M5170: Matematické programování

Petr Zemánek (Masarykova Univerzita, Brno)

Kapitola 2: Základy konvexní analýzy

(verze: 24. září 2020)

Konvexnost má ohromně bohatou strukturu a početné využití. Na druhou stranu mohou být téměř všechny „konvexní“ pojmy vysvětleny pomocí dvoudimenzionálního obrázku.

Alexandre Barvinok, *A Course in Convexity*, AMS (2002).

Definice 2.1.1

Nechť $X \subset \mathbb{R}^n$. Množina X se nazývá <u>konvexní</u>, jestliže pro všechna $x_1, x_2 \in X$ a pro každé $\lambda \in [0,1]$ platí

$$\lambda x_1 + (1-\lambda) x_2 \in X.$$

Obrázek.

$X = \emptyset$ je konvexní?

Konvexnost vs. sjednocení, průniku a sčítání (odčítání)?

Věta 2.1.2

(i) Nechť I je libovolná indexová množina a množiny $X_i \subseteq \mathbb{R}^n$ jsou konvexní pro všechna $i \in I$. Pak množina $\bigcap_{i \in I} X_i$ je také konvexní.

(ii) Nechť $X_1, \dots, X_m \subseteq \mathbb{R}^n$ jsou konvexní množiny a $\alpha_1, \dots, \alpha_m \in \mathbb{R}$. Pak množina

$$\alpha_1 X_1 + \dots + \alpha_m X_m := \left\{ x \in \mathbb{R}^n \mid x = \sum_{i=1}^{m} \alpha_i x_i \text{ pro nějaká } x_i \in X_i \right\}$$

je opět konvexní.

Co je $X - X$?

Důkaz Věty 2.1.2

(i) Nechť I je konečná nebo nekonečná indexová množina a nechť X_i jsou konvexní množiny pro všechna $i \in I$. Je-li množina $X := \bigcap_{i \in I} X_i = \emptyset$ nebo jednoprvková, pak je tvrzení triviální. Uvažme tedy alespoň dvouprvkovou množinu X. Nechť $x_1, x_2 \in X$ jsou libovolné. Pak nutně $x_1, x_2 \in X_i$ pro všechna $i \in I$. Jenže všechny množiny X_i jsou konvexní, takže nutně $x := \lambda x_1 + (1-\lambda) x_2 \in X_i$ pro všechna $\lambda \in [0,1]$ a všechna $i \in I$. To ale znamená, že také $x \in X$, tj. $\bigcap_{i \in I} X_i$ je konvexní.

(ii) Nechť X je konvexní množina. Nejdříve ukážeme, že také množina

$$Z = \alpha X := \left\{ z \in \mathbb{R}^n \mid z = \alpha x \text{ pro nějaké } x \in X \right\}$$

je konvexní pro libovolné $\alpha \in \mathbb{R}$. Je-li $X = \emptyset$ nebo jednoprvková, je tvrzení triviální. Uvažme tedy alespoň dvouprvkovou množinu X. Jestliže $z_1, z_2 \in \alpha X$ jsou libovolné, pak $z_1 = \alpha x_1$ a $z_2 = \alpha x_2$ pro nějaké $x_1, x_2 \in X$. Proto pro každé $\lambda \in [0,1]$ máme

$$\lambda z_1 + (1-\lambda) z_2 = \alpha \underbrace{[\lambda x_1 + (1-\lambda) x_2]}_{\in X} \in \alpha X = Z,$$

tj. αX je konvexní.

Důkaz Věty 2.1.2 (pokr.)

Nechť nyní X, Y jsou konvexní množiny. Ukážeme, že také množina

$$Z = X + Y := \left\{ z \in \mathbb{R}^n \mid z = x + y \text{ pro nějaké } x \in X \text{ a } y \in Y \right\}$$

je konvexní. Jestliže X, Y jsou nejvýše jednoprvkové, je tvrzení triviální. Uvažme tedy, že alespoň jedna z těchto množin je alespoň dvouprvková. Jestliže $z_1, z_2 \in X + Y$, pak $z_1 = x_1 + y_1$ a $z_2 = x_2 + y_2$ pro nějaká $x_1, x_2 \in X$ a $y_1, y_2 \in Y$. Proto pro každé $\lambda \in [0,1]$ máme

$$\begin{aligned} \lambda z_1 + (1-\lambda) z_2 &= \lambda (x_1 + y_1) + (1-\lambda)(x_2 + y_2) \\ &= \underbrace{[\lambda x_1 + (1-\lambda) x_2]}_{\in X} + \underbrace{[\lambda y_1 + (1-\lambda) y_2]}_{\in Y} \in X + Y = Z, \end{aligned}$$

tj. $X + Y$ je konvexní. Snadnou úvahou pak obdržíme samotné tvrzení věty. ■

Definice 2.1.3

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

(i) kužel, jestliže pro každé $x \in X$ a pro každé $\lambda \in [0, \infty)$ je také $\lambda x \in X$;

(ii) konvexní kužel, jestliže je množina X konvexní a současně kuželem;

(iii) afinní (též lineární varieta), jestliže pro každé $x_1, x_2 \in X$ a pro každé $\lambda \in \mathbb{R}$ platí

$$\lambda x_1 + (1 - \lambda) x_2 \in X.$$

✎ Příklady a další komentář...

Definice 2.1.4

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

(i) konvexní, jestliže $\lambda_1, \dots, \lambda_m \geqslant 0$ a $\sum_{i=1}^m \lambda_i = 1$;

(ii) nezáporná, jestliže $\lambda_1, \dots, \lambda_m \geqslant 0$;

(iii) afinní, jestliže $\sum_{i=1}^m \lambda_i = 1$.

✎ Obrázek.

S využitím Definice 2.1.4 ihned dostaneme:

- 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 výseč určenou polopřímkami vycházejícími z počátku a procházejícími těmito body) je *konvexní kužel*.
- Množina obsahující všechny konvexní kombinace libovolných *dvou* svých bodů (tj. s libovolnými dvěma body obsahuje i úsečku spojující tyto body) je *konvexní*.

A pro více než dvojice?

Věta 2.1.5

Nechť $X \subseteq \mathbb{R}^n$. Množina X je konvexní (konvexní kužel/afinní) právě tehdy, když libovolná konvexní (nezáporná/afinní) kombinace prvků z X je opět prvkem množiny X, tj. je uzavřená na všechny konvexní (nezáporné/afinní) kombinace.

Technicistní (vnitřní) vs. „umělecký“ (vnější) popis konvexních množin.

Důkaz Věty 2.1.5

Tvrzení dokážeme pouze pro *konvexní* kombinace, neboť zbývající dvě varianty lze dokázat s použitím analogických argumentů.

„$\Longleftarrow$“ Triviální.

„$\Longrightarrow$“ Nechť množina X je konvexní. Je-li $X = \emptyset$ nebo jednoprvková je tvrzení triviální. Nechť tedy X obsahuje alespoň dva různé body. Tvrzení dokážeme pomocí matematické indukce. Z definice ihned vyplývá (viz výše), že libovolná konvexní kombinace prvků z X je opět v X pro $m = 2$. Připusťme, že každá konvexní kombinace m-tice prvků z X je opět v X pro nějaké $m \in \mathbb{N}\backslash\{1\}$. Uvažme libovolnou konvexní kombinaci $m+1$ prvků z X, tj. $x := \sum_{i=1}^{m+1} \lambda_i x_i$ pro libovolná $x_1, \dots, x_{m+1} \in X$ a $\lambda_1, \dots, \lambda_{m+1} \geqslant 0$ splňující $\sum_{i=1}^{m+1} \lambda_i = 1$. Je-li $\lambda_{m+1} = 1$, pak $\lambda_1 = \cdots = \lambda_m = 0$ a zjevně $x \in X$. Je-li $\lambda_{m+1} \in [0,1)$, pak můžeme psát

$$x = (1 - \lambda_{m+1}) \sum_{i=1}^{m} \frac{\lambda_i}{1 - \lambda_{m+1}} x_i + \lambda_{m+1} x_{m+1}.$$

Ovšem $\sum_{i=1}^{m} \frac{\lambda_i}{1-\lambda_{m+1}} = 1$ a $\sum_{i=1}^{m} \frac{\lambda_1}{1-\lambda_{m+1}} x_i \in X$ podle indukčního předpokladu. Pak ale x máme vyjádřeno jako konvexní kombinaci dvou bodů z X, z čehož díky konvexnosti X ihned vyplývá $x \in X$, což dokazuje, že X je konvexní množina. ■

Definice 2.1.6

Nechť $X \subseteq \mathbb{R}^n$.

(i) Průnik všech konvexních množin obsahujících množinu X se nazývá konvexní obal množiny X a značí se $\operatorname{conv} X$.

(ii) Průnik všech konvexních kuželů obsahujících množinu X se nazývá kónický (kuželový) obal množiny X a značí se $\operatorname{cone} X$.

(iii) Průnik všech afinních množin obsahujících množinu X se nazývá afinní obal množiny X a značí se $\operatorname{aff} X$. Jeho zaměření se nazývá lineární obal množiny X a značí se $\operatorname{Lin} X$. Dimenze afinního obalu množiny X se značí $\dim X$ a klademe $\dim X := \dim \operatorname{Lin} X$.

Lin vs. span.

Věta 2.1.2(i) & Definice 2.1.6 …

Obrázek.

Věta 2.1.7

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

$$\operatorname{conv} X = \Big\{x \in \mathbb{R}^n \mid x = \sum_{i=1}^{m} \lambda_i x_i \text{ pro libovolné } m \in \mathbb{N} \text{ a } \\ x_1, \dots, x_m \in X,\ \lambda_1, \dots, \lambda_m \geqslant 0,\ \sum_{i=1}^{m} \lambda_i = 1\Big\},$$

$$\operatorname{cone} X = \Big\{x \in \mathbb{R}^n \mid x = \sum_{i=1}^{m} \lambda_i x_i \text{ pro libovolné } m \in \mathbb{N} \text{ a } \\ x_1, \dots, x_m \in X,\ \lambda_1, \dots, \lambda_m \geqslant 0\Big\},$$

$$\operatorname{aff} X = \Big\{x \in \mathbb{R}^n \mid x = \sum_{i=1}^{m} \lambda_i x_i \text{ pro libovolné } m \in \mathbb{N} \text{ a } \\ x_1, \dots, x_m \in X,\ \sum_{i=1}^{m} \lambda_i = 1\Big\}.$$

Důkaz Věty 2.1.7

Opět dokážeme pouze první část, neboť zbývající dvě lze ukázat pomocí analogických argumentů.
Nechť

$$Z := \Big\{x \in \mathbb{R}^n \mid x = \sum_{i=1}^{m} \lambda_i x_i \text{ pro libovolné } m \in \mathbb{N} \text{ a } \\ x_1, \dots, x_m \in X,\ \lambda_1, \dots, \lambda_m \geqslant 0,\ \sum_{i=1}^{m} \lambda_i = 1\Big\}.$$

Postupně ukážeme, že $\operatorname{conv} X \subseteq Z$ a $\operatorname{conv} X \supseteq Z$, z čeho ihned vyplyne tvrzení věty.

„$\subseteq$" Je zřejmé, že $X \subseteq Z$. Proto $\operatorname{conv} X \subseteq \operatorname{conv} Z$. Jenže množina Z je konvexní (!!V.2.1.5!! X vs. Z). Vskutku, jsou-li $z_1, z_2 \in Z$ libovolné, tj.

$$z_1 = \sum_{i=1}^{m} \alpha_i x_i \quad \& \quad z_2 = \sum_{i=1}^{k} \beta_i y_i$$

pro nějaká $m, k \in \mathbb{N}$, $x_1, \dots, x_m, y_1, \dots, y_k \in X$, $\alpha_1, \dots, \alpha_m, \beta_1, \dots, \beta_k \geqslant 0$ splňující $\sum_{i=1}^{m} \alpha_i = 1 = \sum_{i=1}^{k} \beta_i$. BÚNO můžeme předpokládat, že $m \leqslant k$. Proto položme $\alpha_{m+1} = \cdots = \alpha_k = 0$.

Důkaz Věty 2.1.7 (pokr.)

Pak pro libovolné $\lambda \in [0,1]$ máme

$$z \coloneqq \lambda z_1 + (1-\lambda) z_2 = \sum_{i=1}^{k} (\lambda \alpha_i x_i + (1-\lambda) \beta_i y_i) \in Z$$

podle definice množiny Z, neboť $\lambda \alpha_i, (1-\lambda) \beta_i \geqslant 0$ a současně

$$\sum_{i=1}^{k} (\lambda \alpha_i + (1-\lambda) \beta_i) = \lambda + (1-\lambda) = 1.$$

Tedy $Z = \operatorname{conv} Z$, a tudíž $\operatorname{conv} X \subseteq \operatorname{conv} Z = Z$.

„$\supseteq$“ Nechť Y je libovolná konvexní množina taková, že $X \subseteq Y$. Podle Věty 2.1.5 je libovolná konvexní kombinace prvků z Y prvkem Y, což vzhledem k inkluzi $X \subseteq Y$ implikuje $Z \subseteq Y$. Ovšem $Y \supseteq X$ byla zvolena libovolně, takže nutně platí

$$Z \subseteq \bigcap_{\substack{Y \supseteq X \\ Y \text{ konvexní}}} Y \overset{\text{Dfn. 2.1.6}}{=} \operatorname{conv} X.$$

■

Je skutečně skutečně nutné charakterizovat $\operatorname{conv} X$ pomocí všech možných m-tic?

Lemma 2.1.8

Nechť $X \subseteq \mathbb{R}^n$ a $x \in \operatorname{cone} X$ je libovolné. Pak existují body $x_1, \dots, x_n \in X$ a čísla $\lambda_1, \dots, \lambda_n \geqslant 0$ taková, že $x = \lambda_1 x_1 + \cdots + \lambda_n x_n$.

Důkaz. Nechť $x \in \operatorname{cone} X$ je libovolné. Podle Věty 2.1.7 existuje $m \in \mathbb{N}$, body $x_1, \dots, x_m \in X$ a čísla $\lambda_1, \dots, \lambda_m \geqslant 0$ taková, že $x = \sum_{i=1}^{m} \lambda_i x_i$. Jsou-li body $x_1, \dots, x_m$ „lineárně nezávislé“, pak nutně $m \leqslant n$ a tvrzení věty platí. Je-li $m > n$, pak jsou tyto body nutně „lineárně závislé“. Proto existují $\mu_1, \dots, \mu_m \in \mathbb{R}$ taková, že $\sum_{i=1}^{m} \mu_i x_i = 0$, přičemž alespoň jedno $\mu_i > 0$. Nechť $\alpha \in \mathbb{R}$ je libovolné. Potom

$$x = \sum_{i=1}^{m} \lambda_i x_i - \alpha \sum_{i=1}^{m} \mu_i x_i = \sum_{i=1}^{m} (\lambda_i - \alpha \mu_i) x_i. \tag{2.1.1}$$

Zvolme nyní $\alpha \coloneqq \min_{\mu_i > 0} \frac{\lambda_i}{\mu_i}$. Pak $\lambda_i - \alpha \mu_i \geqslant 0$ pro všechny indexy $i \in \{1, \dots, m\}$, přičemž alespoň pro jeden index i platí $\lambda_i - \alpha \mu_i = 0$, tj. podle (2.1.1) lze x vyjádřit jako nezápornou kombinaci nejvýše $m-1$ bodů. Je-li $m-1 = n$, je důkaz hotov. V opačném případě můžeme celou konstrukci zopakovat ještě $(m-n-1)$-krát, tj. dokud nedostaneme x jako nezápornou kombinaci nejvýše n bodů. ■

Z předchozího tvrzení snadno získáme analogický výsledek pro konvexní obal.

Věta 2.1.9 (Carathéodory)

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

$$x = \lambda_1 x_1 + \cdots + \lambda_{n+1} x_{n+1}.$$

Důkaz. Nechť $A \subseteq \mathbb{R}^{n+1}$ je množina bodů tvaru $A := \{[x, 1] \mid x \in X\}$. Pak platí, že $y = [x, 1] \in \operatorname{cone} A$ právě tehdy, když $x \in \operatorname{conv} X$. Vskutku, nechť $x \in \operatorname{conv} X$, tj. $x = \sum_{i=1}^m \lambda_i x_i$ pro nějaké $m \in \mathbb{N}$, $x_1, \dots, x_m \in X$ a $\lambda_1, \dots, \lambda_m \geqslant 0$ splňující $\sum_{i=1}^m \lambda_i = 1$. Pak $[x, 1] \in \operatorname{cone} A$, neboť

$$[x, 1] = \left[\sum_{i=1}^m \lambda_i x_i, 1\right] = \left[\sum_{i=1}^m \lambda_i x_i, \sum_{i=1}^m \lambda_i\right] = \sum_{i=1}^m \lambda_i [x_i, 1],$$

což je nezáporná kombinace bodů z A. Naopak, je-li $[x, 1] \in \operatorname{cone} A$, pak máme $[x, 1] = \sum_{i=1}^m \lambda_i [x_i, 1]$ pro nějaké $m \in \mathbb{N}$, $x_1, \dots, x_m \in X$ a $\lambda_1, \dots, \lambda_m \geqslant 0$, tj. $x = \sum_{i=1}^m \lambda_i x_i$ a $\sum_{i=1}^m \lambda_i = 1$ neboli $x \in \operatorname{conv} X$. Jenže my navíc z Lemma 2.1.8 víme, že nám k předchozímu vyjádření stačí nejvýše $n+1$ bodů, čímž je tvrzení dokázáno. ■

Konvexní obal vs. kompaktnost? A jiné vlastnosti?

Důsledek 2.1.10

Nechť $X \subseteq \mathbb{R}^n$ je kompaktní množina. Pak $\operatorname{conv} X$ je také kompaktní.

Vnitřní bod?

Definice 2.1.11

Nechť $X \subseteq \mathbb{R}^n$. Bod $x^* \in X$ se nazývá <u>relativně vnitřním</u> bodem množiny X, jestliže existuje okolí $\mathcal{O}(x^*)$ bodu x^* takové, že

$$\mathcal{O}(x^*) \cap \operatorname{aff} X \subseteq X.$$

Množinu všech relativně vnitřních bodů množiny X nazýváme <u>relativním vnitřkem</u> množiny X a značíme $\operatorname{ri} X$. Množina $\operatorname{r\partial} X := \overline{X} \setminus \operatorname{ri} X$ se nazývá <u>relativní hranice</u> množiny X.

Obrázek.

$\operatorname{ri} X$ vs. $\operatorname{int} X$.

$X_1 \subseteq X_2 \implies \operatorname{int} X_1 \subseteq \operatorname{int} X_2$ a $\operatorname{ri} X_1$ vs. $\operatorname{ri} X_2$?

$\operatorname{ri} X$ pro afinní množinu X?

Základní vztah: $\operatorname{ri} X \subseteq X \subseteq \overline{X} \subseteq \operatorname{aff} X$ ($\subseteq$ vs. $\subset$).

Konvexní množiny

Věta 2.1.12

Nechť $X \subseteq \mathbb{R}^n$ je neprázdná konvexní množina.

(i) Pokud $x \in \operatorname{ri} X$ a $y \in \overline{X}$, pak pro každé $\lambda \in (0, 1]$ platí

$$(1-\lambda)y + \lambda x \in \operatorname{ri} X.$$

(ii) Množina $\operatorname{ri} X$ je neprázdná, konvexní a $\operatorname{aff} X = \operatorname{aff}(\operatorname{ri} X)$. Je-li navíc $\dim X = m > 0$, pak existují $z_0, \dots, z_m \in \operatorname{ri} X$ taková, že $z_1 - z_0, \dots, z_m - z_0$ tvoří bázi $\operatorname{Lin} X$.

(iii) Platí $x \in \operatorname{ri} X$ právě tehdy, když pro každé $\tilde{x} \in X$ existuje $\gamma > 1$ takové, že

$$x + (\gamma - 1)(x - \tilde{x}) = \gamma x + (1-\gamma)\tilde{x} \in X.$$

(iv) Platí $\overline{X} = \overline{\operatorname{ri} X}$, $\operatorname{ri} X = \operatorname{ri} \overline{X}$, množina $\overline{X}$ je konvexní, $\operatorname{aff} X = \operatorname{aff} \overline{X}$ a je-li Y jiná konvexní množina, pak následující tvrzení jsou ekvivalentní
  (a) $\operatorname{ri} X = \operatorname{ri} Y$,
  (b) $\overline{X} = \overline{Y}$,
  (c) $\operatorname{ri} X \subseteq Y \subseteq \overline{X}$.

Důsledky…

Konvexní funkce

Konvexní funkce v $\mathbb{R}$?

Definice 2.2.1

Nechť $X \subseteq \mathbb{R}^n$ je konvexní množina. Funkce $f : X \to \mathbb{R}$ se nazývá

(i) konvexní na X, jestliže pro všechna $x_1, x_2 \in X$ a každé $\lambda \in [0, 1]$ platí

$$f(\lambda x_1 + (1-\lambda) x_2) \leqslant \lambda f(x_1) + (1-\lambda) f(x_2); \tag{2.2.1}$$

(ii) ostře (ryze) konvexní na X, jestliže nerovnost (2.2.1) je ostrá pro všechna $x_1, x_2 \in X$, $x_1 \neq x_2$ a každé $\lambda \in (0, 1)$;

(iii) silně konvexní na X s konstantou silné konvexnosti $\vartheta > 0$, jestliže pro všechna $x_1, x_2 \in X$ a každé $\lambda \in [0, 1]$ platí

$$f(\lambda x_1 + (1-\lambda) x_2) \leqslant \lambda f(x_1) + (1-\lambda) f(x_2) - \vartheta \lambda (1-\lambda) \|x_1 - x_2\|^2. \tag{2.2.2}$$

Konvexnost v „praxi“?

Příklady a další jednoduché vlastnosti…

Věta 2.2.2

Nechť $X \subseteq \mathbb{R}^n$ je konvexní množina a nechť $f : X \to \mathbb{R}$. Funkce f je konvexní na X právě tehdy, když její *nadgraf* (epigraf)

$$\operatorname{epi} f := \left\{ [x, \beta] \in \mathbb{R}^{n+1} \mid x \in X,\ \beta \geqslant f(x) \right\}$$

je konvexní množina.

Důkaz Věty 2.2.2

„$\Longrightarrow$" Nechť f je konvexní funkce na X. Ukážeme, že s libovolnými dvěma body z $\operatorname{epi} f$ obsahuje $\operatorname{epi} f$ také celou úsečku spojující tyto body. Nechť tedy $\lambda \in [0,1]$ a $[x_1, \beta_1], [x_2, \beta_2] \in \operatorname{epi} f$, tj. $\beta_1 \geqslant f(x_1)$ a $\beta_2 \geqslant f(x_2)$. Potom

$$\lambda[x_1, \beta_1] + (1-\lambda)[x_2, \beta_2] = \big[\underbrace{\lambda x_1 + (1-\lambda) x_2}_{\in X}, \lambda\beta_1 + (1-\lambda)\beta_2\big]$$

a současně

$$\lambda\beta_1 + (1-\lambda)\beta_2 \geqslant \lambda f(x_1) + (1-\lambda) f(x_2) \geqslant f(\lambda x_1 + (1-\lambda) x_2),$$

což znamená, že také $\lambda[x_1, \beta_1] + (1-\lambda)[x_2, \beta_2] \in \operatorname{epi} f$, tj. $\operatorname{epi} f$ je konvexní.

„$\Longleftarrow$" Nechť $\operatorname{epi} f$ je konvexní množina. Ukážeme, že pak pro libovolné dva body z X platí (2.2.1). Nechť $x_1, x_2 \in X$ a $\lambda \in [0,1]$. Pak pro $[x_1, f(x_1)], [x_2, f(x_2)] \in \operatorname{epi} f$ je také

$$\lambda[x_1, f(x_1)] + (1-\lambda)[x_2, f(x_2)] = \big[\lambda x_1 + (1-\lambda) x_2, \lambda f(x_1) + (1-\lambda) f(x_2)\big] \in \operatorname{epi} f,$$

což ale znamená, že $\lambda f(x_1) + (1-\lambda) f(x_2) \geqslant f(\lambda x_1 + (1-\lambda) x_2)$, tj. f je konvexní na X. ■

Věta 2.2.3

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

Příklady.

Nadgraf vs. vrstevnice.

Věta 2.2.4

Nechť $X \subseteq \mathbb{R}^n$ je konvexní množina a $f : X \to \mathbb{R}$ je konvexní funkce na X. Pak pro libovolné $K \in \mathbb{R}$ je odpovídající *dolní vrstevnicová množina*

$$V_K := \big\{x \in X \mid f(x) \leqslant K\big\}$$

také konvexní.

Důkaz. Je-li $V_K = \emptyset$ nebo jednoprvková množina, opak je tvrzení triviální. Nechť je tedy V_K alespoň dvouprvková množina a $x_1, x_2 \in V_k$ jsou libovolné. Pak pro libovolné $\lambda \in [0,1]$ máme

$$f(\lambda x_1 + (1-\lambda) x_2) \leqslant \lambda f(x_1) + (1-\lambda) f(x_2) \leqslant \lambda K + (1-\lambda) K = K,$$

tj. také $\lambda x_1 + (1-\lambda) x_2 \in V_k$, a tedy V_K je skutečně konvexní množina. ■

Konvexnost vs. *kvazikonvexnost*.

Věta 2.2.5 (Jensen)

Nechť $X \subseteq \mathbb{R}^n$ je konvexní množina a funkce $f: X \to \mathbb{R}$ je konvexní na X. Pak pro libovolné $m \in \mathbb{N}$, $x_1, \dots, x_m \in X$ a čísla $\lambda_1, \dots, \lambda_m \geqslant 0$ splňující $\sum_{i=1}^m \lambda_i = 1$ platí

$$f\left(\sum_{i=1}^m \lambda_i x_i\right) \leqslant \sum_{i=1}^m \lambda_i f(x_i). \tag{2.2.3}$$

Je-li funkce f navíc ostře konvexní a $\lambda_1, \dots, \lambda_m \in (0,1)$, pak rovnost v (2.2.3) nastane právě tehdy, když $x_1 = \dots = x_m$.

Ekvivalence v první části?

A ve druhé části? Obměna!

$$\begin{aligned}(A \wedge B) \Longrightarrow (C \Longleftrightarrow D) &\equiv \neg(C \Longleftrightarrow D) \Longrightarrow \neg(A \wedge B)\\ &\equiv \big((C \wedge \neg D) \vee (\neg C \wedge D)\big) \Longrightarrow (\neg A \vee \neg B)\end{aligned}$$

Jestliže existuje $m \in \mathbb{N}$ takové, že buď v (2.2.3) nastane rovnost pro různé body nebo nastane ostrá nerovnost pro stejné body, pak funkce f není ostře konvexní nebo $\lambda_i \in \{0,1\}$ pro nějaké $i \in \{1 \dots, m\}$.

Přehlednější: Nechť $m \in \mathbb{N}$ je libovolné. Pak pro $\lambda_1, \dots, \lambda_m \in [0,1]$ splňující $\sum_{i=1}^m \lambda_i = 1$ a m-tici $x_1, \dots, x_m \in X$ obsahující alespoň dva různé body s odpovídajícími kladnými koeficienty λ_p a λ_q nastane v (2.2.3) ostrá nerovnost právě tehdy, když funkce f je ostře konvexní.

Wikipedia: rovnost v (2.2.3) $\Longleftrightarrow$ f je lineární nebo $x_1 = \dots = x_m$?!

Důkaz Věty 2.2.5

Tvrzení dokážeme pomocí matematické indukce. Pro $m = 1$ je nerovnost triviální, zatímco pro $m = 2$ je ekvivalentní s definicí konvexností funkce f na X. Připusťme nyní, že tvrzení platí pro nějaké $m \in \mathbb{N} \setminus \{1\}$. Ukážeme, že pak platí i pro $m+1$. Je-li $\lambda_{m+1} = 1$, potom $\lambda_1 = \dots = \lambda_m = 0$ a nerovnost (2.2.3) je splněna triviálně. Uvažme tedy $\lambda_{m+1} < 1$. Pak

$$\begin{aligned}&f(\lambda_1 x_1 + \dots + \lambda_{m+1} x_{m+1})\\ &\quad = f\left((1-\lambda_{m+1})\left[\frac{\lambda_1}{1-\lambda_{m+1}} x_1 + \dots + \frac{\lambda_m}{1-\lambda_{m+1}} x_m\right] + \lambda_{m+1} x_{m+1}\right),\end{aligned}$$

přičemž $\frac{\lambda_1}{1-\lambda_{m+1}}, \dots, \frac{\lambda_m}{1-\lambda_{m+1}} \in [0,1]$ a $\sum_{i=1}^m \frac{\lambda_i}{1-\lambda_{m+1}} = 1$. Proto platí $\widetilde{x} := \frac{\lambda_1}{1-\lambda_{m+1}} x_1 + \dots + \frac{\lambda_m}{1-\lambda_{m+1}} x_m \in X$, a tedy z konvexnosti funkce f na X plyne

Důkaz Věty 2.2.5 (pokr.)

$$
\begin{aligned}
f(\lambda_1 x_1 + \cdots + \lambda_{m+1} x_{m+1}) &= f\big((1-\lambda_{m+1})\widetilde{x} + \lambda_{m+1} x_{m+1}\big) \\
&\overset{\text{konvexnost}}{\leqslant} (1-\lambda_{m+1}) f(\widetilde{x}) + \lambda_{m+1} f(x_{m+1}) \\
&= (1-\lambda_{m+1}) f\left(\frac{\lambda_1}{1-\lambda_{m+1}} x_1 + \cdots + \frac{\lambda_m}{1-\lambda_{m+1}} x_m\right) + \lambda_{m+1} f(x_{m+1}) \\
&\overset{\text{„indukce“}}{\leqslant} (1-\lambda_{m+1}) \left[\frac{\lambda_1}{1-\lambda_{m+1}} f(x_1) + \cdots + \frac{\lambda_m}{1-\lambda_{m+1}} f(x_m)\right] + \lambda_{m+1} f(x_{m+1}) \\
&= \lambda_1 f(x_1) + \cdots + \lambda_{m+1} f(x_{m+1}),
\end{aligned}
$$

což dokazuje platnost nerovnosti (2.2.3) pro libovolné $m \in \mathbb{N}$.

Nechť je nyní navíc f ostře konvexní a $\lambda_1, \ldots, \lambda_m \in (0,1)$. Je-li $x_1 = \cdots = x_m =: x \in X$, pak zřejmě platí

$$
f(\lambda_1 x_1 + \cdots + \lambda_m x_m) = f\left(x \sum_{i=1}^{m} \lambda_i\right) = f(x) = f(x) \sum_{i=1}^{m} \lambda_i = \sum_{i=1}^{m} \lambda_i f(x) = \sum_{i=1}^{m} \lambda_i f(x_i),
$$

tedy v (2.2.3) nastává rovnost.

Důkaz Věty 2.2.5 (pokr.)

Naopak, nechť $x_1, \ldots, x_m \in X$ jsou taková, že v (2.2.3) nastává rovnost. Opět využijeme matematickou indukci. Nechť $m = 2$ a $\lambda_1, \lambda_2 \in (0,1)$ splňují $\lambda_1 + \lambda_2 = 1$. Pak rovnost (2.2.3) je tvaru

$$
f(\lambda_1 x_1 + \lambda_2 x_2) = \lambda_1 f(x_1) + \lambda_2 f(x_2).
$$

Z definice ostré konvexnosti pak ale nutně plyne, že $x_1 = x_2$ (proč?). Nechť toto platí pro nějaké $m \in \mathbb{N} \setminus \{1\}$ a nechť

$$
f\left(\sum_{i=1}^{m+1} \lambda_i x_i\right) = \sum_{i=1}^{m+1} \lambda_i f(x_i), \tag{2.2.4}
$$

kde $\lambda_1, \ldots, \lambda_{m+1} \in (0,1)$ splňují $\sum_{i=1}^{m+1} \lambda_i = 1$. Položme $\widetilde{x} \coloneqq \sum_{i=1}^{m} \frac{\lambda_i}{1-\lambda_{m+1}} x_i$. Potom

$$
\begin{aligned}
f(\lambda_1 x_1 + \cdots + \lambda_m x_m) &= f\big((1-\lambda_{m+1})\widetilde{x} + \lambda_{m+1} x_{m+1}\big) \\
&\overset{\text{konvexnost}}{\leqslant} (1-\lambda_{m+1}) f(\widetilde{x}) + \lambda_{m+1} f(x_{m+1}) \\
&= (1-\lambda_{m+1}) f\left(\frac{\lambda_1}{1-\lambda_{m+1}} x_1 + \cdots + \frac{\lambda_m}{1-\lambda_{m+1}} x_m\right) + \lambda_{m+1} f(x_{m+1}) \\
&\overset{(2.2.3)}{\leqslant} \lambda_1 f(x_1) + \cdots + \lambda_{m+1} f(x_{m+1}).
\end{aligned}
$$

Důkaz Věty 2.2.5 (pokr.)

Jenže vzhledem k indukčnímu předpokladu (2.2.4) je zřejmé, že se nerovnosti v předchozím výpočtu ve skutečnosti realizují jako rovnosti. Pak první z nich implikuje $\widetilde{x} = x_{m+1}$ (ze stejného důvodu jako pro $m = 2$) a poslední $x_1 = \cdots = x_m$ (indukční předpoklad). Pak z definice $\widetilde{x}$ plyne

$$\begin{aligned} x_{m+1} = \widetilde{x} &= \sum_{i=1}^{m} \frac{\lambda_i}{1-\lambda_{m+1}}\, x_i = x_1 \sum_{i=1}^{m} \frac{\lambda_i}{1-\lambda_{m+1}} \\ &= \frac{x_1}{1-\lambda_{m+1}} \sum_{i=1}^{m} \lambda_i = \frac{x_1}{1-\lambda_{m+1}} (1-\lambda_{m+1}) = x_1, \end{aligned}$$

což dohromady dává $x_1 = \cdots = x_m = x_{m+1}$ a důkaz je hotov. ■

Aplikace Jensenovy nerovnosti (AG-nerovnost a jiné).

Věta 2.2.6

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

(i) Libovolné lokální minimum funkce f na X je současně globálním minimem.

(ii) Množina bodů množiny X, v nichž funkce f nabývá svého minima na X, je konvexní. Je-li funkce f dokonce ostře konvexní, pak je tato množina nejvýše jednoprvková.

(iii) Je-li funkce f diferencovatelná na otevřené množině $\mathcal{U} \supseteq X$ a $x^* \in X$ je jejím stacionárním bodem, tj. $\operatorname{grad} f(x^*) = \mathbb{0}$, pak x^* je bodem globálního minima funkce f na množině X.

„Konstantnost“ v části (ii)

Ekvivalence v části (iii)?

Důsledek

Je-li $f : X \to \mathbb{R}$ (ostře) konvexní a *spojitá* funkce na konvexní a kompaktní množině $X \subseteq \mathbb{R}^n$, pak f má na X (právě jedno) globální minimum.

Důkaz. Weisestrassova věta. ■

Minimalizace & silná konvexnost

Je-li funkce $f : X \to \mathbb{R}$ silně konvexní s konstantou silné konvexnosti $\vartheta > 0$ a *spojitá* na uzavřené konvexní množině $X \subseteq \mathbb{R}^n$, potom pro libovolné $K \in \mathbb{R}$ jsou dolní vrstevnicové množiny V_K ohraničené.

Proto má f na X právě jedno globální minimum ($\Longleftarrow$ opět „Weierstrass“).

$\rightsquigarrow$

Silně konvexní funkce jsou ideálním objektem pro minimalizační úlohy, což se bohužel ne vždy stává... Ovšem na druhou stranu silná konvexnost pochopitelně není nutná pro existenci jediného globálního minima (viz níže).

$X = \mathbb{R}^n$

Pro funkci $f : \mathbb{R}^n \to \mathbb{R}$ platí:

(i) je-li f konvexní na $\mathbb{R}^n$ a v $x^* \in \mathbb{R}^n$ nastává lokální minimum, pak x^* je globální minimum;

(ii) je-li f ostře konvexní na $\mathbb{R}^n$ a v $x^* \in \mathbb{R}^n$ nastává lokální minimum, pak x^* je jediné globální minimum;

(iii) je-li f silně konvexní na $\mathbb{R}^n$, pak *existuje jediné globální minimum.*

✎ Příklady.

Důkaz Věty 2.2.6

(i) Důkaz provedeme sporem. Nechť x^* je lokální minimum funkce f na X, tj. existuje $\mathcal{O}(x^*)$ takové, že pro všechna $x \in \mathcal{O}(x^*) \cap X$ platí $f(x) \geqslant f(x^*)$. Nechť ale současně x^* není globálním minimem funkce f na X, tj. existuje $\widehat{x} \in X$ takové, že $f(\widehat{x}) < f(x^*)$. Pak z konvexnosti funkce f a množiny X plyne

$$f(\underbrace{\lambda x^* + (1-\lambda)\widehat{x}}_{\in X}) \leqslant \lambda f(x^*) + (1-\lambda) f(\widehat{x}) < \lambda f(x^*) + (1-\lambda) f(x^*) = f(x^*)$$

pro všechny $\lambda \in [0,1)$. To je ale spor s tím, že x^* je lokálním minimem, neboť pro λ dostatečně blízko 1 musí být $\lambda x^* + (1-\lambda)\widehat{x} \in \mathcal{O}(x^*) \cap X$ (✎obrázek).

Důkaz Věty 2.2.6 (pokr.)

(ii) Označme jako X^* množinu všech bodů, ve kterých funkce f nabývá svého minima na X, tj. $X^* := \{x \in X \mid f(x) = f^* = \inf_{x \in X} f(x)\}$. Je-li množina X^* prázdná nebo jednoprvková, je tvrzení triviální. Nechť tedy X^* je alespoň dvouprvková, tj. máme různé $x_1, x_2 \in X^*$. Pak pro libovolné $\lambda \in [0, 1]$ platí

$$f(\lambda x_1 + (1-\lambda) x_2) \leqslant \lambda f(x_1) + (1-\lambda) f(x_2) = f^*. \tag{$*$}$$

Jenže z definice f^* vyplývá, že také $f^* \leqslant f(\lambda x + (1-\lambda) y)$ pro libovolné $x, y \in X$. To znamená, že v $(*)$ nastává rovnost, tj. $\lambda x_1 + (1-\lambda) x_2 \in X^*$ pro libovolné $\lambda \in [0, 1]$, a tedy množina X^* je konvexní. Je-li navíc funkce f ostře konvexní, pak pro $\lambda \in (0, 1)$ nastane v $(*)$ rovnost právě tehdy, když $x_1 = x_2$, tj. množina X^* je nejvýše jednoprvková.

(iii) Viz důsledky Věty 2.4.2 (později) nebo přímo.

Minimalizace konvexní funkce na **ne**konvexní množině?

A maximalizace konvexní funkce na konvexní množině?

Věta

(i) Nechť $f : X \to \mathbb{R}$ je konvexní funkce na konvexní množině $X \subseteq \mathbb{R}^n$. Jestliže f nabývá svého maxima na množině X v bodě $x^* \in \operatorname{ri} X$, pak f je konstantní funkce, tj. nekonstantní konvexní funkce může nabývat svého globálního maxima pouze v bodech množiny $X \setminus \operatorname{ri} X$ (tj. na $\operatorname{r\partial} X \cap X$ — zejména je-li X otevřená, pak f svého maxima na X nenabývá).

(ii) Základní věta konvexního programování: Máme-li konvexní funkci $f : X \to \mathbb{R}$ na polytopu $X := \operatorname{conv}\{x_1, \dots, x_m\} \subseteq \mathbb{R}^n$, pak je maximum funkce f na X dosaženo v některém z bodů $x_1, \dots, x_m$ (případně ve více z nich $\rightsquigarrow$ ✎ na celé hraně??).

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).

Důsledek=Základní věta lineárního programování: Je-li funkce f afinní, pak *globální minimum* nastává v některém z bodů $x_1, \dots, x_m$ („vrcholů“ polytopu).

✎ Motivace.

Definice 2.3.1 Neprázdné množiny $X_1, X_2 \subseteq \mathbb{R}^n$ se nazývají

(i) oddělitelné, jestliže existuje $p \in \mathbb{R}^n \setminus \{0\}$ takový, že

$$\langle p, x_1 \rangle \geqslant \langle p, x_2 \rangle \tag{2.3.1}$$

pro každé $x_1 \in X_1$ a $x_2 \in X_2$;

(ii) vlastně oddělitelné, jestliže jsou oddělitelné a zároveň existují body $x_1^* \in X_1$ a $x_2^* \in X_2$ splňující

$$\langle p, x_1^* \rangle > \langle p, x_2^* \rangle\,;$$

(iii) silně oddělitelné, jestliže existuje $p \in \mathbb{R}^n \setminus \{0\}$ takový, že

$$\inf_{x_1 \in X_1} \langle p, x_1 \rangle > \sup_{x_2 \in X_2} \langle p, x_2 \rangle\,.$$

Je-li navíc $\beta \in [\sup_{x_2 \in X_2} \langle p, x_2 \rangle, \inf_{x_1 \in X_1} \langle p, x_1 \rangle]$, nadrovina

$$H_{p,\beta} := \{x \in \mathbb{R}^n \mid \langle p, x \rangle = \beta\}$$

se nazývá oddělující nadrovinou množin X_1 a X_2.

✎ Vlastnosti z definice a příklady…

Definice 2.3.2

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 $\Pi_X(x)$, jestliže

$$\|\Pi_X(x) - x\| \leqslant \|y - x\| \quad \text{pro každé } y \in X.$$

Příklady (existence a jednoznačnost).

Lemma 2.3.3

Nechť $X \subseteq \mathbb{R}^n$ je neprázdná, uzavřená a konvexní množina. Pak pro libovolný bod $x^* \in \mathbb{R}^n$ existuje jediná projekce $\Pi_X(x^*) \in X$ a pro každé $x \in X$ platí

$$\langle \Pi_X(x^*) - x^*, \, x - \Pi_X(x^*) \rangle \geqslant 0, \tag{2.3.2}$$

$$\langle \Pi_X(x^*) - x^*, \, x - x^* \rangle \geqslant \|\Pi_X(x^*) - x^*\|^2 \geqslant 0. \tag{2.3.3}$$

Důkaz Lemma 2.3.3

Je-li $x^* \in X$, pak tvrzení je triviální, neboť nutně $\Pi_X(x^*) = x^*$.

Nechť tedy $x^* \notin X$. Nejdříve si uvědomme, že projekce bodu x^* na množinu X je vlastně řešením úlohy

$$f(x) := \|x - x^*\| \to \min, \quad x \in X. \tag{2.3.4}$$

Pak funkce f odpovídá kvadratické formě s maticí $A = I$ plus afinní část. To znamená, že f je ostře konvexní, a z Věty 2.2.6(ii) vyplývá, že úloha (2.3.4) má nejvýše jedno řešení. Položme nyní

$$Y := X \cap \{x \in \mathbb{R}^n \mid \|x - x^*\| \leqslant R\},$$

kde $R \in \mathbb{R}$ dostatečně velké (jak? viz níže). Pak množina Y je kompaktní, a tedy existuje řešení úlohy

$$f(x) \to \min, \quad x \in Y.$$

Jenže řešení této úlohy je totožné s řešením (2.3.4) (proč? $\rightsquigarrow$ volba R), tj. projekce $\Pi_X(x^*)$ existuje a je jednoznačně určena.

Důkaz Lemma 2.3.3 (pokr.)

Z konvexnosti množiny X platí pro libovolné $x \in X$ a $\lambda \in (0,1]$, že

$$\begin{aligned}
\|\Pi_X(x^*) - x^*\|^2 &\overset{\Pi_X(x^*)\text{ řeší (2.3.4)}}{\leqslant} \|\underbrace{\lambda x + (1-\lambda)\Pi_X(x^*)}_{\in X\ \&\ \neq \Pi_X(x^*)} - x^*\|^2 = \\
&= \big\|\Pi_X(x^*) - x^* + \lambda\,[x - \Pi_X(x^*)]\big\|^2 = \\
&= \langle \Pi_X(x^*) - x^* + \lambda\,[x - \Pi_X(x^*)],\, \Pi_X(x^*) - x^* + \lambda\,[x - \Pi_X(x^*)]\rangle = \\
&= \langle \Pi_X(x^*) - x^*,\, \Pi_X(x^*) - x^*\rangle + \lambda\,\langle \Pi_X(x^*) - x^*,\, x - \Pi_X(x^*)\rangle + \\
&\qquad + \lambda\,\langle x - \Pi_X(x^*),\, \Pi_X(x^*) - x^*\rangle + \lambda^2\,\langle x - \Pi_X(x^*),\, x - \Pi_X(x^*)\rangle \\
&= \|\Pi_X(x^*) - x^*\|^2 + 2\lambda\,\langle \Pi_X(x^*) - x^*,\, x - \Pi_X(x^*)\rangle + \lambda^2 \|x - \Pi_X(x^*)\|^2.
\end{aligned}$$

Vydělením číslem $\lambda \in (0,1]$ dostáváme

$$0 \leqslant 2\,\langle \Pi_X(x^*) - x^*,\, x - \Pi_X(x^*)\rangle + \lambda \|x - \Pi_X(x^*)\|^2,$$

z čehož limitním přechodem pro $\lambda \to 0^+$ obdržíme

$$\langle \Pi_X(x^*) - x^*,\, x - \Pi_X(x^*)\rangle \geqslant 0,$$

což je nerovnost (2.3.2). Navíc odtud plyne

$$\begin{aligned}
0 &\leqslant \langle \Pi_X(x^*) - x^*,\, x - \Pi_X(x^*) + x^* - x^*\rangle \\
&= \langle \Pi_X(x^*) - x^*,\, x - x^*\rangle + \langle \Pi_X(x^*) - x^*,\, x^* - \Pi_X(x^*)\rangle \\
&= \langle \Pi_X(x^*) - x^*,\, x - x^*\rangle - \|\Pi_X(x^*) - x^*\|^2,
\end{aligned}$$

tj. platí $\langle \Pi_X(x^*) - x^*,\, x - x^*\rangle \geqslant \|\Pi_X(x^*) - x^*\|^2 \geqslant 0$, což je nerovnost (2.3.3). ■

Díky předchozímu výsledku „snadno“ odvodíme nutnou a postačující podmínku silné oddělitelnosti.

Věta 2.3.4

Neprázdné konvexní množiny $X_1, X_2 \subseteq \mathbb{R}^n$ jsou silně oddělitelné právě tehdy, když mají nenulovou vzdálenost, tj.

$$\rho(X_1, X_2) := \inf_{x_1 \in X_1,\, x_2 \in X_2} \|x_1 - x_2\| > 0,$$

což je ekvivalentní s podmínkou $0 \notin \overline{X_1 - X_2}$.

Toto ukazuje rozdíl mezi vlastní a silnou oddělitelností dvou množin.

Poznámka

Podmínky Věty 2.3.4 jsou splněny zejména v případě

(i) $X_1 = \{x\}$ a $x \notin \overline{X_2}$,

(ii) $\overline{X_1} \cap \overline{X_2} = \emptyset$ a X_1 nebo X_2 je ohraničená.

Pozor! Nestačí, aby množiny X_1, X_2 byly konvexní, disjunktní a uzavřené, např. ✎, viz následující důsledek.

Důkaz Věty 2.3.4

„$\Longrightarrow$“ Nechť X_1, X_2 jsou silně oddělitelné (nemusí být konvexní), pak

$$\begin{aligned} 0 < \varepsilon &=: \inf_{x_1\in X_1} \langle p, x_1\rangle - \sup_{x_2\in X_2} \langle p, x_2\rangle \\ &\overset{\inf A - \sup B = \inf(A-B)}{=} \inf_{(x_1,x_2)\in X_1\times X_2} \langle p, x_1 - x_2\rangle \\ &\overset{\text{C–S}}{\leqslant} \inf_{(x_1,x_2)\in X_1\times X_2} \|p\|\|x_1 - x_2\| = \|p\|\rho(X_1, X_2), \end{aligned}$$

a tedy $\rho(X_1, X_2) \geqslant \varepsilon/\|p\| > 0$.

Důkaz Věty 2.3.4 (pokr.)

„$\Longleftarrow$“ Nechť X_1, X_2 jsou konvexní množiny a platí $\rho(X_1, X_2) > 0$. Označme $X := \overline{X_1 - X_2}$ (pozor, platí pouze $\overline{X_1} - \overline{X_2} \subseteq \overline{X_1 - X_2}$). Pak X je konvexní a uzavřená (viz V.2.1.2(ii) a V.2.12(iv)). Protože $\rho(X_1, X_2) > 0 \Longleftrightarrow \rho(\overline{X_1}, \overline{X_2}) > 0$, máme $0 \notin X$ (✎vskutku?). Pak ale $p := \Pi_X(0) \neq 0$ a ukážeme, že právě tento vektor splňuje požadavky silné oddělitelnosti (tj. p je normálový vektor /silně/ oddělující nadroviny). Podle Lemma 2.3.3 je vektor p určen jednoznačně a z nerovnosti (2.3.3) máme při $x^* = 0$ vztah

$$\langle p, x\rangle \geqslant \|p\|^2 > 0$$

pro všechna $x \in X$, z čehož ihned vyplývá $\inf_{x\in X} \langle p, x\rangle > 0$. Současně pro libovolné $x_1 \in X_1$ a $x_2 \in X_2$ je $x := x_1 - x_2 \in X$ a platí

$$\inf_{x\in X} = \inf_{(x_1,x_2)\in X_1\times X_2)} \langle p, x_1 - x_2\rangle = \inf_{x_1\in X_1} \langle p, x_1\rangle - \sup_{x_2\in X_2} \langle p, x_2\rangle .$$

Tudíž

$$\inf_{x_1\in X_1} \langle p, x_1\rangle - \sup_{x_2\in X_2} \langle p, x_2\rangle > 0,$$

z čehož plyne tvrzení. ■

Z Věty 2.3.4 ihned vyplývá následující tvrzení (důkaz?).

Důsledek 2.3.5

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

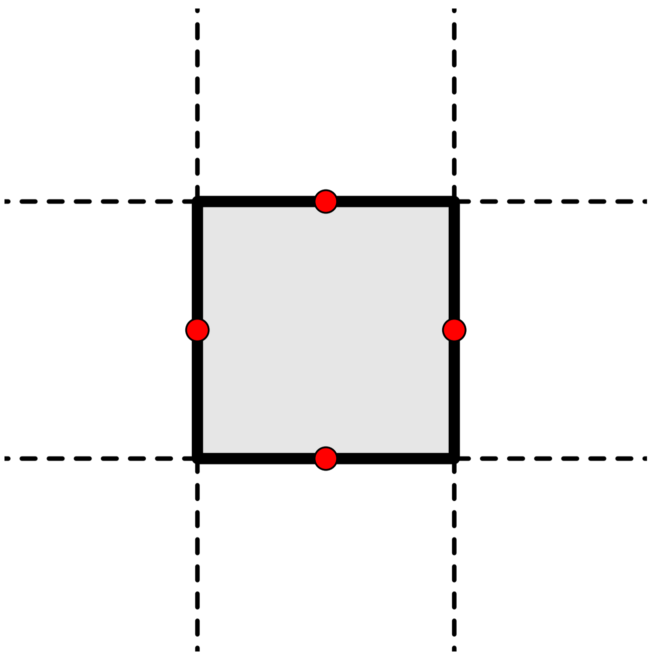
A nyní již můžeme podat „umělecký" (vnější) popis uzavřených konvexních množin (tzv. geometrická Hahnova–Banachova věta $\rightsquigarrow$ základní princip konvexní geometrie).

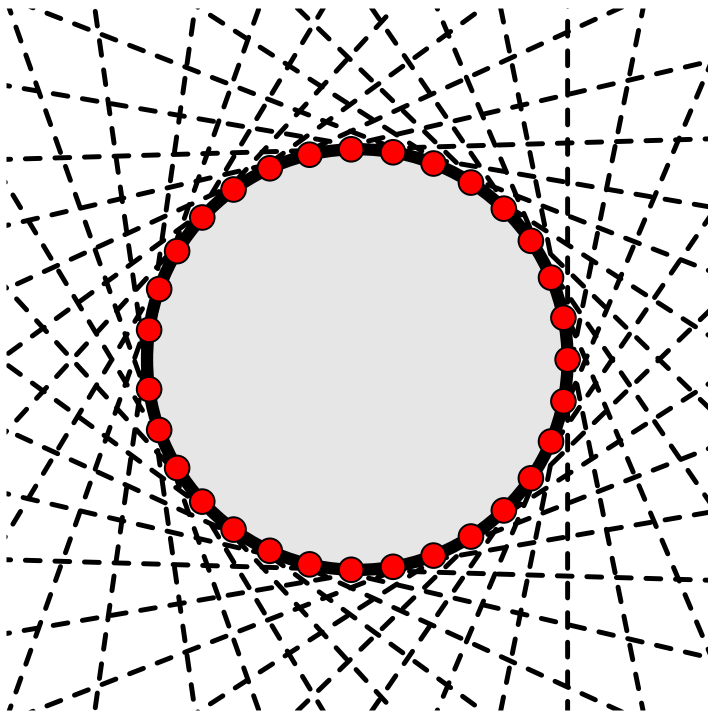
Věta 2.3.5a

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

(Geometricky: každá uzavřená konvexní množina $X \subsetneq \mathbb{R}^n$ je průnikem uzavřených poloprostorů, konkrétně všech uzavřených poloprostorů obsahujících X.)

„Důkaz" Věty 2.3.5a





■

Mezi všemi uzavřenými poloprostory, které obsahují uzavřenou konvexní a netriviální (tj. $\neq \emptyset$ a $\neq \mathbb{R}^n$) množinu X jsou velmi zajímavé ty „extrémní“, tj. takové jejichž hraniční nadrovina se dotýká množiny X.

Tato terminologie je použitelná pro libovolnou (nikoli nutně uzavřenou) konvexní množinu, což je obsahem následující definice.

S pomocí těchto nadrovin odvodíme postačující podmínku pro oddělitelnost dvou konvexních množin (V.2.3.7a) a ekvivalentní charakterizaci vlastní oddělitelnosti dvou konvexních množin (V.2.3.8).

Definice 2.3.6

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

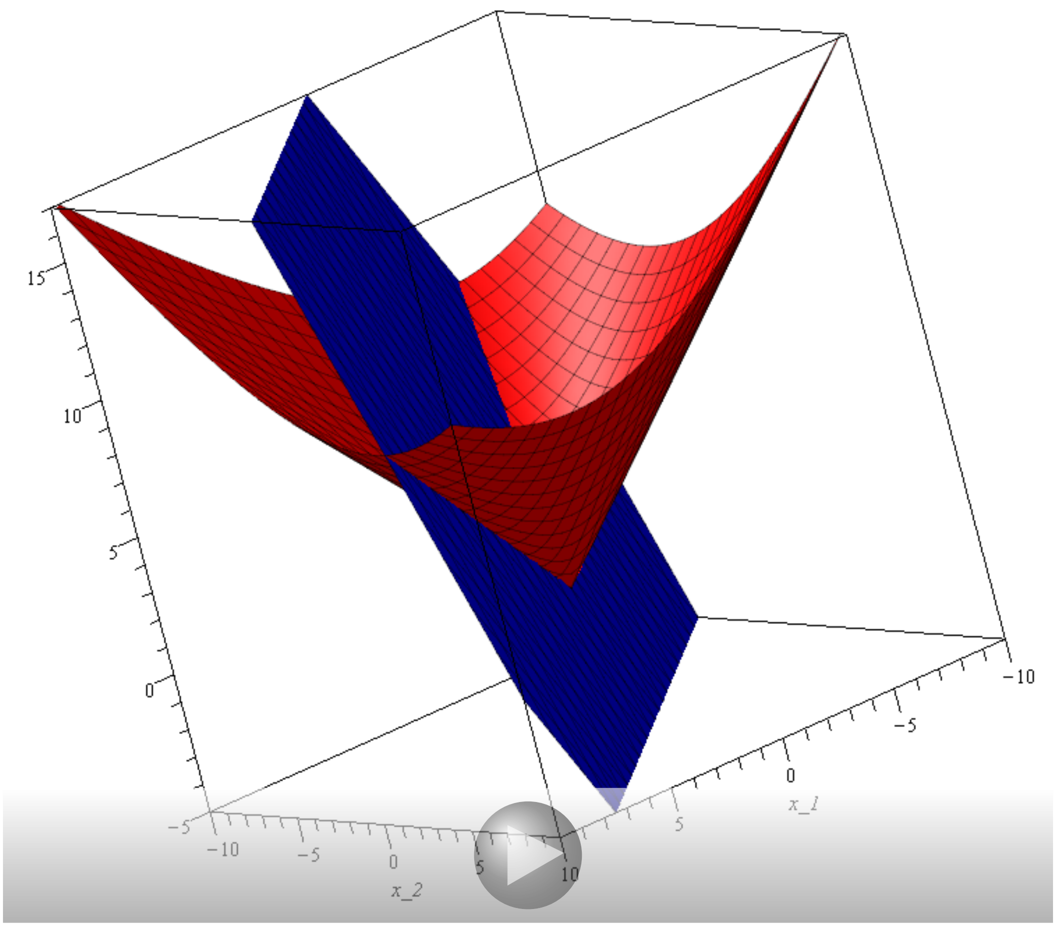
(i) <u>opěrnou nadrovinou</u> množiny X v bodě a, jestliže

$$\langle p, x \rangle \geqslant \beta = \langle p, a \rangle \quad \text{pro každé } x \in X;$$

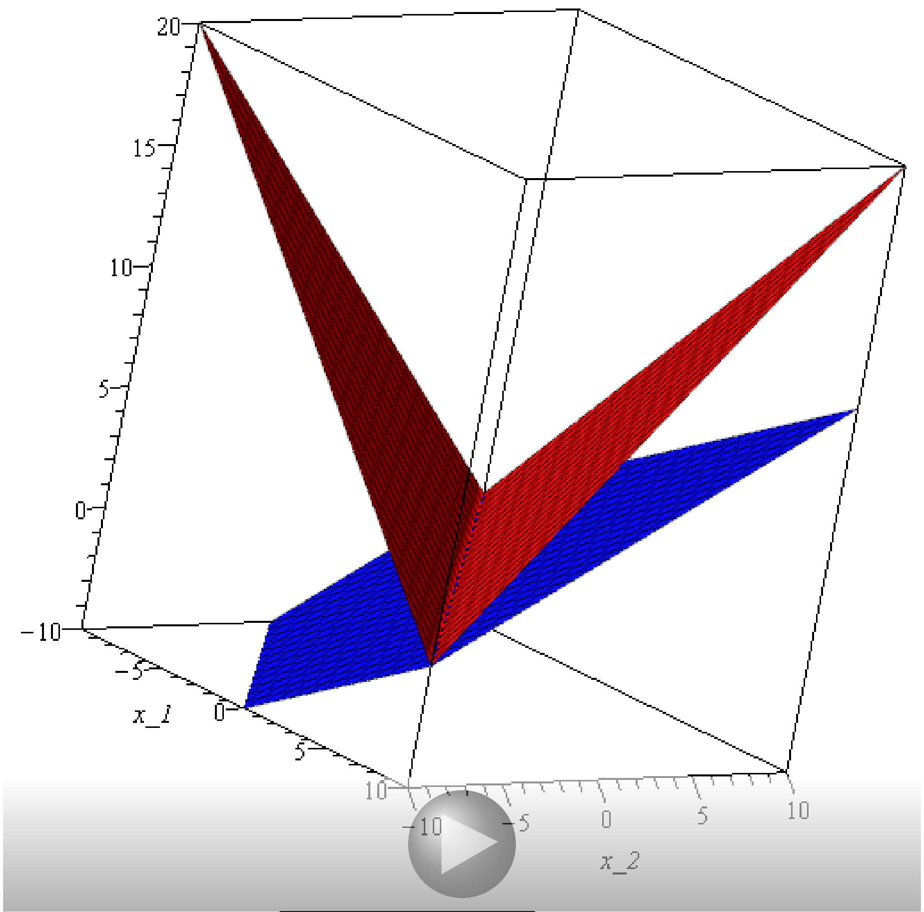
(ii) <u>vlastní opěrnou nadrovinou</u> množiny X, jestliže je opěrnou nadrovinou množiny X a existuje-li $x^* \in X$ takové, že

$$\langle p, x^* \rangle > \beta.$$

✎ Příklady.

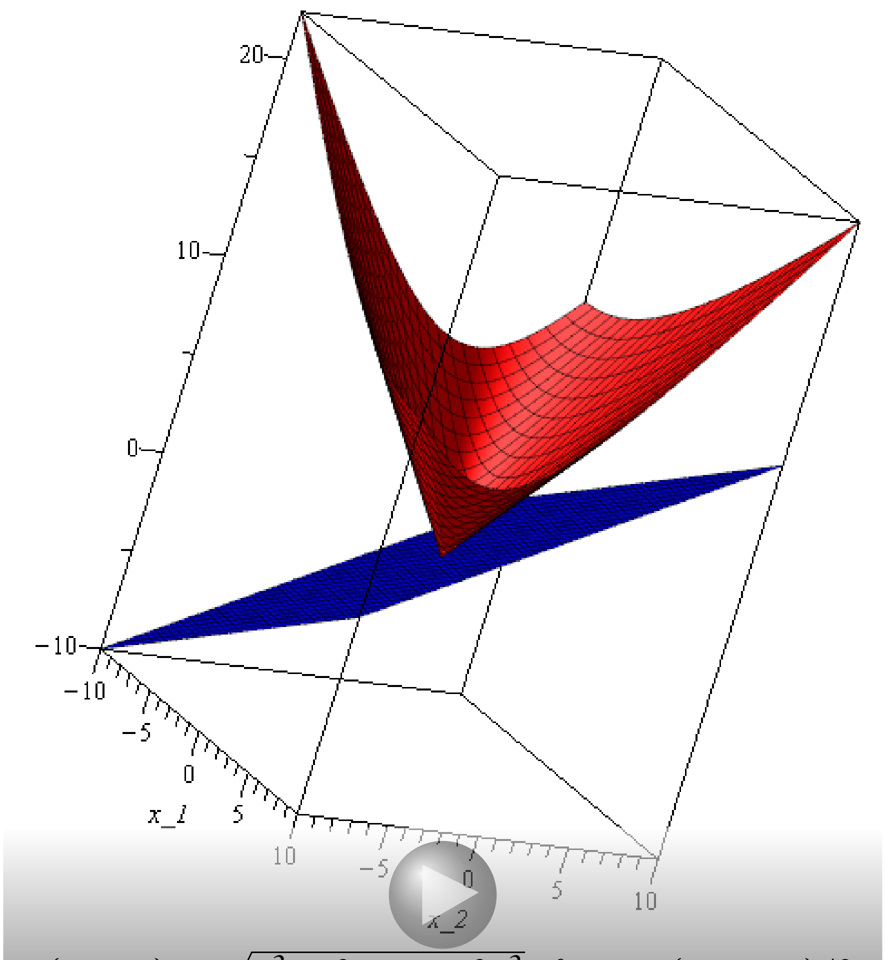


$$f(x_1, x_2) = \sqrt{x_1^2 - x_1 x_2 + x_2^2} \quad \& \quad z = 2(x_1 - x_2)$$



$$f(x_1, x_2) = \sqrt{x_1^2 - 2x_1x_2 + x_2^2} \quad \& \quad z = (x_1 - x_2)/2$$

Oddělování konvexních množin a jeho důsledky



$$f(x_1, x_2) = \sqrt{x_1^2 + 2x_1x_2 + 2x_2^2} \quad \& \quad z = (x_1 + x_2)/2$$

Věta 2.3.7

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

Poznámka

(i) $r\partial X$ vs. ∂X.

(ii) Tvrzení Věty 2.3.7 můžeme ještě *rozšířit*: pro existenci opěrné nadroviny stačí $a \in \partial X$.
Důkaz. Je-li $\operatorname{int} X = \emptyset$, pak $\operatorname{aff} X \neq \mathbb{R}^n$ a $\operatorname{aff} X$ je průsečík několika nadrovin, vždyť $\operatorname{aff} X = \{x \in \mathbb{R}^n \mid Ax = b\}$. Potom ale nutně některá z těchto nadrovin je právě opěrnou (ne nutně vlastní) nadrovinou pro $a \in \partial X \subseteq \overline{X}$. Je-li $\operatorname{int} X \neq \emptyset$, pak $\partial X = r\partial X$ a tvrzení plyne přímo z Věty 2.3.7.

(iii) $X = \{a\}$?

Důkaz Věty 2.3.7

Popíšeme konstrukci normálového vektoru opěrné nadroviny. Nechť $a \in r\partial X$ je libovolný bod. Pak existuje posloupnost $\{a_k\}_{k\in\mathbb{N}}$ taková, že $a_k \in \operatorname{aff} X \backslash X$ a $a_k \to a$ pro $k \to \infty$ (✎obrázek). Potom podle Věty 2.3.3 existuje $\Pi_X(a_k)$ pro každé $k \in \mathbb{N}$ a platí $\Pi_X(a_k) \neq a_k$. Položme

$$p_k := \frac{\Pi_X(a_k) - a_k}{\|\Pi_X(a_k) - a_k\|}.$$

Pak $\|p_k\| = 1$ a $p_k \in \operatorname{Lin} X$ (viz Definici 2.1.6). Množina $\{y \in \mathbb{R}^n \mid \|y\| = 1\}$ je kompaktní, můžeme z posloupnosti $\{p_k\}_{k\in\mathbb{N}}$ vybrat konvergentní podposloupnost. BÚNO proto předpokládejme, že $p_k \to p$, přičemž $\|p\| = 1$ a $p \in \operatorname{Lin} X$ (z uzavřenosti $\operatorname{Lin} X$). Je toto hledaný normálový vektor? Z nerovnosti (2.3.3) plyne pro libovolné $x \in X$, že

$$\begin{aligned}\langle p_k, x\rangle - \langle p_k, a_k\rangle = \langle p_k, x - a_k\rangle &= \frac{1}{\Pi_X(a_k) - a_k} \langle \Pi_X(a_k) - a_k, x - a_k\rangle \\ &\overset{(2.3.3)}{\geqslant} \frac{1}{\Pi_X(a_k) - a_k} \|\Pi_X(a_k) - a_k\|^2 = \|\Pi_X(a_k) - a_k\| > 0,\end{aligned}$$

neboť $\Pi_X(a_k) \neq a_k$. Limitním přechodem pro $k \to \infty$ dostáváme

$$\langle p, x\rangle \geqslant \langle p, a\rangle =: \beta \quad \text{pro každé } x \in X,$$

což znamená, že $H_{p,\beta}$ je hledaná opěrná nadrovina.

A je vlastní?

Důkaz Věty 2.3.7 (pokr.)

Podle Věty 2.1.12(ii) je $\operatorname{ri} X \neq \emptyset$, a tedy existuje $x_1 \in \operatorname{ri} X$, tj. $\mathcal{O}_\varepsilon(x_1) \cap \operatorname{aff} X \subseteq X$ pro nějaké $\varepsilon > 0$. Položme $\overline{x} := x_1 - \overline{\varepsilon} p$ pro nějaké dostatečně malé $\overline{\varepsilon} > 0$. Protože $p \in \operatorname{Lin} X$, je $\overline{x} \in \operatorname{aff} X$. Z definice $\operatorname{ri} X$ navíc vyplývá, že pro dostatečně malé $\overline{\varepsilon} > 0$ je dokonce $\overline{x} \in \operatorname{ri} X$, a tedy také $\overline{x} \in X$. Proto

$$\beta \leqslant \langle p, \overline{x} \rangle = \langle p, x_1 - \overline{\varepsilon} p \rangle = \langle p, x_1 \rangle - \overline{\varepsilon} \langle p, p \rangle = \langle p, x_1 \rangle - \overline{\varepsilon},$$

tj. $\langle p, x_1 \rangle \geqslant \beta + \overline{\varepsilon} > \beta$, což dokazuje, že nadrovina $H_{p,\beta}$ je vlastní opěrnou nadrovinou. ∎

Tento výsledek nyní využijeme v důkazech tvrzení o oddělitelnosti konvexních množin. Nejdříve začneme se samotnou oddělitelností (srovnej s Větou 2.3.4).

Věta 2.3.7a Nechť $X_1, X_2 \subseteq \mathbb{R}^n$ jsou neprázdné, konvexní a disjunktní množiny. Pak pro tyto množiny existuje oddělující nadrovina.

Důkaz. Uvažme konvexní množinu

$$X := X_1 - X_2 = \{x \in \mathbb{R}^n \mid x = x_1 - x_2 \text{ pro nějaké } x_1 \in X_1,\ x_2 \in X_2\}.$$

Protože $X_1 \cap X_2 = \emptyset$, máme $0 \notin X$ a mohou tedy nastat pouze dvě možnosti:

(i) $0 \notin \overline{X}$. Pak množiny $\{0\}$ a X jsou silně oddělitelné, tj. existuje $p \in \mathbb{R}^n \setminus \{0\}$ takový, že

$$\inf_{x \in X} \underbrace{\langle p, x \rangle}_{=\langle p, x_1 - x_2 \rangle} > \sup_{y \in \{0\}} \langle p, y \rangle = 0,$$

tj. $\inf_{x_1 \in X_1} \langle p, x_1 \rangle > \sup_{x_2 \in X_2} \langle p, x_2 \rangle$, a tedy množiny X_1, X_2 jsou dokonce silně oddělitelné.

(ii) $0 \in \overline{X} \setminus X \subseteq \partial X$. Pak podle poznámky za Větou 2.3.7 existuje opěrná nadrovina, tj. existuje $p \in \mathbb{R}^n \setminus \{0\}$ takový, že

$$\langle p, x \rangle \geqslant \beta := \langle p, 0 \rangle = 0$$

pro všechna $x \in X$, tj. $\langle p, x_1 \rangle \geqslant \langle p, x_2 \rangle$ pro všechna $x_1 \in X_1$ a $x_2 \in X_2$. ∎

Ekvivalence? A je-li X_1 otevřená? Bez otevřenosti?

Věta 2.3.8 Neprázdné konvexní množiny $X_1, X_2 \subseteq \mathbb{R}^n$ jsou vlastně oddělitelné právě tehdy, když $\operatorname{ri} X_1 \cap \operatorname{ri} X_2 = \emptyset$.

Důkaz. „$\Longrightarrow$“ Sporem. Nechť X_1, X_2 jsou vlastně oddělitelné nadrovinou $H_{p,\beta}$ a předpokládejme, že existuje $x \in \operatorname{ri} X_1 \cap \operatorname{ri} X_2$. Z definice vlastní oddělitelnosti vyplývá, že existují $\overline{x_1} \in X_1$ a $\overline{x_2} \in X_2$ takové, že

$$\langle p, \overline{x_1} \rangle > \langle p, \overline{x_2} \rangle, \quad \text{tj.} \quad \langle p, \overline{x_1} - \overline{x_2} \rangle > 0. \tag{$*$}$$

Pro $\alpha \in [-1, 0]$ položme

$$\widetilde{x}_1 := x - \alpha(\overline{x_1} - x) = -\alpha\overline{x_1} + (1+\alpha)x \in X_1,$$
$$\widetilde{x}_2 := x - \alpha(\overline{x_2} - x) = -\alpha\overline{x_2} + (1+\alpha)x \in X_2.$$

Důkaz Věty 2.3.8 (pokr.)

Navíc podle Věty 2.1.12(iii) máme také $\widetilde{x}_1 \in X_1$ a $\widetilde{x}_2 \in X_2$ pro $\alpha > 0$ dostatečně malé, neboť

$$\widetilde{x}_1 = x + (\gamma - 1)(x - \overline{x_1}) \quad \text{a} \quad \widetilde{x}_2 = x + (\gamma - 1)(x - \overline{x_2})$$

pro $\gamma := 1 + \alpha$. Pak ale pro toto $\alpha > 0$ dostaneme

$$\langle p, \widetilde{x}_1 \rangle - \langle p, \widetilde{x}_2 \rangle = \langle p, \alpha(x - \overline{x_1}) - \alpha(x - \overline{x_2}) \rangle = -\alpha \langle p, \overline{x_1} - \overline{x_2} \rangle \overset{(*)}{<} 0,$$

což je spor s oddělitelností množin X_1 a X_2.

„$\Longleftarrow$“ Nechť $\operatorname{ri} X_1 \cap \operatorname{ri} X_2 = \emptyset$. Položme $X := \operatorname{ri} X_1 - \operatorname{ri} X_2$ (jako pro Větu 2.3.7a). Pak X je konvexní množina a $0 \notin X$, přičemž mohou nastat dvě možnosti:

(i) $0 \notin \overline{X}$. Potom podle Věty 2.3.4 jsou množiny $\{0\}$ a $\overline{X}$ silně oddělitelné a tato oddělující nadrovina vlastně odděluje i množiny X_1 a X_2.

(ii) $0 \in \overline{X} \setminus X \subseteq \operatorname{r\partial} X$. Potom podle Věty 2.3.7 existuje v bodě 0 vlastní opěrná nadrovina množiny X a právě tato nadrovina vlastně odděluje množiny X_1 a X_2. ■

Teorie soustav lineárních nerovností…

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:

$$Ax = b, \quad x \geqslant 0, \tag{2.3.5}$$

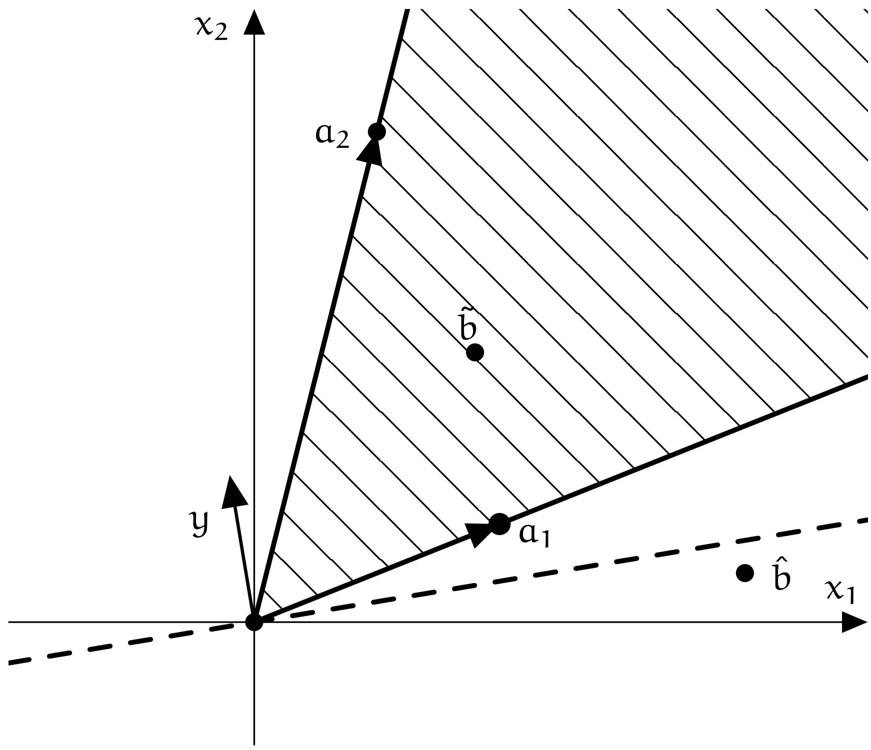
$$A^\top y \geqslant 0, \quad \langle y, b\rangle < 0. \tag{2.3.6}$$

„Obměna“

Alternativní formulace:

Soustava (2.3.5) má řešení právě tehdy, když pro všechna $y \in \mathbb{R}^m$ taková, že $A^\top y \geqslant 0$, platí $\langle y, b\rangle \geqslant 0$.

„Důkaz“ Věty 2.3.9



V literatuře lze najít celou řadu různých vět o alternativě, které různými způsoby rozšiřují/zobecňují Větu 2.3.9. My jsme uvedli právě toto tvrzení kvůli jeho důležitosti v otázce řešitelnosti úloh lineárního programování.

Věta 2.3.11

Nechť $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ a $c \in \mathbb{R}^n$ jsou dány. Jestliže (přípustná) množina
$$X \coloneqq \{x \in \mathbb{R}^n \mid Ax \geqslant b\}$$
je neprázdná a funkce $f(x) \coloneqq \langle c, x\rangle$ je zdola ohraničená na X, pak úloha lineárního programování
$$\langle c, x\rangle \to \min, \quad x \in X, \tag{2.3.7}$$
je řešitelná.

Důkaz. $\rightsquigarrow$ M0160.

Následující tvrzení a jeho „regulární“ modifikace budou hrát velmi důležitou roli při dokazování základní věty matematického programování.

Věta 2.3.12 (Fan & Glicksburg & Hoffman)

Nechť množina $X \subseteq \mathbb{R}^n$ je konvexní, funkce $f_1, \dots, f_k : X \to \mathbb{R}$ jsou konvexní a funkce $f_{k+1}, \dots, f_m$ afinní, tj. pro $j \in \{k+1, \dots, m\}$ máme $f_j(x) = \langle a_j, x\rangle + \beta_j$ pro vhodná $a_j \in \mathbb{R}^n$ a $\beta_j \in \mathbb{R}$. Jestliže systém nerovností a rovností
$$\left.\begin{aligned} f_i(x) < 0, &\quad i \in \{1, \dots, k\}, \\ f_j(x) = 0, &\quad j \in \{k+1, \dots, m\}, \end{aligned}\right\} \tag{2.3.8}$$
nemá řešení na X, pak existují takové konstanty
$$y_1, \dots, y_k \geqslant 0 \quad \text{a} \quad y_{k+1}, \dots, y_m \in \mathbb{R},$$
že alespoň pro jedno $\ell \in \{1, \dots, m\}$ je $y_\ell \neq 0$ a pro všechna $x \in X$ platí
$$\sum_{i=1}^{m} y_i f_i(x) \geqslant 0. \tag{2.3.9}$$

Důkaz Věty 2.3.12

Definujme množiny

$$A := \Big\{u \in \mathbb{R}^m \mid \text{ existuje } \overline{x} \in X \text{ tak, že } f_i(\overline{x}) \leqslant u_i,\ i \in \{1, \dots, k\},\\ \& \ f_j(\overline{x}) = u_j,\ j \in \{k+1, \dots, m\}\Big\},$$

$$B := \Big\{v \in \mathbb{R}^m \mid v_i < 0,\ i \in 1, \dots, k,\ \& \ v_j = 0,\ j \in \{k+1, \dots, m\}\Big\}.$$

Jsou-li $\overline{u}, \tilde{u} \in A$ a $\overline{x}, \widetilde{x} \in X$ odpovídající prvky z X z definice A, pak pro $\lambda \in [0, 1]$ platí, že $\lambda \overline{u} + (1-\lambda)\tilde{u} \in A$, tj. množina A je konvexní. Vskutku, toto plyne z předpokladů na f_i a

$$f_i(\lambda \overline{x} + (1-\lambda)\widetilde{x}) \leqslant \lambda f_i(\overline{x}) + (1-\lambda) f_i(\widetilde{x}) \leqslant \lambda \overline{u}_i + (1-\lambda)\tilde{u}_i, \quad i \in \{1, \dots, k\}$$

$$\begin{aligned} f_j(\lambda \overline{x} + (1-\lambda)\widetilde{x}) &= \langle a_j, \lambda \overline{x} + (1-\lambda)\widetilde{x}\rangle + \beta_j = \\ &= \lambda(\langle a_j, \overline{x}\rangle + \beta_j) + (1-\lambda)(\langle a_j, \widetilde{x}\rangle + \beta_j) = \\ &= \lambda \overline{u}_j + (1-\lambda)\tilde{u}_j, \quad j \in \{k+1, \dots, m\}. \end{aligned}$$

Oddělování konvexních množin a jeho důsledky

Důkaz Věty 2.3.12 (pokr.)

Množina B je také konvexní (zřejmé). Protože systém (2.3.8) nemá řešení na X, platí $A \cap B = \emptyset$. To znamená, že množiny A, B jsou oddělitelné, tj. existuje $y = (y_1, \dots, y_m)^\top \neq 0$ takové, že $\langle y, u\rangle \geqslant \langle y, v\rangle$ pro všechna $u \in A$ a $v \in B$, tj.

$$\sum_{i=1}^{m} y_i u_i \geqslant \sum_{i=1}^{k} y_i v_i.$$

Pokud $v_1, \dots, v_k \to -\infty$ (levá strana je daná), vidíme, že předchozí nerovnost je splněna pouze tehdy, když $y_1, \dots, y_m \geqslant 0$. Dosadíme-li do této nerovnosti $u_i := f_i(\overline{x})$ pro libovolné $\overline{x} \in X$, $i \in \{1, \dots, m\}$, a jestliže $v_1, \dots, v_k \to 0^-$, dostaneme požadované tvrzení. ■

Tvrzení Věty 2.3.12 nevylučuje $y_1 = \dots = y_k = 0$!

Nyní si ukážeme jednu regulární modifikaci Věty 2.3.12. Jejím obsahem jsou tzv. podmínky regularity, které zajišťují kladnost jistého význačného koeficientu y_i v (2.3.9). Po vydělení tohoto vztahu číslem y_i se příliš nezmění, a tak můžeme BÚNO brát $y_i = 1$. Navíc BÚNO můžeme brát, že tento význačný index odpovídá $i = 0$.

Věta 2.3.13

Nechť množina $X \subseteq \mathbb{R}^n$ je konvexní a funkce $f_0, \dots, f_m : X \to \mathbb{R}$ jsou konvexní. Jestliže systém nerovností

$$f_0(x) < 0, \tag{2.3.10}$$

$$f_i(x) < 0, \quad i \in \{1, \dots, m\}, \tag{2.3.11}$$

nemá řešení na X a podsystém (2.3.11) má řešení na X, pak existují $y_1, \dots, y_m \geqslant 0$ taková, že pro všechna $x \in X$ platí

$$f_0(x) + \sum_{i=1}^{m} y_i\, f_i(x) \geqslant 0. \tag{2.3.12}$$

Oddělování konvexních množin a jeho důsledky

Důkaz Věty 2.3.13

Využijeme tvrzení Věty 2.3.12, kde $k = m$ (tj. nemáme afinní část) a $i \in \{0, \dots, k\}$. Pak podle této věty existují $y_0, \dots, y_m \geqslant 0$ taková, že

$$y_0\, f_0(x) + \sum_{i=1}^{m} y_i\, f_i(x) \geqslant 0. \tag{$*$}$$

Je-li $y_0 = 0$, pak nutně alespoň jedno z čísel $y_1, \dots, y_m$ musí být kladné. Je-li $\overline{x} \in X$ řešením (2.3.11), pak v $(*)$ dostáváme

$$\underbrace{y_0\, f_0(\overline{x})}_{=0} + \sum_{i=1}^{m} y_i\, \underbrace{f_i(\overline{x})}_{<0} < 0,$$

což je ale spor s $(*)$, tj. nutně $y_0 > 0$. Vydělením nerovnosti $(*)$ kladným číslem y_0 a přeznačením $y_i / y_0 \rightsquigarrow y_i$ pro $i \in \{1, \dots, m\}$ dostáváme (2.3.12). ■

Nyní si ukážeme několik zajímavých (a důležitých vlastností) konvexních funkcí.

Začneme se spojitostí.

Věta 2.4.1

Nechť $X \subseteq \mathbb{R}^n$ je konvexní množina a funkce $f : X \to \mathbb{R}$ je konvexní na X. Pak f je spojitá pro každé $x \in \operatorname{ri} X$.

A pro body $x \in \operatorname{r\partial} X$?

$X = \mathbb{R}^n$?

Opačný směr: Nechť $X \subseteq \mathbb{R}^n$ je konvexní množina a funkce $f : X \to \mathbb{R}$ spojitá na X. Jestliže pro každé $x, y \in X$, $x \neq y$, existuje číslo $\lambda \in (0, 1)$ takové, že

$$f(\lambda x + (1 - \lambda) y) \leqslant \lambda f(x) + (1 - \lambda) f(y),$$

je funkce f konvexní na X?

Důkaz Věty 2.4.1

Nechť $x^* \in \operatorname{ri} X$ je libovolný a uvažme funkci $F(x) := f(x^* - x) - f(x^*)$. Pak je zřejmé, že funkce f je spojitá v x^* právě tehdy, když F je spojitá v 0 ($\lim_{x\to 0} F(x) = \lim_{x\to x^*} f(x) - f(x^*)$). Ukážeme, že F je skutečně spojitá v 0.

Uvažme nejdříve, že $\operatorname{int} X \neq \emptyset$. Označme

$$K_r := \{x \in \mathbb{R}^n \mid \|x\|_\infty < r\},$$

kde $\|x\|_\infty := \max_{1\leqslant i\leqslant n}\{|x_i|\}$ pro $x = (x_1, \dots, x_n^\top$ je „maximální norma“ a $r > 0$ je takové, že $K_r \subsetneqq X$ (tj. do X vkládám /hyper-/krychli s hranami délky $2r$). Označme jako $X_1, \dots, X_m$ pro $m = 2^n$ vrcholy krychle K_r (tj. souřadnice X_i jsou pouze $\pm r$) a položme

$$\alpha := \max_{1\leqslant i\leqslant m} F(X_i).$$

Každý bod $y \in K_r$ lze vyjádřit jako konvexní kombinaci bodů $X_1, \dots, X_m$ (vždyť $K_r = \operatorname{conv}\{X_1, \dots, X_m\}$), tj. existují $\lambda_i \geqslant 0$ splňující $\sum_{i=1}^m \lambda_i = 1$ a $y = \sum_{i=1}^m \lambda_i X_i$ (dokonce stačí pouze $n+1$ bodů, viz Větu 2.1.9). Poněvadž funkce F je také konvexní, plyne z Jensenova nerovnosti (V.2.2.8)

$$F(y) = F\Big(\sum_{i=1}^m \lambda_i X_i\Big) \leqslant \sum_{i=1}^m \lambda_i F(X_i) \leqslant \alpha \sum_{i=1}^m \lambda_i = \alpha. \tag{$*$}$$

Důkaz Věty 2.4.1 (pokr.)

Nechť nyní $\varepsilon \in (0,1]$ je libovolné. Označme $\tilde{K}_\varepsilon := \varepsilon K_r$. Pak pro každé $x \in \tilde{K}_\varepsilon$ platí, že $\pm x/\varepsilon \in K_r$, a tedy

$$F(x) = F\big(\varepsilon \tfrac{x}{\varepsilon} + (1-\varepsilon)0\big) \overset{\text{F je konvexní}}{\leqslant} \varepsilon F(\underbrace{x/\varepsilon}_{\in K_r}) + (1-\varepsilon)\underbrace{F(0)}_{=0} \overset{(*)}{\leqslant} \varepsilon\,\alpha.$$

Na druhou stranu máme

$$0 = F(0) = F\big(\tfrac{1}{1+\varepsilon}x + \tfrac{\varepsilon}{1+\varepsilon}(-x/\varepsilon)\big) \overset{\text{F je konvexní}}{\leqslant} \frac{1}{1+\varepsilon}F(x) + \frac{\varepsilon}{1+\varepsilon}F(\underbrace{-x/\varepsilon}_{\in K_r}) \overset{(*)}{\leqslant}$$

$$\overset{(*)}{\leqslant} \frac{1}{1+\varepsilon}F(x) + \frac{\varepsilon}{1+\varepsilon}\alpha,$$

což znamená, že $F(x) \geqslant -\varepsilon\,\alpha$. Celkem tedy dostáváme $|F(x)| \leqslant \varepsilon\,\alpha$ (a tedy $\alpha \geqslant 0$), což vzhledem k libovolnosti ε znamená, že funkce F je spojitá v 0 (vskutku?).

Pokud $\operatorname{int} X = \emptyset$, pak postupujeme stejně jako v předchozí části s tím, že místo n-dimenzionálních krychle uvažujeme pouze $(\dim X)$-dimenzionální krychli. ■

$f : \mathbb{R} \to \mathbb{R}$

Z kurzu Matematické analýzy I (snad) známe několik podmínek zaručujících konvexnost diferencovatelné funkce $f : \mathbb{R} \to \mathbb{R}$:

(i) 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;

(ii) 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

$$f(x) \overset{(>)}{\geqslant} f(x^*) + f'(x^*)\,(x - x^*),$$

tj. graf funkce f na I leží nad tečnou sestrojenou v libovolném bodě;

(iii) má-li f vlastní druhou derivaci v otevřeném intervalu I, pak f je konvexní na I právě tehdy, když funkce $f''(x) \geqslant 0$ (je-li $f''(x) > 0$ na I, pak je ostře konvexní).

Vlastnosti konvexních funkcí

Nyní si tato tvrzení zobecní pro $f : \mathbb{R}^n \to \mathbb{R}$. Navíc je formulujeme pro silně konvexní funkce, což se při volbě $\vartheta = 0$ redukuje na „obyčejnou“ konvexnost.

Věta 2.4.2

Nechť $X \subseteq \mathbb{R}^n$ je konvexní množina a funkce f diferencovatelná na otevřené množině $\mathcal{U} \supseteq X$. Pak f je silně konvexní na X s konstantou silné konvexnosti $\vartheta \geqslant 0$ právě tehdy, když pro každé $x, x^* \in X$ platí

$$f(x) \geqslant f(x^*) + \langle \operatorname{grad} f(x^*), x - x^* \rangle + \vartheta \|x - x^*\|^2. \tag{2.4.1}$$

Ostrá konvexnost ve Větě 2.4.2.

✎ Důkaz pro $\vartheta = 0$.

Některé zajímavé důsledky nerovnosti (2.4.1):

Poznámka

(i) Diferencovatelnost funkce f v bodě x^* je ekvivalentní s existencí tečné nadroviny ke grafu funkce f v bodě $[x^*, f(x^*)]$, která má rovnici

$$z = f(x^*) + \langle \operatorname{grad} f(x^*),\, x - x^* \rangle\,.$$

Ze vztahu (2.4.1) s $\vartheta = 0$ pak vidíme, že funkce diferencovatelná f je konvexní na X právě tehdy, když $z \leqslant f(x)$, tj. její graf leží nad tečnou nadrovinou sestrojenou v libovolném bodě $x \in X$. Navíc, položme

$$p := \big(-\operatorname{grad}^\top f(x^*), 1\big)^\top \in \mathbb{R}^{n+1} \quad \text{a} \quad \alpha := f(x^*) - \langle \operatorname{grad} f(x^*),\, x^* \rangle\,.$$

Pak nadrovina $H_{p,\alpha} = \{x \in \mathbb{R}^n \mid \langle p,\, x \rangle = \alpha\}$ je opěrnou nadrovinou k nadgrafu funkce f, tj. množině $\operatorname{epi} f$, v bodě $[x^*, f(x^*)]$, tj.

$$\left\langle p,\, (x, y)^\top \right\rangle \geqslant \alpha = \left\langle p,\, \big(x^*, f(x^*)\big)^\top \right\rangle$$

pro každé $[x, y] \in \operatorname{epi} f$.

Poznámka (pokr.)

(ii) Je-li x^* stacionárním bodem funkce f, tj. $\operatorname{grad} f(x^*) = 0$, pak (2.4.1) dává

$$f(x) \geqslant f(x^*) + \vartheta \|x - x^*\|^2.$$

Přitom $z = f(x^*) + \vartheta \|x - x^*\|^2$ je rotační paraboloid s vrcholem v bodě $[x^*, f(x^*)]$, což znamená, že graf funkce f leží uvnitř tohoto rotačního paraboloidu. ✎Obrázek.

Z Věty 2.4.2 plyne několik podstatných důsledků – nejdříve ty, které se týkají optimalizačních problémů:

Poznámka

(i) Je-li $f : \mathbb{R}^n \to \mathbb{R}$ diferencovatelná konvexní funkce a $\operatorname{grad} f(x^*) = 0$, pak x^* je globální minimum funkce f na $\mathbb{R}^n$ (srovnej s Větou 2.2.6(iii)) — klasická nutná podmínka pro nepodmíněnou optimalizaci pro libovolnou funkci, která se pro konvexní funkce stává i postačující ($n = 1 \rightsquigarrow$ Fermat /cca 1637/).

(ii) Podobně podmínka

$$\langle \operatorname{grad} f(x^*), x - x^* \rangle \geqslant 0 \quad \text{pro všechna } x \in X$$

implikuje díky (2.4.1), že x^* minimalizuje diferencovatelnou konvexní funkci f na konvexní množině X. Tato postačující podmínka optimality je současně nutná. Vskutku, připusťme, že tomu tak není, tj. x^* minimalizuje f na X a $\langle \operatorname{grad} f(x^*), x - x^* \rangle < 0$ pro nějaké $x \in X$. Potom

$$\lim_{\lambda \to 0+} \frac{f(x^* + \lambda(x - x^*)) - f(x^*)}{\lambda} = \overbrace{f_{x-x^*}(x^*)}^{\text{směrová derivace}} = \\ = \langle \operatorname{grad} f(x^*), x - x^* \rangle < 0,$$

což znamená, že pro dostatečně malé $\lambda > 0$ je $f(x^* + \lambda(x - x^*))$ klesající, což je spor s tím, že x^* minimalizuje f na X (srovnej s Větou 4.1.3 později).

Důsledek 2.4.4

Nechť $X \subseteq \mathbb{R}^n$ je konvexní množina a funkce f diferencovatelná na otevřené množině $\mathcal{U} \supseteq X$. Pak funkce f je silně konvexní na X s konstantou silné konvexnosti $\vartheta \geqslant 0$ právě tehdy, když pro každé $x, x^* \in X$ platí

$$\langle \operatorname{grad} f(x) - \operatorname{grad} f(x^*), x - x^* \rangle \geqslant 2\vartheta \|x - x^*\|^2. \tag{2.4.2}$$

Jinými slovy, $\operatorname{grad} f(x)$ je tzv. silně monotónní funkce (F je silně monotónní, jestliže existuje číslo $c > 0$ takové, že pro každé x, y platí $\langle F(x) - F(y), x - y \rangle \geqslant c\|x - y\|^2$).

Ostrá konvexnost v Důsledku 2.4.4?

A ještě zbývá zobecnění posledního tvrzení pro funkce jedné proměnné – velmi důležité kritérium (ostré/silné) konvexnosti.

Důsledek 2.4.5

Nechť $X \subseteq \mathbb{R}^n$ je konvexní množina splňující $\operatorname{int} X \neq \emptyset$. Nechť funkce $f : X \to \mathbb{R}$ je dvakrát spojitě diferencovatelná na otevřené množině $\mathcal{U} \supseteq X$ s maticí druhých derivací $\nabla^2 f(x)$. Pak f je silně konvexní na množině X s konstantou silné konvexnosti $\vartheta \geqslant 0$ právě tehdy, když pro každé $x \in X$ a $h \in \mathbb{R}^n$ platí

$$\left\langle \nabla^2 f(x)\, h,\, h \right\rangle \geqslant 2\vartheta \|h\|^2, \tag{2.4.3}$$

jinými slovy $\nabla^2 f(x) \geqslant 2\vartheta I$ pro všechna $x \in X$.

Důkaz Důsledku 2.4.5

„$\Longrightarrow$“ Nechť funkce f je silně konvexní s konstantou silné konvexnosti $\vartheta \geqslant 0$ a nechť $x^* \in \operatorname{int} X$ je libovolné. Pak $x = x^* + \lambda h \in X$ pro $\lambda > 0$ dostatečně malé a libovolné $h \in \mathbb{R}^n$. Protože f má spojité parciální derivace druhého řádu, platí

$$f(x) = f(x^*) + \langle \operatorname{grad} f(x^*),\, x - x^* \rangle + \frac{\lambda^2}{2} \left\langle \nabla^2 f(x^*)\, h,\, h \right\rangle + \omega(\underbrace{x - x^*}_{=\lambda h}), \tag{$*$}$$

přičemž funkce $\omega(z)$ splňuje $\lim_{\|z\| \to 0} \frac{\omega(z)}{\|z\|^2} = 0$. Odtud dostáváme

$$\tfrac{1}{2} \left\langle \nabla^2 f(x^*)\, h,\, h \right\rangle + \frac{\omega(\lambda h)}{\lambda^2 \|h\|^2} \|h\|^2 \overset{(*)}{=} \frac{1}{\lambda^2} \Big[f(x) - f(x^*) - \langle \operatorname{grad} f(x^*),\, x - x^* \rangle \Big] \geqslant$$

$$\overset{(2.4.1)}{\geqslant} \frac{1}{\lambda^2} \vartheta \|x - x^*\|^2 = \vartheta \|h\|^2,$$

z čehož limitním přechodem pro $\lambda \to 0^+$ plyne

$$\left\langle \nabla^2 f(x^*)\, h,\, h \right\rangle \geqslant 2\vartheta \|h\|^2,$$

tj. (2.4.3) platí.

Důkaz Důsledku 2.4.5 (pokr.)

Je-li nyní $x^* \in X \setminus \operatorname{int} X$, pak $x^* \in \partial X$ a tedy (mohl by x^* být izolovaný bod?) existuje posloupnost $\{x_k\}_{k\in\mathbb{N}}$ taková, že $x_k \in \operatorname{int} X$ pro všechna $k \in \mathbb{N}$ a $x_k \to x^*$ pro $k \to \infty$. Pak z předchozí části plyne, že $\left\langle \nabla^2 f(x_k)\, h,\, h \right\rangle \geqslant 2\vartheta \|h\|^2$ pro každé $h \in \mathbb{R}^n$ a $k \in \mathbb{N}$. Poněvadž je však $\nabla^2 f(x)$ spojitá, můžeme „jít s limitou dovnitř", čímž dostaneme $\left\langle \nabla^2 f(x^*)\, h,\, h \right\rangle \geqslant 2\vartheta \|h\|^2$ i v tomto případě.

„$\Longleftarrow$" Nechť platí (2.4.3) pro všechna $x \in X$ a $h \in \mathbb{R}^n$. Nechť jsou $x, x^* \in X$ libovolná a položme $h := x - x^*$, tj. $x = x^* + h$. Pak s využitím Taylorova rozvoje druhého řádu dostaneme

$$f(x) - f(x^*) = \langle \operatorname{grad} f(x^*),\, x - x^* \rangle + \frac{1}{2} \left\langle \nabla^2 f\underbrace{(x^* + \lambda h)}_{\lambda\in(0,1) \Longrightarrow x^* + \lambda h \in X},\, h \right\rangle \geqslant$$

$$\overset{(2.4.3)}{\geqslant} \langle \operatorname{grad} f(x^*),\, x - x^* \rangle + \frac{1}{2} 2\vartheta \|h\|^2 = \langle \operatorname{grad} f(x^*),\, x - x^* \rangle + \vartheta \|h\|^2,$$

což s přihlédnutím k (2.4.1) ale znamená, že funkce f je silně konvexní na X s konstantou silné konvexnosti ϑ. ■

Poznámka

(i) Všimněte si, že pro důkaz implikace „$\Longleftarrow$" není požadavek $\operatorname{int} X \neq \emptyset$ potřebný.

(ii) Nicméně opačná implikace již bez tohoto požadavku nemusí být pravdivá. Uvažme např. ✎funkci $f(x, y) = x^2 - y^2$ na množině $X = \{[x, y] \in \mathbb{R}^2 \mid x \in \mathbb{R},\ y = 0\}$. Podmínku $\operatorname{int} X \neq \emptyset$ lze vypustit za cenu tu, že se omezíme pouze na $h \in \operatorname{Lin} X$. V literatuře bývá mnohdy předpoklad $\operatorname{int} X \neq \emptyset$ nahrazen silnější podmínkou vyžadující, že X je *otevřená* množina.

(iii) Z Důsledku 2.4.5 každopádně plynou následující tři implikace:
- $\nabla^2 f(x) \geqslant 0$ pro všechna $x \in X \Longrightarrow f$ je konvexní na X;
- $\nabla^2 f(x) > 0$ pro všechna $x \in X \Longrightarrow f$ je ostře konvexní na X;
- $\operatorname{int} X \neq \emptyset$ a f je konvexní na $X \Longrightarrow \nabla^2 f(x) \geqslant 0$ pro všechna $x \in X$.

(iv) Opačná implikace pro ostrou konvexnost?

✎ Příklady.

Poznámka

Je-li $A \in \mathbb{R}^{n\times n}$ symetrická matice, pak můžeme definovat tzv. Rayleighyho podíl $\rho_A : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}$ předpisem

$$\rho_A(h) := \frac{\langle Ah, h\rangle}{\langle h, h\rangle}.$$

(Ekvivalentně lze brát $\rho_A(h) = \langle Ah, h\rangle$ pro $\|h\| = 1$.) Pak platí tzv. *Min–Max věta:*

Má-li matice A vlastní čísla $\lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_n$, pak

$$\lambda_k = \min_{\dim S_k = n-k+1} \Big\{ \max_{h\in S_k\setminus\{0\}} \rho_A(h) \Big\},$$
$$\lambda_k = \max_{\dim S_k = k} \Big\{ \min_{h\in S_k\setminus\{0\}} \rho_A(h) \Big\}.$$

Zejména,

$$\lambda_{\max}(A) = \lambda_n(A) = \max_{h\in\mathbb{R}^n\setminus\{0\}} \rho_A(h),$$
$$\lambda_{\min}(A) = \lambda_1(A) = \min_{h\in\mathbb{R}^n\setminus\{0\}} \rho_A(h).$$

Poznámka

Důsledek 2.4.5 lze zformulovat také takto:

(i) Funkce f je konvexní na X splňující $\operatorname{int} X \neq \emptyset$ právě tehdy když, $\lambda_{\min}(\nabla^2 f(x)) \geqslant 0$ pro všechna $x \in X$;

(ii) Jestliže $\lambda_{\min}(\nabla^2 f(x)) > 0$ pro všechna $x \in X$, pak funkce f je ostře konvexní na X;

(iii) Funkce f je na X splňující $\operatorname{int} X \neq \emptyset$ silně konvexní s konstantou silné konvexnosti $\vartheta > 0$ právě tehdy, když $\lambda_{\min}(\nabla^2 f(x)) \geqslant 2\vartheta$ pro každé $x \in X$, tj. $\inf_{x\in X} \big(\lambda_{\min}(\nabla^2 f(x))\big) > 0$. A jaká bude největší možná hodnota konstanty ϑ?

Až doposud jsme pracovali s diferencovatelnými funkcemi. Z MAII víme, že tato vlastnost implikuje existenci *směrové derivace* v libovolném směru, avšak opačná implikace již neplatí. Nyní se proto podíváme na jednostranné směrové derivace ve směru vektoru h a v bodě x^*, čímž pro funkci $f : \mathbb{R}^n \to \mathbb{R}$, bod $x^* \in \mathbb{R}^n$ a vektor $h \in \mathbb{R}^n$ rozumíme

$$f'_h(x^*) \coloneqq \lim_{t \to 0+} \frac{f(x^* + th) - f(x^*)}{t}.$$

Věta 2.4.7

Nechť $X \subseteq \mathbb{R}^n$ je konvexní množina, funkce $f : X \to \mathbb{R}$ konvexní na X a $x^* \in \operatorname{ri} X$. Pak po všechna $h \in \operatorname{Lin} X$ existuje konečná $f'_h(x^*)$.

Tvrzení Věty 2.4.7 nemusí být pravdivé pro $x^* \in \operatorname{r\partial} X$. ✎Viz např. opět $X = [-1, 1]$ a

$$f(x) = \begin{cases} 0, & x \in (-1, 1), \\ 1, & x = \pm 1. \end{cases}$$

Pro řešení minimalizačních úloh s nediferencovatelnou funkcí f, tj. pro tzv. *nehladkou optimalizační úlohu*, jsou velmi užitečným nástrojem tzv. *subgradient* a *subdiferenciál* funkce f, které hrají podobně důležitou roli jako gradient v případě diferencovatelných funkcí. Motivace: pro diferencovatelnou konvexní funkci $f : X \to \mathbb{R}$ podle Věty 2.4.2 platí

$$f(x) \geqslant f(x^*) + \langle \operatorname{grad} f(x^*),\, x - x^* \rangle \quad \text{pro všechna } x, x^* \in X,$$

tj. $\operatorname{grad} f(x^*)$ udává „nevertikální“ opěrnou nadrovinu k množině $\operatorname{epi} f$ v bodě $[x^*, f(x^*)]$, viz poznámku za Větou 2.4.2. Ovšem co kdyby f nebyla diferencovatelná?

Definice 2.5.1

Nechť $X \subseteq \mathbb{R}^n$ je konvexní množina. Vektor $a \in \mathbb{R}^n$ se nazývá <u>subgradient</u> funkce $f : X \to \mathbb{R}$ v bodě $x^* \in X$, jestliže

$$f(x) - f(x^*) \geqslant \langle a,\, x - x^* \rangle \tag{2.5.1}$$

pro každé $x \in X$. Množina všech subgradientů funkce f v bodě x^* se nazývá <u>subdiferenciál</u> funkce f v bodě x^* a značí se $\partial f(x^*)$. Funkce f se nazývá <u>subdiferencovatelná</u> v bodě x^*, jestliže $\partial f(x^*) \neq \emptyset$.

Poznámka

(i) Jistě platí $\operatorname{grad} f(x^*) \in \partial f(x^*)$ díky Větě 2.4.2. Více? Viz Větu 2.5.7.

(ii) (zobecnění Fermatovy věty) Pro funkci $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ platí

$$f(x^*) = \min_{x \in X} f(x)$$

právě tehdy, když $0 \in \partial f(x^*)$.

(iii) Je-li množina $X \subseteq \mathbb{R}^n$ konvexní, funkce $f_1, \dots, f_m : X \to \mathbb{R}$ konvexní na X a $\alpha_1, \dots, \alpha_m \geqslant 0$, pak

$$\partial\big(\alpha_1 f_1(x^*) + \dots + \alpha_m f_m(x^*)\big) = \alpha_1\, \partial f_1(x^*) + \dots + \alpha_m\, \partial f_m(x^*),$$

tj. „je to lineární“.

Příklady.

Poznámka

Je-li $f : X \subseteq \mathbb{R} \to \mathbb{R}$ konvexní a $x^* \in \operatorname{ri} X$, pak podle Věty 2.4.7 a následné poznámky existují jednostranné derivace $f'_+(x^*)$ a $f'_-(x^*)$, přičemž platí $f'_-(x^*) \leqslant f'_+(x^*)$. V tomto případě pak máme $\partial f(x^*) = [f'_-(x^*), f'_+(x^*)]$.

Poznámka

Podmínka (2.5.1) (podobně jako (2.4.1) ve Větě 2.4.2) znamená, že graf funkce f

$$G_f := \{[x, \beta] \in \mathbb{R}^{n+1} \mid \beta = f(x) \text{ pro } x \in X\}$$

neleží pod grafem

$$H := \{[x, \beta] \in \mathbb{R}^{n+1} \mid \beta = \ell(x) \text{ pro } x \in \mathbb{R}^n\}$$

afinní funkce $\ell(x) := f(x^*) + \langle a, x - x^* \rangle$. Současně $H = H_{p,\beta}$, kde $H_{p,\beta}$ je nadrovina s $p := (-a^\top, 1)^\top \in \mathbb{R}^{n+1}$ a $\beta := f(x^*) - \langle a, x^* \rangle$, je opěrnou nadrovinou k nadgrafu funkce f v bodě $[x^*, f(x^*)]$. Neboli

$$a \in \partial f(x^*) \text{ právě tehdy, když } H_{p,\beta} \text{ je opěrná nadrovina.}$$

Věta 2.5.4

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

(i) Je-li funkce f konvexní a $x^* \in \operatorname{ri} X$, pak $\partial f(x^*)$ je neprázdná, uzavřená a konvexní množina.

(ii) Je-li $\partial f(x)$ neprázdná pro každé $x \in X$, pak f je konvexní na X.

Poznámka

(i) Tvrzení ve Větě 2.5.4(i) nelze obrátit, tj. nahradit ve Větě 2.5.4(ii) „každé $x \in X$“ za „každé $x \in \operatorname{ri} X$“. Např.

$$X = [0,1] \times [0,1] \subseteq \mathbb{R}^2 \quad \& \quad f(x,y) = \begin{cases} 0, & 0 \leqslant x \leqslant 1,\ 0 < y \leqslant 1, \\ \frac{1}{2} - |x - \frac{1}{2}|, & 0 \leqslant x \leqslant 1,\ y = 0. \end{cases}$$

Pak funkce f má jediný subgradient pro každé $x \in \operatorname{int} X = \operatorname{ri} X$ (vskutku?), ale funkce f není konvexní (vskutku?).

(ii) Požadavek $x^* \in \operatorname{ri} X$ je pro neprázdnost $\partial f(x^*)$ klíčový. ✎Uvažme např. $X = [-1, 1]$ a $f(x) = -\sqrt{1 - x^2}$.

(iii) ✎Ostrá konvexnost?

(iv) Dá se ukázat, že pro konvexní funkci je $\partial f(x^*)$ dokonce kompaktní množina.

Důkaz Věty 2.5.4

(i) Uzavřenost plyne přímo z definice stejně jako konvexnost, neboť v případě alespoň dvoubodové množiny $\partial f(x^*)$ máme pro $a, b \in \partial f(x^*)$ nerovnosti

$$f(x) - f(x^*) \geqslant \langle a, x - x^* \rangle, \quad f(x) - f(x^*) \geqslant \langle b, x - x^* \rangle .$$

Pak pro libovolné $\lambda \in [0, 1]$ dostaneme sečtením λ-násobku první nerovnosti a $(1-\lambda)$-násobku druhé nerovnosti vztah

$$\underbrace{\lambda \big(f(x) - f(x^*)\big) + (1-\lambda) \big(f(x) - f(x^*)\big)}_{f(x)-f(x^*)} \geqslant \underbrace{\lambda \langle a, x - x^* \rangle + (1-\lambda) \langle b, x - x^* \rangle}_{\langle \lambda a + (1-\lambda) b, x - x^* \rangle},$$

tj. $\lambda a + (1-\lambda) b \in \partial f(x^*)$. Zbývá ukázat, že $\partial f(x^*) \neq \emptyset$. Pro jednoduchost se omezíme pouze na případ $x^* \in \operatorname{int} X$ (situace $x^* \in \operatorname{ri} X$ je technicky náročnější). Protože $[x^*, f(x^*)] \in \partial(\operatorname{epi} f)$ a současně $\operatorname{epi} f$ je konvexní množina, můžeme využít Větu 2.3.7 k existenci opěrné nadroviny k $\operatorname{epi} f$ v bodě $[x^*, f(x^*)]$, tj. existuje vektor $\big(p^\top, \beta\big) \in \mathbb{R}^{n+1} \setminus \{0\}$ s $a \in \mathbb{R}^n$ a $\beta \in \mathbb{R}$ takový, že

$$\Big\langle \big(\begin{smallmatrix} p \\ \beta \end{smallmatrix}\big), \big(\begin{smallmatrix} x \\ y \end{smallmatrix}\big) \Big\rangle \leqslant \Big\langle \big(\begin{smallmatrix} p \\ \beta \end{smallmatrix}\big), \big(\begin{smallmatrix} x^* \\ f(x^*) \end{smallmatrix}\big) \Big\rangle \tag{2.5.2}$$

pro všechna $\big(x^\top, y\big)^\top \in \operatorname{epi} f$, kde $x \in X$ a $y \in \mathbb{R}$. Nerovnost (2.5.2) je ekvivalentní s

$$\langle p, x \rangle + \beta y \leqslant \langle p, x^* \rangle + \beta f(x^*) \tag{2.5.3}$$

pro každé $x \in X$ a $y \geqslant f(x)$.

Důkaz Věty 2.5.4 (pokr.)

Ukážeme, že $\beta < 0$. Připusťme, že $\beta > 0$. Pak výraz na pravé straně (2.5.3) je konstantní, takže v tomto případě snadno obdržíme spor v (2.5.3), pokud vezmeme y dostatečně velké.
Jestliže $\beta = 0$, pak $p \neq 0$ a (2.5.3) se redukuje na

$$\langle p, x \rangle \leqslant \langle p, x^* \rangle \tag{2.5.4}$$

pro všechna $x \in X$. Ovšem $x^* \in \operatorname{int} X$, takže existuje $\varepsilon > 0$ takové, že $x_1 := x^* + \varepsilon p \in X$. Dosazením x_1 do (2.5.4) dostaneme

$$\langle p, x^* \rangle + \varepsilon \langle p, p \rangle \leqslant \langle p, x^* \rangle,$$

tj. $\varepsilon \|p\|^2 \leqslant 0$. Ovšem toto je možné pouze pro $\|p\| = 0$ neboli $p = 0$, což je spor. Tedy nutně $\beta < 0$. Pokud (2.5.3) vydělíme číslem β a nahradíme-li $y = f(x)$, máme

$$\tfrac{1}{\beta} \langle p, x \rangle + f(x) \geqslant \tfrac{1}{\beta} \langle p, x^* \rangle + f(x^*)$$

neboli

$$f(x) - f(x^*) \geqslant \Big\langle \underbrace{\tfrac{1}{\beta} p}_{=:a}, x - x^* \Big\rangle \tag{2.5.5}$$

pro každé $x \in X$. Pak tedy $a := \frac{1}{\beta} p \in \partial f(x^*)$, neboť pak je (2.5.5) tvaru (2.5.1).

Důkaz Věty 2.5.4 (pokr.)

(ii) Jestliže X je prázdná nebo jednobodová, tak tvrzení je splněno triviálně. Nechť tedy X je alespoň dvoubodová a zvolme $x_1, x_2 \in X$ a $\lambda \in [0,1]$ libovolně. Položme $x^* := \lambda x_1 + (1-\lambda) x_2 \in X$. Podle předpokladů existuje $a \in \partial f(x^*)$, takže

$$f(x_1) - f(x^*) \geqslant \langle a,\, x_1 - x^* \rangle \quad \& \quad f(x_2) - f(x^*) \geqslant \langle a,\, x_2 - x^* \rangle\,.$$

Sečteme-li λ-násobek první nerovnosti a $(1-\lambda)$-násobek druhé nerovnosti dostaneme

$$\lambda f(x_1) + (1-\lambda) f(x_2) - (\lambda + 1 - \lambda) f(x^*) \geqslant \lambda \langle a,\, x_1 - x^* \rangle + (1-\lambda) \langle a,\, x_2 - x^* \rangle = 0,$$

tj. $\lambda f(x_1) + (1-\lambda) f(x_2) \geqslant f(x^*) = f(\lambda x_1 + (1-\lambda) x_2)$, což dokazuje konvexnost funkce f na X. ■

Subdiferenciál je také velmi úzce spojen s jednostrannou směrovou derivací. Je-li $x^* \in \operatorname{ri} X$, pak nerovnost (2.5.1) je ekvivalentní s

$$\frac{f(x^* + th) - f(x^*)}{t} \geqslant \langle a,\, h \rangle$$

pro všechna $h \in \operatorname{Lin} X$ a $t > 0$ dostatečně malé ($x^* \in \operatorname{ri} X \;\Rightarrow\; x^* + th \in X$). Pak podíl na levé straně konverguje pro $t \to 0^+$ k $f'_h(x^*)$, viz Větu 2.4.7, takže nerovnost (2.5.1) je ekvivalentní s $f'_h(x^*) \geqslant \langle a,\, h \rangle$. Tedy platí

$$a \in \partial f(x^*) \quad \Longleftrightarrow \quad f'_h(x^*) \geqslant \langle a,\, h \rangle \quad \text{pro všechna } h \in \operatorname{Lin} X,$$

takže nutně také

$$f'_h(x^*) \geqslant \max_{a \in \partial f(x^*)} \langle a,\, h \rangle\,. \tag{2.5.6}$$

Jenže ono platí ještě více.

Věta 2.5.6 Nechť $X \subseteq \mathbb{R}^n$ je konvexní množina, funkce $f : X \to \mathbb{R}$ konvexní na X a $x^* \in \operatorname{ri} X$. Pak platí

$$\partial f(x^*) = \left\{ a \in \mathbb{R}^n \mid \langle a,\, h \rangle \leqslant f'_h(x^*) \text{ pro všechna } h \in \operatorname{Lin} X \right\} \tag{2.5.7}$$

a pro každé $h \in \operatorname{Lin} X$ máme

$$f'_h(x^*) = \max_{a \in \partial f(x^*)} \langle a,\, h \rangle\,. \tag{2.5.8}$$

Na závěr si ještě upřesníme vztah mezi gradientem a subgradientem (viz první část poznámky za Definicí 2.5.1).

Věta 2.5.7

Nechť $X \subseteq \mathbb{R}^n$ je konvexní množina, funkce $f : X \to \mathbb{R}$ konvexní a $x^* \in \operatorname{int} X$.

(i) Je-li f diferencovatelná v x^*, pak $\partial f(x^*)$ je jednoprvková a platí $\partial f(x^*) = \{\operatorname{grad} f(x^*)\}$, tj. gradient je jediným subgradientem.

(ii) Je-li $\partial f(x^*) = \{a\}$ jednoprvková množina, pak je funkce f diferencovatelná v bodě x^* a platí $\operatorname{grad} f(x^*) = a$.

Důkaz Věty 2.5.7

(i) Jestliže existuje $x^* \in \operatorname{int} X$, pak nutně $\operatorname{Lin} X = \mathbb{R}^n$. Již víme, že $\operatorname{grad} f(x^*) \in \partial f(x^*)$. Nechť $a \in \partial f(x^*)$ je libovolný vektor. Protože funkce f je diferencovatelná, pro libovolné $h \in \operatorname{Lin} X$ platí

$$f'_h(x^*) = \langle \operatorname{grad} f(x^*),\, h\rangle\,,$$

což dohromady s (2.5.6) dává

$$\langle a,\, h\rangle \leqslant f'_h(x^*) = \langle \operatorname{grad} f(x^*),\, h\rangle\,.$$

Proto pro každé $h \in \operatorname{Lin} X$ je tedy

$$\langle \operatorname{grad} f(x^*) - a,\, h\rangle \geqslant 0,$$

což při volbě $h = a - \operatorname{grad} f(x^*) \in \operatorname{Lin} X = \mathbb{R}^n$ dává $\|\operatorname{grad} f(x^*) - a\|^2 \leqslant 0$, a tedy nutně $a = \operatorname{grad} f(x^*)$.

(ii) Jestliže $x^* \in \operatorname{int} X$, existuje $r > 0$ takové, že $x^* + B(x^*, r) \subset \operatorname{int} X$. Uvažme nyní funkci

$$\phi(t,h) := \frac{f(x^* + th) - f(x^*)}{t} - \langle a,\, h\rangle\,, \qquad (*)$$

kde $t \in (0, 1]$ a $h \in B(x^*, r)$. Jelikož $\partial f(x^*) = \{a\}$, plyne z (2.5.8), že $f'_h(x^*) = \langle a,\, h\rangle$. Pak podle (důkazu) Věty 2.4.7 pro jakékoli $h \in B(x^*, r)$ je funkce $\phi(\cdot, h)$ nerostoucí na $(0, 1]$ a konverguje k 0 pro $t \to 0^+$. Současně (vzhledem k Větě 2.4.1) je funkce $\phi(t, h)$ spojitá pro pevné $t \in (0, 1]$ vzhledem k h z kompaktní množiny $B(x^*, r)$.

Důkaz Věty 2.5.7 (pokr.)

To znamená, že konvergence $\varphi(t,h) \to 0$ je stejnoměrná díky Diniho větě ($\rightsquigarrow \{f_n\}$ monotónně rostoucí posloupnost, tj. $f_n \leqslant f_{n+1}$, spojitých funkcí, které konvergují bodově ke spojité funkci $\Rightarrow$ konvergence je stejnoměrná), tj. pro každé $\varepsilon > 0$ existuje $\delta \in (0, 1]$ takové, že

$$0 \leqslant \varphi(t,h) \leqslant \varepsilon$$

pro všechna $t \in (0,\delta)$ a pevně zvolené $h \in B(x^*, r)$. Pro libovolné $h \in B(x^*, \delta r)$ položme $\tilde{t} := \|h\|/r$ a $\tilde{h} := rh/\|h\|$. Potom $\tilde{t} \in (0,\delta)$ a $\tilde{h} \in B(x^*, r)$, takže

$$0 \leqslant \varphi(\tilde{t},\tilde{h}) \leqslant \varepsilon$$

neboli podle $(*)$ máme

$$0 \leqslant \frac{f(x^*+h)-f(x^*)}{\|h\|/r} - \left\langle a,\, \tfrac{h}{\|h\|/r} \right\rangle \leqslant \varepsilon$$

neboli

$$0 \leqslant \frac{f(x^*+h)-f(x^*)-\langle a,\, h\rangle}{\|h\|} \leqslant \varepsilon/r.$$

Odtud pro $\|h\| \to 0$ plyne

$$\lim_{\|h\|\to 0} \frac{f(x^*+h)-f(x^*)-\langle a,\, h\rangle}{\|h\|} = 0,$$

a tedy funkce f je diferencovatelná v x^* s diferenciálem $\mathrm{d}f_h(x^*) = \langle a,\, h\rangle$, a tudíž $a = \operatorname{grad} f(x^*)$. ■

V posledním tématu konvexní analýzy se podíváme transformaci, která dané funkci $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ přiřadí konvexní funkci $f^* : \mathbb{R}^n \to \mathbb{R}$.

Při studiu úloh matematického programování se často stává, že je užitečnější/výhodnější se věnovat jinému (podobnému) problému, který je přirozeně přidružen původní úloze. Toto je tzv. *duální problém*, kterému se budeme podrobněji věnovat v jedné části následující+1 kapitole. V jejím závěru si naznačíme roli Fenchelovy transformace v této oblasti.

Motivace: uvažme libovolnou funkci $f : \mathbb{R}^n \to \mathbb{R}$ a množinu všech afinních funkcí h, pro které je f majorantou (neboli h minorizují f), tj.

$$h(x) \leqslant f(x)$$

pro všechna $x \in \mathbb{R}^n$. Jelikož funkce $h(x)$ jsou tvaru $h(x) = \langle y, x\rangle - \alpha$ pro nějaké $y \in \mathbb{R}^n$ a $\alpha \in \mathbb{R}$, požadujeme

$$\langle y, x\rangle - \alpha \leqslant f(x) \quad \text{pro všechna } x \in \mathbb{R}^n$$

neboli

$$\alpha \geqslant \langle y, x\rangle - f(x) \quad \text{pro všechna } x \in \mathbb{R}^n,$$

takže při pevně zvoleném $y \in \mathbb{R}^n$ musíme nutně brát

$$\alpha = \sup\{\langle y, x\rangle - f(x) \mid x \in \mathbb{R}^n\},$$

viz Definici 2.6.7 a Větu 2.6.8.

Definice 2.6.1

Nechť $f : \mathbb{R}^n \to \mathbb{R}$. Funkce

$$f^*(y) := \sup_{x \in \mathbb{R}^n} [\langle x, y\rangle - f(x)] \tag{2.6.1}$$

se nazývá Fenchelovou transformací (též /konvexně/ konjugovanou funkcí) funkce f.

Poznámka

(i) Je zřejmé, že $f^* : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\} \rightsquigarrow$ tzv. efektivní definiční obor $D^*(f) = \{x \in D(f) \mid f(x) < \infty\}$.

(ii) Fenchelovu transformaci jsme zavedli pro funkce definované na celém $\mathbb{R}^n$. Pokud $X = D(f) \subsetneqq \mathbb{R}^n$, tak můžeme upravit (2.6.1) tak, že budeme brát supremum pouze přes $x \in X$. Druhou možností je dodefinování funkce f na $\mathbb{R}^n \setminus X$ jako $f(x) = \infty$. Potom pro takto dodefinovanou funkci je hodnota suprema přes $\mathbb{R}^n$ totožná s hodnotou suprema pouze přes X (vskutku?).

Poznámka (pokr.)

(iii) Uvažme diferencovatelnou funkci f a y takové, že hodnota suprema v (2.6.1) je menší než ∞. Pak nutná podmínka pro extrém je

$$\frac{\partial}{\partial x_i}\big[\langle x, y\rangle - f(x)\big] = 0, \quad i \in \{1, \dots, n\},$$

tedy musí platit $y = \operatorname{\mathbf{grad}} f(x)$. Je-li navíc stacionární bod x^* funkce $\langle x, y\rangle - f(x)$ zároveň i bodem maxima této funkce (kdy se to může stát?), pak platí

$$f^*(y) = \langle x^*, y\rangle - f(x^*),$$

přičemž $y = \operatorname{\mathbf{grad}} f(x^*)$. V takovém případě je Fenchelova transformace známa jako Legendreova transformace.

(iv) Z Definice 2.6.1 očividně plyne

$$f^*(0) = -\inf_{x\in X} f(x).$$

(v) Ekonomická interpretace $f^*(y)$.

Příklad.

Geometrická ilustrace $f^*(y)$.

V předchozím příkladu platilo $f = f^{**}$ — náhoda? Uvidíme…

Lemma 2.6.3

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

(i) Funkce f^* je konvexní na množině $Y := \{y \in \mathbb{R}^n \mid f^*(y) < \infty\}$.

(ii) Pro každé $x \in X$ a $y \in \mathbb{R}^n$ platí tzv. *Fenchelova(–Youngova) nerovnost*

$$f(x) + f^*(y) \geqslant \langle x, y\rangle,$$

přičemž rovnost nastane právě tehdy, když $y \in \partial f(x)$.

(iii) Je-li $f(x) \geqslant g(x)$ na X, pak $f^*(y) \leqslant g^*(y)$ pro všechna $y \in \mathbb{R}^n$.

Poznámka

Aplikováním Fenchelovy nerovnosti na funkci $f(x) = \frac{1}{\alpha}\sum_{i=1}^n |x_i|^\alpha$ dostaneme

$$\langle x, y\rangle \leqslant \frac{\sum_{i=1}^n |x_i|^\alpha}{\alpha} + \frac{\sum_{i=1}^n |y_i|^\beta}{\beta}$$

pro $\beta \in \mathbb{R}$ splňující $1/\alpha + 1/\beta = 1$. Zejména v případě $n = 1$ obdržíme tzv. *Youngovu nerovnost*

$$xy \leqslant \frac{|x|^\alpha}{\alpha} + \frac{|y|^\beta}{\beta}.$$

Poznámka

Z Lemma 2.6.3(ii) navíc vyplývá:

(i) Je-li $\partial f(x) \neq \emptyset$, pak $f(x) = f^{**}(x)$ pro každé $x \in X$. Naopak, je-li $f(x) = f^{**}(x)$ pro $x \in X$, pak $\partial f(x) = \partial f^{**}(x)$.

(ii) Je-li $y \in \partial f(x)$, pak $x \in \partial f^*(y)$. Je-li navíc $f(x) = f^{**}(x)$, pak $y \in \partial f(x)$ právě tehdy, když $x \in \partial f^*(y)$.

Lemma 2.6.5

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

(i) Je-li $a \in \mathbb{R} \backslash \{0\}$ a

$$\widetilde{f}(x) := \begin{cases} f(ax), & ax \in X, \\ \infty, & \text{jinak}, \end{cases}$$

potom platí

$$\widetilde{f}^*(y) = f^*(y/a).$$

Lemma 2.6.5 (pokr.)

(ii) Je-li $b \in \mathbb{R}^n$ a

$$\widetilde{f}(x) := \begin{cases} f(x+b), & x+b \in X, \\ \infty, & \text{jinak}, \end{cases}$$

potom platí

$$\widetilde{f}^*(y) = f^*(y) - \langle b, y \rangle .$$

(iii) Pro každé $\alpha > 0$ platí

$$(\alpha f)^*(y) = \alpha f^*(y/\alpha).$$

(iv) Pro každé $\gamma \in \mathbb{R}$ platí

$$(f+\gamma)^*(y) = f^*(y) - \gamma.$$

(v) Je-li $g(x) := \langle a, x \rangle$ pro $a \in \mathbb{R}^n$, potom

$$(f+g)^*(y) = f^*(y-a).$$

Důkaz Lemma 2.6.5

Jednotlivá tvrzení plynou přímo z Definice 2.6.1 a vlastností suprema.

(i) Platí
$$\widetilde{f}^*(y) = \sup_{x\in\mathbb{R}^n} \{\langle x, y\rangle - g(x)\} = \sup_{x\in\mathbb{R}^n} \{\langle x, y\rangle - f(ax)\} =$$
$$= \sup_{x\in\mathbb{R}^n} \{\langle ax, y/a\rangle - f(ax)\} = f^*(y/a).$$

(ii) Platí
$$\widetilde{f}^*(y) = \sup_{x\in\mathbb{R}^n} \{\langle x, y\rangle - g(x)\} = \sup_{x\in\mathbb{R}^n} \{\langle x, y\rangle - f(x+b)\} =$$
$$= \sup_{x\in\mathbb{R}^n} \{\langle x+b, y\rangle - f(x+b) - \langle b, y\rangle\} =$$
$$= \sup_{x\in\mathbb{R}^n} \{\langle x+b, y\rangle - f(x+b)\} - \langle b, y\rangle = f^*(y) - \langle b, y\rangle.$$

(iii) Platí
$$(\alpha f)^*(y) = \sup_{x\in\mathbb{R}^n} \{\langle x, y\rangle - \alpha f(x)\} = \alpha\{\langle x, y/\alpha\rangle - f(x)\} = \alpha f^*(y/\alpha).$$

(iv) Platí
$$(f+c)^*(y) = \sup_{x\in\mathbb{R}^n} \{\langle x, y\rangle - (f+c)(x)\} = \sup_{x\in\mathbb{R}^n} \{\langle x, y\rangle - f(x)\} - c = f^*(y) - c.$$

(v) Platí
$$(f+g)^*(y) = \sup_{x\in\mathbb{R}^n} \{\langle x, y\rangle - f(x) - \langle a, x\rangle\} = \sup_{x\in\mathbb{R}^n} \{\langle x, y-a\rangle - f(x)\} = f^*(y-a).$$
■

✎ Obrázek.

V důkazu prvního tvrzení v Poznámce za Lemma 2.6.3 jsme ukázali, že pro každé $x \in \mathbb{R}^n$ platí
$$f(x) \geqslant f^{**}(x).$$

A opačná nerovnost?

Věta 2.6.6 (Fenchel & Moreau)

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

Důkaz – nemáme dostatečné nástroje...

Poznámka

Zkombinováním Věty 2.6.6 a druhého tvrzení v Poznámce za Lemma 2.6.3 snadno zjistíme, že pro každou konvexní funkci $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ a libovolné $x \in \operatorname{ri} X$ platí
$$y \in \partial f(x) \quad \text{právě tehdy, když} \quad x \in \partial f^*(y).$$

✎ A Fenchelovy transformace „vyššího řádu“?

Na závěr ještě jeden pojem, který dává dohromady konvexní funkce a Fenchelovu transformaci.

Definice 2.6.7

Nechť $X \subseteq \mathbb{R}^n$ je konvexní množina a $f : X \to \mathbb{R}$. Potom funkce

$$g(x) := \sup\left\{h(x) \mid h \text{ je konvexní a } h(x) \leqslant f(x) \text{ pro každé } x \in X\right\}$$

se nazývá konvexní obal (obálka) funkce f a značí se $\operatorname{co} f$.

Jinými slovy, $\operatorname{co} f$ je největší konvexní funkce, která je majorizována funkcí f ($\operatorname{co} f$ pro konvexní funkci?).

Známe-li $\operatorname{epi} f$ pro funkci $f : X \to \mathbb{R}$, pak ji můžeme „obnovit“, neboť

$$f(x) = \inf\left\{\alpha \in \mathbb{R} \mid x \in X,\ [x, \alpha] \in \operatorname{epi} f\right\},$$

přičemž konvexnost $\operatorname{epi} f$ zaručuje konvexnost f. A jak získat $\operatorname{co} f$?

Věta 2.6.8

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

$$\operatorname{co} f(x) = \inf\left\{\alpha \in \mathbb{R} \mid [x, \alpha] \in \operatorname{conv}(\operatorname{epi} f)\right\},$$

a tedy

$$\operatorname{co} f(x) = \inf\left\{\sum_{i=1}^{n+1} \lambda_i f(x_i) \mid x_i \in X,\ \sum_{i=1}^{n+1} \lambda_i x_i = x,\ \lambda_i \geqslant 0,\ \sum_{i=1}^{n+1} \lambda_i = 1\right\} \quad (*)$$

Příklady.

Poznámka

Platí

$$D(\operatorname{co} f) = \operatorname{conv}(D(f)), \tag{2.6.2}$$

což dává $\operatorname{conv}(\operatorname{epi} f) \subseteq \operatorname{epi}(\operatorname{co} f)$. Vlastní podmnožiny?

Předchozí věta nám sice dává bližší objasnění $\operatorname{co} f$, ovšem z praktického hlediska stále není příliš užitečná. Jak vlastně spočítat $\operatorname{co} f$? V $(*)$ dostáváme minimalizační problém, který ale již v případě funkce jedné proměnné obsahuje čtyři proměnné $x_1, x_2 \in X$ a $\lambda_1, \lambda_2 \in [0, 1]$ a dvě omezení

$$\lambda_1 x_1 + \lambda_2 x_2 = x \quad \& \quad \lambda_1 + \lambda_2 = 1.$$

Věta 2.6.9

Nechť $X \subseteq \mathbb{R}^n$ je konvexní množina a $f : X \to \mathbb{R}$. Potom pro každé $x \in \operatorname{ri} X$ platí

$$\operatorname{co} f(x) = f^{**}(x).$$

Důkaz Věty 2.6.9

Z Fenchelovy nerovnosti plyne

$$f(x) \geqslant \langle x, y\rangle - f^*(y)$$

pro všechna $x, y \in \mathbb{R}^n$ (přičemž pro $x \notin D(f)$ a $y \notin D(f^*)$ klademe $f(x) = \infty = f(y)$), a tedy

$$f(x) \geqslant \sup_{y \in \mathbb{R}^n} \{\langle x, y\rangle - f^*(y)\} = f^{**}(y).$$

Proto (s využitím Věty 2.6.8) platí $\operatorname{co} f(x) \geqslant \operatorname{co} f^{**}(x)$. Jenže podle Lemma 2.6.3(i) je f^{**} konvexní, tj. $\operatorname{co} f^{**}(x) = f^{**}(x)$, což celkem dává $\operatorname{co} f(x) \geqslant f^{**}(x)$.
Nyní ukážeme opačnou nerovnost. Funkce $\operatorname{co} f(x)$ je konvexní, a tedy podle Věty 2.4.1 je spojitá pro každé $x \in \operatorname{ri} X$. Pak podle Věty 2.6.6 platí $(\operatorname{co} f)^{**}(x) = \operatorname{co} f(x)$ pro každé $x \in \operatorname{ri} X$. Navíc podle definice platí $\operatorname{co} f(x) \leqslant f(x)$ pro každé $x \in X$, což implikuje $(\operatorname{co} f)^{**}(x) \leqslant f^{**}(x)$, viz Lemma 2.6.3(iii). Proto v každém bodě $x \in \operatorname{ri} X$ platí

$$\operatorname{co} f(x) = (\operatorname{co} f)^{**}(x) \leqslant f^{**}(x),$$

což dohromady s první částí dává požadované tvrzení. ■

Konec.