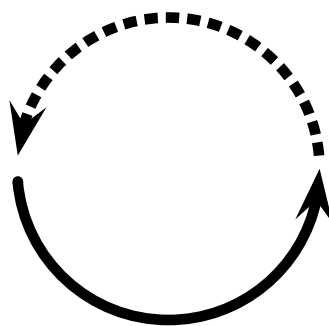


Milan Hladík

# Diskrétní a Spojitá Optimalizace

text k přednášce

17. května 2022





Tento text zatím slouží jako pomocný text přednášky Diskrétní a spojitá optimalizace na studiu informatiky MFF UK v Praze, a to pro část spojité optimalizace.

Autor čerpal především z knih Bazaraa et al. [2006]; Boyd and Vandenberghe [2004]; Luenberger and Ye [2008].

Případné připomínky a chyby zasílejte prosím na adresu [hladik@kam.mff.cuni.cz](mailto:hladik@kam.mff.cuni.cz).



# Obsah

<b>Obsah</b>	<b>5</b>
<b>1 Úvod</b>	<b>7</b>
1.1 Motivační úlohy . . . . .	8
1.2 Spojitá optimalizace: První pojmy, klasifikace a úpravy . . . . .	9
1.3 Lineární regrese . . . . .	12
<b>2 Optimalizace bez omezení</b>	<b>17</b>
<b>3 Konvexita</b>	<b>19</b>
3.1 Konvexní množiny . . . . .	19
3.2 Konvexní funkce . . . . .	21
3.3 Charakterizace konvexity funkce prvního a druhého řádu . . . . .	23
3.4 Další pravidla pro zjišťování konvexity funkce . . . . .	25
<b>4 Konvexní optimalizace</b>	<b>27</b>
4.1 Základní vlastnosti . . . . .	27
4.2 Kvadratické programování . . . . .	29
4.3 Kuželové programování . . . . .	30
4.3.1 Dualita v kuželovém programování . . . . .	32
4.3.2 Kuželové kvadratické programování . . . . .	35
4.3.3 Semidefinitní programování . . . . .	35
4.4 Výpočetní složitost . . . . .	36
4.4.1 Dobré zprávy – elipsoidová metoda . . . . .	37
4.4.2 Špatné zprávy – kópositivní programování . . . . .	37
4.5 Aplikace . . . . .	38
4.5.1 Robustní PCA . . . . .	38
4.5.2 Minimální opsaný elipsoid . . . . .	39
<b>5 Karush–Kuhn–Tuckerovy podmínky optimality</b>	<b>41</b>
<b>6 Metody</b>	<b>47</b>
6.1 Line search . . . . .	47
6.2 Úlohy bez omezení . . . . .	49
6.3 Úlohy s omezeními . . . . .	50
6.3.1 Metody přípustných směrů . . . . .	51
6.3.2 Metody aktivních podmínek . . . . .	51
6.3.3 Penalizační a bariérové metody . . . . .	52
6.4 Metoda sdružených gradientů . . . . .	54
<b>7 Vybraná témata</b>	<b>57</b>
7.1 Robustní optimalizace . . . . .	57
7.2 Konkávní programování . . . . .	60

Příloha	61
Značení	63
Literatura	65

# Kapitola 1

## Úvod

Úlohou optimalizace (matematického programování) se rozumí úloha

$$\min f(x) \text{ za podm. } x \in M,$$

kde  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  je *účelová funkce* a  $M \subseteq \mathbb{R}^n$  je *množina přípustných řešení*. Obecně je tato úloha nerozhodnutelná, tedy neexistuje algoritmus, který by ji řešil [Zhu, 2006], podobně jako známý *halting problém*. Na druhou stranu, existují třídy úloh, které jsou řešitelné a efektivně.

V závislosti na charakteru množiny přípustných řešení  $M$  rozlišujeme dva základní typy úloh:

- **diskrétní optimalizace**

Množina  $M$  je konečná, nicméně stále dost velká na to, abychom prozkoumaly všechna přípustná řešení. Typicky  $|M| \geq 2^n$ .

Typickými úlohami jsou hledání nejkratší cesty, minimální kostry či minimální párování v grafu. Tyto úlohy jsou efektivně řešitelné. Na druhou stranu, některé úlohy jsou NP-těžké: celočíselné lineární programování, nebo jeho speciální podúlohy jako je problém obchodního cestujícího, problém batohu či maximální řez v grafu.

- **spojitá optimalizace**

Zde předpokládáme, že množina  $M$  je nekonečná, dokonce nespočetná. To kupodivu může někdy ulehčit práci: lineární programování je efektivně řešitelné, ale s podmínkami na celočíselnost je rázem NP-těžké.

Typickými úlohami jsou lineární programování a různé typy nelineárního programování (konvexní, kvadratické aj.)

### Vztah diskrétní a spojité optimalizace

Diskrétní a spojité optimalizace nejsou od sebe disjunktně odděleny, jak by se mohlo na první pohled zdát. Naopak, techniky a znalosti z jednoho oboru se promítají do druhého. To je nejlépe vidět na *celočíslném programování*: metody na řešení jsou často založeny na iterativním řešení relaxovaných spojitých úloh.

Na druhou stranu, celočíselnou podmínku snadno převedeme na spojitou. Například podmínku  $x \in \{0, 1\}$  můžeme ekvivalentně přepsat jako  $x = x^2$  (v praxi to ale vesměs výhodné nebývá).

Vztah obou oborů ilustrujeme na příkladu toků v sítích.

**Příklad 1.1** (Toky v sítích). Mějme orientovaný grafu  $G = (V, E)$ , kde  $s \in V$  je počáteční vrchol a  $t \in V$  koncový. Dále mějme kapacity hran  $u: E \mapsto \mathbb{R}^+$ . Cílem je najít maximální tok od  $s$  do  $t$ . Pro každý vrchol musí platit podmínka rovnováhy (přítok = odtok).

Přidáme-li do grafu symbolickou hranu  $(t, s)$ , pak úlohu maximálního toku můžeme formulovat elegantně jako hledání maximálního toku skrze přidanou hranu

$$\begin{aligned} \max x_{ts} \text{ za podm. } & \sum_{j:(i,j) \in E} x_{ij} - \sum_{j:(j,i) \in E} x_{ji} = 0, \forall i \in V \\ & 0 \leq x_{ij} \leq u_{ij}, \forall (i, j) \in E, \end{aligned}$$

což je úloha lineárního programování. Je-li  $A$  matice incidence grafu  $G$ , pak tato úloha má maticový tvar

$$\max x_{t,s} \text{ za podm. } Ax = 0, 0 \leq x \leq u.$$

Nejlepší známé algoritmy na řešení této úlohy využívají diskrétní (grafové) povahy problému. Na druhou stranu, formulace pomocí lineárního programu má také řadu výhod. Protože matice  $A$  je totálně unimodulární, výsledné řešení je vždy celočíselné, pokud kapacity jsou celočíselné. Proto je úloha efektivně řešitelná pomocí lineárního programování i s podmínkami celočíselnosti.

Forma lineárního programu má i tu výhodu, že je snadno modifikovatelná pro různé varianty úlohy. Předpokládejme variantu, kdy hledáme nejlevnější toku (*minimum-cost flow*). Známe  $c_{ij}$ , cenu přepravy jednotky zboží po hraně  $(i, j) \in E$ , a minimální požadovaný tok  $d > 0$ . Úloha lze pak formulovat jako lineární program

$$\min \sum_{(i,j) \in E} c_{ij} x_{ij} \text{ za podm. } Ax = 0, 0 \leq x \leq u, x_{ts} \geq d. \quad \square$$

## 1.1 Motivační úlohy

**Příklad 1.2** (Teorie – vlastní čísla). Buď  $A \in \mathbb{R}^{n \times n}$  symetrická a  $\lambda_1 \geq \dots \geq \lambda_n$  její (reálná) vlastní čísla. Uvažujme jednotkovou kouli v prostoru  $\mathbb{R}^n$ , která je popsána  $B := \{x \in \mathbb{R}^n; \|x\|_2 \leq 1\}$ . Pak největší vlastní číslo  $\lambda_1$  se nabyde jako největší hodnota kvadratické formy  $x^T A x$  na kouli  $B$ , a nejmenší vlastní číslo  $\lambda_n$  se nabyde jako nejmenší hodnota kvadratické formy  $x^T A x$  na  $B$ . Formálně:

$$\lambda_1 = \max_{x: \|x\|_2 \leq 1} x^T A x, \quad \lambda_n = \min_{x: \|x\|_2 \leq 1} x^T A x.$$

Toto je znění *Rayleigh–Ritzovy věty*. Důkaz pro  $\lambda_1$ :

Nerovnost „ $\leq$ “: Buď  $x_1$  vlastní vektor příslušný k  $\lambda_1$  normovaný  $\|x_1\|_2 = 1$ . Pak  $Ax_1 = \lambda_1 x_1$ . Přenásobením  $x_1^T$  zleva dostaneme

$$\lambda_1 = \lambda_1 x_1^T x_1 = x_1^T A x_1 \leq \max_{x: \|x\|_2 = 1} x^T A x.$$

Nerovnost „ $\geq$ “: Buď  $x \in \mathbb{R}^n$  libovolný vektor takový, že  $\|x\|_2 = 1$ . Buď  $A = Q\Lambda Q^T$  spektrální rozklad matice  $A$ . Označme  $y := Q^T x$ , pak  $\|y\|_2 = 1$  a dostaneme:

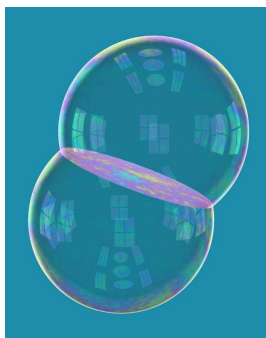
$$x^T A x = x^T Q\Lambda Q^T x = y^T \Lambda y = \sum_{i=1}^n \lambda_i y_i^2 \leq \sum_{i=1}^n \lambda_1 y_i^2 = \lambda_1 \|y\|_2^2 = \lambda_1. \quad \square$$

**Příklad 1.3** (Funkcionální optimalizace). Principiálně nemusí být konečný počet proměnných. V tom případě se typicky jedná o funkcionální úlohu, kde chceme najít funkci, splňující daná omezení a která je v jistém kritériu nejlepší. Pro ilustraci si představme plánování dráhy rakety se Země na Mars; proměnnou je zde tvar dráhy popsáný funkcí a kritérium je minimalizace času přepravy. Jednoduché funkcionální úlohy lze vyřešit analyticky, ale obecně se často neznámá funkce diskretizuje a aplikují techniky klasické optimalizace.

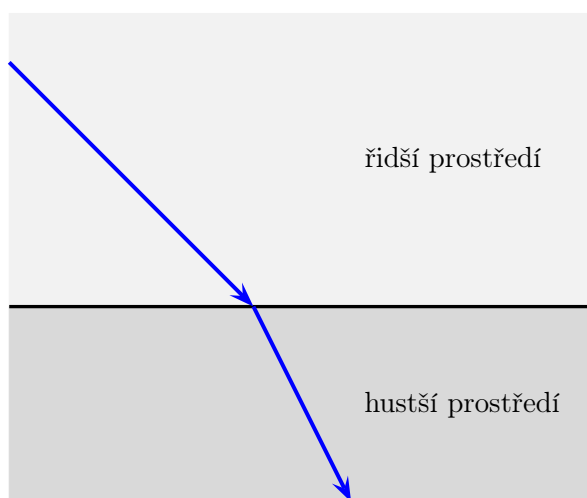
Do této třídy úloh patří i izoperimetrické úlohy. Je všeobecně známo, že při daném objemu má ze všech geometrických těles minimální povrch koule. Jak je to ale se dvěma objemy, pokud minimalizujeme celkový povrch (sjednocení plus dotyk)? Tato úloha je známá jako *double bubble* a vyřešil jí až Hutchings et al. [2002]. Výsledný optimální tvar je složen ze dvou částí koulí, dotýkajících se v úhlu  $120^\circ$ . Koule se dotýkají také ve tvaru sféry; ve speciálním případě dvou rovných objemů se jedná o kruh. Viz obrázek 1.1.  $\square$

**Příklad 1.4** (Když příroda optimalizuje). Snellův zákon říká, jak se světlo lomí, když přechází z jednoho prostředí do druhého. Optimalizační pohled na věc je ten, že paprsek světla volí nejrychlejší trajektorii. V hustším prostředí pohybuje pomaleji a v řidším naopak rychleji, takže výsledná lomená trajektorie je taková, aby paprsek od zdroje k cíli dorazil co nejrychleji (tzv. Fermatův princip). Viz obrázek 1.2.  $\square$





Obrázek 1.1: (Example 1.3) Úloha „The double bubble problem“.



Obrázek 1.2: (příklad 1.4) Snellův zákon a Fermatův princip.

## 1.2 Spojitá optimalizace: První pojmy, klasifikace a úpravy

### Lokální a globální minima

Řešení lze kategorizovat na několik typů, viz obrázek 1.3. Bod  $x^* \in M$  se nazývá:

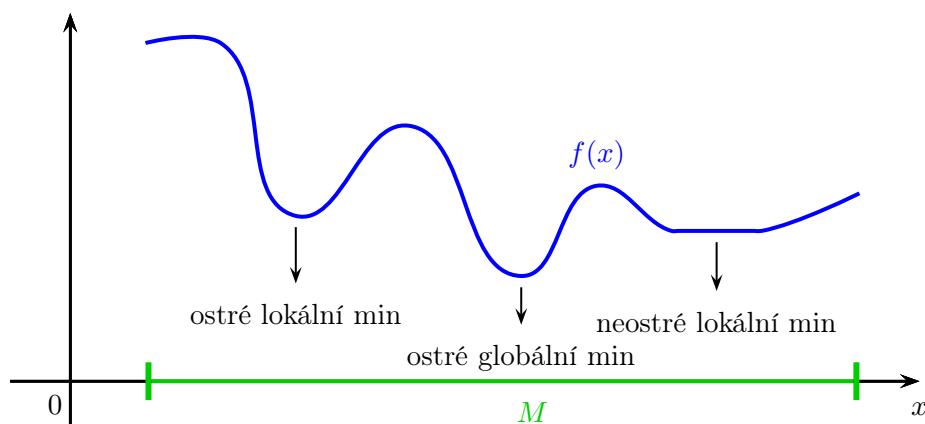
- *globální minimum* pokud  $f(x^*) \leq f(x)$  pro všechna  $x \in M$ ,
- *ostré (globální) minimum* pokud  $f(x^*) < f(x)$  pro všechna  $x^* \neq x \in M$ ,
- *lokální minimum* pokud  $f(x^*) \leq f(x)$  pro všechna  $x \in M \cap \mathcal{O}_\varepsilon(x^*)$ ,
- *ostré lokální minimum* pokud  $f(x^*) < f(x)$  pro všechna  $x^* \neq x \in M \cap \mathcal{O}_\varepsilon(x^*)$ .

Minimum funkce  $f(x)$  se na množině  $M$  vůbec nemusí nabýt. Uvažme například úlohu  $\min_{x \in \mathbb{R}} x$ , která je zdola neomezená, nebo úlohu  $\min_{x \in \mathbb{R}} e^x$ , která je zdola omezená. Postačující podmínku pro existenci minima podává známá Weierstrassova věta.

**Věta 1.5** (Weierstrass). *Je-li funkce  $f(x)$  spojitá a množina  $M$  kompaktní, pak minimum se nabyde.*

I když se globální minimum nabyde, nemáme ještě vyhráno díky problému lokálních minim. Základní metody na řešení optimalizačních úloh jsou iterativní. Začnou v počátečním bodě a postupují ve směru klesající účelové funkce. Tímto způsobem snadno uvíznou v lokálním minimu. V lineárním programování, nebo obecněji v konvexní optimalizaci, tento problém nenastává, protože každé lokální minimum je automaticky globálním minimem (viz věta 3.15).

Poznamenejme, že pojem lokálního minima lze adaptovat i v diskrétní optimalizaci. Například v úloze minimální kostry můžeme okolí definovat jako ty kostry, které se od zadané liší v nanejvýš jedné hraně.



Obrázek 1.3: Lokální a globální minima.

## Klasifikace

Množina  $M$  je často popsána soustavou

$$\begin{aligned} g_j(x) &\leq 0, & j &= 1, \dots, J, \\ h_\ell(x) &= 0, & \ell &= 1, \dots, L, \end{aligned}$$

kde  $g_j(x), h_\ell(x): \mathbb{R}^n \rightarrow \mathbb{R}$ . Zkráceně budeme psát omezení ve tvaru

$$g(x) \leq 0, \quad h(x) = 0,$$

kde  $g: \mathbb{R}^n \rightarrow \mathbb{R}^J$  a  $h: \mathbb{R}^n \rightarrow \mathbb{R}^L$ . Podle typu účelové funkce a množiny přípustných řešení rozdělujeme obecnou úlohu na několik tříd:

- *Lineární programování.* Funkce  $f(x)$ ,  $g_j(x)$ ,  $h_\ell(x)$  jsou lineární. Předpokládáme, že čtenář má základní znalosti o lineárním programování.
- *Optimalizace bez omezení.* Zde  $M = \mathbb{R}^n$ .
- *Konvexní programování.* Funkce  $f(x)$ ,  $g_j(x)$  jsou konvexní a  $h_\ell(x)$  jsou lineární.

## Základní úpravy

Pokud hledáme maximum funkce  $f(x)$  na množině  $M$ , tak úlohu snadno převedeme na minimalizační, neboť

$$\max_{x \in M} f(x) = - \min_{x \in M} -f(x).$$

Rovnicová omezení můžeme převést na nerovnosti, neboť  $h(x) = 0$  je ekvivalentní s  $h(x) \leq 0$ ,  $h(x) \geq 0$ , obvykle to ale pro výpočetní postupy výhodné není.

**Transformace funkcí.** Úlohu

$$\min f(x) \text{ za podm. } g(x) \leq 0, \quad h(x) = 0$$

můžeme ekvivalentně přepsat na úlohu

$$\min \varphi(f(x)) \text{ za podm. } \psi(g(x)) \leq 0, \quad \eta(h(x)) = 0,$$

pokud

- $\varphi(z)$  je rostoucí funkce na daném definičním oboru, např.  $z^k$ ,  $z^{1/k}$ ,  $\log(z)$ ;

- $\psi(z)$  zachovává nekladnost, tj.  $z \leq 0 \Leftrightarrow \psi(z) \leq 0$ , např.  $z^3$ ;
- $\eta(z)$  zachovává nulovost, tj.  $z = 0 \Leftrightarrow \eta(z) = 0$ , např.  $z^2$ .

Ekvivalence úloh zde znamená, že obě úlohy mají stejná optima. Optimální hodnoty se ale liší.

**Příklad 1.6** (Geometrické programování). Výše zmíněný postup se s výhodou použije například v geometrickém programování. Pro ilustraci uvažujme konkrétní příklad

$$\min x^2y \text{ za podm. } 5xy^3 \leq 1, 7x^{-3}y \leq 1, x, y > 0.$$

Zlogaritmováním funkcí dostaneme úlohu

$$\min 2\log(x) + \log(y) \text{ za podm. } \log(5) + \log(x) + 3\log(y) \leq 0, \log(7) - 3\log(x) + \log(y) \leq 0.$$

Substituce  $x' := \log(x)$ ,  $y' := \log(y)$  pak vede na úlohu lineárního programování

$$\min 2x' + y' \text{ za podm. } \log(5) + x' + 3y' \leq 0, \log(7) - 3x' + y' \leq 0. \quad \square$$

**Účelová funkce do omezení.** Častá transformace je přepsání účelové funkce do omezení, tj. úlohu  $\min_{x \in M} f(x)$  přepíšeme jako

$$\min z \text{ za podm. } f(x) \leq z, x \in M.$$

Účelová funkce je nyní lineární a veškeré „obtíže“ jsou skryté v popisu omezení.

**Příklad 1.7** (Konečný minimax). Uvažujme úlohu

$$\min_{x \in M} \max_{i=1, \dots, s} f_i(x).$$

Obecně úlohy minimaxového typu jsou velice obtížné. Vzhledem k tomu, že vnější funkce je maximum na konečně množině, můžeme úlohu přepsat jako

$$\min z \text{ za podm. } f_i(x) \leq z, i = 1, \dots, s, x \in M.$$

Tento přepis má ještě další výhodu, protože v původní úloze účelová funkce  $\max_{i=1, \dots, s} f_i(x)$  nebyla hladká.  $\square$

Poznamenejme, že někdy je výhodný i druhý směr, tj. omezení přepsat do účelové funkce, viz sekce 6.3.3.

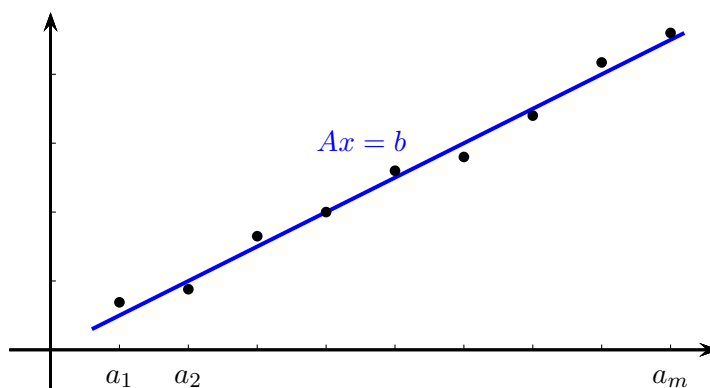
**Eliminace rovnic a proměnných.** Uvažujme úlohu typu

$$\min f(x) \text{ za podm. } g(x) \leq 0, Ax = b,$$

kde  $A \in \mathbb{R}^{m \times n}$  má lineárně nezávislé řádky. Postup nyní spočívá ve vyřešení soustavy rovnic  $Ax = b$ . Předpokládejme, že množina řešení je neprázdná. Pak je rovna  $x^0 + \text{Ker}(A)$ , kde  $x^0$  je jedno libovolné řešení a  $\text{Ker}(A)$  je jádro matice. Sestavíme-li matici  $B \in \mathbb{R}^{n \times (n-m)}$  po sloupcích z báze  $\text{Ker}(A)$ , pak řešení soustavy  $Ax = b$  mají popis  $x = x^0 + Bz$ , kde  $z \in \mathbb{R}^{n-m}$ . Substitucí za  $x$  nyní optimalizační úloha získá tvar

$$\min f(x^0 + Bz) \text{ za podm. } g(x^0 + Bz) \leq 0.$$

Tímto způsobem jsme eliminovali rovnice a snížili dimenzi problému (počet proměnných) o  $m$ .



Obrázek 1.4: Lineární regrese.

### 1.3 Lineární regrese

Úloha lineární regrese spočívá v hledání lineární závislosti v datech  $(a_1, b_1), \dots, (a_m, b_m) \in \mathbb{R}^{n+1}$ , viz obrázek 1.4. Lineární regrese se používá napříč obory, od ekonomie, biologie až po informatiku – uvažme například strojové učení pro rozpoznávání vzorků v datech za účelem predikování a automatického rozhodování (filtrování spamu, doporučování knih/videí, rozpoznávání obličejů na snímcích, zneužití kreditních karet). Samozřejmě, že skutečná závislost nemusí být lineární a existují modely pro nelineární regrese, my se ale omezíme jen na lineární model.

Sestavme matici  $A \in \mathbb{R}^{m \times n}$  s řádky  $a_1, \dots, a_m$ . Pak úlohu lze formulovat jako hledání vektoru  $x \in \mathbb{R}^n$  tak, aby  $Ax \approx b$ . Protože typicky  $m \gg n$ , tak soustava rovnic  $Ax = b$  je přeuredená a nemá řešení. Proto se musíme spokojit s přibližným řešením.

Matematicky můžeme problém nalezení vhodného řešení vyjádřit jako optimalizační úlohu hledání  $x \in \mathbb{R}^n$  takového, aby rozdíl mezi levou a pravou stranou byl v jisté normě co nejmenší:

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|.$$

Geometrická interpretace této úlohy je projekce vektoru  $b \in \mathbb{R}^m$  v prostoru  $\mathbb{R}^m$  do sloupcového prostoru  $\mathcal{S}(A)$  matice  $A$ . Možné volby normy:

- *Eukleidovská.* Úloha je pak ekvivalentní  $\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2 = \min_{x \in \mathbb{R}^n} \sum_{i=1}^m (A_{i*}x - b_i)^2$ , čili jedná se o metodu nejmenších čtverců. Pokud jsou sloupce matice  $A$  lineárně nezávislé, máme jednoznačné řešení  $x^* = (A^T A)^{-1} A^T b$ . Tento přístup má dobrou statistickou motivaci. Pokud předpokládáme, že závislost je skutečně lineární a hodnoty ve vektoru  $b$  jsou zatíženy nezávisle s chybou s normálním rozdělením, pak  $x^*$  je nejlepší nestranný odhad a také nejvěrohodnější odhad.

- *Součtová.* Úloha  $\min_{x \in \mathbb{R}^n} \|Ax - b\|_1$  se dá přepsat jako lineární program

$$\min e^T z \quad \text{za podm.} \quad -z \leq Ax - b \leq z, \quad z \in \mathbb{R}^m, \quad x \in \mathbb{R}^n.$$

I tento případ má statistickou interpretaci. Optimum úlohy dává nejvěrohodnější odhad, pokud chyby mají Laplaceovo rozdělení.

- *Maximová.* Úloha  $\min_{x \in \mathbb{R}^n} \|Ax - b\|_\infty$  se dá také přepsat jako lineární program

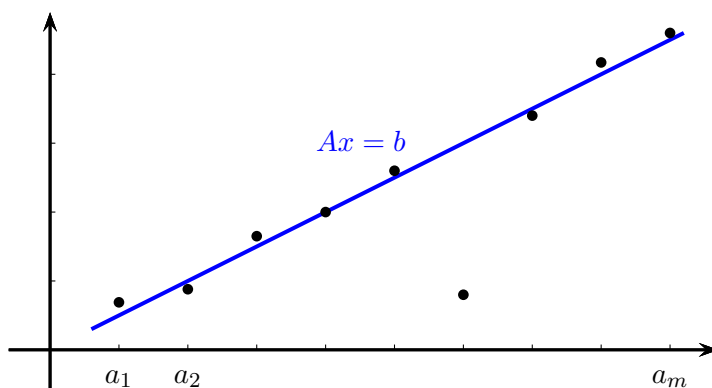
$$\min z \quad \text{za podm.} \quad -ze \leq Ax - b \leq ze, \quad z \in \mathbb{R}, \quad x \in \mathbb{R}^n.$$

**Outliery.** Outlier je význačná odchylka v pozorování, typicky způsobená chybou, viz obrázek 1.5. Outlier tak „kazí“ lineární tendenci a můžeme dostat zkreslený výsledek. Z výše uvedených norem je ta součtová relativně méně citlivá na outliery, ale i tak mohou outliery působit potíže.

Pokud tušíme, že by v datech mohlo být obsaženo  $k \ll m$  outlierů, pak můžeme řešit úlohu takto:

$$\min \|A_I x - b_I\| \quad \text{za podm.} \quad x \in \mathbb{R}^n, \quad I \subseteq \{1, \dots, m\}, \quad |I| \geq m - k,$$

kde  $A_I, b_I$  reprezentuje podmatice indexované množinou  $I$ . Dostali jsme ale velice těžký kombinatorický optimalizační problém.



Obrázek 1.5: Lineární regrese s outlierem.

**Kardinalita.** Kardinalitou vektoru  $x \in \mathbb{R}^n$  rozumíme počet jeho nenulových složek, a značíme

$$\|x\|_0 = |\{i; x_i \neq 0\}|.$$

Toto značení připomíná vektorovou  $\ell_p$ -normu  $\|x\|_p = \sqrt[p]{\sum_{i=1}^n |x_i|^p}$ . Skutečně, kardinalitu dostaneme limitním přechodem pokud zanedbáme odmocninu:

$$\|x\|_0 = \lim_{p \rightarrow 0^+} \sum_{i=1}^n |x_i|^p.$$

Nicméně,  $\|x\|_0$  již vektorovou normu netvoří.

V regresi typicky chceme vysvětlit  $b$  pomocí malého počtu regresorů, tedy chceme zároveň minimalizovat  $\|x\|_0$ . Dvě kritéria spojíme v jedno pomocí váženého součtu, což vede na formulaci

$$\min \|Ax - b\|_2 + \gamma \|x\|_0,$$

kde  $\gamma > 0$  je zvolená konstanta. To je opět těžký kombinatorický problém. Proto se  $\|x\|_0$  aproximuje součtovou normou (dá se ukázat, že to je v jistém smyslu nejlepší aproximace). Tím dostaneme efektivně řešitelnou úlohu

$$\min \|Ax - b\|_2 + \gamma \|x\|_1.$$

**Příklad 1.8** (Rekonstrukce signálu). Uvažujme problém rekonstrukce signálu. Nechť vektor  $\tilde{x} \in \mathbb{R}^n$  značí neznámý signál, a nechť  $y = \tilde{x} + \text{err}$  značí naměřený signál se šumem. Chceme zašuměný signál vyhladit a tedy najít dobrou aproximaci  $\tilde{x}$ . Budeme tedy hledat takový vektor  $x \in \mathbb{R}^n$ , aby byl dost blízko  $y$  a zároveň aby byl odpovídající signál vyhlazený, tj. nedocházelo k velkým skokům mezi sousedními hodnotami.

To vede přirozeně na více-kriteriální úlohu

$$\min_{x \in \mathbb{R}^n} \|x - y\|_2, |x_{i+1} - x_i| \forall i.$$

Váženým součtem kritérií dostaneme skalarizaci

$$\min_{x \in \mathbb{R}^n} \|x - y\|_2 + \gamma \sum_{i=1}^{n-1} |x_{i+1} - x_i|,$$

kde  $\gamma > 0$  je daný parametr. Menší hodnota  $\gamma$  upřednostňuje první kritérium, tedy signál je blíže naměřenému, zatímco větší  $\gamma$  vede na vyhlazenější signál.

Zavedeme-li diferenční matici  $D \in \mathbb{R}^{(n-1) \times n}$  s prvky  $D_{ii} = 1$ ,  $D_{i,i+1} = -1$  a s nulami jinak, tak úloha lze přepsat jako

$$\min_{x \in \mathbb{R}^n} \|x - y\|_2 + \gamma \|Dx\|_1.$$

To lze opět nahlížet jako na aproximaci úlohy kardinality

$$\min_{x \in \mathbb{R}^n} \|x - y\|_2 + \gamma \|Dx\|_0,$$

v níž hledáme aproximaci signálu ve tvaru po částech konstantní funkce. Tento přístup se nazývá *total variation reconstruction*, a najde uplatnění hlavně u digitálních signálů.

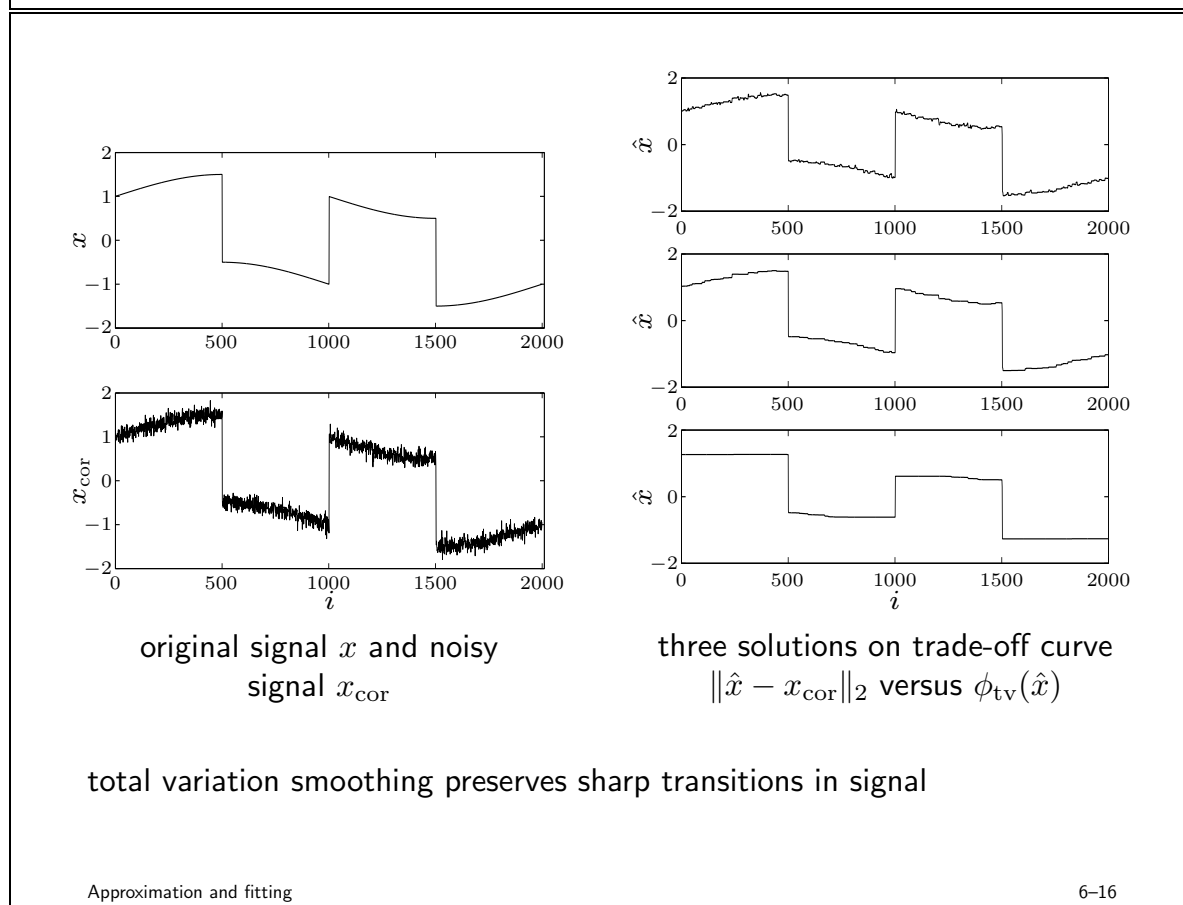
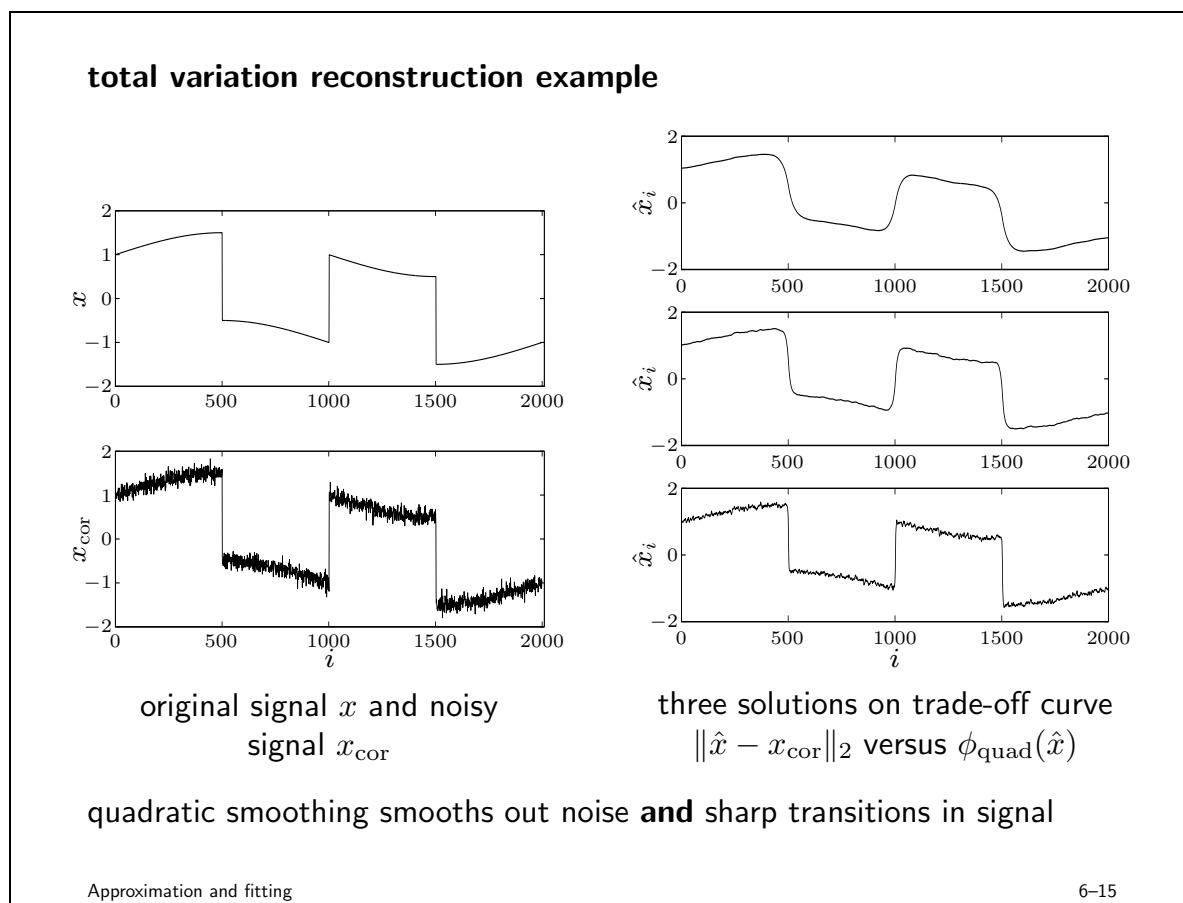
Grafické porovnání je k dispozici na obrázku 1.6, který pochází z webových stránek:

<http://stanford.edu/class/ee364a/lectures/approx.pdf>

Podobný přístup se používá i pro analýzu a zpracování obrázků, pro problémy odstranění neostrotí a rozmazanosti, rekonstrukci poničených obrázků atp. Viz webové stránky

<http://www.imm.dtu.dk/~pcha/mxTV/>

□



Obrázek 1.6: Příklad 1.8: U obou obrázků vlevo původní signál a dole se šumem. Napravo očištěný signál se zmenšující se vahou  $\gamma$ . V prvním obrázku nahoře použití kvadratického vyhlazování (tj.,  $\|Dx\|_2$  namísto  $\|Dx\|_1$ ), dole total variation reconstruction, která lépe aproximuje digitální signál.





## Kapitola 2

# Optimalizace bez omezení

Optimalizace bez omezení řeší úlohu

$$\min f(x) \text{ za podm. } x \in \mathbb{R}^n.$$

Na účelovou funkci  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  případně doplníme určité požadavky na diferencovatelnost.

Nejprve zmíníme známou nutnou podmínku optimality prvního řádu.

**Věta 2.1** (Nutná podmínka optimality prvního řádu). *Bud'  $f(x)$  diferencovatelná a bud'  $x^* \in \mathbb{R}^n$  lokální extrém. Pak  $\nabla f(x^*) = o$ .*

*Důkaz.* Bez újmy na obecnosti necht'  $x^*$  je lokální minimum. Připomeňme, že pro libovolné  $i = 1, \dots, n$  je

$$\nabla_i f(x) = \frac{\partial f(x^*)}{\partial x_i} = \lim_{h \rightarrow 0} \frac{f(x_1^*, \dots, x_{i-1}^*, x_i^* + h, x_{i+1}^*, \dots, x_n^*) - f(x^*)}{h}.$$

Limita se musí rovnat limitě zprava i zleva. V prvním případě dostáváme

$$\nabla_i f(x) = \lim_{h \rightarrow 0^+} \frac{f(x_1^*, \dots, x_{i-1}^*, x_i^* + h, x_{i+1}^*, \dots, x_n^*) - f(x^*)}{h} \geq 0,$$

a v druhém analogicky  $\nabla_i f(x) \leq 0$ . Tudíž  $\nabla_i f(x) = 0$ . □

Je jasné, že podmínka z věty je pouze nutnou pro optimalitu, protože neumí rozeznat minimum od maxima, nemluvě od inflexního bodu, viz obrázek 2.1. Bod s nulovým gradientem se nazývá *stacionární bod*.

Podmínky optimality druhého řádu máme dvě, nutnou a postačující.

**Věta 2.2** (Nutná podmínka optimality druhého řádu). *Bud'  $f(x)$  spojitě dvakrát diferencovatelná a bud'  $x^* \in M$  lokální minimum. Pak Hessián  $\nabla^2 f(x^*)$  je pozitivně semidefinitní.*

*Důkaz.* Ze spojitosti druhých parciálních derivací pro každé  $\lambda \in \mathbb{R}$  a  $y \in \mathbb{R}^n$  existuje  $\theta \in (0, 1)$  tak, že

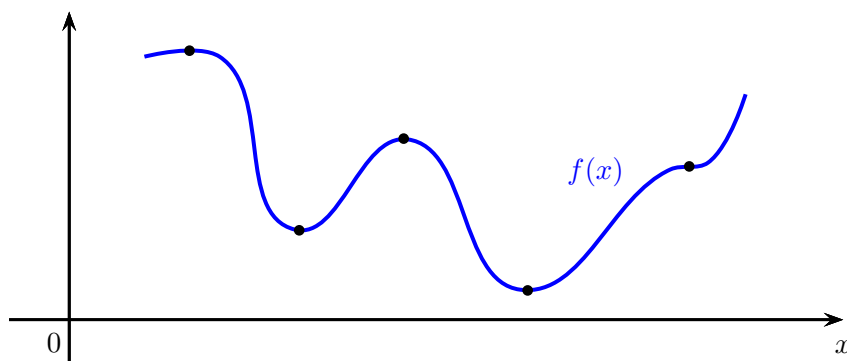
$$f(x^* + \lambda y) = f(x^*) + \lambda \nabla f(x^*)^T y + \frac{1}{2} \lambda^2 y^T \nabla^2 f(x^* + \theta \lambda y) y. \quad (2.1)$$

Toto je v zásadě vyjádření Taylorova rozvoje s Lagrangeovým tvarem zbytku. Z minimality  $x^*$  je  $f(x^* + \lambda y) \geq f(x^*)$  a dle věty 2.1 je  $\nabla f(x^*) = o$ . Tudíž

$$\lambda^2 y^T \nabla^2 f(x^* + \theta \lambda y) y \geq 0.$$

Limitním přechodem  $\lambda \rightarrow 0$  dostaneme  $y^T \nabla^2 f(x^*) y \geq 0$ . □

**Věta 2.3** (Postačující podmínka optimality druhého řádu). *Bud'  $f(x)$  spojitě dvakrát diferencovatelná. Je-li  $\nabla f(x^*) = o$  a  $\nabla^2 f(x^*)$  je pozitivně definitní pro určité  $x^* \in M$ , pak  $x^*$  je ostré lokální minimum.*



Obrázek 2.1: Stacionární body funkce  $f(x)$  jsou jak lokální minima a maxima, tak i inflexní bod.

*Důkaz.* Jako v důkazu věty 2.2, v rovnici (2.1) je pro  $\lambda \neq 0$ ,  $y \neq 0$  a dost malé  $\theta$

$$\lambda \nabla f(x^*)^T y = 0, \quad \frac{1}{2} \lambda^2 y^T \nabla^2 f(x^* + \theta \lambda y) y > 0.$$

Tudíž  $f(x^* + \lambda y) > f(x^*)$ . □

Vidíme, že rozdíl mezi nutnou a postačující podmínkou je celkem těsný. Přesto je nenulový, což dosvědčuje například funkce  $f(x) = -x^4$ . V bodě  $x = 0$  se nabývá ostré lokální maximum a nikoliv minimum, a přesto postačující podmínka (pochopitelně) selže, a nutná podmínka uspěje.

**Příklad 2.4** (Metoda nejmenších čtverců). Uvažujme soustavu  $Ax = b$ , kde  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$  a matice  $A$  má hodnost  $n$  (srov. sekce 1.3). Typicky  $m$  je mnohem větší než  $n$ . Protože soustava nemusí mít řešení, hledáme přibližné řešení optimalizační úlohou

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2.$$

Zde je snaha najít takový vektor  $x$ , aby chyba mezi levou a pravou stranou byla v eukleidovské normě co nejmenší. Protože je druhá mocnina rostoucí funkce, optimum se nabyde ve stejném bodě jako pro úlohu

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2 = (Ax - b)^T (Ax - b) = x^T A^T A x - 2b^T A x + b^T b.$$

Zkusme prověřit podmínky věty 2.3. Gradient účelové funkce je roven  $2A^T A x - 2A^T b$  (viz příloha na str. 61). Pokud má být nulový, dostáváme podmínku  $A^T A x = A^T b$ , z čehož  $x = (A^T A)^{-1} A^T b$ . Hessián účelové funkce je  $2A^T A$ , což je pozitivně definitní matice. Tudíž je bod  $x = (A^T A)^{-1} A^T b$  ostrým lokálním minimem. Protože je účelová funkce konvexní (vyplyne později, viz věta 4.4), je tento bod ve skutečnosti globálním minimem.

Pokud by matice  $A$  neměla plnou sloupcovou hodnost, pak kandidátem na optimum je jakékoliv řešení rovnice  $A^T A x = A^T b$ . Dá se ukázat, že řešení rovnice je nekonečně mnoho a všechna tvoří optimum naší úlohy. □

# Kapitola 3

## Konvexita

Konvexní množiny a funkce se začaly intenzivně studovat před více než 100 lety a mezi průkopníky patřily osobnosti jako Hölder (1889), Jensen (1906) či Minkowski (1910).

### 3.1 Konvexní množiny

**Definice 3.1.** Množina  $M \subseteq \mathbb{R}^n$  je *konvexní*, pokud pro každé  $x_1, x_2 \in M$  a každé  $\lambda_1, \lambda_2 \geq 0$ ,  $\lambda_1 + \lambda_2 = 1$ , je konvexní kombinace  $\lambda_1 x_1 + \lambda_2 x_2 \in M$ .

Příklad:  $\emptyset$  nebo  $\{x\}$  jsou konvexní.

Geometricky konvexita říká, že s každými dvěma body obsahuje množina  $M$  i celou úsečku spojující tyto body. Pro úsečku s konci  $x_1, x_2$  budeme používat značení

$$u(x_1, x_2) := \{x \in \mathbb{R}^n; x = \lambda_1 x_1 + \lambda_2 x_2, \lambda_1, \lambda_2 \geq 0, \lambda_1 + \lambda_2 = 1\}.$$

Ekvivalentně můžeme konvexity množiny charakterizovat pomocí konvexních kombinací všech  $k$ -tic bodů.

**Věta 3.2.** *Bud'  $k \geq 2$ . Množina  $M \subseteq \mathbb{R}^n$  je konvexní právě tehdy, když pro libovolné  $x_1, \dots, x_k \in M$  a  $\lambda_1, \dots, \lambda_k \geq 0$ ,  $\sum_{i=1}^k \lambda_i = 1$ , platí  $\sum_{i=1}^k \lambda_i x_i \in M$ .*

*Důkaz.* Za cvičení. □

Evidentně, konvexita není uzavřená na sjednocení. Na druhou stranu, je uzavřená na libovolné průniky.

**Věta 3.3.** *Jsou-li  $M_i \subseteq \mathbb{R}^n$ ,  $i \in I$  konvexní, pak také  $\bigcap_{i \in I} M_i$  je konvexní.*

*Důkaz.* Bud'  $x_1, x_2 \in \bigcap_{i \in I} M_i$ . Pak pro každé  $i \in I$  je  $x_1, x_2 \in M_i$ , a tedy i konvexní kombinace  $\lambda_1 x_1 + \lambda_2 x_2 \in M_i$ . □

Tato věta nás opravňuje zavést konvexní obal množiny  $M$  jako nejmenší (co do inkluze) konvexní množinu obsahující  $M$ .

**Definice 3.4.** *Konvexním obalem množiny  $M \subseteq \mathbb{R}^n$  je průnik všech konvexních množin v  $\mathbb{R}^n$  obsahujících  $M$ . Značíme  $\text{conv}(M)$ .*

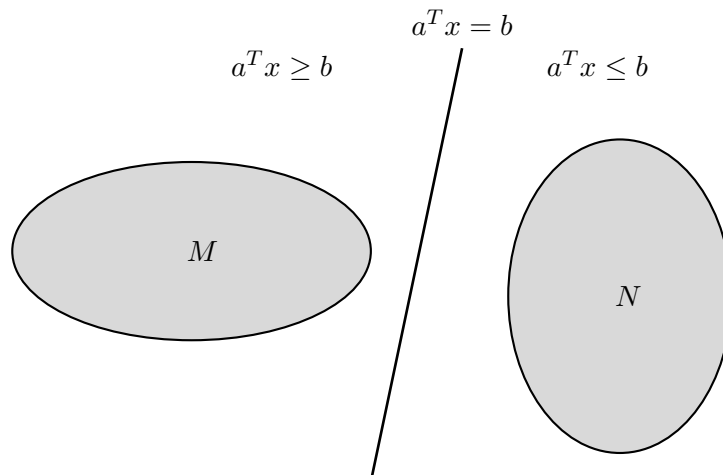
Nyní máme další charakterizaci konvexity množiny.

**Věta 3.5.** *Množina  $M \subseteq \mathbb{R}^n$  je konvexní právě tehdy, když  $M = \text{conv}(M)$ .*

*Důkaz.* „ $\Rightarrow$ “ Protože je  $M$  konvexní, je to jedna z množin se kterými děláme průnik z definice  $\text{conv}(M)$ .

„ $\Leftarrow$ “ Protože  $\text{conv}(M)$  je konvexní dle věty 3.3, je  $M$  také. □

Připomeňme, že *relativní vnitřek* množiny  $M \subseteq \mathbb{R}^n$  je vnitřek  $M$  pokud se omezíme na nejmenší afinní podprostor obsahující  $M$ . Značíme  $\text{ri}(M)$ .



Obrázek 3.1: Oddělitelnost množin  $M$  a  $N$  pomocí nadroviny  $a^T x = b$ .

**Věta 3.6.** *Je-li  $M \subseteq \mathbb{R}^n$  konvexní, pak také  $\text{ri}(M)$  je konvexní.*

*Důkaz.* Buď  $x_1, x_2 \in \text{ri}(M)$ . Pak existují relativní  $\varepsilon$ -okolí  $\mathcal{O}_\varepsilon(x_1), \mathcal{O}_\varepsilon(x_2) \subseteq M$ . Uvažme konvexní kombinaci  $x := \lambda_1 x_1 + \lambda_2 x_2$  a bod  $y \in \mathcal{O}_\varepsilon(x)$ . Pak libovolný bod v  $\mathcal{O}_\varepsilon(x)$  má tvar  $x + y = \lambda_1 x_1 + \lambda_2 x_2 + y = \lambda_1(x_1 + y) + \lambda_2(x_2 + y)$ , což náleží do  $M$  díky tomu, že  $x_1 + y, x_2 + y \in M$ .  $\square$

Důležitá vlastnost, týkající se konvexních množin, z pohledu optimalizace je oddělitelnost, viz obrázek 3.1.

**Definice 3.7.** Neprázdňné množiny  $M, N \subseteq \mathbb{R}^n$  jsou *oddělitelné* pokud existuje vektor  $a \neq 0 \in \mathbb{R}^n$  a číslo  $b \in \mathbb{R}$  takové, že

$$\begin{aligned} a^T x &\leq b \quad \forall x \in M, \\ a^T x &\geq b \quad \forall x \in N, \end{aligned}$$

a zároveň neplatí

$$a^T x = b \quad \forall x \in M \cup N.$$

Níže uvádíme jednu z variant vět o oddělitelnosti konvexních množin. Důkaz neuvádíme, je součástí jiného kursu.

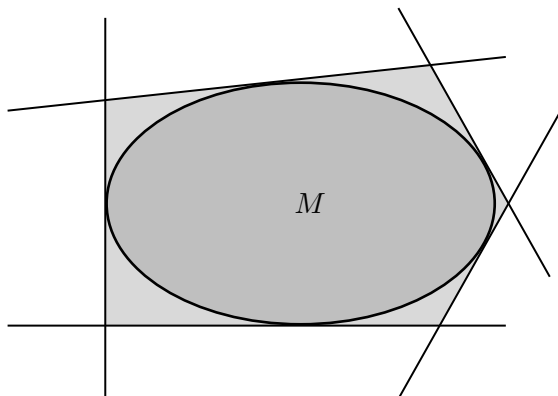
**Věta 3.8** (O oddělitelnosti). *Budťe  $M, N \subseteq \mathbb{R}^n$  neprázdňné a konvexní. Pak jsou oddělitelné právě tehdy, když  $\text{ri}(M) \cap \text{ri}(N) = \emptyset$ .*

Buď  $M \subseteq \mathbb{R}^n$  konvexní a uzavřená. Pomocí oddělitelnosti pak jednoduše můžeme zavést tzv. *tečnou nadrovinu* v hraničním bodě  $x^* \in M$  k množině  $M$  jako tu nadrovinu  $a^T x = b$ , která odděluje  $M$  a  $x^*$ . Pak platí  $a^T x^* = b$  (nadrovina prochází bodem  $x^*$ ) a množina  $M$  leží v kladném poloprostoru, vytyčeném nadrovinou, tj.  $a^T x \leq b$  pro všechna  $x \in M$ .

**Tvrzení 3.9.** *Buď  $M \subseteq \mathbb{R}^n$  konvexní a uzavřená. Pak  $M$  je rovna průniku poloprostorů určených tečnými nadrovinami.*

*Důkaz.* Z vlastnosti  $a^T x \leq b \quad \forall x \in M$  musí  $M$  ležet v průniku všech poloprostorů. Opačnou část dokážeme sporem. Pokud existuje  $x^* \notin M$  v průniku poloprostorů, pak jej (přesněji řečeno, jeho malé okolí) lze oddělit od  $M$  nadrovinou tečnou k  $M$ . Našli jsme tedy poloprostor, kde bod  $x^*$  neleží, spor.  $\square$

Tvrzení nemá jen teoretický rámec. Tečnými nadrovinami můžeme množinu  $M$  zapouzdřit do konvexního polyedru s libovolnou přesností, viz obrázek 3.2. Těto vlastnosti se využívá i v některých algoritmech, které začínou s počátečním výběrem tečných nadrovin a iterativně přidávají další podle potřeby, zejména pokud je třeba oddělit určitý bod polyedru od množiny  $M$ .

Obrázek 3.2: Vnější aproximace množiny  $M$  pomocí tečných nadrovin.

## 3.2 Konvexní funkce

Konvexní může být nejenom množina, ale i funkce.

**Definice 3.10.** Buď  $M \subseteq \mathbb{R}^n$  konvexní množina. Potom funkce  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  je *konvexní* na  $M$  pokud pro každé  $x_1, x_2 \in M$  a každé  $\lambda_1, \lambda_2 \geq 0$ ,  $\lambda_1 + \lambda_2 = 1$ , platí

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

Pokud platí podmínka

$$f(\lambda_1 x_1 + \lambda_2 x_2) < \lambda_1 f(x_1) + \lambda_2 f(x_2)$$

pro všechny konvexní kombinace s  $x_1 \neq x_2$  a  $\lambda_1, \lambda_2 > 0$ , pak je funkce *ryze konvexní* na  $M$ .

Podobně můžeme definovat *konkávní* funkci jako takovou funkci  $f(x)$ , pro níž  $-f(x)$  je konvexní. Evidentně, funkce je lineární (přesněji afinní) právě tehdy, když je zároveň konvexní i konkávní.

**Příklad 3.11.** Typickým příkladem konvexní funkce je jakákoliv norma, protože z definice je pro  $x_1, x_2 \in \mathbb{R}^n$  a  $\lambda_1, \lambda_2 \geq 0$ ,  $\lambda_1 + \lambda_2 = 1$ ,

$$\|\lambda_1 x_1 + \lambda_2 x_2\| \leq \|\lambda_1 x_1\| + \|\lambda_2 x_2\| = \lambda_1 \|x_1\| + \lambda_2 \|x_2\|.$$

Konvexní funkcí je tedy nejenom hladká eukleidovská norma  $\|x\|_2$ , ale i nehladké normy  $\|x\|_1$ ,  $\|x\|_\infty$  nebo jakákoliv maticová norma.

Analogicky jako ve větě 3.2 lze charakterizovat konvexní funkci pomocí konvexních kombinací  $k$ -tic bodů.

**Věta 3.12** (Jensenova nerovnost). *Buď  $k \geq 2$  a  $M \subseteq \mathbb{R}^n$  konvexní. Pak funkce  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  je konvexní na  $M$  právě tehdy, když pro libovolné  $x_1, \dots, x_k \in M$  a  $\lambda_1, \dots, \lambda_k \geq 0$ ,  $\sum_{i=1}^k \lambda_i = 1$ , platí*

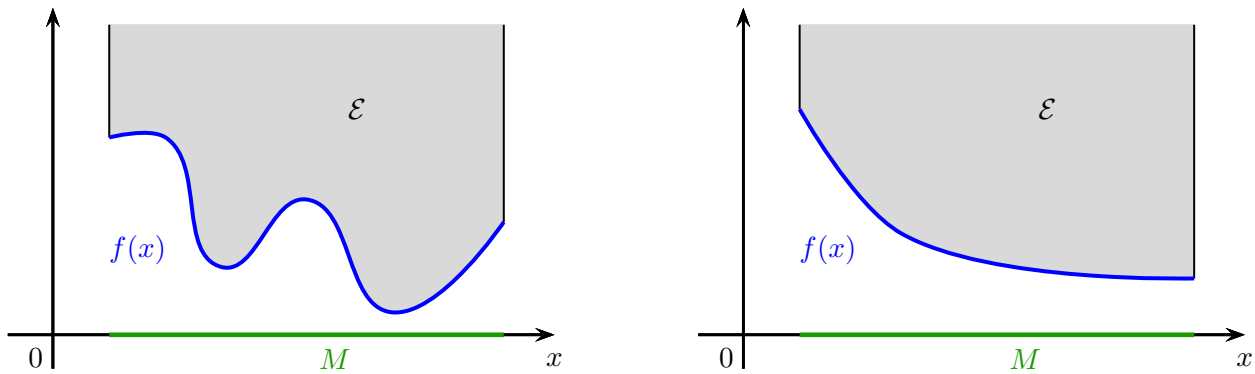
$$f\left(\sum_{i=1}^k \lambda_i x_i\right) \leq \sum_{i=1}^k \lambda_i f(x_i).$$

*Důkaz.* Matematickou indukcí podle  $k$ . Pro  $k = 2$  jasné, tak se zaměříme na indukční krok. Definujme  $\alpha := \sum_{i=1}^{k-1} \lambda_i$ . Protože  $\alpha + \lambda_k = 1$  a  $\sum_{i=1}^{k-1} \alpha^{-1} \lambda_i = 1$ , máme podle indukčního předpokladu

$$\begin{aligned} f\left(\sum_{i=1}^k \lambda_i x_i\right) &= f\left(\alpha \sum_{i=1}^{k-1} \alpha^{-1} \lambda_i x_i + \lambda_k x_k\right) \leq \alpha f\left(\sum_{i=1}^{k-1} \alpha^{-1} \lambda_i x_i\right) + \lambda_k f(x_k) \\ &\leq \alpha \sum_{i=1}^{k-1} \alpha^{-1} \lambda_i f(x_i) + \lambda_k f(x_k) = \sum_{i=1}^k \lambda_i f(x_i). \end{aligned}$$

□

Pro praktické ověřování konvexity funkce je pak velmi užitečné následující pozorování.



Obrázek 3.3: Epigraf  $\mathcal{E}$  nekonvexní (vlevo) a konvexní funkce (vpravo).

**Věta 3.13.** Funkce  $f(x)$  je konvexní na  $M$  právě tehdy, když je konvexní na každé přímce. To jest, funkce  $g(t) = f(x + ty)$  je konvexní na odpovídajícím kompaktním intervalu proměnné  $t$  pro každé  $x \in M$  a každé  $y$  velikosti 1.

Jiná užitečná charakterizace je pomocí epigrafu, viz obrázek 3.3.

**Definice 3.14.** Epigraf funkce  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  na množině  $M \subseteq \mathbb{R}^n$  je množina

$$\{(x, z) \in \mathbb{R}^{n+1}; x \in M, z \geq f(x)\}.$$

**Věta 3.15** (Fenchel, 1951). Buď  $M \subseteq \mathbb{R}^n$  konvexní množina. Potom funkce  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  je konvexní právě tehdy, když je její epigraf konvexní množina.

*Důkaz.* „ $\Rightarrow$ “ Označme jako  $\mathcal{E}$  epigraf funkce  $f(x)$  na  $M$ , a buď  $(x_1, z_1), (x_2, z_2) \in \mathcal{E}$  jeho libovolné body. Uvažme jejich konvexní kombinaci

$$\lambda_1(x_1, z_1) + \lambda_2(x_2, z_2) = (\lambda_1 x_1 + \lambda_2 x_2, \lambda_1 z_1 + \lambda_2 z_2).$$

Z konvexity  $M$  je  $\lambda_1 x_1 + \lambda_2 x_2 \in M$  a z konvexity  $f(x)$  je

$$f(\lambda_1 x_1 + \lambda_2 x_2) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2) \leq \lambda_1 z_1 + \lambda_2 z_2.$$

„ $\Leftarrow$ “ Buď  $\mathcal{E}$  konvexní. Pro libovolné  $x_1, x_2 \in M$  je  $(x_1, f(x_1)), (x_2, f(x_2)) \in \mathcal{E}$ . Uvažme konvexní kombinaci  $\lambda_1 x_1 + \lambda_2 x_2 \in M$ . Z konvexity  $\mathcal{E}$  je

$$\lambda_1(x_1, f(x_1)) + \lambda_2(x_2, f(x_2)) = (\lambda_1 x_1 + \lambda_2 x_2, \lambda_1 f(x_1) + \lambda_2 f(x_2)) \in \mathcal{E},$$

tudíž  $f(\lambda_1 x_1 + \lambda_2 x_2) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2)$ . □

Následující vlastnost, ilustrovaná na obrázku 3.4, je často používaná v optimalizaci. Jak uvidíme v kapitole 4, množina přípustných řešení  $M$  optimalizační úlohu  $\min_{x \in M} f(x)$  je často popsána nerovnostmi  $g_j(x) \leq 0$ ,  $j = 1, \dots, J$ . Pokud jsou funkce  $g_j$  konvexní, množina  $M$  je pak také konvexní.

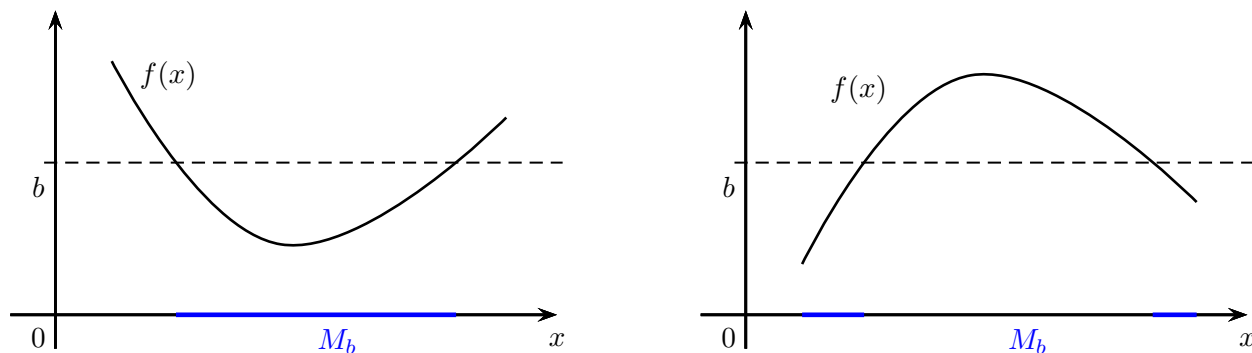
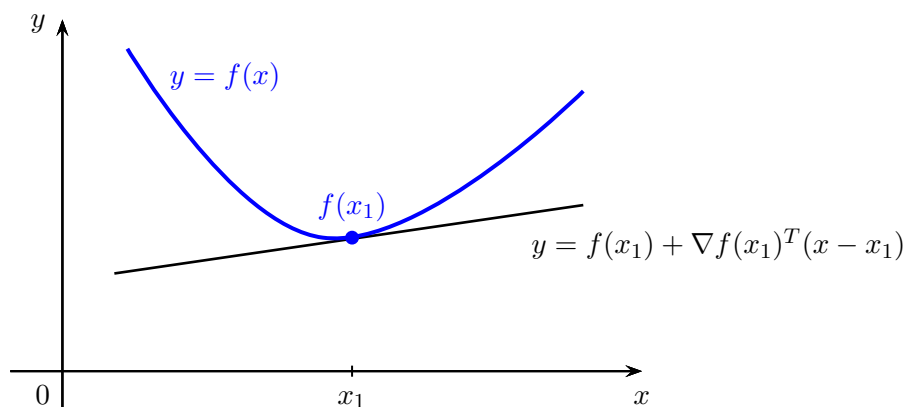
**Věta 3.16.** Buď  $M \subseteq \mathbb{R}^n$  konvexní množina a  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  konvexní funkce. Pro libovolné  $b \in \mathbb{R}$  je množina  $\{x \in M; f(x) \leq b\}$  konvexní.

*Důkaz.* Pro libovolné body  $x_1, x_2 \in \{x \in M; f(x) \leq b\}$  uvažme konvexní kombinaci  $\lambda_1 x_1 + \lambda_2 x_2 \in M$ . Z konvexity funkce  $f(x)$  máme

$$f(\lambda_1 x_1 + \lambda_2 x_2) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2) \leq \lambda_1 b + \lambda_2 b = b. \quad \square$$

Pěkná vlastnost konvexních funkcí je i ta, že jsou spojité. Větu uvádíme bez důkazu, který je možný najít např. v Lange [2016].

**Věta 3.17.** Buď  $M \subseteq \mathbb{R}^n$  neprázdná konvexní množina dimenze  $n$ , a buď funkce  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  konvexní. Pak  $f(x)$  je spojitá a lokálně lipschitzovská na int  $M$ .

Obrázek 3.4: Množina  $M_b := \{x \in M; f(x) \leq b\}$  pro konvexní a nekonvexní funkci (viz věta 3.16).Obrázek 3.5: Tečna grafu konvexní funkce  $f(x)$  v bodě  $(x_1, f(x_1))$ .

### 3.3 Charakterizace konvexity funkce prvního a druhého řádu

Charakterizace konvexity funkce prvního řádu má pěknou geometrickou interpretaci, viz obrázek 3.5. Tečna vytvořená v každém bodě  $(x_1, f(x_1))$  grafu funkce musí ležet pod grafem funkce, tedy  $f(x) \geq f(x_1) + \nabla f(x_1)^T(x - x_1)$ .

**Věta 3.18** (Charakterizace konvexity funkce prvního řádu, Avriel 1976, Mangasarian 1969). *Buď  $\emptyset \neq M \subseteq \mathbb{R}^n$  konvexní množina a buď  $f(x)$  diferencovatelná funkce na otevřené množině obsahující  $M$ . Pak  $f(x)$  je konvexní na  $M$  právě tehdy, když pro každé  $x_1, x_2 \in M$  platí*

$$f(x_2) - f(x_1) \geq \nabla f(x_1)^T(x_2 - x_1). \quad (3.1)$$

*Důkaz.* „ $\Rightarrow$ “ Buď  $x_1, x_2 \in M$  a  $\lambda \in (0, 1)$  libovolné. Pak

$$\begin{aligned} f((1 - \lambda)x_1 + \lambda x_2) &\leq (1 - \lambda)f(x_1) + \lambda f(x_2), \\ f((1 - \lambda)x_1 + \lambda x_2) - f(x_1) &\leq \lambda(f(x_2) - f(x_1)), \\ \frac{f(x_1 + \lambda(x_2 - x_1)) - f(x_1)}{\lambda} &\leq f(x_2) - f(x_1). \end{aligned}$$

Limitním přechodem  $\lambda \rightarrow 0$  dostaneme (3.1) podle pravidel o derivování složené funkce  $g(\lambda) = f(x_1 + \lambda(x_2 - x_1))$  podle  $\lambda$ .

„ $\Leftarrow$ “ Buď  $x_1, x_2 \in M$  a uvažujme konvexní kombinaci  $x = \lambda_1 x_1 + \lambda_2 x_2$ . Podle (3.1) máme

$$\begin{aligned} f(x_1) - f(x) &\geq \nabla f(x)^T(x_1 - x) = \nabla f(x)^T(x_1 - (\lambda_1 x_1 + \lambda_2 x_2)) = \lambda_2 \nabla f(x)^T(x_1 - x_2), \\ f(x_2) - f(x) &\geq \nabla f(x)^T(x_2 - x) = \nabla f(x)^T(x_2 - (\lambda_1 x_1 + \lambda_2 x_2)) = \lambda_1 \nabla f(x)^T(x_2 - x_1). \end{aligned}$$

První nerovnici vynásobíme  $\lambda_1$ , druhou  $\lambda_2$  a sečtením máme

$$\lambda_1(f(x_1) - f(x)) + \lambda_2(f(x_2) - f(x)) \geq 0,$$

neboli  $\lambda_1 f(x_1) + \lambda_2 f(x_2) \geq f(x)$ . □

**Poznámka 3.19.** Pro ryzí konvexitu máme analogickou charakterizaci

$$\forall x_1, x_2 \in M, x_1 \neq x_2 : f(x_2) - f(x_1) > \nabla f(x_1)^T(x_2 - x_1). \quad (3.2)$$

**Věta 3.20** (Charakterizace konvexity funkce druhého řádu, Fenchel, 1951). *Bud'  $\emptyset \neq M \subseteq \mathbb{R}^n$  konvexní otevřená množina dimenze  $n$  a necht' funkce  $f: M \rightarrow \mathbb{R}$  má spojitě druhé parciální derivace na  $M$ . Pak  $f(x)$  je konvexní na  $M$  právě tehdy, když Hessián  $\nabla^2 f(x)$  je pozitivně semidefinitní pro každé  $x \in M$ .*

*Důkaz.* Bud'  $x^* \in M$  libovolné pevné. Ze spojitosti druhých parciálních derivací pro každé  $\lambda \in \mathbb{R}$  a  $y \in \mathbb{R}^n$ ,  $x^* + \lambda y \in M$ , existuje  $\theta \in (0, 1)$  takové, že

$$f(x^* + \lambda y) = f(x^*) + \lambda \nabla f(x^*)^T y + \frac{1}{2} \lambda^2 y^T \nabla^2 f(x^* + \theta \lambda y) y. \quad (3.3)$$

„ $\Rightarrow$ “ Podle věty 3.18 je

$$f(x^* + \lambda y) \geq f(x^*) + \lambda \nabla f(x^*)^T y,$$

tudíž z (3.3) máme

$$y^T \nabla^2 f(x^* + \theta \lambda y) y \geq 0.$$

Limitním přechodem  $\lambda \rightarrow 0$  dostaneme  $y^T \nabla^2 f(x^*) y \geq 0$ .

„ $\Leftarrow$ “ Z pozitivní semidefinitnosti Hessiánu máme  $y^T \nabla^2 f(x^* + \theta \lambda y) y \geq 0$  ve výrazu (3.3), tudíž

$$f(x^* + \lambda y) \geq f(x^*) + \lambda \nabla f(x^*)^T y,$$

což podle věty 3.18 ukazuje konvexitu  $f(x)$ . □

**Poznámka 3.21.** Pro ryzí konvexitu máme takovouto charakterizaci:

- (1) Je-li  $f$  ryze konvexní, pak Hessián  $\nabla^2 f(x)$  je pozitivně definitní na  $M$  až na množinu míry 0, kde je pozitivně semidefinitní.
- (2) Je-li Hessián  $\nabla^2 f(x)$  pozitivně definitní na  $M$ , pak  $f$  je ryze konvexní.

Důvod, proč v první vlastnosti není Hessián pozitivně definitní všude na  $M$ , je ten, že při adaptaci důkazu věty 3.20 na ryzí konvexitu může limitním přechodem  $\lambda \rightarrow 0$  ostrá nerovnost přejít v neostrou.

**Příklad 3.22.**

1. Funkce  $f(x) = x^4$  je ryze konvexní na  $\mathbb{R}$ , ale Hessián  $f(x)'' = 12x^2$  je nulový v bodě  $x = 0$ .
2. Funkce  $f(x) = x^{-2}$  má kladné druhé derivace všude na  $\mathbb{R} \setminus \{0\}$ , ale není tam konvexní. Důvod je nejen ten, že  $\mathbb{R} \setminus \{0\}$  není konvexní množina, ale základní vlastnost konvexní funkce nebude splněna ani když budeme vynechávat nulu v konvexní kombinaci. Proto je nutné, aby nosná množina byla konvexní. Tudíž  $f(x)$  je konvexní třeba na  $(0, \infty)$  nebo na  $(-\infty, 0)$ . □

**Příklad 3.23.** Uvažujme kvadratickou funkci  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  danou předpisem  $f(x) = x^T A x + b^T x + c$ , kde  $A \in \mathbb{R}^{n \times n}$  je symetrická,  $b \in \mathbb{R}^n$ ,  $c \in \mathbb{R}$  (viz příloha na str. 61). Pak

- $f(x)$  je konvexní právě tehdy, když  $A$  je pozitivně semidefinitní,
- $f(x)$  je ryze konvexní právě tehdy, když  $A$  je pozitivně definitní. □



### 3.4 Další pravidla pro zjišťování konvexity funkce

Zde ukážeme jak se konvexita funkcí chová při součinu, dělení, skládání a pod.

**Věta 3.24.** Mějme funkce  $f, g: \mathbb{R} \rightarrow \mathbb{R}$

- (1) Jsou-li  $f(x), g(x)$  konvexní, nezáporné a neklesající (nebo obě nerostoucí), pak  $f(x) \cdot g(x)$  je konvexní.
- (2) Je-li  $f(x)$  konvexní, nezáporná a neklesající, a  $g(x)$  konkávní, kladná a nerostoucí, pak  $f(x)/g(x)$  je konvexní.

*Důkaz.* Ukážeme první vlastnost, druhá se dokáže analogicky.

Nyní  $(fg)'' = f''g + 2f'g' + fg'' \geq 0$ , protože každý sčítanec je nezáporný. Tudíž podle věty 3.20 je  $fg$  konvexní.  $\square$

**Příklad 3.25.** Funkce  $f(x) = x$ ,  $g(y) = y$  jsou obě konvexní, ale jejich součin  $h(x, y) = xy$  už není konvexní ani když se omezíme na  $(x, y) \in [0, \infty)^2$ , protože je konkávní na úsečce mezi body  $(1, 0)$  a  $(0, 1)$ . Proto je nutno, aby obě funkce byly stejné proměnné. Funkce  $f(x) \cdot g(x) = x^2$  už zřejmě konvexní jest.  $\square$

**Věta 3.26.** Mějme funkce  $f: \mathbb{R}^n \rightarrow \mathbb{R}^k$ ,  $g: \mathbb{R}^k \rightarrow \mathbb{R}$ .

- (1) Jsou-li  $f_i(x)$  konvexní pro každé  $i = 1, \dots, k$  a je-li  $g(y)$  konvexní a neklesající v každé složce, pak  $(g \circ f)(x) = g(f(x))$  je konvexní.
- (2) Jsou-li  $f(x)$  konkávní pro každé  $i = 1, \dots, k$  a je-li  $g(y)$  konvexní a nerostoucí v každé složce, pak  $(g \circ f)(x) = g(f(x))$  je konvexní.

*Důkaz.* Ukážeme první vlastnost, druhá vlastnost se dokáže analogicky.

Buď  $x_1, x_2 \in \mathbb{R}^n$ . Pak pro konvexní kombinaci  $\lambda_1 x_1 + \lambda_2 x_2$  máme

$$g(f(\lambda_1 x_1 + \lambda_2 x_2)) \leq g(\lambda_1 f(x_1) + \lambda_2 f(x_2)) \leq \lambda_1 g(f(x_1)) + \lambda_2 g(f(x_2)),$$

kde první nerovnost plyne z konvexity  $f$  a monotonie  $g$ , a druhá nerovnost plyne z konvexity  $g$ .  $\square$

**Příklad 3.27.**

1. Je-li  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  konvexní, pak  $e^{f(x)}$  je konvexní. Tedy např.  $e^{e^x}$ ,  $e^{x^2-x}$ ,  $e^{x_1-x_2}$ , ...
2. Je-li  $f(x) \geq 0$  a konvexní,  $p \geq 1$ , pak  $f(x)^p$  je konvexní.
3. Je-li  $f(x) \geq 0$  a konkávní, pak  $-\log(f(x))$  je konvexní.  $\square$

**Příklad 3.28.** Bez monotonie věta 3.26 neplatí. Např. pro konvexní funkce  $f(x) = x^2 - 1$ ,  $g(y) = y^2$  není  $g(f(x)) = (x^2 - 1)^2$  konvexní.  $\square$

**Poznámka 3.29.** Testování konvexity je obecně těžký problém. Ahmadi et al. [2013] ukázali, že je to NP-těžký problém i pro třídu polynomů (více proměnných) stupně nanejvýš 4. Pro „obecné“ funkce je zatím otevřená i otázka jestli je konvexita rozhodnutelný problém.



# Kapitola 4

## Konvexní optimalizace

Úlohou *konvexní optimalizace* rozumíme úlohu

$$\min f(x) \text{ za podm. } x \in M,$$

kde  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  je konvexní funkce a  $M \subseteq \mathbb{R}^n$  je konvexní množina. Často se uvažuje množina přípustných řešení  $M$  ve tvaru

$$M = \{x \in \mathbb{R}^n; g_j(x) \leq 0, j = 1, \dots, J\},$$

kde  $g_j(x): \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $j = 1, \dots, J$ , jsou konvexní funkce. Podle věty 3.16 je tedy množina  $M$  konvexní. V této kapitole ale budeme pracovat s obecnou konvexní množinou  $M$ .

**Příklad 4.1.** Úlohou konvexní optimalizace je například úloha

$$\min x_1 + x_2 \text{ za podm. } x_1^2 + x_2^2 \leq 2$$

nebo úloha

$$\min x_1^2 + x_2^2 + 2x_2 \text{ za podm. } x_1^2 + x_2^2 \leq 2. \quad \square$$

### 4.1 Základní vlastnosti

**Věta 4.2** (Fenchel, 1951). *Pro úlohu konvexní optimalizace platí:*

- (1) Každé lokální minimum je globálním minimem.
- (2) Množina optimálních řešení je konvexní.
- (3) Je-li  $f(x)$  ryze konvexní funkce, pak optimální řešení je nanejvýš jedno.

*Důkaz.*

- (1) Buď  $x^0 \in M$  lokální minimum a pro spor nechť existuje  $x^* \in M$  takové, že  $f(x^*) < f(x^0)$ . Uvažujme konvexní kombinaci  $x = \lambda x^* + (1 - \lambda)x^0 \in M$ ,  $\lambda \in (0, 1)$ . Pak

$$f(x) \leq \lambda f(x^*) + (1 - \lambda)f(x^0) < \lambda f(x^0) + (1 - \lambda)f(x^0) = f(x^0).$$

To je spor s lokální minimalitou  $x^0$  neboť pro libovolně malé  $\lambda > 0$  je  $f(x) < f(x^0)$ .

- (2) Buďte  $x_1, x_2 \in M$  optimální řešení, a označme  $z = f(x_1) = f(x_2)$  optimální hodnotu. Pak pro konvexní kombinaci  $x = \lambda_1 x_1 + \lambda_2 x_2 \in M$  máme

$$f(x) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2) = \lambda_1 z + \lambda_2 z = z,$$

tedy i  $x$  je optimální řešení.

- (3) Pro spor předpokládejme, že  $x_1, x_2 \in M$ ,  $x_1 \neq x_2$ , jsou optimální řešení. Označme  $z = f(x_1) = f(x_2)$  optimální hodnotu. Pak pro konvexní kombinaci  $x = \lambda_1 x_1 + \lambda_2 x_2 \in M$ ,  $\lambda_1, \lambda_2 > 0$ , máme

$$f(x) < \lambda_1 f(x_1) + \lambda_2 f(x_2) = \lambda_1 z + \lambda_2 z = z,$$

tedy  $x$  je lepší než optimální řešení, spor.  $\square$

Poznamenejme, že konvexní úloha nemusí mít žádné optimum, například  $\min_{x \in \mathbb{R}} e^x$ . A to ani na kompaktní množině:

**Příklad 4.3.** Definujme funkci  $f: [1, 2] \rightarrow \mathbb{R}$  předpisem

$$f(x) = \begin{cases} x & \text{pro } 1 < x \leq 2, \\ 2 & \text{pro } x = 1. \end{cases}$$

Funkce je konvexní, ale ne spojitá a díky tomu se minimum na  $[1, 2]$  nenabyde.  $\square$

**Věta 4.4.** Bud'  $\emptyset \neq M \subseteq \mathbb{R}^n$  otevřená konvexní a  $f: M \rightarrow \mathbb{R}$  konvexní a diferencovatelná na  $M$ . Pak  $x^* \in M$  je optimální řešení právě tehdy, když  $\nabla f(x^*) = o$ .

*Důkaz.* „ $\Rightarrow$ “ Bud'  $x^* \in M$  optimální řešení. Pak je lokálním minimem a podle věty 2.1 je  $\nabla f(x^*) = o$ .

„ $\Leftarrow$ “ Bud'  $\nabla f(x^*) = o$ . Podle věty 3.18 pro každé  $x \in M$  je  $f(x) - f(x^*) \geq \nabla f(x^*)^T (x - x^*) = 0$ , tedy  $f(x) \geq f(x^*)$ , a proto je  $x^*$  optimálním řešením.  $\square$

Poznamenejme, že otevřenost množiny  $M$  z předchozí věty je nutná. Například pro úlohu  $\min_{x \in [1, 2]} x$  je  $M = [1, 2]$  konvexní a účelová funkce  $f(x) = x$  je diferencovatelná na  $\mathbb{R}$ , ale v optimu  $x^* = 1$  je její derivace  $f'(1) = 1$ .

Větu můžeme zobecnit na ne nutně otevřené množiny:

**Věta 4.5.** Bud'  $\emptyset \neq M \subseteq \mathbb{R}^n$  konvexní a  $f: M' \rightarrow \mathbb{R}$  konvexní a diferencovatelná na otevřené  $M' \supseteq M$ . Pak  $x^* \in M$  je optimální řešení právě tehdy, když pro každé  $y \in M$  je  $\nabla f(x^*)^T (y - x^*) \geq 0$ .

*Důkaz.* „ $\Rightarrow$ “ Pro spor předpokládejme, že existuje  $y \in M$  takové, že  $\nabla f(x^*)^T (y - x^*) < 0$ . Uvažujme konvexní kombinaci  $x_\lambda = \lambda y + (1 - \lambda)x^* = x^* + \lambda(y - x^*) \in M$ . Pak

$$0 > \nabla f(x^*)^T (y - x^*) = \lim_{\lambda \rightarrow 0^+} \frac{f(x^* + \lambda(y - x^*)) - f(x^*)}{\lambda} = \lim_{\lambda \rightarrow 0^+} \frac{f(x_\lambda) - f(x^*)}{\lambda}.$$

Tudíž pro dost malé  $\lambda > 0$  je  $f(x_\lambda) < f(x^*)$ , spor.

„ $\Leftarrow$ “ Podle věty 3.18 pro každé  $y \in M$  je  $f(y) - f(x^*) \geq \nabla f(x^*)^T (y - x^*) \geq 0$ . Tedy  $f(y) \geq f(x^*)$ , a proto je  $x^*$  optimálním řešením.  $\square$

Podmínka z věty uspěje speciálně, když  $\nabla f(x^*) = o$ . Tedy každý stacionární bod je automaticky globálním minimem.

**Příklad 4.6.** První úloha z příkladu 4.1

$$\min x_1 + x_2 \quad \text{za podm. } x_1^2 + x_2^2 \leq 2$$

má zřejmě optimum v bodě  $x^* = (-1, -1)^T$ . Pomocí věty 4.5 to snadno ověříme. Protože  $\nabla f(x^*) = (1, 1)^T$ , musíme ukázat, že pro každé přípustné řešení  $y$  platí

$$\nabla f(x^*)^T (y - x^*) = (1, 1) \begin{pmatrix} y_1 + 1 \\ y_2 + 1 \end{pmatrix} \geq 0,$$

čili  $y_1 + y_2 \geq -2$ . To je zjevně pravda.

V druhé úloze z příkladu 4.1 řešíme

$$\min x_1^2 + x_2^2 + 2x_2 \quad \text{za podm. } x_1^2 + x_2^2 \leq 2.$$

Spočítáme  $\nabla f(x^*) = (2x_1, 2x_2 + 2)^T$  a tento gradient je nulový v bodě  $x^* = (0, -1)$ . Protože bod  $x^*$  splňuje omezení úlohy, je to hledané optimum.  $\square$

**Příklad 4.7** (Systém hodnocení výkonnosti). Existuje řada metod, jak vyjádřit hodnocení entit na základě známých výsledků - například výkonnost hráče či sportovního týmu na základě odehraných utkání. Uvažujme následující metodu [Langville and Meyer, 2012]. Mějme  $n$  týmů, které chceme ohodnotit pomocí čísel  $r_1, \dots, r_n \in \mathbb{R}^n$ . Nechť  $A \in \mathbb{R}^{n \times n}$  je známá skórovací matice, která na pozici  $a_{ij}$  udává bodový zisk týmu  $i$  nad týmem  $j$ . Tato matice je antisymetrická, tj.  $A = -A^T$ , neboť  $a_{ii} = 0$  a  $a_{ij} = -a_{ji}$ . Chceme, aby hodnocení  $r = (r_1, \dots, r_n)^T$  reflektovalo bodové zisky, takže ideálně  $a_{ij} = r_i - r_j$ , neboli  $A = re^T - er^T$ . To se nemusí podařit přesně, ale můžeme se snažit o co nejlepší aproximaci. To nás vede na optimalizační úlohu

$$\min_{x \in \mathbb{R}^n} f(x) = \|A - (xe^T - ex^T)\|^2.$$

Jako maticovou normu zvolíme Frobeniovu normu, definovanou pro matici  $M \in \mathbb{R}^{n \times n}$  jako  $\|M\| = \sqrt{\sum_{i,j} m_{ij}^2} = \sqrt{\text{tr}(M^T M)}$ ; proto minimalizujeme druhou mocninu normy. Účelovou funkci pak můžeme vyjádřit

$$\begin{aligned} f(x) &= \text{tr}((A - (xe^T - ex^T))^T (A - (xe^T - ex^T))) \\ &= \text{tr}(A^T A) - \text{tr}(A^T(xe^T - ex^T) + (xe^T - ex^T)^T A) + \text{tr}((xe^T - ex^T)^T (xe^T - ex^T)) \\ &= \text{tr}(A^T A) - 4x^T A e + 2n(x^T x) - 2(e^T x)^2. \end{aligned}$$

Gradient a Hessián účelové funkce jsou

$$\nabla f(x) = -4Ae + 4nx - 4ee^T x, \quad \nabla^2 f(x) = 4(nI_n - ee^T).$$

Hessián je pozitivně semidefinitní, proto je  $f(x)$  konvexní. Podmínka optimality  $\nabla f(x) = 0$  pak vede na soustavu

$$(nI_n - ee^T)x = Ae.$$

Matice má hodnotu  $n - 1$  a tudíž je množina řešení celá přímka  $x = \frac{1}{n}Ae + \alpha e$ ,  $\alpha \in \mathbb{R}$ . Na této přímce je funkce  $f(x)$  konstantní, proto jsou všechny její body optimální. V praxi se vektor hodnocení normuje tak, aby  $e^T r = 0$ , což nám dá výsledný vzoreček pro hodnocení  $r = \frac{1}{n}Ae$ .  $\square$

Speciální úlohy konvexní optimalizace mají specifické vlastnosti. V následujících sekcích probereme několik takových speciálních případů.

## 4.2 Kvadratické programování

Úlohou *kvadratického programování* je úloha

$$\min x^T C x + d^T x \text{ za podm. } x \in M,$$

kde  $C \in \mathbb{R}^{n \times n}$  je symetrická,  $d \in \mathbb{R}^n$  a  $M \subseteq \mathbb{R}^n$  je konvexní polyedr. Pokud matice  $C$  je pozitivně semidefinitní, pak jde o konvexní úlohu, která se nazývá *konvexní kvadratické programování*.

Konvexní kvadratické programování je efektivně řešitelné v polynomiálním čase [Floudas and Pardalos, 2009, sekce „Complexity Theory: Quadratic Programming“]. V případě, že  $C$  není pozitivně semidefinitní, problém se stává NP-těžkým, dokonce i nalezení lokálního minima. Zajímavé je, že NP-těžkost platí i jen pokud jediné vlastní číslo matice  $C$  je záporné [Pardalos and Vavasis, 1991; Vavasis, 1991]. NP-těžkost speciálního případu s negativně definitní ukážeme dole; z estetických důvodů úlohu formulujeme ekvivalentně jako maximalizaci s pozitivně definitní maticí, neboť  $\max_{x \in M} x^T C x = -\min_{x \in M} -x^T C x$ .

**Věta 4.8.** *Úloha  $\max_{x \in M} x^T C x$  je NP-těžká i pro  $C$  pozitivně definitní.*

*Důkaz.* Redukcí z NP-úplné úlohy SET-PARTITIONING: Lze množinu daných čísel  $\alpha_1, \dots, \alpha_n \in \mathbb{N}$  rozdělit do dvou skupin se stejným součtem? Ekvivalentně, existuje  $x \in \{\pm 1\}^n$  tak, že  $\sum_{i=1}^n \alpha_i x_i = 0$ ? Tuto úlohu můžeme přeformulovat jako

$$\max \sum_{i=1}^n x_i^2 \text{ za podm. } \sum_{i=1}^n \alpha_i x_i = 0, \quad x \in [-1, 1]^n.$$

Optimální hodnota této úlohy je  $n$  právě tehdy, když SET-PARTITIONING má řešení. Optimalizační úloha odpovídá předpisu, protože podmínky jsou lineární a účelová funkce má tvar  $x^T C x + d^T x$  pro  $C = I_n$ ,  $d = o$ .  $\square$

**Příklad 4.9** (Problém výběru portfolia). Jedná se o učebnicový příklad použití konvexního kvadratického programování. Průkopníkem v této oblasti byl Harry Markowitz, nositel Nobelovy ceny za ekonomii z roku 1990, i když jeho výsledek je z už roku 1952.

Základní úloha zní: Máme kapitál  $K$  a můžeme investovat to  $n$  investic, investice  $i$  vynese  $c_i$  jednotek peněz. Problém nalezení optimálního portfolia investic vede na triviální úlohu lineárního programování

$$\max c^T x \text{ za podm. } e^T x = K, x \geq o.$$

Výnos investic ale nebývá známý dopředu a modeluje se jako náhodná veličina. Nechť náhodný vektor  $c$  má střední hodnotu  $\tilde{c} := E c$  a kovarianční matici  $\Sigma := \text{cov } c = E(c - \tilde{c})(c - \tilde{c})^T$ , která je pozitivně semidefinitní (*důkaz*: pro každé  $x \in \mathbb{R}^n$  je  $x^T \Sigma x = x^T (E(c - \tilde{c})(c - \tilde{c})^T) x = E x^T (c - \tilde{c})(c - \tilde{c})^T x = E((c - \tilde{c})^T x)^2 \geq 0$ ). Pak pro reálný vektor  $x \in \mathbb{R}^n$  máme, že střední hodnota  $c^T x$  je  $E(c^T x) = \tilde{c}^T x$  a rozptyl  $c^T x$  je  $\text{var}(c^T x) = x^T \Sigma x$ .

Pokud bychom chtěli maximalizovat střední hodnotu výnosu, povede to na úlohu

$$\max \tilde{c}^T x \text{ za podm. } e^T x = K, x \geq o.$$

Chceme-li vzít v úvahu riziko investic, modelujeme úlohu jako konvexní kvadratický program

$$\max \tilde{c}^T x - \gamma x^T \Sigma x \text{ za podm. } e^T x = K, x \geq o,$$

kde  $\gamma > 0$  je tzv. parametr averze vůči riziku.  $\square$

**Příklad 4.10** (Plánování tras kvadroptér). Cílem je naplánovat trasu kvadroptérám tak, aby nedošlo k jejich kolizi a aby se přesunuli z počáteční do koncové pozice s co nejmenším úsilím. V našem modelu diskretizujeme čas na úseky délky  $h$ . Proměnné budou pozice  $p_i(k)$ , rychlost  $v_i(k)$  a zrychlení  $a_i(k)$  pro kvadroptéru  $i$  v čase  $k$ . Omezení jsou:

- fyzikální: vztah zrychlení a rychlosti, rychlosti a pozice, ... (např.  $v_i(k) = v_i(k-1) + h \cdot a_i(k-1)$ ,  $p_i(k) = p_i(k-1) + h \cdot v_i(k-1)$ , ...)
- omezení na maximální rychlost, zrychlení a změnu zrychlení,
- počáteční a koncová pozice,
- zamezení kolize je nekonvexní omezení ( $\|p_i(k) - p_j(k)\|_2 \geq r \forall i \neq j$ ), proto se musí linearizovat

Účelová funkce je pak dána součtem norem zrychlení v jednotlivých časech ( $\sum_{i,k} \|a_i(k) + g\|_2^2$ ). Více podrobností viz:

- <https://www.youtube.com/watch?v=wwK7WvvUv1I>
- F. Augugliaro, A.P. Schoellig, and R. D'Andrea, *Generation of collision-free trajectories for a quadcopter fleet: A sequential convex programming approach*, IEEE/RSJ International Conference on Intelligent Robots and Systems, 2012: pp. 1917–1922.

Jakkoli by se mohlo zdát, že tento problém je trochu umělý, plánování bezkolizních tras je zrovna aktuální problém při řízení letového provozu nad letišti.

### 4.3 Kuželové programování

Tato sekci vychází zejména z knihy Ben-Tal and Nemirovski [2001]. Motivace ke kuželovému programování je následující. Úloha lineárního programování

$$\min c^T x \text{ za podm. } Ax \geq b$$

se dá zobecňovat různými způsoby. V sekci 4.2 jsme nahradili lineární funkce kvadratickými. Jiný způsob je zobecnit relaci „ $\geq$ “.

**Definice 4.11.** Množina  $\emptyset \neq \mathcal{K} \subseteq \mathbb{R}^n$  tvoří *konvexní kužel* pokud platí dvě podmínky:

- (1) pro každé  $\alpha \geq 0$  a  $x \in \mathcal{K}$  je  $\alpha x \in \mathcal{K}$ ,
- (2) pro každé  $x, y \in \mathcal{K}$  je  $x + y \in \mathcal{K}$ .

Dále, kužel je *bodový*, pokud neobsahuje přímku.

**Tvrzení 4.12.** Je-li  $\mathcal{K}$  bodový konvexní kužel, pak definuje

- (1) *částečné uspořádání* předpisem  $x \geq_{\mathcal{K}} y$  právě tehdy, když  $x - y \in \mathcal{K}$ ,
- (2) *částečné ostré uspořádání* předpisem  $x >_{\mathcal{K}} y$  právě tehdy, když  $x - y \in \text{int } \mathcal{K}$ .

Nadále budeme uvažovat bodový, konvexní, uzavřený kužel  $\mathcal{K}$ , mající neprázdný vnitřek.

**Příklad 4.13** (Příklady kuželů). Typickými příklady kuželů jsou

- *nezáporný ortant*  $\mathbb{R}_+^n = \{x \in \mathbb{R}^n; x \geq 0\}$ . Odpovídající uspořádání je standardní nerovnost  $\geq$  pro vektory.
- *Lorentzův kužel (zmrzlinový kornout)*  $\mathcal{L} = \{x \in \mathbb{R}^n; x_n \geq \sqrt{\sum_{i=1}^{n-1} x_i^2}\} = \{x \in \mathbb{R}^n; x_n \geq \|(x_1, \dots, x_{n-1})\|_2\}$ .
- *Zobecněný Lorentzův kužel*  $\mathcal{L} = \{x \in \mathbb{R}^n; x_n \geq \|(x_1, \dots, x_{n-1})\|\}$ , kde  $\|\cdot\|$  je libovolná norma.
- *Polyedrický kužel* je popsán soustavou nerovnic  $Ax \leq 0$ . Do této kategorie spadá například *nezáporný ortant* nebo *zobecněný Lorentzův kužel* při použití maximové či součtové normy.
- *Kužel pozitivně semidefinitních matic.*

Nyní máme vše připraveno k tomu vyjádřit, co to je kuželové programování. Úloha *kuželového programování* je

$$\min c^T x \text{ za podm. } Ax \geq_{\mathcal{K}} b. \quad (4.1)$$

**Příklad 4.14** (Příklady kuželových programů).

- Pro  $\mathcal{K} = \mathbb{R}_+^n$  dostaneme klasické lineární programování.
- Použijeme-li Lorentzův kužel, dostaneme zajímavější příklad

$$\min c^T x \text{ za podm. } \|Bx - a\|_2 \leq d^T x + f. \quad (4.2)$$

Tato úloha nelze jednoduše přepsat na úlohu s konvexními kvadratickými omezeními, neboť umocněním obou stran nerovnice dostaneme sice kvadratickou funkci, ale nebude konvexní (konvexita se zrušila umocněním).

- Kuželová omezení jsou navíc kombinovat, takže můžeme uvažovat i úlohy typu

$$\min c^T x \text{ za podm. } Ax \geq b, \|Bx - a\|_2 \leq d^T x + f.$$

Důvod je ten, že kartézský součin kuželů je opět kužel. V tomto případě jsme použili kužel  $\mathcal{K} := \mathbb{R}_+^n \times \mathcal{L}$ .

Tento typ úloh se nazývá kuželové kvadratické programování, viz sekce 4.3.2.

- Kužel pozitivně semidefinitních matic vede na úlohy typu

$$\min c^T x \text{ za podm. } \sum_{k=1}^n x_k A^{(k)} - B \text{ je pozitivně semidefinitní,}$$

kde  $A_1, \dots, A_n, B$  jsou symetrické matice. Tyto úlohy se nazývají úlohami semidefinitního programování, více viz sekce 4.3.3.

### 4.3.1 Dualita v kuželovém programování

**Motivace.** Pro lineární program  $\min\{c^T x; Ax \geq b\}$  se dualita odvozovala takto: Buď  $x$  přípustné a pro každé  $y \geq 0$  platí jako důsledek  $y^T Ax \geq y^T b$ . Pokud náhodou platí  $y^T A = c^T$ , pak dostáváme  $c^T x = y^T Ax \geq y^T b$ . Jinými slovy,  $y^T b$  je dolní odhad na minimální hodnotu pro všechna  $y \geq 0$  s vlastností  $A^T y = c$ . To vede k duální úloze a slabé dualitě

$$\min\{c^T x; Ax \geq b\} \geq \max\{b^T y; A^T y = c, y \geq 0\}.$$

Jakou relací nahradit  $y \geq 0$  v případě kuželového programování (4.1)? Ukazuje se, že  $y \geq 0$  ani  $y \geq_{\mathcal{K}} 0$  nemusí fungovat správně. Zajímá nás vlastně množina takových  $y$ , že pro každé  $a \geq_{\mathcal{K}} 0$  je  $y^T a \geq 0$ . Je zřejmé, že tato množina tvoří kužel a tento kužel se nazývá duální k  $\mathcal{K}$ .

**Definice 4.15.** Buď  $\mathcal{K} \subseteq \mathbb{R}^n$  kužel. Pak *duální kužel* ke  $\mathcal{K}$  je kužel

$$\mathcal{K}^* = \{y \in \mathbb{R}^n; y^T a \geq 0 \forall a \in \mathcal{K}\}.$$

Pomocí duálního kužele pak stanovíme duální úlohu k (4.1) jako

$$\max b^T y \text{ za podm. } A^T y = c, y \geq_{\mathcal{K}^*} 0. \quad (4.3)$$

Slabou dualitu jsme již víceméně dokázali.

**Věta 4.16** (Slabá dualita). *Platí:*  $\min\{c^T x; Ax \geq_{\mathcal{K}} b\} \geq \max\{b^T y; A^T y = c, y \geq_{\mathcal{K}^*} 0\}$ .

*Důkaz.* Pro každé  $y \geq_{\mathcal{K}^*} 0$  splňující  $A^T y = c$  a pro každé  $x$  splňující  $Ax \geq_{\mathcal{K}} b$  platí

$$c^T x = y^T Ax \geq y^T b.$$

Jinými slovy, hodnota účelové funkce každého přípustného řešení primární úlohy je horní mezí na hodnotu účelové funkce každého přípustného řešení duální úlohy. Tudíž nerovnost platí i pro mezní hodnoty.  $\square$

Ukažme si některé základní příklady a vlastnosti duálních kuželů. Obrázek 4.1 je pak graficky ilustruje. Například vlastnost tvrzení 4.18(4) je ilustrována obrázky 4.1a a 4.1c.

**Příklad 4.17.**

- Nezáporný ortant je samoduální, to jest  $(\mathbb{R}_+^n)^* = \mathbb{R}_+^n$  (viz obrázek 4.1a).
- Lorentzův kužel je také samoduální,  $\mathcal{L}^* = \mathcal{L}$  (viz obrázek 4.1b).
- Kužel pozitivně semidefinitních matic je také samoduální; zde skalární součin pozitivně semidefinitních matic definujeme jako  $\langle A, B \rangle := \text{tr}(AB) = \sum_{i,j} a_{ij} b_{ij}$ .  $\square$

**Tvrzení 4.18.** *Platí:*

- (1)  $\mathcal{K}^*$  je uzavřený konvexní kužel.
- (2) Je-li  $\mathcal{K}$  uzavřený konvexní kužel, pak  $(\mathcal{K}^*)^* = \mathcal{K}$ .
- (3) Jsou-li  $\mathcal{K}_1, \mathcal{K}_2$  kužele, pak  $\mathcal{K}_1 \times \mathcal{K}_2$  je kužel a  $(\mathcal{K}_1 \times \mathcal{K}_2)^* = \mathcal{K}_1^* \times \mathcal{K}_2^*$ .
- (4) Jsou-li  $\mathcal{K}_1 \subseteq \mathcal{K}_2$  kužele, pak  $\mathcal{K}_1^* \supseteq \mathcal{K}_2^*$ .

Z těchto vlastností hned vidíme, že úlohou konvexního kuželového programování je i například

$$\min c^T x \text{ za podm. } Ax \geq b, Bx \geq_{\mathcal{K}} d,$$

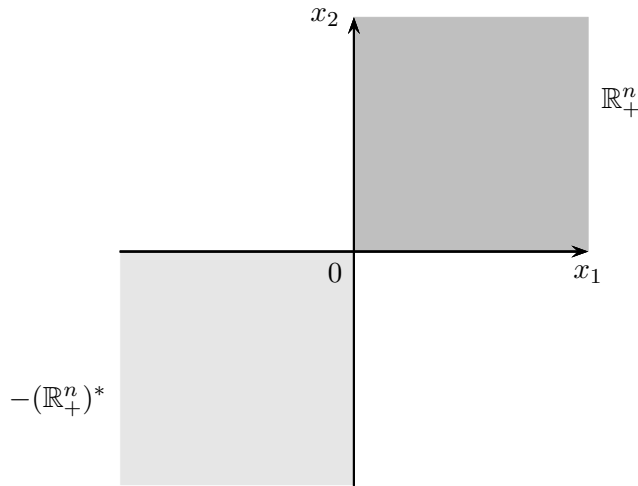
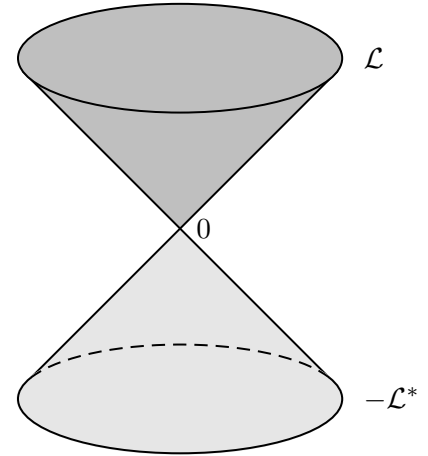
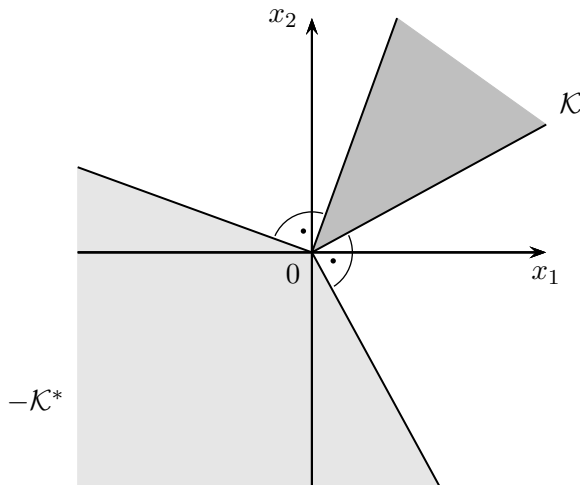
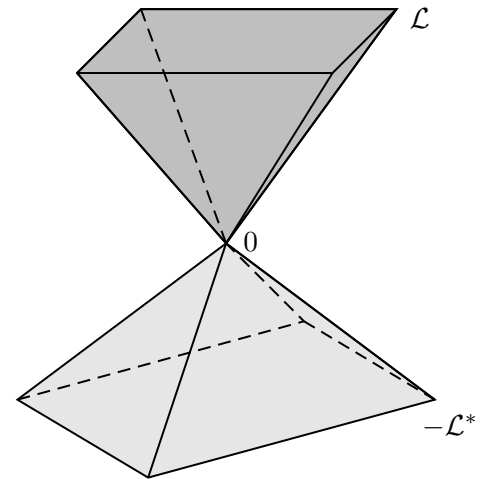
jehož duální úlohou je

$$\max b^T y + d^T z \text{ za podm. } A^T y + B^T z = c, y \geq 0, z \geq_{\mathcal{K}^*} 0.$$

Podobně můžeme uvažovat i jiné kombinace kuželových omezení.

Jak je to se silnou dualitou v konvexním kuželovém programování? V celé obecnosti neplatí a je nutno se vyhnout patologickým případům. Za celkem obecných předpokladů lze silnou dualitu odvodit.



(a) Nezáporný ortant  $\mathbb{R}_+^n$  a jeho duál  $(\mathbb{R}_+^n)^*$  pro  $n = 2$ .(b) Lorentzův kužel  $\mathcal{L}$  a jeho duál  $\mathcal{L}^*$ .(c) Kužel a jeho duál v  $\mathbb{R}^2$ .(d) Zobecněný Lorentzův kužel  $\mathcal{L}$  a jeho duál  $\mathcal{L}^*$ .

Obrázek 4.1: Kužele a duální kužele (pro účely zobrazení jsou duální kužele vynásobené  $-1$ , tedy otočeny dle počátku).

**Věta 4.19** (Silná dualita). *Optimální hodnoty primární a duální úlohy jsou stejné pokud platí jedna z podmínek*

- (1) *primární úloha je striktně přípustná, tj. existuje  $x$  takové, že  $Ax >_{\mathcal{K}} b$ ,*
- (2) *duální úloha je striktně přípustná, tj. existuje  $y >_{\mathcal{K}^*} 0$  takové, že  $A^T y = c$ .*

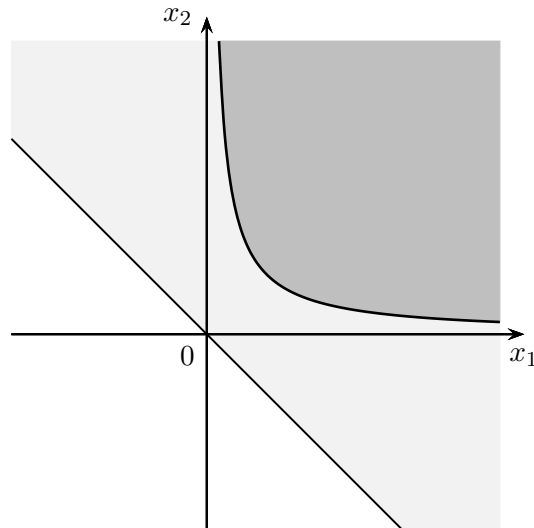
*Důkaz.* Ukážeme jen základní myšlenku pro (1) bez technických detailů; díky dualitě je (2) analogické.

Označme jako  $f^*$  optimální hodnotu a předpokládejme, že  $c \neq 0$  (jinak  $f^* = 0$  a pro  $y = 0$  nastává silná dualita). Definujme množinu

$$\mathcal{M} := \{y = Ax - b; x \in \mathbb{R}^n, c^T x \leq f^*\}.$$

Dá se snadno nahlédnout, že  $\mathcal{M} \cap \text{int}(\mathcal{K}) = \emptyset$ . Jinak totiž existuje  $x$  takové, že  $Ax >_{\mathcal{K}} b$  a  $c^T x \leq f^*$ , a proto malou změnou vektoru  $x$  ve směru  $-c$  docílíme super-optimální hodnoty.

Protože jsou obě množiny  $\mathcal{M}, \mathcal{K}$  konvexní, jdou oddělit nadrovinou  $\lambda^T y = 0$  (nula na pravé straně je důsledek toho, že  $\mathcal{K}$  je kužel). Protože  $\mathcal{K}$  náleží do kladného poloprostoru, tak  $\lambda^T y \geq 0$  pro všechna  $y \in \mathcal{K}$ , z čehož  $\lambda \in \mathcal{K}^*$ . Protože  $\mathcal{M}$  náleží do záporného poloprostoru, tak  $\lambda^T y \leq 0$  pro všechna  $y \in \mathcal{M}$ , čili  $\lambda^T Ax \leq \lambda^T b$  pro všechna  $x$  taková, že  $c^T x \leq f^*$ . To nastane jen když normály  $A^T \lambda, c$  jsou lineárně závislé, tedy  $A^T \lambda = \mu c$  pro  $\mu \geq 0$ .



Obrázek 4.2: (Příklad 4.20) Kuželový kvadratický program, pro který se optimální hodnota nenabyde.

Dá se nahlédnout, že musí  $\mu > 0$ . Pokud totiž  $\mu = 0$ , tak  $A^T \lambda = 0$  a také  $\lambda^T b \geq 0$ . Díky striktní přípustnosti primární úlohy máme  $\tilde{x}$  takové, že  $A\tilde{x} >_{\mathcal{K}} b$ . Protože  $\lambda \geq_{\mathcal{K}^*} 0$ ,  $\lambda \neq 0$ , tak přenásobením získáváme  $\lambda^T(A\tilde{x} - b) > 0$ , neboli  $\lambda^T b < 0$ , spor.

Po normalizaci  $\mu \equiv 1$  máme  $A^T \lambda = c$ . To nám dává duálně přípustné řešení  $\lambda$ , protože splňuje  $A^T \lambda = c$ ,  $\lambda \geq_{\mathcal{K}^*} 0$ . Navíc víme  $\lambda^T b \geq A\lambda^T x = c^T x$  pro všechna  $x$  taková, že  $c^T x \leq f^*$  (tedy i pro  $x$  takové, že  $c^T x = f^*$ ), z čehož  $\lambda^T b \geq f^*$ . Spolu se slabou dualitou máme rovnost.  $\square$

Poznamenejme, že i když platí silná dualita a navíc jsou optimální hodnoty obou úloh (stejně a) konečné, tak se může stát, že optimální hodnota se nenabyde (formálně bychom měli psát „inf“ místo „min“). To ilustruje následující příklad.

**Příklad 4.20.** Uvažujme úlohu typu kuželového kvadratického programování, tj. typu (4.2)

$$\min x_1 \text{ za podm. } \sqrt{(x_1 - x_2)^2 + 1} \leq x_1 + x_2.$$

Umocněním obou stran nerovnice a upravením přepíšeme úlohu na

$$\min x_1 \text{ za podm. } 4x_1x_2 \leq 1, x_1 + x_2 > 0.$$

Přestože je úloha striktně přípustná, optimální hodnota 0 se nenabyde, viz obrázek 4.2.  $\square$

Další příklad ukazuje situaci, kdy předpoklady věty 4.19 nejsou splněny a silná dualita nenastává.

**Příklad 4.21.** Uvažujme úlohu typu kuželového kvadratického programování

$$\min x_2 \text{ za podm. } \sqrt{x_1^2 + x_2^2} \leq x_1.$$

Úpravou podmínek dostaneme tvar

$$\min x_2 \text{ za podm. } x_2 = 0, x_1 \geq 0,$$

ze kterého je patrné, že optimální hodnota je 0 a optimální řešení jsou všechna přípustná řešení, to jest nezáporná první souřadná osa.

Abychom sestavili duální program, přepíšeme primární úlohu do základního tvaru

$$\min x_2 \text{ za podm. } (x_1, x_2, x_1)^T \geq_{\mathcal{L}} 0.$$

Duální úloha pak má tvar

$$\max 0 \text{ za podm. } y_1 + y_3 = 0, y_2 = 1, y \geq_{\mathcal{L}} 0.$$

Nerovnost  $y \geq_{\mathcal{L}} 0$  přepíšeme na  $y_3 \geq \sqrt{y_1^2 + y_2^2}$ , což s podmínkou  $y_1 + y_3 = 0$  vede na  $y_2 = 1$  a spor. Tudíž duální úloha nemá přípustné řešení. A to přestože primární úloha má konečné optimum.  $\square$

### 4.3.2 Kuželové kvadratické programování

Pod pojmem kuželové kvadratické programování rozumíme úlohu kuželového programování s lineárními podmínkami a s podmínkami odpovídajícími Lorentzově kuželi. Pro jednoduchost uvažujme jen jeden Lorentzův kužel a úlohu

$$\min c^T x \text{ za podm. } Ax \geq b, Bx \geq_{\mathcal{L}} d. \quad (4.4)$$

Pokud vyjádříme

$$(B \mid d) = \left( \begin{array}{c|c} D & f \\ \hline p^T & q \end{array} \right)$$

tak podmínku  $Bx \geq_{\mathcal{L}} d$  můžeme přepsat jako  $\|Dx - f\|_2 \leq p^T x - q$ . Tudíž dostáváme explicitní popis úlohy (4.4)

$$\min c^T x \text{ za podm. } Ax \geq b, \|Dx - f\|_2 \leq p^T x - q. \quad (4.5)$$

Připomeňme, že úloha se nedá přímočaře přepsat jako konvexní kvadratický program, ani s konvexními kvadratickými podmínkami. Představuje tak novou třídu úloh, které jsou efektivně řešitelné, ale přitom mají velkou formulační schopnost. Do tvaru (4.5) lze totiž přepsat mnoho zajímavých funkcí a nelineárních podmínek.

**Příklad 4.22** (Příklady kuželového kvadratického programování).

- *Kvadratické podmínky.* Například podmínka  $x^T x \leq z$  se přepíše na  $x^T x + \frac{1}{4}(z-1)^2 \leq \frac{1}{4}(z+1)^2$ , z čehož odmocněním plyne vyjádření  $\|(x^T, \frac{1}{2}(z-1))\|_2 \leq \frac{1}{2}(z+1)$ .
- *Hyperbola.* Podmínka  $x \cdot y \geq 1$  na  $y \geq 0$  se přepíše do tvaru  $\frac{1}{4}(x+y)^2 \geq 1 + \frac{1}{4}(x-y)^2$ , z čehož odmocněním (jež je na  $y \geq 0$  ekvivalentní úpravou) dostaneme  $\|(1, \frac{1}{2}(x-y))\|_2 \leq \frac{1}{2}(x+y)$ .

Na druhou stranu, kuželovou kvadratickou podmínkou není například  $e^x \leq z$  atp.

Duální úloha

$$\max b^T y + d^T z \text{ za podm. } A^T y + B^T z = c, y \geq 0, z \geq_{\mathcal{L}^*} 0$$

pak při vyjádření  $z = (u^T, v)^T$  dostane tvar

$$\max b^T y + d^T z \text{ za podm. } A^T y + B^T z = c, y \geq 0, v \geq \|u\|_2.$$

Duální úloha tak představuje rovněž kuželový kvadratický program.

### 4.3.3 Semidefinitní programování

Použijeme-li kužel pozitivně semidefinitních matic v úloze kuželového programování (4.1), dostaneme třídu úloh tzv. *semidefinitního programování*

$$\min c^T x \text{ za podm. } \sum_{k=1}^n x_k A^{(k)} \succeq B, \quad (4.6)$$

kde  $c \in \mathbb{R}^n$ , matice  $A^{(1)}, \dots, A^{(n)}, B \in \mathbb{R}^{m \times m}$  jsou symetrické a relace  $A \succeq B$  značí, že  $A - B$  je pozitivně semidefinitní matice.<sup>1)</sup>

Jak bude vypadat duální úloha? Podle našeho vzoru (4.3) bude mít úloha  $m^2$  proměnných, čili představují neznámou matici  $Y \in \mathbb{R}^{m \times m}$ . Duální účelová funkce je  $\sum_{i,j} b_{ij} y_{ij}$ , rovnice mají tvar  $\sum_{i,j} a_{ij}^{(k)} y_{ij} = c_k$  a podmínka  $Y \succeq_{\mathcal{K}^*} 0$  je pak zřejmě tvaru  $Y \succeq 0$ . Celkem dostáváme duální úlohu jako

$$\max \text{tr}(BY) \text{ za podm. } \text{tr}(A^{(k)}Y) = c_k, k = 1, \dots, n, Y \succeq 0. \quad (4.7)$$

<sup>1)</sup>Relace  $\succeq$  určuje částečné uspořádání, které se někdy nazývá Löwnerovo uspořádání podle amerického matematika českého původu Karla Löwnera.

**Příklad 4.23** (Příklady semidefinitního programování).

- *Lineární podmínky.* Lineární podmínky  $Ax \leq b$  můžeme přepsat jako semidefinitní podmínky takto:

$$\begin{pmatrix} b_1 - A_{1*}x & 0 & \dots & 0 \\ 0 & b_2 - A_{2*}x & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & b_m - A_{m*}x \end{pmatrix} \succeq 0.$$

- *Kuželové kvadratické podmínky.* Jdou přepsat jako semidefinitní podmínky. Stačí to ukázat pro podmínku  $\|x\|_2 \leq z$ , ostatní dostaneme lineární substitucí. Platí

$$\|x\|_2 \leq z \Leftrightarrow \begin{pmatrix} z \cdot I_n & x \\ x^T & z \end{pmatrix} \succeq 0. \quad (4.8)$$

*Důkaz.* Pro  $z = 0$  ekvivalence platí, tudíž se omezíme na případ  $z > 0$ . Matici budeme nahlížet jako matici kvadratické formy a upravíme ji na blokově diagonální tvar pomocí řádkových a analogicky sloupcových úprav. Odečtením  $\frac{1}{z}x^T$ -násobku prvního blokového řádku od druhého a stejně na sloupce dostaneme

$$\begin{pmatrix} z \cdot I_n & x \\ x^T & z \end{pmatrix} \sim \begin{pmatrix} z \cdot I_n & 0 \\ 0 & z - \frac{1}{z}x^T x \end{pmatrix}.$$

Tato matice je pozitivně semidefinitní právě tehdy, když  $z > 0$  a  $x^T x \leq z^2$ , neboli po odmocnění  $\|x\|_2 \leq z$ .

- *Vlastní čísla.* Řada vztahů okolo vlastních čísel se dá vyjádřit jako semidefinitní podmínka. Například největší vlastní číslo  $\lambda_{\max}$  symetrické matice  $A \in \mathbb{R}^{n \times n}$ :

$$\lambda_{\max} = \min z \text{ za podm. } z \cdot I_n \succeq A.$$

**Příklad 4.24.** Uvažujme opět problém výběru portfolia (příklad 4.9)

$$\max c^T x \text{ za podm. } e^T x = K, x \geq 0,$$

kde  $c$  je náhodný vektor se střední hodnotou  $\tilde{c} := E c$  a kovarianční maticí  $\Sigma := \text{cov } c = E(c - \tilde{c})(c - \tilde{c})^T$ . Předpokládejme, že portfolio  $\tilde{x}$  je již vybráno, ale pro kovarianční matici máme jen intervalový odhad  $\Sigma_1 \leq \Sigma \leq \Sigma_2$ . Jaké je riziko portfolia  $\tilde{x}$ ? To je dáno rozptylem výnosu  $c^T \tilde{x}$ , jenž je roven  $\tilde{x}^T \Sigma \tilde{x}$ . Tudíž největší rozptyl spočítáme semidefinitním programem

$$\max \tilde{x}^T \Sigma \tilde{x} \text{ za podm. } \Sigma_1 \leq \Sigma \leq \Sigma_2, \Sigma \succeq 0.$$

Účelová funkce je lineární v proměnných  $\Sigma$  a podmínky snadno přepíšeme do tvaru (4.7) podle příkladu 4.23.  $\square$

## 4.4 Výpočetní složitost

Úlohy konvexní optimalizace jsou považovány vesměs za efektivně řešitelné. Vskutku, jak naznačíme v sekci 4.4.1, za celkem obecných předpokladů jsou konvexní úlohy řešitelné v polynomiálním čase. Na druhou stranu, neplatí to pro všechny konvexní úlohy. V sekci 4.4.2 se seznámíme s jednou výpočetně složitou třídou úloh.

### 4.4.1 Dobré zprávy – elipsoidová metoda

Úloha konvexní optimalizace  $\min_{x \in M} f(x)$  je řešitelná polynomiálně pomocí elipsoidové metody za celkem obecných předpokladů. Tento výsledek je spíše teoretický, k praktickému řešení se spíše hodí jiné polynomiální metody, jako jsou metody vnitřních bodů.

Elipsoidová metoda slouží k nalezení přípustného řešení, ale to je v zásadě použitelné i pro nalezení optima. Základní myšlenka, jak nalézt bod  $x \in M$  nebo určit, že  $M$  je prázdná, je následující. Nejprve nalezneme dostatečně velký elipsoid  $\mathcal{E}$ , který obsahuje celou množinu  $M$ . Pak otestujeme, zda střed  $c$  elipsoidu  $\mathcal{E}$  se nalézá v  $M$ . Pokud ne, což je ten zajímavější případ, tak sestrojíme nadrovinu, která prochází bodem  $c$  a neprotíná  $M$  (tedy něco jako tečnou nadrovinu, ale nemusí být přesně tečná, ale může být posunutá až do  $c$ ), tudíž  $M$  se nalézá v poloprostoru  $a^T x \leq b$ . Pak sestrojíme co do velikosti menší elipsoid, který obsahuje průnik  $\mathcal{E} \cap \{x; a^T x \leq b\}$ . Tento postup opakujeme, a protože se velikost elipsoidu exponenciálně zmenšuje, tak po polynomiální počtu kroků nalezneme přípustný bod nebo musí  $M = \emptyset$ .

Aby výše popsany algoritmus byl korektní a polynomiální, musí být splněny určité předpoklady, jako je

- Množina přípustných řešení  $M$  nesmí být moc placatá ani moc velká. Musí existovat „rozumě“ velká čísla  $r, R > 0$  taková, že  $M$  obsahuje kruh o poloměru  $r$  a přitom  $M$  je obsaženo v kruhu  $\{x; \|x\|_2 \leq R\}$ .
- *Separáční vlastnost.* Pro každé  $x^* \in \mathbb{R}^n$  musíme umět rozhodnout v polynomiálním čase zda  $x^* \in M$ . Pokud  $x^* \notin M$ , tak musíme navíc umět najít vektor  $a \neq 0$  takový, že  $a^T x^* \geq \sup_{x \in M} a^T x$ . To nám dá nadrovinu  $a^T x = a^T x^*$ , která splňuje  $a^T x \leq a^T x^*$  pro všechna  $x \in M$ .

Separáční vlastnost lze většinou implementovat takto. Pokud popis  $M$  obsahuje podmínku  $g(x) \leq 0$ , která není v  $x^*$  splněná, pak lze volit  $a := \nabla g(x^*)$ , neboť z předpokladu  $g(x^*) > 0$  a z konvexity  $g$  je  $a^T(x - x^*) = \nabla g(x^*)^T(x - x^*) \leq g(x) - g(x^*) < g(x) \leq 0$  pro všechna  $x \in M$ .

Elipsoidová metoda řeší problém s danou přesností. Optimum může být například  $\sqrt{2}$ , což nelze přesně reprezentovat v klasickém výpočetním modelu, proto ani přesné optimum nemůžeme obecně požadovat. Pro vyjádření přesnosti je však ještě zapotřebí určit míru nepřipustnosti. Pro lineární podmínky  $Ax \leq b$  lze použít například míru

$$\min z \text{ za podm. } Ax - ze \leq b, z \geq 0,$$

a pro semidefinitní podmínku  $\sum_{k=1}^n x_k A^{(k)} \succeq B$  lze použít míru

$$\min z \text{ za podm. } \sum_{k=1}^n x_k A^{(k)} + zI_m \succeq B, z \geq 0.$$

**Příklad 4.25.** V určitých případech dává elipsoidová metoda polynomiální algoritmus i pro úlohy s exponenciálně či dokonce nekonečně mnoha omezeními. Buď například  $M$  jednotková koule popsaná tečnými nadrovinami

$$M = \{x \in \mathbb{R}^n; a^T x \leq 1, \forall a: \|a\|_2 = 1\}.$$

Abychom ověřili, zda daný bod  $x^* \in \mathbb{R}^n$  náleží do množiny  $M$ , není potřeba procházet všech nekonečně mnoho omezení, ale stačí zkontrolovat to případně porušené, což je případ  $a = \frac{1}{\|x^*\|_2} x^*$ .  $\square$

### 4.4.2 Špatné zprávy – kositivní programování

Ne každá konvexní úloha je výpočetně jednoduchá (polynomiální). Zde předvedeme konvexní úlohu, která je NP-těžká. Označme jako

$$\mathcal{C} := \{A \in \mathbb{R}^{n \times n}; A = A^T, x^T A x \geq 0 \forall x \geq 0\}$$

konvexní kužel *kositivních matic* a jako

$$\mathcal{C}^* := \text{conv}\{xx^T; x \geq 0\}$$

k němu duální kužel *úplně kositivních matic*. Zřejmě množina  $\mathcal{C}$  obsahuje nezáporné symetrické matice i pozitivně semidefinitní matice, nicméně obsahuje i další. Podobně množina  $\mathcal{C}^*$  obsahuje matice, které jsou

nutně nezáporné pozitivně semidefinitní, ale ne každá takováto matice náleží do  $\mathcal{C}^*$ . Poznamenejme, že dokonce už jen testování, zda daná matice je kopolitivní, je co-NP-úplný problém [Murty and Kabadi, 1987]. Testování, zda daná matice je úplně pozitivní je NP-těžké [Dickinson and Gijben, 2014], ale příslušnost třídy NP dosud nebyla prokázána.

Uvažujme úlohu tzv. *kopolitivního programování* [Dür, 2010]

$$\min \operatorname{tr}(CX) \text{ za podm. } \operatorname{tr}(A_i X) = b_i, \quad i = 1, \dots, m, \quad X \in \mathcal{C}, \quad (4.9)$$

kde  $C, A_1, \dots, A_k \in \mathbb{R}^{n \times n}$  a  $b_1, \dots, b_k \in \mathbb{R}$ . Účelová funkce

$$\operatorname{tr}(CX) = \sum_{i,j} c_{ij} x_{ij}$$

je obecná lineární funkce v proměnných  $X$  a analogicky rovnice jsou lineární. Jedinou nelineární podmínkou je  $X \in \mathcal{C}$ , což z úlohy dělá konvexní kuželovou úlohu. Pro srovnání uvažujme i konvexní úlohu s podmínkou na úplnou pozitivnost matice  $X$ :

$$\min \operatorname{tr}(CX) \text{ za podm. } \operatorname{tr}(A_i X) = b_i, \quad i = 1, \dots, m, \quad X \in \mathcal{C}^*. \quad (4.10)$$

Obě úlohy jsou konvexní, ale přesto NP-těžké. Předvedeme důkaz pro druhou z nich.

**Věta 4.26.** *Úloha (4.10) je NP-těžká.*

Důkaz plyne z toho, že úlohou můžeme zformulovat problém nalezení velikosti maximální nezávislé množiny v grafu. Buď  $G = (V, E)$  graf s  $n$  vrcholy a necht'  $\alpha$  značí velikost maximální nezávislé množiny v grafu  $G$ , tj. velikost největší (co do kardinality) množiny  $I \subseteq V$  takové, že  $i, j \in I \Rightarrow \{i, j\} \notin E$ .

**Věta 4.27.** *Platí*

$$\alpha = \max \operatorname{tr}(ee^T X) \text{ za podm. } x_{ij} = 0 \quad \forall \{i, j\} \in E, \quad \operatorname{tr}(X) = 1, \quad X \in \mathcal{C}^*. \quad (4.11)$$

*Důkaz.* Uvažujme konvexní kužel

$$\{X \in \mathcal{C}^*; x_{ij} = 0 \quad \forall \{i, j\} \in E\}.$$

Jeho extrémní hrany jsou matice tvaru  $xx^T$ , kde  $x \geq 0$  a support vektoru  $x$  (tj. počet kladných složek) odpovídá nezávislé množině v grafu  $G$ . Podmínka  $\operatorname{tr}(X) = 1$  v úloze (4.11) pak jen normuje vektory tohoto kužele. Protože účelová funkce úlohy (4.11) je lineární, optimální řešení se musí nabýt v extrémním bodě. Tudíž můžeme předpokládat, že optimum  $X^*$  je tvaru

$$X^* = x^* x^{*T}, \quad x^* \geq 0, \quad \|x^*\| = 1,$$

neboť podmínka  $\operatorname{tr}(X^*) = 1$  je ekvivalentní s  $1 = \sqrt{\operatorname{tr}(X^*)} = \sqrt{\operatorname{tr}(x^* x^{*T})} = \sqrt{\operatorname{tr}(x^{*T} x^*)} = \sqrt{x^{*T} x^*} = \|x^*\|$ . Support vektoru  $x^*$  odpovídá nezávislé množině o velikosti  $\alpha(x^*)$ . Označme  $\tilde{x} \in \mathbb{R}^{\alpha(x^*)}$  restrikci vektoru  $x$  jen na kladné složky, nulové jsou vynechány. Pak optimální hodnota  $h$  úlohy (4.11) se dá vyjádřit jako

$$h = \max_{\|\tilde{x}\|=1} (e^T \tilde{x})^2, \quad \tilde{x} \geq 0,$$

neboť  $\operatorname{tr}(ee^T X) = \operatorname{tr}(ee^T xx^T) = \operatorname{tr}(e^T xx^T e) = (e^T x)^2$ . Není těžké nahlédnout, že optimálním řešením této úlohy je vektor stejných složek, tedy  $\tilde{x}^* = \alpha(x^*)^{-1/2} e$ . Tudíž  $h = (e^T \tilde{x}^*)^2 = (\alpha(x^*)^{-1/2} \alpha(x^*))^2 = \alpha(x^*)$ . Proto  $h$  bude rovno velikosti největší nezávislé množiny v grafu  $G$ .  $\square$

## 4.5 Aplikace

### 4.5.1 Robustní PCA

Buď  $A \in \mathbb{R}^{m \times n}$  matice reprezentující data. Naše úloha je nyní určit této matice klíčovou podstatu dat; pokud matice reprezentuje obrázek, může se jednat o kompresi, rozpoznání vzorku (např. obličeje) v obrázku, rekonstrukce poškozeného obrázku, atp.

K tomuto může posloužit SVD rozklad matice  $A$ , nicméně pro některé účely je nedostatečný. My si zformulujeme úlohu jako problém robustní PCA (principal component analysis):

→ Rozděli matici na součet  $A = L + S$ , kde  $L$  má malou hodnotu a  $S$  je řídká.

Pak  $L$  reprezentuje podstatnou část dat a  $S$  typicky chybná měření. Takto je úloha vágně formulovaná a těžká k řešení, proto se aproximuje jako optimalizační problém

$$\min \|L\|_* + \|S\|_{\ell_1} \quad \text{za podm. } A = L + S, \quad (4.12)$$

kde  $\|S\|_{\ell_1}$  je součtová norma definovaná jako

$$\|S\|_{\ell_1} := \sum_{i,j} |s_{ij}|,$$

a  $\|L\|_*$  je nukleární norma definovaná jako součet singulárních čísel

$$\|L\|_* := \sum_i \sigma_i(L).$$

Z možných norem se pro aproximaci hodnoty matice používá nukleární norma, protože je nejlepším konvexním dolním odhadem pro hodnotu uvnitř jednotkové koule. Podobně součtová norma je dobrou aproximací řídkosti matice, tj. počtu jejích nenulových složek.

Úloha (4.12) je úlohou konvexní optimalizace, neboť norma je konvexní funkcí. Tudíž se dá řešit poměrně efektivně, i když popsat nejlepší algoritmy jednoduché není.

### Rozpoznávání pozadí a popředí sekvence obrázků

Jak se může popsaná technika použít na rozpoznávání pozadí videa či sekvence obrázků? Necht' sloupce matice  $A$  odpovídají jednotlivým obrázkům. Pak se dá čekat, že matice  $L$  bude zhruba odpovídat pozadí, protože to se nepohybuje a matice tak má malou hodnotu. Matice  $S$  pak zachycuje popředí.

Více viz:

- <http://sites.google.com/site/rpcaforegrounddetection/>
- E. Candes, X. Li, Y. Ma, J. Wright, *Robust Principal Component Analysis?*, J. ACM 58(3), 2011

### 4.5.2 Minimální opsaný elipsoid

Tzv. *minimum volume enclosing ellipsoid* je úloha nalezení elipsoidu s minimálním objemem a který obsahuje daný konvexní polyedr [Todd, 2016].

Necht'  $x_1, \dots, x_m \in \mathbb{R}^n$  jsou vrcholy (omezeného) konvexního polyedru, který chceme opsat elipsoidem. Pro jednoduchost se omezíme na plnodimenzionální elipsoid se středem v počátku. Ten má popis  $x^T H x \leq 1$ , kde  $H \in \mathbb{R}^{n \times n}$  je pozitivně definitní matice (zkráceně píšeme  $H \succ 0$ ). Objem elipsoidu je nepřímo úměrný  $\det(H)$ . Proto můžeme úlohu formulovat jako

$$\min -\det(H) \quad \text{za podm. } H \succ 0, x_i^T H x_i \leq 1, i = 1, \dots, m,$$

kde neznámou je matice  $H$ . Tato úloha není konvexní, proto účelovou funkci přeškálujeme logaritmem (optimální řešení se nezmění)

$$\min -\log \det(H) \quad \text{za podm. } H \succ 0, x_i^T H x_i \leq 1, i = 1, \dots, m.$$

Funkce  $-\log \det(H)$  je ryze konvexní na množině pozitivně definitních matic, čili úloha je již konvexní a je efektivně řešitelná vhodnými technikami.





## Kapitola 5

# Karush–Kuhn–Tuckerovy podmínky optimality

V této kapitole uvažujeme úlohou optimalizace

$$\min f(x) \text{ za podm. } x \in M$$

kde  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  je diferencovatelná a množina přípustných řešení  $M \subseteq \mathbb{R}^n$  je popsána soustavou

$$\begin{aligned} g_j(x) &\leq 0, & j = 1, \dots, J, \\ h_\ell(x) &= 0, & \ell = 1, \dots, L, \end{aligned}$$

kde  $g_j(x), h_\ell(x): \mathbb{R}^n \rightarrow \mathbb{R}$ .

Víme z věty 2.1, že pokud  $M = \mathbb{R}^n$ , tak nutnou podmínkou pro  $x \in \mathbb{R}^n$  aby bylo optimálním řešením je  $\nabla f(x) = 0$ ; pokud je  $f(x)$  konvexní, tak je to i postačující podmínka (věta 4.4).

Tato kapitola se zabývá zobecněním této podmínky na případ optimalizace s omezeními, tzv. Karush–Kuhn–Tuckerovy podmínky. Nejprve připomeneme rovnicový případ a pak se budeme zabývat obecným tvarem.

### Rovnicový případ

Na chvíli uvažujme úlohu pouze s rovnicemi

$$\min f(x) \text{ za podm. } h(x) = 0. \tag{5.1}$$

Buď  $x^*$  přípustný bod a ptáme se, za jakých podmínek je optimálním. Nejprve prozkoumáme situaci, kdy funkce z omezení jsou lineární.

**Tvrzení 5.1.** *Je-li bod  $x^* \in \mathbb{R}^n$  lokálním optímem úlohy*

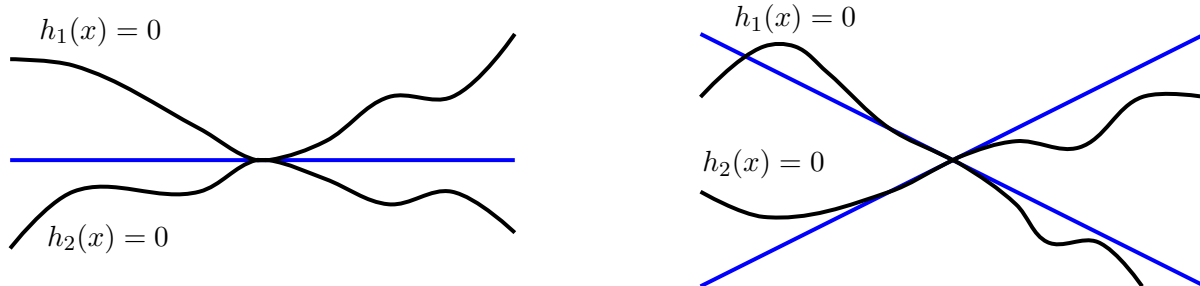
$$\min f(x) \text{ za podm. } Ax = b,$$

*pak  $\nabla f(x^*) \in \mathcal{R}(A)$ .*

*Důkaz.* Množina přípustných řešení, to jest řešení rovnice  $Ax = b$ , představuje afinní podprostor  $x^* + \text{Ker}(A)$ . Tvoří-li sloupce matice  $B$  nějakou bázi  $\text{Ker}(A)$ , pak řešení rovnice  $Ax = b$  se dají také vyjádřit jako  $x = x^* + Bv$ ,  $v \in \mathbb{R}^k$ . Dosazením dostaneme úlohu bez omezení

$$\min f(x^* + Bv) \text{ za podm. } v \in \mathbb{R}^k.$$

Nutnou podmínkou pro lokální minimum v bodě  $v = 0$  je podle věty 2.1 nulový gradient, to jest  $\nabla f(x^*)^T B = 0^T$ . Jinými slovy  $\nabla f(x^*) \in \text{Ker}(A)^\perp = \mathcal{R}(A)$ .  $\square$



(a) Degenerovaný případ: průnik křivek je bod, ale průnik tečen je přímka.

(b) Regulární případ: průnik křivek je bod, průnik tečen rovněž.

Obrázek 5.1: Linearizace nelineárních podmínek – křivky nahradíme tečnami.

Základní myšlenka nyní spočívá v linearizaci nelineárních funkcí  $h_\ell$ . Rovnici  $h_\ell(x) = 0$  nahradíme tečnou nadrovinou v bodě  $x^*$ :

$$\nabla h_\ell(x^*)^T(x - x^*) = 0,$$

takže linearizované podmínky souhrnně můžeme vyjádřit jako  $A(x - x^*) = 0$ . Aby bylo  $x^*$  optimální, musí být gradient účelové funkce  $\nabla f(x^*)$  kolmý na průnik těchto tečen, tedy gradient  $\nabla f(x^*)$  musí být lineární kombinací gradientů tečných nadrovin  $\nabla h_\ell(x^*)$ . Podle Tvzení 5.1 tedy musí  $\nabla f(x^*) \in \mathcal{R}(A)$ . To vede na podmínku

$$\nabla f(x^*) + \sum_{\ell=1}^L \nabla h_\ell(x^*) \mu_\ell = 0.$$

Jak ilustruje obrázek 5.1, tento postup je obecně *špatně*, protože může nastat patologický případ jako na obrázku. Proto potřebujeme zaručit, že takovýto případ nenastane. A právě to zaručuje například podmínka lineární nezávislosti gradientů  $\nabla h_\ell(x^*)$ .

**Věta 5.2.** *Budte  $\nabla h_\ell(x^*)$ ,  $\ell = 1, \dots, L$ , lineárně nezávislé. Je-li  $x^*$  lokální optimum, pak existuje  $\mu \in \mathbb{R}^L$  takové, že*

$$\nabla f(x^*) + \nabla h(x^*)\mu = 0.$$

*Důkaz.* Viz základní přednáška z matematické analýzy. □

Koeficienty  $\mu_1, \dots, \mu_L$  se nazývají *Lagrangeovy multiplifikátory*. Podmínka optimality z věty je nutnou podmínkou, což je výhodné, protože nám umožní omezit množinu přípustných řešení na menší (v ideálním případě jednorvkovou) množinu kandidátů na optimum.

## Obecný případ

Nyní uvažujeme úlohu optimalizace s rovnicemi i nerovnicemi. Pro přípustný bod  $x$  definujeme *množinu aktivních podmínek* jako množinu těch nerovnic, které jsou splněny s rovností

$$I(x) = \{j; g_j(x) = 0\}.$$

**Věta 5.3** (KKT-podmínky). *Budte  $\nabla h_\ell(x^*)$ ,  $\ell = 1, \dots, L$ ,  $\nabla g_j(x^*)$ ,  $j \in I(x^*)$ , dohromady lineárně nezávislé. Je-li  $x^*$  lokální optimum, pak existují  $\lambda \in \mathbb{R}^J$ ,  $\lambda \geq 0$ , a  $\mu \in \mathbb{R}^L$  takové, že*

$$\nabla f(x^*) + \nabla h(x^*)\mu + \nabla g(x^*)\lambda = 0, \quad (5.2)$$

$$\lambda^T g(x^*) = 0. \quad (5.3)$$

*Poznámka.* Podmínka (5.3) se nazývá *podmínkou komplementarity*, protože ekvivalentně říká, že pro každé  $j = 1, \dots, J$  je  $\lambda_j = 0$  nebo  $g_j(x^*) = 0$ . Pokud  $g_j(x^*) < 0$ , tak musí  $\lambda_j = 0$ , a tím pádem proměnná  $\lambda_j$  z podmínek vypadává, což odpovídá tomu, že  $x^*$  není na hranici omezení této funkce. Naopak, pokud  $g_j(x^*) = 0$ , tak podmínka komplementarity neklade na  $\lambda_j$  žádné omezení. Souhrnem tedy podmínka komplementarity říká, že uvažujeme Lagrangeovy multiplikátory  $\lambda_j$  pouze pro aktivní podmínky.

*Důkaz. (Hlavní myšlenka.)* Úlohu linearizujeme v tom smyslu, že účelovou funkci a funkce z podmínek nahradíme tečnými nadrovinami v bodě  $x^*$ . To vede na úlohu lineárního programování

$$\begin{aligned} \min \quad & \nabla f(x^*)^T x \quad \text{za podm.} \quad \nabla g_j(x^*)^T (x - x^*) \leq 0, \quad j \in I(x^*), \\ & \nabla h_\ell(x^*)^T (x - x^*) = 0, \quad \ell = 1, \dots, L. \end{aligned}$$

Díky předpokladu na lineární nezávislost gradientů zůstane  $x^*$  optimálním (tento krok by vyžadoval podrobnější vysvětlení). Duální úloha k této úloze lineárního programování je

$$\begin{aligned} \max \quad & \sum_{\ell=1}^L (\nabla h_\ell(x^*)^T x^*) \mu_\ell + \sum_{j \in I(x^*)} (\nabla g_j(x^*)^T x^*) \lambda_j \quad \text{za podm.} \\ & \nabla f(x^*) + \sum_{\ell=1}^L \nabla h_\ell(x^*) \mu_\ell + \sum_{j \in I(x^*)} \nabla g_j(x^*) \lambda_j = 0, \\ & \lambda_j \geq 0, \quad j \in I(x^*). \end{aligned}$$

Protože primární úloha má optimum, musí být duální přípustná. Pro  $j \notin I(x^*)$  definujeme  $\lambda_j := 0$  a dostáváme, že i úloha

$$\begin{aligned} \max \quad & (x^*)^T \nabla g(x^*) \lambda + (x^*)^T \nabla h(x^*) \mu \quad \text{za podm.} \\ & \nabla f(x^*) + \nabla h(x^*) \mu + \nabla g(x^*) \lambda = 0, \\ & \lambda \geq 0 \end{aligned}$$

je přípustná. Tudíž existují  $\lambda \geq 0, \mu$  splňující (5.2). Podmínka (5.3) platí jednoduše proto, že pro  $j \in I(x^*)$  je  $g_j(x^*) = 0$  z definice, a pro  $j \notin I(x^*)$  jsme dodefinovali  $\lambda_j = 0$ .  $\square$

Podmínky (5.2)–(5.3) se nazývají *Karush–Kuhn–Tuckerovy podmínky* [Karush, 1939; Kuhn and Tucker, 1951], zkráceně *KKT-podmínky*.<sup>1)</sup>

Jelikož podmínka na lineární nezávislost gradientů se špatně ověřuje (pro neznámé  $x^*$ ), existují jiné varianty; většinou jde o jednodušší ale silnější předpoklady. Jedna z nich je *Slaterova podmínka*

$$\exists x^0 \in M : g(x^0) < 0.$$

**Věta 5.4.** *Uvažujme úlohu*

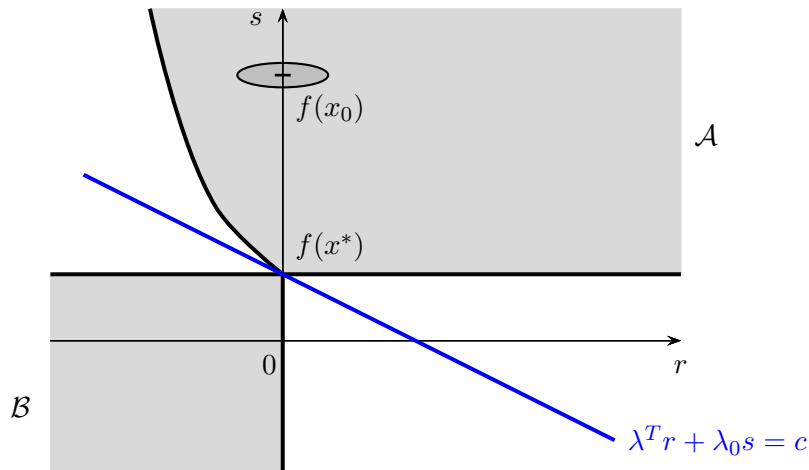
$$\min f(x) \quad \text{za podm.} \quad g(x) \leq 0, \quad x \in M,$$

kde  $f(x), g_j(x)$  jsou konvexní funkce a  $M$  konvexní množina. Nechť platí Slaterova podmínka. Je-li  $x^*$  minimem původní úlohy, pak existuje  $\lambda \geq 0$  takové, že  $x^*$  je minimem úlohy

$$\min f(x) + \lambda^T g(x) \quad \text{za podm.} \quad x \in M \tag{5.4}$$

a navíc platí  $\lambda^T g(x^*) = 0$ .

<sup>1)</sup>Podmínky zavedli a zpopularizovali Harold W. Kuhn a Albert W. Tucker roku 1951. Nezávisle na nich je (s ne tak silnými výsledky) objevil už dříve roku 1939 William Karush ve své diplomové práci, která se zabývala konečně-dimenzionální úlohou variačního počtu. Poté, co se Kuhn roku 1974 dozvěděl o Karushově práci, tak se zasadil o přidání jeho jména do názvu těchto podmínek.



Obrázek 5.2: Ilustrace k důkazu věty 5.4.

*Důkaz.* Definujme množiny

$$\begin{aligned}\mathcal{A} &:= \{(r, s) \in \mathbb{R}^J \times \mathbb{R}; r \geq g(x), s \geq f(x), x \in M\}, \\ \mathcal{B} &:= \{(r, s) \in \mathbb{R}^J \times \mathbb{R}; r \leq 0, s \leq f(x^*)\},\end{aligned}$$

viz obrázek 5.2. Obě množiny jsou konvexní a mají prázdný průnik vnitřků, protože jinak by existoval bod  $x \in M$  takový, že  $g(x) < 0$  a  $f(x) < f(x^*)$ . Můžeme je tedy oddělit nadrovinou tvaru  $\lambda^T r + \lambda_0 s = c$ , kde  $(\lambda, \lambda_0) \neq 0$ . Oddělující vlastnost znamená:

$$\begin{aligned}\forall (r, s) \in \mathcal{A} : \lambda^T r + \lambda_0 s &\geq c, \\ \forall (r, s) \in \mathcal{B} : \lambda^T r + \lambda_0 s &\leq c.\end{aligned}$$

Protože  $(0, f(x^*)) \in \mathcal{A} \cap \mathcal{B}$ , leží tento bod i na oddělující nadrovině, z čehož  $c = \lambda_0 f(x^*)$ . Podobně  $(g(x^*), f(x^*)) \in \mathcal{A} \cap \mathcal{B}$ , leží tento bod tedy na oddělující nadrovině. Dosazením

$$\lambda^T g(x^*) + \lambda_0 f(x^*) = c = \lambda_0 f(x^*),$$

což dává podmínku komplementarity  $\lambda^T g(x^*) = 0$ .

Protože pro každé  $i$  je  $(-e_i, f(x^*)) \in \mathcal{B}$ , musí tento bod ležet v záporném poloprostoru, tedy  $-\lambda^T e_i + \lambda_0 f(x^*) \leq c$ , z čehož  $\lambda_i \geq 0$ . To znamená  $\lambda \geq 0$ . Analogicky odvodíme  $\lambda_0 \geq 0$ : protože  $(0, f(x^*) - 1) \in \mathcal{B}$ , dostáváme  $\lambda^T 0 + \lambda_0(f(x^*) - 1) \leq c$ , z čehož  $\lambda_0 \geq 0$ .

Protože  $g(x^0) < 0$ , platí  $(r, f(x^0)) \in \mathcal{A}$  pro všechna  $r$  z okolí 0. Tudíž oddělující nadrovina nemůže být vertikální, což znamená  $\lambda_0 \neq 0$ ; bez újmy na obecnost normujme tak, že  $\lambda_0 = 1$ . Formální důkaz této vlastnosti: Pro spor předpokládejme  $\lambda_0 = 0$ . Nyní tedy  $c = 0$  a díky  $\lambda \neq 0$  existuje  $i$  takové, že  $\lambda_i > 0$ . Pak ale pro bod  $(-\varepsilon e_i, f(x^0)) \in \mathcal{A}$ , kde  $\varepsilon > 0$  je dost malé, dostáváme dosazením do nerovnice  $-\varepsilon \lambda_i \geq 0$ , spor.

Pro každé  $x \in M$  je  $(g(x), f(x)) \in \mathcal{A}$ , tedy splňuje

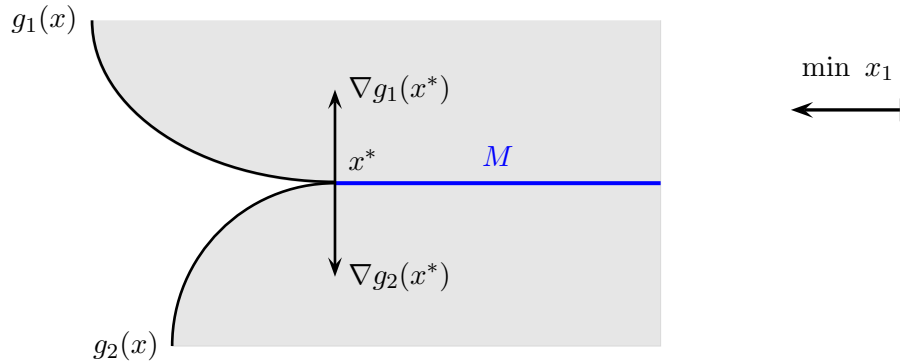
$$\lambda^T g(x) + f(x) \geq c = \lambda^T g(x^*) + f(x^*).$$

Právě jsme dokázali, že  $x^*$  je minimem úlohy (5.4). □

Aplikujeme-li podmínky optimality z věty 2.1 na úlohu (5.4), dostáváme požadované KKT-podmínky:

**Důsledek 5.5.** *Pokud pro konvexní úlohu*

$$\min f(x) \text{ za podm. } g(x) \leq 0$$



Obrázek 5.3: (Příklad 5.7) Věta o KKT podmínkách (důsledek 5.5) selže bez Slaterovy podmínky.

platí Slaterova podmínka a  $x^*$  je minimem, pak existuje  $\lambda \geq 0$  takové, že platí KKT-podmínky

$$\nabla f(x^*) + \nabla g(x^*)\lambda = 0, \quad (5.5a)$$

$$\lambda^T g(x^*) = 0. \quad (5.5b)$$

Nebo obecněji s lineárními rovnicemi.

**Důsledek 5.6.** Pokud pro konvexní úlohu

$$\min f(x) \text{ za podm. } g(x) \leq 0, \quad Ax = b$$

platí Slaterova podmínka a  $x^*$  je minimem, pak existují  $\lambda \geq 0$  a  $\mu$  takové, že platí KKT-podmínky

$$\nabla f(x^*) + \nabla g(x^*)\lambda + A^T \mu = 0,$$

$$\lambda^T g(x^*) = 0.$$

**Příklad 5.7.** Bez Slaterovy podmínky věta o KKT podmínkách (důsledek 5.5) obecně neplatí. Uvažujme příklad úlohy  $\min_{x \in M} x_1$  ilustrovaný na obrázku 5.3. Dvě omezení definují jako množinu přípustných řešení polopřímku, vycházející z bodu  $x^*$ . Bod  $x^*$  je optimem. KKT podmínky zde mají tvar  $-\nabla f(x^*) = \nabla g(x^*)\lambda$ , ale bod  $x^*$  je nespĺňuje, protože gradienty  $\nabla g_1(x^*) = (0, -1)^T$ ,  $\nabla g_2(x^*) = (0, 1)^T$  generují vertikální přímku, ve které se opačný gradient účelové funkce  $-\nabla f(x^*) = (-1, 0)^T$  nenachází.  $\square$

V optimalizaci jsou prakticky důležitější nutné podmínky optimality než postačující, protože obvykle výrazně zužují množinu potenciálních kandidátů na optimum a pomáhají tak urychlit nalezení optima. Pro úplnost ukážeme, že KKT podmínky představují i postačující podmínku pro optimalitu za celkem obecných předpokladů.

**Věta 5.8** (postačující KKT-podmínky). Bud'  $x^* \in \mathbb{R}^n$  přípustný bod úlohy

$$\min f(x) \text{ za podm. } g(x) \leq 0,$$

buď  $f(x)$  konvexní a buďte  $g_j(x)$ ,  $j \in I(x^*)$ , konvexní. Pokud pro  $x^*$  platí KKT podmínky (5.5) pro nějaké  $\lambda \geq 0$ , pak je optimálním řešením.

*Důkaz.* Z konvexity funkce  $f(x)$  plyne  $f(x) - f(x^*) \geq \nabla f(x^*)^T(x - x^*)$  podle věty 3.18. Analogicky máme pro funkce  $g_j(x)$ ,  $j \in I(x^*)$ , vztah  $g_j(x) - g_j(x^*) \geq \nabla g_j(x^*)^T(x - x^*)$ . Podle KKT podmínek platí  $\nabla f(x^*) = -\nabla g(x^*)\lambda$ , z čehož přenásobením  $(x - x^*)$  dostaneme dohromady s předchozím

$$\begin{aligned} f(x) - f(x^*) &\geq \nabla f(x^*)^T(x - x^*) = -\lambda^T \nabla g(x^*)^T(x - x^*) \\ &= - \sum_{j \in I(x^*)} \lambda_j \nabla g_j(x^*)^T(x - x^*) \geq - \sum_{j \in I(x^*)} \lambda_j (g_j(x) - g_j(x^*)) \\ &= - \sum_{j \in I(x^*)} \lambda_j g_j(x) \geq 0. \end{aligned}$$

Tedy  $f(x^*)$  je optimální hodnota a  $x^*$  je optimální řešení.  $\square$



# Kapitola 6

## Metody

Vyřešit optimalizační úlohu je v obecnosti velmi těžký úkol – dokonce nerozhodnutelný (prokazatelně neexistuje algoritmus, který by to uměl)! Nemůžeme proto doufat, že se danou úlohu vždy podaří vyřešit optimálně. Řada metod proto vydává pouze přibližné řešení – bod, splňující KKT podmínky, lokální minimum atp. Pokud je úloha těžká a velká, pak nezbývá často nic jiného než použít heuristické algoritmy (genetické a evoluční algoritmy, simulované žíhání, tabu search, . . .), u nichž garance optimality či konvergence je velmi obtížná. Na druhou stranu, mnohé těžké úlohy optimalizace jdou vyřešit ke skutečnému optimu za pomoci technik globální optimalizace, ale pouze pro malé rozměry, protože tyto metody jsou velmi časově náročné. Volba konkrétní metody tudíž pak závisí nejen na typu optimalizační úlohy, ale také na jejím rozměru, časovému omezení aj.

V následujících odstavcích představíme nejznámější typy metod pro základní třídy optimalizačních úloh.

### 6.1 Line search

Line search je minimalizace funkce jedné proměnné  $f(x) : \mathbb{R} \rightarrow \mathbb{R}$ , tedy případ  $n = 1$ . Studuje se podrobněji, protože se používá často jako pomocná procedura v obecnějším případě.

Cílem je najít lokální minimum (či jeho jistou aproximaci) v okolí od aktuálního od bodu. Nastíníme dvě metody: Armijovo pravidlo, které slouží k tomu přejít do bodu s potenciálně výrazně menší hodnotou účelové funkce, a Newtonovu metodu, která konverguje za jistých předpokladů k lokálnímu minimu.

#### Armijovo pravidlo

Předpokládáme, že  $f(x)$  je diferencovatelná a  $f'(0) < 0$ , čili lokálně funkce klesá v  $x = 0$ . Cílem je najít bod napravo od bodu  $x = 0$  s výrazně menší hodnotou účelové funkce (tzn. pokud možno neuváznout v nejbližším okolí aktuálního bodu, ale posunout se dál).

Definujme podmínku

$$f(x) \leq f(0) + \varepsilon \cdot f'(0) \cdot x, \quad (6.1)$$

kde  $0 < \varepsilon < 1$  je daný parametr, typicky se volí  $\varepsilon \approx 0.2$ .

Podmínka se pak používá takto: Zvolme parametr  $\beta > 0$  (např.  $\beta = 2$  nebo  $\beta = 10$ ) a libovolné  $x > 0$ . Nyní

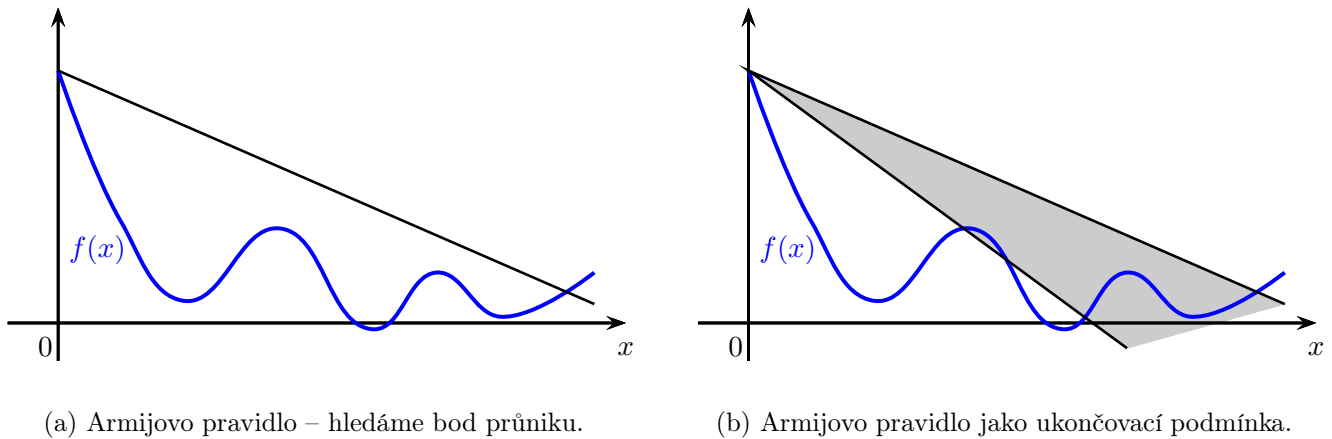
- pokud podmínka (6.1) platí, tak uprav  $x := \beta x$  a postup opakuj dokud podmínka platí;
- pokud (6.1) neplatí, tak uprav  $x := x/\beta$  a postup opakuj dokud podmínka nebude platit.

Tento postup zaručí, že se posuneme do bodu s menší hodnotou účelové funkce a přitom se posuneme so nejdál; viz obrázek 6.1a.

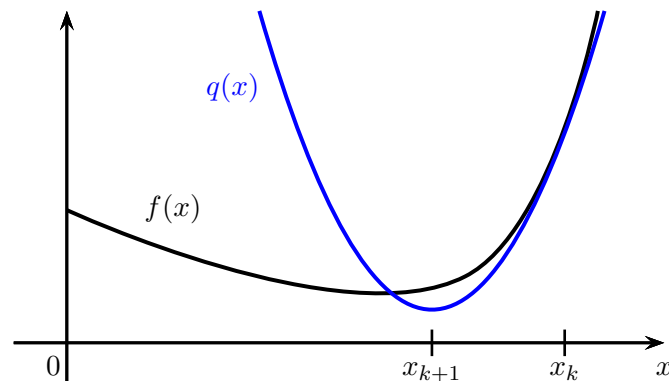
Armijovo pravidlo se používá i jako ukončovací podmínka pro jiné line search metody: Nesmíme porušit podmínku (6.1) (to vynutí, aby  $x$  nebylo moc velké) a zároveň musí platit opačná nerovnost

$$f(x) \geq f(0) + \varepsilon' \cdot f'(0) \cdot x,$$

pro daný parametr  $\varepsilon' > \varepsilon$  (to vynutí, aby  $x$  nebylo moc malé); viz obrázek 6.1b.



Obrázek 6.1: Armijovo pravidlo

Obrázek 6.2: Newtonova metoda: aproximace  $f(x)$  v bodě  $x_k$  kvadratickou funkcí  $q(x)$  a přesun do jejího minima  $x_{k+1}$ .

## Newtonova metoda

Jde o klasickou Newtonovu metodu na hledání kořene  $f'(x) = 0$ . Zde potřebujeme, aby  $f(x)$  byla dvakrát diferencovatelná.

Metoda je iterativní, kdy vytváříme posloupnost bodů  $x_0 = 0, x_1, x_2, \dots$ , které za určitých předpokladů konvergují k lokálnímu minimu. Základní myšlenka je aproximovat funkci  $f(x)$  funkcí  $q(x)$  tak, aby v aktuálním bodě  $x_k$  (v  $k$ -té iteraci) se obě funkce shodovali ve funkční hodnotě a hodnotách první a druhé derivace. Tudíž chceme  $q(x_k) = f(x_k)$ ,  $q'(x_k) = f'(x_k)$  a  $q''(x_k) = f''(x_k)$ ; viz obrázek 6.2. To naznačuje, že bude vhodné volit  $q(x_k)$  jako kvadratický polynom. Ten existuje jediný a má předpis

$$q(x) = f(x_k) + f'(x_k)(x - x_k) + \frac{1}{2}f''(x_k)(x - x_k)^2.$$

(Důkaz: dosad'  $x := x_k$ .) Minimum kvadratické funkce  $q(x_k)$  je pak v bodě, kde je nulová derivace, tedy

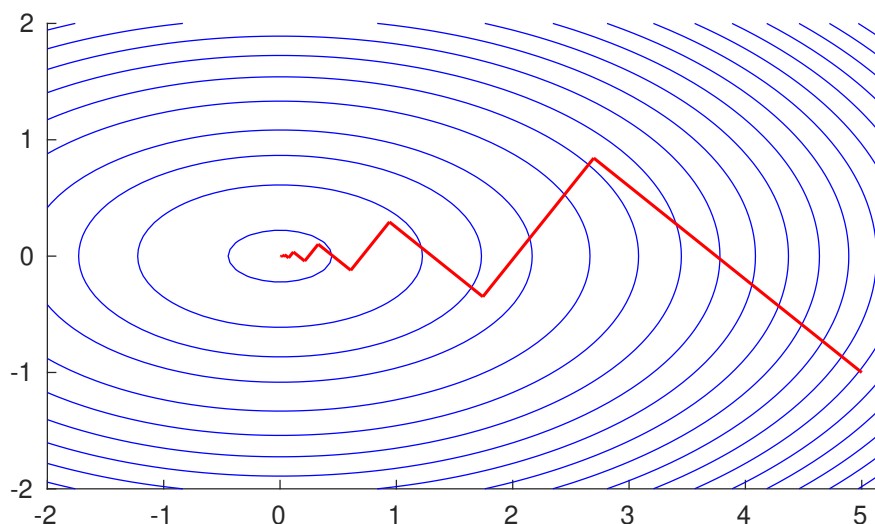
$$0 = f'(x_k) + f''(x_k)(x - x_k).$$

Z toho nám vychází

$$x = x_k - \frac{f'(x_k)}{f''(x_k)},$$

což bereme jako další bod  $x_{k+1}$ .





Obrázek 6.3: Modře: vrstevnice konvexní kvadratické funkce  $f(x) = x_1^2 + 4x_2^2$ . Červeně: iterace metody největšího spádu s počátečním bodem  $(5, -1)^T$ .

## 6.2 Úlohy bez omezení

Uvažujme úlohu

$$\min f(x) \text{ za podm. } x \in \mathbb{R}^n,$$

kde účelová funkce  $f(x)$  je diferencovatelná.

Základní metoda je iterativní, kdy vytváříme posloupnost bodů  $x_0, x_1, x_2, \dots$ , které za určitých předpokladů konvergují k lokálnímu minimu. Počáteční bod  $x_0$  se volí náhodně, pokud nemáme nějakou informaci navíc. Iterační postup ukončujeme, když hodnoty účelové funkce v bodech  $x_k$  ustálují.

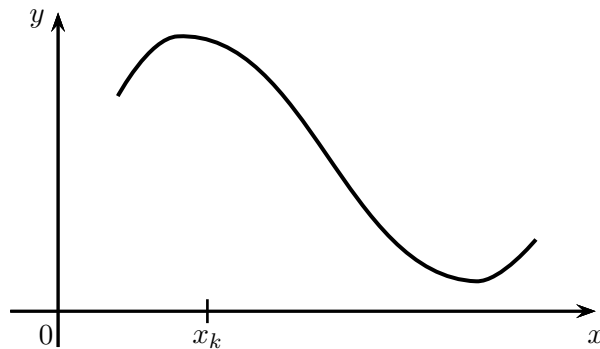
### Gradientní metody<sup>1)</sup>

V  $k$ -té iteraci se nacházíme v bodě  $x_k$  a nalezneme směr  $d_k$ , ve kterém funkce  $f(x)$  klesá, to jest,  $\nabla f(x_k)^T d_k < 0$ . Nyní zavoláme line search metodu na funkci  $\varphi(\alpha) := f(x_k + \alpha d_k)$  a necht' vrátí hodnotu  $\alpha_k$ . Pak další bod volíme jako  $x_{k+1} := x_k + \alpha_k d_k$ .

Jak volit  $d_k$ ? Nejjednodušší *metoda největšího spádu* volí  $d_k := -\nabla f(x_k)$ , tedy směr, ve kterém účelová funkce nejvíce klesá. Toto nemusí být nejlepší volba, viz obrázek 6.3, který ilustruje funkci pomalou konvergenci metody i pro jednoduchou konvexní kvadratickou funkci  $f(x) = x_1^2 + 4x_2^2$ . Sofistikovanější metody využívají také Hessiánu  $\nabla^2 f(x_k)$  či jeho aproximaci a kombinují směr největšího spádu se směry z předchozích iterací (viz také metoda sdružených gradientů, sekce 6.4).

**Příklad 6.1** (Učení neuronové sítě). Metoda největšího spádu se v podstatě používá při učení umělých neuronových sítí (pěkný úvod viz Higham and Higham [2019]). Cílem učení je nastavit váhy vstupů jednotlivých neuronů tak, aby se neuronová síť chovala co nejlépe na trénovacích datech. Matematictěji vyjádřeno, proměnnými jsou váhy vstupů jednotlivých neuronů a účelová funkce, kterou minimalizujeme, je vzdálenost výstupního vektoru od vektoru hodnot, které by síť měla vydat. Protože je tato optimalizační úloha nelineární a vysokodimensionální, je těžké najít optimum. Proto se úloha řeší přibližně a iterativně, kdy v každém kroku vylepšujeme váhy metodou největšího spádu podle gradientu účelové funkce. Protože trénovacích dat je mnoho a určení gradientu je výpočetně náročné, zjednodušuje se problém dále tak, že aproximujeme gradient dílčí hodnotou gradientu v náhodně vybraném trénovacím vzorku. Této variantě se pak říká *stochastic gradient descent*.  $\square$

<sup>1)</sup>Historie gradientních metod se datuje do roku 1847, kdy L.A. Cauchy uvedl jakousi jejich variantu pro řešení astronomického problému spočítání orbity nebeského tělesa.



Obrázek 6.4: Hessián  $\nabla^2 f(x_k)$  není pozitivně definitní.

**Příklad 6.2.** Optimalizační techniky se využívají i pro řešení úloh, které svou podstatou přímo optimalizační nejsou. Uvažujme například soustavu lineárních rovnic  $Ax = b$ , kde  $A$  je pozitivně definitní matice. Potom optimum úlohy konvexního kvadratického programování

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} x^T A x - b^T x$$

je přesně bod  $A^{-1}b$ , tedy řešení soustavy, protože v tomto bodě je nulový gradient  $\nabla f(x) = Ax - b$  účelové funkce  $f(x) = \frac{1}{2} x^T A x - b^T x$ . Tudíž můžeme hledat řešení soustavy pomocí optimalizačních technik. Toho se využívá pro velmi velké a řídké soustavy rovnic. Pro vektor  $d_k$  existuje řada různých možností. Známa metoda sdružených gradientů sofistikovaným způsobem kombinuje gradient a předchozí směr, tedy používá lineární kombinaci  $\nabla f(x_k)$  a  $d_{k-1}$ . Více viz sekce 6.4.  $\square$

## Newtonova metoda

Podobně jako v jednorozměrném případě aproximujeme účelovou funkci kvadratickou, pro kterou snadno určíme minimum, jež je výchozím bodem další iterace.

Podrobně,  $k$ -tá iterace pracuje s aktuálním bodem  $x_k$  a v něm aproximujeme  $f(x)$  pomocí Taylorova rozvoje jako

$$f(x) \approx f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{1}{2} (x - x_k)^T \nabla^2 f(x_k) (x - x_k).$$

Tím dostaneme kvadratickou funkci, a je-li Hessián  $\nabla^2 f(x_k)$  pozitivně definitní, tak její minimum je jednoznačné a je to bod, kde je gradient nulový. To vede na rovnici

$$\nabla f(x_k) + \nabla^2 f(x_k)(x - x_k) = 0,$$

z níž vyjádříme řešení

$$x = x_k - (\nabla^2 f(x_k))^{-1} \nabla f(x_k).$$

Tento bod pak tvoří další bod iterace  $x_{k+1}$ .

*Poznámka.* Výraz  $y := (\nabla^2 f(x_k))^{-1} \nabla f(x_k)$  prakticky počítáme jako řešení soustavy lineárních rovnic  $\nabla^2 f(x_k)y = \nabla f(x_k)$ .

Výhoda metody je její rychlá konvergence (blízko minima). Nevýhodou je, že Hessián  $\nabla^2 f(x_k)$  nemusí být pozitivně definitní; viz obrázek 6.4. Také výpočet Hessiánu je výpočetně náročný. Proto je třeba metodu pro praktické použití upravit (aproximovat a regularizovat Hessián atp.).

## 6.3 Úlohy s omezeními

Uvažujme úlohu

$$\min f(x) \text{ za podm. } x \in M,$$

kde  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  je diferencovatelná a množina přípustných řešení  $M \subseteq \mathbb{R}^n$  je popsána soustavou

$$\begin{aligned} g_j(x) &\leq 0, & j &= 1, \dots, J, \\ h_\ell(x) &= 0, & \ell &= 1, \dots, L, \end{aligned}$$

kde  $g_j(x), h_\ell(x): \mathbb{R}^n \rightarrow \mathbb{R}$ .

Metody jsou opět iterativní, kdy vytváříme posloupnost bodů  $x_0, x_1, x_2, \dots$ . Počáteční bod  $x_0$  se volí náhodně, pokud nemáme nějakou informaci navíc. Iterační postup ukončujeme, když hodnoty účelové funkce v bodech  $x_k$  ustálují.

### 6.3.1 Metody přípustných směrů

Tyto metody přirozeně zobecňují spádové metody z optimalizace bez omezení. Fungují na stejném principu s tím rozdílem, že při jednorozměrném line search se nesmíme dostat mimo  $M$ . Navíc dá jistou práci zohlednit omezení ve tvaru rovnic  $h(x) = 0$ .

Tyto metody jsou výhodné pokud je  $M$  konvexní polyedr, čili nadále v této sekci předpokládáme, že  $M = \{x \in \mathbb{R}^n; Ax \leq b\}$ .

#### Metoda Frank and Wolfe [1956]<sup>2)</sup>

Nechť v  $k$ -té iteraci jsme v přípustném bodě  $x_k$ . Přípustný spádový směr  $d_k$  najdeme pomocí úlohy lineárního programování

$$\min \nabla f(x_k)^T x \text{ za podm. } Ax \leq b.$$

Označíme-li  $x_k^*$  optimální řešení, pak volíme  $d_k := x_k^* - x_k$ . Tento směr je přípustný, protože  $x_k^*$  je přípustné. Navíc  $d_k$  a odpovídá směru velkého poklesu, neboť účelová funkce  $\nabla f(x_k)^T(x - x_k)$  udává derivaci funkce  $f$  v bodě  $x_k$  ve směru  $x - x_k$  (člen  $\nabla f(x_k)^T x_k$  lze vynechat, je konstantní).

#### Metoda Zoutendijk [1960]

Podobné předchozímu, ale pomocná úloha má tvar

$$\min \nabla f(x_k)^T x \text{ za podm. } Ax \leq b, \|x - x_k\| \leq 1.$$

Pokud použijeme eukleidovskou normu, hledáme vlastně přípustný směr největšího spádu. Aby ale pomocná úloha byla snadno řešitelná, směr aproximujeme volbou maximové či součtové normy. Podmínku se součtovou normou lze přepsat pomocí lineárních podmínek jako

$$e^T z \leq 1, \quad x - x_k \leq z, \quad -x + x_k \leq z.$$

### 6.3.2 Metody aktivních podmínek

Jde o metody, redukcující problém na sekvenci úloh s omezeními pouze ve tvaru rovnic.

Buď  $x_k$  aktuální řešení a

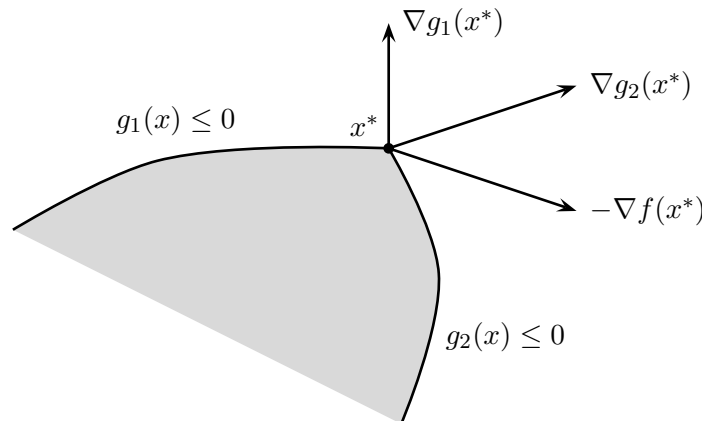
$$W := \{j; g_j(x_k) = 0\}$$

množina aktivních podmínek. Pak řešíme pomocnou úlohu

$$\min f(x) \text{ za podm. } h(x) = 0, \quad g_j(x) = 0, \quad j \in W.$$

Pokud se během výpočtu dostaneme na hranici  $M$  a další podmínka se stane aktivní, přidáme ji do  $W$ . Pokud se během výpočtu dospějeme k lokálnímu minimu  $x^*$  této pomocné úlohy, předpokládejme, že v něj jsou splněny KKT-podmínky, tj. existuje  $\lambda$  takově, že

$$\nabla f(x^*) + \nabla h(x^*)\mu + \sum_{j \in W} \lambda_j \nabla g_j(x^*) = 0.$$



Obrázek 6.5: V bodě  $x^*$  je  $\nabla f(x^*) - \nabla g_1(x^*) + \nabla g_2(x^*) = 0$ , tudíž index 1 vypustíme z aktivních podmínek a index 2 ponecháme.

Nyní, pokud  $\lambda_j \geq 0$ , tak  $j$  ponecháme v množině  $W$ ; v opačném případě index  $j$  z množiny  $W$  odstraníme. To odpovídá interpretaci Lagrangeových multiplikátorů jakožto záporných derivací účelové funkce při změně pravých stran. Tudíž pro  $\lambda_j < 0$  vede zmenšení  $g_j(x)$  lokálně ke zmenšení  $f(x)$ ; viz obrázek 6.5.

Schema této metody připomíná simplexovou metodu v lineárním programování, ve které přecházením mezi přípustnými bázemi také dynamicky měníme množinu aktivních podmínek. Proto metoda aktivních podmínek má využití především pro úlohy s lineárními omezeními.

### 6.3.3 Penalizační a bariérové metody

Tyto metody transformují úlohu tak, že omezení přepíší do účelové funkce a celý problém redukuje na optimalizaci bez omezení. Transformace omezení do účelové funkce funguje tak, že buďto platíme penále ve formě vyšší účelové funkce pro nepřipustné body (penalizační metody) nebo zabráníme dostat se mimo množinu přípustných řešení bariérou ve formě neomezeně velké účelové funkce (bariérové metody).

#### Penalizační metody

Uvažujme úlohu

$$\min f(x) \text{ za podm. } x \in M,$$

kde  $f(x)$  je spojitá funkce a  $M \neq \emptyset$  uzavřená množina. *Penalizační funkce* je spojitá nezáporná funkce  $q: \mathbb{R}^n \rightarrow \mathbb{R}$ , která splňuje podmínky

- $q(x) = 0$  pro všechna  $x \in M$ ,
- $q(x) > 0$  pro všechna  $x \notin M$ .

Penalizační metody jsou pak založené na transformaci původní úlohy na optimalizační úlohu bez omezení

$$\min f(x) + c \cdot q(x) \text{ za podm. } x \in \mathbb{R}^n,$$

kde  $c > 0$  je daný parametr.

Prakticky se penalizační metody implementují tak, že  $c$  není konstanta, ale postupně se zvětšuje. Na začátku nesmí být moc velké  $c$ , protože by pak úloha byla špatně podmíněná. Volí se tedy hodnoty z vhodné posloupnosti  $c_k > 0$ , kde  $c_k \rightarrow_{k \rightarrow \infty} \infty$ .

**Věta 6.3.** *Bud'  $x_k$  optimální řešení úlohy*

$$\min f(x) + c_k \cdot q(x) \text{ za podm. } x \in \mathbb{R}^n.$$

*Pokud  $x_k \rightarrow_{k \rightarrow \infty} x^*$ , tak  $x^*$  je optimální řešení původní úlohy  $\min_{x \in M} f(x)$ .*

<sup>2)</sup>Marguerite Franková (\*1927) a Philip Wolfe (1927–2016), jedni z průkopníků matematického programování.

*Důkaz.* Je-li  $x^* \notin M$ , tak pro dost velké  $k^*$  jsou  $x_k \notin M$ ,  $k \geq k^*$ , a tedy hodnoty účelových funkcí rostou nade všechny meze. Proto  $f(x^*) + c_k \cdot q(x^*) \rightarrow_{k \rightarrow \infty} \infty$  a též  $f(x_k) + c_k \cdot q(x_k) \rightarrow_{k \rightarrow \infty} \infty$ , což je spor s optimalitou  $x_k$ .

Uvažujme nyní případ  $x^* \in M$  a pro spor předpokládejme, že  $x^*$  není optimem. Pak existuje bod  $x' \in M$  takový, že  $f(x') < f(x^*)$ . Protože na množině  $M$  je penalizace nulová, dostáváme

$$f(x') + c_k \cdot q(x') < f(x^*) + c_k \cdot q(x^*)$$

pro každé  $k \in \mathbb{N}$ . Ze spojitosti pak pro dost velké  $k$  platí

$$f(x') + c_k \cdot q(x') < f(x_k) + c_k \cdot q(x_k),$$

což je ve sporu s optimalitou  $x_k$  úlohy v  $k$ -tém kroku. □

**Příklad 6.4.** Pro omezení typu  $g(x) \leq 0$  se často používá penalizace ve tvaru

$$q(x) := \sum_{j=1}^J (g_j(x)^+)^2 = \sum_{j=1}^J \max(0, g_j(x))^2,$$

která zachovává hladkost účelové funkce, a pro omezení typu  $h(x) = 0$  penalizace

$$q(x) := \sum_{\ell=1}^L h_{\ell}(x)^2.$$

## Bariérové metody

Uvažujme opět úlohu

$$\min f(x) \text{ za podm. } x \in M,$$

kde  $f(x)$  je spojitá funkce. Necht'  $M$  je souvislá množina splňující  $M = \text{cl}(\text{int } M)$ , tedy rovna uzávěru svého vnitřku. *Bariérová funkce* je spojitá nezáporná funkce  $q: \text{int } M \rightarrow \mathbb{R}$ , která splňuje podmínku  $q(x) \rightarrow \infty$  pro  $x \rightarrow \partial M$ . Tudíž jak se  $x$  blíží k hranici  $M$  tak bariérová funkce roste nade všechny meze.

Původní úlohu pak transformujeme na úlohu bez omezení

$$\min f(x) + \frac{1}{c} q(x) \text{ za podm. } x \in \mathbb{R}^n, \quad (6.2)$$

kde  $c > 0$  je daný parametr.

Algoritmus pak je podobný penalizačním metodám, tedy hledáme iterativně řešení pomocných úloh pro  $c \rightarrow \infty$ . Nevýhoda této metody je, že musíme znát počáteční přípustné řešení. Výhoda je jednoduchost redukce na optimalizaci bez omezení.

Za pionýry tohoto přístupu se považují Fiacco and McCormick [1968].

**Příklad 6.5.** Pro omezení typu  $g(x) \leq 0$  se často používá bariérová funkce ve tvaru

$$q(x) := - \sum_{j=1}^J \frac{1}{g_j(x)}$$

nebo ve tvaru

$$q(x) := - \sum_{j=1}^J \log(-g_j(x)).$$

Ta druhá se vyskytuje u populárních metod vnitřních bodů, na nichž jsou založené polynomiální metody na řešení lineárního programování a některých konvexních úloh jako konvexní kvadratické programování. Tudíž například úloha lineárního programování

$$\min c^T x \text{ za podm. } Ax \leq b$$

se transformuje na úlohu

$$\min c^T x - \frac{1}{c} \sum_{i=1}^m \log(b_i - A_{i*}x).$$

Bariérová funkce pro semidefinitní podmínku  $X \succeq 0$  se používá například ve tvaru

$$q(X) := -\log(\det(X)).$$

Za určitých obecných předpokladů optimální řešení úloh se zvětšující se hodnotou  $c$  konvergují k optimu původní úlohy.

**Věta 6.6.** *Bud'  $c_k > 0$  posloupnost čísel taková, že  $c_k \rightarrow_{k \rightarrow \infty} \infty$ . Bud'  $x_k$  optimální řešení úlohy*

$$\min f(x) + \frac{1}{c_k} q(x) \text{ za podm. } x \in \mathbb{R}^n.$$

*Pokud  $x_k \rightarrow_{k \rightarrow \infty} x^*$ , tak  $x^*$  je optimální řešení původní úlohy  $\min_{x \in M} f(x)$ .*

*Důkaz.* Sporem předpokládejme, že  $x^*$  není optimum, tedy existuje  $x' \in M$  takové, že  $f(x') < f(x^*)$ . Ze spojitosti funkce  $f(x)$  pak existuje  $x'' \in \text{int } M$  takové, že  $f(x'') < f(x^*)$ . Potom pro dost velké  $k$  je

$$f(x'') + \frac{1}{c_k} q(x'') < f(x^*) + \frac{1}{c_k} q(x^*).$$

Pro dost velké  $k$  pak platí rovněž

$$f(x'') + \frac{1}{c_k} q(x'') < f(x_k) + \frac{1}{c_k} q(x_k),$$

což je spor s optimalitou  $x_k$  k  $k$ -tém kroku. □

Pro konvexní úlohu za celkem obecných předpokladů (ryze konvexní bariérová funkce a omezená  $M$ ) je pak optimální řešení  $x(c)$  úlohy (6.2) jednoznačné a body  $x(c)$ ,  $c > 0$ , vykreslují hladkou křivku, nazývanou *centrální cesta*, jejíž limita pro  $c \rightarrow \infty$  je v optimu původní úlohy.

Na podobném principu fungují i algoritmy: pro postupně zvětšující se  $c$  hledají přibližná optima  $x(c)$ . Protože s malou změnou  $c$  se optimum změní spojitě, je rychlé přejít k novému optimu. Pro teoretickou analýzu polynomiality speciálních konvexních úloh se volí krátké kroky, ale prakticky jsou výhodnější trochu delší. Typicky se zvětšuje s faktorem cca 1.1.

Zde je přirozená otázka, proč rovnou nevolit  $c$  co největší? Jedním důvodem jsou numerické potíže. Druhým je, že to není rychlejší. Newtonova metoda (nebo i jiná, kterou řešíme úlohu (6.2)) je hodně pomalá pokud začneme daleko od optima. Tudíž sledování centrální cesty rychlými kroky je výhodnější. Sice na začátku řešíme podobný problém, ale to lze obejít.

## 6.4 Metoda sdružených gradientů

Metoda sdružených gradientů (ang. conjugate gradients) byla vyvinuta pro řešení soustavy lineárních rovnic  $Ax = b$  s pozitivně definitní maticí  $A \in \mathbb{R}^{n \times n}$ . Pochází od autorů Hestenes and Stiefel [1952] a je standardní součástí optimalizačních učebnic [Luenberger and Ye, 2008] stejně jako metod numerické matematiky [Liesen and Strakoš, 2013]. Přestože je metoda povahou iterační, konverguje po nanejvýš  $n$  krocích. Je paměťově nenáročná, matici  $A$  nijak neupravuje, a proto je výhodná zejména, když  $n$  je velmi velké.

Základní myšlenka je uvažovat kvadratickou funkci

$$f(x) = \frac{1}{2}x^T Ax - b^T x.$$

Protože  $A$  je pozitivně definitní, funkce je ryze konvexní a nabývá jednoznačného minima v bodě s nulovým gradientem  $\nabla f(x) = Ax - b$ . Tedy minimum funkce  $f(x)$  je přesně řešením rovnice  $Ax = b$ . Tím jsme problém hledání řešení rovnice převedli na optimalizační úlohu.

Popíšeme metodu v trochu zjednodušené podobě. Namísto standardní báze prostoru  $\mathbb{R}^n$  budeme uvažovat ortonormální bázi  $d_1, \dots, d_n$  a skalární součin  $\langle x, y \rangle := x^T Ay$  namísto toho standardního; kolmost v tomto skalárním součinu nazýváme A-kolmost a ortonormální bázi jako A-ortonormální. Ke konkrétní volbě báze se dostaneme později. Jako  $x^* := A^{-1}b$  označíme hledané řešení a jako  $x_k$  přibližné řešení v  $k$ -té iteraci. Na začátku zvolíme počáteční vektor  $x_1$  libovolně.

**Základní schéma.** Vektor  $x^* - x_1$  vyjádříme pomocí báze

$$x^* - x_1 = \sum_{k=1}^n \alpha_k d_k.$$

Základní schéma metody je v zásadě jednoduché, jako bychom se pohybovali od jednoho vrcholu kváдру k protilehlému po kolmých hranách:

$$\text{Iteruj } x_{k+1} := x_k + \alpha_k d_k, \quad k = 1, \dots$$

K implementaci metody potřebujeme ještě určit bázi  $d_1, \dots, d_n$  a jak efektivně počítat koeficienty  $\alpha_k$ . Ještě označíme  $g_k := \nabla f(x_k) = Ax_k - b$ , což udává nejen gradient v bodu  $x_k$  v  $k$ -té iteraci, ale i jeho residuum, tedy rozdíl mezi levou a pravou stranou (kdyby residuum byla 0, tak máme řešení). Všimněme si rovněž, že pro jakékoli  $j \in \{1, \dots, k\}$  platí

$$x_{k+1} = x_k + \alpha_k d_k = x_{k-1} + \alpha_k d_k + \alpha_{k-1} d_{k-1} = \dots = x_j + \sum_{i=j}^k \alpha_i d_i.$$

**Výpočet  $\alpha_k$ .** Protože  $d_1, \dots, d_n$  je A-ortonormální báze, souřadnice  $\alpha_k$  jsou vlastně Fourierovy koeficienty a získáme je snadno jako  $\alpha_k = \langle d_k, x^* - x_1 \rangle$ . Jenomže vektor  $x^*$  neznáme. Protože  $x_k - x_1 = \sum_{i=1}^{k-1} \alpha_i d_i$ , je vektor  $x_k - x_1$  A-koľmý na  $d_k$ , čili  $\langle d_k, x_k - x_1 \rangle = 0$ . Odvodíme

$$\begin{aligned} \alpha_k &= \langle d_k, x^* - x_1 \rangle = \langle d_k, x^* - x_k + x_k - x_1 \rangle = \langle d_k, x^* - x_k \rangle + \langle d_k, x_k - x_1 \rangle \\ &= \langle d_k, x^* - x_k \rangle = d_k^T A(x^* - x_k) = d_k^T (b - Ax_k) = -d_k^T g_k. \end{aligned}$$

**Tvrzení 6.7.** Vektor  $x_{k+1}$  je minimem  $f(x)$  na afinním podprostoru  $x_1 + \text{span}\{d_1, \dots, d_k\}$ , čili  $g_{k+1}^T d_j = 0$  pro  $j = 1, \dots, k$  (tzn.  $g_{k+1} \perp d_j$  ve smyslu eukleidovské kolmosti).

*Důkaz.* Stačí ukázat, že  $\nabla f(x_{k+1}) = g_{k+1}$  je kolmý na podprostor  $x_1 + \text{span}\{d_1, \dots, d_k\}$ , tedy kolmý na všechny vektory  $d_1, \dots, d_k$ . Upravíme

$$g_{k+1} = Ax_{k+1} - b = A \left( x_j + \sum_{i=j}^k \alpha_i d_i \right) - b$$

Pro libovolné  $j \in \{1, \dots, k\}$  pak

$$\begin{aligned} d_j^T g_{k+1} &= d_j^T \left( A \left( x_j + \sum_{i=j+1}^k \alpha_i d_i \right) - b \right) = d_j^T (Ax_j - b) + d_j^T A \left( \sum_{i=j}^k \alpha_i d_i \right) \\ &= d_j^T g_j + \alpha_j = 0. \end{aligned} \quad \square$$

**Volba báze  $d_1, \dots, d_n$ .** Bázi volíme tak, aby  $\text{span}\{d_1, \dots, d_k\} = \text{span}\{g_1, \dots, g_k\}$  pro všechna  $k = 1, \dots, n$ . Na začátku volíme celkem přirozeně  $d_1 := -g_1 / \sqrt{\langle g_1, g_1 \rangle}$ . V  $(k+1)$ -ní iteraci pak vektor  $d_{k+1}$  sestojíme z vektoru  $-g_{k+1}$  nakolmením k podprostoru  $\text{span}\{d_1, \dots, d_k\}$ .

**Tvrzení 6.8.** Platí  $\text{span}\{g_1, \dots, g_k\} = \text{span}\{g_1, Ag_1, \dots, A^{k-1}g_1\}$ .

*Důkaz.* Matematickou indukcí podle  $k$ . Z definice a indukčního předpokladu je  $g_k = Ax_k - b$ , kde

$$x_k \in x_1 + \text{span}\{g_1, \dots, g_{k-1}\} = x_1 + \text{span}\{g_1, Ag_1, \dots, A^{k-2}g_1\}.$$

Tedy

$$g_k \in Ax_1 - b + \text{span}\{Ag_1, A^2g_1, \dots, A^{k-1}g_1\} \subseteq \text{span}\{g_1, Ag_1, A^2g_1, \dots, A^{k-1}g_1\}.$$

Rovnost pak nastane díky tomu, že  $g_k$  nenáleží do  $\text{span}\{g_1, \dots, g_{k-1}\}$ . V opačném případě, protože podle Tvrzení 6.7 je  $g_k$  kolmý na tento podprostor, muselo by být  $g_k = Ax_k - b$  nulové, a tím pádem by  $x_k$  již bylo řešením.  $\square$

Protože  $g_{k+1}$  je kolmý (v eukleidovském smyslu) na vektory  $d_1, \dots, d_k$ , je kolmý i na  $g_1, \dots, g_k$ , a podle Tvzení 6.8 je tudíž A-kolmý na vektory  $g_1, \dots, g_{k-1}$ . Tím pádem k výpočtu  $d_{k+1}$  stačí pouze nakolmit  $-g_{k+1}$  k vektoru  $d_k$ . To ukazuje následující tvrzení; poznamenejme, že výslednou hodnotu  $d_{k+1}$  je třeba poté ještě znormovat.

**Tvrzení 6.9.** Platí  $d_{k+1} = -g_{k+1} + \beta_{k+1}d_k$ , kde  $\beta_{k+1} = \langle d_k, g_{k+1} \rangle$ .

*Důkaz.* Jak jsme již nahlédli,  $\langle g_{k+1}, d_i \rangle = 0$  pro  $i = 1, \dots, k-1$ . Tudíž  $d_{k+1}$  má tvar  $d_{k+1} = -g_{k+1} + \beta_{k+1}d_k$  pro nějaké  $\beta_{k+1}$ . Z rovnosti  $0 = \langle d_k, d_{k+1} \rangle = d_k^T A(-g_{k+1} + \beta_{k+1}d_k)$  odvodíme hodnotu pro  $\beta_{k+1} = \frac{d_k^T A g_{k+1}}{d_k^T A d_k} = \frac{\langle d_k, g_{k+1} \rangle}{\langle g_k, g_k \rangle} = \langle d_k, g_{k+1} \rangle$ .  $\square$

**Shrnutí.** Tímto máme již všechny ingredience pro výsledný algoritmus. Popis dole technicky sumarizuje jednotlivé kroky:

- 1: zvol  $x_1 \in \mathbb{R}^n$  a polož  $d_0 := 0$ ,
- 2: pro  $k = 1, \dots, n$  proved

$$\begin{aligned} g_k &:= Ax_k - b, \\ \beta_k &:= d_{k-1}^T A g_k, \\ d_k &:= -g_k + \beta_k d_{k-1}, \quad d_k := d_k / \sqrt{d_k^T A d_k}, \\ \alpha_k &:= -d_k^T g_k, \\ x_{k+1} &:= x_k + \alpha_k d_k, \end{aligned}$$

**Poznámka 6.10.** Několik poznámek k metodě sdružených gradientů:

- (1) Metoda je paměťově nenáročná a nijak neupravuje matici  $A$ . Výhodná je zejména, pokud matice  $A$  je velká a řídká. Pak výpočetní čas jednotlivé iterace je relativně nízký. Navíc ne vždy je nutné proběhnout všech  $n$  iterací – k řešení nebo jeho dostatečně přesné aproximaci můžeme dříve podstatně dříve.
- (2) Obvykle se metoda prezentuje bez normování vektoru  $d_k$ . To se ale pak musí zohlednit v ostatních výrazech, kde se  $d_k$  vyskytuje.
- (3) Pokud volíme  $x_1 = 0$ , pak  $\text{span}\{d_1, \dots, d_k\} = \text{span}\{b, Ab, A^2b, \dots, A^{k-1}b\}$  se nazývá Krylovův podprostor a teorie za nimi je velmi zajímavá [Liesen and Strakoš, 2013].

Myšlenky metody sdružených gradientů lze použít i k minimalizaci obecné nelineární funkce  $f(x)$  na prostoru  $\mathbb{R}^n$ . Zde jako klíčová je myšlenka sestavit směr postupu  $d_k$  jako lineární kombinaci gradientu  $g_k$  a předchozího směru  $d_{k-1}$ . Vektor  $g_k$  bude opět gradient funkce  $f(x)$  v bodě  $x_k$  a koeficienty se přepočítají analogicky. Výsledná metoda je známá jako metoda Fletcher–Reeves (1964). Existuje ale řada variant, v nichž koeficienty  $\beta_k$  mají trochu jinou podobu.

Existují metody využívající Krylovovy podprostory i pro řešení soustav  $Ax = b$ , kde matice  $A$  není symetrická pozitivně definitní. Mezi ně patří například GMRES (Generalized minimal residual method, Saad & Schultz, 1986), která v  $k$  té iteraci spočítá vektor  $x_k$ , který minimalizuje eukleidovskou normu residua (tedy  $\|Ax - b\|$ ) na podprostoru  $\text{span}\{b, Ab, A^2b, \dots, A^{k-1}b\}$ .



# Kapitola 7

## Vybraná témata

### 7.1 Robustní optimalizace

Data jsou často nepřesná a to nás motivuje k tomu, hledat řešení, které bude tzv. *robustní*. To znamená (přesná definice neexistuje), aby bylo přípustné a co nejlepší i při malých perturbacích v datech; viz Ben-Tal et al. [2009]. Ukážeme dva přístupy k robustnosti, intervalový a elipsoidový.

#### Intervalová nejistota (I)

Uvažujme nejprve úlohu lineárního programování ve tvaru

$$\min c^T x \text{ za podm. } Ax \leq b, x \geq 0.$$

Předpokládejme, že hodnoty  $A, b$  nejsou přímo známy, ale pro každou složku známe interval, ve kterém se nachází. Tudiž známe matici intervalů  $[\underline{A}, \overline{A}]$  a vektor intervalů nejistoty pravých stran  $[\underline{b}, \overline{b}]$ . Pak říkáme, že vektor  $x$  je robustním přípustným řešením, pokud splňuje nerovnost  $Ax \leq b$  pro každé  $A \in [\underline{A}, \overline{A}]$  a  $b \in [\underline{b}, \overline{b}]$ . Díky nezápornosti proměnných je  $x$  přípustné právě tehdy, když  $\overline{A}x \leq \underline{b}$ . Robustní verze lineárního programu pak je

$$\min c^T x \text{ za podm. } \overline{A}x \leq \underline{b}, x \geq 0.$$

**Příklad 7.1** (Výživový plán sumců). Tento příklad převzatý z <http://www.fao.org/3/x5738e/x5738e0h.htm> (a zjednodušený) spočívá v nalezení optimálního plánu na krmení sumců v Thajsku tak, abychom při daných omezení minimalizovali cenu. Matematická formulace úlohy je

$$\min c^T x \text{ za podm. } Ax \geq b, x \geq 0, \quad (7.1)$$

kde proměnná  $x_j$  udává množství živiny  $j$ ,  $c_j$  je cena za jednotku živiny  $j$  a  $i$ -tá nerovnice udává  $i$ -té nutriční omezení. Data jsou uvedena v Tabulce 7.1. V našem případě tedy máme

$$A = \begin{pmatrix} 9 & 65 & 44 & 12 & 0 \\ 1.10 & 3.90 & 2.57 & 1.99 & 0 \\ 0.02 & 3.7 & 0.3 & 0.1 & 38.0 \end{pmatrix}, \quad b = \begin{pmatrix} 30 \\ 250 \\ 0.5 \end{pmatrix}, \quad c = \begin{pmatrix} 2.15 \\ 8.0 \\ 6.0 \\ 2.0 \\ 0.4 \end{pmatrix}.$$

Protože výživové hodnoty nejsou známy zcela přesně, předpokládejme, že jsou známy s přesností na 5%. Tudiž každá složka matice má ve skutečnosti hodnotu v intervalu  $[0.95 \cdot a_{ij}, 1.05 \cdot a_{ij}]$ . Podle postupu nahoře získáme robustní řešení tak, že matici  $A$  nahradíme maticí horních mezí intervalů  $\underline{A}$ , tedy

$$\underline{A} = \begin{pmatrix} 8.550 & 61.75 & 41.800 & 11.400 & 0.00 \\ 1.045 & 3.705 & 2.4415 & 1.8905 & 0.00 \\ 0.019 & 3.515 & 0.2850 & 0.0950 & 36.1 \end{pmatrix}. \quad \square$$

	cena (THB/kg)	bílkoviny (%)	energie (Mcal/kg)	vápník (%)
kukuřice	2.15	9	1.10	0.02
rybky	8.0	65	3.90	3.7
sója	6.0	44	2.57	0.3
rýžové otruby	2.0	12	1.99	0.1
vápenec	0.4	0	0	38.0
požadavek	min	30	250	0.5

Tabulka 7.1: (Příklad 7.1) Krmení sumců: nutriční hodnoty jednotlivých živin a požadavky na ně.

## Intervalová nejistota (II)

Uvažujme nyní úlohu lineárního programování ve tvaru bez nezápornosti proměnných

$$\min c^T x \text{ za podm. } Ax \leq b.$$

Buď  $a^T x \leq d$  jedna z nerovností. Nechť jsou dány intervaly  $[\underline{a}, \bar{a}] = ([\underline{a}_1, \bar{a}_1], \dots, [\underline{a}_n, \bar{a}_n])^T$ ,  $[\underline{d}, \bar{d}]$ . Nyní  $x$  je robustní řešení dané nerovnice, pokud splňuje

$$a^T x \leq d \quad \forall a \in [\underline{a}, \bar{a}], \quad \forall d \in [\underline{d}, \bar{d}],$$

neboli

$$\max_{a \in [\underline{a}, \bar{a}]} a^T x \leq \underline{d}.$$

**Lemma 7.2.** Označme  $a_\Delta = \frac{1}{2}(\bar{a} - \underline{a})$  vektor poloměřů a  $a_c = \frac{1}{2}(\underline{a} + \bar{a})$  vektor středů. Pak platí

$$\max_{a \in [\underline{a}, \bar{a}]} a^T x = a_c^T x + a_\Delta^T |x|.$$

*Důkaz.* Pro každé  $a \in [\underline{a}, \bar{a}]$  platí

$$a^T x = a_c^T x + (a - a_c)^T x \leq a_c^T x + |a - a_c|^T |x| \leq a_c^T x + a_\Delta^T |x|.$$

Nerovnost se nabyde jako rovnost pro vhodné  $a \in [\underline{a}, \bar{a}]$ . Je-li  $x \geq 0$ , pak  $a_c^T x + a_\Delta^T |x| = a_c^T x + a_\Delta^T x = \bar{a}^T x$ . Je-li  $x \leq 0$ , pak  $a_c^T x + a_\Delta^T |x| = a_c^T x - a_\Delta^T x = \underline{a}^T x$ . Jinak aplikujeme tento postup po složkách, čili nerovnost se nabyde jako rovnost pro  $a$ , jehož každá složka je buďto horní nebo dolní kraj intervalu.  $\square$

S pomocí lemmatu pak robustní omezení přepíšeme na tvar

$$a_c^T x + a_\Delta^T |x| \leq \underline{d}.$$

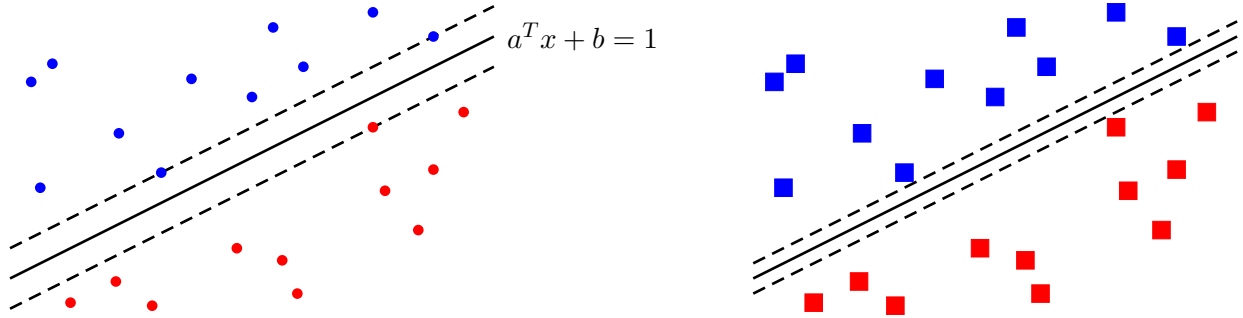
Toto omezení nevypadá na první pohled pěkně, protože funkce vlevo je sice konvexní, ale nehladká díky absolutní hodnotě. Nicméně jednoduchým trikem se zavedením nové proměnné  $y \in \mathbb{R}^n$  lze omezení přepsat jako lineární

$$a_c^T x + a_\Delta^T y \leq \underline{d}, \quad x \leq y, \quad -x \leq y.$$

Tudíž se linearita zachovává, i robustní řešení intervalových lineárních programů jsou popsána lineárními podmínkami.

**Příklad 7.3** (Robustní klasifikace). Mějme data rozdělená podle určité vlastnosti do dvou skupin, první obsahuje množinu bodů  $x_1, \dots, x_p \in \mathbb{R}^n$  a druhá body  $y_1, \dots, y_q \in \mathbb{R}^n$ . Cílem je zkonstruovat klasifikátor, který bude umět odhadnout příslušnost ke skupině pro nová data. Základní lineární klasifikátor může být založen na oddělení dat co nejširším pásem. Matematicky vyjádřeno, hledáme nadrovinu  $a^T x + b = 1$  takovou, aby první část dat byla v kladném poloprostoru, druhá část v druhém a příslušný oddělovací pás byl co nejširší. To vede na úlohu konvexního kvadratického programování (viz obrázek 7.1a)

$$\min \|a\|_2 \text{ za podm. } a^T x_i + b \geq 1 \quad \forall i, \quad a^T y_j + b \leq -1 \quad \forall j.$$



(a) Nejširší oddělující pás pro reálná data.

(b) Nejširší oddělující pás pro intervalová data.

Obrázek 7.1: (Příklad 7.3) Klasifikace pro reálná data a robustní klasifikace pro intervalová data.

Předpokládejme nyní, že data nejsou změřena přesně, ale známe je jen s určitou intervalovou přesností. Tudíž máme dány vektory intervalů  $[\underline{x}_i, \bar{x}_i] = [(x_c)_i - (x_\Delta)_i, (x_c)_i + (x_\Delta)_i]$ ,  $i = 1, \dots, p$ , a  $[\underline{y}_j, \bar{y}_j] = [(y_c)_j - (y_\Delta)_j, (y_c)_j + (y_\Delta)_j]$ ,  $j = 1, \dots, q$ , obsahující skutečná data. Podle naší kuchařky pak robustní model zní (viz obrázek 7.1b)

$$\min \|a\|_2 \text{ za podm. } (x_c)_i^T a + (x_\Delta)_i^T a' + b \leq 1 \quad \forall i, \quad (y_c)_j^T a + (y_\Delta)_j^T a' + b \leq -1 \quad \forall j, \quad \pm a \leq a'.$$

Opět se jedná o úlohu konvexního kvadratického programování, nyní s proměnnými  $a, a' \in \mathbb{R}^n$ ,  $b \in \mathbb{R}$ .  $\square$

### Elipsoidová nejistota

Uvažujme úlohu lineárního programování ve tvaru bez nezápornosti proměnných

$$\min c^T x \text{ za podm. } Ax \leq b.$$

Buď  $a^T x \leq d$  jedna z nerovností. Uvažujme elipsoid

$$\mathcal{E} = \{a \in \mathbb{R}^n; a = p + Pu, \|u\|_2 \leq 1\},$$

který je dán jako obraz jednotkové koule při lineárním (či přesněji afinním) zobrazení. Nyní  $x$  je robustní řešení dané nerovnice, pokud splňuje

$$a^T x \leq d \quad \forall a \in \mathcal{E}$$

neboli

$$\max_{a \in \mathcal{E}} a^T x \leq d.$$

**Lemma 7.4.** Platí

$$\max_{a \in \mathcal{E}} a^T x = p^T x + \|P^T x\|_2.$$

*Důkaz.* Upravíme

$$\begin{aligned} \max_{a \in \mathcal{E}} a^T x &= \max_{\|u\|_2 \leq 1} (p + Pu)^T x = p^T x + \max_{\|u\|_2 \leq 1} (P^T x)^T u \\ &= p^T x + (P^T x)^T \frac{1}{\|P^T x\|_2} P^T x = p^T x + \|P^T x\|_2. \end{aligned} \quad \square$$

Lemma pak robustní omezení umožní přepsat na tvar

$$p^T x + \|P^T x\|_2 \leq d.$$

Vlevo je konvexní hladká funkce, tedy výsledná úloha bude konvexní kuželovou optimalizační úlohou.

**Příklad 7.5.** Uvažujme opět problém výběru portfolia (příklad 4.9)

$$\max c^T x \quad \text{za podm. } e^T x = K, x \geq o, \quad (7.2)$$

kde  $c$  je náhodný vektor s vícerozměrným normálním rozdělením, se střední hodnotou  $\tilde{c} := E c$  a kovarianční maticí  $\Sigma := \text{cov } c = E(c - \tilde{c})(c - \tilde{c})^T$ . Vrstevnice hustoty tohoto rozdělení představují elipsoidy, takže je přirozené pracovat s nimi. Pro náhodný vektor  $c$  platí, že pravděpodobnost  $P(c - \tilde{c} \in \mathcal{E}_\eta) = \eta$ , kde  $\mathcal{E}_\eta$  je jistý elipsoid (konkrétně,  $\mathcal{E}_\eta = \{d \in \mathbb{R}^n; d = F^{-1}(\eta)\sqrt{\Sigma}u, \|u\|_2 \leq 1\}$ , kde  $F^{-1}(\eta)$  je kvantilová funkce normálního rozdělení a  $\sqrt{\Sigma}$  je pozitivně semidefinitní odmocnina z matice  $\Sigma$ , tzn.  $(\sqrt{\Sigma})^2 = \Sigma$ ).

Jeden z možných způsobů, jak řešit úlohu (7.2), je uvažovat deterministickou verzi

$$\max z \quad \text{za podm. } P(c^T x \geq z) \geq \eta, e^T x = K, x \geq o,$$

kde  $\eta \in [\frac{1}{2}, 1]$  je pevná hodnota, např.  $\eta = 0.95$ . Podmínka  $P(c^T x \geq z) \geq \eta$  je zřejmě splněná, pokud  $d^T x \geq z$  platí pro každé  $d \in \mathcal{E}_\eta + \tilde{c}$ , tudíž úlohu můžeme aproximovat jako

$$\max z \quad \text{za podm. } d^T x \geq z \quad \forall d \in \mathcal{E}_\eta + \tilde{c}, e^T x = K, x \geq o.$$

Toto je již úloha s elipsoidovou nejistotou, která je ekvivalentní úloze

$$\max z \quad \text{za podm. } \tilde{c}^T x - F^{-1}(\eta)\|\sqrt{\Sigma}x\|_2 \geq z, e^T x = K, x \geq o.$$

Protože  $F^{-1}(\eta) \geq 0$  pro  $\eta \geq \frac{1}{2}$ , jedná se o úlohu kuželového kvadratického programování.

## 7.2 Konkávní programování

Konkávním programováním se rozumí minimalizace konkávní funkce na konvexní množině, neboli ekvivalentně

$$\max f(x) \quad \text{za podm. } x \in M,$$

kde  $M \subseteq \mathbb{R}^n$  je konvexní množina a  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  konvexní funkce.

**Věta 7.6.** *Buď  $M$  omezený konvexní polyedr. Pak optimum existuje a nabyde se v alespoň jednom vrcholu  $M$ .*

*Poznámka.* Věta lze zobecnit na tvar: Spojitá konvexní funkce na kompaktní množině  $M$  nabývá svého maxima v extrémním bodě  $M$ , viz Avriel et al. [1988]. Vlastnost platí dokonce obecněji i pokud  $f(x)$  je tzv. kvazikonvexní.

*Důkaz.* Buďte  $v_1, \dots, v_m$  vrcholy  $M$  a necht' bez újmy na obecnost  $f(v_1) = \max_{i=1, \dots, m} f(v_i)$ . Pak každý bod  $x \in M$  lze vyjádřit jako konvexní kombinaci  $x = \sum_{i=1}^m \alpha_i v_i$ , kde  $\alpha_i \geq 0$  a  $\sum_{i=1}^m \alpha_i = 1$ . Nyní

$$f(x) = f(\sum_{i=1}^m \alpha_i v_i) \leq \sum_{i=1}^m \alpha_i f(v_i) \leq \sum_{i=1}^m \alpha_i f(v_1) = f(v_1).$$

Tudíž  $v_1$  je optimem. □

Tato vlastnost je shodná s vlastností v lineárním programování, ale pro hledání optima to nemusí být výhodné, protože  $M$  může obsahovat mnoho vrcholů a nevíme, který je optimální. Z věty 4.8 navíc víme, že tato úloha je NP-těžká.

Typické úlohy, vedoucí na konkávní programování, jsou:

- *Fixed charged problems.* Účelová funkce je tvaru  $f(x) = \sum_{i=1}^k f_i(x_i)$ , kde  $f_i(x_i) = 0$  pro  $x_i = 0$  a  $f_i(x_i) = c_i + g_i(x_i)$  pro  $x_i > 0$ . Zde  $f_i(x_i)$  reprezentuje cenu (např. za dopravu zboží o velikosti  $x_i$ ). Tedy přirozeně je cena nulová při  $x_i = 0$ . Při  $x_i > 0$  platíme pevné náklady  $c_i$  plus cenu  $g_i(x_i)$  v závislosti na velikosti  $x_i$ . Přirozeně je  $g_i(x_i)$  konkávní, protože s větším množstvím můžeme předpokládat množstevní slevu.
- *Multiplikativní programování.* Účelová funkce je tvaru  $f(x) = \prod_{i=1}^k x_i$ . To obecně sice není konkávní funkce, ale zlogaritmováním dostaneme konkávní funkci  $\log(f(x)) = \sum_{i=1}^k \log(x_i)$ . Takovéto úlohy se vyskytují například v geometrických problémech, kdy minimalizujeme objem tělesa (kvádr) za určitých podmínek (kvádr obsahuje danou množinu bodů).

# Příloha

**Derivace maticových výrazů.** Buď  $A \in \mathbb{R}^{n \times n}$ ,  $b \in \mathbb{R}^n$ ,  $c \in \mathbb{R}$  a uvažujme kvadratickou funkci  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  definovanou předpisem

$$f(x) = x^T A x + b^T x + c.$$

Její gradient je

$$\nabla f(x) = (A + A^T)x + b$$

a Hessián

$$\nabla^2 f(x) = A + A^T.$$

Speciálně, pokud je matice  $A$  symetrická, pak

$$\nabla f(x) = 2Ax + b, \quad \nabla^2 f(x) = 2A.$$

*Důkaz.* Začneme s lineárním členem:

$$\frac{\partial}{\partial x_k} b^T x = \frac{\partial}{\partial x_k} \sum_{i=1}^n b_i x_i = b_k,$$

z čehož  $\nabla b^T x = b$ .

Pro kvadratický člen máme

$$\begin{aligned} \frac{\partial}{\partial x_k} x^T A x &= \frac{\partial}{\partial x_k} \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j = \frac{\partial}{\partial x_k} \left( a_k x_k^2 + \sum_{i \neq k} (a_{ik} + a_{ki}) x_i x_k + \sum_{i, j \neq k} a_{ij} x_i x_j \right) \\ &= 2a_k x_k + \sum_{i \neq k} (a_{ik} + a_{ki}) x_i = \sum_{i=1}^n (a_{ik} + a_{ki}) x_i = ((A + A^T)x)_i. \end{aligned}$$

Tudíž jeho gradient je  $\nabla x^T A x = (A + A^T)x$ . Jelikož tento gradient je lineární funkcí, jednotlivé složky derivujeme podle již odvozeného pravidla o podobě gradientu lineárního členu a dostáváme  $\nabla^2 x^T A x = A + A^T$ .  $\square$



# Značení

## Množiny a čísla

$\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$	množina přirozených, celých, racionálních a reálných čísel
$U + V$	součet množin, $U + V = \{u + v; u \in U, v \in V\}$
$U + v$	součet množiny a vektoru, $U + v = U + \{v\} = \{u + v; u \in U\}$
$\text{conv}(M)$	konvexní obal množiny $M$
$\text{int}(M)$	topologický vnitřek množiny $M$
$\mathcal{K}^*$	duální kužel ke kuželi $\mathcal{K}$ , viz definice 4.15
$r^+$	kladná část reálného čísla, $r^+ = \max(r, 0)$

## Matice a vektory

$\text{tr}(A)$	stopa matice $A$ , $\text{tr}(A) = \sum_i a_{ii}$
$\mathcal{S}(A)$	sloupcový prostor matice $A$
$A^T$	transpozice matice $A$
$A \geq B$	nezápornost matice $A - B$ , tj. $a_{ij} \geq b_{ij}$
$A > B$	kladnost matice $A - B$ , tj. $a_{ij} > b_{ij}$
$A \succeq B$	matice $A - B$ je pozitivně semidefinitní
$A \succ B$	matice $A - B$ je pozitivně definitní
$A_{i*}$	$i$ -tý řádek matice $A$
$A_{*j}$	$j$ -tý sloupec matice $A$
$\text{diag}(v)$	diagonální matice s diagonálními prvky $v_1, \dots, v_n$
$0_n, 0$	nulová matice (všechny složky jsou rovny 0)
$I_n, I$	jednotková matice (diagonální s jedničkami na diagonále)
$e_i$	jednotkový vektor, $e_i = I_{*i} = (0, \dots, 0, 1, \dots, 0)^T$
$e$	vektor ze samých jedniček, $e = (1, \dots, 1)^T$

## Funkce

$\ x\ _p$	$\ell_p$ -norma vektoru $x \in \mathbb{R}^n$ , $\ x\ _p = \left(\sum_{i=1}^n  x_i ^p\right)^{\frac{1}{p}}$
$\ x\ _1$	součtová norma vektoru $x \in \mathbb{R}^n$ , $\ x\ _1 = \sum_{i=1}^n  x_i $
$\ x\ _2$	eukleidovská norma vektoru $x \in \mathbb{R}^n$ , $\ x\ _2 = \sqrt{\sum_{i=1}^n x_i^2}$
$\ x\ _\infty$	maximová (Čebyševova) norma vektoru $x \in \mathbb{R}^n$ , $\ x\ _\infty = \max_{i=1, \dots, n}  x_i $
$P(c)$	pravděpodobnost náhodného jevu $c$





# Literatura

- A. A. Ahmadi, A. Olshevsky, P. A. Parrilo, and J. N. Tsitsiklis. NP-hardness of deciding convexity of quartic polynomials and related problems. *Math. Program.*, 137(1-2):453–476, 2013.  
<https://arxiv.org/pdf/1012.1908>. 25
- M. Avriel, W. E. Diewert, S. Schaible, and I. Zang. *Generalized concavity*, volume 36 of *Mathematical Concepts and Methods in Science and Engineering*. Plenum Press, New York, 1988. 60
- M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. *Nonlinear Programming. Theory and Algorithms. 3rd ed.* John Wiley & Sons., NJ, 2006. 3
- A. Ben-Tal and A. Nemirovski. *Lectures on modern convex optimization. Analysis, algorithms, and engineering applications*. SIAM, Philadelphia, PA, 2001.  
[https://www2.isye.gatech.edu/~nemirovs/Lect\\_ModConvOpt.pdf](https://www2.isye.gatech.edu/~nemirovs/Lect_ModConvOpt.pdf). 30
- A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust Optimization*. Princeton University Press, 2009.  
<https://www2.isye.gatech.edu/~nemirovs/FullBookDec11.pdf>. 57
- S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.  
<http://web.stanford.edu/~boyd/cvxbook/>. 3
- S. Boyd, S.-J. Kim, L. Vandenberghe, and A. Hassibi. A tutorial on geometric programming. *Optim. Eng.*, 8(1):67, 2007.
- P. J. Dickinson and L. Gijben. On the computational complexity of membership problems for the completely positive cone and its dual. *Comput. Optim. Appl.*, 57(2):403–415, 2014.  
<http://dx.doi.org/10.1007/s10589-013-9594-z>. 38
- M. Dür. Copositive programming – a survey. In M. Diehl, F. Glineur, E. Jarlebring, and W. Michiels, editors, *Recent Advances in Optimization and its Applications in Engineering: The 14th Belgian-French-German Conference on Optimization*, pages 3–20. Springer, Berlin, Heidelberg, 2010.  
[http://www.optimization-online.org/DB\\_FILE/2009/11/2464.pdf](http://www.optimization-online.org/DB_FILE/2009/11/2464.pdf). 38
- A. Fiacco and G. McCormick. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. Wiley, New York, 1968. 53
- C. A. Floudas and P. M. Pardalos, editors. *Encyclopedia of Optimization. 2nd ed.* Springer, New York, 2009. 29
- M. Frank and P. Wolfe. An algorithm for quadratic programming. *Naval Res. Logist. Quart.*, 3(1-2): 95–110, 1956. 51
- M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *J. Res. Natl. Bur. Stand.*, 49(6):409–436, 1952.  
<https://doi.org/10.6028/jres.049.044>. 54
- C. F. Higham and D. J. Higham. Deep learning: An introduction for applied mathematicians. *SIAM Rev.*, 61(4):860–891, 2019. 49

- M. Hutchings, F. Morgan, M. Ritoré, and A. Ros. Proof of the double bubble conjecture. *Ann. Math.*, 155(2):459–489, 2002.  
<https://arxiv.org/pdf/math/0406017>. 8
- W. Karush. Minima of functions of several variables with inequalities as side constraints. M.Sc. dissertation, Department of Mathematics, University of Chicago, Chicago, IL, USA, 1939. 43
- H. W. Kuhn and A. W. Tucker. Nonlinear programming. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950*, pages 481–492, Berkeley, 1951. University of California Press. 43
- K. Lange. *MM Optimization Algorithms*, volume 147 of *Other Titles Appl. Math.* SIAM, Philadelphia, PA, 2016. 22
- A. N. Langville and C. D. Meyer. *Who's #1? The science of rating and ranking*. Princeton University Press, Princeton, NJ, 2012. 29
- J. Liesen and Z. Strakoš. *Krylov Subspace Methods, Principles and Analysis*. Oxford University Press, Oxford, 2013. 54, 56
- D. Luenberger and Y. Ye. *Linear and Nonlinear Programming*. Springer, New York, third edition, 2008. 3, 54
- K. G. Murty and S. N. Kabadi. Some NP-complete problems in quadratic and nonlinear programming. *Math. Program.*, 39(2):117–129, 1987.  
<https://doi.org/10.1007/BF02592948>. 38
- Y. Nesterov and A. Nemirovskii. *Interior-Point Polynomial Algorithms in Convex Programming*. SIAM, Philadelphia, 1994.
- P. M. Pardalos and S. A. Vavasis. Quadratic programming with one negative eigenvalue is NP-hard. *J. Glob. Optim.*, 1(1):15–22, 1991.  
<https://doi.org/10.1007/BF00120662>. 29
- M. J. Todd. *Minimum-Volume Ellipsoids: Theory and Algorithms*, volume 23 of *MOS-SIAM Series on Optimization*. SIAM & Mathematical Optimization Society, Philadelphia, PA, 2016. 39
- S. A. Vavasis. *Nonlinear Optimization: Complexity Issues*. Oxford University Press, New York, 1991. 29
- W. Zhu. Unsolvability of some optimization problems. *Appl. Math. Comput.*, 174(2):921–926, 2006.  
<https://doi.org/10.1016/j.amc.2005.05.025>. 7
- G. Zoutendijk. *Methods of feasible directions: A study in linear and nonlinear programming*. PhD thesis, University of Amsterdam, Amsterdam, Netherlands, 1960. 51