Optimalizace

Elektronická skripta předmětu B0B33OPT,
verze **23. září 2024**.

Tomáš Werner
`werner@fel.cvut.cz`

Katedra kybernetiky
Fakulta elektrotechnická
České vysoké učení technické v Praze

Obsah

Kapitola 1

Značení a základní pojmy

Matematické výrazy na zvláštním řádku (rovnice apod.) jsou ve skriptech číslované číslem v závorce, takže např. (3.1) je vždy odkaz na rovnici. Na matematické věty se odkazujeme např. jako Věta 3.1, podobně na tvrzení, důsledky a lemmata. Odkazy na kapitoly, sekce a podsekce jsou značeny znakem paragraf, např. §3 je odkaz na kapitolu 3 a §3.1 je odkaz na sekci 1 v kapitole 3. Na řešené příklady se odkazujeme např. jako Příklad 3.1. Na konci každé kapitoly jsou cvičení, na která se odkazujeme např. jako Cvičení 3.1. Když tedy napíšeme třeba 'podívejte se na Příklad 3.1', znamená to něco *úplně jiného* než 'podívejte se na Cvičení 3.1'. Pokud potkáte slovo či sousloví vysázené **tučně**, jde o nově zavedený pojem, který máte pochopit a zapamatovat si. Ve skriptech tedy nepotkáte číslované definice, neboť všechny tučně vysázené pojmy jsou definice. Všechny tyto pojmy (a nějaké další) jsou v *rejstříku*. Slova vysázená *kurzívou* znamenají buď zdůraznění, nebo poprvé zmíněný důležitý avšak známý pojem. Odstavce, věty, důkazy, příklady a cvičení označené hvězdičkou ($\star$) jsou rozšiřující (a často zajímavé) ale nebudou se zkoušet. Je-li za nějakým tvrzením v závorce pokyn *odvoďte!, proveďte!, ověřte!* či otázka *proč?*, znamená to, že správnost tvrzení vám nemusí být jasná na první pohled, ale snadno ho dokážete, když se nad ním zamyslíte nebo si to napíšete na papír (pokud to nesvedete, musíte se zastavit a prostudovat příslušnou pasáž znovu).

1.1 Značení

Zopakujme matematické značení, které se používá v celých skriptech a které by student měl bezpečně ovládat.

1.1.1 Množiny

Názvy množin budeme psát velkými skloněnými písmeny, např. A nebo X. Budeme používat standardní množinové značení:

$\{a_1, \ldots, a_n\}$	množina s prvky $a_1, \ldots, a_n$
$a \in A$	prvek a patří do množiny A (neboli a je prvkem A)
$A \subseteq B$	množina A je podmnožinou množiny B, tj. každý prvek z A patří do B
$A = B$	množina A je rovna množině B, platí zároveň $A \subseteq B$ a $B \subseteq A$
$\{\, a \in A \mid \varphi(a) \,\}$	množina prvků a z množiny A, které splňují logický výrok $\varphi(a)$
$A \cup B$	sjednocení množin, množina $\{\, a \mid a \in A$ nebo $a \in B \,\}$
$A \cap B$	průnik množin, množina $\{\, a \mid a \in A,\ a \in B \,\}$

$A \setminus B$	rozdíl množin, množina $\{\, a \mid a \in A,\ a \notin B \,\}$
$\lvert A \rvert$	velikost (tj. počet prvků) množiny A (definováno jen pro konečné množiny)
$(a_1, \ldots, a_n)$	uspořádaná n-tice prvků $a_1, \ldots, a_n$
$A_1 \times \cdots \times A_n$	kartézský součin množin, množina $\{\, (a_1, \ldots, a_n) \mid a_1 \in A_1, \ldots, a_n \in A_n \,\}$
A^n	kartézská mocnina množiny, $A^n = A \times \cdots \times A$ (n-krát)
$\emptyset$	prázdná množina

Číselné množiny budeme značit takto:

$\mathbb{N}$	množina všech přirozených čísel, tj. množina $\{1, 2, 3, \ldots\}$
$\mathbb{Z}$	množina všech celých čísel
$\mathbb{Q}$	množina všech racionálních čísel
$\mathbb{R}$	množina všech reálných čísel
$\mathbb{R}_+$	množina nezáporných reálných čísel $\{\, x \in \mathbb{R} \mid x \geq 0 \,\}$
$\mathbb{R}_{++}$	množina kladných reálných čísel $\{\, x \in \mathbb{R} \mid x > 0 \,\}$
$[x_1, x_2]$	uzavřený reálný interval $\{\, x \in \mathbb{R} \mid x_1 \leq x \leq x_2 \,\}$
(x_1, x_2)	otevřený reálný interval $\{\, x \in \mathbb{R} \mid x_1 < x < x_2 \,\}$
$[x_1, x_2)$	polouzavřený reálný interval $\{\, x \in \mathbb{R} \mid x_1 \leq x < x_2 \,\}$
$\mathbb{C}$	množina komplexních čísel

Desetinná čísla budeme psát anglicky s desetinnou *tečkou*, např. 1.23 místo českého 1,23.

1.1.2 Zobrazení

Binární *relace* (angl. *relation*, tedy ‘vztah’) f množin A a B je podmnožina jejich kartézského součinu, tedy

$$f \subseteq A \times B. \tag{1.1}$$

Jestliže $(a, b) \in f$ a $(a, b') \in f$ implikuje $b = b'$ (tj. každý prvek množiny A je v relaci s nejvýše jedním prvkem množiny B), pak relaci nazýváme *zobrazení* z množiny A do množiny B a místo (1.1) ho značíme

$$f\colon A \to B \qquad \text{nebo (méně často)} \qquad A \xrightarrow{f} B \tag{1.2}$$

a místo $(a, b) \in f$ píšeme $b = f(a)$. Zobrazení si neformálně můžeme představit jako ‘černou skříňku’, která prvku $a \in A$ (*vzoru*) přiřadí právě jeden prvek $b = f(a) \in B$ (*obraz* neboli *hodnotu* prvku a v zobrazení f):

$$a \in A \longrightarrow \boxed{\ f\ } \longrightarrow f(a) \in B$$

Slovo *funkce* (angl. *function*) formálně znamená přesně to samé, jako zobrazení (angl. *mapping* or *map*), často se však používá pro zobrazení do číselných množin (tedy $B = \mathbb{R}$, $\mathbb{Z}$, $\mathbb{C}$ apod.).

Obraz množiny $A' \subseteq A$ v zobrazení $f\colon A \to B$ značíme

$$f(A') = \{\, f(a) \mid a \in A' \,\} = \{\, b \in B \mid \exists a \in A'\colon b = f(a) \,\}. \tag{1.3}$$

Např. je-li $A' = \{1, 3, 4, -1\} \subseteq \mathbb{Z}$ a zobrazení $f\colon \mathbb{Z} \to \mathbb{Z}$ má hodnoty $f(a) = a^2$, je $f(A') = \{\, a^2 \mid a \in A' \,\} = \{1, 9, 16\}$. Je-li $A' = \{\, a \in A \mid \varphi(a) \,\}$ a $f\colon A \to B$, používáme zkratku

$$f(A') = \{\, f(a) \mid a \in A,\ \varphi(a) \,\} \tag{1.4}$$

nebo jen $\{ f(a) \mid \varphi(a) \}$, je-li A jasné z kontextu. Např. $\{ x^2 \mid x \in \mathbb{R},\ -1 < x < 1 \} = [0, 1)$.

Množina A se nazývá *obor hodnot* (angl. *domain*), množina B *cílová množina* (angl. *co-domain*)[1] a množina $f(A)$ *obor hodnot* zobrazení (1.2). Vždy je $f(A) \subseteq B$, ale obecně nemusí být $f(A) = B$: např. zobrazení $f\colon \mathbb{Z} \to \mathbb{Z}$ s hodnotami $f(a) = a^2$ má cílovou množinu $\mathbb{Z}$ ale obor hodnot $f(\mathbb{Z}) = \{0, 1, 2, \ldots\}$.

Složení dvou zobrazení $f\colon A \to B$ a $g\colon B \to C$ (přehledněji píšeme $A \xrightarrow{f} B \xrightarrow{g} C$) je zobrazení $g \circ f\colon A \to C$ s hodnotami $(g \circ f)(a) = g(f(a))$ pro každé $a \in A$. Skládání je *asociativní*: pro $A \xrightarrow{f} B \xrightarrow{g} C \xrightarrow{h} D$ platí $(h \circ g) \circ f = h \circ (g \circ f)$. Proto můžeme závorky vynechat a psát jen $h \circ g \circ f$. Takto lze definovat složení libovolného počtu zobrazení.

Je-li $A' \subseteq A$, *restrikce* (neboli *zúžení*) zobrazení $f\colon A \to B$ na množinu A' je zobrazení $f|_{A'}\colon A' \to B$, které má na A' stejné hodnoty jako f, tj. $f|_{A'}(a) = f(a)$ pro každé $a \in A'$.

Zobrazení $f\colon A \to B$ se nazývá

- *injektivní* (neboli *prosté*), jestliže každý vzor má jiný obraz, tj. $f(a) = f(a') \Rightarrow a = a'$.
- *surjektivní* (neboli *na*), jestliže každý obraz má aspoň jeden vzor, tj. $f(A) = B$.
- *bijektivní* (neboli *vzájemně jednoznačné*), jestliže je injektivní a surjektivní. K bijektivnímu zobrazení existuje *inverzní zobrazení* $g\colon B \to A$ tak, že $f(x) = y \Leftrightarrow x = g(y)$ pro každé $x \in A$ a $y \in B$ (pak píšeme $g = f^{-1}$).

Zobrazení nějaké množiny do sebe, tedy $f\colon A \to A$, se často říká také *transformace*[2]. Transformace se nazývá

- *idempotentní*, jestliže $f \circ f = f$, tedy pro každé $a \in A$ platí $f(f(a)) = f(a)$,
- *involuce*, jestliže $f \circ f$ je identické zobrazení (tedy pro každé $a \in A$ platí $f(f(a)) = a$, neboli f je inverzí sebe sama),
- *permutace*, jestliže zobrazení f je bijekce a množina A je konečná.

1.1.3 Funkce a zobrazení více reálných proměnných

Symbol

$$\mathbb{R}^n = \underbrace{\mathbb{R} \times \cdots \times \mathbb{R}}_{n\text{-krát}}$$

značí kartézský součin množiny $\mathbb{R}$ se sebou n-krát (jak jsme zavedli v §1.1.1). Každý prvek množiny $\mathbb{R}^n$ je tedy uspořádaná n-tice reálných čísel $\mathbf{x} = (x_1, \ldots, x_n)$, které budeme říkat ($n$-rozměrný) **vektor**[3]. Zápis

$$\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m \tag{1.5}$$

označuje zobrazení, které vektoru $\mathbf{x} \in \mathbb{R}^n$ přiřadí vektor

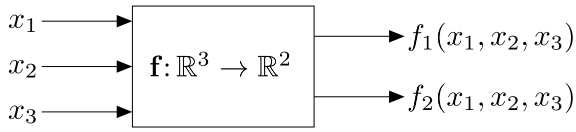
$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(x_1, \ldots, x_n) = (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x})) = (f_1(x_1, \ldots, x_n), \ldots, f_m(x_1, \ldots, x_n)) \in \mathbb{R}^m,$$

kde $f_1, \ldots, f_m\colon \mathbb{R}^n \to \mathbb{R}$ jsou *složky* zobrazení. Píšeme také $\mathbf{f} = (f_1, \ldots, f_m)$.

[1] Zatímco v angličtině se slovo ‘co-domain’ používá běžně, v češtině výraz ‘cílová množina’ není ustálený a pisatelé se mu spíše vyhýbají.

[2] Ne všichni autoři používají slovo ‘transformace’ v tomto smyslu, je to jen široce používaná konvence spíše než definice. Používá se zejména tehdy, má-li f geometrickou interpretaci.

[3] I když slovo *vektor* má v lineární algebře obecnější význam jako prvek libovolného (axiomy definovaného) vektorového prostoru, ve skriptech potkáte pouze vektory z $\mathbb{R}^n$ a občas $\mathbb{C}^n$. Takové vektory se obvykle píší tučně $\mathbf{x}$ spíše než $\vec{x}$. V pokročilejších textech se vektory značí jednoduše kurzívou x, tedy stejně jako skaláry.



Pro $m = 1$ jsou hodnotami zobrazení skaláry, proto budeme v tom případě psát jeho jméno kurzívou jako f. Pro $m > 1$ jsou hodnotami zobrazení vektory, proto jeho jméno budeme psát tučně jako $\mathbf{f}$. I když slova ‘funkce’ a ‘zobrazení’ znamenají jedno a to samé, budeme často pro $m = 1$ mluvit o *funkci* a pro $m > 1$ o *zobrazení*.

Je-li $n \in \{1, 2, 3\}$, často budeme proměnné v zobrazení $f\colon \mathbb{R}^n \to \mathbb{R}^m$ značit x, y, z místo x_1, x_2, x_3. Např. pro $f\colon \mathbb{R}^2 \to \mathbb{R}$ je pohodlnější psát $f(x, y) = x^2 - xy$ místo $f(x_1, x_2) = x_1^2 - x_1 x_2$.

1.2 Extrémy funkce na množině

Mějme množinu X', její podmnožinu $X \subseteq X'$ a funkci $f\colon X' \to \mathbb{R}$. Nechť $x^* \in X$ je takové, že $f(x^*) \leq f(x)$ pro všechna $x \in X$. Pak x^* nazveme *minimum* (přesněji *argument minima*, angl. *minimizer*) funkce f na množině X, nebo také říkáme, že funkce f *nabývá minima* na množině X v prvku x^*. Číslo $f(x^*)$ nazýváme *minimální hodnotou* funkce f na množině X, značíme ho

$$\min_{x \in X} f(x). \tag{1.6}$$

Pokud navíc je $f(x^*) < f(x)$ pro všechna $x \in X \setminus \{x^*\}$, mluvíme o *ostrém minimu*. Množinu všech argumentů minima funkce f na množině X značíme

$$\operatorname*{argmin}_{x \in X} f(x). \tag{1.7}$$

Na minimum funkce na množině se lze podívat i poněkud abstraktněji. Nechť Y je nějaká množina reálných čísel, tedy $Y \subseteq \mathbb{R}$. Prvek $y^* \in Y$ nazveme *nejmenší* (nebo také *minimální*) *prvek* množiny Y, jestliže $y^* \leq y$ pro všechna $y \in Y$. Tento nejmenší prvek značíme $\min Y$. Ne každá množina $Y \subseteq \mathbb{R}$ má nejmenší prvek (např. interval $(0, 1]$ ho nemá). Pokud ho ovšem má, má pouze jeden.

Označme

$$f(X) = \{\, f(x) \mid x \in X \,\} \subseteq \mathbb{R}$$

obraz množiny X funkcí f (viz §1.1.2). Pokud množina $f(X)$ má nejmenší prvek, platí tedy

$$\min_{x \in X} f(x) = \min\{\, f(x) \mid x \in X \,\} = \min f(X). \tag{1.8}$$

Budeme používat oba zápisy v (1.8).

Funkce nemusí mít na množině minimum, což plyne z toho, že ne každá množina $Y \subseteq \mathbb{R}$ má minimální prvek. V tom případě je množina (1.7) prázdná.

Podobně definujeme maximum funkce na množině a symboly ‘max’ a ‘argmax’. Minima a maxima funkce se souhrnně nazývají její *extrémy* nebo *optima*. Pokud odkaz na množinu X chybí, myslí se $X = X'$. Podotkněme, že hledání maxima funkce na množině lze snadno převést na hledání minima, protože samozřejmě

$$\max_{x \in X} f(x) = -\min_{x \in X}(-f(x)), \qquad \operatorname*{argmax}_{x \in X} f(x) = \operatorname*{argmin}_{x \in X}(-f(x)). \tag{1.9}$$

Příklad 1.1.

- Nechť $X' = X = [1, \infty)$ a $f(x) = 1/x$. Máme $f(X) = (0, 1]$. Ale množina $(0, 1]$ nemá minimální prvek, proto funkce f na množině X nemá minimum.
- $\min_{x \in \mathbb{R}} |x-1| = \min\{\, |x-1| \mid x \in \mathbb{R} \,\} = \min \mathbb{R}_+ = 0, \ \operatorname{argmin}_{x \in \mathbb{R}} |x-1| = \{1\}$
- Nechť $f(x) = \max\{|x|, 1\}$. Pak $\operatorname{argmin}_{x \in \mathbb{R}} f(x) = [-1, 1]$.
- Nechť $(a_1, a_2, \ldots, a_5) = (1, 2, 3, 2, 3)$. Pak[4] $\max_{i=1}^{5} a_i = 3, \ \operatorname{argmax}_{i=1}^{5} a_i = \{3, 5\}$. ♦

1.3 Úloha spojité optimalizace

Matematická optimalizace se obecně zabývá úlohami, ve kterých hledáme minima nějaké funkce $f \colon X' \to \mathbb{R}$ na nějaké množině $X \subseteq X'$, ve smyslu minulého odstavce §1.2. Kromě pojmů z §1.2 se v matematické optimalizaci používají navíc následující pojmy:

- Prvky množiny X se nazývají *přípustná řešení* úlohy. Je to množina všech možností, ze kterých můžeme vybírat.
- Funkce f se nazývá *účelová* (také pokutová, cenová, kriteriální) funkce. Pro každou dovolenou možnost (přípustné řešení) z množiny X kvantifikuje, jak špatná je tato možnost.
- Prvky množiny $\operatorname{argmin}_{\mathbf{x} \in X} f(\mathbf{x})$ se nazývají *optimální řešení* nebo krátce *optima* úlohy.
- Číslo $\min_{\mathbf{x} \in X} f(\mathbf{x})$ se nazývá *optimální hodnota* úlohy.

Podotkněme, že zápisem (1.6) se někdy rozumí *minimální hodnota* funkce f na množině X (tj. reálné číslo) a někdy se jím rozumí *úloha* hledání minim funkce f na množině X. Pak tedy mluvíme např. o přípustných řešeních nebo optimální hodnotě úlohy (1.6).

Formulace (1.6) je velmi obecná, neboť množina X může být libovolná. Obvykle se rozlišují tři široké třídy úloh:

- Pokud je množina X konečná (i když třeba hodně velká), mluvíme o *kombinatorické optimalizaci*. Její prvky mohou být např. reálné vektory, cesty v grafu, konfigurace Rubikovy kostky, nebo textové řetězce konečné délky. Příkladem je hledání nejkratší cesty v grafu nebo problém obchodního cestujícího.
- Pokud množina X obsahuje nespočetné množství reálných vektorů (tedy $X \subseteq \mathbb{R}^n$) a je navíc popsatelná způsobem uvedeným dále, mluvíme o *spojité optimalizaci*. Příkladem je úloha lineárního programování.
- Pokud množina X obsahuje reálné funkce, mluvíme o *variačním počtu*. Příkladem je hledání rovinné křivky, která sama sebe neprotíná a při dané délce obepíná co největší plochu.

Tento kurs se zabývá *spojitou optimalizací*, ve které $X' = \mathbb{R}^n$ a množina X má nespočetný počet prvků a je popsána jako množina řešení soustavy rovnic a nerovnic. Tedy X je množina všech vektorů $(x_1, \ldots, x_n) \in \mathbb{R}^n$ splňujících

$$g_i(x_1, \ldots, x_n) \le 0, \quad i = 1, \ldots, m \tag{1.10a}$$
$$h_i(x_1, \ldots, x_n) = 0, \quad i = 1, \ldots, l \tag{1.10b}$$

[4]Místo $\max_{i=1}^{5} a_i$ se obvykle píše $\max_{i=1,\ldots,5} a_i$. Používáme zde první způsob v analogii se značením $\sum_{i=1}^{5} a_i$.

pro dané funkce $g_1, \dots, g_m, h_1, \dots, h_l \colon \mathbb{R}^n \to \mathbb{R}$. Funkce f, g_i, h_i jsou obvykle spojité a často i diferencovatelné a množina X je obvykle nespočetná a souvislá (nebo je aspoň sjednocením malého počtu souvislých množin). Říkáme také, že minimalizujeme funkci f *za podmínek* (1.10) příp. (1.12). To se zapisuje také jako

$$\begin{aligned} \min \quad & f(x_1, \dots, x_n) \\ \text{za podmínek} \quad & g_i(x_1, \dots, x_n) \le 0, \quad i = 1, \dots, m \\ & h_i(x_1, \dots, x_n) = 0, \quad i = 1, \dots, l \\ & x_1, \dots, x_n \in \mathbb{R} \end{aligned} \tag{1.11}$$

Podmínka $x_1, \dots, x_n \in \mathbb{R}$ v (1.11) není vlastně omezení, často se ale píše, aby bylo jasné, co jsou proměnné úlohy. Toto je tedy *úloha spojité optimalizace v obecném tvaru.*

Soustavu (1.10) lze napsat kratčeji ve vektorovém značení jako

$$\mathbf{g}(\mathbf{x}) \le \mathbf{0}, \tag{1.12a}$$
$$\mathbf{h}(\mathbf{x}) = \mathbf{0}, \tag{1.12b}$$

kde $\mathbf{g} \colon \mathbb{R}^n \to \mathbb{R}^m$, $\mathbf{h} \colon \mathbb{R}^n \to \mathbb{R}^l$ a $\mathbf{0}$ značí nulové vektory příslušné dimenze. Tedy

$$X = \{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{g}(\mathbf{x}) \le \mathbf{0},\ \mathbf{h}(\mathbf{x}) = \mathbf{0} \,\}. \tag{1.13}$$

Ve shodě se značením (1.4) se pak úloha (1.11) může zapsat jako

$$\min\{\, f(\mathbf{x}) \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{g}(\mathbf{x}) \le \mathbf{0},\ \mathbf{h}(\mathbf{x}) = \mathbf{0} \,\}. \tag{1.14}$$

Zápis (1.14) může opět označovat nejen úlohu minimalizace funkce f na množině (1.13), ale také minimální hodnotu funkce f na množině (1.13).

Užívají se (ne zcela jednotně) tyto pojmy:

- Rovnice a nerovnice (1.10) se nazývají *omezující podmínky*, krátce *omezení.*
- Omezení (1.10a) příp. (1.10b) se nazývají omezení *typu nerovnosti* příp. *typu rovnosti.*
- Minimům funkce f na množině (1.13) se také říká *minima funkce f vázané podmínkami* (1.10). Podobně pro maxima a extrémy (tj. minima nebo maxima).
- Pokud omezení chybí ($m = l = 0$, tedy $X = \mathbb{R}^n$), mluvíme o *volných minimech* funkce f a o úloze *bez omezení.*
- Pokud je omezení typu nerovnosti $g_i(\mathbf{x}) \le 0$ v nějakém bodě $\mathbf{x} \in \mathbb{R}^n$ splněno s rovností, tedy $g_i(\mathbf{x}) = 0$, říkáme, že toto omezení je v bodě $\mathbf{x}$ *aktivní.*
- Pokud $X \neq \emptyset$, úloha se nazývá *přípustná*, v opačném případě ($X = \emptyset$) je *nepřípustná.*

1.3.1 Vybrané příklady

Dále uvedeme několik příkladů formulace (a často i řešení) úloh ve tvaru (1.11). Některé z nich byste měli být schopni řešit znalostmi a dovednostmi ze střední školy příp. z analýzy, jde tedy o opakování. Jiné předesílají, co přijde v pozdějších kapitolách.

Příklad 1.2. Pastevec vlastní 100 metrů pletiva a chce z něj udělat ohradu pro ovce o co největším obsahu. Ohrada bude mít tvar obdélníka, z něhož tři strany budou tvořeny plotem a zbylá strana řekou (ovce neplavou, řeka tedy slouží jako plot).

Označme strany obdélníka jako x, y. Obsah obdélníka je xy a jeho obvod (bez strany tvořené řekou) $2x + y$. Řešíme úlohu

$$\begin{aligned} \max \quad & xy \\ \text{za podmínek} \quad & 2x + y = 100 \\ & x, y \geq 0 \end{aligned}$$

neboli

$$\max\{\, xy \mid x \geq 0,\ y \geq 0,\ 2x + y = 100 \,\}.$$

To je úloha s $n = 2$ proměnnými, $m = 2$ omezeními typu nerovnosti a $l = 1$ omezením typu rovnosti.

I když jde o úlohu s omezeními, bylo by velmi nešikovné snažit se ji řešit formalismem Lagrangeových multiplikátorů či KKT podmínek (pokud je někdo zná). Místo toho z podmínky $2x + y = 100$ vyjádříme $y = 100 - 2x$ (přesněji: rovnice $2x + y = 100$ je ekvivalentní rovnici $y = 100 - 2x$) a dosadíme do původní úlohy. Při eliminaci proměnné y nesmíme zapomenout na nerovnosti: podmínka $y \geq 0$ spolu s $y = 100 - 2x$ implikují $x \leq 50$. Tím dostaneme úlohu s jednou proměnnou

$$\max\{\, x(100 - 2x) \mid 0 \leq x \leq 50 \,\},$$

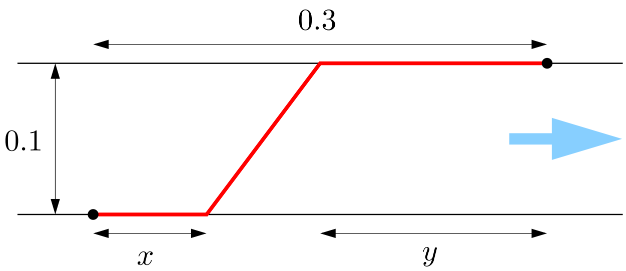
což někdo raději píše jako $\max\limits_{0 \leq x \leq 50} x(100 - 2x)$ nebo $\max\limits_{x \in [0,50]} x(100 - 2x)$.

Tu snadno vyřešíme metodami analýzy funkcí jedné proměnné. Optimální řešení může být buď ve vnitřním bodě nebo v jednom z krajních bodů intervalu $[0, 50]$. Maximum výrazu $x(100 - 2x)$ na množině $\mathbb{R}$ snadno najdeme pomocí derivace, nabývá se v bodě $x = 25$. Ten zároveň splňuje podmínku $0 \leq x \leq 50$. Body na krajích intervalu mají menší hodnotu kritéria, tedy nejsou optimální. Dosazením dostaneme $y = 100 - 2x = 50$.

Úlohu lze ovšem řešit i jednodušeji následující úvahou. Představíme si, že ohrada symetricky pokračuje i za řekou, tedy se jedná o obdélník rozpůlený řekou. Taková ohrada bude mít dvakrát větší obsah i obvod než původní ohrada. Nyní použijeme skutečnost (která je intuitivně jasná, i když jsme ji neověřili výpočtem), že obdélník o daném obvodu a největším možném obsahu je čtverec. Tedy původní ohrada musí být půlka čtverce, tj. mít poměr stran $1 : 2$. ♦

Příklad 1.3. Jste na pravém břehu řeky široké 0.1 km a chcete se dostat ke stanu na levém břehu, který je 0.3 km po proudu od bodu, který je na levém břehu nejblíže vám. Řeka teče pomalu, zanedbatelnou rychlostí. Plavete rychlostí 1 km/h a chodíte rychlostí 3 km/h (běhat odmítáte, protože je vedro). Jaký je nejkratší čas, za který se dokážete dostat ke stanu?

Optimální dráha bude mít tvar jako na obrázku:



Tedy nejdřív jdeme kus x po pravém břehu, pak přeplaveme šikmo do bodu na levém břehu vzdáleném y od stanu, nakonec dojdeme kus y po levém břehu.

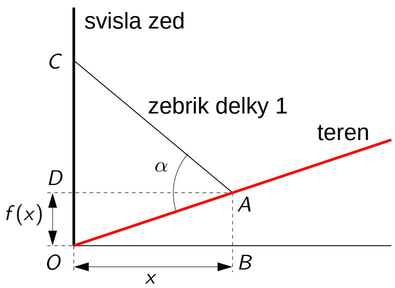
Může mít dráha jiný tvar? Těžko, pokud přijmeme, že dráha, kterou se dostaneme nejrychleji z bodu do bodu za předpokladu, že v každém místě prostoru se pohybujeme stejně rychle, je úsečka spojující tyto dva body.

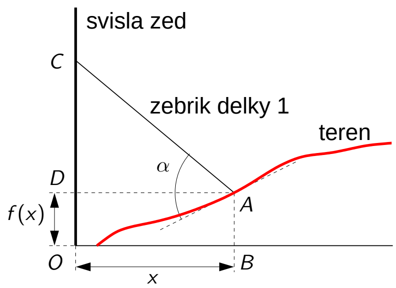
Celkový čas je $t = \frac{1}{3}x + \sqrt{0.1^2 + (0.3 - x - y)^2} + \frac{1}{3}y$ (hodin). Ten chceme minimalizovat pro proměnné $x, y \in \mathbb{R}$. Všimněte si, že nepotřebujeme omezení $x, y \geq 0$ ani $x + y \leq 0.3$, neboť i bez nich dostaneme přípustné dráhy. Ovšem tyto dráhy očividně nejsou optimální, protože pro $x < 0$ bychom na začátku hloupě kráčeli pryč od stanu, pro $y < 0$ bychom doplavali až za stan, a pro $x + y > 0.3$ bychom zase plavali pryč od stanu. Takže kdybychom tato omezení přidali, úloha by se nezměnila (tj. omezení jsou redundantní).

Máme tedy úlohu se dvěma proměnnými bez omezení. Ovšem proměnné se vyskytují jen v součtu, tedy označíme-li $z = x + y$, je $t = \frac{1}{3}z + \sqrt{0.1^2 + (0.3 - z)^2}$ (samozřejmě vidíme i z obrázku, že x nebo y můžeme uvažovat nulové bez újmy na obecnosti). Pomocí derivace získáme stacionární bod $z = 0.3 - \sqrt{2}/40$ km ≈ 264.6 m. Minimální čas je $t \approx 0.19428$ hodin, tedy asi 11.7 minut. ♦

Příklad 1.4. Zloděj má žebřík délky 1 a potřebuje se dostat přes svislou zeď ze strany, kde terén (v obrázku červeně) stoupá s konstantní (kladnou) směrnicí k. Jak daleko od zdi musí zloděj žebřík zapíchnout do země, aby jeho druhý konec dosáhl co nejvýše na zeď? Jaký bude v tom případě úhel mezi žebříkem a terénem?

Nakreslíme situaci (levý obrázek):





Hledáme hodnotu x, která maximalizuje výšku vršku žebříku na zdi $|OC| = kx + \sqrt{1 - x^2}$, za podmínky $x \geq 0$ (protože žebřík nemůžeme zapíchnout za zeď). Tuto podmínku ale můžeme ignorovat, protože zjevně žádná poloha s $x < 0$ nebude optimální (vršek bude níž než pro jakékoliv $x \geq 0$). Tedy opět analýza funkcí jedné proměnné: podmínka stacionarity je $k - x/\sqrt{1 - x^2} = 0$, z toho $x = 1/\sqrt{1 + 1/k^2}$. Jedná se o maximum, což můžeme ověřit pomocí druhé derivace (musí být záporná).

Ukážeme, že pro optimální x bude žebřík kolmý k terénu (tj. $\alpha = \pi/2$). Z podmínky $k = x/\sqrt{1 - x^2}$ vidíme, že $|AD|/|CD| = k$, protože $|AD| = x$ a $|CD| = \sqrt{1 - x^2}$. Ale z definice směrnice je také $kx/x = |DO|/|AD| = |AB|/|BO| = k$. Tedy pravoúhlé trojúhelníky DCA, BOA a DAO jsou si podobné. Protože součet úhlů v trojúhelníku je π, úhel CAO (tj. α) musí být pravý.

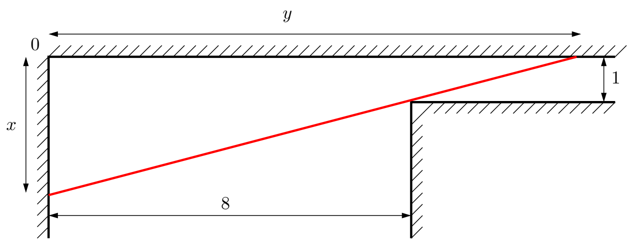
Tento výsledek jsme mohli uhodnout následující úvahou: pokud $\alpha < \frac{\pi}{2}$, pak zmenšením x vršek žebříku C povyleze vzhůru. Podobně pro $\alpha > \frac{\pi}{2}$ a zvětšení x. Když ale $\alpha = \frac{\pi}{2}$, pak zmenšení i zvětšení x způsobí pokles bodu C. Jinými slovy, z obrázku vidíme, že $\alpha = \frac{\pi}{2}$ je lokální maximum.

Úlohu nyní zobecníme tak, že terén nemá tvar přímky $f(x) = kx$, ale obecné neklesající diferencovatelné funkce $f(x)$ (viz obrázek nahoře vpravo). Dosahuje-li žebřík na zdi (bod C) nejvýše, jaký bude úhel (opět označený α) mezi žebříkem a terénem? Minimalizujeme výraz

$f(x) + \sqrt{1-x^2}$. Stacionární podmínka zní $f'(x) - x/\sqrt{1-x^2} = 0$. To opět znamená, že je žebřík kolmý k terénu (odvození podobné jako minule). Ovšem v tomto případě bod splňující stacionární podmínku nemusí být globální maximum, ale může to být jen lokální maximum nebo dokonce lokální minimum. ♦

Příklad 1.5. Jdete chodbou o šířce 8 m, která zatáčí do pravého úhlu a zároveň se zužuje na šířku 1 m (viz obrázek). Jaká je největší možná délka rovné tuhé tyče, se kterou lze zatáčkou projít? Požaduje se, abyste tyč drželi stále vodorovně.

Nejdelší tyč, se kterou projedeme, se určitě bude dotýkat vnitřního rohu chodby:



Uvažujme pravoúhlý trojúhelník, jehož přepona je tyč (na obrázku červeně) dotýkající se tohoto rohu a odvěsny jsou označeny x a y. Potřebujeme najít délky x a y, pro které bude délka odvěsny minimální (to bude chvíle, kdy bychom už delší tyč nepronesli). Trojúhelník má vrcholy $(0,0)$, $(x,0)$ a $(0,y)$. Z podobnosti trojúhelníků máme $(x-1)/1 = 8/(y-8)$. Za této podmínky minimalizujeme čtverec délky odvěsny x^2+y^2. Měli bychom přidat ještě podmínky $x, y \geq 0$, ale opět se snadno ukáže, že v optimálním řešení budou splněny.

Z podmínky vyjádříme $y = (8x)/(x-1) = 8/(1-1/x)$ a dosadíme do kritéria, takže minimalizujeme $x^2 + 64/(1-1/x)^2$. Stacionární podmínka je $2x(x^3-3x^2+3x-65)/(x-1)^3 = 0$, což je ekvivalentní $2x(x^3-3x^2+3x-65) = 0$ za předpokladu $x \neq 1$. Bohužel, to je rovnice 4. stupně. Její jediné nenulové reálné řešení je $x = 5$, což se dá uhodnout nebo použít vhodný numerický algoritmus. Z toho $y = 10$. Tedy největší možná délka tyče je $\sqrt{x^2+y^2} = \sqrt{5^2+10^2} = 5\sqrt{5} \approx 11.18034$.

Existuje jiný, jednodušší způsob řešení, kterým se navíc vyhneme rovnici 4. stupně. Délku tyče lze napsat jako $8/\cos\alpha + 1/\sin\alpha$ kde α je úhel mezi tyčí a osou x. Hledáme α, pro které je tato délka minimální. Stacionarní podmínka je $2\sin\alpha = \cos\alpha$, tedy $\tan\alpha = \frac{1}{2}$. Lze ověřit, že jde o minimum. ♦

Příklad 1.6. Na zahradě vám rostou melouny. Dnes je jich tam 200 kg. Každý den melouny přirostou o 5 kg, ale cena 1 kg melounů na trhu klesne o 10 haléřů. Pokud dnešní cena 1 kg melounů na trhu je 9 Kč, jak dlouho máte čekat se sklizní, abyste dosáhli co největšího zisku? Předpokládejte, že melouny sklidíte a prodáte tentýž den.

Cena na trhu za t dnů je $(200+5t)(9-t/10) = -t^2/2 + 25t + 1800$. Chceme najít takové t, pro které tato funkce bude maximální. Porovnáním derivace s nulou dostaneme stacionární bod $t = 25$, který je maximem.

Ovšem pozor: správně jsme měli požadovat, aby t bylo nezáporné a celočíselné, tedy $t \geq 0$ a $t \in \mathbb{Z}$. V našem případě jsme měli štěstí a optimální řešení t tyto podmínky splňuje. I kdyby je nesplňovalo, snadno bychom z něj spočetli optimální řešení, které by je splňovalo (jak?). ♦

K poslednímu příkladu poznamenejme, že žádný skutečný zelinář by takovou úlohu nepočítal. Hmotnost melounů ani jejich výkupní cena nejsou lineární funkce času a skutečné funkce není snadné odhadnout. I kdyby je zelinář znal, jeho kritériem není jen velikost zisku z melounů, protože např. pěstuje i spoustu jiných věcí a v den optimální sklizně melounů zrovna musí okopávat okurky nebo se dívat na fotbal. Většina slovních úloh v těchto skriptech bude takto nerealisticky zjednodušená (půjde o tzv. *toy problems*). Skutečné úlohy z optimalizační praxe jsou obvykle příliš složité na to, abychom je v tomto kursu dokázali uvést a vysvětlit. Tato složitost je ovšem často jen kvantitavního rázu (více proměnných a více omezení), *typ* úlohy je stejný jako u našich zjednodušených úloh.

Příklad 1.7. Řešme tuto soustavu třech lineárních rovnic o dvou neznámých:

$$\begin{aligned} x + 2y &= 6 \\ -x + y &= 3 \\ x + y &= 4 \end{aligned}$$

Tato soustava nemá řešení (je přeurčená). Chceme najít alespoň přibližné řešení.

Termín ‘přibližné řešení’ není jednoznačný, různí lidé jím mohou myslet různé věci. Velmi často užívaná formulace je minimalizovat součet čtverců[5] zbytků (residuí) jednotlivých rovnic. Tedy minimalizujeme funkci

$$f(x,y) = (x + 2y - 6)^2 + (-x + y - 3)^2 + (x + y - 4)^2$$

na množině $\mathbb{R}^2$. Je jasné, že $f(x,y) > 0$ pro všechna $x, y \in \mathbb{R}$, kdyby totiž $f(x,y) = 0$ pak by x, y bylo řešení soustavy. Minimum funkce f najdeme snadno, protože f je kvadratická funkce dvou proměnných, zdola omezená nulou. Nutná a postačující podmínka na minimum je nulovost parciálních derivací.

Jiný způsob, jak formalizovat termín ‘přibližné řešení’, je minimalizovat funkci

$$f(x,y) = |x + 2y - 6| + |-x + y - 3| + |x + y - 4|.$$

Zde už nám derivace nepomohou. Později uvidíme, že tato funkce je konvexní a její minimalizaci lze převést na úlohu lineárního programování. ♦

Příklad 1.8. Nechť reálná veličina y závisí na reálné veličině x kvadraticky, tj. $y = ax^2 + bx + c$. Neznáme ovšem koeficienty a, b, c. Naměřili jsme hodnoty veličiny y pro m hodnot veličiny x, tj. máme dvojice $(x_i, y_i)_{i=1}^m$. Jak najdeme koeficienty a, b, c?

To je úloha na lineární regresi. Lze ji formulovat ve smyslu nejmenších čtverců jako minimalizaci funkce

$$f(a,b,c) = \sum_{i=1}^m (ax_i^2 + bx_i + c - y_i)^2$$

přes proměnné $a, b, c \in \mathbb{R}$. Minimalizace funkce f je příkladem lineární úlohy nejmenších čtverců. Nutná (a zde i postačující) podmínka na minimum je opět nulovost parciálních derivací. ♦

Příklad 1.9. Jsou dány body $\mathbf{a}_1 \neq \mathbf{b}_1$ a $\mathbf{a}_2 \neq \mathbf{b}_2$ v prostoru $\mathbb{R}^n$. Najděte vzdálenost přímky procházející body $\mathbf{a}_1, \mathbf{b}_1$ od přímky procházející body $\mathbf{a}_2, \mathbf{b}_2$ (tj. vzdálenost mimoběžek, neboli délka příčky mimoběžek).

[5] ‘Čtverec’ čísla znamená jeho druhou mocninu. Tento historický termín je jednak hezký a jednak se používá tak často, že ho budeme používat i my.

Víme, že i-tá přímka (kde $i = 1, 2$) je množina $\{\, \mathbf{a}_i + \alpha(\mathbf{b}_i - \mathbf{a}_i) \mid \alpha \in \mathbb{R} \,\}$. Hledáme bod na první přímce a bod na druhé přímce tak, aby jejich vzdálenost byla co nejmenší. Tedy minimalizujeme funkci

$$f(\alpha_1, \alpha_2) = \|\mathbf{a}_1 + \alpha_1(\mathbf{b}_1 - \mathbf{a}_1) - \mathbf{a}_2 - \alpha_2(\mathbf{b}_2 - \mathbf{a}_2)\|^2, \tag{1.15}$$

kde

$$\|\mathbf{x}\| = (x_1^2 + \cdots + x_n^2)^{1/2}$$

označuje eukleidovskou normu vektoru $\mathbf{x} \in \mathbb{R}^n$. Funkce (1.15) je opět kvadratická funkce dvou proměnných. ♦

Příklad 1.10. Pán[6] u stánku prodává lupínky za 120 Kč/kg a hranolky za 76 Kč/kg. Na výrobu 1 kg lupínků se spotřebuje 2 kg brambor a 0.4 kg oleje. Na výrobu 1 kg hranolků se spotřebuje 1.5 kg brambor a 0.2 kg oleje. Je nakoupeno 100 kg brambor a 16 kg oleje. Kolik má pán vyrobit lupínků a kolik hranolků, aby co nejvíce utržil? Přitom nepočítáme cenu surovin a předpokládáme, že všechny výrobky se prodají a nevyužité suroviny se po pracovní době vyhodí.

Tuto úlohu lze formalizovat takto:

$$\begin{array}{rrcl} \max & 120l + 76h & & \\ \text{za podmínek} & 2l + 1.5h & \leq & 100 \\ & 0.4l + 0.2h & \leq & 16 \\ & l, h & \geq & 0 \end{array}$$

Tato úloha patří do třídy lineárního programování, ve kterém minimalizujeme nebo maximalizujeme lineární funkci za podmínek lineárních nerovností a rovností. Optimální řešení je $l = 20$ kg lupínků a $h = 40$ kg hranolků. Podotkněme, že v tomto příkladě jsou omezení $x, y \geq 0$ zbytečná – kdybychom je vynechali, optimální řešení by se nezměnilo (to je vidět z toho, že ani jedna z podmínek $x, y \geq 0$ není v optimu aktivní). Ovšem kdyby druhů surovin a výrobků bylo více, podmínky by zbytečné nebyly. ♦

Příklad 1.11. Najděte nejmenší n-rozměrnou kouli obsahující dané body $\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^n$.

Koule se středem $\mathbf{x}$ a poloměrem r je množina $\{\, \mathbf{a} \in \mathbb{R}^n \mid \|\mathbf{x} - \mathbf{a}\| \leq r \,\}$. Úlohu lze tedy napsat jako

$$\begin{array}{rl} \min & r \\ \text{za podmínek} & \|\mathbf{x} - \mathbf{a}_i\| \leq r, \quad i = 1, \ldots, m \\ & \mathbf{x} \in \mathbb{R}^n \\ & r \in \mathbb{R} \end{array} \tag{1.16}$$

Mohla by tam být ještě podmínka $r \geq 0$, ta je ale zbytečná, protože norma je vždy nezáporná. Úloha (1.16) je konvexní a patří do třídy programování na kuželu druhého řádu (SOCP), což si ovšem řekneme až mnohem později.

Protože minimalizaci přes proměnnou r lze udělat explicitně, úloha (1.16) je ekvivalentní minimalizaci funkce

$$f(\mathbf{x}) = \max_{i=1}^{m} \|\mathbf{x} - \mathbf{a}_i\| \tag{1.17}$$

[6]Úloha převzata z elektronických skript *Petr Hliněný: Optimalizační úlohy. Katedra informatiky VŠB-TU Ostrava, 2006.*

na množině $\mathbb{R}^n$ (tj. bez omezení). To lze napsat také jako

$$\min_{\mathbf{x}\in\mathbb{R}^n} \max_{i=1}^{m} \|\mathbf{x}-\mathbf{a}_i\|. \tag{1.18}$$

Umocněním podmínek na druhou a substitucí $r^2 = y$ můžeme úlohu napsat ještě v jiném ekvivalentním tvaru

$$\begin{aligned} \min \quad & y \\ \text{za podmínek} \quad & \|\mathbf{x}-\mathbf{a}_i\|^2 \le y, \quad i=1,\dots,m \\ & \mathbf{x} \in \mathbb{R}^n \\ & y \in \mathbb{R} \end{aligned} \tag{1.19}$$

neboli (po eliminaci y)

$$\min_{\mathbf{x}\in\mathbb{R}^n} \max_{i=1}^{m} \|\mathbf{x}-\mathbf{a}_i\|^2. \tag{1.20}$$

♦

Úloha (1.19) je konvexní úloha kvadratického programování (QCQP), o které si také později řekneme.

Na konvexní QCQP a SOCP existují numerické metody, které bychom mohli mechanicky použít. Ovšem na tuto úlohu byly vymyšleny i speciální algoritmy (hlavně pro případy $n = 2$ a $n = 3$), které jsou rychlejší.

Příklad 1.12. Jsou dány body $\mathbf{a}_1,\dots,\mathbf{a}_m \in \mathbb{R}^n$ a hledáme vektor $\mathbf{x} = (x_1,\dots,x_n) \in \mathbb{R}^n$, který minimalizuje (bez omezujících podmínek) funkci

$$f(\mathbf{x}) = \sum_{i=1}^{m} \|\mathbf{x}-\mathbf{a}_i\|^2. \tag{1.21}$$

Označíme-li $\mathbf{a}_i = (a_{i1},\dots,a_{in})$, je $\sum_{i=1}^m \|\mathbf{x}-\mathbf{a}_i\|^2 = \sum_{i=1}^m \sum_{j=1}^n (x_j - a_{ij})^2$, tedy účelová funkce je součtem n funkcí, z nichž každá závisí jen na jedné souřadnici x_j. Minimum funkce f lze tedy najít tak, že najdeme minimum každé funkce zvlášť (úloha se nám tak 'rozpadla' na n nezávislých optimalizačních úloh). Jak snadno spočtete (a jistě jste dávno uhodli), minimum se nabývá v *těžišti* $(\mathbf{a}_1 + \cdots + \mathbf{a}_m)/m$ daných bodů. ♦

Příklad 1.13. Jsou dána čísla $a_1,\dots,a_m \in \mathbb{R}$. Najděte minimum funkce

$$f(x) = \sum_{i=1}^{m} |x - a_i|. \tag{1.22}$$

Uvidíme později, že tato úloha se dá převést na úlohu lineárního programování. Ovšem její řešení lze nalézt i elementární úvahou. Seřadíme čísla $a_1,\dots,a_m$ vzestupně, tedy předpokládáme $a_1 \le a_2 \le \cdots \le a_{m-1} \le a_m$. Funkce f je po částech lineární (přesněji: afinní), je diferencovatelná všude kromě bodů a_i. Derivace je konstantní pro každý interval (a_{i-1}, a_i). Najděte hodnoty těchto derivací. K tomu si nejprve ujasněte, jak vypadá funkce $|x - a_i|$ pro jediné i a jaká je její derivace (derivaci můžete napsat pomocí funkce signum).

Z derivací na intervalech usoudíme, kde je funkce f klesající, kde rostoucí a kde konstantní. Nyní je jasné, kde f nabývá minima: pro m liché je to v bodě $a_{(m+1)/2}$, pro m sudé na intervalu $[a_{m/2}, a_{m/2+1}]$. Argument minima se označuje jako *medián* čísel $a_1,\dots,a_m$. Přesněji: pro liché m je mediánem prostřední z čísel $a_1,\dots,a_m$, tedy $a_{(m+1)/2}$, pro sudé m se za medián považuje číslo $(a_{m/2} + a_{m/2+1})/2$. ♦

Příklad 1.14. Jsou dány body $\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^n$. Najděte minimum funkce

$$f(\mathbf{x}) = \sum_{i=1}^{m} \|\mathbf{x} - \mathbf{a}_i\|. \tag{1.23}$$

Pro $n = 1$ se funkce (1.23) redukuje na (1.22). Pro $n \geq 2$ je řešení úlohy známo jako *geometrický medián*. Ovšem na rozdíl od obyčejného mediánu se minimum obecně nenabývá v žádném z bodů $\mathbf{a}_1, \ldots, \mathbf{a}_m$. Neexistuje algoritmus, který by pro $n \geq 2$ našel minimum funkce (1.23) v konečném počtu kroků. Dobrá zpráva ale je, že úloha je konvexní (patří do třídy SOCP, o které jsme se už zmínili).

Pro případ $n = 2$ má úloha jednoduchý mechanický model[7]. Do vodorovného prkna vyvrtáme díry o souřadnicích $\mathbf{a}_i$. Každou dírou provlečeme provázek. Provázky jsou nahoře svázané uzlem do jednoho bodu a dole mají závaží o stejné hmotnosti. Poloha uzlu je $\mathbf{x}$. Hodnota $f(\mathbf{x})$ je potenciální energie soustavy a ustálený stav odpovídá minimu $f(\mathbf{x})$. ♦

Příklad 1.15. Mějme m bodů $\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^n$ v n-rozměrném prostoru. Úkolem je rozmístit dalších k bodů $\mathbf{x}_1, \ldots, \mathbf{x}_k \in \mathbb{R}^n$ tak, aby průměrná vzdálenost bodu $\mathbf{a}_i$ od nejbližšího bodu $\mathbf{x}_j$ byla co nejmenší. Tedy minimalizujeme účelovou funkci

$$f(\mathbf{x}_1, \ldots, \mathbf{x}_k) = \sum_{i=1}^{m} \min_{j=1}^{k} \|\mathbf{a}_i - \mathbf{x}_j\| \tag{1.24}$$

pro neznámé vektory $\mathbf{x}_1, \ldots, \mathbf{x}_k \in \mathbb{R}^n$. Úloha je známá jako *shlukování* (angl. *clustering*). Jako motivaci si představme optimální rozmístění cisteren ve vesnici, kde občas neteče voda. Zde máme $n = 2$, $\mathbf{a}_i$ jsou souřadnice domů a $\mathbf{x}_j$ jsou souřadnice cisteren. Chceme, aby průměrná vzdálenost obyvatele od nejbližší cisterny byla co nejmenší.

Není (a pravděpodobně nikdy nebude) znám algoritmus, který by pro libovolný vstup (tedy libovolné $n, m, k \in \mathbb{N}$ a $\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^n$) našel globální minimum funkce (1.24) za prakticky přijatelný čas. Lze totiž dokázat, že úloha je tzv. NP-těžká[8]. V praktické situaci často použijeme algoritmus, který najde pouze přibližné (typicky lokální) optimum, např. *k-means*. ♦

Příklad 1.16. Nechť (V, E) je ohodnocený neorientovaný graf, tj. V je konečná množina, $E \subseteq \binom{V}{2}$ je množina dvouprvkových podmnožin V, a je dáno zobrazení $c\colon E \to \mathbb{R}$ (tedy každá hrana $\{i, j\} \in E$ má přiřazené číslo c_{ij}). Úloha na *maximální řez* (*maximum cut*)[9] v grafu zní

$$\begin{aligned} \max \quad & \sum_{\{i,j\} \in E} c_{ij} x_i x_j \\ \text{za podmínek} \quad & x_i^2 = 1, \quad i \in V \\ & x_i \in \mathbb{R}, \quad i \in V \end{aligned} \tag{1.25}$$

[7]Toto mechanické zařízení je známé jako *Varignon frame* a v minulosti se opravdu používalo na řešení úlohy. Úloha má bohatou historii, je známa také jako Fermatův-Weberův problém.

[8]Tento pojem patří do *teorie složitosti algoritmů*, kterou jste ještě nebrali. Zde jen řekneme, že pro NP-těžkou úlohu je jen malá naděje, že bude někdy nalezen algoritmus, který by úlohu řešil v polynomiálním čase. Algoritmus řeší úlohu v *polynomiálním čase*, jestliže existuje polynom p takový, že pro každý vstup najde řešení v čase menším než $p(L)$, kde L je počet bitů potřebných k zápisu vstupu.

[9]Úloha byla intenzivně studována nejen v kombinatorické optimalizaci, ale také ve statistické fyzice pod názvem hledání *minimální energie Isingova modelu*.

Účelová funkce je kvadratická (dokonce bilineární) a máme $|V|$ kvadratických omezení typu rovnosti. Každé omezení $x_i^2 = 1$ je ekvivalentní $x_i \in \{-1, 1\}$, tedy[10] množina přípustných řešení má konečný počet ($2^{|V|}$) prvků a jedná se tedy o kombinatorickou úlohu. Jedná se o klasickou NP-těžkou úlohu, proto už pro některé poměrně malé úlohy je prakticky nemožné najít globální optimum. ♦

1.4 Cvičení

1.1. Vyřešte následující úlohy, přičemž slovní úlohy nejdříve formulujte ve tvaru (1.11). Stačí vám k tomu papír, tužka, zdravý rozum a analýza funkcí jedné proměnné. Všimněte si, že některé úlohy lze převést na hledání extrémů funkce jedné proměnné na intervalu, což umíte z analýzy funkcí jedné proměnné.

a) $\min\{\, x^2 + y^2 \mid x \geq 0,\ xy \geq 1 \,\}$

b) $\min\{\, (x-2)^2 + (y - \frac{1}{2})^2 \mid x^2 \leq 1,\ y^2 \leq 1 \,\}$

c) $\min\{\, x \mid x \in \mathbb{R},\ x \geq a_i\ \forall i = 1, \ldots, n \,\}$ pro daná $a_1, \ldots, a_n \in \mathbb{R}$

d) Máte vyrobit papírovou krabici o objemu 72 litrů, jejíž délka je dvojnásobek její šířky. Krabice má všech šest stěn. Jaké budou její rozměry, má-li se na ni spotřebovat co nejméně papíru? Tloušťka stěn je zanedbatelná.

e) Jaké má rozměry válec s jednotkovým objemem a nejmenším povrchem?

f) Najděte rozměry půllitru (tvaru válce), na jehož výrobu je třeba co nejméně skla. Tloušťka stěn je zanedbatelná.

g) Najděte obsah největšího obdélníka vepsaného do kružnice s poloměrem 1.

h) Obdélník v rovině má jeden roh v počátku a druhý na křivce $y = x^2 + x^{-2}$, přičemž jeho strany jsou rovnoběžné se souřadnicovými osami. Pro jaké x bude jeho obsah minimální? Může být jeho obsah libovolně veliký?

i) Najděte bod v rovině na parabole s rovnicí $y = x^2$ nejblíže bodu $(3, 0)$.

j) Hektarová oblast obdélníkového tvaru se má obehnat ze tří stran živým plotem, který stojí 1000 korun na metr, a ze zbývající strany obyčejným plotem, který stojí 500 korun na metr. Jaké budou rozměry oblasti při nejmenší ceně plotu?

k) x, y jsou čísla v intervalu $[1, 5]$ taková, že jejich součet je 6. Najděte tato čísla tak, aby xy^2 bylo (a) co nejmenší, (b) co největší.

l) Hledá se n-tice čísel $x_1, \ldots, x_n \in \{-1, 1\}$ tak, že jejich součin je kladný a jejich součet minimální. Jako výsledek napište (co nejjednodušší) vzorec, který udává hodnotu tohoto minimálního součtu pro obecné n.

m) *Potkaní biatlon.* Potkan stojí na břehu kruhového jezírka o poloměru 1 a potřebuje se dostat na protilehlý bod břehu. Potkan plave rychlostí v_1 a běží rychlostí v_2. Chce se do cíle dostat co nejrychleji, přičemž může běžet, plavat, nebo zvolit kombinaci obojího. Jakou dráhu zvolí? Strategie potkana může být různá pro různé hodnoty v_1 a v_2, vyřešte pro všechny kombinace těchto hodnot.

n) Normální rozdělení se střední hodnotou μ a směrodatnou odchylkou σ má hustotu pravděpodobnosti $p_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. Chceme odhadnout parametry μ a σ z

[10] Všimněte si, že kdybychom omezení napsali jako $x_i \in \{-1, 1\}$, tak by úloha nebyla ve tvaru (1.11). Když ho ovšem napíšeme jako $x_i^2 = 1$, tak úloha ve tvaru (1.11) je.

i.i.d. (*independent, identically distributed*) vzorku $x_1, \dots, x_n$ na základě principu maximální věrohodnosti, tedy chceme maximalizovat věrohodnost $\prod_{i=1}^n p_{\mu,\sigma}(x_i)$.

o) ($\star\star$) Do mezikruží (množinový rozdíl dvou soustředných kruhů o poloměrech $R > r > 0$) máme vepsat elipsu s co největším obsahem. Celá elipsa (hranice i vnitřek) musí ležet v mezikruží. (Motivací úlohy byl návrh pracovního prostoru robotu.)

1.2. Nechť X je libovolná množina a $f\colon X \to \mathbb{R}$. Nechť $g\colon \mathbb{R} \to \mathbb{R}$ je rostoucí funkce. Dokažte, že $\operatorname*{argmin}_{x\in X} f(x) = \operatorname*{argmin}_{x\in X} g(f(x))$.

1.3. ($\star$) Nechť X je libovolná množina, $Y \subseteq X$, a $f\colon X \to \mathbb{R}$. Najděte co nejobecnější podmínku, za které platí $\operatorname*{argmin}_{x\in Y} f(x) = Y \cap \operatorname*{argmin}_{x\in X} f(x)$.

1.4. Dokažte, že pro libovolné funkce $f_1, \dots, f_m\colon \mathbb{R}^n \to \mathbb{R}$ platí

$$\min_{\mathbf{x}_1,\dots,\mathbf{x}_m\in\mathbb{R}^n} [f_1(\mathbf{x}_1) + \cdots + f_m(\mathbf{x}_m)] = \min_{\mathbf{x}\in\mathbb{R}^n} f_1(\mathbf{x}) + \cdots + \min_{\mathbf{x}\in\mathbb{R}^n} f_m(\mathbf{x}), \tag{1.26}$$

tedy minimalizace funkce $f(\mathbf{x}_1, \dots, \mathbf{x}_m) = f_1(\mathbf{x}_1) + \cdots + f_m(\mathbf{x}_m)$ se rozpadne na m nezávislých minimalizací.

($\star$) Platí tvrzení i tehdy, když nahradíme součet maximem nebo minimem? Obecněji, pro jaké funkce $g\colon \mathbb{R}^m \to \mathbb{R}$ platí

$$\min_{\mathbf{x}_1,\dots,\mathbf{x}_m\in\mathbb{R}^n} g(f_1(\mathbf{x}_1), \dots, f_m(\mathbf{x}_m)) = g(\min_{\mathbf{x}\in\mathbb{R}^n} f_1(\mathbf{x}), \dots, \min_{\mathbf{x}\in\mathbb{R}^n} f_m(\mathbf{x}))\,? \tag{1.27}$$

Nápověda a řešení

1.1.a) Vyřešíme úvahou s pomocí obrázku. Množina přípustných řešení je kladná větev hyperboly $xy = 1$ a oblast nad ní. Účelová funkce $f(x, y) = x^2 + y^2$ je čtverec (= druhá mocnina) vzdálenosti bodu (x, y) od počátku, její vrstevnice jsou kružnice se středem v počátku $(0, 0)$. Protože odmocnina je rostoucí funkce, úloha (přesněji, její množina argumentů minima) by se nezměnila (viz Cvičení 1.2), kdybychom účelovou funkce změnili na $\sqrt{x^2 + y^2}$, což je vzdálenost od počátku. Optimální hodnota se proto nabývá v bodě hyperboly nejblíž počátku, tedy v bodě $(x, y) = (1, 1)$.

1.1.b) Vyřešíme opět obrázkem. Protože $x^2 \le 1$ je ekvivalentní $-1 \le x \le 1$, množina přípustných řešení je čtverec (i s vnitřkem) se středem v počátku a stranou délky 2. Účelová funkce je čtverec vzdálenosti od bodu $(2, \frac{1}{2})$. Minimum se nabývá v bodě čtverce nejblíže bodu $(2, \frac{1}{2})$, tj. v bodě $(1, \frac{1}{2})$

1.1.c) Optimální řešení je nejmenší číslo x, které není menší než žádné z čísel $a_1, \dots, a_n$. Očividně, takové číslo je $x = \max_i a_i$.

1.1.d) Krabice je kvádr se stranami $x, 2x, y$ (šířka, délka, výška v decimetrech). Minimalizujeme jeho povrch $2(2x^2 + xy + 2xy) = 2x(2x + 3y)$ za podmínek $2x^2 y = 72$ a $x, y \ge 0$. Z této úlohy eliminujeme proměnnou y. První podmínka je ekvivalentní $y = 36/x^2$. Všimneme si, že podmínka $y = 36/x^2 \ge 0$ je automaticky splněna pro $x > 0$. Po dosazení do kriteria tedy minimalizujeme $2x(2x + 108/x^2) = 4x^2 + 216/x$ za podmínky $x > 0$. To spadá do analýzy funkcí jedné proměnné: položení derivace rovné nule dá stacionární bod $x = \sqrt[3]{27} = 3$, což splňuje $x \ge 0$. Mohli bychom ověřit, např. pomocí druhé derivace nebo úvahou, že je to globální minimum na intervalu $(0, \infty)$.

1.1.h) Nakreslete si obrázek. Minimalizujeme obsah xy obdélníka za podmínky $y = x^2 + x^{-2}$ a $x > 0$ přes proměnné $x, y \in \mathbb{R}$ (podmínka $x > 0$ vynutí kladný obsah obdélníka; všimněte si, že řešení pro $-x$ by bylo symetrické a tedy ho nemusíme uvažovat). Dosazením za y převedeme na minimalizaci funkce jedné proměnné $x(x^2 + x^{-2}) = x^3 + 1/x$ na intervalu $(0, \infty)$).

1.1.i) Minimalizujeme čtverec vzdálenosti $(x-3)^2 + y^2$ za podmínky $y = x^2$ přes proměnné $x, y \in \mathbb{R}$. Dosazením za y převedeme na minimalizaci funkce $(x-3)^2+x^4$ bez omezení. Podmínka stacionarity je $2(x-3) + 4x^3 = 0$, po vydělení dvěma $2x^3 + x - 3 = 0$. Uhodneme jedno řešení $x = 1$, což odpovídá $y = x^2 = 1$. Z obrázku (nakreslete si) je patrné, že nejbližší bod nemůže mít $x < 0$. Protože funkce $2x^3 + x - 3$ na intervalu $[0, \infty)$ striktně roste (spočtěte její derivaci), nemůže na něm mít více než jeden kořen. Tedy nejbližší bod je $(1, 1)$.

1.1.k) Hledáme extrém funkce xy^2 za podmínek $x+y = 6$ a $x, y \in [1, 5]$. Tuto úlohu převedeme na úlohu s jedinou proměnnou eliminací např. y. Z $x + y = 6$ vyjádříme $y = 6 - x$. Všimneme si (ověřte podrobně!), že podmínka $6 - x \in [1, 5]$ je ekvivalentní $x \in [1, 5]$, což už ale máme. Tedy původní úloha je ekvivalentní úloze s jednou proměnnou, ve které hledáme extrémy funkce $x(6 - x)^2$ na intervalu $x \in [1, 5]$. To známe z analýzy.

1.1.l) Minimalizujeme $x_1 + \cdots + x_n$ za podmínek $x_1 \cdots x_n > 0$ a $x_1, \ldots, x_n \in \{-1, 1\}$. To není úloha ve tvaru (1.11), protože omezení $x_i \in \{-1, 1\}$ nejsou v (1.11) dovolena. To ovšem snadno spravíme, protože je můžeme nahradit ekvivalentními omezeními $x_i^2 = 1$, které už tvar (1.11) dovoluje. Stejně ale jde o netypickou úlohu spojité optimalizace, protože každá proměnná může nabývat jen dvou hodnot, tedy množina přípustných řešení má konečný počet prvků (jedná se tedy spíše o úlohu kombinatorické optimalizace).

Úlohu vyřešíme jednoduchou úvahou. Aby $x_1 \cdots x_n > 0$, zápornou hodnotu smí mít sudý počet proměnných. Chceme-li minimalizovat $x_1+\cdots+x_n$, nastavíme všechny (pro sudé n) příp. všechny kromě jedné (pro liché n) proměnné na -1. Optimální hodnota bude tedy $-n$ (pro n sudé) a $2-n$ (pro n liché).

1.1.m) Optimální dráha potkana bude složena ze dvou úseků: z přímočarého plavání ze startu do nějakého bodu A na hranici jezírka, a z běhu po hranici jezírka z bodu A do cíle. Mohli bychom si myslet, že lepšího času lze dosáhnout kombinací více než jednoho plaveckého a/nebo více než jednoho běžeckého úseku, ale takovou dráhu lze vždycky 'přeskládat' do dráhy sestávající z nejvýše jednoho plaveckého a nejvýše jednoho běžeckého úseku, která bude mít stejný nebo lepší čas (rozmyslete). Tedy úloha se redukuje na nalezení bodu A odpovídajícímu minimálnímu času.

Nechť φ je úhel při středu jezírka v trojúhelníku start – střed jezírka – bod A. Celkový čas potkana je $t(\varphi) = \frac{1}{v_1}\sqrt{2 - 2\cos\varphi} + \frac{1}{v_2}(\pi - \varphi)$, kde první člen je čas plavání ze startu do bodu A (odvoďte např. z kosinové věty!) a druhý člen je čas běhu po obvodu jezírka z bodu A do cíle. Tuto funkci potřebujeme minimalizovat na intervalu $[0, \pi]$.

Je $t'(\varphi) = \frac{1}{v_1}\frac{\sin\varphi}{\sqrt{2-2\cos\varphi}} - \frac{1}{v_2} = \frac{1}{v_1}\sqrt{\frac{1+\cos\varphi}{2}} - \frac{1}{v_2}$, kde jsme použili identitu $\sin\varphi = \sqrt{1 - \cos^2\varphi}$ (která platí na intervalu $[0, \pi]$, na jiném intervalu bychom museli dávat pozor na znaménko odmocniny). Spočítáme i druhou derivaci $t''(\varphi) = -\frac{1}{v_1}\frac{\sin\varphi}{2\sqrt{2+2\cos\varphi}}$ a vidíme, že je na intervalu $(0, \pi)$ záporná, tedy účelová funkce na tomto intervalu nemůže nabývat lokálního minima. Tedy účelová funkce může nabývat minima pouze v jednom z hraničních bodů intervalu. Jinými slovy, aby dosáhl minimálního času, musí potkan buď celou dobu plavat ($\varphi = 0$) nebo celou dobu běžet ($\varphi = \pi$). Máme $t(0) = \frac{\pi}{v_2}$ a $t(\pi) = \frac{2}{v_1}$, tedy potkan poplave jestliže $\frac{v_1}{v_2} \geq \frac{\pi}{2}$, jinak poběží.

1.1.n) Maximalizujeme funkci $\prod_{i=1}^n p_{\mu,\sigma}(x_i) = \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$ přes proměnné $\mu \in \mathbb{R}$ a $\sigma > 0$. Je pohodlné místo toho minimalizovat záporný logaritmus této funkce (což je stejné, neboť záporný logaritmus je striktně klesající funkce, srov. Cvičení 1.2). Navíc faktory $\frac{1}{\sqrt{2\pi}}$ můžeme vynechat bez změny úlohy. Tedy minimalizujeme funkci $\sum_{i=1}^n (\ln\sigma + \frac{(x_i-\mu)^2}{2\sigma^2}) = n\ln\sigma + \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu)^2$. Položením parciální derivace podle μ rovné nule dostaneme $-\frac{1}{2\sigma^2}\sum_{i=1}^n 2(x_i - \mu) = 0$, z toho $\mu = \frac{1}{n}\sum_{i=1}^n x_i$. Tedy odhad μ je aritmetický průměr čísel $x_1, \ldots, x_n$. To se dalo čekat.

Položením parciální derivace podle σ rovné nule dostaneme $\frac{n}{\sigma} - \frac{1}{\sigma^3}\sum_{i=1}^n (x_i - \mu)^2 = 0$, tedy $\sigma^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \mu)^2$. Dosadíme výše spočtené $\mu = \frac{1}{n}\sum_{i=1}^n x_i$ a po úpravě (proveďte sami a nespleťte to!) dostaneme $\sigma^2 = \frac{1}{n}\sum_{i=1}^n x_i^2 - (\frac{1}{n}\sum_{i=1}^n x_i)^2$.

1.2. Úkolem je dokázat, že pro každé $x^* \in X$ platí následující ekvivalence: pro každé $x \in X$ platí $f(x^*) \le f(x)$ právě tehdy, když pro každé $x \in X$ platí $g(f(x^*)) \le g(f(x))$. To je ale zjevné, protože pro každou rostoucí funkci $g\colon \mathbb{R} \to \mathbb{R}$ a každé $a, b \in \mathbb{R}$ platí $a \le b \Leftrightarrow g(a) \le g(b)$.

Část I

Použití lineární algebry v optimalizaci

Kapitola 2

Maticová algebra

Cílem této kapitoly je zopakovat a prohloubit si schopnost manipulovat s výrazy obsahujícími matice a vektory, aniž bychom zatím používali abstraktnější pojmy z lineární algebry jako např. lineární nezávislost.

Reálná **matice** rozměru $m\times n$, kde $m, n \in \mathbb{N}$, je tabulka reálných čísel s m řádky a n sloupci[1],

$$\mathbf{A} = [a_{ij}] = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix},$$

kde a_{ij} jsou **prvky** matice[2]. Množinu všech reálných matic rozměru $m \times n$ značíme $\mathbb{R}^{m\times n}$.

Používají se tyto názvy:

- Pro $m = n$ se matice nazývá **čtvercová** a pro $m \neq n$ **obdélníková**, přičemž pro $m < n$ je **široká** a pro $m > n$ je **úzká**.
- **Diagonální prvky** matice jsou prvky a_{ii} pro $i = 1, \dots, p$ kde $p = \min\{m, n\}$. Matice je **diagonální**, když všechny nediagonální prvky jsou nulové, tedy $a_{ij} = 0$ pro všechna $i \neq j$. Všimněte si, že diagonální matice nemusí být čtvercová. Čtvercová ($m = n$) diagonální matice s diagonálními prvky $d_1, \dots, d_n$ se se značí $\operatorname{diag}(d_1, \dots, d_n)$. Např.

$$\operatorname{diag}(1, 3, -2) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -2 \end{bmatrix}.$$

- **Nulová matice** má všechny prvky nulové. Značíme ji $\mathbf{0}_{m,n}$ (pokud jsou rozměry m, n jasné z kontextu, pak pouze $\mathbf{0}$).
- **Jednotková matice** je čtvercová diagonální, jejíž diagonální prvky jsou jedničky. Značíme ji $\mathbf{I}_n$ (pokud je rozměr n jasný z kontextu, pak pouze $\mathbf{I}$). Prvky jednotkové matice můžeme označit symbolem 'Kroneckerovo delta',

$$\delta_{ij} = \begin{cases} 0 & \text{když } i \neq j, \\ 1 & \text{když } i = j. \end{cases} \tag{2.1}$$

[1] Formálněji můžeme matici definovat jako zobrazení $\{1, \dots, m\} \times \{1, \dots, n\} \to \mathbb{R}$, podobně jako vektor $(x_1, \dots, x_n) \in \mathbb{R}^n$ lze považovat za zobrazení $\{1, \dots, n\} \to \mathbb{R}$.

[2] Někdo značí prvky matice $\mathbf{A}$ jako A_{ij}, což má své výhody, my se ale budeme držet častějšího a_{ij}.

- **Horní [dolní] trojúhelníková matice** má $a_{ij} = 0$ pro všechna $i > j$ [$i < j$]. Všimněte si, že horní/dolní trojúhelníková matice nemusí být čtvercová.
- **Permutační matice** je čtvercová matice, která má v každém sloupci právě jednu jedničku, v každém řádku právě jednu jedničku, a ostatní jsou nuly.

2.1 Operace s maticemi

2.1.1 Sčítání a násobení skalárem

Matice lze násobit reálným číslem a sčítat po prvcích:

- Součin čísla $\alpha \in \mathbb{R}$ a matice $\mathbf{A} \in \mathbb{R}^{m\times n}$ je matice $\alpha\mathbf{A} = \mathbf{A}\alpha = [\alpha a_{ij}] \in \mathbb{R}^{m\times n}$.

  Součin $\dfrac{1}{\alpha}\mathbf{A}$ můžeme psát krátce jako $\dfrac{\mathbf{A}}{\alpha}$ nebo $\mathbf{A}/\alpha$. Součin $(-1)\mathbf{A}$ píšeme krátce jako $-\mathbf{A}$.
- Součet matic $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times n}$ je matice $\mathbf{A} + \mathbf{B} = [a_{ij} + b_{ij}] \in \mathbb{R}^{m\times n}$.

  Rozdíl matic je $\mathbf{A} - \mathbf{B} = \mathbf{A} + (-\mathbf{B})$.

Pro pevná $m, n \in \mathbb{N}$ tvoří množina $\mathbb{R}^{m\times n}$ vybavená těmito dvěma operacemi lineární prostor nad tělesem $\mathbb{R}$. Prvky tohoto tělesa, tedy čísla $\alpha \in \mathbb{R}$, se nazývají **skaláry**. Zdůrazněme, že slovo *skalár* nemá význam samostatně, ale vždy jen v kontextu nějakého lineárního prostoru.

2.1.2 Maticový součin

Maticový součin matic $\mathbf{A} \in \mathbb{R}^{m\times p}$ a $\mathbf{B} \in \mathbb{R}^{p\times n}$ je matice $\mathbf{AB} = \mathbf{C} \in \mathbb{R}^{m\times n}$ s prvky

$$c_{ij} = \sum_{k=1}^{p} a_{ik} b_{kj}, \qquad i = 1, \dots, m,\ j = 1, \dots, n. \tag{2.2}$$

Všimněte si, že násobit lze jen matice, které mají vnitřní rozměr (p) stejný. Vlastnosti maticového součinu:

- $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$ (asociativita)
- $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$ a $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$ (distributivita se sčítáním)
- $\mathbf{AI} = \mathbf{A} = \mathbf{IA}$ (neutralita jednotkové matice)
- $(\alpha\mathbf{A})\mathbf{B} = \mathbf{A}(\alpha\mathbf{B}) = \alpha(\mathbf{AB})$

Obecně neplatí $\mathbf{AB} = \mathbf{BA}$ (maticový součin není komutativní)!

Zdůrazněme, že skalár nelze považovat za ‘matici’ rozměru 1×1, i když to k tomu svádí. Výraz $\alpha\mathbf{A}$ není maticový součin dvou matic už proto, že vnitřní rozměr matic by byl obecně různý. Násobení matice skalárem je zkrátka jiná operace než maticový součin.

Pro čtvercovou matici $\mathbf{A}$ se maticový součin

$$\mathbf{A}^k = \underbrace{\mathbf{AA}\cdots\mathbf{A}}_{k\text{-krát}} \tag{2.3}$$

nazývá k-tá **mocnina** matice. Definujeme $\mathbf{A}^0 = \mathbf{I}$. Odlišujte od kartézského součinu *množiny* A se sebou k-krát, který také značíme A^k (viz §1.1.1) ovšem množinu píšeme netučným fontem.

2.1.3 Transpozice

Transpozici matice $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{m\times n}$ značíme $\mathbf{A}^T = [a_{ji}] \in \mathbb{R}^{n\times m}$. Vlastnosti transpozice:

- $(\alpha\mathbf{A})^T = \alpha\mathbf{A}^T$
- $(\mathbf{A}^T)^T = \mathbf{A}$
- $(\mathbf{A}+\mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$
- $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$

Čtvercová matice se nazývá

- **symetrická**, když $\mathbf{A}^T = \mathbf{A}$, tj. $a_{ij} = a_{ji}$,
- **antisymetrická**, když $\mathbf{A}^T = -\mathbf{A}$, tj. $a_{ij} = -a_{ji}$ (z čehož plyne $a_{ii} = 0$).

2.1.4 Inverze

Když pro matice $\mathbf{A} \in \mathbb{R}^{n\times m}$ a $\mathbf{B} \in \mathbb{R}^{m\times n}$ platí

$$\mathbf{AB} = \mathbf{I}_n, \tag{2.4}$$

matice $\mathbf{B}$ se nazývá **pravá inverze** matice $\mathbf{A}$ a matice $\mathbf{A}$ se nazývá **levá inverze** matice $\mathbf{B}$. Matice může mít obě tyto inverze, nebo jen jednu z nich, nebo nemusí mít ani jednu. Když má matice pravou nebo levou inverzi, tato nemusí být jediná. Následující tvrzení budeme schopni dokázat[3] až v §3.2.3, zatím si je pamatujte:

- Pravá inverze matice existuje, právě když má matice lineárně nezávislé řádky. Speciálně, úzká matice nemůže mít pravou inverzi.
- Levá inverze matice existuje, právě když má matice lineárně nezávislé sloupce. Speciálně, široká matice nemůže mít levou inverzi.

Má-li matice pravou i levou inverzi, pak se obě inverze rovnají a jsou jediné. Opravdu, z $\mathbf{AB} = \mathbf{I}$ a $\mathbf{CA} = \mathbf{I}$ plyne $\mathbf{C} = \mathbf{CAB} = \mathbf{B}$. Pak mluvíme pouze o **inverzi** matice a značíme ji $\mathbf{A}^{-1}$. Z tvrzení výše navíc plyne, že $\mathbf{A}$ i $\mathbf{A}^{-1}$ jsou čtvercové. Matice, která má inverzi, se nazývá **invertovatelná** nebo **regulární**. Čtvercová matice, která nemá inverzi, se nazývá **singulární**. Vlastnosti inverze:

- $\mathbf{AA}^{-1} = \mathbf{I} = \mathbf{A}^{-1}\mathbf{A}$
- $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$
- $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$
- $(\alpha\mathbf{A})^{-1} = \alpha^{-1}\mathbf{A}^{-1}$
- $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$, což krátce značíme $\mathbf{A}^{-T}$.

2.1.5 Determinant

Determinant čtvercové matice $\mathbf{A} \in \mathbb{R}^{n\times n}$ se definuje vzorcem

$$\det \mathbf{A} = \sum_\sigma \operatorname{sgn}\sigma \prod_{i=1}^n a_{i\,\sigma(i)}, \tag{2.5}$$

[3] Je zajímavé, že pouze z definice (2.4) (tedy bez použití pojmů jako lineární nezávislost, hodnost nebo lineární zobrazení) nejspíš nelze dokázat, že úzká matice nemůže mít pravou inverzi. Schválně, zkuste najít důkaz!

kde sčítáme přes všechny permutace σ prvků $1, \dots, n$ (tj. bijekce $\sigma\colon \{1, \dots, n\} \to \{1, \dots, n\}$), přičemž $\operatorname{sgn} \sigma$ označuje znaménko permutace. Některé vlastnosti determinantu:

- $\det \mathbf{I} = 1$
- $\det(\mathbf{AB}) = (\det \mathbf{A})(\det \mathbf{B})$
- $\det \mathbf{A}^{-1} = (\det \mathbf{A})^{-1}$ (plyne z předchozího pro $\mathbf{B} = \mathbf{A}^{-1}$)
- $\det \mathbf{A}^T = \det \mathbf{A}$
- $\det \mathbf{A} = 0$ právě tehdy, když $\mathbf{A}$ je singulární
- Determinant je multilineární funkce sloupců matice, tj. je lineární funkcí libovolného sloupce, jsou-li všechny ostatní sloupce konstantní.
- Determinant je alternující funkce sloupců matice, tj. prohození libovolných dvou sloupců změní znaménko determinantu.

2.1.6 Stopa

Stopa (angl. *trace*) čtvercové matice $\mathbf{A} \in \mathbb{R}^{n\times n}$ je součet jejích diagonálních prvků, značí se

$$\operatorname{tr} \mathbf{A} = a_{11} + \cdots + a_{nn}. \tag{2.6}$$

Vlastnosti (dokažte!):

- $\operatorname{tr}(\mathbf{A} + \mathbf{B}) = \operatorname{tr} \mathbf{A} + \operatorname{tr} \mathbf{B}$
- $\operatorname{tr}(\alpha\mathbf{A}) = \alpha \operatorname{tr} \mathbf{A}$
- $\operatorname{tr}(\mathbf{A}^T) = \operatorname{tr} \mathbf{A}$
- $\operatorname{tr}(\mathbf{AB}) = \operatorname{tr}(\mathbf{BA})$ pro každé $\mathbf{A} \in \mathbb{R}^{m\times n}$ a $\mathbf{B} \in \mathbb{R}^{n\times m}$

Z poslední rovnosti plyne např. $\operatorname{tr}(\mathbf{ABC}) = \operatorname{tr}(\mathbf{CAB})$ (pokud matice mají kompatibilní rozměry pro násobení), protože $\operatorname{tr}(\mathbf{DC}) = \operatorname{tr}(\mathbf{CD})$, kde $\mathbf{D} = \mathbf{AB}$. Podobně $\operatorname{tr}(\mathbf{ABCD}) = \operatorname{tr}(\mathbf{DABC}) = \operatorname{tr}(\mathbf{CDAB}) = \operatorname{tr}(\mathbf{BCDA})$. Vidíme, že stopa součinu několika matic je rovna stopě součinu jejich cyklické permutace (tzv. *cykličnost stopy*). Obecně ale neplatí např. $\operatorname{tr}(\mathbf{ABC}) = \operatorname{tr}(\mathbf{CBA})$.

2.1.7 Standardní skalární součin a norma

Standardní **skalární součin matic** $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m\times n}$ se definuje jako

$$\langle \mathbf{A}, \mathbf{B} \rangle = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ij}. \tag{2.7}$$

Standardní skalární součin lze počítat pomocí stopy a naopak:

$$\langle \mathbf{A}, \mathbf{B} \rangle = \operatorname{tr}(\mathbf{A}^T\mathbf{B}) \tag{2.8}$$

$$\operatorname{tr} \mathbf{A} = \langle \mathbf{I}, \mathbf{A} \rangle \tag{2.9}$$

(pro důkaz (2.8) si napište diagonální prvky matice $\mathbf{A}^T\mathbf{B}$).

Z definice (2.7) snadno dokážeme následující vlastnosti skalárního součinu:

- $\langle \mathbf{A}, \mathbf{B} \rangle = \langle \mathbf{B}, \mathbf{A} \rangle$ (komutativita)
- $\langle \mathbf{A}, \mathbf{B} \rangle = \langle \mathbf{A}^T, \mathbf{B}^T \rangle$

- $\langle \alpha\mathbf{A}, \mathbf{B}\rangle = \alpha\langle \mathbf{A}, \mathbf{B}\rangle = \langle \mathbf{A}, \alpha\mathbf{B}\rangle$
- $\langle \mathbf{A}+\mathbf{B}, \mathbf{C}\rangle = \langle \mathbf{A}, \mathbf{C}\rangle + \langle \mathbf{B}, \mathbf{C}\rangle$ a $\langle \mathbf{A}, \mathbf{B}+\mathbf{C}\rangle = \langle \mathbf{A}, \mathbf{B}\rangle + \langle \mathbf{A}, \mathbf{C}\rangle$

Méně očividné je chování skalárního součinu vzhledem k maticovému součinu[4]:

$$\langle \mathbf{AB}, \mathbf{C}\rangle = \langle \mathbf{B}, \mathbf{A}^T\mathbf{C}\rangle = \langle \mathbf{A}, \mathbf{CB}^T\rangle. \tag{2.10}$$

První identitu dokážeme např. takto:

$$\langle \mathbf{AB}, \mathbf{C}\rangle = \langle \mathbf{C}, \mathbf{AB}\rangle = \operatorname{tr}(\mathbf{C}^T\mathbf{AB}) = \operatorname{tr}((\mathbf{A}^T\mathbf{C})^T\mathbf{B}) = \langle \mathbf{A}^T\mathbf{C}, \mathbf{B}\rangle = \langle \mathbf{B}, \mathbf{A}^T\mathbf{C}\rangle.$$

Druhou identitu dokážeme podobně (proveďte!).

Skalární součin (2.7) indukuje **normu matice** $\mathbf{A} \in \mathbb{R}^{m\times n}$ jako

$$\|\mathbf{A}\| = \sqrt{\langle \mathbf{A}, \mathbf{A}\rangle} = \sqrt{\operatorname{tr}(\mathbf{A}^T\mathbf{A})} = \sqrt{\operatorname{tr}(\mathbf{A}\mathbf{A}^T)} = \left(\sum_{i=1}^m\sum_{j=1}^n a_{ij}^2\right)^{1/2}. \tag{2.11}$$

Existují i jiné normy matic (které neuvádíme). Norma (2.11) se nazývá **Frobeniova norma** a pokud ji chceme odlišit od jiných norem matic, značíme ji $\|\cdot\|_{\mathrm{F}}$.

2.2 Matice s jedním sloupcem nebo jedním řádkem

Matici s jediným sloupcem (tedy prvku $\mathbb{R}^{n\times 1}$) se také říká **sloupcový vektor** a matici s jediným řádkem (prvku $\mathbb{R}^{1\times n}$) **řádkový vektor**. Lineární prostor $\mathbb{R}^{n\times 1}$ matic s jediným sloupcem je 'skoro stejný' (je s ním *isomorfní*) jako lineární prostor $\mathbb{R}^n$ uspořádaných n-tic $(x_1,\ldots,x_n)$. Proto je zvykem tyto dva prostory ztotožnit a bez upozornění přecházet mezi oběma významy. Prvkům

$$\mathbf{x} = \underbrace{(x_1,\ldots,x_n)}_{\text{uspořádaná } n\text{-tice}} = \underbrace{\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}}_{\text{matice } n\times 1} \in \mathbb{R}^n = \mathbb{R}^{n\times 1}$$

tohoto prostoru budeme říkat krátce **vektory**. Slovem *vektor* (bez přívlastku) tedy rozumíme *sloupcový vektor* (tj. matici s jedním sloupcem) nebo uspořádanou n-tici čísel. Zopakujme, že uspořádané n-tice (prvky kartézského součinu) značíme *kulatými* závorkami (viz §1.1.3), kdežto matice značíme *hranatými* závorkami.

Totéž bychom samozřejmě mohli udělat s řádky (tj. ztotožnit prostory $\mathbb{R}^n$ a $\mathbb{R}^{1\times n}$), ale převládající konvence je dělat to se sloupci[5]. Chceme-li tedy napsat řádkový vektor, můžeme napsat $\mathbf{x} \in \mathbb{R}^{1\times n}$, ale to se většinou nedělá, neboť malá tučná písmena obvykle označují prvky $\mathbb{R}^n$. Řádkový vektor proto raději píšeme jako $\mathbf{x}^T$ kde $\mathbf{x} \in \mathbb{R}^n$, z čehož plyne $\mathbf{x}^T \in \mathbb{R}^{1\times n}$.

Popsané ztotožnění je důležité, protože nyní se k vektorům $\mathbf{x} \in \mathbb{R}^n$ můžeme chovat jako k maticím, tj. můžeme je maticově násobit jinými maticemi, transponovat, apod. Uveďme důležité případy, kdy se v maticovém součinu vyskytují vektory:

[4]V obecném lineárním prostoru se skalárním součinem vlastnost $\langle f(x), y\rangle = \langle x, f^*(y)\rangle$ definuje *adjungované* lineární zobrazení f^* k lineárnímu zobrazení f. V našem jednoduchém případě, kdy máme zobrazení $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x}$ a standardní skalární součin, je $\mathbf{f}^*(\mathbf{y}) = \mathbf{A}^T\mathbf{y}$, tedy adjunkce splývá s maticovou transpozicí.

[5]Výjimkou jsou např. počítačoví grafici, pro které 'vektor' bez přívlastku znamená řádkový vektor.

- Pro $\mathbf{A} \in \mathbb{R}^{m\times n}$ a $\mathbf{x} \in \mathbb{R}^n$ je výraz $\mathbf{A}\mathbf{x}$ maticový součin matice $m \times n$ a matice $n \times 1$, což je (sloupcový) vektor délky m. Je to vlastně lineární kombinace sloupců matice $\mathbf{A}$ s koeficienty $\mathbf{x}$ (viz §3.2).
- Pro $\mathbf{A} \in \mathbb{R}^{m\times n}$ a $\mathbf{x} \in \mathbb{R}^m$ je výraz $\mathbf{x}^T\mathbf{A}$ maticový součin matice $1 \times m$ a matice $m \times n$, což je řádkový vektor délky n. Je to vlastně lineární kombinace řádků matice $\mathbf{A}$ s koeficienty $\mathbf{x}$.
- Pro $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ je výraz
$$\mathbf{x}^T\mathbf{y} = x_1y_1 + \cdots + x_ny_n = \langle \mathbf{x}, \mathbf{y} \rangle \tag{2.12}$$
maticový součin řádkového vektoru $\mathbf{x}^T$ a sloupcového vektoru $\mathbf{y}$. Výsledek je skalár, známý jako **standardní skalární součin vektorů x a y**. Druhá rovnost říká (ověřte!), že je to speciální případ skalárního součinu matic (2.7), ve kterém považujeme $\mathbf{x}, \mathbf{y}$ za matice s jedním sloupcem. Pro standardní skalární součin vektorů budeme v dalším textu vždy používat značení $\mathbf{x}^T\mathbf{y}$, tj. nebudeme používat $\langle \mathbf{x}, \mathbf{y} \rangle$.
- Pro $\mathbf{x} \in \mathbb{R}^m$ a $\mathbf{y} \in \mathbb{R}^n$ je výraz
$$\mathbf{x}\mathbf{y}^T = \begin{bmatrix} x_1y_1 & \cdots & x_1y_n \\ \vdots & \ddots & \vdots \\ x_my_1 & \cdots & x_my_n \end{bmatrix} \in \mathbb{R}^{m\times n} \tag{2.13}$$
matice $m \times n$, které se někdy říká **vnější součin** vektorů $\mathbf{x}$ a $\mathbf{y}$ nebo **dyáda**[6].

Symbol $\mathbf{1}_n = (1, \dots, 1) \in \mathbb{R}^n$ značí (sloupcový) vektor s jedničkovými složkami. Pokud n plyne z kontextu, píšeme jen $\mathbf{1}$. Příklad: pro $\mathbf{x} \in \mathbb{R}^n$ je $\mathbf{1}^T\mathbf{x} = x_1 + \cdots + x_n$. Jiný příklad: jednotkovou matici můžeme značit (i když se to tak nedělá) $\mathbf{I} = \operatorname{diag} \mathbf{1}$.

Symbol $\mathbf{e}_i = (0, \dots, 0, 1, 0, \dots, 0) \in \mathbb{R}^n$ (jednička na i-tém místě) značí i-tý (sloupcový) vektor standardní báze, kde počet n složek vektoru $\mathbf{e}_i$ je určen kontextem. Standardní báze tvoří sloupce jednotkové matice, $\begin{bmatrix} \mathbf{e}_1 & \cdots & \mathbf{e}_n \end{bmatrix} = \mathbf{I}_n$.

2.3 Matice sestavené z bloků

Matici je možno sestavit z několika jejích **podmatic** (zvaných též **bloky**) jako
$$\begin{bmatrix} \mathbf{A}_{11} & \cdots & \mathbf{A}_{1n} \\ \vdots & \ddots & \vdots \\ \mathbf{A}_{m1} & \cdots & \mathbf{A}_{mn} \end{bmatrix} \in \mathbb{R}^{m\times n} \tag{2.14}$$
kde $\mathbf{A}_{ij} \in \mathbb{R}^{m_i\times n_j}$ a $m = \sum_i m_i$ a $n = \sum_j n_j$. Např.
$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}, \quad \begin{bmatrix} \mathbf{A} & \mathbf{B} \end{bmatrix}, \quad \begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix}, \quad \begin{bmatrix} \mathbf{A} & \mathbf{I} \\ \mathbf{0} & \mathbf{D} \end{bmatrix}, \tag{2.15}$$
kde v prvním příkladu musí mít např. matice $\mathbf{A}, \mathbf{B}$ stejný počet řádků a matice $\mathbf{A}, \mathbf{C}$ stejný počet sloupců. V posledním příkladu jsou rozměry jednotkové matice $\mathbf{I}$ a nulové matice $\mathbf{0}$ určeny rozměry matic $\mathbf{A}, \mathbf{D}$. V Matlabu matice (2.15) napíšeme jako

`[A B; C D]`, `[A B]`, `[A;B]`, `[A eye(m); zeros(n) D]` (kde `[m,n]=size(A)`).

[6] Dyáda se dá považovat za tensorový součin dvou vektorů v maticové notaci. Ovšem *tensorový součin* je abstraktnější a obecnější pojem.

Při násobení matic sestavených z bloků lze neformálně užít obvyklý vzorec (2.2) pro násobení matic, ve kterém si místo prvků matice představíme bloky. Formálně,

$$\begin{bmatrix} \mathbf{A}_{11} & \cdots & \mathbf{A}_{1p} \\ \vdots & \ddots & \vdots \\ \mathbf{A}_{m1} & \cdots & \mathbf{A}_{mp} \end{bmatrix} \begin{bmatrix} \mathbf{B}_{11} & \cdots & \mathbf{B}_{1n} \\ \vdots & \ddots & \vdots \\ \mathbf{B}_{p1} & \cdots & \mathbf{B}_{pn} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_{11} & \cdots & \mathbf{C}_{1n} \\ \vdots & \ddots & \vdots \\ \mathbf{C}_{m1} & \cdots & \mathbf{C}_{mn} \end{bmatrix} \tag{2.16}$$

kde

$$\mathbf{C}_{ij} = \sum_{k=1}^{p} \mathbf{A}_{ik}\mathbf{B}_{kj}, \qquad i = 1, \ldots, m, \ j = 1, \ldots, n. \tag{2.17}$$

Příklady:

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} = \begin{bmatrix} \mathbf{AX} + \mathbf{BY} \\ \mathbf{CX} + \mathbf{DY} \end{bmatrix}, \quad \begin{bmatrix} \mathbf{A} & \mathbf{B} \end{bmatrix} \begin{bmatrix} \mathbf{C} \\ \mathbf{D} \end{bmatrix} = \mathbf{AC} + \mathbf{BD}, \quad \begin{bmatrix} \mathbf{A} \\ \mathbf{C} \end{bmatrix} \begin{bmatrix} \mathbf{E} & \mathbf{F} \end{bmatrix} = \begin{bmatrix} \mathbf{AE} & \mathbf{AF} \\ \mathbf{CE} & \mathbf{CF} \end{bmatrix}.$$

Často je užitečné vnímat matici $\mathbf{A} \in \mathbb{R}^{m\times n}$ jako blokovou matici, kde bloky jsou sloupce matice, tj.

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}, \tag{2.18}$$

kde sloupcové vektory $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$ jsou sloupce matice $\mathbf{A}$. Matici lze také vnímat jako sestavenou z řádků, tj.

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_m^T \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_m \end{bmatrix}^T, \tag{2.19}$$

kde řádkové vektory[7] $\mathbf{a}_1^T, \ldots, \mathbf{a}_m^T$ jsou řádky matice $\mathbf{A}$, přičemž $\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^n$.

Vyjádříme-li matici $\mathbf{A} \in \mathbb{R}^{m\times p}$ pomocí sloupců a matici $\mathbf{B} \in \mathbb{R}^{p\times n}$ pomocí řádků, je

$$\mathbf{AB} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_p \end{bmatrix} \begin{bmatrix} \mathbf{b}_1^T \\ \vdots \\ \mathbf{b}_p^T \end{bmatrix} = \mathbf{a}_1\mathbf{b}_1^T + \cdots + \mathbf{a}_p\mathbf{b}_p^T. \tag{2.20}$$

Vyjádříme-li naopak matici $\mathbf{A} \in \mathbb{R}^{m\times p}$ pomocí řádků a matici $\mathbf{B} \in \mathbb{R}^{p\times n}$ pomocí sloupců, je

$$\mathbf{AB} = \begin{bmatrix} \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_m^T \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_n \end{bmatrix} = \begin{bmatrix} \mathbf{Ab}_1 & \cdots & \mathbf{Ab}_n \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1^T\mathbf{B} \\ \vdots \\ \mathbf{a}_m^T\mathbf{B} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1^T\mathbf{b}_1 & \cdots & \mathbf{a}_1^T\mathbf{b}_n \\ \vdots & \ddots & \vdots \\ \mathbf{a}_m^T\mathbf{b}_1 & \cdots & \mathbf{a}_m^T\mathbf{b}_n \end{bmatrix}. \tag{2.21}$$

2.4 Co je soustava lineárních rovnic?

Rovnice s n neznámými je *lineární*, jestliže se v ní neznámé nevyskytují ve vyšší než první mocnině ani v součinech. Neboli pravá i levá strana rovnice je polynom nejvýše prvního stupně. Pojmenujeme-li neznámé jako $x_1, \ldots, x_n$, lineární rovnici lze psát jako

$$a_1x_1 + \cdots + a_nx_n = b$$

[7] Vektory $\mathbf{a}_i$ samozřejmě označují *jiné* vektory v (2.18) a v (2.19).

nebo také vektorově jako $\mathbf{a}^T\mathbf{x} = b$, kde $\mathbf{a} = (a_1, \ldots, a_n)$ a $\mathbf{x} = (x_1, \ldots, x_n)$. Lineární rovnice je **homogenní**, jestliže v ní chybí absolutní člen (tedy $b = 0$), jinak je **nehomogenní**.

Soustavu m lineárních rovnic s n neznámými $x_1, \ldots, x_n$ lze zapsat jako

$$\begin{aligned} a_{11}x_1 \ + \cdots + \ a_{1n}x_n \ &= b_1, \\ &\vdots \\ a_{m1}x_1 \ + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

nebo maticově jako

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \tag{2.22}$$

kde $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{x} \in \mathbb{R}^n$. Soustava je **homogenní**, jsou-li všechny její rovnice homogenní (tj. $\mathbf{b} = \mathbf{0}$) a **nehomnogenní**, je-li aspoň jedna její rovnice nehomogenní (tj. $\mathbf{b} \neq \mathbf{0}$).

Chceme-li řešit soustavu lineárních rovnic na počítači, příslušné algoritmy často vyžadují soustavu ve tvaru (2.22), tj. všechny neznámé jsou soutředěné do jediného vektoru. Např. v Matlabu se řešení soustavy (2.22) spočítá jednoduše jako `x=A\b` (zde předpokládáme, že soustava má právě jedno řešení). Ovšem ne vždy dostaneme lineární soustavu v tomto tvaru. Přepisujeme-li takové soustavy do tvaru (2.22), můžeme snadno udělat chybu. Existují způsoby, jak to dělat elegantněji a obecněji[8], viz následující příklady a Cvičení 2.6.

Příklad 2.1. Soustava

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \tag{2.23a}$$
$$\mathbf{A}^T\mathbf{y} = \mathbf{x}, \tag{2.23b}$$

kde $\mathbf{x} \in \mathbb{R}^n$ a $\mathbf{y} \in \mathbb{R}^m$ jsou neznámé a $\mathbf{A} \in \mathbb{R}^{m\times n}$ a $\mathbf{b} \in \mathbb{R}^m$ jsou dány. Je to lineární soustava $m+n$ rovnic s $m+n$ neznámými. Tuto soustavu můžeme napsat ve tvaru (2.22) (kde samozřejmě písmena $\mathbf{A}, \mathbf{b}$ znamenají něco jiného než v naší soustavě) takto:

$$\begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{I} & -\mathbf{A}^T \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix}.$$

Všimněte si, že soustavu (2.23) lze vyřešit i jinak. Dosadíme $\mathbf{x}$ z druhé rovnice do první, což dá $\mathbf{A}\mathbf{A}^T\mathbf{y} = \mathbf{b}$. Tuto rovnici vyřešíme pro $\mathbf{y}$ a spočítáme $\mathbf{x}$ z druhé rovnice. ♦

Příklad 2.2. Homogenní soustava lineárních rovnic

$$\mathbf{a}_i^T\mathbf{x} = \alpha_i b_i \quad \forall i = 1, \ldots, m$$

kde $\mathbf{x} \in \mathbb{R}^n$ a $\alpha_1, \ldots, \alpha_m$ jsou neznámé a $\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^n$ a $b_1, \ldots, b_m \in \mathbb{R}$ jsou dány. Označíme-li $\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_m \end{bmatrix}^T \in \mathbb{R}^{m\times n}$, $\mathbf{b} = (b_1, \ldots, b_m)$ a $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)$, pak soustavu můžeme psát jako $\mathbf{A}\mathbf{x} = (\operatorname{diag}\mathbf{b})\boldsymbol{\alpha}$, neboli

$$\begin{bmatrix} \mathbf{A} & -\operatorname{diag}\mathbf{b} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \boldsymbol{\alpha} \end{bmatrix} = \mathbf{0}.$$

♦

[8] Abychom řekli přesně, co znamená ‘obecněji’, musíme rozlišovat mezi *instancí* soustavy lineárních rovnic, což je jedna konkrétní soustava, a *nekonečnou množinou instancí* soustav rovnic v určitém tvaru (kterou někdy krátce nazýváme *problém* nebo *úloha*; to jsou pojmy z *teorie výpočetní složitosti*). Např. soustava lineárních rovnic $\{\, x_1 - 2x_2 = 1, \ -x_1 + x_2 = 3 \,\}$ je instance, ale (2.23) je nekonečná množina lineárních soustav, jejíž každý prvek je určen konkrétní dvojicí $(\mathbf{A}, \mathbf{b})$. Převodem úlohy (2.23) do tvaru (2.22) pak rozumíme *algoritmus*, který pro libovolnou dvojici $(\mathbf{A}, \mathbf{b})$ najde dvojici $(\mathbf{A}', \mathbf{b}')$ tak, že soustava (2.23) bude ekvivalentní soustavě $\mathbf{A}'\mathbf{x}' = \mathbf{b}'$ kde $\mathbf{x}' = (\mathbf{x}, \mathbf{y})$.

Příklad 2.3. Soustava

$$\sum_{j=1}^{n} x_{ij} = a_i, \quad i = 1, \ldots, m, \tag{2.24a}$$

$$\sum_{i=1}^{m} x_{ij} = b_j, \quad j = 1, \ldots, n, \tag{2.24b}$$

kde $x_{11}, \ldots, x_{mn} \in \mathbb{R}$ jsou neznámé a $a_1, \ldots, a_m, b_1, \ldots, b_n \in \mathbb{R}$ jsou dány. Je to soustava $m+n$ lineárních rovnic s mn neznámými. Tato soustava jde napsat v maticovém tvaru jako

$$\begin{aligned} \mathbf{X}\mathbf{1} &= \mathbf{a}, \\ \mathbf{X}^T\mathbf{1} &= \mathbf{b}, \end{aligned}$$

kde $\mathbf{a} = (a_1, \ldots, a_m) \in \mathbb{R}^m$, $\mathbf{b} = (b_1, \ldots, b_n) \in \mathbb{R}^n$ a $\mathbf{X} = [x_{ij}] \in \mathbb{R}^{m \times n}$. To ale není tvar (2.22), neboť proměnné jsou opět soustředěné do matice $\mathbf{X}$ a nikoliv do vektoru. Soustavu ale lze přepsat takto (ověřte roznásobením matic!):

$$\begin{bmatrix} \mathbf{I}_m & \cdots & \mathbf{I}_m \\ \mathbf{1}_m^T & & \\ & \ddots & \\ & & \mathbf{1}_m^T \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{bmatrix} = \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \tag{2.25}$$

kde $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^m$ jsou sloupce matice $\mathbf{X}$. ♦

2.5 Elementární maticové úpravy

Z Gaussovy eliminace znáte **elementární řádkové úpravy** matice. Každá elementární řádková úprava matice $\mathbf{A} \in \mathbb{R}^{m \times n}$ odpovídá jejímu vynásobení *zleva* nějakou regulární maticí $\mathbf{E} \in \mathbb{R}^{m \times m}$, která se nazývá **elementární matice**. Jsou tři:

1. Prohození dvou řádků i, j. To odpovídá elementární matici

$$\mathbf{E} = \mathbf{I} - (\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)^T = \begin{array}{c} \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & & 1 & & \\ & & & \ddots & & & \\ & & 1 & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix} \begin{array}{c} \\ \\ i \\ \\ j \\ \\ \\ \end{array} \\ \begin{array}{ccccccc} & & i & & j & & \end{array} \end{array} \tag{2.26a}$$

2. Vynásobení řádku i nenulovým skalárem α. To odpovídá elementární matici

$$\mathbf{E} = \mathbf{I} - (1-\alpha)\mathbf{e}_i\mathbf{e}_i^T = \begin{array}{cc} \begin{bmatrix} 1 &&&&&& \\ & \ddots &&&&& \\ && 1 &&&& \\ &&& \alpha &&& \\ &&&& 1 && \\ &&&&& \ddots & \\ &&&&&& 1 \end{bmatrix} & \begin{matrix} \\ \\ \\ i \\ \\ \\ \\ \end{matrix} \\ \begin{matrix} & & & i & & & \end{matrix} & \end{array} \tag{2.26b}$$

3. Přičtení α-násobku řádku i k řádku j. To odpovídá elementární matici

$$\mathbf{E} = \mathbf{I} + \alpha\mathbf{e}_j\mathbf{e}_i^T = \begin{array}{cc} \begin{bmatrix} 1 &&&&&& \\ & \ddots &&&&& \\ && 1 &&&& \\ &&& \ddots &&& \\ && \alpha && 1 && \\ &&&&& \ddots & \\ &&&&&& 1 \end{bmatrix} & \begin{matrix} \\ \\ i \\ \\ j \\ \\ \\ \end{matrix} \\ \begin{matrix} & & i & & j & & \end{matrix} & \end{array} \tag{2.26c}$$

Ověřte rovnosti ve vzorečcích (2.26), kde $\mathbf{e}_i$ značí i-tý vektor standarní báze prostoru $\mathbb{R}^m$ (viz §2.2) – všimněte si k tomu, že $\mathbf{e}_i\mathbf{e}_j^T \in \mathbb{R}^{m\times m}$ je matice se samými nulami kromě prvku $a_{ij} = 1$. Transpozice a inverze elementární matice je opět elementární matice (viz Cvičení 2.24). Lze dokázat (důkaz neuvádíme), že každá regulární matice je součinem konečné posloupnosti elementárních matic, tj. lze ji získat aplikací konečné posloupnosti elementárních řádkových úprav na jednotkovou matici $\mathbf{I}$.

Zcela analogicky lze definovat **elementární sloupcové úpravy**, které odpovídají násobení (transponovanou) elementární maticí *zprava*.

2.5.1 Gaussova eliminace a LU rozklad

Elementární řádkové úpravy se často používají v algoritmech numerické lineární algebry. Základní takový algoritmus je *Gaussova eliminace* na řešení soustavy lineárních rovnic $\mathbf{Ax} = \mathbf{b}$. Koeficienty soustavy zapíšeme do matice ('tabulky') $\begin{bmatrix}\mathbf{A} & \mathbf{b}\end{bmatrix} \in \mathbb{R}^{m\times(n+1)}$ (tzv. *rozšířená matice soustavy*). Výsledkem elementární úpravy na této matici je matice $\mathbf{E}\begin{bmatrix}\mathbf{A} & \mathbf{b}\end{bmatrix} = \begin{bmatrix}\mathbf{EA} & \mathbf{Eb}\end{bmatrix}$, která odpovídá soustavě $\mathbf{EAx} = \mathbf{Eb}$. Tato soustava má stejnou množinu řešení jako původní soustava, protože matice $\mathbf{E}$ je regulární (tedy vynásobením soustavy inverzní maticí $\mathbf{E}^{-1}$ zleva bychom dostali opět $\mathbf{Ax} = \mathbf{b}$).

V Gaussově eliminaci zvolíme posloupnost řádkových úprav $\mathbf{E}_1, \dots, \mathbf{E}_k$ tak, abychom matici $\mathbf{A}$ postupně převedli na horní trojúhelníkovou matici, tj. naši soustavu převedeme na

$$\underbrace{\mathbf{E}_k\mathbf{E}_{k-1}\cdots\mathbf{E}_2\mathbf{E}_1\mathbf{A}}_{\mathbf{U}}\,\mathbf{x} = \underbrace{\mathbf{E}_k\mathbf{E}_{k-1}\cdots\mathbf{E}_2\mathbf{E}_1\mathbf{b}}_{\mathbf{b}'}, \tag{2.27}$$

kde matice $\mathbf{U}$ je horní trojúhleníková. Tato soustava se pak levně vyřěší zpětnou substitucí.

Tento proces lze zapsat i jinak. Rovnost $\mathbf{E}_k\mathbf{E}_{k-1}\cdots\mathbf{E}_2\mathbf{E}_1\mathbf{A} = \mathbf{U}$ lze (díky regularitě matic $\mathbf{E}_i$) psát jako

$$\mathbf{A} = \mathbf{LU} \tag{2.28}$$

kde $\mathbf{L} = (\mathbf{E}_k\mathbf{E}_{k-1}\cdots\mathbf{E}_2\mathbf{E}_1)^{-1} = \mathbf{E}_1^{-1}\cdots\mathbf{E}_k^{-1}$. Matice $\mathbf{L}$ jde levně spočítat, protože inverze $\mathbf{E}_i^{-1}$ jsou opět elementární matice a tedy odpovídají elementárním řádkovým úpravám. Jestliže při Gaussově eliminaci neprohazujeme řádky (tedy nepoužíváme matice (2.26a)), jsou matice $\mathbf{E}_i$ dolní trojúhelníkové, tedy i matice $\mathbf{L}$ je dolní trojúhelníková (viz Cvičení 2.23). Používáme-li během algoritmu také matice (2.26a), je $\mathbf{L}$ součinem dolní trojúhelníkové matice a permutační matice. Vzorci (2.28) se říká **LU-rozklad** matice $\mathbf{A}$. LU rozklad lze tedy vidět jako 'maticový popis' Gaussovy eliminace. Studujte matlabskou funkci `lu` pomocí příkazu `help lu`!

2.6 Maticové zločiny

Při manipulaci s maticovými výrazy a rovnicemi dělají někteří studenti hrubé chyby, kterých se lze při alespoň minimální snaze vyhnout. Takové chyby jsou *neomluvitelné*. Uveďme typické příklady těchto zločinů.

Výraz je nesmyslný kvůli rozměrům matic

Jako první uveďme chyby, kdy výraz nemá smysl kvůli rozměrům matic a vektorů. Např.:

- Pokud $\mathbf{A} \in \mathbb{R}^{2\times 3}$ a $\mathbf{B} \in \mathbb{R}^{3\times 3}$, tak následující výrazy jsou chybné:

$$\mathbf{A}+\mathbf{B}, \quad \mathbf{A}=\mathbf{B}, \quad [\mathbf{A}\ \mathbf{B}], \quad \mathbf{A}^T\mathbf{B}, \quad \mathbf{A}^{-1}, \quad \det\mathbf{A}, \quad \mathbf{A}^2.$$

- Zcela odstrašující je použití zlomku pro matice, např. $\dfrac{\mathbf{A}}{\mathbf{B}}$. 'Zlomková čára' není pro matice definována. Nebyla by totiž jednoznačná, protože může znamenat buď $\mathbf{AB}^{-1}$ nebo $\mathbf{B}^{-1}\mathbf{A}$.

  Abyste měli jasnou představu o rozměrech matic v maticových výrazech, pod matice a vektory si malujte obdélníčky s rozměry matic!
- Předpoklad, že existuje pravá inverze úzké matice, např. $\mathbf{QQ}^T = \mathbf{I}$ pro $\mathbf{Q} \in \mathbb{R}^{5\times 3}$.
- Inverze čtvercové, ale evidentně singulární matice, např. $(\mathbf{AB})^{-1}$ pro úzkou $\mathbf{A}$. Kdyby totiž inverze existovala, bylo by $\mathbf{AB}(\mathbf{AB})^{-1} = \mathbf{I}$, tedy $\mathbf{A}$ by měla pravou inverzi $\mathbf{B}(\mathbf{AB})^{-1}$. Ale to není možné, protože $\mathbf{A}$ je úzká. Extrémnější příklad stejného zločinu je $(\mathbf{ww}^T)^{-1}$ kde $\mathbf{w}$ je vektor.

Příklad 2.4. Vidíme-li výraz $(\mathbf{A}^T\mathbf{B})^{-1}$, musí nám ve zlomku vteřiny hlavou proběhnout tyto úvahy o rozměrech matic:

- Aby matice bylo možné násobit, musí mít stejný počet řádků.
- Protože nelze invertovat obdélníkovou matici, musí být součin $\mathbf{A}^T\mathbf{B}$ čtvercový. Tedy obě matice musí mít stejný počet sloupců. Teď víme, že obě matice musí mít stejný rozměr.
- Pokud by $\mathbf{A}^T$ byla úzká nebo $\mathbf{B}$ široká, $\mathbf{A}^T\mathbf{B}$ by byla singulární, a tedy by neměla inverzi. Tedy musí být obě matice čtvercové nebo úzké. ♦

Abyste nedělali chyby v rozměrech matic, pod maticové výrazy si pište rozměry matic, např.

$$\underbrace{\underbrace{\underbrace{\mathbf{A}}_{m\times n}{}^T}_{n\times m}\underbrace{\mathbf{B}}_{m\times n}}_{n\times n}.$$

Použití neexistujících maticových identit

Pro manipulaci s maticovými výrazy je užitečné mít v paměti zásobu maticových identit. Ovšem nesmí být chybné. Typické příklady:

- $(\mathbf{AB})^T = \mathbf{A}^T\mathbf{B}^T$ (pokud v maticovém součinu $\mathbf{A}^T\mathbf{B}^T$ je vnitřní rozměr různý, je to chyba už kvůli rozměrům matic)
- $(\mathbf{AB})^{-1} = \mathbf{A}^{-1}\mathbf{B}^{-1}$ (podobná situace)
- $(\mathbf{A}+\mathbf{B})^2 = \mathbf{A}^2 + 2\mathbf{AB} + \mathbf{B}^2$. Tato identita vyplývá z neexistující (avšak velice užitečné) identity $\mathbf{AB} = \mathbf{BA}$. Správně je $(\mathbf{A}+\mathbf{B})^2 = \mathbf{A}^2 + \mathbf{AB} + \mathbf{BA} + \mathbf{B}^2$.

Pracujte nejen s papírem, ale i s Matlabem! Hypotézy o platnosti či neplatnosti maticových rovnic lze často testovat na náhodných maticích. Např. nejste-li si jisti rovností $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$, zkuste `A=randn(5,3); B=randn(3,6); (A*B)'-B'*A'` (což ale samozřejmě není důkaz).

Neekvivalentní úpravy rovnic

Zde pachatel udělá *neekvivalentní úpravu* rovnice, ale myslí si, že udělal ekvivalentní. Ekvivalentní a neekvivalentní úpravy skalárních rovnic známe již ze základní školy. Např. úprava ‘přičti k rovnici jedničku’ je ekvivalentní, neboť $a = b \iff a+1 = b+1$. Úprava ‘umocni rovnici na druhou’ je neekvivalentní, neboť sice $a = b \implies a^2 = b^2$, ale $a^2 = b^2 \not\Rightarrow a = b$. Příklady:

- Student z rovnice $\mathbf{a}^T\mathbf{x} = \mathbf{a}^T\mathbf{y}$ odvodí rovnici $\mathbf{x} = \mathbf{y}$ (neplatí, ani když vektor $\mathbf{a}$ je nenulový).
- Student si myslí, že pokud $\mathbf{A} \in \mathbb{R}^{3\times 5}$ a $\mathbf{AX} = \mathbf{AY}$, pak $\mathbf{X} = \mathbf{Y}$ (není pravda, protože $\mathbf{A}$ nemá lineárně nezávislé sloupce, tedy nemá levou inverzi).
- Student z $\mathbf{A}^T\mathbf{A} = \mathbf{B}^T\mathbf{B}$ odvodí $\mathbf{A} = \mathbf{B}$ (neplatí dokonce ani pro skaláry).

Soustavy rovnic (lineárních či nelineárních) se často snažíme řešit (nebo aspoň zjednodušit) eliminací některých proměnných. To obvykle znamená, že v některé rovnici osamostatníme jednu proměnnou a dosadíme za tuto proměnnou do ostatních rovnic. Např. ze soustavy

$$x^2 + y^2 = 1, \quad x - 2y = 1 \tag{2.29}$$

eliminujeme proměnnou x tak, že druhou rovnici přepíšeme na $x = 2y + 1$ a pravou stranu dosadíme do první, což dá

$$(2y+1)^2 + y^2 = 1. \tag{2.30}$$

Ovšem je nutné, aby upravená soustava byla ekvivalentní původní soustavě, tj. pro všechna $x, y \in \mathbb{R}$ musí platit (2.29) $\Leftrightarrow$ (2.30). To je pravda, protože $x - 2y = 1 \iff x = 2y + 1$. Pokud bychom ale chtěli “vyjádřit” x z první rovnice jako $x = \sqrt{1-y^2}$ a dosadit do druhé rovnice, sice $x = \sqrt{1-y^2} \implies x^2 + y^2 = 1$ ale $x^2 + y^2 = 1 \not\Rightarrow x = \sqrt{1-y^2}$. Následkem toho by výsledná rovnice $\sqrt{1-y^2} - 2y = 1$ nebyla ekvivalentní původní soustavě (2.29).

Toto dobře znáte ze skalární algebry, ale ne tak dobře v maticové algebře. Student řeší např. soustavu rovnic

$$\mathbf{a}^T\mathbf{x} = 0, \quad \mathbf{x}^T\mathbf{x} = 1$$

(kde $\mathbf{a} \in \mathbb{R}^n$ je dáno a $\mathbf{x} \in \mathbb{R}^n$ je neznámá) tak, že se snaží "vyjádřit" $\mathbf{x}$ z první rovnice a dosadit ho do druhé. To je těžký zločin, protože rovnice $\mathbf{a}^T\mathbf{x} = 0$ má nekonečně mnoho řešení (pro $n > 1$) a tedy z ní neplyne, že $\mathbf{x}$ je rovno jakémukoliv jednomu vektoru.

2.7 Cvičení

2.1. Vyřešte tyto rovnice a soustavy rovnic pro neznámou matici $\mathbf{X}$ (předpokládejte, že každá potřebná inverze existuje):

a) $\mathbf{AX} + \mathbf{B} = \mathbf{A}^2\mathbf{X}$

b) $\mathbf{X} - \mathbf{A} = \mathbf{XB}$

c) $2\mathbf{X} - \mathbf{AX} + 2\mathbf{A} = \mathbf{0}$

2.2. Řešíme soustavu rovnic $\mathbf{b}_i = \mathbf{X}\mathbf{a}_i$ (kde $i = 1, \ldots, k$) pro neznámou matici $\mathbf{X} \in \mathbb{R}^{m\times n}$. Napište soustavu jako jedinou maticovou rovnici. Jaké musí být k, aby soustava měla stejný počet (skalárních) rovnic jako neznámých? Za jaké podmínky má soustava jediné řešení?

2.3. Chceme vyřešit soustavu rovnic

$$\begin{aligned} \mathbf{Ax} + (\mathbf{y}^T\mathbf{B})^T &= \alpha\mathbf{1} \\ \mathbf{Ay} + \mathbf{c} &= \mathbf{0} \end{aligned}$$

kde $\mathbf{A}$, $\mathbf{B}$ jsou známé matice, $\mathbf{c}$ je známý vektor, $\mathbf{x}$, $\mathbf{y}$ jsou neznámé vektory a α je neznámý skalár. Soustavu přepište do tvaru $\mathbf{Pu} = \mathbf{q}$, kde matice $\mathbf{P}$ a vektor $\mathbf{q}$ obsahují známé konstanty a vektor $\mathbf{u}$ obsahuje všechny neznámé.

2.4. Mějme soustavu rovnic pro neznámé vektory $\mathbf{x}$ a $\mathbf{y}$:

$$\begin{aligned} \mathbf{Ax} + \mathbf{By} &= \mathbf{a} \\ \mathbf{Cx} + \mathbf{Dy} &= \mathbf{b} \end{aligned}$$

a) Vyjádřete soustavu ve tvaru $\mathbf{Pu} = \mathbf{q}$.

b) Jestliže $\mathbf{a}, \mathbf{x} \in \mathbb{R}^m$ a $\mathbf{b}, \mathbf{y} \in \mathbb{R}^n$, eliminujte ze soustavy neznámý vektor $\mathbf{y}$ a najděte vzorec pro neznámý vektor $\mathbf{x}$. Předpokládejte přitom, že potřebné inverze existují.

2.5. V následujících soustavách rovnic malá písmena značí vektory a velká matice. Jaké jsou nejobecnější rozměry matic a vektorů, aby rovnice byly syntakticky správně? Jaký je počet rovnic a neznámých v každé soustavě? Které z těchto soustav rovnic jsou lineární?

a) $\mathbf{Ax} = \mathbf{b}$, neznámá $\mathbf{x}$.

b) $\mathbf{x}^T\mathbf{Ax} = 1$, neznámá $\mathbf{x}$.

c) $\mathbf{a}^T\mathbf{Xb} = 0$, neznámá $\mathbf{X}$.

d) $\mathbf{AX} + \mathbf{XA}^T = \mathbf{C}$, neznámá $\mathbf{X}$

e) $\{\, \mathbf{X}^T\mathbf{Y} = \mathbf{A},\ \mathbf{XY}^T = \mathbf{B} \,\}$, neznámé $\mathbf{X}, \mathbf{Y}$

2.6. Zobrazení $\operatorname{vec}: \mathbb{R}^{m\times n} \to \mathbb{R}^{mn}$ (*vektorizace* matice, v Matlabu označeno `A(:)`) je definováno tak, že vektor $\operatorname{vec}\mathbf{A} \in \mathbb{R}^{mn}$ je matice $\mathbf{A} \in \mathbb{R}^{m\times n}$ přerovnaná po sloupcích do vektoru. Např. $\operatorname{vec}\begin{bmatrix}1 & 2\\ 3 & 4\end{bmatrix} = (1,3,2,4) = \begin{bmatrix}1 & 3 & 2 & 4\end{bmatrix}^T$. *Kroneckerův součin* dvou matic je definován jako

$$\mathbf{A}\otimes\mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B}\\ \vdots & \ddots & \vdots\\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B}\end{bmatrix},$$

v Matlabu `kron(A,B)`. Pro libovolné matice (s kompatibilními velikostmi) platí

$$\operatorname{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\operatorname{vec}\mathbf{B}. \tag{2.31}$$

Použijte tohoto vzorce pro transformaci následujících soustav rovnic s neznámou maticí $\mathbf{X}$ do tvaru (2.22). Předpokládejte, že matice a vektory mají nejobecnější možné rozměry, které jsou kompatibilní s maticovými operacemi v soustavě.

a) Soustava $\mathbf{b}_i^T\mathbf{X}\mathbf{a}_i = 0\ \forall i = 1,\dots,k$, kde neznámá je matice $\mathbf{X}$.
b) Soustava $\mathbf{AX} + \mathbf{X}^T\mathbf{B}^T = \mathbf{C}$, kde neznámá je matice $\mathbf{X}$.
c) Soustava z Příkladu 2.3.

2.7. *Komutátor* dvou matic je matice $[\mathbf{A},\mathbf{B}] = \mathbf{AB} - \mathbf{BA}$. Dokažte:

a) Komutátor symetrických matic je antisymetrická matice.
b) $[\mathbf{A},[\mathbf{B},\mathbf{C}]] + [\mathbf{B},[\mathbf{C},\mathbf{A}]] + [\mathbf{C},[\mathbf{A},\mathbf{B}]] = \mathbf{0}$ (*Jacobiho identita*)
c) $[\mathbf{A},\mathbf{BC}] = [\mathbf{A},\mathbf{B}]\mathbf{C} + \mathbf{B}[\mathbf{A},\mathbf{C}]$

2.8. Dokažte, že matice $\mathbf{A}$ je sama sobě inverzí právě tehdy, když $\mathbf{A}^2 = \mathbf{I}$.

2.9. Dokažte pro regulární matice $\mathbf{A},\mathbf{B}$, že $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.

2.10. Dokažte, že inverze regulární symetrické matice je symetrická matice.

2.11. Za předpokladu, že invertované matice jsou invertovatelné, dokažte identity:

a) $\mathbf{A}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{B} = (\mathbf{A}^{-1}+\mathbf{B}^{-1})^{-1} = \mathbf{B}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{A}$
b) $(\mathbf{I}+\mathbf{AB})^{-1} = \mathbf{I} - \mathbf{A}(\mathbf{I}+\mathbf{BA})^{-1}\mathbf{B}$
c) $(\mathbf{I}+\mathbf{AB})^{-1}\mathbf{A} = \mathbf{A}(\mathbf{I}+\mathbf{AB})^{-1}$
d) $\mathbf{A} - \mathbf{A}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{A} = \mathbf{B} - \mathbf{B}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{B}$

2.12. Dokažte

a) *vzorec Shermana a Morrisonové*

$$(\mathbf{A}+\mathbf{u}\mathbf{v}^T)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}}{1+\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}}$$

za předpokladu, že $\mathbf{A}$ je regulární a $\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}+1 \neq 0$.

b) *vzorec Shermana, Morrisonové a Woodburyho* neboli *lemma o maticové inverzi*

$$(\mathbf{A}+\mathbf{UCV}^T)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{C}^{-1}+\mathbf{V}^T\mathbf{A}^{-1}\mathbf{U})^{-1}\mathbf{V}^T\mathbf{A}^{-1}$$

za předpokladu, že $\mathbf{A}$, $\mathbf{C}$ a $\mathbf{C}^{-1}+\mathbf{V}^T\mathbf{A}^{-1}\mathbf{U}$ jsou regulární.

c) Vzorec pro inverzi blokové matice

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{Q}^{-1} & -\mathbf{Q}^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}\mathbf{Q}^{-1} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}\mathbf{Q}^{-1}\mathbf{B}\mathbf{D}^{-1} \end{bmatrix} \quad \text{kde} \quad \mathbf{Q} = \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}$$

za předpokladu, že $\mathbf{D}$ a $\mathbf{Q}$ jsou regulární. (Matice $\mathbf{Q}$ je známa jako *Schurův doplněk* invertované blokové matice.) Srov. Cvičení 2.4.

2.13. Kdy je diagonální matice regulární? Co je inverzí diagonální matice?

2.14. Ukažte, že čtvercové diagonální matice komutují (tj. $\mathbf{AB} = \mathbf{BA}$).

2.15. Jistý student tvrdí, že když matice $\mathbf{A}$ má levou inverzi (označme ji $\mathbf{B}$), tak soustavu rovnic $\mathbf{Ax} = \mathbf{b}$ lze vyřešit jednoduše vynásobením zleva maticí $\mathbf{B}$. Dostaneme $\mathbf{BAx} = \mathbf{Bb}$, a jelikož $\mathbf{BA} = \mathbf{I}$, máme tedy $\mathbf{x} = \mathbf{Bb}$. Má student pravdu? Vysvětlete.

2.16. Dokažte, že pro každou čtvercovou matici $\mathbf{A}$ platí:

a) $\mathbf{A} + \mathbf{A}^T$ je symetrická,
b) $\mathbf{A} - \mathbf{A}^T$ je antisymetrická,
c) existuje právě jedna symetrická $\mathbf{B}$ a právě jedna antisymetrická $\mathbf{C}$ tak, že $\mathbf{A} = \mathbf{B} + \mathbf{C}$,
d) $\mathbf{A}^T\mathbf{A}$ je symetrická.

2.17. Pro vektory $\mathbf{a}, \mathbf{b}$ a matice $\mathbf{A}, \mathbf{B}$ dokažte:

a) $\|\mathbf{a} + \mathbf{b}\|^2 = \|\mathbf{a}\|^2 + 2\mathbf{a}^T\mathbf{b} + \|\mathbf{b}\|^2$
b) $(\mathbf{a} + \mathbf{b})^T(\mathbf{a} - \mathbf{b}) = \|\mathbf{a}\|^2 - \|\mathbf{b}\|^2$
c) $\|\mathbf{A} + \mathbf{B}\|^2 = \|\mathbf{A}\|^2 + 2\langle \mathbf{A}, \mathbf{B} \rangle + \|\mathbf{B}\|^2$
d) $\langle \mathbf{A} + \mathbf{B}, \mathbf{A} - \mathbf{B} \rangle = \|\mathbf{A}\|^2 - \|\mathbf{B}\|^2$

2.18. ($\star$) Nechť čtvercové matice $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ jsou takové, že $\mathbf{AB}^T$ a $\mathbf{CD}^T$ jsou symetrické a platí $\mathbf{AD}^T - \mathbf{BC}^T = \mathbf{I}$. Dokažte, že $\mathbf{A}^T\mathbf{D} - \mathbf{C}^T\mathbf{B} = \mathbf{I}$.

2.19. Dokažte, že pro každé $\mathbf{A} \in \mathbb{R}^{m \times n}$ a $\mathbf{B} \in \mathbb{R}^{n \times m}$ má matice

$$\mathbf{L} = \begin{bmatrix} \mathbf{I} - \mathbf{BA} & \mathbf{B} \\ 2\mathbf{A} - \mathbf{ABA} & \mathbf{AB} - \mathbf{I} \end{bmatrix}$$

vlastnost $\mathbf{L}^2 = \mathbf{I}$ (kde $\mathbf{L}^2$ je zkratka pro $\mathbf{LL}$). Matice s touto vlastností se nazývá *involuce* (viz §1.1.2).

2.20. Dokažte, že rovnice $\mathbf{XY} - \mathbf{YX} = \mathbf{I}$ nemá řešení pro žádné matice $\mathbf{X}, \mathbf{Y}$.

2.21. Dokažte identitu (2.10) bez použití (2.8), tj. přímo z definic (2.7) a (2.2).

2.22. Máte matici $\mathbf{A} \in \mathbb{R}^{n \times n}$ a znáte $\det \mathbf{A}$. Jak najdete $\det(-\mathbf{A})$?

2.23. Dokažte, že:

a) Inverze horní/dolní trojúhelníkové matice je horní/dolní trojúhelníková matice.
b) Součin horních/dolních trojúhelníkových matic je horní/dolní trojúhelníková matice.

2.24. Dokažte tyto jednoduché vlastnosti elementárních matic (2.26):

a) Transpozice elementárních matic jsou opět elementární matice. Jakým elementárním řádkovým úpravám odpovídají?
b) Inverze elementárních matic jsou opět elementární matice. Jakým elementárním řádkovým úpravám odpovídají? Inverze najděte nejprve pomocí úvah o elementárních řádkových úpravách a pak pomocí vzorce Shermana a Morrisonové ze Cvičení 2.12.

Nápověda a řešení

2.1.a) $\mathbf{X} = (\mathbf{A}^2 - \mathbf{A})^{-1}\mathbf{B} = (\mathbf{A} - \mathbf{I})^{-1}\mathbf{A}^{-1}\mathbf{B}$

2.1.b) $\mathbf{X} = \mathbf{A}(\mathbf{I} - \mathbf{B})^{-1}$

2.1.c) $\mathbf{X} = 2(\mathbf{A} - 2\mathbf{I})^{-1}\mathbf{A} = (\mathbf{A}/2 - \mathbf{I})^{-1}\mathbf{A}$

2.2. Lze napsat jako $\mathbf{B} = \mathbf{XA}$, kde $\mathbf{a}_i \in \mathbb{R}^n$ jsou sloupce $\mathbf{A} \in \mathbb{R}^{n\times k}$ a $\mathbf{b}_i \in \mathbb{R}^m$ jsou sloupce $\mathbf{B} \in \mathbb{R}^{m\times k}$. Neznámých je $m \times n$, rovnic je $m \times k$, tedy musí být $n = k$. Pro jediné řešení musí být vektory $\mathbf{a}_i$ lineárně nezávislé.

2.3. $\mathbf{u} = \begin{bmatrix}\mathbf{x}\\ \mathbf{y}\\ \alpha\end{bmatrix}$, $\mathbf{P} = \begin{bmatrix}\mathbf{A} & \mathbf{B}^T & -\mathbf{1}\\ \mathbf{0} & \mathbf{A} & \mathbf{0}\end{bmatrix}$, $\mathbf{q} = \begin{bmatrix}\mathbf{0}\\ -\mathbf{c}\end{bmatrix}$

2.4.a) $\mathbf{P} = \begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{bmatrix}$, $\mathbf{q} = \begin{bmatrix}\mathbf{a}\\ \mathbf{b}\end{bmatrix}$, $\mathbf{u} = \begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix}$

2.4.b) Díky rozměrům vektorů $\mathbf{a}, \mathbf{b}, \mathbf{x}, \mathbf{y}$ víme, že matice $\mathbf{A}, \mathbf{D}$ jsou čtvercové. Tedy ze druhé rovnice vyjádříme $\mathbf{y}$ a dosadíme do první rovnice. Dostaneme $\mathbf{x} = (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}(\mathbf{a} - \mathbf{B}\mathbf{D}^{-1}\mathbf{b})$.

2.5.a) Rovnic je m, neznámých n, kde $\mathbf{A} \in \mathbb{R}^{m\times n}$. Je lineární.

2.5.b) Rovnice je jedna, neznámých je n, kde $\mathbf{x} \in \mathbb{R}^n$. Není lineární.

2.5.c) Rovnice je jedna, neznámých je mn, kde $\mathbf{X} \in \mathbb{R}^{m\times n}$. Je lineární.

2.5.d) Všechny tři matice $\mathbf{A}, \mathbf{C}, \mathbf{X}$ musí být čtvercové velikosti $n \times n$. Rovnic i neznámých je n^2.

2.5.e) Rovnic je $m^2 + n^2$, neznámých je $2mn$, kde $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{m\times n}$. Není lineární.

2.6.a) Je $\mathbf{b}_i^T\mathbf{X}\mathbf{a}_i = \operatorname{vec}(\mathbf{b}_i^T\mathbf{X}\mathbf{a}_i) = (\mathbf{a}_i^T \otimes \mathbf{b}_i^T)\operatorname{vec}\mathbf{X}$. Tedy soustavu lze napsat jako $\begin{bmatrix}\mathbf{a}_1^T \otimes \mathbf{b}_1^T\\ \vdots\\ \mathbf{a}_k^T \otimes \mathbf{b}_k^T\end{bmatrix}\operatorname{vec}\mathbf{X} = \mathbf{0}$.

2.6.b) Musí být $\mathbf{X} \in \mathbb{R}^{m\times n}$ je neznámá a $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times m}$ a $\mathbf{C} \in \mathbb{R}^{n\times n}$ jsou dány. Jedná se tedy o lineární soustavu n^2 rovnic s mn neznámými. Neznámé ovšem nejsou soustředěné do vektoru, ale do matice, tedy soustava není ve tvaru (2.22). Je $\operatorname{vec}(\mathbf{AX}) = \operatorname{vec}(\mathbf{AXI}_n) = (\mathbf{I}_n \otimes \mathbf{A})\operatorname{vec}\mathbf{X}$ a $\operatorname{vec}(\mathbf{XB}^T) = \operatorname{vec}(\mathbf{I}_m\mathbf{XB}^T) = (\mathbf{B} \otimes \mathbf{I}_m)\operatorname{vec}\mathbf{X}$. Tedy soustavu můžeme napsat jako $((\mathbf{I}_n \otimes \mathbf{A}) + (\mathbf{B} \otimes \mathbf{I}_m))\operatorname{vec}\mathbf{X} = \operatorname{vec}\mathbf{C}$.

2.6.c) Je $\operatorname{vec}(\mathbf{X}\mathbf{1}_n) = \operatorname{vec}(\mathbf{I}_m\mathbf{X}\mathbf{1}_n) = (\mathbf{1}_n^T \otimes \mathbf{I}_m)\operatorname{vec}\mathbf{X}$ a $\operatorname{vec}(\mathbf{1}_m^T\mathbf{X}) = \operatorname{vec}(\mathbf{1}_m\mathbf{X}\mathbf{I}_n) = (\mathbf{I}_n \otimes \mathbf{1}_m)\operatorname{vec}\mathbf{X}$. Tedy soustavu můžeme psát jako $\begin{bmatrix}\mathbf{1}_n^T \otimes \mathbf{I}_m\\ \mathbf{I}_n \otimes \mathbf{1}_m\end{bmatrix}\operatorname{vec}\mathbf{X} = \begin{bmatrix}\mathbf{a}\\ \mathbf{b}\end{bmatrix}$. Zkontrolujte vypočítáním výrazů $\mathbf{1}_n^T \otimes \mathbf{I}_m$ a $\mathbf{I}_n \otimes \mathbf{1}_m$, že je to totéž jako soustava (2.25).

2.7.a) $[\mathbf{A}, \mathbf{B}]^T = (\mathbf{AB} - \mathbf{BA})^T = \mathbf{B}^T\mathbf{A}^T - \mathbf{A}^T\mathbf{B}^T = \mathbf{BA} - \mathbf{AB} = -[\mathbf{A}, \mathbf{B}]$

2.7.c) Máme dokázat, že $\mathbf{ABC} - \mathbf{BCA} = (\mathbf{AB} - \mathbf{BA})\mathbf{C} + \mathbf{B}(\mathbf{AC} - \mathbf{CA})$, což zjevně platí.

2.8. Máme dokázat, že $\mathbf{A}^{-1} = \mathbf{A}$ právě když $\mathbf{A}^2 = \mathbf{I}$. Zjevně $\mathbf{A}^{-1} = \mathbf{A}$ implikuje $\mathbf{A}^2 = \mathbf{I}$, protože $\mathbf{A}^2 = \mathbf{AA} = \mathbf{AA}^{-1} = \mathbf{I}$. Dokažme, že $\mathbf{A}^2 = \mathbf{I}$ implikuje $\mathbf{A}^{-1} = \mathbf{A}$. Protože z $\mathbf{A}^2 = \mathbf{I}$ plyne, že $\mathbf{A}$ je regulární, a tedy můžeme rovnici $\mathbf{A}^2 = \mathbf{I}$ vynásobit maticí $\mathbf{A}^{-1}$, čímž dostaneme $\mathbf{A} = \mathbf{A}^{-1}$.

2.9. Z definice inverze stačí ukázat, že $\mathbf{ABB}^{-1}\mathbf{A}^{-1} = \mathbf{I}$. To je očividné, protože $\mathbf{BB}^{-1} = \mathbf{AA}^{-1} = \mathbf{I}$.

2.10. Nechť $\mathbf{A}$ je symetrická (tedy je čtvercová a platí $\mathbf{A}^T = \mathbf{A}$) a regulární. Nechť $\mathbf{BA} = \mathbf{I}$, tedy $\mathbf{B}$ je levá inverze $\mathbf{A}$. Z toho $\mathbf{A}^T\mathbf{B}^T = \mathbf{AB}^T = \mathbf{I}$, tedy $\mathbf{B}^T$ je pravá inverze $\mathbf{A}$. Ale pro regulární matici jsou si levá a pravá inverze rovny (viz §2.1.4), tedy $\mathbf{B}^T = \mathbf{B}$.

2.11.a) Stačí dokázat první rovnost, druhá platí symetricky. Máme ukázat $\mathbf{A}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{B}(\mathbf{A}^{-1}+\mathbf{B}^{-1}) = \mathbf{I}$. Je $\mathbf{A}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{B}(\mathbf{A}^{-1}+\mathbf{B}^{-1}) = \mathbf{A}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{BA}^{-1} + \mathbf{A}(\mathbf{A}+\mathbf{B})^{-1} = \mathbf{A}(\mathbf{A}+\mathbf{B})^{-1}(\mathbf{BA}^{-1}+\mathbf{I})$. Vynásobení rovnosti $\mathbf{A}(\mathbf{A}+\mathbf{B})^{-1}(\mathbf{BA}^{-1}+\mathbf{I}) = \mathbf{I}$ zleva maticí $(\mathbf{A}+\mathbf{B})\mathbf{A}^{-1}$ (což je ekvivalentní úprava) dá $(\mathbf{BA}^{-1}+\mathbf{I}) = (\mathbf{A}+\mathbf{B})\mathbf{A}^{-1}$, což zjevně platí.

2.12.a) Máme dokázat, že následující součin je roven $\mathbf{I}$, což je cvičení na úpravu maticových výrazů:

$$\begin{aligned}(\mathbf{A}+\mathbf{u}\mathbf{v}^T)\Big(\mathbf{A}^{-1}-\frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}}{1+\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}}\Big) &= \mathbf{I}+\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}-\frac{\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}+\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}}{1+\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}}\\ &= \mathbf{I}+\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}-\frac{\mathbf{u}(1+\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u})\mathbf{v}^T\mathbf{A}^{-1}}{1+\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}}\\ &= \mathbf{I}+\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}-\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}=\mathbf{I}\end{aligned}$$

2.13. Matice $\operatorname{diag}(a_1,\ldots,a_n)$ je regulární, právě když $a_1,\ldots,a_n$ jsou všechny nenulové. Je $\operatorname{diag}(a_1,\ldots,a_n)^{-1}=\operatorname{diag}(1/a_1,\ldots,1/a_n)$. Obojí plyne přímo z definic regularity, inverze a diagonální matice.

2.14. Plyne snadno z definice diagonální matice a maticového součinu.

2.15. Nápověda: Nemá pravdu, protože (za podmínky $\mathbf{BA}=\mathbf{I}$) sice platí $\mathbf{Ax}=\mathbf{b} \;\Rightarrow\; \mathbf{x}=\mathbf{Bb}$, ale neplatí $\mathbf{Ax}=\mathbf{b} \;\Leftarrow\; \mathbf{x}=\mathbf{Bb}$. Zkuste si to třeba pro $\mathbf{A}=\begin{bmatrix}1\\2\end{bmatrix}$ a $\mathbf{b}=\begin{bmatrix}2\\1\end{bmatrix}$.

2.16.a) Máme dokázat, že $(\mathbf{A}+\mathbf{A}^T)^T=\mathbf{A}+\mathbf{A}^T$. To je jasné, protože $(\mathbf{A}^T)^T=\mathbf{A}$, tedy $(\mathbf{A}+\mathbf{A}^T)^T=\mathbf{A}^T+\mathbf{A}$.

2.16.b) Analogické jako minule.

2.16.c) Máme dokázat, že pro libovolnou danou čtvercovou matici $\mathbf{A}$ má soustava $\mathbf{A}=\mathbf{B}+\mathbf{C}$, $\mathbf{B}=\mathbf{B}^T$, $\mathbf{C}=-\mathbf{C}^T$ právě jedno řešení pro $\mathbf{B},\mathbf{C}$. Tuto soustavu tří maticových rovnic snadno vyřešíme např. takto. Transpozicí první rovnice získáme $\mathbf{A}^T=\mathbf{B}^T+\mathbf{C}^T=\mathbf{B}-\mathbf{C}$. Sečtení a odečtení s první rovnicí dá $\mathbf{A}+\mathbf{A}^T=2\mathbf{B}$ a $\mathbf{A}-\mathbf{A}^T=2\mathbf{C}$, z toho $\mathbf{B}=\frac{1}{2}(\mathbf{A}+\mathbf{A}^T)$ a $\mathbf{C}=\frac{1}{2}(\mathbf{A}-\mathbf{A}^T)$.

2.16.d) Plyne okamžitě z identity $(\mathbf{AB})^T=\mathbf{B}^T\mathbf{A}^T$ (viz §2.1.3).

2.19. Opět pouhé cvičení na úpravy maticových výrazů.

2.20. Použijte cykličnost stopy.

2.22. Máme $\det(-\mathbf{A})=\det(-\mathbf{I}_n\mathbf{A})=\det(-\mathbf{I}_n)\det\mathbf{A}=(-1)^n\det\mathbf{A}$, kde jsme užili vzorec pro determinant součinu matic (§2.1.5). Platí $\det(-\mathbf{I}_n)=(-1)^n$, neboť determinant diagonální matice $-\mathbf{I}_n$ je roven součinu prvků na diagonále.

Think geometrically, prove algebraically.
— John Tate

Kapitola 3

Linearita

Množina $\mathbb{R}^n$ spolu s operacemi sčítání vektorů a násobení vektorů skalárem tvoří *lineární prostor* (také zvaný *vektorový prostor*)[1] nad tělesem $\mathbb{R}$. Protože budeme pracovat pouze v prostoru $\mathbb{R}^n$ (a občas v prostoru $\mathbb{R}^{m\times n}$), nebudeme uvádět definici (pomocí axiomů) lineárního prostoru ani odkazovat se na ni – určitě si ji ale najděte a zopakujte! To nám sice občas zabrání využít celou sílu a krásu lineární algebry, ale zkrátí a zjednoduší to výklad.

3.1 Podprostory

Lineární kombinace vektorů $\mathbf{x}_1,\dots,\mathbf{x}_k \in \mathbb{R}^n$ je vektor

$$\mathbf{x} = \alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k \tag{3.1}$$

pro nějaké skaláry $\alpha_1,\dots,\alpha_k \in \mathbb{R}$. Vektory $\mathbf{x}_1,\dots,\mathbf{x}_k$ jsou **lineárně nezávislé**, jestliže

$$\alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k = \mathbf{0} \quad\Longrightarrow\quad \alpha_1 = \cdots = \alpha_k = 0. \tag{3.2}$$

V opačném případě (tedy když pro nějaká $\alpha_1,\dots,\alpha_k$, aspoň jedno z nich nenulové, platí $\alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k = \mathbf{0}$) jsou vektory $\mathbf{x}_1,\dots,\mathbf{x}_k$ **lineárně závislé**. Lze dokázat (viz Cvičení 3.4), že jestliže jsou vektory lineárně závislé, tak je aspoň jeden z nich lineární kombinací ostatních.

Tvrzení 3.1. Jsou-li vektory $\mathbf{x}_1,\dots,\mathbf{x}_k$ lineárně nezávislé, koeficienty $\alpha_1,\dots,\alpha_k$ v (3.1) jsou určeny jednoznačně (tj. soustava (3.1) s neznámými $\alpha_1,\dots,\alpha_k$ má právě jedno řešení).

Důkaz. Nechť kromě (3.1) platí také $\mathbf{x} = \beta_1\mathbf{x}_1 + \cdots + \beta_k\mathbf{x}_k$. Odečtením obou rovnic máme $\mathbf{0} = (\alpha_1-\beta_1)\mathbf{x}_1 + \cdots + (\alpha_k-\beta_k)\mathbf{x}_k$. Ale z (3.2) plyne $\alpha_i - \beta_i = 0$, tedy $\alpha_i = \beta_i$. ∎

Lineární obal vektorů $\mathbf{x}_1,\dots,\mathbf{x}_k$ je množina

$$\operatorname{span}\{\mathbf{x}_1,\dots,\mathbf{x}_k\} = \{\, \alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k \mid \alpha_1,\dots,\alpha_k \in \mathbb{R}\,\}$$

[1] Jaký je vlastně rozdíl mezi *množinou* a *prostorem*? Množina je 'pytel' prvků ('věcí') bez jakékoliv struktury. Prostor je neformální název pro algebraickou strukturu, která má nějaký geometrický význam. *Algebraická struktura* je množina (někdy i více množin) spolu s operacemi definovanými na této množině (příklady: grupa, těleso, lineární prostor, lineární prostor se skalárním součinem). Operace a jejich vlastnosti určují vlastnosti algebraické struktury, tedy i prostoru. Operace se ale v matematickém textu často explicitně nezmiňují (je z kontextu jasné, které to jsou) a tak se místo o algebraické struktuře mluví jen o množině.

všech jejich lineárních kombinací. Zde mlčky předpokládáme, že vektorů je konečný počet (lineární obal nekonečné množiny vektorů se definuje trochu jinak, více o tom později v §13.2).

Neprázdná množina $X \subseteq \mathbb{R}^n$ se nazývá **lineární podprostor**[2] (často se přívlastek 'lineární' vynechává a říkáme tedy jen **podprostor**) lineárního prostoru $\mathbb{R}^n$, jestliže každá lineární kombinace každé (konečné) množiny vektorů z X leží v X (neboli množina X je uzavřená na lineární kombinace):

$$\mathbf{x}_1, \dots, \mathbf{x}_k \in X, \ \ \alpha_1, \dots, \alpha_k \in \mathbb{R} \quad \Longrightarrow \quad \alpha_1 \mathbf{x}_1 + \dots + \alpha_k \mathbf{x}_k \in X. \tag{3.3}$$

Tvrzení (3.3) je ekvivalentní dvěma tvrzením

$$\mathbf{x}, \mathbf{y} \in X \quad \Longrightarrow \quad \mathbf{x} + \mathbf{y} \in X, \tag{3.4a}$$
$$\mathbf{x} \in X, \ \alpha \in \mathbb{R} \quad \Longrightarrow \quad \alpha \mathbf{x} \in X, \tag{3.4b}$$

kde implikace (3.3) $\Rightarrow$ (3.4) je očividná a implikace (3.4) $\Rightarrow$ (3.3) by se dokázala např. indukcí. Snadno se ukáže, že lineární obal libovolné množiny vektorů je lineární podprostor.

Každý podprostor je neprázdný, protože sám o sobě to musí být lineární *prostor*, a ten ze své definice musí obsahovat nulový vektor $\mathbf{0}$ (kterému se z historických důvodů také říká *počátek*). Podprostor $X = \{\mathbf{0}\}$ se nazývá **triviální**..

Báze lineárního podprostoru $X \subseteq \mathbb{R}^n$ je lineárně nezávislá množina[3] vektorů, jejichž lineární obal je X. Platí tato zásadní tvrzení (důkazy nejsou krátké a neuvádíme je, najdete je v každé učebnici lineární algebry):

Věta 3.2 (vlastnosti báze).

- Každý lineární podprostor má (alespoň jednu) bázi.
- Každá báze lineárního podprostoru má stejný počet vektorů.
- Z každé množiny vektorů lze vybrat bázi jejich lineárního obalu.
- Každou lineárně nezávislou množinu vektorů z lineárního podprostoru lze doplnit na jeho bázi.

Počet vektorů báze lineárního podprostoru X se nazývá jeho **dimenze**, značíme ji $\dim X$. Je-li $\mathbf{x}_1, \dots, \mathbf{x}_k$ báze podprostoru X a $\mathbf{x} \in X$, pak (jednoznačně určené) skaláry $\alpha_1, \dots, \alpha_k$ v (3.1) se nazývají **souřadnice** vektoru $\mathbf{x}$ v bázi $\mathbf{x}_1, \dots, \mathbf{x}_k$.

Tvrzení 3.3. Pro každé podprostory $X, Y \subseteq \mathbb{R}^n$ platí:

- $X \subseteq Y$ implikuje $\dim X \leq \dim Y$.
- $X \subseteq Y$ a $\dim X = \dim Y$ implikuje $X = Y$.

Důkaz. Jestliže $X \subseteq Y$, každá báze podprostoru X patří do Y. Dle Věty 3.2 lze tuto bázi doplnit na bázi podprostoru Y, odtud první tvrzení. Jestliže navíc $\dim X = \dim Y$, každá báze X je už bází Y (tedy doplnění nepřidá žádný vektor), odtud druhé tvrzení. ∎

[2] Krátce budeme říkat, že $X \subseteq \mathbb{R}^n$ je podprostor. To nemůže způsobit zmatení, neboť aby množina byla podprostor prostoru $\mathbb{R}^n$, musí všechny její prvky patřit do $\mathbb{R}^n$. Např. množina $X \subseteq \mathbb{R}^3$ tedy nemůže být podprostor prostoru $\mathbb{R}^2$, to by byl nesmysl.

[3] Někdy (např. když mluvíme o souřadnicích vektoru vzhledem k bázi) nám záleží na pořadí prvků báze. Pak musíme bázi považovat ne za množinu, ale za *uspořádanou* množinu nebo *posloupnost* vektorů.

Příklad 3.1. Uveďme jednoduché příklady podprostorů prostorů $\mathbb{R}^n$:

- Triviálně, množina $X = \mathbb{R}^3$ je podprostorem prostoru $\mathbb{R}^3$. Jeho dimenze je 3. Jeho báze je např. množina $\{(1,0,0),(0,1,0),(0,0,1)\}$ (standardní báze) nebo $\{(1,1,1),(1,-1,0),(2,0,0)\}$.
- Množina $X = \operatorname{span}\{(1,2,3)\} = \{\,\alpha(1,2,3) \mid \alpha \in \mathbb{R}\,\} \subseteq \mathbb{R}^3$ je podprostor dimenze 1 prostoru $\mathbb{R}^3$. Jeho báze je např. množina $\{(1,2,3)\}$, jiná báze je $\{(2,4,6)\}$. Je to přímka procházející počátkem.
- Množina $X = \operatorname{span}\{(1,2,3),(1,0,-1)\} = \{\,\alpha(1,2,3)+\beta(1,0,-1) \mid \alpha,\beta \in \mathbb{R}\,\}$ je podprostor dimenze 2 prostoru $\mathbb{R}^3$. Jeho báze je např. množina $\{(1,2,3),(1,0,-1)\}$, jiná báze je $\{(3,2,1),(0,2,4)\}$. Je to rovina procházející počátkem. Všimněte si, že tato množina nemá žádnou 'přirozenou' nebo 'standardní' bázi, jako mají prostory $\mathbb{R}^n$.
- Množina všech trojic $(x_1,x_2,x_3) \in \mathbb{R}^3$ splňujících rovnici $x_1 + 2x_2 - x_3 = 0$ je podprostor dimenze 2 prostoru $\mathbb{R}^3$. Jeho báze je libovolná dvojice lineárně nezávislých vektorů $\{(a_1,a_2,a_2),(b_1,b_2,b_3)\}$ splňujících $a_1 + 2a_2 - a_3 = 0$ a $b_1 + 2b_2 - b_3 = 0$, např. $\{(1,0,1),(0,1,2)\}$. Je to rovina procházející počátkem.
- Množina $X = \operatorname{span}\{(1,2,3,0),(0,1,2,-1),(0,2,4,-2)\} \subseteq \mathbb{R}^4$ je podprostor $\mathbb{R}^4$ dimenze 2. Všimněte si, že vektory jsou lineárně závislé. Báze podprostoru X je např. množina $\{(1,2,3,0),(0,1,2,-1)\}$.
- Všechny možné podprostory prostoru $\mathbb{R}^3$ jsou počátek $\mathbf{0}$ (dimenze 0), všechny přímky procházející počátkem (dimenze 1), všechny roviny procházející počátkem (dimenze 2), a konečně celý prostor $\mathbb{R}^3$ (dimenze 3).
- Množina $X = \{\,(1+\alpha,\alpha) \mid \alpha \in \mathbb{R}\,\} \subseteq \mathbb{R}^2$ (přímka neprocházející počátkem) není podprostor, protože např. $(1,0) \in X$ ale $2(1,0) = (2,0) \notin X$. ♦

3.2 Lineární zobrazení

Zobrazení $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$ je **lineární**, jestliže pro každé $\mathbf{x}_1,\dots,\mathbf{x}_k \in \mathbb{R}^n$ a $\alpha_1,\dots,\alpha_k \in \mathbb{R}$ platí

$$\mathbf{f}(\alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k) = \alpha_1\mathbf{f}(\mathbf{x}_1) + \cdots + \alpha_k\mathbf{f}(\mathbf{x}_k). \tag{3.5}$$

Tato podmínka je ekvivalentní dvěma podmínkám, pro každé $\mathbf{x},\mathbf{y} \in X$ a $\alpha \in \mathbb{R}$,

$$\mathbf{f}(\mathbf{x}+\mathbf{y}) = \mathbf{f}(\mathbf{x}) + \mathbf{f}(\mathbf{y}), \tag{3.6a}$$
$$\mathbf{f}(\alpha\mathbf{x}) = \alpha\mathbf{f}(\mathbf{x}). \tag{3.6b}$$

Věta 3.4. Zobrazení $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$ je lineární, právě když existuje matice $\mathbf{A} \in \mathbb{R}^{m\times n}$ splňující[4]

$$\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x} \tag{3.7}$$

pro všechna $\mathbf{x} \in \mathbb{R}^n$. V tom případě je matice $\mathbf{A}$ zobrazením $\mathbf{f}$ určena jednoznačně.

[4]Možná se ptáte, proč tedy nedefinujeme lineární zobrazení rovnou vztahem (3.7). Pro zobrazení $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$ je to opravdu možné a definici (3.5) vlastně nepotřebujeme. Přesto definici (3.5) uvádíme, protože jednak jste na ni zvyklí z lineární algebry a jednak funguje i pro zobrazení mezi obecnými (tj. axiomy definovanými) lineární prostory.

Důkaz. Zobrazení (3.7) je lineární, protože $\mathbf{f}(\mathbf{x}+\mathbf{y}) = \mathbf{A}(\mathbf{x}+\mathbf{y}) = \mathbf{A}\mathbf{x}+\mathbf{A}\mathbf{y} = \mathbf{f}(\mathbf{x})+\mathbf{f}(\mathbf{y})$ a $\mathbf{f}(\alpha\mathbf{x}) = \mathbf{A}(\alpha\mathbf{x}) = \alpha\mathbf{A}\mathbf{x} = \alpha\mathbf{f}(\mathbf{x})$.

Nechť $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$ je lineární zobrazení. Nechť $\mathbf{e}_1, \dots, \mathbf{e}_n \in \mathbb{R}^n$ je standardní báze prostoru $\mathbb{R}^n$. Pro každé $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ máme $\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n$. Z (3.5) plyne

$$\mathbf{f}(\mathbf{x}) = \mathbf{f}(x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n) = x_1\mathbf{f}(\mathbf{e}_1) + \cdots + x_n\mathbf{f}(\mathbf{e}_n) = \underbrace{\begin{bmatrix}\mathbf{f}(\mathbf{e}_1) & \cdots & \mathbf{f}(\mathbf{e}_n)\end{bmatrix}}_{\mathbf{A}}\mathbf{x}.$$

■

Platí-li (3.7), říkáme, že matice $\mathbf{A}$ *reprezentuje* lineární zobrazení $\mathbf{f}$.

Pokud $m = 1$, lineární zobrazení je funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ a obvykle se mu říká *lineární funkce* nebo *lineární forma*. Pak je zvykem ji psát jako

$$f(\mathbf{x}) = \mathbf{a}^T\mathbf{x} = a_1x_1 + \cdots + a_nx_n, \tag{3.8}$$

což je skalární součin vektorů $\mathbf{a}, \mathbf{x} \in \mathbb{R}^n$ neboli maticový součin matice $\mathbf{A} = \mathbf{a}^T \in \mathbb{R}^{1\times n}$ a vektoru $\mathbf{x}$.

Příklad 3.2. Zobrazení $\mathbf{f}\colon \mathbb{R}^2 \to \mathbb{R}^3$ definované jako $\mathbf{f}(x_1, x_2) = (x_1 + x_2, x_1 - x_2, 2x_1)$ je lineární. To bychom dokázali ověřením podmínky (3.5). Ovšem je to patrné na první pohled, protože jej lze vyjádřit ve tvaru (3.7):

$$\mathbf{f}(x_1, x_2) = \begin{bmatrix}1 & 1\\ 1 & -1\\ 2 & 0\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix} = \begin{bmatrix}x_1 + x_2\\ x_1 - x_2\\ 2x_1\end{bmatrix} = (x_1 + x_2, x_1 - x_2, 2x_1).$$

♦

Výraz $\mathbf{A}\mathbf{x}$ ve vzorci (3.7) je maticový součin matice $m \times n$ s maticí $n \times 1$, tj. se sloupcovým vektorem (viz §2.2). Je důležité se s tímto výrazem beze zbytku seznámit, protože ho v různém značení budeme potkávat znovu a znovu. Označíme-li $\mathbf{y} = \mathbf{A}\mathbf{x}$, podle (2.2) je

$$\begin{aligned} y_1 &= a_{11}x_1 + \cdots + a_{1n}x_n,\\ &\;\;\vdots\\ y_m &= a_{m1}x_1 + \cdots + a_{mn}x_n. \end{aligned} \tag{3.9}$$

Napíšeme-li matici $\mathbf{A}$ pomocí sloupců, máme

$$\mathbf{A}\mathbf{x} = \begin{bmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_n\end{bmatrix}\begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix} = x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n, \tag{3.10}$$

tedy vektor $\mathbf{A}\mathbf{x}$ je *lineární kombinace sloupců* matice $\mathbf{A}$ s koeficienty $\mathbf{x}$. Tohle si dobře uvědomte! Plyne z toho například, že definice lineární nezávislosti (3.2) pro sloupce matice $\mathbf{A}$ zní

$$\mathbf{A}\mathbf{x} = \mathbf{0} \implies \mathbf{x} = \mathbf{0} \tag{3.11}$$

(nedejte se poplést tím, že $\mathbf{x}$ v (3.11) označuje koeficienty lineární kombinace, zatímco v (3.2) označuje kombinované vektory!).

Napíšeme-li matici $\mathbf{A}$ pomocí řádků, máme

$$\mathbf{A}\mathbf{x} = \begin{bmatrix}\mathbf{a}_1^T\\ \vdots\\ \mathbf{a}_m^T\end{bmatrix}\mathbf{x} = \begin{bmatrix}\mathbf{a}_1^T\mathbf{x}\\ \vdots\\ \mathbf{a}_m^T\mathbf{x}\end{bmatrix}, \tag{3.12}$$

tedy složky vektoru $\mathbf{A}\mathbf{x}$ jsou *skalární součiny řádků* matice $\mathbf{A}$ a vektoru $\mathbf{x}$. Všimněte si, že (3.10) a (3.12) jsou speciální případy (2.21) a (2.20).

Zde je snadný ale důležitý fakt o skládání (viz §1.1.2) lineárních zobrazení:

Tvrzení 3.5. Složení lineárních zobrazení je opět lineární zobrazení, přičemž matice složeného zobrazení je součinem matic jednotlivých zobrazení.

Důkaz. Tvrzení stačí dokázat pro dvě zobrazení, pro více plyne z asociativity skládání zobrazení. Je-li $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x}$ a $\mathbf{g}(\mathbf{y}) = \mathbf{B}\mathbf{y}$, pak $(\mathbf{g} \circ \mathbf{f})(\mathbf{x}) = \mathbf{g}(\mathbf{f}(\mathbf{x})) = \mathbf{B}(\mathbf{A}\mathbf{x}) = (\mathbf{B}\mathbf{A})\mathbf{x} = \mathbf{B}\mathbf{A}\mathbf{x}$. ■

Všimněte si, že v důkazu jsme použili asociativitu maticového součinu (viz §2.1.2). Protože je skládání zobrazení asociativní, musí být i maticový součin. To je velmi přirozený důvod, proč maticový součin musí být asociativní.

3.2.1 Prostor obrazů a nulový prostor

S lineárním zobrazením jsou spjaty dva lineární podprostory, prostor obrazů a nulový prostor. Je-li zobrazení reprezentováno maticí jako $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x}$, hovoříme o prostoru obrazů a nulovém prostoru matice $\mathbf{A}$.

Prostor obrazů (angl. *range*) matice $\mathbf{A} \in \mathbb{R}^{m\times n}$ (příp. zobrazení $\mathbf{f}$) je množina[5]

$$\operatorname{rng} \mathbf{A} = \{\, \mathbf{A}\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n \,\}. \tag{3.13}$$

Z této definice plyne, že prostor obrazů je

- množina $\mathbf{f}(\mathbb{R}^n)$ všech hodnot, jichž může zobrazení $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x}$ nabýt,
- množina všech vektorů $\mathbf{y}$, pro které má soustava rovnic $\mathbf{A}\mathbf{x} = \mathbf{y}$ řešení,
- lineární obal sloupců matice $\mathbf{A}$ (viz (3.10)), tedy je to lineární podprostor prostoru $\mathbb{R}^m$.

Nulový prostor (ang. *null space*) matice $\mathbf{A}$ (příp. zobrazení $\mathbf{f}$) je množina

$$\operatorname{null} \mathbf{A} = \{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} = \mathbf{0} \,\}. \tag{3.14}$$

Z této definice plyne, že nulový prostor je

- množina všech vektorů, které se zobrazí do nulového vektoru.
- množina všech vektorů, které jsou ortogonální na každý řádek matice $\mathbf{A}$ (viz (3.12)). Z toho je vidět, že je to lineární podprostor $\mathbb{R}^n$.

Pro prostor obrazů a nulový prostor se často používají i jiné názvy. Prostoru $\operatorname{null} \mathbf{A}$ se říká také *jádro* nebo *kernel*, což se častěji používá pro lineární zobrazení mezi obecnými (axiomy definovanými) lineárními prostory. Prostoru $\operatorname{rng} \mathbf{A}$ se říká také *sloupcový prostor* matice $\mathbf{A}$, protože je to lineární obal sloupců. Lineárnímu obalu řádků $\operatorname{rng}(\mathbf{A}^T)$ se pak říká *řádkový prostor* matice $\mathbf{A}$. Podobně, prostorům $\operatorname{null} \mathbf{A}$ resp. $\operatorname{null}(\mathbf{A}^T)$ se přesněji říká *pravý nulový prostor* resp. *levý nulový prostor* matice $\mathbf{A}$. Čtyřem podprostorům

$$\operatorname{rng} \mathbf{A}, \quad \operatorname{rng}(\mathbf{A}^T), \quad \operatorname{null} \mathbf{A}, \quad \operatorname{null}(\mathbf{A}^T) \tag{3.15}$$

se někdy říká *základní* (nebo *fundamentální*) *podprostory* generované maticí $\mathbf{A}$. Uvidíme později v §4.2, že tyto podprostory jsou po dvojicích svázány operací ortogonálního doplňku.

[5]Někdo raději píše $\operatorname{rng}(\mathbf{A})$ než $\operatorname{rng} \mathbf{A}$, ale přece je normální psát např. $\sin x$ a ne $\sin(x)$.

3.2.2 Hodnost

Hodnost matice je definována jako dimenze lineárního obalu jejích sloupců,

$$\operatorname{rank} \mathbf{A} = \dim \operatorname{rng} \mathbf{A}. \tag{3.16}$$

> **Věta 3.6.** Pro každou matici $\mathbf{A} \in \mathbb{R}^{m \times n}$ hodnosti r existují matice $\mathbf{B} \in \mathbb{R}^{m \times r}$ a $\mathbf{C} \in \mathbb{R}^{r \times n}$ tak, že $\mathbf{A} = \mathbf{BC}$.

Důkaz. Zvolme libovolnou bázi prostoru $\operatorname{rng} \mathbf{A}$. Nechť tato báze tvoří sloupce matice $\mathbf{B} \in \mathbb{R}^{m \times r}$, tedy $r = \operatorname{rank} \mathbf{A} = \dim \operatorname{rng} \mathbf{A}$. Nyní j-tý sloupec $\mathbf{a}_j$ matice $\mathbf{A}$ je lineární kombinací sloupců matice $\mathbf{B}$, neboli $\mathbf{a}_j = \mathbf{B}\mathbf{c}_j$ pro nějaké $\mathbf{c}_j \in \mathbb{R}^r$. Je tedy $\mathbf{A} = \mathbf{BC}$, kde matice $\mathbf{C} \in \mathbb{R}^{r \times n}$ má sloupce $\mathbf{c}_1, \ldots, \mathbf{c}_n$. ■

Rozkladu dle Věty 3.6 se říká **rozklad matice podle hodnosti** (angl. *rank factorization*). Lze jej vidět jako 'kompresi' matice $\mathbf{A}$, což při $r \ll \min\{m, n\}$ může být značná úspora. Např.:

- Uložení matice $\mathbf{A}$ do paměti zabere mn čísel, uložení matic $\mathbf{B}$ a $\mathbf{C}$ jen $(m + n)r$ čísel.
- Přímý výpočet maticového součinu $\mathbf{Ax}$ vyžaduje $O(mn)$ operací. Spočítáme-li ale nejdřív vektor $\mathbf{z} = \mathbf{Cx}$ a pak vektor $\mathbf{y} = \mathbf{Bz}$, potřebujeme jen $O((m + n)r)$ operací.
- Představme si zobrazení $\mathbf{f}(\mathbf{x}) = \mathbf{Ax}$ jako přenosový kanál, jehož vstupem jsou proměnné $x_1, \ldots, x_n$ a výstupem $y_1, \ldots, y_m$. Přenos můžeme realizovat jen po r 'drátech' $z_1, \ldots, z_r$.

S pomocí Věty 3.6 dokážeme zásadní fakt, že dimenze lineárního obalu sloupců je rovna dimenzi lineárního obalu řádků:

> **Věta 3.7.** Pro každou matici $\mathbf{A}$ platí $\operatorname{rank} \mathbf{A} = \operatorname{rank}(\mathbf{A}^T)$.

Důkaz. Pišme $\mathbf{A} = \mathbf{BC}$ ve Větě 3.6 jako $\mathbf{A}^T = \mathbf{C}^T\mathbf{B}^T$. Z Tvrzení 3.10 je $\operatorname{rng}(\mathbf{A}^T) \subseteq \operatorname{rng}(\mathbf{C}^T)$. Máme tedy

$$\operatorname{rank}(\mathbf{A}^T) = \dim \operatorname{rng}(\mathbf{A}^T) \le \dim \operatorname{rng}(\mathbf{C}^T) \le r = \operatorname{rank} \mathbf{A}, \tag{3.17}$$

kde první nerovnost plyne z Věty 3.3 a druhá nerovnost z toho, že matice $\mathbf{C}^T$ má r sloupců.

Nerovnost (3.17) platí pro každou matici $\mathbf{A}$. Můžeme ji tedy použít na matici $\mathbf{A}^T$ místo $\mathbf{A}$, čímž dostaneme $\operatorname{rank} \mathbf{A} \le \operatorname{rank}(\mathbf{A}^T)$. Obě nerovnosti dohromady dají $\operatorname{rank} \mathbf{A} = \operatorname{rank}(\mathbf{A}^T)$. ■

Z Věty 3.2 plyne, že dimenze lineárního obalu množiny vektorů nemůže být větší než počet těchto vektorů. To spolu s Větou 3.7 znamená, že pro každou matici $\mathbf{A} \in \mathbb{R}^{m \times n}$ je

$$\operatorname{rank} \mathbf{A} \le \min\{m, n\}. \tag{3.18}$$

Když $\operatorname{rank} \mathbf{A} = \min\{m, n\}$, říkáme, že matice má **plnou hodnost**. Je $\operatorname{rank} \mathbf{A} = n$, právě když $\mathbf{A}$ má lineárně nezávislé sloupce, a $\operatorname{rank} \mathbf{A} = m$, právě když $\mathbf{A}$ má lineárně nezávislé řádky.

Z Věty (3.18) plyne, že každá množina více než n vektorů v $\mathbb{R}^n$ je lineárně závislá. Tento známý fakt se může zdát očividný, ale pro jeho důkaz jsme potřebovali (netriviální) Větu 3.7.

Je užitečné si pamatovat horní mez na hodnost součinu matic (důkaz viz Cvičení 3.15):

$$\operatorname{rank}(\mathbf{AB}) \le \min\{\operatorname{rank} \mathbf{A}, \operatorname{rank} \mathbf{B}\}. \tag{3.19}$$

3.2.3 Matice s plnou hodností

Matice s plnou hodností, tj. s lineárně nezávislými sloupci nebo lineárně nezávislými řádky, jsou důležité a 'hezké'. Uvedeme řadu vlastností, které jsou tomu ekvivalentní. Důkazy vět jsou jednoduché, spíše jde o to shrnout a uspořádat dosavadní poznatky (a nezamotat se v tom).

Věta 3.8. Pro matici $\mathbf{A} \in \mathbb{R}^{m\times n}$ jsou následující výroky ekvivalentní:

1. $\operatorname{rng} \mathbf{A} = \mathbb{R}^m$
2. Soustava $\mathbf{Ax} = \mathbf{y}$ má řešení pro každé $\mathbf{y}$.
3. $\operatorname{rank} \mathbf{A} = m$
4. Zobrazení $\mathbf{f}(\mathbf{x}) = \mathbf{Ax}$ je surjektivní, tj. $\mathbf{f}(\mathbb{R}^n) = \mathbb{R}^m$ (viz §1.1.2).
5. Řádky matice $\mathbf{A}$ jsou lineárně nezávislé.
6. Matice $\mathbf{A}$ má pravou inverzi, tj. $\mathbf{AB} = \mathbf{I}$ pro nějakou matici $\mathbf{B}$.
7. Matice $\mathbf{AA}^T \in \mathbb{R}^{m\times m}$ je regulární.

Důkaz.

- $1 \Leftrightarrow 2 \Leftrightarrow 3 \Leftrightarrow 4$ plyne přímo z definic.
- $3 \Leftrightarrow 5$ plyne z definice hodnosti a z Věty 3.7.
- $2 \Rightarrow 6$ platí, neboť soustava $\mathbf{Ab}_i = \mathbf{e}_i$ má řešení $\mathbf{b}_i$ pro každé i (kde $\mathbf{e}_i$ resp. $\mathbf{b}_i$ je i-tý sloupec matice $\mathbf{I}$ resp. $\mathbf{B}$). Pro důkaz $6 \Rightarrow 2$ položíme $\mathbf{x} = \mathbf{By}$.
- Tvrzení $1 \Leftrightarrow 7$ plyne z rovnosti (5.6a), uvedené později. ∎

Věta 3.9. Pro matici $\mathbf{A} \in \mathbb{R}^{m\times n}$ jsou následující výroky ekvivalentní:

1. $\operatorname{null} \mathbf{A} = \{\mathbf{0}\}$ (tj. nulový prostor je triviální).
2. Soustava $\mathbf{Ax} = \mathbf{0}$ má jediné řešení $\mathbf{x} = \mathbf{0}$.
3. $\operatorname{rank} \mathbf{A} = n$.
4. Zobrazení $\mathbf{f}(\mathbf{x}) = \mathbf{Ax}$ je injektivní (viz §1.1.2).
5. Sloupce matice $\mathbf{A}$ jsou lineárně nezávislé.
6. Matice $\mathbf{A}$ má levou inverzi, tj. $\mathbf{BA} = \mathbf{I}$ pro nějakou matici $\mathbf{B}$.
7. Matice $\mathbf{A}^T\mathbf{A} \in \mathbb{R}^{n\times n}$ je regulární.

Důkaz.

- Ekvivalence $1 \Leftrightarrow 2$ plyne z definice nulového prostoru (3.14).
- Ekvivalence $2 \Leftrightarrow 5$ plyne z definice lineární nezávislosti (3.2), tj. $\mathbf{Ax} = \mathbf{0} \;\Rightarrow\; \mathbf{x} = \mathbf{0}$.
- Ekvivalence $3 \Leftrightarrow 5$ plyne z definice hodnosti (3.16).
- Tvrzení 4 říká, že pro každé $\mathbf{x}, \mathbf{y}$ je $\mathbf{Ax} = \mathbf{Ay} \;\Rightarrow\; \mathbf{x} = \mathbf{y}$, tj. $\mathbf{A}(\mathbf{x} - \mathbf{y}) = \mathbf{0} \;\Rightarrow\; \mathbf{x} - \mathbf{y} = \mathbf{0}$. Ale to je definice (3.2) lineární nezávislosti sloupců $\mathbf{A}$. Tedy platí $2 \Leftrightarrow 4$.
- Tvrzení 6 je ekvivalentní tomu, že matice $\mathbf{A}^T$ má pravou inverzi, tj. $\mathbf{A}^T\mathbf{B}^T = \mathbf{I}$. Tedy $3 \Leftrightarrow 6$ plyne z ekvivalence $3 \Leftrightarrow 6$ ve Větě 3.8.
- Tvrzení $1 \Leftrightarrow 7$ je plyne z rovnosti (5.6b), uvedené později. ∎

Z těchto vět plyne, že čtvercová matice má plnou hodnost, právě když je regulární (viz §2.1.4).

Nakonec uvedeme snadné tvrzení, které se bude později mockrát hodit:

Tvrzení 3.10. Pro libovolné matice $\mathbf{A}, \mathbf{B}$ platí:

- $\operatorname{rng}(\mathbf{AB}) \subseteq \operatorname{rng} \mathbf{A}$.
- $\operatorname{rng}(\mathbf{AB}) = \operatorname{rng} \mathbf{A}$, jestliže řádky matice $\mathbf{B}$ jsou lineárně nezávislé.
- $\operatorname{null}(\mathbf{AB}) \supseteq \operatorname{null} \mathbf{B}$.
- $\operatorname{null}(\mathbf{AB}) = \operatorname{null} \mathbf{B}$, jestliže sloupce matice $\mathbf{A}$ jsou lineárně nezávislé.

Důkaz.

- Máme dokázat, že jestliže soustava $\mathbf{z} = \mathbf{ABx}$ má řešení, pak soustava $\mathbf{z} = \mathbf{Ay}$ má řešení. To je ale jasné, protože vezmeme $\mathbf{y} = \mathbf{Bx}$.
- Druhé tvrzení navíc říká, že jestliže má $\mathbf{A}$ lineárně nezávislé řádky a soustava $\mathbf{z} = \mathbf{Ay}$ má řešení, pak soustava $\mathbf{z} = \mathbf{ABx}$ má řešení. To platí, neboť dle Věty 3.8 má soustava $\mathbf{Bx} = \mathbf{y}$ řešení pro každé $\mathbf{y}$.
- Vynásobíme-li rovnost $\mathbf{Bx} = \mathbf{0}$ maticí $\mathbf{A}$ zleva, dostaneme $\mathbf{ABx} = \mathbf{0}$.
- Druhé tvrzení říká, že když $\mathbf{A}$ má lineárně nezávislé sloupce, pak $\mathbf{ABx} = \mathbf{0} \Rightarrow \mathbf{Bx} = \mathbf{0}$. To platí, neboť dle Věty 3.9 platí $\mathbf{Ay} = \mathbf{0} \Rightarrow \mathbf{y} = \mathbf{0}$. ■

3.2.4 Věta o hodnosti a nulitě

Zatímco dimenze prostoru obrazů matice se jinak nazývá hodnost, dimenzi nulového prostoru matice se někdy říká **nulita** matice. Zopakujeme nyní veledůležitou větu o vztahu hodnosti a nulity, známou jako *rank-plus-nullity theorem*.

Věta 3.11. Pro každou matici $\mathbf{A} \in \mathbb{R}^{m\times n}$ platí

$$\underbrace{\dim \operatorname{rng} \mathbf{A}}_{\operatorname{rank} \mathbf{A}} + \dim \operatorname{null} \mathbf{A} = n. \tag{3.20}$$

Důkaz. Nechť báze prostoru $\operatorname{rng} \mathbf{A}$ jsou sloupce matice $\mathbf{B} \in \mathbb{R}^{m\times r}$. Tedy existuje matice $\mathbf{C} \in \mathbb{R}^{n\times r}$ tak, že $\mathbf{B} = \mathbf{AC}$. Nechť báze prostoru $\operatorname{null} \mathbf{A}$ jsou sloupce matice $\mathbf{D} \in \mathbb{R}^{n\times q}$.

Protože sloupce $\mathbf{B}$ jsou báze $\operatorname{rng} \mathbf{A}$, pro každý vektor $\mathbf{x} \in \mathbb{R}^n$ existuje právě jeden vektor $\mathbf{y} \in \mathbb{R}^r$ tak, že $\mathbf{Ax} = \mathbf{By}$, to jest $\mathbf{Ax} = \mathbf{ACy}$, to jest $\mathbf{A}(\mathbf{x} - \mathbf{Cy}) = \mathbf{0}$. To ale znamená $\mathbf{x} - \mathbf{Cy} \in \operatorname{null} \mathbf{A}$, proto existuje právě jeden vektor $\mathbf{z} \in \mathbb{R}^q$ tak, že $\mathbf{x} - \mathbf{Cy} = \mathbf{Dz}$.

A jsme hotovi. Ukázali jsme totiž, že pro každé $\mathbf{x} \in \mathbb{R}^n$ existuje právě jedno $\mathbf{y} \in \mathbb{R}^r$ a právě jedno $\mathbf{z} \in \mathbb{R}^q$ tak, že $\mathbf{x} = \mathbf{Cy} + \mathbf{Dz}$. To ale znamená, že sloupce matice $\begin{bmatrix} \mathbf{C} & \mathbf{D} \end{bmatrix} \in \mathbb{R}^{n\times(r+q)}$ jsou báze prostoru $\mathbb{R}^n$. Tedy musí být $r + q = n$, což je rovnost (3.20). ■

Interpretace věty 3.11:

- Každá dimenze na vstupu zobrazení $\mathbf{f}(\mathbf{x}) = \mathbf{Ax}$ se buď 'splácne' do nulového vektoru nebo se objeví na výstupu.
- Počet lineárně nezávislých řešení lineární homogenní soustavy $\mathbf{Ax} = \mathbf{0}$ je $n - \operatorname{rank} \mathbf{A}$.

- Dle Věty 3.7 je rovnost (3.20) ekvivalentní rovnosti

$$\dim \operatorname{rng}(\mathbf{A}^T) + \dim \operatorname{null} \mathbf{A} = n. \tag{3.21}$$

Zde $\operatorname{rng}(\mathbf{A}^T)$ je lineární obal řádků matice $\mathbf{A}$ (řádkový prostor $\mathbf{A}$) a $\operatorname{null}\mathbf{A}$ je množina všech vektorů ortogonálních na řádky $\mathbf{A}$. V §4.2 ukážeme, že tyto dva podprostory jsou ortogonální doplněk jeden druhého. Rovnost (3.21) tedy říká, že součet dimenzí podprostoru a jeho ortogonálního doplňku je n (Věta 4.1).

3.3 Afinní podprostor a zobrazení

Afinní kombinace vektorů $\mathbf{x}_1, \dots, \mathbf{x}_k \in \mathbb{R}^n$ je lineární kombinace $\alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k$, ve které koeficienty kombinace splňují

$$\alpha_1 + \cdots + \alpha_k = 1.$$

Afinní obal vektorů $\mathbf{x}_1, \dots, \mathbf{x}_k$ je množina všech jejich afinních kombinací. Množina $A \subseteq \mathbb{R}^n$ je **afinní podprostor**[6] lineárního prostoru $\mathbb{R}^n$, jestliže každá afinní kombinace každé (konečné) množiny vektorů z A leží v A (neboli množina A je uzavřená na afinní kombinace):

$$\mathbf{x}_1, \dots, \mathbf{x}_k \in A,\ \ \alpha_1 + \cdots + \alpha_k = 1 \quad \Longrightarrow \quad \alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k \in A. \tag{3.22}$$

Afinní kombinace *nezávisí na počátku*. To znamená, že afinní kombinace vektorů posunutých o libovolný vektor $\mathbf{x}_0$ je rovna afinní kombinaci vektorů posunuté o $\mathbf{x}_0$. To snadno dokážeme:

$$\alpha_1(\mathbf{x}_1-\mathbf{x}_0)+\cdots+\alpha_k(\mathbf{x}_k-\mathbf{x}_0) = \alpha_1\mathbf{x}_1+\cdots+\alpha_k\mathbf{x}_k-(\alpha_1+\cdots+\alpha_k)\mathbf{x}_0 = (\alpha_1\mathbf{x}_1+\cdots+\alpha_k\mathbf{x}_k)-\mathbf{x}_0.$$

Naproti tomu obecná lineární kombinace na počátku závisí.

Neformálně lze tedy afinní podprostor vnímat jako 'podprostor, u něhož jsme zapomněli počátek'. Proto si jeho prvky stačí představovat/kreslit jako body (puntíky). Poloha počátku není důležitá, protože afinní kombinace na něm nezávisí. Afinní kombinaci bodů na papíře lze sestrojit pomocí pravítka a měřítka bez znalosti polohy počátku. Naproti tomu, jestliže s prvky lineárního (pod)prostoru provádíme operace lineární kombinace, tyto prvky si představujeme jako šipky spojující počátek s koncovým bodem (přesněji jako *volné vektory*), protože lineární kombinaci vektorů můžeme sestrojit pouze se znalostí polohy počátku. To je důvod, proč se prvkům afinního (pod)prostoru říká *body*, zatímto prvkům pouhého lineárního (pod)prostoru *vektory*. Často lze ovšem prostor, ve kterém pracujeme (obvykle $\mathbb{R}^n$), nazírat obojím způsobem.

Příklad 3.3. Afinní obal dvou bodů $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^3$ je množina

$$\operatorname{aff}\{\mathbf{x}_1, \mathbf{x}_2\} = \{\, \alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 \mid \alpha_1 + \alpha_2 = 1 \,\}.$$

Pokud $\mathbf{x}_1 \neq \mathbf{x}_2$, je touto množinou přímka procházející body $\mathbf{x}_1, \mathbf{x}_2$. Tato přímka je afinní podprostor $\mathbb{R}^3$. Na obrázku dole vlevo je několik bodů na přímce a příslušné koeficienty (α_1, α_2).

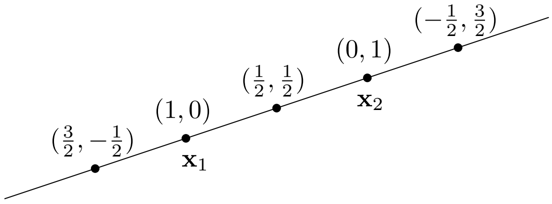
Naproti tomu, lineární obal dvou vektorů $\mathbf{x}_1, \mathbf{x}_2$ (pokud jsou lineárně nezávislé) je rovina procházející těmito dvěma body a počátkem $\mathbf{0}$.

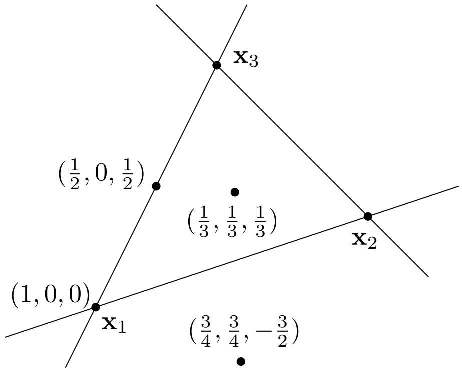
Afinní obal tří bodů $\mathbf{x}_1, \mathbf{x}_2. \mathbf{x}_3 \in \mathbb{R}^3$ je množina

$$\operatorname{aff}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\} = \{\, \alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + \alpha_3\mathbf{x}_3 \mid \alpha_1 + \alpha_2 + \alpha_3 = 1 \,\}.$$

[6] Všimněte si, že definujeme afinní *podprostor* lineárního prostoru, ale už ne afinní *prostor* sám o sobě. Definice afinního prostoru bez odkazu k nějakému lineárnímu prostoru (tj. pomocí axiomů) existuje, ale neuvádíme ji.

Pokud body $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ neleží v přímce, je touto množinou rovina jimi procházející. Tato rovina je afinní podprostor $\mathbb{R}^3$. Na obrázku vpravo je několik bodů v této rovině a příslušné koeficienty $(\alpha_1, \alpha_2, \alpha_3)$.





(Protože afinní kombinace nezávisí na poloze počátku, do obrázků jsme počátek ani nekreslili a prvky prostoru $\mathbb{R}^3$ jsme kreslili jako body, viz poznámka výše.) ♦

Pro množinu $X \subseteq \mathbb{R}^n$ a vektor $\mathbf{x} \in \mathbb{R}^n$ označme

$$X + \mathbf{x} = \mathbf{x} + X = \{\, \mathbf{x} + \mathbf{y} \mid \mathbf{y} \in X \,\}. \tag{3.23}$$

Věta 3.12.

- Je-li X lineární podprostor $\mathbb{R}^n$ a $\mathbf{x} \in \mathbb{R}^n$, pak množina $X + \mathbf{x}$ je afinní podprostor $\mathbb{R}^n$.
- Je-li A afinní podprostor $\mathbb{R}^n$ a $\mathbf{x} \in A$, pak množina $A - \mathbf{x}$ je lineární podprostor $\mathbb{R}^n$.

Důkaz. První tvrzení: Chceme dokázat, že afinní kombinace bodů z množiny $X + \mathbf{x}$ leží v této množině. Nechť tedy $\alpha_1 + \cdots + \alpha_k = 1$ a chceme dokázat implikaci

$$\mathbf{x}_1, \ldots, \mathbf{x}_k \in X + \mathbf{x} \quad \Longrightarrow \quad \alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k \in X + \mathbf{x},$$

neboli[7]

$$\mathbf{x}_1 - \mathbf{x}, \ldots, \mathbf{x}_k - \mathbf{x} \in X \quad \Longrightarrow \quad \alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k - \mathbf{x} = \alpha_1(\mathbf{x}_1 - \mathbf{x}) + \cdots + \alpha_k(\mathbf{x}_k - \mathbf{x}) \in X.$$

Tato implikace platí díky linearítě X.

Druhé tvrzení: Chceme dokázat, že lineární kombinace vektorů z množiny $A - \mathbf{x}$ leží v této množině. Nechť tedy $\alpha_1, \ldots, \alpha_k \in \mathbb{R}$ a chceme dokázat implikaci

$$\mathbf{x}_1, \ldots, \mathbf{x}_k \in A \quad \Longrightarrow \quad \alpha_1(\mathbf{x}_1 - \mathbf{x}) + \cdots + \alpha_k(\mathbf{x}_k - \mathbf{x}) \in A - \mathbf{x}.$$

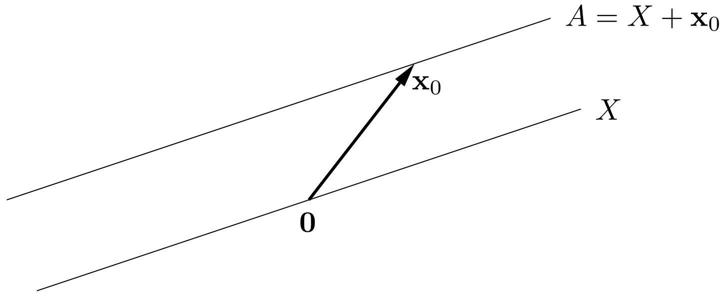
Pravá strana této implikace jde psát jako

$$\alpha_1(\mathbf{x}_1 - \mathbf{x}) + \cdots + \alpha_k(\mathbf{x}_k - \mathbf{x}) + \mathbf{x} = \alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k + (1 - \alpha_1 - \cdots - \alpha_k)\mathbf{x} \in A.$$

To je ale pravda, neboť $\alpha_1 + \cdots + \alpha_k + (1 - \alpha_1 - \cdots - \alpha_k) = 1$, a tedy poslední výraz je afinní kombinace vektorů z A, která podle předpokladu leží v A. ■

[7] Zde používáme skutečnost, že k výroku typu $\mathbf{x} \in X$ můžeme přičíst libovolný vektor a výrok se tím nezmění. Přesněji, pro každé $X \subseteq \mathbb{R}^n$ a $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ platí $\mathbf{x} \in X \;\Leftrightarrow\; \mathbf{x} + \mathbf{y} \in X + \mathbf{y}$. To snadno plyne z (3.23).

Věta 3.12 říká, že afinní podprostor není nic jiného, než 'posunutý' lineární podprostor (tedy nemusí procházet počátkem, na rozdíl of lineárního podprostoru). Třetí tvrzení věty navíc ukazuje, že tento lineární podprostor je afinním prostorem určen jednoznačně:



Dimenze neprázdného[8] **afinního podprostoru** je dimenze tohoto lineárního podprostoru. Afinnímu podprostoru $\mathbb{R}^n$ dimenze 0, 1, 2 a $n-1$ se říká po řadě **bod**, **přímka**, **rovina**, a **nadrovina**.

Věta 3.13. Množina $A \subseteq \mathbb{R}^n$ je afinní podprostor právě tehdy, když je množinou řešení nějaké soustavy lineárních rovnic, tj. existují $\mathbf{A}$ a $\mathbf{b}$ splňující $A = \{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{Ax} = \mathbf{b} \,\}$.

Důkaz. Předpokládejme, že množina A je neprázdná (pro $A = \emptyset$ věta zjevně platí, protože prázdná množina je afinní podprostor a je to také řešení nějaké soustavy lineárních rovnic).

Důkaz $\Leftarrow$: Nechť $A = \{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{Ax} = \mathbf{b} \,\}$. Nechť $\mathbf{x}_1, \dots, \mathbf{x}_k \in A$ a $\alpha_1 + \dots + \alpha_k = 1$. Dokažme, že $\alpha_1\mathbf{x}_1 + \dots + \alpha_k\mathbf{x}_k \in A$:

$$\mathbf{A}(\alpha_1\mathbf{x}_1 + \dots + \alpha_k\mathbf{x}_k) = \alpha_1\mathbf{Ax}_1 + \dots + \alpha_k\mathbf{Ax}_k = \alpha_1\mathbf{b} + \dots + \alpha_k\mathbf{b} = \mathbf{b}.$$

Jiný důkaz $\Leftarrow$: Nechť $X = \operatorname{null}\mathbf{A}$ a $\mathbf{x}_0$ je libovolný bod splňující $\mathbf{Ax}_0 = \mathbf{b}$ (tzv. *partikulární řešení* soustavy $\mathbf{Ax} = \mathbf{b}$). Pak

$$A = \{\, \mathbf{x} \mid \mathbf{Ax} = \mathbf{b} \,\} = \{\, \mathbf{x} \mid \mathbf{A}(\mathbf{x} - \mathbf{x}_0) = \mathbf{0} \,\} = \{\, \mathbf{x}' + \mathbf{x}_0 \mid \mathbf{Ax}' = \mathbf{0} \,\} = X + \mathbf{x}_0. \tag{3.24}$$

Tedy dle Věty 3.12 je A afinní podprostor.

Důkaz $\Rightarrow$: Nechť A je afinní podprostor a nechť $\mathbf{x}_0 \in A$. Pak (dle Věty 3.12) je množina $X = A - \mathbf{x}_0$ lineární podprostor, tedy (viz poznámka v §3.2.1) existuje matice $\mathbf{A}$ tak, že $X = \operatorname{null}\mathbf{A}$. Nechť $\mathbf{b} = \mathbf{Ax}_0$. Pak dle (3.24) je $X + \mathbf{x}_0 = \{\, \mathbf{x} \mid \mathbf{Ax} = \mathbf{b} \,\}$. ∎

Díky Větě 3.13 se afinnímu podprostoru říká také *lineární varieta*. Je to speciální případ *algebraické variety*, což je množina řešení soustavy polynomiálních rovnic. Jejich studiem se zabývá *algebraická geometrie*.

Shrňme: každý lineární podprostor lze reprezentovat buď jako $\operatorname{rng}\mathbf{B}$ pro nějakou matici $\mathbf{B}$, nebo jako $\operatorname{null}\mathbf{A}$ (tj. jako množinu řešení soustavy $\mathbf{Ax} = \mathbf{0}$) pro nějakou (jinou!) matici $\mathbf{A}$. Každý afinní podprostor lze reprezentovat buď jako $\mathbf{x} + X$ pro nějaký vektor $\mathbf{x}$ a lineární podprostor X, nebo jako množinu řešení soustavy $\mathbf{Ax} = \mathbf{b}$ pro nějaké $\mathbf{A}, \mathbf{b}$.

Body $\mathbf{x}_1, \dots, \mathbf{x}_k \in \mathbb{R}^n$ jsou **afinně nezávislé**, jestliže žádný není afinní kombinací ostatních. Afinní nezávislost lze ovšem definovat i jinak (důkaz věty neuvádíme, i když není těžký):

[8] Na rozdíl od lineárního podprostoru může být afinní podprostor prázdný. Dimenze prázdného afinního podprostoru se někdy konvencí definuje jako -1.

Věta 3.14. Pro body $\mathbf{x}_1, \dots, \mathbf{x}_k \in \mathbb{R}^n$ jsou následující výroky ekvivalentní:

1. Žádný bod není roven afinní kombinaci ostatních bodů.
2. Platí (srov. se (3.2))
$$\alpha_1 + \cdots + \alpha_k = 0, \quad \alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k = \mathbf{0} \quad \Longrightarrow \quad \alpha_1 = \cdots = \alpha_k = 0. \tag{3.25}$$
3. Vektory $\mathbf{x}_2 - \mathbf{x}_1, \mathbf{x}_3 - \mathbf{x}_1, \dots, \mathbf{x}_k - \mathbf{x}_1$ jsou lineárně nezávislé.
4. Vektory $\begin{bmatrix}\mathbf{x}_1\\1\end{bmatrix}, \dots, \begin{bmatrix}\mathbf{x}_k\\1\end{bmatrix}$ jsou lineárně nezávislé (kde $\begin{bmatrix}\mathbf{x}\\1\end{bmatrix} \in \mathbb{R}^{n+1}$ značí vektor $\mathbf{x}$ s přidanou jedničkou jako $(n+1)$-ní souřadnicí[9]).

Příklad 3.4. Dva body $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$ jsou afinně nezávislé, právě když $\mathbf{x}_1 \neq \mathbf{x}_2$ (neboli nejsou identické). Tři body $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3 \in \mathbb{R}^n$ jsou afinně nezávislé, právě když neleží v jedné přímce (neboli nejsou *kolineární*). Viz Příklad 3.3. Čtyři body $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4 \in \mathbb{R}^n$ jsou afinně nezávislé, právě když neleží v jedné rovině (neboli nejsou *koplanární*). ♦

Zobrazení $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$ nazveme **afinní**, pokud (3.5) platí pro všechna $\alpha_1+\cdots+\alpha_k = 1$. Lze dokázat (proveďte!), že zobrazení $\mathbf{f}$ je afinní, právě když existuje matice $\mathbf{A} \in \mathbb{R}^{m\times n}$ a vektor $\mathbf{b} \in \mathbb{R}^m$ tak, že
$$\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}, \tag{3.26}$$
Pro $m = 1$ se zobrazení (3.26) nazývá také **afinní funkce**[10] $f\colon \mathbb{R}^n \to \mathbb{R}$ a má tvar
$$f(\mathbf{x}) = \mathbf{a}^T\mathbf{x} + b = a_1x_1 + \cdots + a_nx_n + b, \tag{3.27}$$
kde $\mathbf{a} \in \mathbb{R}^n$ a $b \in \mathbb{R}$.

Příklad 3.5. Zobrazení $\mathbf{f}\colon \mathbb{R}^2 \to \mathbb{R}^3$ definované jako $\mathbf{f}(x_1, x_2) = (x_1 + x_2 + 1, x_1 - x_2 - 2, 2x_1)$ je afinní. To bychom mohli dokázat ověřením podmínek (3.5) pro $\alpha_1 + \cdots + \alpha_k = 1$. Ale je to patrné i z toho, že zobrazení lze vyjádřit ve tvaru (3.26):
$$\mathbf{f}(x_1, x_2) = \begin{bmatrix}1 & 1\\1 & -1\\2 & 0\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} + \begin{bmatrix}1\\-2\\0\end{bmatrix}.$$
♦

Příklad 3.6. Často potkáme zobrazení ve tvaru
$$\mathbf{f}(\mathbf{x}) = \mathbf{A}(\mathbf{x} - \mathbf{x}_0). \tag{3.28}$$
To je afinní zobrazení (3.26), kde $\mathbf{b} = -\mathbf{A}\mathbf{x}_0$. Zatímco v (3.26) eplicitně vidíme posunutí v hodnotě (vektor $\mathbf{b}$), v (3.28) explicitně vidíme posunutí v argumentu. ♦

[9] Tento vektor dobře znají počítačoví grafici jako *homogenní souřadnice* bodu $\mathbf{x}$.

[10] V lineární algebře znamená ‘lineární funkce’ něco jiného než v matematické analýze. Např. funkci jedné proměnné $f(x) = ax + b$ znáte ze základní školy jako lineární, v lineární algebře však lineární není – je afinní. Ovšem soustavě rovnic $\mathbf{A}\mathbf{x} = \mathbf{b}$ se říká ‘lineární’ i v lineární algebře.

Příklad 3.7. Pro dané body $\mathbf{x}_1, \ldots, \mathbf{x}_k \in \mathbb{R}^n$ a $\mathbf{y}_1, \ldots, \mathbf{y}_k \in \mathbb{R}^m$ máme za úkol najít afinní zobrazení $\mathbf{f}$ takové, aby platilo $\mathbf{y}_i = \mathbf{f}(\mathbf{x}_i)$ pro všechna $i = 1, \ldots, k$. Řešíme tedy lineární soustavu $\mathbf{y}_i = \mathbf{A}\mathbf{x}_i + \mathbf{b}$, $i = 1, \ldots, k$, pro neznámé $\mathbf{A} \in \mathbb{R}^{m \times n}$ a $\mathbf{b} \in \mathbb{R}^m$. Tuto soustavu můžeme napsat v maticovém tvaru

$$\begin{bmatrix}\mathbf{y}_1 & \cdots & \mathbf{y}_k\end{bmatrix} = \begin{bmatrix}\mathbf{A} & \mathbf{b}\end{bmatrix} \begin{bmatrix}\mathbf{x}_1 & \cdots & \mathbf{x}_k \\ 1 & \cdots & 1\end{bmatrix}.$$

Je-li $k = n + 1$ a body $\mathbf{x}_1, \ldots, \mathbf{x}_k$ jsou afinně nezávislé, matice $\begin{bmatrix}\mathbf{x}_1 & \cdots & \mathbf{x}_k \\ 1 & \cdots & 1\end{bmatrix}$ je regulární, tedy soustava má jediné řešení. ♦

Shrňme: každé lineární zobrazení lze reprezentovat jako $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x}$ pro nějakou $\mathbf{A}$ a každé afinní zobrazení jako $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$ pro nějaké $\mathbf{A}, \mathbf{b}$.

3.4 Cvičení

3.1. Rozhodněte, zda následující množiny jsou lineární nebo afinní podprostory $\mathbb{R}^n$ a když ano, určete jejich dimenze:

a) $\{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}^T\mathbf{x} = 0 \,\}$ pro dané $\mathbf{a} \in \mathbb{R}^n$

b) $\{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}^T\mathbf{x} = b \,\}$ pro dané $\mathbf{a} \in \mathbb{R}^n$, $b \in \mathbb{R}$

c) $\{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{x}^T\mathbf{x} = 1 \,\}$

d) $\{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}\mathbf{x}^T = \mathbf{I} \,\}$ pro dané $\mathbf{a} \in \mathbb{R}^n$

e) $\left\{\, \mathbf{x} \in \mathbb{R}^n \;\middle|\; \sum_{i=1}^n x_i = 0 \,\right\}$

f) $\mathbf{a} + \operatorname{span}\{\mathbf{b}, \mathbf{c}\}$, kde $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^n$ jsou známé vektory takové, že $\mathbf{b}, \mathbf{c}$ jsou lineárně nezávislé.

3.2. Je množina $\{\, (x_1, x_2, x_3, x_4) \in \mathbb{R}^4 \mid x_1 + x_3 = 0 \,\}$ lineární podprostor? Pokud ano, najděte jeho libovolnou bázi.

3.3. Máme podprostor $X \subseteq \mathbb{R}^3$ s bází $\{(1, 2, 3), (-1, 0, 1)\}$. Nechť $\mathbf{x} = (2, 2, 2)$. Platí $\mathbf{x} \in X$? Pokud ano, najděte souřadnice vektoru $\mathbf{x}$ v této bázi.

3.4. Dokažte, že jestliže jsou vektory $\mathbf{x}_1, \ldots, \mathbf{x}_k$ lineárně závislé, pak je aspoň jeden z nich lineární kombinací ostatních.

3.5. Nechť X je lineární podprostor $\mathbb{R}^n$ a $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$ je lineární zobrazení. Je množina $\mathbf{f}(X)$ lineární podprostor $\mathbb{R}^m$? Odpověď dokažte.

3.6. Je dáno zobrazení $\mathbf{f}(\mathbf{x}) = \mathbf{x} \times \mathbf{y}$, kde $\mathbf{y} \in \mathbb{R}^3$ je pevný vektor a $\times$ označuje vektorový součin dvou vektorů[11]. Jde tedy o zobrazení $\mathbb{R}^3 \to \mathbb{R}^3$. Je toto zobrazení lineární? Pokud ano, najděte matici $\mathbf{A}$ tak, aby $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x}$. Čemu je rovno $\mathbf{A}^T$? Jakou hodnost má $\mathbf{A}$?

3.7. Máme zobrazení $\mathbf{f}\colon \mathbb{R}^2 \to \mathbb{R}^3$ definované jako $\mathbf{f}(x, y) = (x + y,\, 2x - 1,\, x - y)$. Je toto zobrazení lineární? Pokud ano, napište ho ve formě (3.7). Je toto zobrazení afinní? Pokud ano, napište ho ve formě (3.26). Obě odpovědi dokažte z definic.

[11] Tedy $\times$ zde neoznačuje kartézský součin množin, jako jinde ve skriptech.

3.8. Máme nehomogenní soustavu lineárních rovnic

$$\begin{aligned} x + 2y + z &= 1 \\ -x + y + 2z &= 2 \end{aligned}$$

dvou rovnic o třech neznámých. Napište množinu řešení soustavy jako $X + \mathbf{x}_0$, kde $X \subseteq \mathbb{R}^3$ je lineární podprostor (napište jeho bázi) a $\mathbf{x}_0 \in \mathbb{R}^3$.

3.9. Zjistěte, zda existuje lineární funkce f splňující tyto podmínky:

a) $f(1,2) = 2,\ f(3,4) = 3$.
b) $f(1,2) = 2,\ f(3,4) = 3,\ f(5,6) = 4$.
c) $f(1,0,1) = -1,\ f(0,1,2) = 1,\ f(1,1,3) = 2$.

3.10. Najděte bázi prostoru obrazů a bázi nulového prostoru následujících lineárních zobrazení:

a) $\mathbf{f}(x_1, x_2, x_3) = (x_1 - x_2,\ x_2 - x_3 + 2x_1)$
b) $\mathbf{f}(x_1, x_2) = (2x_1 + x_2,\ x_1 - x_2,\ 2x_2 + x_1)$

3.11. Tvrdíme, že sloupce matice $\mathbf{A} \in \mathbb{R}^{n\times m}$ jsou báze podprostoru $X \subseteq \mathbb{R}^n$, právě když $\dim X = m$ a $X = \operatorname{rng} \mathbf{A}$. Je to pravda? Proč?

3.12. Nechť matice $\mathbf{A}$ a $\mathbf{B}$ mají stejné rozměry a obě mají lineárně nezávislé sloupce. Dokažte, že sloupce matic tvoří báze stejného podprostoru (tj. $\operatorname{rng} \mathbf{A} = \operatorname{rng} \mathbf{B}$), právě když $\mathbf{B} = \mathbf{AC}$ pro nějakou regulární matici $\mathbf{C}$.

3.13. Pro matice $\mathbf{A}, \mathbf{B}$ se stejným počtem řádků dokažte tvrzení:

a) $\operatorname{rng} \mathbf{A} \subseteq \operatorname{rng}[\mathbf{A}\ \mathbf{B}]$
b) $\operatorname{rng} \mathbf{A} = \operatorname{rng}[\mathbf{A}\ \mathbf{B}]$ jestliže $\operatorname{rng} \mathbf{B} \subseteq \operatorname{rng} \mathbf{A}$
c) $\operatorname{rng} \mathbf{A} = \operatorname{rng}[\mathbf{A}\ \mathbf{B}]$ jestliže $\operatorname{rank} \mathbf{A} = \operatorname{rank}[\mathbf{A}\ \mathbf{B}]$
d) $\operatorname{rng} \mathbf{A} = \operatorname{rng} \mathbf{B}$ právě když $\operatorname{rank} \mathbf{A} = \operatorname{rank}[\mathbf{A}\ \mathbf{B}] = \operatorname{rank} \mathbf{B}$

3.14. Máme matici $\mathbf{A}$ a vektor $\mathbf{x} \neq \mathbf{0}$ takové, že $\mathbf{Ax} = \mathbf{0}$. Může mít matice $\mathbf{A}$ lineárně nezávislé řádky? Může mít lineárně nezávislé sloupce? Odpovědi dokažte z definice lineární nezávislosti.

3.15. Dokažte, že pro libovolné dvě matice platí nerovnost (3.19).

3.16. Navrhněte postup, jak spočítat afinní zobrazení $\mathbf{f}\colon \mathbb{R}^2 \to \mathbb{R}^2$, které zobrazí trojúhelník s vrcholy $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3 \in \mathbb{R}^2$ na trojúhelník s vrcholy $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3 \in \mathbb{R}^2$. Zobrazení má zobrazit bod $\mathbf{p}_1$ do bodu $\mathbf{q}_1$ atd.

3.17. Které z těchto výroků jsou pravdivé? Každý výrok dokažte nebo najděte protipříklad. Některé výroky mohou platit jen pro určité rozměry matic – najděte co nejobecnější podmínky na rozměry matic, aby výroky byly pravdivé.

a) Pokud $\mathbf{AB}$ má plnou hodnost, pak $\mathbf{A}$ a $\mathbf{B}$ mají plnou hodnost.
b) Pokud $\mathbf{A}$ a $\mathbf{B}$ mají plnou hodnost, pak $\mathbf{AB}$ má plnou hodnost.
c) Pokud $\mathbf{A}$ a $\mathbf{B}$ mají triviální nulový prostor, pak $\mathbf{AB}$ má triviální nulový prostor.
d) ($\star$) Pokud $\mathbf{A}$ a $\mathbf{B}$ jsou úzké s plnou hodností a platí $\mathbf{A}^T\mathbf{B} = \mathbf{0}$, pak matice $[\mathbf{A}\ \mathbf{B}]$ je úzká s plnou hodností.
e) ($\star$) Pokud matice $\begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix}$ má plnou hodnost, pak $\mathbf{A}$ i $\mathbf{B}$ mají plnou hodnost.

3.18. V §2.1.4 jsme zmínili a v §3.2.3 dokázali, že jestliže nějaké matice $\mathbf{A}, \mathbf{B}$ splňují $\mathbf{AB} = \mathbf{I}$, pak $\mathbf{A}$ má lin. nezávislé řádky a $\mathbf{B}$ má lin. nezávislé sloupce. Dokažte obecnější tvrzení: jestliže matice $\mathbf{AB}$ je regulární, pak $\mathbf{A}$ má lin. nezávislé řádky a $\mathbf{B}$ má lin. nezávislé sloupce.

3.19. Chceme vynásobit dvě matice, kde první má velikost $m \times n$ a hodnost r a druhá má velikost $n \times k$ a hodnost s. Jak byste to udělali s co nejmenší výpočetní složitostí? Použijte rank factorization.

3.20. ($\star$) Nechť X, Y jsou podprostory $\mathbb{R}^n$. Definujme $X + Y = \{\, \mathbf{x} + \mathbf{y} \mid \mathbf{x} \in X,\ \mathbf{y} \in Y \,\}$ (tzv. *Minkowského součet* vektorů). Dokažte:

   a) $X \subseteq Y \ \Rightarrow\ X^\perp \supseteq Y^\perp$.
   b) $X + Y$ je generován sjednocením libovolné báze X a libovolné báze Y. Tedy je-li $\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ báze X a $\{\mathbf{y}_1, \ldots, \mathbf{y}_l\}$ báze Y, pak $X+Y = \operatorname{span}\{\mathbf{x}_1, \ldots, \mathbf{x}_k, \mathbf{y}_1, \ldots, \mathbf{y}_l\}$.
   c) $(X + Y)^\perp = X^\perp \cap Y^\perp$.
   d) $(X \cap Y)^\perp = X^\perp + Y^\perp$

3.21. ($\star$) Mějme vektory $\mathbf{x}_1, \ldots, \mathbf{x}_k$. Vektor $\mathbf{x}_i$ nazveme *klíčový*, není-li lineární kombinací vektorů $\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}$. Dokažte, že množina klíčových vektorů je báze podprostoru $\operatorname{span}\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$. Všimněte si, že vlastně dokazujete první tvrzení Věty 3.2.

3.22. ($\star$) Máme lineárně nezávislé vektory $\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^n$ a hledáme vektory $\mathbf{a}_{m+1}, \ldots, \mathbf{a}_n \in \mathbb{R}^n$ tak, aby vektory $\mathbf{a}_1, \ldots, \mathbf{a}_n$ byly lineárně nezávislé. Dokažte, že to jde udělat následovně. Nechť $\mathbf{A} \in \mathbb{R}^{m \times n}$ je matice s řádky $\mathbf{a}_1^T, \ldots, \mathbf{a}_m^T$. Vybereme množinu $J \subseteq \{1, \ldots, n\}$ lineárně nezávislých sloupců matice $\mathbf{A}$ (viz předchozí cvičení). Pak zvolíme vektory $\mathbf{a}_{m+1}, \ldots, \mathbf{a}_n$ jako $n - m$ vektorů standardní báze $\mathbf{e}_j \in \mathbb{R}^n$ kde $j \in \{1, \ldots, n\} \setminus J$.

3.23. ($\star$) Dokažte Větu 3.14.

Nápověda a řešení

3.1.a) Lineární podprostor, dimenze $n - 1$ pro $\mathbf{a} \neq \mathbf{0}$ a n pro $\mathbf{a} = \mathbf{0}$.

3.1.b) Afinní podprostor dimenze $n - 1$ pro $\mathbf{a} \neq \mathbf{0}$ a n pro $\mathbf{a} = \mathbf{0}$, $b = 0$. Pro $\mathbf{a} = \mathbf{0}$, $b \neq 0$ je množina prázdná (tedy není afinní podprostor).

3.1.c) Není lineární ani afinní podprostor (je to sféra).

3.1.d) Pro $n = 1$ a $\mathbf{a} \neq \mathbf{0}$ je množinou jediný bod, tedy afinní podprostor prostoru $\mathbb{R}$. Pro $n = 1$ a $\mathbf{a} = \mathbf{0}$ je množina prázdná (tedy není afinní podprostor). Pro $n > 1$ je množina také prázdná, protože soustava $\mathbf{a}\mathbf{x}^T = \mathbf{I}$ nemá řešení pro žádné $\mathbf{a}, \mathbf{x}$ (možný důkaz: je $\operatorname{rank} \mathbf{I} = n$, ale $\operatorname{rank}(\mathbf{a}\mathbf{x}^T) \leq 1$).

3.1.e) Lineární podprostor dimenze $n - 1$.

3.1.f) Je to rovina v $\mathbb{R}^n$, která prochází prochází body $\mathbf{a}$, $\mathbf{a} + \mathbf{b}$ a $\mathbf{a} + \mathbf{c}$. Tyto tři body jsou různé (protože vektory $\mathbf{b}$ a $\mathbf{c}$ jsou lineárně nezávislé), tedy rovinu definují. Tato rovina je afinní (pro $\mathbf{a} = \mathbf{0}$ dokonce lineární) podprostor dimenze 2.

3.2. Ano. Je to vlastně množina $\{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}^T\mathbf{x} = 0 \,\}$ (tedy nadrovina procházející počátkem s normálovým vektorem $\mathbf{a}$), kde $n = 4$ a $\mathbf{a} = (1, 0, 1, 0)$. Bázi získáme řešením homogenní soustavy $\mathbf{a}^T\mathbf{x} = x_1 + x_3 = 0$. Báze má tři prvky, např. $(1, 0, -1, 0)$, $(0, 1, 0, 0)$, $(0, 0, 0, 1)$.

3.3. Je $\mathbf{x} \in X$, právě když $\mathbf{x} = \alpha\mathbf{a} + \beta\mathbf{b}$ pro nějaké $\alpha, \beta \in \mathbb{R}$. Pokud tato soustava lineárních rovnic má řešení, pak (α, β) jsou souřadnice vektoru $\mathbf{x}$ v bázi $(\mathbf{a}, \mathbf{b})$ (zde je drobná nesrovnalost: někdy jsme o bázi mluvili jako o *množině* vektorů, protože nám na pořadí vektorů báze nezáleželo.

3.4. Jsou-li vektory lin. závislé, pak $\alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k = \mathbf{0}$ kde aspoň jedno z čísel $\alpha_1, \ldots, \alpha_k$ je nenulové. Je-li např. $\alpha_1 \neq 0$, pak $\alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k = \mathbf{0}$ lze napsat jako $\mathbf{x}_1 = \alpha'_2\mathbf{x}_2 + \cdots + \alpha'_k\mathbf{x}_k$ kde $\alpha'_i = \alpha_i/\alpha_1$.

3.5. Ano. Důkaz: Dle (3.3) máme dokázat, že pro každé $\mathbf{y}_1, \ldots, \mathbf{y}_k \in \mathbf{f}(X)$ platí $\sum_i \alpha_i\mathbf{y}_i \in \mathbf{f}(X)$. Protože $\mathbf{f}(X) = \{\, \mathbf{f}(\mathbf{x}) \mid \mathbf{x} \in X \,\}$, je $\mathbf{y} \in \mathbf{f}(X)$, právě když existuje $\mathbf{x} \in X$ tak, že $\mathbf{y} = \mathbf{f}(\mathbf{x})$. Tedy máme dokázat, že pro každé $\mathbf{x}_1, \ldots, \mathbf{x}_k \in X$ existuje $\mathbf{x} \in X$ tak, že $\sum_i \alpha_i\mathbf{f}(\mathbf{x}_i) = \mathbf{f}(\mathbf{x})$. To je ale pravda, protože $\sum_i \alpha_i\mathbf{f}(\mathbf{x}_i) = \mathbf{f}(\sum_i \alpha_i\mathbf{x}_i)$ a $\mathbf{x} = \sum_i \alpha_i\mathbf{x}_i \in X$.

3.6. Je lineární, $\mathbf{A} = \begin{bmatrix} 0 & y_3 & -y_2 \\ -y_3 & 0 & y_1 \\ y_2 & -y_1 & 0 \end{bmatrix}$ je antisymetrická hodnosti 2 pro $\mathbf{y} \neq \mathbf{0}$.

3.7. Je afinní. Je $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$, kde $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 2 & 0 \\ 1 & -1 \end{bmatrix}$ $\mathbf{b} = (0, -1, 0) = \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}$, $\mathbf{x} = (x, y) = \begin{bmatrix} x \\ y \end{bmatrix}$.

3.8. Např. $(1, -1, 2) + \operatorname{span}\{(1, -1, 1)\} = \{\, (1 + \alpha, -1 - \alpha, 2 + \alpha) \mid \alpha \in \mathbb{R} \,\}$.

3.9.a) Je $f(x_1, x_2) = a_1x_1 + a_2x_2$. Řešíme soustavu $a_1 + 2a_2 = 2$, $3a_1 + 4a_2 = 3$. Tato soustava má řešení, tedy lineární funkce existuje.

3.9.b) Ano.

3.9.c) Ne.

3.10.a) $\operatorname{rng}\mathbf{A} = \mathbb{R}^2$, tedy báze $\operatorname{rng}\mathbf{A}$ je např. $(0, 1), (1, 0)$. Báze $\operatorname{null}\mathbf{A}$ je např. $(1, 1, 3)$.

3.10.b) Báze $\operatorname{rng}\mathbf{A}$ je např. $(2, 1, 1), (1, -1, 2)$. Báze $\operatorname{null}\mathbf{A}$ je např. $(0, 0)$.

3.12. Implikace $\Leftarrow$: Nechť $\mathbf{C}$ je regulární, tedy má l.n. řádky. Pak dle Tvrzení 3.10 je $\operatorname{rng}\mathbf{B} = \operatorname{rng}(\mathbf{AC}) = \operatorname{rng}\mathbf{A}$.

Implikace $\Rightarrow$: Nechť $\mathbf{A}, \mathbf{B}$ mají l.n. sloupce a $\operatorname{rng}\mathbf{A} = \operatorname{rng}\mathbf{B}$. To znamená, že sloupce $\mathbf{A}, \mathbf{B}$ tvoří dvě různé báze toho samého podprostoru. Matice $\mathbf{C}$ je tedy matice přechodu mezi těmito dvěma bázemi. Máme dokázat (což byste měli vědět z lin. algebry), že matice přechodu existuje a je regulární. Abychom ji našli, řešíme rovnici $\mathbf{B} = \mathbf{AC}$ pro neznámou $\mathbf{C}$. Tu rozepíšeme po sloupcích jako $\mathbf{a}_j = \mathbf{B}\mathbf{c}_j$ kde $\mathbf{a}_j, \mathbf{c}_j$ je j-tý sloupec matice $\mathbf{A}, \mathbf{C}$. Každá tato rovnice má jediné řešení pro $\mathbf{c}_j$, protože $\mathbf{a}_j \in \operatorname{rng}\mathbf{B}$ (což plyne z $\operatorname{rng}\mathbf{A} = \operatorname{rng}\mathbf{B}$) a $\operatorname{null}\mathbf{B} = \{\mathbf{0}\}$ (protože $\mathbf{B}$ má l.n. sloupce). Zbývá dokázat, že $\mathbf{C}$ je regulární: protože $\operatorname{rank}\mathbf{A} = \operatorname{rank}\mathbf{B} = n$ kde n je počet sloupců matic $\mathbf{A}, \mathbf{B}, \mathbf{C}$, dle (3.19) nemůže být $\operatorname{rank}\mathbf{C} < n$.

3.13.a) Víme (viz §3.2.1), že množina $\operatorname{rng}\mathbf{A}$ je lineární obal sloupců matice $\mathbf{A}$. Inkluze plyne okamžitě z toho, že $\operatorname{rng}[\mathbf{A}\ \mathbf{B}]$ je lineární obal sloupců matice $\mathbf{A}$ a navíc sloupců matice $\mathbf{B}$.

3.13.b) Předpokládejme $\operatorname{rng}\mathbf{B} \subseteq \operatorname{rng}\mathbf{A}$. Dle definice prostoru obrazů (3.13) to znamená, že pro každé $\mathbf{y}$ existuje $\mathbf{x}$ tak, že $\mathbf{B}\mathbf{y} = \mathbf{A}\mathbf{x}$. Máme dokázat $\operatorname{rng}\mathbf{A} = \operatorname{rng}[\mathbf{A}\ \mathbf{B}]$, tj. že pro každé $\mathbf{u}, \mathbf{v}$ existuje $\mathbf{z}$ tak, že $[\mathbf{A}\ \mathbf{B}]\begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} = \mathbf{A}\mathbf{u} + \mathbf{B}\mathbf{v} = \mathbf{A}\mathbf{z}$. Z předpokladu víme, že existuje $\mathbf{x}$ tak, že $\mathbf{B}\mathbf{v} = \mathbf{A}\mathbf{x}$. Tedy zbývá dokázat, že pro každé $\mathbf{u}$ existuje $\mathbf{z}$ tak, že $\mathbf{A}\mathbf{u} + \mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{z}$. To platí: položíme $\mathbf{z} = \mathbf{u} + \mathbf{x}$.

3.13.c) Označme pro přehlednost $X = \operatorname{rng}\mathbf{A}$ a $Z = \operatorname{rng}[\mathbf{A}\ \mathbf{B}]$. Z podúkolu (a) víme, že $X \subseteq Z$. Protože $\dim X = \dim\operatorname{rng}\mathbf{A} = \operatorname{rank}\mathbf{A} = \operatorname{rank}[\mathbf{A}\ \mathbf{B}] = \dim\operatorname{rng}[\mathbf{A}\ \mathbf{B}] = \dim Z$ (kde jsme užili definici hodnosti (3.16)), dle Tvrzení 3.3 máme $X = Z$.

3.13.d) Plyne snadno z podúkolů (a) a (b). Označme $X = \operatorname{rng}\mathbf{A}$, $Y = \operatorname{rng}\mathbf{B}$, $Z = \operatorname{rng}[\mathbf{A}\ \mathbf{B}]$. Máme dokázat, že $X = Y \iff \dim X = \dim Z = \dim Y$.

Důkaz implikace $\Rightarrow$: Z podúkolu (b) víme, že $Y \subseteq X \Rightarrow X = Z$. Jestliže $X = Y$, pak tedy $X = Z = Y$ a tedy i $\dim X = \dim Z = \dim Y$.

Důkaz implikace $\Leftarrow$: jestliže $\dim X = \dim Z = \dim Y$, pak z podúkolu (a) máme $X = Z = Y$.

3.14. Víme, že vektor $\mathbf{Ax}$ je lineární kombinace sloupců matice $\mathbf{A}$ s koeficienty $\mathbf{x}$. Definice lineární nezávislosti (3.2) říká, že lineární nezávislost sloupců implikuje $\mathbf{x} = \mathbf{0}$. Protože $\mathbf{x} \neq \mathbf{0}$, sloupce musí být lineárně závislé.

Řádky ovšem mohou být lineárně nezávislé, příkladem je $\mathbf{A} = [1\ 1]$ a $\mathbf{x} = (1, -1)$.

3.15. Dle Tvrzení 3.10 je $\operatorname{rng}(\mathbf{AB}) \subseteq \operatorname{rng} \mathbf{A}$, z toho $\operatorname{rank}(\mathbf{AB}) \leq \operatorname{rank} \mathbf{A}$. Dle Věty 3.7 je $\operatorname{rank}(\mathbf{AB}) = \operatorname{rank}((\mathbf{AB})^T) = \operatorname{rank}(\mathbf{B}^T\mathbf{A}^T) \leq \operatorname{rank}(\mathbf{B}^T) = \operatorname{rank} \mathbf{B}$.

3.16. Viz Příklad 3.7.

3.19. Uděláme rank factorization obou matic a pak násobíme nejdříve dvě prostřední matice. Složitost bude $rns + \min\{mrs + nsk, rsk + mrk\}$ násobení (a o něco méně sčítání).

3.20.c) Plyne z (b).

3.20.d) Plyne z (c) s použitím $(X^\perp)^\perp = X$.

Kapitola 4

Ortogonalita

4.1 Délky, úhly, vzdálenosti

Na prostoru $\mathbb{R}^n$ přirozeně máme *standardní skalární součin*

$$\mathbf{x}^T\mathbf{y} = x_1y_1 + \cdots + x_ny_n = \mathbf{y}^T\mathbf{x}. \tag{4.1}$$

Ten už jsme uvedli v (2.12) a pro matice v §2.1.7, kde jsme ho značili $\langle \mathbf{x}, \mathbf{y} \rangle$. Značení $\langle \cdot, \cdot \rangle$ je nutné, když jde o obecný (axiomy definovaný) skalární součin na obecném (axiomy definovaném) lineárním prostoru (který se pak nazývá *lineární prostor se skalárním součinem*)[1], na prostoru $\mathbb{R}^n$ ale můžeme psát jednoduše $\mathbf{x}^T\mathbf{y}$. Standardní skalární součin indukuje **eukleidovskou normu**

$$\|\mathbf{x}\| = \sqrt{\mathbf{x}^T\mathbf{x}} = (x_1^2 + \cdots + x_n^2)^{1/2}. \tag{4.2}$$

Norma měří *délku* vektoru $\mathbf{x}$. *Úhel* φ dvojice vektorů se spočítá ze vzorce

$$\cos\varphi = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\,\|\mathbf{y}\|}. \tag{4.3}$$

Vektory jsou navzájem **ortogonální**[2], jestliže $\mathbf{x}^T\mathbf{y} = 0$, což značíme také $\mathbf{x} \perp \mathbf{y}$. Eukleidovská norma indukuje **eukleidovskou metriku**

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|, \tag{4.4}$$

která měří *vzdálenost* bodů $\mathbf{x}$ a $\mathbf{y}$.

Protože pro $n = 3$ takto definované pojmy délky, úhlu a vzdálenosti dobře modelují prostor, ve kterém žijeme, prostoru $\mathbb{R}^n$ se standardním skalárním součinem se často říká *Eukleidovský prostor*.

[1] Standardní skalární součin na $\mathbb{R}^n$, definovaný vzorcem (4.1), se anglicky obvykle nazývá *dot product*, zatímco obecný skalární součin na obecném lineárním prostoru je *inner product*.

[2] *Ortogonální vektory* znamená skoro totéž jako *kolmé vektory* (angl. *perpendicular*). Relace ortogonality se definuje pro dvojice prvků z libovolného lineárního prostoru (tedy např. pro dvojice funkcí, posloupností, atd.). Relace kolmosti se původně definovala v geometrii pro dvojice *přímek*. Je možné ji rozšířit na dvojice *nenulových* vektorů $\mathbf{x}, \mathbf{y}$, které vlastně chápeme jako přímky span$\{\mathbf{x}\}$ a span$\{\mathbf{y}\}$. Vektory ale musejí být nenulové, neboť span$\{\mathbf{0}\}$ není přímka. Naproti tomu nulový vektor $\mathbf{0}$ je ortogonální s každým vektorem, neboť $\mathbf{x}^T\mathbf{0} = 0$ pro každé $\mathbf{x}$.

4.2 Ortogonální podprostory

Vektor $\mathbf{y} \in \mathbb{R}^n$ je **ortogonální** k podprostoru $X \subseteq \mathbb{R}^n$, je-li $\mathbf{y} \perp \mathbf{x}$ pro každé $\mathbf{x} \in X$. Značíme $\mathbf{y} \perp X$ nebo $X \perp \mathbf{y}$. Pro testování této podmínky stačí ověřit, že $\mathbf{y}$ je ortogonální na každý bázový vektor podprostoru X, neboť (dokažte!)

$$\mathbf{y} \perp \operatorname{span}\{\mathbf{x}_1, \dots, \mathbf{x}_k\} \iff \mathbf{y} \perp \mathbf{x}_1, \ \dots, \ \mathbf{y} \perp \mathbf{x}_k. \tag{4.5}$$

Podprostory $X, Y \subseteq \mathbb{R}^n$ jsou **ortogonální**, je-li $\mathbf{x} \perp \mathbf{y}$ pro každé $\mathbf{x} \in X$ a $\mathbf{y} \in Y$. Značíme $X \perp Y$ (přičemž zjevně $X \perp Y \Leftrightarrow Y \perp X$). Platí

$$X \perp Y \implies X \cap Y = \{\mathbf{0}\}, \tag{4.6}$$

neboť jediný vektor ortogonální sám k sobě je nulový vektor $\mathbf{0}$.

Ortogonální doplněk podprostoru $X \subseteq \mathbb{R}^n$ je množina

$$X^\perp = \{\, \mathbf{y} \in \mathbb{R}^n \mid \mathbf{y} \perp X \,\} \tag{4.7}$$

všech vektorů z $\mathbb{R}^n$ ortogonálních na podprostor X. Rozdíl mezi tvrzeními $Y \perp X$ a $Y = X^\perp$ je tedy v tom, že v prvním případě Y nemusí obsahovat *všechny* vektory kolmé na X a ve druhém případě ano. Množina (4.7) je podprostor prostoru $\mathbb{R}^n$ (dokažte!).

Příklad 4.1. Následující příklady jsou v prostoru $\mathbb{R}^3$. Dvě k sobě kolmé přímky procházející počátkem jsou ortogonální podprostory, jedna přímka není ale ortogonálním doplňkem druhé přímky. Ortogonální doplněk k přímce procházející počátkem je rovina procházející počátkem, která je na tuto přímku kolmá. Ortogonální doplněk celého $\mathbb{R}^3$ (míněného jako podprostor prostoru $\mathbb{R}^3$) je triviální podprostor $\{\mathbf{0}\}$. ♦

Příklad 4.2. Pozor, stěna místnosti není ortogonální k podlaze. Opravdu, existuje dvojice vektorů, jeden v podlaze a jeden ve stěně, které nejsou ortogonální (kde jsou?)[3]. ♦

Věta 4.1. Pro každý podprostor $X \subseteq \mathbb{R}^n$ platí $\dim X + \dim(X^\perp) = n$.

Důkaz. S ortogonálním doplňkem jsme se vlastně už setkali v definici nulového prostoru: nulový prostor matice je tvořen vektory ortogonálními na všechny řádky matice. Dle (4.5) je vektor ortogonální na řádky matice právě tehdy, když je ortogonální na lineární obal řádků matice. Tedy pro každou $\mathbf{A} \in \mathbb{R}^{m \times n}$ je

$$(\operatorname{rng}(\mathbf{A}^T))^\perp = \operatorname{null} \mathbf{A}. \tag{4.8}$$

Dokazovaná rovnost je tedy rovnost (3.21) pro $X = \operatorname{rng}(\mathbf{A}^T)$. ■

[3] Jak jsme už poznamenali, kolmost se v elementární geometrii definuje pro dvojice přímek. Tento pojem se v prostorové (v $\mathbb{R}^3$) rozšiřuje i na dvojice přímka-rovina či rovina-rovina. V tomto pojetí stěna a podlaha k sobě kolmé jsou (s čímž by jistě souhlasil např. pan zedník). Vidíme tedy, že prostorech $\mathbb{R}^n$ relace kolmosti a ortogonality splývají pro dvojice podprostorů dimenze 1 (tj. přímek), ale liší se pro dvojice podprostorů jiných dimenzí.

Věta 4.2. Pro každý podprostor $X \subseteq \mathbb{R}^n$ platí $(X^\perp)^\perp = X$.

Důkaz. Každý vektor je zřejmě ortogonální na všechny vektory k němu ortogonální, tedy platí $X \subseteq (X^\perp)^\perp$ (rozmyslete!). Z Věty 4.1 použité jednou na X a jednou na $X^\perp$ máme

$$\dim X + \dim X^\perp = n = \dim X^\perp + \dim(X^\perp)^\perp.$$

Z toho plyne $\dim X = \dim(X^\perp)^\perp$. Z Věty 3.3 proto máme $X = (X^\perp)^\perp$. ∎

Věta 4.2 ukazuje, že relace 'býti ortogonálním doplňkem' je symetrická: $Y = X^\perp \Leftrightarrow X = Y^\perp$. Proto můžeme říkat, že podprostory X a Y jsou *ortogonálním doplňkem jeden druhého.*

4.2.1 Vztah k prostoru obrazů a nulovému prostoru

Rozviňme naše pozorování (4.8) z důkazu Věty 4.1. Vzpomeňme z §3.2.1, že každá matice $\mathbf{A} \in \mathbb{R}^{m\times n}$ definuje čtyři základní podprostory:

- $\operatorname{rng} \mathbf{A} = \{\, \mathbf{Ax} \mid \mathbf{x} \in \mathbb{R}^n \,\} \subseteq \mathbb{R}^m$ je lineární obal sloupců $\mathbf{A}$,
- $\operatorname{null} \mathbf{A} = \{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{Ax} = \mathbf{0} \,\} \subseteq \mathbb{R}^n$ je prostor všech vektorů kolmých na řádky $\mathbf{A}$,
- $\operatorname{rng}(\mathbf{A}^T) = \{\, \mathbf{A}^T\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^m \,\} \subseteq \mathbb{R}^n$ je lineární obal řádků $\mathbf{A}$,
- $\operatorname{null}(\mathbf{A}^T) = \{\, \mathbf{x} \in \mathbb{R}^m \mid \mathbf{A}^T\mathbf{x} = \mathbf{0} \,\} \subseteq \mathbb{R}^m$ je prostor všech vektorů kolmých na sloupce $\mathbf{A}$.

Věta 4.3. Pro každou matici $\mathbf{A}$ platí

$$(\operatorname{rng} \mathbf{A})^\perp = \operatorname{null}(\mathbf{A}^T), \tag{4.9a}$$

$$(\operatorname{null} \mathbf{A})^\perp = \operatorname{rng}(\mathbf{A}^T). \tag{4.9b}$$

Důkaz. Rovnost (4.9a) plyne z prvního a čtvrtého řádku, je to vlastně rovnost (4.8) použitá na matici $\mathbf{A}^T$. Rovnost (4.9b) získáme použitím (4.9a) na matici $\mathbf{A}^T$ a rovnosti $(X^\perp)^\perp = X$. ∎

4.3 Ortonormální vektory

Vektor $\mathbf{u} \in \mathbb{R}^n$ nazveme **normalizovaný**, pokud má jednotkovou délku, tj. $\|\mathbf{u}\| = 1$. Množinu vektorů $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ nazveme[4] **ortonormální**[5], jestliže každý vektor z této množiny je normalizovaný a každá dvojice vektorů z této množiny je ortogonální, tedy

$$\mathbf{u}_i^T\mathbf{u}_j = \delta_{ij} = \begin{cases} 0 & \text{když } i \neq j, \\ 1 & \text{když } i = j. \end{cases} \tag{4.10}$$

[4]Pro vektory z ortonormální množiny je obvyklé používat písmena $\mathbf{u}, \mathbf{v}, \mathbf{q}$ místo $\mathbf{x}, \mathbf{y}, \mathbf{a}$, jak jste zvyklí. Podobně, matice ortonormálními sloupci (§4.4) je obvyklé značit $\mathbf{U}, \mathbf{V}, \mathbf{Q}$ místo $\mathbf{A}, \mathbf{B}$. To ale *neznamená*, že toto značení používá každý vždycky!

[5]Často říkáme zkráceně, že *vektory* $\mathbf{u}_1, \dots, \mathbf{u}_k$ *jsou ortonormální.* To je ale nepřesné, protože ortonormalita není vlastnost jednoho vektoru, ale množiny vektorů.

Tvrzení 4.4. Ortonormální množina vektorů je lineárně nezávislá.

Důkaz. Vektory $\mathbf{u}_1, \ldots, \mathbf{u}_k$ jsou lineárně nezávislé, je-li splněna podmínka (3.2), kde tedy místo $\mathbf{x}_i$ máme $\mathbf{u}_i$. Vynásobme levou stranu implikace (3.2) skalárně vektorem $\mathbf{u}_i$, což s použitím (4.10) dá

$$\alpha_1 \mathbf{u}_i^T \mathbf{u}_1 + \cdots + \alpha_k \mathbf{u}_i^T \mathbf{u}_k = \alpha_i = 0.$$

To platí pro každé i, tedy $\alpha_1 = \cdots = \alpha_k = 0$. To je pravá strana implikace (3.2). ∎

Tvrzení 4.5. Nechť vektor $\mathbf{x}$ je lineární kombinací ortonormálních vektorů $\mathbf{u}_1, \ldots, \mathbf{u}_k$:

$$\mathbf{x} = \alpha_1 \mathbf{u}_1 + \cdots + \alpha_k \mathbf{u}_k. \tag{4.11}$$

Pak koeficienty lineární kombinace se spočítají jednoduše jako $\alpha_i = \mathbf{u}_i^T \mathbf{x}$ $(i = 1, \ldots, k)$.

Důkaz. Vynásobením rovnice (4.11) skalárně vektorem $\mathbf{u}_i$ dostaneme $\mathbf{u}_i^T \mathbf{x} = \mathbf{u}_i^T \mathbf{u}_i \alpha_i = \alpha_i$, kde jsme opět použili (4.10). ∎

Skaláry $\alpha_1, \ldots, \alpha_k$ v Tvrzení 4.5 jsou souřadnice vektoru $\mathbf{x}$ v ortonormální bázi $\mathbf{u}_1, \ldots, \mathbf{u}_k$ podprostoru $\operatorname{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$. Později v §5.2 uvidíme (znáte to už ze střední školy), že α_i je délka (se znaménkem) ortogonální projekce vektoru $\mathbf{x}$ do přímky $\operatorname{span}\{\mathbf{u}_i\}$, za předpokladu $\|\mathbf{u}_i\| = 1$. Uvědomte si výhodu oproti situaci, kdy vektory $\mathbf{u}_1, \ldots, \mathbf{u}_k$ jsou lineárně nezávislé ale ne ortonormální (Tvrzení 3.1): pak koeficienty α_i musíme pracně počítat řešením soustavy lineárních rovnic.

Z Tvrzení 4.5 plyne, že pro každé $\mathbf{x} \in \operatorname{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ platí

$$\mathbf{x} = (\mathbf{u}_1^T \mathbf{x})\mathbf{u}_1 + \cdots + (\mathbf{u}_k^T \mathbf{x})\mathbf{u}_k. \tag{4.12}$$

Toto má souvislost s ortogonálními projektory, o kterých se dozvíte později.

Ortonormální báze lineárního (pod)prostoru odpovídá tomu, co ze základní školy znáte pod pojmem *kartézská souřadnicová soustava*. Souřadnice bodu vůči jeho ortonormální bázi (viz §3.1) pak znáte jako jeho *kartézské souřadnice*.

4.4 Matice s ortonormálními sloupci

Nechť sloupce matice $\mathbf{U} \in \mathbb{R}^{m \times n}$ tvoří ortonormální množinu vektorů. Dle Tvrzení 4.4 jsou sloupce $\mathbf{U}$ lineárně nezávislé, tedy nutně $m \geq n$ (tj. $\mathbf{U}$ je čtvercová nebo úzká). Podmínku ortonormality (4.10) sloupců matice $\mathbf{U}$ lze psát stručně jako

$$\mathbf{U}^T \mathbf{U} = \mathbf{I}. \tag{4.13}$$

Opravdu, skalární součin i-tého sloupce a j-tého sloupce matice $\mathbf{U}$ je (i, j)-tý prvek matice $\mathbf{U}^T \mathbf{U}$ (ověřte roznásobením blokových matic) a (i, j)-tý prvek matice $\mathbf{I}$ je δ_{ij}. Rovnost (4.13) také říká, že $\mathbf{U}^T$ je levá inverze matice $\mathbf{U}$ a $\mathbf{U}$ je pravá inverze matice $\mathbf{U}^T$ (viz §2.1.4).

Lineární zobrazení $\mathbf{f}(\mathbf{x}) = \mathbf{U}\mathbf{x}$ (zobrazení $\mathbb{R}^n \to \mathbb{R}^m$) zachovává skalární součin, neboť

$$\mathbf{f}(\mathbf{x})^T \mathbf{f}(\mathbf{y}) = (\mathbf{U}\mathbf{x})^T (\mathbf{U}\mathbf{y}) = \mathbf{x}^T \mathbf{U}^T \mathbf{U}\mathbf{y} = \mathbf{x}^T \mathbf{y}. \tag{4.14}$$

Pro $\mathbf{x} = \mathbf{y}$ dostaneme $\|\mathbf{f}(\mathbf{x})\| = \|\mathbf{U}\mathbf{x}\| = \|\mathbf{x}\|$, tedy zobrazení zachovává také eukleidovskou normu. Zobrazení tedy zachovává vzdálenosti a úhly (viz §2.1.7). Taková zobrazení se nazývají **isometrie**[6].

Obecněji, skalární součin matic (viz §2.1.7) je invariantní vůči skládání s isometrií zprava i zleva: jsou-li $\mathbf{A}, \mathbf{B}$ libovolné matice a $\mathbf{U}, \mathbf{V}$ matice s ortonormálními sloupci, z (2.10) máme

$$\langle \mathbf{U}\mathbf{A}\mathbf{V}^T, \mathbf{U}\mathbf{B}\mathbf{V}^T \rangle = \langle \mathbf{A}, \mathbf{U}^T\mathbf{U}\mathbf{B}\mathbf{V}^T\mathbf{V} \rangle = \langle \mathbf{A}, \mathbf{B} \rangle. \tag{4.15}$$

Tedy i Frobeniova norma (2.11) je invariantní vůči isometrii,

$$\|\mathbf{U}\mathbf{A}\mathbf{V}^T\| = \|\mathbf{A}\|. \tag{4.16}$$

Tvrzení 4.6. Pro každou čtvercovou matici $\mathbf{U}$ platí

$$\mathbf{U}^T\mathbf{U} = \mathbf{I} \iff \mathbf{U}^T = \mathbf{U}^{-1} \iff \mathbf{U}\mathbf{U}^T = \mathbf{I}. \tag{4.17}$$

Důkaz. Nechť $\mathbf{U}^T\mathbf{U} = \mathbf{I}$. Pak jsou sloupce $\mathbf{U}$ ortonormální a tedy lineárně nezávislé. To ale znamená, že $\mathbf{U}$ je regulární, protože je čtvercová. Vynásobením rovnosti $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ maticí $\mathbf{U}^{-1}$ zprava získáme $\mathbf{U}^T = \mathbf{U}^{-1}$. Vynásobením rovnosti $\mathbf{U}^T = \mathbf{U}^{-1}$ maticí $\mathbf{U}$ zleva získáme $\mathbf{U}\mathbf{U}^T = \mathbf{I}$. Zbylé implikace dokážeme podobně. ■

Věta říká, že má-li čtvercová matice ortonormální sloupce, má ortonormální i řádky, a inverze takové matice se spočítá jednoduše transpozicí. Čtvercové matici splňující podmínky (4.17) se říká **ortogonální matice**.

Zdůrazněme, že pokud $\mathbf{U}$ je obdélníková s ortonormálními sloupci, neplatí $\mathbf{U}\mathbf{U}^T = \mathbf{I}$. Dále, pokud má $\mathbf{U}$ ortogonální (ne však ortonormální) sloupce, nemusí mít ortogonální řádky[7].

Nechť $\mathbf{U}$ je ortogonální matice. Vezmeme-li determinant obou stran rovnice $\mathbf{U}^T\mathbf{U} = \mathbf{I}$, máme $\det(\mathbf{U}^T\mathbf{U}) = \det(\mathbf{U}^T)\det\mathbf{U} = (\det\mathbf{U})^2 = 1$. Tedy $\det\mathbf{U} \in \{-1, 1\}$.

- Pokud $\det\mathbf{U} = 1$, matici se říká **speciální ortogonální** nebo také **rotační**, protože transformace $\mathbf{f}(\mathbf{x}) = \mathbf{U}\mathbf{x}$ (zobrazení z $\mathbb{R}^n$ do sebe) znamená *otočení* vektoru $\mathbf{x}$ okolo počátku. Každou rotaci v prostoru $\mathbb{R}^n$ lze jednoznačně reprezentovat rotační maticí.
- Pokud $\det\mathbf{U} = -1$, transformace $\mathbf{f}$ je složením otočení a *ortogonální reflexe* neboli zrcadlení (viz §5.2.2).

Příklad 4.3. Všechny rotační matice 2×2 lze napsat jako

$$\mathbf{U} = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix}$$

pro nějaké φ. Násobení vektoru touto maticí odpovídá otočení vektoru v rovině o úhel φ. Zkontrolujte si, že $\mathbf{U}^T\mathbf{U} = \mathbf{I} = \mathbf{U}\mathbf{U}^T$ a $\det\mathbf{U} = 1$. ♦

[6]Obecněji, *isometrie* je zobrazení mezi dvěma metrickými (ne nutně lineárními) prostory, které zachovává vzdálenosti. V našem případě bychom mohli přesněji mluvit o *lineární isometrii*, tedy o isometrii, která je zároveň lineární zobrazení.

[7]To je možná důvod, proč se čtvercové matici s ortonormálními sloupci (tedy i řádky) neříká 'ortonormální' ale 'ortogonální'. Obdélníková matice s ortonormálními sloupci a čtvercová matice s ortogonálními (ne však ortonormálními) sloupci zvláštní jména nemají.

Příklad 4.4. Zrcadlení (neboli reflexe) v $\mathbb{R}^2$ kolem přímky procházející počátkem a směrnicí $\tan(\varphi/2)$ je reprezentováno ortogonální maticí

$$\begin{bmatrix} \cos\varphi & \sin\varphi \\ \sin\varphi & -\cos\varphi \end{bmatrix}.$$

♦

Příklad 4.5. Permutační matice je čtvercová matice, jejíž sloupce jsou permutované vektory standardní báze. Např.

$$\begin{bmatrix} \mathbf{e}_3 & \mathbf{e}_1 & \mathbf{e}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

Permutační matice je ortogonální (dokažte!) a její determinant je rovný znaménku permutace.♦

Na závěr uvedeme jedno hezké tvrzení, které nám bude užitečné později. Jeho druhá část má úzký vztah k Tvrzení 5.3 o ortogonálních projektorech z příští kapitoly.

Tvrzení 4.7. Nechť $\begin{bmatrix} \mathbf{U} & \mathbf{V} \end{bmatrix} \in \mathbb{R}^{m\times m}$ je ortogonální matice vodorovně rozdělená do dvou bloků $\mathbf{U} \in \mathbb{R}^{m\times n}$ a $\mathbf{V} \in \mathbb{R}^{m\times(m-n)}$. Pak $(\operatorname{rng}\mathbf{U})^\perp = \operatorname{rng}\mathbf{V}$ a $\mathbf{U}\mathbf{U}^T + \mathbf{V}\mathbf{V}^T = \mathbf{I}$.

Důkaz. Napíšeme rovnosti z Věty 4.6 a roznásobíme blokové matice:

$$\begin{bmatrix} \mathbf{U}^T\mathbf{U} & \mathbf{U}^T\mathbf{V} \\ \mathbf{V}^T\mathbf{U} & \mathbf{V}^T\mathbf{V} \end{bmatrix} = \begin{bmatrix} \mathbf{U}^T \\ \mathbf{V}^T \end{bmatrix} \begin{bmatrix} \mathbf{U} & \mathbf{V} \end{bmatrix} = \mathbf{I} = \begin{bmatrix} \mathbf{U} & \mathbf{V} \end{bmatrix} \begin{bmatrix} \mathbf{U}^T \\ \mathbf{V}^T \end{bmatrix} = \mathbf{U}\mathbf{U}^T + \mathbf{V}\mathbf{V}^T. \tag{4.18}$$

Porovnáním bloků matice vlevo s bloky jednotkové matice $\mathbf{I}$ máme $\mathbf{U}^T\mathbf{U} = \mathbf{I}$, $\mathbf{V}^T\mathbf{V} = \mathbf{I}$ a $\mathbf{U}^T\mathbf{V} = \mathbf{0}$. Třetí rovnost znamená $\operatorname{rng}\mathbf{U} \perp \operatorname{rng}\mathbf{V}$, neboli $\operatorname{rng}\mathbf{V} \subseteq (\operatorname{rng}\mathbf{U})^\perp$. Z rozměrů matic a Věty 4.1 je ale $\dim\operatorname{rng}\mathbf{V} = \dim(\operatorname{rng}\mathbf{U})^\perp$, tedy dle Tvrzení 3.3 platí $\operatorname{rng}\mathbf{V} = (\operatorname{rng}\mathbf{U})^\perp$. ■

4.5 Gramova-Schmidtova ortonormalizace

Gramova-Schmidtova ortonormalizace je algoritmus, který pro dané lineárně nezávislé vektory $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$ najde vektory $\mathbf{q}_1, \ldots, \mathbf{q}_n \in \mathbb{R}^m$ takové, že

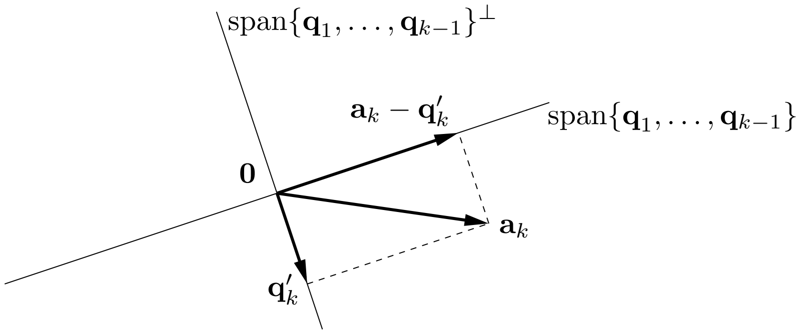
- $\mathbf{q}_1, \ldots, \mathbf{q}_n$ jsou ortonormální,
- pro každé $k = 1, \ldots, n$ platí $\operatorname{span}\{\mathbf{q}_1, \ldots, \mathbf{q}_k\} = \operatorname{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$.

Myšlenka algoritmu je jednoduchá. Předpokládejme, že máme ortonormální vektory $\mathbf{q}_1, \ldots, \mathbf{q}_{k-1}$ splňující $\operatorname{span}\{\mathbf{q}_1, \ldots, \mathbf{q}_{k-1}\} = \operatorname{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_{k-1}\}$. Spočítáme vektor

$$\mathbf{q}'_k = \mathbf{a}_k - \sum_{i=1}^{k-1} (\mathbf{q}_i^T\mathbf{a}_k)\mathbf{q}_i. \tag{4.19}$$

Tento vektor je kolmý na každý z vektorů $\mathbf{q}_1, \ldots, \mathbf{q}_{k-1}$, protože když skalárně vynásobíme výraz (4.19) jakýmkoliv z nich, tak díky ortonormalitě dostaneme nulu[8] (ověřte!). Podle předpokladu je $\mathbf{a}_k \notin \operatorname{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_{k-1}\} = \operatorname{span}\{\mathbf{q}_1, \ldots, \mathbf{q}_{k-1}\}$, tedy $\mathbf{q}'_k \neq \mathbf{0}$. Z (4.19) plyne $\mathbf{q}'_k \in \operatorname{span}\{\mathbf{q}_1, \ldots, \mathbf{q}_{k-1}, \mathbf{a}_k\} = \operatorname{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$. Teď stačí vektor $\mathbf{q}'_k$ znormalizovat, $\mathbf{q}_k = \mathbf{q}'_k/\|\mathbf{q}'_k\|$.

[8] Později uvidíme, že suma v (4.19) je ortogonální projekce (5.16) vektoru $\mathbf{a}_k$ na ortogonální doplněk podprostoru $\operatorname{span}\{\mathbf{q}_1, \ldots, \mathbf{q}_{k-1}\}$.



Algoritmus provede iteraci (4.19) postupně pro $k = 1,\dots,n$. Zde je běh algoritmu pro $n = 3$:

$$
\begin{aligned}
\mathbf{q}'_1 &= \mathbf{a}_1, & \mathbf{q}_1 &= \mathbf{q}'_1/\|\mathbf{q}'_1\| \\
\mathbf{q}'_2 &= \mathbf{a}_2 - (\mathbf{q}_1^T\mathbf{a}_2)\mathbf{q}_1, & \mathbf{q}_2 &= \mathbf{q}'_2/\|\mathbf{q}'_2\| \\
\mathbf{q}'_3 &= \mathbf{a}_3 - (\mathbf{q}_1^T\mathbf{a}_3)\mathbf{q}_1 - (\mathbf{q}_2^T\mathbf{a}_3)\mathbf{q}_2, & \mathbf{q}_3 &= \mathbf{q}'_3/\|\mathbf{q}'_3\|
\end{aligned}
$$

Gramův-Schmidtův algoritmus dokazuje nepřekvapivé ale nesamozřejmé vlastnosti ortonormální báze (analogie Věty 3.2):

Věta 4.8 (vlastnosti ortonormální báze).

- Každý lineární podprostor má (alespoň jednu) ortonormální bázi.
- Každou ortonormální množinu vektorů z lineárního podprostoru lze doplnit na ortonormální bázi tohoto podprostoru.

Důkaz. Dle Věty 3.2 má každý podprostor bázi, označme ji $\mathbf{a}_1,\dots,\mathbf{a}_n$. Algoritmus z těchto vektorů vyrobí vektory $\mathbf{q}_1,\dots,\mathbf{q}_n$, které tvoří ortonormální bázi stejného podprostoru.

Pro druhé tvrzení uvažujme ortonormální množinu vektorů $\mathbf{a}_1,\dots,\mathbf{a}_k$ z nějakého podprostoru. Dle Věty 3.2 tuto množinu doplníme na (ne nutně ortonormální) bázi $\mathbf{a}_1,\dots,\mathbf{a}_n$ tohoto podprostoru, kde $n \ge k$. Algoritmus z vektorů $\mathbf{a}_1,\dots,\mathbf{a}_n$ vyrobí vektory $\mathbf{q}_1,\dots,\mathbf{q}_n$, kde ale podle definice algoritmu bude $\mathbf{q}_i = \mathbf{a}_i$ pro každé $i \le k$. ■

4.5.1 QR rozklad

Věta 4.9 (QR rozklad). Každou matici $\mathbf{A} \in \mathbb{R}^{m\times n}$ lze rozložit na součin

$$\mathbf{A} = \mathbf{Q}\mathbf{R}, \tag{4.20}$$

kde matice $\mathbf{Q} \in \mathbb{R}^{m\times m}$ je ortogonální a matice $\mathbf{R} \in \mathbb{R}^{m\times n}$ je horní trojúhelníková.

Rozkladu matice (4.20) se říká **QR rozklad** matice $\mathbf{A}$. Gramova-Schmidtova ortogonalizace umožňuje dokázat Větu 4.9 pro případ, kdy matice $\mathbf{A}$ má lineárně nezávislé sloupce:

Důkaz. Rovnosti (4.19) lze napsat jako

$$\mathbf{a}_k = \sum_{i=1}^{k-1} r_{ik}\mathbf{q}_i + r_{kk}\mathbf{q}_k = \sum_{i=1}^{k} r_{ik}\mathbf{q}_i, \tag{4.21}$$

kde $r_{ik} = \mathbf{q}_i^T\mathbf{a}_k$ pro $i < k$ a $r_{kk} = \|\mathbf{q}'_k\|$. Soustava rovnic (4.21) pro $k = 1,\dots,n$ se dá napsat v maticovém tvaru jako $\mathbf{A} = \mathbf{Q}\mathbf{R}$, kde $\mathbf{a}_1,\dots,\mathbf{a}_n$ jsou sloupce matice $\mathbf{A}$, $\mathbf{q}_1,\dots,\mathbf{q}_n$ jsou sloupce matice $\mathbf{Q}$ a matice $\mathbf{R} = [r_{ik}]$ je horní trojúhelníková (rozmyslete!). ■

Gramův-Schmidtův algoritmus je silný teoretický nástroj, ale v uvedené podobě je nestabilní vůči zaokrouhlovacím chybám. Jeho modifikacemi lze tuto nevýhodu odstranit a navíc ho zobecnit na počítání QR rozkladu libovolné matice. Častěji se ovšem QR rozklad počítá algoritmy založenými na *Householderových reflexích* nebo *Givensových rotacích* (neuvádíme).

Rozlišujeme dvě verze QR rozkladu:

- Věta 4.9 popisuje *plnou* verzi QR rozkladu.
- Pokud $m > n$, je posledních $m - n$ řádků matice $\mathbf{R}$ nulových (protože $\mathbf{R}$ je horní trojúhelníková). Tyto řádky jsou násobeny posledními $m - n$ sloupci matice $\mathbf{Q}$. Můžeme tedy vynechat z matice $\mathbf{R}$ posledních $m - n$ řádků a z matice $\mathbf{Q}$ posledních $m - n$ sloupců, tedy bude $\mathbf{Q} \in \mathbb{R}^{m \times n}$ a $\mathbf{R} \in \mathbb{R}^{n \times n}$. Toto je *redukovaná* verze QR rozkladu.

Pro $m \leq n$ obě verze splývají. V Matlabu spočítáte plný QR rozklad příkazem `[Q,R]=qr(A)` a redukovaný příkazem `[Q,R]=qr(A,0)`. Zkoumejte v Matlabu příkaz `help qr`!

Příklad 4.6. Zde je příklad plného a redukovaného QR rozkladu matice 3×2:

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & \sqrt{2/3} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{2} & 1/\sqrt{6} & -1/\sqrt{3} \end{bmatrix} \begin{bmatrix} \sqrt{2} & 1/\sqrt{2} \\ 0 & \sqrt{3/2} \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & \sqrt{2/3} \\ 1/\sqrt{2} & -1/\sqrt{6} \\ 1/\sqrt{2} & 1/\sqrt{6} \end{bmatrix} \begin{bmatrix} \sqrt{2} & 1/\sqrt{2} \\ 0 & \sqrt{3/2} \end{bmatrix}.$$

♦

QR rozklad je velmi užitečný. Uveďme jeho použití na řešení lineárních soustav. Řešíme-li soustavu $\mathbf{A}\mathbf{x} = \mathbf{b}$, rozložíme $\mathbf{A} = \mathbf{Q}\mathbf{R}$ a vynásobíme soustavu zleva $\mathbf{Q}^T$, což dá

$$\mathbf{R}\mathbf{x} = \mathbf{Q}^T\mathbf{b}. \tag{4.22}$$

Pokud jsme použili plný QR rozklad, je toto ekvivalentní úprava, neboť $\mathbf{Q}$ je regulární. Ale protože je $\mathbf{R}$ trojúhelníková, soustavu jsme velmi zjednodušili. Např. pokud je $\mathbf{A}$ čtvercová regulární, jediné řešení soustavy (4.22) lze levně najít zpětnou substitucí.

4.6 Cvičení

4.1. Máme vektory $\mathbf{x} = (1, 2, 3)$ a $\mathbf{y} = (-1, 0, 1)$. Spočítejte (a) délku vektoru $\mathbf{x}$, (b) vzdálenost bodů $\mathbf{x}$ a $\mathbf{y}$, (c) úhel mezi vektory $\mathbf{x}$ a $\mathbf{y}$.

4.2. Dokažte, že eukleidovská norma (4.2) splňuje **trojúhelníkovou nerovnost** $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$. Pro jaké vektory $\mathbf{x}, \mathbf{y}$ platí tato nerovnost s rovností, tj. $\|\mathbf{x}+\mathbf{y}\| = \|\mathbf{x}\|+\|\mathbf{y}\|$?

4.3. Najděte bázi ortogonálního doplňku prostoru $\operatorname{span}\{(0, 1, 1), (1, 2, 3)\}$.

4.4. Jsou dány množiny $X = \operatorname{span}\{(1, 0, 1, 0)\}$, $Y = \{\, (x_1, x_2, x_3, x_4) \in \mathbb{R}^4 \mid x_1 + x_3 = 0 \,\}$ a $Z = \{\, (x_1, x_2, x_3, x_4) \in \mathbb{R}^4 \mid x_1 + x_3 = 0,\ x_2 = x_4 \,\}$. Jsou to lineární podprostory? Pokud ano, najděte jejich dimenze. Rozhodněte, které z následujících výroků platí: $X \perp Y$, $X \perp Z$, $Y \perp Z$, $X = Y^\perp$, $Y = X^\perp$, $X = Z^\perp$. Najděte libovolnou bázi podprostoru $Z^\perp$.

4.5. Pro dva vektory $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ dokažte následující tvrzení, nakreslete obrázek a uvědomte si, jaké známé středoškolské poučky jste to vlastně dokázali.

   a) Jestliže $\|\mathbf{x}\| = \|\mathbf{y}\|$, pak $(\mathbf{x} + \mathbf{y}) \perp (\mathbf{x} - \mathbf{y})$.
   b) Jestliže $\mathbf{x} \perp \mathbf{y}$, pak $\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 = \|\mathbf{x} - \mathbf{y}\|^2$.
   c) Jestliže $\mathbf{x} \perp \mathbf{y}$, pak $\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 = \|\mathbf{x} + \mathbf{y}\|^2$.

d) Jestliže vektory $\mathbf{x}_1, \ldots, \mathbf{x}_k$ jsou po dvojicích ortogonální (přesně: každá dvojice různých vektorů je ortogonální), pak $\|\mathbf{x}_1 + \cdots + \mathbf{x}_k\|^2 = \|\mathbf{x}_1\|^2 + \cdots + \|\mathbf{x}_k\|^2$.

4.6. Najděte ortonormální bázi podprostoru span$\{\,(1,1,1,-1),\ (2,-1,-1,1),\ (-1,2,2,1)\,\}$ pomocí QR rozkladu (použijte Matlab).

4.7. Dokažte, že součin ortogonálních matic je ortogonální matice.

4.8. Pro jaké n je matice $\operatorname{diag}(-\mathbf{1}_n)$ (tedy diagonální matice se samými minus jedničkami na diagonále) rotační?

4.9. Za jakých podmínek na čísla a, b je matice $\begin{bmatrix} a+b & b-a \\ a-b & b+a \end{bmatrix}$ ortogonální?

4.10. Existuje isometrie $\mathbf{f}\colon \mathbb{R}^3 \to \mathbb{R}^4$ tak, že $\mathbf{f}(1,-1,2) = (1,2,-1,1)$ a $\mathbf{f}(1,1,0) = (0,1,-1,0)$?

4.11. Isometrie (rotace+zrcadlení) v prostoru $\mathbb{R}^n$ jsou reprezentovány ortogonálními maticemi. Pro $n > 3$ ovšem nemáme o rotacích a zrcadleních intuitivní představu.

a) Počet nezávislých parametrů (‘stupňů volnosti’) ortogonální matice je rozdíl počtu prvků matice a počtu nezávislých (v našem případě různých) rovnic v podmínce $\mathbf{U}^T\mathbf{U} = \mathbf{I}$. Neformálně řečeno udává, kolika ‘knoflíky’ můžeme nezávisle ‘kroutit’ při rotaci v $\mathbb{R}^n$. Jaké je toto číslo pro $n = 2, 3, 4$? Najděte vzorec pro obecné n.

b) Je známo, že velmi malé rotace jsou v prvním řádu aproximovány maticí $\mathbf{I} + \mathbf{S}$ pro nějakou antisymetrickou matici $\mathbf{S}$. Ukažte, že to tak je. Spočítejte počet nezávislých parametrů antisymetrické matice $n \times n$ a ověřte, že je stejný jako v podúkolu (a).

4.12. Máme vektory $\mathbf{x}_1 = (1,-1,0,2)$, $\mathbf{x}_2 = (1,1,1,0)$, $\mathbf{x}_3 = (-1,-1,2,0)$.

a) Ověřte, že vektory $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ jsou po dvojicích ortogonální.

b) Najděte libovolnou bázi podprostoru $\operatorname{span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}^\perp$.

4.13. Najděte dva ortogonální vektory $\mathbf{x}, \mathbf{y}$ takové, že $\operatorname{span}\{\mathbf{x}, \mathbf{y}\} = \operatorname{span}\{(0,1,1), (1,2,3)\}$.

4.14. Spočtěte co nejjednodušším způsobem inverzi matice $\dfrac{1}{3}\begin{bmatrix} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{bmatrix}$.

4.15. Jak byste levně spočítali absolutní hodnotu determinantu matice z jejího QR rozkladu?

4.16. Pro některé (čtvercové) matice platí $(\operatorname{rng}\mathbf{A})^\perp = \operatorname{null}\mathbf{A}$. Dokažte, že to platí pro (a) symetrické matice, (b) antisymetrické matice.

4.17. Nechť $\mathbf{U}^T\mathbf{U} = \mathbf{I} = \mathbf{V}^T\mathbf{V}$, kde matice $\mathbf{U}$ a $\mathbf{V}$ mají stejný rozměr. Dokažte, že $\operatorname{rng}\mathbf{U} = \operatorname{rng}\mathbf{V}$ (tedy sloupce $\mathbf{U}$ a $\mathbf{V}$ jsou ortonormální báze téhož podprostoru) právě tehdy, když $\mathbf{V} = \mathbf{U}\mathbf{C}$ pro nějakou ortogonální matici $\mathbf{C}$.

4.18. Nechť $\mathbf{U}^T\mathbf{U} = \mathbf{I}$. Dokažte, že

a) Prvky matice $\mathbf{U}$ splňují $|a_{ij}| \le 1$.

b) Eukleidovská norma každého řádku matice $\mathbf{U}$ není větší než 1.

4.19. Nechť zobrazení $\mathbf{F}\colon \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ má hodnoty $\mathbf{F}(\mathbf{X}) = (\mathbf{I}-\mathbf{X})(\mathbf{I}+\mathbf{X})^{-1}$. Předpokládejte, že $\mathbf{X}$ je taková, že $\mathbf{I}+\mathbf{X}$ je regulární. Dokažte, že:

a) Pro každou $\mathbf{X}$ je $(\mathbf{I}-\mathbf{X})(\mathbf{I}+\mathbf{X}) = (\mathbf{I}+\mathbf{X})(\mathbf{I}-\mathbf{X})$.

b) Pro každou $\mathbf{X}$ je $(\mathbf{I}-\mathbf{X})(\mathbf{I}+\mathbf{X})^{-1} = (\mathbf{I}+\mathbf{X})^{-1}(\mathbf{I}-\mathbf{X})$.

c) Pro každou antisymetrickou $\mathbf{X}$ je $\mathbf{F}(\mathbf{X})$ ortogonální.

d) Pro každou ortogonální $\mathbf{X}$ je $\mathbf{F}(\mathbf{X})$ antisymetrická.

e) Zobrazení $\mathbf{F}$ je inverzí sama sebe (tedy je to *involuce*, viz §1.1.2), tj. $\mathbf{F}(\mathbf{F}(\mathbf{X})) = \mathbf{X}$ pro každou $\mathbf{X}$.

Před důkazy na papíře se přesvědčte v Matlabu, že tvrzení platí pro náhodné matice.

Nápověda a řešení

4.1. (a) $\|\mathbf{x}\| = \sqrt{14}$, (b) $\|\mathbf{x}-\mathbf{y}\| = 2\sqrt{3}$, (c) ≈ 1.1832 radiánů

4.2. Aby platila s rovností, vektory musí být rovnoběžné a shodně orientované, tj. jeden musí být nezáporným násobkem druhého. Tento výsledek byste měli také uhodnout úvahami nad obrázkem. Algebraicky to lze dokázat úpravou rovnost $\|\mathbf{x}+\mathbf{y}\| = \|\mathbf{x}\| + \|\mathbf{y}\|$ a porovnáním výsledku se vzorcem (4.3).

4.3. Např. $(1, 1, -1)$

4.4. Řešení ‘hrubou silou’ by bylo najít báze podprostorů X, Y, Z a jejich ortogonálních doplňků a pomocí nich na dané otázky odpovědět. To by ale zabralo zbytečně mnoho práce. Jednodušší je využít vztahů (4.9). Z definice rng a null je $X = \operatorname{rng}([\,1\;0\;1\;0\,]^T)$, $Y = \operatorname{null}[\,1\;0\;1\;0\,]$ a $Z = \operatorname{null}\left[\begin{smallmatrix}1&0&1&0\\0&1&0&-1\end{smallmatrix}\right]$. Zřejmě je $\dim X = 1$, $\dim Y = 3$ a $\dim Z = 2$. Dle (4.9) tedy $Y^\perp = (\operatorname{null}[\,1\;0\;1\;0\,])^\perp = \operatorname{rng}([\,1\;0\;1\;0\,]^T) = X$. Toto je dle Věty 4.2 ekvivalentní $X^\perp = Y$. Z definice ort. doplňku ihned plyne $X \perp Y$. Protože podprostor Z je omezen jednou rovnicí navíc oproti Y, je $Z \subseteq Y$. Z toho ihned $X \perp Z$ (protože když je každý vektor z Y kolmý na každý vektor z X, tak určitě je každý vektor z Z kolmý na každý vektor z X). Ale už není $X = Z^\perp$, protože to by dle Věty 4.1 muselo platit $\dim X = 4 - \dim Z$, což neplatí. Protože $\{\mathbf{0}\} \neq Z \subseteq Y$, nemůže být $Z \perp Y$, protože Z a Y určitě sdílí nějaký nenulový vektor (a nenulový vektor není kolmý sám na sebe). Je $Z^\perp = \left(\operatorname{null}\left[\begin{smallmatrix}1&0&1&0\\0&1&0&-1\end{smallmatrix}\right]\right)^\perp = \operatorname{rng}\left(\left[\begin{smallmatrix}1&0&1&0\\0&1&0&-1\end{smallmatrix}\right]^T\right)$, tedy báze $Z^\perp$ je $(1,0,1,0), (0,1,0,-1)$.

4.5.a) Dokazujeme, že úhlopříčky kosočtverce jsou na sebe kolmé.

4.5.b) $\|\mathbf{x}-\mathbf{y}\|^2 = (\mathbf{x}-\mathbf{y})^T(\mathbf{x}-\mathbf{y}) = \mathbf{x}^T\mathbf{x} - \mathbf{y}^T\mathbf{x} - \mathbf{x}^T\mathbf{y} + \mathbf{y}^T\mathbf{y} = \|\mathbf{x}\|^2 - 2\mathbf{x}^T\mathbf{y} + \|\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$, protože $\mathbf{x}^T\mathbf{y} = 0$. Dokázali jsme Pythagorovu větu.

4.5.d) Tvrzení je zobecnění Pythagorovy věty.

4.6. Báze vyjde $\{\, (1,1,1,-1)/2,\ (3,-1,-1,1)/\sqrt{12},\ (0,1,1,2)/\sqrt{6}\,\}$.

4.7. Nechť $\mathbf{U}^T\mathbf{U} = \mathbf{I} = \mathbf{V}^T\mathbf{V}$ a $\mathbf{W} = \mathbf{U}\mathbf{V}$. Pak $\mathbf{W}^T\mathbf{W} = \mathbf{V}^T\mathbf{U}^T\mathbf{U}\mathbf{V} = \mathbf{V}^T\mathbf{V} = \mathbf{I}$.

4.8. Musí být $\det\operatorname{diag}(-\mathbf{1}_n) = (-1)^n > 0$, tedy pro sudá n.

4.10. Ne, protože isometrie zachovává eukleidovskou normu, ale $\|(1,-1,2)\| \neq \|(1,2,-1,1)\|$.

4.11.a) Podmínka $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ je soustava rovnic $\mathbf{a}_i^T\mathbf{a}_j = \delta_{ij}$ pro $i,j = 1,\dots,n$, kde $\mathbf{a}_1,\dots,\mathbf{a}_n$ jsou sloupce matice $\mathbf{U}$. Tato soustava obsahuje jen $\binom{n}{2} + n$ různých rovnic (napište si ji např. pro $n = 4$). Tedy počet stupňů volnosti je $n^2 - \binom{n}{2} - n = \binom{n}{2} = n(n-1)/2$.

4.11.b) Nápověda: dokažte, že matice $\mathbf{I} + t\mathbf{S}$ je pro velmi malé t v prvním řádu blízká rotační, tedy $(\mathbf{I}+t\mathbf{S})^T(\mathbf{I}+t\mathbf{S}) \approx \mathbf{I}$.

4.13. Zvolíme $\mathbf{y} = (1,2,3) - r\mathbf{x}$, kde $r \in \mathbb{R}$ spočítáme z $\mathbf{x}^T\mathbf{y} = 0$. Tedy $r = \frac{5}{2}$ a $\mathbf{y} = (1, -\frac{1}{2}, \frac{1}{2})$. (V duchu Gramovy-Schmidtovy ortogonalizace.)

4.14. Není matice náhodou ortogonální?

4.16. Dle (4.9a) je $(\operatorname{rng}\mathbf{A})^\perp = \operatorname{null}(\mathbf{A}^T)$. Ale pro symetrické matice je $\operatorname{null}(\mathbf{A}^T) = \operatorname{null}\mathbf{A}$, pro antisymetrické je $\operatorname{null}(\mathbf{A}^T) = \operatorname{null}(-\mathbf{A}) = \operatorname{null}\mathbf{A}$.

4.17. Implikace $\Leftarrow$ plyne z Tvrzení 3.10, protože $\mathbf{C}$ je ortogonální a tedy regulární.

4.18.b) Doplňme matici $\mathbf{U}$ na ortogonální matici $\mathbf{C} = \begin{bmatrix}\mathbf{U} & \mathbf{V}\end{bmatrix}$. i-tý řádek této matice je $\mathbf{c}_i^T = \begin{bmatrix}\mathbf{a}_i^T & \mathbf{b}_i^T\end{bmatrix}$. Z ortogonality $\mathbf{C}$ je však $\mathbf{c}_i^T\mathbf{c}_i = \mathbf{a}_i^T\mathbf{a}_i + \mathbf{b}_i^T\mathbf{b}_i = 1$. Z toho $\mathbf{a}_i^T\mathbf{a}_i \leq 1$.

Kapitola 5

Lineární úloha nejmenších čtverců

Mějme soustavu m lineárních rovnic o n neznámých

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \tag{5.1}$$

kde $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$. Soustava má (aspoň jedno) řešení, právě když $\mathbf{b} \in \operatorname{rng} \mathbf{A}$ (tedy $\mathbf{b}$ je lineární kombinací sloupců $\mathbf{A}$), což lze psát také jako $\operatorname{rank} \begin{bmatrix} \mathbf{A} & \mathbf{b} \end{bmatrix} = \operatorname{rank} \mathbf{A}$ (Frobeniova věta[1]). Množina řešení soustavy je afinní podprostor $\mathbb{R}^n$ (dle Věty 3.13).

V této kapitole se zaměříme pouze na nehomogenní soustavy (tj. $\mathbf{b} \neq \mathbf{0}$). Rozlišme tři případy:

- Soustava nemá řešení. To nastane právě tehdy, když $\mathbf{b} \notin \operatorname{rng} \mathbf{A}$. Taková soustava se nazývá **přeurčená**. V tom případě můžeme chtít řešit soustavu přibližně, což je tématem §5.1.
- Soustava má právě jedno řešení. To nastane právě tehdy, když $\mathbf{b} \in \operatorname{rng} \mathbf{A}$ a matice $\mathbf{A}$ má lineárně nezávislé sloupce (tedy její nulový prostor je triviální).
- Soustava má nekonečně mnoho řešení. To nastane právě tehdy, když $\mathbf{b} \in \operatorname{rng} \mathbf{A}$ a matice $\mathbf{A}$ má lineárně závislé sloupce. Taková soustava se nazývá **nedourčená**. V tom případě můžeme chtít z množiny řešení vybrat jediné, čímž se budeme zabývat v §5.4.

5.1 Přibližné řešení ve smyslu nejmenších čtverců

Když soustava (5.1) nemá řešení, často je užitečné nalézt takové $\mathbf{x}$, aby rovnost (5.1) platila aspoň přibližně (což můžeme zapsat jako $\mathbf{A}\mathbf{x} \approx \mathbf{b}$). Přesněji, hledejme takové $\mathbf{x}$, aby eukleidovská norma vektoru $\mathbf{r} = \mathbf{b} - \mathbf{A}\mathbf{x}$ zbytků (neboli *reziduí*) byla co nejmenší (jednotlivá residua jsou $r_i = \mathbf{a}_i^T\mathbf{x} - b_i$ kde $\mathbf{a}_i^T$ jsou řádky matice $\mathbf{A}$ a b_i jsou složky vektoru $\mathbf{b}$). Úloha se nezmění (viz Cvičení 1.2), když místo eukleidovské normy budeme minimalizovat její čtverec $\|\mathbf{r}\|^2 = \mathbf{r}^T\mathbf{r} = r_1^2 + \cdots + r_m^2$. Tuto úlohu lze napsat jako

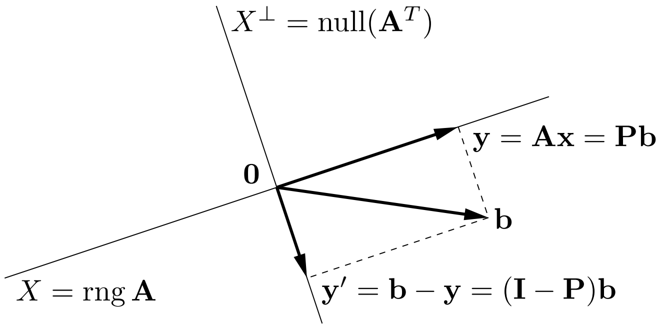
$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2. \tag{5.2}$$

Protože minimalizujeme součet čtverců reziduí, mluvíme o přibližném řešení soustavy (5.1) **ve smyslu nejmenších čtverců** (*least squares solution*)[2].

Řešení úlohy (5.2) najdeme následující úvahou, patrnou z obrázku (zatím ignorujte výrazy s $\mathbf{P}$, ty budeme potřebovat později):

[1] Frobeniova věta je v jiných zemích než Česku známa pod názvem *Rouché–Capelli theorem* nebo *Kronecker–Capelli theorem*.

[2] Přesně by se mělo říkat *least sum of squares*, protože existuje i metoda založená na *least median of squares*.



Pro různá $\mathbf{x}$ výraz $\mathbf{Ax}$ nabývá všech hodnot z podprostoru $\mathrm{rng}\,\mathbf{A}$ (viz definice (3.13) prostoru obrazů). Substitucí $\mathbf{y} = \mathbf{Ax}$ tedy můžeme úlohu (5.2) zapsat jako

$$\min_{\mathbf{y} \in \mathrm{rng}\,\mathbf{A}} \|\mathbf{y} - \mathbf{b}\|^2. \tag{5.3}$$

Vzdálenost $\|\mathbf{y} - \mathbf{b}\|$ bodů $\mathbf{y}$ a $\mathbf{b}$ je minimální, právě když vektor $\mathbf{b} - \mathbf{y}$ je kolmý na prostor $\mathrm{rng}\,\mathbf{A}$,[3] tedy na každý sloupec matice $\mathbf{A}$. Tuto podmínku lze zapsat jako $\mathbf{A}^T(\mathbf{Ax} - \mathbf{b}) = \mathbf{0}$, tj.

$$\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}. \tag{5.4}$$

Soustava (5.4) má n rovnic a n neznámých a nazývá se soustava **normálních rovnic** (protože ‘normála’ znamená ‘kolmice’).

Poznamenejme, že řešení (5.4) lze také snadno odvodit pomocí analýzy, položením parciálních derivací (diferencovatelné) funkce $\|\mathbf{Ax} - \mathbf{b}\|^2$ rovných nule. To uděláme až později v Příkladu 10.5, protože nyní ještě neumíte úsporně derivovat maticové výrazy (srov. také Příklad 6.15). Jednoduché instance úlohy (jako Příklad 5.1 nebo Příklad 1.9) je ovšem pohodlnější řešit derivacemi než převodem do mativové formy a použitím vzorce (5.4).[4].

Příklad 5.1. Mějme soustavu třech rovnic o dvou neznámých

$$\begin{aligned} x + 2y &= 6 \\ -x + y &= 3 \\ x + y &= 4 \end{aligned}$$

což lze napsat jako $\mathbf{Ax} = \mathbf{b}$ kde

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ -1 & 1 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 6 \\ 3 \\ 4 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}.$$

Tato soustava je přeurčená, tedy nemá řešení. Vyřešit ji přibližné ve smyslu nejmenších čtverců znamená najít taková čísla x, y, která minimalizují funkci

$$\|\mathbf{Ax} - \mathbf{b}\|^2 = (x + 2y - 6)^2 + (-x + y - 3)^2 + (x + y - 4)^2. \tag{5.5}$$

[3] Toto tvrzení (někdy nazývané *Věta o kolmici*) je sice intuitivně zřejmé, ale mělo by se dokázat. Jeho důkaz by ovšem vedl na důkaz toho, že optimální řešení úlohy (5.2) jsou právě řešení soustavy (5.4), což, jak jsme řekli, se snadno dokáže pomocí analýzy.

[4] Mnoha studentům tohle z nějakého důvodu uniká, zamotají se do matic a vektorú a jednoduchou lineární úlohu nějmenších čtverců nevyřeší.

Soustava (5.4) zní (po výnásobení matic)

$$\begin{aligned} 3x + 2y &= \ \ 7 \\ 2x + 6y &= 19 \end{aligned}$$

což také dostaneme položením parciálních derivací funkce (5.5) rovných nule (proveďte!). Tato soustava má právě jedno řešení $(x, y) = (\frac{2}{7}, \frac{43}{14})$. ♦

Pojďme zkoumat řešitelnost této soustavy. K tomu potřebovat následující větu[5]:

Věta 5.1. Pro každou matici $\mathbf{A}$ platí

$$\operatorname{rng}(\mathbf{A}^T\mathbf{A}) = \operatorname{rng}(\mathbf{A}^T), \tag{5.6a}$$

$$\operatorname{null}(\mathbf{A}^T\mathbf{A}) = \operatorname{null}\mathbf{A}. \tag{5.6b}$$

Důkaz. Dokažme nejprve rovnost (5.6b), tj. $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{0} \iff \mathbf{A}\mathbf{x} = \mathbf{0}$. Implikace $\Leftarrow$ je snadná, vynásobením $\mathbf{A}\mathbf{x} = \mathbf{0}$ zleva maticí $\mathbf{A}^T$. Implikace $\Rightarrow$ se dokáže takto:

$$\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{0} \quad\Longrightarrow\quad \mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x} = (\mathbf{A}\mathbf{x})^T(\mathbf{A}\mathbf{x}) = 0 \quad\Longrightarrow\quad \mathbf{A}\mathbf{x} = \mathbf{0},$$

neboť pro libovolný vektor $\mathbf{y}$ platí $\mathbf{y}^T\mathbf{y} = 0 \ \Rightarrow\ \mathbf{y} = \mathbf{0}$ (proč?).

Dokažme (5.6a). Z definice (3.13) je jasné (viz Tvrzení 3.10), že $\operatorname{rng}(\mathbf{A}^T\mathbf{A}) \subseteq \operatorname{rng}(\mathbf{A}^T)$. Nyní použijeme Větu 3.11 jednou na matici $\mathbf{A}$ a jednou na matici $\mathbf{A}^T\mathbf{A}$:

$$\dim\operatorname{rng}\mathbf{A} + \dim\operatorname{null}\mathbf{A} = n = \dim\operatorname{rng}(\mathbf{A}^T\mathbf{A}) + \dim\operatorname{null}(\mathbf{A}^T\mathbf{A}).$$

Díky (5.6b) z toho máme $\dim\operatorname{rng}(\mathbf{A}^T\mathbf{A}) = \dim\operatorname{rng}\mathbf{A} = \dim\operatorname{rng}(\mathbf{A}^T)$, kde druhá rovnost plyne z Věty 3.7. Z Věty 3.3 tedy máme (5.6a). ∎

Teď snadno ukážeme, že soustava (5.4) má (aspoň jedno) řešení pro každé $\mathbf{A}, \mathbf{b}$. Opravdu, soustava má řešení právě tehdy, když $\mathbf{A}^T\mathbf{b} \in \operatorname{rng}(\mathbf{A}^T\mathbf{A}) = \operatorname{rng}(\mathbf{A}^T)$, kde rovnost je (5.6a). Ale z definice prostoru obrazů je $\mathbf{A}^T\mathbf{b} \in \operatorname{rng}(\mathbf{A}^T)$ pro libovolná $\mathbf{A}, \mathbf{b}$.

Zkombinujeme-li (5.6a) a Větu 3.7, máme

$$\operatorname{rank}(\mathbf{A}\mathbf{A}^T) = \operatorname{rank}\mathbf{A} = \operatorname{rank}(\mathbf{A}^T) = \operatorname{rank}(\mathbf{A}^T\mathbf{A}). \tag{5.7}$$

Dle (5.7) je matice $\mathbf{A}^T\mathbf{A}$ regulární, právě když $\mathbf{A}$ má lineárně nezávislé sloupce. V tom případě můžeme soustavu (5.4) řešit pomocí inverze. Řešením je vektor $\mathbf{x} = \mathbf{A}^+\mathbf{b}$, kde

$$\mathbf{A}^+ = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T. \tag{5.8}$$

Matice (5.8) se nazývá **pseudoinverze** matice $\mathbf{A}$ s lineárně nezávislými sloupci. Je to jedna z levých inverzí matice $\mathbf{A}$, neboť $\mathbf{A}^+\mathbf{A} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{A} = \mathbf{I}$. Později ukážeme (v Příkladu 11.5), že $\mathbf{A}^+$ má ze všech levých inverzí matice $\mathbf{A}$ nejmenší (Frobeniovu) normu[6] $\|\mathbf{A}^+\|$.

Má-li matice $\mathbf{A}$ lineárně závislé sloupce, matice $\mathbf{A}^T\mathbf{A}$ nemá inverzi a proto vzorec (5.8) nelze použít. Potom soustava (5.4) a tedy i úloha (5.2) mají nekonečně mnoho (afinní podprostor) optimálních řešení (pozor, to je něco jiného, než že soustava (5.1) má nekonečně mnoho řešení!).

[5]Matice tvaru $\mathbf{A}^T\mathbf{A}$ či $\mathbf{A}\mathbf{A}^T$ se často objevují v různých situacích. Označme jako $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$ sloupce matice $\mathbf{A}$. Matici $\mathbf{A}^T\mathbf{A} \in \mathbb{R}^{n\times n}$ se říká *Gramova matice* nebo *Gramián* vektorů $\mathbf{a}_1, \ldots, \mathbf{a}_n$ a její prvky jsou skalární součiny $\mathbf{a}_i^T\mathbf{a}_j$ (viz (2.21)). Matici $\mathbf{A}\mathbf{A}^T \in \mathbb{R}^{m\times m}$ lze zase vidět (až na skalární násobek) jako empirickou *kovarianční matici* n pozorování $\mathbf{a}_1, \ldots, \mathbf{a}_n$ m-tice náhodných proměnných.

[6]Toto tvrzení je důsledek tzv. *Gaussovy-Markovovy* věty o LS estimátoru.

5.1.1 Řešení pomocí QR rozkladu

I když má matice $\mathbf{A}$ lineárně nezávislé sloupce, řešení pomocí explicitního řešení normální rovnice (5.4) nebo pomocí vzorce (5.8) není nejvhodnější pro numerické výpočty. Jeden důvod je, že výpočet součinu $\mathbf{A}^T\mathbf{A}$ může být pracný. Druhý důvod je, že v aritmetice s konečnou přesností (kterou na počítačích používáme) může při výpočtu součinu $\mathbf{A}^T\mathbf{A}$ dojít ke zbytečné ztrátě přesnosti.

Příklad 5.2. Řešme soustavu $\mathbf{Ax} = \mathbf{b}$ pro

$$\mathbf{A} = \begin{bmatrix} 3 & 6 \\ 1 & 2.01 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 9 \\ 3.01 \end{bmatrix}.$$

Matice $\mathbf{A}$ je regulární. Dejme tomu, že používáme aritmetiku s pohyblivou řádovou čárkou s přesností na 3 platné cifry. Gaussova eliminace najde přesné řešení soustavy $\mathbf{x} = (1, 1)$. Pokud ovšem v této aritmetice zformulujeme normální rovnici $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$, dostaneme

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} 10 & 20 \\ 20 & 40 \end{bmatrix}, \quad \mathbf{A}^T\mathbf{b} = \begin{bmatrix} 30 \\ 60.1 \end{bmatrix}.$$

I když v přesné aritmetice je matice $\mathbf{A}^T\mathbf{A}$ regulární, v naší přibližné aritmetice došlo v součinu $\mathbf{A}^T\mathbf{A}$ k zaokrouhlení a výsledná matice je singulární. Tedy soustava $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$ nemá řešení. ♦

Numericky vhodnější způsob je řešit normální rovnici *bez* explicitního výpočtu součinu $\mathbf{A}^T\mathbf{A}$. To lze udělat pomocí redukovaného QR rozkladu $\mathbf{A} = \mathbf{QR}$. Po dosazení do normální rovnice máme $\mathbf{R}^T\mathbf{Q}^T\mathbf{QRx} = \mathbf{R}^T\mathbf{Q}^T\mathbf{b}$ neboli $\mathbf{R}^T\mathbf{Rx} = \mathbf{R}^T\mathbf{Q}^T\mathbf{b}$. Jestliže $\mathbf{A}$ má lineárně nezávislé sloupce, matice $\mathbf{R}$ je regulární. Vynásobení maticí $\mathbf{R}^{-T}$ zleva (což je ekvivalentní úprava) dá

$$\mathbf{Rx} = \mathbf{Q}^T\mathbf{b}. \tag{5.9}$$

Zdůrazněme, že pokud $\mathbf{A}$ není čtvercová, pak soustava (5.9) není ekvivalentní původní soustavě $\mathbf{Ax} = \mathbf{b}$.

Jestliže sloupce $\mathbf{A}$ jsou lineárně závislé, postup je trochu složitější, ale také stojí na QR rozkladu. V Matlabu je řešení nehomogenní soustavy lineárních rovnic implementováno v operátoru \ (*zpětné lomítko*). Pokud je soustava přeurčená, výsledkem je přibližné řešení ve smyslu nejmenších čtverců, přičemž použitý algoritmus používá QR rozklad. Pochopte všechny funkce operátorů *lomítko* a *zpětné lomítko* pomocí studia příkazů `help mrdivide` a `help mldivide`!

5.2 Ortogonální projekce na podprostor

Naše úvaha o řešení úlohy (5.2) nám umožňuje zavést následující důležitý pojem: vektor $\mathbf{y} \in X$ splňující $(\mathbf{b} - \mathbf{y}) \perp X$ se nazývá **ortogonální projekce** bodu $\mathbf{b}$ na podprostor X. Vektoru $\mathbf{y}' = \mathbf{b} - \mathbf{y}$ se někdy říká **ortogonální rejekce** bodu $\mathbf{b}$ z podprostoru X. Rejekce z podprostoru X je zjevně totéž jako projekce na jeho ortogonální doplněk $X^\perp$ a naopak. Vše promyslete na obrázku v §5.1.

Nechť podprostor X je reprezentován bází tvořící sloupce matice $\mathbf{A}$, tj. $\mathbf{A}$ má lineárně nezávislé sloupce a $X = \operatorname{rng} \mathbf{A}$. Pak z (5.8) máme $\mathbf{y} = \mathbf{Ax} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$, kde $\mathbf{x}$ je (jediné) řešení rovnice (5.4). To lze napsat jako $\mathbf{y} = \mathbf{Pb}$, kde matici

$$\mathbf{P} = \mathbf{AA}^+ = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \tag{5.10}$$

říkáme **ortogonální projektor** na podprostor X. Vztah $\mathbf{y} = \mathbf{P}\mathbf{b}$ říká, že ortogonální projekce je lineární zobrazení, reprezentované maticí $\mathbf{P}$. Lze ukázat (ve Cvičení 5.13), že matice (5.10) nezávisí na bázi podprostoru X. Tedy ortogonální projekce existuje a je jediná[7].

Ze vzorce (5.10) plyne (ověřte!), že každý ortogonální projektor splňuje rovnosti

$$\mathbf{P}^2 = \mathbf{P} = \mathbf{P}^T. \tag{5.11}$$

Rovnost $\mathbf{P}^2 = \mathbf{P}$ říká očividnou věc: když vektor jednou promítneme na nějaký podprostor, tak opětovné promítnutí na tentýž podprostor ho už nezmění. Jinými slovy, lineární zobrazení $\mathbf{f}(\mathbf{b}) = \mathbf{P}\mathbf{b}$ je idempotentní, tj. $\mathbf{f} \circ \mathbf{f} = \mathbf{f}$ (viz §1.1.2). Rovnost $\mathbf{P}^T = \mathbf{P}$ říká, že matice $\mathbf{P}$ je symetrická.

Co je prostorem obrazů a nulovým prostorem ortogonálního projektoru? Úvahou (Kam se promítne každý vektor? Které vektory se promítnou do počátku?) snadno vidíme, že

$$\operatorname{rng} \mathbf{P} = X, \qquad \operatorname{null} \mathbf{P} = X^\perp. \tag{5.12}$$

Algebraicky to plyne z Věty 3.10.

Existence a jednoznačnost ortogonální projekce má důležitý důsledek:

Tvrzení 5.2. Pro každý vektor $\mathbf{b} \in \mathbb{R}^m$ existuje právě jedna dvojice vektorů $\mathbf{y} \in X$ a $\mathbf{y}' \in X^\perp$ tak, že $\mathbf{y} + \mathbf{y}' = \mathbf{b}$.

Důkaz. Vše je vidět z obrázku výše. Rovnost $\mathbf{b} = \mathbf{y} + \mathbf{y}'$ napíšeme jako $\mathbf{y}' = \mathbf{b} - \mathbf{y}$. Podmínku $\mathbf{y}' \in X^\perp$ lze psát jako $\mathbf{y}' \perp X$ neboli $(\mathbf{b} - \mathbf{y}) \perp X$. Tedy $\mathbf{y}$ je ortogonální projekce bodu $\mathbf{b}$ na X, která existuje a je jediná. ■

Protože $\mathbf{y} \in X$, je $\mathbf{y} \perp X^\perp$ a tedy $(\mathbf{b} - \mathbf{y}') \perp X^\perp$. Z definice ortogonální projekce je tedy vektor $\mathbf{y}'$ ortogonální projekce vektoru $\mathbf{b}$ na podprostor $X^\perp$. Z toho ihned tato užitečná věc:

Tvrzení 5.3. Je-li $\mathbf{P}$ ortogonální projektor na podprostor X, pak ortogonální projektor na podprostor $X^\perp$ je $\mathbf{I} - \mathbf{P}$.

Důkaz. Okamžitě z rovností $\mathbf{y}' = \mathbf{b} - \mathbf{y} = \mathbf{b} - \mathbf{P}\mathbf{b} = (\mathbf{I} - \mathbf{P})\mathbf{b}$. ■

V některých speciálních případech se vzorec (5.10) zjednoduší:

- Má-li matice $\mathbf{A}$ jediný sloupec, který označme $\mathbf{A} = \mathbf{a} \in \mathbb{R}^m$, promítáme na jednorozměrný podprostor (přímku procházející počátkem) $X = \operatorname{span}\{\mathbf{a}\}$. Pak

  $$\mathbf{P} = \frac{\mathbf{a}\mathbf{a}^T}{\mathbf{a}^T\mathbf{a}}. \tag{5.13}$$

  Všimněte si, že výraz $\mathbf{a}\mathbf{a}^T \in \mathbb{R}^{m\times m}$ je matice (dyáda) zatímco $\mathbf{a}^T\mathbf{a} \in \mathbb{R}$ je skalár (viz §2.2). Ortogonální projekce bodu $\mathbf{b}$ na přímku $\operatorname{span}\{\mathbf{a}\}$ je

  $$\mathbf{y} = \mathbf{P}\mathbf{b} = \frac{\mathbf{a}\mathbf{a}^T}{\mathbf{a}^T\mathbf{a}}\mathbf{b} = \frac{\mathbf{a}^T\mathbf{b}}{\mathbf{a}^T\mathbf{a}}\mathbf{a}. \tag{5.14}$$

  Vidíme, že $\mathbf{y}$ je skalární násobek vektoru $\mathbf{a}$. To je vzoreček pro průmět vektoru na přímku, který znáte ze střední školy. Skalár $\alpha = (\mathbf{a}^T\mathbf{b})/(\mathbf{a}^T\mathbf{a})$ je délka průmětu, srov. Tvrzení 4.5.

[7] Tím neříkáme, že normální rovnice (5.4) má jediné řešení (což nastane právě tehdy, když $\mathbf{A}$ má lin. nezávislé sloupce). Ale pro každé takové řešení $\mathbf{x}$ je vektor $\mathbf{y} = \mathbf{A}\mathbf{x}$ stejný (což plyne z (5.6b)).

- Je-li vektor navíc normalizovaný, tj. $X = \operatorname{span}\{\mathbf{u}\}$ kde $\mathbf{u}^T\mathbf{u} = 1$, máme $\mathbf{P} = \mathbf{u}\mathbf{u}^T$ a[8] $\mathbf{y} = \mathbf{u}\mathbf{u}^T\mathbf{b} = (\mathbf{u}^T\mathbf{b})\mathbf{u}$.
- Pokud $X = \operatorname{rng}\mathbf{U}$ kde $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ (tj. podprostor X je reprezentován ortonormální bází tvořenou sloupci matice $\mathbf{U}$) vzorec (5.10) se zjednoduší na

$$\mathbf{P} = \mathbf{U}\mathbf{U}^T. \tag{5.15}$$

Označíme-li sloupce matice $\mathbf{U}$ jako $\mathbf{u}_1, \dots, \mathbf{u}_n \in \mathbb{R}^m$, ortogonální projekce je

$$\mathbf{y} = \mathbf{U}\mathbf{U}^T\mathbf{b} = \mathbf{u}_1\mathbf{u}_1^T\mathbf{b} + \cdots + \mathbf{u}_n\mathbf{u}_n^T\mathbf{b} = (\mathbf{u}_1^T\mathbf{b})\mathbf{u}_1 + \cdots + (\mathbf{u}_n^T\mathbf{b})\mathbf{u}_n = \alpha_1\mathbf{u}_1 + \cdots + \alpha_n\mathbf{u}_n, \tag{5.16}$$

kde $\alpha_i = \mathbf{u}_i^T\mathbf{y}$ jsou souřadnice vektoru $\mathbf{y}$ v ortonormální bázi $\mathbf{u}_1, \dots, \mathbf{u}_n$ (viz Tvrzení 4.5). Vidíme, že každý vektor $\mathbf{u}_i\mathbf{u}_i^T\mathbf{b} = (\mathbf{u}_i^T\mathbf{b})\mathbf{u}_i = \alpha_i\mathbf{u}_i$ je sám o sobě ortogonální projekce vektoru $\mathbf{b}$ na přímku $\operatorname{rng}\mathbf{u}_i = \operatorname{span}\{\mathbf{u}_i\}$. Uvědomte si souvislost s Tvrzením 4.5: (5.16) je vlastně zkrácená suma (4.12).

5.2.1 Vzdálenost bodu od podprostoru

Uvažujme úlohu

$$\min_{\mathbf{y} \in X} \|\mathbf{y} - \mathbf{b}\| \tag{5.17}$$

kde $\mathbf{b} \in \mathbb{R}^m$ a $X \subseteq \mathbb{R}^m$ je libovolná množina. Jestliže má úloha optimální řešení (což nemusí, viz §1.2, najděte příklad!), pak její optimální hodnota je *vzdálenost* bodu $\mathbf{b}$ od množiny X. Optimální řešení (argument) úlohy (5.17) pak přirozeně můžeme nazvat *ortogonální projekcí* bodu $\mathbf{b}$ na množinu X. Tento název je ale smysluplný jen tehdy, jestliže optimální řešení je právě jedno. To nastane např. když X je podprostor nebo afinní podprostor[9]. Zaměřme se nyní na tyto dva případy.

Je-li X podprostor, vzdálenost bodu $\mathbf{b}$ od podprostoru X je rovna délce vektoru $\mathbf{y}' = \mathbf{b} - \mathbf{y}$ a vzdálenost bodu $\mathbf{y}$ od podprostoru $X^\perp$ je délka vektoru $\mathbf{y}$, kde $\mathbf{y}$ příp. $\mathbf{y}'$ je ortogonální projekce bodu $\mathbf{b}$ na X příp. $X^\perp$. Je-li $\mathbf{P}$ projektor na podprostor X, pak tedy

- vzdálenost bodu $\mathbf{b}$ od podprostoru X (neboli délka ortogonální projekce bodu $\mathbf{b}$ na podprostor $X^\perp$) je rovna $\|\mathbf{y}'\| = \|\mathbf{b} - \mathbf{y}\| = \|(\mathbf{I} - \mathbf{P})\mathbf{b}\|$,
- vzdálenost bodu $\mathbf{b}$ od podprostoru $X^\perp$ (neboli délka ortogonální projekce bodu $\mathbf{b}$ na podprostor X) je rovna $\|\mathbf{y}\| = \|\mathbf{P}\mathbf{b}\|$.

Je-li podprostor X reprezentován ortonormální bází tvořenou sloupci matice $\mathbf{U}$ (tj. $X = \operatorname{rng}\mathbf{U}$ kde $\mathbf{U}^T\mathbf{U} = \mathbf{I}$), pro vzdálenost bodu $\mathbf{b}$ od podprostoru $X^\perp = (\operatorname{rng}\mathbf{U})^\perp = \operatorname{null}(\mathbf{U}^T)$ dostaneme obzvlášť jednoduchý a užitečný vzorec

$$\|\mathbf{y}\| = \|\mathbf{U}\mathbf{U}^T\mathbf{b}\| = \|\mathbf{U}^T\mathbf{b}\|, \tag{5.18}$$

kde první rovnost plyne z (5.15) a druhá rovnost platí, protože isometrie zachovává eukleidovskou normu (viz §4.4). Má-li matice $\mathbf{U}$ jediný sloupec, tedy $\mathbf{U} = \mathbf{u} \in \mathbb{R}^m$ kde $\mathbf{u}^T\mathbf{u} = 1$, máme $\|\mathbf{y}\| = \|\mathbf{u}^T\mathbf{b}\|$, což je tedy vzdálenost bodu $\mathbf{b}$ od nadroviny $\{\, \mathbf{x} \in \mathbb{R}^m \mid \mathbf{u}^T\mathbf{x} = 0 \,\}$ s normálou $\mathbf{u}$.

[8] Závorka ve výrazu $(\mathbf{u}^T\mathbf{b})\mathbf{u}$ je nutná, protože součin výrazů $\mathbf{u}^T\mathbf{b}$ a $\mathbf{u}$ není maticový součin, ale násobení vektoru skalárem (viz poznámka v §2.1.2). Oproti tomu, výraz $\mathbf{u}\mathbf{u}^T\mathbf{b}$ je maticový součin tří matic, takže závorky v něm být nemusí.

[9] Obecněji platí, že úloha (5.17) má právě jedno optimální řešení, jestliže X je uzavřená konvexní množina.

Chceme-li spočítat vzdálenost bodu $\mathbf{b} \subseteq \mathbb{R}^m$ od *afinního* podprostoru $X \subseteq \mathbb{R}^m$, zvolíme libovolný vektor $\mathbf{b}_0 \in X$ a bod $\mathbf{b}$ i podprostor X posuneme o vektor $-\mathbf{b}_0$. Tím se vzdálenost nezmění, ale úlohu tím převedeme na výpočet vzdálenosti bodu $\mathbf{b} - \mathbf{b}_0$ od množiny $X - \mathbf{b}_0$, která je dle Věty 3.12 lineární podprostor.

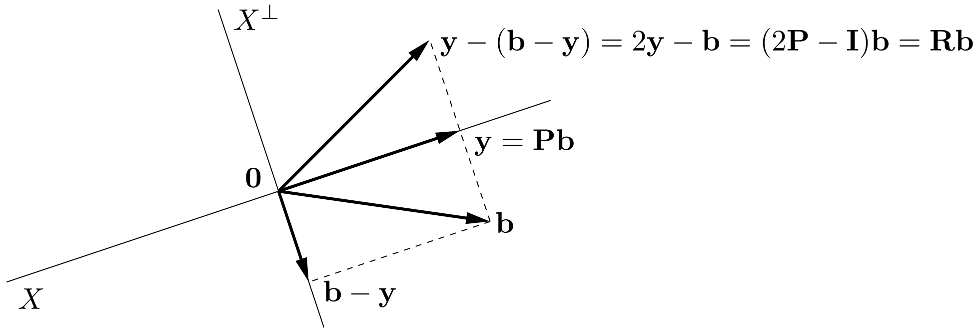
Je-li náš afinní podprostor množina řešení soustavy $\mathbf{U}^T\mathbf{x} = \mathbf{c}$ (srov. Věta 3.13), pak $\mathbf{b}_0$ zvolíme jako libovolné partikulární řešení soustavy, tedy splňující $\mathbf{U}^T\mathbf{b}_0 = \mathbf{c}$. Je-li $\mathbf{U}^T\mathbf{U} = \mathbf{I}$, vzdálenost bodu $\mathbf{b}$ od tohoto afinního podprostoru dostaneme použitím vzorce (5.18) jako

$$\|\mathbf{U}^T(\mathbf{b} - \mathbf{b}_0)\| = \|\mathbf{U}^T\mathbf{b} - \mathbf{c}\|. \tag{5.19}$$

Speciálně, nechť náš afinní podprostor je nadrovina s rovnicí $\mathbf{u}^T\mathbf{x} = c$ kde $\|\mathbf{u}\| = 1$ (což vždy jde zajistit vydělením celé rovnice skalárem $\|\mathbf{u}\|$). Pak vzdálenost bodu $\mathbf{b}$ od nadroviny je $|\mathbf{u}^T\mathbf{b} - c|$. Vidíme zde geometrický význam vektoru $\mathbf{u}$ a čísla c: $\mathbf{u}$ je (normalizovaný) normálový vektor nadroviny a $|c|$ je vzdálenost nadroviny od počátku.

5.2.2 ($\star$) Ortogonální zrcadlení a reflektory

Je-li $\mathbf{y} \in X$ ortogonální projekce vektoru $\mathbf{b} \in \mathbb{R}^m$ na podprostor $X \subseteq \mathbb{R}^m$, pak vektor $2\mathbf{y} - \mathbf{b}$ nazveme **ortogonální reflexí (zrcadlením)** vektoru $\mathbf{b}$ okolo podprostoru X. Promyslete si na obrázku, jak tento vektor vznikne:



Je-li $\mathbf{P}$ ortogonální projektor na X, máme $2\mathbf{y} - \mathbf{b} = 2\mathbf{P}\mathbf{b} - \mathbf{b} = (2\mathbf{P} - \mathbf{I})\mathbf{b}$. Matice

$$\mathbf{R} = 2\mathbf{P} - \mathbf{I} \tag{5.20}$$

se nazývá **ortogonální reflektor**. Z podmínek (5.11) lze ověřit, že reflektor splňuje

$$\mathbf{R}^2 = \mathbf{I}, \qquad \mathbf{R}^T = \mathbf{R}. \tag{5.21}$$

Opravdu: $\mathbf{R}^2 = (2\mathbf{P} - \mathbf{I})^2 = 4\mathbf{P}^2 - 2\mathbf{P} - 2\mathbf{P} + \mathbf{I} = 4\mathbf{P} - 2\mathbf{P} - 2\mathbf{P} + \mathbf{I} = \mathbf{I}$, symetrii dokážeme ještě snadněji. Rovnost $\mathbf{R}^2 = \mathbf{I}$ říká, že lineární zobrazení $\mathbf{f}(\mathbf{b}) = \mathbf{R}\mathbf{b}$ je *involuce* (viz §1.1.2), tj. $\mathbf{f} \circ \mathbf{f}$ je identické zobrazení. To má přirozené vysvětlení: jestliže ozrcadlíme libovolný vektor okolo podprostoru, pak opětovné ozrcadlení okolo stejného podprostoru dá původní vektor. Podmínky (5.21) charakterizují ortogonální reflektory: každá matice $\mathbf{R}$ splňující (5.21) je ortogonální reflektor (okolo nějakého podprostoru).

Rovnice (5.20) tedy přiřazuje každému ortogonálnímu projektoru ortogonální reflektor. Obráceně ověřte, je-li $\mathbf{R}$ libovolná matice splňující (5.21), pak matice $\mathbf{P} = \frac{1}{2}(\mathbf{I} + \mathbf{R})$ splňuje (5.11), tedy je ortogonální projektor.

Splňuje-li nějaká matice $\mathbf{R}$ podmínky (5.21), pak matice $-\mathbf{R}$ je také splňuje. Z obrázku je patrno, že je-li $\mathbf{R}$ ortogonální reflektor okolo podprostoru X, pak $-\mathbf{R}$ je ortogonální reflektor okolo podprostoru $X^\perp$.

Nejznámější je případ, kdy ortogonálně zrcadlíme okolo nadroviny. Má-li tato nadrovina normálový vektor $\mathbf{a}$ s jednotkovou normou, pak ortog. projektor na nadrovinu je $\mathbf{P} = \mathbf{I} - \mathbf{a}\mathbf{a}^T$ a ortog. reflektor kolem nadroviny je tedy $\mathbf{R} = \mathbf{I} - 2\mathbf{a}\mathbf{a}^T$. Tato matice je známa jako *elementární reflektor* neboli *Householderova matice*.

5.2.3 ($\star$) Obecné projektory a reflektory

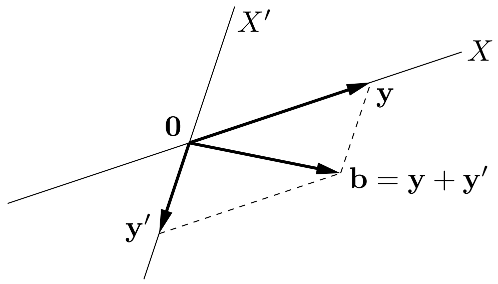
Ortogonální projekce a projektory lze elegantním způsobem zobecnit, což zde uvedeme jako rozšiřující nepovinnou poznámku.

Dvojice podprostorů $X, X' \subseteq \mathbb{R}^m$ se nazývá *komplementární*, splňuje-li podmínky

$$X \cap X' = \{\mathbf{0}\}, \tag{5.22a}$$

$$\dim X + \dim X' = m. \tag{5.22b}$$

V tom případě pro každý vektor $\mathbf{b} \in \mathbb{R}^m$ existuje právě jedna dvojice vektorů $\mathbf{y} \in X$ a $\mathbf{y}' \in X'$ tak, že $\mathbf{b} = \mathbf{y} + \mathbf{y}'$ (důkaz tohoto tvrzení neuvádíme). Vektor $\mathbf{y}$ nazveme (obecná) *projekce* vektoru $\mathbf{b}$ na podprostor X ve směru podprostoru X'. Symetricky, $\mathbf{y}'$ je projekce vektoru $\mathbf{b}$ na X' ve směru X.



Příkladem projekce v $\mathbb{R}^3$ je stín vržený na zem předmětem osvětleným sluncem. Podprostor X je rovina země a podprostor X' je přímka rovnoběžná se slunečními paprsky (předpokládáme, že slunce je nekonečně daleko a tedy všechny paprsky jsou rovnoběžné). V tomto příkladu zanedbáváme zákryty, představte si např., že předmět je mrak izolovaných bodů.

(Obecný) *projektor* se v lineární algebře definuje jako čtvercová matice $\mathbf{P}$ splňující

$$\mathbf{P}^2 = \mathbf{P}. \tag{5.23}$$

Platí následující tvrzení o projektorech (uvádíme bez důkazů, pokuste se o ně!):

- Pro každý projektor jsou podprostory $\operatorname{rng} \mathbf{P}$ a $\operatorname{null} \mathbf{P}$ komplementární a platí $\operatorname{null} \mathbf{P} = \operatorname{rng}(\mathbf{I} - \mathbf{P})$.
- Tedy $\mathbf{y} = \mathbf{P}\mathbf{b}$ je projekcí vektoru $\mathbf{b}$ na podprostor $\operatorname{rng} \mathbf{P}$ ve směru podprostoru $\operatorname{null} \mathbf{P}$.
- Pro každou komplementární dvojici podprostorů X, X' existuje právě jeden projektor tak, že $\operatorname{rng} \mathbf{P} = X$ a $\operatorname{null} \mathbf{P} = X'$.
- Matice $\mathbf{P}$ je projektor, právě když $\mathbf{P} = \mathbf{B}\mathbf{A}$ pro nějaké matice $\mathbf{A}, \mathbf{B}$ splňující $\mathbf{A}\mathbf{B} = \mathbf{I}$. V tom případě platí $\operatorname{rng} \mathbf{P} = \operatorname{rng} \mathbf{B}$ a $\operatorname{null} \mathbf{P} = \operatorname{null} \mathbf{A}$.

Podobně jako projektory lze zobecnit i reflektory. Je-li $\mathbf{P}$ projektor na podprostor X ve směru podprostoru X', pak matice (5.20) je (obecný) *reflektor* okolo X ve směru X'. Reflektory jsou charakterizovány podmínkou

$$\mathbf{R}^2 = \mathbf{I}. \tag{5.24}$$

Tedy zrcadlení je *involutorní lineární transformace.*

Speciálním případem projekce a reflexe jsou ortogonální projekce a reflexe, kdy $X' = X^\perp$. Ortogonální projektor je definován jako projektor splňující $\operatorname{null}\mathbf{P} \perp \operatorname{rng}\mathbf{P}$. Pro každý projektor $\mathbf{P}$ platí, že $\operatorname{null}\mathbf{P} \perp \operatorname{rng}\mathbf{P}$ právě když $\mathbf{P}^T = \mathbf{P}$. Z toho plyne, že projektor je ortogonální právě když kromě (5.23) navíc splňuje $\mathbf{P}^T = \mathbf{P}$. Projekce, která není ortogonální, se pak nazývá *šikmá* (ang. *oblique*).

5.3 Některé aplikace úlohy nejmenších čtverců

5.3.1 Lineární regrese

Regrese je modelování funkční závislosti nějaké proměnné na jiné proměnné. Modelujme závislost proměnné $y \in \mathbb{R}$ na proměnné $x \in X$ (kde X je libovolná množina) regresní funkcí

$$y = f(x, \boldsymbol{\theta}),$$

která je známa až na parametry $\boldsymbol{\theta} \in \mathbb{R}^n$. Je dán soubor dvojic (x_i, y_i), $i = 1, \ldots, m$, kde měření $y_i \in \mathbb{R}$ jsou zatížena chybou. Úkolem je najít parametry $\boldsymbol{\theta}$, aby $y_i \approx f(x_i, \boldsymbol{\theta})$ pro všechna i. Minimalizujeme součet čtverců reziduí, tedy řešíme úlohu

$$\min_{\boldsymbol{\theta} \in \mathbb{R}^n} \sum_{i=1}^m (y_i - f(x_i, \boldsymbol{\theta}))^2. \tag{5.25}$$

Často je regresní funkce taková, že pro každé x je lineární funkcí parametrů $\boldsymbol{\theta}$. V tom případě mluvíme o **lineární regresi**. Taková funkce jde vždy napsat jako lineární kombinace

$$f(x, \boldsymbol{\theta}) = \theta_1 \varphi_1(x) + \cdots + \theta_n \varphi_n(x) = \boldsymbol{\varphi}(x)^T \boldsymbol{\theta} \tag{5.26}$$

nějakých daných funkcí[10] $\varphi_j \colon X \to \mathbb{R}$. Pak

$$\sum_{i=1}^m (y_i - f(x_i, \boldsymbol{\theta}))^2 = \|\mathbf{y} - \mathbf{A}\boldsymbol{\theta}\|^2,$$

kde $\mathbf{y} = (y_1, \ldots, y_m)$ a

$$\mathbf{A} = \begin{bmatrix} \varphi_1(x_1) & \cdots & \varphi_n(x_1) \\ \vdots & \ddots & \vdots \\ \varphi_1(x_m) & \cdots & \varphi_n(x_m) \end{bmatrix}$$

Tedy převedli jsme úlohu (5.25) na tvar (5.2).

Příklad 5.3. Nejjednodušší případ je pro $n = 1$ a konstantní funkci $\varphi_i(x) = 1$. Funkce (5.26) je tedy $f(x, \theta) = \theta$. Úloha (5.25) zní $\min_{\theta \in \mathbb{R}} \sum_i (y_i - \theta)^2$. Snadno spočítáme (udělejte!), že řešením je aritmetický průměr $\theta = \frac{1}{m} \sum_{i=1}^m y_i$ čísel $y_1, \ldots, y_m$. ♦

[10] Funkce φ_j se často nazývají *bázové funkce* (pokud jsou ovšem lineárně nezávislé).

Příklad 5.4. *Proložení bodů polynomem*[11]. Nechť $X = \mathbb{R}$ a $\varphi_j(x) = x^{j-1}$. Pak regresní funkce

$$f(x, \boldsymbol{\theta}) = \theta_1 + \theta_2 x + \theta_3 x^2 + \cdots + \theta_n x^{n-1}$$

je polynom stupně $n-1$ proměnné x. Matice

$$\mathbf{A} = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ & & & \vdots & \\ 1 & x_m & x_m^2 & \cdots & x_m^{n-1} \end{bmatrix}$$

je známá jako *Vandermondova matice.*

Tento případ jde snadno zobecnit na polynomy více proměnných: máme $X = \mathbb{R}^d$ a bázové funkce jsou monomy proměnných $x_1, \ldots, x_d$ (viz §6) až do nějakého stupně. ♦

5.3.2 Vícekriteriální nejmenší čtverce, regularizace

V některých úlohách se hodí minimalizovat více kritérií tvaru $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|$ 'současně'. K tomu se dá přistoupit tak, že minimalizujeme (nezáporně) vážený součet kritérií[12], tedy funkci[13]

$$\mu_1 \|\mathbf{A}_1\mathbf{x} - \mathbf{b}_1\|^2 + \cdots + \mu_k \|\mathbf{A}_k\mathbf{x} - \mathbf{b}_k\|^2 \tag{5.27}$$

kde $\mu_i \geq 0$, $\mathbf{A}_i \in \mathbb{R}^{m_i \times n}$ a $\mathbf{b}_i \in \mathbb{R}^{m_i}$. Minimalizace této funkce není nic nového pod sluncem, protože se dá převést na tvar (5.2). Opravdu, výraz (5.27) je roven (viz Cvičení 5.17)

$$\left\| \begin{bmatrix} \sqrt{\mu_1}(\mathbf{A}_1\mathbf{x} - \mathbf{b}_1) \\ \vdots \\ \sqrt{\mu_k}(\mathbf{A}_k\mathbf{x} - \mathbf{b}_k) \end{bmatrix} \right\|^2 = \left\| \begin{bmatrix} \sqrt{\mu_1}\mathbf{A}_1 \\ \vdots \\ \sqrt{\mu_k}\mathbf{A}_k \end{bmatrix} \mathbf{x} - \begin{bmatrix} \sqrt{\mu_1}\mathbf{b}_1 \\ \vdots \\ \sqrt{\mu_k}\mathbf{b}_k \end{bmatrix} \right\|^2 = \|\mathbf{A}'\mathbf{x} - \mathbf{b}'\|^2, \tag{5.28}$$

kde $\mathbf{A}' \in \mathbb{R}^{m' \times n}$ a $\mathbf{b}' \in \mathbb{R}^{m'}$ kde $m' = m_1 + \cdots + m_k$. Jestliže jsou sloupce matice $\mathbf{A}'$ lineárně nezávislé, optimální $\mathbf{x}$ je rovno (ověřte roznásobením blokových matic!)

$$\mathbf{x} = (\mathbf{A}'^T\mathbf{A}')^{-1}\mathbf{A}'^T\mathbf{b}' = (\mu_1\mathbf{A}_1^T\mathbf{A}_1 + \cdots + \mu_k\mathbf{A}_k^T\mathbf{A}_k)^{-1}(\mu_1\mathbf{A}_1^T\mathbf{b}_1 + \cdots + \mu_k\mathbf{A}_k^T\mathbf{b}_k). \tag{5.29}$$

Speciálně, někdy chceme přibližně řešit soustavu $\mathbf{A}\mathbf{x} = \mathbf{b}$ a zároveň chceme, aby norma řešení $\mathbf{x}$ nebyla moc velká. To lze formulovat jako

$$\min_{\mathbf{x} \in \mathbb{R}^n} \big(\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 + \mu\|\mathbf{x}\|^2 \big). \tag{5.30}$$

pro zvolenou váhu $\mu > 0$. Přidání členu $\mu\|\mathbf{x}\|^2$ se říká (Tichonovova) **regularizace** úlohy (5.2). Dosazením do vzorečku (5.29) ukážeme (proveďte!), že optimální řešení je rovno $\mathbf{x} = \mathbf{A}_\mu^+\mathbf{b}$ kde

$$\mathbf{A}_\mu^+ = (\mathbf{A}^T\mathbf{A} + \mu\mathbf{I})^{-1}\mathbf{A}^T \tag{5.31}$$

je 'regularizovaná pseudoinverze'[14] matice $\mathbf{A}$. Důležité je, že matice $\mathbf{A}^T\mathbf{A} + \mu\mathbf{I}$ regulární pro každé $\mu > 0$ a $\mathbf{A} \in \mathbb{R}^{m \times n}$ (viz Cvičení 5.18), tedy $\mathbf{A}_\mu^+$ je vždy definována.

---

[11] Nedejte se zmást tím, že polynom není lineární funkce a přesto jde o lineární regresi. Důležité je, že regresní funkce (5.26) je lineární v parametrech $\boldsymbol{\theta}$.

[12] Minimalizací více kritérií současně se zabývá obor *vícekriteriální optimalizace (multiobjective optimization)*.

[13] Matematicky elegantnější by samozřejmě bylo 'schovat' skaláry μ_i do matic $\mathbf{A}_i$ a vektorů $\mathbf{b}_i$ a tedy je tam nepsat. Odvození minima funkce (5.27) by pak bylo kratší.

[14] Symbol $\mathbf{A}_\mu^+$ zde neoznačuje pseudoinverzi nějaké matice $\mathbf{A}_\mu$, taková matice $\mathbf{A}_\mu$ totiž neexistuje. Zachováváme ale značení křížkem v horním indexu, abychomn zdůraznili souvislost s pseudoinverzí.

5.4 Řešení s nejmenší normou

Předpokládejme nyní, že soustava (5.1) je nedourčená, tj. má nekonečně mnoho řešení. Je často užitečné z této množiny řešení vybrat jediné podle nějakého kritéria. Přirozeným kritériem je minimalizovat eukleidovskou normu řešení (tedy jeho vzdálenost od počátku), což vede na úlohu

$$\min\{\, \|\mathbf{x}\|^2 \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{A}\mathbf{x} = \mathbf{b}\,\}. \tag{5.32}$$

Místo normy $\|\mathbf{x}\|$ opět minimalizujeme její čtverec. Tato úloha je známa jako řešení nehomogenní soustavy lineárních rovnic **s nejmenší normou** (*least norm solution*). Podotkněme, že někdy je vhodné použít jiná kritéria než nejmenší eukleidovskou normu, viz např. Cvičení 11.18.

Příklad 5.5. Soustava dvou rovnic o třech neznámých

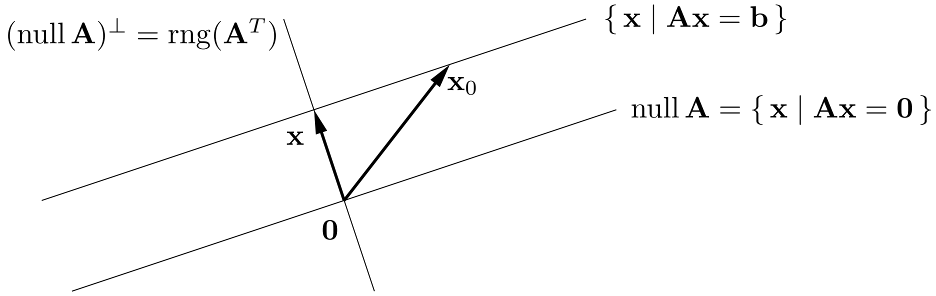
$$\begin{aligned} x + 2y + z &= 1 \\ -x + y + 2z &= 2 \end{aligned}$$

je nedourčená, tj. má nekonečně mnoho řešení. Její řešení s nejmenší normou je takové řešení, které minimalizuje číslo $x^2 + y^2 + z^2$. ♦

Množinu řešení soustavy (5.1) lze psát (viz důkaz Věty 3.13) jako

$$\{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} = \mathbf{b}\,\} = \operatorname{null}\mathbf{A} + \mathbf{x}_0, \tag{5.33}$$

kde $\mathbf{x}_0$ je libovolné (partikulární) řešení soustavy, tedy $\mathbf{A}\mathbf{x}_0 = \mathbf{b}$. Množina (5.33) je afinní podprostor $\mathbb{R}^n$, je to lineární podprostor $\operatorname{null}\mathbf{A}$ posunutý o $\mathbf{x}_0$. Viz obrázek:



Vektory $\mathbf{x}$ a $\mathbf{x}_0$ jsou dvě různá řešení soustavy, ale pouze $\mathbf{x}$ má nejmenší normu. Řešení $\mathbf{x}$ má nejmenší normu právě tehdy, když $\mathbf{x} \perp \operatorname{null}\mathbf{A}$, tj. $\mathbf{x} \in (\operatorname{null}\mathbf{A})^\perp = \operatorname{rng}(\mathbf{A}^T)$ (viz (4.9b)). Neboli musí existovat vektor $\mathbf{y} \in \mathbb{R}^m$ tak, že $\mathbf{x} = \mathbf{A}^T\mathbf{y}$. Pro vyřešení úlohy (5.32) tedy musíme vyřešit soustavu rovnic

$$\mathbf{A}^T\mathbf{y} = \mathbf{x}, \tag{5.34a}$$
$$\mathbf{A}\mathbf{x} = \mathbf{b}. \tag{5.34b}$$

To je soustava $m + n$ rovnic o $m + n$ neznámých $\mathbf{x}, \mathbf{y}$.

Vyřešme tuto soustavu za předpokladu, že matice $\mathbf{A}$ má lineárně nezávislé řádky. Dosazením $\mathbf{x}$ do druhé rovnice obdržíme $\mathbf{A}\mathbf{A}^T\mathbf{y} = \mathbf{b}$. Tato soustava má řešení, jestliže má soustava $\mathbf{A}\mathbf{x} = \mathbf{b}$ řešení, protože pak $\mathbf{b} \in \operatorname{rng}\mathbf{A} = \operatorname{rng}(\mathbf{A}\mathbf{A}^T)$. Protože $\mathbf{A}$ má lineárně nezávislé řádky, dle (5.7) je matice $\mathbf{A}\mathbf{A}^T$ regulární a tedy $\mathbf{y} = (\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b}$. Dosazením do první rovnice dostaneme $\mathbf{x} = \mathbf{A}^+\mathbf{b}$, kde

$$\mathbf{A}^+ = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1} \tag{5.35}$$

se nazývá **pseudoinverze** matice $\mathbf{A}$ s lineárně nezávislými řádky. Je to jedna z pravých inverzí matice $\mathbf{A}$ (ověřte!).

Je poučné odvodit tento výsledek i trochu jinak. Z obrázku je patrno, že řešení $\mathbf{x}$ má nejmenší normu právě tehdy, když je ortogonální projekcí vektoru $\mathbf{x}_0$ na podprostor $(\operatorname{null}\mathbf{A})^\perp = \operatorname{rng}(\mathbf{A}^T)$. Ortogonální projektor na podprostor reprezentovaný svou bází je dán vztahem (5.10), zde ovšem promítáme na $\operatorname{rng}(\mathbf{A}^T)$, tedy musíme vzorec použít s $\mathbf{A}^T$ místo s $\mathbf{A}$. Tedy

$$\mathbf{x} = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{x}_0 = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b} = \mathbf{A}^+\mathbf{b}. \tag{5.36}$$

Protože ortogonální projekce je určena jednoznačně, vidíme, že úloha (5.32) má právě jedno optimální řešení (pokud je přípustná, tj. soustava $\mathbf{A}\mathbf{x} = \mathbf{b}$ má aspoň jedno řešení).

Zmíníme ještě třetí úvahu, jak dospět ke vzorečku (5.35). Úlohu (5.32) si lze neformálně představit jako minimalizaci výrazu $\|\mathbf{x}\|^2 + \frac{1}{\mu}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$ pro velmi malé kladné μ (tedy velmi velké $\frac{1}{\mu}$). To je ale totéž, jako regularizovaná úloha nejmenších čtverců (5.30) pro velmi malé kladné μ. Tedy můžeme očekávat, že

$$\mathbf{A}^+ = \lim_{\mu\to 0^+} \mathbf{A}^+_\mu, \tag{5.37}$$

kde matice $\mathbf{A}^+_\mu$ je definována v (5.31). Ale pro každé $\mu > 0$ platí

$$\mathbf{A}^+_\mu = (\mathbf{A}^T\mathbf{A} + \mu\mathbf{I})^{-1}\mathbf{A}^T = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T + \mu\mathbf{I})^{-1}. \tag{5.38}$$

To dokážeme snadno: pro $\mu > 0$ jsou obě matice $\mathbf{A}^T\mathbf{A} + \mu\mathbf{I}$ a $\mathbf{A}\mathbf{A}^T + \mu\mathbf{I}$ regulární (dle Cvičení 5.18) a tedy jimi můžeme rovnost (5.38) vynásobit; zbytek důkazu je pak jen roznásobení závorek. Má-li ovšem matice $\mathbf{A}$ lineárně nezávislé řádky, je matice $\mathbf{A}\mathbf{A}^T$ regulární a limita (5.37) je proto rovna (5.35).

Slovem 'pseudoinverze' jsme dosud označili dvě různé operace: jednu pro matice s lineárně nezávislými sloupci (tj. hodností n) dle vzorce (5.8) a druhou pro matice s lineárně nezávislými řádky (tj. hodností m) dle vzorce (5.35). Tyto dva vzorce tedy dohromady definují pseudoinverzi libovolné matice s plnou hodností $\min\{m, n\}$:

$$\mathbf{A}^+ = \begin{cases} (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T & \text{jestliže } \operatorname{rank}\mathbf{A} = n, \\ \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1} & \text{jestliže } \operatorname{rank}\mathbf{A} = m. \end{cases} \tag{5.39}$$

Pssudoinverzi lze ovšem definovat pro matici libovolné hodnosti, kde se pak (5.39) jeví jako speciální případy. To zmíníme (nepovinně) v §5.5.

5.5 ($\star$) Pseudoinverze obecné matice

Zatím jsme odděleně diskutovali případy, kdy soustava má žádné, jedno, nebo nekonečně mnoho řešení. Překvapivě, tyto tři případy lze spojit do jediné formulace. Zopakujme, že optimální řešení úlohy (5.2) jsou právě řešení soustavy normálních rovnic (5.4). Co když je ale sama soustava (5.4) nedourčená? Pak můžeme hledat její řešení s nejmenší normou, tj. řešit úlohu

$$\min\{\, \|\mathbf{x}\|^2 \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b} \,\}. \tag{5.40}$$

Jelikož soustava (5.4) má vždy řešení, úloha (5.40) má vždy právě jedno řešení (jak jsme ukázali v §5.4). Toto řešení, $\mathbf{x}^*$, má následující vlastnosti:

- Má-li soustava $\mathbf{Ax} = \mathbf{b}$ jediné řešení, $\mathbf{x}^*$ je toto řešení.
- Nemá-li soustava $\mathbf{Ax} = \mathbf{b}$ řešení, $\mathbf{x}^*$ je její přibližné řešení ve smyslu nejmenších čtverců, tj. optimální řešení úlohy (5.2). Pokud ovšem úloha (5.2) má více než jedno (tedy nekonečně mnoho) optimálních řešení, $\mathbf{x}^*$ je optimální řešení úlohy (5.2) které má nejmenší normu.
- Má-li soustava $\mathbf{Ax} = \mathbf{b}$ nekonečně mnoho řešení, $\mathbf{x}^*$ je řešení této soustavy s nejmenší normou, tj. optimální řešení problému (5.32).

Podle §5.4 je $\mathbf{x}$ optimální řešení úlohy (5.40) právě tehdy, když existuje $\mathbf{y}$ takové, že

$$\mathbf{A}^T\mathbf{Ay} = \mathbf{x}, \tag{5.41a}$$
$$\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}. \tag{5.41b}$$

Tuto soustavu obecně nejde řešit pomocí inverze, protože $\mathbf{A}^T\mathbf{A}$ nemusí být regulární. Z tvaru soustavy (5.41) ale plyne, že $\mathbf{x} = \mathbf{A}^+\mathbf{b}$ pro nějakou matici $\mathbf{A}^+$, která nezávisí na $\mathbf{b}$. Tutu matici nazýváme **pseudoinverze** (přesněji *Moore-Penroseova pseudoinverze*) matice $\mathbf{A}$. Když $\mathbf{A}$ má lineárně nezávislé sloupce, $\mathbf{A}^+$ je (5.8). Když $\mathbf{A}$ má lineárně nezávislé řádky, $\mathbf{A}^+$ je (5.35). Když ovšem $\mathbf{A}$ nemá plnou hodnost, $\mathbf{A}^+$ je nutno počítat jinak. Elegantně se to udělá pomocí singulárního rozkladu (SVD), což ukážeme už zde (vraťte se k tomu po přečtení §7.4):

Věta 5.4. Nechť $\mathbf{A} = \mathbf{USV}^T$ je rank-minimální SVD matice $\mathbf{A}$, tj. $\mathbf{U}, \mathbf{V}$ mají ortonormální sloupce a $\mathbf{S}$ je diagonální regulární. Pak

$$\mathbf{A}^+ = \mathbf{VS}^{-1}\mathbf{U}^T. \tag{5.42}$$

Důkaz. Stačí dokázat, že $\mathbf{x} = \mathbf{A}^+\mathbf{b} = \mathbf{VS}^{-1}\mathbf{U}^T\mathbf{b}$ a $\mathbf{y} = \mathbf{VS}^{-2}\mathbf{V}^T\mathbf{x}$ splňují (5.41):

$$\begin{aligned} \mathbf{A}^T\mathbf{A} &= \mathbf{VS}^T\mathbf{U}^T\mathbf{USV}^T = \mathbf{VS}^2\mathbf{V}^T, \\ \mathbf{A}^T\mathbf{Ay} &= \mathbf{VS}^2\mathbf{V}^T\mathbf{y} = \mathbf{VS}^2\mathbf{V}^T\mathbf{VS}^{-2}\mathbf{V}^T\mathbf{x} = \mathbf{x}, \\ \mathbf{A}^T\mathbf{Ax} &= \mathbf{VS}^2\mathbf{V}^T\mathbf{x} = \mathbf{VS}^2\mathbf{V}^T\mathbf{VS}^{-1}\mathbf{U}^T\mathbf{b} = \mathbf{VSU}^T\mathbf{b} = \mathbf{A}^T\mathbf{b}. \end{aligned}$$ ■

K pseudoinverzi obecné matice lze dospět také pomocí úvah o regularizaci. Všimněte si, že regularizovaná úloha (5.30) je ‘něco mezi’ úlohami (5.2) a (5.32). Lze dokázat (což uděláte později ve Cvičení 7.18), že pseudoinverze obecné matice je rovna limitě (5.37), kde $\mathbf{A}^+_\mu$ je definováno v (5.38). Zdůrazněme rozdíl oproti §5.4, kde jsme předpokládali lineární nezávislost řádků matice $\mathbf{A}$ a tedy regularitu matice $\mathbf{AA}^T$. Zde nově tvrdíme, že limita existuje i v případě, kdy matice $\mathbf{A}$ nemá plnou hodnost a tedy ani jedna z matic $\mathbf{A}^T\mathbf{A}$ a $\mathbf{AA}^T$ není regulární.

Nakonec uveďme, že pseudoinverze obecné matice $\mathbf{A}$ je jednoznačně určená podmínkami

$$\mathbf{AA}^+\mathbf{A} = \mathbf{A}, \quad \mathbf{A}^+\mathbf{AA}^+ = \mathbf{A}^+, \quad (\mathbf{AA}^+)^T = \mathbf{AA}^+, \quad (\mathbf{A}^+\mathbf{A})^T = \mathbf{A}^+\mathbf{A}. \tag{5.43}$$

Toto tvrzení nebudeme dokazovat, ale aspoň ověřte, že matice (5.8), (5.35) a (5.42) tyto podmínky splňují (více viz Cvičení 5.21).

Pro praktické řešení lineárních soustav se pseudoinverze obecné matice moc nepoužívá, protože $\mathbf{A}^+$ je nespojitá funkce matice $\mathbf{A}$.

5.6 Cvičení

5.1. Máme soustavu $\mathbf{Ax} = \mathbf{b}$, kde $\mathbf{A} \in \mathbb{R}^{m\times n}$ a $\mathbf{b} \neq \mathbf{0}$. Dokažte nebo vyvraťte následující tvrzení:

a) Pokud $m < n$, pak soustava má vždy řešení.
b) Pokud $m > n$, pak soustava nemá nikdy řešení.
c) Pokud $m < n$ a $\mathbf{A}$ má plnou hodnost, pak soustava má vždy nekonečně mnoho řešení.

5.2. Vyřešte (možno použít počítač) soustavu

$$\begin{bmatrix} 1 & 0 & -1 \\ 1 & 2 & 1 \\ 1 & 1 & -3 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

přibližně ve smyslu nejmenších čtverců pomocí (a) pseudoinverze, (b) QR rozkladu.

5.3. Formulujte jako přibližné řešení soustavy $\mathbf{Pu} = \mathbf{q}$ ve smyslu nejmenších čtverců, tedy jako úlohu $\min_{\mathbf{u}} \|\mathbf{Pu} - \mathbf{q}\|^2$. Jako výsledek napište matice $\mathbf{P}, \mathbf{q}, \mathbf{u}$. Pokud existuje jednoduchý vzorec pro řešení (jak pro optimální hodnotu tak optimální argument), napište je.

a) Příklad 1.12.
b) Hledá se vzdálenost bodu $\mathbf{y} \in \mathbb{R}^n$ od přímky $\{\, \mathbf{a} + t\mathbf{s} \mid t \in \mathbb{R} \,\}$ kde $\mathbf{a}, \mathbf{s} \in \mathbb{R}^n$.
c) Příklad 1.9.
d) Máme množinu m přímek v $\mathbb{R}^n$, kde i-tá přímka je množina $\{\, \mathbf{a}_i + t\mathbf{s}_i \mid t \in \mathbb{R} \,\}$ pro dané $\mathbf{a}_i, \mathbf{s}_i \in \mathbb{R}^n$. Hledá se bod $\mathbf{y} \in \mathbb{R}^n$, jehož součet čtverců vzdáleností od přímek je minimální.
e) Máme m nadrovin v prostoru $\mathbb{R}^n$, kde i-tá nadrovina má rovnici $\mathbf{a}_i^T \mathbf{x} = b_i$ pro dané $\mathbf{a}_i \in \mathbb{R}^n$ a $b_i \in \mathbb{R}$. Hledá se bod $\mathbf{y} \in \mathbb{R}^n$, který minimalizuje součet čtverců vzdáleností od jednotlivých nadrovin.
f) V prknu je n děr o souřadnicích $x_1, \ldots, x_n \in \mathbb{R}$, všechny v jedné přímce. Naměříme metrem vzdálenosti $d_{ij} = x_j - x_i$ pro vybrané dvojice $(i, j) \in E$, kde množina $E \subseteq \{1, \ldots, n\} \times \{1, \ldots, n\}$ je dána. Přitom dvojice jsou vybrané tak, že vždy $x_j > x_i$. Ze vzdáleností d_{ij} chceme spočítat souřadnice $x_1, \ldots, x_n$. Odpovězte dále na otázky:

   1. Kolik řešení má soustava $\mathbf{Pu} = \mathbf{q}$? Odpověď dokažte a interpretujte.
   2. Jsou sloupce $\mathbf{P}$ lineárně nezávislé?

   Diskutujte obě otázky pro případ, že měření jsou přesná, a pro případ, že měření jsou zatížená nepřesnostmi.
g) Závislost výkonu P kotle na průtoku G plynu a průřezu S díry na přívod vzduchu je modelována funkcí $\hat{P}(G, S) = G(a_1 + a_2 S + a_3 10^{G+S} + a_4 10^{-S})$. Odhadujeme koeficienty $a_1, \ldots, a_4$ z naměřených trojic $(G_1, S_1, P_1), \ldots, (G_n, S_n, P_n)$.
h) Známý průběh ceny akcie jisté firmy po dnech je daný posloupností $p_1, \ldots, p_k$. Chceme předpovídat cenu akcie den dopředu. Tuto cenu modelujeme *autoregresní funkcí* $\hat{p}_{t+1} = \beta_1 + \beta_2 p_t + \beta_3 p_{t-1}$. Odhadněte koeficienty β_i tak, aby celková chyba predikce $\sum_{t=3}^{k} (p_t - \hat{p}_t)^2$ byla na onom známém průběhu ceny minimální.

5.4. V problému *vážených nejmenších čtverců* chceme najít $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ minimalizující funkci

$$f(\mathbf{x}) = \sum_{i=1}^{m} w_i \Big(\sum_{j=1}^{n} a_{ij} x_j - b_i \Big)^2$$

kde w_i jsou nezáporné váhy. Napište funkci v maticovém tvaru, k čemuž zaveďte diagonální matici $\mathbf{W} = \operatorname{diag}(w_1, \ldots, w_m)$. Napište normální rovnici a pseudoinverzi pro tento případ.

5.5. Čtvercová matice $\mathbf{A}$ se nazývá *normální*, když $\mathbf{A}^T\mathbf{A} = \mathbf{A}\mathbf{A}^T$. Příkladem je symetrická nebo antisymetrická matice. Dokažte, že pro normální matice platí $(\operatorname{rng} \mathbf{A})^\perp = \operatorname{null} \mathbf{A}$.

5.6. Jaké bude řešení normálních rovnic (5.4) v případě, že $\mathbf{A}$ má lineárně nezávislé sloupce a $\mathbf{b} \perp \operatorname{rng} \mathbf{A}$? Vyřešte geometrickou úvahou (vzpomeňte si na ortogonální projekci a hleďte na obrázek v §5.1!) a pak zkuste dokázat algebraicky.

5.7. Ortogonální projekci $\mathbf{y}$ vektoru $\mathbf{b}$ na přímku $\operatorname{span}\{\mathbf{a}\}$ můžete naivně spočítat tak, že nejdříve spočítáme projektor $\mathbf{P}$ vzorečkem (5.13) a pak spočítáme $\mathbf{y} = \mathbf{P}\mathbf{b}$. Je to dobrý nápad, když vektor $\mathbf{b}$ má mnoho komponent? Kolik potřebujeme paměti a FLOPů? Jde to dělat úsporněji? Jak s tím souvisí rovnost $(\mathbf{a}\mathbf{a}^T)\mathbf{b} = (\mathbf{a}^T\mathbf{b})\mathbf{a}$? Zobecněte na případ, kdy promítáme na podprostor $\operatorname{rng} \mathbf{A}$ s použitím projektoru (5.10) případně (5.15).

5.8. Máme vektory $\mathbf{u} = (2, 1, -3)$ a $\mathbf{v} = (1, -1, 1)$. Najděte ortogonální projekci vektoru $(2, 0, 1)$ na podprostor (a) $\operatorname{span}\{\mathbf{u}\}$, (b) $(\operatorname{span}\{\mathbf{u}\})^\perp$, (c) $\operatorname{span}\{\mathbf{u}, \mathbf{v}\}$, (d) $(\operatorname{span}\{\mathbf{u}, \mathbf{v}\})^\perp$.

5.9. Nechť $X = \operatorname{span}\{\, (-\frac{3}{5}, 0, \frac{4}{5}, 0), (0, 0, 0, 1), (\frac{4}{5}, 0, \frac{3}{5}, 0) \,\}$. Najděte projektory na podprostor X a podprostor $X^\perp$.

5.10. Nechť $\mathbf{A} = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 4 & 1 \\ 1 & 2 & 0 \end{bmatrix}$. Najděte ortogonální projekci vektoru $(1, 1, 1)$ na podprostor (a) $\operatorname{rng} \mathbf{A}$, (b) $\operatorname{null} \mathbf{A}$, (c) $\operatorname{rng}(\mathbf{A}^T)$, (d) $\operatorname{null}(\mathbf{A}^T)$.

5.11. Nulový prostor projektoru je typicky netriviální, tedy projektor $\mathbf{P}$ je singulární matice. Kdy je $\mathbf{P}$ regulární? Jaká je v tom případě matice $\mathbf{A}$ ve vzorci (5.10) a podprostor $X = \operatorname{rng} \mathbf{A}$? Jaký je geometrický význam této situace?

5.12. Najděte projektor na podprostor $\operatorname{null} \mathbf{A}$, kde $\mathbf{A}$ má lin. nezávislé řádky.

5.13. Báze podprostoru, na který promítá ortogonální projektor (5.10), jsou sloupce matice $\mathbf{A}$. Dokažte, že projektor se nezmění, vezmeme-li jinou bázi podprostoru.

5.14. Je složení dvou ortogonálních projekcí ortogonální projekce? Odpověď dokažte.

5.15. Nechť $\mathbf{P} \in \mathbb{R}^{n \times n}$ je ortogonální projektor. Dokažte, že

a) Prvky matice $\mathbf{P}$ splňují $|p_{ij}| \le 1$.
b) Diagonální prvky matice $\mathbf{P}$ jsou nezáporné.
c) $\|\mathbf{P}\|^2 = n$ (kde $\|\cdot\|$ je Frobeniova norma matice).[15]

5.16. V tomto cvičení předpokládejte, že matice $\mathbf{A}$ má lineárně nezávislé řádky a matice $\mathbf{U}$ má ortonormální sloupce (tj. $\mathbf{U}^T\mathbf{U} = \mathbf{I}$). Najděte (co možná jednoduchý) vzorec pro

a) vzdálenost bodu $\mathbf{v}$ od afinního podprostoru $A = \mathbf{y} + X$, známe-li ortogonální projektor na podprostor X,

[15] Tato vlastnost je poněkud nepřirozená. Je způsobená tím, že Frobeniova norma nesplňuje axiom maticové normy $\|\mathbf{A}\mathbf{B}\| \le \|\mathbf{A}\|\|\mathbf{B}\|$ (viz zponámka pod čarou v §2.1.7). Pro maticové normy lze ukázat, že $\|\mathbf{P}\| = 1$.

b) vzdálenost počátku $\mathbf{0}$ od nadroviny $\{\, \mathbf{x} \mid \mathbf{a}^T\mathbf{x} = b \,\}$,

c) vzdálenost počátku $\mathbf{0}$ od afinního podprostoru $\{\, \mathbf{x} \mid \mathbf{A}\mathbf{x} = \mathbf{b} \,\}$, kde $\mathbf{A}$ má lin. nezávislé řádky,

d) vzdálenost bodu $\mathbf{v}$ od nadroviny $\{\, \mathbf{x} \mid \mathbf{a}^T\mathbf{x} = b \,\}$,

e) vzdálenost bodu $\mathbf{v}$ od podprostoru $\operatorname{rng} \mathbf{U}$,

f) vzdálenost bodu $\mathbf{v}$ od podprostoru $\operatorname{null}(\mathbf{U}^T) = \{\, \mathbf{x} \mid \mathbf{U}^T\mathbf{x} = \mathbf{0} \,\}$,

g) vzdálenost počátku $\mathbf{0}$ od afinního podprostoru $\{\, \mathbf{x} \mid \mathbf{U}^T\mathbf{x} = \mathbf{b} \,\}$,

h) vzdálenost bodu $\mathbf{v}$ od afinního podprostoru $\{\, \mathbf{x} \mid \mathbf{U}^T\mathbf{x} = \mathbf{b} \,\}$.

5.17. Pro vektory $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ ukažte, že $\left\| \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \right\|^2 = \|\mathbf{a}\|^2 + \|\mathbf{b}\|^2$.

5.18. Dokažte, že matice $\mathbf{A}^T\mathbf{A} + \mu\mathbf{I}$ je regulární pro každou matici $\mathbf{A}$ a každé $\mu > 0$.

5.19. Spočítejte pseudoinverzi (a) nenulového skaláru (tj. matice s jedním řádkem a jedním sloupcem), (b) nenulového sloupcového vektoru (tj. matice s jedním sloupcem), (c) nenulového řádkového vektoru (tj. matice s jedním řádkem).

5.20. Afinní podprostor $A \subseteq \mathbb{R}^m$ lze reprezentovat jako $A = \mathbf{x} + \operatorname{rng} \mathbf{U}$, kde $\mathbf{x} \in A$ a $\mathbf{U} \in \mathbb{R}^{m\times n}$. Bod $\mathbf{x}$ ovšem není touto reprezentací určen jednoznačně: zvolíme-li libovolný jiný bod $\mathbf{y} \in A$, dle Věty 3.12 je $\mathbf{y} + \operatorname{rng} \mathbf{U} = A$. Tuto nejednoznačnost můžeme odstranit přidáním podmínky, že norma $\|\mathbf{x}\|$ je minimální, tj. bod $\mathbf{x}$ je průmětem počátku na podprostor A. Jsou-li dány $\mathbf{y} \in A$ a $\mathbf{U} \in \mathbb{R}^{m\times n}$, najděte co nejjednodušší vzorec pro takový bod $\mathbf{x}$.

5.21. Dokažte následující vlastnosti pseudoinverze ze vztahů (5.8) a (5.35) pro libovolné (úzké, široké nebo čtvercové) matice plné hodnosti:

a) $\mathbf{A}^+ = \mathbf{A}^{-1}$, právě když $\mathbf{A}$ je čtvercová

b) $(\mathbf{A}^+)^+ = \mathbf{A}$

c) $(\mathbf{A}^T)^+ = (\mathbf{A}^+)^T$

d) $\mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A}$, $\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+$

e) $(\mathbf{A}\mathbf{A}^+)^T = \mathbf{A}\mathbf{A}^+$, $(\mathbf{A}^+\mathbf{A})^T = \mathbf{A}^+\mathbf{A}$

f) $\mathbf{A}^T = \mathbf{A}^T\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+\mathbf{A}\mathbf{A}^T$

g) $(\mathbf{A}^T\mathbf{A})^+ = \mathbf{A}^+(\mathbf{A}^T)^+$, $(\mathbf{A}\mathbf{A}^T)^+ = (\mathbf{A}^T)^+\mathbf{A}^+$

5.22. V této kapitole se nám objasnil význam druhé části Tvrzení 4.7 z předešlé kapitoly. Vysvětlete její souvislost se vzorcem (5.15) a Tvrzením 5.3.

5.23. Dokažte následující tvrzení: soustava rovnic $\mathbf{A}\mathbf{x} = \mathbf{b}$ má řešení právě tehdy, když neexistuje vektor $\mathbf{y}$ tak, že $\mathbf{y}^T\mathbf{A} = \mathbf{0}$ a $\mathbf{y}^T\mathbf{b} \neq 0$.

Nápověda a řešení

5.1.a) Neplatí. Příklad: $m = 1$, $n = 2$, $\mathbf{A} = \begin{bmatrix} 0 & 0 \end{bmatrix}$, $\mathbf{b} = 1$.

5.1.b) Neplatí. Příklad: $m = 2$, $n = 1$, $\mathbf{A} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

5.1.c) Platí. Matice $\mathbf{A}$ má hodnost m, tedy lineárně nezávislé řádky, tedy $\operatorname{rng} \mathbf{A} = \mathbb{R}^m$, tedy soustava má řešení. Navíc má $\mathbf{A}$ netriviální nulový prostor, tedy má nekonečně mnoho řešení.

5.2. $(x_1, x_2, x_3) = (2, 1, 0)/3$

5.3.a) Máme napsat funkci (1.21) ve tvaru $f(\mathbf{u}) = \|\mathbf{Pu} - \mathbf{q}\|^2$. Dle Cvičení 5.17 je $\sum_{i=1}^m \|\mathbf{x} - \mathbf{a}_i\|^2 = \left\| \begin{bmatrix} \mathbf{x} - \mathbf{a}_1 \\ \vdots \\ \mathbf{x} - \mathbf{a}_m \end{bmatrix} \right\|^2 = \left\| \begin{bmatrix} \mathbf{I}_n \\ \vdots \\ \mathbf{I}_n \end{bmatrix} \mathbf{x} - \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_m \end{bmatrix} \right\|^2$, tedy $\mathbf{u} = \mathbf{x}$, $\mathbf{P} = \begin{bmatrix} \mathbf{I}_n \\ \vdots \\ \mathbf{I}_n \end{bmatrix} \in \mathbb{R}^{mn \times n}$, $\mathbf{q} = \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_m \end{bmatrix} \in \mathbb{R}^{mn}$.

5.3.c) Máme napsat funkci (1.15) ve tvaru $f(\mathbf{u}) = \|\mathbf{Pu} - \mathbf{q}\|^2$. To je opět snadné: $\mathbf{u} = (\alpha_1, \alpha_2)$, $\mathbf{P} = \begin{bmatrix} \mathbf{b}_1 - \mathbf{a}_1 & \mathbf{a}_2 - \mathbf{b}_2 \end{bmatrix}$, $\mathbf{q} = \mathbf{a}_1 - \mathbf{a}_2$.

5.3.d) Minimalizujte přes proměnné $\mathbf{y}, t_1, \dots, t_m$.

5.3.e) Nejprve si vzpomeňte či odvoďte, jak se spočítá vzdálenost bodu $\mathbf{y}$ od nadroviny $\mathbf{a}^T\mathbf{x} = b$.

5.4. $f(\mathbf{x}) = (\mathbf{Ax} - \mathbf{b})^T\mathbf{W}(\mathbf{Ax} - \mathbf{b}) = \|\mathbf{W}^{1/2}\mathbf{Ax} - \mathbf{W}^{1/2}\mathbf{b}\|^2$.

5.5. Dle (4.9a) a (5.6b) je $(\operatorname{rng} \mathbf{A})^\perp = \operatorname{null}(\mathbf{A}^T) = \operatorname{null}(\mathbf{AA}^T) = \operatorname{null}(\mathbf{A}^T\mathbf{A}) = \operatorname{null} \mathbf{A}$. Jedná se o zobecnění Cvičení 4.16.

5.6. $\mathbf{x} = \mathbf{0}$

5.8. (a) $(2, 1, -3)/14$, (b) $(26, -1, 17)/14$, (c) $(62, -35, 17)/38$, (d) $(14, 35, 21)/38$

5.9. Pokud si všimneme, že vektory jsou ortonormální, hodně nám to ulehčí práci. Projektor na X je $\mathbf{P} = \operatorname{diag}(1, 0, 1, 1)$. Projektor na $X^\perp$ je $\mathbf{P} = \operatorname{diag}(0, 1, 0, 0)$.

5.10. (a) $(1, 1, 1)$, (b) $(0.4, -0.2, 0)$, (c) $(0.6, 1.2, 1)$, (d) $(0, 0, 0)$. Pozor, $\mathbf{A}$ nemá plnou hodnost.

5.11. $\mathbf{A}$ je regulární, tedy $X = \mathbb{R}^m$. Projektor je identita.

5.12. Označíme-li $\operatorname{null} \mathbf{A} = X$, dle Věty 4.3 je $X^\perp = (\operatorname{null} \mathbf{A})^\perp = \operatorname{rng}(\mathbf{A}^T)$. Ortogonální projektor na podprostor $X^\perp$ je $\mathbf{P} = \mathbf{A}^T(\mathbf{AA}^T)^{-1}\mathbf{A}$ dle vzorce (5.10), do kterého jsme ovšem místo $\mathbf{A}$ museli dosadit $\mathbf{A}^T$. Projektor na podprostor $X = (X^\perp)^\perp$ je dle Tvrzení 5.3 tedy $\mathbf{I} - \mathbf{P} = \mathbf{I} - \mathbf{A}^T(\mathbf{AA}^T)^{-1}\mathbf{A}$.

5.13. Dle Cvičení 3.12 jsou všechny možné báze podprostoru X dány sloupci matice $\mathbf{AC}$ pro všechny možné regulární matice $\mathbf{C} \in \mathbb{R}^{n \times n}$ (tedy $\mathbf{C}$ je matice přechodu k jiné bázi). Pak $\tilde{\mathbf{A}}(\tilde{\mathbf{A}}^T\tilde{\mathbf{A}})^{-1}\tilde{\mathbf{A}}^T = \mathbf{AC}(\mathbf{C}^T\mathbf{A}^T\mathbf{AC})^{-1}\mathbf{C}^T\mathbf{A}^T = \mathbf{ACC}^{-1}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{C}^{-T}\mathbf{C}^T\mathbf{A}^T = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$

5.14. Ne vždy. Jako důkaz této odpovědi stačí najít matice $\mathbf{P}, \mathbf{Q}$ takové, že $\mathbf{P}^2 = \mathbf{P}$, $\mathbf{Q}^2 = \mathbf{Q}$ a $(\mathbf{PQ})^2 \neq \mathbf{PQ}$.

5.15.a) Z (5.15) máme $p_{ij} = \mathbf{a}_i^T\mathbf{a}_j$, kde $\mathbf{a}_i$ je i-tý řádek matice $\mathbf{A}$ s ortonormálními sloupci. Protože $\|\mathbf{a}_i\| \leq 1$ (viz Cvičení 4.18), musí být $|\mathbf{a}_i^T\mathbf{a}_j| \leq 1$.

5.15.b) Je $p_{ii} = \mathbf{a}_i^T\mathbf{a}_i \geq 0$

5.15.c) Dosaďte do $\|\mathbf{P}\|^2 = \langle \mathbf{P}, \mathbf{P} \rangle$ z (5.16) a upravujte.

5.16.a) Všechno (tj. bod $\mathbf{v}$ i afinní podprostor A) posuneme o vektor $-\mathbf{y}$, čímž se vzdálenost nezmění a z afinního podprostoru A stane lineární podprostor X (dle Věty 3.12). Hledaná vzdálenost bude tedy vzdálenost bodu $\mathbf{v} - \mathbf{y}$ od podprostoru X. Ta je rovna délce ort. projekce na podprostor $X^\perp$, tj. $\|(\mathbf{I} - \mathbf{P})(\mathbf{v} - \mathbf{y})\|$.

5.16.b) $|b|/\|\mathbf{a}\|$

5.16.c) Čtverec vzdálenosti je roven optimální hodnotě úlohy (5.32), tedy $(\mathbf{A}^+\mathbf{b})^T(\mathbf{A}^+\mathbf{b}) = \mathbf{b}^T(\mathbf{AA}^T)^{-T}\mathbf{AA}^T(\mathbf{AA}^T)^{-1}\mathbf{b} = \mathbf{b}^T(\mathbf{AA}^T)^{-1}\mathbf{b}$.

5.16.d) $|\mathbf{a}^T\mathbf{v} - b|/\|\mathbf{a}\|$

5.16.e) $\|(\mathbf{I} - \mathbf{UU}^T)\mathbf{v}\|$, víc to už zjednodušit nejde.

5.16.f) $\|\mathbf{U}^T\mathbf{v}\|$, viz §5.2.1.

5.16.g) $\|\mathbf{b}\|$, viz §5.2.1.

5.16.h) $\|\mathbf{U}^T\mathbf{v} - \mathbf{b}\|$, viz §5.2.1.

5.17. Ihned plyne ze vzorce $\|\mathbf{a}\|^2 = a_1^2 + \cdots + a_n^2$. Jen si takto rozepište levou a pravou stranu dokazované identity.

5.18. Je $\mathbf{A}^T\mathbf{A} + \mu\mathbf{I} = \mathbf{B}^T\mathbf{B}$, kde $\mathbf{B} = \begin{bmatrix} \mathbf{A} \\ \mu^{1/2}\mathbf{I} \end{bmatrix}$. Matice $\mathbf{B}$ má l.n. sloupce, protože už matice $\mu^{1/2}\mathbf{I}$ je má. Tedy dle (5.7) má $\mathbf{B}^T\mathbf{B}$ plnou hodnost, tedy je regulární.

5.20. $\mathbf{x} = (\mathbf{I} - \mathbf{U}\mathbf{U}^T)\mathbf{y}$

5.21.b) Když $\mathbf{A}$ má l.n. sloupce, dle (5.8) je $\mathbf{A}^+ = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$. Protože $(\mathbf{A}^T\mathbf{A})^{-1}$ je regulární, $\mathbf{A}^+$ má l.n. řádky. Dle (5.35) tedy $\mathbf{A}^{++} = \mathbf{A}^{+T}(\mathbf{A}^+\mathbf{A}^{+T})^{-1} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-T}[(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-T}]^{-1} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-T}[(\mathbf{A}^T\mathbf{A})^{-T}]^{-1} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-T}(\mathbf{A}^T\mathbf{A})^T = \mathbf{A}$. Když $\mathbf{A}$ má l.n. řádky, udělá se to podobně.

5.23. Důkaz jedné implikace je snadný: existuje-li takový vektor $\mathbf{y}$, pak vynásobením soustavy $\mathbf{A}\mathbf{x} = \mathbf{b}$ vektorem $\mathbf{y}$ zleva dostaneme $0 = \mathbf{y}^T\mathbf{A}\mathbf{x} = \mathbf{y}^T\mathbf{b} \neq 0$, tedy soustava nemá řešení.

Pro důkaz druhé implikace předpokládejme, že soustava $\mathbf{A}\mathbf{x} = \mathbf{b}$ nemá řešení (tj. $\mathbf{b} \notin \operatorname{rng}\mathbf{A}$), a hledejme vektor $\mathbf{y}$ s požadovanými vlastnostmi. Tvrdíme, že jeden takový vektor $\mathbf{y}$ je ortogonální projekce vektoru $\mathbf{b}$ na podprostor $(\operatorname{rng}\mathbf{A})^\perp$. Protože $\mathbf{y} \in (\operatorname{rng}\mathbf{A})^\perp$, je $\mathbf{y} \perp \operatorname{rng}\mathbf{A}$ neboli $\mathbf{y}^T\mathbf{A} = \mathbf{0}$. Protože $\mathbf{b} \notin \operatorname{rng}\mathbf{A}$, vektor $\mathbf{b}$ není ortogonální na podprostor $(\operatorname{rng}\mathbf{A})^\perp$ a tedy ani na vektor $\mathbf{y}$, tedy $\mathbf{y}^T\mathbf{b} \neq 0$. Vše vyjasní pohled na obrázek v §5.1, kde ovšem náš vektor $\mathbf{y}$ je označen jako $\mathbf{y}'$.

Kapitola 6

Kvadratické formy a funkce

Ze základní školy znáte polynomy jedné proměnné. Co jsou ale polynomy více proměnných? **Monom** (angl. *monomial*) n proměnných je výraz

$$x_1^{k_1} \cdots x_n^{k_n},$$

kde $k_1, \ldots, k_n \in \mathbb{N} \cup \{0\}$. **Stupeň monomu** je pak číslo $k_1 + \cdots + k_n$. **Polynom** (angl. *polynomial*) n proměnných je lineární kombinace monomů, přičemž **stupeň polynomu** je stupeň jeho monomu (s nenulovým koeficientem) nejvyššího stupně. Např. funkce

$$f(x, y) = x^2 y + xy - 2x + 1 \tag{6.1}$$

je polynom dvou proměnných třetího stupně, kde např. $x^2 y$ je monom třetího stupně a xy je monom druhého stupně. Polynom je **homogenní**, pokud stupně všech jeho monomů jsou stejné. Polynom (6.1) není homogenní, ale např. polynom $f(x, y) = x^2 y - 5y^3$ je homogenní stupně tři.

Vidíme, že afinní funkce (3.27) je jen jiný název pro polynom prvního stupně a lineární funkce (3.8) (také zvaná lineární forma) je jiný název pro homogenní polynom prvního stupně[1]. Polynom druhého stupně se nazývá *kvadratická funkce* a homogenní polynom druhého stupně *kvadratická forma*[2]. Cílem této kapitoly je porozumět extrémům kvadratických forem a funkcí.

6.1 Kvadratická forma

Kvadratická forma na $\mathbb{R}^n$ je homogenní polynom $f\colon \mathbb{R}^n \to \mathbb{R}$ druhého stupně. Je pohodlné ji zapsat v maticovém tvaru

$$f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} = \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij} x_i x_j \tag{6.2}$$

kde $\mathbf{A} \in \mathbb{R}^{n \times n}$. Reprezentace funkce f maticí $\mathbf{A}$ je ovšem redundantní, tj. různé matice mohou definovat stejnou funkci. Protože $x_i x_j = x_j x_i$ (násobení čísel je komutativní), máme

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij} x_i x_j = \frac{1}{2} \sum_{j=1}^{n} \sum_{i=1}^{n} (a_{ij} + a_{ji}) x_i x_j = \frac{1}{2} \mathbf{x}^T (\mathbf{A} + \mathbf{A}^T) \mathbf{x}. \tag{6.3}$$

[1]Toto platí jen pro funkce na lineárním prostoru $\mathbb{R}^n$. I když polynomy lze definovat (i když složitějším způsobem) na abstraktním (tedy definovaným axiomy) lineárním prostoru, obvykle se jim tak neříká.

[2]Názvosloví není zcela konzistentní, což je opět dáno tím, že některá jména pocházejí z lineární algebry a některá z matematické analýzy.

Vidíme, že funkce f závisí jen na součtech $a_{ij} + a_{ji}$. Je proto zvykem předpokládat $a_{ij} = a_{ji}$, neboli že matice $\mathbf{A}$ je symetrická ($\mathbf{A}^T = \mathbf{A}$). V tom případě tedy $\frac{1}{2}(\mathbf{A} + \mathbf{A}^T) = \mathbf{A}$. Symetrická matice $\mathbf{A}$ je již kvadratickou formou f určena jednoznačně.

Podívejme se na rovnost (6.3) ještě jinak. Každou čtvercovou matici lze jednoznačně napsat jako součet symetrické a antisymetrické části (viz Cvičení 2.16):

$$\mathbf{A} = \underbrace{\frac{1}{2}(\mathbf{A} + \mathbf{A}^T)}_{\text{symetrická}} + \underbrace{\frac{1}{2}(\mathbf{A} - \mathbf{A}^T)}_{\text{antisymetrická}}.$$

Ale pro každé $\mathbf{x}$ máme

$$\mathbf{x}^T(\mathbf{A} - \mathbf{A}^T)\mathbf{x} = \mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{A}^T\mathbf{x} = \mathbf{x}^T\mathbf{A}\mathbf{x} - (\mathbf{x}^T\mathbf{A}\mathbf{x})^T = 0,$$

kde jsme použili skutečnost, že transpozice skaláru je tentýž skalár. Tedy když $\mathbf{A}$ není symetrická, můžeme ji nahradit její symetrickou částí a kvadratická forma se nezmění.[3]

Příklad 6.1. Příkladem kvadratické formy dvou proměnných je funkce

$$f(x,y) = 2x^2 - 2xy + y^2 = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 2 & -2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

Všimněte si, že první matice není symetrická a druhá ano. ♦

6.1.1 Definitnost (matice) kvadratické formy

Čtvercová matice $\mathbf{A}$ se nazývá

- **positivně [negativně] semidefinitní**, když pro každé $\mathbf{x}$ platí $\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$ [$\mathbf{x}^T\mathbf{A}\mathbf{x} \leq 0$],
- **positivně [negativně] definitní**, když pro každé $\mathbf{x} \neq \mathbf{0}$ platí $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ [$\mathbf{x}^T\mathbf{A}\mathbf{x} < 0$],
- **indefinitní**, když existují $\mathbf{x}$ a $\mathbf{y}$ tak, že $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ a $\mathbf{y}^T\mathbf{A}\mathbf{y} < 0$.

Všimněte si, že matice může mít i několik těchto vlastností najednou. Tak positivně definitní matice je zároveň positivně semidefinitní. Nulová matice je zároveň positivně i negativně semidefinitní. Matice, která není positivně semidefinitní ani negativně semidefinitní, je nutně indefinitní. Dále je jasné, že $\mathbf{A}$ je positivně [semi]definitní, právě když $-\mathbf{A}$ je negativně [semi]definitní.

I když definice dává smysl pro libovolné čtvercové matice, obvykle je zvykem hovořit o těchto vlastnostech jen pro symetrické matice. Někdy se tyto vlastnosti definují ne pro matici, ale abstraktněji pro kvadratickou formu.

Z definitnosti matice ihned plyne, jestli má kvadratická forma extrém a případně jaký:

Tvrzení 6.1. Nechť funkce f je dána jako $f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x}$.

- Je-li $\mathbf{A}$ positivně [negativně] semidefinitní, pak f v bodě $\mathbf{0}$ nabývá minimum [maximum].
- Je-li $\mathbf{A}$ positivně [negativně] definitní, pak f v bodě $\mathbf{0}$ nabývá ostré minimum [maximum].
- Je-li $\mathbf{A}$ indefinitní, pak f nemá minimum ani maximum.

Důkaz. Je-li $\mathbf{A}$ positivně semidefinitní, funkce f není nikde záporná a zároveň pro $\mathbf{x} = \mathbf{0}$ je nulová, proto v bodě $\mathbf{x} = \mathbf{0}$ (i když možná i jinde) nabývá svého minima. Je-li $\mathbf{A}$ positivně definitní, je forma nulová jen v počátku a všude jinde kladná, tedy počátek je ostré minimum. Je-li $\mathbf{A}$ indefinitní a např. $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$, bod $\mathbf{x}$ nemůže být maximum, protože $(2\mathbf{x})^T\mathbf{A}(2\mathbf{x}) > \mathbf{x}^T\mathbf{A}\mathbf{x}$, a zároveň $\mathbf{x}$ nemůže být minimum, protože pro nějaké $\mathbf{y}$ je $\mathbf{y}^T\mathbf{A}\mathbf{y} < 0$. ■

[3] Ještě jinak řečeno: je-li matice $\mathbf{A}$ antisymetrická, je (6.2) nulová funkce.

6.2 Definitnost ze znamének hlavních minorů

Jak určit definitnost (symetrické) matice? Jedno možné kritérium (Věta 6.2) využívá znaménka determinantů jistých jejich podmatic. Determinanty čtvercových podmatic nějaké matice se nazývají její **minory**. My budeme potřebovat jen některé minory. Pro matici $\mathbf{A} \in \mathbb{R}^{n\times n}$ a neprázdnou množinu $I \subseteq \{1,\ldots,n\}$, symbolem $\mathbf{A}_I = [a_{ij}]_{i,j\in I} \in \mathbb{R}^{|I|\times|I|}$ označíme matici vytvořenou z prvků matice $\mathbf{A}$ v řádcích a sloupcích s indexy I. Definujme:

- **Hlavní minor** matice $\mathbf{A}$ je číslo $\det \mathbf{A}_I$ pro nějakou neprázdnou $I \subseteq \{1,\ldots,n\}$.
- **Vůdčí hlavní minor** matice $\mathbf{A}$ je její hlavní minor pro nějaké $I = \{1,\ldots,k\}$ a $1 \le k \le n$.

Příklad 6.2. Příklady matic $\mathbf{A}_I$ pro matici $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{4\times 4}$:

$$\underset{I=\{1,2,3,4\}\ \text{(tj. } \mathbf{A}_I = \mathbf{A})}{\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}} \quad \underset{I=\{1,2,3\}}{\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}} \quad \underset{I=\{1,2\}}{\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}} \quad \underset{I=\{1,3,4\}}{\begin{bmatrix} a_{11} & a_{13} & a_{14} \\ a_{31} & a_{33} & a_{34} \\ a_{41} & a_{43} & a_{44} \end{bmatrix}} \quad \underset{I=\{2,3\}}{\begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}}$$

Determinanty všech těchto matic jsou hlavní minory matice $\mathbf{A}$. Determinanty první, druhé a třetí matice jsou navíc vůdčí. ♦

Věta 6.2. Symetrická matice je

- positivně semidefinitní, právě když všechny její hlavní minory jsou nezáporné,
- positivně definitní, právě když všechny její vůdčí hlavní minory jsou kladné (toto je známo jako **Sylvestrovo kritérium**).

Důkaz (částečný). Plný důkaz věty nebudeme uvádět, dokážeme pouze její snadnou část.

Začneme pozorováním, že determinant každé symetrické positivně [semi]definitní matice je kladný [nezáporný]. To lze dokázat různými způsoby[4], např. viz Cvičení 6.21 a 6.7.

Dále snadno vidíme, že je-li matice $\mathbf{A}$ positivně [semi]definitní, pak pro každou neprázdnou množinu $I \subseteq \{1,\ldots,n\}$ je podmatice $\mathbf{A}_I$ také positivně [semi]definitní. Opravdu, jestliže např. $\mathbf{x}^T\mathbf{A}\mathbf{x} \ge 0$ platí pro všechny vektory $\mathbf{x} \in \mathbb{R}^n$, pak to platí speciálně pro vektory $\mathbf{x}$, které mají složky x_i s indexy $i \notin I$ nulové. Pro takové vektory se ale výraz $\mathbf{x}^T\mathbf{A}\mathbf{x}$ zjednoduší na $\mathbf{x}_I^T\mathbf{A}_I\mathbf{x}_I$, kde $\mathbf{x}_I = (x_i)_{i\in I} \in \mathbb{R}^{|I|}$ označuje vektor tvořený složkami vektoru $\mathbf{x}$ s indexy I. Tedy máme $\mathbf{x}_I^T\mathbf{A}_I\mathbf{x}_I \ge 0$ pro všechna $\mathbf{x}_I \in \mathbb{R}^{|I|}$, což znamená, že matice $\mathbf{A}_I$ je positivně semidefinitní.

Kombinací uvedených dvou pozorování dostaneme, že positivně [semi]definitní matice $\mathbf{A}$ má všechny sve hlavní minory kladné [nezáporné]. To je část Věty 6.2. ■

Pozor ovšem: neplatí (a věta to neříká), že symetrická matice je

[4]($\star$) Uvedeme elegantní důkaz od studenta Václava Voráčka (2018), který bohužel používá některé pojmy, které neznáte. Positivně definitní matice mají nenulový determinant (protože jsou regulární) a tvoří souvislou množinu (konvexní kužel, viz Věta 17.2). Determinant je spojitá funkce (polynom). Proto musí mít všechny positivně definitní matice stejné znaménko determinantu. Ale jednotková matice je positivně definitní a má kladný determinant, tedy to znaménko musí být kladné. Množina positivně *semi*definitních matic je uzávěr množiny positivně definitních matic, proto nemohou mít jiné znaménko determinantu než positivně definitní matice, ale mohou mít navíc determinant nulový, protože mohou být singulární.

- positivně semidefinitní, jestliže všechny její vůdčí hlavní minory jsou nezáporné (např. matice $\left[\begin{smallmatrix}0&0&0\\0&0&1\\0&1&0\end{smallmatrix}\right]$ je indefinitní a má všechny vůdčí hlavní minory nezáporné),
- negativně definitní, jestliže všechny její vůdčí hlavní minory jsou záporné (např. matice $\left[\begin{smallmatrix}-2&-1&-1\\-1&2&0\\-1&0&2\end{smallmatrix}\right]$ je indefinitní a má všechny vůdčí hlavní minory záporné),
- negativně semidefinitní, jestliže všechny její hlavní minory jsou nekladné (např. matice $\left[\begin{smallmatrix}0&-1\\-1&0\end{smallmatrix}\right]$ je indefinitní a má všechny hlavní minory nekladné).

Přesto věta umožňuje rozhodnout všech pět případů definitnosti (jak?).

Pro různé množiny I Věta 6.2 poskytuje různé podmínky nutné (ale ne postačující) pro positivní (semi)definitnost matice $\mathbf{A}$. Tak pro $|I| = 1$ (tj. jednoprvkové množiny I) máme:

Důsledek 6.3. Diagonální prvky positivně [semi]definitní matice jsou kladné [nezáporné]. Diagonální prvky negativně [semi]definitní matice jsou záporné [nekladné].

Důkaz. Jestliže I má jen jeden prvek, $I = \{i\}$ pro nějaké $i \in \{1, \ldots, n\}$, pak $\mathbf{A}_I$ je jednoduše diagonální prvek a_{ii} matice $\mathbf{A}$ a jeho determinant je tento prvek sám. Zbytek plyne z toho, že matice $\mathbf{A}$ je negativně [semi]definitní, právě když $-\mathbf{A}$ je positivně semidefinitní.

Toto lze ovšem vidět i bez Věty 6.2. Např. pro positivně semidefinitní matici platí $\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$ pro každé $\mathbf{x}$, což speciálně pro $\mathbf{x} = \mathbf{e}_i$ (i-tý vektor standarní báze) zní $\mathbf{x}^T\mathbf{A}\mathbf{x} = a_{ii} \geq 0$. ■

Pro $|I| = 2$ dostaneme také zajímavou podmínku:

Důsledek 6.4. Nechť symetrická matice $\mathbf{A} = [a_{ij}]$ je positivně semidefinitní nebo negativně semidefinitní. Pak pro každé $i, j \in \{1, \ldots, n\}$ platí $a_{ii}a_{jj} \geq a_{ij}^2$.

Důkaz. Pro $I = \{i, j\}$ (kde $i \neq j$) je $\mathbf{A}_I = \left[\begin{smallmatrix}a_{ii}&a_{ij}\\a_{ji}&a_{jj}\end{smallmatrix}\right] \in \mathbb{R}^{2\times 2}$. Je-li $\mathbf{A}$ positivně nebo negativně semidefinitní, pak $\mathbf{A}_I$ má také tuto vlastnost, tedy $\det \mathbf{A}_I = \det(-\mathbf{A}_I) = a_{ii}a_{jj} - a_{ij}^2 \geq 0$. ■

Speciálně, pro $a_{ii} = 0$ nerovnost $a_{ii}a_{jj} \geq a_{ij}^2$ implikuje $a_{ij} = a_{ji} = 0$. Tedy když symetrická positivně nebo negativně semidefinitní matice má nějaký diagonální prvek nulový, pak všechny prvky ve stejném řádku a ve stejném sloupci má také nulové.

Příklad 6.3. Okamžitě vidíme, že matice

$$\mathbf{A} = \begin{bmatrix}9&0&1\\0&0&1\\1&1&9\end{bmatrix}$$

je indefinitní. Opravdu, protože $a_{22} = 0$, pro positivně nebo negativně semidefinitní matici by muselo být $a_{21} = a_{12} = a_{23} = a_{32} = 0$, což není. Tedy matice musí být indefinitní. ♦

Věta 6.2 je užitečný teoretický výsledek, ale nevede na efektivní algoritmus na určení definitnosti matice. Počet všech vůdčích hlavních minorů matice $n \times n$ je n, avšak počet všech hlavních minorů je $2^n - 1$. Proto nelze na počítači spočítat všechny hlavní minory pro větší n. Navíc počítat každý (vůdčí) hlavní minor zvlášť je hloupé, je efektivnější dělat to rekurzivně – ale to je pak už lepší použít metody popsané dále.

6.3 Diagonalizace matice kvadratické formy

Zkoumejme, jaký vliv na kvadratickou formu má lineární tranformace souřadnic. To znamená, že ve vzorci (6.2) nahradíme vektor $\mathbf{x}$ vektorem $\mathbf{x} = \mathbf{C}\mathbf{y}$, kde $\mathbf{C} \in \mathbb{R}^{n\times n}$ je nějaká regulární matice, a zkoumáme výslednou funkci vektoru $\mathbf{y}$:

$$f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{y}^T\mathbf{C}^T\mathbf{A}\mathbf{C}\mathbf{y} = \mathbf{y}^T\mathbf{B}\mathbf{y} = g(\mathbf{y}) = f(\mathbf{C}\mathbf{y}). \tag{6.4}$$

Vidíme, že nová funkce g je kvadratická forma, reprezentovaná maticí $\mathbf{B} = \mathbf{C}^T\mathbf{A}\mathbf{C}$.

To vede na následující definici: matice $\mathbf{B}$ se nazývá **kongruentní**[5] matici $\mathbf{A}$, jestliže existuje regulární matice $\mathbf{C}$ tak, že

$$\mathbf{B} = \mathbf{C}^T\mathbf{A}\mathbf{C}. \tag{6.5}$$

Snadno ukážeme, že je-li $\mathbf{B}$ kongruentní $\mathbf{A}$, pak $\mathbf{A}$ je kongruentní $\mathbf{B}$ (viz Cvičení 6.13). Proto má smysl říkat, že matice $\mathbf{A}, \mathbf{B}$ jsou (*navzájem*) kongruentní.

Tvrzení 6.5. Kongruentní matice mají stejnou definitnost.

Důkaz. Matice $\mathbf{C}$ je regulární, tj. reprezentuje bijektivní transformaci (viz §1.1.2). Proto

$$f(\mathbf{x}) \ge 0 \;\; \forall \mathbf{x} \in \mathbb{R}^n \qquad \Longleftrightarrow \qquad f(\mathbf{C}\mathbf{y}) \ge 0 \;\; \forall \mathbf{y} \in \mathbb{R}^n,$$

tedy $\mathbf{A}$ je positivně semidefinitní, právě když $\mathbf{C}^T\mathbf{A}\mathbf{C}$ je positivně semidefinitní. Podobně pro positivní a negativní (semi)definitnost a indefinitnost. ■

Předpokládejme, že se nám nějak podařilo najít regulární matici $\mathbf{C}$ tak, že matice (6.5) je diagonální – neboli *diagonalizovat* matici $\mathbf{A}$. Definitnost diagonální matice se už určí snadno:

Tvrzení 6.6. Diagonální matice je

- positivně [negativně] semidefinitní, právě když má všechny diagonální prvky nezáporné [nekladné],
- positivně [negativně] definitní, právě když má všechny diagonální prvky kladné [záporné],
- indefinitní, právě když má aspoň jeden kladný a aspoň jeden záporný diagonální prvek.

Důkaz. Je-li matice $\mathbf{B} \in \mathbb{R}^{n\times n}$ diagonální (s diagonálními prvky $b_{11}, \dots, b_{nn}$), pak

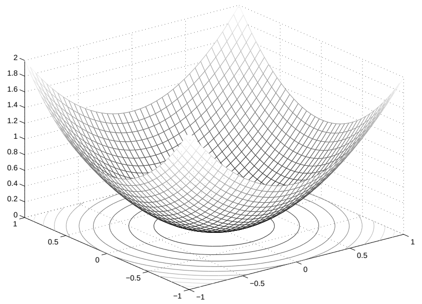
$$g(\mathbf{y}) = \mathbf{y}^T\mathbf{B}\mathbf{y} = b_{11}y_1^2 + \cdots + b_{nn}y_n^2. \tag{6.6}$$

Tento výraz je pouhá lineární kombinace čtverců proměnných $y_1, \dots, y_n$. Je jasné, že výraz je např. nezáporný pro všechna $\mathbf{y} \in \mathbb{R}^n$, právě když koeficienty $b_{11}, \dots, b_{nn}$ jsou nezáporné. ■

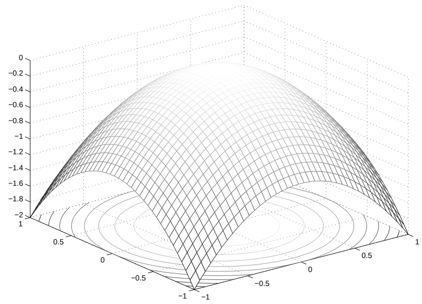
[5] Z lineární algebry znáte pojem příbuzný kongruenci: matice $\mathbf{B}$ je *podobná* matici $\mathbf{A}$, existuje-li regulární matice $\mathbf{C}$ tak, že $\mathbf{B} = \mathbf{C}^{-1}\mathbf{A}\mathbf{C}$. Tento pojem také souvisí s lineární transformací souřadnic, zde ovšem transformujeme prostor vzorů+obrazů lineární transformace $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^n$ reprezentované maticí $\mathbf{A}$, tedy $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x}$. Opravdu, transformujeme-li tento prostor lineární transformací $\mathbf{x} = \mathbf{C}\mathbf{y}$, dostaneme $\mathbf{C}\mathbf{f}(\mathbf{C}\mathbf{y}) = \mathbf{A}\mathbf{C}\mathbf{y}$, tedy $\mathbf{f}(\mathbf{C}\mathbf{y}) = \mathbf{g}(\mathbf{y}) = \mathbf{C}^{-1}\mathbf{A}\mathbf{C}\mathbf{y} = \mathbf{B}\mathbf{y}$. Na rozdíl od toho, v případě kvadratické formy jsme transformovali prostor vzorů funkce $f\colon \mathbb{R}^n \to \mathbb{R}$. Vidíme, že pojem 'diagonalizace matice' závisí na tom, co tato matice reprezentuje. V *tensorovém počtu* se toto rozlišuje: matice je sice tensor druhého řádu, ovšem reprezentuje-li lineární transformaci pak má jeden index kovariantní a jeden kontravariantní, zatímco reprezentuje-li kvadratickou formu pak má oba indexy kovariantní – podle toho poznáme, jak se má tensor transformovat.

Vlastnosti kvadratické formy jsou mnohem lépe patrny z jejího diagonálního tvaru g než z původního f. 'Kvalitativní' tvar grafu a vrstevnic funkce g je dán znaménky čísel na diagonále. Pro $n = 2$ si je snadno představíme (viz obrázek):

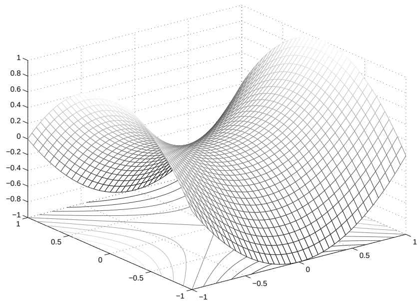
- Je-li kvadratická forma positivně definitní (oba koeficienty kladné), její graf vypadá jako 'ďolík' a její vrstevnice (kladné výšky) jsou elipsy (nebo dokonce kružnice).
- Je-li forma negativně definitní (oba koeficienty záporné), graf vypadá jako 'kopec' a vrstevnice (záporné výšky) jsou opět elipsy.
- Je-li forma indefinitní (jeden koeficient kladný a druhý záporný), graf vypadá jako 'sedlo' a vrstevnice (nenulové výšky) jsou hyperboly.



$g(y_1, y_2) = y_1^2 + y_2^2$



$g(y_1, y_2) = -y_1^2 - y_2^2$



$g(y_1, y_2) = y_1^2 - y_2^2$

Tvrzení 6.5 a 6.6 napovídají, jak určit definitnost obecné symetrické matice $\mathbf{A}$: nejprve najdeme regulární matici $\mathbf{C}$ tak, že matice $\mathbf{B} = \mathbf{C}^T\mathbf{A}\mathbf{C}$ je diagonální, a pak zjistíme definitnost matice $\mathbf{A}$ ze znamének diagonálních prvků matice $\mathbf{B}$. V následující dvou sekcích popíšeme dvě metody, jak takovou matici $\mathbf{C}$ nalézt (čímž zároveň dokážeme, že taková matice vždy existuje):

- symetrickou Gaussovu eliminaci (§6.3.1), která je rychlá na výpočet (a trvá jen konečný počet iterací),
- spektrální rozklad (§6.3.2), který je výpočetně náročnější (navíc k nalezení nekonečně přesného řešení potřebuje nekonečný počet iterací), ale najde matici $\mathbf{C}$, která je nejen regulární ale dokonce ortogonální.

6.3.1 Symetrická Gaussova eliminace

Víte, že na libovolné matici můžeme provádět *ekvivalentní řádkové úpravy* (přečtěte si znovu sekci §2.5!). Každá taková úprava odpovídá násobení nějakou *elementární maticí* zleva. Analogicky můžeme dělat ekvivalentní *sloupcové* úpravy, které odpovídají násobení elementární maticí zprava. My ale chceme těmito elementárními úpravami udržet kongruenci s původní maticí. Proto vždy, když na matici aplikujeme nějakou řádkovou ekvivalentní úpravu, hned poté na ní také aplikujeme analogickou sloupcovou ekvivalentní úpravu. Tím se z matice $\mathbf{A}$ stane matice $\mathbf{E}^T\mathbf{A}\mathbf{E}$, kde $\mathbf{E}$ označuje elementární matici reprezentující sloupcovou úpravu (a $\mathbf{E}^T$ reprezentuje řádkovou úpravu). Nová matice je kongruentní k matici $\mathbf{A}$; navíc bude symetrická, pokud byla $\mathbf{A}$ symetrická. Všimněte si, že nezáleží na tom, jestli jsme nejdříve udělali řádkovou úpravu a pak sloupcovou nebo naopak, protože díky asociativitě maticového násobení máme $\mathbf{E}^T(\mathbf{A}\mathbf{E}) = (\mathbf{E}^T\mathbf{A})\mathbf{E}$. Převedeme-li matici $\mathbf{A}$ posloupností k takových úprav na diagonální, získáme diagonální matici

$$\mathbf{B} = \underbrace{\mathbf{E}_k^T \dots \mathbf{E}_1^T}_{\mathbf{C}^T} \mathbf{A} \underbrace{\mathbf{E}_1 \dots \mathbf{E}_k}_{\mathbf{C}}. \tag{6.7}$$

Základní verze symetrické Gaussovy eliminace postupně pro $i = 1, \ldots, n-1$ opakuje následující iteraci: vynuluj (pomocí řádkových+sloupcových úprav) všechny prvky matice $\mathbf{A}$ pod diagonálním prvkem a_{ii} (kterému říkáme **pivot**) a napravo od něj, tedy prvky $a_{ij} = a_{ji}$ pro všechna $j > i$.

Příklad 6.4. Uveďme příklad běhu algoritmu. Každá iterace je označena číslem nad šipkou, pivot je vždy v rámečku. Iterace .L vždy značí řádkovou a .R analogickou sloupcovou úpravu. První matice je naše vstupní symetrická matice $\mathbf{A}$:

$$\begin{bmatrix} \boxed{1} & 3 & -1 \\ 3 & 4 & 1 \\ -1 & 1 & 2 \end{bmatrix} \xrightarrow{1\mathrm{L}} \begin{bmatrix} \boxed{1} & 3 & -1 \\ 0 & -5 & 4 \\ 0 & 4 & 1 \end{bmatrix} \xrightarrow{1\mathrm{R}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \boxed{-5} & 4 \\ 0 & 4 & 1 \end{bmatrix} \xrightarrow{2\mathrm{L}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \boxed{-5} & 4 \\ 0 & 0 & \frac{21}{5} \end{bmatrix} \xrightarrow{2\mathrm{R}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -5 & 0 \\ 0 & 0 & \frac{21}{5} \end{bmatrix}$$

V iteraci 1L vynulujeme prvky pod pivotem tak, že od druhého řádku odečteme trojnásobek prvního řádku a ke třetímu řádku přičteme první řádek. V iteraci 1R vynulujeme prvky vpravo od pivotu stejnými sloupcovými úpravami, tedy od druhého sloupce odečteme trojnásobek prvního sloupce a ke třetímu sloupci přičteme první sloupec. Výsledná (třetí) matice je opět symetrická. V iteraci 2L vynulujeme prvek pod novým pivotem tak, že k třetímu řádku přičteme $\frac{5}{4}$ druhého a k třetímu sloupci přičteme $\frac{4}{5}$ druhého. V iteraci 2R uděláme totéž se sloupci. ♦

Kdybychom chtěli spočítat i matici $\mathbf{C} = \mathbf{E}_1, \ldots, \mathbf{E}_k$, můžeme to zařídit jako při invertování matice Gaussovou-Jordanovou eliminací: sloupcové úpravy budeme místo na matice $\mathbf{A}$ aplikovat na matici $\begin{bmatrix} \mathbf{A} \\ \mathbf{I} \end{bmatrix}$.[6] Protože $\begin{bmatrix} \mathbf{A} \\ \mathbf{I} \end{bmatrix} \mathbf{C} = \begin{bmatrix} \mathbf{AC} \\ \mathbf{C} \end{bmatrix}$, po skončení algoritmu se matice $\mathbf{I}$ změní na matici $\mathbf{C}$.

Příklad 6.5. Zde je stejný příklad jako minulý, ale s přidanou jednotkovou maticí:

$$\left[\begin{array}{ccc} \boxed{1} & 3 & -1 \\ 3 & 4 & 1 \\ -1 & 1 & 2 \\ \hline 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right] \xrightarrow{1\mathrm{L}} \left[\begin{array}{ccc} \boxed{1} & 3 & -1 \\ 0 & -5 & 4 \\ 0 & 4 & 1 \\ \hline 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right] \xrightarrow{1\mathrm{R}} \left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & \boxed{-5} & 4 \\ 0 & 4 & 1 \\ \hline 1 & -3 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right] \xrightarrow{2\mathrm{L}} \left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & \boxed{-5} & 4 \\ 0 & 0 & \frac{21}{5} \\ \hline 1 & -3 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right] \xrightarrow{2\mathrm{R}} \left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & -5 & 0 \\ 0 & 0 & \frac{21}{5} \\ \hline 1 & -3 & \frac{7}{5} \\ 0 & 1 & \frac{4}{5} \\ 0 & 0 & 1 \end{array}\right]$$

Zkontrolujeme, že opravdu

$$\mathbf{C}^T\mathbf{AC} = \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ \frac{7}{5} & \frac{4}{5} & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & -1 \\ 3 & 4 & 1 \\ -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & -3 & \frac{7}{5} \\ 0 & 1 & \frac{4}{5} \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -5 & 0 \\ 0 & 0 & \frac{21}{5} \end{bmatrix} = \mathbf{B}.$$

♦

Tato základní verze algoritmu funguje, jestliže v každé iteraci je pivot nenulový, $a_{ii} \neq 0$. K tomu postačuje, aby vstupní matice $\mathbf{A}$ byla positivně nebo negativně definitní, neboť positivně [negativně] definitní matice má všechny diagonální prvky kladné [záporné] a definitnost matice se během algoritmu zachovává. Ukážeme, jak ošetřit případy, kdy je pivot nulový.

Jestliže v nějaké iteraci algoritmu je pivot a_{ii} nulový a $a_{jj} \neq 0$ pro nějaké $j > i$, pak můžeme prohodit řádky+sloupce i, j (prohazujeme vždy zároveň řádky i sloupce, aby se zachovala symetrie matice).

[6] Nebo bychom samozřejmě mohli aplikovat řádkové úpravy na matici $\begin{bmatrix} \mathbf{A} & \mathbf{I} \end{bmatrix}$.

Příklad 6.6. V první matici nemůžeme vynulovat prvky pod pivotem a_{22} přičtením vhodného násobku pivotového řádku ke třetímu a čtvrtému řádku, protože pivot je nulový:

$$\begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & \boxed{0} & 3 & -1 \\ 0 & 3 & 2 & 4 \\ 0 & -1 & 4 & 4 \end{bmatrix} \xrightarrow{\text{1L}} \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & \boxed{-1} & 4 & 4 \\ 0 & 3 & 2 & 4 \\ 0 & 0 & 3 & -1 \end{bmatrix} \xrightarrow{\text{1R}} \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & \boxed{4} & 4 & -1 \\ 0 & 4 & 2 & 3 \\ 0 & -1 & 3 & 0 \end{bmatrix}$$

Protože např. $a_{44} \neq 0$, prohodíme druhý a čtvrtý řádek (iterace 1L) a druhý a čtvrtý sloupec (iterace 1R). Nyní je pivot nenulový. ♦

Jestliže je pivot a_{ii} nulový a prvky a_{jj} pro všechna $j > i$ jsou také nulové, pak prohazování řádků+sloupců nemůže učinit pivot nenulový. To nám ale nevadí. Jestliže je totiž prvek $a_{ij} = a_{ji}$ nenulový pro libovolné $j > i$, pak matice je nutně indefinitní (viz Příklad 6.3) a tedy není třeba diagonalizaci dokončit (protože chceme jen určit definitnost matice). V opačném případě ($a_{ij} = a_{ji} = 0$ pro všechna $j > i$) je matice již diagonální.

Příklad 6.7. Jestliže v nějaké iteraci algoritmu vypadá aktuální matice jako první matice níže (pivot je opět v rámečku), tak je dle Důsledku 6.4 nutně indefinitní, protože v pivotovém řádku+sloupci je aspoň jeden prvek nenulový. V opačném případě (všechny prvky v pivotovém řádku+sloupci nulové) je již matice diagonální (druhá matice).

$$\begin{bmatrix} -2 & 0 & 0 \\ 0 & \boxed{0} & 1 \\ 0 & 1 & 0 \end{bmatrix} \qquad \begin{bmatrix} -2 & 0 & 0 \\ 0 & \boxed{0} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

♦

LDL a Choleského rozklad

Symetrická Gaussova eliminace je základem pro algoritmy na rozklady symetrických matic.

LDL rozklad rozloží symetrickou matici $\mathbf{A} \in \mathbb{R}^{n \times n}$ jako

$$\mathbf{A} = \mathbf{L}\mathbf{D}\mathbf{L}^T \tag{6.8}$$

kde $\mathbf{L} \in \mathbb{R}^{n \times n}$ je regulární dolní trojúhelníková a $\mathbf{D} \in \mathbb{R}^{n \times n}$ je diagonální. To je zhruba rozklad, který jsme popsali výše: rozklad (6.7) napíšeme jako $\mathbf{A} = \mathbf{C}^{-T}\mathbf{B}\mathbf{C}^{-1}$ a označíme $\mathbf{L} = \mathbf{C}^{-T}$ a $\mathbf{D} = \mathbf{B}$. Rozdíl je jednak v detailech algoritmu (kvůli zajištění numerické stability) a jednak v tom, že v našem algoritmu ne vždy vyjde $\mathbf{C}$ horní trojúhelníková (tedy $\mathbf{L}$ dolní trojúhelníková). Matice $\mathbf{C}$ vyjde horní trojúhelníková tehdy, když během algoritmu nepotkáme nulový pivot a nemusíme tedy prohazovat dvojice řádků+sloupců – v tom případě totiž používáme jen ekvivalentní úpravy odpovídající dolním trojúhelníkovým elementárním maticím (viz §2.5). Např. v Příkladu 6.4 jsme nulový pivot nepotkali a proto $\mathbf{C}$ vyšla horní trojúhelníková. Jak jsme podotkli výše, nenulový pivot určitě nepotkáme v případě, že $\mathbf{A}$ je positivně definitní.

Choleského rozklad rozloží symetrickou positivně definitní matici $\mathbf{A} \in \mathbb{R}^{n \times n}$ jako

$$\mathbf{A} = \mathbf{L}\mathbf{L}^T \tag{6.9}$$

kde $\mathbf{L} \in \mathbb{R}^{n \times n}$ je regulární dolní trojúhelníková. Tento rozklad se ihned dostane z LDL rozkladu $\mathbf{A} = \bar{\mathbf{L}}\mathbf{D}\bar{\mathbf{L}}^T$: stačí položit $\mathbf{L} = \bar{\mathbf{L}}\mathbf{D}^{1/2}$, kde $\mathbf{D}^{1/2}$ značí diagonální matici, která má na diagonále odmocniny z diagonálních prvků matice $\mathbf{D}$ (diagonální prvky $\mathbf{D}$ jsou kladné díky positivní definitnosti $\mathbf{A}$). Pak tedy $\mathbf{L}\mathbf{L}^T = \bar{\mathbf{L}}\mathbf{D}^{1/2}(\mathbf{D}^{1/2})^T\bar{\mathbf{L}}^T = \bar{\mathbf{L}}\mathbf{D}^{1/2}\mathbf{D}^{1/2}\bar{\mathbf{L}}^T = \bar{\mathbf{L}}\mathbf{D}\bar{\mathbf{L}}^T = \mathbf{A}$.

Choleského rozklad lze také získat jednoduchou modifikací výše popsané základní verze synetrické Gaussovy eliminace: po každé iteraci bychom vynásobili pivotový řádek a sloupec takovým číslem, abychom pivot a_{ii} učinili roven 1. Tj. museli bychom pivotový řádek a sloupec vydělit číslem $\sqrt{a_{ii}}$ (kde $a_{ii} > 0$ díky positivní definitnosti).

V Matlabu se LDL a Choleského rozklad spočítá příkazy `ldl` a `chol`. Prostudujte nápovědu k těmto příkazům a vyzkoušejte si je na náhodných symetrických maticích!

Typické použití LDL/Choleského rozkladu je pro řešení soustavy $\mathbf{A}\mathbf{x} = \mathbf{b}$ se symetrickou positivně definitní maticí $\mathbf{A}$. Pro takové soustavy je tato metoda rychlejší a/nebo numericky stabilnější než Gaussova eliminace / LU rozklad (viz §2.5) i QR rozklad. Takové soustavy se vyskytují např. v každé iteraci některých numerických iteračních algoritmů (§10). Soustavu $\mathbf{A}\mathbf{x} = \mathbf{L}\mathbf{L}^T\mathbf{x} = \mathbf{b}$ vyřešíme levně tak, že nejdřív spočteme $\mathbf{y}$ ze soustavy $\mathbf{L}\mathbf{y} = \mathbf{b}$ zpětnou substitucí a pak spočteme $\mathbf{x}$ ze soustavy $\mathbf{L}^T\mathbf{x} = \mathbf{y}$ opět zpětnou substitucí. (V případě LDL rozkladu bychom dostali soustavu $\mathbf{L}\mathbf{D}\mathbf{L}^T\mathbf{x} = \mathbf{b}$, jak byste ji levně vyřešili?)

6.3.2 Spektrální rozklad symetrické matice

Zde popíšeme diagonalizaci symetrické matice pomocí tzv. spektrálního rozkladu. Ten úzce souvisí s teorií vlastních čísel a vektorů. Tato teorie je obsáhlá a má mnoho různých použití, my z ní ale budeme potřebovat jen relativně malou část týkající se symetrických reálných matic.

Vlastní čísla a vektory

Nechť pro čtvercovou matici $\mathbf{A} \in \mathbb{R}^{n\times n}$, *nenulový* vektor $\mathbf{v} \in \mathbb{C}^n$ a skalár $\lambda \in \mathbb{C}$ platí[7]

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}. \tag{6.10}$$

Pak λ se nazývá **vlastní číslo** matice a $\mathbf{v}$ **vlastní vektor** matice příslušný vlastnímu číslu λ. Vlastní čísla a vektory mohou být obecně komplexní, i když $\mathbf{A}$ má všechny prvky reálné. Množině všech vlastních čísel matice se také říká její *spektrum*. Všimněte si, že vlastní čísla a vektory lze opravdu definovat jen pro čtvercové matice, protože jinak by rovnice (6.10) nedávala smysl (proč?).

Rovnost (6.10) je soustava n *nelineárních* rovnic s $n+1$ neznámými λ a $\mathbf{v}$. Z této soustavy lze eliminovat $\mathbf{v}$ známým obratem, čímž dostaneme jedinou rovnici pro λ. Přepíšeme (6.10) jako

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}, \tag{6.11}$$

což pro pevné λ je soustava homogenních lineárních rovnic pro $\mathbf{v}$. Protože vlastní vektory nesmí být nulové, λ splňuje (6.10) právě tehdy, když soustava (6.11) má netriviální (tj. nenulové) řešení. To nastane právě tehdy, když její matice $\mathbf{A} - \lambda\mathbf{I}$ je singulární, neboli

$$p_{\mathbf{A}}(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}) = 0. \tag{6.12}$$

Dle definice determinantu (2.5) je funkce $p_{\mathbf{A}} : \mathbb{R} \to \mathbb{R}$ polynom stupně n. Nazývá se **charakteristický polynom** matice $\mathbf{A}$. Jeho kořeny[8] jsou tedy právě vlastní čísla matice $\mathbf{A}$.

[7] Zopakujme, že $\mathbb{C}$ je množina komplexních čísel, tj. čísel ve tvaru $a + bi$ kde $a, b \in \mathbb{R}$ a $i = \sqrt{-1}$.

[8] Rozlišujte pojmy *kořen* a *řešení*: kořen polynomu $p\colon \mathbb{R} \to \mathbb{R}$ je řešení rovnice $p(x) = 0$. Sousloví 'kořen rovnice' nebo 'řešení polynomu' jsou nesmysly.

Podle tzv. *základní věty algebry* má každý polynom stupně n právě n komplexních kořenů, kde počítáme každý kořen tolikrát, kolik je jeho násobnost. Tedy

$$p_{\mathbf{A}}(\lambda) = \prod_{i=1}^{n} (\lambda_i - \lambda), \tag{6.13}$$

kde $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ je seznam kořenů, přičemž každý kořen je v seznamu tolikrát, kolik je jeho násobnost. V tomto smyslu má tedy každá matice velikosti $n \times n$ právě n vlastních čísel (reálných či komplexních), kde každé vlastní číslo počítáme tolikrát, kolik je jeho násobnost[9].

Vlastní vektory příslušné nějakému vlastnímu číslo λ tvoří množinu řešení homogenní lineární soustavy (6.11), kromě počátku (protože vlastní vektory nesmějí být nulové). Tato množina je podprostor $\operatorname{null}(\mathbf{A} - \lambda\mathbf{I})$, kterému říkáme **vlastní podprostor** příslušný vlastnímu číslu λ.

Příklad 6.8. Vlastní čísla matice $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}$ jsou kořeny kvadratického polynomu

$$p_{\mathbf{A}}(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}) = \det \begin{bmatrix} -\lambda & 1 \\ -2 & -3-\lambda \end{bmatrix} = \lambda^2 + 3\lambda + 2 = (\lambda + 1)(\lambda + 2).$$

Ten má dva reálné kořeny $\lambda_1 = -1$ a $\lambda_2 = -2$, každý s násobností 1.

Vlastní vektory příslušné vlastnímu číslu λ_1 najdeme řešením homogenní lineární soustavy

$$(\mathbf{A} - \lambda_1\mathbf{I})\mathbf{v} = (\mathbf{A} + \mathbf{I})\mathbf{v} = \begin{bmatrix} 1 & 1 \\ -2 & -2 \end{bmatrix} \mathbf{v} = \mathbf{0}.$$

Jeden takový vektor je $\mathbf{v} = (-1, 1)$. Vlastní podprostor příslušný vlastnímu číslu λ_1 je přímka $\operatorname{null}(\mathbf{A} + \lambda_1\mathbf{I}) = \operatorname{span}\{(-1, 1)\}$. ♦

Příklad 6.9. Vlastní čísla matice $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ jsou kořeny polynomu

$$p_{\mathbf{A}}(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}) = (1 - \lambda)^2.$$

Ten má jediný kořen $\lambda = 1$ s násobností 2. Příslušný vlastní podprostor je

$$\operatorname{null}(\mathbf{A} - \lambda\mathbf{I}) = \operatorname{null} \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = \operatorname{span}\{(0, 1)\}.$$

♦

Příklad 6.10. Vlastní čísla matice $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$ jsou kořeny polynomu

$$p_{\mathbf{A}}(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}) = \det \begin{bmatrix} 1-\lambda & 1 \\ -1 & 1-\lambda \end{bmatrix} = \lambda^2 - 2\lambda + 2 = (\lambda - 1 - \mathrm{i})(\lambda - 1 + \mathrm{i}).$$

Ten má dva komplexní kořeny $\lambda_1 = 1 + \mathrm{i}$ a $\lambda_2 = 1 - \mathrm{i}$, každý s násobností 1. ♦

[9] Násobnost kořene λ charakteristického polynomu se přesněji nazývá *algebraická* násobnost vlastního čísla λ. Kromě ní existuje i *geometrická* násobnost vlastního čísla λ, což je dimenze prostoru $\operatorname{null}(\mathbf{A} - \lambda\mathbf{I})$. Platí věta (její důkaz je netriviální), že geometrická násobnost není nikdy větší než algebraická.

Nechť $\lambda_1, \ldots, \lambda_n$ jsou vlastní čísla matice $\mathbf{A}$ (včetně násobností) a $\mathbf{v}_1, \ldots, \mathbf{v}_n$ nějaké k nim příslušné vlastní vektory, tedy platí

$$\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i, \qquad i = 1, \ldots, n. \tag{6.14}$$

Tuto soustavu můžeme napsat maticově jako

$$\mathbf{A}\mathbf{V} = \mathbf{V}\boldsymbol{\Lambda} \tag{6.15}$$

kde

$$\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n) = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}, \qquad \mathbf{V} = [\mathbf{v}_1 \;\cdots\; \mathbf{v}_n],$$

tedy $\boldsymbol{\Lambda} \in \mathbb{R}^{n\times n}$ je diagonální s vlastními čísly na diagonále a $\mathbf{V} \in \mathbb{R}^{n\times n}$ je čtvercová s vlastními vektory jako sloupci. (Rozmyslete, proč (6.15) je stejné jako (6.14)!)

Jak se počítají vlastní čísla a vektory? Charakteristický polynom je hlavně teoretický nástroj a přímé hledání jeho kořenů se hodí jen pro malé matice[10]. Pro větší matice se používají numerické iterační algoritmy. Zásadní rozdíl oproti např. symetrické Gaussově eliminaci je v tom, že vlastní čísla obecně nelze spočítat konečným počtem operací sčítání, odčítání, dělení a k-té (kde k je libovolné přirozené číslo) odmocniny. To plyne z toho, že kořeny polynomu většího než čtvrtého stupně obecně nelze spočítat konečným počtem těchto operací. V Matlabu se algoritmus na vlastní čísla a vektory volá funkcí `[V,D]=eig(A)`, která vrátí matici $\boldsymbol{\Lambda}$ (označenou $\mathbf{D}$) a matici $\mathbf{V}$ splňující (6.15).

Vlastní čísla a vektory symetrické matice

Pro některé matice lze vybrat vlastní vektory $\mathbf{v}_1, \ldots, \mathbf{v}_n$ (připomeňme, že jednomu vlastnímu číslu přísluší více vlastních vektorů) tak, že jsou lineárně nezávislé. To není možné vždy: např. v Příkladu 6.9 má vlastní prostor příslušný jedinému vlastnímu číslu dimenzi 1, proto nemůže obsahovat dva lineárně nezávislé vektory. Příznivá situace nastane pro symetrické matice:

Věta 6.7. Každá symetrická matice $\mathbf{A} \in \mathbb{R}^{n\times n}$ má všechna vlastní čísla $\lambda_1, \ldots, \lambda_n$ reálná a jim odpovídající vlastní vektory $\mathbf{v}_1, \ldots, \mathbf{v}_n$ lze zvolit tak, že jsou po dvojicích ortogonální.

Důkaz. Uvádíme ho nepovinně níže v této sekci. ∎

Důsledek 6.8. Každou symetrickou matici $\mathbf{A} \in \mathbb{R}^{n\times n}$ lze rozložit jako

$$\mathbf{A} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T = \lambda_1 \mathbf{v}_1 \mathbf{v}_1^T + \cdots + \lambda_n \mathbf{v}_n \mathbf{v}_n^T \tag{6.16}$$

kde matice $\boldsymbol{\Lambda} \in \mathbb{R}^{n\times n}$ je diagonální (její diagonální prvky jsou $\lambda_1, \ldots, \lambda_n$) a $\mathbf{V} \in \mathbb{R}^{n\times n}$ je ortogonální (její sloupce jsou vektory $\mathbf{v}_1, \ldots, \mathbf{v}_n$).

Důkaz. Okamžitě z rovnosti (6.15) a Věty 6.7. Protože jsou vlastní čísla reálná, je $\boldsymbol{\Lambda} \in \mathbb{R}^{n\times n}$. Zvolíme-li jim příslušné vlastní vektory po dvojicích ortogonální, po normalizaci každého vektoru (což je možné, protože vlastní vektory jsou nenulové) tvoří ortonormální množinu, sloupce ortogonální matice $\mathbf{V}$. Tedy máme $\mathbf{V}^{-1} = \mathbf{V}^T$ a (6.15) je proto ekvivalentní (6.16). ∎

[10]Naopak, hledání kořenů libovolného polynomu lze převést na hledání vlastních čísel matice, která se nazývá *doprovodná matice (companion matrix)* polynomu.

Rovnosti (6.16) se říká **spektrální rozklad** symetrické matice. Zároveň jsme v (6.16) uvedli i druhou formu rozkladu jako součet dyád (viz §2.2; ověřte druhou rovnost v (6.16) roznásobením výrazu $\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ dle pravidel pro násobení blokových matic). Zavedeme konvenci, že vlastní čísla v rozkladu (6.16) jsou *sestupně*[11] seřazena[12],

$$\lambda_1 \geq \cdots \geq \lambda_n, \tag{6.17}$$

což lze vždy zařídit vhodnou permutací sloupců matice $\mathbf{V}$ a diagonálních prvků matice $\mathbf{\Lambda}$.

Hodnost matice $\mathbf{A}$ je rovna hodnosti matice $\mathbf{\Lambda}$ (což plyne např. z Tvrzení 3.10, použité dvakrát na výraz $\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$). Ale hodnost diagonální matice $\mathbf{\Lambda}$ je počet jejích nenulových prvků, tedy nenulových vlastních čísel. Je-li $\operatorname{rank} \mathbf{A} = r < n$, pak $n - r$ vlastních čísel je tedy nulových a v rozkladu (6.16) můžeme vynechat jim odpovídající sloupce+řádky matice $\mathbf{\Lambda}$ a sloupce matice $\mathbf{V}$. Tedy $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ pro nějaké $\mathbf{V} \in \mathbb{R}^{n\times r}$, $\mathbf{V}^T\mathbf{V} = \mathbf{I}$ a $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_r)$. Tomuto rozkladu budeme říkat **rank-minimální**[13] spektrální rozklad symetrické matice. Odpovídá to vynechání $n - r$ nulových sčítanců v sumě dyád v (6.16).

Příklad 6.11. Spektrální rozklad (6.16) symetrické matice $\mathbf{A} \in \mathbb{R}^{3\times 3}$ hodnosti 2, nejdříve plný, pak rank-minimální, a nakonec ve formě součtu dyád. Vlastní čísla jsou sestupně seřazena:

$$\begin{aligned}
\begin{bmatrix} 2 & 3 & -2 \\ 3 & 2 & 2 \\ -2 & 2 & -8 \end{bmatrix} &= \begin{bmatrix} 1/\sqrt{2} & -2/3 & 1/\sqrt{18} \\ 1/\sqrt{2} & 2/3 & -1/\sqrt{18} \\ 0 & 1/3 & 4/\sqrt{18} \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -9 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -2/3 & 2/3 & 1/3 \\ 1/\sqrt{18} & -1/\sqrt{18} & 4/\sqrt{18} \end{bmatrix} \\
&= \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{18} \\ 1/\sqrt{2} & -1/\sqrt{18} \\ 0 & 4/\sqrt{18} \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & -9 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 1/\sqrt{18} & -1/\sqrt{18} & 4/\sqrt{18} \end{bmatrix} \\
&= 5 \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \end{bmatrix} - 9 \begin{bmatrix} 1/\sqrt{18} \\ -1/\sqrt{18} \\ 4/\sqrt{18} \end{bmatrix} \begin{bmatrix} 1/\sqrt{18} & -1/\sqrt{18} & 4/\sqrt{18} \end{bmatrix} \quad \blacklozenge
\end{aligned}$$

Spektrální rozklad symetrické matice je speciálním případem její diagonalizace pomocí maticové kongruence (viz §6.3), protože rovnost (6.16) lze také napsat jako $\mathbf{\Lambda} = \mathbf{V}^T\mathbf{A}\mathbf{V}$.[14] Napišme pro tento případ analogii transformace (6.4):

$$f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{y}^T\mathbf{V}^T\mathbf{A}\mathbf{V}\mathbf{y} = \mathbf{x}^T\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T\mathbf{x} = \mathbf{y}^T\mathbf{\Lambda}\mathbf{y} = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2 = g(\mathbf{y}) \tag{6.18}$$

Substituce $\mathbf{x} = \mathbf{V}\mathbf{y}$ (tj. $\mathbf{y} = \mathbf{V}^T\mathbf{x}$) diagonalizovala matici $\mathbf{A}$ kvadratické formy. Protože matice $\mathbf{V}$ je ortogonální[15] (tedy nejen regulární jako v §6.3), transformace $\mathbf{x} = \mathbf{V}\mathbf{y}$ je isometrie. Tedy funkce f a g se liší jen otočením+zrcadlením prostoru vzorů. Totéž lze říct v jazyce bází: zatímco ve standardní bázi má kvadratická forma matici $\mathbf{A}$, v ortonormální bázi tvořené vlastními vektory (sloupci matice $\mathbf{V}$) má forma diagonální matici $\mathbf{\Lambda}$. Např. vrstevnice (výšky např. 1) positivně definitní kvadratické formy (tj. množina řešení rovnice $\mathbf{x}^T\mathbf{A}\mathbf{x} = 1$) je elipsa se středem v počátku a směry jejích hlavních os jsou vlastní vektory:

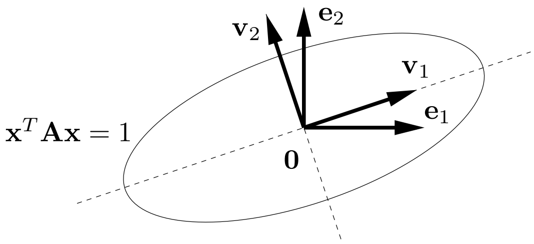
[11] Bohužel, matlabská funkce `eig` je řadí *vzestupně*.

[12] Vlastní čísla obecné (tj. ne nutně symetrické) matice nejdou takto přirozeně uspořádat, protože to jsou obecně komplexní čísla.

[13] Tento název není široce používaný, podobně jako 'rank-minimální SVD' zmíněné později.

[14] Protože $\mathbf{V}^T = \mathbf{V}^{-1}$, matice $\mathbf{A}$ a $\mathbf{\Lambda}$ jsou zároveň kongruentní i podobné (viz footnote 5 v této kapitole).

[15] Proto se spektrálnímu rozkladu symetrické matice také někdy říká její *ortogonální diagonalizace*.



Tvrzení 6.5 a 6.6 mají jednoduchý důsledek:

Důsledek 6.9. Symetrická matice je

- positivně [negativně] semidefinitní, právě když má všechna vlastní čísla nezáporná [nekladná]
- positivně [negativně] definitní, právě když má všechna vlastní čísla kladná [záporná]
- indefinitní, právě když má alespoň jedno kladné a alespoň jedno záporné vlastní číslo.

($\star$) Důkaz Věty 6.7

Tento důkaz je nepovinný, ale není těžký a je poučné se jím prokousat. Rozdělíme ho na dvě části: Věta 6.10 ukáže, že symetrická matice má všechna vlastní čísla reálná, a Věty 6.11 a 6.12 ukáží, že z vlastních vektorů symetrické matice lze vybrat ortonormální množinu n vektorů.

K důkazu první věty budeme potřebovat základní pojmy z komplexní lineární algebry[16]:

- *Komplexně sdružené* číslo ke komplexnímu číslu $x = a + bi \in \mathbb{C}$ je číslo $\overline{x} = a - bi$. Číslo x je reálné, právě když $\overline{x} = x$. Pro matici $\mathbf{A}$ definujeme $\overline{\mathbf{A}}$ po prvcích.
- *Absolutní hodnota* čísla $x \in \mathbb{C}$ je reálné číslo $|x| = (\overline{x}x)^{1/2} = (a^2 + b^2)^{1/2}$.
- *Adjungovaná* matice k matici $\mathbf{A} \in \mathbb{C}^{m\times n}$ je matice $\mathbf{A}^* = \overline{\mathbf{A}}^T \in \mathbb{C}^{n\times m}$. Matice $\mathbf{A}$ je *samoadjungovaná* (neboli *hermitovská*), je-li $\mathbf{A}^* = \mathbf{A}$. Pro reálné matice je $\mathbf{A}^* = \mathbf{A}^T$, tedy samoadjungovanost splývá se symetrií.
- *Skalární součin* vektorů $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$ je $\mathbf{x}^*\mathbf{y} = \overline{x}_1 y_1 + \cdots + \overline{x}_n y_n$ (všimněte si, že $\mathbf{x}^*\mathbf{y} = \overline{\mathbf{y}^*\mathbf{x}}$, tedy skalární součin v komplexním oboru není komutativní).
- *Norma* vektoru $\mathbf{x} \in \mathbb{C}^n$ je $\|\mathbf{x}\| = (\mathbf{x}^*\mathbf{x})^{1/2} = (|x_1|^2 + \cdots + |x_n|^2)^{1/2}$.

Věta 6.10. Každá hermitovská matice má všechna vlastní čísla reálná.

Důkaz. Nechť $\mathbf{A}^* = \mathbf{A}$ a $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$. Pak

$$\lambda\|\mathbf{v}\|^2 = \lambda\mathbf{v}^*\mathbf{v} = \mathbf{v}^*\mathbf{A}\mathbf{v} = \mathbf{v}^*\mathbf{A}^*\mathbf{v} = (\mathbf{A}\mathbf{v})^*\mathbf{v} = (\lambda\mathbf{v})^*\mathbf{v} = \overline{\lambda}\mathbf{v}^*\mathbf{v} = \overline{\lambda}\|\mathbf{v}\|^2.$$

Jelikož $\mathbf{v} \neq \mathbf{0}$, je tedy $\lambda = \overline{\lambda}$, tedy $\lambda \in \mathbb{R}$. ∎

Existenci ortonormální množiny n vlastních vektorů nejprve dokážeme ve Větě 6.11 pro speciální případ, kdy matice má n různých vlastních čísel: v tomto případě je důkaz velmi krátký. Obecný případ (kdy některá vlastní čísla mohou být násobná), dokazuje Věta 6.12.

[16] Je mrzuté, že pro důkaz, že *reálná* symetrická matice má *reálná* vlastní čísla, potřebujeme komplexní čísla. Existuje i důkaz používající jen reálná čísla, ale je složitější a vyžaduje pokročilé nástroje matematické analýzy.

Věta 6.11. Nechť $\lambda_i \neq \lambda_j$ jsou dvě různá vlastní čísla symetrické reálné matice $\mathbf{A}$ a $\mathbf{v}_i, \mathbf{v}_j$ jsou jim příslušné vlastní vektory. Pak $\mathbf{v}_i \perp \mathbf{v}_j$.

Důkaz. Máme
$$\lambda_j \mathbf{v}_i^T \mathbf{v}_j = \mathbf{v}_i^T(\mathbf{A}\mathbf{v}_j) = (\mathbf{A}\mathbf{v}_i)^T \mathbf{v}_j = \lambda_i \mathbf{v}_i^T \mathbf{v}_j.$$
Z toho $(\lambda_i - \lambda_j)\mathbf{v}_i^T \mathbf{v}_j = 0$, z toho $\mathbf{v}_i^T \mathbf{v}_j = 0$. ∎

K důkazu další věty opět budeme potřebovat nové pojmy. I když jsme vlastní čísla a vektory definovali pro matici $\mathbf{A} \in \mathbb{R}^{n\times n}$, lze je obecněji definovat pro lineární transformaci $\mathbf{f}\colon X \to X$ na libovolném (netriviálním) lineárním podprostoru $X \subseteq \mathbb{R}^n$ podmínkou

$$\mathbf{f}(\mathbf{v}) = \lambda \mathbf{v}, \tag{6.19}$$

(kde tedy $\mathbf{v}$ musí být v X). Protože zobrazení $\mathbf{f}$ lze reprezentovat maticí vzhledem k nějaké bázi podprostoru X, dle základní věty algebry má $\mathbf{f}$ aspoň jedno vlastní číslo a vektor.

Podprostor $X \subseteq \mathbb{R}^n$ nazveme *invariantní podprostor* lineární transformace $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^n$ jestliže pro každé $\mathbf{x} \in X$ je $\mathbf{f}(\mathbf{x}) \in X$, tedy $\mathbf{f}$ zobrazuje množinu X do sebe. V tom případě je restrikce (viz §1.1.2) $\mathbf{f}|_X$ zobrazení $\mathbf{f}$ na množinu X sama o sobě lineární transformace na X, tedy máme $\mathbf{f}|_X\colon X \to X$ místo pouhého $\mathbf{f}|_X\colon X \to \mathbb{R}^n$. Jestliže je X netriviální, má tedy zobrazeni $\mathbf{f}|_X$ aspoň jedno vlastní číslo a vektor.

Věta 6.12. Z vlastních vektorů každé symetrické matice $\mathbf{A} \in \mathbb{R}^{n\times n}$ lze vybrat ortonormální množinu n vektorů.

Důkaz. Použijeme indukci: ukážeme, že když pro nějaké $k < n$ již máme vlastní vektory $\mathbf{v}_1, \ldots, \mathbf{v}_k$, pak existuje vlastní vektor, který je ortogonální na každý z nich.

Dokážeme nejprve, že jestliže $\mathbf{A}$ je symetrická, podprostor $X = \operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}^\perp$ je invariantní vůči lineární transformaci $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x}$. Nechť $\mathbf{x} \in X$, tedy $\mathbf{x}^T\mathbf{v}_1 = \cdots = \mathbf{x}^T\mathbf{v}_k = 0$. Máme ukázat, že $\mathbf{A}\mathbf{x} \in X$. To platí, neboť $\mathbf{x}^T\mathbf{v}_i = 0 \;\Rightarrow\; (\mathbf{A}\mathbf{x})^T\mathbf{v}_i = \mathbf{x}^T\mathbf{A}^T\mathbf{v}_i = \mathbf{x}^T\mathbf{A}\mathbf{v}_i = \lambda_i\mathbf{x}^T\mathbf{v}_i = 0$.

Protože X je invariantní vůči $\mathbf{f}$, restrikce $\mathbf{f}|_X$ zobrazení $\mathbf{f}$ na množinu X je lineární transformace na X. Protože $k < n$, je $X \neq \{\mathbf{0}\}$. Tedy $\mathbf{f}|_X$ má aspoň jeden vlastní vektor. Protože tento vektor patří do X, je ortogonální na $\mathbf{v}_1, \ldots, \mathbf{v}_k$. ∎

I když jsme Věty 6.11 a 6.12 dokázali jen pro reálné symetrické matice, platí obecněji pro komplexní hermitovské matice (definujeme-li ortogonalitu komplexních vektorů podmínkou $\mathbf{x}^*\mathbf{y} = 0$). Zkuste tyto věty a jejich důkazy takto zobecnit!

6.4 Kvadratická funkce

Kvadratická funkce je polynom (ne nutně homogenní) druhého stupně. Lze jej psát v maticovém tvaru[17]

$$f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} + c, \tag{6.20}$$

[17] Jak už jsme zmínili v §2.4 pro afinní zobrazení, kvadratickou funkci nemusíme vždy dostat zadanou přímo ve tvaru (6.20). Příkladem jsou funkce $f(\mathbf{x}, \mathbf{y}) = \mathbf{y}^T(\mathbf{A}\mathbf{x} - \mathbf{b})$ či $f(\mathbf{X}) = \operatorname{tr}((\mathbf{A}\mathbf{X} + \mathbf{B})^T(\mathbf{A}\mathbf{X} + \mathbf{B}))$. Převést takové funkce do tvaru (6.20) vyžaduje jisté dovednosti nad rámec skript.

kde $\mathbf{A}^T = \mathbf{A} \in \mathbb{R}^{n\times n}$, $\mathbf{b} \in \mathbb{R}^n$ a $c \in \mathbb{R}$. Oproti kvadratické formě tedy přibyl lineární a konstantní člen. Všimněte si, že pro $n = 1$ je (6.20) známá kvadratická funkce jedné proměnné $f(x) = ax^2 + bx + c$.

Jak nalézt extrémy kvadratické funkce? Extrémy lze hledat mechanicky pomocí derivací, to však ukážeme až v pozdější kapitole. Jiný způsob je převést kvadratickou funkci na kvadratickou formu posunutím počátku. Tento způsob popíšeme nyní.

6.4.1 Doplnění na čtverec

Někdy lze najít $\mathbf{x}_0 \in \mathbb{R}^n$ a $y_0 \in \mathbb{R}$ takové, že pro každé $\mathbf{x} \in \mathbb{R}^n$ platí

$$\mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} + c = (\mathbf{x} - \mathbf{x}_0)^T\mathbf{A}(\mathbf{x} - \mathbf{x}_0) + y_0. \tag{6.21}$$

Výraz na pravé straně je kvadratická forma s počátkem posunutým do bodu $\mathbf{x}_0$ (neboli homogenní polynom druhého stupně v proměnných $\mathbf{x} - \mathbf{x}_0$), plus konstanta. Této úpravě se říká **doplnění na čtverec**. Viděli jste ji (ale asi dávno zapomněli) pro případ $n = 1$, neboť tak se na základní škole odvozuje vzorec pro řešení kvadratické rovnice jedné proměnné. Zkusme spočíst $\mathbf{x}_0, y_0$ z daných $\mathbf{A}, \mathbf{b}, c$. Roznásobením pravé strany dostaneme

$$\begin{aligned}\mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} + c = (\mathbf{x} - \mathbf{x}_0)^T\mathbf{A}(\mathbf{x} - \mathbf{x}_0) + y_0 &= \mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{A}\mathbf{x}_0 - \mathbf{x}_0^T\mathbf{A}\mathbf{x} + \mathbf{x}_0^T\mathbf{A}\mathbf{x}_0 + y_0 \\ &= \mathbf{x}^T\mathbf{A}\mathbf{x} \underbrace{- 2\mathbf{x}_0^T\mathbf{A}}_{\mathbf{b}^T}\mathbf{x} + \underbrace{\mathbf{x}_0^T\mathbf{A}\mathbf{x}_0 + y_0}_{c}\end{aligned}$$

(kde jsme použili identitu $\mathbf{x}^T\mathbf{A}\mathbf{x}_0 = (\mathbf{x}^T\mathbf{A}\mathbf{x}_0)^T = \mathbf{x}_0^T\mathbf{A}^T\mathbf{x} = \mathbf{x}_0^T\mathbf{A}\mathbf{x}$, za předpokladu $\mathbf{A}^T = \mathbf{A}$)). Aby toto platilo pro každé $\mathbf{x} \in \mathbb{R}^n$, musí být koeficienty u stejných monomů rovny (proč?):

$$\mathbf{b} = -2\mathbf{A}\mathbf{x}_0, \tag{6.22a}$$
$$c = \mathbf{x}_0^T\mathbf{A}\mathbf{x}_0 + y_0. \tag{6.22b}$$

Pokud soustava (6.22a) má řešení, spočítáme z ní $\mathbf{x}_0$ a pak z rovnice (6.22b) snadno získáme y_0. Pokud soustava (6.22a) nemá řešení, doplnění na čtverec není možné.

Pokud je doplnění na čtverec možné, vyšetření extrémů kvadratické funkce se neliší od vyšetření extrémů kvadratické formy, protože rozdíl je jen v posunutí $\mathbf{x}_0$. Pokud doplnění na čtverec možné není, kvadratická funkce extrém nemá (toto tvrzení zde uvádíme bez důkazu, dokázalo by se snadno pomocí derivací). Extrémy kvadratických funkcí lze také (a pohodlněji) hledat pomocí derivací, ale to si ukážeme až později.

Příklad 6.12. Předveďme nejprve zmíněný případ $n = 1$, tedy $f(x) = ax^2 + bx + c$. Rovnice (6.22a) zní $b = -2ax_0$, z toho $x_0 = -b/(2a)$. Z (6.22b) tedy $y_0 = c - ax_0^2 = c - b^2/(4a)$. Ukázali jsme, že pro každé $x \in \mathbb{R}$ platí

$$ax^2 + bx + c = a(x - x_0)^2 + y_0 = a\Big(x + \frac{b}{2a}\Big)^2 + c - \frac{b^2}{4a}. \tag{6.23}$$

Kdybychom nyní chtěli najít kořeny tohoto polynomu (což už není součástí doplnění funkce f na čtverec!), je to snadné díky tomu, že v pravém výrazu v (6.23) se x vyskytuje na jediném místě. Položením výrazu rovným nule a úpravami získáme známý vzorec

$$x = -\frac{b}{2a} \pm \sqrt{\frac{b^2}{4a^2} - \frac{c}{a}} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

♦

Příklad 6.13. Máme kvadratickou funkci

$$f(x,y) = 2x^2 - 2xy + y^2 - 2y + 3 = \begin{bmatrix} x \\ y \end{bmatrix}^T \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 0 \\ -2 \end{bmatrix}^T \begin{bmatrix} x \\ y \end{bmatrix} + 3.$$

Její doplnění na čtverec je

$$f(x,y) = 2(x-1)^2 - 2(x-1)(y-2) + (y-2)^2 + 1 = \begin{bmatrix} x-1 \\ y-2 \end{bmatrix}^T \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x-1 \\ y-2 \end{bmatrix} + 1,$$

tedy máme $\mathbf{x}_0 = (1,2)$, $y_0 = 1$. Jelikož matice $\mathbf{A}$ je positivně definitní (ověřte!), má kvadratická funkce minimum v bodě $\mathbf{x}_0$. ♦

Příklad 6.14. Kvadratická funkce

$$f(x,y) = x^2 + y = \begin{bmatrix} x \\ y \end{bmatrix}^T \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}^T \begin{bmatrix} x \\ y \end{bmatrix}$$

doplnit na čtverec nejde. Funkce tedy nemá na $\mathbb{R}^2$ extrém – což je ale jasné, protože při konstantním x můžeme změnou y dosáhnout libovolně malé i libovolně velké hodnoty funkce. Načrtněte si graf této funkce! ♦

Příklad 6.15. Vraťme se k lineární úloze nejmenších čtverců (5.2). Její účelová funkce je kvadratická funkce ve speciálním tvaru: je to součet druhých mocnin ('čtverců') afinních funkcí $\mathbf{a}_i^T\mathbf{x} - b_i$ (z toho ihned plyne, že je zdola omezená nulou). Upravme ji (roznásobením závorek) do tvaru (6.20):

$$\begin{aligned} f(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{b}\|^2 &= (\mathbf{Ax} - \mathbf{b})^T (\mathbf{Ax} - \mathbf{b}) \\ &= (\mathbf{x}^T\mathbf{A}^T - \mathbf{b}^T)(\mathbf{Ax} - \mathbf{b}) \\ &= \mathbf{x}^T\mathbf{A}^T\mathbf{Ax} - \mathbf{x}^T\mathbf{A}^T\mathbf{b} - \mathbf{b}^T\mathbf{Ax} + \mathbf{b}^T\mathbf{b} \\ &= \mathbf{x}^T\mathbf{A}^T\mathbf{Ax} - 2\mathbf{b}^T\mathbf{Ax} + \mathbf{b}^T\mathbf{b}, \end{aligned} \tag{6.24}$$

kde jsme použili rovnost $\mathbf{b}^T\mathbf{Ax} = (\mathbf{b}^T\mathbf{Ax})^T = \mathbf{x}^T\mathbf{A}^T\mathbf{b}$ (neboť skalár je roven své transpozici). Extrém této funkce můžeme najít doplněním na čtverec. Soustava (6.22a) bude mít tvar $\mathbf{A}^T\mathbf{A}\mathbf{x}_0 = \mathbf{A}^T\mathbf{b}$ ($\mathbf{A}, \mathbf{b}$ zde samozřejmě označuje něco jiného než v (6.22a)), tedy dostali jsme normální rovnici (5.4). Zároveň je jasné, že matice $\mathbf{A}^T\mathbf{A}$ je positivně semidefinitní, neboť pro každé $\mathbf{x} \in \mathbb{R}^n$ máme

$$\mathbf{x}^T\mathbf{A}^T\mathbf{Ax} = (\mathbf{Ax})^T\mathbf{Ax} = \|\mathbf{Ax}\|^2 \geq 0. \tag{6.25}$$

Tedy v bodě $\mathbf{x}_0$ bude minimum.

Ukažme tuto úpravu pro funkci (5.5) z Příkladu 5.1:

$$\begin{aligned} f(x,y) &= (x+2y-6)^2 + (-x+y-3)^2 + (x+y-4)^2 \\ &= 3x^2 + 4xy + 6y^2 - 14x - 38y + 61 \\ &= 3(x - \tfrac{2}{7})^2 + 4(x - \tfrac{2}{7})(y - \tfrac{43}{14}) + 6(y - \tfrac{43}{14})^2 + \tfrac{9}{14}. \end{aligned}$$

V prvním kroku jsme umocnili závorky a zjednodušili, v druhém kroku doplnili na čtverec výše uvedeným postupem (podrobnosti obou kroků vynecháváme). ♦

6.4.2 Kvadrika

Vrstevnice kvadratické funkce se nazývá **kvadrika**. Tedy kvadrika je množina[18]

$$\{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} + c = 0 \,\} \tag{6.26}$$

všech řešení kvadratické rovnice, neboli množina všech kořenů kvadratické funkce.

Když $\mathbf{b} = \mathbf{0}$, množina (6.26) je vrstevnice kvadratické formy a kvadrika má tedy střed v počátku. Když $\mathbf{A}$ je diagonální, kvadrika má osy rovnoběžné se souřadnicovými osami. Když $\mathbf{A} = \mathbf{0}$, množina (6.26) je pouhá nadrovina. Když $\mathbf{A}$ nemá plnou hodnost, kvadrika je degenerovaná. Množina (6.26) může být i prázdná.

Jestliže kvadratická funkce dovoluje doplnění na čtverec, můžeme množinu (6.26) psát jako

$$\{\, \mathbf{x} \in \mathbb{R}^n \mid (\mathbf{x} - \mathbf{x}_0)^T\mathbf{A}(\mathbf{x} - \mathbf{x}_0) + y_0 = 0 \,\}, \tag{6.27}$$

což je vrstevnice kvadratické formy posunuté o $\mathbf{x}_0$. V tom případě je *typ kvadriky* určen jednoduše znaménky vlastních čísel matice $\mathbf{A}$. Speciálně, když všechna vlastní čísla jsou kladná (tedy $\mathbf{A}$ je positivně definitní), jde o povrch **elipsoidu**[19]. Když některá vlastní čísla jsou nulová (tedy $\mathbf{A}$ nemá plnou hodnost), kvadrika je degenerovaná. Toto ale nevyčerpává všechny typy degenerace: další typy degenerace nastanou, když $\mathbf{A}$ nemá plnou hodnost a funkce nedovoluje doplnění na čtverec.

Předpokládejme, že y_0 je takové, že množina (6.27) obsahuje nekonečný počet bodů. Pro $n = 2$ se pak kvadrika nazývá **kuželosečka** (angl. *conic*). Je-li matice $\mathbf{A}$ positivně či negativně definitní maticí, kuželosečka je *elipsa* (speciálně, pokud $\lambda_1 = \lambda_2$ pak je to kružnice), je-li $\mathbf{A}$ indefinitní, je to *hyperbola*.

6.5 Cvičení

6.1. Pro každou z těchto funkcí určete, zda je to polynom. Pokud ano, určete počet proměnných a stupeň polynomu a rozhodněte, jestli je polynom homogenní.

a) $f\colon \mathbb{R}^2 \to \mathbb{R},\ f(x, y) = (x^2 + y^2)(x - y) + xy - x - y$
b) $f\colon \mathbb{R}^n \to \mathbb{R},\ f(\mathbf{x}) = \mathbf{a}^T\mathbf{x}$ kde $\mathbf{a}$ je dáno
c) $f\colon \mathbb{R}^n \to \mathbb{R},\ f(\mathbf{x}) = \|\mathbf{x}\|$
d) $f\colon \mathbb{R}^n \to \mathbb{R},\ f(\mathbf{x}) = \|\mathbf{A}\mathbf{x} + \mathbf{b}\|^2$ kde $\mathbf{A}, \mathbf{b}$ jsou dány
e) $f\colon \mathbb{R}^{2n} \to \mathbb{R},\ f(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T\mathbf{y}$
f) $f\colon \mathbb{R}^{n\times n} \to \mathbb{R},\ f(\mathbf{X}) = \mathbf{a}^T\mathbf{X}\mathbf{b}$ kde $\mathbf{a}, \mathbf{b}$ jsou dány
g) $f\colon \mathbb{R}^{n\times n} \to \mathbb{R},\ f(\mathbf{X}) = \det \mathbf{X}$

6.2. Spočítejte vlastní čísla a vlastní vektory matic $\begin{bmatrix} 1 & 2 \\ -1 & -3 \end{bmatrix}$, $\begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix}$.

6.3. Napište rovnici, jejímiž kořeny jsou vlastní čísla matice $\begin{bmatrix} 2 & 0 & 3 \\ 0 & -2 & -1 \\ 3 & -1 & 2 \end{bmatrix}$.

[18] Bez ztráty obecnosti uvažujeme vrstevnici výšky nula (tj. množina kořenů) kvadratické funkce (6.20), neboť c můžeme zvolit libovolně.

[19] Někteří autoři myslí elipsoidem množinu i s vnitřkem, někteří jen její hranici. Rozdíl je stejný jako mezi sférou a koulí.

6.4. Jaká jsou vlastní čísla a vlastní vektory (a) nulové, (b) jednotkové, (c) diagonální matice? Jaká jsou vlastní čísla trojúhelníkové matice?

6.5. Známe vlastní čísla a vektory matice $\mathbf{A}$. Jaká jsou vlastní čísla a vektory matice $\mathbf{A}+\alpha\mathbf{I}$?

6.6. Nechť $\mathbf{A} \in \mathbb{R}^{m\times n}$ a $\mathbf{B} \in \mathbb{R}^{n\times m}$. Dokažte, že nenulová vlastní čísla matic $\mathbf{AB}$ a $\mathbf{BA}$ jsou stejná. Jaký je vztah vlastních vektorů odpovídajících stejným nenulovým vlastním číslům?

6.7. Každá čtvercová matice splňuje

$$\det \mathbf{A} = \lambda_1 \cdots \lambda_n, \tag{6.28a}$$

$$\operatorname{tr} \mathbf{A} = \lambda_1 + \cdots + \lambda_n. \tag{6.28b}$$

Dokažte tyto rovnosti pro symetrické matice (pro nesymetrické je dokazovat nemusíte).

6.8. Určete definitnost těchto symetrických matic. Zvolte si vhodný způsob, případně více způsobů.

$$\underset{\text{(a)}}{\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}} \quad \underset{\text{(b)}}{\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}} \quad \underset{\text{(c)}}{\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}} \quad \underset{\text{(d)}}{\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}} \quad \underset{\text{(e)}}{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}} \quad \underset{\text{(f)}}{\begin{bmatrix} 3 & -2 & 1 \\ -2 & 3 & 0 \\ 1 & 0 & 3 \end{bmatrix}} \quad \underset{\text{(g)}}{\begin{bmatrix} -2 & 0 & 1 \\ 0 & 3 & -1 \\ 1 & -1 & -2 \end{bmatrix}}$$

6.9. Máme matici $\mathbf{A} = \begin{bmatrix} 1 & -3 \\ 2 & -4 \end{bmatrix}$. Která z následujících tvrzení jsou pravdivá?

a) Výraz $\mathbf{x}^T\mathbf{A}\mathbf{x}$ je nezáporný pro každé $\mathbf{x} \in \mathbb{R}^2$.
b) Výraz $\mathbf{x}^T\mathbf{A}\mathbf{x}$ je nekladný pro každé $\mathbf{x} \in \mathbb{R}^2$.
c) Funkce $f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x}$ má v bodě $\mathbf{x} = \mathbf{0}$ extrém.

6.10. Máme kvadratickou formu dvou proměnných $f(x, y) = 3x^2 + 2xy + 3y^2$.

a) Napište ji ve tvaru $f(x, y) = \begin{bmatrix} x & y \end{bmatrix} \mathbf{A} \begin{bmatrix} x \\ y \end{bmatrix}$ se symetrickou $\mathbf{A}$.
b) Najděte $a, b \in \mathbb{R}$ a ortogonální $\mathbf{U}$ tak, že $f(x, y) = au^2 + bv^2$, kde $\begin{bmatrix} u \\ v \end{bmatrix} = \mathbf{U} \begin{bmatrix} x \\ y \end{bmatrix}$.
c) Nakreslete množinu bodů (u, v) splňujících $au^2 + bv^2 = 1$.
d) Transformujte tuto množinu do souřadnic (x, y) a nakreslete.

6.11. Je množina $\{ (x, y) \in \mathbb{R}^2 \mid x^2 - 3xy + y^2 = 1 \}$ elipsa nebo hyperbola? Odůvodněte.

6.12. ($\star$) Napište v Matlabu funkci `ellipse(A)`, která vykreslí elipsu s rovnicí $\mathbf{x}^T\mathbf{A}\mathbf{x} = 1$ pro positivně definitní $\mathbf{A}$. Zamyslete se, jak byste postupovali při návrhu funkce `conic(Q)`, která vykreslí kuželosečku $\mathbf{x}^T\mathbf{A}\mathbf{x} = 1$ pro $\mathbf{A}$ libovolné definitnosti (nezapomeňte, že obecná kuželosečka může být neomezená, tedy je nutno ji oříznout do daného obdélníku).

6.13. Kongruence je příkladem binární relace na množině matic $\mathbb{R}^{n\times n}$, označme ji symbolem $\sim$. Dokažte, že tato binární relace je ekvivalence, tj. je rexlexívní (tj. $\mathbf{A} \sim \mathbf{A}$), symetrická ($\mathbf{A} \sim \mathbf{B}$ implikuje $\mathbf{B} \sim \mathbf{A}$) a transitivní ($\mathbf{A} \sim \mathbf{B}$ a $\mathbf{B} \sim \mathbf{C}$ implikuje $\mathbf{A} \sim \mathbf{C}$).

6.14. Následující kvadratické funkce napište ve tvaru (6.20) se symetrickou $\mathbf{A}$, doplňte na čtverec, najděte extrémy a určete typ každého extrému.

a) $f(x, y) = x^2 + 4xy - 2y^2 + 3x - 6y + 5$
b) $f(x, y) = x^2 + 2y^2 - 2xy + 2x - y$

6.15. Máme neorientovaný graf (V, E) s množinou vrcholů $V = \{1, \dots, n\}$ a množinou hran[20] $E \subseteq \binom{V}{2}$. Každému vrcholu $i \in V$ je přiřazeno číslo $x_i \in \mathbb{R}$, tato čísla tvoří vektor $\mathbf{x} \in \mathbb{R}^n$. Je dána funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ s hodnotami

$$f(\mathbf{x}) = \sum_{\{i,j\} \in E} (x_i - x_j)^2.$$

a) Ukažte, že f je kvadratická forma.
b) Jaká je definitnost této kvadratické formy?
c) Pro jaká $\mathbf{x}$ platí $f(\mathbf{x}) = 0$?
d) Nechť $\mathbf{A} \in \{0, 1\}^{n \times n}$ je matice sousednosti grafu, tj. $a_{ij} = 1$ právě když $\{i, j\} \in E$. Ukažte, že $f(\mathbf{x}) = \mathbf{x}^T \mathbf{L} \mathbf{x}$ kde matice $\mathbf{L} \in \mathbb{R}^{n \times n}$ je definována jako $\mathbf{L} = \mathbf{D} - \mathbf{A}$ kde $\mathbf{D} = \operatorname{diag}(d_1, \dots, d_n) = \operatorname{diag}(\mathbf{A1})$ je diagonální matice a d_i je stupeň (tedy počet incidentních hran) vrcholu i.
e) Nechť $\mathbf{B} \in \{-1, 0, 1\}^{n \times m}$ (kde $m = |E|$) je incidenční matice orientovaného grafu vytvořeného tak, že pro každou hranu neorientovaného grafu (V, E) zvolíme libovolnou orientaci. Ukažte, že $f(\mathbf{x}) = \|\mathbf{B}^T \mathbf{x}\|^2$, tedy $\mathbf{L} = \mathbf{B}\mathbf{B}^T$, kde tyto výrazy nezávisejí na zvolených orientacích hran.

Poznamenejme, že funkci f (příp. matici $\mathbf{L}$) se říká *Laplacián* grafu (V, E).

6.16. Co se dá říct o definitnosti symetrické matice (ne nutně diagonální), známe-li znaménka jejích diagonálních prvků? Konkrétně, co se dá říct, jestliže její diagonální prvky jsou (a) $1, 2, 0$, (b) $1, 2, 3$, (c) $-4, -2, -1$, (d) $-1, 2, 0$.

6.17. Dokažte, že matice $\mathbf{A}^T \mathbf{A} + \mu \mathbf{I}$ je positivně definitní pro každou matici $\mathbf{A}$ a každé $\mu > 0$.

6.18. Dokažte, že symetrická matice je positivně definitní, právě když je positivně semidefinitní a invertovatelná.

6.19. Dokažte, že každá symetrická positivně definitní matice je invertovatelná a její inverze je také positivně definitní. Dokažte

a) s použitím maticové kongruence,
b) bez použití maticové kongruence.

6.20. Dokažte:

a) Pro každou matici $\mathbf{A}$ je matice $\mathbf{A}^T \mathbf{A}$ positivně semidefinitní.
b) Matice $\mathbf{A}^T \mathbf{A}$ je positivně definitní, právě když $\mathbf{A}$ má lineárně nezávislé sloupce,
c) Pro každou positivně semidefinitní $\mathbf{B}$ existuje $\mathbf{A}$ tak, že $\mathbf{B} = \mathbf{A}^T \mathbf{A}$.
d) Pro každou positivně definitní $\mathbf{B}$ existuje regulární $\mathbf{A}$ tak, že $\mathbf{B} = \mathbf{A}^T \mathbf{A}$.

6.21. Dokažte pomocí maticové kongruence, že determinant positivně [semi]definitní matice je kladný [nezáporný].

6.22. ($\star$) Positivně semidefinitní symetrické matice lze vidět jako zobecnění nezáporných čísel. Proto se někdy positivní semidefinitnost značí $\mathbf{A} \succeq \mathbf{0}$. Zápis $\mathbf{A} \preceq \mathbf{B}$ je pak zkratkou pro $\mathbf{B} - \mathbf{A} \succeq \mathbf{0}$.

a) Dokažte, že relace $\preceq$ je částečné uspořádání (tj. reflexivní, tranzitivní a antisymetrická) na množině symetrických matic $n \times n$.

[20] $\binom{V}{k}$ značí množinu všech k-prvkových podmnožin množiny V.

Na základě podobnosti relace $\preceq$ a relace $\leq$ na množině $\mathbb{R}$ bychom očekávali, že:

b) Pokud $\mathbf{A} \preceq \mathbf{B}$ a $\mathbf{C} \preceq \mathbf{D}$, potom $\mathbf{A} + \mathbf{C} \preceq \mathbf{B} + \mathbf{D}$.
c) Pokud $\mathbf{A} \succeq \mathbf{0}$ a $\alpha \geq 0$, potom $\alpha\mathbf{A} \succeq \mathbf{0}$.
d) Pokud $\mathbf{A} \succeq \mathbf{0}$, potom $\mathbf{A}^2 \succeq \mathbf{0}$.
e) Pokud $\mathbf{A} \succeq \mathbf{0}$ a $\mathbf{B} \succeq \mathbf{0}$, potom $\mathbf{AB} \succeq \mathbf{0}$.
f) Pokud $\mathbf{A} \succeq \mathbf{0}$ a $\mathbf{B} \succeq \mathbf{0}$, potom $\mathbf{ABA} \succeq \mathbf{0}$.

Které z těchto tvrzení platí a která neplatí? Odpovědi dokažte.

6.23. ($\star$) Geometrickou úvahou najděte aspoň dva vlastní vektory a příslušná vlastní čísla reflektoru.

6.24. ($\star$) Je známo, že libovolnou rotaci ve třírozměrném prostoru lze realizovat jako rotaci kolem jisté přímky (jdoucí počátkem) o jistý úhel. Geometrickou úvahou zjistěte co nejvíce o vlastních číslech a vektorech rotační matice rozměru 3×3.

6.25. Dokažte, že každý ortogonální projektor je positivně semidefinitní.

6.26. ($\star$) Dokažte, že projektor má vždy vlastní číslo 1, někdy má i vlastní číslo 0 a jiná vlastní čísla mít nemůže.

6.27. Jaké je znaménko determinantu negativně (semi)definitní matice?

6.28. Na Wikipedii vyhledejte hesla *conic section* (kuželosečka) a *quadric surface* (kvadrika). Všechny typy kuželoseček a kvadrik (nedegenerovaných i denenerovaných) tam uvedené zkuste napsat jako vrstevnice vhodných kvadratických forem nebo funkcí.

6.29. ($\star$) Elipsu lze definovat jako množinu bodů v rovině, které mají od dané dvojice bodů (ohnisek) konstantní součet vzdáleností[21]. Pokud elipsa má střed v počátku, je to tedy množina bodů $\mathbf{x}$ splňujících $\|\mathbf{x} - \mathbf{a}\| + \|\mathbf{x} + \mathbf{a}\| = 1$. Člověk by si mohl myslet, že v $\mathbb{R}^n$ bude tato množina elipsoid se středem v počátku, tedy bude popsána rovnicí $\mathbf{x}^T\mathbf{A}\mathbf{x} = 1$ pro nějaké positivně definitní $\mathbf{A}$. Je to pravda? Pokud ano, jak se najde $\mathbf{A}$ známe-li $\mathbf{a}$?

Nápověda a řešení

6.5. Vlastní čísla se zvětší o α. Vlastní vektory jsou stejné.

6.6. Nechť $\mathbf{ABv} = \lambda\mathbf{v}$ kde $\mathbf{v} \neq \mathbf{0}$ (vlastní vektory nesmí být nulové) a $\lambda \neq 0$ (předpoklad). Z toho plyne $\mathbf{BABv} = \lambda\mathbf{Bv}$ a $\mathbf{Bv} \neq \mathbf{0}$, tedy $\mathbf{BAu} = \lambda\mathbf{u}$, kde $\mathbf{u} = \mathbf{Bv} \neq \mathbf{0}$.

6.7. Nechť $\mathbf{A}$ je symetrická. Ze spektrálního rozkladu je $\det \mathbf{A} = \det(\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T) = \det \mathbf{V} \det \mathbf{\Lambda} \det(\mathbf{V}^T) = \det \mathbf{V} \det(\mathbf{V}^T) \det \mathbf{\Lambda} = \det(\mathbf{V}\mathbf{V}^T) \det \mathbf{\Lambda} = \det \mathbf{I} \det \mathbf{\Lambda} = \det \mathbf{\Lambda} = \lambda_1 \cdots \lambda_n$, kde jsme dvakrát použili vzorec pro determinant součinu matic (§2.1.5).

6.8. (a) Symetrická Gaussova eliminace: od druhého řádku/sloupce odečteme dvojnásobek prvního, čímž dostaneme diagonální matici $\left[\begin{smallmatrix} 1 & 0 \\ 0 & -3 \end{smallmatrix}\right]$. Protože na diagonále jsou čísla s různými znaménky, dle Tvrzení 6.5 a 6.6 je vstupní matice indefinitní.

Důkaz indefinitnosti výpočtem vlastních čísel (viz Tvrzení 6.9) je pracnější, vyjdou -1 a 3.

Pomocí hlavních minorů: Označme vstupní matici jako $\mathbf{A} = \left[\begin{smallmatrix} 1 & 2 \\ 2 & 1 \end{smallmatrix}\right]$. Její hlavní minory jsou $1, -3, 1$, tedy matice $\mathbf{A}$ není positivně semidefinitní. Hlavní minory matice $-\mathbf{A}$ jsou $-1, -3, -1$, tedy $-\mathbf{A}$ není positivně semidefinitní, tedy $\mathbf{A}$ není negativně semidefinitní. Z toho plyne, že $\mathbf{A}$ je indefinitní.

[21] Možná víte, jak pomocí této definice nakreslit elipsu. Potřebujete papír, dva špendlíky (hřebíky...) a tužku. Špendlíky zapíchnete do papíru (jejich vzdálenost musí být menší než délka provázku) a přivážete k nim oba konce provázku. Pak už jen tužkou jedete po papíru tak, aby udržovala provázek stále napnutý.

(b) positivně definitní. Sym. Gaussova eliminace dá matici $\begin{bmatrix} 2 & 0 \\ 0 & 3/2 \end{bmatrix}$. Vl. čísla jsou $1, 3$.

(c) Díky Důsledku 6.4 (viz Příklad 6.3) okamžitě vidíme, že matice je indefinitní.

Sym. Gaussova eliminace: nejprve musíme udělat ekv. úpravu, která učiní pivot a_{11} nenulový, např. přičteme druhý řádek/sloupec k prvnímu, což dá $\begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix}$. Pak odečteme polovinu prvního řádku/sloupce od druhého, což dá $\begin{bmatrix} 2 & 0 \\ 0 & -1/2 \end{bmatrix}$.

Vl. čísla jsou $-1, 1$.

(d) positivně semidefinitní. Matice je již diagonální, takže nemusíme nic dělat. Vl. čísla jsou její diagonální prvky, $0, 1$.

(e) indefinitní, což ihned vidíme díky Důsledku 6.4.

Kdybychom přesto chtěli dělat sym. Gaussovu eliminace, pak ta v první iteraci dělat nic, protože prvky pod a vpravo od pivotu a_{11} jsou už nulové. Ve druhé iteraci musí učinit pivot a_{22} nenulový např. přičtením třetího řádku/sloupce k druhému. Pak vynuluje prvky a_{32} a a_{23}. Výsledek je $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1/2 \end{bmatrix}$.

Výpočet vl. čísel dá více práce (kubická rovnice), vyjdou ale náhodou hezky: $-1, 1, 1$.

(f) positivně definitní. Sym. Gaussova eliminace dá postupně $\begin{bmatrix} 3 & 0 & 0 \\ 0 & 5/3 & 2/3 \\ 0 & 2/3 & 8/3 \end{bmatrix}$ a $\begin{bmatrix} 3 & 0 & 0 \\ 0 & 5/3 & 0 \\ 0 & 0 & 12/5 \end{bmatrix}$.

(g) indefinitní, což ihned vidíme ze znamének diagonálních prvků (viz Cvičení 6.16).

6.9. Matice $\mathbf{A}$ není symetrická. Musíme ji nejdříve symetrizovat, tj. vzít matici $\frac{1}{2}(\mathbf{A} + \mathbf{A}^T)$, a pak teprv počítat vlastní čísla. Tím zjistíme, že žádné tvrzení neplatí.

6.10.a) $\mathbf{A} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$

6.10.b) $a = 2$, $b = 4$, $\mathbf{U} = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$

6.11. Hyperbola, neboť $\mathbf{A}$ má nenulová vlastní čísla opačných znamének.

6.13. Rexlexivnost: každá matice je kongruentní sama sobě, neboť v (6.5) lze zvolit $\mathbf{C} = \mathbf{I}$. Symetrie: $\mathbf{A} \sim \mathbf{B}$ znamená, že (6.5) platí pro nějakou regulární $\mathbf{C}$. Pak ale matice $\mathbf{C}^{-1}$ je také regulární a platí $\mathbf{A} = \mathbf{C}^{-T}\mathbf{B}\mathbf{C}^{-1}$, tedy z definice kongruence máme $\mathbf{B} \sim \mathbf{A}$. Transitivnost dokažte sami.

6.14.a) Převod na tvar (6.20): $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$, $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 3 \\ -6 \end{bmatrix}$, $c = 5$. Doplnění na čtverec existuje (protože $\mathbf{A}$ je regulární). Funkce nemá extrém (protože $\mathbf{A}$ je indefinitní), má sedlo v bodě $(\frac{1}{2}, -1)$.

6.14.b) Má minimum v bodě $-(3, 1)/2$.

6.15.a) Je $(x_i - x_j)^2 = x_i^2 - 2x_i x_j + x_j^2$. Tedy f je homogenní polynom stupně dva, což je kvadratická forma.

6.15.b) Funkce je součet čtverců (druhých mocnin), tedy $f(\mathbf{x}) \geq 0$ pro všechna $\mathbf{x} \in \mathbb{R}^n$, tedy f je positivně semidefinitní. Zároveň $f(\mathbf{x}) = 0$, když všechny složky x_i jsou stejné (tedy $\mathbf{x} = \alpha\mathbf{1}$ pro libovolné $\alpha \in \mathbb{R}$), tedy není positivně definitní.

6.15.c) Zjevně $f(\mathbf{x}) = 0$, právě když $x_i = x_j$ pro všechna $\{i, j\} \in E$. Když graf je souvislý, nastane to právě tehdy, když všechna x_i jsou stejná. Jinak to nastane právě tehdy, když jsou x_i stejná v každé komponentě grafu.

6.15.d) $\displaystyle\sum_{\{i,j\}} (x_i - x_j)^2 = \frac{1}{2} \sum_{\{i,j\}} x_i^2 - \sum_{\{i,j\}} x_i x_j + \frac{1}{2} \sum_{\{i,j\}} x_j^2 = \sum_{\{i,j\}} x_i^2 - \sum_{\{i,j\}} x_i x_j = \mathbf{x}^T\mathbf{D}\mathbf{x} - \mathbf{x}^T\mathbf{A}\mathbf{x}$

6.16. Viz Důsledek 6.3. Výsledky: (a) matice nemůže být positivně definitní ani negativně semidefinitní, (b) nemůže být negativně semidefinitní, (c) nemůže být positivně semidefinitní, (d) nemůže být positivně indefinitní ani negativně indefinitní, tudíž je indefinitní.

6.17. $\mathbf{x}^T(\mathbf{A}^T\mathbf{A}+\mu\mathbf{I})\mathbf{x} = \mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x}+\mu\mathbf{x}^T\mathbf{I}\mathbf{x} = \|\mathbf{A}\mathbf{x}\|^2+\mu\|\mathbf{x}\|^2 > 0$ pro každé $\mathbf{x}\neq\mathbf{0}$.

6.18. Nechť $\mathbf{A}=\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$. Matice $\mathbf{A}$ je positivně semidefinitní, právě když jsou všechny diagonální prvky matice $\mathbf{\Lambda}$ (což jsou vlastní čísla matice $\mathbf{A}$) nezáporné. Jak jsme řekli v §6.3.2, $\mathbf{A}$ je invertovatelná právě když $\mathbf{\Lambda}$ je invertovatelná, neboli $\mathbf{\Lambda}$ má na diagonále nenulové prvky. Ale $\mathbf{A}$ je positivně definitní, právě když její vl. čísla jsou kladná, tj. nezáporná a nenulová.

6.19.a) V §6.3.1 jsme ukázali, že každá symetrická matice je kongruentní nějaké diagonální matici, tedy existuje regulární $\mathbf{C}$ a diagonální $\mathbf{B}$ tak, že $\mathbf{A}=\mathbf{C}^T\mathbf{B}\mathbf{C}$. Dle Tvrzení 6.5 je $\mathbf{B}$ positivně definitní, tedy dle Tvrzení 6.6 má na diagonále kladné prvky a je tedy regulární. Je tedy $\mathbf{A}^{-1}=(\mathbf{C}^T\mathbf{B}\mathbf{C})^{-1}=\mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{C}^{-T}$. Protože inverze diagonální matice $\mathbf{B}$ je jednoduše $\mathbf{B}^{-1}=\operatorname{diag}(1/b_{11},\ldots,1/b_{nn})$, je $\mathbf{B}^{-1}$ také positivně definitní a dle Tvrzení 6.6 tedy i $\mathbf{A}^{-1}$.

6.19.b) Nechť $\mathbf{A}$ je positivně definitní, tedy $\mathbf{x}^T\mathbf{A}\mathbf{x}>0$ pro každé $\mathbf{x}\neq\mathbf{0}$. Kdyby $\mathbf{A}$ nebyla invertovatelná (tj. regulární), dle Věty 3.9 by měla netriviální nulový prostor, tedy $\mathbf{A}\mathbf{x}=\mathbf{0}$ pro nějaké $\mathbf{x}\neq\mathbf{0}$. Ale to nelze, protože pak by také $\mathbf{x}^T\mathbf{A}\mathbf{x}=0$.

Nechť $\mathbf{A}$ je invertovatelná. Položme $\mathbf{y}=\mathbf{A}\mathbf{x}$ a $\mathbf{x}=\mathbf{A}^{-1}\mathbf{y}$. Víme, že $\mathbf{x}^T\mathbf{A}\mathbf{x}>0$ pro každé $\mathbf{x}\neq\mathbf{0}$. Z toho $\mathbf{x}^T\mathbf{A}\mathbf{x}=\mathbf{y}^T\mathbf{A}^{-T}\mathbf{A}\mathbf{A}^{-1}\mathbf{y}=\mathbf{y}^T\mathbf{A}^{-1}\mathbf{y}>0$. Protože zobrazení $\mathbf{f}(\mathbf{x})=\mathbf{A}\mathbf{x}$ je bijekce, je $\mathbf{y}^T\mathbf{A}^{-1}\mathbf{y}>0$ pro každé $\mathbf{y}\neq\mathbf{0}$.

6.20.a) $\mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x}=(\mathbf{A}\mathbf{x})^T\mathbf{A}\mathbf{x}=\|\mathbf{A}\mathbf{x}\|^2\geq 0$

6.20.b) Víme, že $\mathbf{A}^T\mathbf{A}$ je p.s.d. pro každou $\mathbf{A}$. Tedy stačí dokázat, že $\mathbf{A}^T\mathbf{A}$ je invertovatelná, právě když $\mathbf{A}$ má l.n. sloupce. To jsme dokázali v (5.7).

6.20.c) Ve spektrálním rozkladu $\mathbf{B}=\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ má $\mathbf{\Lambda}$ nezáporné diagonální prvky. Položme $\mathbf{A}=\mathbf{\Lambda}^{1/2}\mathbf{V}^T$.

6.20.d) Ve spektrálním rozkladu $\mathbf{B}=\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ má $\mathbf{\Lambda}$ kladné diagonální prvky. Položme $\mathbf{A}=\mathbf{\Lambda}^{1/2}\mathbf{V}^T$.

6.21. Víme, že ke každé symetrické matici $\mathbf{A}$ existuje kongruentní diagonální $\mathbf{B}$, tj. $\mathbf{A}=\mathbf{C}^T\mathbf{B}\mathbf{C}$ pro nějakou regulární $\mathbf{C}$. Máme $\det\mathbf{A}=\det(\mathbf{C}^T\mathbf{B}\mathbf{C})=\det(\mathbf{C}^T)\det\mathbf{B}\det\mathbf{C}=\det\mathbf{C}\det\mathbf{B}\det\mathbf{C}=(\det\mathbf{C})^2\det\mathbf{B}$. Ale je-li $\mathbf{A}$ positivně semidefinitní, pak je i $\mathbf{B}$ a tedy $\det\mathbf{B}\geq 0$. Zároveň $\det\mathbf{C}\neq 0$ protože $\mathbf{C}$ je regulární. Srov. Cvičení 6.7.

6.22.a) Relace je reflexivní, když $\mathbf{A}\preceq\mathbf{A}$ pro každou $\mathbf{A}$, neboli $\mathbf{A}-\mathbf{A}\succeq\mathbf{0}$, což platí.

Pro další si všimněte, že $\mathbf{A}\preceq\mathbf{B}$ znamená $(\forall\mathbf{x})(\mathbf{x}^T\mathbf{A}\mathbf{x}\leq\mathbf{x}^T\mathbf{B}\mathbf{x})$. Relace je antisymetrická, když $\mathbf{A}\preceq\mathbf{B}\preceq\mathbf{A}\ \Rightarrow\ \mathbf{A}=\mathbf{B}$. Ale $\mathbf{x}^T\mathbf{A}\mathbf{x}\leq\mathbf{x}^T\mathbf{B}\mathbf{x}\leq\mathbf{x}^T\mathbf{A}\mathbf{x}$ implikuje $\mathbf{x}^T\mathbf{A}\mathbf{x}=\mathbf{x}^T\mathbf{B}\mathbf{x}$. To platí pro všechna $\mathbf{x}$, protože $\mathbf{A},\mathbf{B}$ jsou symetrické.

Relace je tranzitivní, když $\mathbf{A}\preceq\mathbf{B}\preceq\mathbf{C}\ \Rightarrow\ \mathbf{A}\preceq\mathbf{C}$. To platí, neboť $\mathbf{x}^T\mathbf{A}\mathbf{x}\leq\mathbf{x}^T\mathbf{B}\mathbf{x}\leq\mathbf{x}^T\mathbf{C}\mathbf{x}\ \Rightarrow\ \mathbf{x}^T\mathbf{A}\mathbf{x}\leq\mathbf{x}^T\mathbf{C}\mathbf{x}$.

6.25. Chceme dokázat, že pro každý ortog. projektor $\mathbf{P}$ a každý vektor $\mathbf{x}$ platí $\mathbf{x}^T\mathbf{P}\mathbf{x}\geq 0$. Protože ortog. projektor splňuje $\mathbf{P}^2=\mathbf{P}=\mathbf{P}^T$, máme $\mathbf{P}=\mathbf{P}^T\mathbf{P}$ a tedy $\mathbf{x}^T\mathbf{P}\mathbf{x}=\mathbf{x}^T\mathbf{P}^T\mathbf{P}\mathbf{x}=\|\mathbf{P}\mathbf{x}\|$, což je jistě nezáporné číslo. Geometricky to znamená, že úhel mezi vektory $\mathbf{x}$ a $\mathbf{P}\mathbf{x}$ je pravý nebo ostrý, dle (4.3)).

6.26. Projektor $\mathbf{P}$ splňuje $\mathbf{P}^2=\mathbf{P}$. Je-li λ vlastní číslo $\mathbf{P}$, pak $\mathbf{P}\mathbf{v}=\lambda\mathbf{v}$ pro nějaké $\mathbf{v}\neq\mathbf{0}$. Vynásobením maticí $\mathbf{P}$ zleva z toho máme $\mathbf{P}^2\mathbf{v}=\mathbf{P}\mathbf{v}=\lambda\mathbf{P}\mathbf{v}$. Jestliže $\mathbf{P}\mathbf{v}\neq\mathbf{0}$, pak $\lambda=1$. Jestliže $\mathbf{P}\mathbf{v}=\mathbf{0}$, pak $\lambda=0$ (neboť $\mathbf{v}\neq\mathbf{0}$).

Osvětleme tento výsledek také geometricky. Vlastní vektory $\mathbf{v}$ příslušné $\lambda=0$ splňují $\mathbf{P}\mathbf{v}=\mathbf{0}$, tedy tvoří podprostor $\operatorname{null}\mathbf{P}$ bez počátku. To geometricky odpovídá tomu, že všechny vektory z $\operatorname{null}\mathbf{P}$ se promítají do počátku. Vlastní vektory příslušné $\lambda=1$ splňují $\mathbf{P}\mathbf{v}=\mathbf{v}$, což (díky vlastnostem projektoru) platí jen pro $\mathbf{v}\in\operatorname{rng}\mathbf{P}$, tedy tvoří podprostor $\operatorname{rng}\mathbf{P}$ bez počátku. To odpovídá tomu, že vektory z $\operatorname{rng}\mathbf{P}$ se projekcí nezmění.

6.27. Ze §6.2 víme, že positivně [semi]definitní matice má determinant kladný [nezáporný]. Dále víme, že matice $\mathbf{A}$ je negativně [semi]definitní, právě když $-\mathbf{A}$ je positivně [semi]definitní. Pro matici

$\mathbf{A} \in \mathbb{R}^{n\times n}$ platí $\det(-\mathbf{A}) = (-1)^n \det \mathbf{A}$ (viz Cvičení 2.22). Tedy negativně [semi]definitní matice má determinant kladný [nezáporný] pro sudé n a záporný [nekladný] pro liché n.

Kapitola 7

PCA a SVD

7.1 Úloha na největší/nejmenší stopu

Ukážeme nyní, že vlastní čísla a vektory symetrické matice úzce souvisí s jistými jednoduchými optimalizačními úlohami. Na základě těchto teoretických výsledků pak v §7.2 formulujeme algoritmus na proložení bodů podprostorem (PCA).

Nechť $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ je spektrální rozklad symetrické matice $\mathbf{A} \in \mathbb{R}^{n\times n}$, tedy matice $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ je diagonální a $\mathbf{V} = \begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_n\end{bmatrix}$ je ortogonální. Předpokládáme opět, že vlastní čísla jsou *sestupně* seřazena, $\lambda_1 \geq \cdots \geq \lambda_n$.

Věta 7.1. Optimální hodnota úlohy

$$\max\{\, \mathbf{x}^T\mathbf{A}\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{x}^T\mathbf{x} = 1 \,\} \tag{7.1}$$

je rovna λ_1 (tedy největšímu vlastnímu číslu matice $\mathbf{A}$) a nabývá se pro[1] $\mathbf{x} = \mathbf{v}_1$.

Důkaz. Myšlenka důkazu je jednoduchá: dosadíme-li za matici $\mathbf{A}$ její spektrální rozklad, řešení úlohy se stane očividným. Zopakujme rovnost (6.18),

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{x}^T\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T\mathbf{x} = \mathbf{y}^T\mathbf{\Lambda}\mathbf{y} = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2,$$

kde jsme substitucí $\mathbf{x} = \mathbf{V}\mathbf{y}$ (neboli $\mathbf{y} = \mathbf{V}^T\mathbf{x}$) diagonalizovali kvadratickou formu. Protože $\mathbf{V}$ je ortogonální, je $\mathbf{x}^T\mathbf{x} = \mathbf{y}^T\mathbf{V}^T\mathbf{V}\mathbf{y} = \mathbf{y}^T\mathbf{y}$, tedy podmínka $\mathbf{x}^T\mathbf{x} = 1$ je ekvivalentní podmínce $\mathbf{y}^T\mathbf{y} = 1$. Tedy úloha (7.1) má stejnou optimální hodnotu jako úloha

$$\max\{\, \mathbf{y}^T\mathbf{\Lambda}\mathbf{y} \mid \mathbf{y}^T\mathbf{y} = 1,\ \mathbf{y} \in \mathbb{R}^n \,\} = \max\{\, \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2 \mid y_1^2 + \cdots + y_n^2 = 1,\ y_1, \ldots, y_n \in \mathbb{R} \,\}. \tag{7.2}$$

Elementární úvahou (viz Cvičení 7.3 a Příklad 12.4) vidíme, že maximum nastane pro $y_1 = 1$ a $y_2 = \cdots = y_n = 0$, tj. pro $\mathbf{y} = \mathbf{e}_1$ (první vektor standardní báze). To odpovídá $\mathbf{x} = \mathbf{V}\mathbf{e}_1 = \mathbf{v}_1$ a $\mathbf{y}^T\mathbf{\Lambda}\mathbf{y} = \lambda_1$. ■

Úlohu (7.1) lze zobecnit:

[1] Netvrdíme, že se optimum nabývá *pouze* v bodě $\mathbf{v}_n$. Je-li vlastní číslo λ_1 násobné, nabývá se ve více bodech.

Věta 7.2. Nechť $k \leq n$. Optimální hodnota úlohy[2]

$$\max\{\, \mathrm{tr}(\mathbf{X}^T\mathbf{A}\mathbf{X}) \mid \mathbf{X} \in \mathbb{R}^{n\times k},\ \mathbf{X}^T\mathbf{X} = \mathbf{I}\,\} \tag{7.3}$$

je rovna $\lambda_1 + \cdots + \lambda_k$ (tedy součtu k největších vlastních čísel matice $\mathbf{A}$) a nabývá se pro $\mathbf{X} = \begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_k\end{bmatrix}$.

Před důkazem Věty 7.2 osvětlíme její význam. Označíme-li sloupce matice $\mathbf{X}$ jako $\mathbf{x}_1, \ldots, \mathbf{x}_k \in \mathbb{R}^n$, podmínka $\mathbf{X}^T\mathbf{X} = \mathbf{I}$ říká, že tyto vektory musejí být ortonormální. Dále platí (roznásobte blokové matice a použijte definici stopy, viz Cvičení 7.6)

$$\mathrm{tr}(\mathbf{X}^T\mathbf{A}\mathbf{X}) = \mathbf{x}_1^T\mathbf{A}\mathbf{x}_1 + \cdots + \mathbf{x}_k^T\mathbf{A}\mathbf{x}_k. \tag{7.4}$$

Tedy v úloze (7.3) hledáme ortonormální vektory $\mathbf{x}_1, \ldots, \mathbf{x}_k$, které maximalizují součet kvadratických forem (7.4). Věta říká, že takové vektory jsou právě $\mathbf{v}_1, \ldots, \mathbf{v}_k$, tedy (normalizované) vlastní vektory příslušné k největším vlastním číslům. Nyní důkaz věty:

Důkaz. Substitucí $\mathbf{X} = \mathbf{V}\mathbf{Y}$ (neboli $\mathbf{Y} = \mathbf{V}^T\mathbf{X}$) opět převedeme úlohu na jednodušší tvar. Je $\mathrm{tr}(\mathbf{X}^T\mathbf{A}\mathbf{X}) = \mathrm{tr}(\mathbf{X}^T\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T\mathbf{X}) = \mathrm{tr}(\mathbf{Y}^T\boldsymbol{\Lambda}\mathbf{Y})$. Protože $\mathbf{X}^T\mathbf{X} = \mathbf{Y}^T\mathbf{V}^T\mathbf{V}\mathbf{Y} = \mathbf{Y}^T\mathbf{Y}$, podmínka $\mathbf{X}^T\mathbf{X} = \mathbf{I}$ je ekvivalentní podmínce $\mathbf{Y}^T\mathbf{Y} = \mathbf{I}$. Tedy úloha (7.3) má stejnou optimální hodnotu jako úloha

$$\max\{\, \mathrm{tr}(\mathbf{Y}^T\boldsymbol{\Lambda}\mathbf{Y}) \mid \mathbf{Y} \in \mathbb{R}^{n\times k},\ \mathbf{Y}^T\mathbf{Y} = \mathbf{I}\,\}. \tag{7.5}$$

Řešme úlohu (7.5). Z vlastností stopy (viz §2.1.6) máme

$$\mathrm{tr}(\mathbf{Y}^T\boldsymbol{\Lambda}\mathbf{Y}) = \mathrm{tr}(\mathbf{Y}\mathbf{Y}^T\boldsymbol{\Lambda}) = \mathrm{tr}(\mathbf{P}\boldsymbol{\Lambda}) = \lambda_1 p_{11} + \cdots + \lambda_n p_{nn},$$

kde čísla $p_{11}, \ldots, p_{nn}$ jsou diagonální prvky matice (ortogonálního projektoru) $\mathbf{P} = \mathbf{Y}\mathbf{Y}^T$. Tato čísla splňují rovnost

$$p_{11} + \cdots + p_{nn} = \mathrm{tr}\,\mathbf{P} = \mathrm{tr}(\mathbf{Y}\mathbf{Y}^T) = \mathrm{tr}(\mathbf{Y}^T\mathbf{Y}) = \mathrm{tr}\,\mathbf{I}_k = k.$$

Dále splňují nerovnosti $0 \leq p_{ii} \leq 1$ pro každé $i = 1, \ldots, n$ (viz Cvičení 5.15). Uvažujme úlohu[3]

$$\max\{\, \lambda_1 p_{11} + \cdots + \lambda_n p_{nn} \mid p_{11} + \cdots + p_{nn} = k,\ 0 \leq p_{ii} \leq 1\ \forall i = 1, \ldots, n\,\}. \tag{7.6}$$

Tuto úlohu vyřešíme snadno úvahou (viz Příklad 12.5): maximální hodnota je $\lambda_1 + \cdots + \lambda_k$ a nabývá se pro $p_{11} = \cdots = p_{kk} = 1$ a $p_{k+1,k+1} = \cdots = p_{nn} = 1$. Ale matici $\mathbf{P} = \mathbf{Y}\mathbf{Y}^T$ s těmito diagonálními prvky lze realizovat volbou $\mathbf{Y} = \begin{bmatrix}\mathbf{I}_k \\ \mathbf{0}\end{bmatrix}$. Z toho $\mathbf{X} = \mathbf{V}\mathbf{Y} = \begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_k\end{bmatrix}$. ∎

Uveďme ještě 'bezsouřadnicový' pohled na úlohu 7.3. Je-li $X \subseteq \mathbb{R}^m$ podprostor dimenze k, pak výraz (7.4) je stejný pro všechny ortonormální báze $\mathbf{X} = \begin{bmatrix}\mathbf{x}_1 & \cdots & \mathbf{x}_k\end{bmatrix}$ podprostoru X: opravdu, z cykličnosti stopy máme $\mathrm{tr}(\mathbf{X}^T\mathbf{A}\mathbf{X}) = \mathrm{tr}(\mathbf{X}\mathbf{X}^T\mathbf{A}) = \mathrm{tr}(\mathbf{P}\mathbf{A})$, kde $\mathbf{P} = \mathbf{X}\mathbf{X}^T$ je ortogonální projektor na podprostor X, o němž z §5.2 víme, že nezávisí na bázi podprostoru. Proto můžeme výraz (7.4) označit $\mathrm{tr}_X\,\mathbf{A}$ a nazvat ho *stopa matice* $\mathbf{A}$ *na podprostoru* X. V úloze 7.2 pak hledáme podprostor $X \subseteq \mathbb{R}^m$ dimenze k, na kterém je stopa $\mathrm{tr}_X\,\mathbf{A}$ matice $\mathbf{A}$ největší.

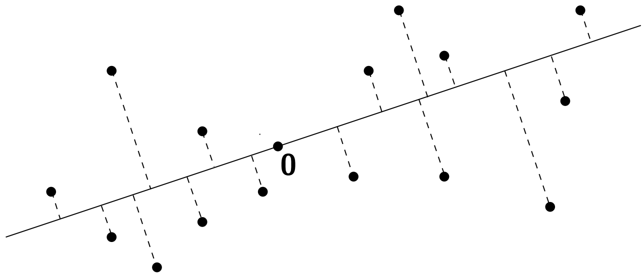
Kdybychom maximum v úloze (7.3) nahradili minimem, analogicky by se dokázalo, že optimální hodnota je $\lambda_{n-k+1} + \cdots + \lambda_n$ (součet k nejmenších vlastních čísel) a nabývá se pro $\mathbf{X} = \begin{bmatrix}\mathbf{v}_{n-k+1} & \cdots & \mathbf{v}_n\end{bmatrix}$.

[2] Namítnete, že úloha (7.3) není ve tvaru (1.14), protože optimalizujeme přes množinu *matic* a ne vektorů. To je ale jen záležitost značení, neboť matici rozměru $m \times n$ lze vidět jako vektor s mn složkami.

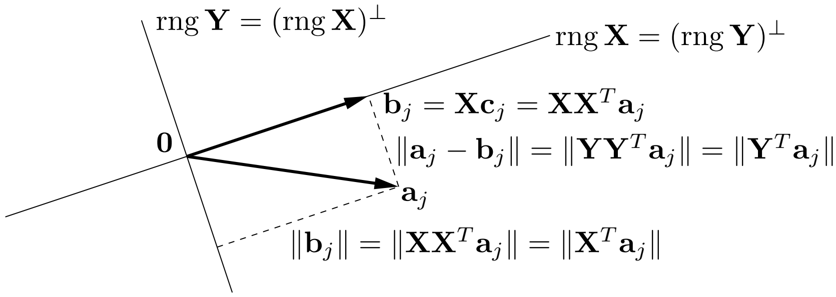
[3] Úloha (7.5) je zobecněním úlohy (7.1). Abyste to lépe viděli, uvědomte si, že úlohu (7.2) lze psát jako $\max\{\, \lambda_1 p_1 + \cdots + \lambda_n p_n \mid p_1 + \cdots + p_n = 1,\ 0 \leq p_i \leq 1\ \forall i = 1, \ldots, n\,\}$, kde jsme označili $p_i = y_i^2$.

7.2 Proložení bodů podprostorem

Nechť jsou dány body $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$ a přirozené číslo $k \le m$. Hledejme lineární podprostor dimenze k prostoru $\mathbb{R}^m$, který minimalizuje součet čtverců vzdáleností k bodům $\mathbf{a}_1, \ldots, \mathbf{a}_n$. Pro pohodlí uložíme dané body jako sloupce do matice $\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} \in \mathbb{R}^{m \times n}$. Obrázek ukazuje příklad bodů (plná kolečka) a optimálního podprostoru (plná čára) pro $n = 15$, $m = 2$, $k = 1$ (minimalizujeme součet druhých mocnin délek přerušovaných úseček):



Reprezentujme hledaný podprostor jeho ortonormální bází, která nechť tvoří sloupečky matice $\mathbf{X} \in \mathbb{R}^{m \times k}$, tedy $\mathbf{X}^T\mathbf{X} = \mathbf{I}$. Uvažujme i ortogonální doplněk tohoto podprostoru, který dle Věty 4.1 má dimenzi $m - k$. Budeme ho reprezentovat ortonormální bází, tvořící sloupce matice $\mathbf{Y} \in \mathbb{R}^{m \times (m-k)}$, tedy máme $\mathbf{Y}^T\mathbf{Y} = \mathbf{I}$ a $(\operatorname{rng} \mathbf{X})^\perp = \operatorname{rng} \mathbf{Y}$. Viz obrázek (zatím na obrázku ignorujte symboly s $\mathbf{b}_j$, které budeme potřebovat až později):



Vzdálenost bodu $\mathbf{a}_j$ od podprostoru $\operatorname{rng} \mathbf{X}$ je délka projekce na jeho ortogonální doplněk, tj. $\|\mathbf{Y}\mathbf{Y}^T\mathbf{a}_j\| = \|\mathbf{Y}^T\mathbf{a}_j\|$ (viz §5.2.1). Chceme minimalizovat součet čtverců těchto vzdáleností

$$\|\mathbf{Y}^T\mathbf{a}_1\|^2 + \cdots + \|\mathbf{Y}^T\mathbf{a}_n\|^2 = \|\mathbf{Y}^T\mathbf{A}\|^2, \tag{7.7}$$

kde $\|\cdot\|$ ve druhém výrazu značí Frobeniovu maticovou normu (2.11). Řešíme tedy úlohu

$$\min\{\, \|\mathbf{Y}^T\mathbf{A}\|^2 \mid \mathbf{Y} \in \mathbb{R}^{m \times (m-k)},\ \mathbf{Y}^T\mathbf{Y} = \mathbf{I} \,\}. \tag{7.8}$$

Lze se na to ale podívat i jinak. Z Pythagorovy věty máme

$$\|\mathbf{X}^T\mathbf{a}_j\|^2 + \|\mathbf{Y}^T\mathbf{a}_j\|^2 = \|\mathbf{a}_j\|^2, \tag{7.9}$$

kde $\|\mathbf{X}\mathbf{X}^T\mathbf{a}_j\| = \|\mathbf{X}^T\mathbf{a}_j\|$ je vzdálenost bodu $\mathbf{a}_j$ od podprostoru $\operatorname{rng} \mathbf{Y}$ (viz obrázek). Po dosazení $\|\mathbf{Y}^T\mathbf{a}_j\|^2$ ze (7.9) vidíme, že výraz (7.7) je, až na konstantu $\|\mathbf{a}_1\|^2 + \cdots + \|\mathbf{a}_n\|^2 = \|\mathbf{A}\|^2$, roven zápornému výrazu

$$\|\mathbf{X}^T\mathbf{a}_1\|^2 + \cdots + \|\mathbf{X}^T\mathbf{a}_n\|^2 = \|\mathbf{X}^T\mathbf{A}\|^2 = \operatorname{tr}(\mathbf{X}^T\mathbf{A}\mathbf{A}^T\mathbf{X}). \tag{7.10}$$

Ukázali jsme, že minimalizace součtu čtverců vzdáleností bodů od podprostoru je ekvivalentní maximalizaci součtu čtverců vzdáleností bodů od jeho ortogonálního doplňku. To vede na úlohu

$$\max\{\, \|\mathbf{X}^T\mathbf{A}\|^2 \mid \mathbf{X} \in \mathbb{R}^{m\times k},\ \mathbf{X}^T\mathbf{X} = \mathbf{I}\,\}. \tag{7.11}$$

Z vlastností Frobeniovy normy a stopy (viz §2.1.6 a §2.1.7) plyne $\|\mathbf{X}^T\mathbf{A}\|^2 = \operatorname{tr}(\mathbf{X}^T\mathbf{A}\mathbf{A}^T\mathbf{X})$ a $\|\mathbf{Y}^T\mathbf{A}\|^2 = \operatorname{tr}(\mathbf{Y}^T\mathbf{A}\mathbf{A}^T\mathbf{Y})$. Tedy úloha (7.8) resp. (7.11) je úloha na nejmenší resp. největší stopu[4]. Řešení obou úloh získáme ze spektrálního rozkladu matice $\mathbf{A}\mathbf{A}^T$:

Tvrzení 7.3. Úlohu na proložení bodů podprostorem vyřešíme tak, že spočítáme spektrální rozklad $\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T = \mathbf{A}\mathbf{A}^T$ matice $\mathbf{A}\mathbf{A}^T$ (kde vlastní čísla jsou sestupně seřazena) a ortogonální matici $\mathbf{V} = \begin{bmatrix}\mathbf{X} & \mathbf{Y}\end{bmatrix} \in \mathbb{R}^{m\times m}$ rozdělíme do dvou bloků $\mathbf{X} \in \mathbb{R}^{m\times k}$ a $\mathbf{Y} \in \mathbb{R}^{m\times(m-k)}$. Pak:

- Matice $\mathbf{X}$ (tj. vlastní vektory odpovídající k největším vlastním číslům) jsou řešením úlohy (7.11), její sloupce jsou tedy ortonormální báze hledaného podprostoru.
- Matice $\mathbf{Y}$ (tj. vlastní vektory odpovídající $m-k$ nejmenším vlastním číslům) je řešením úlohy (7.8), její sloupce jsou tedy ortonormální báze ortogonálního doplňku hledaného podprostoru.

Důkaz. Plyne okamžitě z Věty 7.2 a Tvrzení 4.7. ■

Optimální hodnota úlohy (7.8) je chyba proložení našich bodů $\mathbf{a}_1, \ldots, \mathbf{a}_n$ podprostorem dimenze k, neboli je to míra, do jaké tyto body neleží v (nějakém) podprostoru dimenze k. Dle §7.1 je tato optimální hodnota rovna

$$\lambda_{k+1} + \cdots + \lambda_m, \tag{7.12}$$

tj. součet $m-k$ nejmenších vlastních čísel matice $\mathbf{A}\mathbf{A}^T$.

Příklad 7.1. Jsou dány body $\mathbf{a}_1, \ldots, \mathbf{a}_n$ v prostoru $\mathbb{R}^3$, jež tvoří sloupce matice $\mathbf{A} \in \mathbb{R}^{3\times n}$. Nechť $\mathbf{A}\mathbf{A}^T = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ je spektrální rozklad matice $\mathbf{A}\mathbf{A}^T$, kde $\mathbf{V} = \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3\end{bmatrix} \in \mathbb{R}^{3\times 3}$ a $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$. Přímka procházející počátkem (tedy $k=1$), která minimalizuje součet čtverců vzdáleností od bodů, je množina

$$\operatorname{span}\{\mathbf{v}_1\} = (\operatorname{span}\{\mathbf{v}_2, \mathbf{v}_3\})^\perp = \{\, \mathbf{x} \in \mathbb{R}^3 \mid \mathbf{v}_2^T\mathbf{x} = \mathbf{v}_3^T\mathbf{x} = 0\,\}.$$

Číslo $\lambda_2 + \lambda_3$ říká, do jaké míry body neleží v (nějaké) přímce procházející počátkem.

Rovina procházející počátkem (tedy $k=2$), která minimalizuje součet čtverců vzdáleností od bodů, je množina

$$\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\} = (\operatorname{span}\{\mathbf{v}_3\})^\perp = \{\, \mathbf{x} \in \mathbb{R}^3 \mid \mathbf{v}_3^T\mathbf{x} = 0\,\}.$$

Číslo λ_3 říká, do jaké míry body neleží v (nějaké) rovině procházející počátkem. ♦

[4] Snad je jasné, že $\mathbf{A}$ v §(7.11) a (7.8) je jiná matice než v §7.1. Podobně, $\mathbf{Y}$ v úloze (7.8) označuje jinou matici než $\mathbf{Y}$ v důkazu Věty 7.2.

7.2.1 Jiný pohled: nejbližší matice nižší hodnosti

Úlohu na optimální proložení bodů podprostorem lze formulovat i jiným, ekvivalentním způsobem (viz obrázek výše): hledáme body $\mathbf{b}_1, \ldots, \mathbf{b}_n \in \mathbb{R}^m$, které leží v (nějakém neznámém) podprostoru dimenze k a minimalizují součet čtverců vzdáleností

$$\|\mathbf{a}_1 - \mathbf{b}_1\|^2 + \cdots + \|\mathbf{a}_n - \mathbf{b}_n\|^2 = \|\mathbf{A} - \mathbf{B}\|^2 \tag{7.13}$$

k daným bodům $\mathbf{a}_1, \ldots, \mathbf{a}_n$. Výraz (7.13) jsme napsali i v maticové formě, kde $\mathbf{B} = \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_n \end{bmatrix} \in \mathbb{R}^{m \times n}$ je matice se sloupci $\mathbf{b}_1, \ldots, \mathbf{b}_n$. Body $\mathbf{b}_1, \ldots, \mathbf{b}_n$ leží v podprostoru dimenze k, právě když[5]

$$\dim \operatorname{span}\{\mathbf{b}_1, \ldots, \mathbf{b}_n\} = \dim \operatorname{rng} \mathbf{B} = \operatorname{rank} \mathbf{B} \leq k. \tag{7.14}$$

Úlohu lze tedy napsat jako[6]

$$\min\{\, \|\mathbf{A} - \mathbf{B}\|^2 \mid \mathbf{B} \in \mathbb{R}^{m \times n},\ \operatorname{rank} \mathbf{B} \leq k \;\}. \tag{7.15}$$

To ale znamená, že k dané matici $\mathbf{A}$ hledáme nejbližší (ve smyslu Frobeniovy normy) matici $\mathbf{B}$ hodnosti nejvýše k. Tato úloha je angl. známa jako *low rank approximation*.

Tvrzení 7.4. Úlohy (7.15) a (7.8) mají stejnou optimální hodnotu.
Optimální řešení úlohy (7.15) je $\mathbf{B} = \mathbf{X}\mathbf{X}^T\mathbf{A}$ kde $\mathbf{X}$ je matice z Tvrzení 7.3.

Důkaz. Důkaz snadno plyne z úvahy výše. Ať už je podprostor, ve kterém body $\mathbf{b}_j$ leží, jakýkoliv, tak v optimu budou body $\mathbf{b}_j$ určitě ortogonálními projekcemi bodů $\mathbf{a}_j$ na tento podprostor. Jinak by totiž výraz (7.13) šel zmenšit posunutím některého bodu v tomto podprostoru. Tedy řešení úlohy (7.15) najdeme tak, že nejprve najdeme optimální podprostor $\operatorname{rng} \mathbf{X}$ dle Tvrzení 7.3 a pak na něj promítneme body $\mathbf{a}_j$.

Ortogonální projektor na tento optimální podprostor je $\mathbf{X}\mathbf{X}^T$ a na jeho ortogonální doplněk $\mathbf{Y}\mathbf{Y}^T = \mathbf{I} - \mathbf{X}\mathbf{X}^T$ (viz Tvrzení 4.7). Vzdálenost bodu $\mathbf{a}_j$ od optimálního podprostoru je $\|\mathbf{a}_j - \mathbf{b}_j\| = \|\mathbf{a}_j - \mathbf{X}\mathbf{X}^T\mathbf{a}_j\| = \|(\mathbf{I} - \mathbf{X}\mathbf{X}^T)\mathbf{a}_j\| = \|\mathbf{Y}\mathbf{Y}^T\mathbf{a}_j\| = \|\mathbf{Y}^T\mathbf{a}_j\|$. Sečtením přes všechna j dostaneme účelové funkce úloh (7.15) a (7.8). ■

Projekci $\mathbf{b} = \mathbf{X}\mathbf{X}^T\mathbf{a}$ bodu $\mathbf{a}$ na optimální podprostor můžeme psát jako $\mathbf{b} = \mathbf{X}\mathbf{c}$, kde $\mathbf{c} = \mathbf{X}^T\mathbf{a} \in \mathbb{R}^k$ jsou souřadnice bodu $\mathbf{b}$ v ortonormální bázi tvořené sloupci matice $\mathbf{X}$. Tedy $\mathbf{B} = \mathbf{X}\mathbf{X}^T\mathbf{A}$ lze psát jako $\mathbf{B} = \mathbf{X}\mathbf{C}$, kde sloupce $\mathbf{c}_1, \ldots, \mathbf{c}_n$ matice $\mathbf{C} = \mathbf{X}^T\mathbf{A} \in \mathbb{R}^{k \times n}$ jsou souřadnice bodů $\mathbf{b}_1, \ldots, \mathbf{b}_n$ v této bázi. Pro $n \gg m$ řešení úlohy (7.15) můžeme vidět jako kompresi dat uložených v matici $\mathbf{A}$: body $\mathbf{a}_1, \ldots, \mathbf{a}_n$ chceme nahradit body $\mathbf{c}_1, \ldots, \mathbf{c}_n$ (které zaberou v paměti daleko méně čísel, přičemž velikost báze $\mathbf{X}$ je zanedbatelná) tak, aby chyba aproximace (7.13) byla co nejmenší. Srovnej s Větou 3.6 o rank factorization!

7.2.2 Co když prokládáme afinním podprostorem?

Změňme nyní úlohu tak, že místo (lineárního) podprostoru hledáme *afinní* podprostor dimenze k, který minimalizuje součet čtverců vzdáleností od bodů $\mathbf{a}_1, \ldots, \mathbf{a}_n$.

[5] V (7.14) je $\operatorname{rank} \mathbf{B} \leq k$ a ne $\operatorname{rank} \mathbf{B} = k$ (jak byste asi čekali), protože když množina bodů leží v podprostoru nějaké dimenze, pak leží i v (nějakém) podprostoru libovolné větší dimenze (např. když množina bodů v $\mathbb{R}^3$ leží na přímce, tak leží také v rovině).

[6] Pozor, v úloze (7.15) je neznámá matice značena jako $\mathbf{B}$, na což asi nejste zvyklí.

Tvrzení 7.5. Afinní podprostor dimenze $k \leq m$ minimalizující součet čtverců vzdáleností od bodů $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^m$ prochází jejich těžištěm $\bar{\mathbf{a}} = \frac{1}{n}(\mathbf{a}_1 + \dots + \mathbf{a}_n)$.

Důkaz. Hledaný afinní podprostor parametrizujme jako (viz Věta 3.13)

$$\{ \mathbf{b} \in \mathbb{R}^m \mid \mathbf{Y}^T \mathbf{b} = \mathbf{y} \}, \tag{7.16}$$

kde[7] $\mathbf{Y} \in \mathbb{R}^{m \times (m-k)}$ je matice s ortonormálními sloupci a $\mathbf{y} \in \mathbb{R}^n$. Vzdálenost bodu $\mathbf{a} \in \mathbb{R}^m$ od tohoto podprostoru je $\|\mathbf{Y}^T \mathbf{a} - \mathbf{y}\|$ (viz §5.2.1). Součet čtverců vzdáleností všech bodů od podprostoru je

$$\|\mathbf{Y}^T \mathbf{a}_1 - \mathbf{y}\|^2 + \dots + \|\mathbf{Y}^T \mathbf{a}_n - \mathbf{y}\|^2.$$

Tento výraz tedy musíme minimalizovat přes proměnné $\mathbf{Y}$ a $\mathbf{y}$ za podmínky $\mathbf{Y}^T \mathbf{Y} = \mathbf{I}$. Pokud je $\mathbf{Y}$ pevné a minimalizujeme pouze přes $\mathbf{y}$, minimum se nabývá v bodě $\mathbf{y} = \mathbf{Y}^T \bar{\mathbf{a}}$ (viz Příklad 1.12). Z toho plyne (proč?), že podprostor (7.16) obsahuje bod $\bar{\mathbf{a}}$. ∎

Nyní je řešení jasné: nejprve posuneme body $\mathbf{a}_1, \dots, \mathbf{a}_n$ tak, aby jejich těžiště leželo v počátku, a potom najdeme (lineární) podprostor minimalizující součet čtverců vzdáleností od posunutých bodů dle Tvrzení 7.3. Jeho posunutím zpět pak získáme kýžený afinní podprostor.

Úloha na proložení bodů afinním podprostorem dimenze k je statistice známá jako *rozvoj podle hlavních komponent* (angl. *principal component analysis, PCA)* nebo *Karhunen-Loewův rozvoj*. Podotkněme, že úlohu nelze nijak převést na přibližné řešení soustavy lineárních rovnic ve smyslu nejmenších čtverců, tedy úlohu (5.2). Za předpokladu $\bar{\mathbf{a}} = \mathbf{0}$ je ve statistice matice $\frac{1}{n}\mathbf{A}\mathbf{A}^T = \frac{1}{n}(\mathbf{a}_1\mathbf{a}_1^T + \dots + \mathbf{a}_n\mathbf{a}_n^T)$ interpretována jako *empirická kovarianční matice* vzorku $\mathbf{a}_1, \dots, \mathbf{a}_n$ m-rozměrné náhodné veličiny (viz poznámka u Věty 5.1).

7.3 Přeurčené homogenní lineární soustavy

Řešme homogenní soustavu lineárních rovnic

$$\mathbf{A}\mathbf{x} = \mathbf{0}, \tag{7.17}$$

kde $\mathbf{A} \in \mathbb{R}^{m \times n}$. Její množinou řešení je $\operatorname{null} \mathbf{A}$, což je lineární podprostor $\mathbb{R}^n$ dimenze $n - \operatorname{rank} \mathbf{A}$ viz (3.20). Může být homogenní soustava 'přeurčená'? Přeurčenost není rozumné definovat jako u nehomogenní soustavy (viz §5.1), protože homogenní soustava má vždy triviální řešení $\mathbf{x} = \mathbf{0}$. Kdybychom tedy 'přeurčenou' soustavu (7.17) zkusili přibližně řešit jako minimalizaci $\|\mathbf{A}\mathbf{x}\|$, dostali bychom triviální optimální řešení $\mathbf{x} = \mathbf{0}$.

Abychom se tomu vyhnuli, můžeme navíc požadovat $\mathbf{x}^T\mathbf{x} = 1$. To vede na úlohu

$$\min\{ \|\mathbf{A}\mathbf{x}\| \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{x}^T\mathbf{x} = 1 \}. \tag{7.18}$$

Protože $\|\mathbf{A}\mathbf{x}\|^2 = \mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x}$, řešením této úlohy vlastní vektor matice $\mathbf{A}^T\mathbf{A}$ příslušný jejímu nejmenšímu vlastnímu číslu.

Příklad 7.2 (přibližné proložení bodů kuželosečkou). Kuželosečka (viz §6.4.2) je množina

$$K = \{ (x, y) \in \mathbb{R}^2 \mid Q(x, y) = a_1x^2 + a_2xy + a_3y^2 + a_4x + a_5y + a_6 = 0 \}.$$

[7] Porovnejte množinu (7.16) s množinou $\operatorname{rng} \mathbf{X} = (\operatorname{rng} \mathbf{Y})^\perp = \operatorname{null}(\mathbf{Y}^T) = \{ \mathbf{b} \in \mathbb{R}^m \mid \mathbf{Y}^T\mathbf{b} = \mathbf{0} \}$.

Parametry $\mathbf{a} = (a_1, \ldots, a_6) \in \mathbb{R}^6$ určují kuželosečku. Všimněte si, že ne pro každý vektor $\mathbf{a}$ je množina K kuželosečka: pro $\mathbf{a} = \mathbf{0}$ bude $K = \mathbb{R}^2$. Dále, vynásobíme-li vektor $\mathbf{a}$ nenulovým skalárem, množina k se nezmění.

Je dáno m bodů $(x_1, y_1), \ldots, (x_m, y_m) \in \mathbb{R}^2$, o kterých víme, že mají ležet na kuželosečce. Body jsou ale zatíženy šumem, tedy obecně nemusí existovat kuželosečka, která jimi prochází. Hledejme tedy kuželosečku, která je 'nejbližší' daným bodům. Jedna možná formulace této úlohy je, že hledáme parametry $\mathbf{a}$, které minimalizují součet čtverců vzdáleností bodů od křivky. Vyřešit přesně tuto úlohu je ale obtížné.

Formulujme proto úlohu přibližně: hledejme vektor $\mathbf{a}$, který minimalizují součet

$$Q(x_1, y_1)^2 + \cdots + Q(x_m, y_m)^2. \tag{7.19}$$

Tato formulace ale nevyjadřuje to, co chceme, protože minimum výrazu (7.19) se nabývá pro $\mathbf{a} = \mathbf{0}$. V tomto případě množina K není křivka, ale celá rovina $\mathbb{R}^2$, což rozhodně nechceme. Ve skutečnosti navíc potřebujeme, aby aspoň jedno z čísel $a_1, \ldots, a_6$ bylo nenulové. Toho dosáhneme uvalením podmínky

$$\mathbf{a}^T\mathbf{a} = a_1^2 + \cdots + a_6^2 = 1. \tag{7.20}$$

Všimněte si, že pro každou kuželosečku můžeme dosáhnout splnění této podmínky, protože vynásobení vektoru $\mathbf{a}$ libovolným nenulovým číslem nezmění množinu K.

Minimalizace (7.19) za podmínky (7.20) se dá psát jako minimalizace $\|\mathbf{Ba}\|^2$ za podmínky $\mathbf{a}^T\mathbf{a} = 1$, kde

$$\mathbf{B} = \begin{bmatrix} x_1^2 & x_1y_1 & y_1^2 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_m^2 & x_my_m & y_m^2 & x_m & y_m & 1 \end{bmatrix}.$$

♦

Obecněji, soustavu (7.17) nazveme přeurčenou tehdy, když dimenze jejího prostoru řešení je nižší než nějaká předem známá dimenze $n-k$. Přibližné řešení soustavy pak vede na úlohu (7.8), která je ovšem ekvivalentní úloze (7.15). Úloha (7.15) odpovídá tomu, že co nejméně změníme matici $\mathbf{A}$, aby prostor řešení soustavy (7.17) měl kýženou dimenzi $n-k$. Neboli nejprve najdeme matici $\mathbf{B}$ s hodností k nejbližší matici $\mathbf{A}$, a potom řešíme soustavu $\mathbf{Bx} = \mathbf{0}$, jejíž prostor řešení již má dimenzi $n - k$.

Poznámka: vztah k nehomogennímu případu. V §5.1 jsme formulovali přibližné řešení nehomogenní ($\mathbf{b} \neq \mathbf{0}$) soustavy $\mathbf{Ax} = \mathbf{b}$ jako úlohu $\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{Ax} - \mathbf{b}\|$. Může se zdát, že tato formulace je úplně odlišná od formulace přibližného řešení homogenní soustavy, kterou jsme uvedli zde. Ale tak tomu není. Formulujme přibližné řešení nehomogenní soustavy takto: pokud soustava $\mathbf{Ax} = \mathbf{b}$ nemá řešení, změňme vektor $\mathbf{b}$ co nejméně tak, aby soustava řešení měla. Přesněji, hledáme vektor $\mathbf{c}$ tak, aby pro nějaké $\mathbf{x}$ platilo $\mathbf{Ax} = \mathbf{c}$ a přitom číslo $\|\mathbf{b} - \mathbf{c}\|$ bylo co nejmenší. Tuto úlohu lze napsat jako

$$\min\{\, \|\mathbf{b} - \mathbf{c}\| \mid \mathbf{Ax} = \mathbf{c},\ \mathbf{x} \in \mathbb{R}^n,\ \mathbf{c} \in \mathbb{R}^m \,\}.$$

Zde minimalizujeme přes proměnné $\mathbf{x}$ a $\mathbf{c}$ (nevadí, že $\mathbf{x}$ se nevyskytuje v účelové funkci). Ale tato úloha jde zjednodušit: dosadíme $\mathbf{c} = \mathbf{Ax}$ do účelové funkce $\|\mathbf{b} - \mathbf{c}\|$, čímž dostaneme $\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{Ax} - \mathbf{b}\|$. Shrňme:

- V přibližném řešení nehomogenní soustavy $\mathbf{Ax} = \mathbf{b}$ chceme změnit vektor $\mathbf{b}$ co nejméně tak, aby soustava měla řešení.

- V přibližném řešení homogenní soustavy $\mathbf{Ax} = \mathbf{0}$ chceme změnit matici $\mathbf{A}$ co nejméně tak, aby soustava měla prostor řešení dané dimenze.

7.4 Singulární rozklad (SVD)

Nyní představíme další důležitý maticový rozklad (faktorizaci), jehož existence sice snadno plyne ze spektrálního rozkladu symetrické matice, ale dovoluje některá tvrzení formulovat elegantněji a existují na něj efektivní numerické algoritmy.

Věta 7.6. Každou matici $\mathbf{A} \in \mathbb{R}^{m\times n}$ lze rozložit jako

$$\mathbf{A} = \mathbf{USV}^T = s_1\mathbf{u}_1\mathbf{v}_1^T + \cdots + s_p\mathbf{u}_p\mathbf{v}_p^T \tag{7.21}$$

kde $p = \min\{m, n\}$, matice $\mathbf{S} = \operatorname{diag}(s_1, \ldots, s_p) \in \mathbb{R}^{p\times p}$ je diagonální a matice $\mathbf{U} = \begin{bmatrix}\mathbf{u}_1 & \cdots & \mathbf{u}_p\end{bmatrix} \in \mathbb{R}^{m\times p}$ a $\mathbf{V} = \begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_p\end{bmatrix} \in \mathbb{R}^{n\times p}$ mají ortonormální sloupce.

Rozklad (7.21) se nazývá **singulární rozklad** (*singular value decomposition, SVD*) matice $\mathbf{A}$. Napsali jsme ho v maticové formě i jako součet dyád (srov. (6.16)). Diagonální prvky $s_1, \ldots, s_p$ matice $\mathbf{S}$ se nazývají **singulární čísla** matice $\mathbf{A}$. Singulární čísla je zvykem (i Matlab to tak dělá) volit nezáporná a sestupně je seřadit (podobně jako vlastní čísla symetrické matice),

$$s_1 \geq \cdots \geq s_p \geq 0, \tag{7.22}$$

což lze vždy zajistit případným vynásobením čísla s_i a jednoho z vektorů $\mathbf{u}_i, \mathbf{v}_i$ minus jedničkou a permutací sloupců $\mathbf{U}, \mathbf{V}$ a diagonálních prvků $\mathbf{S}$. Sloupce $\mathbf{u}_1, \ldots, \mathbf{u}_p$ matice $\mathbf{U}$ se nazývají levé **singulární vektory** a sloupce $\mathbf{v}_1, \ldots, \mathbf{v}_p$ matice $\mathbf{V}$ pravé singulární vektory matice $\mathbf{A}$.

Různé verze SVD. Součin $\mathbf{USV}$ se nezmění, přidáme-li k matici $\mathbf{U}$ libovolné sloupce a k matici $\mathbf{S}$ stejný počet nulových řádků, protože přidané sloupce matice $\mathbf{U}$ budou v součinu násobeny nulami (nakreslete si matice a rozmyslete!). Podobně, je-li několik posledních řádků matice $\mathbf{S}$ nulových (což se stane, když jsou některá singulární čísla nulová), můžeme tyto řádky vynechat a vynechat odpovídajicí sloupce matice $\mathbf{U}$. Totéž lze dělat se sloupci matice $\mathbf{V}$ a sloupci matice $\mathbf{S}$. Díky tomu se rozlišují různé verze SVD, ve kterých vždy platí $\mathbf{A} = \mathbf{USV}^T$ kde $\mathbf{S}$ je diagonální[8] a $\mathbf{U}, \mathbf{V}$ mají ortonormální sloupce, jen rozměry matic se liší:

- **Plné SVD** má $\mathbf{S} \in \mathbb{R}^{m\times n}$, $\mathbf{U} \in \mathbb{R}^{m\times n}$ a $\mathbf{V} \in \mathbb{R}^{n\times n}$. Je-li $\mathbf{A}$ široká příp. úzká, získáme ho doplněním matice $\mathbf{U}$ příp. $\mathbf{V}$ přidáním sloupců na ortogonální matici a přidáním odpovídajícího počtu nulových řádků příp. sloupců matice $\mathbf{S}$.
- **Redukované SVD** má $\mathbf{S} \in \mathbb{R}^{p\times p}$, $\mathbf{U} \in \mathbb{R}^{m\times p}$ a $\mathbf{V} \in \mathbb{R}^{n\times p}$, tj. jako ve Větě 7.6.
- **Rank-minimální SVD**[9] má $\mathbf{S} \in \mathbb{R}^{r\times r}$, $\mathbf{U} \in \mathbb{R}^{m\times r}$ a $\mathbf{V} \in \mathbb{R}^{n\times r}$ kde $r = \operatorname{rank}\mathbf{A}$. Platí totiž $\operatorname{rank}\mathbf{A} = \operatorname{rank}\mathbf{S}$ (plyne z Tvrzení 3.10), tedy hodnost matice je rovna počtu jejích nenulových singulárních čísel. Pokud $r < p$, pak tedy posledních $p - r$ diagonálních prvků $\mathbf{S}$ je nulových ($s_{r+1} = \cdots = s_p = 0$) a můžeme vynechat posledních posledních $p - r$ řádků+sloupců matice $\mathbf{S}$ a posledních $p - r$ sloupců matic $\mathbf{U}$ a $\mathbf{V}$.

[8] Připomeňme, že diagonální matice nemusí být čtvercová (viz §2), diagonální matice rozměru $m \times n$ má $p = \min\{m, n\}$ diagonálních prvků.

[9] Tato nejmenší verze SVD nemá ustálené jméno, někdy se jí také říká *kompaktní SVD*.

Přechod od plného k rank-minimálnímu SVD lze napsat jako

$$\mathbf{A} = \begin{bmatrix}\mathbf{U} & \mathbf{U}'\end{bmatrix}\begin{bmatrix}\mathbf{S} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{V}^T\\ \mathbf{V}'^T\end{bmatrix} = \mathbf{U}\mathbf{S}\mathbf{V}^T = s_1\mathbf{u}_1\mathbf{v}_1^T + \cdots + s_r\mathbf{u}_r\mathbf{v}_r^T \tag{7.23}$$

kde $\mathbf{U} \in \mathbb{R}^{m\times r}$, $\mathbf{U}' \in \mathbb{R}^{m\times(m-r)}$, $\mathbf{S} \in \mathbb{R}^{r\times r}$, $\mathbf{V} \in \mathbb{R}^{n\times r}$, $\mathbf{V}' \in \mathbb{R}^{n\times(n-r)}$ a rozměry nulových matic plynou z rozměrů ostatních matic (některé mohou být i 'prázdné').

Uvědomte si, že různé verze SVD má smysl rozlišovat jen v maticové formě SVD. Napíšeme-li SVD jako sumu dyád v (7.21), rozdíl mezi verzemi se stane triviální: sčítance $s_i\mathbf{u}_i\mathbf{v}_i^T$ pro $i > p$ v této sumě nejsou vůbec, a pro $r < p$ je posledních $p - r$ sčítanců nulových (viz (7.23)).

Příklad 7.3. Příklad plného a redukovaného (zde i rank-minimálního) SVD matice 2×3:

$$\begin{aligned}\begin{bmatrix}3 & 2 & 2\\ 2 & 3 & -2\end{bmatrix} &= \begin{bmatrix}1/\sqrt{2} & 1/\sqrt{2}\\ 1/\sqrt{2} & -1/\sqrt{2}\end{bmatrix}\begin{bmatrix}5 & 0 & 0\\ 0 & 3 & 0\end{bmatrix}\begin{bmatrix}1/\sqrt{2} & 1/\sqrt{2} & 0\\ 1/\sqrt{18} & -1/\sqrt{18} & 4/\sqrt{18}\\ 2/3 & -2/3 & 1/3\end{bmatrix}\\ &= \begin{bmatrix}1/\sqrt{2} & 1/\sqrt{2}\\ 1/\sqrt{2} & -1/\sqrt{2}\end{bmatrix}\begin{bmatrix}5 & 0\\ 0 & 3\end{bmatrix}\begin{bmatrix}1/\sqrt{2} & 1/\sqrt{2} & 0\\ 1/\sqrt{18} & -1/\sqrt{18} & 4/\sqrt{18}\end{bmatrix}\end{aligned}$$

♦

7.4.1 SVD ze spektrálního rozkladu

Singulární rozklad matice $\mathbf{A} \in \mathbb{R}^{m\times n}$ má vztah ke spektrálním rozkladům symetrických matic

$$\mathbf{A}\mathbf{A}^T \in \mathbb{R}^{m\times m}, \tag{7.24a}$$

$$\mathbf{A}^T\mathbf{A} \in \mathbb{R}^{n\times n}, \tag{7.24b}$$

$$\begin{bmatrix}\mathbf{0} & \mathbf{A}^T\\ \mathbf{A} & \mathbf{0}\end{bmatrix} \in \mathbb{R}^{(m+n)\times(m+n)}. \tag{7.24c}$$

Konkrétně, nenulová singulární čísla matice $\mathbf{A}$ jsou odmocniny z nenulových vlastních čísel matic (7.24a) a (7.24b) a zároveň jsou to kladná vlastní čísla matice (7.24c) (zkuste si to v Matlabu na náhodných maticích $\mathbf{A}$!). Singulární vektory lze pak také snadno spočítat z vlastních vektorů těchto matic.

Je snadné to ukázat pro matici (7.24b) (pro matici (7.24a) to je analogické). Bez ztráty obecnosti stačí uvažovat rank-minimální SVD, kdy matice $\mathbf{S}$ má všechny diagonální prvky kladné. Je-li $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$, pak

$$\mathbf{A}^T\mathbf{A} = \mathbf{V}\mathbf{S}^T\mathbf{U}^T\mathbf{U}\mathbf{S}\mathbf{V}^T = \mathbf{V}\mathbf{S}^T\mathbf{S}\mathbf{V}^T = \mathbf{V}\mathbf{S}^2\mathbf{V}^T. \tag{7.25}$$

Ale to je rank-minimální spektrální rozklad (viz §6.3.2) symetrické positivně semidefinitní (viz Cvičení 6.20) matice $\mathbf{A}^T\mathbf{A}$. Vidíme, že nenulová vlastní čísla matice $\mathbf{A}^T\mathbf{A}$ jsou diagonální prvky matice $\mathbf{S}^2$, tedy druhé mocniny nenulových singulárních čísel matice $\mathbf{A}$. Lze navíc dokázat (ve Cvičení 7.20), že získáme-li matice $\mathbf{V}$ a $\mathbf{S}$ z rozkladu (7.25) a položíme-li $\mathbf{U} = \mathbf{A}\mathbf{V}\mathbf{S}^{-1}$ (tj. $\mathbf{u}_i = \mathbf{A}\mathbf{v}_i/s_i$) pak takto získané matice tvoří SVD matice $\mathbf{A}$, tedy platí $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ a $\mathbf{U}\mathbf{S}\mathbf{V}^T = \mathbf{A}$.

Důkaz pro matici (7.24c) je předmětem Cvičení 7.21.

Souvislost SVD a spektrálního rozkladu matice (7.24a) příp. (7.24b) je hlavně teoretický výsledek a počítat takto SVD není numericky vhodné, protože výpočet součinu $\mathbf{A}\mathbf{A}^T$ příp. $\mathbf{A}^T\mathbf{A}$ vede ke zbytečným zaokrouhlovacím chybám (jak jsme ukázali v §5.1.1). Praktické algoritmy na SVD se proto buď počítají pomocí spektrálního rozkladu matice (7.24c) nebo používají úplně jiné metody. Tedy kdykoliv chcete počítat spektrální rozklad matice $\mathbf{A}\mathbf{A}^T$ nebo $\mathbf{A}^T\mathbf{A}$, *je numericky vhodnější použít algoritmus na SVD matice* $\mathbf{A}$. V Matlabu příkaz `[U,S,V]=svd(A)` počítá plné SVD a `[U,S,V]=svd(A,'econ')` počítá redukované SVD.

7.4.2 Nejbližší matice nižší hodnosti z SVD

V §7.2.1 jsme odvodili řešení úlohy (7.15) pomocí spektrálního rozkladu matice $\mathbf{A}\mathbf{A}^T$. Následující klasická věta formuluje toto řešení elegantně pomocí SVD matice $\mathbf{A}$.

Věta 7.7 (Eckart-Young). Nechť $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$ je SVD matice $\mathbf{A}$ a $k \leq p = \min\{m, n\}$. Řešení úlohy (7.15) je

$$\mathbf{B} = \mathbf{U}\mathbf{S}_k\mathbf{V}^T = s_1\mathbf{u}_1\mathbf{v}_1^T + \cdots + s_k\mathbf{u}_k\mathbf{v}_k^T \qquad \text{kde} \quad \mathbf{S}_k = \begin{bmatrix}\mathbf{I}_k & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{S} \tag{7.26}$$

(tedy matice $\mathbf{S}_k$ se získá vynulováním $p-k$ nejmenších diagonálních prvků matice $\mathbf{S}$).

Důkaz. Důkaz plyne ze vztahu singulárního a spektrálního rozkladu a z Věty 7.4. Dle §7.4.1 můžeme matici $\mathbf{U}$ ve spektrálním rozkladu $\mathbf{A}\mathbf{A}^T = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$ získat jako matici $\mathbf{U}$ v singulárním rozkladu $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$. Protože jsou vlastní i singulární čísla řazeny sestupně, blok této matice odpovídající k největším vlastním číslům matice $\mathbf{A}\mathbf{A}^T$ je $\mathbf{X} \in \mathbb{R}^{m\times k}$, kde $\mathbf{U} = \begin{bmatrix}\mathbf{X} & \mathbf{Y}\end{bmatrix}$. Dle Věty 7.4 je tedy optimální řešení úlohy (7.15) rovno $\mathbf{B} = \mathbf{X}\mathbf{X}^T\mathbf{A}$, což rozepíšeme jako $\mathbf{B} = \mathbf{X}\mathbf{C}$ a $\mathbf{C} = \mathbf{X}^T\mathbf{A}$. Zbytek je cvičení na úpravy maticových výrazů:

$$\begin{aligned}
\mathbf{C} &= \mathbf{X}^T\mathbf{A} = \mathbf{X}^T\mathbf{U}\mathbf{S}\mathbf{V}^T = \mathbf{X}^T\begin{bmatrix}\mathbf{X} & \mathbf{Y}\end{bmatrix}\mathbf{S}\mathbf{V}^T = \begin{bmatrix}\mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Y}\end{bmatrix}\mathbf{S}\mathbf{V}^T = \begin{bmatrix}\mathbf{I}_k & \mathbf{0}\end{bmatrix}\mathbf{S}\mathbf{V}^T\\
\mathbf{B} &= \mathbf{X}\mathbf{C} = \mathbf{X}\begin{bmatrix}\mathbf{I}_k & \mathbf{0}\end{bmatrix}\mathbf{S}\mathbf{V}^T = \mathbf{X}\begin{bmatrix}\mathbf{X} & \mathbf{Y}\end{bmatrix}\begin{bmatrix}\mathbf{I}_k & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{S}\mathbf{V}^T = \mathbf{U}\begin{bmatrix}\mathbf{I}_k & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{S}\mathbf{V}^T = \mathbf{U}\mathbf{S}_k\mathbf{V}^T.
\end{aligned}$$

Rozměry nulových matic se liší podle toho, zda je SVD plné nebo redukované (věta platí pro obě verze SVD). ∎

Všimněte si, že suma dyád v (7.26) je vlastně zkrácená suma v (7.21).

7.4.3 Extrémy lineární funkce isometrie

Zde ukážeme, jak optimalizovat lineární funkci matice s ortonormálními sloupci (reprezentující tedy isometrii). Lineárni funkce matice $\mathbf{X} \in \mathbb{R}^{m\times n}$ má obecný tvar $f(\mathbf{X}) = \langle \mathbf{A}, \mathbf{X}\rangle = \sum_{i,j} a_{ij}x_{ij}$ (viz §2.1.7), kde $\mathbf{A} \in \mathbb{R}^{m\times n}$. Tato úloha (7.27) je jednodušší než (7.3), je to nejjednodušší smysluplná úloha s omezením na ortonormalitu (protože co je jednoduššího než lineární funkce?).

Věta 7.8. Nechť $m \geq n$ a $\mathbf{A} \in \mathbb{R}^{m\times n}$. Úloha

$$\max\{\, \langle \mathbf{A}, \mathbf{X}\rangle \mid \mathbf{X} \in \mathbb{R}^{m\times n},\ \mathbf{X}^T\mathbf{X} = \mathbf{I}\,\} \tag{7.27}$$

má optimální řešení $\mathbf{X} = \mathbf{U}\mathbf{V}^T$, kde $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$ je redukované SVD matice $\mathbf{A}$.

Důkaz. Uvažujme nejdříve *plné* SVD $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$. Platí $\langle \mathbf{A}, \mathbf{X} \rangle = \langle \mathbf{U}\mathbf{S}\mathbf{V}^T, \mathbf{X} \rangle = \langle \mathbf{S}, \mathbf{U}^T\mathbf{X}\mathbf{V} \rangle$ (viz §2.1.7). Zavedeme substituci $\mathbf{Y} = \mathbf{U}^T\mathbf{X}\mathbf{V}$. Protože $\mathbf{U}$ a $\mathbf{V}$ jsou ortogonální, $\mathbf{X}^T\mathbf{X} = \mathbf{I}$ platí právě když $\mathbf{Y}^T\mathbf{Y} = \mathbf{I}$. Tedy úloha (7.27) je ekvivalentní úloze

$$\max\{\, \langle \mathbf{S}, \mathbf{Y} \rangle \mid \mathbf{Y} \in \mathbb{R}^{m\times n},\ \mathbf{Y}^T\mathbf{Y} = \mathbf{I} \,\}. \tag{7.28}$$

Protože $m \geq n$ a $\mathbf{S}$ je diagonální, máme $\langle \mathbf{S}, \mathbf{Y} \rangle = s_1 y_{11} + \cdots + s_n y_{nn}$. Protože $\mathbf{Y}^T\mathbf{Y} = \mathbf{I}$, prvky matice $\mathbf{Y}$ splňují nerovnosti $-1 \leq y_{ij} \leq 1$ (viz Cvičení 4.18). Protože $s_1, \ldots, s_n \geq 0$, výraz $\langle \mathbf{S}, \mathbf{Y} \rangle$ proto nemůže být větší než $s_1 + \cdots + s_n$ a tuto hodnotu nabývá pro $y_{11} = \cdots = y_{nn} = 1$. Existuje právě jedna matice $\mathbf{Y} \in \mathbb{R}^{m\times n}$ splňující $y_{11} = \cdots = y_{nn} = 1$ a $\mathbf{Y}^T\mathbf{Y} = \mathbf{I}$, a to $\mathbf{Y} = \begin{bmatrix} \mathbf{I}_n \\ \mathbf{0} \end{bmatrix}$. To odpovídá $\mathbf{X} = \mathbf{U}\mathbf{Y}\mathbf{V}^T = \mathbf{U}\begin{bmatrix} \mathbf{I}_n \\ \mathbf{0} \end{bmatrix}\mathbf{V}^T$. Ale matice $\mathbf{U}\begin{bmatrix} \mathbf{I}_n \\ \mathbf{0} \end{bmatrix}$ je tvořená prvními n sloupci matice $\mathbf{U}$, tedy $\mathbf{X} = \mathbf{U}\mathbf{V}^T$ kde $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$ je redukované SVD matice $\mathbf{A}$. ■

Nejbližší isometrie

Tento výsledek lze použít pro nalezení matice s ortonormálními sloupci nejbližší dané matici $\mathbf{A} \in \mathbb{R}^{m\times n}$ s $m \geq n$, tedy pro řešení úlohy

$$\min\{\, \|\mathbf{X} - \mathbf{A}\|^2 \mid \mathbf{X} \in \mathbb{R}^{m\times n},\ \mathbf{X}^T\mathbf{X} = \mathbf{I} \,\}. \tag{7.29}$$

Máme

$$\|\mathbf{X} - \mathbf{A}\|^2 = \langle \mathbf{X} - \mathbf{A}, \mathbf{X} - \mathbf{A} \rangle = \langle \mathbf{X}, \mathbf{X} \rangle - \langle \mathbf{X}, \mathbf{A} \rangle - \langle \mathbf{A}, \mathbf{X} \rangle + \langle \mathbf{A}, \mathbf{A} \rangle = \|\mathbf{X}\|^2 - 2\langle \mathbf{A}, \mathbf{X} \rangle + \|\mathbf{A}\|^2.$$

Ale výrazy $\|\mathbf{A}\|^2$ a $\|\mathbf{X}\|^2 = \operatorname{tr}(\mathbf{X}^T\mathbf{X}) = \operatorname{tr}\mathbf{I} = n$ nezávisí na $\mathbf{X}$, tedy minimalizace $\|\mathbf{X} - \mathbf{A}\|^2$ je totéž jako maximalizace $\langle \mathbf{A}, \mathbf{X} \rangle$.

Ortogonální Prokrustův problém

Dále lze výsledek použít na úlohu[10], kterou často můžeme potkat např. v robotice. Jsou dány body $\mathbf{a}_1, \ldots, \mathbf{a}_k \in \mathbb{R}^n$ a $\mathbf{b}_1, \ldots, \mathbf{b}_k \in \mathbb{R}^m$, tvořící sloupce matic $\mathbf{A} \in \mathbb{R}^{n\times k}$ a $\mathbf{B} \in \mathbb{R}^{m\times k}$, kde $m \geq n$. Hledáme takovou isometrickou transformaci $\mathbf{X}$ bodů $\mathbf{a}_1, \ldots, \mathbf{a}_k$, aby byly co nejblíže bodům $\mathbf{b}_1, \ldots, \mathbf{b}_k$ ve smyslu nejmenších čtverců, tedy aby výraz

$$\sum_{i=1}^{k} \|\mathbf{X}\mathbf{a}_i - \mathbf{b}_i\|^2 = \|\mathbf{X}\mathbf{A} - \mathbf{B}\|^2$$

byl minimální. Řešíme tedy úlohu

$$\min\{\, \|\mathbf{X}\mathbf{A} - \mathbf{B}\|^2 \mid \mathbf{X} \in \mathbb{R}^{m\times n},\ \mathbf{X}^T\mathbf{X} = \mathbf{I} \,\}. \tag{7.30}$$

Opět máme

$$\|\mathbf{X}\mathbf{A} - \mathbf{B}\|^2 = \langle \mathbf{X}\mathbf{A} - \mathbf{B}, \mathbf{X}\mathbf{A} - \mathbf{B} \rangle = \|\mathbf{X}\mathbf{A}\|^2 - 2\langle \mathbf{X}\mathbf{A}, \mathbf{B} \rangle + \|\mathbf{B}\|^2.$$

Ale výrazy $\|\mathbf{X}\mathbf{A}\|^2 = \|\mathbf{A}\|^2$ (viz (4.16)) a $\|\mathbf{B}\|^2$ jsou konstanty, tedy minimalizace $\|\mathbf{X}\mathbf{A} - \mathbf{B}\|^2$ je totéž jako maximalizace $\langle \mathbf{X}\mathbf{A}, \mathbf{B} \rangle = \langle \mathbf{X}, \mathbf{B}\mathbf{A}^T \rangle = \langle \mathbf{B}\mathbf{A}^T, \mathbf{X} \rangle$. Tedy jsme úlohu převedli na tvar (7.27).

Pokud dovolíme body $\mathbf{a}_1, \ldots, \mathbf{a}_k$ kromě otočení/ozrcadlení také posunout, podobnou úvahou jako v §7.2.2 zjistíme, že optimální posunutí je rozdíl těžišť bodů $\mathbf{b}_1, \ldots, \mathbf{b}_k$ a $\mathbf{a}_1, \ldots, \mathbf{a}_k$.

[10] Procrústés (neboli napínač) byl loupežník z řecké mytologie, který měl nepříjemný zvyk: zbloudilého poutníka donutil ulehnout ve své lesní chatě na postel a když byl nešťastník kratší než postel, natáhl ho na délku postele, a když přečuhoval, tak ho zase od nohou kousek usekl.

7.5 Cvičení

7.1. Jsou dána čísla $c_{ij} \in \mathbb{R}$ pro $i, j = 1, \ldots, n$. Minimalizujte výraz $\sum_i \sum_j c_{ij} x_i x_j$ přes proměnné $x_1, \ldots, x_n \in \mathbb{R}$ za podmínky $\sum_i x_i^2 = 1$.

7.2. Dokázali jsme, že optimální hodnota úlohy 7.1 je největší vlastní číslo matice $\mathbf{A}$. Bude toto tvrzení platit, i když matice $\mathbf{A}$ nebude symetrická?

7.3. Řekli jsme, že maximální hodnota výrazu $\mathbf{y}^T \mathbf{\Lambda} \mathbf{y} = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2$ za podmínky $\mathbf{y}^T \mathbf{y} = y_1^2 + \cdots + y_n^2 = 1$ je λ_1 (kde předpokládáme (6.17)). Dokažte to přesně.

7.4. Pro čtvercovu matici $\mathbf{A}$ je výraz $f(\mathbf{x}) = \dfrac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$ znám jako *Rayleighův kvocient*[11]. Dokažte, že $\max\limits_{\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}} f(\mathbf{x}) = \max\{\, \mathbf{x}^T \mathbf{A} \mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{x}^T \mathbf{x} = 1 \,\}$.

7.5. Přidejme k úloze (7.1) omezení, že $\mathbf{x}$ musí být kolmé na vlastní vektory $\mathbf{v}_1, \ldots, \mathbf{v}_k$. Tedy maximalizujeme $\mathbf{x}^T \mathbf{A} \mathbf{x}$ přes $\mathbf{x} \in \mathbb{R}^n$ za podmínek $\mathbf{x}^T \mathbf{x} = 1$ a $\mathbf{v}_1^T \mathbf{x} = \cdots = \mathbf{v}_k^T \mathbf{x} = 0$. Dokažte, že optimální hodnota této úlohy je λ_{k+1} a optimum se nabývá pro $\mathbf{x} = \mathbf{v}_{k+1}$.

7.6. Jsou-li $\mathbf{x}_1, \ldots, \mathbf{x}_k$ sloupce matice $\mathbf{X}$, dokažte tyto rovnosti (srov. (7.4)):

$$\operatorname{tr}(\mathbf{X}^T \mathbf{A} \mathbf{X}) = \langle \mathbf{A}\mathbf{X}, \mathbf{X} \rangle = \langle \mathbf{A}, \mathbf{X}\mathbf{X}^T \rangle = \mathbf{x}_1^T \mathbf{A} \mathbf{x}_1 + \cdots + \mathbf{x}_k^T \mathbf{A} \mathbf{x}_k.$$

7.7. ($\star$) Můžete namítnout, že Věta 7.2 plyne ihned z výsledku Cvičení 7.5, neboť maximalizaci součtu $\mathbf{x}_1^T \mathbf{A} \mathbf{x}_1 + \cdots + \mathbf{x}_k^T \mathbf{A} \mathbf{x}_k$ za podmínky ortonormality vektorů $\mathbf{x}_1, \ldots, \mathbf{x}_k$ lze provést takto. V prvním kroku maximalizujeme $\mathbf{x}_1^T \mathbf{A} \mathbf{x}_1$ za podmínky $\mathbf{x}_1^T \mathbf{x}_1 = 1$, což má řešení $\mathbf{x}_1 = \mathbf{v}_1$. V druhém kroku $\mathbf{x}_1$ zafixujeme a maximalizujeme $\mathbf{x}_1^T \mathbf{A} \mathbf{x}_1 + \mathbf{x}_2^T \mathbf{A} \mathbf{x}_2$ za podmínky $\mathbf{x}_2^T \mathbf{x}_2 = 1$ a $\mathbf{x}_1^T \mathbf{x}_2 = 0$, což má řešení $\mathbf{x}_2 = \mathbf{v}_2$. Atd.

Takovému typu algoritmu se říká 'hladový' (angl. *greedy*). Obecný hladový algoritmus je popsán takto. Maximalizujeme účelovou funkci $f_1(x_1) + \cdots + f_k(x_k)$ za podmínky $(x_1, \ldots, x_k) \in X$, kde X je daná množina přípustných řešení. Algoritmus v prvním kroku maximalizuje $f_1(x_1)$ za podmínky $(x_1, \ldots, x_k) \in X$. Ve druhém kroku zafixuje proměnnou x_1 a maximalizuje $f_2(x_2)$ za podmínky $(x_1, \ldots, x_k) \in X$. Ve třetím kroku zafixuje x_1, x_2 a maximalizuje $f_3(x_3)$ za podmínky $(x_1, \ldots, x_k) \in X$, atd.

a) Hladový algoritmus obecně nemusí najít globální optimum (i když pro úlohu (7.3) ho najde). Najděte příklad (tedy konkrétní k, X a $f_1, \ldots, f_k$), kdy ho nenajde.
b) Zformulujte podobným způsobem hladový algoritmus pro úlohu (7.27) (můžete ho i naprogramovat). Najde vždy globální minimum?

7.8. Jsou dány čtyři body $\mathbf{a}_1 = (3, -3, 4)$, $\mathbf{a}_2 = (-2, -3, -2)$, $\mathbf{a}_3 = (1, 0, -1)$, $\mathbf{a}_4 = (3, 1, 0)$ v $\mathbb{R}^3$. Najděte množinu $X \subseteq \mathbb{R}^3$, která minimalizuje součet čtverců kolmých vzdáleností bodů od podprostoru X, kde X je

a) přímka procházející počátkem,
b) rovina procházející počátkem,
c) přímka, která nemusí procházet počátkem.

Vždy najděte vektor $\mathbf{x}_0$ a ortonormální bázi lineárního podprostoru X' tak, aby $X = \mathbf{x}_0 + X'$. Dále spočítejte ortogonální projekce bodů na podprostor X (je snad jasné, co myslíme projekcí na afinní podprostor) a souřadnice těchto projekcí v ortonormální bázi podprostoru X'. Použijte (A) spektrální rozklad, (B) SVD. Můžete použít počítač.

[11] 'Kvocient' znamená 'podíl' nebo 'zlomek'.

7.9. Je dán spektrální rozklad dané symetrické matice. Jak byste z něho jednoduše nalezli SVD této matice?

7.10. Pro matici $\mathbf{A} = \frac{1}{15}\begin{bmatrix} -13 & 2 & -22 \\ -16 & 14 & -4 \end{bmatrix}$ najděte matici $\mathbf{B}$ hodnosti jedna takovou, že $\|\mathbf{A}-\mathbf{B}\|$ je minimální. Pak spočtěte $\|\mathbf{A}-\mathbf{B}\|$. Spočtěte pomocí (a) spektrálního rozkladu, (b) SVD. Použijte počítač.

7.11. Hledáme redukované SVD matice $\mathbf{A}$ ze Cvičení 7.10 ručně, bez počítače. Někdo nám napoví, že $\mathbf{U} = \frac{1}{5}\begin{bmatrix} -3 & 4 \\ 4 & 3 \end{bmatrix}$. Jak byste spočetli matice $\mathbf{S}, \mathbf{V}$?

7.12. Máme $n = 100$ bodů $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^{10^6}$. Chceme najít podprostor dimenze k (kde $k < 100$) minimalizující součet čtverců vzdáleností od bodů. Jak to uděláme co nejefektivněji? Nápověda: použijte výsledek Cvičení 6.6.

7.13. Komutuje operace ortogonální projekce s operací těžiště? Tj., když body $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$ ortogonálně promítneme na podprostor $X \subseteq \mathbb{R}^m$ a pak najdeme těžiště výsledných bodů, dostaneme stejný bod, jako když nejdříve spočítáme těžiště bodů $\mathbf{a}_1, \ldots, \mathbf{a}_n$ a pak ho promítneme na podprostor X? Odpověď dokažte.

7.14. Zjednodušme Příklad 7.2 tak, že místo kuželosečky chceme dané body proložit přímkou $\{\, (x, y) \in \mathbb{R}^2 \mid ax + by + c = 0 \,\}$. Děláme to opět tak, že minimalizujeme $\sum_i (ax_i + by_i + c)^2$ za podmínky $a^2 + b^2 + c^2 = 1$.

   a) Minimalizuje formulovaná úloha součet čtverců vzdáleností bodů od přímky, nebo jen jeho aproximaci (jako pro kuželosečku)?
   b) Co když místo omezení $a^2 + b^2 + c^2 = 1$ použijeme omezení $a^2 + b^2 = 1$? Co úloha vyjadřuje a jak ji vyřešíme?
   c) ($\star$) Když v Příkladě 7.2 změníme omezení (7.20) na $a^2 + b^2 + c^2 + d^2 + e^2 = 1$ příp. na $a^2 + b^2 + c^2 = 1$, bude úloha minimalizovat přesný součet čtverců vzdáleností bodů od kuželosečky?

7.15. ($\star$) Dokažte, že $\max\{\, \mathbf{x}^T\mathbf{A}\mathbf{y} \mid \mathbf{x} \in \mathbb{R}^m,\ \mathbf{y} \in \mathbb{R}^n,\ \mathbf{x}^T\mathbf{x} = \mathbf{y}^T\mathbf{y} = 1 \,\} = s_1$, kde s_1 je největší singulární číslo matice $\mathbf{A} \in \mathbb{R}^{m\times n}$, a optimální argument je $(\mathbf{x}, \mathbf{y}) = (\mathbf{u}_1, \mathbf{v}_1)$, tedy levý a pravý singulární vektor příslušný singulárnímu číslu s_1.

7.16. Uvažujme plné SVD matice $\mathbf{A} \in \mathbb{R}^{m\times n}$ hodnosti r ve tvaru (7.23) kde $\mathbf{U} \in \mathbb{R}^{m\times r}$, $\mathbf{U}' \in \mathbb{R}^{m\times(m-r)}$, $\mathbf{V} \in \mathbb{R}^{n\times r}$, $\mathbf{V}' \in \mathbb{R}^{n\times(n-r)}$ a $\mathbf{S} \in \mathbb{R}^{r\times r}$. Dokažte, že sloupce matic $\mathbf{U}, \mathbf{U}', \mathbf{V}, \mathbf{V}'$ tvoří ortonormální báze čtyř základních podprostorů matice $\mathbf{A}$ (viz §4.2.1), tedy

$$\operatorname{rng}\mathbf{U} = \operatorname{rng}\mathbf{A}, \quad \operatorname{rng}\mathbf{U}' = \operatorname{null}(\mathbf{A}^T), \quad \operatorname{rng}\mathbf{V} = \operatorname{rng}(\mathbf{A}^T), \quad \operatorname{rng}\mathbf{V}' = \operatorname{null}\mathbf{A}.$$

7.17. V Matlabu se ortonormální báze nulového prostoru matice najde funkcí `null`. Vypište si implementaci této funkce matlabským příkazem `edit null` a pochopte ji. Najděte souvislost se Cvičením 7.16.

7.18. ($\star$) Dokažte, že pro libovolnou matici $\mathbf{A}$ existuje limita (5.37) a je rovna pseudoinverzi $\mathbf{A}^+$ matice $\mathbf{A}$ (dle Věty 5.4). Postupujte takto:

   a) Dokažte, že $\mathbf{A}^+_\mu = \mathbf{V}\mathbf{S}^+_\mu\mathbf{U}^T$, kde $\mathbf{A}^+_\mu$ je definované vzorcem (5.31), $\mathbf{S}^+_\mu = (\mathbf{S}^2 + \mu\mathbf{I})^{-1}\mathbf{S}$ a $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$ je rank-minimální SVD matice $\mathbf{A}$.
   b) Dokažte, že $\mathbf{S}^+ = \lim\limits_{\mu\to 0^+} \mathbf{S}^+_\mu = \mathbf{S}^{-1}$.

7.19. Do této chvíle jsme potkali již několik rozkladů matic: QR, spektrální rozklad, Choleského rozklad, SVD. Návrh algoritmů na operace s maticemi, řešení soustav lineárních rovnic a rozklady matic je předmětem *numerické lineární algebry*. Jejím cílem je nalézt rychlé algoritmy, které jsou odolné vůči zaokrouhlovacím chybám. Existují volně dostupné softwarové balíky na numerickou lineární algebru. Matlab je postaven na balíku LAPACK. Jiný oblíbený balík je BLAS, který obsahuje funkce nižší úrovně než LAPACK. Najděte na internetu dokumentaci k balíkům LAPACK a BLAS a prostudujte ji!

7.20. ($\star$) Nechť $\mathbf{A}^T\mathbf{A} = \mathbf{V}\mathbf{S}^2\mathbf{V}^T$ je rank-minimální spektrální rozklad matice $\mathbf{A}^T\mathbf{A}$. Nechť $\mathbf{U} = \mathbf{A}\mathbf{V}\mathbf{S}^{-1}$ (tj. $\mathbf{u}_i = \mathbf{A}\mathbf{v}_i/\sigma_i$). Dokažte, že $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ a $\mathbf{U}\mathbf{S}\mathbf{V}^T = \mathbf{A}$.

7.21. ($\star$) Nenulová singulární čísla matice $\mathbf{A} \in \mathbb{R}^{m\times n}$ jsou kladná vlastní čísla symetrické matice $\mathbf{H} = \begin{bmatrix} \mathbf{0} & \mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{(m+n)\times(m+n)}$. Toto tvrzení dokažte ve dvou krocích. Ukažte, že:

a) 'Rank-minimální' spektrální rozklad matice $\mathbf{H}$ lze psát jako $\mathbf{H} = \mathbf{W}\mathbf{\Lambda}\mathbf{W}^T$ kde

$$\mathbf{W} = \tfrac{1}{\sqrt{2}}\begin{bmatrix} \mathbf{V} & \mathbf{V} \\ \mathbf{U} & -\mathbf{U} \end{bmatrix} \in \mathbb{R}^{(m+n)\times(2r)}, \qquad \mathbf{\Lambda} = \begin{bmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & -\mathbf{S} \end{bmatrix} \in \mathbb{R}^{(2r)\times(2r)}$$

kde $r = \operatorname{rank}\mathbf{A}$, $\mathbf{U} \in \mathbb{R}^{m\times r}$ a $\mathbf{V} \in \mathbb{R}^{n\times r}$ splňují $\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I}_r$, a $\mathbf{S} \in \mathbb{R}^{r\times r}$ je diagonální s kladnými prvky.

b) Ze spektrálního rozkladu $\mathbf{H} = \mathbf{W}\mathbf{\Lambda}\mathbf{W}^T$ plyne $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$.

c) Najděte optimální hodnotu úlohy (7.27).

Nápověda a řešení

7.1. To je úloha 7.1 (v minimalizační formě), ale v jiném značení než jste zvyklí. Úloha jde napsat jako $\min\{\, \mathbf{x}^T\mathbf{C}\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{x}^T\mathbf{x} = 1 \,\}$, kde $\mathbf{C} = [c_{ij}]$. Opt. hodnota je nejmenší vlastní číslo symetrické části matice $\mathbf{C}$.

7.2. Samozřejmě nebude, protože zatímco kvadratická forma se nezmění nahrazením matice $\mathbf{A}$ její symetrickou částí $\frac{1}{2}(\mathbf{A}+\mathbf{A}^T)$, vlastní čísla těchto dvou matic budou jiná. Vlastní číslo nesymetrické matice ani nemusí být reálné.

7.3. Jeden způsob důkazu: Máme dokázat, že pro každá $y_1, \ldots, y_n \in \mathbb{R}$ splňující $y_1^2 + \cdots + y_n^2 = 1$ platí $\lambda_1 \geq \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2$. Vyjádříme z podmínky $y_1^2 = 1 - y_2^2 - \cdots - y_n^2$ a dosadíme do kritéria. Teď máme tedy dokázat, že pro každá $y_2, \ldots, y_n \in \mathbb{R}$ platí $\lambda_1 \geq \lambda_1(1 - y_2^2 - \cdots - y_n^2) + \cdots + \lambda_n y_n^2$. To už je snadné.

Druhý způsob důkazu: Pro jednoduchost předpokládáme $\lambda_1 > \lambda_2$ (obecnější případ $\lambda_1 \geq \lambda_2$ si rozmyslete potom sami). Dokážeme, že v optimu musí nutně být $y_1 = 1$ a $y_2 = \cdots = y_n = 0$. Jinými slovy dokažte, že kdyby bylo $y_1 < 1$ (a tedy $y_2^2 + \cdots + y_n^2 > 0$), tak byste mohli přičíst malé číslo $\delta > 0$ k y_1 a odečíst stejné číslo od y_2, čímž by podmínka $y_1^2 + \cdots + y_n^2 = 1$ zůstala v platnosti ale výraz $\lambda_1 y_1^2 + \cdots + \lambda_n y_n^2$ by se zvětšil.

7.4. Klíčové je si všimnout, že $f(\alpha\mathbf{x}) = f(\mathbf{x})$ pro každé $\alpha \neq 0$ a $\mathbf{x} \neq \mathbf{0}$. To nám dovolí dokázat rovnost množin $\{\, f(\mathbf{x}) \mid \mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\} \,\} = \{\, \mathbf{x}^T\mathbf{A}\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{x}^T\mathbf{x} = 1 \,\}$. Každý prvek z pravé množiny zřejmě patří také do levé množiny. Ale také každý prvek z levé množiny patří do pravé množiny, protože když vektor $\mathbf{x}$ vydělíme číslem $\|\mathbf{x}\|$, bude $\mathbf{x}^T\mathbf{x} = 1$ a hodnota $f(\mathbf{x})$ se nezmění.

7.5. Obdobné důkazu Věty 7.1. Spektrálním rozkladem a substitucí $\mathbf{x} = \mathbf{V}\mathbf{y}$ úlohu převedeme na maximalizaci $\lambda_1 y_1^2 + \cdots + \lambda_n y_n^2$ za podmínek $y_1^2 + \cdots + y_n^2 = 1$ a $y_1 = \cdots = y_k = 0$.

7.7.a) $k = 2$, $X = \{\, (x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 = 1 \,\}$ je kružnice, $f_1(x) = f_2(x) = -x$. Globální maximum je $x_1 = x_2 = -1/\sqrt{2}$ (nakreslete si obrázek). Hladový algoritmus ovšem v prvním kroku maximalizuje $-x_1$ za podmínky $x_1^2 + x_2^2 = 1$, což dá $x_1 = -1$. To už zůstane zafixované.

7.8.a) $\mathbf{x}_0 = \mathbf{0}$, ortonorm. báze X' je např. $\{(0.6912, -0.2181, 0.6889)\}$ (zaokrouhleno).

7.8.b) $\mathbf{x}_0 = \mathbf{0}$, ortonorm. báze X' je např. $\{(0.6912, -0.2181, 0.6889), (-0.3771, -0.9221, 0.0864)\}$.

7.8.c) $\mathbf{x}_0 = (1.25, -1.25, 0.25)$, ortonorm. báze X' je např. $\{(0.6599, 0.0280, 0.7508)\}$.

7.10. Na (a) použijte Větu 7.4, tedy nejdříve spočteme matici $\mathbf{Y}$ z Tvrzení 7.3 a pak spočteme $\mathbf{B} = \mathbf{X}\mathbf{X}^T\mathbf{A}$. Vzdálenost $\|\mathbf{A} - \mathbf{B}\|$ teď spočteme dosazením; ovšem dle Věty 7.4 je také rovna odmocnině z výrazu (7.12), tj. $\sqrt{\lambda_1}$ kde λ_1 je nejmenší vlastní číslo matice $\mathbf{A}\mathbf{A}^T$.

Na (b) použijeme Věty 7.7. Vzdálenost $\|\mathbf{A} - \mathbf{B}\|$ můžeme spočítat opět dosazením, bude ale také rovna s_2, tj. nejmenšímu singulárnímu číslu matice $\mathbf{A}$.

7.11. Hledáme diagonální $\mathbf{S} \in \mathbb{R}^{2\times 2}$ a $\mathbf{U} \in \mathbb{R}^{2\times 2}$ s ortonorm. sloupci tak, aby $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$ neboli $\mathbf{U}^T\mathbf{A} = \mathbf{S}\mathbf{V}^T$. Diagonální prvky $\mathbf{S}$ v součinu $\mathbf{S}\mathbf{V}^T$ násobí řádky $\mathbf{V}^T$, tedy řádky matice $\mathbf{S}\mathbf{V}^T$ by měly být ortogonální, což ověříme. Diagonální prvky matice $\mathbf{S}$ pak musíme zvolit tak, aby řádky $\mathbf{V}^T$ byly ortonormální, tedy $\mathbf{S} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \end{bmatrix}$.

7.13. Ano, komutují. Nápověda k důkazu: Označíme-li $\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix} \in \mathbb{R}^{m\times n}$, pak těžiště bodů je bod $\mathbf{A}\mathbf{t}$ kde $\mathbf{t} = \frac{1}{n}\mathbf{1}_n$ a projekce bodů je $\mathbf{P}\mathbf{A}$ kde $\mathbf{P}$ je projektor. Důkaz dokončete sami.

7.14.a) Je to jen aproximace vzdálenosti od přímce. Skutečná vzdálenost bodu (x_i, y_i) od přímky s rovnicí $ax + by + c = 0$ je $|ax_i + by_i + c|$ za předpokladu $a^2 + b^2 = 1$ (viz Cvičení 5.16).

7.14.b) Nyní úloha popisuje hledání přímky, která minimalizuje součet čtverců vzdáleností od daných bodů, viz bod (a). Vyřešíme ji dle §7.2.2.

7.14.c) Nebude, budou to opět jen aproximace. Zkuste zformulovat úlohu na hledání vzdálenosti bodu od kuželosečky. Zjistíte, že je složitější než zde popsané aproximace.

7.15. Dosaďte $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$ ($\mathbf{U}, \mathbf{V}$ ortogonální, $\mathbf{S} \in \mathbb{R}^{m\times n}$). Řešíme $\max\{\, \bar{\mathbf{x}}^T\mathbf{S}\bar{\mathbf{y}} \mid \bar{\mathbf{x}}^T\bar{\mathbf{x}} = \bar{\mathbf{y}}^T\bar{\mathbf{y}} = 1 \,\}$. Máme $\bar{\mathbf{x}}^T\mathbf{S}\bar{\mathbf{y}} = \sum_{i=1}^p s_i \bar{x}_i \bar{y}_i = \sum_{i=1}^p s_i z_i$, kde $z_i = \bar{x}_i\bar{y}_i$. Platí $|z_i| \le 1$ a $\left|\sum_i z_i\right| \le 1$ (rozmyslete). Úloha $\max\{\, \sum_i s_i z_i \mid |z_i| \le 1,\ \left|\sum_i z_i\right| \le 1 \,\}$ má optimální hodnotu s_1 a optimální argument je $z_1 = 1$ a $z_i = 0$ pro $i \ne 1$. Taková čísla $z_i = \bar{x}_i\bar{y}_i$ lze ale realizovat volbou $\bar{\mathbf{x}} = \mathbf{e}_1$ a $\bar{\mathbf{y}} = \mathbf{e}_1$. Optimální argument původní úlohy je $\mathbf{x} = \mathbf{U}\bar{\mathbf{x}} = \mathbf{u}_1$ a $\mathbf{y} = \mathbf{V}\bar{\mathbf{y}} = \mathbf{v}_1$.

7.16. Dokažme $\operatorname{rng}\mathbf{U} = \operatorname{rng}\mathbf{A}$ a $\operatorname{rng}\mathbf{V} = \operatorname{rng}(\mathbf{A}^T)$. První rovnost plyne z Tvrzení 3.10, neboť $\mathbf{V}^T$ a $\mathbf{S}$ mají l.n. řádky a proto $\operatorname{rng}\mathbf{A} = \operatorname{rng}(\mathbf{U}\mathbf{S}\mathbf{V}^T) = \operatorname{rng}(\mathbf{U}\mathbf{S}) = \operatorname{rng}\mathbf{U}$. Druhá rovnost je první rovnost použitá na matici $\mathbf{A}^T$. Zbývající dvě rovnosti plynou z Tvrzení 3.10 a 4.7.

7.20. Platí

$$\mathbf{U}^T\mathbf{U} = \mathbf{S}^{-1}\mathbf{V}^T\mathbf{A}^T\mathbf{A}\mathbf{V}\mathbf{S}^{-1} = \mathbf{S}^{-1}\mathbf{V}^T\mathbf{V}\mathbf{S}^2\mathbf{V}^T\mathbf{V}\mathbf{S}^{-1} = \mathbf{S}^{-1}\mathbf{S}^2\mathbf{S}^{-1} = \mathbf{I}, \tag{7.31a}$$

$$\mathbf{U}\mathbf{S}\mathbf{V}^T = \mathbf{A}\mathbf{V}\mathbf{S}^{-1}\mathbf{S}\mathbf{V}^T = \mathbf{A}\mathbf{V}\mathbf{V}^T = \mathbf{A}. \tag{7.31b}$$

Jediné, co zde není očividné, je rovnost $\mathbf{A}\mathbf{V}\mathbf{V}^T = \mathbf{A}$, neboť sice $\mathbf{V}^T\mathbf{V} = \mathbf{I}$ ale už ne nutně $\mathbf{V}\mathbf{V}^T = \mathbf{I}$ (protože matice $\mathbf{V}$ není čtvercová). Dokažme tuto rovnost. Platí

$$\operatorname{rng}(\mathbf{A}^T) = \operatorname{rng}(\mathbf{A}^T\mathbf{A}) = \operatorname{rng}(\mathbf{V}\mathbf{S}^2\mathbf{V}^T) = \operatorname{rng}\mathbf{V}, \tag{7.32}$$

kde první rovnost plyne z (5.6a) a třetí rovnost z Tvrzení 3.10 (protože matice $\mathbf{S}^2\mathbf{V}^T$ má lineárně nezávislé řádky). Matice $\mathbf{V}\mathbf{V}^T$ je ortogonální projektor na podprostor $\operatorname{rng}\mathbf{V} = \operatorname{rng}(\mathbf{A}^T)$ a proto pro každé $\mathbf{x} \in \operatorname{rng}(\mathbf{A}^T)$ platí $\mathbf{V}\mathbf{V}^T\mathbf{x} = \mathbf{x}$. Protože každý sloupec matice $\mathbf{A}^T$ patří do $\operatorname{rng}(\mathbf{A}^T)$, máme $\mathbf{V}\mathbf{V}^T\mathbf{A}^T = \mathbf{A}^T$ neboli $\mathbf{A}\mathbf{V}\mathbf{V}^T = \mathbf{A}$.

Lze to vidět i jinak. Dle (7.32) je každý sloupec $\mathbf{A}^T$ lineární kombinací sloupců $\mathbf{V}$, což lze psát jako $\mathbf{A}^T = \mathbf{V}\mathbf{B}$. Dosazením máme $\mathbf{A}\mathbf{V}\mathbf{V}^T = \mathbf{B}^T\mathbf{V}^T\mathbf{V}\mathbf{V}^T = \mathbf{B}^T\mathbf{V}^T = \mathbf{A}$.

7.21.a) Pro každá $\lambda, \mathbf{u}, \mathbf{v}$ platí implikace (dokažte roznásobením blokových matic)

$$\begin{bmatrix} \mathbf{0} & \mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{v} \\ \mathbf{u} \end{bmatrix} = \lambda \begin{bmatrix} \mathbf{v} \\ \mathbf{u} \end{bmatrix} \quad \Longrightarrow \quad \begin{bmatrix} \mathbf{0} & \mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{v} \\ -\mathbf{u} \end{bmatrix} = -\lambda \begin{bmatrix} \mathbf{v} \\ -\mathbf{u} \end{bmatrix},$$

tedy má-li matice $\mathbf{H}$ vlastní číslo λ s vlastním vektorem $(\mathbf{v}, \mathbf{u})$, pak má také vlastní číslo $-\lambda$ s vlastním vektorem $(\mathbf{v}, -\mathbf{u})$. Proto $\mathbf{H}$ musí mít sudý počet nenulových vlastních čísel, označme ho $2r$. Rozdělíme-li těchto $2r$ nenulových vlastních čísel matice $\mathbf{H}$ na r kladných a r záporných (diagonální prvky matic $\mathbf{S}$ a $-\mathbf{S}$), dostaneme rozklad $\mathbf{H} = \mathbf{W}\mathbf{\Lambda}\mathbf{W}^T$.

Rovnosti $\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I}_r$ plynou z rovnosti $\mathbf{W}^T\mathbf{W} = \mathbf{I}_{2r}$ po vynásobení blokových matic a porovnáním s bloky matice $\mathbf{I}_{2r}$.

7.21.b) Ve výrazu $\mathbf{W}\mathbf{\Lambda}\mathbf{W}^T$ roznásobte blokové matice,

$$\mathbf{W}\mathbf{\Lambda}\mathbf{W}^T = \tfrac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{V} & \mathbf{V} \\ \mathbf{U} & -\mathbf{U} \end{bmatrix} \begin{bmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & -\mathbf{S} \end{bmatrix} \tfrac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{V} & \mathbf{V} \\ \mathbf{U} & -\mathbf{U} \end{bmatrix}^T = \begin{bmatrix} \mathbf{0} & \mathbf{V}\mathbf{S}\mathbf{U}^T \\ \mathbf{U}\mathbf{S}\mathbf{V}^T & \mathbf{0} \end{bmatrix},$$

a porovnejte bloky s bloky matice $\mathbf{H}$.

7.21.c) Musíme dosadit optimální řešení $\mathbf{X} = \mathbf{U}\mathbf{V}^T$ úlohy do kritéria $\langle \mathbf{A}, \mathbf{X} \rangle$ a upravit, přičemž také dosadíme za $\mathbf{A}$ její singulární rozklad: $\langle \mathbf{A}, \mathbf{X} \rangle = \operatorname{tr}(\mathbf{X}^T\mathbf{A}) = \operatorname{tr}(\mathbf{V}\mathbf{U}^T\mathbf{U}\mathbf{S}\mathbf{V}^T) = \operatorname{tr}(\mathbf{V}\mathbf{S}\mathbf{V}^T) = \operatorname{tr}(\mathbf{V}^T\mathbf{V}\mathbf{S}) = \operatorname{tr}\mathbf{S}$. Tedy optimální hodnota je součet singulárních čísel matice $\mathbf{A}$.

Část II

Nelineární optimalizace

Kapitola 8

Derivace

8.1 Nelineární zobrazení

Dosud jsme potkali lineární a afinní zobrazení a kvadratické funkce. Nyní přidáme libovolné diferencovatelné funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ a zobrazení $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$ (zopakujte funkce a zobrazení z §1.1.3). Pro ty, kdo prošli vícerozměrnou analýzou, bude většina této kapitoly opakování, i když možná v trochu jiném značení.

Většinu definic a vět vyslovíme pro funkce a zobrazení, jejichž definičním oborem je celé $\mathbb{R}^n$. To samozřejmě obecně neplatí, protože součástí definice funkce je i její definiční obor. Např. funkce $f\colon \mathbb{R} \to \mathbb{R}$ daná vzorečkem $f(x) = x^2$ a funkce $g\colon [0,1] \to \mathbb{R}$ daná tím samým vzorečkem $g(x) = x^2$ jsou dvě různé funkce. Podobně, funkce (s reálnými hodnotami) daná vzorcem $f(x) = \sqrt{1-x^2}$ není pro některá $x \in \mathbb{R}$ definována a jejím největším možným definičním oborem je interval $[-1,1]$. Toto zjednodušení však usnadní výklad a vždy bude očividné, jak by se látka zobecnila pro jiný definiční obor.

Pro funkci $f\colon X \to \mathbb{R}$ s definičním oborem $X \subseteq \mathbb{R}^n$ užíváme tyto pojmy:

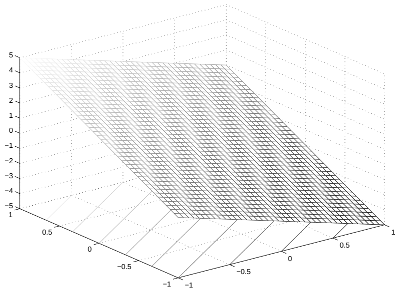
- **Graf** funkce f je množina[1] $\{\, (\mathbf{x}, y) \in \mathbb{R}^{n+1} \mid f(\mathbf{x}) = y \,\} = \{\, (\mathbf{x}, f(\mathbf{x})) \mid \mathbf{x} \in \mathbb{R}^n \,\}$.
- **Vrstevnice** funkce f výšky $y \in \mathbb{R}$ je množina $\{\, \mathbf{x} \in \mathbb{R}^n \mid f(\mathbf{x}) = y \,\}$.

Příklad 8.1. Graf funkce $f\colon \mathbb{R} \to \mathbb{R}$ s hodnotami $f(x) = x^2$ je množina $\{\, (x,y) \mid x^2 = y \,\}$. Vrstevnice funkce f výšky 1 je množina $\{\, x \mid x^2 = 1 \,\} = \{-1, 1\}$.

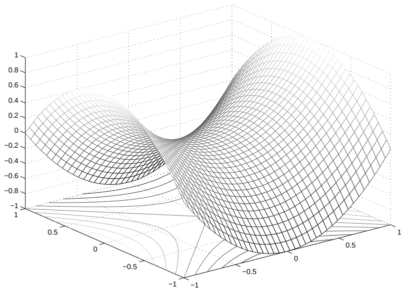
Graf funkce $f\colon \mathbb{R}^2 \to \mathbb{R}$ s hodnotani $f(x,y) = x^2 + y^2$ je množina $\{\, (x,y,z) \mid x^2 + y^2 = z \,\}$. Vrstevnice funkce f výšky 1 je množina $\{\, (x,y) \mid x^2 + y^2 = 1 \,\}$. ♦

Příklad 8.2. Obrázek ukazuje část (na obdélníku $[-1,1]^2$) grafu a vrtevnic funkcí $f\colon \mathbb{R}^2 \to \mathbb{R}$:

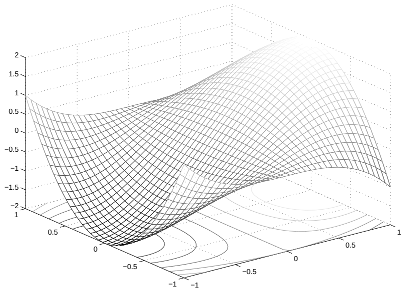
[1] Formálně vzato, graf funkce je přesně to samé, jako funkce. To proto, že funkce (zobrazení) $f\colon A \to B$ se definuje jako binární relace $f \subseteq A \times B$ taková, že pro každé $a \in A$ je nejvýše jedno $b \in B$ takové, že $(a,b) \in f$ (viz §1.1.2). Graf funkce je přesně tato binární relace, kde v našem případě $A = \mathbb{R}^n$ a $B = \mathbb{R}$.



$f(x, y) = -2x + 3y$



$f(x, y) = x^2 - y^2$



$f(x, y) = 3x - x^3 - 3xy^2$

♦

Příklad 8.3. Příklady funkcí a zobrazení reálných proměnných:

1. $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$
2. $f\colon \mathbb{R}^2 \to \mathbb{R}$, $f(x, y) = x^2 - y^2$
3. $f\colon \mathbb{R}^n \to \mathbb{R}$, $f(\mathbf{x}) = a$ kde $a \in \mathbb{R}$ (konstantní funkce n proměnných)
4. $f\colon \mathbb{R}^n \to \mathbb{R}$, $f(\mathbf{x}) = x_1$ (i když $x_2, \dots, x_n$ chybí, f se rozumí jako funkce n proměnných)
5. $f\colon \mathbb{R}^n \to \mathbb{R}$, $f(\mathbf{x}) = \mathbf{a}^T\mathbf{x} = a_1x_1 + \cdots + a_nx_n$ kde $\mathbf{a} \in \mathbb{R}^n$ (lineární funkce)
6. $f\colon \mathbb{R}^n \to \mathbb{R}$, $f(\mathbf{x}) = \mathbf{a}^T\mathbf{x} + b = a_1x_1 + \cdots + a_nx_n + b$ kde $\mathbf{a} \in \mathbb{R}^n$, $b \in \mathbb{R}$ (afinní funkce)
7. $f\colon \mathbb{R}^n \to \mathbb{R}$, $f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} + c$ kde $\mathbf{A} \in \mathbb{R}^{n\times n}$, $\mathbf{b} \in \mathbb{R}^n$, $c \in \mathbb{R}$ (kvadratická funkce)
8. $f\colon \mathbb{R}^n \to \mathbb{R}$, $f(\mathbf{x}) = \mathrm{e}^{-\mathbf{x}^T\mathbf{x}}$ (nenormalizované Gaussovo rozdělení)
9. $f\colon \mathbb{R}^n \to \mathbb{R}$, $f(\mathbf{x}) = \max\limits_{i=1}^{n} x_i$
10. $\mathbf{f}\colon \mathbb{R} \to \mathbb{R}^2$, $\mathbf{f}(t) = (\cos t,\, \sin t)$ (parametrizace kružnice, množina $\mathbf{f}([0, 2\pi))$ je kružnice)
11. $\mathbf{f}\colon \mathbb{R} \to \mathbb{R}^3$, $\mathbf{f}(t) = (\cos t,\, \sin t,\, at)$ (parametrizace šroubovice neboli helixu)
12. $\mathbf{f}\colon \mathbb{R} \to \mathbb{R}^n$, $\mathbf{f}(t) = \mathbf{x} + t\mathbf{v}$ (parametrizace přímky jdoucí bodem $\mathbf{x}$ ve směru $\mathbf{v}$)
13. *Poloha bodu/částice v čase* ve fyzice je popsána zobrazením $\mathbf{f}\colon \mathbb{R} \to \mathbb{R}^3$ kde $\mathbf{f}(t) \in \mathbb{R}^3$ je poloha bodu čase t
14. $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^n$, $\mathbf{f}(\mathbf{x}) = \mathbf{x}$ (identické zobrazení neboli identita)
15. $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$, $\mathbf{f}(\mathbf{x}) = \mathbf{a}$ kde $\mathbf{a} \in \mathbb{R}^m$ (konstantní zobrazení)
16. $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$, $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x}$ kde $\mathbf{A} \in \mathbb{R}^{m\times n}$
    (lineární zobrazení; je to také parametrizace lineárního podprostoru $\mathbf{f}(\mathbb{R}^n) = \operatorname{rng} \mathbf{A} \subseteq \mathbb{R}^m$)
17. $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$, $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$ kde $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$
    (afinní zobrazení; je to také parametrizace afinního podprostoru $\mathbf{f}(\mathbb{R}^n) = \mathbf{b} + \operatorname{rng} \mathbf{A} \subseteq \mathbb{R}^m$)
18. $\mathbf{f}\colon \mathbb{R}^2 \to \mathbb{R}^3$, $\mathbf{f}(u, v) = (\,(R + r\cos v)\cos u,\, (R + r\cos v)\sin u,\, r\sin v\,)$
    (parametrizace toru neboli anuloidu, množina $\mathbf{f}([0, 2\pi) \times [0, 2\pi))$ je torus)
19. *Skalární pole* (pojem z fyziky) je funkce $\mathbb{R}^3 \to \mathbb{R}$. Např. teplotní pole přiřadí každému bodu prostoru teplotu.
20. *Vektorové pole* (opět z fyziky) je zobrazení $\mathbb{R}^3 \to \mathbb{R}^3$. Např. intenzita elektrického pole přiřadí každému bodu v $\mathbb{R}^3$ vektor z $\mathbb{R}^3$. ♦

8.2 Spojitost

Spojitost zobrazení $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$ lze definovat několika ekvivalentními způsoby. Uvedeme tento: zobrazení $\mathbf{f}$ je **spojité** v bodě $\mathbf{x} \in \mathbb{R}^n$, jestliže

$$\lim_{\mathbf{y}\to\mathbf{x}} \mathbf{f}(\mathbf{y}) = \mathbf{f}(\mathbf{x}). \tag{8.1}$$

Tato definice není pohodlná, protože počítat limity funkcí více proměnných je obtížné. Uvedeme proto postačující (avšak nikoliv nutnou) podmínku pro spojitost, která v praxi většinou stačí. Přitom předpokládáme, že čtenář pouze dokáže ověřit spojitost funkcí jedné proměnné. Věta vlastně ukazuje, jak lze rekurzivně sestrojit spojitá zobrazení více proměnných ze spojitých funkcí jedné proměnné. Důkaz věty vynecháme.

Věta 8.1.

1. Nechť funkce $f\colon \mathbb{R} \to \mathbb{R}$ je spojitá v bodě x. Nechť $k \in \{1,\dots,n\}$ a nechť funkce $g\colon \mathbb{R}^n \to \mathbb{R}$ je dána jako $g(x_1,\dots,x_n) = f(x_k)$ (tedy g závisí jen na proměnné x_k). Pak funkce g je spojitá v každém bodě $(x_1,\dots,x_n)$ takovém, že $x_k = x$.
2. Nechť funkce $f, g\colon \mathbb{R}^n \to \mathbb{R}$ jsou spojité v bodě $\mathbf{x}$. Pak funkce $f+g$, $f-g$ a fg jsou spojité v bodě $\mathbf{x}$. Pokud $g(\mathbf{x}) \neq 0$, je funkce f/g spojitá v bodě $\mathbf{x}$.
3. Nechť $f\colon \mathbb{R}^n \to \mathbb{R}$ je spojitá v bodě $\mathbf{x}$ a $g\colon \mathbb{R} \to \mathbb{R}$ je spojitá v bodě $y = f(\mathbf{x})$. Pak složená funkce $g \circ f\colon \mathbb{R}^n \to \mathbb{R}$ je spojitá v bodě $\mathbf{x}$.
4. Nechť funkce $f_1,\dots,f_m\colon \mathbb{R}^n \to \mathbb{R}$ jsou spojité v bodě $\mathbf{x}$. Pak zobrazení $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$ definované jako $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}),\dots,f_m(\mathbf{x}))$ je spojité v bodě $\mathbf{x}$.

Příklad 8.4. Z věty snadno ukážeme, že např. funkce dvou proměnných

$$f(x,y) = \sin(x+y^2) \tag{8.2}$$

je spojitá. Podle 1 je x spojitá funkce dvou proměnných (x,y). Podobně, y^2 je spojitá funkce proměnných (x,y). Podle 2 je proto funkce $x+y^2$ spojitá. Protože funkce sin je spojitá, podle 3 je funkce $\sin(x+y^2)$ spojitá. Takto jsme 'rekurzivně' dokázali spojitost celé funkce. ♦

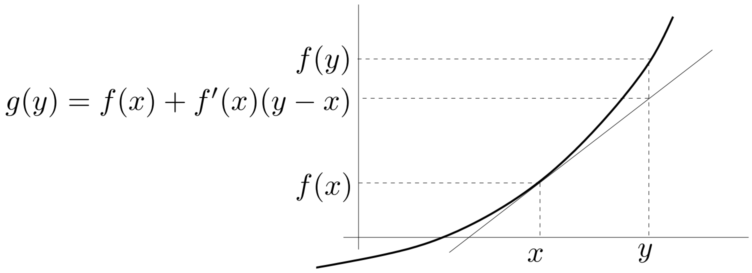
8.3 Derivace funkce jedné proměnné

Zopakujme pojmy diferencovatelnosti a derivace funkce jedné proměnné. Funkce $f\colon \mathbb{R} \to \mathbb{R}$ je *diferencovatelná* v bodě $x \in \mathbb{R}$ tehdy, existuje-li limita

$$\lim_{y\to x} \frac{f(y)-f(x)}{y-x} = \lim_{\alpha\to 0} \frac{f(x+\alpha)-f(x)}{\alpha}. \tag{8.3}$$

Pak se hodnota limity (8.3) nazývá *derivace* funkce f v bodě x. Tuto derivaci značíme jedním ze symbolů $\dfrac{\mathrm{d}f(x)}{\mathrm{d}x}$ (Leibnizovo značení) nebo $f'(x)$ (Lagrangeovo značení). Symbol $\dfrac{\mathrm{d}f}{\mathrm{d}x}\colon \mathbb{R} \to \mathbb{R}$ nebo $f'\colon \mathbb{R} \to \mathbb{R}$ označuje funkci, která bodu $x \in \mathbb{R}$ přiřadí derivaci funkce f v bodě x.

Této derivaci se přesněji říká *oboustraná derivace*. *Jednostranná* diferencovatelnost a derivace funkce $f\colon \mathbb{R} \to \mathbb{R}$ v bodě $x \in \mathbb{R}$ se definuje obdobně, pouze limita (8.3) se změní na limitu zprava či zleva. Značíme je $f'_+(x)$ (derivace zprava) a $f'_-(x)$ (derivace zleva). Oboustranná derivace existuje právě tehdy, když existují derivace zprava i zleva a jsou si rovny.



Diferencovatelnost a derivaci lze definovat i jiným, ekvivalentním způsobem: funkce f je diferencovatelná v bodě x tehdy, lze-li ji v okolí bodu x 'dobře' aproximovat afinní funkcí. Upřesněme tento výrok. Označme naši afinní funkci jako $g\colon \mathbb{R} \to \mathbb{R}$ a pišme ji ve tvaru[2]

$$g(y) = f(x) + a(y - x), \tag{8.4}$$

Samozřejmě, funkce g je obecně jiná[3] pro každé $x \in \mathbb{R}$. Funkce g je opravdu afinní (viz (3.7)), protože $g(y) = ay + b$ kde $b = f(x) - ax$. Tedy funkce f je diferencovatelná v bodě x právě tehdy, když existuje číslo a takové, že chyba aproximace $f(y) - g(y)$ se pro $y \to x$ v prvním řádu blíží nule, tedy

$$\lim_{y\to x} \frac{f(y) - g(y)}{|y - x|} = \lim_{y\to x} \frac{f(y) - f(x) - a(y - x)}{|y - x|} = 0. \tag{8.5}$$

Pak se číslu a říká *derivace* funkce f v bodě x. Dokážeme, že obě definice jsou ekvivalentní:

Tvrzení 8.2. Limita (8.3) existuje a je rovna a, právě když platí (8.5).

Důkaz. Oboustranná limita existuje, právě když existují obě jednostranné limity a jsou si rovny. Tyto jednostranné limity je ale možné zjednodušit:

$$\begin{aligned}
\lim_{y\to x^+} \frac{f(y) - f(x) - a(y - x)}{|y - x|} &= \lim_{y\to x^+} \frac{f(y) - f(x) - a(y - x)}{y - x} = \lim_{y\to x^+} \frac{f(y) - f(x)}{y - x} - a,\\
\lim_{y\to x^-} \frac{f(y) - f(x) - a(y - x)}{|y - x|} &= \lim_{y\to x^-} \frac{f(y) - f(x) - a(y - x)}{-y + x} = a - \lim_{y\to x^-} \frac{f(y) - f(x)}{y - x}.
\end{aligned}$$

Porovnáme-li pravé strany s (8.3), dostaneme dokazované tvrzení. ■

8.4 Parciální derivace

Parciální derivaci funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ podle i-té proměnnné v bodě $\mathbf{x} = (x_1, \ldots, x_n)$ spočítáme tak, že všechny proměnné x_j, $j \neq i$, považujeme za konstanty a zderivujeme funkci podle jediné proměnné x_i. Značíme ji jedním ze symbolů $\dfrac{\partial f(\mathbf{x})}{\partial x_i}$ nebo $f_{x_i}(\mathbf{x})$. Symbol $\dfrac{\partial f}{\partial x_i}\colon \mathbb{R}^n \to \mathbb{R}$ nebo $f_{x_i}\colon \mathbb{R}^n \to \mathbb{R}$ pak označuje funkci, která přiřazuje bodu $\mathbf{x} \in \mathbb{R}^n$ parciální derivaci funkce f podle i-té proměnné v bodě $\mathbf{x}$.

[2]Někdo raději píše $g(x + \alpha) = f(x) + a\alpha$, což je samozřejmě stejné jako (8.4) dosazením $y = x + \alpha$.
[3]Takže bychom přesněji mohli psát $g_x(y)$ nebo $g(x, y)$.

Příklad 8.5. Parciální derivace funkce (8.2) jsou

$$\frac{\partial f(x,y)}{\partial x} = f_x(x,y) = \cos(x+y^2), \tag{8.6a}$$

$$\frac{\partial f(x,y)}{\partial y} = f_y(x,y) = 2y\cos(x+y^2). \tag{8.6b}$$

♦

8.5 Totální derivace

Jak se dají pojmy diferencovatelnosti a derivace zobecnit z funkce $f\colon \mathbb{R} \to \mathbb{R}$ na zobrazení $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$? Mohli bychom zkusit toho dosáhnout zobecněním limity (8.3)[4]. Obvyklejší způsob ovšem vede přes alternativní definici diferencovatelnosti (8.5). Zkusme zobrazení v blízkosti daného bodu $\mathbf{x} \in \mathbb{R}^n$ aproximovat afinním zobrazením

$$\mathbf{g}(\mathbf{y}) = \mathbf{f}(\mathbf{x}) + \mathbf{A}(\mathbf{y} - \mathbf{x}), \tag{8.7}$$

kde $\mathbf{A} \in \mathbb{R}^{m\times n}$. Zobrazení $\mathbf{f}$ se nazývá **totálně diferencovatelné** (nebo jen **diferencovatelné**) v bodě $\mathbf{x}$, jestliže existuje matice $\mathbf{A}$ taková, že chyba aproximace $\mathbf{f}(\mathbf{y}) - \mathbf{g}(\mathbf{y})$ se pro $\mathbf{y} \to \mathbf{x}$ v prvním řádu blíží nule, tedy

$$\lim_{\mathbf{y}\to\mathbf{x}} \frac{\mathbf{f}(\mathbf{y}) - \mathbf{g}(\mathbf{y})}{\|\mathbf{y} - \mathbf{x}\|} = \lim_{\mathbf{y}\to\mathbf{x}} \frac{\mathbf{f}(\mathbf{y}) - \mathbf{f}(\mathbf{x}) - \mathbf{A}(\mathbf{y} - \mathbf{x})}{\|\mathbf{y} - \mathbf{x}\|} = \mathbf{0}. \tag{8.8}$$

Tato limita nejde nijak zjednodušit, jako jsme to udělali s limitou (8.5).

Definice (8.8) není pro ověřování diferencovatelnosti pohodlná, protože počítání limit funkcí více proměnných je obtížné. Uvedeme proto postačující (ne však nutnou) podmínku pro diferencovatelnost, která je mnohem pohodlnější. Zobrazení $\mathbf{f}$ nazveme **spojitě diferencovatelné** v bodě $\mathbf{x}$, jestliže v bodě $\mathbf{x}$ existují všechny parciální derivace $\partial f_i(\mathbf{x})/\partial x_j$ a funkce $\partial f_i/\partial x_j$ jsou v něm spojité.

Věta 8.3. Je-li zobrazení v nějakém bodě spojitě diferencovatelné, pak je v tomto bodě totálně diferencovatelné.

Příklad 8.6. Obě parciální derivace (8.6) funkce (8.2) jsou spojité funkce na celém $\mathbb{R}^2$ (neboť splňují předpoklady Věty 8.1). Proto je funkce (8.2) spojitě diferencovatelná na $\mathbb{R}^2$, a tedy dle Věty 8.3 je diferencovatelná na $\mathbb{R}^2$. ♦

Pozor: pouhá existence všech parciálních derivací pro diferencovatelnost nestačí.

V případě, že je zobrazení $\mathbf{f}$ je v bodě $\mathbf{x}$ diferencovatelné, má matice $\mathbf{A}$ přirozený tvar: její složky jsou parciální derivace všech složek zobrazení podle všech proměnných:

$$\mathbf{A} = \frac{\mathrm{d}\mathbf{f}(\mathbf{x})}{\mathrm{d}\mathbf{x}} = \mathbf{f}'(\mathbf{x}) = \begin{bmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_1(\mathbf{x})}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_m(\mathbf{x})}{\partial x_n} \end{bmatrix} \in \mathbb{R}^{m\times n}. \tag{8.9}$$

[4] Tato cesta je možná a vede na zajímavou definici totální derivace podle Carathéodoryho.

Matice (8.9) se nazývá **totální derivace** (nebo krátce jen **derivace**) zobrazení $\mathbf{f}$ v bodě $\mathbf{x}$. Z historických důvodů se jí říká také **Jacobiho matice**[5]. Funkci, která bodu $\mathbf{x} \in \mathbb{R}^n$ přiřadí (totální) derivaci zobrazení $\mathbf{f}$ v bodě $\mathbf{x}$, značíme $\dfrac{\mathrm{d}\mathbf{f}}{\mathrm{d}\mathbf{x}}\colon \mathbb{R}^n \to \mathbb{R}^{m\times n}$ nebo $\mathbf{f}'\colon \mathbb{R}^n \to \mathbb{R}^{m\times n}$.

Pamatujte speciální případy:

- Pro $f\colon \mathbb{R} \to \mathbb{R}$ je $f'(x)$ skalár a splývá s obyčejnou derivací (8.3) (dle Tvrzení 8.2).
- Pro $\mathbf{f}\colon \mathbb{R} \to \mathbb{R}^m$ je $\mathbf{f}'(x) = \begin{bmatrix} f_1'(x) \\ \vdots \\ f_m'(x) \end{bmatrix}$ sloupcový vektor, jehož složky jsou derivace složek $f_1, \dots, f_m\colon \mathbb{R} \to \mathbb{R}$ zobrazení $\mathbf{f}$. Funkce je diferencovatelná, právě když jsou všechny její složky diferencovatelné, protože rovnost (8.8) je ekvivalentní m rovnostem (8.5).
- Pro $f\colon \mathbb{R}^n \to \mathbb{R}$ je $f'(\mathbf{x}) = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix}$ řádkový vektor.

8.5.1 Derivace složeného zobrazení

Známé 'řetízkové pravidlo' pro derivaci složených funkcí jedné proměnné lze přirozeným způsobem rozšířit na zobrazení. Důkaz následující věty není krátký a nebudeme ho uvádět.

Věta 8.4 (řetízkové pravidlo). Nechť zobrazení $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$ je diferencovatelnné v bodě $\mathbf{x} \in \mathbb{R}^n$ a zobrazení $\mathbf{g}\colon \mathbb{R}^m \to \mathbb{R}^l$ je diferencovatelné v bodě $\mathbf{f}(\mathbf{x})$. Pak složené zobrazení $\mathbf{g} \circ \mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^l$ je diferencovatelné v bodě $\mathbf{x}$ a jeho derivace je rovna

$$(\mathbf{g} \circ \mathbf{f})'(\mathbf{x}) = (\mathbf{g}' \circ \mathbf{f})(\mathbf{x})\,\mathbf{f}'(\mathbf{x}) \qquad \text{neboli} \qquad \frac{\mathrm{d}\mathbf{g}(\mathbf{f}(\mathbf{x}))}{\mathrm{d}\mathbf{x}} = \mathbf{g}'(\mathbf{f}(\mathbf{x}))\,\mathbf{f}'(\mathbf{x}). \tag{8.10}$$

Dimenze zúčastněných prostorů lze přehledně znázornit zápisem (viz §1.1.2)

$$\mathbb{R}^n \xrightarrow{\mathbf{f}} \mathbb{R}^m \xrightarrow{\mathbf{g}} \mathbb{R}^l.$$

Pokud píšeme $\mathbf{u} = \mathbf{f}(\mathbf{x})$ a $\mathbf{y} = \mathbf{g}(\mathbf{u})$, pravidlo lze psát také v Leibnizově značení jako

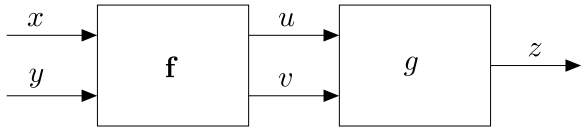
$$\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}\mathbf{x}} = \frac{\mathrm{d}\mathbf{y}}{\mathrm{d}\mathbf{u}}\frac{\mathrm{d}\mathbf{u}}{\mathrm{d}\mathbf{x}}, \tag{8.11}$$

což se dobře pamatuje, protože d$\mathbf{u}$ se 'jakoby vykrátí' (což ale není důkaz ničeho!). Zdůrazněme, že tato rovnost je *násobení matic*. Výraz na levé straně je matice $l \times n$, první výraz na pravé straně je matice $l \times m$ a druhý výraz na pravé straně je matice $m \times n$. Pro $l = m = n = 1$ dostaneme řetízkové pravidlo pro derivaci složení funkcí jedné proměnné. Pravidlo se indukcí dá zobecnit na složení více než dvou zobrazení (řetízek): *Jacobiho matice složeného zobrazení je součinem Jacobiho matic jednotlivých zobrazení.*

Příklad 8.7. Nechť $g\colon \mathbb{R}^2 \to \mathbb{R}$ je diferencovatelná funkce. Najděte derivaci funkce $h(x,y) = g(x+y, xy)$ podle vektoru (x,y), tedy řádkový vektor $h'(x,y) = \begin{bmatrix} h_x(x,y) & h_y(x,y) \end{bmatrix} \in \mathbb{R}^{1\times 2}$.

Máme $\mathbb{R}^2 \xrightarrow{\mathbf{f}} \mathbb{R}^2 \xrightarrow{g} \mathbb{R}$, kde $\mathbf{f}(x,y) = (u,v) = (x+y, xy)$. Označme $z = g(u,v)$. Viz obrázek:

[5] V optimalizaci se této matici často říká také *jacobián*, v matematické analýze toto slovo ale označuje determinant Jacobiho matice (je-li čtvercová).



Derivace zobrazení g podle vektoru (u, v) je řádkový vektor $g'(u, v) = \begin{bmatrix} g_u(u, v) & g_v(u, v) \end{bmatrix}$. Derivace zobrazení $\mathbf{f}$ podle vektoru (x, y) je matice 2×2

$$\mathbf{f}'(x, y) = \frac{\mathrm{d}\mathbf{f}(x, y)}{\mathrm{d}(x, y)} = \begin{bmatrix} \frac{\partial(x+y)}{\partial x} & \frac{\partial(x+y)}{\partial y} \\ \frac{\partial(xy)}{\partial x} & \frac{\partial(xy)}{\partial y} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ y & x \end{bmatrix}.$$

Derivace zobrazení $h = g \circ \mathbf{f} \colon \mathbb{R}^2 \to \mathbb{R}$ podle vektoru (x, y) je řádkový vektor

$$\begin{aligned} \frac{\mathrm{d}h(x, y)}{\mathrm{d}(x, y)} = \frac{\mathrm{d}g(\mathbf{f}(x, y))}{\mathrm{d}(x, y)} &= g'(u, v)\, \mathbf{f}'(x, y) \\ &= \begin{bmatrix} g_u(u, v) & g_v(u, v) \end{bmatrix} \begin{bmatrix} 1 & 1 \\ y & x \end{bmatrix} \\ &= \begin{bmatrix} g_u(u, v) + y g_v(u, v) & g_u(u, v) + x g_v(u, v) \end{bmatrix}, \end{aligned}$$

kde $u = x + y$ a $v = xy$. ♦

Příklad 8.8. Ukažme dva způsoby, jak najít parciální derivaci h_x funkce $h(x, y) = \mathrm{e}^{(x+y)^2 + (xy)^2}$:

- Považujeme y za konstantu a derivujeme h jako funkci jedné proměnné x:

$$h_x(x, y) = [2(x + y) + 2(xy)y]\mathrm{e}^{(x+y)^2+(xy)^2} = 2(x + y + xy^2)\mathrm{e}^{(x+y)^2+(xy)^2}.$$

- Položíme $u = x + y$, $v = xy$, $g(u, v) = \mathrm{e}^{u^2+v^2}$. Z Příkladu 8.7 máme $h_x = g_u + y g_v$. Jelikož

$$g_u(u, v) = 2u\mathrm{e}^{u^2+v^2}, \quad g_v(u, v) = 2v\mathrm{e}^{u^2+v^2},$$

  máme $h_x(x, y) = g_u + y g_v = 2u\mathrm{e}^{u^2+v^2} + y(2v)\mathrm{e}^{u^2+v^2} = 2(x + y + xy^2)\mathrm{e}^{(x+y)^2+(xy)^2}$. ♦

Příklad 8.9. Máme diferencovatelnou funkci $g \colon \mathbb{R}^2 \to \mathbb{R}$ a chceme spočítat derivaci funkce $g(t + t^2, \sin t)$ podle t.

Máme $\mathbb{R} \xrightarrow{\mathbf{f}} \mathbb{R}^2 \xrightarrow{g} \mathbb{R}$, kde $\mathbf{f}(t) = (u, v) = (t + t^2, \sin t)$. Je

$$\frac{\mathrm{d}g(t + t^2, \sin t)}{\mathrm{d}t} = g'(u, v)\mathbf{f}'(t) = \begin{bmatrix} g_u(u, v) & g_v(u, v) \end{bmatrix} \begin{bmatrix} 1 + 2t \\ \cos t \end{bmatrix} = g_u(u, v)(1 + 2t) + g_v(u, v) \cos t.$$

♦

8.5.2 Maticový kalkulus

Je-li nějaké zobrazení zadáno výrazem obsahujícím vektory a matice, jeho derivaci lze obvykle napsat jako výraz obsahující tytéž vektory a matice. To znamená, že při počítání derivací považujeme matice a vektory za *nedělitelné objekty*. Část matematiky zabývající se derivacemi zobrazení s vektory a maticemi je známa jako *maticový (diferenciální) kalkulus*[6].

[6] Asi víte, že *diferenciální počet* (angl. *differential calculus*) se zabývá derivacemi funkcí. Zatímco *matematická analýza* je široká disciplína studující funkce reálných či komplexních proměnných, slovem *calculus* se obvykle myslí jen soubor pravidel pro derivování funkcí.

Příklad 8.10. Najděme derivaci afinního zobrazení $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$, kde $\mathbf{A} \in \mathbb{R}^{m \times n}$, tedy

$$f_i(\mathbf{x}) = \sum_{j=1}^{n} a_{ij} x_j + b_i, \qquad i = 1, \ldots, m.$$

Je $\partial f_i(\mathbf{x})/\partial x_j = a_{ij}$, tedy dle (8.9) je $\mathbf{f}'(\mathbf{x}) = \mathbf{A}$. Tomu se nedivíme, neboť zobrazení $\mathbf{f}$ již je afinní a proto jeho afinní aproximace (8.7) musí být ono samo. ♦

Příklad 8.11. Najděme derivaci kvadratické formy $f\colon \mathbb{R}^n \to \mathbb{R}$,

$$f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} = \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij} x_i x_j, \tag{8.12}$$

kde $\mathbf{A} = [a_{ij}]$ je (ne nutně symetrická) čtvercová matice. Parciální derivace této funkce podle k-té proměnné je (rozmyslete, derivujte člen po členu!)

$$\frac{\partial f(\mathbf{x})}{\partial x_k} = \sum_{i=1}^{n} (a_{ik} + a_{ki}) x_i$$

Tedy (opět rozmyslete!) $f'(\mathbf{x}) = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix} = \mathbf{x}^T(\mathbf{A} + \mathbf{A}^T)$.. ♦

Příklad 8.12. Odvoďme vektorovou formu známého pravidla pro derivaci součinu. Nechť

$$f(\mathbf{x}) = \mathbf{g}(\mathbf{x})^T \mathbf{h}(\mathbf{x}) = \sum_{i=1}^{m} g_i(\mathbf{x}) h_i(\mathbf{x}), \tag{8.13}$$

kde $\mathbf{g}, \mathbf{h}\colon \mathbb{R}^n \to \mathbb{R}^m$. Podle pravidla pro derivaci součinu skalárních funkcí je

$$\frac{\partial f(\mathbf{x})}{\partial x_j} = \sum_{i=1}^{m} \left(g_i(\mathbf{x}) \frac{\partial h_i(\mathbf{x})}{\partial x_j} + h_i(\mathbf{x}) \frac{\partial g_i(\mathbf{x})}{\partial x_j} \right).$$

Tedy $f'(\mathbf{x}) = \mathbf{g}(\mathbf{x})^T \mathbf{h}'(\mathbf{x}) + \mathbf{h}(\mathbf{x})^T \mathbf{g}'(\mathbf{x})$. ♦

Příklad 8.13. Najděme derivaci zobrazení $\mathbf{f}(\mathbf{A}\mathbf{x} + \mathbf{b})$ podle $\mathbf{x}$, kde $\mathbf{f}\colon \mathbb{R}^l \to \mathbb{R}^m$ a $\mathbf{A} \in \mathbb{R}^{l \times n}$. V řetízkovém pravidle máme $\mathbb{R}^n \xrightarrow{\mathbf{g}} \mathbb{R}^l \xrightarrow{\mathbf{f}} \mathbb{R}^m$, kde $\mathbf{g}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$. Je $\mathbf{g}'(\mathbf{x}) = \mathbf{A}$, tedy

$$\frac{\mathrm{d}\mathbf{f}(\mathbf{A}\mathbf{x} + \mathbf{b})}{\mathrm{d}\mathbf{x}} = \mathbf{f}'(\mathbf{A}\mathbf{x} + \mathbf{b})\mathbf{A}.$$

♦

Tato a některá další pravidla uvedeme v tabulkách. Nejprve pro zobrazení $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$:

$\mathbf{f}(\mathbf{x})$	$\mathbf{f}'(\mathbf{x})$	poznámka
$\mathbf{x}$	$\mathbf{I}$	
$\mathbf{A}\mathbf{x} + \mathbf{b}$	$\mathbf{A}$	$\mathbf{A} \in \mathbb{R}^{m \times n},\ \mathbf{b} \in \mathbb{R}^m$
$\mathbf{A}\mathbf{g}(\mathbf{x})$	$\mathbf{A}\mathbf{g}'(\mathbf{x})$	$\mathbf{A} \in \mathbb{R}^{m \times l},\ \mathbf{g}\colon \mathbb{R}^n \to \mathbb{R}^l$
$\mathbf{g}(\mathbf{A}\mathbf{x} + \mathbf{b})$	$\mathbf{g}'(\mathbf{A}\mathbf{x} + \mathbf{b})\mathbf{A}$	$\mathbf{A} \in \mathbb{R}^{l \times n},\ \mathbf{g}\colon \mathbb{R}^l \to \mathbb{R}^m$

A dále pro některé funkce $f\colon \mathbb{R}^n \to \mathbb{R}$:

$f(\mathbf{x})$	$f'(\mathbf{x})$	poznámka
$\mathbf{a}^T\mathbf{x} + b$	$\mathbf{a}^T$	$\mathbf{a} \in \mathbb{R}^n,\ b \in \mathbb{R}$
$\mathbf{x}^T\mathbf{x} = \|\mathbf{x}\|^2$	$2\mathbf{x}^T$	
$\mathbf{x}^T\mathbf{A}\mathbf{x}$	$\mathbf{x}^T(\mathbf{A} + \mathbf{A}^T)$	$\mathbf{A} \in \mathbb{R}^{n\times n}$
$\|\mathbf{x}\|$	$\mathbf{x}^T/\|\mathbf{x}\|$	
$\mathbf{g}(\mathbf{x})^T\mathbf{g}(\mathbf{x})$	$2\mathbf{g}(\mathbf{x})^T\mathbf{g}'(\mathbf{x})$	$\mathbf{g}\colon \mathbb{R}^n \to \mathbb{R}^m$
$\mathbf{g}(\mathbf{x})^T\mathbf{h}(\mathbf{x})$	$\mathbf{g}(\mathbf{x})^T\mathbf{h}'(\mathbf{x}) + \mathbf{h}(\mathbf{x})^T\mathbf{g}'(\mathbf{x})$	$\mathbf{g}, \mathbf{h}\colon \mathbb{R}^n \to \mathbb{R}^m$

Všechny vzorečky v tabulce se vlastně dají odvodit z derivace součinu a řetízkového pravidla (poslední dva řádky). Např. $\mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{g}(\mathbf{x})^T\mathbf{h}(\mathbf{x})$ kde $\mathbf{g}(\mathbf{x}) = \mathbf{x}$ a $\mathbf{h}(\mathbf{x}) = \mathbf{A}\mathbf{x}$. Rozmyslete i pro ostatní řádky tabulky!

($\star$) Zobrazení z/do prostorů matic

Občas se setkáme s funkcemi $\mathbb{R} \to \mathbb{R}^{m\times n}$, které skalárům přiřazují matice, a s funkcemi $\mathbb{R}^{m\times n} \to \mathbb{R}$, které maticím přiřazují skaláry (příkladem druhé funkce je stopa, determinant nebo maticová norma). Přísně formálně, naši definici (8.8) totální derivace na taková zobrazení nelze použít, protože je formulována pouze pro zobrazení typu $\mathbb{R}^n \to \mathbb{R}^m$. Ovšem lze ji použít, když si matici $m \times n$ si představíme ‘přerovnanou’ do vektoru délky mn.

Složky funkce $\mathbf{F}\colon \mathbb{R} \to \mathbb{R}^{m\times n}$ jsou jednoduše funkce $f_{ij}\colon \mathbb{R} \to \mathbb{R}$ jediné skalární proměnné. Funkce je diferencovatelná, právě když všechny její složky jsou diferencovatelné. Derivaci této funkce v bodě $x \in \mathbb{R}$ je přirozené definovat jako matici

$$\mathbf{F}'(x) = \begin{bmatrix} f'_{11}(x) & \cdots & f'_{1n}(x) \\ \vdots & \ddots & \vdots \\ f'_{m1}(x) & \cdots & f'_{mn}(x) \end{bmatrix} \in \mathbb{R}^{m\times n}. \tag{8.14}$$

Pro $n = 1$ je tato matice Jacobiho matice (8.9) zobrazení $\mathbb{R} \to \mathbb{R}^m$, tedy sloupcový vektor.

Derivaci diferencovatelné funkce $f\colon \mathbb{R}^{m\times n} \to \mathbb{R}$ v bodě $\mathbf{X} = [x_{ij}] \in \mathbb{R}^{m\times n}$ je přirozené definovat jako matici

$$f'(\mathbf{X}) = \begin{bmatrix} \frac{\partial f(\mathbf{X})}{\partial x_{11}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{m1}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(\mathbf{X})}{\partial x_{1n}} & \cdots & \frac{\partial f(\mathbf{X})}{\partial x_{mn}} \end{bmatrix} \in \mathbb{R}^{n\times m}. \tag{8.15}$$

Matice (8.15) má rozměr $n \times m$ (tedy obráceně, než čekáte), aby pro případ $n = 1$ souhlasila s Jacobiho maticí (8.9). Zde jsou bez důkazů vzorce pro derivace některých funkcí $f\colon \mathbb{R}^{m\times n} \to \mathbb{R}$:

$f(\mathbf{X})$	$f'(\mathbf{X})$	poznámka
$\operatorname{tr}\mathbf{X} = \langle \mathbf{I}, \mathbf{X}\rangle$	$\mathbf{I}$	$\mathbf{X}$ čtvercová
$\operatorname{tr}(\mathbf{AX}) = \langle \mathbf{A}^T, \mathbf{X}\rangle$	$\mathbf{A}$	
$\|\mathbf{X}\|^2 = \langle \mathbf{X}, \mathbf{X}\rangle = \operatorname{tr}(\mathbf{X}^T\mathbf{X})$	$2\mathbf{X}^T$	
$\operatorname{tr}(\mathbf{X}^T\mathbf{AX}) = \langle \mathbf{X}, \mathbf{AX}\rangle$	$\mathbf{X}^T(\mathbf{A}+\mathbf{A}^T)$	
$\mathbf{a}^T\mathbf{X}\mathbf{b}$	$\mathbf{b}\mathbf{a}^T$	
$\operatorname{tr}(\mathbf{AXB})$	$\mathbf{BA}$	
$\operatorname{tr}(\mathbf{X}^k)$	$k\mathbf{X}^{k-1}$	$k \in \mathbb{N}$
$\operatorname{tr}(\mathbf{X}^{-1})$	$-\mathbf{X}^{-1}\mathbf{X}^{-1} = -\mathbf{X}^{-2}$	$\mathbf{X}$ regulární
$\det\mathbf{X}$	$(\det\mathbf{X})\mathbf{X}^{-1}$	$\mathbf{X}$ regulární
$\ln\det\mathbf{X}$	$\mathbf{X}^{-1}$	$\mathbf{X}$ positivně definitní
$g(\mathbf{X}^T)$	$g'(\mathbf{X}^T)^T$	$g\colon \mathbb{R}^{m\times n} \to \mathbb{R}$

Tuto tabulku se neučte nazpaměť, je to spíše motivace pro vaše hlubší studium maticového kalkulu. Obdivujte, jak pravidla v tabulce zobecňují známá pravidla pro derivování funkcí jedné proměnné, jako např. $(ax)' = a$, $(x^k)' = kx^{k-1}$, $(1/x)' = -1/x^2$, $(\ln x)' = 1/x$, a pravidla z tabulky pro funkce $\mathbb{R}^n \to \mathbb{R}$.

($\star$) Derivace zobrazení mezi obecnými lineárními prostory

Definici derivace (8.8) lze zobecnit na zobrazení $f\colon U \to V$ mezi obecnými normovanými lineárními prostory U, V: zobrazení $f\colon U \to V$ je *diferencovatelné* v bodě $x \in U$, jestliže existuje lineární zobrazení $a\colon U \to V$ tak, že

$$\lim_{y\to x} \frac{\|f(y) - f(x) - a(y-x)\|}{\|y-x\|} = 0, \tag{8.16}$$

kde $\|\cdot\|$ značí ve jmenovateli normu na prostoru U a v čitateli normu na prostoru V. Pak se zobrazení $a = f'(x)$ nazývá totální (nebo Fréchetova) derivace zobrazení f v bodě x. Řetízkové pravidlo zní

$$(g\circ f)'(x) = (g'\circ f)(x) \circ f'(x) = g'(f(x)) \circ f'(x) \tag{8.17}$$

kde $U \xrightarrow{f} V \xrightarrow{g} W$ a $U \xrightarrow{f'(x)} V \xrightarrow{g'(f(x))} W$. Tedy zatímco v (8.10) násobíme Jacobiho matice, v (8.17) skládáme lineární zobrazení.

Každý ze zúčastněných lineárních prostorů U, V, W může být prostor skalárů, vektorů nebo matic. Problém je, jak kompaktně reprezentovat lineární zobrazení mezi těmito prostory. Lineární zobrazení $\mathbb{R} \to \mathbb{R}^m$, $\mathbb{R}^n \to \mathbb{R}$, $\mathbb{R}^n \to \mathbb{R}^m$, $\mathbb{R}^{m\times n} \to \mathbb{R}$ a $\mathbb{R} \to \mathbb{R}^{k\times l}$ lze reprezentovat maticemi: pak násobení matice vektorem lze vidět jako konvenci pro reprezentaci lineárních zobrazení $\mathbb{R}^n \to \mathbb{R}^m$ a maticový součin jako konvenci pro skládání takových zobrazení. Ovšem lineární zobrazení $\mathbb{R}^{m\times n} \to \mathbb{R}^l$, $\mathbb{R}^l \to \mathbb{R}^{m\times n}$ a $\mathbb{R}^{m\times n} \to \mathbb{R}^{k\times l}$ nelze reprezentovat maticemi a pro manipulaci s takovými lineárními zobrazeními je pohodlnější použít jiné konvence.

8.6 Směrová derivace

Řez zobrazení $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$ v bodě $\mathbf{x} \in \mathbb{R}^n$ ve směru[7] $\mathbf{v} \in \mathbb{R}^n$ je zobrazení $\boldsymbol{\varphi}\colon \mathbb{R} \to \mathbb{R}^m$

[7] Někdy se řez a směrová derivace definují jen pro *normalizované* vektory $\mathbf{v}$ (tj. $\|\mathbf{v}\| = 1$), protože ty lépe zachycují pojem 'směr'. Většina autorů (a my také) však dovoluje libovolný vektor $\mathbf{v}$.

s hodnotami

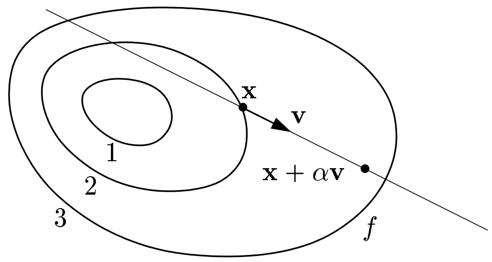
$$\boldsymbol{\varphi}(\alpha) = \mathbf{f}(\mathbf{x} + \alpha\mathbf{v}). \tag{8.18}$$

Směrová derivace[8] zobrazení $\mathbf{f}$ v bodě $\mathbf{x}$ ve směru $\mathbf{v}$ je vektor

$$\mathbf{f}_{\mathbf{v}}(\mathbf{x}) = \boldsymbol{\varphi}'_+(0) = \lim_{\alpha\to 0^+} \frac{\boldsymbol{\varphi}(\alpha) - \boldsymbol{\varphi}(0)}{\alpha} = \lim_{\alpha\to 0^+} \frac{\mathbf{f}(\mathbf{x} + \alpha\mathbf{v}) - \mathbf{f}(\mathbf{x})}{\alpha}, \tag{8.19}$$

kde $\boldsymbol{\varphi}'_+(0)$ označuje derivaci zobrazení $\boldsymbol{\varphi}$ v bodě $\alpha = 0$ zprava, tedy vektor se složkami $\varphi'_{1+}(0), \ldots, \varphi'_{m+}(0)$, kde $\varphi'_{i+}(0)$ označuje derivaci funkce φ_i v bodě $\alpha = 0$ zprava. Směrová derivace existuje, právě když existují derivace všech těchto složek. Pak také říkáme, že zobrazení $\mathbf{f}$ je *diferenovatelné v bodě* $\mathbf{x}$ *ve směru* $\mathbf{v}$.

Směrová derivace se geometricky snadněji představí pro případ $m = 1$, tedy pro funkci $f\colon \mathbb{R}^n \to \mathbb{R}$. Obrázek ilustruje situaci pro funkci $f\colon \mathbb{R}^2 \to \mathbb{R}$:



Příklad 8.14. Spočítejme směrovou derivaci $f_{(u,v)}(x, y)$ funkce

$$f(x, y) = \sin(x + y^2) \tag{8.20}$$

v bodě (x, y) ve směru (u, v):

$$\begin{aligned} \varphi(\alpha) &= \sin(x + \alpha u + (y + \alpha v)^2), \\ \varphi'_+(\alpha) = \varphi'(\alpha) &= (u + 2v(y + \alpha v)) \cos(x + \alpha u + (y + \alpha v)^2), \\ f_{(u,v)}(x, y) = \varphi'_+(0) = \varphi'(0) &= (u + 2vy) \cos(x + y^2). \end{aligned}$$

♦

Parciální derivace funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ není nic jiného než její směrová derivace ve směru i-tého vektoru standardní báze $\mathbf{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0) \in \mathbb{R}^n$ (jednička na i-tém místě).

Jestliže je zobrazení totálně diferencovatelné, směrová derivace má jednodušší vyjádření:

Věta 8.5. Jestliže je zobrazení $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$ v bodě $\mathbf{x} \in \mathbb{R}^n$ totálně diferencovatelné, pak je v tomto bodě diferencovatelné v každém směru $\mathbf{v} \in \mathbb{R}^n$ a platí

$$\mathbf{f}_{\mathbf{v}}(\mathbf{x}) = \mathbf{f}'(\mathbf{x})\mathbf{v}. \tag{8.21}$$

Důkaz. Zobrazení $\mathbf{y} = \boldsymbol{\varphi}(\alpha) = \mathbf{f}(\mathbf{x} + \alpha\mathbf{v})$ je složením zobrazení $\mathbf{y} = \mathbf{f}(\mathbf{u})$ a $\mathbf{u} = g(\alpha) = \mathbf{x} + \alpha\mathbf{v}$, tedy $\mathbb{R} \xrightarrow{g(\alpha)=\mathbf{u}} \mathbb{R}^n \xrightarrow{\mathbf{f}(\mathbf{u})=\mathbf{y}} \mathbb{R}^m$. Je $g'(\alpha) = \mathbf{v}$. Protože $\mathbf{f}$ má derivaci v bodě $\mathbf{x}$, podle řetízkového pravidla má také $\boldsymbol{\varphi}$ (oboustrannou) derivaci v bodě $\mathbf{x}$ a

$$\boldsymbol{\varphi}'(\alpha) = \mathbf{f}'(g(\alpha))g'(\alpha) = \mathbf{f}'(g(\alpha))\mathbf{v}.$$

Ale $g(0) = \mathbf{x}$ a proto $\mathbf{f}_{\mathbf{v}}(\mathbf{x}) = \boldsymbol{\varphi}'_+(0) = \boldsymbol{\varphi}'(0) = \mathbf{f}'(\mathbf{x})\mathbf{v}$. ■

[8] Někteří autoři definují směrovou derivaci jako oboustrannou, kde se místo jednostranné limity (8.19) zprava vezme limita oboustranná. Pokud existují derivace zprava $\boldsymbol{\varphi}'_+(0)$ i derivace zleva $\boldsymbol{\varphi}'_-(0)$ a jsou si rovny, pak jednostranná i oboustranná směrová derivace jsou si rovny.

Příklad 8.15. Spočítejme směrovou derivaci $f_{(u,v)}(x,y)$ funkce (8.20) podle Věty 8.5:

$$f_{(u,v)}(x,y) = uf_x(x,y) + vf_y(x,y) = u\cos(x+y^2) + 2vy\cos(x+y^2) = (u+2vy)\cos(x+y^2).$$

To je stejný výsledek, jaký nám vyšel v Příkladě 8.14. ♦

Rovnost (8.21) říká, že směrová derivace zobrazení $\mathbf{f}$ v pevném bodě $\mathbf{x}$ je *lineární zobrazení* směru $\mathbf{v}$, reprezentované Jacobiho maticí $\mathbf{f}'(\mathbf{x})$. Speciálně je $\mathbf{f}_{-\mathbf{v}}(\mathbf{x}) = -\mathbf{f}_{\mathbf{v}}(\mathbf{x})$. Zdůrazněme, že nic takového obecně neplatí, pokud zobrazení $\mathbf{f}$ není (totálně) diferencovatelné:

Příklad 8.16. Nechť $f\colon \mathbb{R}^n \to \mathbb{R}$ je $f(\mathbf{x}) = \|\mathbf{x}\|$ (tj. eukleidovská norma). Tato funkce v bodě $\mathbf{x} = \mathbf{0}$ není (totálně) diferencovatelná (pro $n = 1$ je $f(x) = |x|$, nakreslete si graf funkce pro $n = 2$). Spočítejme její směrové derivace v tomto bodě podle definice (8.19). Řez funkce f v bodě $\mathbf{x} = \mathbf{0}$ ve směru $\mathbf{v} \in \mathbb{R}^n$ je

$$\varphi(\alpha) = f(\mathbf{x} + \alpha\mathbf{v}) = f(\alpha\mathbf{v}) = \|\alpha\mathbf{v}\| = |\alpha| \cdot \|\mathbf{v}\|.$$

Funkce φ v bodě $\alpha = 0$ nemá oboustrannou derivaci, ale má obě jednostranné derivace. Směrová derivace funkce f v bodě $\mathbf{x} = \mathbf{0}$ ve směru $\mathbf{v}$ je $f_{\mathbf{v}}(\mathbf{0}) = \varphi'_+(0) = \|\mathbf{v}\|$, protože derivace funkce $|\alpha|$ v bodě $\alpha = 0$ zprava je 1. Všimněte si, že $\|\mathbf{v}\|$ není lineární funkcí vektoru $\mathbf{v}$. ♦

8.7 Gradient

Sloupcový vektor parciálních derivací funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ se nazývá její **gradient** a značí se

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix} = f'(\mathbf{x})^T$$

(∇ čteme 'nabla'). Je to tedy transpozice Jacobiho matice $f'(\mathbf{x})$, což je řádkový vektor[9].

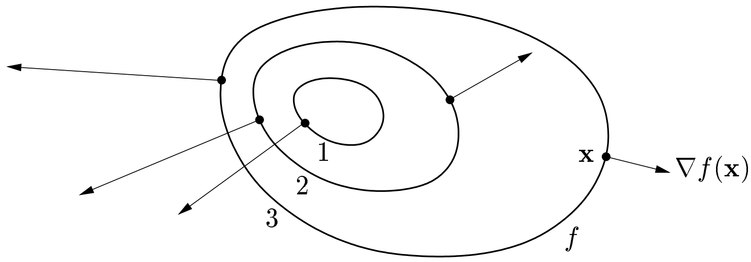
Zkoumejme směrovou derivaci $f_{\mathbf{v}}(\mathbf{x})$ diferencovatelné funkce f v pevném bodě $\mathbf{x}$ pro různé normalizované směry $\mathbf{v}$ (tedy $\|\mathbf{v}\| = 1$). Dle Věty 8.5 je $f_{\mathbf{v}}(\mathbf{x}) = f'(\mathbf{x})\mathbf{v} = \nabla f(\mathbf{x})^T\mathbf{v}$, což je skalární součin gradientu v bodě $\mathbf{x}$ a vektoru $\mathbf{v}$. Je jasné (ale promyslete!), že:

- Směrová derivace je největší ve směru $\mathbf{v} = \nabla f(\mathbf{x})/\|\nabla f(\mathbf{x})\|$, tedy když je $\mathbf{v}$ rovnoběžný s gradientem a stejně orientovaný. Tedy směr gradientu je *směr největšího růstu* funkce.
- Velikost gradientu $\|\nabla f(\mathbf{x})\|$ je strmost funkce ve směru největšího růstu.
- Směrová derivace ve směru kolmém na gradient je nulová.

Dále lze ukázat, že za jistých podmínek (viz §11.2.1) je gradient vždy *kolmý k vrstevnici*.

Obrázek ukazuje tři vrstevnice funkce $f\colon \mathbb{R}^2 \to \mathbb{R}$ a její gradienty v několika bodech:

[9] Zavedení nového slova pro transpozici derivace se zdá zbytečné – důvod je ale v tom, že totální diferenciál je *lineární forma*, kdežto gradient je *vektor*. Toto rozlišení by bylo důležité, kdybychom uvažovali derivaci funkce $f\colon V \to \mathbb{R}$, kde V je obecný (normovaný) lineární prostor (tj. ne nutně $\mathbb{R}^n$). Různí autoři bohužel nejsou jednotní v rozlišování gradientu a (totální) derivace funkce $f\colon \mathbb{R}^n \to \mathbb{R}$. Někdo obojí ztotožňuje a značí $\nabla f(\mathbf{x})$. To ale vede mj. k nekonzistenci s reprezentací totální derivace pomocí Jacobiho matice (8.9), neboť derivace funkce $\mathbb{R}^n \to \mathbb{R}$ pak není speciálním případem derivace zobrazení $\mathbb{R}^n \to \mathbb{R}^m$ pro $m = 1$, což je řádkový vektor.



8.8 Parciální derivace druhého řádu

Zderivujeme-li funkci $f\colon \mathbb{R}^n \to \mathbb{R}$ nejdříve podle proměnné x_i a pak podle proměnné x_j, dostaneme parciální derivaci druhého řádu, kterou značíme

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

Je-li $i = j$, píšeme zkráceně

$$\frac{\partial}{\partial x_i}\frac{\partial f(\mathbf{x})}{\partial x_i} = \frac{\partial^2 f(\mathbf{x})}{\partial x_i^2}.$$

Věta 8.6. Nechť $f\colon \mathbb{R}^n \to \mathbb{R}$ a $\mathbf{x} \in \mathbb{R}^n$. Jestliže druhé parciální derivace

$$\frac{\partial^2 f(\mathbf{x})}{\partial x_i\,\partial x_j}, \qquad \frac{\partial^2 f(\mathbf{x})}{\partial x_j\,\partial x_i}$$

v bodě $\mathbf{x}$ existují a jsou v tomto bodě spojité, pak jsou si rovny.

Příklad 8.17. Určeme všechny druhé parciální derivace funkce $f(x, y) = \sin(x + y^2)$ z Příkladu 8.4. První derivace už tam máme. Nyní druhé derivace:

$$\begin{aligned}
\frac{\partial^2 f(x,y)}{\partial x^2} &= \frac{\partial}{\partial x}(\cos(x+y^2)) &&= -\sin(x+y^2),\\
\frac{\partial^2 f(x,y)}{\partial x\,\partial y} &= \frac{\partial}{\partial y}(\cos(x+y^2)) &&= -2y\sin(x+y^2),\\
\frac{\partial^2 f(x,y)}{\partial y\,\partial x} &= \frac{\partial}{\partial x}(2y\cos(x+y^2)) &&= -2y\sin(x+y^2),\\
\frac{\partial^2 f(x,y)}{\partial y^2} &= \frac{\partial}{\partial y}(2y\cos(x+y^2)) &&= 2\cos(x+y^2) - 4y^2\sin(x+y^2).
\end{aligned}$$

Vidíme, že vskutku nezáleží na pořadí derivování podle x a podle y. ♦

Pro funkci $f\colon \mathbb{R}^n \to \mathbb{R}$ budeme značit matici všech druhých parciálních derivací

$$f''(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_n} \end{bmatrix} \in \mathbb{R}^{n\times n}.$$

Dle Věty 8.6 je tato matice symetrická. Často se jí říká **Hessova matice**[10].

Zde jsou druhé derivace některých funkcí $f\colon \mathbb{R}^n \to \mathbb{R}$:

$f(\mathbf{x})$	$f''(\mathbf{x})$	poznámka
$\mathbf{a}^T\mathbf{x} + b$	$\mathbf{0}$	
$\mathbf{x}^T\mathbf{x} = \|\mathbf{x}\|^2$	$2\mathbf{I}$	
$\mathbf{x}^T\mathbf{A}\mathbf{x}$	$\mathbf{A} + \mathbf{A}^T$	
$g(\mathbf{A}\mathbf{x} + \mathbf{b})$	$\mathbf{A}^T g''(\mathbf{A}\mathbf{x} + \mathbf{b})\mathbf{A}$	$\mathbf{A} \in \mathbb{R}^{m\times n},\ g\colon \mathbb{R}^m \to \mathbb{R}$

Co by byla druhá derivace zobrazení $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$? Nebyla by to již dvojrozměrná tabulka (tedy matice) velikosti $n \times n$, nýbrž třírozměrná tabulka velikosti $m \times n \times n$.

Shrnutí značení derivací

Teď je správná chvíle shrnout značení všech derivací, které jsme v této kapitole potkali:

symbol		význam
$f'(x), \frac{\mathrm{d}f(x)}{\mathrm{d}x}$	$\in \mathbb{R}$	(oboustranná) derivace funkce $f\colon \mathbb{R} \to \mathbb{R}$ v bodě $x \in \mathbb{R}$
$f'_+(x), f'_-(x)$	$\in \mathbb{R}$	derivace funkce $f\colon \mathbb{R} \to \mathbb{R}$ v bodě $x \in \mathbb{R}$ zprava/zleva
$f_{x_i}(\mathbf{x}), \frac{\partial f(\mathbf{x})}{\partial x_i}$	$\in \mathbb{R}$	parciální derivace funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ v bodě $\mathbf{x} \in \mathbb{R}^n$ podle proměnné x_i
$\mathbf{f}_\mathbf{v}(\mathbf{x})$	$\in \mathbb{R}^m$	směrová derivace zobrazení $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$ v bodě $\mathbf{x} \in \mathbb{R}^n$ ve směru $\mathbf{v} \in \mathbb{R}^n$
$\mathbf{f}'(\mathbf{x}), \frac{\mathbf{d}\mathbf{f}(\mathbf{x})}{\mathbf{d}\mathbf{x}}$	$\in \mathbb{R}^{m\times n}$	(totální) derivace (Jacobiho matice) zobrazení $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$ v bodě $\mathbf{x} \in \mathbb{R}^n$
$\nabla f(\mathbf{x})$	$\in \mathbb{R}^n$	gradient funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ v bodě $\mathbf{x} \in \mathbb{R}^n$
$\mathbf{F}'(x), \frac{\mathrm{d}\mathbf{F}(x)}{\mathrm{d}x}$	$\in \mathbb{R}^{m\times n}$	derivace zobrazení $\mathbf{F}\colon \mathbb{R} \to \mathbb{R}^{m\times n}$ v bodě $x \in \mathbb{R}$
$f'(\mathbf{X}), \frac{\mathrm{d}f(\mathbf{X})}{\mathrm{d}\mathbf{X}}$	$\in \mathbb{R}^{n\times m}$	derivace zobrazení $f\colon \mathbb{R}^{m\times n} \to \mathbb{R}$ v bodě $\mathbf{X} \in \mathbb{R}^{m\times n}$
$f''(\mathbf{x})$	$\in \mathbb{R}^{n\times n}$	druhá derivace (Hessova matice) funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ v bodě $\mathbf{x} \in \mathbb{R}^n$

Podotkněme, že značení derivací nejsou zcela ustálená a jiní lidé mohou používat trošku jiná značení. Zvolili jsme značení, která jsou nejčastěji používaná a dohromady konzistentní.

8.9 Taylorův polynom

Nechť funkce $f\colon \mathbb{R} \to \mathbb{R}$ má v bodě x derivace až do řádu k. Její **Taylorův polynom** stupně k v bodě x je polynom $T_x^k\colon \mathbb{R} \to \mathbb{R}$ takový, že v bodě x má všechny derivace až do řádu k stejné jako funkce f. V tomto smyslu je polynom T_x^k aproximací funkce f v okolí bodu x.

Taylorův polynom je tímto požadavkem definován jednoznačně a má tvar (odvoďte!)

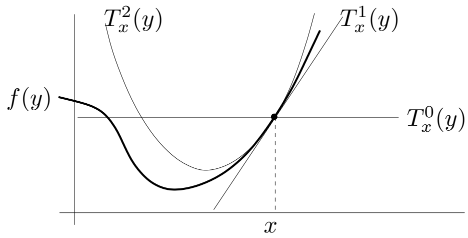
$$T_x^k(y) = \sum_{i=0}^{k} \frac{1}{i!} f^{(i)}(x)\,(y-x)^i, \tag{8.22}$$

[10]Podobně jako u Jacobiho matice, v optimalizaci se Hessova matice často nazývá *hessián*, což ale jinde označuje determinant Hessovy matice.

kde $f^{(i)}$ označuje i-tou derivaci funkce f (kde nultá derivace je funkce sama, $f^{(0)} = f$) a kde klademe $0! = 1$. Tvary polynomu až do stupně dva:

$$\begin{aligned}
T_x^0(y) &= f(x), \\
T_x^1(y) &= f(x) + f'(x)\,(y - x), \\
T_x^2(y) &= f(x) + f'(x)\,(y - x) + \tfrac{1}{2} f''(x)\,(y - x)^2.
\end{aligned}$$

Taylorův polynom nultého stupně T_x^0 je hodně špatná aproximace, rovná jednoduše konstantní funkci. Polynom prvního stupně T_x^1 je vlastně afinní funkce g v (8.4). Polynom druhého stupně T_x^2 je parabola, která má s funkcí f v bodě x společnou hodnotu a první dvě derivace. Viz obrázek:



Jak zobecnit Taylorův polynom pro funkci více proměnných $f\colon \mathbb{R}^n \to \mathbb{R}$? Definujme Taylorův polynom k-tého stupně (funkce f v okolí bodu $\mathbf{x}$) jako polynom $T_{\mathbf{x}}^k\colon \mathbb{R}^n \to \mathbb{R}$, který má v bodě $\mathbf{x}$ všechny parciální derivace až do řádu k stejné jako funkce f. Nebudeme uvádět vzorec pro polynom libovolného stupně, budou nám stačit jen polynomy do stupně dva:

$$T_{\mathbf{x}}^0(\mathbf{y}) = f(\mathbf{x}), \tag{8.23a}$$

$$T_{\mathbf{x}}^1(\mathbf{y}) = f(\mathbf{x}) + f'(\mathbf{x})\,(\mathbf{y} - \mathbf{x}), \tag{8.23b}$$

$$T_{\mathbf{x}}^2(\mathbf{y}) = f(\mathbf{x}) + f'(\mathbf{x})\,(\mathbf{y} - \mathbf{x}) + \tfrac{1}{2}(\mathbf{y} - \mathbf{x})^T f''(\mathbf{x})\,(\mathbf{y} - \mathbf{x}). \tag{8.23c}$$

Zde $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $f'(\mathbf{x}) \in \mathbb{R}^{1\times n}$ je Jacobiho matice (řádkový vektor) a $f''(\mathbf{x}) \in \mathbb{R}^{n\times n}$ je Hessova matice. Funkce (8.23b) je afinní a funkce (8.23c) je kvadratická.

Taylorův polynom lze zobecnit na zobrazení $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$ tak, že vezmeme Taylorovy polynomy všech složek $f_1, \dots, f_m$. Polynom prvního stupně tak vede na zobrazení

$$\mathbf{T}_{\mathbf{x}}^1(\mathbf{y}) = \mathbf{f}(\mathbf{x}) + \mathbf{f}'(\mathbf{x})\,(\mathbf{y} - \mathbf{x}), \tag{8.24}$$

což je vlastně afinní zobrazení $\mathbf{g}$ v (8.7). Polynom druhého stupně vede na zobrazení $\mathbf{T}_{\mathbf{x}}^2$, jehož složky jsou funkce (8.23c). To nejde napsat v maticové formě, protože všech $m \times n \times n$ druhých parciálních derivací se 'nevejde' do matice.

Příklad 8.18. Najděte Taylorův polynom druhého stupně funkce $f(x, y) = x^{-1} + y^{-1} + xy$ v bodě $(x_0, y_0) = (2, 1)$. Máme

$$\begin{aligned}
f(x_0, y_0) &= x^{-1} + y^{-1} + xy\big|_{(x,y)=(2,1)} = \frac{7}{2}, \\
f'(x_0, y_0) &= \begin{bmatrix} y - x^{-2} & x - y^{-2} \end{bmatrix}\Big|_{(x,y)=(2,1)} = \begin{bmatrix} \frac{3}{4} & 1 \end{bmatrix}, \\
f''(x_0, y_0) &= \begin{bmatrix} 2x^{-3} & 1 \\ 1 & 2y^{-3} \end{bmatrix}\Bigg|_{(x,y)=(2,1)} = \begin{bmatrix} \frac{1}{4} & 1 \\ 1 & 2 \end{bmatrix}.
\end{aligned}$$

Dle (8.23c) je (pozor, naše proměnné jsou označené jinak než v (8.23c))

$$
\begin{aligned}
T^2_{(x_0,y_0)}(x,y) &= f(x_0,y_0) + f'(x_0,y_0)\begin{bmatrix} x-x_0 \\ y-y_0 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} x-x_0 \\ y-y_0 \end{bmatrix}^T f''(x_0,y_0)\begin{bmatrix} x-x_0 \\ y-y_0 \end{bmatrix} \\
&= \frac{7}{2} + \begin{bmatrix} \frac{3}{4} & 1 \end{bmatrix}\begin{bmatrix} x-2 \\ y-1 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} x-2 \\ y-1 \end{bmatrix}^T \begin{bmatrix} \frac{1}{4} & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} x-2 \\ y-1 \end{bmatrix} \\
&= \tfrac{1}{8}x^2 + xy + y^2 - \tfrac{3}{4}x - 3y + \tfrac{9}{2}.
\end{aligned}
$$
♦

8.10 Cvičení

8.1. Načrtněte několik vrstevnic (připište k nim výšky) těchto funkcí dvou proměných:

a) $f(x_1,x_2) = 2x_1 + x_2$
b) $f(x_1,x_2) = x_1 - 3x_2 + 1$
c) $f(x_1,x_2) = x_1^2$
d) $f(x_1,x_2) = x_1^2 + 4x_2^2$
e) $f(x_1,x_2) = x_1^2 - x_2^2 = (x_1+x_2)(x_1-x_2)$
f) $f(x_1,x_2) = x_1x_2$
g) $f(x_1,x_2) = \sqrt{x_1^2+x_2^2}$ (tj. eukleidovská norma $f(\mathbf{x}) = \|\mathbf{x}\|$ pro $n=2$)

8.2. Najděte parametrizaci válce. Přesněji, najděte zobrazení $\mathbf{f}\colon \mathbb{R}^2 \to \mathbb{R}^3$ tak, aby množina $\mathbf{f}([0,2\pi)\times\mathbb{R})$ byla povrch nekonečného válce o poloměru r bez podstav.

8.3. Máme funkci $f\colon \mathbb{R}^2 \to \mathbb{R}$ s hodnotami $f(x,y) = \ln(1+xy)$. Máme bod $(x_0,y_0) = (1,2)$.

a) Je funkce f v bodě (x_0,y_0) spojitá?
b) Je funkce f v bodě (x_0,y_0) spojitě diferencovatelná?
c) Je funkce f v bodě (x_0,y_0) diferencovatelná?
d) Najděte totální derivaci (Jacobiho matici) $f'(x,y)$ funkce f v bodě (x_0,y_0).
e) Najděte gradient $\nabla f(x,y)$ funkce f v bodě (x_0,y_0).
f) Najděte řez a směrovou derivaci funkce f v bodě (x_0,y_0) ve směru $(1,-1)$.
g) Najděte Hessovu matici funkce f v bodě (x_0,y_0).

8.4. Funkce $f\colon \mathbb{R}^2 \to \mathbb{R}$ má hodnoty $f(x,y) = \max\{x,y\}$. Ve kterých bodech je tato funkce spojitě diferencovatelná? Odpověď odůvodněte.

8.5. Najděte totální derivaci (Jacobiho matici) zobrazení 1 až 18 (u kterých je to možné) z Příkladu 8.3.

8.6. Nechť $f(x,y)$ je diferencovatelná funkce dvou proměnných.

a) Spočítej derivaci f podle polárních souřadnic (φ,r), kde $x = r\cos\varphi$, $y = r\sin\varphi$.
b) Bod (x,y) se v čase t pohybuje po křivce dané rovnicí $(x,y) = (t^2+2t, \ln(t^2+1))$. Najděte derivaci f podle času.

8.7. Nechť $f(x,y)$ je diferencovatelá funkce dvou proměnných. Spočítej derivaci funkce $g(\mathbf{u}) = f(\mathbf{a}^T\mathbf{u}, \|\mathbf{u}\|)$ podle vektoru $\mathbf{u}$.

8.8. Odvoď co nejjednodušší vzorec pro totální derivaci (Jacobiho matici) těchto zobrazení. Kde je to možné, odvoď nejdříve přímým výpočtem a pak řetízkovým pravidlem.

a) $f(\mathbf{x}) = \mathbf{x}^T\mathbf{x} = \|\mathbf{x}\|^2$
b) $f(\mathbf{x}) = \|\mathbf{x}\| = \sqrt{\mathbf{x}^T\mathbf{x}}$
c) $f(\mathbf{x}) = \mathbf{g}(\mathbf{x})^T\mathbf{h}(\mathbf{x})$ (kde $\mathbf{g}, \mathbf{h}\colon \mathbb{R}^n \to \mathbb{R}^m$ jsou dána)
d) $f(\mathbf{x}) = \|\mathbf{g}(\mathbf{x})\|$
e) $f(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T\mathbf{y}$ (kde tedy $f\colon \mathbb{R}^{2n} \to \mathbb{R}$ a derivujeme podle vektoru $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{2n}$)

8.9. Odvoď totální derivaci kvadratické formy, místo postupu z Příkladu 8.11 ale použij vzorec $(\mathbf{g}(\mathbf{x})^T\mathbf{h}(\mathbf{x}))' = \mathbf{g}(\mathbf{x})^T\mathbf{h}'(\mathbf{x}) + \mathbf{h}(\mathbf{x})^T\mathbf{g}'(\mathbf{x})$.

8.10. Nadmořská výška krajiny je dána vzorcem $h(d, s) = 2s^2 + 3sd - d^2 + 5$, kde d je zeměpisná délka (zvětšuje se od západu k východu) a s je zeměpisná šířka (zvětšuje se od jihu k severu). V bodě $(d, s) = (-1, 1)$ určete (a) směr nejstrmějšího stoupání terénu, (b) strmost terénu v jihovýchodním směru.

8.11. Dokažte, že je-li funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ diferencovatelná v bodě $\mathbf{x} \in \mathbb{R}^n$ ve směru $\mathbf{v} \in \mathbb{R}^n$, pak pro každé $\alpha \geq 0$ platí $\mathbf{f}_{\alpha\mathbf{v}}(\mathbf{x}) = \alpha\mathbf{f}_{\mathbf{v}}(\mathbf{x})$. Jestliže funkce není totálně diferencovatelná, ukažte, že toto nemusí platit pro $\alpha < 0$.

8.12. Spočítejte druhou derivaci $f''(\mathbf{x})$ (Hessovu matici) těchto funkcí:
a) $f(x, y) = \mathrm{e}^{-x^2-y^2}$
b) $f(x, y) = \ln(\mathrm{e}^x + \mathrm{e}^y)$
c) $f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x}$ (matice $\mathbf{A}$ je dána, nemusí být symetrická)
d) $f(\mathbf{x}) = \mathbf{g}(\mathbf{x})^T\mathbf{g}(\mathbf{x})$ (zobrazení $\mathbf{g}\colon \mathbb{R}^n \to \mathbb{R}^m$ je dáno)
e) $f(\mathbf{x}) = \|\mathbf{x}\|$

8.13. Je dána funkce $f(x, y) = 6xy^2 - 2x^3 - 3y^3$. V bodě $(x_0, y_0) = (1, -2)$ najděte Taylorův polynom nultého, prvního a druhého stupně.

8.14. Čemu je roven Taylorův polynom prvního stupně polynomu prvního stupně (tj. afinní funkce)? A čemu je roven Taylorův polynom druhého stupně polynomu druhého stupně (tj. kvadratické funkce)? Proč?

8.15. *Metoda konečných diferencí* počítá derivaci funkce přibližně jako

$$f'(x) \approx \frac{f(x + \alpha) - f(x)}{\alpha},$$

kde α je malé číslo (dobrá volba je $h = \sqrt{\varepsilon}$, kde ε je strojová přesnost). Toto jde použít i na parciální derivace. Vymyslete si dvě zobrazení $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$ a $\mathbf{g}\colon \mathbb{R}^m \to \mathbb{R}^l$ pro nějaké navzájem různé dimenze $n, m, l > 1$. Zvolte bod $\mathbf{x} \in \mathbb{R}^n$. Spočítejte přibližně totální derivace (Jacobiho matice) $\mathbf{f}'(\mathbf{x})$ a $\mathbf{g}'(\mathbf{f}(\mathbf{x}))$ v Matlabu metodou konečných diferencí. Potom spočítejte derivaci složeného zobrazení $(\mathbf{g} \circ \mathbf{f})'(\mathbf{x})$ jednak metodou konečných diferencí a jednak vynásobením matic $\mathbf{f}'(\mathbf{x})$ a $\mathbf{g}'(\mathbf{f}(\mathbf{x}))$. Porovnejte.

8.16. ($\star$) Nechť zobrazení $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$ je definováno výrazem $\mathbf{f}(\mathbf{x})$ obsahujícím vektor $\mathbf{x}$, konstantní matice a vektory, a operace součet matic, maticový součin a transpozice (např. $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{x}\mathbf{b}^T\mathbf{x}$). Zamyslete se nad tím, zda lze derivaci $\mathbf{f}'(\mathbf{x})$ vždy vyjádřit výrazem obsahujícím vektor $\mathbf{x}$, ty samé konstantní matice a vektory, a operace součet matic, maticový součin a transpozice?

8.17. ($\star$) Najděte derivaci funkce $f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x}$ a derivaci funkce $g(\mathbf{A}) = \mathbf{x}^T\mathbf{A}\mathbf{x}$.

8.18. Najděte první derivaci (Jacobiho matici) a druhou derivaci (Hessovu matici) funkce $f(\mathbf{x}) = \|\mathbf{x}\|$. Dále najděte Taylorův polynom prvního a druhého stupně funkce f v okolí bodu $\mathbf{x} \neq \mathbf{0}$.

8.19. ($\star$) Nechť je funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ v nějakém bodě a směru lineární, tedy $f(\mathbf{x} + \alpha\mathbf{v}) = f(\mathbf{x}) + \alpha f(\mathbf{v})$ pro nějaké $\mathbf{x}, \mathbf{v} \in \mathbb{R}^n$ a pro všechna $\alpha \in \mathbb{R}$. Lze z toho usoudit, že Hessova matice $f''(\mathbf{x})$ této funkce v bodě $\mathbf{x}$ je singulární?

Nápověda a řešení

8.2. $\mathbf{f}(\varphi, h) = (r\cos\varphi,\ r\sin\varphi,\ h)$

8.3.a) Ano, dle Věty 8.1.

8.3.b) Ano, protože její parciální derivace $\partial f(x,y)/\partial x = y/(1+xy)$ a $\partial f(x,y)/\partial y = x/(1+xy)$ jsou (dle Věty 8.1) spojité funkce.

8.3.c) Ano, dle Věty 8.3.

8.3.d) $f'(x,y) = \begin{bmatrix} y/(1+xy) & x/(1+xy) \end{bmatrix} \in \mathbb{R}^{1\times 2}$, tedy $f'(x_0,y_0) = \begin{bmatrix} 2 & 1 \end{bmatrix}/3$.

8.3.e) $\nabla f(x,y) = [f'(x,y)]^T = \begin{bmatrix} y/(1+xy) \\ x/(1+xy) \end{bmatrix} \in \mathbb{R}^{2\times 1}$, tedy $\nabla f(x_0,y_0) = (2,1)/3 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}/3$.

8.3.f) Řez je $\varphi(\alpha) = \ln[1 + (1+\alpha)(2-\alpha)]$. Směrová derivace je $f_{(1,-1)}(x_0,y_0) = 1/3$.

8.4. Je spoj. diferencovatelná na množině $\mathbb{R}^2 \setminus \{\, (x,y) \mid x = y \,\}$.

8.6.a) $f_\varphi(x,y) = -f_x(x,y) r\sin\varphi + f_y(x,y) r\cos\varphi$, $f_r(x,y) = f_x(x,y)\cos\varphi + f_y(x,y)\sin\varphi$

8.6.b) $f_t(x,y) = 2f_x(x,y)(t+1) + 2tf_y(x,y)/(t^2+1)$

8.8.e) Nejjednodušší je elementární postup. Výsledek bude řádkový vektor, jehož prvky budou parciální derivace výrazu $\mathbf{x}^T\mathbf{y} = x_1y_1 + \cdots + x_ny_n$ podle $x_1, \ldots, x_n, y_i, \ldots, y_n$. Tedy $f'(\mathbf{x},\mathbf{y}) = \begin{bmatrix} \mathbf{y}^T & \mathbf{x}^T \end{bmatrix}$.

Druhá možnost je uvědomit si, že $\mathbf{x}^T\mathbf{y} = \mathbf{z}^T\mathbf{A}\mathbf{z}$ kde $\mathbf{z} = (\mathbf{x},\mathbf{y})$ a $\mathbf{A} = \frac{1}{2}\begin{bmatrix} \mathbf{0} & \mathbf{I} \\ \mathbf{I} & \mathbf{0} \end{bmatrix}$ a tedy $f'(\mathbf{x},\mathbf{y}) = f'(\mathbf{z}) = 2\mathbf{z}^T\mathbf{A} = 2\begin{bmatrix} \mathbf{x}^T & \mathbf{y}^T \end{bmatrix}\mathbf{A} = \begin{bmatrix} \mathbf{y}^T & \mathbf{x}^T \end{bmatrix}$.

8.9. Dosaď $\mathbf{g}(\mathbf{x}) = \mathbf{x}$ a $\mathbf{h}(\mathbf{x}) = \mathbf{A}\mathbf{x}$.

8.10. (a) $(5,1)/\sqrt{26}$, (b) $(5,1)^T(1,-1)/\sqrt{2} = 2\sqrt{2}$

8.11. pro $\alpha \geq 0$ dokažte přímo z definice (8.19). Pro $\alpha < 0$ viz Příklad 8.16.

8.12.a) $2e^{-x^2-y^2}\begin{bmatrix} 2x^2-1 & 2xy \\ 2xy & 2y^2-1 \end{bmatrix}$

8.12.b) $\frac{e^{x+y}}{(e^x+e^y)^2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$

8.12.c) $f''(\mathbf{x}) = \mathbf{A} + \mathbf{A}^T$

8.12.d) $f''(\mathbf{x}) = 2\mathbf{g}'(\mathbf{x})^T\mathbf{g}'(\mathbf{x}) + 2\sum_{i=1}^{m} g_i(\mathbf{x})g_i''(\mathbf{x})$

8.13. $T^0_{(1,-2)}(x,y) = 46$, $T^1_{(1,-2)}(x,y) = 18x - 60y - 92$, $T^2_{(1,-2)}(x,y) = -6x^2 - 24xy - 18x + 24y^2 + 60y + 46$

8.14. V obou případech je roven původnímu polynomu. Ty přece určitě mají nulté a prvé (a příp. druhé) derivace stejné jako původní polynom.

8.17. Ve funkci f je matice $\mathbf{A}$ konstanta a derivujeme podle vektoru $\mathbf{x}$. Derivace je $f'(\mathbf{x}) = (\mathbf{A} + \mathbf{A}^T)\mathbf{x}^T$. Ve funkci g je vektor $\mathbf{x}$ konstanta a derivujeme podle matice $\mathbf{A}$. Derivace je $g'(\mathbf{A}) = \mathbf{x}\mathbf{x}^T$ (viz poslední tabulka v §8.5.2).

8.18. První derivaci jsme odvodili dříve, $f'(\mathbf{x}) = \mathbf{x}^T/\|\mathbf{x}\|$ (řádkový vektor).

i-tá složka Jacobiho matice je $f'(\mathbf{x})_i = x_i/\|\mathbf{x}\|$. Prvek (i,j) Hessovy matice je parciální derivace tohoto výrazu podle x_j, což je (s použitím derivace součinu a řetízkového pravidla) $f''(\mathbf{x})_{ij} = \delta_{ij}/\|\mathbf{x}\| - x_i x_j/\|\mathbf{x}\|^3$. Maticově hessián zapíšeme jako $f''(\mathbf{x}) = \mathbf{I}/\|\mathbf{x}\| - \mathbf{x}\mathbf{x}^T/\|\mathbf{x}\|^3$.

Taylorův polynom prvního stupně je $T^1_{\mathbf{x}}(\mathbf{y}) = f(\mathbf{x}) + f'(\mathbf{x})(\mathbf{y}-\mathbf{x}) = \|\mathbf{x}\| + \mathbf{x}^T(\mathbf{y}-\mathbf{x})/\|\mathbf{x}| = \mathbf{x}^T\mathbf{y}$.

Taylorův polynom druhého stupně $T^2_{\mathbf{x}}(\mathbf{y}) = T^1_{\mathbf{x}}(\mathbf{y}) + \frac{1}{2}(\mathbf{y}-\mathbf{x})^T f''(\mathbf{x})(\mathbf{y}-\mathbf{x})$ vyjde jako dlouhý výraz který, zdá se, nejde příliš zjednodušit.

Kapitola 9

Extrémy funkce na množině

Extrémy funkce na množině jsme už zmínili v §1.2, zde si o nich řekneme více. Kapitola je opakování některých základních věcí z matematické analýzy a něco navíc.

9.1 Vnitřek a hranice množiny

Označme jako

$$B_\varepsilon(\mathbf{x}) = \{\, \mathbf{y} \in \mathbb{R}^n \mid \|\mathbf{x} - \mathbf{y}\| < \varepsilon \,\} \tag{9.1}$$

n-rozměrnou kouli bez hranice[1] se středem $\mathbf{x} \in \mathbb{R}^n$ a poloměrem $\varepsilon > 0$. Všimněte si, že pro speciální případ $n = 1$ je množina (9.1) interval $B_\varepsilon(x) = (x - \varepsilon, x + \varepsilon)$.

Mějme nyní množinu $X \subseteq \mathbb{R}^n$. Bod $\mathbf{x} \in \mathbb{R}^n$ se nazývá její

- **vnitřní bod**, jestliže existuje $\varepsilon > 0$ takové, že $B_\varepsilon(\mathbf{x}) \subseteq X$,
- **hraniční bod**, jestliže pro každé $\varepsilon > 0$ je $B_\varepsilon(\mathbf{x}) \cap X \neq \emptyset$ a $B_\varepsilon(\mathbf{x}) \cap (\mathbb{R}^n \setminus X) \neq \emptyset$.

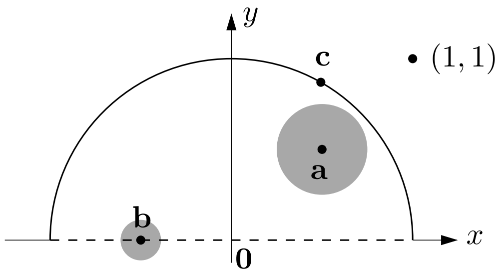
Všimněte si, že hraniční bod množiny nemusí patřit do této množiny. Pokud leží bod v množině, je buď vnitřní nebo hraniční, ale ne obojí najednou (dokažte!). **Vnitřek** [**hranice**] množiny je množina všech jejích vnitřních [hraničních] bodů.

Množina se nazývá **otevřená**, jestliže všechny její body jsou vnitřní. Množina se nazývá **uzavřená**, jestliže obsahuje každý svůj hraniční bod. Lze dokázat, že množina X je uzavřená [otevřená], právě když její doplněk $\mathbb{R}^n \setminus X$ je otevřený [uzavřený]. Otevřenost a uzavřenost se nevylučují: množiny $\emptyset$ a $\mathbb{R}^n$ jsou zároveň otevřené i uzavřené. Naopak, některé množiny nejsou ani otevřené ani uzavřené, např. interval $(0, 1]$.

Množina X je **omezená**, jestliže existuje $r \in \mathbb{R}$ takové, že $\|\mathbf{x} - \mathbf{y}\| < r$ pro všechna $\mathbf{x}, \mathbf{y} \in X$. Jinými slovy, množina se 'vejde' do koule konečného průměru.

Příklad 9.1. Máme množinu $\{\, (x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1,\ y > 0 \,\} \cup \{(1, 1)\}$ na obrázku:

[1] Mohli bychom použít i kouli s hranicí $\{\, \mathbf{y} \in \mathbb{R}^n \mid \|\mathbf{x} - \mathbf{y}\| \leq \varepsilon \,\}$. Podobně, norma v (9.1) může být eukleidovská, ale i libovolná jiná vektorová p-norma (viz §12.4.1). Vnitřek a hranice každé množiny by se tím nezměnila.



Bod **a** je vnitřní bod množiny, protože existuje koule (s nenulovým poloměrem!) se středem **a**, která celé leží v množině. Bod **b** je hraniční, protože každá koule se středem **b** má neprázdný průnik s množinou i s jejím doplňkem. Všimněte si, že **b** nepatří do množiny. Bod **a** není hraniční a bod **b** není vnitřní. Bod **c** není vnitřní, je hraniční a patří do množiny. Bod $(1,1)$ (který patří do množiny, viz její definice výše) je hraniční.

Množina není otevřená, protože např. bod **c** není vnitřní. Není ani uzavřená, protože např. bod **b** je hraniční, ale nepatří do množiny. Množina je omezená. ♦

Na nalezení vnitřku a hranice množiny obecně neexistuje mechanický postup (ten by navíc závisel na způsobu, jakým je množina popsána). Často je lze ale uhodnout na základě geometrické představy (množinu si načrtneme/představíme) a pak ověřit správnost podle definice.

Příklad 9.2. Bod $\frac{1}{2}$ je vnitřní bod polouzavřeného intervalu $(0,1] \subseteq \mathbb{R}$ a body 0 a 1 jsou hraniční. Vnitřek intervalu $(0,1]$ je tedy otevřený interval $(0,1)$ a hranice je množina $\{0,1\}$. ♦

Příklad 9.3. Množina $[0,1] \times \{1\} = \{\, (x,y) \mid 0 \le x \le 1,\ y = 1 \,\} \subseteq \mathbb{R}^2$ (úsečka v rovině) nemá žádné vnitřní body, tedy její vnitřek je prázdná množina. Všechny její body jsou hraniční, je tedy sama svou vlastní hranicí. Není otevřená, je uzavřená, je omezená. ♦

Všimněte si v minulém příkladu, že nejde říct, zda 'úsečka' např. má nebo nemá vnitřní body – záleží totiž na tom, v jakém prostoru je 'umístěna'. Úsečka v prostoru $\mathbb{R}$ (tedy interval) vnitřní body má, úsečka v prostoru $\mathbb{R}^2$ žádné nemá.

Příklad 9.4. Kružnice v rovině $\{\, (x,y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1 \,\}$ nemá žádné vnitřní body, všechny její body jsou hraniční. Podobně pro n-rozměrnou sféru $\{\, \mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\| = 1 \,\}$. ♦

Příklad 9.5. Mějme kruh bez hranice $\{\, (x,y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1 \,\}$. Všechny jeho body jsou vnitřní, takže je otevřený. Jeho hranice je kružnice $\{\, (x,y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1 \,\}$. Podobně pro n-rozměrnou kouli bez hranice $\{\, \mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\| < 1 \,\}$. ♦

9.2 Existence globálních extrémů

Zopakujme (§1.2), že funkce $f \colon \mathbb{R}^n \to \mathbb{R}$ nabývá na množině $X \subseteq \mathbb{R}^n$ v bodě $\mathbf{x} \in X$ svého

- *minima*, jestliže $f(\mathbf{x}) \le f(\mathbf{x}')$ pro všechna $\mathbf{x}' \in X$,
- *ostrého minima*, jestliže $f(\mathbf{x}) < f(\mathbf{x}')$ pro všechna $\mathbf{x}' \in X \setminus \{\mathbf{x}\}$.

Pro odlišení od lokálních minim (viz dále) se tato 'obyčejná' minima často nazývaji *globální minima*. *Hodnota minima* je pak číslo $f(\mathbf{x}) = \min_{\mathbf{x}' \in X} f(\mathbf{x}')$. Tedy hodnota minima je nejmenší prvek množiny (promyslete!)

$$f(X) = \{\, f(\mathbf{x}) \mid \mathbf{x} \in X \,\} \subseteq \mathbb{R},$$

která je obrazem množiny X v zobrazení f (viz §1.1.2). Funkce na množině nemusí mít minimum, neboť množina $f(X)$ nemusí mít nejmenší prvek.

Pro maxima jsou definice analogické, minima a maxima se dohromady nazývají *extrémy*.

Příklad 9.6. Funkce $f(x) = x$ nemá na množině $(0,1) \subseteq \mathbb{R}$ (otevřený interval) minimum, neboť množina $f((0,1)) = (0,1)$ nemá nejmenší prvek. ♦

Příklad 9.7. Funkce $f(x,y) = x + y$ nemá na množině $X = \{\, (x,y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1 \,\}$ minimum, neboť množina $f(X) = (-\sqrt{2}, \sqrt{2})$ nemá nejmenší prvek. ♦

Uvedeme nyní (bez důkazu) následující zásadní skutečnost:

Věta 9.1. Spojité zobrazení uzavřené omezené množiny je uzavřená omezená množina.

Tedy máme-li uzavřenou a omezenou množinu $X \subseteq \mathbb{R}^n$ a spojité zobrazení $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$, pak množina $\mathbf{f}(X) = \{\, \mathbf{f}(\mathbf{x}) \mid \mathbf{x} \in X \,\} \subseteq \mathbb{R}^m$ bude také uzavřená a omezená[2].

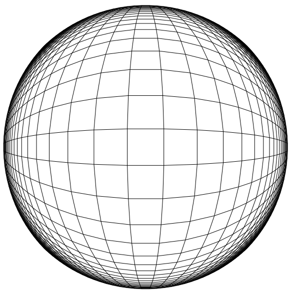
Mohlo by se zdát, že spojité zobrazení bude zachovávat uzavřenost bez omezenosti nebo omezenost bez uzavřenosti. Je snadné najít protipříklady.

Příklad 9.8. Nechť X je interval $[1, +\infty) \subseteq \mathbb{R}$. Tato množina je uzavřená a není omezená. Spojité zobrazení $f(x) = 1/x$ zobrazí tuto množinu na množinu $f(X) = (0,1]$, která není uzavřená a je omezená. ♦

Příklad 9.9. Uvažujme spojité zobrazení $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^n$ s hodnotami

$$\mathbf{f}(\mathbf{x}) = \frac{\mathbf{x}}{\sqrt{1 + \mathbf{x}^T\mathbf{x}}}. \tag{9.2}$$

Množina $\mathbb{R}^n$ je uzavřená, otevřená a neomezená. Množina $\mathbf{f}(\mathbb{R}^n)$ je otevřená a omezená: je to jednotková koule (bez hranice). Pro ilustraci je na obrázku množina $\mathbf{f}(X)$ pro $X = (\mathbb{R} \times \mathbb{Z}) \cup (\mathbb{Z} \times \mathbb{R}) \subseteq \mathbb{R}^2$ (tedy X je mřížka v rovině, rozmyslete!):



♦

Věta 9.1 má důležitý důsledek pro optimalizaci, který je znám jako *věta o extrémní hodnotě* nebo *Weierstrassova věta*.

Důsledek 9.2 (Weierstrassova věta). Spojitá funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ nabývá na uzavřené omezené množině $X \subseteq \mathbb{R}^n$ svého minima.

[2]Množinám, které jsou zároveň uzavřené a omezené, se říká také *kompaktní*.

Důkaz. Pro spojitou funkci $f\colon \mathbb{R}^n \to \mathbb{R}$ je obraz uzavřené omezené množiny $X \subseteq \mathbb{R}^n$ uzavřená omezená množina $f(X) \subseteq \mathbb{R}$. Snadno nyní dokážeme, že každá uzavřená omezená podmnožina množiny $\mathbb{R}$ má nejmenší prvek.

Nechť tedy $Y \subseteq \mathbb{R}$ je uzavřená a omezená. Vždy existuje její infimum, označme ho $a = \inf Y$. Protože Y je omezená, je $a \in \mathbb{R}$, tj. $a > -\infty$. Ukážeme, že $a \in Y$, tj. a je nejmenší bod množiny Y. Protože Y je uzavřená a tedy každý její hraniční bod do ní patří, stačí k tomu stačí dokázat, že a je hraniční bod množiny Y.

Protože Y je uzavřená, každý její hraniční bod do ní patří, tedy $a \in Y$. Z toho plyne, že a je nejmenší bod množiny Y. ∎

Zdůrazněme, že Weierstrassova věta udává pouze postačující (avšak ne nutné) podmínky pro existenci minima funkce na množině. Např. funkce $f(x) = x^2$ má na množině $\mathbb{R}$ minimum, i když množina $\mathbb{R}$ není omezená.

9.3 Lokální extrémy

Funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ nabývá na množině $X \subseteq \mathbb{R}^n$ v bodě $\mathbf{x} \in X$ svého **lokálního minima**, jestliže existuje $\varepsilon > 0$ tak, že funkce f nabývá na množině $X \cap B_\varepsilon(\mathbf{x})$ svého (globálního) minima (tuto definici promyslete, je důležitá!). Analogicky můžeme definovat ostrá lokální minima. Maximum, lokální maximum a jejich ostré verze se definují obdobně. Každé minimum funkce f je zároveň lokální minimum funkce f (proč?), naopak to ale obecně neplatí.

(Můžeme přirozeně mluvit i o lokálních minimech/maximech/optimech optimalizační úlohy ve standardním tvaru (1.11), což jsou lokální minima/maxima účelové funkce f úlohy na množině (1.13) jejích přípustných řešení.)

Každý bod množiny $X \subseteq \mathbb{R}^n$ je buď vnitřní nebo hraniční. Extrém $\mathbf{x}$ (globální či lokální) funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ na množině X se nazývá **volný**, když $\mathbf{x}$ je vnitřní bod množiny X, a **vázaný** (množinou X), když $\mathbf{x}$ je hraniční bod množiny X.

Tvrzení 9.3. Nechť $f\colon \mathbb{R}^n \to \mathbb{R}$ a nechť $\mathbf{x}$ je *vnitřní* bod množiny $X \subseteq \mathbb{R}^n$. Následující výroky jsou ekvivalentní:

- $\mathbf{x}$ je lokální minimum [maximum] funkce f na množině X.
- $\mathbf{x}$ je lokální minimum [maximum] funkce f na množině $\mathbb{R}^n$.

Důkaz. Nechť $\mathbf{x} \in X$ je lokální minimum f na $\mathbb{R}^n$, tedy (globální) minimum f na $B_\varepsilon(\mathbf{x})$. Pak ovšem je $\mathbf{x}$ také (globální) minimum f na množině $X \cap B_\varepsilon(\mathbf{x})$, tedy lokální minimum f na X.

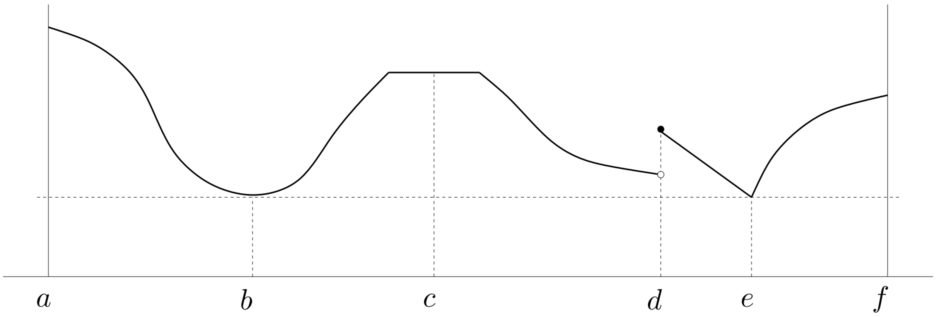
Nechť $\mathbf{x}$ je lokální minimum f na X, tedy (globální) minimum f na množině $X \cap B_\varepsilon(\mathbf{x})$. Protože $\mathbf{x}$ je vnitřní bod množiny X, můžeme ε zmenšit tak, že $B_\varepsilon(\mathbf{x}) \subseteq X$. Pak $\mathbf{x}$ bude stále (globální) minimum f na množině $X \cap B_\varepsilon(\mathbf{x}) = B_\varepsilon(\mathbf{x})$, a tedy lokální minimum f na $\mathbb{R}^n$.

Důkaz zjevně zůstane v platnosti, nahradíme-li minima maximy. ∎

Tvrzení 9.3 ukazuje, že není velký rozdíl mezi volnými lokálními extrémy na množině X a lokálními extrémy na $\mathbb{R}^n$. Říká, že bod je volný lokální extrém f na X, právě když je to lokální extrém f na $\mathbb{R}^n$ a zároveň vnitřní bod množiny X. Tedy abychom našli volné lokální extrémy f na X, stačí najít lokální extrémy f a zahodit ty, které nejsou vnitřními body množiny X.

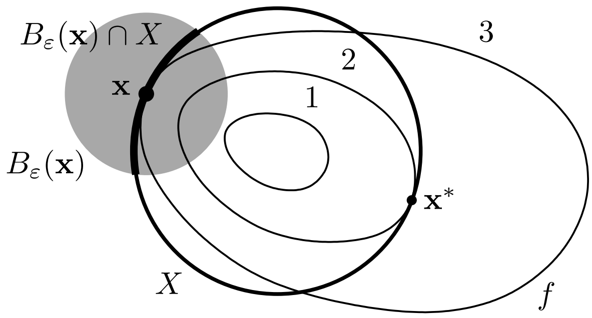
Může nastat speciální situace, kdy bod je lokální minimum [maximum] f na $\mathbb{R}^n$ a zároveň hraniční bod množiny X. Pak, dle definice výše, považujeme tento bod za vázané lokální minimum [maximum] f na X, i když ho množina X vlastně nijak neomezuje.

Příklad 9.10. Funkce jedné proměnné na obrázku nabývá na uzavřeném intervalu $[a, f] \subseteq \mathbb{R}$ v bodě a globálního (a tedy i lokálního) maxima, v bodech b, e globálního (a tedy i lokálního) minima, v bodě c lokálního maxima a zároveň lokálního minima, v bodě d lokálního maxima, v bodě f lokálního maxima. Extrémy v bodech a, b, d, e, f jsou ostré. Extrémy v bodech a, f jsou vázané, v bodech b, c, d, e jsou volné.



♦

Příklad 9.11. Nechť $X \subseteq \mathbb{R}^2$ je kružnice a funkce $f\colon \mathbb{R}^2 \to \mathbb{R}$ má vrstevnice jako na obrázku:



V bodě $\mathbf{x}^*$ nabývá funkce f na množině X globálního (a tedy i lokálního) minima, protože v žádném bodě na kružnici X nemá funkce menší hodnotu než $f(\mathbf{x}^*) = 2$. V bodě $\mathbf{x}$ nabývá funkce f na množině X lokálního minima, protože existuje $\varepsilon > 0$ tak, že funkce f nabývá na části kružnice $B_\varepsilon(\mathbf{x}) \cap X$ svého (globálního) minima. Oba extrémy jsou vázané, protože body $\mathbf{x}^*$ a $\mathbf{x}$ jsou hraniční body množiny X (množina X vnitřní body nemá, viz Příklad 9.4). ♦

Příklad 9.12. Funkce $f(\mathbf{x}) = x_1$ na množině $\{\, \mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\| \leq 1 \,\}$ má v bodě $(-1, 0, \ldots, 0)$ vázané globální (a tedy i lokální) minimum. ♦

Příklad 9.13. Libovolná funkce $f\colon \mathbb{R} \to \mathbb{R}$ má na množině $\mathbb{Z}$ (množina celých čísel) v libovolném bodě $x \in \mathbb{Z}$ lokální minimum i lokální maximum. ♦

9.4 Cvičení

9.1. Máme množiny $X = [-1, 1] \times \{0\}$ a $Y = [-1, 1] \times [-1, 1]$. Načrtněte následující množiny:

a) $\left\{\, \mathbf{x} \in \mathbb{R}^2 \;\middle|\; 1 \geq \min_{\mathbf{y} \in X} \|\mathbf{x} - \mathbf{y}\| \,\right\}$

b) $\left\{\, \mathbf{x} \in \mathbb{R}^2 \;\middle|\; 2 \geq \max_{\mathbf{y} \in X} \|\mathbf{x} - \mathbf{y}\| \,\right\}$

c) vrstevnice výšky 1 funkce $f(\mathbf{x}) = \min_{\mathbf{y} \in Y} \|\mathbf{x} - \mathbf{y}\|$

d) vrstevnice výšky $\sqrt{2}$ funkce $f(\mathbf{x}) = \max_{\mathbf{y} \in Y} \|\mathbf{x} - \mathbf{y}\|$

9.2. Načrtněte následující množiny (jedná se o podmnožiny $\mathbb{R}^2$ nebo $\mathbb{R}^3$):

a) $[-1,0] \times \{1\}$
b) $\mathbb{Z} \times \mathbb{Z}$
c) $\mathbb{R} \times \mathbb{Z}$
d) $(\mathbb{R} \times \mathbb{Z}) \cup (\mathbb{Z} \times \mathbb{R})$
e) $\{ (x,y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1 \} \times \mathbb{R}$
f) $\{ (x,y) \in \mathbb{R}^2 \mid x > 0,\ y > 0,\ xy = 1 \}$
g) $\{ (x,y) \in \mathbb{R}^2 \mid \min\{x,y\} = 1 \}$

9.3. Co je vnitřek a hranice dané množiny? Je množina omezená?

a) množina reálných čísel $\mathbb{R}$
b) uzavřený interval $[a,b] = \{ x \in \mathbb{R} \mid a \leq x \leq b \}$ (kde $a < b$)
c) množina racionálních čísel $\mathbb{Q}$
d) množina (9.1)
e) $\{ (x,y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1,\ y \geq 0 \}$
f) $\{ (x,y) \in \mathbb{R}^2 \mid y = x^2,\ -1 < x \leq 1 \}$
g) $\{ (x,y) \in \mathbb{R}^2 \mid xy < 1,\ x > 0,\ y > 0 \}$
h) $\{ \mathbf{x} \in \mathbb{R}^n \mid \max_{i=1}^n x_i \leq 1 \}$
i) $\{ \mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} = \mathbf{b} \}$ (afinní podprostor $\mathbb{R}^n$)
j) $\{ \mathbf{x} \in \mathbb{R}^n \mid b \leq \mathbf{a}^T\mathbf{x} \leq c \}$, kde $\mathbf{a} \in \mathbb{R}^n$, $b, c \in \mathbb{R}$ (panel, angl. *slab*)
k) ($\star$) $\{ \mathbf{X} \in \mathbb{R}^{n\times n} \mid \det \mathbf{X} = 0 \}$
l) ($\star$) $\{ \mathbf{X} \in \mathbb{R}^{n\times n} \mid \mathbf{X}^T\mathbf{X} = \mathbf{I} \}$

9.4. Každá z následujících množin je sjednocením konečného počtu (otevřených, uzavřených či polouzavřených) intervalů. Najděte tyto intervaly. Příklad: $\{ x^2 \mid x \in \mathbb{R} \} = [0, +\infty)$.

a) $\{ 1/x \mid x \geq 1 \}$
b) $\{ 1/x \mid |x| \geq 1 \}$
c) $\{ \mathrm{e}^{-x^2} \mid x \in \mathbb{R} \}$
d) $\{ x + y \mid x^2 + y^2 < 1 \}$
e) $\{ x + y \mid x^2 + y^2 = 1 \}$
f) $\{ x - y \mid x^2 + y^2 = 1 \}$
g) $\{ |x| + |y| \mid x^2 + y^2 = 1 \}$
h) $\{ |x - y| \mid x \in [0,1],\ y \in (1,2] \}$
i) $\{ x + y \mid x^2 \geq 1,\ y^2 \geq 1 \}$

9.5. ($\star$) Řekli jsme, že množina $\mathbf{f}(\mathbb{R}^n) = \{ \mathbf{f}(\mathbf{x}) \mid \mathbf{x} \in \mathbb{R}^n \}$ obrazů zobrazení (9.2) je otevřená jednotková koule, kterou (dle (9.1)) označme jako $B_1(\mathbf{0}_n)$. Dokažte (viz §1.1.2), že zobrazení $\mathbf{f}\colon \mathbb{R}^n \to B_1(\mathbf{0}_n)$ s hodnotami (9.2) je bijektivní. Najděte zobrazení k němu inverzní.

9.6. Najděte všechny (globální i lokální) extrémy funkce $f(x,y) = x + y$ na množině $\{ (x,y) \in \mathbb{R}^2 \mid y = x^2,\ y \leq 1 \}$.

9.7. Najděte (úvahou, s pomocí náčrtků, bez použití derivací) všechny extrémy funkce

a) $f(\mathbf{x}) = \mathbf{a}^T\mathbf{x}$ (kde nenulový vektor $\mathbf{a} \in \mathbb{R}^n$ je dán)
b) $f(\mathbf{x}) = \mathbf{x}^T\mathbf{x}$

na množině

A) $\mathbb{R}^n$
B) $\{\, \mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\| = 1 \,\}$
C) $\{\, \mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\| \leq 1 \,\}$
D) $\{\, \mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\| < 1 \,\}$
E) daný afinní podprostor prostoru $\mathbb{R}^n$
F) $\{\, \mathbf{x} \in \mathbb{R}^n \mid -1 \leq \mathbf{1}^T\mathbf{x} \leq 1 \,\}$

U každého extrému určete, zda je lokální/globální, volný/vázaný.

9.8. Nechť $f\colon \mathbb{R}^n \to \mathbb{R}$ a $\mathbf{x} \in X' \subseteq X \subseteq \mathbb{R}^n$. Uvažujme dva výroky:

1. Funkce f má v bodě $\mathbf{x}$ lokální minimum na množině X.
2. Funkce f má v bodě $\mathbf{x}$ lokální minimum na množině X'.

Vyplývá druhý výrok z prvního? Vyplývá první výrok z druhého? Odpovědi dokažte.

9.9. Může se stát, že funkce má na množině lokální minimum a nemá na ní globální minimum? Odpověď dokažte.

9.10. Může nekonstantní lineární funkce nabývat na množině lokálního extrému ve vnitřním bodě této množiny? Odpověď dokažte.

9.11. Platilo by Tvrzení 9.3, kdyby se v něm slovo 'lokální' nahradilo slovem 'globální'? Odpověď dokažte.

9.12. Má-li lineární funkce na uzavřené množině extrém, nabývá ho vždy v hraničním bodě množiny. Je tento výrok pravdivý? Platí pro globální extrémy? Platí pro lokální extrémy? Odpovědi dokažte.

9.13. Jsou dány body $\mathbf{a}_1, \dots, \mathbf{a}_m \in \mathbb{R}^2$. Hledáme maxima funkce $f(\mathbf{x}) = \min_{i=1}^{m} \|\mathbf{a}_i - \mathbf{x}\|$ na jednotkovém kruhu $K = \{\, \mathbf{x} \in \mathbb{R}^2 \mid \|\mathbf{x}\| \leq 1 \,\}$.

a) Pochopte a rozmyslete tuto úlohu pro malá m (tedy $m = 1, 2, 3$). Jako první krok si nakreslete vrstevnice funkce f pro různé množiny bodů $\mathbf{a}_1, \dots, \mathbf{a}_m$ (pro malá m).
b) Najděte body $\mathbf{a}_1, \dots, \mathbf{a}_m$ (pro pokud možno malé m) a bod $\mathbf{x}$ tak, aby $\mathbf{x}$ bylo lokální maximum funkce f na kruhu K, které není globální.
c) ($\star$) Rozmyslete zobecnění úlohy, kdy je dána uzavřená množina $X \subseteq \mathbb{R}^2$ a hledáme maximum funkce $f(\mathbf{x}) = \min_{\mathbf{y} \in X} \|\mathbf{y} - \mathbf{x}\|$ (tj. vzdálenost od množiny X) na kruhu K (úloha výše je pak speciální případ pro $X = \{\mathbf{a}_1, \dots, \mathbf{a}_m\}$). Úlohu promyslete pro nějaké zajímavé a přitom jednoduché množiny X (např. kružnice nebo sjednocení malého počtu přímek).

Nápověda a řešení

9.3.a) Vnitřek je $\mathbb{R}$, protože pro každý bod $x \in \mathbb{R}$ existuje koule se středem x a kladným poloměrem ε, která celá leží v $\mathbb{R}$. Hranice je $\emptyset$, protože neexistuje žádný bod $x \in \mathbb{R}$ tak, že každá koule s kladným poloměrem protíná (= má neprázdný průnik) jak s množinou $\mathbb{R}$, tak s jejím doplňkem (tento doplněk je prázdná množina $\emptyset$).

9.3.b) Vnitřek je otevřený interval (a, b). Hranice je dvouprvková množina $\{a, b\}$. Množina je omezená.

9.3.c) Vnitřek je $\emptyset$ a hranice je $\mathbb{R}$. Neomezená.

9.3.d) Je to koule bez hranice, tedy intuitivně je vnitřek je ta samá množina a hranice je prázdná. Formálně bychom to snadno dokázali z definic v §9.1. Je omezená.

9.3.e) Jedná se o půlkružnici i s 'krajními' body. Vnitřek je prázdný (protože v kružnici nemůže ležet žádný kruh o kladném poloměru). Množina je sama svou hranicí. Je omezená.

9.3.f) Je to kus grafu paraboly (konečný, tedy množina je omezená), přičemž levý 'krajní' bod chybí. Vnitřek je prázdný. Hranice je ta samá množina ale i s levým krajním bodem $(-1, 1)$, tedy množina $\{ (x, y) \in \mathbb{R}^2 \mid y = x^2,\ -1 \leq x \leq 1 \}$. Všimněte si, že hraniční bod $(-1, 1)$ neleží v množině (což definice nevyžaduje).

9.3.g) Je to část hyperboly v kladném kvadrantu. Vnitřek je prázdný, je sama svou hranicí. Je neomezená. Všimněte si, že by se vnitřek a hranice nezměnily, kdyby se množina změnila na $\{ (x, y) \in \mathbb{R}^2 \mid xy < 1,\ x \geq 0,\ y \geq 0 \}$.

9.3.h) Množinu si pro představu načrtneme pro $n = 2$, tedy $\{ (x_1, y_1) \mid \max\{x_1, x_2\} \leq 1 \}$: je to záporný kvadrant posunutý vrcholem do bodu $(1, 1)$. Tedy neomezená. Pro větší n to bude podobné. Vnitřek je tentýž kvadrant bez hranice, který lze napsat jako $\{ \mathbf{x} \in \mathbb{R}^n \mid \max_{i=1}^n x_i < 1 \}$. Hranice je $\{ \mathbf{x} \in \mathbb{R}^n \mid \max_{i=1}^n x_i = 1 \}$. (Ovšem nemyslete si, že vnitřek se *vždycky* získá tak, že neostré nerovnosti v definici množiny nahradíme ostrými a pro hranici je nahradíme rovnostmi).

9.3.i) Vnitřek je prázdný, hranice ten samý afinní podprostor. Samozřejmě je neomezený.

9.3.j) Vnitřek je $\{ \mathbf{x} \in \mathbb{R}^n \mid b < \mathbf{a}^T\mathbf{x} < c \}$, hranice jsou dvě nadroviny $\{ \mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}^T\mathbf{x} = b \} \cup \{ \mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}^T\mathbf{x} = c \}$.

9.4. (a) $(0, 1]$, (b) $[-1, 0) \cup (0, 1]$, (c) $(0, 1]$, (d) $(-\sqrt{2}, \sqrt{2})$, (e,f,g) $[-\sqrt{2}, \sqrt{2}]$

9.5. Inverze k $\mathbf{f}$ je zobrazení $\mathbf{f}^{-1}\colon B_1(\mathbf{0}_n) \to \mathbb{R}^n$ s hodnotami $\mathbf{f}^{-1}(\mathbf{y}) = \dfrac{\mathbf{y}}{\sqrt{1 - \mathbf{y}^T\mathbf{y}}}$. To dokážeme ověřením, že $\mathbf{g}(\mathbf{f}(\mathbf{x})) = \mathbf{x}$ pro $\mathbf{x} \in \mathbb{R}^n$ a $\mathbf{f}(\mathbf{g}(\mathbf{y})) = \mathbf{y}$ pro $\mathbf{y} \in B_1(\mathbf{0}_n)$, což uděláme dosazením vzorečků do sebe. Bijektivnost obou zobrazení plyne z existence inverze.

9.6. $(1, 1)$ je glob. maximum, $(-\frac{1}{2}, \frac{1}{4})$ glob. minimum, $(-1, 1)$ lok. maximum. To je vidět z obrázku (nakreslete!). Jiný postup je eliminovat proměnnou y, čímž úlohu převedeme na hledání extrémů funkce $x + x^2$ na intervalu $[-1, 1]$.

9.7. (aA) Funkce je na množině shora i zdola neomezená, globální extrémy neexistují. Lokální také ne.

(aB) Lineární funkce na (hyper)sféře. Extrémy zjistíme např. pomocí náčrtku pro $n = 2$ a zobecníme pro libovolné n. Vrstevnice funkce jsou rovnoběžné nadroviny, kolmé na vektor $\mathbf{a}$. (Globální) maximum se nabývá v bodě $\mathbf{x} = \mathbf{a}/\|\mathbf{a}\|$ (normalizovaný vektor $\mathbf{a}$, neboli jednotkový vektor rovnoběžný s vektorem $\mathbf{a}$), (globální) minimum v bodě $\mathbf{x} = -\mathbf{a}/\|\mathbf{a}\|$. Jsou to extrémy vázané, neboť se nabývají v hraničních bodech množiny. Jiné extrémy nejsou.

(aC) Stejné jako (aB), protože lineární funkce nabývá na (hyper)kouli extrémy v hraničních bodech.

(aD) Maximum ani minimum funkce na množině neexistuje (existuje ale supremum a infimum, které jsou stejné jako v (aB)).

(aE) Řešení uhodneme např. z představy pro $n = 3$ a dimenzi afinního podprostoru 0 (bod), 1 (přímka) nebo 2 (rovina). Jestliže bude vektor $\mathbf{a}$ ortogonální k podprostoru (tedy kolmý k lineárnímu podprostoru vzniklému posunutím afinního podprostoru do počátku), podprostor leží v nějaké vrstevnici funkce, tedy funkce je na podprostoru konstantní (nulová), tedy každý bod podprostoru je vázané maximum i minimum. V opačném případě je funkce na množině shora i zdola neomezená, tedy extrémy neexistují.

(aF) Množina je 'panel', načrtněte si ho pro $n = 2$ a představte si ho pro $n = 3$. Situace je podobná jako (aE). Jestliže je vektor $\mathbf{a}$ kolmý k panelu (tj. $\mathbf{a}$ je rovnoběžný s vektorem $\mathbf{1}$, tj. $\mathbf{a} = \alpha\mathbf{1}$ pro nějaké $\alpha \neq 0$), pak jsou (globální) maximum a minimum v hraničních bodech panelu, v opačném případě extrémy neexistují. Nalézt body maxima a minima vyžaduje trochu zamyšlení: pro $\alpha > 0$ je maximum v bodě $\mathbf{x} = \mathbf{1}/n$ (aby platilo $\mathbf{1}^T\mathbf{x} = 1$) a minimum v bodě $\mathbf{x} = -\mathbf{1}/n$, pro $\alpha < 0$ je to opačně.

(bA) Hodnota funkce $f(\mathbf{x})$ je čtverec vzdálenosti bodu $\mathbf{x}$ od počátku, její vrstevnice jsou kružnice se středem v počátku. Funkce je na množině zdola omezená a shora neomezená. Globální minimum je v počátku $\mathbf{x} = \mathbf{0}$, je vázané. Jiné extrémy nejsou.

(bB) Funkce je na množině (hypersféře) konstantní, tedy v každém bodě množiny je (globální) minimum i maximum. Jsou vázané, protože leží na hranici množiny.

(bC) Volné (globální) minimum je v počátku, vázaná globální maxima jsou na hranici hyperkoule.

(bD) Volné (globální) minimum je v počátku, maxima neexistují.

(bE) Pokud je afinní podprostor bod, minimum a maximum se triviálně nabývají v něm. V opačném případě je na něm funkce shora neomezená, tedy maximum neexistuje. Minimum se nabývá v bodu $\mathbf{x}$ podprostoru nejblíže počátku, což znamená že vektor $\mathbf{x}$ musí být kolmý k podprostoru. Všechny extrémy jsou vázané, jen když je afinní podprostor celé $\mathbb{R}^n$ tak jsou volné.

(bF) Minimum je v počátku. Maximum se nabývá v bodech panelu nejblíže počátku, tj. ve dvou bodech $\mathbf{x} = \pm\mathbf{1}/n$.

9.8. Druhý výrok plyne z prvního. Naopak to ale neplatí, protipříklad je $X = \mathbb{R}$, $X' = [0, 1]$, $f(x) = x$.

9.13.a) Slovy lze úlohu popsat takto: hledáme bod v kruhu, který je co nejdále od nejbližšího z bodů $\mathbf{a}_1, \ldots, \mathbf{a}_m$.

Kapitola 10

Volné lokální extrémy

Zde se budeme věnovat podmínkám a algoritmům na volné lokální extrémy diferencovatelných funkcí. Díky Tvrzení 9.3 místo volných lokálních extrémů funkcí $f: \mathbb{R}^n \to \mathbb{R}$ na množině $X \subseteq \mathbb{R}^n$ budeme uvažovat pouze lokální extrémy funkce f na množině $\mathbb{R}^n$.

10.1 Analytické podmínky

Intuitivně víme, že je-li směrová derivace (viz §8.6) funkce $f: \mathbb{R}^n \to \mathbb{R}$ v bodě $\mathbf{x}$ ve směru $\mathbf{v}$ kladná, tj. $f_{\mathbf{v}}(\mathbf{x}) > 0$, funkce v bodě $\mathbf{x}$ ve směru $\mathbf{v}$ stoupá. Podobně, je-li $f_{\mathbf{v}}(\mathbf{x}) < 0$, funkce v bodě $\mathbf{x}$ ve směru $\mathbf{v}$ klesá. Upřesněme, co to znamená:

Lemma 10.1. Nechť funkce $f: \mathbb{R}^n \to \mathbb{R}$ je diferencovatelná v bodě $\mathbf{x} \in \mathbb{R}^n$ ve směru $\mathbf{v} \in \mathbb{R}^n$.

- Jestliže $f_{\mathbf{v}}(\mathbf{x}) > 0$, pak existuje $\delta > 0$ tak, že pro každé $\alpha \in (0, \delta]$ je $f(\mathbf{x} + \alpha\mathbf{v}) > f(\mathbf{x})$.
- Jestliže $f_{\mathbf{v}}(\mathbf{x}) < 0$, pak existuje $\delta > 0$ tak, že pro každé $\alpha \in (0, \delta]$ je $f(\mathbf{x} + \alpha\mathbf{v}) < f(\mathbf{x})$.

Důkaz. Zopakujme definici směrové derivace (8.19):

$$f_{\mathbf{v}}(\mathbf{x}) = \lim_{\alpha \to 0^+} g(\alpha) = a, \qquad \text{kde} \quad g(\alpha) = \frac{f(\mathbf{x} + \alpha\mathbf{v}) - f(\mathbf{x})}{\alpha}.$$

Dle definice limity zprava (viz každá učebnice analýzy) pro každé $\varepsilon > 0$ existuje $\delta > 0$ tak, že pro každé $\alpha \in (0, \delta]$ platí $|g(\alpha) - a| \leq \varepsilon$. Jestliže $a > 0$, existuje tedy $\delta > 0$ tak, že pro každé $\alpha \in (0, \delta]$ platí $g(\alpha) > 0$, tedy $f(\mathbf{x} + \alpha\mathbf{v}) - f(\mathbf{x}) > 0$. Případ $a < 0$ se dokáže obdobně. ∎

Jestliže $f_{\mathbf{v}}(\mathbf{x}) > 0$ [$f_{\mathbf{v}}(\mathbf{x}) < 0$], směr $\mathbf{v}$ nazveme **vzestupný [sestupný] směr** (angl. *ascent [descent] direction*) funkce f v bodě $\mathbf{x}$. Dle Lemmatu 10.1 můžeme v tom případě hodnotu funkce zvětšit [zmenšit] malým posunutím bodu $\mathbf{x}$ ve směru $\mathbf{v}$. Z toho ihned plyne nutná podmínka na volné lokální extrémy:

Věta 10.2. Nechť funkce $f: \mathbb{R}^n \to \mathbb{R}$ je diferencovatelná v bodě $\mathbf{x} \in \mathbb{R}^n$ ve směru $\mathbf{v} \in \mathbb{R}^n$.

- Jestliže $\mathbf{x}$ je lokální minimum funkce f, pak $f_{\mathbf{v}}(\mathbf{x}) \geq 0$.
- Jestliže $\mathbf{x}$ je lokální maximum funkce f, pak $f_{\mathbf{v}}(\mathbf{x}) \leq 0$.

Důkaz. Jestliže $f_{\mathbf{v}}(\mathbf{x}) < 0$, pak dle Lemmatu 10.1 pro každé $\varepsilon > 0$ existuje $\alpha > 0$ tak, že $\mathbf{x} + \alpha\mathbf{v} \in B_\varepsilon(\mathbf{x})$ a $f(\mathbf{x} + \alpha\mathbf{v}) < f(\mathbf{x})$, tedy $\mathbf{x}$ není lokální minimum funkce f. Obdobně pro maximum. (Viz definice lokálního extrému v §9.3.) ∎

Větu 10.2 moc často nepoužijete, protože málokdy potkáte funkci, která by v bodě byla diferencovatelná v nějakém směru a nebyla v něm (totálně) diferencovatelná. Je-li funkce (totálně) diferencovatelná, podmínka na volné lokální extrémy se zjednoduší:

Věta 10.3. Nechť funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ je v bodě $\mathbf{x} \in \mathbb{R}^n$ (totálně) diferencovatelná. Jestliže $\mathbf{x}$ je lokální extrém (tj. minimum či maximum) funkce f, pak $f'(\mathbf{x}) = \mathbf{0}$.

Důkaz. Protože f je v bodě $\mathbf{x}$ diferencovatelná, dle Věty 8.5 je $f_{\mathbf{v}}(\mathbf{x}) = f'(\mathbf{x})\mathbf{v}$ pro všechny $\mathbf{v} \in \mathbb{R}^n$. Speciálně to platí pro každý směr $\mathbf{v} = \mathbf{e}_i$ (i-tý vektor standardní báze prostoru $\mathbb{R}^n$), tedy $f'(\mathbf{x}) = \mathbf{0}$. ∎

Všimněte si, že $f'(\mathbf{x}) = \mathbf{0}$ říká, že všechny parciální derivace funkce f v bodě $\mathbf{x}$ jsou nulové ($f'(\mathbf{x})\mathbf{e}_i = f_{x_i}(\mathbf{x})$ je parciální derivace podle proměnné x_i, viz §8.6.) Jestliže je funkce f v bodě $\mathbf{x}$ totálně diferencovatelná a platí $f'(\mathbf{x}) = \mathbf{0}$, bod $\mathbf{x}$ se nazývá **stacionární bod** funkce f. Věta 10.3 říká, že stacionární body jsou body 'podezřelé' z volného lokálního extrému. Věta ovšem svádí k použití v situacích, kdy nejsou splněny její předpoklady. Uveďme příklady těchto situací:

Příklad 10.1. Funkce $f(x) = x^3$ má v bodě 0 stacionární bod, ale nemá tam lokální extrém. ♦

Příklad 10.2. V Příkladu 9.10 jsou předpoklady Věty 10.3 splněny pouze pro body b, c, které jsou stacionární a vnitřní. Body a, f jsou hraniční (tedy ne vnitřní) body intervalu $[a, f]$ a v bodech d, e není funkce diferencovatelná. ♦

Příklad 10.3. Funkce $f(\mathbf{x}) = \|\mathbf{x}\|$ má na hyperkrychli $X = \{\, \mathbf{x} \in \mathbb{R}^n \mid -\mathbf{1} \le \mathbf{x} \le \mathbf{1} \,\}$ v bodě $\mathbf{0}$ volné lokální minimum (nakreslete si množinu X a vrstevnice funkce f pro $n = 1$ a pro $n = 2$). Nemá tam ale stacionární bod, protože tam není totálně diferencovatelná, a tedy nelze použít Větu 10.3 (lze ovšem použít Lemma 10.1, protože f je v bodě $\mathbf{0}$ diferencovatelná ve všech směrech). Dále má funkce na množině X vázaná lokální maxima ve všech jejích rozích (např. v bodě $\mathbf{1}$), což jsou její hraniční body. Bod $\mathbf{1}$ ovšem není stacionární bod funkce f. ♦

Věta 10.3 udává podmínku *prvního řádu* na volné lokální extrémy, protože obsahuje první derivace. Následující podmínka *druhého řádu* pomůže zjistit, zda je stacionární bod volným lokálním extrémem, případně jakým:

Věta 10.4. Nechť funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ je v bodě $\mathbf{x} \in \mathbb{R}^n$ dvakrát diferencovatelná.

- Jestliže $\mathbf{x}$ je lokální minimum [maximum] funkce f, pak platí $f'(\mathbf{x}) = \mathbf{0}$ a Hessova matice $f''(\mathbf{x})$ je positivně [negativně] semidefinitní.
- Jestliže platí $f'(\mathbf{x}) = \mathbf{0}$ a Hessova matice $f''(\mathbf{x})$ je positivně [negativně] definitní, pak $\mathbf{x}$ je ostré lokální minimum [maximum] funkce f.

Všimněte si:

- Z Věty 10.4 plyne, že když $f''(\mathbf{x})$ je indefinitní (tedy není ani positivně semidefinitní ani negativně semidefinitní), pak $\mathbf{x}$ není lokální extrém. Bod $\mathbf{x}$, ve kterém $f'(\mathbf{x}) = \mathbf{0}$ a matice $f''(\mathbf{x})$ je indefinitní, se nazývá **sedlový bod** funkce f.
- Je-li matice $f''(\mathbf{x})$ (positivně či negativně) semidefinitní, věta neříká nic o tom, zda funkce v bodě $\mathbf{x}$ má nebo nemá lokální extrém. Příklady jsou funkce x^3 a x^4 v bodě 0.

Větu 10.4 nebudeme dokazovat, uvedeme jen důvod, díky kterému jí snad ochotněji uvěříte. Místo funkce f vyšetřujme v blízkosti bodu $\mathbf{x}$ její Taylorův polynom druhého stupně (8.23c),

$$T^2_{\mathbf{x}}(\mathbf{y}) = f(\mathbf{x}) + \underbrace{f'(\mathbf{x})(\mathbf{y}-\mathbf{x})}_{\mathbf{0}} + \tfrac{1}{2}(\mathbf{y}-\mathbf{x})^T f''(\mathbf{x})\,(\mathbf{y}-\mathbf{x}).$$

Protože $f'(\mathbf{x}) = \mathbf{0}$, lineární člen je nulový a polynom je tedy kvadratická forma posunutá do bodu $\mathbf{x}$. Zda funkce $T^2_{\mathbf{x}}$ má či nemá v bodě $\mathbf{x}$ extrém bychom tedy mohli určit podle Tvrzení 6.1 z definitnosti matice $f''(\mathbf{x})$. Zde ovšem vyšetřujeme funkci f a ne její aproximaci $T^2_{\mathbf{x}}$, proto pro lokální extrém nestačí (positivní či negativní) semidefinitnost $f''(\mathbf{x})$.

Příklad 10.4. Extrémy kvadratické funkce (6.20) umíme hledat pomocí rozkladu na čtverec. Ovšem je to také možné pomocí derivací. Podmínka stacionarity je

$$\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}}(\mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} + c) = 2\mathbf{x}^T\mathbf{A} + \mathbf{b}^T = \mathbf{0}.$$

Po transpozici dostaneme rovnici (6.22a).

Druhá derivace (Hessova matice) funkce f je $2\mathbf{A}$. Všimněte si, že když je matice $\mathbf{A}$ positivně nebo negativně semidefinitní (ale ne definitní), tak nám podmínky druhého řádu (Věta 10.4) o typu extrému nic neřeknou a musíme použít silnější podmínky pro kvadratické funkce z §6.4 nebo se odvolat na konvexitu či konkavitu funkce f. ♦

Příklad 10.5. V lineární úloze nejmenších čtverců (5.2) minimalizujeme kvadratickou (tudíž libovolně-krát diferencovatelnou) funkci ve tvaru $f(\mathbf{x}) = \|\mathbf{A}\mathbf{x}-\mathbf{b}\|^2$ na množině $\mathbb{R}^n$ (kde matice $\mathbf{A}$ samozřejmě označuje něco jiného než v Příkladu 10.4). Funkci zderivujeme např. užitím řetízkového pravidla (f je složení funkcí $\|\mathbf{y}\|^2$ a $\mathbf{y} = \mathbf{A}\mathbf{x}-\mathbf{b}$):

$$f'(\mathbf{x}) = 2(\mathbf{A}\mathbf{x}-\mathbf{b})^T\mathbf{A}.$$

(Nebo můžeme nejprve roznásobit závorky jako v (6.24) a derivovat potom.) Stacionární podmínka $f'(\mathbf{x}) = \mathbf{0}$, po transpozici a snadné úpravě, dostane tvar (5.4). hessián funkce f spočítáme opětovnou derivací funkce $f'(\mathbf{x})^T$, je to $f''(\mathbf{x}) = 2\mathbf{A}^T\mathbf{A}$, což je positivně semidefinitní (ale ne nutně positivně definitní) matice. Věta 10.4 nám tedy opět nemusí pomoct a musíme použít podmínky z §6.4. ♦

Podotkněme, že ověřování podmínek druhého řádu pro funkce mnoha proměnných může být nepříjemné a často se snažíme tomu vyhnout (už nalézt Hessovu matici může být problém, natož ověřovat její definitnost). Místo toho se snažíme vymyslet jiný (jednodušší) důkaz, že daný stacionární bod je/není lokální minimum/maximum.

10.2 Iterační metody na volné lokální extrémy

Dále se budeme věnovat numerickým iteračním metodám[1] na hledání volných lokálních extrémů (budeme uvažovat pouze minima) diferencovatelných funkcí $f\colon \mathbb{R}^n \to \mathbb{R}$. Počínaje počátečním odhadem $\mathbf{x}_0 \in \mathbb{R}^n$ řešení, iterační metoda postupně vytváří posloupnost $\mathbf{x}_1, \mathbf{x}_2, \ldots$, která za příznivých okolností konverguje k řešení úlohy (tj. k lokálnímu minimu).

[1] Schválně píšeme *metody* a ne *algoritmy*, neboť algoritmus by měl skončit po konečném počtu operací, kdežto iterační metoda typicky pouze konverguje v nekonečném počtu iterací.

Bod $\mathbf{x}_{k+1}$ závisí na předchozím bodu $\mathbf{x}_k$ a hodnotě účelové funkce f a příp. (pokud je funkce diferencovatelná) na jejích derivacích v tomto bodě, někdy též na hodnotě příp. derivacích v několika minulých bodech $\mathbf{x}_{k-1}, \mathbf{x}_{k-2}, \ldots$. **Řádem metody** se myslí nejvyšší řád použité derivace: metody *nultého řádu* (také zvané *metody bez derivací, derivative-free methods*) používají jen funkční hodnoty, metody *prvního řádu* navíc první derivace, a metody *druhého řádu* navíc druhé derivace. Metody vyššího než druhého řádu se užívají zřídka.

Základní otázka je, zda metoda konverguje k (nějakému) lokálnímu minimu a když ano, tak jak rychle. Tímto se zabývá *konvergenční analýza* metody a příslušné matematické věty jsou obvykle složité a jejich důkazy ještě složitější. Konvergenční vlastnosti konkrétní metody zjevně závisejí na vlastnostech funkce f a na volbě počátečního odhadu $\mathbf{x}_0$. Metody vyššího řádu obvykle konvergují k lokálnímu minimu pro menší množinu funkcí f a počátečních odhadů $\mathbf{x}_0$, ale když konvergují, tak konvergují obvykle rychleji než metody nižšího řádu.

Zde se zaměříme na třídu metod zvaných **sestupné metody**[2], jejichž iterace má tvar

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{v}_k, \tag{10.1}$$

kde $\mathbf{v}_k \in \mathbb{R}^n$ je sestupný směr v bodě $\mathbf{x}_k$ (tedy $f_{\mathbf{v}_k}(\mathbf{x}_k) < 0$) zvaný **směr hledání** a součinitel $\alpha_k > 0$ ovlivňuje délku kroku[3] v k-té iteraci. Metoda v každé iteraci zvolí vektor $\mathbf{v}_k$ a skalár α_k a provede aktualizaci odhadu dle (10.1). Volba směru hledání je klíčový znak pro rozlišení jednotlivých metod.

10.2.1 Volba součinitele délky kroku

Máme-li směr hledání $\mathbf{v}_k$, zde jsou obvyklé způsoby jak volit součinitel α_k délky kroku:

- *Exact line search.* Je-li směr $\mathbf{v}_k$ sestupný, dle Lemmatu 10.1 existuje délka kroku $\alpha_k > 0$ tak, že $f(\mathbf{x}_{k+1}) < f(\mathbf{x}_k)$. Nejlepší délka kroku α_k je minimum řezové funkce

  $$\varphi(\alpha) = f(\mathbf{x}_k + \alpha \mathbf{v}_k) \tag{10.2}$$

  na intervalu $(0, \infty)$. Tato úloha je v kontextu vícerozměrné optimalizace nazývána *jednorozměrné hledání* (angl. *line search*) a řešíme-li ji přesně (tj. hledáme-li globální minimum), mluvíme o *exact line search*.

- *Approximate line search.* Obvykle není chytré hledat globální minimum (na intervalu $(0, \infty)$) funkce (10.2) (ledaže by šlo najít velmi levně), protože se aktuální odhad $\mathbf{x}_k$ stejně v příští iteraci výrazně změní. Proto se často minimum hledá jen přibližně. Na nějaký algoritmus hledající přibližné minimum funkce jedné proměnné byste snadno přišli. Ovšem ne každý takový algoritmus garantuje dobré konvergenční vlastnosti iterační metody.

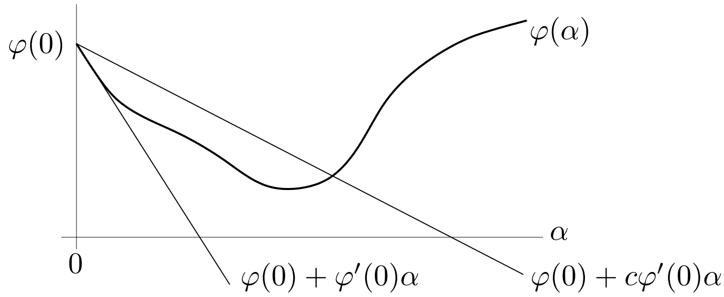
  Oblíbená metoda zaručující dobré konvergenční vlastnosti je *backtracking line search.* Hledáme α splňující tzv. *Armijovo-Goldsteinovo pravidlo*: pro nějakou konstantu $c \in (0,1)$ musí být

  $$f(\mathbf{x}_k + \alpha \mathbf{v}_k) = \varphi(\alpha) \quad \le \quad f(\mathbf{x}_k) + c\alpha f'(\mathbf{x}_k)\mathbf{v}_k = \varphi(0) + c\alpha\varphi'(0). \tag{10.3}$$

  Všimněte si, že výraz $\varphi(0) + \alpha\varphi'(0)$ je Taylorův polynom prvního stupně funkce φ v bodě 0 (tedy afinní funkce proměnné α), tedy $\varphi(0) + c\alpha\varphi'(0)$ je tato afinní funkce s mírnější směrnicí, viz obrázek:

[2] Existují i metody, ve kterých směr $\mathbf{v}_k$ není vždy sestupný (např. *subgradientní metody*).

[3] Skalár α_k se někdy nazývá *délka kroku*, což je ale špatný název, protože délka kroku v k-té iteraci je ve skutečnosti číslo $\|\mathbf{x}_{k+1} - \mathbf{x}_k\| = \alpha_k \|\mathbf{v}_k\|$.



Toto pravidlo samo nestačí, protože každá dostatečně malá hodnota $\alpha > 0$ ho splňuje. Proto se užívá v jednoduchém algoritmu: na začátku zvolíme nějakou maximální hodnotu α a parametry $c, \tau \in (0, 1)$ a pak zmenšujeme α jeho násobením číslem τ, dokud nezačne platit (10.3).

- *Pevná posloupnost.* V tomto případě součinitelé $\alpha_k > 0$ závisejí pouze na k. Obvykle se vyžaduje, aby posloupnost (α_k) splňovala

$$\lim_{k\to\infty} \alpha_k = 0, \qquad \sum_{k=1}^{\infty} \alpha_k = \infty. \tag{10.4}$$

Druhá podmínka je nutná, protože vzdálenost $\|\mathbf{x}_0 - \mathbf{x}^*\|$ mezi počátečním odhadem a hledaným řešením může být libovolně velká. Všimněte si, že i když jsou směry $\mathbf{v}_k$ sestupné, pevná posloupnost nezajišťuje monotónní pokles funkce f (tedy pro nějaká k může být $f(\mathbf{x}_{k+1}) > f(\mathbf{x}_k)$).

- *Konstantní posloupnost.* Součinitel je stejný pro každou iteraci, $\alpha_k = \alpha > 0$ pro každé k.

10.3 Gradientní metoda

Gradientní metoda (také známá jako *gradientní sestup* nebo *metoda největšího spádu*, angl. *gradient descent* nebo *steepest descent*) volí směr sestupu jako záporný gradient funkce f v bodě $\mathbf{x}_k$:

$$\mathbf{v}_k = -f'(\mathbf{x}_k)^T = -\nabla f(\mathbf{x}_k). \tag{10.5}$$

Tento směr je sestupný, což je okamžitě vidět dosazením do (8.21).

Výhodou gradientní metody je spolehlivost, daná tím, že směr (10.5) je vždy sestupný. Nevýhodou je často pomalá konvergence, obecne tzv. sublineární (tedy pomalejší než geometrická řada). Pro speciální funkce f je rychlost konvergence lineární (tj. tedy jako geometrická řada). Konvergence může být pomalá např. tehdy, když funkce v okolí lokálního optima je v některých směrech mnohem 'protaženější' než v jiných (přesněji, když vlastní čísla Hessovy matice $f''(\mathbf{x})$ jsou velmi různá).

Příklad 10.6. Hledejme minimum kvadratické formy $f(x, y) = (ax^2 + y^2)/2$ (kde $a > 0$) gradientní metodou s počátečním bodem $(x_0, y_0) = (1, a)$. Minimum se nabývá v bodě $(x, y) = (0, 0)$. Při přesném řešení problému (10.2) (tj. exact line search) je k-tá iterace rovna (odvoďte!)

$$x_k = \left(-\frac{a-1}{a+1}\right)^k, \quad y_k = a\left(\frac{a-1}{a+1}\right)^k. \tag{10.6}$$

Vidíme, že konvergence je velmi pomalá pro $a \ll 1$ nebo $a \gg 1$. ♦

10.3.1 Závislost na lineární transformaci souřadnic

Transformujme vektor proměnných $\mathbf{x}$ lineární transformací $\mathbf{y} = \mathbf{A}\mathbf{x}$, kde $\mathbf{A}$ je regulární matice. Je jasné, že funkce f původních proměnných $\mathbf{x}$ bude mít stejné extrémy jako funkce

$$g(\mathbf{y}) = f(\mathbf{x}) = f(\mathbf{A}^{-1}\mathbf{y}).$$

Iterace gradientní metody v nových proměnných je

$$\mathbf{y}_{k+1} = \mathbf{y}_k - \alpha_k\, g'(\mathbf{y}_k)^T. \tag{10.7}$$

Zjistíme, jaké iteraci to odpovídá v původních proměnných. Z řetízkového pravidla máme

$$g'(\mathbf{y}) = f'(\mathbf{A}^{-1}\mathbf{y})\mathbf{A}^{-1} = f'(\mathbf{x})\mathbf{A}^{-1}.$$

Dosazením za $\mathbf{y}$ a $g'(\mathbf{y})$ do (10.7) a úpravou (proveďte!) dostaneme

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k\, (\mathbf{A}^T\mathbf{A})^{-1} f'(\mathbf{x}_k)^T. \tag{10.8}$$

To lze napsat ve tvaru (10.1) se směrem hledání

$$\mathbf{v}_k = -(\mathbf{A}^T\mathbf{A})^{-1} f'(\mathbf{x}_k)^T. \tag{10.9}$$

Tento směr se liší od původního směru (10.5) vynásobením maticí $(\mathbf{A}^T\mathbf{A})^{-1}$. Vidíme, že gradientní metoda *není invariantní* vůči lineární transformaci souřadnic.

Nový směr (10.9) je také sestupný, tj. $-f'(\mathbf{x}_k)(\mathbf{A}^T\mathbf{A})^{-1}f'(\mathbf{x}_k)^T < 0$, neboť matice $\mathbf{A}^T\mathbf{A}$ a tedy i její inverze je positivně definitní (viz Cvičení 6.20 a 6.19).

Na vzorec (10.9) se lze dívat ještě obecněji. Je jasné, že směr $\mathbf{v}_k = -\mathbf{C}_k^{-1} f'(\mathbf{x}_k)^T$ je sestupný, je-li matice $\mathbf{C}_k$ positivně definitní. Opačně, každý sestupný směr lze napsat takto. Uvidíme, že metody uvedené dále budou mít vždy tento tvar sestupného směru, ovšem matice $\mathbf{C}_k$ bude jiná v každém kroku.

10.4 Newtonova metoda

Newtonova metoda (přesněji *Newtonova-Raphsonova*, v jednorozměrném případě též *metoda tečen*) je slavná iterační metoda na řešení soustav nelineárních rovnic. Lze ji použít i na minimalizaci funkce tak, že hledáme její bod s nulovým gradientem. Obě použití nyní popíšeme.

10.4.1 Použití na soustavy nelineárních rovnic

Nechť $\mathbf{g}\colon \mathbb{R}^n \to \mathbb{R}^n$ je diferencovatelné zobrazení. Chceme řešit rovnici $\mathbf{g}(\mathbf{x}) = \mathbf{0}$, což je soustava n rovnic s n neznámými. Myšlenka Newtonovy metody je jednoduchá: místo hledání nulového bodu zobrazení $\mathbf{g}$ (což je obecně velmi obtížné) opakujeme iteraci, která najde nulový bod afinní aproximace zobrazení $\mathbf{g}$ v okolí aktuálního odhadu (což je snadné).

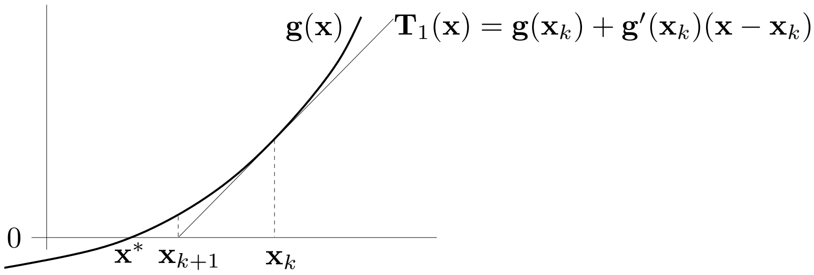
Afinní aproximace zobrazení $\mathbf{g}$ v okolí bodu $\mathbf{x}_k$ je zobrazení (viz (8.7) nebo (8.24))

$$\mathbf{T}^1_{\mathbf{x}_k}(\mathbf{x}) = \mathbf{g}(\mathbf{x}_k) + \mathbf{g}'(\mathbf{x}_k)(\mathbf{x} - \mathbf{x}_k). \tag{10.10}$$

Další odhad $\mathbf{x}_{k+1}$ najdeme řešením nehomogenní soustavy lineárních rovnic $\mathbf{T}^1_{\mathbf{x}_k}(\mathbf{x}_{k+1}) = \mathbf{0}$. Pokud je Jacobiho matice $\mathbf{g}'(\mathbf{x}_k)$ regulární, řešením je

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{g}'(\mathbf{x}_k)^{-1}\mathbf{g}(\mathbf{x}_k). \tag{10.11}$$

Viz obrázek:



Z numerického hlediska samozřejmě není dobré soustavu (10.10) řešit pomocí explicitního výpočtu inverze matice $\mathbf{g}'(\mathbf{x}_k)$. Lépe je nejdříve spočítat vektor $\mathbf{v}_k$ řešením lineární soustavy $\mathbf{g}(\mathbf{x}_k) + \mathbf{g}'(\mathbf{x}_k)\mathbf{v}_k = \mathbf{0}$ (např. Gaussovou eliminací nebo QR rozkladem) a pak provést iteraci (10.11) jako $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{v}_k$.

Hlavní výhodou Newtonovy metody je, že v blízkém okolí řešení obvykle konverguje velmi rychle (tzv. *superlineárně*), mnohem rychleji než gradientní metoda. Nevýhodou je, že je obvykle nutno začít s poměrně přesnou aproximací $\mathbf{x}_0$ skutečného řešení, jinak metoda snadno diverguje.

Příklad 10.7. *Babylónská metoda* na výpočet druhé odmocniny čísla $a \geq 0$ opakuje iteraci

$$x_{k+1} = \frac{1}{2}\Big(x_k + \frac{a}{x_k}\Big).$$

To není nic jiného než Newtonova metoda pro rovnici $0 = g(x) = x^2 - a$. Opravdu,

$$x_{k+1} = x_k - \frac{g(x_k)}{g'(x_k)} = x_k - \frac{x_k^2 - a}{2x_k} = x_k - \frac{1}{2}\Big(x_k - \frac{a}{x_k}\Big) = \frac{1}{2}\Big(x_k + \frac{a}{x_k}\Big).$$ ♦

Příklad 10.8. Hledejme průsečík $(x, y) \in \mathbb{R}^2$ dvou rovinných křivek daných rovnicemi

$$\begin{aligned}(x-1)^2 + y^2 &= 1,\\ x^4 + y^4 &= 1.\end{aligned}$$

Máme

$$\mathbf{g}(x,y) = \begin{bmatrix}(x-1)^2 + y^2 - 1\\ x^4 + y^4 - 1\end{bmatrix} \in \mathbb{R}^2, \qquad \mathbf{g}'(x,y) = \begin{bmatrix}2(x-1) & 2y\\ 4x^3 & 4y^3\end{bmatrix} \in \mathbb{R}^{2\times 2}.$$

Iterace (10.11) je

$$\begin{bmatrix}x_{k+1}\\ y_{k+1}\end{bmatrix} = \begin{bmatrix}x_k\\ y_k\end{bmatrix} - \begin{bmatrix}2(x_k-1) & 2y_k\\ 4x_k^3 & 4y_k^3\end{bmatrix}^{-1} \begin{bmatrix}(x_k-1)^2 + y_k^2 - 1\\ x_k^4 + y_k^4 - 1\end{bmatrix}.$$

Načrtneme-li si obě křivky, vidíme, že mají dva průsečíky, lišící se znaménkem druhé souřadnice. Zvolme počáteční odhad pro horní průsečík $(x_0, y_0) = (1, 1)$. První iterace bude

$$\begin{bmatrix}x_1\\ y_1\end{bmatrix} = \begin{bmatrix}1\\ 1\end{bmatrix} - \begin{bmatrix}0 & 2\\ 4 & 4\end{bmatrix}^{-1} \begin{bmatrix}0\\ 1\end{bmatrix} = \begin{bmatrix}0.75\\ 1\end{bmatrix}.$$

Pro šestou iteraci $(x_6, y_6) = (\,0.671859751039018,\ 0.944629015546222\,)$ jsou rovnice splněny se strojovou přesností. ♦

Příklad 10.9. Funkce $g(x) = x^2 - 1$ má dva nulové body $x = \pm 1$. Pokud v nějaké iteraci bude $x_k = 0$, nastane dělení nulou (derivace g v bodě x_k bude nulová). Pokud bude x_k velmi malé, dělení nulou nenastane, ale iterace x_{k+1} se ocitne velmi daleko od kořene. ♦

Příklad 10.10. Pro funkci $g(x) = x^3 - 2x + 2$ zvolme $x_0 = 0$. Další iterace bude $x_1 = 1$ a další $x_2 = 0$. Metoda bude oscilovat mezi hodnotami 0 a 1, tedy bude divergovat. ♦

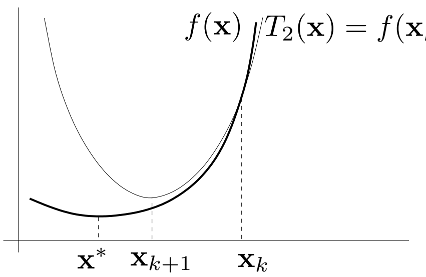
10.4.2 Použití na minimalizaci funkce

Newtonovu metodu lze použít pro hledání lokálního extrému dvakrát diferencovatelné funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ tak, že v metodě (10.11) položíme $\mathbf{g}(\mathbf{x}) = f'(\mathbf{x})^T$. Tím dostaneme iteraci

$$\mathbf{x}_{k+1} = \mathbf{x}_k - f''(\mathbf{x}_k)^{-1} f'(\mathbf{x}_k)^T, \tag{10.12}$$

kde $f''(\mathbf{x}_k)$ je Hessova matice funkce f v bodě $\mathbf{x}_k$. Nepleťte si tato dvě různá použití Newtonovy metody (tj. na hledání kořenů a na hledání lokálních extrémů)!

Iterace (10.11) se odvodila tak, že se zobrazení $\mathbf{g}$ v okolí bodu $\mathbf{x}_k$ aproximovalo afinním zobrazením $\mathbf{T}^1_{\mathbf{x}_k}$ a pak se našel nulový bod tohoto zobrazení. Lze ukázat (viz Cvičení 10.11), že iterace (10.12) lze odvodit také tak, že se funkce f aproximuje Taylorovým polynomem druhého stupně $T^2_{\mathbf{x}_k}$ (tedy kvadratickou funkcí) a pak se najde minimum této kvadratické funkce.



Iteraci (10.12) lze napsat v obecnějším tvaru (10.1), kde

$$\mathbf{v}_k = -f''(\mathbf{x}_k)^{-1} f'(\mathbf{x}_k)^T. \tag{10.13}$$

Oproti tvaru (10.12) v iteraci tedy přibyl koeficient α_k. Výhodou tohoto zobecnění je možnost zvolit optimální (ne nutně jednotkový) součinitel α_k pomocí jednorozměrné minimalizace (10.2). Metodě (10.1) s jednotkovým součinitelem α_k (což odpovídá (10.12)) se pak říká **čistá** Newtonova metoda.

Vektoru (10.13) říkáme **Newtonův směr**. Vidíme, že se od gradientního směru (10.5) liší násobením Hessovou maticí $f''(\mathbf{x}_k)$. Aby to byl sestupný směr, musí být

$$f'(\mathbf{x}_k)\,\mathbf{v}_k = -f'(\mathbf{x}_k) f''(\mathbf{x}_k)^{-1} f'(\mathbf{x}_k)^T < 0.$$

Toto platí, když $f'(\mathbf{x}_k) \neq \mathbf{0}$ (tj. $\mathbf{x}_k$ není stacionární bod) a matice $f''(\mathbf{x}_k)$ je positivně definitní (neboť pak bude positivně definitní i její inverze, viz Cvičení 6.19).

Opět není dobré explicitně počítat inverzi Hessovy matice $f''(\mathbf{x}_k)$, ale je lépe Newtonův směr $\mathbf{v}_k$ počítat řešením lineární soustavy $f'(\mathbf{x}_k) + f''(\mathbf{x}_k)\mathbf{v}_k = \mathbf{0}$. Na rozdíl od §10.4.1 je zde ovšem matice soustavy $f''(\mathbf{x}_k)$ symetrická a typicky positivně definitní, tedy můžeme na řešení soustavy použít efektivnější algoritmy než např. Gaussovu eliminaci (např. Choleského nebo LDL rozklad).

V porovnání s gradientní metodou má Newtonova metoda (použitá na minimalizaci funkce) nevýhodu v tom, že musíme počítat Hessovu matici $f''(\mathbf{x}_k)$ a řešit soustavu $f''(\mathbf{x}_k)\mathbf{v}_k = -f'(\mathbf{x}_k)^T$, což pro velký počet proměnných je pomalé či nemožné.

10.5 Nelineární metoda nejmenších čtverců

Mějme soustavu rovnic $\mathbf{g}(\mathbf{x}) = \mathbf{0}$, kde $\mathbf{g}\colon \mathbb{R}^n \to \mathbb{R}^m$ (tedy je to soustava m rovnic s n neznámými). Soustavu nazveme přeurčenou, jestliže nemá žádné řešení. Chceme takovou přeurčenou soustavu řešit přibližně ve smyslu nejmenších čtverců. Tedy chceme minimalizovat funkci

$$f(\mathbf{x}) = \|\mathbf{g}(\mathbf{x})\|^2 = \mathbf{g}(\mathbf{x})^T\mathbf{g}(\mathbf{x}) = \sum_{i=1}^{m} g_i(\mathbf{x})^2, \tag{10.14}$$

kde g_i jsou složky zobrazení $\mathbf{g}$. Speciálním případem je přibližné řešení lineární nehomogenní soustavy $\mathbf{A}\mathbf{x} = \mathbf{b}$, kde $\mathbf{g}(\mathbf{x}) = \mathbf{b} - \mathbf{A}\mathbf{x}$ (viz §5.1).

Zatímco v §10.3 a §10.4.2 bylo cílem minimalizovat *obecnou* funkci, zde chceme minimalizovat funkci ve speciálním tvaru (10.14). Nyní máme dvě možnosti. Buď můžeme nasadit na funkci (10.14) jednu z metod pro minimalizaci obecné funkce, k čemuž se vrátíme v §10.5.2. Nebo můžeme být chytřejší a využít speciálního tvaru funkce (10.14), což popíšeme v §10.5.1.

10.5.1 Gaussova-Newtonova metoda

Aproximujme opět zobrazení $\mathbf{g}$ v okolí bodu $\mathbf{x}_k$ afinním zobrazením $\mathbf{T}^1_{\mathbf{x}_k}$ dle (10.10). Úloha (10.14) pak vyžaduje minimalizovat $\|\mathbf{T}^1_{\mathbf{x}_k}(\mathbf{x})\|^2$. To je úloha lineárních nejmenších čtverců, kterou již známe z §5.1. Normální rovnice (5.4) má tvar

$$\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}'(\mathbf{x}_k)(\mathbf{x} - \mathbf{x}_k) = -\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}(\mathbf{x}_k). \tag{10.15}$$

Její řešení napišme pomocí pseudoinverze:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{g}'(\mathbf{x}_k)^+\mathbf{g}(\mathbf{x}_k). \tag{10.16}$$

Metoda (10.16) je známa jako (čistá) **Gaussova-Newtonova metoda**. Můžeme ji opět napsat obecněji ve tvaru (10.1) se směrem hledání (Gaussův-Newtonův směr)

$$\mathbf{v}_k = -\mathbf{g}'(\mathbf{x}_k)^+\mathbf{g}(\mathbf{x}_k). \tag{10.17}$$

Všmněte si, že pokud $m = n$ a Jakobiho matice $\mathbf{g}'(\mathbf{x}_k)$ je regulární, je $\mathbf{g}'(\mathbf{x}_k)^+ = \mathbf{g}'(\mathbf{x}_k)^{-1}$, tedy Gaussova-Newtonova metoda (10.16) se redukuje na Newtonovu metodu (10.11).

Pokud má Jacobiho matice $\mathbf{g}'(\mathbf{x}_k)$ lineárně nezávislé sloupce (viz §5.1), výraz (10.17) můžeme napsat jako

$$\mathbf{v}_k = -(\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}'(\mathbf{x}_k))^{-1}\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}(\mathbf{x}_k) = -\tfrac{1}{2}(\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}'(\mathbf{x}_k))^{-1}f'(\mathbf{x}_k)^T, \tag{10.18}$$

kde ve výrazu na pravé straně jsme dosadili derivaci

$$f'(\mathbf{x}) = 2\mathbf{g}(\mathbf{x})^T\mathbf{g}'(\mathbf{x}) \tag{10.19}$$

účelové funkce (10.14) (odvoďte dle §8.5.2!). Vidíme, že Gaussův-Newtonův směr (10.18) se liší od gradientního směru (10.5) pouze násobením maticí $\frac{1}{2}(\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}'(\mathbf{x}_k))^{-1}$. Aby byl tento směr sestupný, musí být

$$f'(\mathbf{x}_k)\,\mathbf{v}_k = -\tfrac{1}{2}f'(\mathbf{x}_k)(\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}'(\mathbf{x}_k))^{-1}f'(\mathbf{x}_k)^T < 0.$$

To platí, když $f'(\mathbf{x}_k) \neq \mathbf{0}$ a matice $\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}'(\mathbf{x}_k)$ je positivně definitní (viz Cvičení 6.19). Matice $\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}'(\mathbf{x}_k)$ je positivně definitní, právě když $\mathbf{g}'(\mathbf{x}_k)$ má lineárně nezávislé sloupce (viz Cvičení 6.20), což ovšem již předpokládáme kvůli existenci inverze. Tedy vidíme, že za přirozených podmínek je Gaussův-Newtonův směr vždy sestupný.

Čistá Gaussova-Newtonova metoda (tj. s $\alpha_k = 1$ pro každé k) může divergovat, a to i když je počáteční odhad $\mathbf{x}_0$ libovolně blízko lokálnímu minimu funkce (10.14). Protože ale Gaussův-Newtonův směr je vždy sestupný, vhodnou volbou součinitelů α_k lze vždy zajistit pokles účelové funkce f v každé iteraci a tedy (jestliže je funkce f zdola omezená) konvergenci.

Příklad 10.11. Hledáme přibližné řešení soustavy tří rovnic o dvou neznámých

$$\begin{aligned}(x-1)^2 + y^2 &= 1,\\ x^4 + y^4 &= 1,\\ x^2 + (y-1)^2 &= 1/2.\end{aligned}$$

Oba průsečíky křivek danými prvními dvěma rovnicemi již známe z Příkladu 10.8. Ani jeden z těchto průsečíků neleží na třetí křivce (i když je jí blízko), tedy soustava je přeurčená. Nezbývá nám tedy, než ji řešit přibližně. Hledáme bod $(x, y) \in \mathbb{R}^2$, který minimalizuje číslo

$$f(x,y) = \mathbf{g}(x,y)^T\mathbf{g}(x,y) = ((x-1)^2 + y^2 - 1)^2 + (x^4 + y^4 - 1)^2 + (x^2 + (y-1)^2 - 1/2)^2$$

kde

$$\mathbf{g}(x,y) = \begin{bmatrix}(x-1)^2 + y^2 - 1\\ x^4 + y^4 - 1\\ x^2 + (y-1)^2 - 1/2\end{bmatrix} \in \mathbb{R}^3, \qquad \mathbf{g}'(x,y) = \begin{bmatrix}2(x-1) & 2y\\ 4x^3 & 4y^3\\ 2x & 2(y-1)\end{bmatrix} \in \mathbb{R}^{3\times 2}.$$

Rozumný počáteční odhad je $(x_0, y_0) = (1, 1)$. První Gaussova-Newtonova iterace (10.16) je

$$\begin{bmatrix}x_1\\ y_1\end{bmatrix} = \begin{bmatrix}1\\ 1\end{bmatrix} - \begin{bmatrix}0 & 2\\ 4 & 4\\ 2 & 0\end{bmatrix}^+ \begin{bmatrix}0\\ 1\\ 1/2\end{bmatrix} = \begin{bmatrix}0.75\\ 1\end{bmatrix}.$$

Po osmé iteraci $(x_8, y_8) = (\,0.691002152515578, 0.940548357857245\,)$ se již hodnota $f(x_8, y_8) = 0.0008674592922855055$ v rámci strojové přesnosti nemění. ♦

Příklad 10.12. V systému GPS máme m satelitů se známými souřadnicemi $\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^n$ a chceme spočítat souřadnice pozorovatele $\mathbf{x} \in \mathbb{R}^n$ z naměřených vzdáleností $y_i = \|\mathbf{a}_i - \mathbf{x}\|$ pozorovatele od satelitů. Měření jsou zatížena chybou, proto obecně tato soustava rovnic nebude mít žádné řešení. Řešme tuto přeurčenou soustavu nelineárních rovnic ve smyslu nejmenších čtverců, tedy minimalizujme funkci

$$f(\mathbf{x}) = \sum_{i=1}^{m} \big(\|\mathbf{x} - \mathbf{a}_i\| - y_i\big)^2.$$

Máme[4] tedy $\mathbf{g} = (g_1, \ldots, g_m)\colon \mathbb{R}^n \to \mathbb{R}^m$, kde $g_i(\mathbf{x}) = \|\mathbf{x} - \mathbf{a}_i\| - y_i$. Derivace složek $\mathbf{g}$ je (pomůže vám §8.5.2, ale udělejte sami!) $g_i'(\mathbf{x}) = (\mathbf{x} - \mathbf{a}_i)^T/\|\mathbf{x} - \mathbf{a}_i\|$. Tedy

$$\mathbf{g}'(\mathbf{x}) = \begin{bmatrix}(\mathbf{x} - \mathbf{a}_1)^T/\|\mathbf{x} - \mathbf{a}_1\|\\ \vdots\\ (\mathbf{x} - \mathbf{a}_m)^T/\|\mathbf{x} - \mathbf{a}_m\|\end{bmatrix} \in \mathbb{R}^{m\times n}.$$

Pak dosadíme do vzorečku (10.16). ♦

[4] Zde ignorujeme, že funkce f není všude diferencovatelná. Přesně, není diferencovatelná v bodech $\mathbf{a}_1, \ldots, \mathbf{a}_m$ (promyslete!), což budeme ignorovat.

10.5.2 Rozdíl oproti Newtonově metodě

Předpokládejme, že bychom optimalizovali naši účelovou funkci (10.14) přímo Newtonovou metodou z §10.4.2. Spočítejme (viz Cvičení 8.12.d) Hessovu matici funkce (10.14):

$$f''(\mathbf{x}) = 2\mathbf{g}'(\mathbf{x})^T\mathbf{g}'(\mathbf{x}) + 2\sum_{i=1}^{m} g_i(\mathbf{x})g_i''(\mathbf{x}). \tag{10.20}$$

Hessova matice je součtem členu obsahujícího derivace prvního řádu a členu obsahujícího derivace druhého řádu. Vidíme, že Gaussův-Newtonův směr (10.18) se liší od Newtonova směru (10.13) zanedbáním členu druhého řádu v Hessově matici (10.20). Jinými slovy, Gaussovu-Newtonovu metodu je možno vnímat jako aproximaci Newtonovy metody na minimalizaci funkce (10.14) spočívající v tom, že zanedbáme členy druhého řádu, tedy skutečná Hessova matice (10.20) aproximujeme výrazem $2\mathbf{g}'(\mathbf{x})^T\mathbf{g}'(\mathbf{x})$.

To se projevuje tím, že Gaussova-Newtonova metoda obvykle konverguje pomaleji než plná Newtonova metoda použitá na funkci (10.14). Ovšem vyhnuli jsme se počítání druhých derivací funkce $\mathbf{g}$, což je hlavní výhoda Gaussovy-Newtonovy metody.

10.5.3 Levenbergova-Marquardtova metoda

Levenbergova-Marquardtova metoda je široce používané vylepšení Gaussovy-Newtonovy metody, které její iteraci

$$\mathbf{x}_{k+1} = \mathbf{x}_k - (\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}'(\mathbf{x}_k))^{-1}\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}(\mathbf{x}_k) \tag{10.21}$$

nahrazuje iterací

$$\mathbf{x}_{k+1} = \mathbf{x}_k - (\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}'(\mathbf{x}_k) + \mu_k\mathbf{I})^{-1}\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}(\mathbf{x}_k) \tag{10.22}$$

kde $\mu_k > 0$. Přidání členu $\mu_k\mathbf{I}$ je vlastně (Tichonovova) regularizace (viz (5.31)). Potom:

- Pro malé μ_k se iterace (10.22) blíží Gaussově-Newtonově iteraci.
- Pro velké μ_k je $(\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}'(\mathbf{x}_k) + \mu_k\mathbf{I})^{-1} \approx \frac{1}{\mu_k}\mathbf{I}$, tedy (10.22) se (po dosazení (10.19)) blíží iteraci $\mathbf{x}_{k+1} = \mathbf{x}_k - \frac{1}{2\mu_k}f'(\mathbf{x}_k)^T$ gradientní metody se součinitelem délky kroku $\alpha_k = \frac{1}{2\mu_k}$.

Tím jsou spojeny výhody Gaussovy-Newtonovy metody (typicky rychlá konvergence v okolí optima) a gradientní metody (spolehlivost i daleko od optima). Volbou parametru μ_k spojitě přecházíme mezi oběma metodami.

Parametr μ_k měníme se zvětšujícím se k pomocí jednoduché heuristiky. Začneme s nějakým velkým μ_0 a pak v každé iteraci:

- Pokud iterace snížila účelovou funkci, iteraci přijmeme a μ_k zmenšíme.
- Pokud iterace nesnížila účelovou funkci, iteraci odmítneme a μ_k zvětšíme.

Zvětšování a zmenšování μ_k děláme násobením a dělením konstantou, třeba 2 nebo 10. Všimněte si, toto nahrazuje optimalizaci součinitele α_k (*line search*).

Motivaci pro přidání regularizace do Gaussovy-Newtonovy iterace (10.21) lze vidět i jinak. Matice $\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}'(\mathbf{x}_k)$ může být singulární (to když sloupce $\mathbf{g}'(\mathbf{x}_k)$ budou lineárně závislé) nebo blízká singulární. Pak její inverze neexistuje nebo je velmi citlivá na malé změny matice $\mathbf{g}'(\mathbf{x}_k)$, což může neblaze ovlivnit konvergenci metody. Matice (10.22) je ale vždy positivně definitní (viz Cvičení 6.17), a tedy regulární.

10.6 Cvičení

10.1. Funkce $f\colon \mathbb{R}^3 \to \mathbb{R}$ má stacionární bod $(2,1,5)$. Co se dá o tomto stacionárním bodě říci, když Hessova matice $f''(2,1,5)$ v něm má vlastní čísla (a) $\{2,3,-1\}$, (b) $\{2,3,0\}$, (c) $\{2,1,1\}$.

10.2. Pro následující funkce najděte stacionární body (dejte pozor při řešení stacionárních podmínek, ať vám nějaká řešení neuniknou). Pro každý stacionární bod určete, zda je to lokální minimum, lokální maximum, či ani jedno. Pokud to určit neumíte, odůvodněte.

a) $f(x,y) = x(1 - \frac{2}{3}x^2 - y^2)$
b) $f(x,y) = 1/x + 1/y + xy$
c) $f(x,y) = \mathrm{e}^y(y^2 - x^2)$
d) $f(x,y) = 3x - x^3 - 3xy^2$
e) $f(x,y) = 6xy^2 - 2x^3 - 3y^4$
f) $f(x,y) = x^4/3 + y^4/2 - 4xy^2 + 2x^2 + 2y^2 + 3$
g) $f(x,y,z) = x^3 + y^3 + 2xyz + z^2$

10.3. Najděte lokální extrémy funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ s hodnotami $f(\mathbf{x}) = \mathbf{a}^T\mathbf{x} - \sum_{i=1}^n x_i \ln x_i$, kde $\mathbf{a}$ je daný vektor.

10.4. Vyšetřete extrémy funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ s hodnotami $f(\mathbf{x}) = \mathbf{a}^T\mathbf{x} + 1/(\mathbf{b}^T\mathbf{x})$, kde $\mathbf{a}$ a $\mathbf{b} \neq \mathbf{0}$ jsou známé vektory. Tj. zjistěte, jaké podmínky musí splňovat vektory $\mathbf{a}$ a $\mathbf{b}$, aby funkce měla aspoň jeden extrém, a za tohoto přepokladu najděte všechny (lokální i globální) extrémy funkce f.

10.5. Najděte všechna řešení rovnice $\sin x = \frac{1}{2}x$ (sinus je v radiánech) na kalkulačce s největší přesností, jakou dokážete.

10.6. Najděte lokální extrém funkce $f(x,y) = x^2 - y + \sin(y^2 - 2x)$ čistou Newtonovou metodou. Počáteční odhad zvolte $(x_0, y_0) = (1,1)$. Můžete použít počítač.

10.7. *Dělení bez dělení.* Pro dané číslo $a \neq 0$ chceme přibližně spočítat číslo $x = 1/a$, přičemž smíme používat jen operace sčítání, odčítání a násobení (dělení ne). Číslo x je řešením rovnice $g(x) = 1/x - a = 0$. Odvoďte iteraci Newtonovy metody pro tuto rovnici a ukažte, že používá pouze tyto operace. Ukažte, že Newtonova iterace pro rovnici $g(x) = 1/a - x = 0$ (která má stejné řešení) by k ničemu užitečnému nevedla, protože by používala dělení.

10.8. Soustavu dvou rovnic o jedné neznámé

$$\begin{aligned} x^2 + x &= 1 \\ x^2 - x &= 1 \end{aligned}$$

chceme řešit přibližně ve smyslu nejmenších čtverců. Napište iteraci (a) čisté Gaussovy-Newtonovy metody, (b) čisté Newtonovy metody. Výsledné vzorce zjednodušte.

10.9. Máme soustavu rovnic

$$\begin{aligned} x + y - 2xy &= 1 \\ -x + y + xy &= -3 \\ x - y + xy &= 1 \end{aligned}$$

Je soustava lineární? Kolik má řešení a proč? Chceme soustavu řešit přibližně ve smyslu nejmenších čtverců, tj. minimalizovat funkci $f(x,y)$ ve tvaru (10.14). Napište iteraci (a) gradientní, (b) Newtonovy, (c) Gaussovy-Newtonovy, (d) Levenbergovy-Marquardtovy metody.

10.10. Chceme najít vzdálenost množiny $\{ (x,y) \in \mathbb{R}^2 \mid x^2 = y \}$ od kružnice s poloměrem 1 a středem v bodě $(2,0)$. Tvrdíme, že tuto úlohu lze řešit tak, že vyřešíme přeurčenou soustavu $\{ x^2 = y,\ (x-2)^2 + y^2 = 1 \}$ přibližně ve smyslu nejmenších čtverců, tedy minimalizujeme funkci $f(x,y) = (x^2 - y)^2 + ((x-2)^2 + y^2 - 1)^2$. Je to pravda, bude minimální hodnota této funkce rovna (čtverci) vzdálenosti mezi množinami? Pokud ne, jak bychom tuto vzdálenost spočítali?

10.11. Čistá Newtonova metoda (10.12) na minimalizaci funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ je Newtonova metoda (10.11) na řešení soustavy $f'(\mathbf{x})^T = \mathbf{0}$. Takto jsme ji odvodili. Ukažte, že iteraci (10.12) lze odvodit také tak, že funkci f aproximujeme okolo bodu $\mathbf{x}_k$ Taylorovým polynomem druhého řádu a najdeme $\mathbf{x}_{k+1}$ jako minimum tohoto polynomu.

10.12. ($\star$) V §10.3.1 jsme ukázali, že iterace gradientní metody není invariantní vůči lineární transformaci souřadnic $\mathbf{y} = \mathbf{A}\mathbf{x}$ (pro regulární $\mathbf{A}$). Ukažte, že iterace Newtonovy metody (10.12) je invariantní vůči této transformaci.

10.13. Máme m bodů v rovině $\mathbf{x}_1, \ldots, \mathbf{x}_m \in \mathbb{R}^2$. Tyto body chceme proložit kružnicí ve smyslu nejmenších čtverců, tj. hledáme kružnici se středem $\mathbf{c} \in \mathbb{R}^2$ a poloměrem $r \in \mathbb{R}$ takovou, aby součet čtverců kolmých vzdáleností bodů od kružnice byl minimální. Zformulujte úlohu matematicky. Napište iteraci (a) gradientní, (b) Gaussovy-Newtonovy, (c) Levenbergovy-Marquardtovy metody.

10.14. ($\star$) Ztížíme předchozí cvičení. Máme m bodů v prostoru $\mathbf{x}_1, \ldots, \mathbf{x}_m \in \mathbb{R}^3$. Tyto body chceme proložit (ve smyslu nejmenších čtverců) kružnicí v prostoru. Kružnice v prostoru je definována středem, poloměrem a rovinou, ve které leží. Formulujte úlohu matematicky. Napište iteraci Levenbergovy-Marquardtovy metody.

10.15. ($\star$) Gaussův-Newtonův směr (10.17) se získá řešením normální rovnice (10.15). Ukázali jsme, že když Jacobiho matice $\mathbf{g}'(\mathbf{x}_k)$ má lineárně nezávislé sloupce, je Gaussův-Newtonův směr sestupný. Má-li $\mathbf{g}'(\mathbf{x}_k)$ lineárně závislé sloupce, rovnice (10.15) má nekonečně mnoho řešení, tj. směrů hledání je nekonečně mnoho. Dokažte, že každý takový směr je sestupný.

Nápověda a řešení

10.1. (a) funkce nemá v tomto bodě lokální extrém, (b) nemůžeme rozhodnout, zda má funkce v tomto bodě lokální extrém, (c) funkce má v tomto bodě lokální minimum

10.2.a) Stacionární body jsou $(0,-1)$ (sedlo), $(0,1)$ (sedlo), $(-1/\sqrt{2},0)$ (minimum), $(1/\sqrt{2},0)$ (maximum).

10.2.c) Stacionární body jsou $(0,0)$ (sedlo), $(0,-2)$ (maximum).

10.2.d) Stacionární body jsou $(-1,0)$ (minimum), $(1,0)$ (maximum), $(0,-1)$ (sedlo), $(0,1)$ (sedlo).

10.2.e) Stacionární body jsou $(0,0)$ (sedlo), $(1,1)$ (maximum), $(1,-1)$ (maximum).

10.2.f) Stacionárních bodů je 5.

10.2.g) Stacionární body jsou $(0,0,0)$ (z podmínek druhého řádu nelze určit, zda je to lok. extrém) a $(3/2,3/2,-9/4)$ (sedlo).

10.3. Funkce je součtem funkcí jedné proměnné, $f(\mathbf{x}) = \sum_i g_i(x_i)$ kde $g_i(x) = a_i x - x\ln x$. Tedy hledání extrémů funkce f se dá převést na nezávislé hledání extrémů funkcí g_i, viz Cvičení 1.4. Přesněji, f bude mít lokální/globální maximum/minimum v bodě $\mathbf{x} = (x_1, \ldots, x_n)$, právě když každá g_i bude mít lokální/globální maximum/minimum v bodě x_i. Je $g_i'(x) = a_i - \ln x - 1 = 0$, tedy $x = \mathrm{e}^{a_i - 1}$. Není těžké ověřit, že jde o globální maximum. Shrnuto: f má jediný lokální a zároveň globální extrém, a to maximum v bodě $\mathbf{x} = (\mathrm{e}^{a_1-1}, \ldots, \mathrm{e}^{a_n-1})$.

10.4. Stacionární podmínka je $f'(\mathbf{x})^T = \mathbf{a} - \mathbf{b}/(\mathbf{b}^T\mathbf{x})^2 = \mathbf{0}$. Tuto rovnici musíme vyřešit, tj. určit, pro jaká $\mathbf{a}, \mathbf{b}$ má řešení a jaká je v tom případě její množina řešení. Nemůže být $\mathbf{a} = \mathbf{b} = \mathbf{0}$, protože $\mathbf{b} = \mathbf{0}$ je zakázáno v zadání (jinak by f nebyla definována). Jistě musí být $\mathbf{b}^T\mathbf{x} \neq \mathbf{0}$, protože jinak by také $f(\mathbf{x})$ nebylo definováno. Pro každé $\alpha \geq 0$ existuje $\mathbf{x}$ takové, že $(\mathbf{b}^T\mathbf{x})^2 = \alpha$ (proč?). Proto má rovnice řešení právě tehdy, když $\mathbf{b} \neq \mathbf{0}$ a existuje skalár $\alpha > 0$ takový, že $\mathbf{a} = \mathbf{b}/\alpha$. Neboli vektory $\mathbf{a}, \mathbf{b}$ jsou nenulové, rovnoběžné a mají stejný směr. (Tento výsledek je intuitivně přijatelný, představíme-li $n = 2$. Pro $\mathbf{a} = (1, 0)$ a $\mathbf{b} = (0, 1)$ (tj. nejsou rovnoběžné) dostaneme funkci $f(x, y) = x + 1/y$, která je očividně neomezená. Pro $\mathbf{a} = (1, 0)$ a $\mathbf{b} = -\mathbf{a}$ (jsou rovnoběžné ale mají opačný směr) dostaneme $f(x, y) = x - 1/x$, a ta je také neomezená (načrtněte si graf).)

Za této podmínky můžeme naši funkci napsat jako $f(\mathbf{x}) = \mathbf{b}^T\mathbf{x}/\alpha + 1/(\mathbf{b}^T\mathbf{x})$, kde $\mathbf{b} \in \mathbb{R}^n$ a $\alpha > 0$ jsou známé. Tato funkce závisí jen na součinu $\mathbf{b}^T\mathbf{x}$ (můžeme se totiž pohybovat jen po přímce dané společným směrem vektorů $\mathbf{a}, \mathbf{b}$). Označíme-li $\mathbf{b}^T\mathbf{x} = y$, je $f(\mathbf{x}) = g(\mathbf{b}^T\mathbf{x})$ kde $g(y) = y/\alpha + 1/y$. Nyní stačí najít extrémy funkce g. Stacionární podmínka je $g'(y) = 1/\alpha - 1/y^2 = 0$. Tato rovnice má dvě řešení $y = \pm\alpha$. Pomocí jednoduchých úvah (načrtnutí grafu, druhá derivace) zjistíme, že kladný kořen je globální minimum a záporný je globální maximum.

Odpověď na otázku v zadání: funkce f má aspoň jeden lokální extrém právě tehdy, když $\mathbf{b} \neq \mathbf{0}$ a existuje skalár $\alpha > 0$ takový, že $\mathbf{a} = \mathbf{b}/\alpha$. Za této podmínky má funkce f lokální a zároveň globální minimum v bodech $\mathbf{x}$ splňujících $\mathbf{b}^T\mathbf{x} = \alpha$ a lokální a zároveň globální maximum v bodech $\mathbf{x}$ splňujících $\mathbf{b}^T\mathbf{x} = -\alpha$.

10.5. Jeden kořen je $x = 0$ a pak jsou dva další lišící se znaménkem. Jeden z nich získáme Newtonovou metodou: $x_{k+1} = x_k - (2\sin x_k - x_k)/(2\cos x_k - 1)$. Načrtneme si grafy funkcí $\sin x$ a $\frac{1}{2}x$ a z toho zvolíme počáteční odhad, např. $x_0 = 2$. Po několika iteracích máme $x_k = 1.895494267033981$.

10.7. Iterace pro $g(x) = 1/x - a = 0$ je $x_{k+1} = x_k - g(x_k)/g'(x_k) = x_k - (1/x_k - a)/(-1/x_k^2) = x_k(2 - ax_k)$.

10.8. (a) Minimalizujeme $\mathbf{g}(x)^T\mathbf{g}(x)$ kde $\mathbf{g}(x) = \begin{bmatrix} x^2 + x - 1 \\ x^2 - x - 1 \end{bmatrix}$. Máme $\mathbf{g}'(x) = \begin{bmatrix} 2x + 1 \\ 2x - 1 \end{bmatrix}$. Iterace je

$$x \leftarrow x - (\mathbf{g}'(x)^T\mathbf{g}'(x))^{-1}\mathbf{g}'(x)^T\mathbf{g}(x) = x - (8x^2+2)^{-1} \begin{bmatrix} 2x + 1 & 2x - 1 \end{bmatrix} \begin{bmatrix} x^2 + x - 1 \\ x^2 - x - 1 \end{bmatrix} = \frac{2x(x^2 + 1)}{4x^2 + 1}.$$

(b) Minimalizujeme $f(x) = \mathbf{g}(x)^T\mathbf{g}(x)) = (x^2 + x - 1)^2 + (x^2 - x - 1)^2 = 2(x^4 - x^2 + 1)$. Máme $f'(x) = 8x^3 - 4x$, $f''(x) = 24x^2 - 4$. Iterace je $x \leftarrow x - f'(x)/f''(x) = (4x^3)/(6x^2 - 1)$. Vidíme, že pro tuto jednoduchou soustavu s jednou proměnnou není Newtonova iterace složitější než Gaussova-Newtonova iterace – obvykle je to ale naopak.

10.9. Soustava je nelineární. Nemá řešení, protože po zavedení proměnné $xy = z$ dostaneme soustavu lineárních rovnic s řešením $(x, y, z) = (0.5, -1.5, -1)$, což je spor.

10.10. Nebude, vzdálenost by se musela počítat jako minimalizace $(x-u)^2+(y-v)^2$ za podmínek $x^2 = y$ a $(u-2)^2+v^2 = 1$. Protože se křivky neprotínají, tato formulace jde zjednodušit: vzdálenost bodu od kružnice je rovna vzdálenosti bodu od středu kružnice, tedy stačí minimalizovat $(x - 2)^2 + y^2$ za podmínky $x^2 = y$, což se zjednoduší na minimalizaci funkce jedné proměnné $(x - 2)^2 + x^4$.

10.11. Po přejmenování proměnných $\mathbf{x}_k, \mathbf{x}_{k+1}$ na $\mathbf{x}, \mathbf{y}$ máme ukázat, že stacionární bod $\mathbf{y}$ Taylorova polynomu (8.23c) je právě bod splňující $\mathbf{g}(\mathbf{x}) + \mathbf{g}'(\mathbf{x})(\mathbf{y} - \mathbf{x}) = \mathbf{0}$ kde $\mathbf{g}(\mathbf{x}) = f'(\mathbf{x})^T$. To se snadno dokáže výpočtem derivace polynomu (s užitím Cvičení 8.12).

10.14. Nechť střed kružnice je $\mathbf{c} \in \mathbb{R}^3$ a poloměr $r \in \mathbb{R}$. Reprezentujme rovinu kružnice jako $\mathbf{a}^T(\mathbf{x}-\mathbf{c}) = 0$ kde $\mathbf{0} \neq \mathbf{a} \in \mathbb{R}^3$ je normála roviny (střed $\mathbf{c}$ leží v této rovině). Jak šikovně spočítat vzdálenost nějakého bodu $\mathbf{x} \in \mathbb{R}^3$ od kružnice. Když tento bod nejprve promítneme do roviny kružnice (dle §5.2), vzdálenost tohoto průmětu od kružnice se už spočítá snadno (jako v předchozím cvičení). Pak minimalizujeme součet čtverců vzdáleností bodů od kružnice přes neznámé $\mathbf{c}, r, \mathbf{a}$.

10.15. Pišme normální rovnici (10.15) jako $\mathbf{A}^T\mathbf{A}\mathbf{v} = \mathbf{A}^T\mathbf{b}$ kde $\mathbf{A} = \mathbf{g}'(\mathbf{x}_k)$, $\mathbf{b} = -\mathbf{g}(\mathbf{x}_k)$ a $\mathbf{v} = \mathbf{x}_{k+1} - \mathbf{x}_k$ je směr hledání. Je-li $\mathbf{v}$ řešení normální rovnice, pak $\mathbf{A}\mathbf{v} = \mathbf{P}\mathbf{b}$ je projekce vektoru $\mathbf{b}$ na podprostor $\operatorname{rng}\mathbf{A}$. Je $f'(\mathbf{x}_k) = -\mathbf{A}^T\mathbf{b}$, tedy podmínka na sestupnost směru $\mathbf{v}$ zní $\mathbf{b}^T\mathbf{A}\mathbf{v} = \mathbf{b}^T\mathbf{P}\mathbf{b} > 0$. Ale $\mathbf{P}$ je positivně semidefinitní (viz Cvičení 6.25), tedy $\mathbf{b}^T\mathbf{P}\mathbf{b} \geq 0$. Zde ovšem rovnost nastane, jen když $\mathbf{b} = \mathbf{0}$ nebo $\mathbf{A}^T\mathbf{b} = \mathbf{0}$. Tedy $\mathbf{v}$ je sestupný, za přirozených předpokladů $\mathbf{g}(\mathbf{x}_k) \neq \mathbf{0}$ a $f'(\mathbf{x}) \neq \mathbf{0}$.

Kapitola 11

Lokální extrémy vázané rovnostmi

Hledejme lokální extrémy funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ na množině

$$X = \{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{g}(\mathbf{x}) = \mathbf{0} \,\}, \tag{11.1}$$

kde $\mathbf{g} = (g_1, \dots, g_m)\colon \mathbb{R}^n \to \mathbb{R}^m$ je zobrazení se složkami $g_1, \dots, g_m$. To odpovídá úloze (1.11) (nebo její maximalizační verzi) s omezeními typu rovnosti:

$$\begin{aligned} \text{min/max} \quad & f(x_1, \dots, x_n) \\ \text{za podmínek} \quad & g_i(x_1, \dots, x_n) = 0, \quad i = 1, \dots, m. \end{aligned} \tag{11.2}$$

Mluvíme o extrémech funkce f *vázaných rovnostmi* $\mathbf{g}(\mathbf{x}) = \mathbf{0}$.

Množina X je množina řešení soustavy m rovnic (obecně nelineárních) o n neznámých. Tato množina obvykle nemá žádné vnitřní body, proto nelze použít podmínky na volné lokální extrémy z §10.1 (viz Tvrzení 9.3). Někdy ovšem lze vyjádřit všechna řešení soustavy $\mathbf{g}(\mathbf{x}) = \mathbf{0}$ parametricky (neboli parametrizovat množinu X) a úlohu tak převést na úlohu bez omezení. To jsme použili v Příkladu 1.2, zde je další příklad:

Příklad 11.1. Řešme úlohu

$$\begin{aligned} \max \quad & x_1 + x_2 \\ \text{za podmínky} \quad & x_1^2 + x_2^2 = 1 \end{aligned} \tag{11.3}$$

tedy máme $m = 1$ a $n = 2$ a maximalizujme účelovou funkci $f(x_1, x_2) = x_1 + x_2$ na kružnici $X = \{\, (x_1, x_2) \in \mathbb{R}^2 \mid g(x_1, x_2) = x_1^2 + x_2^2 - 1 = 0 \,\}$. Množinu X lze parametrizovat jako

$$X = \{\, (\cos\alpha, \sin\alpha) \mid \alpha \in \mathbb{R} \,\}, \tag{11.4}$$

což převede úlohu na maximalizaci funkce $\varphi(\alpha) = f(\cos\alpha, \sin\alpha) = \cos\alpha + \sin\alpha$ na množině $\mathbb{R}$. Je-li α lokální extrém funkce φ, pak dle Věty 10.3 je $\varphi'(\alpha) = -\sin\alpha + \cos\alpha = 0$, což má řešení $\alpha = \frac{\pi}{4} + k\pi$. To odpovídá bodům $(x_1, x_2) = \pm\frac{\sqrt{2}}{2}(1, 1)$. To jsou tedy body podezřelé z lokálního extrému. ♦

Někdy ovšem množinu (11.1) parametrizovat nelze, nebo je to složité nebo nevýhodné.

Příklad 11.2. Zobecněme úlohu (11.3): pro dané $\mathbf{a} \in \mathbb{R}^n$ řešme

$$\begin{aligned} \max \quad & \mathbf{a}^T\mathbf{x} \\ \text{za podmínky} \quad & \mathbf{x}^T\mathbf{x} = 1 \end{aligned} \tag{11.5}$$

Maximalizujeme lineární funkci $f(\mathbf{x}) = \mathbf{a}^T\mathbf{x}$ na množině $X = \{\, \mathbf{x} \in \mathbb{R}^n \mid g(\mathbf{x}) = \mathbf{x}^T\mathbf{x} - 1 = 0 \,\}$, což je n-rozměrná sféra. Zde už není vůbec jasné, jak bychom mohli parametrizovat množinu X podobně jako v (11.4). Přitom řešení úlohy (11.5) po krátkém zamyšlení uhodnete (nakreslete si obrázky pro $n = 1, 2, 3$): maximum se nabývá pro $\mathbf{x} = \mathbf{a}/\|\mathbf{a}\|$.

Asi vás napadne vyjádřit jednu proměnnou (např. x_n) z podmínky $\mathbf{x}^T\mathbf{x} = x_1^2 + \cdots + x_n^2 = 1$ a dosadit ji do účelové funkce $\mathbf{a}^T\mathbf{x} = a_1x_1 + \cdots + a_nx_n$, čímž dostaneme úlohu bez omezení. Musíme ale uvažovat zvlášť dva případy $x_n = \pm\sqrt{1 - x_1^2 - \cdots - x_{n-1}^2}$. Tento způsob řešení je složitý a nehezký už proto, že nás donutil pracovat se složkami vektorů $\mathbf{a}, \mathbf{x}$ (téměř vždy je lépe chovat se k vektorům a maticím jako k nedělitelným objektům). ♦

Dále uvedeme jiné podmínky na lokální extrémy vázané rovnostmi, vyjádřené s pomocí jistých pomocných proměnných, tzv. *Lagrangeových multiplikátorů.*

11.1 Lineární omezení

Uvažujme nejprve důležitý speciální případ, kdy zobrazení $\mathbf{g}$ je afinní, tj. $\mathbf{g}(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b}$, kde $\mathbf{A} \in \mathbb{R}^{m\times n}$. Dle Věty 3.13 je tedy množina (11.1) afinní podprostor $\mathbb{R}^n$. Úlohu můžeme psát jako

$$\min/\max \{\, f(\mathbf{x}) \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{A}\mathbf{x} = \mathbf{b} \,\}, \tag{11.6}$$

tj. minimalizujeme nebo maximalizujeme funkci f za podmínek lineárních rovností. Větu 10.2 snadno zobecníme pro tento případ:

Věta 11.1. Nechť funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ je diferencovatelná v bodě $\mathbf{x}^* \in \mathbb{R}^n$ ve směru $\mathbf{v} \in \operatorname{null} \mathbf{A}$.

- Jestliže $\mathbf{x}^*$ je lokální minimum funkce f vázané podmínkou $\mathbf{A}\mathbf{x} = \mathbf{b}$, pak $f_{\mathbf{v}}(\mathbf{x}^*) \geq 0$.
- Jestliže $\mathbf{x}^*$ je lokální maximum funkce f vázané podmínkou $\mathbf{A}\mathbf{x} = \mathbf{b}$, pak $f_{\mathbf{v}}(\mathbf{x}^*) \leq 0$.

Důkaz. Nechť $\mathbf{A}\mathbf{x}^* = \mathbf{b}$ a $\mathbf{v} \in \operatorname{null} \mathbf{A}$ (tj. $\mathbf{A}\mathbf{v} = \mathbf{0}$). Pak pro každé $\alpha \in \mathbb{R}$ je $\mathbf{A}(\mathbf{x}^* + \alpha\mathbf{v}) = \mathbf{b}$, tedy $\mathbf{x}^* + \alpha\mathbf{v} \in X$. Jestliže $f_{\mathbf{v}}(\mathbf{x}^*) < 0$, pak tedy dle Lemmatu 10.1 pro každé $\varepsilon > 0$ existuje $\alpha > 0$ tak, že $\mathbf{x}^* + \alpha\mathbf{v} \in X \cap B_\varepsilon(\mathbf{x}^*)$ a $f(\mathbf{x}^* + \alpha\mathbf{v}) < f(\mathbf{x}^*)$. Tedy $\mathbf{x}^*$ není lokální minimum funkce f na množině X (viz definice lokálního minima v §9.3). Obdobně pro maximum. ■

Důkaz má jasný geometrický význam: jestliže v nějakém bodě je směr $\mathbf{v} \in \operatorname{null} \mathbf{A}$ sestupný [vzestupný] pro f, pak hodnotu f můžeme zmenšit [zvětšit] malým posunutím bodu ve směru $\mathbf{v}$, čímž neopustíme afinní podprostor X. Proto bod nemůže být lokální minimum [maximum].

Pro (totálně) diferencovatelnou funkci f zobecníme Větu 10.3:

Věta 11.2. Nechť funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ je (totálně) diferencovatelná v bodě $\mathbf{x}^* \in \mathbb{R}^n$. Jestliže $\mathbf{x}^*$ je lokální extrém funkce f vázaný podmínkou $\mathbf{A}\mathbf{x} = \mathbf{b}$, pak $\nabla f(\mathbf{x}^*) \perp \operatorname{null} \mathbf{A}$.

Důkaz. Dle Věty 8.5 je $f_{\mathbf{v}}(\mathbf{x}^*) = f'(\mathbf{x}^*)\mathbf{v}$. Protože $\operatorname{null} \mathbf{A}$ je podprostor, $\mathbf{v} \in \operatorname{null} \mathbf{A}$ platí právě když $-\mathbf{v} \in \operatorname{null} \mathbf{A}$. Jestliže $\mathbf{x}^*$ je lokální minimum f vázané podmínkou $\mathbf{A}\mathbf{x} = \mathbf{b}$, pak dle Věty 11.1 pro každé $\mathbf{v} \in \operatorname{null} \mathbf{A}$ platí $f_{\mathbf{v}}(\mathbf{x}^*) = f'(\mathbf{x}^*)\mathbf{v} \geq 0$ a $f_{-\mathbf{v}}(\mathbf{x}^*) = -f'(\mathbf{x}^*)\mathbf{v} \geq 0$, tedy $f'(\mathbf{x}^*)\mathbf{v} = 0$. To znamená, že vektor $\nabla f(\mathbf{x}^*) = f'(\mathbf{x}^*)^T$ je kolmý na podprostor $\operatorname{null} \mathbf{A}$. ■

Uvedeme ještě jiný důkaz Věty 11.2 (který nám umožní dokázat podmínky druhého řádu):

Důkaz. Bod $\mathbf{x}^*$ splňuje omezení $\mathbf{Ax} = \mathbf{b}$, tedy je (partikulárním) řešením soustavy $\mathbf{Ax} = \mathbf{b}$. Množinu řešení soustavy proto lze psát jako $\mathbf{x}^* + \operatorname{null}\mathbf{A}$ (viz §3.3), tj. lze ji parametrizovat jako $\mathbf{x} = \mathbf{x}^* + \mathbf{By}$ kde sloupce matice $\mathbf{B}$ tvoří bázi podprostoru $\operatorname{null}\mathbf{A}$. Protože $\mathbf{x}^*$ je lokální extrém funkce f za podmínky $\mathbf{Ax} = \mathbf{b}$, musí být bod $\mathbf{y} = \mathbf{0}$ volný lokální extrém funkce

$$\varphi(\mathbf{y}) = f(\mathbf{x}^* + \mathbf{By}). \tag{11.7}$$

Dle Věty 10.3 tedy

$$\varphi'(\mathbf{y}) = f'(\mathbf{x}^* + \mathbf{By})\mathbf{B} = f'(\mathbf{x}^*)\mathbf{B} = \mathbf{0}, \tag{11.8}$$

kde jsme užili řetízkové pravidlo (jako v Příkladě 8.13). To říká, že gradient $\nabla f(\mathbf{x}^*) = f'(\mathbf{x}^*)^T$ je kolmý na podprostor $\operatorname{rng}\mathbf{B} = \operatorname{null}\mathbf{A}$. ■

Podmínka $\nabla f(\mathbf{x}^*) \perp \operatorname{null}\mathbf{A}$ ve Větě 11.2 znamená, že gradient $\nabla f(\mathbf{x}^*)$ je kolmý na afinní podprostor X. Lze ji podrobněji napsat jako (viz Věta 4.3)

$$\nabla f(\mathbf{x}^*) \in (\operatorname{null}\mathbf{A})^\perp = \operatorname{rng}(\mathbf{A}^T). \tag{11.9}$$

To říká, že vektor $\nabla f(\mathbf{x}^*) = f'(\mathbf{x}^*)^T$ je lineární kombinací řádků matice $\mathbf{A}$, neboli $f'(\mathbf{x}^*) = \boldsymbol{\lambda}^T\mathbf{A}$ pro nějaké $\boldsymbol{\lambda} \in \mathbb{R}^m$. Tedy $(\mathbf{x}^*, \boldsymbol{\lambda})$ je řešením soustavy

$$f'(\mathbf{x}) = \boldsymbol{\lambda}^T\mathbf{A}, \tag{11.10a}$$
$$\mathbf{Ax} = \mathbf{b}. \tag{11.10b}$$

Tato soustava má $m + n$ rovnic a $m + n$ neznámých. Zanedlouho uvidíme (v Příkladu 11.15), že prvky vektoru $\boldsymbol{\lambda}$ jsou *Lagrangeovy multiplikátory*.

Příklad 11.3. Vraťme se k úloze (5.32), tedy k hledání řešení nehomogenní soustavy lineárních rovnic s nejmenší normou. Místo funkce $\mathbf{x}^T\mathbf{x}$ budeme minimalizovat funkci $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{x}$, což úlohu nezmění[1]. Je $f'(\mathbf{x}) = \mathbf{x}^T$, tedy rovnost (11.10a) je (po transpozici) $\mathbf{x} = \mathbf{A}^T\boldsymbol{\lambda}$. Soustava (11.10) je tedy soustava (5.34), kterou jsme v §5.4 odvodili úvahou. ◆

Příklad 11.4. Obecněji, řešme úlohu

$$\begin{array}{rl} \min & \frac{1}{2}\|\mathbf{Ax} - \mathbf{b}\|^2 \\ \text{za podmínek} & \mathbf{Cx} = \mathbf{d} \end{array} \tag{11.11}$$

Tedy řešíme úlohu nejmenších čtverců (5.2) s lineárními omezeními (angl. *linearly constrained least squares*). Tato úloha má hodně aplikací.

Je (ověřte!)

$$f'(\mathbf{x}) = \mathbf{x}^T\mathbf{A}^T\mathbf{A} - \mathbf{b}^T\mathbf{A}.$$

Tedy podmínka stacionarity (11.10) je (po transpozici první rovnice)

$$\mathbf{A}^T\mathbf{Ax} - \mathbf{A}^T\mathbf{b} = \mathbf{C}^T\boldsymbol{\lambda} \tag{11.12a}$$
$$\mathbf{Cx} = \mathbf{d}, \tag{11.12b}$$

neboli

$$\begin{bmatrix} \mathbf{A}^T\mathbf{A} & -\mathbf{C}^T \\ \mathbf{C} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^T\mathbf{b} \\ \mathbf{d} \end{bmatrix}.$$

Tuto soustavu lineárních rovnic vyřešíme jednou ze známých metod (k tomu viz Cvičení 11.24). ◆

[1] Přidání činitele $\frac{1}{2}$ ke čtverci eukleidovské normy v kritériu je obvyklý trik. Výhoda je, že derivaci této funkce je jednoduše $\mathbf{x}^T$, tedy bez dvojky.

Příklad 11.5. ($\star$) Při řešení lineární úlohy nejmenších čtverců (5.2) jsme dospěli k pojmu *pseudoinverze* matice $\mathbf{A}$ s lineárně nezávislými sloupci, definované jako $\mathbf{A}^+ = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$. Podotkli jsme, že je to jedna z levých inverzí matice $\mathbf{A}$, neboť $\mathbf{A}^+\mathbf{A} = \mathbf{I}$. Nyní ukážeme, že $\mathbf{A}^+$ má ze všech levých inverzí matice $\mathbf{A}$ nejmenší (Frobeniovu) normu. Hledejme tedy matici $\mathbf{X}$, která minimalizuje funkci $f(\mathbf{X}) = \frac{1}{2}\|\mathbf{X}\|^2$ za podmínky $\mathbf{XA} = \mathbf{I}$. Tato úloha je nezvyklá v tom, že její proměnná je matice a ne vektor (takové úlohy úlohy jsme ovšem již potkali, např. úloha na největší stopu (7.3)).

Dle třetí tabulky v §8.5.2 je $f'(\mathbf{X}) = \mathbf{X}^T$. Stacionární podmínky (11.10) jsou

$$\begin{aligned}\mathbf{X}^T &= \mathbf{A}\mathbf{\Lambda}^T,\\ \mathbf{XA} &= \mathbf{I}.\end{aligned}$$

Lagrangeovy jsou zde v matici $\mathbf{\Lambda}$ (a ne ve vektoru λ) a první rovnici jsme napsali maticově (to vyžaduje zkušenost s lineárními zobrazeními matic, která je nad rámec tohoto kursu, proto je tento příklad nepovinný). Z první rovnice máme $\mathbf{X} = \mathbf{\Lambda}\mathbf{A}^T$, což dosazeno do druhé podmínky dá $\mathbf{\Lambda}\mathbf{A}^T\mathbf{A} = \mathbf{I}$. Z toho $\mathbf{\Lambda} = (\mathbf{A}^T\mathbf{A})^{-1}$, což dosadíme zpět do první rovnice a dostaneme $\mathbf{X} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$. ♦

($\star$) Podmínky druhého řádu

Věta 11.2 udává nutnou podmínku prvního řádu na lokální extrém vázaný lineárními rovnostmi. Uvedeme nyní podmínky druhého řádu. K tomu potřebujeme nový pojem: matice $\mathbf{C} \in \mathbb{R}^{n\times n}$ je *positivně semidefinitní na podprostoru* $Y \subseteq \mathbb{R}^n$, jestliže $\mathbf{x}^T\mathbf{C}\mathbf{x} \geq 0$ pro každé $\mathbf{x} \in Y$. Jak tuto podmínku ověříme? Najdeme libovolnou matici $\mathbf{B} \in \mathbb{R}^{n\times m}$ tak, že $Y = \operatorname{rng}\mathbf{B}$ (např. sloupce $\mathbf{B}$ tvoří bázi podprostoru Y, v tom případě bude tedy $m = \dim Y$). Pak $\mathbf{x} \in Y$ právě když $\mathbf{x} = \mathbf{By}$ pro nějaké $\mathbf{y} \in \mathbb{R}^m$. Protože $\mathbf{x}^T\mathbf{C}\mathbf{x} = \mathbf{y}^T\mathbf{B}^T\mathbf{C}\mathbf{B}\mathbf{y}$, převedli jsme problém na ověřování positivní semidefinitnosti matice $\mathbf{B}^T\mathbf{C}\mathbf{B} \in \mathbb{R}^{m\times m}$ (na celém prostoru $\mathbb{R}^m$). Podobně definujeme positivní a negativní (semi)definitnost a indefinitnost matice na podprostoru.

Věta 11.3. Nechť funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ je dvakrát diferencovatelná v bodě $\mathbf{x}^* \in \mathbb{R}^n$.

- Jestliže $\mathbf{x}^*$ je lokální minimum [maximum] funkce f vázané podmínkou $\mathbf{Ax} = \mathbf{b}$, pak existuje $\boldsymbol{\lambda}$ tak že $(\mathbf{x}^*, \boldsymbol{\lambda})$ splňují soustavu (11.10) a Hessova matice $f''(\mathbf{x}^*)$ je positivně [negativně] semidefinitní na podprostoru $\operatorname{null}\mathbf{A}$.
- Jestliže existuje $\boldsymbol{\lambda}$ tak že $(\mathbf{x}^*, \boldsymbol{\lambda})$ splňují soustavu (11.10) a Hessova matice $f''(\mathbf{x}^*)$ je positivně [negativně] definitní na podprostoru $\operatorname{null}\mathbf{A}$, pak $\mathbf{x}^*$ je ostré lokální minimum [maximum] funkce f vázané podmínkou $\mathbf{Ax} = \mathbf{b}$.

Důkaz. V druhém důkazu Věty 11.2 jsme ukázali, že $\mathbf{x} = \mathbf{x}^* + \mathbf{By}$ je lokální extrém funkce f vázaný podmínkou $\mathbf{Ax} = \mathbf{b}$, právě když $\mathbf{y} = \mathbf{0}$ je lokální extrém funkce φ bez omezení. Je (viz §8.8)

$$\varphi''(\mathbf{y}) = \mathbf{B}^T f''(\mathbf{x}^* + \mathbf{By})\mathbf{B} = \mathbf{B}^T f''(\mathbf{x}^*)\mathbf{B}.$$

Zbytek plyne z podmínek druhého řádu ve Větě 10.4, použité na funkci φ. ∎

Všimněte si, že pro ověření podmínek potřebujeme najít matici $\mathbf{B}$ z parametrizace (11.7). Té jsme se dokázali zbavit při odvození podmínek prvního řádu (ve druhém důkazu Věty 11.1), ale v podmínkách druhého řádu ji potřebujeme. Jinými slovy, Věta 11.3 je vlastně stejná, jako podmínka druhého řádu pro minimalizaci funkce (11.7) bez omezení dle Věty 10.4.

V Příkladech 11.3 a 11.4 byla Hessova matice $f''(\mathbf{x})$ positivně semidefinitní, tedy byl positivně semidefinitní i na každém podprostoru. Zde je příklad, kdy tomu tak není:

Příklad 11.6. Řešme úlohu

$$\begin{aligned} \max \quad & x_1x_2 + x_2x_3 + x_1x_3 \\ \text{za podmínky} \quad & x_1 + x_2 + x_3 = 3. \end{aligned} \tag{11.13}$$

Řešení podmínek prvního řádu (11.10) je $x_1 = x_2 = x_3 = 1$, $\lambda = 2$. Hessova matice

$$f''(x_1, x_2, x_3) = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

je indefinitní (na $\mathbb{R}^3$). Zjistíme její definitnost na podprostoru null $\mathbf{A}$. Máme

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{bmatrix}, \quad \mathbf{B}^T f''(x_1, x_2, x_3)\mathbf{B} = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix}$$

kde sloupce matice $\mathbf{B}$ tvoří (libovolná) bázi podprostoru null $\mathbf{A}$. Poslední matice je negativně definitní, tedy $f''(x_1, x_2, x_3)$ je negativně definitní na podprostoru null $\mathbf{A}$, tedy $(x_1, x_2, x_3) = (1, 1, 1)$ je (ostré) lokální maximum úlohy. ♦

11.2 Nelineární omezení

Přejděme nyní k obecnému případu, kdy $\mathbf{g}$ je libovolné (ne nutně afinní) diferencovatelné zobrazení. Podmínka prvního řádu na lokální extrémy pro tato omezení bude podobná jako pro lineární omezení, ale na rozdíl od Věty 11.2 ji nedokážeme, protože důkaz by byl příliš dlouhý. Pouze vysvětlíme geometrický význam této podmínky a uvedeme příklady.

11.2.1 Tečný prostor

Je-li zobrazení $\mathbf{g}$ v nějakém bodě $\mathbf{x}$ diferencovatelné, můžeme ho v okolí bodu $\mathbf{x}$ aproximovat jeho Taylorovým polynomem prvního stupně (8.24)

$$\mathbf{g}(\mathbf{y}) \approx \mathbf{T}^1_{\mathbf{x}}(\mathbf{y}) = \mathbf{g}(\mathbf{x}) + \mathbf{g}'(\mathbf{x})(\mathbf{y} - \mathbf{x}) = \mathbf{g}'(\mathbf{x})(\mathbf{y} - \mathbf{x}), \tag{11.14}$$

kde $\mathbf{g}(\mathbf{x}) = \mathbf{0}$ neboť $\mathbf{x} \in X$. Množina X se tím změní na

$$\{\, \mathbf{y} \in \mathbb{R}^n \mid \mathbf{g}'(\mathbf{x})(\mathbf{y} - \mathbf{x}) = \mathbf{0} \,\} \;=\; \{\, \mathbf{x} + \mathbf{y}' \mid \mathbf{y}' \in \mathbb{R}^n,\ \mathbf{g}'(\mathbf{x})\mathbf{y}' = \mathbf{0} \,\} \;=\; \mathbf{x} + \operatorname{null} \mathbf{g}'(\mathbf{x}), \tag{11.15}$$

kde jsme udělali substituci $\mathbf{y}' = \mathbf{y} - \mathbf{x}$. Množina (11.15) je afinní podprostor, je to (lineární) podprostor null $\mathbf{g}'(\mathbf{x})$ posunutý[2] do bodu $\mathbf{x}$ (viz §3.3). Protože z definice totální derivace (viz §8.5) se zobrazení $\mathbf{g}$ v okolí bodu $\mathbf{x}$ podobá afinnímu zobrazení (11.14), mohli bychom doufat, že množina X se v okolí bodu $\mathbf{x}$ bude podobat afinnímu podprostoru (11.15). Tak tomu ale překvapivě být nemusí (jak uvidíme na příkladech), je k tomu třeba následující dodatečná podmínka.

Bod $\mathbf{x} \in \mathbb{R}^n$ nazveme **regulární bod** zobrazení $\mathbf{g}\colon \mathbb{R}^n \to \mathbb{R}^m$, jestliže zobrazení $\mathbf{g}$ je v bodě $\mathbf{x}$ spojitě diferencovatelné a Jacobiho matice $\mathbf{g}'(\mathbf{x})$ má lineárně nezávislé řádky (tj. hodnost m). Připomeňme (viz §8.5), že řádky matice $\mathbf{g}'(\mathbf{x})$ jsou transponované gradienty $\nabla g_1(\mathbf{x}), \ldots, \nabla g_m(\mathbf{x})$ složek zobrazení $\mathbf{g}$ v bodě $\mathbf{x}$.

[2] Zde by vám mělo být jasné, že '+' ve výrazu $\mathbf{x} + \operatorname{null} \mathbf{g}'(\mathbf{x})$ označuje operaci definovanou v (3.23).

Pozorování 11.4. Je-li $\mathbf{x} \in X$ regulární bod zobrazení $\mathbf{g}$, pak množina X je v okolí bodu $\mathbf{x}$ podobná afinnímu podprostoru
$$\mathbf{x} + \operatorname{null} \mathbf{g}'(\mathbf{x}) \tag{11.16}$$
dimenze $n - m$. Tento podprostor je tečný k množině X v bodě $\mathbf{x}$.

Tento fakt uvádíme jen jako neformální pozorování[3], pro jeho formalizaci bychom předtím museli definovat pojmy 'podobná' a 'tečný'. Jeho význam intuitivně pochopíte na příkladech. Obvykle se za **tečný prostor** k množině X v bodě $\mathbf{x}$ považuje přímo prostor $\operatorname{null} \mathbf{g}'(\mathbf{x})$ a posunutí $\mathbf{x}$ se neuvádí (protože každému je jasné, že tečný prostor prochází bodem $\mathbf{x}$).

Příklad 11.7. Nechť
$$g\colon \mathbb{R}^n \to \mathbb{R}, \qquad g(\mathbf{x}) = \mathbf{x}^T\mathbf{x} - 1. \tag{11.17}$$
Množina X je jednotková n-rozměrná sféra. Pro každé $\mathbf{x} \in X$ je $g'(\mathbf{x})^T = \nabla g(\mathbf{x}) = 2\mathbf{x} \neq \mathbf{0}$, tedy každý bod na sféře je regulární bod[4] funkce g. Tečný prostor k množině X v bodě $\mathbf{x} \in X$ je nadrovina $\operatorname{null}(\mathbf{x}^T) = \{\, \mathbf{y} \in \mathbb{R}^n \mid \mathbf{x}^T\mathbf{y} = 0 \,\}$. Posuneme-li ho do bodu $\mathbf{x}$, je to nadrovina $\mathbf{x} + \operatorname{null}(\mathbf{x}^T) = \{\, \mathbf{y} \in \mathbb{R}^n \mid \mathbf{x}^T\mathbf{y} = 1 \,\}$ (odvoďte dle (11.15)!).

Pro $n = 1$ množina $X = \{-1, 1\}$ obsahuje jen dva body a tečný prostor v kterémkoliv z těchto bodů je tento bod sám. Pro $n = 2$ je množina X kružnice v $\mathbb{R}^2$ a tečný prostor v bodě $\mathbf{x} = (x_1, x_2) \in X$ je tečna k této kružnici. Pro $n = 3$ je množina X obyčejná sféra v $\mathbb{R}^3$ a tečný prostor v bodě $\mathbf{x} = (x_1, x_2, x_3) \in X$ je tečná rovina ke sféře v tomto bodě. ♦

Příklad 11.8. Nechť
$$\mathbf{g}\colon \mathbb{R}^3 \to \mathbb{R}^2, \qquad \mathbf{g}(x_1, x_2, x_3) = (\, x_1^2 + x_2^2 + x_3^2 - 1,\ (x_1 - 1)^2 + x_2^2 + x_3^2 - 1 \,).$$
Množina X je průnik dvou jednotkových sfér v $\mathbb{R}^3$ se středy $(0,0,0)$ a $(1,0,0)$. Tento průnik je kružnice v $\mathbb{R}^3$. Z geometrie je vidět (algebraicky to nebudeme dokazovat), že pro každý bod $(x_1, x_2, x_3) \in X$ jsou vektory $\nabla g_1(x_1, x_2, x_3) = 2(x_1, x_2, x_3)$ a $\nabla g_2(x_1, x_2, x_3) = 2(x_1 - 1, x_2, x_3)$ lineárně nezávislé, tedy všechny body na X jsou regulární body zobrazení $\mathbf{g}$. Tečný prostor k množině X v bodě $(x_1, x_2, x_3) \in X$ je přímka
$$\operatorname{null} \begin{bmatrix} x_1 & x_2 & x_3 \\ x_1 - 1 & x_2 & x_3 \end{bmatrix} = \operatorname{span}\{(0, x_3, -x_2)\}$$
tečná ke kružnici. Posunutá do bodu (x_1, x_2, x_3) je tato přímka afinní podprostor
$$(x_1, x_2, x_3) + \operatorname{span}\{(0, x_3, -x_2)\} = \{\, (x_1, x_2 + \alpha x_3, x_3 - \alpha x_2) \mid \alpha \in \mathbb{R} \,\}.$$
Tato přímka je průnik tečných rovin k oběma sférám v bodě (x_1, x_2, x_3). ♦

Ukažme nyní příklady, kdy bod $\mathbf{x} \in X$ není regulární bod zobrazení $\mathbf{g}$.

[3] To, že množina X je v okolí bodu $\mathbf{x} \in X$ 'podobná' afinnímu podprostoru, intuitivně znamená, že malinký mravenec lezoucí po množině X v blízkosti bodu $\mathbf{x}$ by nerozlišil, zda leze po ('zakřivené') množině X nebo po 'plochém' afinním podprostoru (11.16). Jsou-li pro nějaké $\varepsilon > 0$ všechny body množiny $X \cap B_\varepsilon(\mathbf{x})$ regulární, tato množina je 'hladký povrch' v $\mathbb{R}^n$ dimenze $n - m$. Studiem 'hladkých povrchů' se zabývá *diferenciální geometrie*.

[4] Uvědomte si, že lineární nezávislost řádků matice $g'(\mathbf{x}) \in \mathbb{R}^{1\times m}$ znamená, že její jediný řádek je nenulový.

Příklad 11.9. Nechť

$$\mathbf{g}\colon \mathbb{R}^2 \to \mathbb{R}^2, \qquad \mathbf{g}(x_1, x_2) = (\,(x_1+1)^2 + x_2^2 - 1,\ (x_1-1)^2 + x_2^2 - 1\,). \tag{11.18}$$

Množina X je průnik dvou jednotkových kružnic se středy $(-1, 0)$ a $(1, 0)$. Ty se protínají v jediném bodě $(x_1, x_2) = (0, 0)$. V tomto bodě jsou vektory $\nabla g_1(x_1, x_2) = 2(x_1+1, x_2) = (2, 0)$ a $\nabla g_2(x_1, x_2) = 2(x_1-1, x_2) = (-2, 0)$ lineárně závislé, tedy to není regulární bod zobrazení $\mathbf{g}$. Množina X obsahuje jediný bod, tedy je to afinní podprostor dimenze $n - m = 0$. Přesto podprostor $\operatorname{null} \mathbf{g}'(x_1, x_2) = \operatorname{span}\{(0, 1)\}$ není tečný k množině X (tečna k bodu je bod sám).♦

Příklad 11.10. Nechť

$$g\colon \mathbb{R}^2 \to \mathbb{R}, \qquad g(x_1, x_2) = x_1 x_2. \tag{11.19}$$

Množina X je sjednocení svislé a vodorovné osy. Bod $(x_1, x_2) = (0, 0)$ není regulární pro funkci g, neboť $\nabla g(x_1, x_2) = (x_2, x_1) = (0, 0)$. Množina X se v okolí tohoto bodu nepodobá žádnému afinnímu podprostoru. Prostor $\operatorname{null} \begin{bmatrix} 0 & 0 \end{bmatrix} = \mathbb{R}^2$ zjevně není tečný k množině X v bodě $(0, 0)$.♦

Může se také stát, že množina X je v okolí nějakého bodu $\mathbf{x} \in X$ podobná afinnímu podprostoru dimenze $n - m$, ale přesto tento bod není regulární a tedy prostor (11.16) v tom bodě není tečný k X.

Příklad 11.11. Nechť $g\colon \mathbb{R}^2 \to \mathbb{R}$ je funkce

$$g(x, y) = (x_1^2 + x_2^2 - 1)^2. \tag{11.20}$$

Protože pro každé číslo $z \in \mathbb{R}$ platí $z = 0 \;\Leftrightarrow\; z^2 = 0$, je

$$X \;=\; \{\,(x_1, x_2) \mid (x_1^2 + x_2^2 - 1)^2 = 0\,\} \;=\; \{\,(x_1, x_2) \mid x_1^2 + x_2^2 - 1 = 0\,\}.$$

Tedy množina X je kružnice (jako v Příkladě 11.7 pro $n = 2$). Všimněte si důležité věci: různá zobrazení $\mathbf{g}$ mohou definovat stejnou množinu X (proto jsme regularitu bodu definovali vzhledem k zobrazení $\mathbf{g}$ a ne vzhledem k množině X). Máme $\nabla g(x_1, x_2) = 4(x_1^2 + x_2^2 - 1)(x_1, x_2)$. Pro každý bod $(x, y) \in X$ je $\nabla g(x_1, x_2) = (0, 0)$, tedy bod (x_1, x_2) není regulární pro funkci g. Prostor $\operatorname{null} \begin{bmatrix} 0 & 0 \end{bmatrix} = \mathbb{R}^2$ zjevně není tečný ke kružnici. ♦

Příklad 11.12. Speciálně pro $m = 1$ je množina X vrstevnice funkce g nulové výšky a $g'(\mathbf{x})^T = \nabla g(\mathbf{x})$. Dostali jsme tedy tvrzení z §8.7, že gradient je kolmý k vrstevnici. ♦

11.2.2 Podmínka prvního řádu

S intuicí osvojenou v §11.2.1 by vás nyní nemělo překvapit, že Větu 11.2 lze zobecnit na (ne nutně afinní) diferencovatelná zobrazení $\mathbf{g}$ tak, že podprostor $\operatorname{null} \mathbf{A}$ nahradíme tečným podprostorem $\operatorname{null} \mathbf{g}'(\mathbf{x})$ k množině X v bodě lokálního extrému $\mathbf{x}$. Větu uvádíme bez důkazu:

Věta 11.5. Nechť funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ je (totálně) diferencovatelná v bodě $\mathbf{x} \in \mathbb{R}^n$.
Nechť $\mathbf{x}$ je regulární bod zobrazení $\mathbf{g}\colon \mathbb{R}^n \to \mathbb{R}^m$.
Jestliže $\mathbf{x}$ je lokální extrém funkce f vázaný podmínkou $\mathbf{g}(\mathbf{x}) = \mathbf{0}$, pak $\nabla f(\mathbf{x}) \perp \operatorname{null} \mathbf{g}'(\mathbf{x})$.

Podmínka $\nabla f(\mathbf{x}) \perp \operatorname{null} \mathbf{g}'(\mathbf{x})$ říká, že gradient $\nabla f(\mathbf{x})$ je kolmý k tečnému prostoru k množině X (tedy vlastně k množině X) v bodě $\mathbf{x}$. Lze ji podrobněji napsat jako (viz Věta 4.3)

$$\nabla f(\mathbf{x}) \in (\operatorname{null} \mathbf{g}'(\mathbf{x}))^\perp = \operatorname{rng}(\mathbf{g}'(\mathbf{x})^T) = \operatorname{span}\{\nabla g_1(\mathbf{x}), \dots, \nabla g_m(\mathbf{x})\}, \tag{11.21}$$

tedy gradient $\nabla f(\mathbf{x}) = f'(\mathbf{x})$ je lineární kombinací řádků matice $\mathbf{g}'(\mathbf{x})$. Lokální extrémy f vázaný podmínkou $\mathbf{g}(\mathbf{x}) = \mathbf{0}$ tedy splňují (za podmínek Věty 11.5)

$$f'(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{g}'(\mathbf{x}) = \mathbf{0}, \tag{11.22a}$$
$$\mathbf{g}(\mathbf{x}) = \mathbf{0} \tag{11.22b}$$

pro nějaká čísla $(\lambda_1, \dots, \lambda_m) = \boldsymbol{\lambda} \in \mathbb{R}^m$ (tzv. **Lagrangeovy multiplikátory**). Soustava (11.22) má $m + n$ rovnic a stejný počet neznámých.

Soustava (11.22) se často zapisuje pomocí **Lagrangeovy funkce** $L\colon \mathbb{R}^{n+m} \to \mathbb{R}$ s hodnotami

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{g}(\mathbf{x}) = f(\mathbf{x}) + \lambda_1 g_1(\mathbf{x}) + \dots + \lambda_m g_m(\mathbf{x}). \tag{11.23}$$

Rovnici (11.22a) pak lze psát jako $L_{\mathbf{x}}(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{0}$ a[5] rovnici (11.22b) jako $L_{\boldsymbol{\lambda}}(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{g}(\mathbf{x})^T = \mathbf{0}$. Tedy řešení $(\mathbf{x}, \boldsymbol{\lambda})$ soustavy (11.22) jsou stacionární bod Lagrangeovy funkce, tj. soustavu (11.22) lze psát jako $L'(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{0}$.

Příklad 11.13. Řešme znovu Příklad 11.1. Lagrangeova funkce je

$$L(x, y, \lambda) = x + y + \lambda(1 - x^2 - y^2).$$

Její stacionární body (x, y, λ) jsou řešeními soustavy tří rovnic o třech neznámých

$$\begin{aligned} L_x(x, y, \lambda) &= 1 - 2\lambda x &&= 0, \\ L_y(x, y, \lambda) &= 1 - 2\lambda y &&= 0, \\ L_\lambda(x, y, \lambda) &= 1 - x^2 - y^2 &&= 0. \end{aligned}$$

Kvůli podmínce $x^2 + y^2 = 1$ nemůžou být x ani y nulové, tedy (z prvních dvou rovnic) ani λ ne. První dvě rovnice dají $x = y = 1/(2\lambda)$. Dosazením do třetí máme $2/(2\lambda)^2 = 1$, což dá dva kořeny $\lambda = \pm 1/\sqrt{2}$. Stacionární body funkce L jsou dva, $(x, y, \lambda) = \pm(1, 1, 1)/\sqrt{2}$. Tedy máme dva kandidáty na lokální extrémy, $(x, y) = \pm(1, 1)/\sqrt{2}$. ♦

Příklad 11.14. Řešme Příklad 11.1, kde ale omezení změníme na $g(x, y) = (1 - x^2 - y^2)^2 = 0$. Podle Příkladu 11.11 máme $g'(x, y) = (0, 0)$ pro každé $(x, y) \in X$, čekáme tedy potíž.

Stacionární body Lagrangeovy funkce $L(x, y, \lambda) = x + y + \lambda(1 - x^2 - y^2)^2$ musí splňovat

$$\begin{aligned} L_x(x, y, \lambda) &= 1 - 4\lambda x(1 - x^2 - y^2) &&= 0, \\ L_y(x, y, \lambda) &= 1 - 4\lambda y(1 - x^2 - y^2) &&= 0, \\ L_\lambda(x, y, \lambda) &= (1 - x^2 - y^2)^2 &&= 0. \end{aligned}$$

Tyto rovnice si odporují: jelikož $1 - x^2 - y^2 = 0$, tak např. první rovnice říká $1 - 4\lambda x \cdot 0 = 0$, což neplatí pro žádné (x, λ). Závěr je, že lokální extrémy $(x, y) = \pm(1, 1)/\sqrt{2}$ jsme nenašli. ♦

[5] Výraz $L_{\mathbf{x}}(\mathbf{x}, \boldsymbol{\lambda})$ označuje řádkový vektor parciálních derivací funkce L podle $x_1, \dots, x_n$. Podobně $L_{\boldsymbol{\lambda}}(\mathbf{x}, \boldsymbol{\lambda})$.

Příklad 11.15. Získejme podmínky stacionarity (11.10) pro úlohu (11.6) s lineárními omezeními pomocí formalismu s Lagrangeovou funkcí. Máme

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T(\mathbf{b} - \mathbf{A}\mathbf{x})$$

tedy

$$\begin{aligned} L_{\mathbf{x}}(\mathbf{x}, \boldsymbol{\lambda}) &= f'(\mathbf{x}) - \boldsymbol{\lambda}^T\mathbf{A} = \mathbf{0}, \\ L_{\boldsymbol{\lambda}}(\mathbf{x}, \boldsymbol{\lambda}) &= \mathbf{b} - \mathbf{A}\mathbf{x} \qquad = \mathbf{0}. \end{aligned}$$ ♦

Příklad 11.16. Vraťme se k úloze (11.5). Máme

$$L(\mathbf{x}, \lambda) = \mathbf{a}^T\mathbf{x} + \tfrac{1}{2}\lambda(1 - \mathbf{x}^T\mathbf{x}),$$

kde jsme pro pohodlí napsali $\frac{1}{2}\lambda$ místo λ (proč to můžeme?). Je $L_{\mathbf{x}}(\mathbf{x}, \lambda) = \mathbf{a}^T - \lambda\mathbf{x}^T$, stacionární body funkce L tedy splňují soustavu

$$\mathbf{a} = \lambda\mathbf{x}, \quad \mathbf{x}^T\mathbf{x} = 1.$$

Rozlišíme dva případy:

- Je-li $\lambda \neq 0$, z první rovnice máme $\mathbf{x} = \mathbf{a}/\lambda$, což dosazeno do druhé dá $\mathbf{a}^T\mathbf{a}/\lambda^2 = 1$, tedy $\lambda = \pm\|\mathbf{a}\|$, tedy kandidáti na lokální extrémy jsou $\mathbf{x} = \pm\mathbf{a}/\|\mathbf{a}\|$.
- $\lambda = 0$ může nastat jen když $\mathbf{a} = \mathbf{0}$, tedy účelová funkce je konstantní. Pak jsou kandidáti na lokální extrémy všechna $\mathbf{x}$ splňující $\mathbf{x}^T\mathbf{x} = 1$, tedy celá sféra X. ♦

Příklad 11.17. Vraťme se k úloze (7.1), kde maximalizujeme kvadratickou formu $\mathbf{x}^T\mathbf{A}\mathbf{x}$ na sféře $\mathbf{x}^T\mathbf{x} = 1$. Předpokládáme, že $\mathbf{A}$ je symetrická. Máme

$$L(\mathbf{x}, \lambda) = \mathbf{x}^T\mathbf{A}\mathbf{x} + \lambda(1 - \mathbf{x}^T\mathbf{x}).$$

Rovnice $L_{\mathbf{x}}(\mathbf{x}, \lambda) = \mathbf{0}$ dá (po transpozici) $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$. Tedy $(\mathbf{x}, \lambda)$ je stacionární bod funkce L, právě když λ je vlastní číslo matice $\mathbf{A}$ a $\mathbf{x}$ je normalizovaný (kvůli podmínce $\mathbf{x}^T\mathbf{x} = 1$) vlastní vektor příslušný λ.

Pouze z podmínek prvního řádu nelze rozhodnout, které stacionární body funkce L odpovídají minimu úlohy (7.1), které maximu a které ani jednomu. Můžeme to ale udělat úvahou: když $(\lambda, \mathbf{x})$ je stacionární bod funkce L, hodnota účelové funkce v tomto bodě je $\mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{x}^T(\lambda\mathbf{x}) = \lambda\mathbf{x}^T\mathbf{x} = \lambda$. Tedy nejmenší [největší] vlastní číslo odpovídá minimu [maximu] a ostatní vlastní čísla neodpovídají (globálním) extrémům úlohy (ovšem touto úvahou jsme nerozhodli, zda tato čísla odpovídají lokálním exrémům nebo jen sedlovým bodům). ♦

Předchozí příklady a také Příklady 11.4 a 11.3 vyžadují od studenta nejen znalost metody Lagrangeových multiplikátorů, ale i jistou zručnost v manipulaci s maticovými výrazy. Cvičte tuto zručnost ve Cvičeních 11.16–11.18!

11.2.3 Podmínky druhého řádu

Věta 11.5 udává podmínky prvního řádu na extrémy vázané rovnostmi. Říká, že pokud $(\mathbf{x}, \boldsymbol{\lambda})$ je stacionární bod Lagrangeovy funkce, pak bod $\mathbf{x}$ je 'podezřelý' z lokálního extrému funkce f na množině X. Uveďme nyní bez důkazu podmínky druhého řádu[6]. Hlavní rozdíl oproti Větě 11.3 je v tom, že místo druhých derivací funkce f v bodě $\mathbf{x}$ v nich vystupují druhé derivace funkce L podle $\mathbf{x}$ v bodě $(\mathbf{x}, \boldsymbol{\lambda})$, tj.

$$\frac{\partial^2 L(\mathbf{x}, \boldsymbol{\lambda})}{\partial \mathbf{x}^2} = f''(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i \, g_i''(\mathbf{x}). \tag{11.24}$$

Věta 11.6. Nechť $f\colon \mathbb{R}^n \to \mathbb{R}$ a $\mathbf{g}\colon \mathbb{R}^n \to \mathbb{R}^m$ jsou dvakrát diferencovatelné v bodě $\mathbf{x} \in \mathbb{R}^n$. Nechť $\mathbf{x}$ je regulární bod zobrazení $\mathbf{g}$.

- Jestliže $\mathbf{x}$ je lokální minimum [maximum] funkce f vázané podmínkou $\mathbf{g}(\mathbf{x}) = \mathbf{0}$, pak existuje $\boldsymbol{\lambda}$ tak že $L'(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{0}$
  a Hessova matice (11.24) je positivně [negativně] semidefinitní na podprostoru $\operatorname{null} \mathbf{g}'(\mathbf{x})$.
- Jestliže existuje $\boldsymbol{\lambda}$ tak že $L'(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{0}$
  a Hessova matice (11.24) je positivně [negativně] definitní na podprostoru $\operatorname{null} \mathbf{g}'(\mathbf{x})$, pak $\mathbf{x}$ je ostré lokální minimum [maximum] funkce f vázané podmínkou $\mathbf{g}(\mathbf{x}) = \mathbf{0}$.

Příklad 11.18. Hledejme strany kvádru s jednotkovým objemem a minimálním povrchem. Tedy minimalizujeme $xy + xz + yz$ za podmínky $xyz = 1$. Lagrangeova funkce je

$$L(x, y, z, \lambda) = xy + xz + yz + \lambda(1 - xyz).$$

Položením derivací L rovným nule máme soustavu

$$\begin{aligned}
L_x(x, y, z, \lambda) &= y + z - \lambda yz = 0\\
L_y(x, y, z, \lambda) &= x + z - \lambda xz = 0\\
L_z(x, y, z, \lambda) &= x + y - \lambda xy = 0\\
L_\lambda(x, y, z, \lambda) &= xyz - 1 = 0.
\end{aligned}$$

Soustava je zjevně splněna pro $(x, y, z, \lambda) = (1, 1, 1, 2)$. Máme ukázat, že tento bod odpovídá lokálnímu minimu. Máme

$$\frac{\partial^2 L(x, y, z, \lambda)}{\partial (x, y, z)^2} = \begin{bmatrix} 0 & 1-\lambda z & 1-\lambda y \\ 1-\lambda z & 0 & 1-\lambda x \\ 1-\lambda y & 1-\lambda x & 0 \end{bmatrix} = \begin{bmatrix} 0 & -1 & -1 \\ -1 & 0 & -1 \\ -1 & -1 & 0 \end{bmatrix}. \tag{11.25}$$

Máme ukázat, že tato matice je positivně definitní na nulovém prostoru Jacobiho matice

$$g'(x, y, z) = \begin{bmatrix} -yz & -xz & -xy \end{bmatrix} = \begin{bmatrix} -1 & -1 & -1 \end{bmatrix}.$$ ♦

Dále pokračujeme podobně jako v Příkladu 11.6.

[6] Zdůrazněme, že druh lokálního extrému *nelze* zjistit podle definitnosti Hessovy matice $L''(\mathbf{x}, \boldsymbol{\lambda})$, tedy je chybou použít Větu 10.4 na funkci L (z Věty 11.6 nic takového neplyne).

11.3 Cvičení

Následující úlohy vyřešte nejprve libovolným (co možná jednoduchým) způsobem a potom metodou Lagrangeových multiplilátorů. Při tom nemusíte ověřovat podmínky druhého řádu – lze-li ale usoudit na druh extrému nějakou snadnou úvahou, udělejte to.

11.1. Na kružnici $x^2 + y^2 = 1$ najděte lokální extrémy funkce

a) $f(x, y) = 2x - y$
b) $f(x, y) = x(y - 1)$
c) $f(x, y) = x^2 + 2y^2$
d) $f(x, y) = x^2 y$
e) $f(x, y) = x^4 + y^2$
f) $f(x, y) = \mathrm{e}^{xy}$
g) $f(x, y) = \sin(xy)$

11.2. Na sféře $x^2 + y^2 + z^2 = 1$ najděte lokální extrémy funkce

a) $f(x, y, z) = (x + y)(y + z)$
b) $f(x, y, z) = a/x + b/y + c/z$, kde $a, b, c > 0$ jsou dány
c) $f(x, y, z) = x^3 + y^2 + z$
d) ($\star$) $f(x, y, z) = x^3 + y^3 + z^3 - 3xyz$

11.3. Najděte lokální extrémy funkce

a) $f(x, y, z) = x + yz$ za podmínek $x^2 + y^2 + z^2 = 1$ a $z^2 = x^2 + y^2$
b) $f(x, y, z) = xyz$ za podmínek $x^2 + y^2 + z^2 = 1$ a $xy + yz + zx = 1$

11.4. Najděte bod nejblíže počátku na křivce

a) $x + y = 1$
b) $x + 2y = 5$
c) $x^2 y = 1$
d) $x^2 + 2y^2 = 1$

11.5. Spočítejte rozměry tělesa tak, aby mělo při daném objemu nejmenší povrch:

a) kvádr
b) kvádr bez víka (má jednu dolní stěnu a čtyři boční, horní stěna chybí)
c) válec
d) půllitr (válec bez víka)
e) ($\star$) kelímek (komolý kužel bez víka). Objem komolého kužele je $V = \frac{\pi}{3} h(R^2 + Rr + r^2)$ a povrch pláště (bez podstav) je $S = \pi(R + r)\sqrt{(R - r)^2 + h^2}$. Použijte vhodnou numerickou metodu na řešení vzniklé soustavy rovnic.

11.6. Rozložte dané kladné reálné číslo na součin n kladných reálných čísel tak, aby jejich součet byl co nejmenší.

11.7. Najděte vzdálenost množiny $\{\, (x, y) \in \mathbb{R}^2 \mid y = x^2 \,\}$ od kružnice s poloměrem 1 a středem v bodě $(2, 0)$.

11.8. Dokažte, že funkce $f(x, y) = x$ nabývá za podmínky $x^3 = y^2$ minimum pouze v počátku. Ukažte, že metoda Lagrangeových multiplikátorů toto minimum nenajde a tento výsledek vysvětlete.

11.9. Nechť funkce $g\colon \mathbb{R}^n \to \mathbb{R}$ má nenulový gradient na celém $\mathbb{R}^n$. Označme vrstevnici výšky nula funkce g ('hladkou nadplochu') jako $X = \{\, \mathbf{x} \in \mathbb{R}^n \mid g(\mathbf{x}) = 0 \,\}$.

a) Nechť $\mathbf{b} \in \mathbb{R}^n$ je bod neležící na nadploše X. Nechť $\mathbf{x}^* \in X$ je bod nadplochy nejblížší bodu $\mathbf{b}$. Dokažte, že gradient $\nabla g(\mathbf{x}^*)$ je rovnoběžný s vektorem $\mathbf{x}^* - \mathbf{b}$.

b) Nechť $H = \{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}^T\mathbf{x} = b \,\}$ je nadrovina neprotínající nadplochu X. Nechť $\mathbf{x}^* \in X$ je bod nadplochy X nejblížší nadrovině H. Dokažte, že gradient $\nabla g(\mathbf{x}^*)$ je kolmý k nadrovině H.

11.10. Do elipsy o daných délkách os vepište obdélník s maximálním obsahem. Předpokládejte přitom, že strany obdélníku jsou rovnoběžné s osami elipsy.

11.11. *Fermatův princip nejkratšího času* v paprskové optice říká, že cesta mezi libovolnými dvěma body na paprsku má takový tvar, aby ji světlo proběhlo za čas kratší než jí blízké dráhy. Z tohoto principu odvoďte:

a) Zákon odrazu od zrcadla: úhel dopadu se rovná úhlu odrazu.

b) Snellův zákon lomu: na rozhraní dvou prostředí se světlo lomí tak, že

$$\frac{c_1}{c_2} = \frac{\sin \alpha_1}{\sin \alpha_2},$$

kde α_i je úhel paprsku od normály rozhraní a c_i je rychlost světla v prostředí i.

Nápověda: Uvažujte zrcadlo/rozhraní jako křivý povrch $X = \{\, \mathbf{x} \in \mathbb{R}^n \mid g(\mathbf{x}) = 0 \,\}$, kde funkce $g\colon \mathbb{R}^n \to \mathbb{R}$ má v každém bodě nenulový gradient. Uvažujte (libovolné) dva body $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ (v případě lomu každý na jiné straně rozhraní) a napište podmínku na bod $\mathbf{x} \in X$ tak, aby čas letu světla po dráze **a-x-b** byl lokálně extrémní.

Později se zjistilo, že správným kritériem není *nejkratší* ale *extrémní* čas, tedy skutečná dráha paprsku má čas větší nebo menší než jí blízké dráhy. Dokážete najít situaci, kdy skutečná dráha paprsku má čas *větší* než jí blízké dráhy?

11.12. Rozdělení pravděpodobnosti diskrétní náhodné proměnné je funkce $p\colon \{1, \ldots, n\} \to \mathbb{R}_+$ (tj. soubor nezáporných čísel $p(1), \ldots, p(n)$) splňující $\sum_{x=1}^n p(x) = 1$.

a) *Entropie* náhodné proměnné s rozdělením p je rovna $-\sum_{x=1}^n p(x) \ln p(x)$. Najděte rozdělení s maximální entropií.

b) Dokažte *Gibbsovu nerovnost* (též zvanou *informační nerovnost*): pro každé dvě rozdělení p, q platí

$$\sum_{x=1}^n p(x) \ln q(x) \le \sum_{x=1}^n p(x) \ln p(x),$$

přičemž rovnost nastává jen tehdy, když $p = q$.

11.13. ($\star$) Máme trojúhelník se stranami délek a, b, c. Uvažujme bod, který má takovou polohu, že součet čtverců jeho vzdáleností od stran trojúhelníku je nejmenší možný. Jaké budou vzdálenosti x, y, z tohoto bodu od stran trojúhelníku?

11.14. ($\star$) Máme krychli s délkou hrany 2. Do stěny krychle je vepsána kružnice (která má tedy poloměr 1) a okolo sousední stěny je opsána kružnice (která má tedy poloměr $\sqrt{2}$). Najděte nejmenší a největší vzdálenost mezi body na kružnicích.

11.15. ($\star$) Najděte extrémy funkce

$$f(x, y, z, u, v, w) = (1 + x + u)^{-1} + (1 + y + v)^{-1} + (1 + z + w)^{-1}$$

za podmínek $xyz = a^3$, $uvw = b^3$ a $x, y, z, u, v, w > 0$.

11.16. Minimalizujte $\mathbf{x}^T\mathbf{x}$ za podmínky $\mathbf{a}^T\mathbf{x} = 1$. Jaký je geometrický význam úlohy?

11.17. Minimalizujte $\mathbf{x}^T\mathbf{A}\mathbf{x}$ za podmínky $\mathbf{b}^T\mathbf{x} = 1$, kde $\mathbf{A}$ je symetrická positivně definitní.

11.18. Minimalizujte $\mathbf{x}^T\mathbf{C}\mathbf{x}$ za podmínky $\mathbf{A}\mathbf{x} = \mathbf{b}$, kde $\mathbf{A}$ má lineárně nezávislé řádky a $\mathbf{C}$ je symetrická positivně definitní. Najděte vzorec pro optimální $\mathbf{x}$.

11.19. Řešte Cvičení 11.18, kde ale matice $\mathbf{C}$ a $\mathbf{A}$ mohou být libovolné. Nemusíte najít vzorec pro optimální $\mathbf{x}$, ale napište soustavu lineárních rovnic, jejímž řešením je stacionární bod Lagrangeovy funkce (jako v Příkladu 11.4).

11.20. ($\star$) Minimalizujte $\mathbf{a}^T\mathbf{x}$ za podmínky $\mathbf{x}^T\mathbf{C}\mathbf{x} = 1$, kde $\mathbf{C}$ je positivně definitní.

11.21. ($\star$) Minimalizujte $\mathbf{x}^T\mathbf{A}\mathbf{x}$ za podmínky $\mathbf{x}^T\mathbf{B}\mathbf{x} = 1$, kde $\mathbf{A}$ a $\mathbf{B}$ jsou positivně definitní.

11.22. ($\star$) Minimalizujte $\mathbf{x}^T\mathbf{A}\mathbf{x}$ za podmínek $\mathbf{B}\mathbf{x} = \mathbf{0}$ a $\mathbf{x}^T\mathbf{x} = 1$, kde $\mathbf{A}$ je positivně definitní.

11.23. ($\star$) Pro jaké matice $\mathbf{A}$ a vektory $\mathbf{b}$ platí $\max\{\, \|\mathbf{A}\mathbf{x}\| \mid \mathbf{b}^T\mathbf{x} = 0 \,\} = 0$?

11.24. ($\star$) Dokažte, že matice $\begin{bmatrix} \mathbf{A}^T\mathbf{A} & \mathbf{C}^T \\ \mathbf{C} & \mathbf{0} \end{bmatrix}$ je regulární právě tehdy, když matice $\begin{bmatrix} \mathbf{A} \\ \mathbf{C} \end{bmatrix}$ má lineárně nezávislé sloupce a matice $\mathbf{C}$ má lineárně nezávislé řádky (viz Příklad 11.4).

11.25. ($\star$) Nechť funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ je diferencovatelná v bodě $\mathbf{x}^* \in \mathbb{R}^n$. Ukažte, že graf Taylorova polynomu $T^1_{\mathbf{x}^*}(\mathbf{x}) = f(\mathbf{x}^*) + f'(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*)$ prvního stupně (tj. afinní aproximace) funkce f v bodě $\mathbf{x}^*$ je tečný podprostor ke grafu funkce f (zopakujte definici grafu funkce ze zač. §8!) v bodě $\mathbf{x}^*$, přičemž podmínky regularity jsou splněny.

11.26. Pro dané *normalizované* vektory $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ maximalizujte $\mathbf{a}^T\mathbf{x}$ za podmínek $\mathbf{b}^T\mathbf{x} = 0$ a $\mathbf{x}^T\mathbf{x} = 1$.

a) Úlohu vyřešte metodou Lagrangeových multiplikátorů. Výsledkem bude vzorec pro optimální řešení $\mathbf{x}$ a vzorec pro odpovídající optimální hodnotu $\mathbf{a}^T\mathbf{x}$.

b) Úlohu lze řešit jednoduchou geometrickou úvahou. Popište tuto úvahu, tedy popište, jakou geometrickou konstrukcí by se nalezlo optimální řešení $\mathbf{x}$.

11.27. Zjistěte definitnost matice $\begin{bmatrix} 6 & -2 & 2 \\ -2 & 0 & 1 \\ 2 & 1 & 4 \end{bmatrix}$ na podprostoru $\{\, (x_1, x_2, x_3) \in \mathbb{R}^3 \mid x_2 = 0 \,\}$.

Nápověda a řešení

11.1.a) Kandidáti na vázané lokální extrémy (tj. stacionární body Lagrangeovy funkce $L(x, y, \lambda) = f(x, y) + \lambda(x^2 + y^2 - 1)$ po 'zahození' λ) jsou $(x, y) = \pm(-2/\sqrt{5}, 1/\sqrt{5})$. Z úvahy/náčrtku je jasné, zda se jedná o extrémy a jakého typu.

11.1.b) Kandidáti na vázané lokální extrémy jsou $(\pm\sqrt{3}/2, -1/2)$.

11.1.d) Zde ukážeme řešení bez metody Lagrangeových multiplikátorů. Do účelové funkce dosadíme $x^2 = 1 - y^2$ a hledáme lokální extrémy funkce $(1 - y^2)y$. Ale pozor: hledáme je ne na množině $\mathbb{R}$, ale na intervalu $[-1, 1]$. Na něm má funkce dva lokální extrémy $y = \pm\frac{1}{\sqrt{3}}$ uvnitř intervalu a dva extrémy ± 1 v krajních bodech intervalu. To dává celkem šest lokálních extrémů (x, y) původní úlohy.

11.1.f) Pokud je možné úlohu zjednodušit, je dobré to udělat. Zde např. je možno účelovou funkci e^{xy} nahradit funkcí xy, protože funkce e^t je striktně rostoucí.

11.1.g) Opět je možno účelovou funkci $\sin(xy)$ nahradit funkcí xy, protože funkce sin je striktně rostoucí na intervalu $[-\pi/2, \pi/2]$, což nám stačí, protože díky podmínce $x^2 + y^2 = 1$ bude vždy $xy \in [-1, 1]$.

11.2.a) Kandidátů na vázané lok. extrémy je šest: $(x, y, z) = \pm(1, 2, 1)/\sqrt{6}, \pm(-1, 1, -1)/\sqrt{3}, \pm(-1, 0, 1)/\sqrt{2}$.

11.3.a) Kandidáti na vázané lokální extrémy jsou $(x, y, z) = (1/\sqrt{3}, -1/\sqrt{6}, -1/\sqrt{2})$, $(1/\sqrt{3}, 1/\sqrt{6}, 1/\sqrt{2})$, $(-1/\sqrt{3}, -1/\sqrt{6}, 1/\sqrt{2})$, $(-1/\sqrt{3}, 1/\sqrt{6}, -1/\sqrt{2})$.

11.4.c) Jsou dva body nejblíže počátku: $x = \pm 2^{1/6}$, $y = 2^{-1/3}$.

11.6. Formulace: Minimalizujeme $x_1 + \cdots + x_n$ za podmínek $x_1 \cdots x_n = a$ a $x_1, \ldots, x_n > 0$, kde $a > 0$ je dané číslo.

11.7. Ukážeme jen řešení pomocí převodu na volné extrémy: po eliminaci proměnné y minimalizujeme $\|(x, x^2) - (2, 0)\|_2^2 = (x-2)^2 + x^4$. Stacionární podmínka je $2x^3 + x - 2 = 0$. To je kubická rovnice. Pokud neznáme vzorce na řešení kubických rovnic (tzv. Cardanovy vzorce), pomůže Newtonova metoda: její iterace je (odvoďte!) $x_{k+1} \leftarrow x_k - \dfrac{2x_k^3 + x_k - 2}{6x_k^2 + 1} = \dfrac{4x_k^3 + 2}{6x_k^2 + 1}$, počáteční odhad $x_0 = 1$. Po pár iteracích je $x \approx 0.835122348481367$, tedy $d \approx 1.357699386102247$.

11.9.b) Minimalizujeme $\|\mathbf{x} - \mathbf{y}\|^2$ za podmínek $g(\mathbf{x}) = 0$ a $\mathbf{a}^T\mathbf{y} = b$. Dokazované tvrzení vyplývá z podmínek prvního řádu (pomocí Lagrangeových multiplikátorů).

11.11a)) Protože rychlost světla je v tomto případě všude stejná, čas letu po dráze **a-x-b** bude přímo úměrný její délce $\|\mathbf{x}-\mathbf{a}\|+\|\mathbf{x}-\mathbf{b}\|$. Stacoinární body Lagrangeovy funkce $L(\mathbf{x}, \lambda) = \|\mathbf{x}-\mathbf{a}\|+\|\mathbf{x}-\mathbf{b}\| + \lambda g(\mathbf{x})$ splňují $(\mathbf{x}-\mathbf{a})^0 + (\mathbf{x}-\mathbf{b})^0 = \lambda\nabla g(\mathbf{x})$, kde jsme užili zkratku $\mathbf{y}^0 = \mathbf{y}/\|\mathbf{y}\|$ (odvoďte!). Tato rovnost říká, že vektor $\nabla g(\mathbf{x})$ (normála k zrcadlu) leží v jedné rovině s jednotkovými vektory $(\mathbf{x}-\mathbf{a})^0$ a $(\mathbf{x}-\mathbf{b})^0$ a půlí úhel mezi nimi (nakreslete, rozmyslete!).

11.12.b) Maximalizujte levou stranu přes q za podmínky $\sum_x q(x) = 1$. (Podmínka $q(x) \geq 0$ je implicitně zajištěna tím, že když se některá složka q blíží shora nule, účelová funkce se zhoršuje nade všechny meze.)

11.17. $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}/(\mathbf{b}^T\mathbf{A}^{-1}\mathbf{b})$

11.18. $\mathbf{x} = \mathbf{C}^{-1}\mathbf{A}^T(\mathbf{A}\mathbf{C}^{-1}\mathbf{A}^T)^{-1}\mathbf{b}$.

11.25. Toto cvičení se týká §11.2.1. Graf funkce f je množina $X = \{\, (\mathbf{x}, y) \in \mathbb{R}^{n+1} \mid g(\mathbf{x}, y) = 0 \,\}$, kde $g\colon \mathbb{R}^{n+1} \to \mathbb{R}$ je $g(\mathbf{x}, y) = f(\mathbf{x}) - y$. Jacobiho matice $g'(\mathbf{x}, y) = \begin{bmatrix} f'(\mathbf{x}) & -1 \end{bmatrix}$ je vždy nenulový řádkový vektor, tedy každý bod $(\mathbf{x}, f(\mathbf{x})) \in X$ je regulární bod funkce g.

Graf funkce $T^1_{\mathbf{x}^*}$ je afinní podprostor $\{\, (\mathbf{x}, y) \in \mathbb{R}^{n+1} \mid f(\mathbf{x}^*) + f'(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) = y \,\}$. Posuneme-li ho do počátku, stane se z něj (lineární) podprostor $\{\, (\mathbf{x}, y) \in \mathbb{R}^{n+1} \mid f'(\mathbf{x}^*)\mathbf{x} = y \,\}$. Máte ukázat, že tento podprostor je roven tečnému podprostoru (bez posunutí) k množině X v bodě $(\mathbf{x}^*, f(\mathbf{x}^*)) \in X$, tedy null $\begin{bmatrix} f'(\mathbf{x}^*) & -1 \end{bmatrix}$.

11.26.a) Lagrangeova funkce je $L(\mathbf{x}, \lambda, \mu) = \mathbf{a}^T\mathbf{x} - \lambda\mathbf{b}^T\mathbf{x} + \mu(1 - \mathbf{x}^T\mathbf{x})$. Podmínky prvního řádu:

$$
\begin{aligned}
(L(\mathbf{x}, \lambda, \mu)_{\mathbf{x}})^T &= \mathbf{a} - \lambda\mathbf{b} - 2\mu\mathbf{x} = \mathbf{0}, \\
\mathbf{b}^T\mathbf{x} &= 0, \\
\mathbf{x}^T\mathbf{x} &= 1.
\end{aligned}
$$

Tuto soustavu $n+2$ rovnic o $n+2$ neznámých vyřešíme následovně. Z první rovnice vyjádříme $\mathbf{x} = (\mathbf{a} - \lambda\mathbf{b})/(2\mu)$ (všimněte si, že z žádné jiné rovnice vyjádřit nic nejde). To nejprve dosadíme do druhé rovnice, čímž po zjednodušení (s použitím předpokladu $\mathbf{b}^T\mathbf{b} = 1$ normalizovanosti vektoru $\mathbf{b}$) dostaneme $\lambda = \mathbf{a}^T\mathbf{b}$. Pak to dosadíme do třetí rovnice, čímž po dosazení za λ a zjednodušení (s využitím předpokladů $\mathbf{a}^T\mathbf{a} = \mathbf{b}^T\mathbf{b} = 1$) získáme $(2\mu)^2 = 1 - (\mathbf{a}^T\mathbf{b})^2$, tedy $2\mu = \pm\sqrt{1 - (\mathbf{a}^T\mathbf{b})^2}$. Dosazením do výrazu pro $\mathbf{x}$ dostaneme kandidáty na extrémy a odpovídající účelové hodnoty

$$
\mathbf{x} = \frac{\mathbf{a} - \lambda\mathbf{b}}{2\mu} = \pm\frac{\mathbf{a} - (\mathbf{a}^T\mathbf{b})\mathbf{b}}{\sqrt{1 - (\mathbf{a}^T\mathbf{b})^2}}, \qquad \mathbf{a}^T\mathbf{x} = \pm\frac{1 - (\mathbf{a}^T\mathbf{b})(\mathbf{a}^T\mathbf{b})}{\sqrt{1 - (\mathbf{a}^T\mathbf{b})^2}} = \pm\sqrt{1 - (\mathbf{a}^T\mathbf{b})^2}.
$$

Protože maximalizujeme, zvolíme znaménko pro které je výraz $\mathbf{a}^T\mathbf{x}$ větší, tj. plus. Vzorce by příp. šly zjednodušit substitucí $\mathbf{a}^T\mathbf{b} = \cos\theta$ a $\sqrt{1-(\mathbf{a}^T\mathbf{b})^2} = \sin\theta$ kde θ je úhel mezi vektory $\mathbf{a}, \mathbf{b}$.

11.26.b) Množina přípustných řešení úlohy je průnik n-rozměrné jednotkové sféry $\mathbf{x}^T\mathbf{x} = 1$ s nadrovinou $\mathbf{b}^T\mathbf{x} = 0$ procházející počátkem a kolmé na vektor $\mathbf{b}$ (pro případ $n = 3$ by to byla kružnice se středem v počátk ua jednotkovým poloměrem, jejíž rovina je kolmá na vektor $\mathbf{b}$). Kdyby vektor $\mathbf{a}$ ležel v nadrovině $\mathbf{b}^T\mathbf{x} = 0$, optimálním řešením $\mathbf{x}$ by byl vektor na sféře rovnoběžný s vektorem $\mathbf{a}$ a stejně orientovaný, tj. vektor $\mathbf{x} = \mathbf{a}/\|\mathbf{a}\|$. Protože vektor $\mathbf{a}$ v nadrovině obecně neleží, musíme ho nahradit jeho průmětem do této nadroviny, tj. vektorem $\tilde{\mathbf{a}} = \mathbf{a} - (\mathbf{a}^T\mathbf{b})\mathbf{b}$. Optimálním řešením bude pak vektor $\mathbf{x} = \tilde{\mathbf{a}}/\|\tilde{\mathbf{a}}\|$, což po zjednodušení výrazu $\|\tilde{\mathbf{a}}\|$odpovídá výsledku výše.

11.27. Postupujeme podle odstavce nad Větou 11.3. Označme matici jako $\mathbf{A}$ a podprostor jako Y. Nejprve najdeme bázi podprostoru Y, např. $\mathbf{B} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}$ (báze tvoří sloupce této matice). Pak určíme definitnost matice $\mathbf{B}^T\mathbf{A}\mathbf{B} = \begin{bmatrix} 6 & 2 \\ 2 & 4 \end{bmatrix}$. Tato matice je positivně definitní (na rozdíl od matice $\mathbf{A}$, která je indefinitní).

To v našem jednoduchém případě můžeme vidět i příměji. Definitnost matice $\mathbf{A}$ na podprostoru Y se ptá na znaménka výrazu $\mathbf{x}^T\mathbf{A}\mathbf{x}$ pro různá $\mathbf{x} \in Y$. Náš podprostor můžeme zjevně psát jako $Y = \{\, (x_1, 0, x_3) \mid x_1, x_3 \in \mathbb{R} \,\}$. Tedy v součinech $\mathbf{x}^T\mathbf{A}\mathbf{x}$ je druhý sloupec+řádek matice $\mathbf{A}$ vždy násoben nulou, proto je můžeme z matice vynechat, čímž dostaneme matici $\mathbf{B}^T\mathbf{A}\mathbf{B}$.

Část III

Lineární programování

Kapitola 12

Lineární programování

Úloha **lineárního programování**[1] (LP, také zvané lineární optimalizace) je minimalizace nebo maximalizace lineární funkce za omezujících podmínek ve tvaru lineárních rovnic a nerovnic. Zde **lineární rovnicí** rozumíme relaci

$$a_1x_1 + \cdots + a_nx_n = b,$$

neboli krátce $\mathbf{a}^T\mathbf{x} = b$. **Lineární nerovnicí** rozumíme jednu z relací

$$a_1x_1 + \cdots + a_nx_n \leq b, \qquad a_1x_1 + \cdots + a_nx_n \geq b,$$

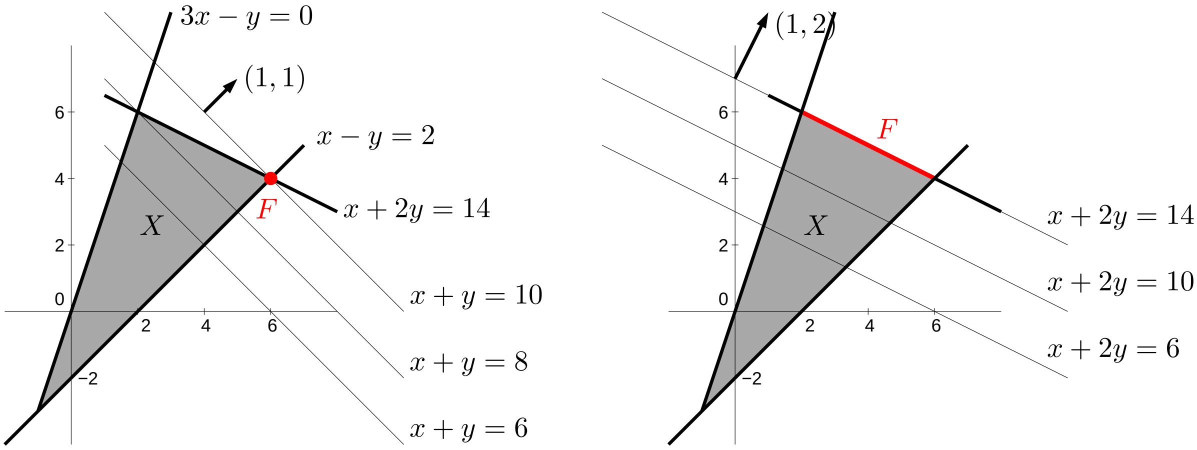
neboli $\mathbf{a}^T\mathbf{x} \leq b$ či $\mathbf{a}^T\mathbf{x} \geq b$. Lineární program je tedy úloha (1.11) ve které je funkce f lineární (tj. tvaru (3.8)) a funkce g_i, h_i jsou afinní (tj. tvaru (3.27)).

Lineární programy s jednou nebo dvěma proměnnými lze řešit graficky.

Příklad 12.1. Příkladem lineárního programu je optimalizační úloha

$$\begin{aligned} \max \quad & x + y \\ \text{za podmínek} \quad & x + 2y \leq 14 \\ & 3x - y \geq 0 \\ & x - y \leq 2 \end{aligned} \tag{12.1}$$

Množina $X = \{\, (x,y) \in \mathbb{R}^2 \mid x + 2y \leq 14,\ 3x - y \geq 0,\ x - y \leq 2 \,\}$ přípustných řešení této úlohy je průnik tří polorovin:



[1] Slovo *programování* zde má trochu jiný význam než znáte: místo tvoření sekvenčního počítačového kódu řešícího daný problém to znamená hledání vhodné účelové funkce a omezujících podmínek tak, aby optimální řešení bylo řešením daného problému. Optimalizaci obecně se také někdy říká *matematické programování*.

Účelová funkce $x + y$, neboli $\mathbf{c}^T\mathbf{x}$ pro $\mathbf{x} = (x, y)$ a $\mathbf{c} = (1, 1)$, má vrstevnice kolmé k vektoru $\mathbf{c}$ a roste ve směru $\mathbf{c}$. Proto (viz obrázek vlevo) účelová funkce na množině X nabývá maxima v bodě $(x, y) = (6, 4)$. Úloha má tedy jediné optimální řešení.

Pokud bychom účelovou funkci úlohy změnili na $x + 2y$, množina optimálních řešení úlohy by byla úsečka spojující body $(2, 6)$ a $(6, 4)$ (viz obrázek vpravo), úloha by tedy měla nekonečně mnoho optimálních řešení. Pokud by účelová funkce byla nulová (tj. $0x + 0y$), množina optimálních řešení úlohy by byla celý trojúhelník X. ♦

Z našich úvah je patrno (přesně ukážeme později), že pro úlohu lineárního programování mohou nastat tři případy:

- úloha má (alespoň jedno) optimální řešení,
- úloha je *nepřípustná* (množina přípustných řešení je prázdná, omezení si odporují),
- úloha je *neomezená* (účelovou funkci lze bez porušení omezení libovolně zlepšovat).

Jestliže úloha má optimální řešení, pak množina optimálních řešení je buď vrchol mnohoúhelníku, nebo jeho hrana, nebo celý mnohoúhelník.

12.1 Transformace úloh LP

Často je užitečné formulovat lineární program v nějakém speciálním tvaru, kdy jsou dovoleny jen určité typy omezení. To obvykle vyžadují např. algoritmy na řešení LP.

Jeden speciální tvar je tvar, kdy minimalizujeme (ne tedy maximalizujeme) a dovolíme pouze omezení typu $\geq$ (větší nebo rovno):

$$\begin{array}{rll} \min & c_1x_1 + \cdots + c_nx_n & \\ \text{za podmínek} & a_{i1}x_1 + \cdots + a_{in}x_n \geq b_i, & i = 1, \ldots, m \end{array} \tag{12.2a}$$

To se pohodlněji napíše v maticovém tvaru

$$\min\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{A}\mathbf{x} \geq \mathbf{b} \,\}, \tag{12.2b}$$

kde $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{c} \in \mathbb{R}^n$. Zápis $\mathbf{A}\mathbf{x} \geq \mathbf{b}$ značí, že pro každé $i = 1, \ldots, m$ není i-tá složka vektoru $\mathbf{A}\mathbf{x}$ menší než i-tá složka vektoru $\mathbf{b}$. Jakýkoliv lineární program snadno převedeme na tvar (12.2) těmito úpravami:

- Maximalizaci funkce $\mathbf{c}^T\mathbf{x}$ nahradíme minimalizací funkce $-\mathbf{c}^T\mathbf{x}$ (viz (1.9)).
- Nerovnost $\mathbf{a}^T\mathbf{x} \leq b$ nahradíme nerovností $-\mathbf{a}^T\mathbf{x} \geq -b$.
- Rovnost $\mathbf{a}^T\mathbf{x} = b$ nahradíme dvěma nerovnostmi $\mathbf{a}^T\mathbf{x} \geq b$, $-\mathbf{a}^T\mathbf{x} \geq -b$.

Jiný často užívaný speciální tvar je **rovnicový tvar**[2], ve kterém všechna omezení jsou rovnosti a všechny proměnné jsou nezáporné, tedy

$$\begin{array}{rll} \min & c_1x_1 + \cdots + c_nx_n & \\ \text{za podmínek} & a_{i1}x_1 + \cdots + a_{in}x_n = b_i, & i = 1, \ldots, m \\ & x_j \geq 0\,, & j = 1, \ldots, n \end{array} \tag{12.3a}$$

neboli

$$\min\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{A}\mathbf{x} = \mathbf{b},\ \mathbf{x} \geq \mathbf{0} \,\}. \tag{12.3b}$$

Tvar (12.2) lze převést na tvar (12.3) dvěma úpravami, ve kterých zavedeme nové proměnné:

[2] Tomuto tvaru se někdy říká *standardní*. Názvosloví bohužel není jednotné, názvy jako ‘standardní tvar’, ‘základní tvar’ či ‘kanonický tvar’ tedy mohou znamenat v různých knihách různé věci.

- Nerovnost $\mathbf{a}^T\mathbf{x} \geq b$ nahradíme dvěma omezeními $\mathbf{a}^T\mathbf{x} - u = b$, $u \geq 0$. Pomocné proměnné u se říká **slacková proměnná**[3]. Podobně bychom převedli nerovnost $\mathbf{a}^T\mathbf{x} \leq b$ na rovnost.
- Neomezenou proměnnou $x \in \mathbb{R}$ rozdělíme na dvě nezáporné proměnné $x^+ \geq 0$, $x^- \geq 0$ přidáním podmínky $x = x^+ - x^-$.

Úloha získaná z původní úlohy pomocí těchto úprav je ekvivalentní původní úloze v tom smyslu, že hodnota jejich optima je stejná a argument optima původní úlohy lze 'snadno a rychle'[4] získat z argumentu optima nové úlohy.

Příklad 12.2. Úlohu (12.1) převeďme na tvar, ve kterém minimalizujeme (místo maximalizujeme) a omezení jsou rovnosti, přičemž některé proměnné mohou být nezáporné. To uděláme zavedením slackových proměnných. Nová úloha je

$$
\begin{array}{rrcrcrcl}
\min & -x & - & y & & & & \\
\text{za podmínek} & x & + & 2y & + & u & = & 14 \\
& 3x & - & y & - & v & = & 0 \\
& x & - & y & + & w & = & 2 \\
& & & & & u,v,w & \geq & 0
\end{array}
$$

Zde proměnné x, y mohou mít libovolné znaménko. Převeďme úlohu na tvar (12.3), kde všechny proměnné jsou nezáporné. Dosadíme $x = x_+ - x_-$ a $y = y_+ - y_-$, kde $x_+, x_-, y_+, y_- \geq 0$. Výsledná úloha je

$$
\begin{array}{rrcrcrcrcrcl}
\min & -x_+ & + & x_- & - & y_+ & + & y_- & & & & \\
\text{za podmínek} & x_+ & - & x_- & + & 2y_+ & - & 2y_- & + & u & = & 14 \\
& 3x_+ & - & 3x_- & - & y_+ & + & y_- & - & v & = & 0 \\
& x_+ & - & x_- & - & y_+ & + & y_- & + & w & = & 2 \\
& & & & & & & \multicolumn{3}{r}{x_+, x_-, y_+, y_-, u, v, w} & \geq & 0
\end{array}
$$

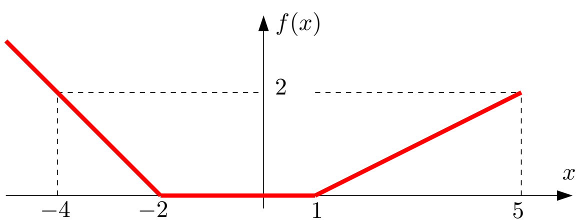
♦

12.1.1 Po částech afinní funkce

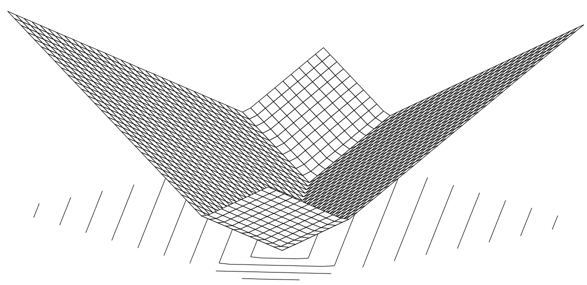
Mějme funkci $f\colon \mathbb{R}^n \to \mathbb{R}$ s hodnotami

$$
f(\mathbf{x}) = \max_{i=1}^{k} (\mathbf{c}_i^T\mathbf{x} + d_i), \tag{12.4}
$$

kde $\mathbf{c}_1, \ldots, \mathbf{c}_k \in \mathbb{R}^n$ a $d_1, \ldots, d_k \in \mathbb{R}$ jsou dány. Tato funkce není lineární ani afinní, je po částech afinní. Na obrázku jsou příklady (vlevo pro $n = 1$ a $k = 3$, vpravo pro $n = 2$ a $k = 4$):



$f(x) = \max\{-x-2,\ 0,\ \frac{1}{2}(x-1)\}$



$f(x,y) = \frac{1}{2}\max\{x+y-1,\ -x+y,\ x-y,\ -x-y-1\}$

[3] *Slack* znamená anglicky např. mezeru mezi zdí a skříní, která není zcela přiražená ke zdi. Termín *slack variable* nemá ustálený český ekvivalent, někdy se překládá jako *skluzová proměnná.*

[4] Přesněji v *lineárním čase*, ve smyslu teorie algoritmů.

Minimalizujme funkci (12.4) za podmínek $\mathbf{Ax} \geq \mathbf{b}$. To není úloha LP, neboť její účelová funkce není lineární. Ovšem lze ji převést na LP zavedením pomocné proměnné.

Obecně, pro každou množinu X a funkci $f\colon X \to \mathbb{R}$ platí (pokud minimum existuje)

$$\min\{\, f(x) \mid x \in X \,\} = \min\{\, y \mid (x,y) \in X \times \mathbb{R},\ f(x) \leq y \,\}. \tag{12.5}$$

To proto, že každé optimální řešení (x, y) pravé úlohy splňuje $f(x) = y$. Kdyby totiž bylo $f(x) < y$, pak by řešení (x, y) bylo sice přípustné ale ne optimální, protože bychom mohli y zmenšit bez porušení omezení a zlepšit tak účelovou funkci.

Teď už naši úlohu snadno převedeme na LP:

$$\begin{aligned}
\min\{\, f(\mathbf{x}) \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{Ax} \geq \mathbf{b} \,\} &= \min\{\, y \mid (\mathbf{x}, y) \in \mathbb{R}^{n+1},\ f(\mathbf{x}) \leq y,\ \mathbf{Ax} \geq \mathbf{b} \,\} && \text{(12.6a)}\\
&= \min\{\, y \mid (\mathbf{x}, y) \in \mathbb{R}^{n+1},\ \max_i(\mathbf{c}_i^T\mathbf{x} + d_i) \leq y,\ \mathbf{Ax} \geq \mathbf{b} \,\} && \text{(12.6b)}\\
&= \min\{\, y \mid (\mathbf{x}, y) \in \mathbb{R}^{n+1},\ (\forall i)(\mathbf{c}_i^T\mathbf{x} + d_i \leq y),\ \mathbf{Ax} \geq \mathbf{b} \,\} && \text{(12.6c)}
\end{aligned}$$

kde (12.6c) je již úloha LP. Rovnost (12.6c) platí proto, že pro libovolná čísla a_i, b zřejmě je

$$\max_i a_i \leq b \iff (\forall i)(a_i \leq b). \tag{12.7}$$

Příklad 12.3. Úloha

$$\begin{aligned}
\min \quad & \max\{\, 3x + 4y + 1,\ 2x - 3y \,\}\\
\text{za podm.} \quad & x + 2y \leq 14\\
& 3x - y \geq 0\\
& x - y \leq 2
\end{aligned}$$

není LP, protože účelová funkce $f(x, y) = \max\{\, 3x + 4y + 1,\ 2x - 3y \,\}$ není lineární ani afinní (načrtněte si na papír několik jejích vrstevnic!). Úlohu lze ale převést na LP

$$\begin{aligned}
\min \quad & z\\
\text{za podm.} \quad & 3x + 4y - z \leq -1\\
& 2x - 3y - z \leq 0\\
& x + 2y \leq 14\\
& 3x - y \geq 0\\
& x - y \leq 2
\end{aligned}$$

♦

Tento převod lze užít i pro funkce obsahující absolutní hodnoty, neboť $|x| = \max\{-x, x\}$, a na různé jiné případy (trénujte to ve Cvičení 12.4!). Buďte ale opatrní: neplatí nic takového jako $(\min_i a_i \leq b) \Leftrightarrow (\forall i)(a_i \leq b)$, tedy máme-li špatnou kombinaci minim/maxim a nerovností, převod na LP není možný.

12.2 Jednoduché úlohy LP

Některé úlohy LP lze řešit úvahou. Je dobré se v tom pocvičit, protože vám to dá intuici potřebnou i pro obecnější optimalizační úlohy. Více viz Cvičení 12.3.

Příklad 12.4. Pro daná čísla $\mathbf{c} = (c_1, \dots, c_n) \in \mathbb{R}^n$ řešíme úlohu

$$\begin{aligned} \max \quad & c_1x_1 + \cdots + c_nx_n \\ \text{za podmínek} \quad & x_1 + \cdots + x_n = 1 \\ & x_1, \dots, x_n \geq 0 \end{aligned}$$

neboli

$$\max\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{1}^T\mathbf{x} = 1,\ \mathbf{x} \geq \mathbf{0} \,\}.$$

Přípustná řešení je možno vidět jako rozdělení pravděpodobnosti diskrétní náhodné veličiny s n stavy. Není těžké uhodnout (zamyslete se!), že optimum úlohy se nabývá právě tehdy, když $x_i = 0$ pro všechna $i \notin \operatorname{argmax}_i c_i$ (a zbývající proměnné jsou libovolné přípustné). Jestliže množina $\operatorname{argmax}_i c_i$ má pouze jeden prvek i^*, pak $x_i = 0$ pro všechna $i \neq i^*$ a $x_{i^*} = 1$. V obou případech je optimální hodnota rovna číslu $c_{\max} = \max_i c_i$, tj. největšímu z čísel $c_1, \dots, c_n$.

Jak dokážeme, že uhodnuté řešení je skutečně optimální? Tak, že ukážeme, že každé jiné přípustné řešení bychom mohli pozměnit tak, že by zůstalo přípustné a účelová funkce by se zlepšila. Kdyby např. bylo $x_1 > 0$ a $c_1 < c_{\max}$, pak bychom mohli x_1 o kousek zmenšit a tento kousek přidat k x_i pro nějaké i splňující $c_i = c_{\max}$, což by zlepšilo (zvýšilo) účelovou funkci.

Uvedeme ještě jiný důkaz, že optimální hodnota úlohy je $c_{\max}$. Hodnota kritéria v libovolném přípustném řešení je *dolní mez* na optimální hodnotu úlohy. Pro naše uhodnuté přípustné řešení máme tedy dolní mez $c_{\max}$. Ale pro každé přípustné řešení $x_1, \dots, x_n$ je

$$c_1x_1 + \cdots + c_nx_n \leq c_{\max}x_1 + \cdots + c_{\max}x_n = c_{\max}(x_1 + \cdots + x_n) = c_{\max}, \tag{12.8}$$

tedy $c_{\max}$ je zároveň *horní mez* na optimální hodnotu úlohy. Protože obě meze splývají, optimální hodnota musí být $c_{\max}$. ♦

Příklad 12.5. Pozměňme předcházející úlohu: jsou dány $\mathbf{c} \in \mathbb{R}^n$ a $k \in \{1, \dots, n\}$ a řešíme

$$\max\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{1}^T\mathbf{x} = k,\ \mathbf{0} \leq \mathbf{x} \leq \mathbf{1} \,\}.$$

Bez ztráty obecnosti předpokládejme, že čísla c_i jsou sestupně seřazena, $c_1 \geq \cdots \geq c_n$.

Není těžké uhodnout, že optimum úlohy se nabývá pro vektor $\mathbf{x}$ splňující

$$x_i = \begin{cases} 0 & \text{jestliže } c_i < c_k, \\ 1 & \text{jestliže } c_i > c_k \end{cases}$$

a zbylé proměnné (tj. x_i pro $c_i = c_k$) jsou libovolné přípustné. Tomu odpovídá optimální hodnota $c_1 + \cdots + c_k$, tedy součet k největších čísel c_i. Speciálně, kdyby $c_{k+1} < c_k$, pak by optimum nastalo právě pro $x_1 = \cdots = x_k = 1$ a $x_{k+1} = \cdots = x_n = 0$.

Optimalitu tohoto řešení dokážeme podobným argumentem jako minule: kdyby pro nějaké i bylo $x_i > 0$ a $c_i < c_k$, pak by nutně (kvůli podmínce $x_1 + \cdots + x_n = k$) bylo $x_{i^*} < 1$ pro nějaké i^* splňující $c_{i^*} \geq c_k$ a mohli bychom tedy x_i o kousek zmenšit a x_{i^*} o stejný kousek zvětšit, což by zvýšilo hodnotu kritéria. Podobně bychom mohli hodnotu kritéria zvýšit, jestliže by pro nějaké i bylo $x_i < 1$ a $c_i > c_k$.

I když důkaz z předchozího odstavce dostačuje, odvodíme i horní mez analogicky k (12.8): pro každé přípustné $\mathbf{x}$ platí (promyslete, proč každá rovnost a nerovnost platí!)

$$\sum_{i=1}^n c_ix_i \leq \sum_{i=1}^k c_ix_i + c_k \sum_{i=k+1}^n x_i = \sum_{i=1}^k c_ix_i + c_k \sum_{i=1}^k (1 - x_i) = \sum_{i=1}^k \underbrace{\big(c_ix_i + c_k(1 - x_i)\big)}_{\leq c_i} \leq \sum_{i=1}^k c_i.$$

♦

12.3 Typické aplikace LP

Ve zbytku této kapitoly uvedeme některá z mnoha použití lineárního programování.

12.3.1 Optimální výrobní program

Z m druhů surovin vyrábíme n druhů výrobků.

- a_{ij} = množství suroviny druhu i potřebné na výrobu výrobku druhu j
- b_i = množství suroviny druhu i, které máme k dispozici
- c_j = zisk z vyrobení jednoho výrobku druhu j
- x_j = počet vyrobených výrobků druhu j

Chceme zjistit, kolik jakých výrobků máme vyrobit, aby zisk byl největši. Tuto úlohu lze formalizovat lineárním programem

$$\begin{aligned} \max \quad & \sum_{j=1}^{n} c_j x_j && && \text{(12.9a)} \\ \text{za podm.} \quad & \sum_{j=1}^{n} a_{ij} x_j \le b_i, && i = 1, \ldots, m && \text{(12.9b)} \\ & x_j \ge 0, && j = 1, \ldots, n && \text{(12.9c)} \end{aligned}$$

což jde úsporněji napsat v maticové formě jako $\max\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{A}\mathbf{x} \le \mathbf{b},\ \mathbf{x} \ge \mathbf{0}\,\}$.

Příkladem úlohy na optimální výrobní program je Příklad 1.10 z Kapitoly 1 (podívejte se na něj!).

12.3.2 Směšovací (výživová) úloha

Z n druhů surovin, z nichž každá je směsí m druhů látek, máme namíchat konečný produkt o požadovaném složení tak, aby cena surovin byla minimální.

- a_{ij} = množství látky druhu i obsažené v jednotkovém množství suroviny druhu j
- b_i = nejmenší požadované množství látky druhu i v konečném produktu
- c_j = jednotková cena suroviny druhu j
- x_j = množství suroviny druhu j

Tuto úlohu lze formalizovat lineárním programem

$$\begin{aligned} \min \quad & \sum_{j=1}^{n} c_j x_j && && \text{(12.10a)} \\ \text{za podm.} \quad & \sum_{j=1}^{n} a_{ij} x_j \ge b_i, && i = 1, \ldots, m && \text{(12.10b)} \\ & x_j \ge 0, && j = 1, \ldots, n && \text{(12.10c)} \end{aligned}$$

neboli $\min\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{A}\mathbf{x} \ge \mathbf{b},\ \mathbf{x} \ge \mathbf{0}\,\}$.

Příklad 12.6. Jste kuchařka v menze a chcete uvařit pro studenty co nejlevnější oběd, ve kterém ovšem kvůli předpisům musí být dané minimální množství živin (cukrů, bílkovin a vitamínů). Oběd vaříte ze tří surovin: brambor, masa a zeleniny. Jsou dány hodnoty v tabulce:

	na jednotku brambor	na jednotku masa	na jednotku zeleniny	min. požadavek na jeden oběd
obsah cukrů	2	1	1	8
obsah bílkovin	2	6	1	16
obsah vitamínů	1	3	6	8
cena	25	50	80	

Kolik je třeba každé suroviny na jeden oběd?

Řešíme LP

$$\begin{aligned} \min \quad & 25b + 50m + 80z \\ \text{za podmínek} \quad & 2b + m + z \geq 8 \\ & 2b + 6m + z \geq 16 \\ & b + 3m + 6z \geq 8 \\ & b, m, z \geq 0 \end{aligned}$$

Optimální řešení je $b = 3.2$, $m = 1.6$, $z = 0$ s hodnotou 160. ♦

12.3.3 Dopravní úloha

Máme m skladů a n spotřebitelů.

- a_i = množství ve skladě i
- b_j = množství zboží požadované spotřebitelem j
- c_{ij} = cena dopravy jednotky zboží ze skladu i ke spotřebiteli j
- x_{ij} = množství zboží vezené ze skladu i ke spotřebiteli j

Chceme co nejlevněji rozvézt zboží ze skladů ke spotřebitelům. Řešením je lineární program

$$\min \quad \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} \tag{12.11a}$$

$$\text{za podm.} \quad \sum_{j=1}^{n} x_{ij} = a_i, \qquad i = 1, \ldots, m \tag{12.11b}$$

$$\sum_{i=1}^{m} x_{ij} = b_j, \qquad j = 1, \ldots, n \tag{12.11c}$$

$$x_{ij} \geq 0, \qquad i = 1, \ldots, m,\ j = 1, \ldots, n \tag{12.11d}$$

Ten jde také napsat jako (rozmyslete!)

$$\min\{\, \langle \mathbf{C}, \mathbf{X} \rangle \mid \mathbf{X} \in \mathbb{R}^{m\times n},\ \mathbf{X}\mathbf{1} = \mathbf{a},\ \mathbf{X}^T\mathbf{1} = \mathbf{b},\ \mathbf{X} \geq \mathbf{0} \,\}, \tag{12.12}$$

kde $\mathbf{a} = (a_1, \ldots, a_m)$, $\mathbf{b} = (b_1, \ldots, b_n)$, $\mathbf{C} = [c_{ij}]$, $\langle \mathbf{C}, \mathbf{X} \rangle$ značí skalární součin matic (2.7) a optimalizujeme přes matice $\mathbf{X} = [x_{ij}]$.

Zadání musí splňovat $\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$ (nabídka je rovna poptávce), jinak je úloha nepřípustná. To snadno vidíme: součet levých stran rovnic (12.11b) je roven součtu levých stran rovnic (12.11c). Úlohu lze modifikovat tak, že dovolíme $\sum_{i=1}^{m} a_i \geq \sum_{j=1}^{n} b_j$, tedy zboží na skladech může být více, než žádají spotřebitelé. Pak by se omezení (12.11b) muselo změnit na $\sum_{j=1}^{n} x_{ij} \leq a_i$ (promyslete!).

12.3.4 Distribuční úloha

Máme m strojů a n druhů výrobků.

- a_i = počet hodin, které jsou k dispozici na stroji i
- b_j = požadované množství výrobku druhu j
- c_{ij} = cena jedné hodiny práce stroje i na výrobku typu j
- k_{ij} = hodinový výkon stroje i při výrobě výrobku druhu j
- x_{ij} = počet hodin, po který bude stroj i vyrábět výrobek druhu j

Pro každý ze strojů máme určit, kolik výrobků se na něm bude vyrábět. Řešení:

$$\min\left\{ \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} \;\middle|\; \sum_{j=1}^{n} x_{ij} \le a_i,\ \sum_{i=1}^{m} k_{ij}x_{ij} = b_j,\ x_{ij} \ge 0 \right\}. \tag{12.13}$$

12.4 Přeurčené soustavy lineárních rovnic

12.4.1 Vektorové normy

Norma formalizuje pojem 'délky' vektoru $\mathbf{x}$. Známe již eukleidovskou normu, ale existují i jiné normy. Obecně pojem norma definujeme takto: funkce $\|\cdot\|\colon \mathbb{R}^n \to \mathbb{R}$ se nazývá vektorová **norma**[5], jestliže splňuje tyto axiomy:

1. Jestliže $\|\mathbf{x}\| = 0$ pak $\mathbf{x} = \mathbf{0}$.
2. $\|\alpha\mathbf{x}\| = |\alpha|\,\|\mathbf{x}\|$ pro každé $\alpha \in \mathbb{R}$ a $\mathbf{x} \in \mathbb{R}^n$ (norma je kladně homogenní).
3. $\|\mathbf{x}+\mathbf{y}\| \le \|\mathbf{x}\| + \|\mathbf{y}\|$ pro každé $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ (trojúhelníková nerovnost).

Z axiomů plynou tyto další vlastnosti normy:

- $\|\mathbf{0}\| = 0$, což plyne z homogenity pro $\alpha = 0$
- $\|\mathbf{x}\| \ge 0$ pro každé $\mathbf{x} \in \mathbb{R}^n$. To jde odvodit tak, že v trojúhelníkové nerovnosti položíme $\mathbf{y} = -\mathbf{x}$ a tedy
$$0 = \|0\| = \|\mathbf{x} - \mathbf{x}\| \le \|\mathbf{x}\| + \|-\mathbf{x}\| = 2\|\mathbf{x}\|,$$
kde na pravé straně jsme použili homogenitu.

Jednotková sféra normy je množina $\{\, \mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\| = 1 \,\}$, tedy vrstevnice normy výšky 1. Díky homogenitě je jednotková sféra středově symetrická a její tvar zcela určuje normu.

Uveďme příklady norem. Důležitou skupinou norem jsou **p-normy**

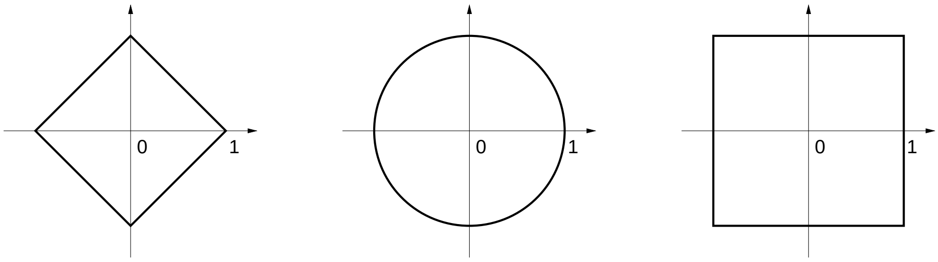
$$\|\mathbf{x}\|_p = (|x_1|^p + \cdots + |x_n|^p)^{1/p}.$$

Musí být $p \ge 1$, jinak neplatí trojúhelníková nerovnost. Nejčastěji narazíte na:

- $\|\mathbf{x}\|_1 = |x_1| + \cdots + |x_n|$. Někdy se jí říká *manhattanská norma*, protože v systému pravoúhlých ulic je vzdálenost mezi body $\mathbf{x}$ a $\mathbf{y}$ rovna $\|\mathbf{x} - \mathbf{y}\|_1$.
- $\|\mathbf{x}\|_2 = \sqrt{x_1^2 + \cdots + x_n^2} = \sqrt{\mathbf{x}^T\mathbf{x}}$. Je to *eukleidovská norma*.
- $\|\mathbf{x}\|_\infty = \lim_{p\to\infty} \|\mathbf{x}\|_p = \max\{|x_1|, \ldots, |x_n|\}$ (spočtěte limitu ve Cvičení 12.13!). Někdy se jí říká *Čebyševova norma* nebo *max-norma*.

[5] Symbolem $\|\cdot\|$ jsme dosud značili eukleidovskou normu. Od teď až do konce skript jím ale budeme značit obecnou vektorovou normu a eukleidovskou normu budeme značit symbolem $\|\cdot\|_2$.

Jednotkové sféry těchto norem v $\mathbb{R}^2$ vypadají takto:



$\|\mathbf{x}\|_1 = 1$ $\qquad$ $\|\mathbf{x}\|_2 = 1$ $\qquad$ $\|\mathbf{x}\|_\infty = 1$

Všimněte si, že pro $n = 1$ se p-norma pro libovolné $p \geq 1$ redukuje na *absolutní hodnotu* $|x|$.
Existují ovšem i normy, které nejsou p-normy, např.

- $\|\mathbf{x}\| = 2|x_1| + \sqrt{x_2^2 + x_3^2} + \max\{|x_4|, |x_5|\}$ je norma na $\mathbb{R}^5$.
- Je-li $\|\mathbf{x}\|$ norma a $\mathbf{A}$ matice s lineárně nezávislými sloupci, pak funkce $f(\mathbf{x}) = \|\mathbf{A}\mathbf{x}\|$ je také norma (tuto normu můžeme označit např. jako $f(\mathbf{x}) = \|\mathbf{x}\|_\mathbf{A} = \|\mathbf{A}\mathbf{x}\|$.

12.4.2 Přibližné řešení lineárních soustav v 1-normě a ∞-normě

Mějme přeurčenou soustavu lineárních rovnic $\mathbf{A}\mathbf{x} = \mathbf{b}$, kde $\mathbf{A} \in \mathbb{R}^{m\times n}$ a $\mathbf{0} \neq \mathbf{b} \in \mathbb{R}^m$. Nalezení jejího přibližného řešení formulujme jako úlohu

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_p. \tag{12.14}$$

Uvažujme tři případy:

- Pro $p = \infty$ hledáme takové $\mathbf{x}$, které minimalizuje výraz

$$\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_\infty = \max_{i=1}^{m} |\mathbf{a}_i^T\mathbf{x} - b_i|, \tag{12.15}$$

  tedy minimalizuje maximální residuum. Toto řešení je známé pod názvem *minimaxní* nebo *Čebyševovo*. Podle §12.1.1 je tato úloha ekvivalentní lineárnímu programu

$$\begin{aligned} \min \quad & y \\ \text{za podm.} \quad & -y \leq \mathbf{a}_i^T\mathbf{x} - b_i \leq y, \quad i = 1, \ldots, m \end{aligned}$$

  což lze napsat úsporněji v maticové formě

$$\min\{\, y \mid \mathbf{x} \in \mathbb{R}^n,\ y \in \mathbb{R},\ -y\mathbf{1} \leq \mathbf{A}\mathbf{x} - \mathbf{b} \leq y\mathbf{1} \,\}. \tag{12.16}$$

- Pro $p = 2$ dostaneme řešení ve smyslu nejmenších čtverců, které jsme odvodili v §5.1.
- Pro $p = 1$ hledáme takové $\mathbf{x}$, které minimalizuje výraz

$$\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_1 = \sum_{i=1}^{m} |\mathbf{a}_i^T\mathbf{x} - b_i|, \tag{12.17}$$

  kde $\mathbf{a}_1^T, \ldots, \mathbf{a}_m^T$ jsou řádky matice $\mathbf{A}$. Úloha je ekvivalentní lineárnímu programu

$$\begin{aligned} \min \quad & y_1 + \cdots + y_m \\ \text{za podm.} \quad & -y_i \leq \mathbf{a}_i^T\mathbf{x} - b_i \leq y_i, \quad i = 1, \ldots, m \end{aligned}$$

  neboli

$$\min\{\, \mathbf{1}^T\mathbf{y} \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{y} \in \mathbb{R}^m,\ -\mathbf{y} \leq \mathbf{A}\mathbf{x} - \mathbf{b} \leq \mathbf{y} \,\}. \tag{12.18}$$

12.4.3 Lineární regrese

Vraťme se k lineární regresi z §5.3.1 (znovu přečtěte!). Funkční závislost přibližně popsanou naměřenými dvojicemi (x_i, y_i), $i = 1, \ldots, m$, jsme aproximovali regresní funkcí

$$f(x, \boldsymbol{\theta}) = \theta_1 \varphi_1(x) + \cdots + \theta_n \varphi_n(x) = \boldsymbol{\varphi}(x)^T \boldsymbol{\theta},$$

kde parametry $\boldsymbol{\theta}$ jsou takové, aby $y_i \approx f(x_i, \boldsymbol{\theta})$ pro všechna i. Přibližné rovnosti $\approx$ jsme chápali ve smyslu nejmenších čtverců, tedy hledali jsme takové $\boldsymbol{\theta}$, které minimalizovalo funkci

$$\sum_{i=1}^{m} (y_i - f(x_i, \boldsymbol{\theta}))^2 = \|\mathbf{y} - \mathbf{A}\boldsymbol{\theta}\|_2, \tag{12.19}$$

kde $\mathbf{y} = (y_1, \ldots, y_m)$ a prvky matice $\mathbf{A}$ jsou $a_{ij} = \varphi_j(x_i)$. Tedy řešíme úlohu (12.14) pro $p = 2$. Můžeme ale použít i jiné normy než eukleidovskou. Pro $p = 1$ minimalizujeme

$$\sum_{i=1}^{m} |y_i - f(x_i, \boldsymbol{\theta})| = \|\mathbf{y} - \mathbf{A}\boldsymbol{\theta}\|_1 \tag{12.20}$$

a pro $p = \infty$ minimalizujeme

$$\max_{i=1}^{m} |y_i - f(x_i, \boldsymbol{\theta})| = \|\mathbf{y} - \mathbf{A}\boldsymbol{\theta}\|_\infty. \tag{12.21}$$

Dále ukážeme, k čemu to může být dobré.

Regrese ve smyslu ∞-normy je vhodná např. při aproximaci funkcí.

Příklad 12.7. Na počítači bez matematického koprocesoru potřebujeme mnohokrát vyhodnocovat funkci sinus na intervalu $[0, \frac{\pi}{2}]$. Protože výpočet hodnot této funkce by trval příliš dlouho, chceme ji aproximovat polynomem třetího stupně $\theta_1 + \theta_2 x + \theta_3 x^2 + \theta_4 x^3$, jehož hodnoty se spočítají rychleji. Spočítejme hodnoty $y_i = \sin x_i$ funkce v dostatečném počtu bodů $x_i = \frac{\pi i}{2m}$ pro $i = 1, \ldots, m$. Koeficienty polynomu hledáme minimalizací Čebyševova kritéria (12.21), neboť to nám dá záruku, že chyba aproximace v žádném z těchto bodů nepřesáhne hodnotu, která je nejmenší možná pro daný stupeň polynomu. ♦

Regrese ve smyslu 1-normy je užitečná tehdy, když je malá část hodnot y_i naměřená úplně chybně (např. se někdo při zapisování výsledků měření spletl v desetinné čárce). Takovým hodnotám se říká **vychýlené hodnoty** (*outliers*). Disciplína zabývající se modelováním funkčních závislostí za přítomnosti vychýlených hodnot se nazývá **robustní regrese**. V tomto případě řešení ve smyslu nejmenších čtverců není vhodné (není 'robustní'), protože i jediný vychýlený bod velmi ovlivní řešení. Regrese ve smyslu 1-normy je vůči vychýleným bodům odolnější.

Ukážeme to na velmi jednoduchém případu regrese: odhadu hodnoty jediného čísla ze souboru jeho nepřesných měření. To je vlastně polynomiální regrese polynomem nultého stupně $f(x, \theta) = \theta$, tj. konstantní funkcí (viz Příklad 5.4). Pro daná čísla $\mathbf{y} = (y_1, \ldots, y_m) \in \mathbb{R}^m$ hledáme $\theta \in \mathbb{R}$ minimalizující výraz

$$\|\mathbf{y} - \mathbf{1}\theta\|_p = \|(y_1 - \theta, \ldots, y_m - \theta)\|_p. \tag{12.22}$$

- Pro $p = \infty$ minimalizujeme $\max_{i=1}^m |y_i - \theta|$. Řešením je $\theta = \frac{1}{2}\big(\min_{i=1}^m y_i + \max_{i=1}^m y_i\big)$, tedy bod v polovině mezi krajními body.

- Pro $p = 2$ minimalizujeme $\sqrt{\sum_{i=1}^m (y_i - \theta)^2}$. Řešením je aritmetický průměr, $\theta = \frac{1}{m}\sum_{i=1}^m y_i$ (viz Příklad 5.4).
- Pro $p = 1$ minimalizujeme $\sum_{i=1}^m |y_i - \theta|$. Řešením je *medián* z čísel $y_1, \dots, y_m$, viz Příklad 1.13.

Předpokládejme nyní, že jedno z čísel, např. y_1, se zvětšuje. V tom případě se řešení θ pro různá p budou chovat různě. Např. aritmetický průměr se bude zvětšovat, a to tak, že zvětšováním hodnoty y_1 dosáhneme *libovolně velké* hodnoty θ. Pro medián to ovšem neplatí – zvětšováním jediného bodu y_1 ovlivníme θ jen natolik, nakolik to změní pořadí bodů. Jeho libovolným zvětšováním nedosáhneme libovolně velké hodnoty θ.

Příklad 12.8. Posuvným měřítkem (‘šuplérou’) změříme průměr ocelové kuličky v několika místech, dostaneme hodnoty $y_1 = 1.02$, $y_2 = 1.04$, $y_3 = 0.99$, $y_4 = 2.03$ (cm). Při posledním měření jsme se na stupnici přehlédli, proto je poslední hodnota úplně špatně. Z těchto měření chceme odhadnout skutečný průměr. Máme

$$\frac{1}{2}\big(\min_{i=1}^{m} y_i + \max_{i=1}^{m} y_i\big) = 1.51, \qquad \frac{1}{m}\sum_{i=1}^{m} y_i = 1.27, \qquad \operatorname*{median}_{i=1}^{m} y_i = 1.03.$$

Je zjevné, že medián je neovlivněn vychýleným bodem, zatímco ostatní odhady ano. ⧫

Ve složitějším případě, např. prokládání dat polynomem jako v Příkladu 5.4, se nedá robustnost řešení ve smyslu 1-normy takto jednoduše formálně ukázat a analýza je obtížnější. Výsledek ale bude podobný: řešení ve smyslu 1-normy je méně citlivé na vychýlené body než řešení ve smyslu 2-normy.

12.5 Celočíselné lineární programování, LP relaxace

Celočíselné lineární programování (angl. *integer linear programming*, ILP) se liší od lineárního programování dodatečným omezením, že proměnné musí nabývat pouze celočíselných hodnot. Nejčastěji používané jsou *binární* proměnné, které mohou nabývat jen dvou hodnot 0 a 1. ILP s binárními proměnnými je tedy úloha

$$\min\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in \{0,1\}^n,\ \mathbf{A}\mathbf{x} \geq \mathbf{b} \,\}. \tag{12.23}$$

Množina přípustných řešení této úlohy obsahuje konečný (avšak obvykle obrovský) počet izolovaných bodů, jedná se o typickou úlohu *kombinatorické optimalizace* (viz §1.3). Zatímco vyřešit obyčejný lineární program je relativně snadné (LP je řešitelné v polynomiálním čase), vyřešit úlohu ILP je obecně velmi obtížné (ILP je tzv. *NP-těžké*).

Mohli bychom zkusit úlohu (12.23) jednoduše nahradit lineárním programem

$$\min\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in [0,1]^n,\ \mathbf{A}\mathbf{x} \geq \mathbf{b} \,\}. \tag{12.24}$$

To je opravdu lineární program, protože omezení $\mathbf{x} \in [0,1]^n$ lze napsat jako $0 \leq x_i \leq 1$ ($i = 1, \dots, n$), což jsou lineární nerovnosti. Úloze (12.24) se říká **LP relaxace** (angl. *relaxation* znamená uvolnění, myšleno jako ‘uvolnění’ omezení) úlohy (12.23). LP relaxace je tedy nahrazení (těžké) úlohy ILP (snadnou) úlohou LP, ve které jsme vypustili omezení na celočíselnost.

Mohli bychom si myslet, že když optimální řešení $\mathbf{x} \in [0,1]^n$ LP relaxace (12.24) nějak šikovně zaokrouhlíme na celočíselný vektor $\mathbf{x}' \in \{0,1\}^n$, dostaneme řešení (nebo aspoň přibližné řešení) úlohy (12.23). To ale bohužel obecně neplatí. Zaokrouhlení totiž může zničit přípustnost, tedy zaokrouhlené proměnné $\mathbf{x}'$ nemusí splňovat podmínky $\mathbf{A}\mathbf{x}' \geq \mathbf{b}$. I to nejšikovnější zaokrouhlení nám nepomůže, neboť už jen odpovědět na otázku, zda úloha (12.23) je přípustná (tedy zda existuje alespoň jedno $\mathbf{x} \in \{0,1\}^n$ splňující $\mathbf{A}\mathbf{x} \geq \mathbf{b}$), je tzv. *NP-úplný problém* (zhruba řečeno, je to stejně těžké jako najít optimální řešení úlohy (12.23)). I kdybychom měli štěstí a zaokrouhlený vektor $\mathbf{x}'$ byl přípustný, hodnota $\mathbf{c}^T\mathbf{x}'$ účelové funkce může být obecně libovolně daleko od optimální hodnoty úlohy (12.23).

Obecně, jediný užitečný vztah mezi optimálním řešením původní a relaxované úlohy je, že optimální hodnota relaxované úlohy není větší než optimální hodnota původní úlohy:

$$\min\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in [0,1]^n,\ \mathbf{A}\mathbf{x} \geq \mathbf{b} \,\} \ \leq\ \min\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in \{0,1\}^n,\ \mathbf{A}\mathbf{x} \geq \mathbf{b} \,\}. \tag{12.25}$$

To proto, že v relaxované úloze minimalizujeme přes *větší* množinu než v původní úloze:

$$\{\, \mathbf{x} \in [0,1]^n \mid \mathbf{A}\mathbf{x} \geq \mathbf{b} \,\} \ \supseteq\ \{\, \mathbf{x} \in \{0,1\}^n \mid \mathbf{A}\mathbf{x} \geq \mathbf{b} \,\}.$$

Přesněji to lze uvidět takto: Nechť $\mathbf{x}^*$ je optimální řešení nerelaxované úlohy (12.23) a nechť $\mathbf{x}$ je optimální řešení relaxované úlohy (12.24). Protože $\mathbf{x}^*$ je přípustné (i když ne nutně optimální) pro relaxovanou úlohu, musí být $\mathbf{c}^T\mathbf{x} \leq \mathbf{c}^T\mathbf{x}^*$. Rozdílu mezi pravou a levou stranou nerovnosti (12.25) se říká *integrality gap* (tedy *mezera celočíselnosti*).

Pro některé konkrétní kombinatorické problémy však je situace příznivější, např. LP relaxace řeší původní úlohu vždy přesně (tedy nerovnost (12.25) platí s rovností), nebo dovoluje sestrojit její přibližné řešení se zárukou přesnosti (dále uvedeme příklady). I když toto štěstí nemáme, dolní mez je často užitečná při hledání přesného nebo přibližného řešení úlohy nějakým jiným způsobem (např. metodou *větví a mezí* nebo metodou *sečných nadrovin*).

12.5.1 Nejlepší přiřazení

V **přiřazovací úloze** (*assignment problem*, též znám jako *bipartitní párování*) jsou dána čísla c_{ij} pro $i,j \in \{1,\dots,n\}$ a cílem je najít bijektivní zobrazení $\pi\colon \{1,\dots,n\} \to \{1,\dots,n\}$ (tedy permutaci množiny $\{1,\dots,n\}$), která minimalizuje součet $\sum_{i=1}^n c_{i,\pi(i)}$.

Každé dvojici $i,j \in \{1,\dots,n\}$ přiřaďme binární proměnnou $x_{ij} \in \{0,1\}$. Tyto proměnné reprezentují permutaci π tak, že $x_{ij} = 1$ právě když $\pi(i) = j$. Protože π přiřazuje každému i právě jedno j, musí být mezi proměnnými $x_{i1},\dots,x_{in}$ právě jedna rovna 1. To zapíšeme jako $\sum_{j=1}^n x_{ij} = 1$. Protože π je bijekce, každý obraz π má právě jeden vzor, tedy $\sum_{i=1}^n x_{ij} = 1$. Nyní úlohu můžeme psát jako ILP

$$\begin{aligned}
\min \quad & \sum_{i=1}^n \sum_{j=1}^n c_{ij} x_{ij} && && (12.26a)\\
\text{za podm.} \quad & \sum_{j=1}^n x_{ij} = 1, && i = 1,\dots,n && (12.26b)\\
& \sum_{i=1}^n x_{ij} = 1, && j = 1,\dots,n && (12.26c)\\
& x_{ij} \in \{0,1\}, && i,j = 1,\dots,n && (12.26d)
\end{aligned}$$

LP relaxaci úlohy (12.26) získáme nahrazením omezení $x_{ij} \in \{0,1\}$ omezeními $0 \leq x_{ij} \leq 1$. Vlastně stačí jen $x_{ij} \geq 0$, neboť (12.26b) a $x_{ij} \geq 0$ implikuje $x_{ij} \leq 1$. Takže LP relaxace je speciální případ dopravního problému (12.11).

Pro tuto úlohu máme stěstí, protože platí toto (důkaz neuvádíme, není krátký):

Věta 12.1. Úloha (12.26) má stejnou optimální hodnotu jako její LP relaxace a mezi optimálními řešeními LP relaxace je aspoň jedno celočíselné.

Vyslovme tuto větu trochu jinak. Zapišme všechny proměnné x_{ij} do matice $\mathbf{X}$ velikosti $n \times n$. Účelová funkce (12.26a) je lineární funkce matice $\mathbf{X}$ (jde psát jako $\langle \mathbf{C}, \mathbf{X} \rangle$, viz §2.1.7). Přípustná řešení úlohy (12.26) jsou všechny matice, které mají v každém řádku a každém sloupci právě jednu jedničku. To jsou *permutační matice* (viz Příklad 4.5). Přípustná řešení LP relaxace jsou matice s nezápornými prvky, ve kterých součet prvků v každém řádku a v každém sloupci je roven 1. Takové matice se nazývají *dvojitě stochastické* matice. Každá permutační matice je dvojitě stochastická, ale naopak to neplatí. Věta 12.1 říká, že minimum lineární funkce na množině permutačních matic je rovno minimu té samé lineární funkce na množině dvojitě stochastických matic. To je klasický výsledek *Birkhoffa a von Neumanna*.

12.5.2 Nejmenší vrcholové pokrytí

Vrcholové pokrytí neorientovaného grafu (V, E) je podmnožina $X \subseteq V$ vrcholů taková, že každá hrana má alespoň jeden vrchol v X. V úloze na **nejmenší vrcholové pokrytí** hledáme pro daný graf vrcholové pokrytí s nejmenším počtem vrcholů. Je to jedna z klasických NP-těžkých úloh kombinatorické optimalizace.

Každému vrcholu $i \in V$ přiřaďme proměnnou $x_i \in \{0, 1\}$ s tímto významem: $x_i = 1$ pro $i \in X$ a $x_i = 0$ pro $i \notin X$. Úlohu můžeme formulovat jako ILP

$$
\begin{aligned}
\min \quad & \sum_{i \in V} x_i && && (12.27a)\\
\text{za podm.} \quad & x_i + x_j \geq 1, && \{i, j\} \in E && (12.27b)\\
& x_i \in \{0, 1\}, && i \in V && (12.27c)
\end{aligned}
$$

Označíme-li *incidenční matici* grafu (V, E) jako $\mathbf{A} \in \{0, 1\}^{n \times m}$ kde $n = |V|$ je počet vrcholů a $m = |E|$ je počet hran[6], úlohu (12.27) můžeme zapsat také jako

$$
\min\{\, \mathbf{1}^T \mathbf{x} \mid \mathbf{x} \in \{0, 1\}^n,\ \mathbf{A}^T \mathbf{x} \geq \mathbf{1} \,\}. \tag{12.28}
$$

LP relaxace úlohy (12.27) se získá tak, že omezení $x_i \in \{0, 1\}$ se nahradí omezeními $0 \leq x_i \leq 1$.

Nechť x_i $(i \in V)$ je optimální řešení relaxované úlohy. Zaokrouhleme toto řešení:

$$
\bar{x}_i = \left\lfloor x_i + \frac{1}{2} \right\rfloor = \begin{cases} 0 & \text{když } x_i < \frac{1}{2}, \\ 1 & \text{když } x_i \geq \frac{1}{2}. \end{cases}
$$

Zaokrouhlené řešení je přípustné pro úlohu (12.27): protože $x_i + x_j \geq 1$, musí být $x_i \geq \frac{1}{2}$ nebo $x_j \geq \frac{1}{2}$, tedy $\bar{x}_i = 1$ nebo $\bar{x}_j = 1$, tedy $\bar{x}_i + \bar{x}_j \geq 1$. Označme optimální hodnotu nerelaxované úlohy (12.27) jako y^*, optimální hodnotu relaxované úlohy jako $y = \sum_{i \in V} x_i$ a hodnotu zaokrouhleného řešení relaxace jako $\bar{y} = \sum_{i \in V} \bar{x}_i$.

Věta 12.2. Pro každý problém minimálního vrcholového pokrytí je $\bar{y} \leq 2y^*$.

[6] Zapakujme z teorie grafů význam prvků incidenční matice $\mathbf{A}$: je $a_{ij} = 1$ když vrchol i je jedním z koncových vrcholů hrany j.

Důkaz. Máme

$$\bar{y} = \sum_{i \in V} \bar{x}_i \le 2 \sum_{i \in V} x_i = 2y \le 2 \sum_{i \in V} x_i^* = 2y^*.$$

První nerovnost plyne z toho, že pro každé číslo $x \ge 0$ je $\lfloor x + \frac{1}{2} \rfloor \le 2x$, tedy $\bar{x}_i \le 2x_i$. Druhá nerovnost plyne z (12.25). ■

Celkově tedy máme tyto nerovnosti (rozmyslete):

$$0 \le y \le y^* \le \bar{y} \le 2y.$$

Z toho plyne $\frac{1}{2}y^* \le y \le y^*$, tedy optimální hodnota LP relaxace nemůže být libovolně-krát menší než optimální hodnota nerelaxované úlohy (mezera celočíselnosti nemůže být libovolná).

Příklad 12.9. Nechť $V = \{1,2,3\}$ a $E = \{\{1,2\},\{1,3\},\{2,3\}\}$, tedy (V,E) je úplný graf se třemi vrcholy. Úloha (12.27) má tři optimální řešení $\mathbf{x} = (1,1,0)$, $\mathbf{x} = (1,0,1)$ a $\mathbf{x} = (0,1,1)$, odpovídající minimálním pokrytím $X = \{1,2\}$, $X = \{1,3\}$ a $X = \{2,3\}$. Optimální hodnota úlohy je 2. LP relaxace má jediné optimální řešení $\mathbf{x} = (\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ s optimální hodnotou $\frac{3}{2}$, což je ve shodě s nerovností (12.25). Zaokrouhlené řešení je $\bar{\mathbf{x}} = (1,1,1)$ s optimální hodnotou 3, což je ve shodě s Větou 12.2. ♦

12.5.3 Největší nezávislá množina

Podmnožina vrcholů $X \subseteq V$ neorientovaného grafu (V,E) se nazývá **nezávislá**, když žádné dva vrcholy z X nejsou spojeny hranou. V úloze na **největší nezávislou množinu** hledáme pro daný graf nezávislou množinu s největším počtem vrcholů.

Formulace této úlohy pomocí ILP je velmi podobná (12.27):

$$\max \quad \sum_{i \in V} x_i \tag{12.29a}$$

$$\text{za podm.} \quad x_i + x_j \le 1, \qquad \{i,j\} \in E \tag{12.29b}$$

$$x_i \in \{0,1\}, \qquad i \in V \tag{12.29c}$$

V LP relaxaci opět nahradíme podmínky $x_i \in \{0,1\}$ nerovnostmi $0 \le x_i \le 1$.

Na rozdíl od úlohy na vrcholové pokrytí tato LP relaxace není příliš užitečná. LP relaxace má vždy přípustné řešení $x_i = \frac{1}{2}$ $(i \in V)$. Tomu odpovídá hodnota účelové funkce $\sum_i x_i = \frac{1}{2}|V|$. Tedy optimální hodnota LP relaxace úlohy nemůže být menší než $\frac{1}{2}|V|$. Bohužel, optimální hodnota y^* nerelaxované úlohy (12.29) může libovolně-krát menší než $\frac{1}{2}|V|$. Např. pro každý úplný graf je $y^* = 1$ (protože každé dva vrcholy jsou spojeny hranou).

Mohli bychom doufat, že např. zaokrouhlením optimálního řešení relaxované úlohy získáme analogii Věty 12.2. Tyto snahy jsou ale beznadějné: bylo dokázáno, že úlohu (12.29) nelze aproximovat s jakoukoliv konstantní (tj. nezávislou na velikosti grafu) zárukou. Přesněji, pro každé číslo $\varepsilon > 0$ je nalezení přípustného řešení $\bar{x}_i \in \{0,1\}$ $(i \in V)$ splňujícího $\sum_{i \in V} \bar{x}_i \ge y^*/\varepsilon$ stejně těžké, jako nalezení optimálního řešení (úloha je tzv. *APX-těžká*).

12.6 Cvičení

12.1. Najděte graficky množinu optimálních řešení úlohy

$$\begin{aligned} \min \quad & c_1x_1 + c_2x_2 + c_3x_3 \\ \text{za podm.} \quad & x_1 + x_2 \geq 1 \\ & x_1 + 2x_2 \leq 3 \\ & x_1 + x_2 \leq 10 \\ & x_1, x_2, x_3 \geq 0 \end{aligned}$$

pro následující případy: (a) $\mathbf{c} = (-1, 0, 1)$, (b) $\mathbf{c} = (0, 1, 0)$, (c) $\mathbf{c} = (0, 0, -1)$.

12.2. Následující úlohy nejprve převeďte na rovnicový tvar (tj. tvar s nezápornými proměnnými a omezeními typu lineární rovnice, viz §12.1). Potom je převeďte do maticové formy $\min\{\, \mathbf{r}^T\mathbf{u} \mid \mathbf{P}\mathbf{u} = \mathbf{q},\ \mathbf{u} \geq \mathbf{0} \,\}$ (výsledkem tedy budou $\mathbf{u}, \mathbf{P}, \mathbf{q}, \mathbf{r}$).

a)
$$\begin{aligned} \min \quad & 2x_1 \phantom{{}- x_2} - 3x_3 + x_4 \\ \text{za podmínek} \quad & x_1 - x_2 - x_3 \geq 0 \\ & -x_1 + 2x_2 - 3x_3 \leq 5 \\ & 2x_1 - x_2 - x_3 + 2x_4 = 6 \\ & x_1, x_2, x_3, x_4 \geq 0 \end{aligned}$$

b) lineární program (12.11)

c) lineární program (12.16)

d) lineární program (12.18)

12.3. Vyřešte úvahou tyto jednoduché lineární programy a napište (jednoduchý) výraz pro optimální hodnotu. Odpovědi dokažte. Vektor $\mathbf{c} \in \mathbb{R}^n$ a číslo $k \in \{1, \ldots, n\}$ jsou dány.

a) $\max\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{0} \leq \mathbf{x} \leq \mathbf{1} \,\}$

b) $\max\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ -\mathbf{1} \leq \mathbf{x} \leq \mathbf{1} \,\}$

c) $\max\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{x} \geq \mathbf{0},\ \mathbf{1}^T\mathbf{x} = 1 \,\}$

d) $\max\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{x} \geq \mathbf{0},\ \mathbf{1}^T\mathbf{x} \leq 1 \,\}$

e) $\max\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ -1 \leq \mathbf{1}^T\mathbf{x} \leq 1 \,\}$

f) $\max\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{x} \geq \mathbf{0},\ \mathbf{1}^T\mathbf{x} = k \,\}$

g) $\max\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{0} \leq \mathbf{x} \leq \mathbf{1},\ \mathbf{1}^T\mathbf{x} \leq k \,\}$

h) $\max\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ 0 \leq x_1 \leq x_2 \leq \cdots \leq x_n \leq 1 \,\}$

i) ($\star$) $\max\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x}, \mathbf{y} \in \mathbb{R}^n,\ -\mathbf{y} \leq \mathbf{x} \leq \mathbf{y},\ \mathbf{1}^T\mathbf{y} = k,\ \mathbf{y} \leq \mathbf{1} \,\}$

j) $\min\{\, \mathbf{1}^T(\mathbf{u} + \mathbf{v}) \mid \mathbf{u}, \mathbf{v} \geq \mathbf{0},\ \mathbf{u} - \mathbf{v} = \mathbf{c} \,\}$

k) ($\star$) $\min\{\, \mathbf{a}^T\mathbf{u} + \mathbf{b}^T\mathbf{v} \mid \mathbf{u}, \mathbf{v} \geq \mathbf{0},\ \mathbf{u} - \mathbf{v} = \mathbf{c} \,\}$, kde předpokládáme $\mathbf{a} \geq \mathbf{b}$

12.4. Pokuste se úlohy transformovat na LP. Pokud to nedokážete, vysvětlete proč.

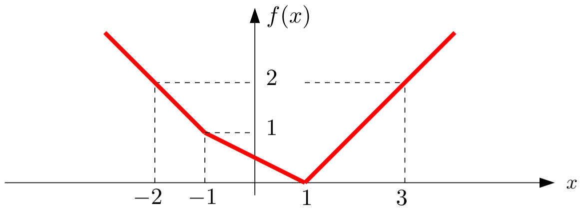
a) $\min\{\, |x_1| + |x_2| \mid x_1, x_2 \in \mathbb{R},\ 2x_1 - x_2 \geq 1,\ -x_1 + 2x_2 \geq 1 \,\}$

b) $\max\{\, |x_1 - c_1| + \cdots + |x_n - c_n| \mid x_1, \ldots, x_n \in \mathbb{R},\ a_1x_1 + \cdots + a_nx_n \geq b \,\}$

c) $\max\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{A}\mathbf{x} = \mathbf{b},\ |\mathbf{d}^T\mathbf{x}| \leq 1,\ \mathbf{x} \geq \mathbf{0} \,\}$

d) $\min\limits_{\mathbf{x} \in \mathbb{R}^n} \sum_{l=1}^{L} \max_{k=1}^{K} (\mathbf{c}_{kl}^T\mathbf{x} + d_{kl})$

e) $\min\limits_{\mathbf{x} \in \mathbb{R}^n} \sum_{i=1}^{m} f(\mathbf{a}_i^T\mathbf{x} - b_i)$, kde po částech afinní funkce f je definována obrázkem



f) $\min\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_\infty \le 1 \,\}$

g) $\min\{\, \|\mathbf{x}\|_1 \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{A}\mathbf{x} = \mathbf{b} \,\}$

h) $\min_{\mathbf{x}\in\mathbb{R}^n}(\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_1 + \|\mathbf{x}\|_\infty)$ (což se dá napsat také jako $\min\{\, \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_1 + \|\mathbf{x}\|_\infty \mid \mathbf{x} \in \mathbb{R}^n \,\}$)

i) ($\star$) $\min\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_1 \le 1 \,\}$

12.5. Máme algoritmus (černou skříňku) na řešení LP, kterou můžeme zavolat i vícekrát. S pomocí tohoto algoritmu vyřešte úlohu $\max\{\, |\mathbf{c}^T\mathbf{x}| \mid \mathbf{A}\mathbf{x} = \mathbf{b},\ \mathbf{x} \ge \mathbf{0} \,\}$.

12.6. Dokažte nebo vyvraťte následující rovnosti. Zde $\mathbf{c} \in \mathbb{R}^n$ a $\mathbf{A} \in \mathbb{R}^{m\times n}$ jsou dány, $\|\cdot\|$ je libovolná norma, a optimalizuje se přes proměnné $\mathbf{x} \in \mathbb{R}^n$.

a) $\max\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ \|\mathbf{x}\| = 1 \,\} = \max\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ \|\mathbf{x}\| \le 1 \,\}$

b) $\min\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ \|\mathbf{x}\| = 1 \,\} = \min\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ \|\mathbf{x}\| \le 1 \,\}$

c) $\max\{\, \|\mathbf{A}\mathbf{x}\| \mid \mathbf{x} \in \mathbb{R}^n,\ \|\mathbf{x}\| = 1 \,\} = \max\{\, \|\mathbf{A}\mathbf{x}\| \mid \mathbf{x} \in \mathbb{R}^n,\ \|\mathbf{x}\| \le 1 \,\}$

12.7. Nechť $X \subseteq \mathbb{R}^n$ je libovolná množina. Dokažte, že

a) $\min\{\, \mathbf{1}^T(\mathbf{u}+\mathbf{v}) \mid \mathbf{u}-\mathbf{v} \in X,\ \mathbf{u},\mathbf{v} \ge \mathbf{0} \,\} = \min_{\mathbf{x}\in X} \|\mathbf{x}\|_1$,

b) $\min\{\, \mathbf{1}^T\mathbf{u} \mid \mathbf{u}-\mathbf{v} \in X,\ \mathbf{u},\mathbf{v} \ge \mathbf{0} \,\} = \min_{\mathbf{x}\in X} \mathbf{x}_+$, kde značíme $\mathbf{x}_+ = \sum_{i=1}^n \max\{0, x_i\}$,

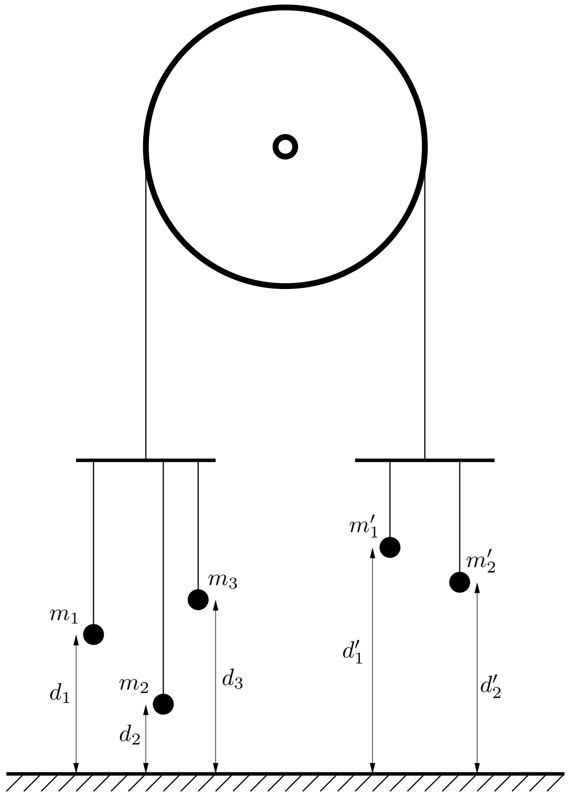
za předpokladu, že všechny úlohy mají optimální řešení. Všimněte si, že první úloha je zobecnění Cvičení 12.3.j pro X obsahující jediný prvek.

12.8. Uvažujme Příklad 1.10 takto pozměněný: pán má pomocníka, kterému platí 10 Kč za každý kg vyrobeného zboží (je jedno, zda to jsou lupínky nebo hranolky). Ovšem pokud se toho vyrobí hodně, chce pomocník větší plat, protože musí zůstat přesčas. Tak za každý kg nad 20 kg vyrobeného zboží si nechá připlatit dalších 10 Kč, a za každý kg nad 30 kg vyrobeného zboží si nechá připlatit dalších 20 Kč (tedy za každý kg nad 30 kg vyrobeného zboží dostane $10+10+20 = 40$ Kč). Kolik má pán vyrobit lupínků a hranolků, aby měl co největší denní zisk (tj. tržbu z prodeje minus plat pomocníkovi)? Zformulujte jako LP.

12.9. Firma na výrobu kánoí má 120 zaměstnanců, z nichž každý pracuje maximálně 30 hodin týdně. Polovina zaměstnanců pracuje v truhlářské dílně, 20 zaměstnanců pracuje v dílně na zpracování plastů a zbytek v kompletační dílně. Firma vyrábí dva typy kánoí: standardní kánoe s čistým ziskem 7 EUR za kus a luxusní kánoe s čistým ziskem 10 EUR za kus. Na výrobu jedné standardní kánoe je třeba 4.5 hodiny práce v truhlářské dílně a dvě hodiny v každé ze zbylých dvou dílen. Jedna luxusní kánoe vyžaduje 5 hodin práce v truhlárně, hodinu v dílně na plasty a 4 hodiny kompletace. Průzkum trhu odhalil, že ne méně jež 1/3 a ne více než 2/3 vyrobených kánoí by měly být luxusní. Kolik kterých kánoí má firma týdně vyrobit, aby byl její čistý zisk maximální? Formalizujte jako optimalizační úlohu, kterou už ale neřešte.

12.10. Armáda má ve dvou skladech uskladněno 6 a 5 tun střeliva. Střelivo je nutno přepravit ke třem střelnicím s požadavky 3, 2 a 2 tuny střeliva. Kvůli minimalizaci rizika je nutné minimalizovat maximální množství střeliva, které se veze po kterékoliv cestě od skladu ke střelnici. Formulujte jako lineární program.

12.11. Máme kladku s provazem, jehož oba konce končí hákem. Na levém háku visí n závaží na provázcích, přičemž i-té závaží má tíhu m_i a jeho výška nad zemí je d_i, pro $i = 1, \ldots, n$. Na pravém háku visí n' závaží na provázcích, přičemž i-té závaží má tíhu m'_i a jeho výška nad zemí je d'_i, pro $i = 1, \ldots, n'$. Výšky d_i a d'_i se měří v poloze, kdy jsou oba háky ve stejné výšce nad zemí. Kladka se pohybuje bez tření, provázky a háky mají nulovou hmotnost. Obrázek ukazuje příklad pro $n = 3$, $n' = 2$.



Soustava má jediný stupeň volnosti daný otáčením kladky. Označme jako x výšku levého háku nad bodem, kdy jsou oba háky ve stejné výšce – tedy pro $x = 0$ jsou oba háky ve stejné výšce a pro $x > 0$ bude levý hák o $2x$ výše než pravý hák. V závislosti na x každé závaží buď visí nad zemí (pak je jeho potenciální energie rovna m_i krát výška nad zemí) nebo leží na zemi (pak je jeho potenciální energie nulová). Soustava bude v rovnováze při minimální celkové potenciální energii.

a) Napište vzorec pro celkovou potenciální energii soustavy jako funkci x.
b) Formulujte hledání minima potenciální energie soustavy jako lineární program.

12.12. Veverka před zimou potřebuje přerovnat zásoby oříšků. Stávající zásoby má v m jamkách, přičemž i-tá jamka má souřadnice $\mathbf{p}_i \in \mathbb{R}^2$ a je v ní a_i oříšků. Potřebuje je přenosit do n nových připravených jamek, přičemž j-tá jamka má souřadnice $\mathbf{q}_j \in \mathbb{R}^2$ a na konci v ní bude y_j oříšků. Veverka unese najednou jen jeden oříšek. Nechť x_{ij} označuje celkový počet oříšků přenesených ze staré jamky i do nové jamky j. Uvažujte dvě úlohy:

a) Čísla y_j jsou dána. Hledají se taková čísla x_{ij}, aby veverka vykonala co nejméně práce, kde práce na přenesení jednoho oříšku je přímo úměrná vzdálenosti (vzdušnou čarou). Běh bez oříšku se za práci nepovažuje.
b) Hledají se čísla x_{ij} a y_j tak, aby veverka vykonala co nejméně práce a navíc byly v nových jamkách oříšky rozloženy co nejrovnoměrněji, čímž minimalizuje škodu způsobenou případnou krádeží. Přesněji, aby rozdíl mezi největším a nejmenším z čísel y_j byl menší než dané číslo t.

Formulujte obě úlohy jako LP. Předpokládejte, že počty oříšků jsou nezáporná reálná čísla, ač ve skutečnosti mohou být pouze nezáporná celá čísla.

12.13. Spočtěte limitu $\lim_{p\to\infty} \|\mathbf{x}\|_p = \lim_{p\to\infty}(|x_1|^p + \cdots + |x_n|^p)^{1/p}$.

12.14. Máme dvě konečné množiny $\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ a $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ bodů z $\mathbb{R}^d$. Hledáme

a) nadrovinu procházející počátkem (tj. lineární podprostor dimenze $d - 1$),

b) nadrovinu (tj. afinní podprostor dimenze $d-1$),

která množiny odděluje (tedy všechny body z první množiny jsou na jedné straně nadroviny a všechny body z druhé množiny jsou na druhé straně). Napište jako lineární program.

12.15. Rozhodněte (a odpověď dokažte), pro jaká n je následující funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ norma:

a) $f(\mathbf{x}) = |\max\{x_1,\ldots,x_n\}|$
b) $f(\mathbf{x}) = \|\mathbf{x}\|_1 + \|\mathbf{x}\|_2$
c) $f(\mathbf{x},\mathbf{y}) = \|\mathbf{x}\|_1 + \|\mathbf{y}\|_2$ (argument této funkce je vektor $(\mathbf{x},\mathbf{y}) \in \mathbb{R}^n$, kde $\mathbf{x} \in \mathbb{R}^k$ a $\mathbf{y} \in \mathbb{R}^{n-k}$)
d) $f(\mathbf{x}) = \max_{i=1}^n x_i - \min_{i=1}^n x_i$

12.16. Dokažte, že p-norma splňuje axiomy vektorové normy.

12.17. Hledáme $\mathbf{x}$, které minimalizuje výraz $f(\mathbf{Ax}-\mathbf{b})$, kde

a) $f(\mathbf{r}) = \|\mathbf{r}\|_\infty = \max_i |r_i|$,
b) $f(\mathbf{r}) = \|\mathbf{r}\|_1 = \sum_i |r_i|$,
c) $f(\mathbf{r}) = \sum_i \max\{-r_i-\theta, 0, r_i-\theta\}$ kde $\theta > 0$ je daná konstanta.

Formulujte jako lineární programy. Kdybychom tyto úlohy interpetovali jako prokládání bodů regresní funkcí (jako v §12.4.3), jaký je geometrický význam těchto úloh?

12.18. Máme m nadrovin v $\mathbb{R}^n$, kde i-tá nadrovina je množina $H_i = \{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}_i^T\mathbf{x} = b_i \,\}$ kde $\mathbf{a}_i \neq \mathbf{0}$. Označme vzdálenost bodu $\mathbf{x} \in \mathbb{R}^n$ od i-té nadroviny jako $d(H_i,\mathbf{x}) = \min\limits_{\mathbf{y}\in H_i} \|\mathbf{x}-\mathbf{y}\|_2$. Hledáme bod $\mathbf{x}$, který minimalizuje výraz

a) $\max_{i=1}^m d(H_i,\mathbf{x})$,
b) $\sum_{i=1}^m d(H_i,\mathbf{x})$.

Formulujte jako lineární program. Rozmyslete geometrický význam úloh, načrtněte si řešení pro $n=2$ a $m=3$.

12.19. ($\star$) Najděte co nejjednodušší algoritmus, který spočte minimum funkce $\mathbf{a}^T\mathbf{X}\mathbf{b}$ na množině dvojitě stochastických matic $\mathbf{X}$, kde $\mathbf{a},\mathbf{b}$ jsou dané vektory.

Nápověda a řešení

12.2.b) Úloha už je v rovnicovém tvaru, takže v prvním kroku nemusíme dělat nic. Úloha jde napsat v maticovém tvaru (12.12), to ale není požadovaný tvar, protože proměnné jsou soustředěné do matice $\mathbf{X}$ a nikoliv do vektoru. Převod do požadovaného tvaru je jasný z Cvičení 2.6.c.

12.3.a) Postupujeme podobně jako v Příkladu 12.4: nejdřív uhodneme optimální řešení, a pak dokážeme jeho optimalitu tak, že ukážeme, že každé jiné řešení jde změnit (bez porušení přípustnosti) tak, že kritérium se zlepší.

Úlohu si můžeme přepsat jako $\max\{\, \sum_i c_i x_i \;\big|\; x_i \in \mathbb{R},\; 0 \le x_i \le 1\; \forall i \,\}$. Všimneme si, že tato úloha jde vlastně rozdělit na n nezávislých úloh $\max\{\, c_i x_i \mid x_i \in \mathbb{R},\; 0 \le x_i \le 1 \,\}$, z nichž každá má jedinou proměnnou x_i. Tuto jednorozměrnou úlohu snadno vyřešíme elementární úvahou: v optimu je $x_i = 0$ pro $c_i < 0$, a $x_i = 1$ pro $c_i > 0$. Kdyby totiž např. bylo $x_i > 0$ a $c_i < 0$, tak bychom mohli kritérium zlepšit položením $x_i = 0$. Optimální hodnota jednorozměrné úlohy je proto $\max\{c_i, 0\}$. Opt. hodnota úvodní úlohy je tedy $\sum_i \max\{c_i, 0\}$.

12.3.b) Postupujeme obdobně. Optimální hodnota je $\sum_{i=1}^n |c_i|$.

12.3.c) Viz Příklad 12.4.

12.3.d) $\max_{i=1}^n \max\{0, c_i\} = \max\{0, \max_{i=1}^n c_i\}$

12.3.e) Když $c_1 = c_2 = \cdots = c_n = a$, tak optimální hodnota je $|a|$. Jinak je úloha neomezená.

12.3.h) Převeďte na jinou úlohu substitucí $y_i = x_i - x_{i-1}$ a pak postupujte jako v předešlých úlohách.

12.3.j) Úlohu si můžeme přepsat jako $\min\{\, \sum_i (u_i + v_i) \;\big|\; u_i, v_i \ge 0,\; u_i - v_i = c_i \;\forall i \,\}$. Tato úloha se opět rozpadá na n nezávislých úloh $\min\{\, u_i + v_i \mid u_i, v_i \ge 0,\; u_i - v_i = c_i \,\}$, z nichž každá má jen dvě proměnné u_i, v_i. Každou tuto jednoduchou úlohu již snadno vyřešíme elementární úvahou: v optimu bude vždy aspoň jedno z čísel u_i, v_i nulové, protože jinak bychom mohli od obou odečíst konstantu, což by neporušilo omezení ale zlepšilo účelovou funkci. Optimum je $(u_i, v_i) = (c_i, 0)$ jestliže $c_i \ge 0$ a $(u_i, v_i) = (0, c_i)$ jestliže $c_i \le 0$. Tedy opt. hodnota jednoduché úlohy je $|c_i|$ a opt. hodnota původní úlohy je $\sum_i |c_i| = \|\mathbf{c}\|_1$.

12.4.a) $\min\{\, z_1 + z_2 \mid x_1, x_2, z_1, z_2 \in \mathbb{R},\; 2x_1 - x_2 \ge 1,\; -x_1 + 2x_2 \ge 1,\; x_1 \le z_1,\; x_2 \le z_2,\; -x_1 \le z_1,\; -x_2 \le z_2 \,\}$

12.4.b) Nejde.

12.4.c) $\max\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{A}\mathbf{x} = \mathbf{b},\; \mathbf{d}^T\mathbf{x} \le 1,\; -\mathbf{d}^T\mathbf{x} \le 1,\; \mathbf{x} \ge \mathbf{0} \,\}$

12.4.d) $\min\{\, \mathbf{1}^T\mathbf{z} \mid \mathbf{c}_{kl}^T\mathbf{x} + d_{kl} \le z_l \;(\forall k, l),\; \mathbf{x} \in \mathbb{R}^n,\; \mathbf{z} \in \mathbb{R}^L \,\}$ (analogické §12.1.1)

12.4.f) $\min\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\; -\mathbf{1} \le \mathbf{A}\mathbf{x} - \mathbf{b} \le \mathbf{1} \,\}$

12.4.h) $\min\{\, \mathbf{1}^T\mathbf{y} + z \mid \mathbf{x} \in \mathbb{R}^n,\; \mathbf{y} \in \mathbb{R}^m,\; z \in \mathbb{R},\; -\mathbf{y} \le \mathbf{A}\mathbf{x} - \mathbf{b} \le \mathbf{y},\; -z\mathbf{1} \le \mathbf{x} \le z\mathbf{1} \,\}$

12.4.i) $\min\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\; \mathbf{z} \in \mathbb{R}^m,\; -\mathbf{z} \le \mathbf{A}\mathbf{x} - \mathbf{b} \le \mathbf{z},\; \mathbf{1}^T\mathbf{z} \le 1 \,\}$. Správnost tohoto převodu bychom museli dokázat podobně jako v Příkladu 12.4 – proveďte! Všimněte si také, že $-\mathbf{z} \le \mathbf{A}\mathbf{x} - \mathbf{b} \le \mathbf{z}$ implikuje $\mathbf{z} \ge \mathbf{0}$, tedy bychom mohli přidat podmínku $\mathbf{z} \ge \mathbf{0}$ a úloha by se nezměnila.

12.5. Postupem uvedeným v §12.1 nedokážeme převést na jedinou úlohu LP. Ale lze vyřešit vypočtením dvou úloh LP: optimální hodnota je $\max\{A, -B\}$, kde $A = \max\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{A}\mathbf{x} = \mathbf{b},\; \mathbf{x} \ge \mathbf{0} \,\}$ a $B = \min\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{A}\mathbf{x} = \mathbf{b},\; \mathbf{x} \ge \mathbf{0} \,\}$.

12.6. Inspirujte se úvahou v §12.1.1.

12.7.a) Dokažte nejprve, že v optimu je buď $\mathbf{u} = \mathbf{0}$ nebo $\mathbf{v} = \mathbf{0}$. Pak např. uvažujte úvahu v §12.1.1 a použijte vhodnou substituci.

12.8. Stejná úloha jako v Příkladu 1.10, jen účelová funkce se změní na $120l + 76h - f(l + h)$, kde $f(t) = \max\{\, 10t,\; 200 + 20(t - 20),\; 400 + 40(t - 30) \,\}$. To převedeme na LP dle §12.1.1.

12.9. $\max 7s + 10l$ z.p. $4.5s + 5l \le 60 \cdot 30$, $2s + l \le 20 \cdot 30$, $2s + 4l \le 40 \cdot 30$, $(s + l)/3 \le l \le 2(s + l)/3$, $s, l \ge 0$ a navíc $s, l \in \mathbb{Z}$.

12.10. Nechť x_{ij} označuje množství střeliva vezeného od skladu i do střelnice j. Minimalizujeme funkci $\max\{x_{11}, x_{12}, x_{13}, x_{21}, x_{22}, x_{23}\}$ za podmínek $x_{11} + x_{12} + x_{13} \le 6$, $x_{21} + x_{22} + x_{23} \le 5$, $x_{11} + x_{21} = 3$, $x_{12} + x_{22} = 2$, $x_{13} + x_{23} = 2$, a $x_{11}, x_{12}, x_{13}, x_{21}, x_{22}, x_{23} \ge 0$. To převedeme na LP dle §12.1.1.

12.11.a) Z fyziky víte, že závaží hmotnosti m ve výšce d má potenciální energii mgd, kde g je gravitační zrychlení. Součinitel g vynecháme, neboli veličinou m rozumíme nikoliv hmotnost, ale tíhovou sílu ('tíhu') závaží. Výška jednoho závaží na levé straně je $d_i - x$, ovšem nesmí být záporná protože závaží není nikdy pod podlahou, tedy vlastně $\max\{d_i - x, 0\}$. Celková potenciální energie je $E(x) = \sum_{i=1}^n m_i \max\{d_i + x, 0\} + \sum_{i=1}^{n'} m'_i \max\{d'_i - x, 0\}$.

12.11.b) Podobně jako v §12.1.1 si zavedeme pomocné proměnné z_i, z'_i. Výsledný lineární program bude $\min\{\, \sum_{i=1}^n m_i z_i + \sum_{i=1}^{n'} m'_i z'_i \mid x, z_i, z'_i \in \mathbb{R},\; z_i \ge d_i + x,\; z'_i \ge d'_i - x,\; z_i \ge 0,\; z'_i \ge 0 \,\}$

12.13. Stačí spočítat $\lim_{p\to\infty} (x_1^p + \cdots + x_n^p)^{1/p}$ pro $x_1, \ldots, x_n \ge 0$. Označte $a = \max\{x_1, \ldots, x_n\}$ a pište $\lim_{p\to\infty} (x_1^p + \cdots + x_n^p)^{1/p} = a \lim_{p\to\infty} ((x_1/a)^p + \cdots + (x_n/a)^p)^{1/p}$. Zbytek je snad jasný.

12.14.a) Podmínku oddělení napište jako soustavu lineárních nerovnic, což je vlastně LP s konstantní (nulovou) účelovou funkcí. Viz Příklad 17.12.

12.15.a) Je to norma pro $n = 1$, protože pak $f(x) = |x|$. Pro $n \geq 2$ to norma není, protože z $f(\mathbf{x}) = 0$ neplyne $\mathbf{x} = \mathbf{0}$ (první axiom), např. pro $\mathbf{x} = (-1, 0, 0, \ldots)$.

12.17. (a) a (b) už znáte: po dosazení $\mathbf{r} = \mathbf{A}\mathbf{x} - \mathbf{b}$ do $f(\mathbf{r})$ jde o přibližné řešení přeurčené lineární soustavy ve smyslu ∞-normy a 1-normy (viz §12.4). Případ (c) lze také vidět jako přibližné řešení přeurčené lineární soustavy, ale tentokrát s 'necitlivým pásmem': pokud residuum r_i je v intervalu $[-\theta, \theta]$, nezapočítáme žádnou pokutu; vně intervalu je to podobné 1-normě. Nakreslete si graf funkce $\max\{-r - \theta, 0, r - \theta\}$ a vše pochopíte.

12.18. Pro pohodlí předpokládejme, že každý vektor $\mathbf{a}_i$ je normalizovaný, což lze vždy zajistit vydělením rovnice $\mathbf{a}_i^T \mathbf{x} = b_i$ číslem $\|\mathbf{a}_i\|$. Pak je vzdálenost bodu $\mathbf{x}$ od i-té nadroviny rovna jednoduše $d(H_i, \mathbf{x}) = |\mathbf{a}_i^T \mathbf{x} - b_i|$ (viz Cvičení 5.16). Tedy podúloha (a) je stejná jako úloha (12.15), tj. minimalizace $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_\infty$. Podobně, podúloha (b) je úloha (12.17), tj. minimalizace $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_1$. Vidíme, že úloha je jen 'zamaskovaná' úloha na přibližné řešení přeurčené soustavy $\mathbf{A}\mathbf{x} = \mathbf{b}$ v ∞-normě nebo 1-normě (viz §12.4), za předpokladu že řádky matice $\mathbf{A}$ jsou normalizované.

12.19. Dle Věty 12.1 se optimální hodnota úlohy nezmění, když budeme minimalizovat přes permutační matice. Tedy hledáme minimum funkce $\sum_i a_i b_{\pi(i)}$ přes všechny permutace π. Zkuste na to vymyslet efektivní algoritmus.

Kapitola 13

Konvexní množiny a mnohostěny

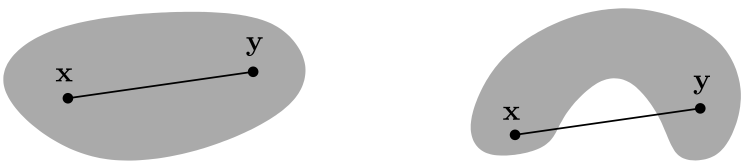
Zde si řekneme více o geometrii lineárního programování. Ke konvexním množinám, zásadnímu to pojmu v optimalizaci, se navíc vrátíme v pozdějších kapitolách.

13.1 Konvexní množiny

Množina $X \subseteq \mathbb{R}^n$ se nazývá **konvexní**, jestliže

$$\mathbf{x} \in X,\ \mathbf{y} \in X,\ 0 \leq \alpha \leq 1 \quad \Longrightarrow \quad (1-\alpha)\mathbf{x} + \alpha\mathbf{y} \in X. \tag{13.1}$$

Množina $\{\, (1-\alpha)\mathbf{x} + \alpha\mathbf{y} \mid 0 \leq \alpha \leq 1 \,\}$ je úsečka spojující body $\mathbf{x}$ a $\mathbf{y}$ (viz Příklad 3.3). Definice tedy říká, že množina je konvexní, jestliže s každými dvěma body obsahuje i úsečku, která je spojuje. Obrázek ukazuje příklad konvexní a nekonvexní množiny v $\mathbb{R}^2$:



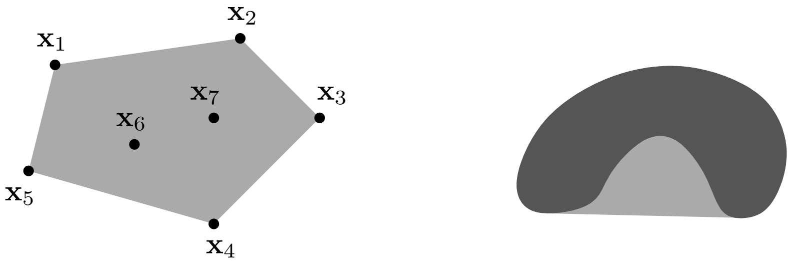
Konvexní kombinace bodů $\mathbf{x}_1, \ldots, \mathbf{x}_k \in \mathbb{R}^n$ je jejich lineární kombinace $\alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k$ taková, že $\alpha_1 + \cdots + \alpha_k = 1$ a $\alpha_1, \ldots, \alpha_k \geq 0$. Lze dokázat, že množina je konvexní právě tehdy, když je uzavřená na konvexní kombinace (neboli každá konvexní kombinace bodů z množiny leží v množině). Všimněte si, že $(1-\alpha)\mathbf{x} + \alpha\mathbf{y}$ pro $0 \leq \alpha \leq 1$ je konvexní kombinací dvou bodů $\mathbf{x}, \mathbf{y}$, neboť $(1-\alpha) + \alpha = 1$, $1 - \alpha \geq 0$, $\alpha \geq 0$.

Konvexní obal bodů $\mathbf{x}_1, \ldots, \mathbf{x}_k$ je množina všech jejich konvexních kombinací, značíme

$$\operatorname{conv}\{\mathbf{x}_1, \ldots, \mathbf{x}_k\} = \{\, \alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k \mid \alpha_1 + \cdots + \alpha_k = 1,\ \alpha_1, \ldots, \alpha_k \geq 0 \,\}. \tag{13.2}$$

Jak ale definovat konvexní obal množiny s *nekonečným* počtem prvků (např. na pravém obrázku výše)? Nelze použít definice (13.2), neboť není jasné, co znamená součet $\alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k$ pro nekonečný počet bodů. **Konvexní obal množiny** $X \subseteq \mathbb{R}^n$ (konečné či nekonečné) definujeme jako průnik všech konvexních množin, které množinu X obsahují. Značíme jej conv X.

Obrázek ukazuje konvexní obal konečné (vlevo) a nekonečné (vpravo) množiny pro $n = 2$:



Tvrzení 13.1. Průnik konvexních množin je konvexní množina.

Důkaz. Dokážeme jen pro dvě množiny, i když platí pro více (i nespočetně mnoho) množin. Nechť $X, Y \subseteq \mathbb{R}^n$ jsou konvexní. Nechť $\mathbf{x}, \mathbf{y} \in X \cap Y$, tedy $\mathbf{x}, \mathbf{y} \in X$ a $\mathbf{x}, \mathbf{y} \in Y$. Proto pro $0 \leq \alpha \leq 1$ je bod $(1-\alpha)\mathbf{x} + \alpha\mathbf{y}$ také v X i Y, tedy je v $X \cap Y$. ■

Sjednocení konvexních množin ale nemusí být konvexní množina.

13.2 Čtyři kombinace, čtyři obaly

Konvexní kombinace je lineární kombinace, jejíž koeficienty splňují omezení $\alpha_1 + \cdots + \alpha_k = 1$ a $\alpha_1, \ldots, \alpha_k \geq 0$. Všimněte si, že když vynecháme druhé omezení, dostaneme afinní kombinaci (viz §3.3). Podle toho, které ze dvou omezení vyžadujeme, dostaneme čtyři druhy kombinací. Udělejme si v nich nyní pořádek.

Vážený součet $\alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k$ vektorů[1] $\mathbf{x}_1, \ldots, \mathbf{x}_k \in \mathbb{R}^n$ se nazývá jejich

lineární kombinace,	jestliže	$\alpha_1, \ldots, \alpha_k \in \mathbb{R}$.		
afinní kombinace,	jestliže	$\alpha_1, \ldots, \alpha_k \in \mathbb{R}$,	$\alpha_1 + \cdots + \alpha_k = 1$.	
nezáporná kombinace,	jestliže	$\alpha_1, \ldots, \alpha_k \in \mathbb{R}$,		$\alpha_1, \ldots, \alpha_k \geq 0$.
konvexní kombinace,	jestliže	$\alpha_1, \ldots, \alpha_k \in \mathbb{R}$,	$\alpha_1 + \cdots + \alpha_k = 1$,	$\alpha_1, \ldots, \alpha_k \geq 0$.

Množina, která je uzavřená na

lineární kombinace,	se nazývá	**lineární podprostor**.
afinní kombinace,	se nazývá	**afinní podprostor**.
nezáporné kombinace,	se nazývá	**konvexní kužel**.
konvexné kombinace,	se nazývá	**konvexní množina**.

K tomu, co již znáte, přibyl pojem nezáporné kombinace a konvexního kuželu.

Lineární [afinní, nezáporný, konvexní] **obal** vektorů $\mathbf{x}_1, \ldots, \mathbf{x}_k$ je množina všech jejich lineárních [afinních, nezáporných, konvexních] kombinací. Obecněji, lineární [afinní, nezáporný, konvexní] obal (konečné či nekonečné) množiny $X \subseteq \mathbb{R}^n$ je průnik všech lineárních podprostorů [afinních podprostorů, konvexních kuželů, konvexních množin] obsahující množinu X.

Příklad 13.1. Mějme tři body v $\mathbb{R}^3$, které neleží v jedné rovině s počátkem. Jejich lineární obal je celé $\mathbb{R}^3$. Jejich afinní obal je rovina jimi procházející. Jejich nezáporný obal je nekonečný trojboký hranol, jehož vrchol je v počátku a jehož hrany jsou tři polopřímky určené počátkem a danými body. Jejich konvexní obal je trojúhelník jimi určený. ♦

[1] Ve světle poznámky v §3.3 tyto vektory také nazýváme *body*, vystupujé-li v afinní nebo konvexní kombinaci.

13.3 Konvexní mnohostěny

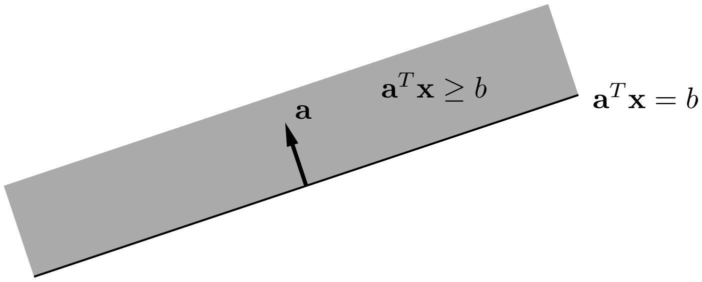
Zopakujme (viz §3.3), že *nadrovina* v $\mathbb{R}^n$ je afinní podprostor dimenze $n-1$. Je to tedy množina

$$\{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}^T\mathbf{x} = b \,\}, \tag{13.3}$$

kde $\mathbf{a} \in \mathbb{R}^n$ a $b \in \mathbb{R}$. Požadujeme $\mathbf{a} \neq \mathbf{0}$, jinak by afinní podprostor (13.3) neměl dimenzi $n-1$ (a nebyl by tedy nadrovinou). Vektor $\mathbf{a}$ je normála nadroviny a $|b|/\|\mathbf{a}\|$ je vzdálenost nadroviny od počátku (viz §5.2.1). **Poloprostor** (přesněji *uzavřený poloprostor*[2]) je množina

$$\{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}^T\mathbf{x} \geq b \,\}. \tag{13.4}$$

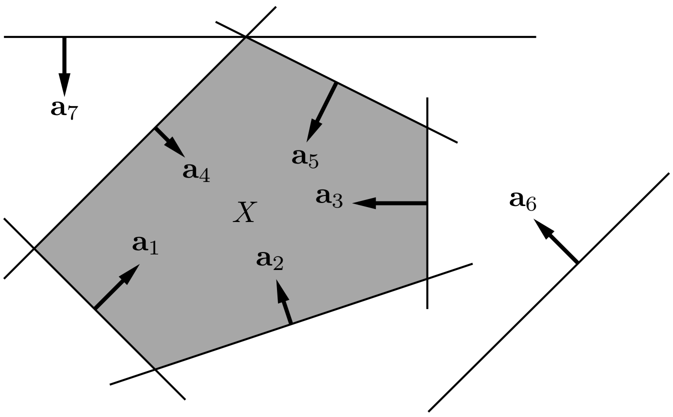
Hranice (viz §9.1) poloprostoru je nadrovina (13.3). Obrázek ilustruje tyto pojmy pro $n = 2$:



Konvexní mnohostěn (krátce jen **mnohostěn**, angl. *polyhedron*) je průnik konečně mnoha uzavřených poloprostorů. Je to tedy množina

$$X = \{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}_i^T\mathbf{x} \geq b_i \;\; \forall i = 1, \ldots, m \,\} = \{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} \geq \mathbf{b} \,\}, \tag{13.5}$$

kde $\mathbf{a}_1^T, \ldots, \mathbf{a}_m^T \in \mathbb{R}^n$ jsou řádky matice $\mathbf{A} \in \mathbb{R}^{m \times n}$ a $b_1, \ldots, b_m \in \mathbb{R}$ jsou složky vektoru $\mathbf{b} \in \mathbb{R}^m$. Obrázek ukazuje příklad pro $n = 2$ a $m = 7$ (všimněte si, že poloprostory 6 a 7 jsou nadbytečné, jejich vynecháním se mnohostěn nezmění):



Poloprostor je očividně konvexní množina, proto dle Tvrzení 13.1 je konvexní mnohostěn konvexní množina. Všimněte si, že konvexní mnohostěn nemusí být omezený.

Dimenze mnohostěnu je dimenze jeho afinního obalu, $\dim X = \dim \operatorname{aff} X$ (afinní obal viz §13.2, dimenze afinního podprostoru viz §3.3). Jinými slovy, je to největší počet afinně nezávislých bodů, které lze do mnohostěnu umístit, minus jedna.

[2] Otevřený poloprostor je množina $\{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}^T\mathbf{x} > b \,\}$.

Příklad 13.2. Množina X z Příkladu 12.1 je konvexní mnohostěn. ♦

Příklad 13.3. Příklady 'jednoduchých' konvexních mnohostěnů v $\mathbb{R}^n$:

1. prázdná množina $\emptyset$
2. celý prostor $\mathbb{R}^n$
3. intervaly typu $(-\infty, a]$, $[a, \infty)$ (tj. poloprostory v $\mathbb{R}$) a jejich průniky $[a, b]$
4. každý afinní podprostor (např. bod, přímka, rovina, nadrovina)
5. polopřímka $\{\, \mathbf{x} + \alpha \mathbf{v} \mid \alpha \geq 0 \,\}$
6. poloprostor
7. množina $\mathbb{R}^n_+$ (kde $\mathbb{R}_+$ značí množinu nezáporných reálných čísel)
8. panel $\{\, \mathbf{x} \in \mathbb{R}^n \mid b_1 \leq \mathbf{a}^T \mathbf{x} \leq b_2 \,\}$ (průnik dvou poloprostorů s opačnými normálami)
9. hyperkrychle $[-1, 1]^n = \{\, \mathbf{x} \in \mathbb{R}^n \mid -1 \leq x_i \leq 1 \ \forall i = 1, \ldots, n \,\} = \{\, \mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\|_\infty \leq 1 \,\}$
10. hyperkvádr ('box') $[a_1, b_1] \times \cdots \times [a_n, b_n] = \{\, \mathbf{x} \in \mathbb{R}^n \mid a_i \leq x_i \leq b_i \ \forall i = 1, \ldots, n \,\}$
11. simplex, což je n-rozměrný mnohostěn který je konvexním obalem svých $n + 1$ vrcholů
12. standardní simplex $\{\, \mathbf{x} \in \mathbb{R}^n \mid x_i \geq 0, \ \sum_{i=1}^n x_i = 1 \,\}$ (tedy $(n - 1)$-rozměrný konvexní mnohostěn, jehož vrcholy jsou vektory $\mathbf{e}_1, \ldots, \mathbf{e}_n$ standardní báze $\mathbb{R}^n$)
13. křížový polytop $\{\, \mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i| \leq 1 \,\}$ (pro $n = 3$ pravidelný osmistěn) ♦

Příklad 13.4. Příklady konvexních množin, které nejsou mnohostěny:

- Koule v $\mathbb{R}^n$ pro $n \geq 2$ je průnikem nekonečně mnoha poloprostorů $\{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}^T \mathbf{x} \leq 1 \,\}$ pro všechna $\mathbf{a} \in \mathbb{R}^n$ splňující $\|\mathbf{a}\|_2 = 1$ (nakreslete a promyslete!). Není to mnohostěn, protože počet poloprostorů není konečný.
- otevřený poloprostor $\{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}^T \mathbf{x} > b \,\}$
- interval $[0, 1)$ ♦

13.3.1 Extremální body

Bod $\mathbf{x} \in X$ se nazývá **extremální bod** konvexní množiny X, jestliže neexistují dva různé body z X takové, že $\mathbf{x}$ je střed úsečky spojující tyto dva body, tj. jestliže platí implikace

$$\mathbf{x}_1, \mathbf{x}_2 \in X, \ \ \mathbf{x} = \tfrac{1}{2}(\mathbf{x}_1 + \mathbf{x}_2) \quad \Longrightarrow \quad \mathbf{x}_1 = \mathbf{x}_2. \tag{13.6}$$

Pro mnohostěn (13.5) a pro množinu $I \subseteq \{1, \ldots, m\}$ budeme symbolem $\mathbf{A}_I$ označovat[3] matici tvořenou řádky matice $\mathbf{A}$ s indexy I a $\mathbf{b}_I$ bude označovat vektor tvořený složkami vektoru $\mathbf{b}$ s indexy I. Tedy $\mathbf{A}_I \mathbf{x} = \mathbf{b}_I$ je jen jiné označení soustavy $\mathbf{a}_i^T \mathbf{x} = b_i \ \forall i \in I$ (pro $I = \emptyset$ definujeme soustavu jako prázdnou a její řešení je libovolné $\mathbf{x} \in \mathbb{R}^n$).

[3] Zde je drobná formální nesrovnalost: I je množina, a tedy na pořadí jejích prvků nezáleží, ale na pořadí řádků matice $\mathbf{A}_I$ záleží. Abychom byli přesní, musíme proto předpokládat, že I je *uspořádaná množina*, kde její uspořádání je zděděné od množiny $\{1, \ldots, m\}$, a řádky matice $\mathbf{A}_I$ jsou seřazené tímto uspořádáním.

Věta 13.2. Pro každý bod $\mathbf{x}$ mnohostěnu (13.5) jsou následující výroky ekvivalentní:

- Bod $\mathbf{x}$ je extremální bod mnohostěnu.
- Existuje neprázdná množina $I \subseteq \{1, \ldots, m\}$ tak, že $\mathbf{A}_I \mathbf{x} = \mathbf{b}_I$ a sloupce matice $\mathbf{A}_I$ jsou lineárně nezávislé.

Důkaz. Nechť X označuje mnohostěn (13.5) a $\mathbf{x} \in X$ je jeho extremální bod. Označme jako $I = \{\, i \mid \mathbf{a}_i^T \mathbf{x} = b_i \,\}$ množinu indexů nerovností v definici (13.5), které jsou v bodě $\mathbf{x}$ aktivní. Dokážeme, že sloupce $\mathbf{A}_I$ jsou lineárně nezávislé. Kdyby totiž nebyly, měla by matice $\mathbf{A}_I$ netriviální nulový prostor (dle Věty 3.9), tedy $\mathbf{A}_I \mathbf{v} = \mathbf{0}$ pro nějaké $\mathbf{v} \neq \mathbf{0}$. Pro každé $i \notin I$ je $\mathbf{a}_i^T \mathbf{x} > b_i$, tedy $\mathbf{a}_i^T \mathbf{x} \pm \varepsilon \geq b_i$ pro nějaké $\varepsilon > 0$. Tedy by bylo $\mathbf{A}(\mathbf{x} \pm \varepsilon \mathbf{v}) \geq \mathbf{b}$ neboli $\mathbf{x} \pm \varepsilon \mathbf{v} \in X$. Dle (13.6) by proto $\mathbf{x}$ nebyl extremální bod.

Nechť I je taková, že $\mathbf{A}_I \mathbf{x} = \mathbf{b}_I$ a sloupce $\mathbf{A}_I$ jsou lineárně nezávislé. Nechť $\mathbf{x} = \frac{1}{2}(\mathbf{x}_1 + \mathbf{x}_2)$ pro nějaké $\mathbf{x}_1, \mathbf{x}_2 \in X$, tedy

$$\begin{aligned} \mathbf{A}_I \mathbf{x}_1 &\geq \mathbf{b}_I, \\ \mathbf{A}_I \mathbf{x}_2 &\geq \mathbf{b}_I, \\ \mathbf{A}_I \mathbf{x} = \tfrac{1}{2}(\mathbf{A}_I \mathbf{x}_1 + \mathbf{A}_I \mathbf{x}_2) &= \mathbf{b}_I. \end{aligned}$$

Z toho plyne $\mathbf{A}_I \mathbf{x}_1 \geq \frac{1}{2}(\mathbf{A}_I \mathbf{x}_1 + \mathbf{A}_I \mathbf{x}_2) \leq \mathbf{A}_I \mathbf{x}_2$. Ale pro libovolná čísla $\alpha_1, \alpha_2 \in \mathbb{R}$ platí implikace

$$\alpha_1 \geq \tfrac{1}{2}(\alpha_1 + \alpha_2) \leq \alpha_2 \quad \Longrightarrow \quad \alpha_1 = \alpha_2. \tag{13.7}$$

Tedy platí $\mathbf{A}_I \mathbf{x}_1 = \mathbf{A}_I \mathbf{x}_2$. Protože však $\mathbf{A}_I$ má lineárně nezávislé sloupce, plyne z toho $\mathbf{x}_1 = \mathbf{x}_2$ (dle Tvrzení 3.1). Tedy $\mathbf{x}$ je extremální bod. ■

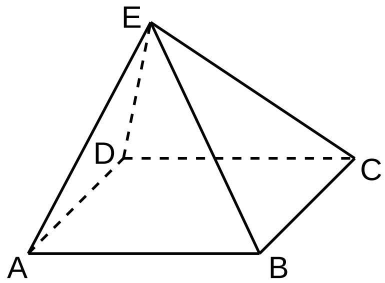
Připomeňme, že soustava lineárních rovnic má právě jedno řešení právě tehdy, když má aspoň jedno řešení a její matice má lineárně nezávislé sloupce. Věta tedy vlastně říká, že $\mathbf{x} \in \mathbb{R}^n$ je extremální bod mnohostěnu, právě když soustava $\mathbf{A}_I \mathbf{x} = \mathbf{b}_I$ má právě jedno řešení $\mathbf{x}$ a toto řešení navíc splňuje $\mathbf{A}\mathbf{x} \geq \mathbf{b}$. Geometricky to znamená, že hranice poloprostorů s indexy I se protínají v právě jednom bodě a ten bod leží v mnohostěnu. To nám dává návod, jak vypsat všechny extremální body mnohostěnu. Procházíme všechny podmnožiny $I \subseteq \{1, \ldots, m\}$. Jestliže soustava $\mathbf{A}_I \mathbf{x} = \mathbf{b}_I$ má právě jedno řešení $\mathbf{x}$ a to navíc splňuje $\mathbf{A}\mathbf{x} \geq \mathbf{b}$, pak $\mathbf{x}$ je extremální bod.

Příklad 13.5. Mějme mnohostěn $\{\, (x, y) \in \mathbb{R}^2 \mid x \geq 0,\ y \geq 0,\ x + y \geq 1 \,\}$ (nakreslete si ho!). Tedy v (13.5) máme $n = 2$, $m = 3$ a

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Zkoumejme podmnožiny $I \subseteq \{1, 2, 3\}$. Pro všechny jednoprvkové podmnožiny ($|I| = 1$) jistě má matice $\mathbf{A}_I$ lineárně závislé sloupce. Pro všechny dvouprvkové podmnožiny ($\{1, 2\}$, $\{1, 3\}$ a $\{2, 3\}$) má matice $\mathbf{A}_I$ lineárně nezávislé sloupce. Např. pro $I = \{1, 2\}$ je řešením soustavy $\mathbf{A}_I \mathbf{x} = \mathbf{b}_I$ bod $\mathbf{x} = (x, y) = (0, 0)$, ten ale nesplňuje $\mathbf{A}\mathbf{x} \geq \mathbf{b}$ (tj. není v mnohostěnu), proto to není extremální bod. Pro $I = \{1, 3\}$ má soustava jediné řešení $(x, y) = (0, 1)$, které patří do mnohostěnu a tudíž je to jeho extremální bod. Pro $I = \{1, 2, 3\}$ nemá soustava $\mathbf{A}_I \mathbf{x} = \mathbf{b}_I$ žádné řešení. ♦

Příklad 13.6. Nechť mnohostěn (13.5) je pyramida v $\mathbb{R}^3$ na obrázku:

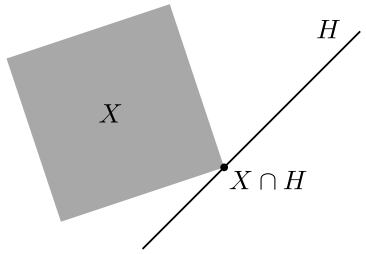


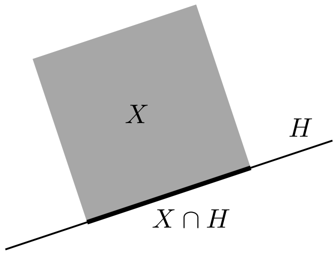
Tento mnohostěn je průnikem pěti poloprostorů, tedy $n = 3$ a $m = 5$. Nechť omezení $\mathbf{a}_i^T\mathbf{x} \geq b_i$ pro $i = 1, 2, 3, 4, 5$ je poloprostor, jehož hranicí je nadrovina určená po řadě body ABCD, ABE, BCE, CDE, ADE. Pro $I = \{1\}$ má soustava $\mathbf{A}_I\mathbf{x} = \mathbf{b}_I$ řešení ale sloupce matice $\mathbf{A}_I$ jsou lineárně závislé, tedy soustava má nekonečně mnoho řešení, a proto žádné její řešení není extremální bod. Pro $I = \{1, 2, 3, 4, 5\}$ nemá soustava žádné řešení. Pro $I = \{1, 2, 3\}$ má soustava řešení a navíc sloupce matice $\mathbf{A}_I$ jsou lineárně nezávislé, tedy soustava má právě jedno řešení řešení. Toto řešení navíc splňuje soustavu $\mathbf{Ax} \geq \mathbf{b}$, tedy je to extremální bod (vrchol B). ♦

Protože podmnožin I je konečně mnoho, má každý konvexní mnohostěn konečně mnoho extremálních bodů. Počet extremálních bodů mnohostěnu popsaného soustavou lineárních nerovnic ovšem může být exponenciální funkcí počtu nerovnic (tedy může být astronomicky velký pro nepříliš velký počet nerovnic). Tomu se není co divit, protože počet všech podmnožin množiny $\{1, \ldots, m\}$ je 2^m. Např. hyperkrychle $\{\, \mathbf{x} \in \mathbb{R}^n \mid -\mathbf{1} \leq \mathbf{x} \leq \mathbf{1} \,\}$ je popsaná $2n$ lineárními nerovnicemi ale má 2^n extremálních bodů. Není tomu tak vždy, např. simplex (viz Příklad 13.3) je popsaný $n+1$ lineárními nerovnicemi a má $n+1$ extremálních bodů.

13.3.2 Stěny mnohostěnu

Opěrná nadrovina konvexní množiny $X \subseteq \mathbb{R}^n$ je nadrovina $H = \{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}^T\mathbf{x} = b \,\}$ taková, že $X \cap H \neq \emptyset$ a pro všechna $\mathbf{x} \in X$ platí $\mathbf{a}^T\mathbf{x} \geq b$. Neboli je to nadrovina, která obsahuje množinu X v jednom ze svých dvou poloprostorů a má s množinou X neprázdný průnik. Obrázky ukazují dvě různé opěrné nadroviny mnohostěnu:





Je-li H opěrná rovina mnohostěnu X, pak množina $X \cap H$ se nazývá **stěna** mnohostěnu. Dle Tvrzení 13.1 je každá stěna konvexního mnohostěnu sama o sobě konvexní mnohostěn. Stěny některých dimenzí mají jméno:

- Stěna dimenze 0 se nazývá **vrchol** (na obrázku vlevo).
- Stěna dimenze 1 se nazývá **hrana** (na obrázku vpravo).
- Stěna dimenze $\dim X - 1$ se nazývá **faseta** (pro množinu X dimenze 2 na obrázku je tedy faseta totéž co hrana).

Lze dokázat (důkaz neuvádíme), že bod mnohostěnu je vrchol, právě když je to extremální bod. V následujícím textu budeme používat pojem extremální bod (se kterým se pracuje lépe než s ekvivalentním pojmem vrchol).

Každým hraničním bodem (viz §9.1) konvexního mnohostěnu prochází alespoň jedna opěrná rovina. To snadno dokážeme: Mnohostěn je průnik konečně mnoha poloprostorů. Hranice každého z těchto poloprostorů je opěrná nadrovina mnohostěnu. Zároveň každý hraniční hod mnohostěnu leží na hranici aspoň jednoho poloprostoru.

13.3.3 Extrémy lineární funkce na mnohostěnu

Lemma 13.3. Je-li H opěrná nadrovina konvexní množiny X, pak každý extremální bod množiny $X \cap H$ je extremální bod množiny X.

Důkaz. Nechť $H = \{\, \mathbf{x} \mid \mathbf{a}^T\mathbf{x} = b \,\}$ a $\mathbf{a}^T\mathbf{x} \geq b$ pro všechna $\mathbf{x} \in X$. Nechť $\mathbf{x} \in X \cap H$ není extremální bod mnohostěnu X, takže $\mathbf{x} = \frac{1}{2}(\mathbf{x}_1 + \mathbf{x}_2)$ pro nějaké $\mathbf{x}_1, \mathbf{x}_2 \in X$ kde $\mathbf{x}_1 \neq \mathbf{x}_2$. Platí

$$\begin{aligned} \mathbf{a}^T\mathbf{x}_1 &\geq b, \\ \mathbf{a}^T\mathbf{x}_2 &\geq b, \\ \mathbf{a}^T\mathbf{x} = \tfrac{1}{2}(\mathbf{a}^T\mathbf{x}_1 + \mathbf{a}^T\mathbf{x}_2) &= b. \end{aligned}$$

Díky (13.7) z toho plyne $\mathbf{a}^T\mathbf{x}_1 = \mathbf{a}^T\mathbf{x}_2 = b$, tedy $\mathbf{x}_1, \mathbf{x}_2 \in H$. Tedy $\mathbf{x}$ není extremální bod mnohostěnu $X \cap H$. ■

Věta tedy říká, že extremální bod stěny mnohostěnu je zároveň extremální bod mnohostěnu.

Ne každý mnohostěn má extremální bod, např. poloprostor nebo afinní podprostor (což je konvexní mnohostěn) nemá žádný. Pro následující větu připomeňme, že *přímka* je afinní podprostor dimenze 1, tedy množina $\{\, \mathbf{x} + \alpha\mathbf{v} \mid \alpha \in \mathbb{R} \,\}$ pro nějaké $\mathbf{x} \in \mathbb{R}^n$ a $\mathbf{0} \neq \mathbf{v} \in \mathbb{R}^n$.

Věta 13.4. Pro každý neprázdný konvexní mnohostěn jsou následující výroky ekvivalentní:

- Mnohostěn má aspoň jeden extremální bod.
- Mnohostěn neobsahuje přímku.

Důkaz. Nejdříve dokážeme, že když mnohostěn X obsahuje přímku, tak nemá ani jeden extremální bod. Víme, že existují $\mathbf{x} \in X$ a $\mathbf{v} \neq \mathbf{0}$ takové, že $\mathbf{x} + \alpha\mathbf{v} \in X$ pro všechna $\alpha \in \mathbb{R}$. Je-li X ve tvaru (13.5), platí $\mathbf{A}\mathbf{x} \geq \mathbf{b}$ a tedy $\mathbf{A}(\mathbf{x}+\alpha\mathbf{v}) \geq \mathbf{b}$ pro všechna $\alpha \in \mathbb{R}$. Z toho plyne $\alpha\mathbf{A}\mathbf{v} \geq \mathbf{0}$ pro $\alpha \in \mathbb{R}$, z čehož plyne $\mathbf{A}\mathbf{v} = \mathbf{0}$. To ale znamená, že pro *každé* $\mathbf{x} \in X$ a každé $\alpha \in \mathbb{R}$ platí $\mathbf{A}(\mathbf{x} + \alpha\mathbf{v}) \geq \mathbf{b}$ neboli $\mathbf{x} + \alpha\mathbf{v} \in X$. Tedy *každým* bodem mnohostěnu prochází přímka, která leží v mnohostěnu. Proto žádný bod mnohostěnu nemůže být extremální bod.

Nyní dokážeme, že jestliže mnohostěn neobsahuje přímku, pak má aspoň jeden extremální bod. Dokážeme to indukcí podle dimenze mnohostěnu. Tvrzení triviálně platí pro mnohostěny dimenze 0, což jsou pouhé body. Předpokládejme, že tvrzení platí pro všechny mnohostěny dimenze menší než n. Pokud neprázdný mnohostěn X dimenze n neobsahuje přímku, pak každá přímka protínající X narazí na hranici X. Jestliže vezmeme přímku, která leží v afinním obalu X, pak v obdrženém hraničním bodě mnohostěnu X existuje opěrná nadrovina H, která mnohostěn neobsahuje (tj. $X \not\subseteq H$). Mnohostěn $X \cap H$ má proto dimenzi menší než n a neobsahuje přímku (protože X ji neobsahuje), tedy podle indukčního předpokladu má aspoň jeden extremální bod. Ten je dle Lemmatu 13.3 extremální bod mnohostěnu X. ■

Věta 13.5. Mějme konvexní mnohostěn, který neobsahuje přímku. Jestliže lineární funkce má na tomto mnohostěnu minimum, pak tato funkce nabývá na mnohostěnu minima aspoň v jednom z jeho extremálních bodů.

Důkaz. Nechť $\mathbf{x}^* \in X$ je bod mnohostěnu X, ve kterém lineární funkce $f(\mathbf{x}) = \mathbf{c}^T\mathbf{x}$ nabývá minima, tedy $\mathbf{c}^T\mathbf{x}^* \leq \mathbf{c}^T\mathbf{x}$ pro všechna $\mathbf{x} \in X$. To znamená, že $H = \{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{c}^T\mathbf{x} = \mathbf{c}^T\mathbf{x}^* \,\}$ je opěrná rovina mnohostěnu X v bodě $\mathbf{x}^*$ (H je vlastně vrstevnice funkce f výšky $\mathbf{c}^T\mathbf{x}^*$). Tedy f nabývá na X minima ve všech bodech množiny $X \cap H$ (což je stěna mnohostěnu X). Mnohostěn $X \cap H$ neobsahuje přímku (protože X ji neobsahuje), tedy dle Věty 13.4 má aspoň jeden extremální bod. Dle Lemmatu 13.3 je tento bod extremální bod mnohostěnu X. ∎

Hledejme nyní minimum lineární funkce na mnohostěnu, neboli řešme lineární program

$$\min\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{A}\mathbf{x} \geq \mathbf{b} \,\}.$$

Jestliže mnohostěn neobsahuje přímku a funkce na něm má minimum, pak dle Věty 13.5 stačí jednoduše projít všechny extremální body mnohostěnu a vybrat z nich ten, ve kterém má funkce nejmenší hodnotu. Tento algoritmus má několik zádrhelů. Nevíme, co dělat s mnohostěny, které obsahují přímku. Nevíme, jak poznat, že účelová funkce má na mnohostěnu minimum. Zdaleka největší zádrhel je ovšem to, že prakticky není možné projít všechny extremální body, neboť jich je příliš mnoho. V příští kapitole popíšeme *simplexový algoritmus*, který téměř všechny tyto potíže řeší.

.

13.4 Cvičení

13.1. Odpovězte, zda následující množiny jsou konvexní a odpověď dokažte z definice konvexní množiny:

a) interval $[a, b]$
b) $\{\, (x, y) \in \mathbb{R}^2 \mid y = x^2 \,\}$
c) $\{\, (x, y) \in \mathbb{R}^2 \mid y \geq x^2 \,\}$
d) $\{\, (x, y) \in \mathbb{R}^2 \mid y^2 \geq x^2 \,\}$
e) $\{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} \leq \mathbf{b},\ \mathbf{C}\mathbf{x} = \mathbf{d} \,\}$
f) $\{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{x}^T\mathbf{x} \leq 1 \,\}$
g) $\mathbb{Z}$ (množina celých čísel)
h) $\{\, \mathbf{x} \in \mathbb{R}^n \mid \max\{x_1, , \ldots, , x_n\} \geq 0 \,\}$
i) $\{\, \mathbf{C}\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{A}\mathbf{x} \geq \mathbf{b} \,\}$ (lineární zobrazení konvexního mnohostěnu)

13.2. Jsou následující množiny konvexní? Odpověď nemusíte dokazovat z definice kovexní množiny, stačí uvést přesvědčivý argument. Jestliže množina není konvexní, napište její konvexní obal jako množinu řešení soustavy (co možná nejjednodušší) rovnic a nerovnic.

a) $\{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} \geq \mathbf{0},\ \sum_{i=1}^n x_i = 1 \,\}$
b) $\{\, \mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\|_2 = 1 \,\}$
c) $\{\, \mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\|_2 < 1 \,\}$
d) $\{\, (x, y) \in \mathbb{R}^2 \mid y = x^2 \,\}$

e) $\{ (x,y) \in \mathbb{R}^2 \mid x \geq 0,\ y \geq 0,\ xy = 1 \}$
f) $\{ (x,y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 2 \} \cap \{ (x,y) \in \mathbb{R}^2 \mid (x-1)^2 + y^2 \leq 2 \}$
g) $\{ (x,y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1,\ x \geq 0,\ y \geq 0 \}$
h) $\{-1, 0, 1\}$
i) $\{(1,1), (2,2)\}$
j) $\{(1,1), (1,2), (3,1)\}$

13.3. ($\star$) Zobrazení $f\colon 2^V \to 2^V$ (kde V je libovolná množina a 2^V značí množinu všech podmnožin množiny V) se nazývá *uzávěr* (angl. *closure*) na množině V, jestliže pro každé $X, Y \subseteq V$ platí

- $X \subseteq f(X)$ (extensivita),
- $f(f(X)) = f(X)$ (idempotence),
- $X \subseteq Y \ \Rightarrow\ f(X) \subseteq f(Y)$ (monotonicita).

Dokažte, že operace lineárního, afinního, nezáporného a konvexního obalu jsou uzávěry. Tedy dokažte tři vlastnosti výše pro $V = \mathbb{R}^n$ a např. $f(X) = \operatorname{conv} X$.

13.4. ($\star$) Navrhněte algoritmus na nalezení konvexního obalu konečné množiny bodů v rovině. Výstupem bude uspořádný seznam bodů, které tvoří extremální body konvexního obalu.

13.5. Nakreslete lineární, afinní, nezáporný a konvexní obal náhodně zvolených k vektorů v $\mathbb{R}^n$ pro všech devět případů $k, n \in \{1, 2, 3\}$.

13.6. Bude Tvrzení 13.1 platit, pokud v ní výraz 'konvexní množina' nahradíme výrazem 'lineární podprostor' (příp. 'afinní podprostor', 'konvexní kužel')? Odpověď dokažte.

13.7. Jsou následující množiny konvexní mnohostěny? Zápornou odpověď odůvodněte. Kladnou odpověď dokažte tak, že množinu napíšete jako množinu řešení soustavy konečně mnoha lineárních nerovnic (tj. jako průnik konečně mnoha poloprostorů).

a) $\{ \mathbf{x} \in \mathbb{R}^n \mid \mathbf{x}^T \mathbf{C} \mathbf{x} \leq 1 \}$, kde $\mathbf{C}$ je positivně definitní
b) $\{ \alpha \mathbf{v} \mid \alpha \in \mathbb{R} \}$, kde $\mathbf{v} \in \mathbb{R}^n$
c) ($\star$) $\{ \mathbf{C}\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ \|\mathbf{x}\|_2 \leq 1 \}$, kde $\mathbf{C} \in \mathbb{R}^{m \times n}$
d) $\{ \mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x} - \mathbf{a}\|_2 \leq \|\mathbf{x} - \mathbf{b}\|_2 \}$, kde $\mathbf{a}, \mathbf{b}$ jsou dány

13.8. Mějme body $\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^n$. Pro každé $i = 1, \ldots, m$ definujeme množinu

$$X_i = \{ \mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x} - \mathbf{a}_i\|_2 \leq \|\mathbf{x} - \mathbf{a}_j\|_2 \ \forall j \neq i \}.$$

Ukažte, že množiny $X_1, \ldots, X_m$ jsou konvexní mnohostěny. Sjednocení hranic těchto množin se nazývá *Voronoiův diagram*. Nakreslete si ho pro $n = 2$ a pro tři případy $m \in \{2, 3, 4\}$ pro různé konfigurace bodů $\mathbf{a}_1, \ldots, \mathbf{a}_m$.

13.9. Hledáme největší (hyper)kouli $\{ \mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x} - \mathbf{c}\|_2 \leq r \}$, která se vejde do mnohostěnu $\{ \mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} \geq \mathbf{b} \}$. Zformulujte jako lineární program. Srov. Cvičení 12.18.

13.10. Chceme najít všechny extremální body mnohostěnu $\{ \mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} \geq \mathbf{b} \}$.

a) Udělejte pro mnohostěn

$$\mathbf{A} = \begin{bmatrix} -1 & -2 \\ 3 & -1 \\ -1 & 1 \\ 2 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -14 \\ 0 \\ -2 \\ 4 \end{bmatrix}.$$

b) Napište program v Matlabu, který vypíše všechny extremální body pro libovolné $\mathbf{A}, \mathbf{b}$.

13.11. Které z následujících mnohostěnů mají aspoň jeden extremální bod? Kladnou odpověď dokažte nalezením libovolného jednoho extremálního bodu. Zápornou odpověď dokažte nalezením libovolné přímky, která leží v mnohostěnu.

a) všechny mnohostěny z Příkladu 13.3

b) $\{ (x, y, z) \in \mathbb{R}^3 \mid x + y \leq 1,\ x \geq 0,\ y \geq 0 \}$

c) $\{ (x, y) \in \mathbb{R}^2 \mid -1 \leq x + y \leq 1 \}$

13.12. ($\star$) Pro mnohostěn $\{ \mathbf{x} \in \mathbb{R}^n \mid \mathbf{Ax} \geq \mathbf{b} \}$ dokažte:

a) Má-li $\mathbf{A}$ lineárně závislé sloupce (tj. $\operatorname{rank} \mathbf{A} < n$), mnohostěn nemá extremální bod.

b) Má-li mnohostěn aspoň jeden extremální bod, je $\operatorname{rank} \mathbf{A} = n$.

Nápověda a řešení

13.1.a) Konvexní, protože pro libovolné $\alpha \in [0, 1]$ je $(1 - \alpha)a + \alpha b \in [a, b]$.

13.1.b) Není konvexní. Např. $\mathbf{x} = (1, 1)$ a $\mathbf{y} = (-1, 1)$ patří do množiny, ale $\mathbf{x}/2 + \mathbf{y}/2 = (0, 1)$ nepatří.

13.1.c) Je konvexní.

13.1.e) Konvexní.

13.1.f) Konvexní.

13.1.g) Nekonvexní. Např. pro $x = 1$, $y = 2$, $\alpha = \frac{1}{2}$ číslo $(1 - \alpha)x + \alpha y = 1.5$ není celé.

13.1.h) Nekonvexní.

13.2.a) průnik poloprostorů a nadroviny, konvexní mnohostěn

13.2.b) Sféra, nekonvexní. Konv. obal je koule $\{ \mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\|_2 \leq 1 \}$.

13.2.c) koule bez hranice, konvexní

13.2.d) Graf paraboly, nekonvexní. Konvexní obal je $\{ (x, y) \in \mathbb{R}^2 \mid y \geq x^2 \}$.

13.2.e) Jedna větev grafu hyperboly, není konvexní. Konv. obal je $\{ (x, y) \in \mathbb{R}^2 \mid x \geq 0,\ y \geq 0,\ xy \geq 1 \}$.

13.2.f) průnik dvou koulí, konvexní

13.2.g) Čtvrt kružnice, nekonvexní. Konv. obal je $\{ (x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1,\ x + y \geq 1 \}$.

13.2.h) Tři body v $\mathbb{R}$, nekonvexní protože např. $(-1 + 0)/2$ nepatří do množiny. Konv. obal je interval $[-1, 1]$.

13.2.i) Dva body v $\mathbb{R}^2$, nekonvexní. Konv. obal je úsečka spojující ty dva body. Tuto úsečka máme napsat jako množinu řešení rovnic a nerovnic, což jde např. takto: $\{ (x, y) \in \mathbb{R}^2 \mid 1 \leq x \leq 2,\ x = y \}$.

13.2.j) Tři body v $\mathbb{R}^2$, nekonvexní. Konv. obal je trojúhelník $\{ (x, y) \in \mathbb{R}^2 \mid x \geq 1,\ y \geq 1,\ x + 2y \leq 5 \}$.

13.7.a) Je to elipsoid, tedy je to konvexní množina. Mnohostěn je to jen pro $n = 1$ (v tom případě je to úsečka).

13.7.b) Přímka jdoucí počátkem (tj. lineární podprostor dimenze 1) se směrovým vektorem $\mathbf{v}$. V uvedeném tvaru není vyjádřena jako průnik poloprostorů. Je-li ale $\mathbf{A} \in \mathbb{R}^{(n-1)\times n}$ matice taková, že $\{ \alpha\mathbf{v} \mid \alpha \in \mathbb{R} \} = \operatorname{span}\{\mathbf{v}\} = \operatorname{null} \mathbf{A}$, přímku jsme vyjádřili jako množinou řešení soustavy $\mathbf{Ax} = \mathbf{0}$, což už je průnik poloprostorů.

13.9. Maximalizujeme r za podmínek $\mathbf{Ac} - \mathbf{b} \geq r\mathbf{1}$ (tedy neznámé jsou $\mathbf{c} \in \mathbb{R}^n$ a $r \geq 0$), přičemž předpokládáme, že v každé nerovnici $\mathbf{a}_i^T\mathbf{x} \geq b_i$ soustavy $\mathbf{Ax} \geq \mathbf{b}$ máme $\|\mathbf{a}_i\| = 1$ (tedy každou nerovnici musíme nejdříve vydělit číslem $\|\mathbf{a}_i\|$).

13.10.a) Extremální body a odpovídající množiny $I \subseteq \{1, \dots, 4\}$:

I	$\{1,2\}$	$\{1,3\}$	$\{2,4\}$	$\{3,4\}$
$\mathbf{x}$	$(2,6)$	$(6,4)$	$(\frac{4}{5}, \frac{12}{5})$	$(2,0)$

13.11.b) nemá extremální bod

13.11.c) nemá extremální bod

13.12.a) Je-li rank $\mathbf{A} < n$, je $\mathbf{Av} = \mathbf{0}$ pro nějaké $\mathbf{v} \neq \mathbf{0}$. Tedy pro každé $\mathbf{x}$ splňující $\mathbf{Ax} \geq \mathbf{b}$, mnohostěn obsahuje přímku $\{\, \mathbf{x} + \alpha\mathbf{v} \mid \alpha \in \mathbb{R} \,\}$. Tedy každým bodem mnohostěnu prochází přímka, která celá leží v mnohostěnu. Tedy žádný bod mnohostěnu není extremální bod. Srov. Věta 13.4.

Kapitola 14

Simplexová metoda

Nyní popíšeme slavný algoritmus na řešení lineárních programů, **simplexovou metodu**. V minulé kapitole jsme ukázali, že má-li mnohostěn aspoň jeden extremální bod, lineární funkce na něm nabývá minima aspoň v jednom extremálním bodě. Tedy můžeme projít všechny extremální body a najít ten s nejlepší účelovou funkcí. Ale extremálních bodů může být mnoho. Simplexová metoda tento naivní algoritmus výrazně zrychluje tak, že postupuje od jednoho extremálního bodu k dalšímu po hranách mnohostěnu tak, aby se účelová funkce zlepšovala nebo aspoň nezhoršovala. Ve zbytku kapitoly tento geometricky popsaný nápad převedeme do algebry.

Pro začátek zapomeňme na účelovou funkci a zkoumejme pouze mnohostěn přípustných řešení LP. Místo mnohostěnu v obecném tvaru (13.5) algoritmus pracuje s mnohostěnem v rovnicovém tvaru (12.3)

$$X = \{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} = \mathbf{b},\ \mathbf{x} \geq \mathbf{0} \,\}, \tag{14.1}$$

kde navíc matice $\mathbf{A} \in \mathbb{R}^{m \times n}$ má lineárně nezávislé řádky (tj. $\operatorname{rank} \mathbf{A} = m$). Množina řešení soustavy $\mathbf{A}\mathbf{x} = \mathbf{b}$ je tedy afinní podprostor dimenze $n - m$.

Zkoumejme soustavu $\mathbf{A}\mathbf{x} = \mathbf{b}$ s dodatečnou podmínkou, že některé proměnné nastavíme na nulu. Tedy zvolíme množinu indexů $J \subseteq \{1, 2, \ldots, n\}$ a řešíme soustavu

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \tag{14.2a}$$

$$x_j = 0 \quad \forall j \in \{1, \ldots, n\} \setminus J. \tag{14.2b}$$

Kdy má tato soustava právě jedno řešení? Dle Tvrzení 3.1 právě tehdy, když sloupce matice $\mathbf{A}$ s indexy J jsou lineárně nezávislé a $\mathbf{b}$ patří do jejich lineárního obalu. Aby sloupce J byly lineárně nezávislé, nutně musí být $|J| \leq m$. Když ovšem $|J| < m$, pak existuje množina $J' \supseteq J$ tak, že $|J'| = m$ a sloupce J' jsou lineárně nezávislé (tedy doplníme sloupce J na bázi prostoru $\operatorname{rng} \mathbf{A}$, což lze díky Větě 3.2). Pak ale, dle Tvrzení 3.1, bude řešení soustavy splňovat $x_j = 0$ pro všechna $j \in J' \setminus J$. Tedy soustava bude mít stejné řešení pro J i J'. Proto, chceme-li najít všechny body $\mathbf{x}$ která jsou jediným řešením soustavy (14.2), stačí soustavu vyřešit jen pro množiny J velikosti $|J| = m$, pro které jsou sloupce J matice $\mathbf{A}$ lineárně nezávislé.

Příklad 14.1. Nechť je soustava $\mathbf{A}\mathbf{x} = \mathbf{b}$ dána tabulkou (blokovou maticí)

$$\begin{bmatrix} \mathbf{A} & \mathbf{b} \end{bmatrix} = \left[\begin{array}{rrrrrr|r} -1 & 1 & 3 & 1 & 0 & 2 & 1 \\ 1 & 0 & 4 & 0 & 1 & 4 & 4 \\ -1 & 0 & 4 & 1 & 1 & 4 & 2 \end{array}\right]. \tag{14.3}$$

Sloupce $J=\{1,4,5\}$ matice $\mathbf{A}$ jsou lineárně nezávislé (ověřte!). Soustava (14.2) vypadá takto

$$\begin{bmatrix} -1 & 1 & 0 \\ 1 & 0 & 1 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 2 \end{bmatrix}, \quad x_2 = x_3 = x_6 = 0 \tag{14.4}$$

a má jediné řešení $\mathbf{x} = (3,0,0,4,1,0)$. Sloupce J tvoří regulární matici velikosti $m \times m$.

Sloupce $J=\{3,4\}$ jsou lineárně nezávislé a soustava

$$\begin{bmatrix} 3 & 1 \\ 4 & 0 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 2 \end{bmatrix}, \quad x_1 = x_2 = x_5 = x_6 = 0$$

má jediné řešení $\mathbf{x} = (0,0,1,-2,0,0)$. Sloupce $J=\{3,4,5\}$ jsou lineárně nezávislé a soustava

$$\begin{bmatrix} 3 & 1 & 0 \\ 4 & 0 & 1 \\ 4 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 2 \end{bmatrix}, \quad x_1 = x_2 = x_6 = 0$$

má jediné řešení $\mathbf{x} = (0,0,1,-2,0,0)$, stejné jako pro $J=\{3,4\}$. Sloupce $J=\{3,4,6\}$ jsou také lineárně nezávislé a soustava má tedy stejné řešení. ♦

Uvedená úvaha motivuje zavedení následujících pojmů. Množina $J \subseteq \{1,\ldots,n\}$ se nazývá **báze**, pokud $|J| = m$ a sloupce J matice $\mathbf{A}$ jsou lineárně nezávislé. Proměnným x_j pro $j \in J$ říkáme **bázové proměnné**. Bod $\mathbf{x}$ je **bázové řešení** příslušné bázi J, jestliže splňuje (14.2). Bázové řešení $\mathbf{x}$ je **přípustné** (tedy leží v mnohostěnu), pokud $\mathbf{x} \geq \mathbf{0}$. Bázové řešení $\mathbf{x}$ je **degenerované**, pokud má více než $m-n$ nulových složek. Protože matice $\mathbf{A}$ má hodnost m, existuje aspoň jedna báze. Každé bázi přísluší právě jedno bázové řešení. Je-li ovšem bázové řešení degenerované, přísluší více než jedné bázi (v Příkladu 14.1 bázové řešení $\mathbf{x} = (0,0,1,-2,0,0)$ přísluší bázím $\{3,4,5\}$ a $\{3,4,6\}$).

Mnohostěn (14.1) neobsahuje přímku, protože už jeho nadmnožina $\{\,\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} \geq \mathbf{0}\,\}$ neobsahuje přímku (proč?). Tedy jestliže je mnohostěn (14.1) neprázdný, má dle Věty 13.4 aspoň jeden extremální bod. Věta 13.2, vyslovená pro mnohostěn v obecném tvaru (13.5), jde vyslovit i pro mnohostěn v rovnicovém tvaru (14.1):

Důsledek 14.1. Bod mnohostěnu (14.1) je extremální, právě když je to přípustné bázové řešení.

Důkaz. Mnohostěn (14.1) lze napsat ve tvaru (13.5) jako $X = \{\,\mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}'\mathbf{x} \geq \mathbf{b}'\,\}$, kde

$$\mathbf{A}' = \begin{bmatrix} \mathbf{A} \\ -\mathbf{A} \\ \mathbf{I}_n \end{bmatrix} \in \mathbb{R}^{(2m+n)\times n}, \qquad \mathbf{b}' = \begin{bmatrix} \mathbf{b} \\ -\mathbf{b} \\ \mathbf{0} \end{bmatrix} \in \mathbb{R}^{2m+n}.$$

Máme dokázat, že pro každé $\mathbf{x} \in \mathbb{R}^n$ jsou následující dva výroky ekvivalentní:

1. $\mathbf{A}'\mathbf{x} \geq \mathbf{b}'$ a existuje množina $I \subseteq \{1,\ldots,2m+n\}$ tak, že $\mathbf{x}$ je jediným řešením soustavy $\mathbf{A}'_I\mathbf{x} = \mathbf{b}'_I$.
2. $\mathbf{x} \geq \mathbf{0}$ a existuje množina $J \subseteq \{1,\ldots,n\}$ tak, že $\mathbf{x}$ je jediným řešením soustavy (14.2).

Implikaci $1 \Leftarrow 2$ dokážeme položením $I = \{1, \ldots, 2m\} \cup \{\, 2m+j \mid j \in \{1, \ldots, n\} \setminus J \,\}$. Implikaci $1 \Rightarrow 2$ dokážeme položením $J = \{\, j \in \{1, \ldots, n\} \mid 2m + j \notin I \,\}$. Rozmyslete, že to tak je! ■

Dvě báze nazveme **sousední**, pokud mají $m - 1$ společných prvků (např. báze $\{3, 4, 5\}$ a $\{3, 4, 6\}$ v Příkladu 14.1). Lze ukázat (důkaz vynecháme), že dvojice sousedních bází odpovídají buď jedinému (degenerovanému) extremálnímu bodu nebo dvojici extremálních bodů spojených hranou. Simplexová metoda přechází mezi sousedními bázemi tak, že bázová řešení jsou stále přípustná a účelová funkce se zlepšuje nebo aspoň nezhoršuje.

V §14.1 popíšeme stavební kameny metody, které pak v §14.2 spojíme do celého algoritmu.

14.1 Stavební kameny algoritmu

14.1.1 Přechod k sousední standardní bázi

Simplexová metoda pracuje pouze se *standardními* bázemi, tj. sloupce J jsou vektory standardní báze. To má výhodu v tom, že (i) nemusíme kontrolovat, zda jsou sloupce J lineárně nezávislé a (ii) nenulové složky bázového řešení $\mathbf{x}$ jsou rovny přímo složkám vektoru $\mathbf{b}$. Na počátku algoritmu se předpokládá, že matice $\mathbf{A}$ obsahuje standardní bázi.

Z lineární algebry znáte *elementární řádkové úpravy* soustavy $\mathbf{Ax} = \mathbf{b}$, reprezentované tabulkou $\begin{bmatrix}\mathbf{A} & \mathbf{b}\end{bmatrix}$ (viz §2.5). Ukážeme, jak přejít od aktuální standardní báze J k sousední standardní bázi, tedy nějaký sloupec $j' \in J$ bázi opustí a nějaký sloupec $j \notin J$ do báze vstoupí. Nechť i je řádek, ve kterém je $a_{ij'} = 1$. Prvek (i, j) matice se nazývá **pivot** (angl. znamená *čep*). Nechť $a_{ij} \neq 0$. Chceme nastavit pivot a_{ij} na jedničku, vynulovat prvky nad i pod pivotem, a nezměnit přitom sloupce $J \setminus \{j'\}$. Toho se dosáhne těmito řádkovými úpravami:

1. Vyděl řádek i číslem a_{ij}.
2. Pro každé $i' \neq i$ odečti $a_{i'j}$-násobek řádku i od řádku i'.

Říkáme, že jsme provedli *ekvivalentní úpravu kolem pivotu* s indexy (i, j).

Příklad 14.2. Mějme soustavu

$$\begin{bmatrix}\mathbf{A} & \mathbf{b}\end{bmatrix} = \left[\begin{array}{cccccc|c} 0 & 2 & 6 & 1 & 0 & 4 & 4 \\ 1 & \boxed{1} & 3 & 0 & 0 & 2 & 3 \\ 0 & -1 & 1 & 0 & 1 & 2 & 1 \end{array}\right] \tag{14.5}$$

se (standardní) bází $J = \{1, 4, 5\}$. Vidíme ihned odpovídající bázové řešení, $\mathbf{x} = (3, 0, 0, 4, 1, 0)$.

Chceme nahradit bázový sloupec $j' = 1$ nebázovým sloupcem $j = 2$, tedy přejít k sousední bázi $\{2, 4, 5\}$. Máme $i = 2$, tedy pivot je prvek a_{22} (v tabulce orámován). Ekvivalentními řádkovými úpravami musíme docílit, aby pivot byl roven jedné a prvky nad ním a pod ním byly nulové. Při tom smíme změnit sloupec 1, ale sloupce 4 a 5 se změnit nesmějí. Toho se docílí vydělením řádku 2 číslem a_{22} (což zde nemá žádný efekt, protože náhodou $a_{22} = 1$) a pak přičtením vhodných násobků řádku 2 k ostatním řádkům. Výsledek:

$$\begin{bmatrix}\mathbf{A} & \mathbf{b}\end{bmatrix} = \left[\begin{array}{cccccc|c} -2 & 0 & 0 & 1 & 0 & 0 & -2 \\ 1 & 1 & 3 & 0 & 0 & 2 & 3 \\ 1 & 0 & 4 & 0 & 1 & 4 & 4 \end{array}\right].$$

Nyní sloupce $\{2, 4, 5\}$ jsou standardní báze. ♦

14.1.2 Kdy je nové bázové řešení přípustné?

Uvedeným způsobem můžeme od aktuální báze přejít k libovolné sousední bázi. Přitom nové bázové řešení může nebo nemusí být přípustné. Je-li aktuální bázové řešení přípustné, jak poznáme, zda i nové bázové řešení bude přípustné?

Protože nenulové složky bázového řešení $\mathbf{x}$ jsou rovny složkám vektoru $\mathbf{b}$, bázové řešení je přípustné právě tehdy, když $\mathbf{b} \geq \mathbf{0}$. Nechť v aktuální tabulce je $\mathbf{b} \geq \mathbf{0}$. Proveďme ekvivalentní úpravu kolem pivotu (i, j). Hledáme podmínky na (i, j), za kterých bude i po úpravě $\mathbf{b} \geq \mathbf{0}$.

Po ekvivalentní úpravě kolem pivotu (i, j) se vektor $\mathbf{b}$ změní takto (viz §14.1.1):

- b_i se změní na b_i/a_{ij},
- pro každé $i' \neq i$ se $b_{i'}$ změní na $b_{i'} - a_{i'j}b_i/a_{ij}$.

Tato čísla musejí zůstat nezáporná. To nastane právě tehdy, když platí následující podmínky:

$$a_{ij} > 0, \tag{14.6a}$$

$$\text{pro každé } i' \neq i \text{ platí } a_{i'j} \leq 0 \text{ nebo } \frac{b_i}{a_{ij}} \leq \frac{b_{i'}}{a_{i'j}}, \tag{14.6b}$$

kde 'nebo' je užito v nevylučovacím smyslu. Podmínka (14.6a) je zřejmá. Podmínka (14.6b) je ekvivalentní podmínce $b_{i'} - a_{i'j}b_i/a_{ij} \geq 0$, neboť $a_{ij} > 0$, $b_i \geq 0$, $b_{i'} \geq 0$ (rozmyslete!).

Příklad 14.3. V tabulce (14.5):

- Ekvivalentní úprava okolo pivotu $(i, j) = (3, 2)$ nepovede k přípustnému bázovému řešení, neboť $a_{ij} = -1 < 0$, což porušuje podmínku (14.6a).
- Ekvivalentní úprava okolo pivotu $(i, j) = (2, 2)$ nepovede k přípustnému bázovému řešení, neboť pro $i' = 1$ je $a_{i'j} > 0$ a $\frac{3}{1} > \frac{4}{2}$, tedy podmínka (14.6b) je porušena.
- Ekvivalentní úprava okolo pivotu $(i, j) = (3, 6)$ povede k přípustnému bázovému řešení. Podmínky (14.6) jsou splněny, neboť $a_{ij} = 2 > 0$ a $\frac{1}{2} \leq \frac{4}{4}$, $\frac{1}{2} \leq \frac{3}{2}$. ♦

14.1.3 Co když je celý sloupec nekladný?

Jestliže jsou všechny prvky v nějakém nebázovém sloupci j nekladné, víme z podmínky (14.6a), že tento sloupec se nemůže stát bázovým. Platí ale navíc, že souřadnice x_j bodu $\mathbf{x}$ se může libovolně zvětšovat a bod $\mathbf{x}$ přesto zůstane v mnohostěnu X. Tedy existuje polopřímka s počátkem v $\mathbf{x}$ ležící celá v mnohostěnu X. Tedy mnohostěn X je neomezený.

Příklad 14.4. Nechť $\begin{bmatrix}\mathbf{A} & \mathbf{b}\end{bmatrix}$ je tabulka

$$\begin{array}{r|rrrrrr|r}
 & 0 & -2 & 6 & 1 & 0 & 4 & 4 \\
 & 1 & -1 & 3 & 0 & 0 & 2 & 3 \\
 & 0 & -1 & 1 & 0 & 1 & 2 & 1 \\
\hline
\mathbf{x} = & 3 & 0 & 0 & 4 & 1 & 0 &
\end{array}$$

s bází $\{1, 4, 5\}$. Pod tabulkou je napsáno bázové řešení $\mathbf{x}$. Když se x_2 bude libovolně zvětšovat, změnu lze kompenzovat současným zvětšováním bázových proměnných x_1, x_4, x_5 tak, že vektor $\mathbf{Ax}$ zůstane nezměněn a tedy roven $\mathbf{b}$. Konkrétně, pro každé $\alpha \geq 0$ bude vektor $\mathbf{x} = (3, 0, 0, 4, 1, 0) + \alpha(1, 1, 0, 2, 1, 0)$ splňovat $\mathbf{Ax} = \mathbf{b}$ a $\mathbf{x} \geq \mathbf{0}$. ♦

14.1.4 Elementární úpravy účelového řádku

Dosud jsme prováděli elementární řádkové úpravy pouze na soustavě $\mathbf{A}\mathbf{x} = \mathbf{b}$ a účelové funkce si nevšímali. Tyto úpravy lze rozšířit na účelovou funkci. Nebudeme účelovou funkci uvažovat ve tvaru $\mathbf{c}^T\mathbf{x}$, ale v mírně obecnějším tvaru $\mathbf{c}^T\mathbf{x} - d$. Tedy řešíme LP

$$\min\{\, \mathbf{c}^T\mathbf{x} - d \mid \mathbf{A}\mathbf{x} = \mathbf{b},\ \mathbf{x} \geq \mathbf{0}\,\}. \tag{14.7}$$

Úlohu budeme reprezentovat **simplexovou tabulkou**

$$\begin{bmatrix} \mathbf{c}^T & d \\ \mathbf{A} & \mathbf{b} \end{bmatrix}. \tag{14.8}$$

Přičtěme k účelovému řádku $\begin{bmatrix}\mathbf{c}^T & \mathrm{d}\end{bmatrix}$ libovolnou lineární kombinaci $\mathbf{y}^T \begin{bmatrix}\mathbf{A} & \mathbf{b}\end{bmatrix}$ ostatních řádků $\begin{bmatrix}\mathbf{A} & \mathbf{b}\end{bmatrix}$, kde $\mathbf{y}$ jsou koeficienty lineární kombinace. Ukážeme, že tato úprava zachová hodnotu účelové funkce $\mathbf{c}^T\mathbf{x} - d$ pro každé $\mathbf{x}$ splňující $\mathbf{A}\mathbf{x} = \mathbf{b}$. Nový účelový řádek bude

$$\begin{bmatrix}\mathbf{c}^T & d\end{bmatrix} + \mathbf{y}^T \begin{bmatrix}\mathbf{A} & \mathbf{b}\end{bmatrix} = \begin{bmatrix}\mathbf{c}^T + \mathbf{y}^T\mathbf{A} & d + \mathbf{y}^T\mathbf{b}\end{bmatrix}.$$

Nová účelová funkce bude tedy

$$(\mathbf{c}^T + \mathbf{y}^T\mathbf{A})\mathbf{x} - (d + \mathbf{y}^T\mathbf{b}) = \mathbf{c}^T\mathbf{x} - d + \mathbf{y}^T(\mathbf{A}\mathbf{x} - \mathbf{b}).$$

Ale to je rovno $\mathbf{c}^T\mathbf{x} - d$ pro každé $\mathbf{x}$ splňující $\mathbf{A}\mathbf{x} = \mathbf{b}$.

14.1.5 Co udělá přechod k sousední bázi s účelovou funkcí?

Nechť J je standardní báze. Přičtěme k účelovému řádku takovou lineární kombinaci ostatních řádků, aby pro všechna $j \in J$ bylo $c_j = 0$ (novému vektoru $\mathbf{c}$ se pak říká *redukované ceny*). Protože bázové řešení $\mathbf{x}$ je v nebázových sloupcích nulové, znamená to $\mathbf{c}^T\mathbf{x} = 0$. Tedy hodnota účelové funkce $\mathbf{c}^T\mathbf{x} - d$ v bázovém řešení $\mathbf{x}$ je rovna jednoduše $-d$.

Navíc je snadno vidět, co udělá přechod k nové bázi s účelovou funkcí. Nechť j' je sloupec opouštějící bázi a j je sloupec vstupující do báze. Při přechodu k nové bázi se číslo $x_{j'}$ stane nulovým a číslo x_j se zvětší z nuly na kladné (nebo se nezmění). Protože $c_{j'} = 0$, číslo $\mathbf{c}^T\mathbf{x} - d$ při $c_j \geq 0$ stoupne (nebo se nezmění) a při $c_j \leq 0$ klesne (nebo se nezmění).

Příklad 14.5. Mějme úlohu se simplexovou tabulkou

$$\begin{bmatrix} \mathbf{c}^T & d \\ \mathbf{A} & \mathbf{b} \end{bmatrix} = \left[\begin{array}{cccccc|c} 1 & -2 & -3 & -1 & 2 & 1 & 4 \\ \hline 0 & 2 & 6 & 1 & 0 & 4 & 4 \\ 1 & 1 & 3 & 0 & 0 & 2 & 3 \\ 0 & -1 & 1 & 0 & 1 & 2 & 1 \end{array}\right],$$

kde $J = \{1, 4, 5\}$. Složky vektoru $\mathbf{c}$ v bázových sloupcích vynulujeme tak, že k účelovému řádku přičteme první řádek, odečteme druhý řádek, a odečteme dvojnásobek třetího řádku:

$$\begin{array}{r|rrrrrr|r} & 0 & 1 & -2 & 0 & 0 & -1 & 3 \\ \hline & 0 & 2 & 6 & 1 & 0 & 4 & 4 \\ & 1 & 1 & 3 & 0 & 0 & 2 & 3 \\ & 0 & -1 & 1 & 0 & 1 & 2 & 1 \\ \hline \mathbf{x} = & 3 & 0 & 0 & 4 & 1 & 0 & \end{array}$$

Pod tabulku jsme napsali bázové řešení $\mathbf{x}$. Nyní je $\mathbf{c}^T\mathbf{x} = 0$, a tedy hodnota účelové funkce v bázovém řešení je $\mathbf{c}^T\mathbf{x} - d = -d = -3$.

Dejme tomu, že chceme přidat do báze nebázový sloupec 2 a vyloučit z ní některý z bázových sloupců $\{1, 4, 5\}$. Po tomto přechodu se x_2 stane kladné nebo zůstane nulové a jedna ze složek x_1, x_4, x_5 se vynuluje. Protože $c_1 = c_4 = c_5 = 0$, změna x_1, x_4, x_5 se na účelové funkci neprojeví a ta se změní o c_2x_2. Kritérium tedy stoupne nebo zůstane stejné, protože $c_2 = 1 > 0$. ♦

Jestliže v některém sloupci j je $c_j < 0$ a $a_{ij} \leq 0$ pro všechna i, proměnnou x_j můžeme libovolně zvětšovat (viz §14.1.3) a účelovou funkci libovolně zmenšovat. Úloha je tedy neomezená.

14.2 Základní algoritmus

Spojením popsaných stavebních kamenů dostaneme iteraci simplexové metody na řešení úlohy (14.7). Iterace přejde k sousední standardní bázi takové, že bázové řešení zůstane přípustné a účelová funkce se nezvětší. Vstupem i výstupem iterace je simplexová tabulka (14.8) s těmito vlastnostmi:

- podmnožina sloupců $\mathbf{A}$ je standardní báze J,
- bázové řešení odpovídající této bázi je přípustné, $\mathbf{b} \geq \mathbf{0}$,
- složky vektoru $\mathbf{c}$ v bázových sloupcích jsou nulové, $c_j = 0$ pro $j \in J$.

Iterace se provede takto:

1. Vyber index j pivotu tak, aby[1] $c_j < 0$ (§14.1.5).
2. Vyber index i pivotu podle podmínek (14.6). Z těchto podmínek plyne (promyslete!)

$$i \in \operatorname*{argmin}_{i' \mid a_{i'j} > 0} \frac{b_{i'}}{a_{i'j}}, \tag{14.9}$$

   kde $\operatorname*{argmin}_{i' \mid a_{i'j} > 0}$ označuje, že se minimalizuje přes všechna i' splňující $a_{i'j} > 0$.
3. Udělej ekvivalentní úpravu okolo pivotu (i, j) (§14.1.1).
4. Ekvivalentní úpravou účelového řádku vynuluj c_j v novém bázovém sloupci j (§14.1.5).

Algoritmus, který opakuje uvedenou iteraci, nazveme **základní simplexový algoritmus**. Algoritmus končí, když už nelze vybrat pivot (i, j) splňující podmínky 1 a 2 výše. To nastane z jednoho z těchto důvodů:

- Všechny koeficienty c_j jsou nezáporné. Pak účelovou funkci nelze zlepšit a jsme v optimu[2]
- Existuje sloupec j, ve kterém $c_j < 0$ a $a_{ij} \leq 0$ pro všechna i. Pak je úloha neomezená.

Výběr indexů (i, j) pivotu v krocích 1 a 2 nemusí být jednoznačný: může existovat více sloupců j s vhodným znaménkem c_j a více řádků i' může splňovat podmínky (14.6) (tedy množina $\{\, b_{i'}/a_{i'j} \mid i' = 1, \ldots, m,\ a_{i'j} > 0 \,\}$ může mít více minimálních prvků). Algoritmus, který vybírá jediný pivot z těchto možností, se nazývá **pivotové pravidlo**.

[1] Když je lineární program ve tvaru maximalizace, samozřejmě ho nemusíte převádět do minimalizačního tvaru (14.7), stačí jen vybrat sloupec s $c_j > 0$.

[2] Pozor: aktuální bázové řešení může být optimální přesto, že nějaký koeficient c_j je záporný. V další iteraci vložíme sloupec j do báze, ale kvůli degeneraci může zůstat $x_j = 0$, tedy účelová funkce se nezmění (viz §14.1.5).

Příklad 14.6. Vyřešte lineární program (14.7) simplexovou metodou, když výchozí simplexová tabulka (14.8) je

$$\begin{array}{rrrrrr|r} 0 & -2 & 1 & 0 & 0 & -3 & 0 \\ \hline 0 & 2 & 6 & 1 & 0 & 4 & 4 \\ 1 & 1 & 3 & 0 & 0 & 2 & 3 \\ 0 & -1 & 1 & 0 & 1 & \boxed{2} & 1 \end{array}.$$

Báze je $J = \{1, 4, 5\}$ a bázové řešení $\mathbf{x} = (3, 0, 0, 4, 1, 0)$.

Účelový řádek budeme nazývat nultý, ostatní pak prvý, druhý atd. První iterace simplexového algoritmu se provede v těchto krocích:

1. Vybereme sloupec j, který vstoupí do báze. To může být libovolný sloupec, který má v nultém řádku záporné číslo. Můžeme vzít např. nejmenší takové číslo, zde -3, tedy $j = 6$.
2. Vybereme řádek i pivotu dle (14.9) nalezením argumentu minima z čísel $\frac{4}{4}, \frac{3}{2}, \frac{1}{2}$. Bude tedy $i = 3$. Výsledný pivot je označen rámečkem. Všimněte si, že řádek $i = 3$ má v aktuální bázi jedničku ve sloupci 5, sloupec 5 tedy bázi opustí.

3, 4. Uděláme ekvivalentní úpravu okolo pivotu (i, j) a zároveň vynulujeme číslo c_j. Neboli chceme, aby se z pivotu a_{ij} stala jednička a nad i pod pivotem byly nuly, a to včetně nultého řádku. Tedy nejprve třetí řádek vydělíme dvěma a potom k nultému řádku přičteme trojnásobek třetího řádku, od prvního řádku odečteme čtyřnásobek třetího řádku, a od druhého řádku odečteme dvojnásobek třetího řádku. Všimněte si: k žádnému řádku nikdy nepřičítáme násobky jiného řádku než pivotového. Výsledek:

$$\begin{array}{rrrrrr|r} 0 & -3.5 & 2.5 & 0 & 1.5 & 0 & 1.5 \\ \hline 0 & \boxed{4} & 4 & 1 & -2 & 0 & 2 \\ 1 & 2 & 2 & 0 & -1 & 0 & 2 \\ 0 & -0.5 & 0.5 & 0 & 0.5 & 1 & 0.5 \end{array}$$

Na konci první iterace máme bázi $J = \{1, 4, 6\}$, bázové řešení $\mathbf{x} = (2, 0, 0, 2, 0, 0.5)$, a hodnotu účelové funkce $-d = -1.5$.

Druhá iterace: pivot je ve sloupci $j = 2$. Jeho řádek najdeme dle (14.9) porovnáním čísel $\frac{2}{4}, \frac{2}{2}$, tedy $i = 1$. Výsledek druhé iterace:

$$\begin{array}{rrrrrr|r} 0 & 0 & 6 & 0.875 & -0.25 & 0 & 3.25 \\ \hline 0 & 1 & 1 & 0.25 & -0.5 & 0 & 0.5 \\ 1 & 0 & 0 & -0.5 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0.125 & \boxed{0.25} & 1 & 0.75 \end{array}$$

Výsledek třetí iterace:

$$\begin{array}{rrrrrr|r} 0 & 0 & 7 & 1 & 0 & 1 & 4 \\ \hline 0 & 1 & 3 & 0.5 & 0 & 2 & 2 \\ 1 & 0 & 0 & -0.5 & 0 & 0 & 1 \\ 0 & 0 & 4 & 0.5 & 1 & 4 & 3 \end{array}$$

Protože všechna čísla v účelovém řádku jsou nezáporná, algoritmus končí. Úloha má optimální řešení s hodnotou -4 v bodě $(x_1, x_2, x_3, x_4, x_5, x_6) = (1, 2, 0, 0, 3, 0)$. ♦

Příklad 14.7. Nechť simplexová tabulka (14.8) je

-2	6	1	0	0	0
-1	-1	-1	1	0	2
$\boxed{2}$	-1	-2	0	1	1

Tabulka po první iteraci je

0	5	-1	0	1	1
0	-1.5	-2	1	0.5	2.5
1	-0.5	-1	0	0.5	0.5

Podle nultého řádku by další pivot měl být ve třetím sloupci. Ale čísla a_{i3} jsou všechna záporná (viz §14.1.3). Tedy úloha je neomezená. V nové tabulce můžeme zvětšovat x_3 libovolně a kompenzovat to vhodným nárůstem x_1 a x_4. Jelikož $c_1 = c_4 = 0$, změny x_1 a x_4 se na účelové funkci neprojeví a jediný vliv na ní bude mít x_3, které ho bude libovolně zmenšovat. ♦

14.2.1 Cyklení

Zřídka se algoritmus může dostat do stavu, kdy cyklicky prochází stále stejnou množinu bází, které odpovídají jedinému degenerovanému bázovému řešení a tedy účelová funkce se nemění. Algoritmus tedy nikdy neskončí. Tomuto chování říkáme **cyklení** algoritmu.

Příklad 14.8. Zde je počáteční simplexová tabulka a dvě iterace simplexového algoritmu:

-2.3	-2.15	13.55	0.4	0	0	0
$\boxed{0.4}$	0.2	-1.4	-0.2	1	0	0
-7.8	-1.4	7.8	0.4	0	1	0

0	-1	5.5	-0.75	5.75	0	0
1	0.5	-3.5	-0.5	2.5	0	0
0	$\boxed{2.5}$	-19.5	-3.5	19.5	1	0

0	0	-2.3	-2.15	13.55	0.4	0
1	0	$\boxed{0.4}$	0.2	-1.4	-0.2	0
0	1	-7.8	-1.4	7.8	0.4	0

Vidíme, že třetí tabulka se liší od první jen rotací sloupců o dva vpravo. Použijeme-li v dalších iteracích pivotové pravidlo, které bude opět vybírat pivoty 0.4 a 2.5, tak za další čtyři iterace získáme počáteční simplexovou tabulku. ♦

Byla objevena pivotová pravidla zaručující, že algoritmus se pro žádný vstup nezacyklí. Nejznámější je *Blandovo anticyklící pravidlo*: při výběru pivotového sloupce vždy vybereme mezi sloupci s $c_j < 0$ ten s nejmenším indexem, při výběru pivotového řádku vybereme z množiny (14.9) řádek s nejmenším indexem. Důkaz správnosti pravidla vynecháme (není krátký).

14.3 Inicializace algoritmu

Zopakujme, že na začátku základního simplexového algoritmu musí být úloha zadána ve tvaru

$$\min\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{A}\mathbf{x} = \mathbf{b},\ \mathbf{x} \geq \mathbf{0} \,\}, \tag{14.10}$$

kde matice $\mathbf{A}$ obsahuje standardní bázi a $\mathbf{b} \geq \mathbf{0}$. Ukážeme, jak lze obecnou úlohu LP převést na tento tvar.

Někdy je převod snadný. Pokud má úloha tvar $\min\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{A}\mathbf{x} \leq \mathbf{b},\ \mathbf{x} \geq \mathbf{0} \,\}$ a platí $\mathbf{b} \geq \mathbf{0}$, přidáním slacků úlohu převedeme na $\min\{\, \mathbf{c}^T\mathbf{u} \mid \mathbf{A}\mathbf{x} + \mathbf{u} = \mathbf{b},\ \mathbf{x} \geq \mathbf{0},\ \mathbf{u} \geq \mathbf{0} \,\}$. Tato úloha má simplexovou tabulkou

$$\begin{bmatrix} \mathbf{c}^T & \mathbf{0} & 0 \\ \mathbf{A} & \mathbf{I} & \mathbf{b} \end{bmatrix},$$

ve které sloupce příslušné proměnným $\mathbf{u}$ jsou standardní báze.

Příklad 14.9. Vyřešte simplexovým algoritmem:

$$\begin{aligned} \min \quad & -3x_1 - x_2 - 3x_3 \\ \text{za podmínek} \quad & 2x_1 + x_2 + x_3 \leq 2 \\ & x_1 + 2x_2 + 3x_3 \leq 5 \\ & 2x_1 + 2x_2 + x_3 \leq 6 \\ & x_1, x_2, x_3 \geq 0 \end{aligned}$$

Přidáme slackové proměnné $u_1, u_2, u_3 \geq 0$, abychom omezení uvedli do tvaru rovností:

$$\begin{aligned} \min \quad & -3x_1 - x_2 - 3x_3 \\ \text{za podmínek} \quad & 2x_1 + x_2 + x_3 + u_1 = 2 \\ & x_1 + 2x_2 + 3x_3 + u_2 = 5 \\ & 2x_1 + 2x_2 + x_3 + u_3 = 6 \\ & x_1, x_2, x_3, u_1, u_2, u_3 \geq 0 \end{aligned}$$

Zde je výchozí simplexová tabulka:

$$\begin{array}{rrrrrr|l} -3 & -1 & -3 & 0 & 0 & 0 & 0 \\ \hline 2 & 1 & 1 & 1 & 0 & 0 & 2 \\ 1 & 2 & 3 & 0 & 1 & 0 & 5 \\ 2 & 2 & 1 & 0 & 0 & 1 & 6 \end{array}$$

♦

14.3.1 Dvoufázová simplexová metoda

Pokud je úloha zadána v obecném tvaru, operacemi z §12.1 ji lze vždy převést do tvaru (14.10). Vynásobením vhodných řádků záporným číslem vždy zajistíme $\mathbf{b} \geq \mathbf{0}$, matice $\mathbf{A}$ ale nemusí obsahovat standardní bázi. Máme dokonce vážnější problém: není vůbec jasné, zda úloha (14.10) je přípustná. V tomto případě nejdříve vyřešíme *pomocnou úlohu* LP, která najde *nějaké* (ne nutně optimální) přípustné řešení. Z něj pak získáme standardní bázi. Pomocná úloha je

$$\min\{\, \mathbf{1}^T\mathbf{u} \mid \mathbf{A}\mathbf{x} + \mathbf{u} = \mathbf{b},\ \mathbf{x} \geq \mathbf{0},\ \mathbf{u} \geq \mathbf{0} \,\} \tag{14.11}$$

a má simplexovou tabulku

$$\begin{bmatrix} \mathbf{0} & \mathbf{1}^T & 0 \\ \mathbf{A} & \mathbf{I} & \mathbf{b} \end{bmatrix}.$$

Pro libovolné $\mathbf{u} \geq \mathbf{0}$ je $\mathbf{1}^T\mathbf{u} \geq 0$, přičemž $\mathbf{1}^T\mathbf{u} = 0$ právě když $\mathbf{u} = \mathbf{0}$. Tedy úloha (14.10) je přípustná, právě když je optimální hodnota úlohy (14.11) rovna 0. Na počátku tvoří sloupce příslušné proměnným $\mathbf{u}$ standardní bázi, lze tedy na ní pustit základní simplexový algoritmus. Ten může skončit dvěma způsoby:

- Pokud je optimum větší než 0, pak úloha (14.10) je nepřípustná.
- Pokud je optimum rovno 0, pak úloha (14.10) je přípustná. Pokud není optimální řešení $(\mathbf{x}, \mathbf{u})$ úlohy (14.11) degenerované, po skončení simplexového algoritmu jsou všechny bázové proměnné kladné. Protože $\mathbf{u} = \mathbf{0}$, proměnné $\mathbf{u}$ budou tedy nebázové. Proto mezi sloupci příslušnými proměnným $\mathbf{x}$ existuje standardní báze.

  Pokud je optimální řešení $(\mathbf{x}, \mathbf{u})$ úlohy (14.11) degenerované, některé proměnné $\mathbf{u}$ mohou být na konci algoritmu bázové. Pak je nutno udělat dodatečné úpravy kolem pivotů ve sloupcích příslušných bázovým proměnným $\mathbf{u}$, abychom tyto sloupce dostali z báze ven. Toto podrobně popisovat nebudeme.

Nalezení nějakého přípustného řešení v pomocné úloze (14.11) se nazývá **první fáze** a řešení původní úlohy pak **druhá fáze** algoritmu, mluvíme tedy o **dvoufázové simplexové metodě**.

Příklad 14.10. Řešte

$$\begin{array}{rrcrcrcl} \min & -20x_1 & - & 30x_2 & - & 40x_3 & & \\ \text{za podmínek} & 3x_1 & + & 2x_2 & + & x_3 & = & 10 \\ & x_1 & + & 2x_2 & + & 2x_3 & = & 15 \\ & & & & & x_1, x_2, x_3 & \geq & 0 \end{array}$$

Máme sice $\mathbf{b} \geq \mathbf{0}$, ale není jasné, zda existuje přípustné $\mathbf{x}$, tím méně není vidět standardní báze. Provedeme první fázi algoritmu. Pomocná úloha bude

$$\begin{array}{rl} \min & u_1 + u_2 \\ \text{za podmínek} & 3x_1 + 2x_2 + x_3 + u_1 = 10 \\ & x_1 + 2x_2 + 2x_3 + u_2 = 15 \\ & x_1, x_2, x_3, u_1, u_2 \geq 0 \end{array}$$

s tabulkou

$$\begin{array}{ccccc|c} 0 & 0 & 0 & 1 & 1 & 0 \\ \hline 3 & 2 & 1 & 1 & 0 & 10 \\ 1 & 2 & 2 & 0 & 1 & 15 \end{array}$$

Sloupce nad přidanými proměnnými jsou standardní báze, můžeme tedy na pomocnou úlohu pustit základní simplexový algoritmus. Po vynulování ceny nad bázovými proměnnými budou kroky algoritmu vypadat takto:

$$\begin{array}{rrrrr|r} -4 & -4 & -3 & 0 & 0 & -25 \\ \hline 3 & \boxed{2} & 1 & 1 & 0 & 10 \\ 1 & 2 & 2 & 0 & 1 & 15 \end{array}$$

$$\begin{array}{rrrrr|r} 2 & 0 & -1 & 2 & 0 & -5 \\ \hline 1.5 & 1 & 0.5 & 0.5 & 0 & 5 \\ -2 & 0 & \boxed{1} & -1 & 1 & 5 \end{array}$$

$$\begin{array}{rrrrr|r} 0 & 0 & 0 & 1 & 1 & 0 \\ \hline 2.5 & 1 & 0 & 1 & -0.5 & 2.5 \\ -2 & 0 & 1 & -1 & 1 & 5 \end{array}$$

Optimum je rovno 0, tedy původní úloha je přípustná. Proměnné u_1, u_2 jsou nebázové a tedy rovny nule, bázové proměnné jsou x_2, x_3. Teď tedy můžeme začít druhou fázi (řešení původní úlohy) s počáteční tabulkou

$$\begin{array}{rrr|r} -20 & -30 & -40 & 0 \\ \hline 2.5 & 1 & 0 & 2.5 \\ -2 & 0 & 1 & 5 \end{array}$$

♦

14.4 Cvičení

14.1. Najděte všechny báze a bázová řešení mnohostěnu $\{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} = \mathbf{b},\ \mathbf{x} \geq \mathbf{0} \,\}$ pro

$$\begin{bmatrix} \mathbf{A} & \mathbf{b} \end{bmatrix} = \left[\begin{array}{rrrr|r} 1 & 0 & 1 & 0 & 1 \\ 2 & -1 & 4 & 1 & 4 \end{array}\right].$$

Která z nich jsou přípustná? Která z nich jsou degenerovaná?

14.2. Protože $\operatorname{rank} \mathbf{A} = m$, dimenze mnohostěnu (14.1) je $\dim X \leq n - m$. Může být $\dim X$ libovolné číslo mezi 0 a $n-m$? Jestliže ano, dokažte to tak, že pro každou dvojici $(n, k) \in \mathbb{N} \times \{0, \dots, n-m\}$ najdete matici $\mathbf{A} \in \mathbb{R}^{m\times n}$ a vektor $\mathbf{b} \in \mathbb{R}^m$ tak, aby $\operatorname{rank} \mathbf{A} = m$ a $\dim X = k$.

14.3. V tabulce zakroužkujte všechny pivoty takové, že ekvivalentní úprava kolem nich povede k přípustnému bázovému řešení:

$$\begin{bmatrix} \mathbf{A} & \mathbf{b} \end{bmatrix} = \left[\begin{array}{rrrrrrrr|r} 0 & 2 & 6 & 1 & 0 & -4 & 3 & 0 & 4 \\ 1 & 1 & -3 & 0 & 0 & 2 & 3 & 0 & 3 \\ 0 & -1 & 1 & 0 & 1 & -2 & -3 & 0 & 1 \\ 0 & -2 & 2 & 0 & 0 & 2 & -1 & 1 & 1 \end{array}\right]$$

14.4. Zapište lineární program

$$\begin{array}{lrcrcrcrcrcrl} \min & -x_1 & & & & & - & x_4 & - & 3x_5 & & & \\ \text{za podmínek} & 2x_1 & & & & & + & x_4 & + & x_5 & + & x_6 & = 2 \\ & -x_1 & + & x_2 & & & + & 2x_4 & + & 3x_5 & & & = 4 \\ & 2x_1 & & & + & x_3 & + & 2x_4 & - & x_5 & & & = 6 \\ & & & & & & & & & \multicolumn{3}{r}{x_1, x_2, x_3, x_4, x_5, x_6} & \geq 0 \end{array}$$

do simplexové tabulky. Předpokládejte, že aktuální báze je $\{2, 3, 6\}$. Jaké je aktuální bázové řešení? Je toto bázové řešení přípustné. Je degenerované? Pokud je to možné, udělejte jeden krok simplexového algoritmu. Pokud to možné není, vysvětlete proč.

14.5. Vyřešte simplexovou metodou (nejdříve ji co nejjednodušeji incializujte):

$$\begin{array}{lrcrcrl} \max & 2x_1 & - & x_2 & - & 3x_3 & \\ \text{za podmínek} & -2x_1 & - & x_2 & + & x_3 & \leq 2 \\ & -x_1 & + & 2x_2 & - & 3x_3 & \leq 5 \\ & -2x_1 & - & 4x_2 & + & x_3 & \leq 6 \\ & & & & \multicolumn{3}{r}{x_1, x_2, x_3} & \geq 0 \end{array}$$

14.6. Vyřešte simplexovou metodou (i když lze řešit úvahou):

$$\begin{aligned} \max \quad & 6x_1 + 9x_2 + 5x_3 + 9x_4 \\ \text{za podmínek} \quad & x_1 + x_2 + x_3 + x_4 = 1 \\ & x_1, x_2, x_3, x_4 \geq 0 \end{aligned}$$

14.7. Nechť úloha (14.7) má více optimálních řešení. Jak se to pozná v simplexové tabulce? Navrhněte algoritmus, který vypíše všechna optimální bázová řešení.

14.8. Mějme lineární program

$$\begin{array}{rrrrrl} \min & 2x_1 & & -\ 3x_3 & +\ \ x_4 & \\ \text{za podmínek} & x_1 & -\ \ x_2 & -\ \ x_3 & & \geq 0 \\ & -x_1 & +\ 2x_2 & -\ 3x_3 & & \leq 5 \\ & 2x_1 & -\ \ x_2 & -\ \ x_3 & +\ 2x_4 & = 6 \\ & & & & x_1, x_2, x_3, x_4 & \geq 0 \end{array}$$

Inicializujte co nejjednodušším způsobem základní simplexový algoritmus. Vyřešte tímto algoritmem. Nepoužívejte dvoufázovou metodu.

14.9. Vyřešte dvoufázovou simplexovou metodou:

$$\begin{array}{rrrrl} \max & 3x_1 & -\ 4x_2 & & \\ \text{za podmínek} & -2x_1 & -\ 5x_2 & \leq & 10 \\ & 3x_1 & +\ \ x_2 & \leq & 3 \\ & -2x_1 & +\ \ x_2 & \leq & -2 \\ & x_1 & & \geq & 0 \\ & & x_2 & \leq & -1 \end{array}$$

14.10. Nechť úloha (14.7) má simplexovou tabulku

$$\begin{bmatrix} \mathbf{c}^T & d \\ \mathbf{A} & \mathbf{b} \end{bmatrix} = \left[\begin{array}{rrrrrr|r} 3 & 0 & 0 & 1 & -3 & 0 & 0 \\ \hline 7 & 0 & 0 & -3 & 3 & 2 & 1 \\ 1 & 1 & 0 & 0 & -1 & 0 & 3 \\ -1 & 0 & 2 & 1 & -1 & 0 & 1 \end{array}\right]$$

Je tato úloha omezená?

14.11. ($\star$) Cvičeni 3.22 má vztah k důkazu Důsledku 14.1. Jaký je tento vztah?

Nápověda a řešení

14.1.

J	$\{1,2\}$	$\{1,3\}$	$\{1,4\}$	$\{2,3\}$	$\{3,4\}$
$\mathbf{x}$	$(1,-2)$	$(0,1)$	$(1,2)$	$(0,1)$	$(1,0)$
příp.	ne	ano	ano	ano	ano
degen.	ne	ano	ne	ano	ano

14.2. Nápověda: mnohostěn $\{\,(x,y) \in \mathbb{R}^2 \mid x+y=1,\ x\geq 0,\ y\geq 0\,\}$ má dimenzi 1, mnohostěn $\{\,(x,y) \in \mathbb{R}^2 \mid x+y=0,\ x\geq 0,\ y\geq 0\,\}$ má dimenzi 0.

14.5. Úloha je neomezená kvůli prvnímu sloupci.

14.8. Např. otočíme znaménko prvního omezení, třetí omezení vydělíme dvěma, přidáme slackové proměnné pro první a druhé omezení. Pak vynulujeme ceny nad bázovými sloupci. Iterace algoritmu:

$$\begin{array}{rrrrrr|r} 1 & 0.5 & -2.5 & 0 & 0 & 0 & -3 \\ \hline -1 & 1 & \boxed{1} & 0 & 1 & 0 & 0 \\ -1 & 2 & -3 & 0 & 0 & 1 & 5 \\ 1 & -0.5 & -0.5 & 1 & 0 & 0 & 3 \end{array} \qquad \begin{array}{rrrrrr|r} -1.5 & 3 & 0 & 0 & 2.5 & 0 & -3 \\ \hline -1 & 1 & 1 & 0 & 1 & 0 & 0 \\ -4 & 5 & 0 & 0 & 3 & 1 & 5 \\ \boxed{0.5} & 0 & 0 & 1 & 0.5 & 0 & 3 \end{array} \qquad \begin{array}{rrrrrr|r} 0 & 3 & 0 & 3 & 4 & 0 & 6 \\ \hline 0 & 1 & 1 & 2 & 2 & 0 & 6 \\ 0 & 5 & 0 & 8 & 7 & 1 & 29 \\ 1 & 0 & 0 & 2 & 1 & 0 & 6 \end{array}$$

Výsledek: $(x_1, x_2, x_3, x_4) = (6, 0, 6, 0)$, hodnota optima -6.

14.9. Optimum je $(x_1, x_2) = (25, -36)/13$.

14.10. Úloha je neomezená. Mohlo by se zdát, že je omezená, protože neexistuje sloupec j takový, že $c_j < 0$ a $a_{ij} \leq 0$ pro všechna i. Ale po jedné či několika iteracích simplexové metody se takový sloupec objeví.

Kapitola 15

Dualita v lineárním programování

Ke každé úloze LP lze sestavit podle jistého postupu jinou úlohu LP. Novou úlohu nazýváme **duální**, původní úlohu nazýváme **primární** či **přímou**. Konstrukce je symetrická (viz Cvičení 15.1): duální úloha k duální úloze je původní úloha. Tedy má smysl říkat, že primární a duální úloha jsou *navzájem* duální. Dvojice duálních úloh je svázána zajímavými vztahy.

15.1 Konstrukce duální úlohy

K úloze LP v obecném tvaru (viz §12.1) se duální úloha získá dle tohoto postupu:

$$
\begin{array}{rlrll}
\min & \sum_{j\in J} c_j x_j & \max & \sum_{i\in I} b_i y_i & \\
\text{za podm.} & \sum_{j\in J} a_{ij}x_j = b_i & \text{za podm.} & y_i \in \mathbb{R}, & i \in I_0 \\
& \sum_{j\in J} a_{ij}x_j \geq b_i & & y_i \geq 0\,, & i \in I_+ \\
& \sum_{j\in J} a_{ij}x_j \leq b_i & & y_i \leq 0\,, & i \in I_- \\
& x_j \in \mathbb{R} & & \sum_{i\in I} a_{ij}y_i = c_j, & j \in J_0 \\
& x_j \geq 0 & & \sum_{i\in I} a_{ij}y_i \leq c_j, & j \in J_+ \\
& x_j \leq 0 & & \sum_{i\in I} a_{ij}y_i \geq c_j, & j \in J_-
\end{array}
\tag{15.1}
$$

V levém sloupci je primární úloha, v prostředním sloupci je z ní vytvořená duální úloha. V pravém sloupci jsou množiny indexů pro obě úlohy: $I = \{1,\dots,m\} = I_0 \cup I_+ \cup I_-$ je indexová množina primárních omezení a duálních proměnných, $J = \{1,\dots,n\} = J_0 \cup J_+ \cup J_-$ je indexová množina primárních proměnných a duálních omezení.

Všimněte si symetrie dvojice úloh: i-tému primárnímu omezení $\sum_j a_{ij}x_j \geq b_i$ odpovídá duální proměnná[1] $y_i \geq 0$. Opačně, j-tá primární proměnná $x_j \geq 0$ odpovídá j-tému duálnímu omezení $\sum_i a_{ij}y_i \leq c_j$. Tak je to pro všechny řádky s tím, že lineární rovnici v primáru odpovídá neomezená proměnná v duálu, nerovnici typu $\geq$ v primáru odpovídá nezáporná proměnná v duálu, a nerovnici typu $\leq$ v primáru odpovídá nekladná proměnná v duálu.

[1] Poznamenejme, že tato proměnná je vlastně *Lagrangeův multiplikátor* příslušného omezení. Podobně, j-tá primární proměnná x_j je Lagrangeův multiplikátor j-tého duálního omezení $\sum_i a_{ij}x_j \leq c_j$.

Příklad 15.1. Následující dvojice lineárních programů je navzájem duální:

$$
\begin{array}{rrcrcrcrcr}
\min & 2x_1 & & & - & 3x_3 & + & x_4 & & \\
\text{za podm.} & 2x_1 & - & x_2 & + & x_3 & + & 2x_4 & = & 6 \\
& -x_1 & + & 2x_2 & - & 3x_3 & & & \leq & 5 \\
& x_1 & - & x_2 & - & x_3 & - & 3x_4 & \geq & 0 \\
& & & & & & & x_1 & \geq & 0 \\
& & & & & & & x_2 & \in & \mathbb{R} \\
& & & & & & & x_3 & \geq & 0 \\
& & & & & & & x_4 & \leq & 0
\end{array}
\qquad
\begin{array}{rrcrcrcr}
\max & 6y_1 & + & 5y_2 & & & & \\
\text{za podm.} & & & & & y_1 & \in & \mathbb{R} \\
& & & & & y_2 & \leq & 0 \\
& & & & & y_3 & \geq & 0 \\
& 2y_1 & - & y_2 & + & y_3 & \leq & 2 \\
& -y_1 & + & 2y_2 & - & y_3 & = & 0 \\
& y_1 & - & 3y_2 & - & y_3 & \leq & -3 \\
& 2y_1 & & & - & 3y_3 & \geq & 1
\end{array}
$$

♦

Pro speciální tvary LP se dvojice duálních úloh přehledněji napíše v maticové formě. Např. pro $I_0 = I_- = J_0 = J_- = \emptyset$ obdržíme

$$
\begin{array}{rr}
\min & \mathbf{c}^T\mathbf{x} \\
\text{za podm.} & \mathbf{A}\mathbf{x} \geq \mathbf{b} \\
& \mathbf{x} \geq \mathbf{0}
\end{array}
\qquad
\begin{array}{rr}
\max & \mathbf{b}^T\mathbf{y} \\
\text{za podm.} & \mathbf{y} \geq \mathbf{0} \\
& \mathbf{A}^T\mathbf{y} \leq \mathbf{c}
\end{array}
\tag{15.2}
$$

15.2 Věty o dualitě

Následující věty platí pro obecný tvar (15.1), ale důkazy uděláme jen pro speciální tvar (15.2).

Věta 15.1 (o slabé dualitě). Nechť $\mathbf{x}$ je přípustné primární řešení a $\mathbf{y}$ přípustné duální řešení. Pak $\mathbf{c}^T\mathbf{x} \geq \mathbf{b}^T\mathbf{y}$.

Důkaz. Díky přípustnosti $\mathbf{x}$ a $\mathbf{y}$ platí $\mathbf{y}^T\mathbf{A} \leq \mathbf{c}^T$ a $\mathbf{x} \geq \mathbf{0}$, z čehož plyne $\mathbf{y}^T\mathbf{A}\mathbf{x} \leq \mathbf{c}^T\mathbf{x}$. Podobně, díky přípustnosti $\mathbf{x}$ a $\mathbf{y}$ platí $\mathbf{A}\mathbf{x} \geq \mathbf{b}$ a $\mathbf{y} \geq \mathbf{0}$, z čehož plyne $\mathbf{y}^T\mathbf{A}\mathbf{x} \geq \mathbf{y}^T\mathbf{b}$. Z toho

$$\mathbf{c}^T\mathbf{x} \geq \mathbf{y}^T\mathbf{A}\mathbf{x} \geq \mathbf{y}^T\mathbf{b}. \tag{15.3}$$

■

Uveďme jeden okamžitý důsledek slabé duality.

Důsledek 15.2. Nechť $\mathbf{x}$ je přípustné primární řešení a $\mathbf{y}$ je přípustné duální řešení. Nechť $\mathbf{c}^T\mathbf{x} = \mathbf{b}^T\mathbf{y}$. Potom $\mathbf{x}$ je optimální pro primární úlohu a $\mathbf{y}$ je optimální pro duální úlohu.

Důkaz. Pro libovolné primární přípustné řešení $\mathbf{x}'$ je dle věty o slabé dualitě $\mathbf{c}^T\mathbf{x}' \geq \mathbf{b}^T\mathbf{y} = \mathbf{c}^T\mathbf{x}$. Toto platí pro *každé* přípustné $\mathbf{x}'$, tedy řešení $\mathbf{x}$ musí být optimální pro primární úlohu.

Optimalita $\mathbf{y}$ pro duální úlohu se dokáže symetricky. ■

Věta 15.3 (o komplementaritě). Nechť $\mathbf{x}$ je přípustné primární řešení a $\mathbf{y}$ přípustné duální řešení. Pak $\mathbf{c}^T\mathbf{x} = \mathbf{b}^T\mathbf{y}$ právě tehdy, když zároveň platí tyto dvě podmínky:

$$\sum_{j\in J} a_{ij}x_j = b_i \quad \text{nebo} \quad y_i = 0 \qquad \forall i \in I, \tag{15.4a}$$

$$x_j = 0 \quad \text{nebo} \quad \sum_{i\in I} a_{ij}y_j = c_j \qquad \forall j \in J. \tag{15.4b}$$

Podmínky (15.4) budeme nazývat *podmínky komplementarity.* Říkají, že na každém řádku ve dvojici duálních úloh je vždy alespoň jedno omezení aktivní, buď primární nebo duální (přičemž omezení typu rovnosti považujeme vždy za aktivní).

Důkaz. Z nerovnosti (15.3) je jasné, že pro přípustná $\mathbf{x}, \mathbf{y}$ platí ekvivalence

$$\mathbf{c}^T\mathbf{x} = \mathbf{y}^T\mathbf{b} \iff \mathbf{c}^T\mathbf{x} = \mathbf{y}^T\mathbf{A}\mathbf{x} = \mathbf{y}^T\mathbf{b}. \tag{15.5}$$

Dvě rovnosti na pravé straně této ekvivalence jde napsat jako

$$\mathbf{y}^T(\mathbf{A}\mathbf{x} - \mathbf{b}) = 0, \tag{15.6a}$$
$$(\mathbf{c}^T - \mathbf{y}^T\mathbf{A})\mathbf{x} = 0. \tag{15.6b}$$

Levé strany těchto rovností jsou skalární součiny nezáporných (díky přípustnosti $\mathbf{x}, \mathbf{y}$) vektorů. Nyní si stačí uvědomit, že pro libovolné nezáporné vektory $\mathbf{u}, \mathbf{v} \geq \mathbf{0}$ platí

$$\mathbf{u}^T\mathbf{v} = 0 \iff \forall i\,(u_i v_i = 0) \iff \forall i\,(u_i = 0 \text{ nebo } v_i = 0). \qquad \blacksquare$$

Uvědomte si, že Důsledek 15.2 a Věta 15.3 neříkají, že rovnost $\mathbf{c}^T\mathbf{x} = \mathbf{b}^T\mathbf{y}$ vůbec někdy nastane. To je předmětem nejdůležitější věty o dualitě:

Věta 15.4 (o silné dualitě). Primární úloha má optimální řešení, právě když má duální úloha optimální řešení. Má-li primární úloha optimální řešení $\mathbf{x}$ a duální úloha optimální řešení $\mathbf{y}$, pak $\mathbf{c}^T\mathbf{x} = \mathbf{b}^T\mathbf{y}$.

Důkaz této věty není jednoduchý a vynecháme jej. Věty o slabé a silné dualitě mají jasnou interpretaci: pro přípustná $\mathbf{x}$ a $\mathbf{y}$ není hodnota duální účelové funkce nikdy větší než hodnota primární účelové funkce a tyto hodnoty se potkají ve společném optimu:

$$\{\, \mathbf{b}^T\mathbf{y} \mid \mathbf{y} \geq \mathbf{0},\ \mathbf{A}^T\mathbf{y} \leq \mathbf{c} \,\} \;\Big|\; \{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{A}\mathbf{x} \geq \mathbf{b},\ \mathbf{x} \geq \mathbf{0} \,\}$$

společné optimum $\mathbf{c}^T\mathbf{x} = \mathbf{y}^T\mathbf{b}$

Příklad 15.2. Mějme dvojici navzájem duálních úloh LP:

$$\begin{array}{rrcrcrcl}
\text{min} & 2x_1 & + & 5x_2 & + & 6x_3 & = & \mathbf{5.4} \\
\mathbf{3} = & 2x_1 & + & x_2 & + & 2x_3 & \geq & 3 \\
\mathbf{2.4} = & x_1 & + & 2x_2 & + & 2x_3 & \geq & 1 \\
\mathbf{3} = & x_1 & + & 3x_2 & + & x_3 & \geq & 3 \\
\mathbf{-0.6} = & -x_1 & + & x_2 & - & 2x_3 & \geq & -1 \\
\mathbf{1.2} = & x_1 & & & & & \geq & 0 \\
\mathbf{0.6} = & & & x_2 & & & \geq & 0 \\
\mathbf{0} = & & & & & x_3 & \geq & 0
\end{array}
\qquad
\begin{array}{rrcrcrcrcl}
\text{max} & 3y_1 & + & y_2 & + & 3y_3 & - & y_4 & = & \mathbf{5.4} \\
\mathbf{0.2} = & y_1 & & & & & & & \geq & 0 \\
\mathbf{0} = & & & y_2 & & & & & \geq & 0 \\
\mathbf{1.6} = & & & & & y_3 & & & \geq & 0 \\
\mathbf{0} = & & & & & & & y_4 & \geq & 0 \\
\mathbf{2} = & 2y_1 & + & y_2 & + & y_3 & - & y_4 & \leq & 2 \\
\mathbf{5} = & y_1 & + & 2y_2 & + & 3y_3 & + & y_4 & \leq & 5 \\
\mathbf{2} = & 2y_1 & + & 2y_2 & + & y_3 & - & 2y_4 & \leq & 6
\end{array}$$

Spočetli jsme optimální řešení obou úloh a dosadili tato řešení do účelových funkcí a do omezení. Hodnoty optimálních řešení $\mathbf{x}^* = (1.2, 0.6, 0)$ a $\mathbf{y}^* = (0.2, 0, 1.6)$ a hodnoty omezení a účelových funkcí v optimech jsou napsané tučně před/za rovnítky. Dle věty o silné dualitě se optima rovnají. Vezmeme-li libovolný řádek (kromě účelového), je na něm alespoň jedno z obou omezení aktivní. Např. ve druhém řádku je primární omezení $2x_1+x_2+2x_3 \geq 3$ aktivní a duální omezení $y_1 \geq 0$ je neaktivní. Podle věty o komplementaritě se nemůže stát, že by na některém řádku byly obě omezení zároveň neaktivní (mohou být obě ale zároveň aktivní, což zde nenastává, ale může to nastat v případě degenerace). ♦

Zopakujme (viz §12), že pro každou úlohu LP mohou nastat 3 možnosti: úloha má optimální řešení, úloha je neomezená, úloha je nepřípustná.

Věta 15.5. Z devíti možností pro dvojici duálních úloh se realizují tyto:

primární/duální	má optimum	neomezená	nepřípustná
má optimum	ano	ne	ne
neomezená	ne	ne	ano
nepřípustná	ne	ano	ano

Důkaz. Snadno najdeme příklady dvojic duálních úloh, které realizují povolené kombinace. Čtyři zakázané kombinace v prvním řádku a prvním sloupci plynou z první části věty o silné dualitě (primární úloha má optimum *právě tehdy*, když duální úloha má optimum).

Poslední zakázaný případ odůvodníme poněkud neformálně. Lze ukázat, že věta o slabé dualitě platí i tehdy, kdy jedna úloha je neomezená, přičemž pro primární [duální] neomezenou úlohu definujeme hodnotu optima (přesněji infima [suprema]) $-\infty$ [$+\infty$]. Pak tato věta zakazuje, aby úlohy byly zároveň neomezené, protože pak bychom měli $-\infty \geq +\infty$. ■

Předložíme-li přípustná primární a duální řešení taková, že se účelové funkce rovnají, dokázali jsme optimalitu obou úloh. Pro velké úlohy to může být nejsnadnější důkaz optimality (tzv. **certifikát optimality**).

Máme-li duální optimální řešení, jak z něj co nejlevněji spočítat primární optimální řešení? Obecně je k tomu nutno vyřešit soustavu lineárních nerovnic (což není o moc snadnější než vyřešit lineární program). Někdy ale postačí vyřešit soustavu rovnic.

Příklad 15.3. Zkuste dokázat bez použití simplexové metody, že $\mathbf{x} = (x_1, x_2, x_3) = (1.2, 0.6, 0)$ je optimální řešení úlohy z Příkladu 15.2 (přičemž optimální duální řešení $\mathbf{y}$ není známo).

Pomocí věty o komplementaritě zkusíme z daného optimálního $\mathbf{x}$ spočítat optimální $\mathbf{y}$. Protože jsou druhé a čtvrté primární omezení neaktivní, z komplementarity plyne $y_2 = y_4 = 0$. Protože $x_1 > 0$ a $x_2 > 0$, z komplementarity musí být první a druhé duální omezení aktivní. Máme tedy soustavu lineárních rovnic

$$\begin{aligned} 2y_1 + y_3 &= 2 \\ y_1 + 3y_3 &= 5 \end{aligned} \tag{15.7}$$

která má jediné řešení $(y_1, y_3) = (0.2, 1.6)$. Tedy $\mathbf{y} = (0.2, 0, 1.6, 0)$. Toto duální řešení je přípustné (tj. splňuje všechna duální omezení). Protože se hodnota primární účelové funkce v bodě $\mathbf{x}$ rovná hodnotě duální účelové funkce v bodě $\mathbf{y}$, musejí být $\mathbf{x}$ a $\mathbf{y}$ optimální řešení.

Tento postup nemusí vést vždy k cíli. Pokud by duální úloha měla nekonečně mnoho optimálních řešení, soustava (15.7) by měla nekonečně mnoho řešení (měla by např. více proměnných než neznámých). Z nich by bylo nutno vybrat přípustná duální řešení, tedy $\mathbf{y} \geq \mathbf{0}$. Museli bychom tedy řešit soustavu rovnic *a nerovnic.* ♦

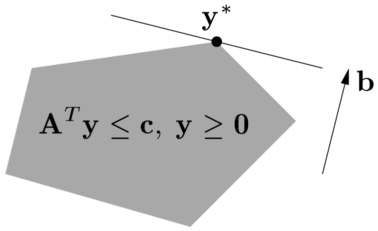
Zkoumejme, jak se změní optimální hodnota úlohy $\min\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{A}\mathbf{x} \geq \mathbf{b},\ \mathbf{x} \geq \mathbf{0} \,\}$, jestliže nepatrně změníme pravé strany omezení $\mathbf{b}$. Odpověď je snadno vidět v duálu.

Věta 15.6 (o stínových cenách). Nechť funkce $f\colon \mathbb{R}^m \to \mathbb{R}$ je definována jako

$$f(\mathbf{b}) = \min\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{A}\mathbf{x} \geq \mathbf{b},\ \mathbf{x} \geq \mathbf{0} \,\} = \max\{\, \mathbf{b}^T\mathbf{y} \mid \mathbf{A}^T\mathbf{y} \leq \mathbf{c},\ \mathbf{y} \geq \mathbf{0} \,\},$$

přičemž předpokládáme, že primární (a tedy i duální) úloha má optimální řešení. Jestliže má duální úloha pro dané $\mathbf{b}$ jediné optimální řešení $\mathbf{y}^*$, pak je funkce f v bodě $\mathbf{b}$ diferencovatelná a platí $f'(\mathbf{b}) = \mathbf{y}^{*T}$, neboli $\partial f(\mathbf{b})/\partial b_i = \mathbf{y}_i^*$.

Důkaz. Je-li $\mathbf{y}^*$ duální optimální řešení pro dané $\mathbf{b}$, je $f(\mathbf{b}) = \mathbf{b}^T\mathbf{y}^*$. Jelikož je toto optimální řešení jediné, nabývá se v extremálním bodě mnohostěnu $\{\, \mathbf{y} \in \mathbb{R}^m \mid \mathbf{A}^T\mathbf{y} \leq \mathbf{c},\ \mathbf{y} \geq \mathbf{0} \,\}$, viz obrázek:



Změníme-li nepatrně $\mathbf{b}$, optimální duální řešení $\mathbf{y}^*$ se nezmění a zůstane jediné (toto odůvodnění není zcela rigorózní, ale geometricky je dostatečně názorné). Tedy při malé změně vektoru $\mathbf{b}$ je hodnota optima stále rovna $f(\mathbf{b}) = \mathbf{b}^T\mathbf{y}^*$, v malém okolí bodu $\mathbf{b}$ je tedy funkce f lineární. Její derivace je $f'(\mathbf{b}) = \mathbf{y}^{*T}$. ■

Zdůrazněme předpoklad jednoznačnosti optimálního řešení. Kdyby množina duálních optimálních řešení byla ne jediný extremální bod, ale stěna vyšší dimenze, po malé změně vektoru $\mathbf{b}$ by se optimální stěna mohla stát extremálním bodem a funkce f by tedy v bodě $\mathbf{b}$ nebyla diferencovatelná.

Protože $\mathbf{b}$ je zároveň vektor pravých stran primární úlohy, optimální duální proměnné $\mathbf{y}^*$ vyjadřují *citlivost* optima primární úlohy na změnu pravých stran primárních omezení $\mathbf{A}\mathbf{x} \geq \mathbf{b}$. Interpretujeme-li naše LP jako optimální výrobní plán (12.9) (pozor, liší se obrácenou nerovností v omezení), pak hodnota y_i^* říká, jak by se náš výdělek zvětšil, kdybychom trochu uvolnili omezení na výrobní zdroje $\mathbf{a}_i^T\mathbf{x} \leq b_i$. V ekonomii se proto duálním proměnným říká **stínové ceny** primárních omezení.

Všimněte si, že věta o stínových cenách je ve shodě s větou o komplementaritě. Pokud $y_i^* = 0$, je dovoleno $\mathbf{a}_i^T\mathbf{x} < b_i$ a tedy malá změna b_i nemá na optimum vliv.

Příklad 15.4. Nechť je známo, že duální úloha v Příkladu 15.2 má jediné optimální řešení. Stínová cena prvního primárního omezení $2x_1+x_2+2x_3 \geq 3$ je $y_1 = 0.2$. Změňme pravou stranu $b_1 = 3$ tohoto omezení o malou hodnotu $h = 0.01$ a zkoumejme, jak se změní optimum. Tato změna nezmění argument $\mathbf{y}^*$ duálního optima, pouze jeho hodnotu $\mathbf{b}^T\mathbf{y}^*$. Podle silné duality hodnota primárního optima musí být rovna hodnotě duálního optima (argument $\mathbf{x}^*$ primárního optima se může nějak změnit, to nás ale nezajímá). Dvojice úloh tedy bude vypadat takto:

$$
\begin{array}{rrcrcrcl}
\min & 2x_1 & + & 5x_2 & + & 6x_3 & = & \mathbf{5.402} \\
& 2x_1 & + & x_2 & + & 2x_3 & \geq & 3.01 \\
& x_1 & + & 2x_2 & + & 2x_3 & \geq & 1 \\
& x_1 & + & 3x_2 & + & x_3 & \geq & 3 \\
& -x_1 & + & x_2 & - & 2x_3 & \geq & -1 \\
& x_1 & & & & & \geq & 0 \\
& & & x_2 & & & \geq & 0 \\
& & & & & x_3 & \geq & 0
\end{array}
\qquad
\begin{array}{rrcrcrcrcl}
\max & 3.01y_1 & + & y_2 & + & 3y_3 & - & y_4 & = & \mathbf{5.402} \\
\mathbf{0.2} = & y_1 & & & & & & & \geq & 0 \\
\mathbf{0} = & & & y_2 & & & & & \geq & 0 \\
\mathbf{1.6} = & & & & & y_3 & & & \geq & 0 \\
\mathbf{0} = & & & & & & & y_4 & \geq & 0 \\
\mathbf{2} = & 2y_1 & + & y_2 & + & y_3 & - & y_4 & \leq & 2 \\
\mathbf{5} = & y_1 & + & 2y_2 & + & 3y_3 & + & y_4 & \leq & 5 \\
\mathbf{2} = & 2y_1 & + & 2y_2 & + & y_3 & - & 2y_4 & \leq & 6
\end{array}
$$

V okolí bodu $\mathbf{b} = (3, 1, 3, -1)$, ve kterém se nemění optimální $\mathbf{y}^*$, bude $f(\mathbf{b}) = \mathbf{b}^T\mathbf{y}^*$ a tedy hodnota společného optima se změní o $h \cdot \partial f(\mathbf{b})/\partial b_1 = h \cdot y_1 = 0.2 \cdot 0.01$ na 5.402. ♦

15.3 Příklady na konstrukci a interpretaci duálních úloh

Dualita umožňuje *vhled* do řešeného problému, často velmi netriviální. Abychom danou úlohu (fyzikální, ekonomickou či jinou) popsanou lineárním programem pochopili do hloubky, je dobré pochopit význam nejen primární úlohy, ale i duální úlohy a vztahy mezi primární a duální úlohou (tj. věty o dualitě). Často se nám podaří dokázat platnost silné duality pro náš konkrétní problém.

Příklad 15.5. Demonstrujme nyní dualitu na velmi jednoduchém lineárním programu

$$\min\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{1}^T\mathbf{x} = 1,\ \mathbf{x} \geq \mathbf{0} \,\} = \min\{\, c_1x_1 + \cdots + c_nx_n \mid x_1 + \cdots + x_n = 1,\ x_i \geq 0 \,\},$$

kde čísla $\mathbf{c} = (c_1, \ldots, c_n) \in \mathbb{R}^n$ jsou dána. Chceme přesně rozumět, proč v tomto případě platí věty o silné dualitě a o komplementaritě.

Úvahou snadno vidíme (viz Cvičení 12.3), že optimální hodnota je $\min_{i=1}^n c_i$ a nabývá se ve vektoru $\mathbf{x}$ jehož všechny složky jsou nulové kromě složek příslušných minimálnímu c_i. Pokud je více minimálních prvků c_i, optimální $\mathbf{x}$ není dáno jednoznačně. Např. pro $\mathbf{c} = (1, 3, 1, 2)$ bude optimálním řešením každé $\mathbf{x} = (x_1, 0, x_3, 0)$ pro $x_1, x_3 \geq 0$ splňující $x_1 + x_3 = 1$.

Podle návodu (15.1) sestrojíme duální úlohu

$$\max\{\, y \in \mathbb{R} \mid y\mathbf{1} \leq \mathbf{c} \,\} = \max\{\, y \in \mathbb{R} \mid y \leq c_i\ \forall i = 1, \ldots, n \,\}.$$

Neboli hledá se největší číslo y, které je menší než všechna čísla c_i. Takové číslo y je zřejmě nejmenší z čísel c_i. Tedy platí silná dualita.

Podmínky komplementarity říkají, že v optimech bude alespoň jedno z odpovídající dvojice primární-duální omezení aktivní. Dvojice omezení $\sum_i x_i = 1$, $y \in \mathbb{R}$ splňuje podmínky komplementarity triviálně. Dvojice omezení $x_i \geq 0$, $y \leq c_i$ je splňuje právě tehdy, když je splněna aspoň jedna z rovností $x_i = 0$, $y = c_i$. To znamená:

- Pokud je v duálu $y < c_i$, v primáru musí být $x_i = 0$. To je ale jasné, protože $y < c_i$ znamená, že c_i není nejmenší ze složek vektoru $\mathbf{c}$ a proto mu v primáru nemůžeme přiřadit nenulovou váhu x_i.
- Obráceně, pokud je v primáru $x_i > 0$, musí být v duálu $y = c_i$. To je jasné, protože pokud jsme v primáru přiřadili číslu c_i nenulovou váhu, musí být nejmenší. ♦

Příklad 15.6. Pro daná $\mathbf{c} \in \mathbb{R}^n$ a $k \in \{1, \dots, n\}$ máme úlohu

$$\max\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{0} \le \mathbf{x} \le \mathbf{1},\ \mathbf{1}^T\mathbf{x} \le k\ \} \tag{15.8}$$

Duální úlohu sestrojíme dle předpisu (15.1):

$$\begin{array}{rrcl}
\max & c_1x_1 + \cdots + c_nx_n & & \\
\text{za podm.} & x_1 + \cdots + x_n & \le & k \\
 & x_i & \le & 1 \\
 & x_i & \ge & 0
\end{array}
\qquad
\begin{array}{rrcll}
\min & ky + z_1 + \cdots + z_n & & & \\
\text{za podm.} & y & \ge & 0 & \\
 & z_i & \ge & 0 & i = 1, \dots, n \\
 & y + z_i & \ge & c_i & i = 1, \dots, n
\end{array}$$

kde vlevo jsme opsali primární úlohu (15.8) a vpravo jsme napsali sestrojenou duální úlohu. Pravá úloha odpovídá levé úloze z dvojice (15.1), protože v úloze (15.8) se maximalizuje. Sestrojit duální úlohu je zde obtížnější než např. v Příkladě 15.2 kvůli přítomnosti indexů $i = 1, \dots, n$.[2] Doporučujeme např. napsat dvojici (15.1) podrobně pro nějaké konkrétní malé n (např. $n = 3$) a pak ji přepsat do obecného tvaru.

Podmínky komplementarity:

- $\sum_i x_i = k$ nebo $y = 0$
- Pro každé i je $x_i = 1$ nebo $z_i = 0$
- Pro každé i je $x_i = 0$ nebo $y + z_i = c_i$

Primární úloha (15.8) se snadno vyřeší úvahou (viz Cvičení 12.3): její optimální hodnota je součet k největších kladných čísel c_i. Ovšem na první pohled není patrné, proč duální úloha má stejnou optimální hodnotu. Zkuste takovou úvahu vymyslet! ♦

Příklad 15.7 (Příklad: Ekonomická interpretace duality). Vraťme se k Příkladu 1.10 o výrobci lupínků a hranolků z brambor a oleje. Napišme k této úloze duální úlohu:

$$\begin{array}{rrcl}
\max & 120l + 76h & & \\
\text{za podm.} & 2l + 1.5h & \le & 100 \\
 & 0.4l + 0.2h & \le & 16 \\
 & l & \ge & 0 \\
 & h & \ge & 0
\end{array}
\qquad
\begin{array}{rrcl}
\min & 100a + 16b & & \\
\text{za podm.} & a & \ge & 0 \\
 & b & \ge & 0 \\
 & 2a + 0.4b & \ge & 120 \\
 & 1.5a + 0.2b & \ge & 76
\end{array}$$

Přijde překupník a chce koupit od výrobce jeho zásoby brambor a oleje. Překupník řeší tuto otázku: Jaké nejnižší ceny musím nabídnout, aby mi výrobce ještě své zásoby prodal? Tvrdíme, že toto je význam duální úlohy.

Vskutku, nechť a, b označují nabízenou cenu za jednotku brambor a oleje. Překupník chce minimalizovat celkovou cenu za suroviny $100a + 16b$. Musí být $2a + 0.4b \ge 120$, neboť jinak by výrobci více vyplatilo vyrobit ze všech brambor a oleje lupínky a prodat je, než prodat suroviny. Ze stejného důvodu musí být $1.5a + 0.2b \ge 76$. Optimální duální řešení je $a = 32$ a $b = 140$.

Toto je další důvod (kromě Věty 15.6), proč se optimálním duálním proměnným někdy říká *stínové ceny* odpovídajících primárních omezení. Např. stínová cena brambor je 32 Kč/kg. ♦

[2] Přesněji řečeno: na rozdíl od v Příkladu 15.2, zde zápis (15.8) nepopisuje jedinou instanci úlohy LP, ale nekonečnou množinu instancí. Každá instance je definovaná parametry $(n, \mathbf{c}, k)$. Viz footnote 8 v §2.4).

Příklad 15.8. Z §12.4 víme, že optimální argument úlohy

$$\min_{x\in\mathbb{R}} \sum_{i=1}^{n} |x-a_i| = \min\{\, y_1+\cdots+y_n \mid y_i\in\mathbb{R},\ x\in\mathbb{R},\ -y_i\le x-a_i\le y_i \,\} \tag{15.9}$$

je medián z čísel $a_1,\ldots,a_n$. Chceme sestrojit duální úlohu a zjednodušit ji. Chceme pro tuto úlohu dokázat platnost silné duality.

Duální úlohu je možné sestrojit podle předpisu (15.1):

$$\begin{array}{rrcl}
\min & \sum_{i=1}^n y_i & & \\
\text{za podm.} & x+y_i & \ge & a_i \\
 & -x+y_i & \ge & -a_i \\
 & y_i & \in & \mathbb{R} \\
 & x & \in & \mathbb{R}
\end{array}
\qquad
\begin{array}{rrcll}
\max & \sum_{i=1}^n (p_i-q_i)a_i & & & \\
\text{za podm.} & p_i & \ge & 0 & i=1,\ldots,n \\
 & q_i & \ge & 0 & i=1,\ldots,n \\
 & p_i+q_i & = & 1 & i=1,\ldots,n \\
 & \sum_{i=1}^n (p_i-q_i) & = & 0 &
\end{array}$$

kde vlevo jsme opsali primární úlohu (15.9) a vpravo jsme napsali sestrojenou duální úlohu. Opět buďte opatrní při konstrukci duální úlohy, protože např. první řádek primárních omezení je vlastně n-tice nerovnic $x+y_i\ge a_i$ pro $i=1,\ldots,n$, viz diskuze v Příkladě 15.6. Pro někoho může být příjemnější nejprve formulovar primár v maticovém tvaru a pak mechanicky sestrojit duál – to ukážeme v Cvičení 15.3.d pro úlohu (12.18), která je zobecněním úlohy (15.9).

Duální úlohu dále zjednodušíme substitucí

$$2p_i = 1+t_i, \quad 2q_i = 1-t_i.$$

Po této substituci je $p_i-q_i=t_i$ a podmínka $p_i+q_i=1$ je splněna automaticky. Podmínka $\sum_i(p_i-q_i)=0$ odpovídá podmínce $\sum_i t_i=0$. Podmínka $p_i\ge 0$ odpovídá $t_i\ge -1$ a podmínka $q_i\ge 0$ odpovídá $t_i\le 1$. Duální úloha s novými proměnnými $\mathbf{t}\in\mathbb{R}^n$ je tedy

$$\max\{\, a_1t_1+\cdots+a_nt_n \mid t_i\in\mathbb{R},\ t_1+\cdots+t_n=0,\ -1\le t_i\le 1 \,\}. \tag{15.10}$$

Primární úloha (15.9) a duální úloha (15.10) spolu zdánlivě vůbec nesouvisejí – avšak podle silné duality jejich optimální hodnoty musejí být stejné. Zkusme dokázat, že tomu tak je.

Nejprve si všimneme, že optimální hodnota primární úlohy (15.9) se nezmění, posuneme-li čísla $a_1,\ldots,a_n$ o libovolnou konstantu $b\in\mathbb{R}$. To je jasné, neboť medián x se posune o stejnou konstantu a je $|(x-b)-(a_i-b)|=|x-a_i|$. Totéž platí pro duální úlohu (15.10), neboť díky podmínce $\sum_i t_i=0$ je $\sum_i(a_i-b)t_i=\sum_i a_it_i$. Proto bez ztráty obecnosti můžeme zvolit $b=\operatorname{median}_i a_i$, neboli posunout body tak, že jejich medián bude $x=0$.

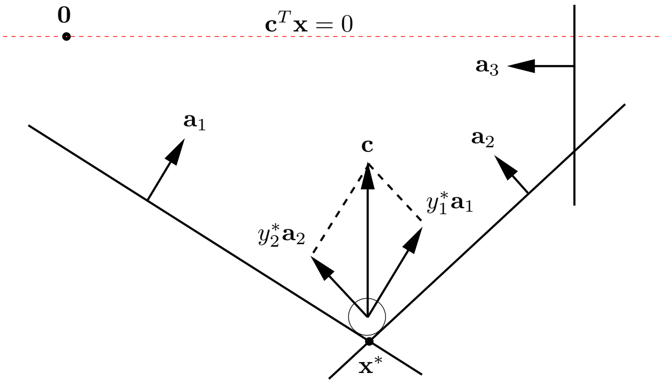
Nyní je primární optimální hodnota rovna jednoduše $\sum_i|x-a_i|=\sum_i|a_i|$. Protože kladných a záporných čísel a_i je stejný počet, duální úloha nabývá optima v takovém vektoru $\mathbf{t}$, kde $t_i=-1$ pro $a_i<0$ a $t_i=1$ pro $a_i>0$ (což splňuje podmínku $\sum_i t_i=0$). Tedy duální optimální hodnota je také $\sum_i a_it_i=\sum_i|a_i|$. ♦

Lineární programy lze modelovat mechanickými (obecněji fyzikálními) modely. Můžeme jim říkat mechanické či analogové počítače. Takové modely se dobře hodí na ilustraci duality.

Příklad 15.9. Uvažujme dvojici duálních úloh

$$\min\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{A}\mathbf{x}\ge\mathbf{b},\, \mathbf{x}\in\mathbb{R}^n \,\} \;=\; \max\{\, \mathbf{b}^T\mathbf{y} \mid \mathbf{A}^T\mathbf{y}=\mathbf{c},\, \mathbf{y}\ge\mathbf{0} \,\}. \tag{15.11}$$

Mějme mnohostěn tvořený poloprostory $\mathbf{a}_i^T\mathbf{x}\ge b_i$ a vektor $\mathbf{c}$ mířici svisle vzhůru:



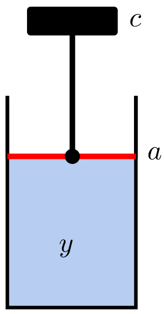
Vhoďme do mnohostěnu maličký míček, na který působí tíhová síla $-\mathbf{c}$. Pro míček se středem v bodě $\mathbf{x}$ je číslo $\mathbf{c}^T\mathbf{x}$ přímo úměrné výšce míčku nad vodorovnou rovinou procházející počátkem, tedy potenciální energii míčku. Míček se zastaví v místě s nejmenší potenciální energií, což je nejnižší vrchol $\mathbf{x}^*$. Proto $\mathbf{x}^*$ je řešením primární úlohy.

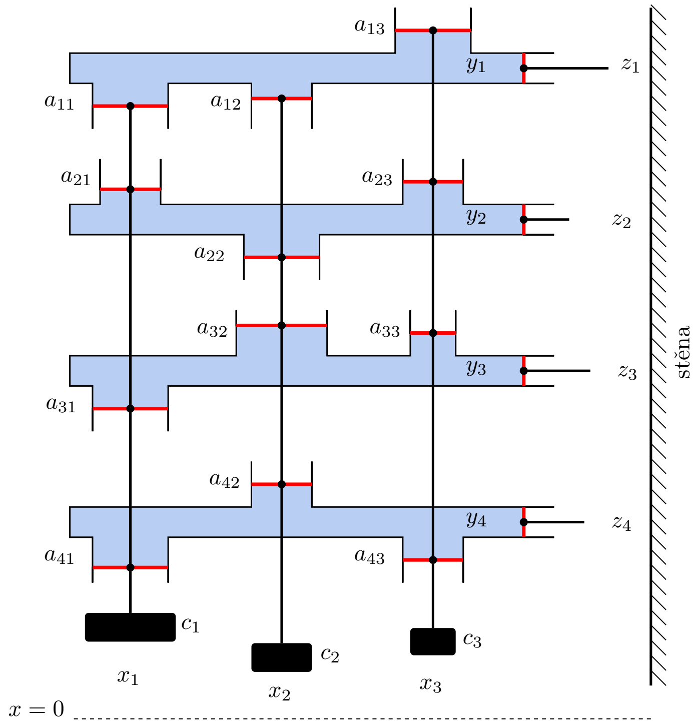
Podívejme se teď na duální úlohu. V bodě $\mathbf{x}^*$ je míček v klidu a proto pro něj platí rovnováha sil: tíha $-\mathbf{c}$ se vyrovnává silami stěn. Tedy existují síly y_i^* tak, že $\mathbf{c} = \sum_i y_i^*\mathbf{a}_i = \mathbf{A}^T\mathbf{y}^*$. Musí být $y_i^* \geq 0$, protože stěny působí silou jen dovnitř mnohostěnu, ne ven. Vidíme, že $\mathbf{y}^*$ je přípustné řešení duální úlohy.

Když $\mathbf{a}_i^T\mathbf{x}^* > b_i$, míček se stěny i nedotýká a tedy síla stěny na míček je nulová, $y_i^* = 0$. Proto $y_i^*(\mathbf{a}_i^T\mathbf{x}^* - b_i) = 0$, což je podmínka komplementarity. Sečtením těchto podmínek získáme $\sum_i y_i^*\mathbf{a}_i^T\mathbf{x}^* - \sum_i b_i\mathbf{x}^* = 0$ neboli $\mathbf{c}^T\mathbf{x}^* = \mathbf{b}^T\mathbf{y}^*$ (to je vlastně Věta 15.3). Tedy platí silná dualita.

Rovnost $\mathbf{c}^T\mathbf{x}^* = \mathbf{b}^T\mathbf{y}^*$ jde vidět i jinak. Potenciální energie $\mathbf{c}^T\mathbf{x}^*$ míčku v bodě $\mathbf{x}^*$ se rovná práci, která by se vykonala posunutím míčku do počátku. Ukážeme, že tato práce je rovna $\mathbf{b}^T\mathbf{y}^*$. Odstraňme nejprve všechny stěny, kterých se míček nedotýká. Posuňme nyní některou stěnu i rovnoběžně tak, aby procházela počátkem. Při tomto posunování se síly stěn na míček nemění. Vzdálenost stěny od počátku je $b_i/\|\mathbf{a}_i\|_2$ (viz Cvičení 5.16). Síla stěny na míček je $y_i^*\mathbf{a}_i$ a působí ve směru posouvání, tedy vykonaná práce je $(b_i/\|\mathbf{a}_i\|_2) \cdot \|y_i^*\mathbf{a}_i\|_2 = b_iy_i^*$. Po odtlačení všech stěn do počátku vykonáme práci $\sum_i b_iy_i^* = \mathbf{b}^T\mathbf{y}^*$. ♦

Příklad 15.10. Známý fyzikální zákon říká, že v nádobě s nestačitelnou kapalinou uzavřenou pístem s povrchem a, na který působí síla c, je tlak $y = c/a$ (obrázek vlevo):





Uvažujme nyní stroj na obrázku vpravo. Stroj se skládá z m nádob s nestlačitelnou kapalinou, z nichž každá je uzavřena n svislými písty a jedním vodorovným pístem. Každá m-tice svislých pístů je spojena tyčí, zakončenou dole závažím. Povrch svislého pístu v nádobě i spojeného se závažím j je $|a_{ij}|$. Pro $a_{ij} > 0$ je píst umístěn nahoře a pro $a_{ij} < 0$ dole. Výška závaží j nad referenční rovinou je x_j. Každý vodorovný píst má jdnotkový povrch a je zakončen tyčí, která nemůže projít stěnou vpravo. Šířka mezery mezi i-tou tyčí a stěnou je z_i, přičemž při $\mathbf{x} = \mathbf{0}$ je $\mathbf{z} = -\mathbf{b}$. Tíha j-tého závaží je c_j. Tlak v i-té nádrži je y_i.

Ukážeme, že stroj 'řeší' levou (primární) úlohu z dvojice (15.11). Ze zachování objemu kapaliny v nádobě i plyne $z_i = a_{i1}x_1 + \cdots + a_{in}x_n - b_i$, tedy $\mathbf{z} = \mathbf{A}\mathbf{x} - \mathbf{b}$. Protože vodorovné tyče nemohou projít stěnou, musí vždy být $z_i \geq 0$, tedy $\mathbf{z} \geq \mathbf{0}$. Potenciální energie závaží j je $c_j x_j$. Když bude stroj v rovnováze, závaží budou v takových výškách, že jejich celková potenciální energie $c_1x_1 + \cdots + c_nx_n = \mathbf{c}^T\mathbf{x}$ bude minimální.

Jaký je význam duálních omezení? Protože povrch vodorovných pístů je jednotkový, tlak y_i v nádrži i se rovná síle vodorovné tyče i na stěnu. Jelikož stěna působí silou vždy od sebe, je $\mathbf{y} \geq \mathbf{0}$. Rovnováha sil pro svislou tyč j zní $a_{1j}y_1 + \cdots + a_{mj}y_m = c_j$, tedy $\mathbf{A}^T\mathbf{y} = \mathbf{c}$.

Dle věty o silné dualitě má být v ustáleném stavu duální kritérium $\mathbf{y}^T\mathbf{b} = y_1b_1 + \cdots + y_mb_m$ rovno potenciální energii $\mathbf{c}^T\mathbf{x}$ závaží. Proč tomu tak je? Potenciální energie závaží je rovna práci, nutné na jejich posunutí do roviny $x = 0$. Tato práce se dá vykonat buď přímo posunutím závaží (což odpovídá primárnímu kritériu $\mathbf{c}^T\mathbf{x}$) nebo posunutím vodorovných tyčí do vzdálenosti $-b_i$ od stěny. Druhý způsob odpovídá duálnímu kritériu. Zafixujeme-li totiž všechny vodorovné tyče kromě tyče i, při odtlačování tyče i se síla, kterou působíme na tyč, nemění. Tedy vykonáme práci y_ib_i. Když takto odtlačíme od stěny postupně všechny tyče, vykonáme práci $\mathbf{y}^T\mathbf{b}$.

Dle věty o komplementaritě v ustáleném stavu pro každé i platí $z_i = 0$ nebo $y_i = 0$. To je jasné, protože když se některá vodorovná tyč nedotýká stěny, je tlak v její nádobě nulový.

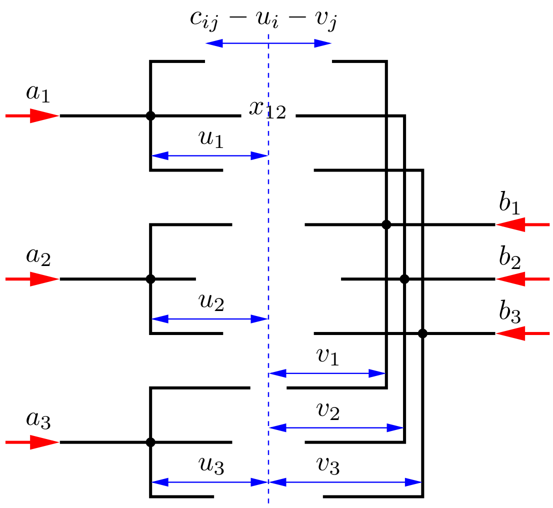
Věta o stínových cenách říká, že se změnou b_i se hodnota minimální potenciální energie mění tím více, čím je větší tlak y_i. To je ale jasné, protože čím je větší tlak, tím větší práce je třeba na posunutí tyče od stěny do vzdálenosti b_i. ♦

Mohlo by se zdát, že mechanické modely z předchozích dvou příkladů dokazují větu o silné dualitě pro dvojici úloh (15.11). Dříve jsme ale řekli, že důkaz této věty je složitý. Jak je to možné? Naše fyzikální úvahy větu o silné dualitě *nedokazují*, předpokládají totiž platnost fyzikálních zákonů, které nelze matematicky dokázat ale pouze experimentálně pozorovat. Naproti tomu matematický důkaz žádné fyzikální zákony nepředpokládá.

Příklad 15.11. Zkoumejme duální úlohu k dopravní úloze (12.11). Konstrukci duálu už nebudeme popisovat prodrobně, výsledná dvojice navzájem duálních úloh je

$$
\begin{array}{rrl}
\min & \sum_{i=1}^n \sum_{j=1}^n c_{ij}x_{ij} & \\
\text{za podm.} & \sum_{j=1}^n x_{ij} = a_i & \\
 & \sum_{i=1}^m x_{ij} = b_j & \\
 & x_{ij} \geq 0 &
\end{array}
\qquad
\begin{array}{rrl}
\max & \sum_{i=1}^m a_i u_i + \sum_{j=1}^n b_j v_j & \\
\text{za podm.} & u_i \in \mathbb{R} & i = 1, \ldots, m \\
 & v_j \in \mathbb{R} & j = 1, \ldots, n \\
 & u_i + v_j \leq c_{ij} & i = 1, \ldots, m,\ j = 1, \ldots, n
\end{array}
$$

Význam dvojice úloh ilustrujeme analogovým počítačem na obrázku (pro $m = n = 3$):



Stroj se skládá z $m + n$ pevných vidlic, z nichž každá se může vodorovně posouvat. Proměnné u_i a v_j jsou posunutí vidlic vzhledem k referenční svislé rovině[3] (na obrázku čárkovaně). Je-li $u_i = v_j = 0$, vzdálenost hrotů vidlic i a j je c_{ij}. Posunutí vidlic je omezeno kontakty dvojic hrotů, neboli podmínkami $c_{ij} - u_i - v_j \geq 0$. Na levé vidlice působí konstantní síly a_i, na pravé b_j (červené šipky). Působením sil se levé vidlice budou přibližovat k pravým, ale jen do té doby, než do sebe některé dvojice hrotů narazí. Proměnné x_{ij} odpovídají silám, které na sebe působí hroty vidlic (na obrázku jsou všechny tyto síly nulové, neboť žádné dva hroty se dosud nedotýkají). Dumejte, čemu odpovídá optimální hodnota úloh a proč platí silná dualita a podmínky komplementarity! ♦

15.4 Cvičení

15.1. Dokažte, že duál duálu se rovná původní úloze. Udělejte pro (a) dvojici (15.2), (b) dvojici (15.1).

15.2. Pro daná čísla $c_1, \ldots, c_n \in \mathbb{R}$ chceme maximalizovat $\sum_{i=1}^n c_i x_i$ za podmínek $-1 \leq x_i \leq 1$.

[3] Chování stroje samozřejme není polohou této referenční roviny ovlivněno. Jak se to projevuje algebraicky v primární a duální úloze?

a) Vyřešte úvahou.
b) Sestrojte duální úlohu a upravte ji do co nejjednoduššího tvaru. Vyřešte duální úlohu úvahou (musí vám vyjít stejná optimální hodnota jako u primární úlohy).
c) Napište podmínky komplementarity.
d) Najděte číselné hodnoty optimálních primárních a duálních proměnných (které si odpovídají přes podmínky komplementarity) pro $n = 3$ a $(c_1, c_2, c_3) = (-2, 3, 4)$.

15.3. Napište duální úlohu a podmínky komplementarity k následujícím úlohám. Pokud úloha není LP, nejdříve převeďte na LP (dle §12.1). Výslednou duální úlohu zjednodušte, je-li to možné. Kde to má smysl, pokuste se interpretovat duální úlohu a věty o dualitě, podobně jako v Příkladu 15.5.

a) lineární program ze Cvičení 14.8
b) $\min_{x\in\mathbb{R}} \max_{i=1}^n |a_i - x|$ (střed intervalu)
c) úloha (12.16) (přibližné řešení přeurčené lin. soustavy ve smyslu max-normy)
d) úloha (12.18) (přibližné řešení přeurčené lin. soustavy ve smyslu 1-normy)
e) všechny úlohy ze Cvičení 12.3
f) úloha vzniklá ve Cvičení 12.11 o kladce se závažími
g) minimalizace maxima afinních funkcí (viz §12.1.1):
   (i) $\min\limits_{x_1,x_2\in\mathbb{R}} \max\{\, 2x_1 - x_2 - 3,\ 1 - x_1,\ x_2 - 2,\ x_1 + x_2 \,\}$
   (ii) $\min\limits_{\mathbf{x}\in\mathbb{R}^n} \max\limits_{i=1}^{m}(\mathbf{a}_i^T\mathbf{x} + b_i)$

15.4. Dokažte bez užití algoritmu na řešení LP, že $\mathbf{x} = (1, 1, 1, 1)$ je optimální řešení úlohy

$$\begin{array}{rl} \min & \begin{bmatrix} 47 & 93 & 17 & -93 \end{bmatrix} \mathbf{x} \\ \text{za podm.} & \begin{bmatrix} -1 & -6 & 1 & 3 \\ -1 & -2 & 7 & 1 \\ 0 & 3 & -10 & -1 \\ -6 & -11 & -2 & 12 \\ 1 & 6 & -1 & -3 \end{bmatrix} \mathbf{x} \le \begin{bmatrix} -3 \\ 5 \\ -8 \\ -7 \\ 4 \end{bmatrix} \end{array}$$

Nápověda a řešení

15.1. Pravou úlohu nejdříve převedeme na tvar levé úlohy a pak k ní napíšeme duální úlohu. Ukážeme jen pro (15.2):

$$\begin{array}{rl} -\min & (-\mathbf{b})^T\mathbf{y} \\ \text{za podm.} & (-\mathbf{A}^T)\mathbf{y} \ge -\mathbf{c} \\ & \mathbf{y} \ge \mathbf{0} \end{array} \qquad \begin{array}{rl} -\max & (-\mathbf{c})^T\mathbf{x} \\ \text{za podm.} & \mathbf{x} \ge \mathbf{0} \\ & (-\mathbf{A})\mathbf{x} \le -\mathbf{b} \end{array} \tag{15.12}$$

Levá úloha (15.12) je ekvivalentní pravé úloze (15.2), pravá úloha (15.12) je ekvivalentní levé úloze (15.2).

15.2.a) Optimální hodnota je $\sum_{i=1}^n |c_i|$ (viz Cvičení 12.3)

15.2.b) Duální úloha je $\min\{\sum_i (u_i + v_i) \mid u_i, v_i \ge 0,\ v_i - u_i = c_i\}$. Vyřešme ji úvahou. Nejdříve uvažujme každé i zvlášť a ukažme, že $\min\{\, u+v \mid u, v \ge 0,\ v-u = c \,\} = |c|$. Podmínka $v - u = c$ zůstává v platnosti, odečteme-li od u, v libovolné číslo. Pokud $u+v$ má být minimální, musí se od u, v odečíst co největší číslo tak, aby platilo $u, v \ge 0$. Tedy jedno z čísel u, v bude nulové a tedy optimální hodnota bude $|c|$. Optimální hodnota celé duální úlohy bude tedy $\sum_i |c_i|$.

15.2.c) Pro každé i platí $x_i = -1$ nebo $u_i = 0$. Pro každé i platí $x_i = 1$ nebo $v_i = 0$.

15.2.d) $(x_1, x_2, x_3) = (-1, 1, 1)$, $(u_1, u_2, u_3) = (2, 0, 0)$, $(v_1, v_2, v_3) = (0, 3, 4)$.

15.3.d) Dvojici úloh lze psát v maticovém tvaru

$$\begin{array}{rrrr} \min & \mathbf{1}^T\mathbf{y} & \max & \mathbf{b}^T(\mathbf{q}-\mathbf{p}) \\ \text{za podm.} & \mathbf{A}\mathbf{x}+\mathbf{y} \geq \mathbf{b} & \text{za podm.} & \mathbf{p} \geq \mathbf{0} \\ & -\mathbf{A}\mathbf{x}+\mathbf{y} \geq -\mathbf{b} & & \mathbf{q} \geq \mathbf{0} \\ & \mathbf{y} \in \mathbb{R}^n & & \mathbf{p}+\mathbf{q} = \mathbf{1} \\ & \mathbf{x} \in \mathbb{R}^m & & \mathbf{A}^T(\mathbf{p}-\mathbf{q}) = \mathbf{0} \end{array} \tag{15.13}$$

neboli

$$\begin{array}{rrrr} \min & \mathbf{h}^T\mathbf{u} & \max & \mathbf{g}^T\mathbf{v} \\ \text{za podm.} & \mathbf{F}\mathbf{u} \geq \mathbf{g} & \text{za podm.} & \mathbf{v} \geq \mathbf{0} \\ & \mathbf{u} \in \mathbb{R}^{m+n} & & \mathbf{F}^T\mathbf{v} = \mathbf{h} \end{array} \tag{15.14}$$

kde

$$\mathbf{F} = \begin{bmatrix} \mathbf{A} & \mathbf{I} \\ -\mathbf{A} & \mathbf{I} \end{bmatrix}, \quad \mathbf{g} = \begin{bmatrix} \mathbf{b} \\ -\mathbf{b} \end{bmatrix}, \quad \mathbf{h} = \begin{bmatrix} \mathbf{0} \\ \mathbf{1} \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} \mathbf{x} \\ \mathbf{z} \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} \mathbf{p} \\ \mathbf{q} \end{bmatrix},$$

přičemž vektor duálních proměnných $\mathbf{v}$ jsme rozdělili na dva bloky $\mathbf{p}, \mathbf{q}$ odpovídající blokům matic $\mathbf{F}$ a $\mathbf{g}$. Při konstrukci duálu tedy postupujeme takto: nejdříve napíšeme úlohu (12.18) jako levou úlohu v (15.13), pak ji napíšeme v maticové formě jako levou úlohu v (15.14), k ní mechanicky napíšeme duál jako pravou úlohu v (15.14), a tu nakonec zjednodušíme roznásobením matic do výsledné duální úlohy napravo v (15.13).

15.3.f) Vlevo je lineární program z Cvičení 12.11, vpravo je jeho duál vytvořený dle návodu (15.1):

$$\begin{array}{rlrrl} \min & \sum_{i=1}^n m_i z_i + \sum_{i=1}^{n'} m'_i z'_i & \max & \sum_{i=1}^n y_i d_i + \sum_{i=1}^{n'} y'_i d'_i & \\ \text{za podm.} & z_i - x \geq d_i & \text{za podm.} & y_i \geq 0 & i = 1, \dots, n \\ & z'_i + x \geq d'_i & & y'_i \geq 0 & i = 1, \dots, n' \\ & x \geq 0 & & \sum_{i=1}^{n'} y'_i - \sum_{i=1}^n y_i = 0 & \\ & z_i \geq 0 & & y_i \leq m_i & i = 1, \dots, n \\ & z'_i \geq 0 & & y'_i \leq m'_i & i = 1, \dots, n' \end{array}$$

Podmínky komplementarity: $z_i(y_i - m_i) = 0$, $z'_i(y'_i - m'_i) = 0$, $(z_i - d_i - x)y_i = 0$, $(z'_i - d'_i + x)y'_i = 0$.

Interpretace duálu: duální proměnné y_i, y'_i odpovídají silám (přičemž zanedbáváme násobení gravitačním zrychlením), kterými závaží napínají vlákna. Každá síla je nezáporná (závaží vždy táhne dolů) ale ne větší než tíha závaží. Duální omezení $\sum_{i=1}^n y_i = \sum_{i=1}^{n'} y'_i$ je rovnováha sil (kdyby neplatila, kladka by se otáčela). Zkuste sami přijít na to, proč hodnoty primárního a duálního kritéria jsou v ustáleném stavu stejné (silná dualita), inspirujte se Příkladem 15.9.

15.3.g) (i) Nejprve převedeme na LP: $\min z$ z.p. $2x_1 - x_2 - 3 \leq z$, $1 - x_1 \leq z$, $x_2 - 2 \leq z$, $x_1 + x_2 \leq z$. Duál k tomuto LP je: $\max -3u_1 + u_2 - 2u_3$ z.p. $2u_1 - u_2 + u_4 = 0$, $-u_1 + u_3 + u_4 = 0$, $u_1 + \cdots + u_4 = 1$, $u_1, \dots, u_4 \geq 0$.

15.4. Postupujte podle Příkladu 15.3. Nejdříve ověřte, že daný vektor $\mathbf{x}$ je přípustný, tj. splňuje daná omezení (kdyby ne, určitě by nebyl optimální). Napište si duální lineární program, jeho proměnné označte $\mathbf{y} \in \mathbb{R}^5$. Dle věty o komplementaritě je dané $\mathbf{x}$ optimální, právě když existuje $\mathbf{y}$, které splňuje duální omezení (tedy je přípustné pro duál) a podmínky komplementarity. Protože z pěti primárních omezení jsou v daném bodě $\mathbf{x} = (1, 1, 1, 1)$ první čtyři splněny s rovností (tj. aktivní) a pátá je splněna s ostrou nerovností (tj. neaktivní), z podmínek komplementarity plyne $y_5 = 0$. Duální omezení příslušné primárním proměnným $\mathbf{x} \in \mathbb{R}^4$ bude tedy soustava čtyř rovnic o čtyřech neznámých y_1, y_2, y_3, y_4. Vyřešením této soustavy zjistíme, že má jediné řešení, které navíc splňuje $y_1, y_2, y_3, y_4 \leq 0$. Tedy jsme získali vektor $\mathbf{y}$, který je přípustný pro dál a splňuje podmínky komplementarity. Závěr: dané $\mathbf{x}$ je optimální.

Část IV

Konvexní optimalizace

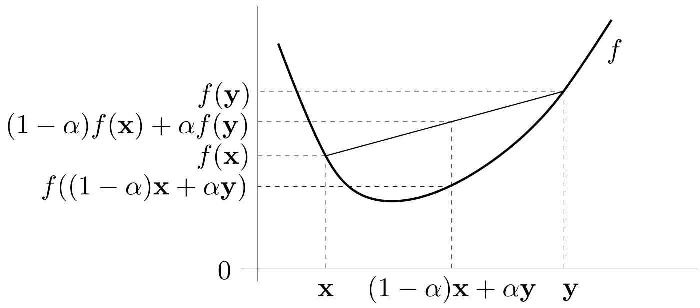
Kapitola 16

Konvexní funkce

Funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ je **konvexní** na konvexní množině $X \subseteq \mathbb{R}^n$, jestliže

$$\mathbf{x} \in X,\ \mathbf{y} \in X,\ 0 \le \alpha \le 1 \quad \Longrightarrow \quad f((1-\alpha)\mathbf{x} + \alpha\mathbf{y}) \le (1-\alpha)f(\mathbf{x}) + \alpha f(\mathbf{y}). \tag{16.1}$$

Geometrický význam definice je ten, že úsečka spojující body $(\mathbf{x}, f(\mathbf{x}))$ a $(\mathbf{y}, f(\mathbf{y}))$ nezasahuje pod graf funkce (rozmyslete, jak podmínka (16.1) odpovídá obrázku!):



Funkce f je **konkávní** na množině X, jestliže je funkce $-f$ konvexní na množině X. Rozlišujte pojmy *konvexní množina* a *konvexní funkce*, jde o různé věci! Všimněte si, že množina X musí být konvexní (konvexita ani konkavita funkce na nekonvexní množině není definována). Pokud X je celý definiční obor funkce f, odkaz na X můžeme vynechat a říkáme jen, že funkce f je konvexní.

Podmínku (16.1) lze zobecnit pro více než dva body: f je konvexní na X, právě když

$$\left.\begin{aligned} \mathbf{x}_1, \dots, \mathbf{x}_k &\in X \\ \alpha_1, \dots, \alpha_k &\ge 0 \\ \alpha_1 + \cdots + \alpha_k &= 1 \end{aligned}\right\} \quad \Longrightarrow \quad f(\alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k) \le \alpha_1 f(\mathbf{x}_1) + \cdots + \alpha_k f(\mathbf{x}_k). \tag{16.2}$$

Podmínka (16.2) je známá jako **Jensenova nerovnost**. Podmínka (16.1) je očividně speciální případ podmínky (16.2). Naopak lze dokázat (neuvádíme), že (16.1) implikuje (16.2). Porovnejte s definicí lineárního zobrazení (3.5) a afinního zobrazení!

Příklad 16.1. Dokažme z podmínky (16.1), že funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ daná jako $f(\mathbf{x}) = \max_{i=1}^n x_i$ (tedy funkční hodnota je maximum ze složek vektoru $\mathbf{x}$) je konvexní. Máme dokázat, že pro každé $\mathbf{x}$, $\mathbf{y}$ a $0 \le \alpha \le 1$ platí

$$f((1-\alpha)\mathbf{x}+\alpha\mathbf{y}) = \max_i((1-\alpha)x_i+\alpha y_i) \tag{16.3a}$$
$$\le \max_i(1-\alpha)x_i + \max_i \alpha y_i \tag{16.3b}$$
$$= (1-\alpha)\max_i x_i + \alpha \max_i y_i \tag{16.3c}$$
$$= (1-\alpha)f(\mathbf{x}) + \alpha f(\mathbf{y}) \tag{16.3d}$$

kde rovnost (16.3c) plyne z nezápornosti čísel α a $1-\alpha$.

Nerovnost (16.3b) plyne z toho, že pro každé $a_1,\dots,a_n,b_1,\dots,b_n \in \mathbb{R}$ platí

$$\max_i(a_i+b_i) \le \max_i a_i + \max_i b_i. \tag{16.4}$$

Nerovnost (16.4) dokážeme například takto. Nechť i^*, j^*, k^* jsou indexy, ve kterých se nabývají maxima, tedy $a_{i^*}+b_{i^*} = \max_i(a_i+b_i)$, $a_{j^*} = \max_i a_i$, $b_{k^*} = \max_i b_i$. Proto $a_{i^*} \le a_{j^*}$ a $b_{i^*} \le b_{k^*}$. Tedy $\max_i(a_i+b_i) = a_{i^*}+b_{i^*} \le a_{j^*}+b_{k^*} = \max_i a_i + \max_i b_i$. ♦

Příklad 16.2. Dokažme z podmínky (16.1), že funkce $f(\mathbf{x}) = \min_{i=1}^n x_i$ není konvexní. Např. volba $n=2$, $\mathbf{x}=(0,2)$, $\mathbf{y}=(2,0)$, $\alpha=\frac{1}{2}$ nesplňuje (16.1), neboť

$$f((\mathbf{x}+\mathbf{y})/2) = f(1,1) = 1 > (f(\mathbf{x})+f(\mathbf{y}))/2 = (0+0)/2 = 0.$$ ♦

Použitím Jensenovy nerovnosti na vhodnou konvexní funkci lze získat mnoho užitečných nerovností.

Příklad 16.3. Funkce ln je konkávní na $\mathbb{R}_{++}$ (kde $\mathbb{R}_{++}$ označuje množinu kladných reálných čísel, viz §1.1.1). Napišme pro tuto funkci Jensenovu nerovnost (16.2) (jelikož funkce je konkávní a ne konvexní, musíme v Jensenově nerovnosti obrátit směr nerovnosti), ve které položíme $\alpha_1 = \cdots = \alpha_n = \frac{1}{n}$:

$$\ln\frac{x_1+\cdots+x_n}{n} \ge \frac{\ln x_1+\cdots+\ln x_n}{n}$$

kde $x_1,\dots,x_n$ jsou kladné. Vezmeme-li exponenciálu každé strany, dostaneme

$$\frac{x_1+\cdots+x_n}{n} \ge (x_1\cdots x_n)^{1/n}.$$

Tato známá nerovnost říká, že aritmetický průměr není nikdy menší než geometrický. ♦

Příklad 16.4. Uveďme často potkávané jednoduché konvexní či konkávní funkce:

1. Exponenciála $f(x) = e^{ax}$ je konvexní na $\mathbb{R}$, pro libovolné $a \in \mathbb{R}$.
2. Mocnina $f(x) = x^a$ je na $\mathbb{R}_{++}$ konvexní pro $a \ge 1$ nebo $a \le 0$ a konkávní pro $0 \le a \le 1$.
3. Mocnina absolutní hodnoty $f(x) = |x|^a$ je pro $a \ge 1$ konvexní na $\mathbb{R}$ (speciálně: absolutní hodnota $|x|$ je konvexní).
4. Logaritmus $f(x) = \ln x$ je konkávní na $\mathbb{R}_{++}$.
5. Funkce $f(x) = x\ln x$ je konvexní na $\mathbb{R}_{++}$ (nebo i na $\mathbb{R}_+$, pokud dodefinujeme $0\ln 0 = 0$, což se často dělá, protože $\lim_{x\to 0^+} x\ln x = 0$). Tato funkce se vyskytuje např. jako jeden člen ve vzorci pro Shannonovu entropii náhodné veličiny, $H(\mathbf{x}) = -\sum_i x_i \ln x_i$.

6. Afinní funkce $f(\mathbf{x}) = \mathbf{a}^T\mathbf{x} + b$ je zároveň konvexní i konkávní.
7. Kvadratická forma $f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x}$ je konvexní pro $\mathbf{A}$ positivně semidefinitní, konkávní pro $\mathbf{A}$ negativně semidefinitní, a nekonvexní a nekonkávní pro $\mathbf{A}$ indefinitní (viz Příklad 16.5).
8. Maximum složek $f(\mathbf{x}) = \max_{i=1}^n x_i = \max\{x_1, \ldots, x_n\}$ je konvexní na $\mathbb{R}^n$.
9. Log-sum-exp funkce $f(\mathbf{x}) = \ln(e^{x_1} + \cdots + e^{x_n})$ je konvexní. Tato funkce se někdy nazývá *měkké maximum*, neboť funkce

$$f_t(\mathbf{x}) = f(t\mathbf{x})/t = \ln(e^{tx_1} + \cdots + e^{tx_n})/t$$

se pro $t \to +\infty$ blíží funkci $\max_{i=1}^n x_i$ (dokažte výpočtem limity!).
10. Geometrický průměr $f(\mathbf{x}) = (x_1 \cdots x_n)^{1/n}$ je konkávní na $\mathbb{R}^n_+$.
11. Každá norma je konvexní funkce, neboť pro každé $\alpha, \beta \geq 0$ máme

$$\|\alpha\mathbf{x} + \beta\mathbf{y}\| \leq \|\alpha\mathbf{x}\| + \|\beta\mathbf{y}\| = \alpha\|\mathbf{x}\| + \beta\|\mathbf{y}\|,$$

kde nerovnost plyne z trojúhelníkové nerovnosti a rovnost z homogenity (viz §12.4.1).

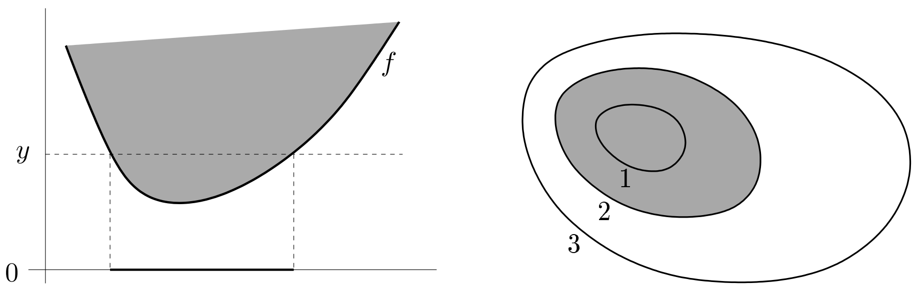
Načrtněte vrstevnice a grafy těchto funkcí (v případě více proměnných jen pro $n \in \{1, 2\}$)! ♦

16.1 Vztah konvexní funkce a konvexní množiny

Zopakujte si pojmy vrstevnice a graf funkce z §1.1.3! Zavedeme dva podobné pojmy, které se liší pouze nahrazením rovnosti nerovností. Pro funkci $f\colon \mathbb{R}^n \to \mathbb{R}$ definujeme:

- **Epigraf** funkce je množina $\{\, (\mathbf{x}, y) \in \mathbb{R}^{n+1} \mid f(\mathbf{x}) \leq y \,\}$.
- **Subkontura**[1] výšky y je množina $\{\, \mathbf{x} \in \mathbb{R}^n \mid f(\mathbf{x}) \leq y \,\}$.

Levý obrázek znázorňuje subkonturu výšky y a epigraf funkce $\mathbb{R} \to \mathbb{R}$, pravý obrázek subkonturu výšky 2 funkce $\mathbb{R}^2 \to \mathbb{R}$:



Existují těsné vztahy mezi konvexitou funkce a konvexitou jejího epigrafu a subkontur (což jsou množiny), dané následujícími větami.

Věta 16.1. Funkce je konvexní, právě když její epigraf je konvexní množina.

Důkaz. Předpokládejme, že funkce f je konvexní. Vezměme dva body $(\mathbf{x}_1, y_1)$ a $(\mathbf{x}_2, y_2)$ z epigrafu, tedy $f(\mathbf{x}_1) \leq y_1$ a $f(\mathbf{x}_2) \leq y_2$. Pro každé $0 \leq \alpha \leq 1$ platí

$$f((1-\alpha)\mathbf{x}_1 + \alpha\mathbf{x}_2) \leq (1-\alpha)f(\mathbf{x}_1) + \alpha f(\mathbf{x}_2) \leq (1-\alpha)y_1 + \alpha y_2,$$

[1] Slovo 'subkontura' je pokus o český překlad anglického 'sublevel set'.

kde první nerovnost plyne z konvexity funkce a druhá nerovnost z $f(\mathbf{x}_1) \leq y_1$ a $f(\mathbf{x}_2) \leq y_2$. Tedy bod $(1-\alpha)(\mathbf{x}_1, y_1) + \alpha(\mathbf{x}_2, y_2)$ patří do epigrafu, který je proto konvexní množina.

Obráceně předpokládejme, že epigraf funkce f je konvexní množina. Tedy pokud body $(\mathbf{x}_1, y_1)$ a $(\mathbf{x}_2, y_2)$ patří do epigrafu, pak také bod $(1-\alpha)(\mathbf{x}_1, y_1) + \alpha(\mathbf{x}_2, y_2)$ patří do epigrafu pro každé $0 \leq \alpha \leq 1$. Volbou $y_1 = f(\mathbf{x}_1)$ a $y_2 = f(\mathbf{x}_2)$ máme

$$f((1-\alpha)\mathbf{x}_1 + \alpha\mathbf{x}_2) \leq (1-\alpha)y_1 + \alpha y_2 = (1-\alpha)f(\mathbf{x}_1) + \alpha f(\mathbf{x}_2),$$

proto je funkce f konvexní. ■

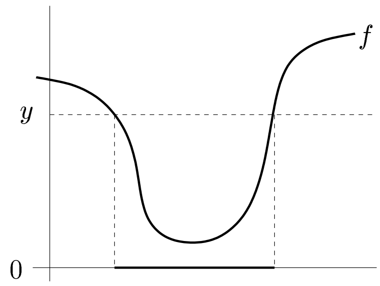
Věta 16.2. Každá subkontura konvexní funkce je konvexní množina.

Důkaz. Předpokládejme, že body $\mathbf{x}_1$ a $\mathbf{x}_2$ patří do subkontury konvexní funkce f, tedy $f(\mathbf{x}_1) \leq y$ a $f(\mathbf{x}_2) \leq y$. Pro každé $0 \leq \alpha \leq 1$ platí

$$f((1-\alpha)\mathbf{x}_1 + \alpha\mathbf{x}_2) \leq (1-\alpha)f(\mathbf{x}_1) + \alpha f(\mathbf{x}_2) \leq (1-\alpha)y + \alpha y = y,$$

kde první nerovnost plyne z konvexity funkce a druhá z nerovností $f(\mathbf{x}_1) \leq y$, $f(\mathbf{x}_2) \leq y$. Tedy bod $(1-\alpha)\mathbf{x}_1 + \alpha\mathbf{x}_2$ patří do subkontury. Dle (13.1) je tedy subkontura konvexní množina. ■

Obrácená implikace ve Větě 16.2 neplatí: existuje funkce, která není konvexní a jejíž každá subkontura je konvexní množina[2]. Např. každá subkontura monotonní (tj. nerostoucí nebo neklesající) funkce jedné proměnné je interval, tedy konvexní množina. Jiný příklad je na obrázku:



16.2 Konvexita diferencovatelných funkcí

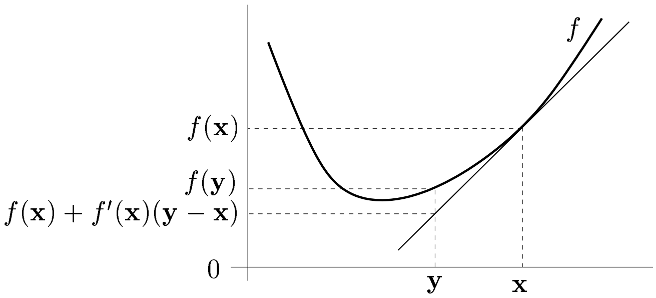
Konvexní funkce nemusí být v každém bodě diferencovatelná (uvažte např. funkci $f(x) = |x|$). Pokud je ale funkce jednou či dvakrát diferencovatelná, její konvexitu lze snadněji než pomocí podmínky (16.1) (které se někdy říká podmínka nultého řádu) charakterizovat pomocí derivací. Následující dvě věty uvedeme bez důkazů.

Věta 16.3 (Podmínka prvního řádu). Diferencovatelná funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ je konvexní na otevřené konvexní množině $X \subseteq \mathbb{R}^n$, právě když

$$\mathbf{x}, \mathbf{y} \in X \quad\Longrightarrow\quad f(\mathbf{y}) \geq f(\mathbf{x}) + f'(\mathbf{x})\,(\mathbf{y} - \mathbf{x}).$$

[2]Funkce, jejíž každá subkontura je konvexní množina, se nazývá *kvazikonvexní*. Kvazikonvexní funkce jsou užitečné, ale zdaleka ne tak jako konvexní funkce.

To znamená, že Taylorův polynom prvního řádu funkce f v každém bodě $\mathbf{x} \in X$ (viz (8.23b)) je všude (tj. pro každé $\mathbf{y}$) menší nebo roven funkci f:



Všimněte si, že věta nic neříká o konvexitě funkcí na konvexních množinách, které nejsou otevřené (na hranici je totiž situace trochu složitější). To ovšem moc nevadí, neboť větu stejně nejčastěji použijeme pro $X = \mathbb{R}^n$ (což je otevřená konvexní množina).

Věta 16.3 má jeden příjemný důsledek. Vzpomeňme (viz §10.1), že nulovost parciálních derivací je obecně jen nutná podmínka na volný lokální extrém. Je-li f konvexní, je tato podmínka ovšem postačující pro globální minimum.

Důsledek 16.4. Nechť funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ je diferencovatelná v bodě $\mathbf{x} \in \mathbb{R}^n$ a konvexní. Jestliže $f'(\mathbf{x}) = \mathbf{0}$, pak $\mathbf{x}$ je globální minimum funkce f.

Důkaz. Je-li $f'(\mathbf{x}) = \mathbf{0}$, dle Věty 16.3 máme $f(\mathbf{y}) \geq f(\mathbf{x})$ pro všechna $\mathbf{y} \in \mathbb{R}^n$. Tedy $\mathbf{x}$ je globální minimum funkce f. (Jak bude vypadat obrázek výše pro tento případ?) ∎

Věta 16.5 (Podmínka druhého řádu). Dvakrát diferencovatelná funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ je konvexní na otevřené konvexní množině $X \subseteq \mathbb{R}^n$, právě když v každém bodě $\mathbf{x} \in X$ je Hessova matice $f''(\mathbf{x})$ positivně semidefinitní.

Příklad 16.5. Nechť $f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x}$, kde $\mathbf{A}$ je symetrická positivně semidefinitní. Ukažme konvexitu této funkce třemi způsoby:

- Dokažme konvexitu z Věty 16.5. To je triviální, protože Hessova matice je $f''(\mathbf{x}) = 2\mathbf{A}$ a tedy je positivně semidefinitní.
- Dokažme konvexitu z Věty 16.3. Protože $f'(\mathbf{x}) = 2\mathbf{x}^T\mathbf{A}$, máme dokázat, že

$$\mathbf{y}^T\mathbf{A}\mathbf{y} \geq \mathbf{x}^T\mathbf{A}\mathbf{x} + 2\mathbf{x}^T\mathbf{A}(\mathbf{y} - \mathbf{x}).$$

  To jde upravit na $\mathbf{x}^T\mathbf{A}\mathbf{x} - 2\mathbf{x}^T\mathbf{A}\mathbf{y} + \mathbf{y}^T\mathbf{A}\mathbf{y} \geq 0$. Platí[3]

$$\mathbf{x}^T\mathbf{A}\mathbf{x} - 2\mathbf{x}^T\mathbf{A}\mathbf{y} + \mathbf{y}^T\mathbf{A}\mathbf{y} = (\mathbf{x} - \mathbf{y})^T\mathbf{A}(\mathbf{x} - \mathbf{y}), \tag{16.5}$$

  což je nezáporné pro každé $\mathbf{x}, \mathbf{y}$, protože $\mathbf{A}$ je positivně semidefinitní.

[3] Všimněte si, že pro $n = 1$ a $\mathbf{A} = 1$ se rovnost (16.5) zjednoduší na známé $x^2 - 2xy + y^2 = (x - y)^2$.

- Dokážme konvexitu z podmínky (16.1). Musíme dokázat, že pro každé $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ a $0 \leq \alpha \leq 1$ platí (16.1), tedy

$$[(1-\alpha)\mathbf{x} + \alpha\mathbf{y}]^T \mathbf{A}[(1-\alpha)\mathbf{x} + \alpha\mathbf{y}] \leq (1-\alpha)\mathbf{x}^T\mathbf{A}\mathbf{x} + \alpha\mathbf{y}^T\mathbf{A}\mathbf{y}$$

  Po roznásobení a převedení všech členů na jednu stranu upravujeme:

$$\begin{aligned}(\alpha - \alpha^2)\mathbf{x}^T\mathbf{A}\mathbf{x} - 2\alpha(1-\alpha)\mathbf{x}^T\mathbf{A}\mathbf{y} + ((1-\alpha) - (1-\alpha)^2)\mathbf{y}^T\mathbf{A}\mathbf{y} &\geq 0 \\ \alpha(1-\alpha)(\mathbf{x}^T\mathbf{A}\mathbf{x} - 2\mathbf{x}^T\mathbf{A}\mathbf{y} + \mathbf{y}^T\mathbf{A}\mathbf{y}) &\geq 0.\end{aligned}$$

  Výraz $\alpha(1-\alpha)$ je pro každé $0 \leq \alpha \leq 1$ nezáporný. Nezápornost výrazu (16.5) jsme již ukázali. ♦

16.3 Operace zachovávající konvexitu funkcí

Operace zachovávající konvexitu funkcí umožňují z jednoduchých konvexních funkcí získat složitější. Konvexitu složitější funkce je často snadnější dokázat pomocí těchto operací než z podmínky (16.1) nebo Vět 16.3 a 16.5. Dále uvedeme příklady takových operací.

16.3.1 Nezáporná lineární kombinace

Tvrzení 16.6. Nechť $g_1, \ldots, g_k \colon \mathbb{R}^n \to \mathbb{R}$ jsou konvexní funkce a nechť $\alpha_1, \ldots, \alpha_k \geq 0$. Pak funkce $f = \alpha_1 g_1 + \cdots + \alpha_k g_k$ je konvexní.

Důkaz je snadný z podmínky (16.1) (ponecháváme jako cvičení). Speciálně, součet konvxních funkcí je konvexní funkce.

Naopak to ale neplatí: může se stát, že f nebo g nejsou konvexní a $f+g$ konvexní je. Např. funkce x^3 není konvexní, ale funkce $x^3 - x^3$ (tedy konstantní nulová funkce) konvexní je.

16.3.2 Skládání funkcí

Složení konvexních funkcí nemusí být konvexní funkce.

Příklad 16.6. Funkce $f, g \colon \mathbb{R} \to \mathbb{R}$ dané jako $f(x) = x^2$ a $g(x) = -x$ jsou konvexní. Funkce $(g \circ f)(x) = g(f(x)) = -x^2$ není konvexní. ♦

Příklad 16.7. Funkce $f, g \colon \mathbb{R} \to \mathbb{R}$ dané jako $f(x) = |x-1|$ a $g(x) = |x|$ jsou konvexní. Funkce $(g \circ f)(x) = g(f(x)) = \big||x| - 1\big|$ není konvexní (nakreslete si její graf!). ♦

Věta 16.7. Nechť funkce $f \colon \mathbb{R} \to \mathbb{R}$ je konvexní. Nechť funkce $g \colon \mathbb{R} \to \mathbb{R}$ je konvexní a neklesající. Pak složená funkce $g \circ f$ (daná předpisem $(g \circ f)(x) = g(f(x))$) je konvexní.

Důkaz. Pro každé $x, y \in \mathbb{R}$ a $0 \leq \alpha \leq 1$ máme

$$g(f((1-\alpha)x + \alpha y)) \leq g((1-\alpha)f(x) + \alpha f(y)) \leq (1-\alpha)g(f(x)) + \alpha g(f(y)).$$

První nerovnost platí, protože f je konvexní a g je neklesající. Druhá nerovnost platí, protože g je konvexní. ■

Obecněji můžeme zkoumat složené zobrazení $(g \circ \mathbf{f})(\mathbf{x}) = g(\mathbf{f}(\mathbf{x}))$, kde $\mathbb{R}^n \xrightarrow{\mathbf{f}} \mathbb{R}^m \xrightarrow{g} \mathbb{R}$. Existuje analogie Věty 16.7 pro tento případ, je ale dosti komplikovaná a nebudeme ji uvádět. Uvedeme jen důležitý případ, kdy $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$ je afinní zobrazení.

Věta 16.8. Nechť funkce $g\colon \mathbb{R}^m \to \mathbb{R}$ je konvexní. Nechť $\mathbf{A} \in \mathbb{R}^{m\times n}$ a $\mathbf{b} \in \mathbb{R}^m$. Pak funkce $h(\mathbf{x}) = g(\mathbf{A}\mathbf{x} + \mathbf{b})$ je konvexní.

Důkaz. Pro každé $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ a $0 \le \alpha \le 1$ platí

$$\begin{aligned} h((1-\alpha)\mathbf{x} + \alpha\mathbf{y}) &= g(\mathbf{A}[(1-\alpha)\mathbf{x} + \alpha\mathbf{y}] + \mathbf{b}) \\ &= g((1-\alpha)(\mathbf{A}\mathbf{x} + \mathbf{b}) + \alpha(\mathbf{A}\mathbf{y} + \mathbf{b})) \\ &\le (1-\alpha)g(\mathbf{A}\mathbf{x} + \mathbf{b}) + \alpha g(\mathbf{A}\mathbf{y} + \mathbf{b}) \\ &= (1-\alpha)h(\mathbf{x}) + \alpha h(\mathbf{y}). \end{aligned}$$

■

Příklad 16.8. Mějme funkci $f\colon \mathbb{R}^{2n} \to \mathbb{R}$ s hodnotami $f(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$, kde $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ a $\|\cdot\|\colon \mathbb{R}^n \to \mathbb{R}$ je libovolná norma. Tato funkce je konvexní funkce argumentu $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{2n}$. Ve Větě 16.8 vezmeme $\mathbf{A} = \begin{bmatrix}\mathbf{I} & -\mathbf{I}\end{bmatrix} \in \mathbb{R}^{n\times(2n)}$ a $\mathbf{b} = \mathbf{0}$. ♦

16.3.3 Maximum

Nejzajímavější operace zachovávající konvexitu funkcí je ovšem maximum.

Věta 16.9. Nechť I je libovolná množina a $g_i\colon \mathbb{R}^n \to \mathbb{R}$, $i \in I$, jsou konvexní funkce. Pak

$$f(\mathbf{x}) = \max_{i\in I} g_i(\mathbf{x}) \tag{16.6}$$

je konvexní funkce, kde předpokládáme, že pro každé $\mathbf{x}$ maximum existuje[4].

Důkaz. Epigraf funkce f průnik epigrafů funkcí g_i, neboť

$$\begin{aligned} \{\, (\mathbf{x}, y) \in \mathbb{R}^{n+1} \mid \mathbf{x} \in \mathbb{R}^n,\ \max_{i\in I} g_i(\mathbf{x}) \le y \,\} &= \{\, (\mathbf{x}, y) \in \mathbb{R}^{n+1} \mid \mathbf{x} \in \mathbb{R}^n,\ g_i(\mathbf{x}) \le y\ \forall i \in I \,\} \\ &= \bigcap_{i\in I} \{\, (\mathbf{x}, y) \in \mathbb{R}^{n+1} \mid \mathbf{x} \in \mathbb{R}^n,\ g_i(\mathbf{x}) \le y \,\}, \end{aligned}$$

kde jsme využili ekvivalence (12.7). Protože funkce g_i jsou konvexní, dle Věty 16.1 jsou jejich epigrafy konvexní množiny. Dle Tvrzení 13.1 je průnik konvexních množin konvexní množina. Tedy epigraf funkce (16.6) je konvexní množina. Dle Věty 16.1 je tedy funkce f konvexní. ■

Příklad 16.9. Konvexitu funkce $f(\mathbf{x}) = \max_{i=1}^n x_i$ jsme již v Příkladu 16.1 dokázali z podmínky (16.1). Ovšem je mnohem pohodlnější použít Větu 16.9. Máme $g_i(\mathbf{x}) = x_i$. Funkce g_i jsou lineární, tedy konvexní. Takže funkce $f(\mathbf{x}) = \max_{i=1}^n g_i(\mathbf{x})$ je konvexní. ♦

Příklad 16.10. Funkce

$$f(\mathbf{x}) = \max_{i=1}^{k}(\mathbf{a}_i^T\mathbf{x} + b_i)$$

je maximem afinních funkcí. Tuto funkci jsme již potkali v §12.1.1. Protože afinní funkce jsou konvexní, je i jejich maximum konvexní. ♦

[4]Pokud pro nějaké $\mathbf{x}$ množina $\{\, g_i(\mathbf{x}) \mid i \in I \,\}$ nemá největší prvek (což se může stát jen tehdy, je-li množina I nekonečná), můžeme maximum v (16.6) nahradit supremem a věta stále platí.

Příklad 16.11. Nechť $C \subseteq \mathbb{R}^n$ je libovolná (ne nutně konvexní) množina. Funkce

$$f(\mathbf{x}) = \max_{\mathbf{y} \in C} \|\mathbf{x} - \mathbf{y}\|$$

udává vzdálenost bodu $\mathbf{x}$ od nejvzdálenějšího bodu množiny C (zde předpokládáme, že maximum existuje). Dle Věty 16.8 je pro každé pevné $\mathbf{y}$ výraz $\|\mathbf{x} - \mathbf{y}\|$ konvexní funkcí $\mathbf{x}$. Tedy výraz $\|\mathbf{x} - \mathbf{y}\|$ lze chápat jako množinu konvexních funkcí $\mathbf{x}$ indexovaných indexem $\mathbf{y}$ (můžeme označit $\|\mathbf{x} - \mathbf{y}\| = g_{\mathbf{y}}(\mathbf{x})$). Jelikož f je maximem těchto funkcí, je i funkce f konvexní. ♦

Příklad 16.12. Mějme funkci

$$f(\mathbf{c}) = \max\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{A}\mathbf{x} \geq \mathbf{b} \,\},$$

která vyjadřuje závislost optimální hodnoty daného lineárního programu na vektoru $\mathbf{c}$ (viz §12). Máme $f(\mathbf{c}) = \max_{\mathbf{x} \in X} \mathbf{c}^T\mathbf{x}$ a $X = \{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} \geq \mathbf{b} \,\}$ (zde předpokládáme, že pro každé $\mathbf{c}$ maximum existuje, neboli množina X je neprázdná a omezená). Je-li $\mathbf{x}$ pevné, je $\mathbf{c}^T\mathbf{x}$ lineární funkce vektoru $\mathbf{c}$. Funkce f je tedy maximum nekonečného množství lineárních funkcí, tedy je konvexní. ♦

Příklad 16.13. Nechť $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^m$, $b_1, \dots, b_n \in \mathbb{R}$ a $\mathbf{w} = (w_1, \dots, w_n) \in \mathbb{R}^n$ je vektor nezáporných vah. Přibližné řešení soustavy $\mathbf{a}_i^T\mathbf{x} = b_i$, $i = 1, \dots, n$, ve smyslu *vážených nejmenších čtverců* (viz Cvičení 5.4) znamená vypočítat

$$f(\mathbf{w}) = \min_{\mathbf{x} \in \mathbb{R}^m} \sum_{i=1}^{n} w_i(\mathbf{a}_i^T\mathbf{x} - b_i)^2,$$

kde jsme označili hodnotu výsledného minima jako funkci vektoru vah. Funkce f je konkávní, protože je minimem lineárních funkcí. ♦

16.4 Cvičení

16.1. Pro každou funkci $f\colon \mathbb{R}^n \to \mathbb{R}$ dokažte z podmínky (16.1), které z těchto čtyř tvrzení platí (a pro jaké n): funkce je konvexní, konkávní, konvexní i konkávní, ani konvexní ani konkávní.

a) $f(\mathbf{x}) = \mathbf{a}^T\mathbf{x} + b$
b) $f(\mathbf{x}) = \mathbf{x}^T\mathbf{x}$
c) $f(\mathbf{x}) =$ aritmetický průměr čísel $x_1, \dots, x_n$
d) $f(\mathbf{x}) = \operatorname{median}_{i=1}^n x_i$ (medián čísel $x_1, \dots, x_n$)
e) $f(\mathbf{x}) = \min_{i=1}^n |x_i|$
f) $f(\mathbf{x}) =$ součet dvou nejmenších čísel z čísel $x_1, \dots, x_n$.

16.2. Dokažte konvexitu či konkavitu funkcí z Příkladu 16.4. Můžete použít podmínku (16.1) a věty z §16.2 a §16.3.

16.3. Pro každou funkci dokažte, které z těchto čtyřech tvrzení platí: funkce je konvexní, konkávní, konvexní i konkávní, ani konvexní ani konkávní. Můžete použít podmínku (16.1) a věty z §16.2 a §16.3.

a) $f(x) = \mathrm{e}^{x^2}$

b) $f(x) = \mathrm{e}^{-x^2}$
c) $f(x, y) = |x - y|$
d) $f(x, y) = -y$
e) $f(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$
f) $f(\mathbf{x}) = \sum_{i=1}^n x_i \ln x_i$ na množině $\mathbb{R}^n_{++}$
g) $f(\mathbf{x}) = \sum_{i=1}^k \ln(b_i - \mathbf{a}_i^T\mathbf{x})$ na množině $X = \{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}_i^T\mathbf{x} < b_i \,, i = 1, \ldots, k \,\}$
h) $f(\mathbf{x}) = \min_{i=1}^n x_i$
i) $f(\mathbf{x}) = \max_{i=1}^n x_i - \min_{i=1}^n x_i$
j) $f(\mathbf{x}) = \max_{i=1}^n x_i + \min_{i=1}^n x_i$
k) $f(\mathbf{x}) = \max\Big\{ \Big|\sum_{i=1}^n x_i - 1\Big|, \sum_{i=1}^n |x_i| - 1 \Big\}$
l) ($\star$) $f(\mathbf{x})$ je součet k největších složek $x_1, \ldots, x_n$ vektoru $\mathbf{x}$ (kde $k \leq n$ je dáno)

16.4. Máme funkci jedné proměnné $f(x) = (x^2 - a)^2$. Pro jaké hodnoty parametru $a \in \mathbb{R}$ je tato funkce konvexní? Načrtněte graf funkce pro nějaké a, pro které funkce není konvexní.

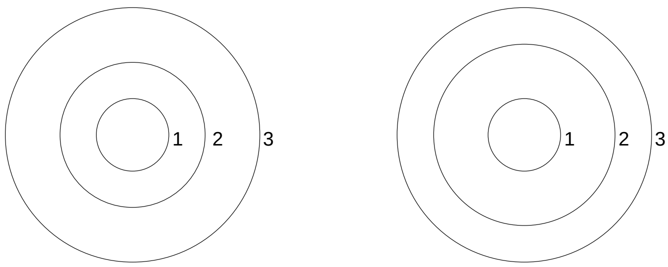
16.5. Může být součet nekonvexních funkcí konvexní funkce? Najděte protipříklad. Je to v rozporu s Tvrzením 16.6?

16.6. Robustní prokládání přímky množinou bodů $(\mathbf{x}_i, y_i) \in \mathbb{R}^n \times \mathbb{R}$ (pro $i = 1, \ldots, m$) s pásmem necitlivosti $\theta > 0$ vyžaduje minimalizaci funkce

$$f(\mathbf{a}, b) = \sum_{i=1}^m \max\{\, -\mathbf{a}^T\mathbf{x}_i + b + y_i - \theta,\ 0,\ \mathbf{a}^T\mathbf{x}_i + b - y_i - \theta \,\},$$

kde $\mathbf{a} \in \mathbb{R}^n$ a $b \in \mathbb{R}$. Dokažte, že $f(\mathbf{a}, b)$ je konvexní funkce. Srov. Cvičení 12.17.c.

16.7. Každý z obrázků zobrazuje některé vrstevnice funkce dvou proměnných a jejich výšky. Je možné, aby funkce, která má tyto vrstevnice, byla konvexní? Dokažte z podmínky (16.1).



16.8. Co je subkontura výšky 2 funkce jedné proměnné $f(x) = x^2 - x$?

16.9. Dokažte, že elipsoid $\{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{x}^T\mathbf{A}\mathbf{x} \leq 1 \,\}$ (kde $\mathbf{A}$ je positivně definitní, viz §6.4.2) je konvexní množina.

16.10. Dokažte, že množina $\{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} \geq \mathbf{0},\ \prod_{i=1}^n x_i \geq 1 \,\}$ je konvexní (symbol $\prod_i x_i$ značí součin čísel $x_1, \ldots, x_n$).

16.11. Dokažte, že následující funkce jsou nekonvexní:

a) Účelová funkce z Příkladu 10.12
b) Účelová funkce z Cvičení 10.13

16.12. ($\star$) Dokažte, že je-li funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ zároveň konvexní i konkávní, pak je nutně afinní.

Nápověda a řešení

16.1.a) Konvexní i konkávní, nerovnost (16.1) platí s rovností.

16.1.b) Je konvexní, není konkávní.

16.1.c) Konvexní i konkávní, nerovnost (16.1) platí s rovností.

16.1.d) Pro $n \leq 2$ konvexní i konkávní, pro $n > 2$ ani konvexní ani konkávní.

16.1.e) Pro $n = 1$ je funkce konvexní. Pro $n = 2$ není konvexní, neboť (16.1) není splněna např. pro $\mathbf{x} = (1, 0)$, $\mathbf{y} = (0, 1)$, $\alpha = \frac{1}{2}$. Pro $n > 2$ také není konvexní, což plyne z nekonvexity pro $n = 2$, protože můžeme zvolit $\mathbf{x} = (1, 0, 0, 0, \ldots)$ a $\mathbf{y} = (0, 1, 1, 1, \ldots)$.

Funkce není konkávní pro žádné n. Pro $n = 1$ dokážeme z (16.1) volbou $\mathbf{x} = -1$, $\mathbf{y} = 1$, $\alpha = \frac{1}{2}$. Pro $n > 1$ můžeme vektory opět doplnit opakováním poslední číslice.

16.1.f) Dokážeme, že pro $n = 3$ funkce není konvexní. Vezmeme $\mathbf{x} = (1, 2, 3)$, $\mathbf{y} = (3, 2, 1)$, $\alpha = \frac{1}{2}$. Pak $f(\alpha\mathbf{x} + (1 - \alpha)\mathbf{y}) = f(2, 2, 2) = 4 \not\leq (3 + 3)/2 = \frac{1}{2}f(1, 2, 3) + \frac{1}{2}f(3, 2, 1) = \frac{1}{2}f(\mathbf{x}) + \frac{1}{2}f(\mathbf{y})$.

16.3.a) Nejrychleji asi pomocí Věty 16.5: druhá derivace $f''(x) = 2(2x^2 + 1)e^{x^2}$ kladná na celém $\mathbb{R}$, takže funkce je konvexní a není konkávníProtože jde o funkci jen jedné proměnné, můžeme si její graf snadno vykreslit např. v Matlabu a uvidíme, že je funkce konvexní (i když důkaz to není).

16.3.b) Funkce je známý Gaussián (až na normalizační konstantu). Zde ovšem druhá derivace $f''(x) = 2(2x^1 - 1)e^{-x^2}$ je někde kladná a někde záporná, takže není konvexní ani konkávní (na $\mathbb{R}$). Přesněji bychom mohli říct (viz footnote k Větě 16.5), že na intervalech $(-\infty, 1/\sqrt{2})$ a $(1/\sqrt{2}, +\infty)$ je funkce konvexní (protože tam je druhá derivace kladná) a na intervalu $(-1/\sqrt{2}, 1/\sqrt{2})$ je konkávní (druhá derivace je tam záporná).

16.3.c) Jde o složení konvexní funkce $g(z) = |z|$ a lineární funkce $z = h(x, y) = x - y$, tedy f je konvexní dle Věty 16.8.

16.3.d) Je to lineární funkce dvou proměnných (nevadí, že nezávisí na x), tedy kovexní i konkávní.

16.3.e) Toto je účelová funkce lineární úlohy nejmenších čtverců, kterou dobře znáte. Měli byste ihned uhodnout, že je konvexní. Formálně to lze dokázat různě. Buď spočítat hessián $f''(\mathbf{x}) = 2\mathbf{A}^T\mathbf{A}$ a vidět, že je positivně semidefinitní. Nebo si uvědomit, že funkce f je kvadratická a zdola omezená (nulou), tedy musí mít positivně semidefinitní hessián.

16.3.f) Funkci lze psát jako $f(\mathbf{x}) = \sum_i g(x_i)$ kde $g(x) = x \ln x$. Funkce g je konvexní na $\mathbb{R}^n_{++}$, neboť $g''(x) = 1/x$. Funkce f je tedy konvexní dle Věty 16.6.

16.3.g) Jde o součet konkávních funkcí, takže je konkávní. Logaritmus je konkávní funkce, tedy funkce $\ln(b_i - \mathbf{a}_i^T\mathbf{x})$ jsou konkávní dle Věty 16.8 (s uvážením, že funkce f je konkávní, právě když $-f$ je konvexní).

16.3.h) Funkci můžeme napsat jako $f(\mathbf{x}) = \min_i x_i = -\max_i(-x_i)$. Funkce $\max_i(-x_i)$ je konvexní funkce vektoru $\mathbf{x}$, protože je to složení konvexní funkce $\max_i x_i$ (což víme z Příkladů 16.1 a 16.9) a lineární funkce $-\mathbf{x}$. Tedy funkce $-\max_i(-x_i)$ je konkávní.

16.3.i) Funkce je konvexní, protože je to rozdíl konvexní a konkávní funkce (tedy součet dvou konvexních funkcí). Není konkávní, např. proto, že jediná funkce, která je zároveň konvexní i konkávní, je afinní funkce.

16.3.j) Zde byste měli uhodnout, že funkce není konvexní ani konkávní, protože je součtem konvexní a konkávní funkce. To ovšem jednak není důkaz, jednak to není pravda pro malá n.

Pro $n = 1$ se funkce zjednoduší na $f(x) = 2x$, která je lineární, tedy konvexní i konkávní.

Pro $n = 2$ je $f(\mathbf{x}) = f(x_1, x_2) = \max\{x_1, x_2\} + \min\{x_1, x_2\}$. Překvapivě, tuto funkci lze také zjednodušit: pro $x_1 \geq x_2$ je $f(x_1, x_2) = x_1 + x_2$ a pro $x_1 \leq x_2$ je $f(x_1, x_2) = x_2 + x_1$, tedy $f(x_1, x_2) = x_1 + x_2$ na celém $\mathbb{R}^2$. Tato funkce je opět konvexní i konkávní.

Pro $n = 3$ funkce není ani konvexní ani konkávní. Dokázat se to musí z definice konvexní funkce: stačí najít jediné $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ a $0 < \alpha < 1$ tak, aby neplatila nerovnost (16.1). Můžeme to udělat buď na počítači, kde generujeme náhodné trojice $(\mathbf{x}, \mathbf{y}, \alpha)$ a kontrolujeme nerovnost (16.1). Bez počítače trojici musíme uhodnout, což může být dost dřina. Funguje např. $\mathbf{x} = (1, 0, -1)$, $\mathbf{y} = (1, -1, 0)$, $\alpha = 1/2$: pak je $f((1-\alpha)\mathbf{x} + \alpha\mathbf{y}) = 1/2$ a $(1-\alpha)f(\mathbf{x}) + \alpha f(\mathbf{y}) = 0$.

Pro $n > 3$ je situace stejná jako pro $n = 3$, neboť vektory $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$ vždy můžeme doplnit dalšími složkami tak, že se maxima ani minima nezmění (zkopírováním poslední složky, např. pro $n = 5$ z vektoru $(1, 0, -1)$ uděláme $(1, 0, -1, -1, -1)$).

16.3.k) Je konvexní. Absolutní hodnota je konvexní funkce, jejich součet také, maximum konvexních funkcí je konvexní funkce.

16.7. V podmínce (16.1) zvolte $\mathbf{x}, \mathbf{y}$ na vrstevnicích výšky 1 a 3. Zvolte chytře α. Odpovědi: ne, ano.

16.8. Interval $[-1, 2]$.

16.9. Funkce $f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x}$ je konvexní, viz Příklad 16.5. Elipsoid je subkontura této funkce, tedy je to konvexní množina.

16.10. Zlogaritmování obou stran podmínky dá $\{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} \geq \mathbf{0},\ \prod_{i=1}^n x_i \geq 1 \,\} = \{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} \geq \mathbf{0},\ \sum_{i=1}^n \ln x_i \geq 0 \,\}$ (logaritmus je rostoucí funkce a tedy nemění znaménko nerovnosti). Podmínka $-\sum_{i=1}^n \ln x_i \leq 0$ definuje konvexní množinu, neboť funkce $f(\mathbf{x}) = -\sum_i \ln x_i$ je na množině $\mathbb{R}^n_{++}$ konvexní a proto každá její subkontura (zde nulové výšky) je konvexní množina dle Věty 16.2.

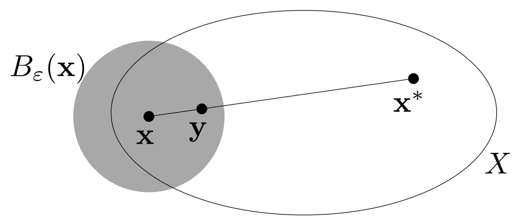
16.12. Začněte porovnáním Jensenovu nerovnost (16.2) a definici afinní funkce v §3.3.

Kapitola 17

Konvexní optimalizační úlohy

Věta 17.1. Nechť funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ je konvexní na konvexní množině $X \subseteq \mathbb{R}^n$. Pak každé lokální minimum funkce f na množině X je zároveň globální.

Důkaz. Nechť $\mathbf{x}$ je lokálním minimem f na X, viz obrázek:



Dle definice lokálního minima (viz §9.3) tedy existuje $\varepsilon > 0$ tak, že $f(\mathbf{x}) \le f(\mathbf{y})$ pro všechna $\mathbf{y} \in B_\varepsilon(\mathbf{x}) \cap X$. Nechť ale $\mathbf{x}$ není globální minimum, tedy existuje $\mathbf{x}^* \in X$ tak, že $f(\mathbf{x}^*) < f(\mathbf{x})$. Ukážeme, že to vede ke sporu. Můžeme totiž zvolit $0 < \alpha < 1$ tak, že bod $\mathbf{y} = (1-\alpha)\mathbf{x} + \alpha\mathbf{x}^*$ leží v $B_\varepsilon(\mathbf{x})$. Protože je množina X konvexní, leží bod $\mathbf{y}$ zároveň i v X. Je

$$f(\mathbf{y}) \;=\; f((1-\alpha)\mathbf{x} + \alpha\mathbf{x}^*) \;\le\; (1-\alpha)f(\mathbf{x}) + \alpha f(\mathbf{x}^*) \;<\; (1-\alpha)f(\mathbf{x}) + \alpha f(\mathbf{x}) \;=\; f(\mathbf{x}).$$

Ale tvrzení $f(\mathbf{y}) < f(\mathbf{x})$ je ve sporu s předpokladem, že $\mathbf{x}$ je lokální minimum. ■

Minimalizaci konvexní funkce na konvexní množině se říká **konvexní optimalizační úloha**. Pro takovou úlohu nám tedy stačí najít libovolné lokální minimum, abychom našli globální minimum.

17.1 Příklady nekonvexních úloh

Najít globální minimum funkce na množině je obvykle mnohem těžší než najít *nějaké* lokální minimum. Mohli bychom si myslet, že globální minimum najdeme tak, že najdeme všechna lokální minima a vybereme to, pro které je účelová funkce nejmenší. Problém je v tom, že nekonvexní úloha může mít lokálních minim velmi mnoho.

Příklad 17.1. Zopakujme úlohu (7.1): pro danou čtvercovou matici $\mathbf{A}$ maximalizujeme (či minimalizujeme) funkci $f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x}$ na množině $\{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{x}^T\mathbf{x} = 1 \,\}$. Tato úloha neni konvexní, neboť množina přípustných řešení není konvexní (je to n-rozměrná sféra). My ale víme, že globální optimum úlohy lze najít pomocí spektrálního rozkladu. ♦

Zde jsme měli štěstí: najít globální optimum úlohy bylo snadné, i když úloha nebyla konvexní. To je ale výjimka – typicky je nalezení globálního minima nekonvexní úlohy velmi těžké.

Příklad 17.2. Uveďme příklad, na kterém bude na první pohled vidět, že nekonvexní úloha může mít velmi mnoho lokálních minim. Řešme úlohu

$$\max\{\, \mathbf{x}^T\mathbf{x} \mid \mathbf{x} \in [-1,1]^n \,\}, \tag{17.1}$$

tedy maximalizujeme konvexní funkci $\mathbf{x}^T\mathbf{x}$ na hyperkrychli $[-1,1]^n$. Je jasné (nakreslete si obrázek pro $n = 2$, tedy pro čtverec!), že funkce má lokální maximum v každém vrcholu hyperkrychle. Jelikož hyperkrychle má 2^n vrcholů, úloha má 2^n lokálních maxim.

V tomto symetrickém případě globální maximum snadno najdeme úvahou. Uvažme však mírně obecnější úlohu

$$\min\{\, \mathbf{x}^T\mathbf{A}\mathbf{x} \mid \mathbf{x} \in [-1,1]^n \,\}, \tag{17.2}$$

kde $\mathbf{A} \in \mathbb{Z}^{n\times n}$ (tedy matice má celočíselné prvky). Je známo, že vyřešení (tj. nalezení globálního maxima) této úlohy pro větší n je prakticky nemožné (přesněji, je NP-těžké, srov. §12.5). ♦

Příklad 17.3. Podívejme se znovu na úlohu shlukování, Příklad 1.15 z úvodní kapitoly. Tam se minimalizuje funkce

$$f(\mathbf{x}) = f(\mathbf{x}_1,\dots,\mathbf{x}_n) = \sum_{i=1}^{m} \min_{j=1}^{n} \|\mathbf{a}_i - \mathbf{x}_j\| \tag{17.3}$$

přes vektory $\mathbf{x}_1,\dots,\mathbf{x}_n \in \mathbb{R}^d$. Máme $f\colon \mathbb{R}^{dn} \to \mathbb{R}$, tedy vlastně minimalizujeme přes jediný vektor $\mathbf{x} = (\mathbf{x}_1,\dots,\mathbf{x}_n) \in \mathbb{R}^{dn}$.

Je funkce (17.3) konvexní? Pro každé i je $\|\mathbf{a}_i - \mathbf{x}_j\|$ funkce vektoru $\mathbf{x}_j$, tedy i vektoru $\mathbf{x}$. Ale funkce $\min_{j=1}^n \|\mathbf{a}_i - \mathbf{x}_j\|$ již konvexní být nemusí (Větu 16.9 nelze použít, ta hovoří o *maximu* konvexních funkcí). Tedy ani funkce f, která je jejich součtem, nemusí být konvexní.

Tím, že se nám nepodařilo dokázat konvexitu funkce f, jsme samozřejmě nedokázali její nekonvexitu. To lze udělat následovně. Vezměme jednoduchý případ $d = 1$, $m = 1$, $n = 2$, $a_1 = 0$. Pak (1.24) má tvar $f(x_1,x_2) = \min\{|x_1|,|x_2|\}$. Tato funkce není konvexní, protože např. její řez $\varphi(t) = f(t,t-1) = \min\{|t|,|t-1|\}$ není konvexní (nakreslete si graf funkce φ). Bez důkazu uveďme, že funkce není konvexní ani pro větší d,m,n. To nás nepřekvapí, protože, jak jsme řekli Příkladu 1.15, že minimalizace funkce (17.3) je NP-těžké. ♦

Příklad 17.4. Úloha celočíselného programování (12.23) je nekonvexní, protože množina $\{\, \mathbf{x} \in \{0,1\}^n \mid \mathbf{A}\mathbf{x} \geq \mathbf{b} \,\}$ jejích přípustných řešení je nekonvexní (srov. Cvičení 13.1.g). ♦

17.2 Konvexní optimalizační úloha ve standardním tvaru

Uvažujme nyní obecnou úlohu spojité optimalizace ve standarním tvaru (1.11),

$$\min\{\, f(\mathbf{x}) \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{g}(\mathbf{x}) \leq \mathbf{0},\ \mathbf{h}(\mathbf{x}) = \mathbf{0} \,\} \tag{17.4}$$

neboli

$$\begin{aligned} \min \quad & f(x_1,\dots,x_n) \\ \text{za podmínek} \quad & g_i(x_1,\dots,x_n) \leq 0, \quad i = 1,\dots,m \\ & h_i(x_1,\dots,x_n) = 0, \quad i = 1,\dots,l \end{aligned}$$

kde $f\colon \mathbb{R}^n \to \mathbb{R}$, $(g_1,\dots,g_m) = \mathbf{g}\colon \mathbb{R}^n \to \mathbb{R}^m$, $(h_1,\dots,h_l) = \mathbf{h}\colon \mathbb{R}^n \to \mathbb{R}^l$. Množina přípustných řešení této úlohy je konvexní, jestliže funkce $f, g_1,\dots,g_m$ jsou konvexní a funkce $h_1,\dots,h_l$ jsou afinní (tedy zobrazení $\mathbf{h}$ je afinní). Tato množina je totiž průnik množin $\{\, \mathbf{x} \in \mathbb{R}^n \mid g_i(\mathbf{x}) \le 0 \,\}$ (které jsou konvexní, neboť jsou to subkontury konvexní funkce g_i) a $\{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{h}(\mathbf{x}) = \mathbf{0} \,\}$ (což je afinní podprostor, tedy také konvexní).

Podmínka, že funkce $g_1,\dots,g_m$ jsou konvexní a zobrazení $\mathbf{h}$ je afinní, je postačující ale nikoliv nutná pro konvexitu množiny přípustných řešení.

Příklad 17.5. Platí (promyslete!)

$$\{\, (x,y) \in \mathbb{R}^2 \mid x/(1+y^2) \le 0,\ (x+y)^2 = 0 \,\} \;=\; \{\, (x,y) \in \mathbb{R}^2 \mid x \le 0,\ x+y = 0 \,\}.$$

Tedy máme dva různé popisy stejné množiny. V prvním tvaru funkce $g(x,y) = x/(1+y^2)$ není konvexní a funkce $h(x,y) = (x+y)^2$ není afinní. Přesto je množina konvexní, což je vidět ze druhého tvaru. ♦

Úloze tvaru (17.4), ve které jsou funkce $f, g_1,\dots,g_m$ konvexní a zobrazení $\mathbf{h}$ afinní, říkáme **konvexní optimalizační úloha ve standardním tvaru**.

17.3 Ekvivalentní transformace úlohy

Dvě úlohy ve tvaru (17.4) nazveme **ekvivalentní**, když se z množiny optimálních řešení jedné dá ‘snadno’ získat množina optimálních řešení druhé a naopak. **Ekvivalentní transformace** je pak každá transformace úlohy, jejímž výsledkem je úloha ekvivalentní. Dále uvedeme příklady ekvivalentních transformací. U každé poznamenáme, zda zachovává konvexitu úlohy:

- *Změna proměnných.* Nechť $\boldsymbol{\varphi}\colon \mathbb{R}^n \to \mathbb{R}^n$ je bijektivní zobrazení (viz §1.1.2). Pak úloha (17.4) je ekvivalentní úloze

  $$\min\{\, f(\boldsymbol{\varphi}(\mathbf{x})) \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{g}(\boldsymbol{\varphi}(\mathbf{x})) \le \mathbf{0},\ \mathbf{h}(\boldsymbol{\varphi}(\mathbf{x})) = \mathbf{0} \,\}.$$

  Tato transformace nemusí zachovat konvexitu úlohy (viz §16.3.2).

- *Monotónní transformace účelové funkce.* Nechť $\psi\colon \mathbb{R} \to \mathbb{R}$ je rostoucí funkce. Pak

  $$\operatorname{argmin}\{\, f(\mathbf{x}) \mid \mathbf{x} \in X \,\} = \operatorname{argmin}\{\, \psi(f(\mathbf{x})) \mid \mathbf{x} \in X \,\}.$$

  Tato transformace nemusí zachovat konvexitu funkce f.

  Příklad 17.6. Tuto transformaci jsme již několikrát použili v nejmenších čtvercích. Máme minimalizovat např. funkci $f(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2$, ale zvolíme $\psi(y) = y^2$ a minimalizujeme funkci $\psi(f(\mathbf{x})) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 = (\mathbf{A}\mathbf{x} - \mathbf{b})^T(\mathbf{A}\mathbf{x} - \mathbf{b})$. Nová funkce má výhodu, že je na rozdíl od staré diferencovatelná, a to při zachování konvexity. ♦

- *Slackové proměnné.* Podobně jako v LP (viz §12.1), úloha (17.4) je ekvivalentní úloze

  $$\min\{\, f(\mathbf{x}) \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{s} \in \mathbb{R}^m,\ \mathbf{s} \ge \mathbf{0},\ \mathbf{g}(\mathbf{x}) + \mathbf{s} = \mathbf{0},\ \mathbf{h}(\mathbf{x}) = \mathbf{0} \,\}.$$

  Tato transformace zachová konvexitu úlohy jen v případě, kdy $\mathbf{g}(\mathbf{x}) + \mathbf{s}$ je afinní zobrazení vektoru $(\mathbf{x}, \mathbf{s})$, tedy kdy zobrazení $\mathbf{g}$ je afinní,

- *Epigrafový tvar.* Platí (už jsme to viděli v (12.5))

$$\min\{\, f(\mathbf{x}) \mid \mathbf{x} \in X \,\} = \min\{\, y \mid \mathbf{x} \in X,\ y \in \mathbb{R},\ f(\mathbf{x}) - y \leq 0 \,\}.$$

  V nové úloze vlastně hledáme nejnižší bod epigrafu funkce f na množině přípustných řešení. Plyne z toho, že každou optimalizační úlohu lze převést na úlohu s lineární účelovou funkcí. Tato transformace zachovává konvexitu úlohy: je-li X konvexní množina a f konvexní funkce na X, pak funkce $f(\mathbf{x}) - y$ proměnných $(\mathbf{x}, y)$ je konvexní na množině $X \times \mathbb{R}$.

17.4 Třídy konvexních optimalizačních úloh

Optimalizační úlohy ve tvaru (17.4) se taxonomizují podle druhu funkcí f, g_i, h_i. Pro každou třídu existují specializované algoritmy schopné najít lokální minimum (v případě konvexní úlohy tedy globální minimum). Přehled dostupných implementací takových algoritmů je možno najít např. na http://www.neos-guide.org.

17.4.1 Lineární programování (LP)

V *lineárním programování* jsou všechny funkce f, g_i, h_i afinní. Jde tedy v jistém smyslu o nejjednoduší případ konvexní optimalizační úlohy. Přesto jsme viděli v Kapitole 12, že již tento jednoduchý případ má velmi mnoho aplikací.

17.4.2 Kvadratické programování (QP)

V *kvadratickém programování* jsou funkce g_i, h_i afinní a funkce f je kvadratická. Tedy je to úloha

$$\begin{aligned} \min \quad & \mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} \\ \text{za podm.} \quad & \mathbf{C}\mathbf{x} \leq \mathbf{d}, \\ & \mathbf{E}\mathbf{x} = \mathbf{f}, \end{aligned}$$

kde $\mathbf{A} \in \mathbb{R}^{n\times n}$, $\mathbf{b} \in \mathbb{R}^n$, $\mathbf{C} \in \mathbb{R}^{m\times n}$, $\mathbf{d} \in \mathbb{R}^m$, $\mathbf{E} \in \mathbb{R}^{l\times n}$, $\mathbf{f} \in \mathbb{R}^l$. Úloha je to konvexní, právě když funkce f je navíc konvexní, neboli matice $\mathbf{A}$ je positivně semidefinitní (viz Příklad 16.5).

Příklad 17.7. Přibližné řešení přeurčené soustavy lineárních rovnic ve smyslu nejmenších čtverců (5.2) je konvexní úloha QP bez omezení, tj. f je kvadratická konvexní (jak jsme ukázali v Příkladě 6.15) a $m = l = 0$. Tuto úlohu lze převést na řešení soustavy lineárních rovnic (5.4). ♦

Příklad 17.8. Řešení soustavy lineárních rovnic s nejmenší normou (§5.4) nebo, obecněji, úloha nejmenších čtverců s omezeními typu rovnosti (Příklad 11.4) jsou příklady konvexního QP s omezeními typu rovnosti (tj. $\mathbf{A}$ je positivně semidefinitní, $m = 0$ a h_i jsou afinní). Tato úloha jde převést na řešení soustavy lineárních rovnic. ♦

Příklad 17.9. Obecněji, minimalizace kvadratické funkce za podmínek typu rovnosti (tj. $m = 0$, $l > 0$ a h_i jsou afinní) se dá převést na řešení soustavy lineárních rovnic, viz Cvičení 11.19. Je-li ovšem $\mathbf{A}$ indefinitní, musíme ověřit podmínky druhého řádu (§11.2.3). ♦

Příklad 17.10. Často je užitečné úlohu nejmenších čtverců (5.2) řešit za omezení $\mathbf{c} \le \mathbf{x} \le \mathbf{d}$, tj. každá proměnná x_j musí být v intervalu $[c_j, d_j]$ (tzv. *box constraints*). To vede na konvexní QP s omezeními typu nerovnosti (tj. $\mathbf{A}$ je positivně semidefinitní a $m > 0$). Tuto úlohu již nelze převést na řešení soustavy lineárních rovnic. ♦

Příklad 17.11. Čtverec vzdálenosti bodu $\mathbf{y} \in \mathbb{R}^n$ od konvexního mnohostěnu $\{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{Ax} \le \mathbf{b} \,\}$ je optimální hodnota kvadratického programu

$$\min\{\, \|\mathbf{y} - \mathbf{x}\|_2^2 \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{Ax} \le \mathbf{b} \,\}.$$ ♦

Příklad 17.12. Je dáno m bodů v $\mathbb{R}^n$, z nichž každý patří do jedné ze dvou tříd, označených -1 a 1. Jinými slovy, je dána množina dvojic $(\mathbf{x}_i, y_i) \in \mathbb{R}^n \times \{-1, 1\}$ pro $i = 1, \dots, m$. V úloze *lineární klasifikace* hledáme nadrovinu, která odděluje body z obou tříd. Tedy hledáme $\mathbf{a} \in \mathbb{R}^n$ a $b \in \mathbb{R}$ takové, aby

$$\begin{aligned} \mathbf{a}^T\mathbf{x}_i - b < 0 \quad & \text{pro } y_i = -1, \\ \mathbf{a}^T\mathbf{x}_i - b > 0 \quad & \text{pro } y_i = 1, \end{aligned}$$

což lze napsat jako

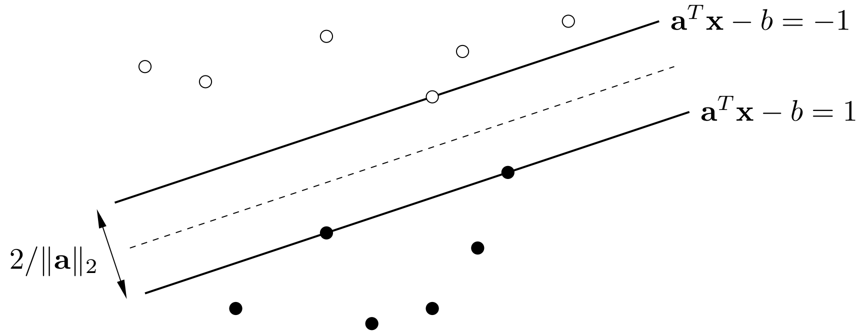
$$y_i(\mathbf{a}^T\mathbf{x}_i - b) > 0, \quad i = 1, \dots, m. \tag{17.5}$$

Označme $\varepsilon_i = y_i(\mathbf{a}^T\mathbf{x}_i - b)$ a vydělme vektor $(\mathbf{a}, b)$ kladným číslem $\min_{i=1}^m \varepsilon_i$. Pak soustavu (17.5) můžeme ekvivalentně psát jako

$$y_i(\mathbf{a}^T\mathbf{x}_i - b) \ge 1, \quad i = 1, \dots, m. \tag{17.6}$$

Hledáme-li libovolnou oddělující nadrovinu, stačí nám najít libovolné řešení soustavy nerovnic (17.6), což vede na úlohu LP.

Soustava ale navíc říká, že body jsou odděleny pásem $\{\, \mathbf{x} \in \mathbb{R}^n \mid -1 \ge \mathbf{a}^T\mathbf{x} - b \ge 1 \,\}$:



Snadno spočítáme (srov. Cvičení 5.16), že šířka pásu je $2/\|\mathbf{a}\|_2$. V úloze *support vector machine* (SVM) hledáme oddělující nadrovinu která maximalizuje šířku pásu, tedy minimalizuje $\|\mathbf{a}\|_2^2 = \mathbf{a}^T\mathbf{a}$ za podmínek (17.6). To je konvexní úloha QP. ♦

17.4.3 Kvadratické programování s kvadratickými omezeními (QCQP)

Obecnější variantou je *kvadratické programování s kvadratickými omezeními* (QCQP, *quadratically constrained quadratic programming*), kde funkce f, g_i jsou kvadratické a funkce h_i jsou afinní. Úloha je konvexní, právě když funkce f, g_i jsou navíc konvexní.

17.4.4 Programování na kuželu druhého řádu (SOCP)

V úloze *programování na kuželu druhého řádu* (SOCP, *second-order cone programming*) jsou funkce f, h_i afinní a funkce g_i mají tvar

$$g_i(\mathbf{x}) = \|\mathbf{A}_i\mathbf{x} + \mathbf{b}_i\|_2 - (\mathbf{c}_i^T\mathbf{x} + d_i). \tag{17.7}$$

Tedy úloha SOCP má tvar (pro jednoduchost neuvažujeme afinní omezení $h_i(\mathbf{x}) = 0$)

$$\begin{aligned} \min \quad & \mathbf{e}^T\mathbf{x} \\ \text{za podmínek} \quad & \|\mathbf{A}_i\mathbf{x} + \mathbf{b}_i\|_2 \leq \mathbf{c}_i^T\mathbf{x} + d_i, \quad i = 1, \ldots, m. \end{aligned}$$

Funkce g_i jsou konvexní (neboť norma je konvexní funkce, viz Příklad 16.4 a dále viz Věta 16.8). Podmínku $g_i(\mathbf{x}) \leq 0$ lze psát také jako $(\mathbf{A}_i\mathbf{x} + \mathbf{b}_i, \mathbf{c}_i^T\mathbf{x} + d_i) \in K_2^n$, kde konvexní množina

$$K_2^n = \{\, (\mathbf{x}, y) \in \mathbb{R}^{n+1} \mid \|\mathbf{x}\|_2 \leq y \,\}$$

je epigraf eukleidovské normy $\|\cdot\|_2$, kterému se také říká *kužel druhého řádu.*

Pro $\mathbf{A}_i = \mathbf{0}$ se podmínka $g_i(\mathbf{x}) \leq 0$ stane lineární nerovnicí. Pro $\mathbf{c}_i = \mathbf{0}$ se podmínka $g_i(\mathbf{x}) \leq 0$ po umocnění na druhou stane konvexní kvadratická. Tedy LP a konvexní QCQP jsou speciální případy SOCP.

Příklad 17.13. Podívejme se znovu na úlohu nalezení geometrického mediánu, Příklad 1.14 z úvodní kapitoly. Tam minimalizujeme funkci

$$f(\mathbf{x}) = \sum_{i=1}^{m} \|\mathbf{x} - \mathbf{a}_i\|_2 \tag{17.8}$$

na množině $\mathbb{R}^n$. Tato funkce je konvexní a po zavedení pomocných proměnných z_i (podobná úprava jako v §12.1.1) lze její minimalizování formulovat jako SOCP:

$$\begin{aligned} \min \quad & z_1 + \cdots + z_m \\ \text{za podmínek} \quad & \|\mathbf{x} - \mathbf{a}_i\|_2 \leq z_i, \quad i = 1, \ldots, m. \end{aligned}$$

Pro případ $n = 2$ má úloha jednoduchý mechanický model[1]. Do vodorovného prkna vyvrtáme díry o souřadnicích $\mathbf{a}_i$. Každou dírou provlečeme provázek. Provázky jsou nahoře svázané uzlem do jednoho bodu a dole mají závaží o stejné hmotnosti. Poloha uzlu je $\mathbf{x}$. Hodnota $f(\mathbf{x})$ je potenciální energie soustavy a ustálený stav odpovídá minimu $f(\mathbf{x})$. ♦

17.4.5 Semidefinitní programování (SDP)

Věta 17.2. Pro každé $n \in \mathbb{N}$ je množina všech positivně semidefinitních matic rozměru $n \times n$ konvexní kužel.

Důkaz. Nechť $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$ jsou takové, že pro každé $\mathbf{x} \in \mathbb{R}^n$ platí $\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$ a $\mathbf{x}^T\mathbf{B}\mathbf{x} \geq 0$. Pak pro každé $\mathbf{x} \in \mathbb{R}^n$ a $\alpha, \beta \geq 0$ platí $\mathbf{x}^T(\alpha\mathbf{A} + \beta\mathbf{B})\mathbf{x} = \alpha\mathbf{x}^T\mathbf{A}\mathbf{x} + \beta\mathbf{x}^T\mathbf{B}\mathbf{x} \geq 0$. ■

[1] Toto mechanické zařízení je známé jako *Varignon frame* a v minulosti se opravdu používalo na řešení úlohy. Úloha má bohatou historii, je známa také jako Fermat-Weberův problém.

Konvexní kužel je konvexní množina. To umožňuje formulovat třídu konvexních úloh známou jako *semidefinitní programování* (SDP). Jednou z možných formulací je[2]

$$\begin{array}{rl} \min & \langle \mathbf{C}, \mathbf{X} \rangle \\ \text{za podmínek} & \mathbf{X} \text{ je positivně semidefinitní} \\ & \langle \mathbf{A}_i, \mathbf{X} \rangle = b_i, \quad i = 1, \ldots, m. \end{array} \tag{17.9}$$

kde matice $\mathbf{C}, \mathbf{A}_i \in \mathbb{R}^{n\times n}$ a skaláry b_i jsou dány a optimalizujeme přes positivně semidefinitní matice $\mathbf{X} \in \mathbb{R}^{n\times n}$. Operace $\langle \mathbf{A}, \mathbf{X} \rangle = \sum_i \sum_j a_{ij} x_{ij}$ označuje skalární součin matic (viz §2.1.7).

SDP je velmi obecná třída konvexních úloh. LP, konvexní QCQP a SOCP jsou speciální případy SDP. Pro ilustraci ukážeme, že pokud matice $\mathbf{C}, \mathbf{A}_i$ jsou diagonální, úloha (17.9) se redukuje na LP. V tom případě v součinech $\langle \mathbf{C}, \mathbf{X} \rangle$ a $\langle \mathbf{A}_i, \mathbf{X} \rangle$ nediagonální prvky matice $\mathbf{X}$ nehrají žádnou roli. Diagonální matice je positivně semidefinitní, právě když všechny její prvky jsou nezáporné (viz Cvičení 6.16). Tedy úloha (17.9) se redukuje na

$$\min\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{x} \geq \mathbf{0},\ \mathbf{a}_i^T\mathbf{x} = b_i \ (i = 1, \ldots, m) \,\},$$

kde vektory $\mathbf{c}, \mathbf{a}_i \in \mathbb{R}^n$ jsou diagonály matic $\mathbf{C}, \mathbf{A}_i$.

Některé konvexní úlohy nepatří do žádné z uvedených tříd.

Příklad 17.14. *Analytický střed mnohostěnu* $\{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} \geq \mathbf{b} \,\}$ je bod uvnitř mnohostěnu, který maximalizuje součin vzdáleností od nadrovin $\mathbf{a}_i^T\mathbf{x} = b_i$. Předpokládáme, že v každé nerovnici $\mathbf{a}_i^T\mathbf{x} \geq b_i$ je $\|\mathbf{a}_i\|_2 = 1$ (to jde vždy zařídit vydělením nerovnice číslem $\|\mathbf{a}_i\|_2$). Vzdálenost (se znaménkem) bodu $\mathbf{x}$ od nadroviny $\mathbf{a}_i^T\mathbf{x} = b_i$ je tedy rovna $\mathbf{a}_i^T\mathbf{x} - b_i$ (viz §5.2.1). Místo součinu vzdáleností maximalizujme jeho logaritmus (což je rostoucí funkce), tedy funkci

$$f(\mathbf{x}) = \ln \prod_{i=1}^{m} (\mathbf{a}_i^T\mathbf{x} - b_i) = \sum_{i=1}^{m} \ln(\mathbf{a}_i^T\mathbf{x} - b_i). \tag{17.10}$$

Tato funkce je konkávní. Funkce je definována jen pro vnitřní body mnohostěnu (tj. body splňující $\mathbf{A}\mathbf{x} > \mathbf{b}$) a blíží se $+\infty$ když se bod $\mathbf{x}$ blíží zevnitř k hranici. Tím je implicitně vynucena podmínka, že optimální bod má ležet uvnitř mnohostěnu. Lze ukázat, že pokud je mnohostěn neprázdný a omezený, úloha má maximum a toto maximum je jediné. ♦

17.5 Konvexní relaxace nekonvexních úloh

Relaxace je technika, kterou lze někdy získat přibližná řešení obtížných úloh. Spočívá na očividné skutečnosti (promyslete!), že pro každou množinu $X \subseteq \mathbb{R}^n$ a funkci $f\colon X \to \mathbb{R}$ platí

$$Y \supseteq X \quad \Longrightarrow \quad \min_{\mathbf{x}\in Y} f(\mathbf{x}) \leq \min_{\mathbf{x}\in X} f(\mathbf{x}). \tag{17.11}$$

Jak toho použít? Předpokládejme, že účelová funkce f je jednoduchá (např. konvexní), ale množina X přípustných řešení je komplikovaná a úloha $\min_{\mathbf{x}\in X} f(\mathbf{x})$ je díky tomu obtížná.

[2] Namítnete, že nemůžeme mluvit o konvexitě úlohy (17.9), protože v této úloze optimalizujeme přes množinu *matic* a konvexitu jsme definovali pro množiny a funkce *vektorů*. Definice konvexity lze ovšem snadno zobecnit na množiny a funkce matic: matici $\mathbb{R}^{m\times n}$ můžeme buď přerovnat do vektoru $\mathbb{R}^{mn}$, nebo (lépe) můžeme konvexitu definovat místo na prostoru $\mathbb{R}^n$ na obecném lineárním prostoru (viz učebnice lineární algebry). Podobně již pro Větu 17.2.

Nahradíme množinu X 'jednodušší' množinou $Y \supseteq X$ a řešíme snadnější úlohu $\min_{\mathbf{x}\in Y} f(\mathbf{x})$. Když budeme mít štěstí, bude nerovnost v (17.11) platit s rovností, tedy obě optima budou stejná. Když ne, získáme alespoň dolní mez na optimální řešení.

Je-li množina X nekonvexní, můžeme ji takto nahradit konvexní množinou $Y \supseteq X$. Pokud funkce f je konvexní, získáme konvexní úlohu. Mluvíme pak o **konvexní relaxaci**. Příkladem konvexní relaxace je LP relaxace, kterou jsme už viděli: lineární program (12.24) je konvexní relaxace lineárního celočíselného programu (12.23).

17.6 Cvičení

17.1. Mějme úlohu

$$\min\{\, f(x,y) \mid x,y \geq 0,\ 2x+y \geq 1,\ x+3y \geq 1\,\}.$$

Nakreslete množinu přípustných řešení. Pro každou z následujících účelových funkcí najděte úvahou množinu optimálních řešení a optimální hodnotu:

a) $f(x,y) = x+y$
b) $f(x,y) = x$
c) $f(x,y) = \min\{x,y\}$
d) $f(x,y) = \max\{x,y\}$
e) $f(x,y) = |x+y|$
f) $f(x,y) = x^2 + 9y^2$

V kterých případech se jedná o konvexní optimalizační úlohu?

17.2. Dokažte, že množina optimálních řešení konvexní optimalizační úlohy je konvexní.

17.3. Význačnou vlastností konvexních funkcí je to, že každé lokální minimum funkce je zároveň globální (Věta 17.1). Ne každá funkce s touto vlastností je ovšem konvexní. Člověk by si mohl myslet, že součet dvou funkcí (ne nutně konvexních) s touto vlastností bude mít tuto vlastnost také. Je toto tvrzení pravdivé?

17.4. Najděte spojitou funkci $f\colon \mathbb{R}^n \to \mathbb{R}$, jejíž každé lokální minimum je zároveň globální a

a) není konvexní
b) žádná její subkontura není konvexní množina.

17.5. Chceme rozestavit n lidí v místnosti čtvercového půdorysu tak, aby 'každý byl od každého co nejdále'. Navrhněte možné formulace této úlohy a u každé určete, zda je konvexní.

17.6. ($\star$) Uvažujme úlohu, známou jako *lineární lomené programování*:

$$\begin{aligned} \min \quad & (\mathbf{c}^T\mathbf{x}+d)/(\mathbf{e}^T\mathbf{x}+f) \\ \text{za podmínek} \quad & \mathbf{A}\mathbf{x} \geq \mathbf{b} \\ & \mathbf{e}^T\mathbf{x}+f > 0 \end{aligned}$$

a) Je účelová funkce na množině přípustných řešení konvexní?
b) Dokažte, že úloha je ekvivalentní lineárnímu programu (s proměnnými $\mathbf{x}$, z)

$$\begin{aligned} \min \quad & \mathbf{c}^T\mathbf{y}+dz \\ \text{za podmínek} \quad & \mathbf{A}\mathbf{y} \geq \mathbf{b}z \\ & \mathbf{e}^T\mathbf{y}+fz = 1 \\ & z \geq 0 \end{aligned}$$

17.7. Najděte explicitní řešení pro následující úlohy QCQP ($\mathbf{A}, \mathbf{B}$ jsou positivně definitní):

a) $\min\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{x}^T\mathbf{A}\mathbf{x} \le 1 \,\}$

b) $\min\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ (\mathbf{x}-\mathbf{b})^T\mathbf{A}(\mathbf{x}-\mathbf{b}) \le 1 \,\}$

c) $\min\{\, \mathbf{x}^T\mathbf{B}\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{x}^T\mathbf{A}\mathbf{x} \le 1 \,\}$

17.8. ($\star$) Formulujte úlohu $\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{A}\mathbf{x}-\mathbf{b}\|_4$ jako konvexní QCQP.

17.9. ($\star$) Dokažte, že pro libovolný vektor $\mathbf{x} \in \mathbb{R}^n$ a skaláry $y \ge 0$, $z \ge 0$ platí

$$\mathbf{x}^T\mathbf{x} \le yz \quad \Longleftrightarrow \quad \|(2\mathbf{x}, y-z)\|_2 \le y+z.$$

Uvažujte úlohu, kdy maximalizujeme harmonický průměr afinních fukcí, tedy funkci

$$f(\mathbf{x}) = \Bigl(\sum_{i=1}^{m} (\mathbf{a}_i^T\mathbf{x} - b_i)^{-1}\Bigr)^{-1}$$

za podmínek $\mathbf{a}_i^T\mathbf{x} > b_i$. Je tato úloha (po možné jednoduché transformaci) konvexní? Vyjádřete úlohu jako SOCP pomocí dokázané ekvivalence.

17.10. Sedumi (http://sedumi.ie.lehigh.edu) je jednou z implementací algoritmu na řešení úloh LP, QP, SOCP a SDP (používá tzv. algoritmus vnitřního bodu). Je k dispozici i pro Matlab. Knihovnu si stáhněte a pak prostudujte a pochopte nápovědu k funkci `sedumi.m`, která je matlabským rozhraním knihovny.

Nápověda a řešení

17.3. Vezměme např. funkce $g_1(x) = -\mathrm{e}^{-(x+1)^2}$ a $g_2(x) = -\mathrm{e}^{-(x-1)^2}$. Každá z těchto funkcí má jediné lokální minimum, ale funkce $g_1 + g_2$ má lokální minima dvě.

17.5. Nechť $\mathbf{x}_i \in \mathbb{R}^2$ jsou souřadnice $i$tého člověka v místnosti. Bez ztráty obecnosti nechť je místnost množina $[-1,1]\times[-1,1]$ (kde $[-1,1]$ značí uzavřený interval). Jedna přirozená formulace je taková, že hledáme takové $\mathbf{x}_1, \ldots, \mathbf{x}_n$ ležící v tomto čtverci, aby číslo $\min_{i\ne j} \|\mathbf{x}_i - \mathbf{x}_j\|$ bylo maximální. Tato úloha není konvexní, neboť její účelová funkce není konvexní.

17.6.a) Ne.

17.6.b) Uvažujte substituci $\mathbf{y} = \mathbf{x}/(\mathbf{e}^T\mathbf{x} + f)$, $z = /(\mathbf{e}^T\mathbf{x} + f)$.

17.7.a) Viz Cvičení 12.6.

17.7.b) Substituujte $\mathbf{y} = \mathbf{x} - \mathbf{b}$.

17.7.c) Optimální hodnota je nula.

17.9. Místo maximalizace funkce $f(\mathbf{x})$ minimalizujme funkci $1/f(\mathbf{x})$, která je konvexní na množině přípustných hodnot. Úloha je ekvivalentní úloze

$$\begin{aligned} \min \quad & t_1 + \cdots + t_m \\ \text{za podmínek} \quad & t_i(\mathbf{a}_i^T\mathbf{x} - b_i) \ge 1, \quad i = 1, \ldots, m \\ & t_i \ge 0, \quad i = 1, \ldots, m. \end{aligned}$$

Použitím dokázané ekvivalence převedeme na SOCP

$$\begin{aligned} \min \quad & t_1 + \cdots + t_m \\ \text{za podmínek} \quad & \|(2, \mathbf{a}_i^T\mathbf{x} - b_i - t_i)\|_2 \ge \mathbf{a}_i^T\mathbf{x} - b_i + t_i, \quad i = 1, \ldots, m \\ & \mathbf{a}_i^T\mathbf{x} \ge b_i, \quad i = 1, \ldots, m \\ & t_i \ge 0, \quad i = 1, \ldots, m. \end{aligned}$$

Kapitola 18

(⋆) Lagrangeova dualita

Zatímco v §15 jsme odvodili dualitu pro lineární programování, zde popíšeme základ teorie duality pro obecné optimalizační úlohy. Dualita v LP se pak bude jevit jako speciální případ.

18.1 Minimaxní nerovnost

Věta 18.1. Pro libovolné množiny X, Y a funkci $L\colon X \times Y \to \mathbb{R}$ platí **minimaxní nerovnost**

$$\min_{x\in X} \underbrace{\max_{y\in Y} L(x,y)}_{F(x)} \geq \max_{y\in Y} \underbrace{\min_{x\in X} L(x,y)}_{G(y)} \tag{18.1}$$

za předpokladu, že všechna minima a maxima existují. Rovnost v (18.1) nastává právě tehdy, existuje-li bod $(x^*, y^*) \in X \times Y$ (**sedlový bod** funkce L) splňující

$$L(x^*, y) \leq L(x^*, y^*) \leq L(x, y^*) \qquad \forall x \in X,\ y \in Y. \tag{18.2}$$

Důkaz. Pro každou funkci $L\colon X \times Y \to \mathbb{R}$ platí (z definice maxima a minima)

$$F(x) \geq L(x,y) \geq G(y) \qquad \forall x \in X,\ y \in Y. \tag{18.3}$$

Z toho plyne

$$\min_{x\in X} F(x) \geq \max_{y\in Y} G(y), \tag{18.4}$$

což je nerovnost (18.1).

Podmínka (18.2) je ekvivalentní (opět z definice maxima a minima) podmínce

$$F(x^*) = L(x^*, y^*) = G(y^*). \tag{18.5}$$

Druhá část věty tvrdí, že rovnost v (18.4) nastane právě tehdy, když platí (18.5). To opět plyne z definice maxima a minima (rozmyslete!). ■

Příklad 18.1. Nechť $X = Y = \{1, 2, 3, 4\}$. Následující dvě tabulky uvádějí dva příklady funkce L (16 čísel uprostřed) a odpovídajících funkcí F (řádková maxima, 4 čísla zcela vpravo) a G (sloupcová minima, 4 čísla zcela dole):

x \ y	1	2	3	4	$F(x)$
1	-1	4	7	4	7
2	4	4	6	-2	6
3	1	5	3	3	5
4	3	5	3	2	5
$G(y)$	-1	4	3	-2	

x \ y	1	2	3	4	$F(x)$
1	-1	4	7	4	7
2	4	4	6	-2	6
3	1	0	$\boxed{3}$	3	3
4	3	3	$\boxed{3}$	2	3
$G(y)$	-1	0	3	-2	

Nerovnosti (18.3) říkají, že maximum libovolného řádku není menší než minimum libovolného sloupce. Funkce L v levé tabulce nemá sedlový bod a tedy $\min_{x\in X} F(x) > \max_{y\in Y} G(y)$. Funkce L v pravé tabulce má sedlový bod (dokonce dva, v rámečcích) a tedy $\min_{x\in X} F(x) = \max_{y\in Y} G(y)$. ♦

Mají-li množiny X, Y nekonečný počet prvků, pak pro některá $x \in X$ nemusí existovat $\max_{y\in Y} L(x, y)$ a pro některá $y \in Y$ nemusí existovat $\min_{x\in X} L(x, y)$. Např. $\min\limits_{x\in(0,+\infty)} \frac{1}{x}$ neexistuje a $\min\limits_{x\in\mathbb{R}} x$ také neexistuje. Nahradíme-li ovšem minima infimy a maxima supremy, budou existovat vždy: např. $\inf\limits_{x\in(0,+\infty)} \frac{1}{x} = 0$ a $\inf\limits_{x\in\mathbb{R}} x = -\infty$. Lze dokázat (důkaz je obdobný), že platí obdoba Věty 18.1, ve které nerovnost (18.1) nahradíme nerovností

$$\min_{x\in X} \underbrace{\sup_{y\in Y} L(x, y)}_{F(x)} \geq \max_{y\in Y} \underbrace{\inf_{x\in X} L(x, y)}_{G(y)}. \tag{18.6}$$

V tom případě ovšem funkce F, G mohou nabývat nekonečných hodnot, tj. máme $F\colon X \to \overline{\mathbb{R}}$ a $G\colon Y \to \overline{\mathbb{R}}$, kde $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ označuje *rozšířenou množinu reálných čísel.*

18.2 Lagrangeova duální úloha

Ke každé optimalizační úloze (kterou nazýváme primární) lze sestrojit jinou optimalizační úlohu (nazývanou duální) tak, že mezi nimi platí jisté užitečné vztahy. Jedna forma duality se získá následovně: chytře zvolíme množiny X, Y a funkci L tak, aby levá strana minimaxní nerovnosti (18.6) byla primární (tedy původní) úloha, v tom případě F bude primární účelová funkce, ve které budou schovaná primární omezení v podobně nekonečných hodnot funkce F. Pravá strana nerovnosti (18.6) pak bude duální úloha, kde G bude duální účelová funkce, ve které duální omezení budou opět schovaná pomocí nekonečných hodnot.

Nechť množina $X \subseteq \mathbb{R}^n$, funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ a zobrazení $\mathbf{g}\colon \mathbb{R}^n \to \mathbb{R}^m$ jsou dány (jak dále uvidíme, reprezentují primární úlohu). Zvolme $Y = \mathbb{R}^m_+$ a definujme **Lagrangeovu funkci** (už jsme ji potkali u metody Lagrangeových multiplikátorů)

$$L(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) + \mathbf{y}^T \mathbf{g}(\mathbf{x}). \tag{18.7}$$

Při této volbě máme

$$F(\mathbf{x}) = \sup_{\mathbf{y}\in Y} L(\mathbf{x}, \mathbf{y}) = \sup_{\mathbf{y}\geq\mathbf{0}} [f(\mathbf{x}) + \mathbf{y}^T \mathbf{g}(\mathbf{x})] = \begin{cases} f(\mathbf{x}) & \text{když } \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \\ +\infty & \text{jinak.} \end{cases} \tag{18.8}$$

Abyste to uviděli, promyslete si, že pro každý vektor $\mathbf{a} \in \mathbb{R}^m$ platí

$$\sup_{\mathbf{y} \geq \mathbf{0}} \mathbf{y}^T \mathbf{a} = \begin{cases} 0 & \text{když } \mathbf{a} \leq \mathbf{0}, \\ +\infty & \text{jinak.} \end{cases}$$

Po dosazení (18.8) do levé strany (18.6) je tedy

$$\min_{\mathbf{x} \in X} F(\mathbf{x}) = \min\{\, f(\mathbf{x}) \mid \mathbf{x} \in X,\ \mathbf{g}(\mathbf{x}) \leq \mathbf{0}\,\}. \tag{18.9}$$

Toto je naše primární úloha. Všimněme si, jak jsme omezující podmínky $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$ zahrnuli do funkce F za cenu, že funkce F může nabývat nekonečných hodnot (tedy hodnot z množiny $\overline{\mathbb{R}}$).

Nyní se podívejme na duální účelovou funkci G:

$$G(\mathbf{y}) = \inf_{\mathbf{x} \in X} L(\mathbf{x}, \mathbf{y}) = \inf_{\mathbf{x} \in X} [f(\mathbf{x}) + \mathbf{y}^T \mathbf{g}(\mathbf{x})]. \tag{18.10}$$

Duální úloha (tzv. **Lagrangeova duální úloha**) bude mít tvar

$$\max_{\mathbf{y} \in Y} G(\mathbf{y}) = \max_{\mathbf{y} \geq \mathbf{0}} G(\mathbf{y}). \tag{18.11}$$

Podobně jako u primární úlohy, pro některá $\mathbf{y} \in Y$ může být $G(\mathbf{y}) = -\infty$, což bude reprezentovat duální omezení. Všimněte si, že pro každé $\mathbf{x}$ je funkce L afinní funkcí proměnné $\mathbf{y}$. Dle Věty 16.9 je tedy duální funkce G konkávní a tedy duální úloha bude maximalizace konkávní funkce, tedy konvexní úloha. To platí vždy, i když primární úloha není konvexní.

Můžeme také zvolit $Y = \mathbb{R}^m$. Pak bude (rozmyslete!)

$$F(\mathbf{x}) = \sup_{\mathbf{y} \in Y} L(\mathbf{x}, \mathbf{y}) = \sup_{\mathbf{y} \in \mathbb{R}^m} [f(\mathbf{x}) + \mathbf{y}^T \mathbf{g}(\mathbf{x})] = \begin{cases} f(\mathbf{x}) & \text{když } \mathbf{g}(\mathbf{x}) = \mathbf{0}, \\ +\infty & \text{jinak} \end{cases} \tag{18.12}$$

a tedy

$$\min_{\mathbf{x} \in X} F(\mathbf{x}) = \min\{\, f(\mathbf{x}) \mid \mathbf{x} \in X,\ \mathbf{g}(\mathbf{x}) = \mathbf{0}\,\}, \tag{18.13}$$

tedy v primární úloze jsme získali omezení typu rovností.

Pokud bychom chtěli sestrojit duální úlohu k úloze s omezeními typu nerovností i rovností (reprezentovanými pomocí zobrazení $\mathbf{g}\colon \mathbb{R}^n \to \mathbb{R}^m$, $\mathbf{h}\colon \mathbb{R}^n \to \mathbb{R}^l$), zvolíme Lagrangeovu funkci jako $L(\mathbf{x}, \mathbf{u}, \mathbf{v}) = f(\mathbf{x}) + \mathbf{u}^T \mathbf{g}(\mathbf{x}) + \mathbf{v}^T \mathbf{h}(\mathbf{x})$ kde $\mathbf{y} = (\mathbf{u}, \mathbf{v}) \in Y = \mathbb{R}^m_+ \times \mathbb{R}^l$. Pak

$$F(\mathbf{x}) = \sup_{(\mathbf{u},\mathbf{v}) \in Y} L(\mathbf{x}, \mathbf{u}, \mathbf{v}) = \sup_{\substack{\mathbf{u} \in \mathbb{R}^m_+ \\ \mathbf{v} \in \mathbb{R}^l}} [f(\mathbf{x}) + \mathbf{u}^T \mathbf{g}(\mathbf{x}) + \mathbf{v}^T \mathbf{h}(\mathbf{x})] = \begin{cases} f(\mathbf{x}) & \text{když } \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \text{ a } \mathbf{h}(\mathbf{x}) = \mathbf{0}, \\ +\infty & \text{jinak} \end{cases} \tag{18.14}$$

a tedy

$$\min_{\mathbf{x} \in X} F(\mathbf{x}) = \min\{\, f(\mathbf{x}) \mid \mathbf{x} \in X,\ \mathbf{g}(\mathbf{x}) \leq \mathbf{0},\ \mathbf{h}(\mathbf{x}) = \mathbf{0}\,\}. \tag{18.15}$$

18.3 Silná dualita

Z nerovnosti (18.6) plyne, že optimální hodnota primární úlohy není menší než optimální hodnota duální úlohy. Tato skutečnost je známa jako věta o **slabé dualitě**. Rozdílu mezi primární a duální optimální hodnotou se říká **dualitní mezera**. Když v nerovnosti (18.1) nastane rovnost, jsou si optimální hodnoty primární a duální úlohy rovny neboli dualitní mezera je nulová. V tom případě říkáme, že pro naší úlohu platí **silná dualita**.

Silná dualita může platit pro velice různé úlohy. Uvedeme nyní, ve Větě 18.2, jednu postačující (avšak nikoliv nutnou) podmínku, za které platí silná dualita. Uvažujeme přitom pouze primírní omezení typu nerovnosti, tedy dvojici úloh (18.9) a (18.11).

Řekneme, že funkce $g_1, \ldots, g_m \colon \mathbb{R}^n \to \mathbb{R}$ na množině X splňují *Slaterovu podmínku*, když existuje vnitřní bod $\mathbf{x}$ množiny X takový, že $\mathbf{g}(\mathbf{x}) < \mathbf{0}$, neboli

$$g_1(\mathbf{x}) < 0, \quad \ldots, \quad g_m(\mathbf{x}) < 0. \tag{18.16}$$

Pokud je prvních $k \leq m$ funkcí g_i afinních, podmínku (18.16) lze změkčit na

$$g_1(\mathbf{x}) \leq 0, \quad \ldots, \quad g_k(\mathbf{x}) \leq 0, \quad g_{k+1}(\mathbf{x}) < 0, \quad \ldots, \quad g_m(\mathbf{x}) < 0. \tag{18.17}$$

Věta 18.2. Nechť
- množina $X \subseteq \mathbb{R}^n$ je konvexní,
- funkce $f \colon \mathbb{R}^n \to \mathbb{R}$ a $g_1, \ldots, g_m \colon \mathbb{R}^n \to \mathbb{R}$ jsou konvexní na X,
- funkce $g_1, \ldots, g_m$ na množině X splňují Slaterovu podmínku.

Pak platí silná dualita, neboli optimální hodnoty úloh (18.9) a (18.11) jsou si rovny.

Dále uveďme obdobu věty o komplementaritě.

Věta 18.3. Nechť $\mathbf{x} \in X$ je optimum primární úlohy a $\mathbf{y} \in \mathbb{R}^m_+$ je optimum duální úlohy a nastává silná dualita. Pak platí *podmínky komplementarity*

$$y_i g_i(x_i) = 0 \qquad \forall i = 1, \ldots, n. \tag{18.18}$$

Důkaz. Platí

$$F(\mathbf{x}) = f(\mathbf{x}) = G(\mathbf{y}) = \min_{\mathbf{x}' \in X} [f(\mathbf{x}') + \mathbf{y}\mathbf{g}(\mathbf{x}')] \leq f(\mathbf{x}) + \mathbf{y}^T \mathbf{g}(\mathbf{x}) \leq f(\mathbf{x}).$$

Druhá rovnost plyne ze silné duality. Třetí rovnost je definice duální úlohy. První nerovnost plyne z definice minima. Druhá nerovnost platí, protože $\mathbf{y} \geq \mathbf{0}$ a $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$.

Ale protože $f(\mathbf{x})$ je na začátku i na konci řetězce nerovností, musí obě nerovnosti být rovnostmi, $f(\mathbf{x}) + \mathbf{y}^T \mathbf{g}(\mathbf{x}) = f(\mathbf{x})$. Z toho plyne

$$\mathbf{y}^T \mathbf{g}(\mathbf{x}) = 0. \tag{18.19}$$

To je ale ekvivalentní podmínkám (18.18), protože $\mathbf{y} \geq \mathbf{0}$ a $\mathbf{g}(\mathbf{x}) \leq \mathbf{0}$. ∎

18.4 Příklady

Příklad 18.2. Nechť $f(\mathbf{x}) = \mathbf{c}^T \mathbf{x}$, $\mathbf{g}(\mathbf{x}) = \mathbf{b} - \mathbf{A}\mathbf{x}$, $X = \mathbb{R}^n$, $Y \in \mathbb{R}^m_+$. Primární úloha je

$$\min_{\mathbf{x} \in \mathbb{R}^n} \sup_{\mathbf{y} \geq \mathbf{0}} L(\mathbf{x}, \mathbf{y}) = \min\{\, \mathbf{c}^T \mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{A}\mathbf{x} \geq \mathbf{b} \,\}. \tag{18.20}$$

Odvoďme duální úlohu. Lagrangeova funkce je

$$L(\mathbf{x},\mathbf{y}) = \mathbf{c}^T\mathbf{x} + \mathbf{y}^T(\mathbf{b}-\mathbf{A}\mathbf{x}) = \mathbf{c}^T\mathbf{x} - \mathbf{y}^T\mathbf{A}\mathbf{x} + \mathbf{y}^T\mathbf{b} = (\mathbf{c}^T - \mathbf{y}^T\mathbf{A})\mathbf{x} + \mathbf{y}^T\mathbf{b}. \tag{18.21}$$

Duální účelová funkce bude

$$G(\mathbf{y}) = \inf_{\mathbf{x}\in\mathbb{R}^n} L(\mathbf{x},\mathbf{y}) = \begin{cases} \mathbf{y}^T\mathbf{b} & \text{když } \mathbf{c}^T - \mathbf{y}^T\mathbf{A} = \mathbf{0}, \\ -\infty & \text{jinak.} \end{cases}$$

Duální úloha tedy je

$$\max_{\mathbf{y}\geq\mathbf{0}} G(\mathbf{y}) = \max\{\, \mathbf{b}^T\mathbf{y} \mid \mathbf{A}^T\mathbf{y} = \mathbf{c},\ \mathbf{y}\geq\mathbf{0}\,\}. \tag{18.22}$$

To je stejný výsledek, jaký bychom dostali podle předpisu (15.1) na konstrukci duálního LP. Konkrétně, primární a duální úloha je dvojice úloh (15.2). ♦

Příklad 18.3. Nechme vše jako v minulém příkladu, jen místo $X = \mathbb{R}^n$ zvolme $X = \mathbb{R}^n_+$. Primární úloha bude

$$\min_{\mathbf{x}\in\mathbb{R}^n_+} \sup_{\mathbf{y}\geq\mathbf{0}} L(\mathbf{x},\mathbf{y}) = \min\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x}\in\mathbb{R}^n,\ \mathbf{A}\mathbf{x}\geq\mathbf{b},\ \mathbf{x}\geq\mathbf{0}\,\}, \tag{18.23}$$

duální účelová funkce bude

$$G(\mathbf{y}) = \inf_{\mathbf{x}\in\mathbb{R}^n_+} L(\mathbf{x},\mathbf{y}) = \begin{cases} \mathbf{y}^T\mathbf{b} & \text{když } \mathbf{c}^T - \mathbf{y}^T\mathbf{A} \geq \mathbf{0}, \\ -\infty & \text{jinak.} \end{cases}$$

a duální úloha

$$\max_{\mathbf{y}\geq\mathbf{0}} G(\mathbf{y}) = \max\{\, \mathbf{b}^T\mathbf{y} \mid \mathbf{A}^T\mathbf{y} \leq \mathbf{c},\ \mathbf{y}\geq\mathbf{0}\,\}. \tag{18.24}$$

To je opět stejný výsledek, jaký bychom dostali podle předpisu (15.1). ♦

Příklad 18.4. Napišme duální úlohu k úloze

$$\min\{\, \mathbf{x}^T\mathbf{x} \mid \mathbf{x}\in\mathbb{R}^n,\ \mathbf{A}\mathbf{x} = \mathbf{b}\,\}.$$

Máme $X = \mathbb{R}^n$ a

$$L(\mathbf{x},\mathbf{y}) = \mathbf{x}^T\mathbf{x} + \mathbf{y}^T(\mathbf{b}-\mathbf{A}\mathbf{x}) = \mathbf{x}^T\mathbf{x} - \mathbf{y}^T\mathbf{A}\mathbf{x} + \mathbf{y}^T\mathbf{b},$$

duální funkce je tedy

$$G(\mathbf{y}) = \min_{\mathbf{x}\in\mathbb{R}^n} L(\mathbf{x},\mathbf{y}) = \min_{\mathbf{x}\in\mathbb{R}^n} (\mathbf{x}^T\mathbf{x} - \mathbf{y}^T\mathbf{A}\mathbf{x} + \mathbf{y}^T\mathbf{b}).$$

Řešení musí splňovat $\partial L(\mathbf{x},\mathbf{y})/\partial\mathbf{x} = \mathbf{0}$, což dá $\mathbf{x} = \mathbf{A}^T\mathbf{y}/2$. Po dosazení dostaneme

$$G(\mathbf{y}) = L(\mathbf{A}^T\mathbf{y}/2,\mathbf{y}) = \mathbf{b}^T\mathbf{y} - \mathbf{y}^T\mathbf{A}\mathbf{A}^T\mathbf{y}/4.$$ ♦

Příklad 18.5. Řešme lineární program

$$\min\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in [0,1]^n,\ \mathbf{A}\mathbf{x} = \mathbf{b} \,\}.$$

Duál bychom mohli sestrojit tak, že bychom omezení $\mathbf{x} \in [0,1]^n$ reprezentovali nerovnostmi $0 \le x_j \le 1\ \forall j$ a pak postupovali dle návodu v kapitole o dualitě v LP. Ale postupujme příměji. Zvolme Lagrangeovu funkci jako (18.20) a $X = [0,1]^n$, $Y = \mathbb{R}^m$. Duální účelová funkce bude

$$\begin{aligned} G(\mathbf{y}) = \inf_{\mathbf{x}\in X} L(\mathbf{x},\mathbf{y}) &= \min_{\mathbf{x}\in[0,1]^n} (\mathbf{c}^T\mathbf{x} - \mathbf{y}^T\mathbf{A}\mathbf{x} + \mathbf{y}^T\mathbf{b}) \\ &= \mathbf{y}^T\mathbf{b} + \min_{\mathbf{x}\in[0,1]^n} (\mathbf{c}^T - \mathbf{y}^T\mathbf{A})\mathbf{x} \\ &= \mathbf{y}^T\mathbf{b} + \min_{\mathbf{x}\in[0,1]^n} \sum_j (c_j - \mathbf{y}^T\mathbf{a}_j)x_j \\ &= \mathbf{y}^T\mathbf{b} + \sum_j \min_{x\in[0,1]} (c_j - \mathbf{y}^T\mathbf{a}_j)x \\ &= \mathbf{y}^T\mathbf{b} + \sum_j \min\{c_j - \mathbf{y}^T\mathbf{a}_j, 0\} \end{aligned}$$

kde $\mathbf{a}_j$ značí j-tý sloupec matice $\mathbf{A}$ a poslední rovnost platí, protože pro každé $d \in \mathbb{R}$ je $\min_{x\in[0,1]} dx = \min\{d, 0\}$. ♦

Příklad 18.6. Řešme celočíselný lineární program

$$\min\{\, \mathbf{c}^T\mathbf{x} \mid \mathbf{x} \in \{0,1\}^n,\ \mathbf{A}\mathbf{x} = \mathbf{b} \,\}.$$

Nechť Lagrangeova funkce je (18.20) a nechť $X = \{0,1\}^n$. Duální úloha je stejná jako v minulém příkladě, neboť $\min_{x\in[0,1]} dx = \min_{x\in\{0,1\}} dx = \min\{d, 0\}$.

Zde předpoklady Věty 18.2 neplatí, protože množina X není konvexní. Opravdu, silná dualita u celočíselného programování obecně neplatí. ♦

Silná dualita může nastat (i když spíše vyjímečně) i pro nekonvexní úlohu. Jednou takovou třídou úloh je libovolné (tedy ne nutně konvexní) QCQP s nejvýše *jedním* omezením.

Příklad 18.7. Úloha QCQP

$$\min\{\, \mathbf{x}^T\mathbf{A}\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{x}^T\mathbf{x} = 1 \,\} \tag{18.25}$$

není konvexní. Víme ale, že řešení se najde pomocí spektrálního rozkladu:

$$\min_{\mathbf{x}^T\mathbf{x}=1} \mathbf{x}^T\mathbf{A}\mathbf{x} = \min_{\mathbf{x}^T\mathbf{x}=1} \mathbf{x}^T\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T\mathbf{x} = \min_{\mathbf{z}^T\mathbf{z}=1} \mathbf{z}^T\boldsymbol{\Lambda}\mathbf{z} = \lambda_{\min}(\mathbf{A}).$$

Ukážeme navíc, že platí silná dualita. Máme

$$L(\mathbf{x},\mathbf{y}) = \mathbf{x}^T\mathbf{A}\mathbf{x} + y(1 - \mathbf{x}^T\mathbf{x}) = \mathbf{x}^T(\mathbf{A} - y\mathbf{I})\mathbf{x} + y.$$

Tedy

$$G(y) = \begin{cases} y & \text{když } \mathbf{A} - y\mathbf{I} \text{ je positivně semidefinitní,} \\ -\infty & \text{jinak.} \end{cases}$$

Ale matice $\mathbf{A} - y\mathbf{I}$ je positivně semidefinitní, právě když nejmenší vlastní číslo matice $\mathbf{A} - y\mathbf{I}$ je nezáporné, neboli (viz Cvičení 6.5) $\lambda_{\min}(\mathbf{A}) \ge y$. Duální úloha tedy je

$$\max\{\, y \in \mathbb{R} \mid \lambda_{\min}(\mathbf{A}) \ge y \,\}.$$

Ta má zřejmé optimální řešení $y = \lambda_{\min}(\mathbf{A})$. ♦

Kapitola 19

($\star$) Vícekriteriální optimalizace

19.1 Uspořádání na množině

Binární relace na množině Y je množina $R \subseteq Y \times Y$. Binární relace je

- **reflexivní**, když $(x, x) \in R$ pro každé $x \in Y$,
- **transitivní**, když $(x, y) \in R$ a $(y, z) \in R$ implikuje $(x, z) \in R$,
- **antisymetrická**, když $(x, y) \in R$ a $(y, x) \in R$ implikuje $x = y$.

Částečné uspořádání (krátce jen **uspořádání**) na množině Y je binární relace na Y, která je reflexivní, transitivní a antisymetrická.

Relace uspořádání se obvykle značí infixově symbolem $\preceq$, tedy místo $(x, y) \in R$ píšeme $x \preceq y$. Pokud potřebujeme rozlišit více různých kvasi-uspořádání, používáme symboly jako $\leq_1, \leq_2, \leq', \preceq'$ atd.

Prvky $x, y \in Y$ jsou **srovnatelné** v uspořádání $\preceq$, když $x \preceq y$ nebo $y \preceq x$ nebo obojí. (Kvasi-)uspořádání je **úplné** (neboli **totální**), když každé dva prvky z Y jsou srovnatelné.

Příklad 19.1.

- $Y = \mathbb{R}$ a $\preceq$ je přirozené uspořádání reálných čísel. Tato relace je úplné uspořádání.
- $Y \subseteq 2^U$ a $\preceq$ je relace inkluze na množině 2^U, tedy $x \preceq y$ právě když $x \subseteq y$. Zde U je libovolná množina a 2^U značí množinu všech jejích podmnožin. Tato relace je uspořádání, ale není úplné.
- $Y = \mathbb{N}$ a $\preceq$ je relace dělitelnosti, tj. $x \preceq y$ právě když x dělí y. Tato relace je uspořádání, není úplné.
- $\mathbf{x} \preceq \mathbf{y}$ právě když $\sum_{i=1}^n x_i \leq \sum_{i=1}^n y_i$. Tato relace není antisymetrická, tedy není uspořádání. ♦

Nás ovšem nejvíce zajímá případ $Y = \mathbb{R}^n$.

Příklad 19.2. Příklady uspořádání na množině $\mathbb{R}^n$:

- Uspořádání 'po složkách': $\mathbf{x} \preceq \mathbf{y}$ právě když $x_i \leq y_i$ pro všechna $i = 1, \dots, n$. Toto uspořádání není úplné: např. pro $n = 2$ jsou vektory $\mathbf{x} = (0, 1)$ a $\mathbf{y} = (1, 0)$ nesrovnatelné.
- Lexikografické ('slovníkové') uspořádání: $\mathbf{x} \preceq \mathbf{y}$ právě když buď $\mathbf{x} = \mathbf{y}$, nebo existuje m takové, že $x_m < y_m$ a pro všechna $i < m$ platí $x_i = y_i$.

  Toto uspořádání je úplné.

- Definujme $\mathbf{x} \preceq \mathbf{y}$ právě když $\sum_{i=1}^n x_i \le \sum_{i=1}^n y_i$. Tato relace není uspořádání, protože není antisymetrická. ♦

  Prvek $x \in Y$ se nazývá (vzhledem ke kvasi-uspořádání $\preceq$)
- **minimální prvek** množiny Y, když $y \preceq x$ implikuje $x \preceq y$, pro všechna $y \in Y$.
- **nejmenší prvek** množiny Y, když $x \preceq y$, pro všechna $y \in Y$.

Pro totální kvasi-uspořádání oba pojmy splývají.

19.2 Úlohy vícekriteriální optimalizace

V klasické optimalizaci jsme se zabývali úlohami typu

$$\min_{x \in X} f(x), \tag{19.1}$$

kde X je množina přípustných řešení a $f\colon X \to \mathbb{R}$ je účelová (neboli kriteriální) funkce. Optimální hodnoty této úlohy jsou minimální prvky množiny $f(X) = \{\, f(x) \mid x \in X \,\} \subseteq \mathbb{R}$. Zde pojem 'minimální prvek' se myslí vzhledem k přirozenému uspořádání na $\mathbb{R}$.

Zobecněme tuto úlohu. Nechť $f\colon X \to Y$ a nechť $\preceq$ je (kvasi-)uspořádání na množině Y. Pak úlohou (19.1) budeme rozumět nalezení minimálních prvků množiny $f(X) \subseteq Y$ vzhledem k uspořádání $\preceq$. Případně pro každý minimální prvek y množiny $f(X)$ chceme najít argument $x \in X$, ve kterém se nabývá, tedy splňujícím $y = f(x)$.

Nejčastěji v aplikacích potkáme případ $Y = \mathbb{R}^n$. Pak mluvíme o *vícekriteriální optimalizaci*[1], nebot vlastně chceme minimalizovat více skalárních kritérií (složek zobrazení $\mathbf{f}\colon X \to \mathbb{R}^n$, hodnoty zobrazení jsou vektory a tedy ho píšeme tučně) najednou. Dále se omezíme pouze na tento případ.

Příklad 19.3. V obchodě nabízejí čtyři druhy aut s těmito vlastnostmi:

		VW Golf	Opel Astra	Ford Focus	Toyota Corolla
cena	[tis. euro]	16	15	14	15
spotřeba	[l/100km]	7.2	7.0	7.5	8.2

Chceme levné auto s malou spotřebou. Která auta je dobré si koupit a která naopak nekoupit?

Máme $X = \{\,$VW Golf, Opel Astra, Ford Focus, Toyota Corolla$\,\}$ a $Y = \mathbb{R}^2$. Tabulka definuje zobrazení $\mathbf{f}$. Není ovšem jasné, jaké (kvasi-)uspořádání na množině $\mathbb{R}^2$ použít pro rozhodování.

Rozhodujme se vzhledem k uspořádání 'po složkách' $\le^2$. Vzhledem k tomuto uspořádání nemá množina $\mathbf{f}(X)$ nejmenší prvek, neboli kritéria jsou v konfliktu. Její minimální prvky jsou $\mathbf{f}(\text{Opel Astra}) = (15, 7.0)$ a $\mathbf{f}(\text{Ford Focus}) = (14, 7.5)$.

Rozhodujme se vzhledem k lexikografickému uspořádání, přesněji nejprve se rozhodujme dle ceny a pak dle spotřeby. Nyní minimální prvek množiny $\mathbf{f}(X)$ je $\mathbf{f}(\text{Ford Focus}) = (14, 7.5)$. ♦

Příklad 19.4. Chceme řešit (přesně či přibližně) nehomogenní soustavu lineárních rovnic $\mathbf{A}\mathbf{x} \approx \mathbf{b}$, ale zároveň chceme, aby velikost vektoru $\mathbf{x}$ byla malá. Řešíme tedy úlohu

$$\min_{\mathbf{x} \in \mathbb{R}^n} (\, \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2,\ \|\mathbf{x}\|_2 \,).$$

[1] Angl. *multicriteria optimization.* Názvosloví ovšem není jednotné, jindy se používají názvy *multiobjective optimization* nebo *vector optimization* (neboť hodnoty zobrazení f jsou vektory).

Jaké jsou minimální prvky množiny $\mathbf{f}(\mathbb{R}^2)$ vzhledem k uspořádání po složkách? Je jich nekonečně mnoho, obrázek. Vezmeme-li lexikografické uspořádání a je-li soustava přeurčená, dostaneme metodu nejmenší normy. ♦

Příklad 19.5. V okrese je n vesnic se souřadnicemi $\mathbf{a}_1, \ldots, a_n \in \mathbb{R}^2$. Do jakého místa $\mathbf{x} \in \mathbb{R}^2$ máme umístit heliport, aby byl ke všem vesnicím blízko?

Máme $X = \mathbb{R}^2$, $Y = \mathbb{R}^n$, a $f_i(\mathbf{x}) = \|\mathbf{a}_i - \mathbf{x}\|_2$. Řešíme úlohu

$$\min_{\mathbf{x}\in\mathbb{R}^2} (\|\mathbf{a}_1 - \mathbf{x}\|_2, \cdots, \|\mathbf{a}_n - \mathbf{x}\|_2).$$

Součástí úlohy není (kvasi-)uspořádání na množině $\mathbb{R}^n$. Jaká (kvasi-)uspořádání jsou vhodná?

Uspořádání po složkách: množina minimálních prvků množiny $\mathbf{f}(\mathbb{R}^2)$ je konvexní obal bodů $\mathbf{a}_1, \ldots, \mathbf{a}_n$. To je intuitivně jasné (i když přesný důkaz vynecháme): neleží-li $\mathbf{x}$ v tomto konvexním obalu, můžeme změnit $\mathbf{x}$ tak, že se vzdálenost od některých bodů $\mathbf{a}_i$ zmenší a od ostatních se nezvětší. Tedy je hloupost umístit heliport mimo tento konvexní obal.

Max-uspořádání: vede na úlohu

$$\min_{\mathbf{x}\in\mathbb{R}^2} \max_{i=1}^{n} \|\mathbf{a}_i - \mathbf{x}\|_2.$$

Toto je úloha klasické (skalární) optimalizace. Této formulaci se někdy říká *minimaxní* formulace. Minimalizujeme vzdálenost heliportu od nejvzdálenější vesnice. ♦

Rejstřík