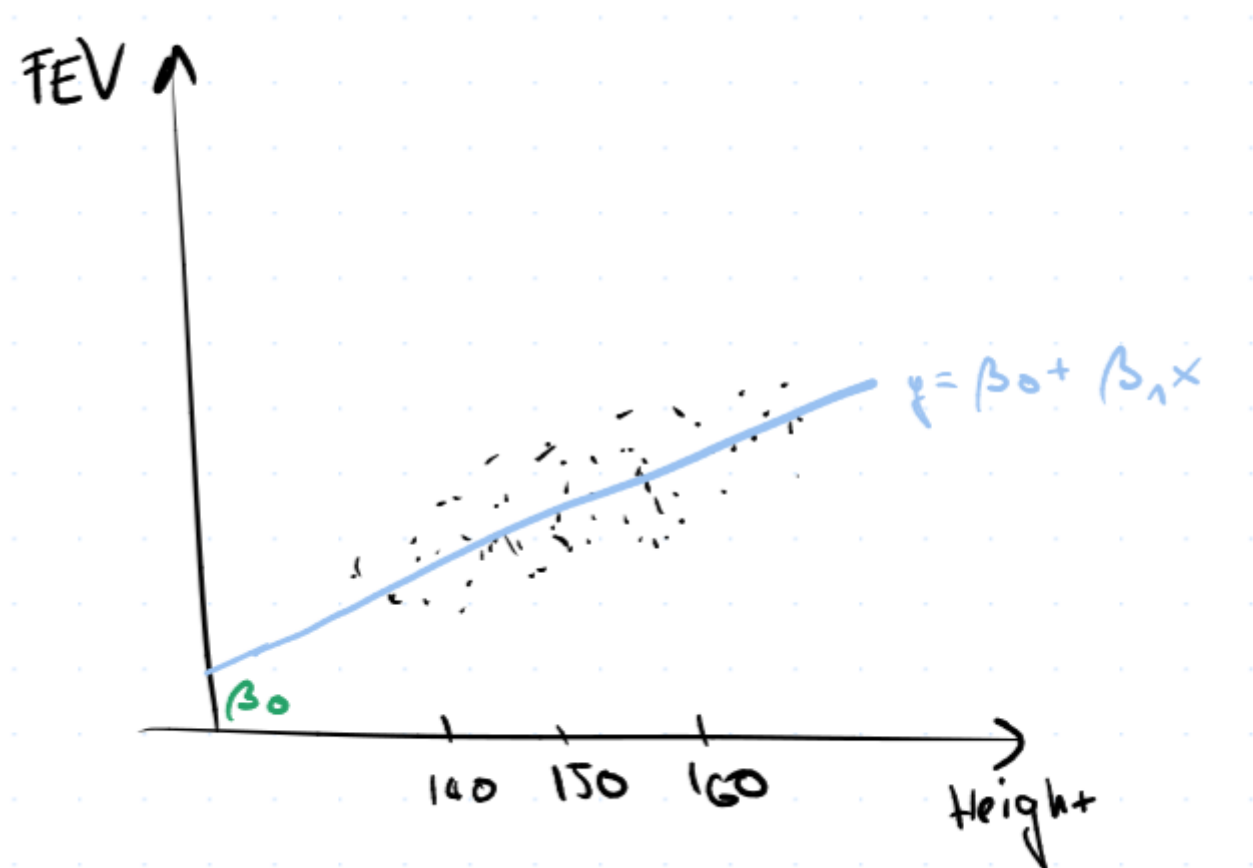


02 / c)

$$FEV_i = \beta_0 + \beta_1 \cdot \text{Height}_i + \epsilon_i$$

Kapacitu plic modelujeme pomocí výšky. Tedy v tomto případě chceme funkci $y = \beta_0 + \beta_1 x$.

A graficky



A matice plánu tentokrát bude $\begin{pmatrix} 1 & \text{Height}_1 & 1 & \text{Height}_2 & \vdots & 1 & \text{Height}_n \end{pmatrix}$, což dosazujeme do (LSM).

Tedy máme model hledáme $y = \beta_0 + \beta_1 x$ a zde

- β_0 ... střední hodnota predikce při nulových hodnotách ostatních prediktorů
- β_1 ... nárůst střední hodnoty predikce při nárůstu výšky (prediktor Height) o 1 cm

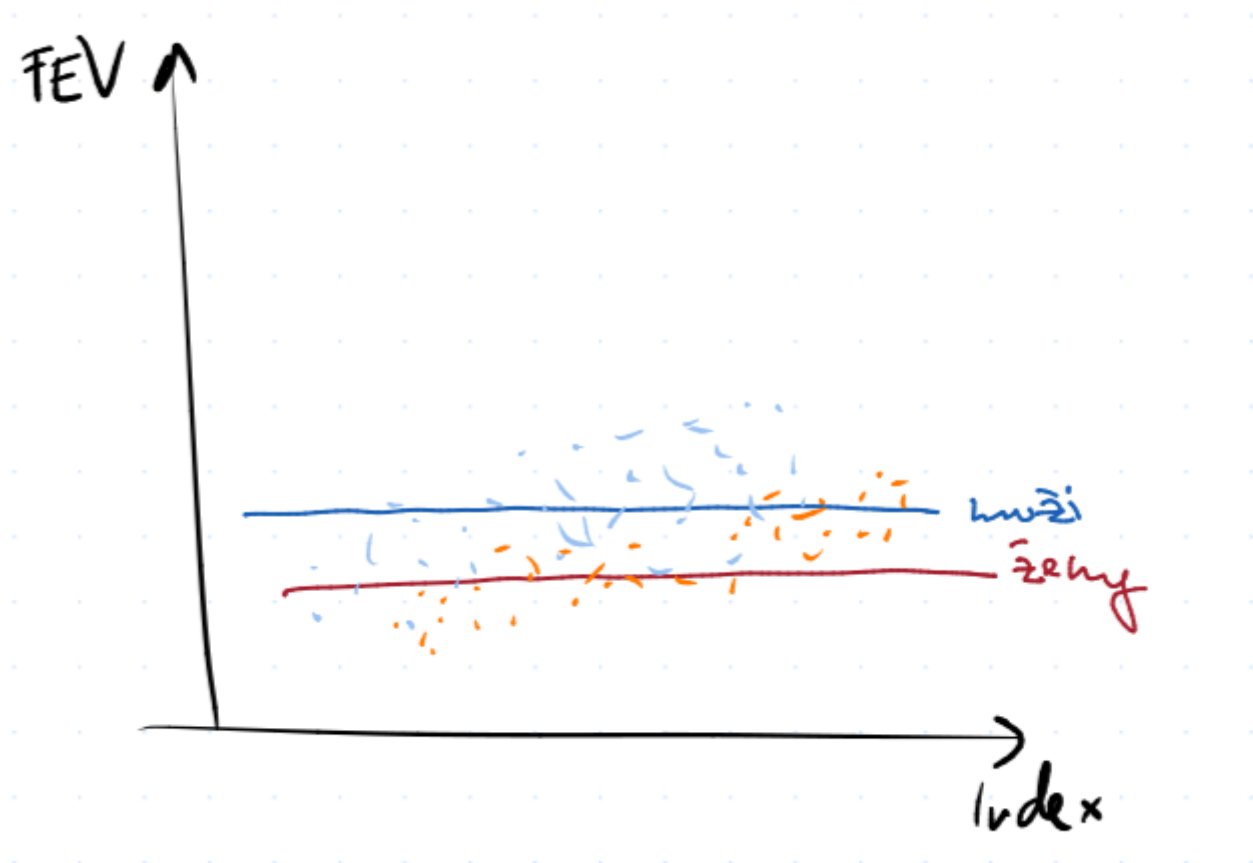
02 / b)

$$FEV_i = \beta_0 + \beta_1 \cdot \text{Sex}_i + \epsilon_i$$

a β_1 zde reprezentuje rozdíl predikce mezi muži a ženami s maticí plánu $\begin{pmatrix} 1 & 1 & 1 & 0 & \vdots & 1 & 1 \end{pmatrix}$,

kde 1 reprezentuje muže.

A graficky



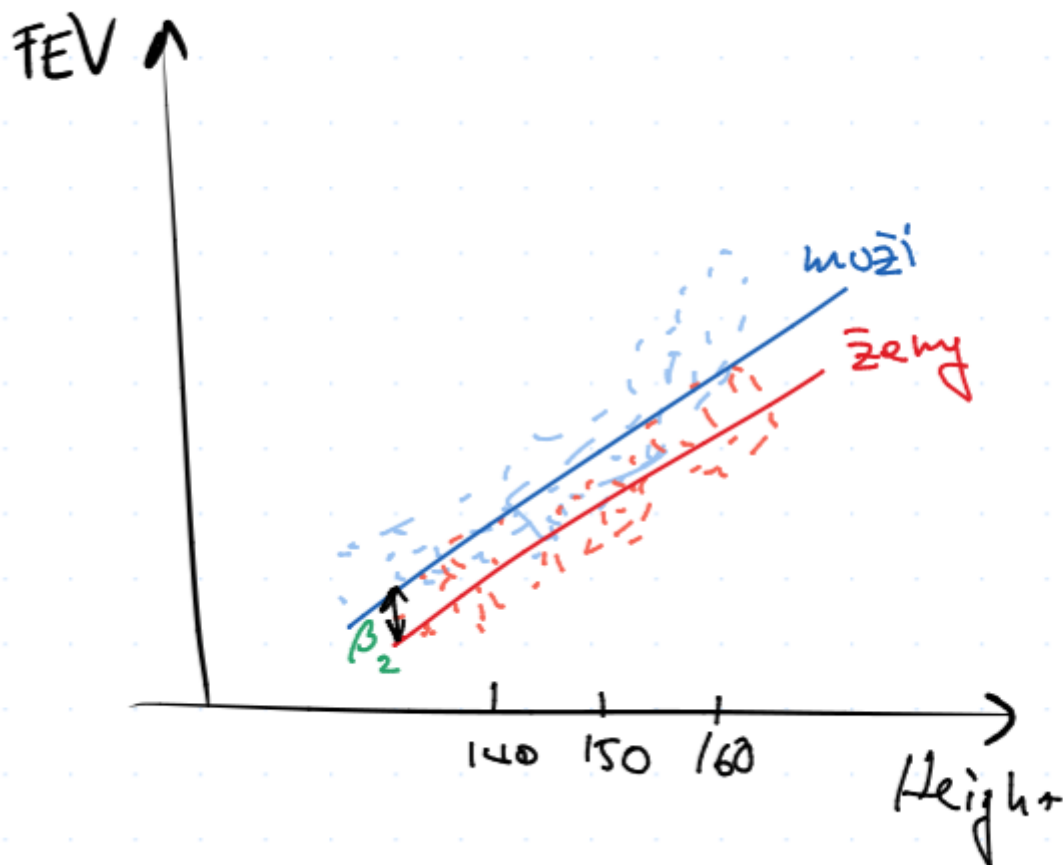
02/ d)

$FEV_i = \beta_0 + \beta_1 \cdot \text{Height}_i + \beta_2 \cdot \text{Height}_i^2 + \epsilon_i$ V tomto případě je β_1, β_2 jsou složité na interpretaci

A matice plánu by v tomto případě byla $\begin{pmatrix} 1 & \text{Height}_1 & \text{Height}_1^2 \\ 1 & \text{Height}_2 & \text{Height}_2^2 \\ \vdots & \vdots & \vdots \\ 1 & \text{Height}_n & \text{Height}_n^2 \end{pmatrix}$, ϵ_i

02 / e)

$FEV_i = \beta_0 + \beta_1 \cdot \text{Height}_i + \beta_2 \cdot \text{Sex}_i + \epsilon_i$ s maticí plánu $\begin{pmatrix} 1 & \text{Height}_1 & 0 \\ 1 & \text{Height}_2 & 1 \\ \vdots & \vdots & \vdots \\ 1 & \text{Height}_n & 1 \end{pmatrix}$, ϵ_i přičemž ve 3. sloupci jsou 1 značí muže.



A hledáme přímku $y = \beta_0 + \beta_1 x + \beta_2 \text{I} \set{\text{Sex} = \text{"male"}}$

s významem koeficientů

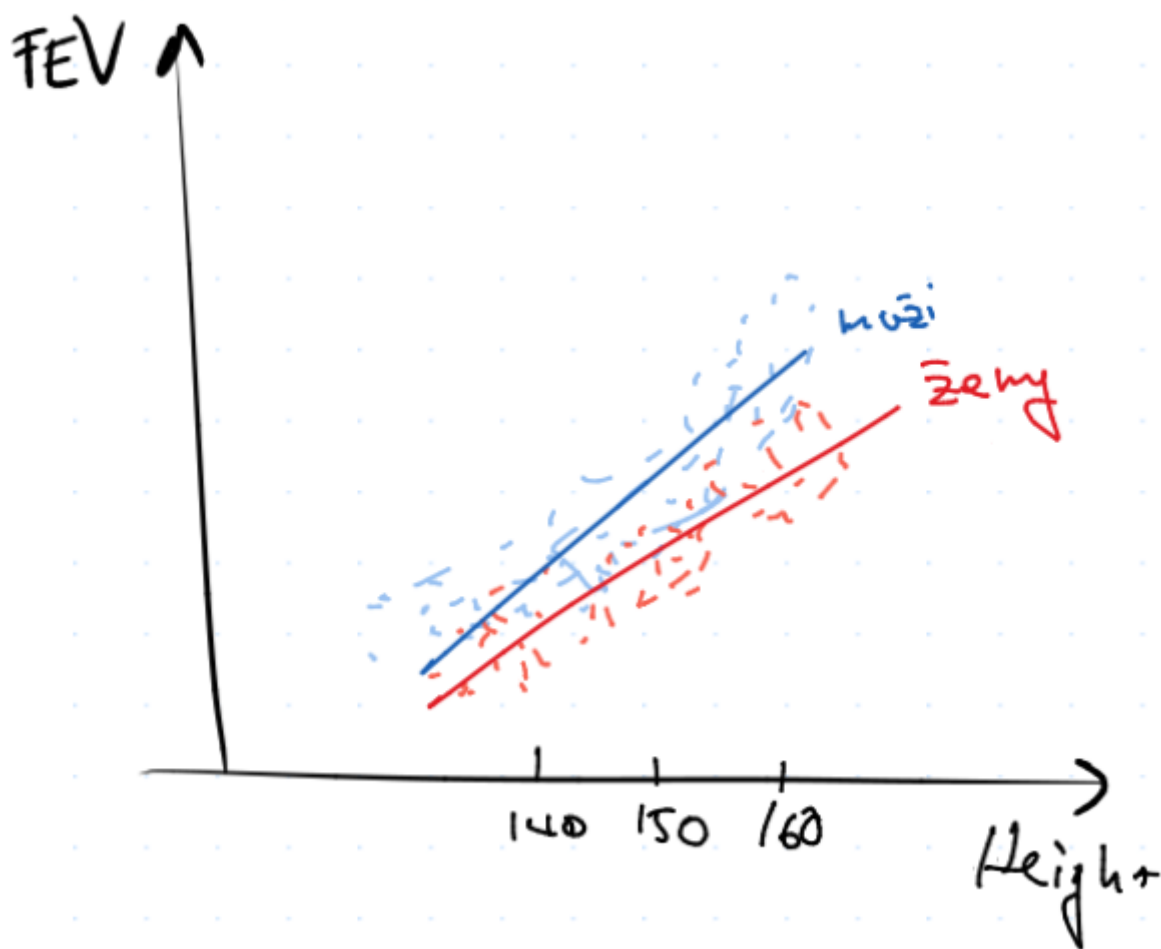
- β_0 ... střední hodnota predikce při nulových hodnotách ostatních prediktorů (nulová výška a ženské pohlaví)
- β_1 ... změna střední hodnoty predikce při nárůstu výšky (prediktor Height) o 1 cm pro ženy
- β_2 ... rozdíl střední hodnoty predikce mezi muži a ženami

02 / g)

$$\text{FEV}_i = \beta_0 + \beta_1 \cdot \text{Height}_i + \beta_2 \text{Sex}_i + \beta_3 (\text{Sex}_i \times \text{Height}_i) + \epsilon_i$$
kde členu $\text{Sex}_i \times \text{Height}_i$ **interakce** a matice plánu bude
$$\begin{pmatrix} 1 & \text{Height}_1 & 0 & 0 \\ 1 & \text{Height}_2 & 1 & \text{Height}_2 \\ 1 & \text{Height}_3 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \text{Height}_n & 1 & \text{Height}_n \end{pmatrix}$$
, přičemž ve 3. sloupci jsou 1 značí muže.

A hledáme přímku $y = \beta_0 + \beta_1 x + \beta_2 \text{I} \set{\text{Sex} = \text{"male"}} + \beta_3 x \text{I} \set{\text{Sex} = \text{"male"}}$, tj.

- žena ... $y = \beta_0 + \beta_1 x$
- muž ... $y = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)x$



Zde β_3 značí rozdíl střední hodnoty predikce mezi muži a ženami při nárůstu výšky o 1 cm

“ rozdíl rychlosti růstu FEV mezi muži a ženami

Interpretation

- “keeping the values of all the other covariates fixed, a unit increase in x_i is associated with a $\hat{\beta}_i$ increase in $E Y$ ”
 - ▶ suitably adapted for categorical predictors and potentially interactions, and depends on the choice of the identifiability conditions
 - ▶ polynomials need a more complex interpretation
- is it meaningful to imagine that a covariate changes while all the other remain fixed?

3. cvičení

$\def\mc{#1}{\mathcal{#1}} \def\sc{#1#2}{\langle #1, #2 \rangle} \def\N{\mathbb N}$
 $\def\R{\mathbb R} \def\Q{\mathbb Q} \def\Z{\mathbb Z} \def\D{\mathbb D}$
 $\def\bm{#1}{\boldsymbol{#1}} \def\vv{#1}{\mathbf{#1}} \def\vp{#1}{\mathbf{#1}}$
 $\def\floor{#1}{\lfloor #1 \rfloor} \def\ceil{#1}{\lceil #1 \rceil} \def\grad{#1}{\mathrm{grad} , #1}$
 $\def\ve{\varepsilon}$

Příklad 1

Máme symetrickou, pozitivně definitní matici $m \times m$ značenou Σ .

a)

Poz. def \iff všechna vlastní čísla jsou kladná $\implies \det(\lambda_1 \dots \lambda_m) > 0$
 \implies inverze bude existovat $\implies h(\Sigma) = m$

b)

Matice Σ je *samoadjungovaný operátor* Inverzi sestojíme pomocí spektrálního rozkladu. $\Sigma = U \Lambda U^T$, kde U je tvořená vlastními vektory a je ortogonální ($U \cdot U^T = I$) a Λ je diagonální matice vlastních čísel.

Pak inverze je $\Sigma^{-1} = U \Lambda^{-1} U^T$

A jako ověření $\Sigma \Sigma^{-1} = U \Lambda U^T U \Lambda^{-1} U^T = I$

c)

Ze spektrálního rozkladu je jisté matice Σ^{-1} pozitivně definitní. Nechť $\forall x \neq 0$ libovolné, pak $\forall x^T \Sigma^{-1} x = \forall x^T U \Lambda^{-1} U^T x = \forall x^T U$

$\Lambda^{-\frac{1}{2}} \Lambda^{-\frac{1}{2}} U^T \mathbf{x}$, kde $\Lambda^{-\frac{1}{2}}$ je matice s převrácenými hodnotami odmocnin vlastních čísel matice Σ na diagonále a tedy $\underbrace{\mathbf{x}^T U \Lambda^{-\frac{1}{2}}}_{\mathbf{y}} \underbrace{\Lambda^{-\frac{1}{2}} U^T \mathbf{x}}_{\mathbf{y}^T} = \|\mathbf{y}\|^2 > 0$

d)

Mějme množinu S_c takovou, že $S_c = \{\mathbf{x}; (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) = c\}$, kde $\boldsymbol{\mu} \in \mathbb{R}^m$ a $c \in \mathbb{R}$.

Pro $c < 0$ je $S_c \equiv \emptyset$.

Dále pro $c = 0$ je řešením pouze $S_c = \{\boldsymbol{\mu}\}$.

Nakonec pro $c > 0$ je $\underbrace{(\mathbf{x} - \boldsymbol{\mu})^T U}_{\mathbf{y}} \Lambda^{-1} U^T (\mathbf{x} - \boldsymbol{\mu}) = \mathbf{y}^T \Lambda^{-1} \mathbf{y}$. A pro $m = 2$ tedy $\begin{pmatrix} y_1 & y_2 \end{pmatrix} \begin{pmatrix} \frac{1}{\lambda_1} & 0 \\ 0 & \frac{1}{\lambda_2} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = c$, což je **rovnice elipsy** se středem (μ_1, μ_2) a směry os jsou právě vlastní vektory Σ . Nakonec délky poloos budou $\sqrt{c \lambda_i}$



Analogicky pro $m > 2$ dostaneme **elipsoid**.

Příklad 4

Matice Σ je poz. **semidef** matice symetrická matice $m \times m$.

a)

$h(\Sigma) = r$, kde $0 < r \leq m$.

Nahradíme U za matici $U_1 = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_r \end{pmatrix}$ a také $\Lambda_1 = \text{diag}(\lambda_1, \dots, \lambda_r)$

b)

Sestrojíme matici $\tilde{\Sigma}$ takovou, že $\tilde{\Sigma} \tilde{\Sigma}^T = \Sigma$ jako $\tilde{\Sigma} = U_1 \Lambda_1^{-\frac{1}{2}}$, když je $h(\Sigma) = r$ a pro $h(\Sigma) = m$ jako $\tilde{\Sigma} = U \Lambda^{-\frac{1}{2}}$.

d)

“ Zde značím pseudoinverzi matice A jako A^\dagger

$\tilde{\Sigma}^\dagger = \Lambda_1^{-\frac{1}{2}} U_1^T$ a pro tuto matici bychom ukázali všechny 4 vlastnosti pseudoinverze, tj.

- $A A^\dagger A = A$
- $A^\dagger A A^\dagger = A^\dagger$
- $(A^\dagger A)^T = A^\dagger A$
- $(A A^\dagger)^T = A A^\dagger$

4. cvičení

$\$ \backslash \mathrm{mcal} \# 1 \{ \backslash \mathrm{mathcal} \{ \# 1 \} \}$ $\backslash \mathrm{def} \backslash \mathrm{scal} \# 1 \# 2 \{ \backslash \mathrm{angle} \# 1, \# 2 \backslash \mathrm{rangle} \}$ $\backslash \mathrm{def} \backslash \mathrm{N} \{ \backslash \mathrm{mathbb} \mathrm{N} \}$
 $\backslash \mathrm{def} \backslash \mathrm{R} \{ \backslash \mathrm{mathbb} \mathrm{R} \}$ $\backslash \mathrm{def} \backslash \mathrm{Q} \{ \backslash \mathrm{mathbb} \{ \mathrm{Q} \} \}$ $\backslash \mathrm{def} \backslash \mathrm{Z} \{ \backslash \mathrm{mathbb} \{ \mathrm{Z} \} \}$ $\backslash \mathrm{def} \backslash \mathrm{D} \{ \backslash \mathrm{mathbb} \{ \mathrm{D} \} \}$
 $\backslash \mathrm{def} \backslash \mathrm{bm} \# 1 \{ \backslash \mathrm{boldsymbol} \{ \# 1 \} \}$ $\backslash \mathrm{def} \backslash \mathrm{vv} \# 1 \{ \backslash \mathrm{mathbf} \{ \# 1 \} \}$ $\backslash \mathrm{def} \backslash \mathrm{vvp} \# 1 \{ \backslash \mathrm{pmb} \{ \# 1 \} \}$
 $\backslash \mathrm{def} \backslash \mathrm{floor} \# 1 \{ \backslash \mathrm{floor} \# 1 \backslash \mathrm{rfloor} \}$ $\backslash \mathrm{def} \backslash \mathrm{ceil} \# 1 \{ \backslash \mathrm{ceil} \# 1 \backslash \mathrm{rceil} \}$ $\backslash \mathrm{def} \backslash \mathrm{grad} \# 1 \{ \mathrm{mathrm} \{ \mathrm{grad} \} , \# 1 \}$
 $\backslash \mathrm{def} \backslash \mathrm{ve} \{ \backslash \mathrm{varepsilon} \}$ $\backslash \mathrm{def} \backslash \mathrm{im} \# 1 \{ \mathrm{mathrm} \{ \mathrm{im} \} (\# 1) \}$ $\backslash \mathrm{def} \backslash \mathrm{tr} \# 1 \{ \mathrm{mathrm} \{ \mathrm{tr} \} (\# 1) \}$
 $\backslash \mathrm{def} \backslash \mathrm{norm} \# 1 \{ \backslash \mathrm{left} \backslash \mathrm{vert} \backslash \mathrm{left} \backslash \mathrm{vert} \# 1 \backslash \mathrm{right} \backslash \mathrm{vert} \backslash \mathrm{right} \backslash \mathrm{vert} \}$ $\backslash \mathrm{def} \backslash \mathrm{scal} \# 1 \# 2 \{ \backslash \mathrm{angle} \# 1, \# 2 \backslash \mathrm{rangle} \}$
 $\$ \$$

1. příklad

Máme symetrickou idempotentní matice P velikosti $n \times n$, což je matice **ortogonální projekce**.

a)

Jelikož je P symetrická, tak existuje spektrální rozklad tvaru $P = U \Lambda U^T$

“ matice U je ortogonální a plné hodnosti

a jelikož je idempotentní $P = PP$ Potom $U \Lambda U^T = P = PP = U \Lambda \underbrace{U^T U}_I \Lambda U^T = U \Lambda^2 U^T$ A celkem dostáváme $\Lambda^2 = \Lambda$. Navíc jelikož je Λ diagonální, tak pro všechna vlastní čísla λ matice Λ platí $\lambda^2 = \lambda \implies \lambda(\lambda - 1) = 0 \implies \lambda = 0 \vee \lambda = 1$

“ Idempotentní matice P je invertibilní právě tehdy, když $P = I$

b)

Jelikož $P = U\Lambda U^T$, pak U je plně hodnostní (je ortogonální), pak $\text{h}(P) = \text{h}(U\Lambda U^T) = \text{h}(\Lambda) = |\{\lambda_i, \lambda_i = 1\}|$

c)

Z části **b)** máme $\text{h}(P) = \sum \lambda_i = \text{tr } \Lambda = \text{tr } \{\Lambda U U^T\}$
A jelikož pro stopu součinu matic platí $\text{tr } \{ABC\} = \text{tr } \{CAB\}$ (invariantnost vůči cyklickým operacím), pak $\text{h}(P) = \text{tr } \{U\Lambda U^T\} = \text{tr } P$

d)

Je-li vektor $y \in \text{im } P$, pak $Py = y$. Pokud $y \in \text{im } P$, pak $\exists z$, že $y = Pz$

Definice obrazu $\text{im } P$: $\text{im } P = \{y \in \mathbb{R}^n \mid \exists z \in \mathbb{R}^n : y = Pz\}$

Pak $Py = P(Pz) = (P^2)z = Pz = y$

e)

Chceme ukázat, že projekce do menšího podprostoru je zároveň projekce do onoho většího prostoru

Nechť $z \in \mathbb{R}^n$ je libovolné, projekce z do menšího prostoru je prvek většího prostoru. $\tilde{P}z \in \text{im } \tilde{P} \leq \text{im } P$. Pak podle **d)** platí $P(\tilde{P}z) = \tilde{P}z$.
 $P(\tilde{P}z) - \tilde{P}z = 0 \in (P - \tilde{P})z = 0$, nicméně vektor z byl libovolný. To tedy znamená, že zobrazení $P - \tilde{P}$ pošle všechny $z \in \mathbb{R}^n$ na nulový vektor, tj. $\ker(P - \tilde{P}) = \mathbb{R}^n$. Z lineární algebry víme, že $\dim \ker(A) + \dim \text{im } A = n$ pro A tvaru $n \times n$. A tedy $\dim \{P - \tilde{P}\} = 0 \implies P - \tilde{P} = 0$. Neboť P, \tilde{P} jsou ortogonální projekce, tak jsou symetrické.

Celkem $\|P\tilde{P} = \tilde{P}^T P^T = (\tilde{P} P)^T = \tilde{P}^T = \tilde{P} = P\tilde{P}$

f)

Nechť $x \in \mathbb{R}^n$ pevné. Mějme $Pz \in \text{im } P$. Vezměme libovolné $y \in \text{im } P$ a spočítáme $\|y - x\|^2$. Pak $\|y - x\|^2 = \|Pz - x + y - Pz\|^2 = \|(Pz - x) + (y - Pz)\|^2 = \langle (Pz - x) + (y - Pz), (Pz - x) + (y - Pz) \rangle$

Pro skalární součin platí $\langle u, v \rangle = \langle v, u \rangle$, $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$, $\langle au, v \rangle = a \langle u, v \rangle$, ; $a \in \mathbb{R}$

Pak dostáváme $\|Pz - x\|^2 + 2 \langle Pz - x, y - Pz \rangle + \|y - Pz\|^2$

A zajímá nás hlavně $\langle Pz - x, y - Pz \rangle$, tedy

$$\langle Pz - x, y - Pz \rangle = (y - Pz)^T (Pz - x) = y^T (Pz - x) - (Pz)^T (Pz - x) = y^T Pz - y^T x - (Pz)^T Pz + (Pz)^T x = y^T Pz - y^T x - \underbrace{(Pz)^T Pz}_{\langle Pz, Pz \rangle} + \underbrace{(Pz)^T x}_{\langle Pz, x \rangle} = y^T Pz - y^T x - \langle Pz, Pz \rangle + \langle Pz, x \rangle$$

Jelikož $y \in \text{im } P$, tak jistě $\exists z \in \mathbb{R}^n$ takové, že $y = Pz$. Z toho plyne

$$y^T Pz - y^T x - \langle Pz, Pz \rangle + \langle Pz, x \rangle = (Pz)^T Pz - (Pz)^T x - \langle Pz, Pz \rangle + \langle Pz, x \rangle = \underbrace{(Pz)^T Pz}_{\langle Pz, Pz \rangle} - \underbrace{(Pz)^T x}_{\langle Pz, x \rangle} - \langle Pz, Pz \rangle + \langle Pz, x \rangle = 0$$

Celkem $\|y - x\|^2$ závisí pouze na $\|y - Pz\|^2$ a $\|Pz - x\|^2 + \underbrace{2 \langle Pz - x, y - Pz \rangle}_{=0} + \|y - Pz\|^2 \geq \|Pz - x\|^2$ a rovnost nastane pouze v případě $Pz = y$.

Zadání v R

Vykreslit si grafy hustot a distribučních funkcí pro

- $N(0,1)$ a spočítat $P(X \leq 2)$
- Studentovo $t(df = 5)$ a spočítat $P(X \leq 2)$
- $\chi^2(df = 10)$ a spočítat $P(X \leq 20)$
- Fisherovo $F(5, 10)$ a spočítat $P(X \leq 2)$

a vypočtete 95% kvantil a zaznačte do grafu

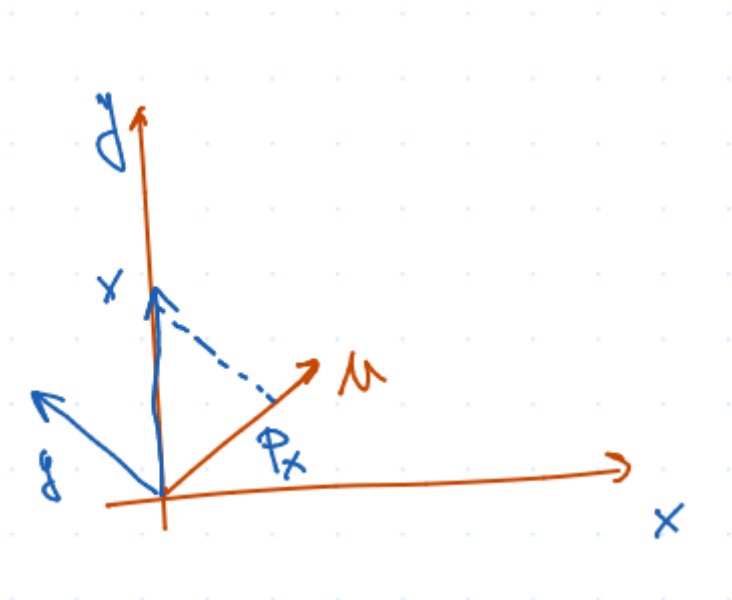
5. cvičení

$\backslash\mathrm{mcal}\{#1\}$ $\backslash\mathrm{scal}\{#1\}\{#2\}$ $\backslash\mathrm{N}\{\mathbb{N}\}$
 $\backslash\mathrm{R}\{\mathbb{R}\}$ $\backslash\mathrm{Q}\{\mathbb{Q}\}$ $\backslash\mathrm{Z}\{\mathbb{Z}\}$ $\backslash\mathrm{D}\{\mathbb{D}\}$
 $\backslash\mathrm{bm}\{#1\}$ $\backslash\mathrm{vv}\{#1\}$ $\backslash\mathrm{vvp}\{#1\}$
 $\backslash\mathrm{floor}\{#1\}$ $\backslash\mathrm{ceil}\{#1\}$ $\backslash\mathrm{grad}\{#1\}$ $\backslash\mathrm{ve}\{\mathrm{varepsilon}\}$ $\backslash\mathrm{im}\{#1\}$ $\backslash\mathrm{tr}\{#1\}$
 $\backslash\mathrm{norm}\{#1\}$ $\backslash\mathrm{ex}\{#1\}$ $\backslash\mathrm{exv}\{#1\}$

4. cvičení / 2. příklad

a) i b)

Mějme vektor $\mathbf{u} \in \mathbb{R}^n$ (takový, že $\|\mathbf{u}\| = 1$), pak matice $P = \frac{\mathbf{u} \mathbf{u}^T}{\|\mathbf{u}\|^2}$ je ortogonální projekce na $\mathrm{im}\{\mathbf{u}\}$.



Pro ortogonální projekci platí

1. $P^2 = P$ - idempotence projekce
2. P je symetrická (z ortogonality)

Pak $P = \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\|\mathbf{u}\|^2}$ a tedy $PP = \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\|\mathbf{u}\|^2} \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\|\mathbf{u}\|^2}$

Z definice skalárního součinu $\langle \mathbf{v}, \mathbf{u} \rangle = \|\mathbf{u}\|^2$ a proto $PP = \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\|\mathbf{u}\|^2} = P$, což zvláště platí pro $\|\mathbf{u}\| = 1$. Pro nějaké $\mathbf{x} \in \mathbb{R}^n$ máme $P \mathbf{x} = \frac{\langle \mathbf{v}, \mathbf{u} \rangle \langle \mathbf{v}, \mathbf{x} \rangle}{\|\mathbf{u}\|^2} = \frac{\langle \mathbf{v}, \mathbf{u} \rangle \langle \mathbf{v}, \mathbf{x} \rangle}{\|\mathbf{u}\|^2} = \underbrace{\frac{\langle \mathbf{v}, \mathbf{u} \rangle \langle \mathbf{v}, \mathbf{x} \rangle}{\|\mathbf{u}\|^2}}_{\in \mathbb{R}} \mathbf{u} \in \text{im } U$

c)

Mějme $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$ ortonormální vektory, pak matice $P = U U^T$, kde $U = (\mathbf{u}_1 \mid \mathbf{u}_2 \mid \dots \mid \mathbf{u}_p)$, je ortogonální projekce na $\text{im } U$.

Ukažme $P \cdot P = U \underbrace{U^T U}_I U^T = U U^T$ a $P \mathbf{x} = U U^T \mathbf{x} = U \begin{pmatrix} \langle \mathbf{u}_1, \mathbf{x} \rangle \\ \langle \mathbf{u}_2, \mathbf{x} \rangle \\ \vdots \\ \langle \mathbf{u}_p, \mathbf{x} \rangle \end{pmatrix} = \underbrace{\mathbf{u}_1 \langle \mathbf{u}_1, \mathbf{x} \rangle}_{\in \mathbb{R}} + \dots + \underbrace{\mathbf{u}_p \langle \mathbf{u}_p, \mathbf{x} \rangle}_{\in \mathbb{R}} \in \text{im } U$, což je lineární kombinace vektorů $\mathbf{u}_1, \dots, \mathbf{u}_p$ a jistě tedy $P \mathbf{x} \in \text{im } U$.

d) i e)

Máme lineárně nezávislé vektory $\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$, pak $P = A(A^T A)^{-1} A^T$ je ortogonální projekce.

Jednou možností by bylo použít spektrální rozklad $A = U \Sigma V$, čehož bychom dostali $P = U U^T$, což jsme ukázali v bodě c).

Druhá možnost je $PP = A(A^T A)^{-1} \underbrace{A^T A(A^T A)^{-1}}_I A^T = P$, což stejně fungovalo i pro pseudoinverzi. Dále $A \mathbf{x} = \mathbf{y} \in \text{im } A$, pak $P \mathbf{y} = (A(A^T A)^{-1} A^T) \mathbf{y} = A \underbrace{(A^T A)^{-1} A^T A}_I \mathbf{x} = A \mathbf{x} = \mathbf{y}$

„ Zde jsme jen ukázali něco o $\mathbf{y} \in \text{im } A$, nikoliv o obecném $\mathbf{x} \in \mathbb{R}^p$

5. cvičení / 1. příklad

Nechť $\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$ a $DX = \text{Var } \mathbf{X} = \text{Cov}(\mathbf{X}, \mathbf{X})$ platí $\text{Cov}(\mathbf{X}, \mathbf{X}) = \begin{pmatrix} \text{Cov}(X_1, X_1) & \text{Cov}(X_1, X_2) & \dots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & \text{Cov}(X_2, X_2) & \dots & \text{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \dots & \text{Cov}(X_n, X_n) \end{pmatrix}$

a)

Ukažme $\mathbf{A} \mathbf{X} + \mathbf{b} = \mathbf{A} \mathbf{X} + \mathbf{b}$, což je analogické k jednorozměrnému případu.

Pro i -tý prvek platí $\sum_{k=1}^n a_{i,k} X_k + b_i = \sum_{k=1}^n X_k + b_i$, což je, co jsme potřebovali.

b)

Ukažme $\text{Var}(\mathbf{A} \mathbf{X} + \mathbf{b}) = \mathbf{A} \text{Var } \mathbf{X} \mathbf{A}^T$ což je opět analogie k $D(aX + b) = a^2 D X$.

Využijeme vlastnost kovariance. Tedy pro (i,j) -tý prvek matice $\mathbf{A} \mathbf{X} + \mathbf{b}$ platí $\text{Cov} \left(\sum_{k=1}^n a_{i,k} X_k + b_i, \sum_{l=1}^n a_{j,l} X_l + b_j \right) = \sum_{k=1}^n \sum_{l=1}^n a_{i,k} a_{j,l} \text{Cov}(X_k, X_l) = \sum_{l=1}^n \left(\sum_{k=1}^n a_{i,k} \text{Cov}(X_k, X_l) \right) a_{j,l}$

c)

Máme ukázat $\mathbf{X}^T \mathbf{X} = \mathbf{X}^T \mathbf{X} + \text{tr}(\text{Var}(\mathbf{X}))$, což je ekvivalentní s $X_1^2 + X_2^2 + \dots + X_n^2 = X_1^2 + \dots + X_n^2 + DX_1 + \dots + DX_n$

Obecně platí $\text{Var}(\mathbf{X}) = \mathbf{X} \mathbf{X}^T - \mathbf{X} \mathbf{X}^T$ a tedy $\mathbf{X}^T \mathbf{X} = \mathbf{X}^T \mathbf{X} + \text{tr}(\mathbf{X} \mathbf{X}^T) = \mathbf{X}^T \mathbf{X} + \text{tr}(\text{Var}(\mathbf{X}) + \mathbf{X} \mathbf{X}^T) = \mathbf{X}^T \mathbf{X} + \text{tr}(\text{Var}(\mathbf{X})) + \text{tr}(\mathbf{X} \mathbf{X}^T) = \mathbf{X}^T \mathbf{X} + \text{tr}(\text{Var}(\mathbf{X}))$

7. cvičení

$\backslash\mathrm{cal}\{#1\}$ $\backslash\mathrm{scal}\{#1\}$ $\backslash\mathrm{angle}\{#1, #2\}$ $\backslash\mathrm{N}\{\}$ $\backslash\mathrm{R}\{\}$ $\backslash\mathrm{Q}\{\}$ $\backslash\mathrm{Z}\{\}$ $\backslash\mathrm{D}\{\}$
 $\backslash\mathrm{bm}\{#1\}$ $\backslash\mathrm{vv}\{#1\}$ $\backslash\mathrm{vvp}\{#1\}$
 $\backslash\mathrm{floor}\{#1\}$ $\backslash\mathrm{ceil}\{#1\}$ $\backslash\mathrm{grad}\{#1\}$ $\backslash\mathrm{ve}\{\}$ $\backslash\mathrm{im}\{#1\}$ $\backslash\mathrm{tr}\{#1\}$
 $\backslash\mathrm{norm}\{#1\}$ $\backslash\mathrm{scal}\{#1\}$ $\backslash\mathrm{ex}\{#1\}$ $\backslash\mathrm{exv}\{#1\}$

Nechť Y jsou data, $\hat{Y} = X \hat{\beta}$, $E \hat{Y} = E Y$ je odhad Y a e je odhad $\hat{Y} - Y$.

A máme **celkovou sumu čtverců** $TSS = \sum_{i=1}^n (Y_i - \overline{Y})^2$ také **vysvětlovanou sumu čtverců** $ESS = \sum_{i=1}^n (\hat{Y}_i - \overline{Y})^2$ a neposlední řadě **reziduální sumu čtverců** $RSS = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$

A platí $TSS = RSS + ESS$

a necht' R^2 je **koeficient determinace** $R^2 = \frac{ESS}{TSS} \in (0, 1]$ a **adjustovaný koeficient determinace** $R^2_{adj} = 1 - \frac{\frac{RSS}{n-p}}{\frac{TSS}{n-1}}$

Dále $\hat{\sigma}^2 = \frac{RSS}{n-p}$ a $\mathrm{var}(\hat{\beta}) = \hat{\sigma}^2 (X^T X)^{-1}$ Přičemž $\mathrm{var}(\hat{\beta})$ dostaneme pomocí `vcov(<model>)`

9. cvičení

a) IS pro β_i : $T_i = \frac{\hat{\beta}_i}{\sqrt{\hat{\sigma}^2 (\mathbf{X}^T \mathbf{X})^{-1}_{i,i}}} \sim t(n-p)$

Pak $P\left(T_i \in \left[t_{\frac{\alpha}{2}(n-p)}, t_{1-\frac{\alpha}{2}(n-p)}\right]\right) = 1 - \alpha$
 $t_{\frac{\alpha}{2}(n-p)} \leq T_i \leq t_{1-\frac{\alpha}{2}(n-p)}$
 $\frac{\hat{\beta}_i - \hat{\beta}_i}{\sqrt{\hat{\sigma}^2 (\mathbf{X}^T \mathbf{X})^{-1}_{i,i}}} \leq t_{1-\frac{\alpha}{2}(n-p)}$
 Z čehož dostaneme $\hat{\beta}_i \in \left(\hat{\beta}_i \pm t_{1-\frac{\alpha}{2}(n-p)} \sqrt{\hat{\sigma}^2 (\mathbf{X}^T \mathbf{X})^{-1}_{i,i}}\right)$

b) $T = \frac{\mathbf{a}^T \hat{\beta}}{\sqrt{\hat{\sigma}^2 \mathbf{a}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{a}}} \sim t(n-p)$

$\left(\mathbf{a}^T \hat{\beta} \pm t_{1-\frac{\alpha}{2}(n-p)} \sqrt{\hat{\sigma}^2 \mathbf{a}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{a}}\right)$

A regresní přímka pro dívky bude mít tvar $y = \hat{\beta}_0 + \hat{\beta}_1 x$ Pro chlapce: $y = \hat{\beta}_0 + \hat{\beta}_2 + (\hat{\beta}_1 + \hat{\beta}_3) x$ IS pro $\hat{\beta}_1 + \hat{\beta}_3$, tedy $\mathbf{a} = (0, 1, 0, 1)$

d) Při počítání predikčního intervalu zohledňujeme chybu u "nového pozorování". Tedy odhad rozptylu je $\hat{\sigma}^2 \mathbf{x}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x} + \hat{\sigma}^2$
 $\implies T = \frac{\mathbf{x}^T \hat{\beta}}{\sqrt{\hat{\sigma}^2 (1 + \mathbf{x}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x})}} \sim t(n-p)$

```
# Confidence interval
predict(..., interval = "confidence")

# Or
predict(..., interval = "prediction")
```

10. cvičení

$\backslash\mathrm{cal}\{#1\}$ $\backslash\mathrm{scal}\{#1\}$ $\backslash\mathrm{angle}\{#1, #2\}$ $\backslash\mathrm{N}\{\mathbb{N}\}$
 $\backslash\mathrm{R}\{\mathbb{R}\}$ $\backslash\mathrm{Q}\{\mathbb{Q}\}$ $\backslash\mathrm{Z}\{\mathbb{Z}\}$ $\backslash\mathrm{D}\{\mathbb{D}\}$
 $\backslash\mathrm{bm}\{#1\}$ $\backslash\mathrm{v}\{#1\}$ $\backslash\mathrm{vp}\{#1\}$
 $\backslash\mathrm{floor}\{#1\}$ $\backslash\mathrm{ceil}\{#1\}$ $\backslash\mathrm{grad}\{#1\}$ $\backslash\mathrm{ve}\{\}$ $\backslash\mathrm{im}\{#1\}$ $\backslash\mathrm{tr}\{#1\}$
 $\backslash\mathrm{norm}\{#1\}$ $\backslash\mathrm{scal}\{#1\}$ $\backslash\mathrm{ex}\{#1\}$ $\backslash\mathrm{exv}\{#1\}$
 $\backslash\mathrm{mtr}\{#1\}$

Scheffeho věta $P\left(\left|\sum_{i=1}^m \mathbf{b}_i^T (\hat{\beta} - \beta)\right|^2 \leq m F_{1-\alpha}(m, n-p)\right) = 1 - \alpha$ $\forall \mathbf{b} \in \mathbb{R}^m$, je-li matice A typu $m \times p$ plně hodnosti.

Příklad $Y_i = \beta_0 + \beta_1 \text{Height}_i + \beta_2 \text{Sex}_i + \beta_3 \text{Height}_i + \beta_4 \text{Sex}_i + \epsilon_i$, $\epsilon_i \sim N(0, \sigma^2)$ a chceme zkonstruovat 95% PS pro chlapce a dívky

1) Napíšeme tvar reg. křivky

- d: $y = \hat{\beta}_0 + \hat{\beta}_1 x$
- ch: $y = \hat{\beta}_0 + \hat{\beta}_2 + (\hat{\beta}_1 + \hat{\beta}_3)x$

2) Zvolíme vhodný tvar \mathbf{b} a A :

- d: $\mathbf{b} = \begin{pmatrix} 1 \\ x \end{pmatrix} \in \mathbb{R}^2$, pak $\overbrace{\begin{pmatrix} 1 & x \end{pmatrix}}^A \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \end{pmatrix}$
- ch: $\mathbf{b} = \begin{pmatrix} 1 \\ x \end{pmatrix}$, pak $\overbrace{\begin{pmatrix} 1 & x \end{pmatrix}}^A \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \end{pmatrix}$

Nejprve počítejme pro dívky, Označme $\mathbf{b}^T A = \mathbf{x}^T = (1, x, 0, 0)$ **3)** Odvodíme tvar pásu spolehlivosti (PS) $P\left(\left|\mathbf{x}^T \hat{\beta} - \underbrace{\mathbf{x}^T \beta}_y\right|^2 \leq 2 F_{1-\alpha}(2, n-4) \sigma^2 \mathbf{x}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}\right) = 1 - \alpha$ kde y je náhodná proměnná. Upravujeme

$P\left(\left|\mathbf{x}^T \hat{\beta} - y\right| \leq \sqrt{2 F_{1-\alpha}(2, n-4) \sigma^2 \mathbf{x}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}}\right) = 1 - \alpha$

- pro $\mathbf{x}^T \hat{\beta} - y > 0$ dostáváme **dolní hranici** $P(y \geq \mathbf{x}^T \hat{\beta} - \sqrt{2 F_{1-\alpha}(2, n-4) \sigma^2 \mathbf{x}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}}) = 1 - \alpha$

- nebo pro $\mathbf{x}^T \hat{\beta} - y < 0$ dostáváme **horní hranici** $P\left(y \leq \mathbf{x}^T \hat{\beta} + \sqrt{2 F_{1-\alpha}(2, n-4)} \sigma^2 \mathbf{x}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}\right) = 1 - \alpha$