Optimalizace

Elektronická skripta předmětu B0B33OPT,
verze ze **17. října 2019**.

Tomáš Werner
`werner@fel.cvut.cz`

Katedra kybernetiky
Fakulta elektrotechnická
České vysoké učení technické

Obsah

Kapitola 1

Značení a základní pojmy

1.1 Značení

Pokud potkáte ve skriptech slovo vysázené **tučně**, jde o nově zavedený pojem, který máte pochopit a zapamatovat si. Ve skriptech tedy nepotkáte slovo ‘definice’, protože všechna tučná slova jsou definice. Slova vysázená *kurzívou* znamenají buď zdůraznění, nebo nový avšak všeobecně známý pojem. Odstavce, věty, důkazy, příklady a cvičení označené hvězdičkou ($\star$) jsou rozšiřující (a často zajímavé) a není nezbytné umět je ke zkoušce.

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

1.1.1 Množiny

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

$\{a_1, \dots, a_n\}$	množina s prvky $a_1, \dots, a_n$
$a \in A$	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 $\{\, a \mid a \in A,\ a \notin B \,\}$
$(a_1, \dots, a_n)$	uspořádaná n-tice prvků $a_1, \dots, a_n$
$A_1 \times \dots \times A_n$	kartézský součin množin, množina $\{\, (a_1, \dots, a_n) \mid a_1 \in A_1, \dots, a_n \in A_n \,\}$
A^n	kartézský součin n stejných množin, $A^n = A \times \dots \times A$ (n-krát)
$\emptyset$	prázdná množina

Číselné množiny budeme značit takto:

$\mathbb{N}$	množina přirozených čísel
$\mathbb{Z}$	množina celých čísel
$\mathbb{Q}$	množina racionálních čísel
$\mathbb{R}$	množina 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 jako v angličtině s desetinnou tečkou, např. 1.23 namísto 1,23.

1.1.2 Zobrazení

Zobrazení (angl. *mapping, map*) z množiny A do množiny B značíme

$$f\colon A \to B \tag{1.1}$$

nebo (méně často) $A \xrightarrow{f} B$. Zobrazení si můžeme představit[1] jako ‘černou skříňku’, která každému prvku $a \in A$ (vzoru) přiřadí právě jeden prvek $b = f(a) \in B$ (obraz). Slovo *funkce* (angl. *function*) znamená přesně totéž jako zobrazení, obvykle se však používá pro zobrazení do číselných množin (tedy $B = \mathbb{R}$, $\mathbb{Z}$, $\mathbb{C}$ apod.). *Transformace* je zobrazení nějaké množiny do sebe, tedy $f\colon A \to A$.

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

$$f(A') = \{\, f(a) \mid a \in A' \,\} = \{\, b \in B \mid b = f(a),\ a \in A' \,\} \tag{1.2}$$

(čárka v definici pravé množiny značí konjunkci, tedy jde o množinu všech $b \in B$ které splňují $b = f(a)$ *a zároveň* $a \in A'$). Např. je-li $A' = \{1, 3, 4, -1\} \subseteq \mathbb{Z}$ a zobrazení $f\colon \mathbb{Z} \to \mathbb{Z}$ je definované předpisem $f(a) = a^2$, 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

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

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

Zobrazení se nazývá

- *injektivní* (neboli prosté), jestliže každý vzor má jiný obraz, tj. $f(a) = f(a') \implies a = a'$.
- *surjektivní* (neboli A na B), jestliže každý obraz má aspoň jeden vzor, tj. $f(A) = B$.
- *bijektivní* (neboli vzájemně jednoznačné), jestliže je zároveň injektivní a surjektivní.

Mějme dvě zobrazení $f\colon A \to B$ a $g\colon B \to C$, neboli $A \xrightarrow{f} B \xrightarrow{g} C$. *Složení* zobrazení f a g je zobrazení $g \circ f\colon A \to C$ definované jako $(g \circ f)(a) = g(f(a))$ pro každé $a \in A$.

[1] Přesná definice je následující: zobrazení $f\colon A \to B$ je podmnožina kartézského součinu $A \times B$ (tedy *relace*) taková, že $(a, b) \in f$ a $(a, b') \in f$ implikuje $b = b'$.

1.1.3 Funkce a zobrazení více reálných proměnných

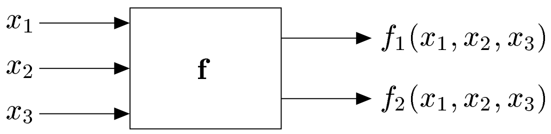
Uspořádané n-tici $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ reálných čísel říkáme (n-rozměrný) **vektor**. Zápis

$$\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m \tag{1.4}$$

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 $\mathbf{f} = (f_1, \ldots, f_m)$. Obrázek ilustruje zobrazení $\mathbf{f}\colon \mathbb{R}^3 \to \mathbb{R}^2$:



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

Je-li $n \in \{1, 2, 3\}$, často budeme proměnné 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 Minimum a infimum

Pro množinu $Y \subseteq \mathbb{R}$ definujeme tyto pojmy (opakování z analýzy):

- **Dolní mez** množiny Y je každé číslo $a \in \mathbb{R}$, pro které $a \le y$ pro všechna $y \in Y$.
- **Infimum** množiny Y je její největší dolní mez. Značíme jej $\inf Y$.
- **Nejmenší prvek** (neboli **minimum**) množiny Y je její dolní mez, která v ní leží. Pokud taková dolní mez existuje, je určena jednoznačně. Značíme $\min Y$.

Horní mez, největší prvek (maximum, $\max Y$) a supremum ($\sup Y$) se definují analogicky.

Minimum či maximum podmnožiny reálných čísel nemusí existovat. Je hlubokou vlastností reálných čísel, že v nich existuje infimum [supremum] každé zdola [shora] omezené podmnožiny. Tato vlastnost množiny $\mathbb{R}$ se nazývá **úplnost**.

Zaveďme dva prvky $-\infty$ a $+\infty$ (které nepatří do množiny $\mathbb{R}$) a definujme $-\infty < a < +\infty$ pro každé $a \in \mathbb{R}$. Pokud množina Y je zdola [shora] neomezená, definujeme $\inf Y = -\infty$ [$\sup Y = +\infty$]. Pro prázdnou množinu definujeme $\inf \emptyset = +\infty$ a $\sup \emptyset = -\infty$.

Příklad 1.1.

1. Množina všech horních mezí intervalu $[0, 1)$ je $[1, +\infty)$.
2. Množina všech horních mezí množiny $\mathbb{R}$ je $\emptyset$.
3. Množina všech horních mezí množiny $\emptyset$ je $\mathbb{R}$.
4. Interval $[0, 1)$ nemá největší prvek, ale má supremum 1.
5. Minimum množiny $\{\, 1/n \mid n \in \mathbb{N} \,\}$ neexistuje, ale infimum je rovno 0.

6. Maximum množiny $\{ x \in \mathbb{Q} \mid x^2 \leq 2 \} = [-\sqrt{2}, \sqrt{2}] \cap \mathbb{Q}$ neexistuje, ale supremum je rovno $\sqrt{2}$.
7. $\max\{1, 2, 3\} = \sup\{1, 2, 3\} = 3$ (minimum a maximum každé konečné množiny existují a jsou rovny infimu a supremu)
8. Množina $\mathbb{R}$ nemá nejmenší prvek, tedy $\min \mathbb{R}$ neexistuje.
9. Prázdná množina $\emptyset$ nemá nejmenší prvek, tedy $\min \emptyset$ neexistuje. $\square$

1.3 Extrémy funkce na množině

Mějme funkci $f\colon X' \to \mathbb{R}$ a množinu $X \subseteq X'$. Nechť $x \in X$ je takové, že $f(x) \leq f(x')$ pro všechna $x' \in X$. Pak x nazveme *minimum funkce* f na množině X, nebo také říkáme, že funkce f *nabývá minima* na množině X v prvku x. Chceme-li být přesnější, takový prvek x nazýváme *minimalizující argument* nebo *argument minima* (angl. *minimizer*) funkce na množině. Číslo $f(x)$ nazýváme *minimální hodnotou* funkce f na množině X a píšeme

$$f(x) = \min_{x' \in X} f(x'). \tag{1.5}$$

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

$$\operatorname*{argmin}_{x \in X} f(x) = \left\{ x \in X \,\middle|\, f(x) \leq f(x') \ \forall x' \in X \right\} \subseteq X. \tag{1.6}$$

Podobně definujeme maximum funkce na množině. Minima a maxima funkce se souhrnně nazývají její *extrémy* nebo *optima*. Pokud odkaz na množinu X chybí, myslí se $X = X'$.

Je užitečné se k minimu funkce na množině postavit poněkud abstraktněji. Nechť $Y \subseteq \mathbb{R}$. Prvek $y \in Y$ nazveme *nejmenší prvek* (nebo také *minimální prvek*) množiny Y, jestliže $y \leq y'$ pro všechna $y' \in Y$. Nejmenší prvek značíme $\min Y$. Ne každá množina $Y \subseteq \mathbb{R}$ má nejmenší prvek (např. interval $(0, 1]$ ho nemá). Na druhou stranu, Y má nejvýše jeden minimální prvek.

Označme nyní

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

obraz množiny X funkcí f. Pokud množina $f(X)$ má nejmenší prvek, definujeme

$$\min_{x \in X} f(x) = \min\{ f(x) \mid x \in X \} = \min f(X). \tag{1.7}$$

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 (1.6) prázdná.

Příklad 1.2.

- 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[2] $\max\limits_{i=1}^{5} a_i = 3$, $\operatorname*{argmax}\limits_{i=1}^{5} a_i = \{3, 5\}$. $\square$

1.4 Úloha spojité optimalizace

Matematickou optimalizací rozumíme hledání minima nějaké funkce $f\colon X' \to \mathbb{R}$ na nějaké množině $X \subseteq X'$. Tato formulace je velmi obecná, neboť množina X může být zcela libovolná. 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

$$g_i(x_1, \ldots, x_n) \le 0, \quad i = 1, \ldots, m \tag{1.8a}$$
$$h_i(x_1, \ldots, x_n) = 0, \quad i = 1, \ldots, l \tag{1.8b}$$

pro dané funkce $g_1, \ldots, g_m, h_1, \ldots, h_l\colon \mathbb{R}^n \to \mathbb{R}$. 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 $f(\mathbf{x})$ za podmínek (1.8) příp. (1.10). To lze psát také jako

$$\begin{array}{rl} \min & f(x_1, \ldots, x_n) \\ \text{za podmínek} & g_i(x_1, \ldots, x_n) \le 0, \quad i = 1, \ldots, m \\ & h_i(x_1, \ldots, x_n) = 0, \quad i = 1, \ldots, l \\ & x_1, \ldots, x_n \in \mathbb{R} \end{array} \tag{1.9}$$

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

Soustavu (1.8) lze napsat kratčeji ve vektorovém značení jako

$$\mathbf{g}(\mathbf{x}) \le \mathbf{0}, \tag{1.10a}$$
$$\mathbf{h}(\mathbf{x}) = \mathbf{0}, \tag{1.10b}$$

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

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

Ve shodě se značením (1.3) se pak úloha (1.9) může zapsat jako

$$\min\{\, f(\mathbf{x}) \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{g}(\mathbf{x}) \le \mathbf{0},\ \mathbf{h}(\mathbf{x}) = \mathbf{0} \,\}. \tag{1.11}$$

V matematické analýze se řešením úlohy (1.9) říká *extrémy funkce f vázané podmínkami* (1.8). Pokud omezení chybí, mluví se o *volných extrémech* funkce f. V matematické optimalizaci se vžilo poněkud odlišné názvosloví:

- Funkce f se nazývá *účelová* (také pokutová, cenová, kriteriální) funkce.

[2] Místo $\max\limits_{i=1}^{5} a_i$ se obvykle píše $\max\limits_{i=1,\ldots,5} a_i$. Používáme zde první způsob v analogii se značením $\sum\limits_{i=1}^{5} a_i$.

- Prvky množiny X se nazývají *přípustná řešení*, což je vlastně protimluv, protože nemusí být řešeními úlohy (1.9). Prvkům množiny $\operatorname{argmin}_{\mathbf{x}\in X} f(\mathbf{x})$ se pak říká *optimální řešení* nebo *optimální argumenty* úlohy. Číslu $\min_{\mathbf{x}\in X} f(\mathbf{x})$ se říká *optimální hodnota* úlohy.
- Rovnice a nerovnice (1.8) se nazývají *omezující podmínky*, krátce *omezení*.
- Omezení (1.8a) příp. (1.8b) se nazývají omezení *typu nerovnosti* příp. *typu rovnosti*. Pokud omezení chybí ($m = l = 0$), jedná se o optimalizaci *bez omezení*.
- Pokud je omezení typu nerovnosti $g_i(\mathbf{x}) \le 0$ splněno s rovností, tedy $g_i(\mathbf{x}) = 0$, říkáme, že je v bodě $\mathbf{x}$ *aktivní*.
- Pokud $X \neq \emptyset$, úloha se nazývá *přípustná*, v opačném případě ($X = \emptyset$) je *nepřípustná*.

Dále uvedeme několik příkladů formulace (a často i řešení) úloh ve tvaru (1.9). 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 1.3. 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

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

neboli

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

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 extrémně nešikovné snažit se ji řešit formalismem Lagrangeových multiplikátorů či KKT podmínek (pokud je někdo zná). Místo toho z podmínky $2x + y = 100$ vyjádříme y (přesněji: rovnice $2x + y = 100$ je ekvivalentní rovnici $y = 100 - 2x$) a dosadíme do původní úlohy. Tím dostaneme úlohu s jednou proměnnou

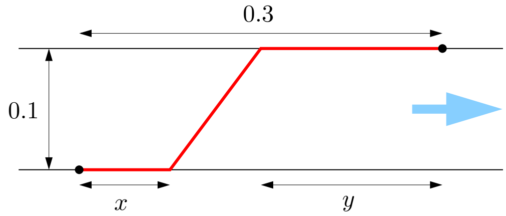
$$\max\{\, x(100 - 2x) \mid 0 \le x \le 50 \,\},$$

což někdo raději píše jako $\max\limits_{0 \le x \le 50} x(100 - 2x)$ 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 $0 \le x \le 50$. Body na krajích intervalu mají menší hodnotu kritéria, tedy nejsou optimální. Dosazením dostaneme $y = 100 - 2x = 50$. □

Příklad 1.4. 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.

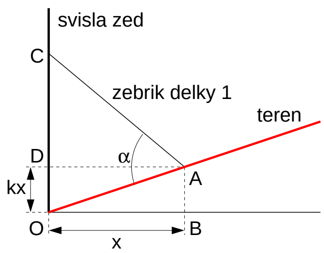
Může mít dráha jiný tvar? Těžko, pokud přijmeme, že dráha, kterou se dostaneme nejrychleji z bodu do bodu za předpokladu, že v každém místě prostoru se pohybujeme stejně rychle, je úsečka spojující tyto dva body.

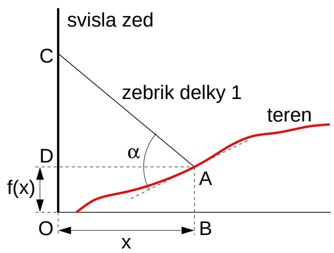
Celkový čas je $t = \frac{1}{3}x + \sqrt{0.1^2 + (0.3 - x - y)^2} + \frac{1}{3}y$ (hodin). Ten chceme minimalizovat pro proměnné $x, y \in \mathbb{R}$. Všimněte si, že nepotřebujeme omezení $x, y \geq 0$ ani $x + y \leq 0.3$, neboť i bez nich dostaneme přípustné dráhy. Ovšem tyto dráhy očividně nejsou 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. $\square$

Příklad 1.5. 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 jedné proměnné: podmínka stacionarity je $k = x/\sqrt{1 - x^2}$, z toho $x = 1/\sqrt{1 + 1/k^2}$. Jedná se o maximum, což můžeme ověřit pomocí druhé derivace (musí být záporná), ale je to i vidět.

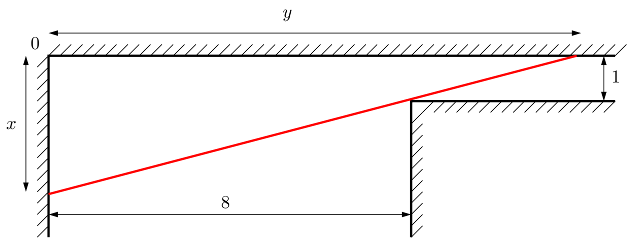
Ukážeme, že pro optimální x bude žebřík kolmý k terénu (tj. $\alpha = \pi/2$). Z podmínky $k = x/\sqrt{1 - x^2}$ vidíme, že $|AD|/|CD| = k$, protože $|AD| = x$ a $|CD| = \sqrt{1 - x^2}$. Ale z definice směrnice je také $kx/x = |DO|/|AD| = |AB|/|BO| = k$. Tedy pravoúhlé trojúhelníky DCA, BOA a DAO jsou 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 je $f'(x) = x/\sqrt{1 - x^2}$. To opět znamená, že je žebřík kolmý k terénu (odvození podobné jako minule). □

Příklad 1.6. 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 (viz obrázek):



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 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. Stacionarni podminka je $2\sin\alpha = \cos\alpha$, tedy $\tan\alpha = \frac{1}{2}$. Lze ověřit, že jde o minimum. □

Příklad 1.7. Na zahradě vám rostou melouny. Dnes je jich tam 200 kg. Každý den melouny narostou 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?).

Prakticky zaměřený čtenář namítne, že žádný zelinář by to takto nedělal. 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.

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 1.8. Najděte nejmenší kruh, který obsahuje dané body $\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^2$.

Kruh je množina $\{\, \mathbf{x} \in \mathbb{R}^2 \mid \|\mathbf{x} - \mathbf{c}\| \leq r \,\}$, kde

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

označuje eukleidovskou normu vektoru $\mathbf{x} \in \mathbb{R}^n$. Úlohu lze napsat jako

$$\begin{aligned} \min \quad & r \\ \text{za podmínek} \quad & \|\mathbf{a}_i - \mathbf{c}\| \leq r, \quad i = 1, \ldots, m \\ & \mathbf{c} \in \mathbb{R}^2 \\ & r \in \mathbb{R} \end{aligned} \tag{1.12}$$

Mohla by tam být ještě podmínka $r \geq 0$, ta je ale zbytečná, protože norma je nezáporná.

Tato úloha je konvexní a patří do třídy programování na kuželu druhého řádu (SOCP), což si ovšem řekneme až mnohem později. Na řešení SOCP existují numerické metody, které bychom mohli mechanicky použít. Algoritmus řešící naši úlohu lze ale vymyslet i se znalostmi, které máte nyní, a to následující úvahou.

Nejmenší kruh bude mít vždy na své hranici (tj. kružnici) aspoň jeden bod, protože jinak by ho bylo možno zmenšit a tedy by nebyl optimální. Optimální kruh může mít na hranici libovolný počet bodů (třeba všechny, když budou dané body ležet na kružnici; nebo jen jeden když bude $m = 1$, pak bude $r = 0$). Zároveň víme, že kružnice je jednoznačně určena třemi (nekolinárními) body. Úlohu tedy řeší tento algoritmus: Projdi všechny jedno-, dvou- a tříprvkové podmnožiny daných bodů. Pro každou podmnožinu najdi nejmenší kružnici, která jí prochází a zjisti, jestli ostatní body leží v tomto kruhu (jehož hranicí je ta kružnice). Nejmenší z těch kruhů je optimální.

Vidíme, že pro tuto úvahu jsme matematický zápis úlohy (1.12) vlastně vůbec nepotřebovali.

Zobecněme nyní úlohu z roviny $\mathbb{R}^2$ do n-rozměrného prostoru $\mathbb{R}^n$. Tedy hledáme nejmenší n-rozměrnou kouli, která obsahuje dané body $\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^n$. Protože n-rozměrná koule je určena $n+1$ body v obecné poloze, museli bychom procházet všechny jedno-, dvou- až $(n+1)$-prvkové podmnožiny. Pro velké n by tento algoritmus už nebyl vhodný, protože množství podmnožin by bylo příliš velké. □

Příklad 1.9. Řešme tuto lineární soustavu třech 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. Nejčastěji užívaná formulace je minimalizovat součet čtverců[3] 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. Nutná (a zde i 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. Uvidíme, že úloha je konvexní a dá se převést na lineární programování. □

Příklad 1.10. Víme, že 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 $a, b, c \in \mathbb{R}$. Tato formulace má statistické odůvodnění (z principu maximální věrohodnosti) za předpokladu, že čísla $x_1, \ldots, x_n$ měříme přesně a nepřesnosti měření čísel $y_1, \ldots, y_m$ jsou i.i.d. normální s nulovou střední hodnotou a stejnou variancí. Minimum kvadratické funkce f opět najdeme porovnáním parciálních derivací s nulou. □

Příklad 1.11. 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).

Víme, že i-tá (kde $i = 1, 2$) přímka 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

$$f(\alpha_1, \alpha_2) = \|\mathbf{a}_1 + \alpha_1(\mathbf{b}_1 - \mathbf{a}_1) - \mathbf{a}_2 - \alpha_2(\mathbf{b}_2 - \mathbf{a}_2)\|^2. \tag{1.13}$$

Opět je to kvadratická funkce dvou proměnných. □

Příklad 1.12. Pán 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.

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

Tuto úlohu lze formalizovat takto:

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

Optimální řešení je $l = 20$ kg lupínků a $h = 40$ kg hranolků. V tomto případě omezení $x, y \geq 0$ nejsou zbytečná. Kdybychom je vynechali, úloha by byla neomezená. □

Příklad 1.13. Jsou dána čísla $c_1, \ldots, c_n \in \mathbb{R}$. Najdi čísla $x_1, \ldots, x_n \in \mathbb{R}$, která maximalizují výraz $\sum_{i=1}^n c_i x_i$ za podmínek $0 \leq x_i \leq 1$ $(i = 1, \ldots, n)$.

Tato úloha patří do třídy lineárního programování. Ovšem dokážeme ji vyřešit jednoduchou úvahou. Tvrdíme, že optimum se nabývá tehdy, když $x_i = 0$ pro $c_i < 0$ a $x_i = 1$ pro $c_i > 0$ (pro $c_i = 0$ je x_i libovolné v intervalu $[0, 1]$). Kdyby to tak totiž nebylo, mohli bychom účelovou funkci zvětšit zmenšením nějakého x_i pro $c_i < 0$ nebo zvětšením pro $c_i > 0$. Optimální hodnota je tedy $\sum_{i=1}^n \max\{0, c_i\}$, tj. součet kladných čísel c_i. □

Příklad 1.14. Jsou dány body $\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^n$ a hledáme vektor $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$, který minimalizuje (bez omezujících podmínek) funkci

$$f(\mathbf{x}) = \sum_{i=1}^m \|\mathbf{x} - \mathbf{a}_i\|^2. \tag{1.14}$$

Označíme-li $\mathbf{a}_i = (a_{i1}, \ldots, a_{in})$, 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 m 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 1.15. jsou dána čísla $a_1, \ldots, a_m \in \mathbb{R}$. Najděte minimum funkce

$$f(x) = \sum_{i=1}^m |x - a_i|. \tag{1.15}$$

Seřadíme čísla $a_1, \ldots, a_m$ vzestupně, tedy předpokládáme $a_1 \leq a_2 \leq \cdots \geq a_{m-1} \leq a_m$. Funkce f je po částech lineární (přesněji: je afinní), je diferencovatelná všude kromě bodů a_i. Derivace je konstantní pro každý interval (a_{i-1}, a_i). Najděte hodnotu těchto derivací (nejpozději teď je nutno si nakreslit obrázek!). Z toho usoudíme, kde je funkce 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 $a_1, \ldots, a_m$ (pro sudé m se konvencí bere číslo $\frac{1}{2}(a_{m/2}, a_{m/2+1})$. □

Příklad 1.16. Jsou dány body $\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^n$. Najděte minimum funkce

$$f(\mathbf{x}) = \sum_{i=1}^m \|\mathbf{x} - \mathbf{a}_i\|. \tag{1.16}$$

Pro $n = 1$ se funkce (1.16) redukuje na (1.15). 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 (1.16) 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[4]. 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 $f(\mathbf{x})$. $\square$

Příklad 1.17. 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 $\mathbf{a}_i$ k nejbližšímu bodu $\mathbf{x}_j$ byla co nejmenší. Tedy minimalizujeme účelovou funkci

$$f(\mathbf{x}_1, \ldots, \mathbf{x}_k) = \sum_{i=1}^{m} \min_{j=1}^{k} \|\mathbf{a}_i - \mathbf{x}_j\| \tag{1.17}$$

po neznámé vektory $\mathbf{x}_1, \ldots, \mathbf{x}_k \in \mathbb{R}^n$. Úloha je známá jako *shlukování*. 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 k nejbližší cisterně byla co nejmenší.

Není znám (a pravděpodobně nikdy nebude) 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 (1.17) za prakticky přijatelný čas. Lze totiž dokázat, že úloha je tzv. NP-těžká[5]. V praktické situaci často použijeme algoritmus, který najde pouze přibližné (typicky lokální) optimum, např. *k-means*. $\square$

Příklad 1.18. 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*)[6] v grafu zní

$$\begin{aligned} \max \quad & \sum_{\{i,j\} \in E} c_{ij} x_i x_j \\ \text{za podmínek} \quad & x_i^2 = 1, \quad i \in V \\ & x_i \in \mathbb{R}, \quad i \in V \end{aligned} \tag{1.18}$$

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 množina přípustných řešení má konečný počet ($2^{|V|}$) prvků a jedná se tedy o kombinatorickou úlohu. Všimněte si že, přísně vzato, omezení nemůžeme psát jako $x_i \in \{-1, 1\}$, protože to není ve tvaru (1.9). Jedná se o klasickou NP-těžkou úlohu, proto už pro některé poměrně malé úlohy je prakticky nemožné najít globální optimum. $\square$

[4] 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.

[5] Tento pojem patří do *teorie složitosti algoritmů*, kterou jste ještě nebrali. Zde jen řekneme, že pro NP-těžkou úlohu jen malá naděje, že bude někdy nalezen algoritmus, který by úlohu shlukování 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.

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

1.5 Cvičení

1.1. Vyřešte následující úlohy, přičemž slovní úlohy nejdříve formulujte ve tvaru (1.9). Stačí vám k tomu zdravý rozum a derivace funkcí jedné proměnné.

a) $\min\{\, x^2 + y^2 \mid x > 0,\ y > 0,\ xy \geq 1 \,\}$
b) $\min\{\, (x-2)^2 + (y-1)^2 \mid x^2 \leq 1,\ y^2 \leq 1 \,\}$
c) $\min\{\, x \mid x \in \mathbb{R},\ x \geq a_i\ (i = 1, \ldots, n) \,\}$ pro daná $a_1, \ldots, a_n \in \mathbb{R}$
d) Máte vyrobit papírovou krabici (včetně víka) o objemu 72 litrů, jejíž délka je dvojnásobek její šířky. Jaké budou její rozměry, má-li se na ní spotřebovat co nejméně papíru? Tloušťka stěn je zanedbatelná.
e) Jaké má rozměry válec s jednotkovým objemem a nejmenším povrchem?
f) Najděte rozměry půllitru, 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 a (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) ($\star$) (na pracovní prostor robotu) Do mezikruží (množinový rozdíl dvou soustředných kruhů o poloměrech $R > r > 0$) máme vepsat elipsu s co největším obsahem. Nejen hranice elipsy, ale i její vnitřek musí ležet v mezikruží.

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

Část I

Použití lineární algebry v optimalizaci

Kapitola 2

Maticová algebra

Cílem této kapitoly je zopakovat si základní maticové pojmy a naučit se manipulovat s maticovými výrazy a rovnicemi, aniž byste zatím museli mnoho vědět o lineární algebře.

Reálná **matice** rozměru $m \times n$ 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. 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 $a_{11}, \ldots, a_{pp}$, 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. Čtvercovou ($m = n$) diagonální matici značíme $\mathbf{A} = \operatorname{diag}(a_{11}, a_{22}, \ldots, a_{nn})$.
- **Nulová matice** má všechny prvky nulové. Značíme ji $\mathbf{0}_{m\times n}$ (pokud jsou rozměry jasné z kontextu, pak pouze $\mathbf{0}$).
- **Jednotková matice** je čtvercová diagonální, jejíž diagonální prvky jsou jedničky. Značíme ji $\mathbf{I}_n$ (pokud jsou rozměry jasné z kontextu, pak pouze $\mathbf{I}$).
- **Horní [dolní] trojúhelníková matice** má $a_{ij} = 0$ pro všechna $i > j$ [$i < j$]. Všimněte si, že horní/dolní trojúhelníková matice nemusí být čtvercová.

2.1 Binární operace s maticemi

V algebře reálných matic se reálná čísla nazývají také **skaláry**[2]. Na maticích jsou definovány následující operace:

[1] Formálněji můžeme matice definovat jako zobrazení $\{1, \ldots, m\} \times \{1, \ldots, n\} \to \mathbb{R}$ (podobně, vektor $(x_1, \ldots, x_n) \in \mathbb{R}^n$ lze považovat za zobrazení $\{1, \ldots, n\} \to \mathbb{R}$). Množinu $\mathbb{R}^{m\times n}$ můžeme vidět jako množinu $(\mathbb{R}^m)^n$. V každém případě ale musíme dodat, že matici značíme tabulkou v hranatých závorkách.

[2] Přesněji, pohlížíme-li na množinu všech matic rozměru $m \times n$ jako na lineární prostor, jedná se o skaláry tohoto lineárního prostoru.

- Součin skaláru $\alpha \in \mathbb{R}$ 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 $\dfrac{1}{\alpha}\mathbf{A}$ píšeme 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 $\mathbf{A} - \mathbf{B} = \mathbf{A} + (-\mathbf{B})$.
- **Maticový součin** matic $\mathbf{A} \in \mathbb{R}^{m\times p}$ a $\mathbf{B} \in \mathbb{R}^{p\times n}$ je matice $\mathbf{C} = \mathbf{AB} \in \mathbb{R}^{m\times n}$ s prvky

$$c_{ij} = \sum_{k=1}^{p} a_{ik} b_{kj}, \quad i = 1, \ldots, m, \ j = 1, \ldots, n. \tag{2.1}$$

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

Vlastnosti maticového součinu:

- $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$
- $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$ a $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$
- $\mathbf{AI}_n = \mathbf{A} = \mathbf{I}_m\mathbf{A}$
- $(\alpha\mathbf{A})\mathbf{B} = \mathbf{A}(\alpha\mathbf{B}) = \alpha(\mathbf{AB})$

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

Poznamenejme, že výraz $\alpha\mathbf{A}$ pro $\alpha \in \mathbb{R}$ nelze považovat za maticový součin 'matice' α rozměru 1×1 a matice $\mathbf{A}$, protože vnitřní rozměr matic by byl obecně různý. Tedy násobení matice skalárem je jiná operace, než maticový součin.

Pro čtvercovou matici $\mathbf{A}$ značí $\mathbf{A}^k$ maticový součin k matic $\mathbf{A}$.

2.2 Transpozice a symetrie

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 $a_{ii} = 0$).

2.3 Hodnost

Hodnost matice je dimenze lineárního obalu jejích sloupců. Značíme ji $\operatorname{rank}\mathbf{A}$. Platí

$$\operatorname{rank}\mathbf{A} = \operatorname{rank}(\mathbf{A}^T), \tag{2.2}$$

tedy[3] hodnost je také rovna dimenzi lineárního obalu řádků. Tato rovnost není zdaleka očividná, dokážeme ji v §3.2.3 (důkaz najdete také v každé učebnici lineární algebry).

[3] Někdo by raději psal $\operatorname{rank}(A)$ než $\operatorname{rank}\mathbf{A}$, ale před rozmožením počítačů bylo běžnější psát např. $\sin x$ než $\sin(x)$.

Z (2.2) plyne, že pro matici $\mathbf{A} \in \mathbb{R}^{m\times n}$ je

$$\operatorname{rank}\mathbf{A} \leq \min\{m, n\}. \tag{2.3}$$

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. Čtvercová matice s plnou hodností se nazývá **regulární**. Čtvercová matice, která nemá plnou hodnost, se nazývá **singulární**.

Je užitečné si pamatovat nerovnost (důkaz viz Cvičení 3.13)

$$\operatorname{rank}(\mathbf{AB}) \leq \min\{\operatorname{rank}\mathbf{A}, \operatorname{rank}\mathbf{B}\}. \tag{2.4}$$

2.4 Inverze

Když platí

$$\mathbf{AB} = \mathbf{I}, \tag{2.5}$$

matice $\mathbf{B}$ se nazývá **pravá inverze** matice $\mathbf{A}$ a matice $\mathbf{A}$ se nazývá **levá inverze** matice $\mathbf{B}$. Pravá či levá inverze nemusí existovat nebo nemusí být jediná. Pravá inverze matice existuje právě tehdy, má-li matice lineárně nezávislé řádky (dokážeme to později ve Větě 3.5, zatím si to pamatujte). Levá inverze matice existuje právě tehdy, má-li matice lineárně nezávislé sloupce (což plyne z předchozího, když si (2.5) napíšeme jako $\mathbf{B}^T\mathbf{A}^T = \mathbf{I}$; viz také Věta 3.7).

Jestliže matice $\mathbf{A}$ je čtvercová a má pravou inverzi, má zároveň i levou inverzi a obě inverze se rovnají a jsou jediné. Opravdu: je-li $\mathbf{AX} = \mathbf{I}$ a $\mathbf{YA} = \mathbf{I}$, pak $\mathbf{YAX} = \mathbf{Y} = \mathbf{X}$. Protože toto platí pro každou levou inverzi a každou pravou inverzi, jsou zároveň jediné. Pak mluvíme pouze o **inverzi** matice a značíme ji $\mathbf{A}^{-1}$. Matice má inverzi, právě když je regulární. 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 $\mathbf{A}^{-T}$.

2.5 Determinant

Determinant je funkce $\mathbb{R}^{n\times n} \to \mathbb{R}$ (tedy přiřazuje čtvercové matici skalár) definovaná jako

$$\det\mathbf{A} = \sum_{\sigma} \operatorname{sgn}\sigma \prod_{i=1}^{n} a_{i\,\sigma(i)}, \tag{2.6}$$

kde sčítáme přes všechny permutace n prvků $\sigma\colon \{1,\ldots,n\} \to \{1,\ldots,n\}$, 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 dvou sloupců změní znaménko determinantu.

2.6 Matice s jedním sloupcem nebo jedním řádkem

Matice s jediným sloupcem (tedy prvek $\mathbb{R}^{m\times 1}$) se také nazývá **sloupcový vektor**[4]. Matice s jediným řádkem (tedy prvek $\mathbb{R}^{1\times n}$) se také nazývá **řádkový vektor**.

Lineární prostor $\mathbb{R}^{m\times 1}$ všech matic s jediným sloupcem je 'skoro stejný' jako lineární prostor $\mathbb{R}^m$ všech uspořádaných m-tic $(x_1,\dots,x_m)$. Proto je zvykem tyto prostory ztotožnit a bez upozornění přecházet mezi oběma významy. Prvkům

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

tohoto prostoru budeme říkat krátce **vektory**. Jinak řečeno, slovem *vektor* (bez přívlastku) budeme rozumět *sloupcový vektor* nebo také uspořánou n-tici čísel[5].

Všimněme si případů, kdy se v maticovém součinu vyskytují vektory:

- Pro matici $\mathbf{A} \in \mathbb{R}^{m\times n}$ a vektor $\mathbf{x} \in \mathbb{R}^n$, výraz $\mathbf{Ax}$ je 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 matici $\mathbf{A} \in \mathbb{R}^{m\times n}$ a vektor $\mathbf{x} \in \mathbb{R}^m$, výraz $\mathbf{x}^T\mathbf{A}$ je 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 $\mathbf{x}^T\mathbf{y} = x_1y_1 + \cdots + x_ny_n$ je maticový součin řádkového vektoru $\mathbf{x}^T$ a sloupcového vektoru $\mathbf{y}$, jehož výsledkem je skalár. Je to standardní skalární součin vektorů $\mathbf{x}$ a $\mathbf{y}$ (více si o něm řekneme §4.1).
- Pro $\mathbf{x} \in \mathbb{R}^m$ a $\mathbf{y} \in \mathbb{R}^n$ je $\mathbf{xy}^T$ matice rozměru $m \times n$, které se někdy říká **vnější součin** vektorů $\mathbf{x}$ a $\mathbf{y}$ nebo **dyáda**.

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 $\mathbf{1}^T\mathbf{x} = x_1 + \cdots + x_n$.

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, $\begin{bmatrix}\mathbf{e}_1 & \cdots & \mathbf{e}_n\end{bmatrix} = \mathbf{I}_n$.

[4] V lineární algebře má slovo *vektor* obecnější význam než v maticové algebře: znamená prvek lineárního prostoru (který se někdy také nazývá *vektorový prostor*).

[5] Totéž bychom samozřejmě mohli udělat s řádky (a někdo to tak i dělá, např. počítačoví grafici).

2.7 Matice sestavené z bloků

Matici je možno sestavit z několika jejích **podmatic** (zvaných též **bloky**), např.

$$\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{bmatrix}, \quad \begin{bmatrix}\mathbf{A} & \mathbf{B}\end{bmatrix}, \quad \begin{bmatrix}\mathbf{A}\\ \mathbf{B}\end{bmatrix}, \quad \begin{bmatrix}\mathbf{A} & \mathbf{I}\\ \mathbf{0} & \mathbf{D}\end{bmatrix}. \tag{2.7}$$

Rozměry jednotlivých bloků musí být slučitelné, 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 $\mathbf{A}$ a $\mathbf{D}$.

Při násobení matic sestavených z bloků je užitečné pravidlo, že lze formálně užít obvyklý postup pro násobení matic, pouze místo prvků matice si představíme bloky.

Příklad 2.1. Jsou-li a, b, c, d, x, y skaláry, máme

$$\begin{bmatrix}a & b\\ c & d\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix} = \begin{bmatrix}ax+by\\ cx+dy\end{bmatrix}.$$

Jsou-li $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{X}, \mathbf{Y}$ matice vhodných rozměrů, máme tedy (ověřte dle vzorce (2.1)!)

$$\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{bmatrix}\begin{bmatrix}\mathbf{X}\\ \mathbf{Y}\end{bmatrix} = \begin{bmatrix}\mathbf{AX}+\mathbf{BY}\\ \mathbf{CX}+\mathbf{DY}\end{bmatrix}.$$

□

Často je užitečné vnímat matici $\mathbf{A} \in \mathbb{R}^{m\times n}$ jako matici sestavenou z bloků

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

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 bloků

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

kde řádkové vektory $\mathbf{a}_1^T, \ldots, \mathbf{a}_m^T$ 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 řádků a matici $\mathbf{B} \in \mathbb{R}^{p\times n}$ pomocí sloupců, je

$$\mathbf{AB} = \begin{bmatrix}\mathbf{a}_1^T\\ \vdots\\ \mathbf{a}_m^T\end{bmatrix}\begin{bmatrix}\mathbf{b}_1 & \cdots & \mathbf{b}_n\end{bmatrix} = \begin{bmatrix}\mathbf{Ab}_1 & \cdots & \mathbf{Ab}_n\end{bmatrix} = \begin{bmatrix}\mathbf{a}_1^T\mathbf{B}\\ \vdots\\ \mathbf{a}_m^T\mathbf{B}\end{bmatrix} = \begin{bmatrix}\mathbf{a}_1^T\mathbf{b}_1 & \cdots & \mathbf{a}_1^T\mathbf{b}_n\\ \vdots & \ddots & \vdots\\ \mathbf{a}_m^T\mathbf{b}_1 & \cdots & \mathbf{a}_m^T\mathbf{b}_n\end{bmatrix}. \tag{2.8}$$

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

$$\mathbf{AB} = \begin{bmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_p\end{bmatrix}\begin{bmatrix}\mathbf{b}_1^T\\ \vdots\\ \mathbf{b}_p^T\end{bmatrix} = \mathbf{a}_1\mathbf{b}_1^T + \cdots + \mathbf{a}_p\mathbf{b}_p^T. \tag{2.9}$$

2.8 Co je soustava lineárních rovnic?

Soustava rovnic je *lineární*, jestliže se v žádné rovnici proměné nevyskytují v mocninách (např. x^2) ani v součinech (např. xy). Tedy, pravá i levá strana každé rovnice je polynom nejvýše prvního stupně (ale ve více proměnných, viz zač. §6). Když pojmenujeme proměnné jako $x_1, \ldots, x_n$, každou lineární soustavu m rovnic s n neznámými jde zapsat jako

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

neboli

$$\sum_{j=1}^{n} a_{ij}x_j = b_i, \quad i = 1, \ldots, m.$$

neboli

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \tag{2.10}$$

kde $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{x} \in \mathbb{R}^n$. Soustava je **homogenní** pokud $\mathbf{b} = \mathbf{0}$ (tj. každé z čísel $b_1, \ldots, b_m$ je nulové), v opačném případě (aspoň jedno z čísel $b_1, \ldots, b_m$ je nenulové) je **nehomogenní**.

Chceme-li řešit lineární soustavu na počítači, příslušné algoritmy většinou vyžadují soustavu ve tvaru (2.10). Ovšem ne vždy dostaneme lineární soustavu v tomto tvaru. Přepsat takové soustavy do tvaru (2.10) vyžaduje cvik, viz následující příklady a Cvičení 2.7.

Příklad 2.2. Soustava $\{\, \mathbf{A}\mathbf{x} = \mathbf{b},\ \mathbf{A}^T\mathbf{y} = \mathbf{x} \,\}$, 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. □

Příklad 2.3. Soustava $\mathbf{A}\mathbf{X} + \mathbf{X}^T\mathbf{B}^T = \mathbf{C}$, kde $\mathbf{X} \in \mathbb{R}^{m\times n}$ je neznámá a $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times m}$ a $\mathbf{C} \in \mathbb{R}^{n\times n}$ jsou dány. Je to lineární soustava n^2 rovnic s mn neznámými. □

Příklad 2.4. Soustava

$$\begin{aligned} \sum_{j=1}^{n} x_{ij} &= a_i, \quad i = 1, \ldots, m, \\ \sum_{i=1}^{m} x_{ij} &= b_j, \quad j = 1, \ldots, n, \end{aligned}$$

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 lineární soustava $m + n$ rovnic s mn neznámými. □

2.9 Maticové zločiny

Při manipulaci s maticovými výrazy a rovnicemi dělají začátečníci někdy hrubé chyby, kterých se lze při alespoň minimální pozornosti vyhnout. Takové chyby jsou neomluvitelné. Uveďme typické příklady těchto zločinů.

2.9.1 Výraz je nesmyslný kvůli rozměrům matic

Jako první příklad 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 $\mathbf{A}\mathbf{B}^{-1}$ nebo $\mathbf{B}^{-1}\mathbf{A}$.
- Inverze čtvercové, ale evidentně singulární matice. Např. $(\mathbf{w}\mathbf{w}^T)^{-1}$, kde $\mathbf{w} \in \mathbb{R}^3$. Singularita matice plyne např. z (2.4).
- Předpoklad, že existuje pravá inverze úzké matice. Ale úzká matice má lineárně závislé řádky, proto nemá pravou inverzi. Napíšeme-li proto $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$ pro matici $\mathbf{Q} \in \mathbb{R}^{5\times 3}$, je to zločin.

Příklad 2.5. Vidíme-li výraz $(\mathbf{A}^T\mathbf{B})^{-1}$, musí nám okamžitě hlavou proběhnout tyto úvahy o rozměrech matic $\mathbf{A} \in \mathbb{R}^{m\times n}$ a $\mathbf{B} \in \mathbb{R}^{k\times p}$:

- Aby byly kompatibilní velikosti matic v násobení, musí být $m = k$. (Výjimkou je případ, kdy $\mathbf{A}$ je skalár nebo $\mathbf{B}$ je skalár – pak by $\mathbf{A}^T\mathbf{B}$ byl sice nezvyklý ale korektní zápis.)
- Jelikož součin $\mathbf{A}^T\mathbf{B}$ je rozměru $n \times p$, musí být $n = p$, protože invertovat můžeme jen čtvercovou matici. Teď tedy víme, že obě matice musí mít stejný rozměr.
- Z nerovnosti (2.4) je jasné, že pokud by $\mathbf{A}^T$ byla úzká nebo $\mathbf{B}$ široká, $\mathbf{A}^T\mathbf{B}$ by byla singulární a tedy by neměla inverzi. Abychom se tomu vyhnuli, musí být obě matice buď čtvercové nebo úzké, $m \geq n$.

Závěr: Aby výraz $(\mathbf{A}^T\mathbf{B})^{-1}$ měl smysl, je nutné, aby matice $\mathbf{A}, \mathbf{B}$ měly stejný rozměr a byly čtvercové nebo úzké. □

2.9.2 Použití neexistujících maticových identit

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

- $(\mathbf{A}\mathbf{B})^T = \mathbf{A}^T\mathbf{B}^T$ (pokud v maticovém součinu $\mathbf{A}^T\mathbf{B}^T$ je vnitřní rozměr různý, je to chyba už kvůli rozměrům matic)
- $(\mathbf{A}\mathbf{B})^{-1} = \mathbf{A}^{-1}\mathbf{B}^{-1}$ (pro nečtvercové matice je to také syntaktická chyba, pro čtvercové ale singulární matice je to také chyba kvůli rozměrům matic)
- $(\mathbf{A}+\mathbf{B})^2 = \mathbf{A}^2 + 2\mathbf{A}\mathbf{B} + \mathbf{B}^2$. Tato identita se opírá o velice užitečnou (avšak neexistující) identitu $\mathbf{A}\mathbf{B} = \mathbf{B}\mathbf{A}$. Správně je $(\mathbf{A}+\mathbf{B})^2 = \mathbf{A}^2 + \mathbf{A}\mathbf{B} + \mathbf{B}\mathbf{A} + \mathbf{B}^2$.

2.9.3 Neekvivalentní úpravy (ne)rovnic

Zde pachatel udělá chybný úsudek při *neekvivalentní úpravě* rovnice či nerovnice. 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 neplatí $a^2 = b^2 \implies a = b$. Příklady:

- Student si myslí, že $\mathbf{a}^T\mathbf{x} = \mathbf{a}^T\mathbf{y} \implies \mathbf{x} = \mathbf{y}$ (není pravda, 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 si myslí, že $\mathbf{A}^T\mathbf{A} = \mathbf{B}^T\mathbf{B} \implies \mathbf{A} = \mathbf{B}$ (není pravda dokonce ani pro skaláry).
- Student řeší soustavu rovnic $\{\,\mathbf{a}^T\mathbf{x} = 0,\ \mathbf{x}^T\mathbf{x} = 1\,\}$, kde $\mathbf{a} \in \mathbb{R}^n$ je dáno a $\mathbf{x} \in \mathbb{R}^n$ je neznámá. Dělá to tak, že 'vyjádří' $\mathbf{x}$ z první rovnice a dosadí ho do druhé rovnice. To je těžký zločin, protože rovnice $\mathbf{a}^T\mathbf{x} = 0$ má nekonečně mnoho řešení (alespoň pro $n > 1$) a tedy z ní neplyne, že $\mathbf{x}$ je rovno jakémukoliv jednomu vektoru.

2.9.4 Další nápady pro práci s maticemi

- Pod výrazy s maticemi a vektory si malujte obdélníčky s rozměry matic, abyste měli jasnou představu o jejich rozměrech.
- Vidíte-li maticovou rovnici či soustavu rovnic, spočítejte si skalární rovnice a neznámé.
- Pracujte nejen s papírem, ale i s Matlabem. Úpravy maticových výrazů lze často ověřit na náhodných maticích. Např. chceme-li ověřit rovnost $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$, zkusíme např. `A=randn(5,3); B=randn(3,6); (A*B)'-B'*A'`. Samozřejmě to není důkaz.

2.10 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. Vyřešte soustavu rovnic $\{\,\mathbf{Ax} = \mathbf{b},\ \mathbf{x} = \mathbf{A}^T\mathbf{y}\,\}$, kde $\mathbf{x}, \mathbf{y}$ jsou neznámé vektory. Předpokládejte, že matice $\mathbf{AA}^T$ je regulární. Najděte jen vztah pro $\mathbf{x}$, vztah pro $\mathbf{y}$ nás nezajímá. Ověřte v Matlabu pro náhodné zadání, které získáte příkazy `A=randn(m,n); b=randn(m,1)`.

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

$$\mathbf{Ax} + \mathbf{By} = \mathbf{a}$$
$$\mathbf{Cx} + \mathbf{Dy} = \mathbf{b}$$

a) Vyjádřete soustavu ve tvaru $\mathbf{Pu} = \mathbf{q}$.
b) Je-li $\mathbf{a}, \mathbf{x} \in \mathbb{R}^m$, $\mathbf{b}, \mathbf{y} \in \mathbb{R}^n$, ukažte, že $\mathbf{x} = (\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C})^{-1}(\mathbf{a} - \mathbf{BD}^{-1}\mathbf{b})$. Jakou to má výpočetní výhodu oproti počítání $\mathbf{u}$ přímo ze soustavy $\mathbf{Pu} = \mathbf{q}$?

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 $\operatorname{vec}\colon \mathbb{R}^{m\times n} \to \mathbb{R}^{mn}$ (‘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. *Kroneckerův součin matic* (v Matlabu `kron(A,B)`) je definován jako

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

Pro libovolné matice (s kompatibilními velikostmi) platí

$$\operatorname{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A}) \operatorname{vec}\mathbf{B}. \tag{2.11}$$

Použijte tohoto vzorce pro transformaci následujících soustav rovnic s neznámou maticí $\mathbf{X}$ do tvaru $\mathbf{Pu} = \mathbf{q}$ s neznámým vektorem $\mathbf{u}$. Předpokládejte, že počet rovnic je rovný počtu neznámých. Předpokládejte, že matice a vektory mají nejobecnější možné rozměry, aby to dávalo smysl.

a) $\{\, \mathbf{b}_i^T\mathbf{X}\mathbf{a}_i = 0,\ i = 1, \ldots, k \,\}$
b) $\mathbf{AX} + \mathbf{XA}^T = \mathbf{C}$

2.7. Následující soustavy rovnic převeďte do tvaru $\mathbf{Pu} = \mathbf{q}$ s neznámým vektorem $\mathbf{u}$ (najděte co nejjednodušší vzorce, které vypočítají matice $\mathbf{P}, \mathbf{q}, \mathbf{u}$ z matic zadané soustavy):

a) Soustava $\mathbf{Ax} = \alpha\mathbf{b}$, kde neznámé jsou $(\mathbf{x}, \alpha)$.
b) Soustava z Příkladu 2.2.
c) Soustava z Příkladu 2.3.
d) Soustava z Příkladu 2.4.

2.8. *Komutátorem* dvou matic rozumíme matici $[\mathbf{A}, \mathbf{B}] = \mathbf{AB} - \mathbf{BA}$. 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) $[\mathbf{A}, \mathbf{BC}] = [\mathbf{A}, \mathbf{B}]\mathbf{C} + \mathbf{B}[\mathbf{A}, \mathbf{C}]$

2.9. Dokažte *vzorec Shermana a Morrisonové* (kde $\mathbf{A}$ je regulární a $\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u} + 1 \neq 0$)

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

a *vzorec Shermana, Morrisonové a Woodburyho* (kde $\mathbf{I} + \mathbf{V}^T\mathbf{A}^{-1}\mathbf{U}$ je regulární)

$$(\mathbf{A} + \mathbf{UV}^T)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{I} + \mathbf{V}^T\mathbf{A}^{-1}\mathbf{U})^{-1}\mathbf{V}^T\mathbf{A}^{-1}.$$

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

2.11. 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.

2.12. 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*.

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. Dokažte, že pokud je $\mathbf{I} - \mathbf{A}$ regulární, pak $\mathbf{A}(\mathbf{I} - \mathbf{A})^{-1} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{A}$.

2.16. Dokažte, že pokud $\mathbf{A}$, $\mathbf{B}$ a $\mathbf{A} + \mathbf{B}$ jsou regulární, pak

$$\mathbf{A}(\mathbf{A} + \mathbf{B})^{-1}\mathbf{B} = \mathbf{B}(\mathbf{A} + \mathbf{B})^{-1}\mathbf{A} = (\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}.$$

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

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 $\mathbf{A}^T\mathbf{D} - \mathbf{C}^T\mathbf{B} = \mathbf{I}$.

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. $\mathbf{x} = \mathbf{A}^T(\mathbf{AA}^T)^{-1}\mathbf{b}$

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

2.5.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.

2.7.d) Soustava jde napsat v maticovém tvaru jako $\{\, \mathbf{X1} = \mathbf{a},\ \mathbf{X}^T\mathbf{1} = \mathbf{b} \,\}$ 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 kýžený tvar, neboť proměnné jsou soustředěné do matice $\mathbf{X}$ a nikoliv do vektoru. Kýžený tvar získáme volbou

$$\mathbf{u} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{bmatrix} \in \mathbb{R}^{mn}, \quad \mathbf{P} = \begin{bmatrix} \mathbf{I}_m & \cdots & \mathbf{I}_m \\ \mathbf{1}_m^T & & \\ & \ddots & \\ & & \mathbf{1}_m^T \end{bmatrix} \in \mathbb{R}^{(m+n)\times(mn)}, \quad \mathbf{q} = \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \in \mathbb{R}^{m+n},$$

kde $\mathbf{x}_1, \ldots, \mathbf{x}_n$ jsou sloupce matice $\mathbf{X} = [x_{ij}]$, tj. $\mathbf{u} = (x_{11}, \ldots, x_{m1}, x_{12}, \ldots, x_{m2}, \ldots, \ldots, x_{mn})$.

2.17. Nechť $\mathbf{A}$ je regulární a symetrická (tedy čtvercová). 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 §2.4), tedy $\mathbf{B}^T = \mathbf{B}$.

Kapitola 3

Linearita

3.1 Lineární podprostory

Množina $\mathbb{R}^n$ spolu s operacemi sčítaní vektorů a násobení vektorů skalárem tvoří *lineární prostor* nad tělesem $\mathbb{R}$. Zopakujte si z lineární algebry pojem lineárního prostoru!

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

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

pro nějaké skaláry $\alpha_1, \ldots, \alpha_k \in \mathbb{R}$. Vektory jsou **lineárně nezávislé**, když platí implikace

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

V opačném případě jsou **lineárně závislé**.

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

Důkaz. Nechť kromě rovnice (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 $\mathbf{x}_1, \ldots, \mathbf{x}_k$ je množina

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

všech jejich lineárních kombinací (zde předpokládáme, že vektorů je konečný počet).

Neprázdná množina $X \subseteq \mathbb{R}^n$ se nazývá **lineární podprostor** (nebo jen **podprostor**) lineárního prostoru $\mathbb{R}^n$, jestliže každá lineární kombinace každé (konečné) množiny vektorů z X leží v X (neboli že množina X je uzavřená vůči lineárním kombinacím):

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

Snadno se ukáže, že lineární obal libovolné množiny vektorů je lineární podprostor.

Báze lineárního podprostoru[1] $X \subseteq \mathbb{R}^n$ je lineárně nezávislá množina vektorů, jejíž lineární obal je X. Platí následující netriviální tvrzení (důkazy najdete v učebnicích lineární algebry):

[1] Výrok ‘množina $X \subseteq \mathbb{R}^n$ je lineární podprostor’ je totéž jako ‘množina X je lineární podprostor prostoru $\mathbb{R}^n$’.

Věta 3.2.

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

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

Věta 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. $\square$

Příklad 3.1. Triviálně, prostor $\mathbb{R}^3$ je svým vlastním podprostorem. Jeho dimenze je 3. Jeho báze je např. $\{(1,0,0),(0,1,0),(0,0,1)\}$ (standardní) nebo $\{(1,1,1),(1,-1,0),(2,0,0)\}$. $\square$

Příklad 3.2. Množina $X = \operatorname{span}\{(1,2,3)\} = \{\, \alpha(1,2,3) \mid \alpha \in \mathbb{R}\,\} \subseteq \mathbb{R}^3$ je podprostor $\mathbb{R}^3$ dimenze 1. Je to přímka procházející počátkem. Její báze je např. množina $\{(1,2,3)\}$, jiná báze je množina $\{(2,4,6)\}$. $\square$

Příklad 3.3. 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 $\{(1,2,3,0),(0,1,2,-1)\}$. $\square$

Příklad 3.4. Všechny možné podprostory prostoru $\mathbb{R}^3$ jsou tyto: 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). $\square$

Příklad 3.5. Množina $X = \{\,(1+\alpha,\alpha) \mid \alpha \in \mathbb{R}\,\} \subseteq \mathbb{R}^2$ (přímka neprocházející počátkem) není podprostor $\mathbb{R}^2$, protože např. $(1,0) \in X$ ale $2(1,0) = (2,0) \notin X$. $\square$

3.2 Lineární zobrazení

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

$$\mathbf{f}(\alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k) = \alpha_1\mathbf{f}(\mathbf{x}_1) + \cdots + \alpha_k\mathbf{f}(\mathbf{x}_k). \tag{3.4}$$

Věta 3.4. *Zobrazení $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$ je lineární, právě když existuje matice $\mathbf{A} \in \mathbb{R}^{m\times n}$ taková, že*

$$\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x} \tag{3.5}$$

pro všechna $\mathbf{x} \in \mathbb{R}^n$. Matice $\mathbf{A}$ je navíc zobrazením $\mathbf{f}$ určena jednoznačně.

Říkáme proto, že matice $\mathbf{A}$ *reprezentuje* lineární zobrazení $\mathbf{f}$.

Důkaz. Důkaz jedné implikace je snadný: zobrazení (3.5) je lineární, neboť

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

Dokažme opačnou implikaci. Nechť $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^m$ je lineární zobrazení. Nechť $\mathbf{e}_1, \ldots, \mathbf{e}_n \in \mathbb{R}^n$ je standardní báze prostoru $\mathbb{R}^n$. Pro každé $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ máme $\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n$. Z (3.4) plyne

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

Nyní označíme $\begin{bmatrix}\mathbf{f}(\mathbf{e}_1) & \cdots & \mathbf{f}(\mathbf{e}_n)\end{bmatrix} = \mathbf{A}$. Tedy $\mathbf{A}$ je matice se sloupci $\mathbf{f}(\mathbf{e}_1), \ldots, \mathbf{f}(\mathbf{e}_n)$.

Dokažme jednoznačnost matice $\mathbf{A}$. Platí-li $\mathbf{A}\mathbf{x} = \mathbf{B}\mathbf{x}$ pro každé $\mathbf{x} \in \mathbb{R}^n$, pak samozřejmě platí $\mathbf{A} = \mathbf{B}$, protože stačí dosadit za $\mathbf{x}$ postupně vektory $\mathbf{e}_1, \ldots, \mathbf{e}_n$. □

Příklad 3.6. 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 (3.4). Ovšem je to patrné na první pohled, protože jej lze vyjádřit ve tvaru (3.5):

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

□

Pokud $m = 1$, lineární zobrazení je funkce $f\colon \mathbb{R}^n \to \mathbb{R}$ tvaru

$$f(\mathbf{x}) = \mathbf{a}^T\mathbf{x} = a_1x_1 + \cdots + a_nx_n, \tag{3.6}$$

kde $\mathbf{a} \in \mathbb{R}^n$. Této funkci se též říká *lineární forma*.

Podívejme se blíže na vzorec (3.5). Výraz $\mathbf{A}\mathbf{x}$ je maticový součin matice $m \times n$ maticí $n \times 1$ (viz §2.6). Označíme-li $\mathbf{y} = \mathbf{A}\mathbf{x}$, je tedy podle (2.1)

$$y_i = \sum_{j=1}^{n} a_{ij}x_j, \quad i = 1, \ldots, m. \tag{3.7}$$

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

$$\mathbf{A}\mathbf{x} = \begin{bmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_n\end{bmatrix}\begin{bmatrix}x_1\\ \vdots\\ x_n\end{bmatrix} = x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n, \tag{3.8}$$

tedy vektor $\mathbf{A}\mathbf{x}$ je *lineární kombinace sloupců* matice $\mathbf{A}$. Naopak, vyjádříme-li matici $\mathbf{A}$ pomocí řádků, máme

$$\mathbf{A}\mathbf{x} = \begin{bmatrix}\mathbf{a}_1^T\\ \vdots\\ \mathbf{a}_m^T\end{bmatrix}\mathbf{x} = \begin{bmatrix}\mathbf{a}_1^T\mathbf{x}\\ \vdots\\ \mathbf{a}_m^T\mathbf{x}\end{bmatrix}, \tag{3.9}$$

tedy složky vektoru $\mathbf{A}\mathbf{x}$ jsou *skalární součiny řádků* matice $\mathbf{A}$ a vektoru $\mathbf{x}$. Všimněte si, že (3.8) a (3.9) jsou speciální případy (2.8) a (2.9).

Složení lineárních zobrazení je opět lineární zobrazení. Pro $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x}$ a $\mathbf{g}(\mathbf{y}) = \mathbf{B}\mathbf{y}$ máme

$$(\mathbf{g} \circ \mathbf{f})(\mathbf{x}) = \mathbf{g}(\mathbf{f}(\mathbf{x})) = \mathbf{B}(\mathbf{A}\mathbf{x}) = (\mathbf{B}\mathbf{A})\mathbf{x} = \mathbf{B}\mathbf{A}\mathbf{x},$$

tedy $\mathbf{B}\mathbf{A}$ je matice složeného zobrazení $\mathbf{g} \circ \mathbf{f}$. Tedy matice složeného zobrazení je součinem matic jednotlivých zobrazení. To je hlavní důvod, proč je rozumné definovat maticové násobení jako (2.1): *násobení matic odpovídá skládání lineárních zobrazení* reprezentovaných těmito maticemi.

3.2.1 Prostor obrazů

S lineárním zobrazením jsou spjaty dva lineární podprostory, prostor obrazů a nulový prostor (jádro). 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 matice $\mathbf{A} \in \mathbb{R}^{m\times n}$ je množina

$$\operatorname{rng} \mathbf{A} = \{\, \mathbf{A}\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n \,\} \subseteq \mathbb{R}^m. \tag{3.10}$$

Interpretace prostoru obrazů:

- Je to množina $\mathbf{f}(\mathbb{R}^n)$ všech hodnot, jichž může zobrazení $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x}$ nabýt.
- Je to množina všech vektorů $\mathbf{y}$, pro které má lineární soustava $\mathbf{A}\mathbf{x} = \mathbf{y}$ řešení.
- Dle (3.8) je to lineární obal sloupců matice $\mathbf{A}$. Tedy je to lineární podprostor $\mathbb{R}^m$.

Dimenze lineárního obalu sloupců se nazývá hodnost matice (už jsme ji potkali v §2.3),

$$\operatorname{rank} \mathbf{A} = \dim \operatorname{rng} \mathbf{A}. \tag{3.11}$$

Věta 3.5. *Pro matici $\mathbf{A} \in \mathbb{R}^{m\times n}$ jsou následující tvrzení ekvivalentní:*

1. $\operatorname{rng} \mathbf{A} = \mathbb{R}^m$
2. *Soustava $\mathbf{A}\mathbf{x} = \mathbf{y}$ má řešení pro každé $\mathbf{y}$.*
3. $\operatorname{rank} \mathbf{A} = m$
4. *Zobrazení $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x}$ 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. existuje $\mathbf{B}$ tak, že $\mathbf{A}\mathbf{B} = \mathbf{I}$.*
7. *Matice $\mathbf{A}\mathbf{A}^T \in \mathbb{R}^{m\times m}$ je regulární.*

Důkaz.

- $1 \Leftrightarrow 2 \Leftrightarrow 3 \Leftrightarrow 4$ plyne přímo z definic.
- $3 \Leftrightarrow 5$ plyne z definice hodnosti a z (2.2).
- $2 \Rightarrow 6$ platí, neboť soustava $\mathbf{A}\mathbf{b}_i = \mathbf{e}_i$ 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 $\mathbf{x} = \mathbf{B}\mathbf{y}$.
- Tvrzení $1 \Leftrightarrow 7$ plyne z rovnosti (5.4a), uvedené později. □

Věta 3.6. *Pro libovolné matice $\mathbf{A}, \mathbf{B}$ platí*

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

Důkaz. První tvrzení říká, že jestliže soustava $\mathbf{z} = \mathbf{A}\mathbf{B}\mathbf{x}$ má řešení, pak i soustava $\mathbf{z} = \mathbf{A}\mathbf{y}$ má řešení. To je ale jasné, protože můžeme položit $\mathbf{y} = \mathbf{B}\mathbf{x}$. Druhé tvrzení navíc říká, že jestliže $\mathbf{B}$ má lineárně nezávislé řádky a $\mathbf{z} = \mathbf{A}\mathbf{y}$ má řešení, pak i $\mathbf{z} = \mathbf{A}\mathbf{B}\mathbf{x}$ má řešení. To platí, neboť dle implikace $5 \Rightarrow 2$ ve Větě 3.5 má soustava $\mathbf{B}\mathbf{x} = \mathbf{y}$ řešení pro každé $\mathbf{y}$. □

3.2.2 Nulový prostor

Nulový prostor matice $\mathbf{A}$ (také se nazývá *jádro* zobrazení $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x}$) je množina

$$\operatorname{null} \mathbf{A} = \{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} = \mathbf{0} \,\} \subseteq \mathbb{R}^n. \tag{3.12}$$

Interpretace nulového prostoru:

- Je to množina všech vektorů, které se zobrazí na nulový vektor.
- Podle (3.9) je to množina všech vektorů, které jsou kolmé na každý řádek matice $\mathbf{A}$. Z toho je vidět, že je to lineární podprostor $\mathbb{R}^n$.

Platí navíc, že každý lineární podprostor prostoru $\mathbb{R}^n$ je nulovým prostorem nějaké matice $\mathbf{A} \in \mathbb{R}^{m\times n}$ (toto tvrzení není samozřejmé, lze ho dokázat např. z pozdější Věty 4.1). Tedy lineární podprostor X lze reprezentovat (např. v počítači) dvěma způsoby:

- jako lineární obal konečného počtu vektorů (tedy maticí $\mathbf{A}$ takovou, že $X = \operatorname{rng} \mathbf{A}$),
- jako množinu řešení homogenní lineární soustavy (tedy maticí $\mathbf{A}$ takovou, že $X = \operatorname{null} \mathbf{A}$).

Věta 3.7. *Pro matici $\mathbf{A} \in \mathbb{R}^{m\times n}$ jsou následující tvrzení ekvivalentní:*

1. $\operatorname{null} \mathbf{A} = \{\mathbf{0}\}$ *(tj. nulový prostor je triviální).*
2. *Soustava $\mathbf{A}\mathbf{x} = \mathbf{0}$ má jediné řešení $\mathbf{x} = \mathbf{0}$.*
3. $\operatorname{rank} \mathbf{A} = n$.
4. *Zobrazení $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x}$ je injektivní (viz §1.1.2).*
5. *Sloupce matice $\mathbf{A}$ jsou lineárně nezávislé.*
6. *Matice $\mathbf{A}$ má levou inverzi, tj. existuje $\mathbf{B}$ tak, že $\mathbf{B}\mathbf{A} = \mathbf{I}$.*
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 (3.12).
- Ekvivalence $2 \Leftrightarrow 5$ plyne z definice lineární nezávislosti (3.2), tj. $\mathbf{A}\mathbf{x} = \mathbf{0} \;\Rightarrow\; \mathbf{x} = \mathbf{0}$.
- Ekvivalence $3 \Leftrightarrow 5$ plyne z definice hodnosti (3.11).
- Tvrzení 4 říká, že pro každé $\mathbf{x}, \mathbf{y}$ je $\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{y} \;\Rightarrow\; \mathbf{x} = \mathbf{y}$, 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 ekvivalence $3 \Leftrightarrow 6$ ve Větě 3.5.
- Tvrzení $1 \Leftrightarrow 7$ je plyne z rovnosti (5.4b), uvedené později. □

Věta 3.8. *Pro libovolné matice $\mathbf{A}, \mathbf{B}$ platí:*

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

Důkaz. První tvrzení říká, že $\mathbf{B}\mathbf{x} = \mathbf{0} \implies \mathbf{A}\mathbf{B}\mathbf{x} = \mathbf{0}$, což platí vynásobením maticí $\mathbf{A}$ zleva. Druhé tvrzení říká, že když $\mathbf{A}$ má lineárně nezávislé sloupce, pak $\mathbf{A}\mathbf{B}\mathbf{x} = \mathbf{0} \implies \mathbf{B}\mathbf{x} = \mathbf{0}$. To platí, neboť dle implikace $5 \Rightarrow 2$ ve Větě 3.7 máme $\mathbf{A}\mathbf{y} = \mathbf{0} \implies \mathbf{y} = \mathbf{0}$. □

3.2.3 Dvě věty o dimenzích

Zde dokážeme dvě obtížnější věty o dimenzích prostoru obrazů a nulového prostoru.

Jako první uvedeme slíbený důkaz rovnosti (2.2). K tomu nejdřív uvedeme pomocnou větu, která je zajímavá sama o sobě.

Věta 3.9. *Pro každou matici* $\mathbf{A} \in \mathbb{R}^{m\times n}$ *hodnosti* r *existují* $\mathbf{B} \in \mathbb{R}^{m\times r}$ *a* $\mathbf{C} \in \mathbb{R}^{r\times n}$ *tak, že*

$$\mathbf{A} = \mathbf{BC}. \tag{3.13}$$

Důkaz. Zvolme libovolnou bázi prostoru $\operatorname{rng} \mathbf{A}$. Nechť vektory této báze tvoří sloupce matice $\mathbf{B} \in \mathbb{R}^{m\times r}$, kde $r = \dim \operatorname{rng} \mathbf{A}$. Nyní j-tý sloupec $\mathbf{a}_j$ matice $\mathbf{A}$ je lineární kombinací sloupců matice $\mathbf{B}$, neboli existuje vektor $\mathbf{c}_j$ tak že $\mathbf{a}_j = \mathbf{B}\mathbf{c}_j$. Platí tedy (3.13), kde matice $\mathbf{C} \in \mathbb{R}^{r\times n}$ má sloupce $\mathbf{c}_1, \dots, \mathbf{c}_n$. □

Rozkladu (3.13) se někdy říká **rozklad matice podle hodnosti** (angl. *rank factorization*). Lze na něj pohlížet jako na '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 mn operací. Spočítáme-li ale nejdřív vektor $\mathbf{z} = \mathbf{Cx}$ a pak vektor $\mathbf{y} = \mathbf{Bz}$, potřebujeme jen $(m + n)r$ operací.
- Představme si zobrazení $\mathbf{f}(\mathbf{x}) = \mathbf{Ax}$ jako přenosový kanál, jehož vstupem jsou proměnné $x_1, \dots, x_n$ a výstupem $y_1, \dots, y_m$. Přenos můžeme realizovat jen po r 'drátech' $z_1, \dots, z_r$.

A nyní můžeme dokázat rovnost (2.2), tj. že dimenze lineárního obalu sloupců je rovna dimenzi lineárního obalu řádků:

Věta 3.10. *Pro každou matici* $\mathbf{A}$ *platí*

$$\operatorname{rank} \mathbf{A} = \dim \operatorname{rng} \mathbf{A} = \dim \operatorname{rng}(\mathbf{A}^T) = \operatorname{rank}(\mathbf{A}^T). \tag{3.14}$$

Důkaz. Pišme (3.13) jako $\mathbf{A}^T = \mathbf{C}^T\mathbf{B}^T$. Dle Věty 3.6 je $\operatorname{rng}(\mathbf{A}^T) \subseteq \operatorname{rng}(\mathbf{C}^T)$. Dle Věty 3.3 tedy $\dim \operatorname{rng}(\mathbf{A}^T) \le \dim \operatorname{rng}(\mathbf{C}^T)$. Protože $\mathbf{C}^T$ má r sloupců, je $\dim \operatorname{rng}(\mathbf{C}^T) \le r = \dim \operatorname{rng} \mathbf{A}$. Tedy

$$\dim \operatorname{rng}(\mathbf{A}^T) \le \dim \operatorname{rng} \mathbf{A}. \tag{3.15}$$

Ukázali jsme, že pro každou matici $\mathbf{A}$ platí nerovnost (3.15). Nyní tuto nerovnost použijeme na matici $\mathbf{A}^T$, což dá $\dim \operatorname{rng} \mathbf{A} \le \dim \operatorname{rng}(\mathbf{A}^T)$. Obě nerovnosti dohromady dají (3.14). □

Jako druhou uvedeme větu o vztahu dimenze prostoru obrazů, nulového prostoru, a 'vstupního' prostoru $\mathbb{R}^n$ lineárního zobrazení. Někdy se jí říká *rank-plus-nullity theorem*, kde *nullity* označuje dimenzi nulového prostoru matice:

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

$$\dim \operatorname{rng} \mathbf{A} + \dim \operatorname{null} \mathbf{A} = n. \tag{3.16}$$

Důkaz. ($\star$) Nechť bázi prostoru $\operatorname{rng} \mathbf{A}$ tvoří 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 bázi prostoru $\operatorname{null} \mathbf{A}$ tvoří sloupce matice $\mathbf{D} \in \mathbb{R}^{n\times q}$.

Protože sloupce $\mathbf{B}$ tvoří bázi $\operatorname{rng} \mathbf{A}$, 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 každý $\mathbf{x} \in \mathbb{R}^n$ existuje právě jeden $\mathbf{y} \in \mathbb{R}^r$ a právě jeden $\mathbf{z} \in \mathbb{R}^q$ tak, že $\mathbf{x} = \mathbf{Cy} + \mathbf{Dz}$. To ale znamená, že sloupce matice $\begin{bmatrix} \mathbf{C} & \mathbf{D} \end{bmatrix} \in \mathbb{R}^{n\times(r+q)}$ tvoří bázi prostoru $\mathbb{R}^n$. Tedy musí být $r + q = n$, což je rovnost (3.16). □

Interpretace věty 3.11:

- Každá dimenze na vstupu zobrazení $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x}$ se buď 'smáčkne' do nulového vektoru nebo se objeví na výstupu.
- Počet lineárně nezávislých řešení lineární homogenní soustavy $\mathbf{A}\mathbf{x} = \mathbf{0}$ je $n - \operatorname{rank} \mathbf{A}$.
- Protože $\operatorname{rank} \mathbf{A} = \operatorname{rank}(\mathbf{A}^T)$, rovnost (3.16) lze psát také jako

$$\dim \operatorname{rng}(\mathbf{A}^T) + \dim \operatorname{null} \mathbf{A} = n. \tag{3.17}$$

  Protože $\operatorname{rng}(\mathbf{A}^T)$ je lineární obal řádků matice $\mathbf{A}$ a $\operatorname{null} \mathbf{A}$ je množina všech vektorů kolmých na řádky $\mathbf{A}$, rovnost (3.17) říká, že součet dimenzí těchto dvou podprostorů prostoru $\mathbb{R}^n$ je n. O tom si více řekneme v části o ortogonálním doplňku (§4.2).

3.3 Afinní podprostor a zobrazení

Afinní kombinace vektorů $\mathbf{x}_1, \ldots, \mathbf{x}_k \in \mathbb{R}^n$ 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 $\mathbf{x}_1, \ldots, \mathbf{x}_k$ je množina všech jejich afinních kombinací. Množina $A \subseteq \mathbb{R}^n$ je **afinní podprostor**[2] lineárního prostoru $\mathbb{R}^n$, jestliže každá afinní kombinace každé (konečné) množiny vektorů z A leží v A (neboli množina A je uzavřená vůči afinním kombinacím):

$$\mathbf{x}_1, \ldots, \mathbf{x}_k \in A, \;\; \alpha_1 + \cdots + \alpha_k = 1 \quad \Longrightarrow \quad \alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k \in A\,. \tag{3.18}$$

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

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

Na rozdíl od toho, obecná lineární kombinace na počátku závisí.

K tomu (mírně neformální) poznámka: Jestliže s prvky vektorového (pod)prostoru provádíme operace afinní kombinace, tyto prvky si stačí představovat/kreslit jako body. 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 vektorového (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 vektorového (= lineárního) (pod)prostoru *vektory*.

Příklad 3.7. 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).

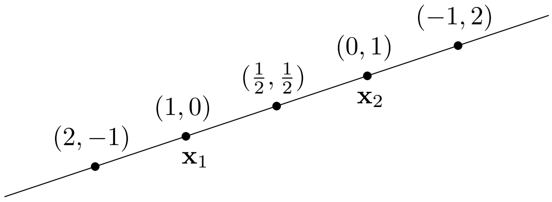
[2] Všimněte si, že definujeme afinní *podprostor* lineárního prostoru, ale už ne afinní *prostor* sám o sobě. Definice afinního prostoru bez odkazu k nějakému lineárnímu prostoru (tj. pomocí axiomů) existuje, ale neuvádíme ji.

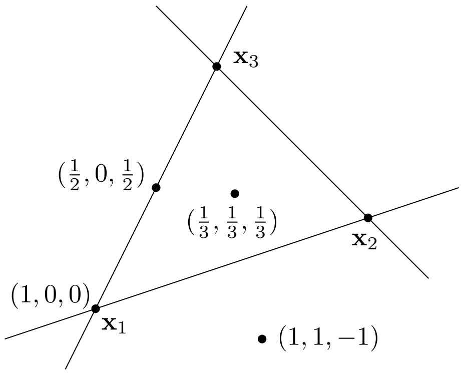
Naproti tomu, lineární obal dvou vektorů $\mathbf{x}_1, \mathbf{x}_2$ (pokud jsou lineárně nezávislé) je rovina procházející těmito dvěma body a počátkem $\mathbf{0}$.

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

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

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.) $\square$

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

$$X + \mathbf{x} = \{\, \mathbf{x} + \mathbf{y} \mid \mathbf{y} \in X \,\}. \tag{3.19}$$

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 $\mathbb{R}^n$.*
- *Je-li A afinní podprostor a $\mathbf{x}, \mathbf{y} \in A$, pak $A - \mathbf{x} = A - \mathbf{y}$.*

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[3]

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

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

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

[3] Zde jsme použili 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 $\mathbf{x} \in X \iff \mathbf{x} + \mathbf{y} \in X + \mathbf{y}$. To snadno plyne z (3.19).

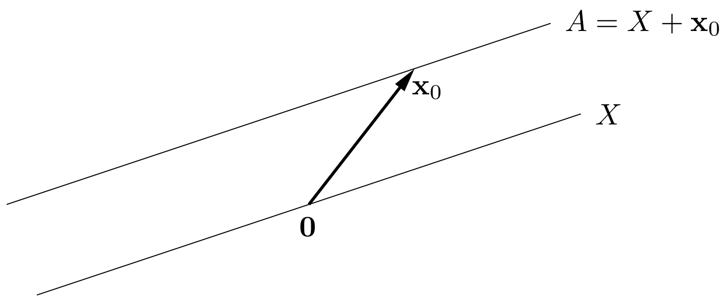
Pravá strana této implikace jde psát jako

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

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.

Třetí tvrzení: Označme $X = A - \mathbf{x}$. Jelikož $\mathbf{y} \in A$, je také $\mathbf{y} - \mathbf{x} \in A - \mathbf{x} = X$. Nyní pišme $A - \mathbf{y} = A - \mathbf{x} + \mathbf{x} - \mathbf{y} = X - (\mathbf{y} - \mathbf{x}) = X$, kde poslední rovnost plyne z linearity X (kterou jsme dokázali v druhém tvrzení). □

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 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 lineární soustavy*[4], *tj. existují $\mathbf{A}$ a $\mathbf{b}$ splňující $A = \{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} = \mathbf{b} \,\}$.*

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 lineární soustavy).

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

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

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

$$A = \{\, \mathbf{x} \mid \mathbf{A}\mathbf{x} = \mathbf{b} \,\} = \{\, \mathbf{x} \mid \mathbf{A}(\mathbf{x} - \mathbf{x}_0) = \mathbf{0} \,\} = \{\, \mathbf{x}' + \mathbf{x}_0 \mid \mathbf{A}\mathbf{x}' = \mathbf{0} \,\} = X + \mathbf{x}_0. \tag{3.20}$$

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 §3.2.2) existuje matice $\mathbf{A}$ tak, že $X = \operatorname{null}\mathbf{A}$. Nechť $\mathbf{b} = \mathbf{A}\mathbf{x}_0$. Pak dle (3.20) je $X + \mathbf{x}_0 = \{\, \mathbf{x} \mid \mathbf{A}\mathbf{x} = \mathbf{b} \,\}$. □

[4]Proto 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 rng $\mathbf{A}$ pro nějakou matici $\mathbf{A}$, nebo jako null $\mathbf{A}$ (tj. množinou 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 kombinace ostatních. Afinní nezávislost lze ovšem definovat i jinými způsoby (důkaz věty neuvádíme):

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

1. *Žádný bod není roven afinní kombinaci ostatních bodů.*
2. *Platí (srov. se (3.2))*

$$\alpha_1 + \cdots + \alpha_k = 0, \quad \alpha_1\mathbf{x}_1 + \cdots + \alpha_k\mathbf{x}_k = \mathbf{0} \quad \implies \quad \alpha_1 = \cdots = \alpha_k = 0. \tag{3.21}$$

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[5]).*

Příklad 3.8. 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 3.7. 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 (3.4) 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

$$\mathbf{f}(\mathbf{x}) = \mathbf{Ax} + \mathbf{b}. \tag{3.22}$$

Pro $m = 1$ se zobrazení (3.22) nazývá také **afinní funkce**[6] $f\colon \mathbb{R}^n \to \mathbb{R}$ a má tvar

$$f(\mathbf{x}) = \mathbf{a}^T\mathbf{x} + b = a_1x_1 + \cdots + a_nx_n + b, \tag{3.23}$$

kde $\mathbf{a} \in \mathbb{R}^n$ a $b \in \mathbb{R}$.

Příklad 3.9. 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 (3.4) pro $\alpha_1 + \cdots + \alpha_k = 1$. Ale je to patrné i z toho, že zobrazení lze vyjádřit ve tvaru (3.22):

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

□

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

[6] 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 $\mathbf{Ax} = \mathbf{b}$ se říká ‘lineární’ i v lineární algebře.

Příklad 3.10. Pro dané body $\mathbf{x}_1, \dots, \mathbf{x}_k \in \mathbb{R}^n$ a $\mathbf{y}_1, \dots, \mathbf{y}_k \in \mathbb{R}^m$ máme za úkol najít afinní zobrazení $\mathbf{f}$ takové, aby platilo $\mathbf{y}_i = \mathbf{f}(\mathbf{x}_i)$ pro všechna $i = 1, \dots, k$. Řešíme tedy lineární soustavu $\mathbf{y}_i = \mathbf{A}\mathbf{x}_i + \mathbf{b}$, $i = 1, \dots, k$, 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 $\mathbf{x}_1, \dots, \mathbf{x}_k$ jsou afinně nezávislé, matice $\begin{bmatrix}\mathbf{x}_1 & \cdots & \mathbf{x}_k \\ 1 & \cdots & 1\end{bmatrix}$ je regulární a 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) $\left\{\, \mathbf{x} \in \mathbb{R}^n \;\middle|\; \sum_{i=1}^n x_i = 0 \,\right\}$

3.2. 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. 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 $\mathbf{A}$?

3.3. 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 (3.5). Je toto zobrazení afinní? Pokud ano, napište ho ve formě (3.22). Obě odpovědi dokažte z definic.

3.4. Mějme nehomogenní lineární soustavu

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

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 $\mathbf{x}_0 \in \mathbb{R}^3$.

3.5. 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) $\mathbf{f}(x_1, x_2) = (2x_1 + x_2, x_1 - x_2, 2x_2 + x_1)$

3.6. Máte matice $\mathbf{A}$ a $\mathbf{B}$ se stejným počtem řádků. Jak byste ověřili, zda $\operatorname{rng} \mathbf{A} = \operatorname{rng} \mathbf{B}$, umíte-li spočítat hodnost libovolné matice?

3.7. 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.

3.8. 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.

3.9. 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.

3.10. Zjisti, 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) $f(1,0,1) = -1,\ f(0,1,2) = 1,\ f(1,1,3) = 2.$

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

3.12. Nechť $\mathbf{A}$ má lineárně nezávislé sloupce. Ukažte, že každá báze podprostoru $\operatorname{rng}\mathbf{A}$ je tvořena sloupci matice $\mathbf{AB}$ pro nějakou regulární matici $\mathbf{B}$.

3.13. Dokažte, že pro libovolné dvě matice platí $\operatorname{rank}(\mathbf{AB}) \le \min\{\operatorname{rank}\mathbf{A}, \operatorname{rank}\mathbf{B}\}$.

3.14. Dokažte, že 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 $\mathbf{A} = \mathbf{BC}$. Poznamenejme, že tomuto rozkladu se říká **rozklad matice podle hodnosti** (*rank factorization*). Jestliže $r \ll \min\{m,n\}$, můžeme tím v některých situacích ušetřit paměť nebo čas. Porovnejte velikost paměti potřebné pro uložení matice $\mathbf{A}$ a matic $\mathbf{B}, \mathbf{C}$. Jak byste efektivně spočítali součin $\mathbf{Ax}$?

3.15. ($\star$) Mějme vektory $\mathbf{x}_1, \ldots, \mathbf{x}_k$. Vektor $\mathbf{x}_i$ nazveme *klíčový*, je-li lineárně nezávislý na předchozích vektorech $\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}$. Dokažte, že množina klíčových vektorů je báze podprostoru $\operatorname{span}\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$. Všimněte si, že vlastně dokazujete třetí tvrzení Věty 3.2.

3.16. ($\star$) 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 $j \in \{1, \ldots, n\} \setminus J$. Všimněte si, že vlastně dokazujete čtvrté tvrzení Věty 3.2.

3.17. ($\star$) 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 $n - 1$.

3.2. 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 $\mathbf{y} \neq \mathbf{0}$.

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

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

3.5.a) $\operatorname{rng} \mathbf{A} = \mathbb{R}^2$ a tedy báze $\operatorname{rng} \mathbf{A}$ je např. $\mathbf{I}_2$. Báze $\operatorname{null} \mathbf{A}$ je např. $(1, 1, 3)$.

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

3.6. $\operatorname{rng} \mathbf{A} = \operatorname{rng} \mathbf{B}$ je ekvivalentní $\operatorname{rank} \mathbf{A} = \operatorname{rank}[\mathbf{A}\ \mathbf{B}] = \operatorname{rank} \mathbf{B}$.

3.9. Viz Příklad 3.10.

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

3.10.b) Ano.

3.10.c) Ne.

3.13. Z Věty 3.6 je $\operatorname{rank}(\mathbf{A}\mathbf{B}) \leq \operatorname{rank} \mathbf{A}$. Dle (2.2) je $\operatorname{rank}(\mathbf{A}\mathbf{B}) = \operatorname{rank}((\mathbf{A}\mathbf{B})^T) = \operatorname{rank}(\mathbf{B}^T \mathbf{A}^T) \leq \operatorname{rank}(\mathbf{B}^T) = \operatorname{rank} \mathbf{B}$.

3.14. Nechť sloupce $\mathbf{B}$ tvoří bázi podprostoru $\operatorname{rng} \mathbf{A}$. Pak každý sloupec $\mathbf{a}_j$ matice $\mathbf{A}$ je lineární kombinací sloupců $\mathbf{B}$, tedy $\mathbf{a}_j = \mathbf{B}\mathbf{c}_j$. Vektory $\mathbf{c}_j$ tvoří sloupce matice $\mathbf{C}$.

Kapitola 4

Ortogonalita

4.1 Standardní skalární součin

Na prostoru $\mathbb{R}^n$ lze přirozeně definovat **standardní skalární součin**[1]

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

Skalární součin splňuje **Cauchy-Schwarzovu nerovnost** $(\mathbf{x}^T\mathbf{y})^2 \leq (\mathbf{x}^T\mathbf{x})(\mathbf{y}^T\mathbf{y})$.

Standardní skalární součin indukuje **eukleidovskou normu**

$$\|\mathbf{x}\| = \sqrt{\mathbf{x}^T\mathbf{x}} = (x_1^2 + \cdots + x_n^2)^{1/2}, \tag{4.1}$$

Norma splňuje **trojúhelníkovou nerovnost** $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$. Ta snadno plyne z Cauchy-Schwarzovy nerovnosti (umocněte a roznásobte). Norma měří *délku* vektoru $\mathbf{x}$.

Úhel φ dvojice vektorů se spočítá jako

$$\cos\varphi = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\,\|\mathbf{y}\|}. \tag{4.2}$$

Speciálně, vektory jsou **ortogonální** neboli **kolmé**, jestliže $\mathbf{x}^T\mathbf{y} = 0$, což značíme také $\mathbf{x} \perp \mathbf{y}$.

Eukleidovská norma indukuje **eukleidovskou metriku**

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|, \tag{4.3}$$

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*.

4.2 Ortogonální podprostory

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

$$\mathbf{y} \perp \operatorname{span}\{\mathbf{x}_1, \ldots, \mathbf{x}_k\} \quad\Longleftrightarrow\quad \mathbf{y} \perp \mathbf{x}_i,\ \ldots,\ \mathbf{y} \perp \mathbf{x}_k. \tag{4.4}$$

[1] Pro skalární součin prvků $\mathbf{x}, \mathbf{y}$ z abstraktního vektorového prostoru se užívají značení $\mathbf{x} \cdot \mathbf{y}$, $\langle \mathbf{x}, \mathbf{y} \rangle$ nebo $(\mathbf{x}, \mathbf{y})$. Protože my ale máme $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, vystačíme s maticovým zápisem $\mathbf{x}^T\mathbf{y}$.

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

$$X \perp Y \quad \Longrightarrow \quad X \cap Y = \{\mathbf{0}\}. \tag{4.5}$$

neboť jediný vektor kolmý sám na sebe je $\mathbf{0}$ (rozmyslete!).

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

$$X^\perp = \{\, \mathbf{y} \in \mathbb{R}^n \mid \mathbf{y} \perp X \,\} \tag{4.6}$$

všech vektorů z $\mathbb{R}^n$ kolmých na podprostor X. Množina (4.6) je podprostor $\mathbb{R}^n$ (dokažte!).

Příklad 4.1. Dvě na sebe kolmé přímky v $\mathbb{R}^3$ procházející počátkem jsou ortogonální podprostory, nejsou ale ortogonální doplněk jeden druhého. Ortogonální doplněk k přímce v $\mathbb{R}^3$ procházející počátkem je rovina procházející počátkem, která je na tuto přímku kolmá. □

Příklad 4.2. Pozor, stěna místnosti není kolmá na podlahu. Opravdu, existuje dvojice vektorů, jeden z podlahy a jeden ze stěny, které nejsou na sebe kolmé (kde jsou?). □

S ortogonálním doplňkem jsme se vlastně už setkali v §3.2.2: nulový prostor je tvořen všemi vektory kolmými na řádky matice. Tedy je-li $X = \operatorname{rng}(\mathbf{A}^T)$, pak $X^\perp = \operatorname{null} \mathbf{A}$. Přesněji by se toto tvrzení dokázalo z (4.4).

Příklad 4.3. Ortogonální doplněk podprostoru $X = \operatorname{span}\{(1,2,3),(0,1,-1)\}$ je množina všech vektorů (x_1, x_2, x_3) splňujících $x_1 + 2x_2 + 3x_3 = x_2 - x_3 = 0$, neboli $X^\perp = \operatorname{null} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & -1 \end{bmatrix}$. □

Věta 4.1. *Pro každé podprostory $X, Y \subseteq \mathbb{R}^n$ platí:*

1. $\dim X + \dim X^\perp = n$,
2. $X \perp Y$ *a* $\dim X + \dim Y = n$ *implikuje* $Y = X^\perp$,
3. $(X^\perp)^\perp = X$.

Důkaz.

1. Tvrzení 1 je rovnost (3.17), kde $X = \operatorname{rng}(\mathbf{A}^T)$ a $X^\perp = \operatorname{null} \mathbf{A}$.
2. Platí $X \perp Y \iff Y \subseteq X^\perp$ (neboť $Y \subseteq X^\perp$ znamená, že každý vektor z Y je kolmý na X). Dle tvrzení 1 tedy $\dim(X^\perp) = n - \dim X = \dim Y$. Dle Věty 3.3 proto $Y = X^\perp$.
3. Nechť $Y = X^\perp$. Z toho plyne $X \perp Y$ a $\dim X + \dim Y = n$. Z toho ale dle tvrzení 2 plyne $X = Y^\perp = (X^\perp)^\perp$. □

Dle tvrzení 3 je $Y = X^\perp \iff X = Y^\perp$. Proto říkáme, že podprostory X a Y jsou *ortogonálním doplňkem jeden druhého*.

Zformulujme podrobněji souvislost ortogonálního doplňku a nulového prostoru:

Věta 4.2. *Pro každou matici* $\mathbf{A}$ *platí*

$$(\operatorname{rng} \mathbf{A})^\perp = \operatorname{null}(\mathbf{A}^T). \tag{4.7a}$$

$$(\operatorname{null} \mathbf{A})^\perp = \operatorname{rng}(\mathbf{A}^T), \tag{4.7b}$$

Důkaz. Stačí si uvědomit, že pro každou $\mathbf{A} \in \mathbb{R}^{m \times n}$ platí:

- $\operatorname{rng} \mathbf{A} = \{ \mathbf{A}\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n \} \subseteq \mathbb{R}^m$ je lineární obal sloupců $\mathbf{A}$,
- $\operatorname{null} \mathbf{A} = \{ \mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} = \mathbf{0} \} \subseteq \mathbb{R}^n$ 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 $\mathbf{A}$.

Rovnost (4.7a) plyne přímo z toho (pro přesný důkaz bychom užili (4.4)). Rovnost (4.7b) se získá použitím rovnosti (4.7a) na matici $\mathbf{A}^T$ a rovnosti $(X^\perp)^\perp = X$. □

4.3 Ortonormální množina vektorů

Vektor $\mathbf{u} \in \mathbb{R}^m$ nazveme **normalizovaný**, pokud má jednotkovou délku ($\|\mathbf{u}\| = 1 = \mathbf{u}^T\mathbf{u}$). Množinu vektorů $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ nazveme[2] **ortonormální** (trochu nepřesně[3] říkáme, že *vektory* $\mathbf{u}_1, \dots, \mathbf{u}_n$ *jsou ortonormální*), jestliže každý vektor z této množiny je normalizovaný a každá dvojice vektorů z této množiny je ortogonální, tedy

$$\mathbf{u}_i^T\mathbf{u}_j = \delta_{ij} = \begin{cases} 0 & \text{když } i \neq j, \\ 1 & \text{když } i = j. \end{cases} \tag{4.8}$$

Věta 4.3. *Ortonormální množina vektorů je lineárně nezávislá.*

Důkaz. Vynásobme levou stranu implikace (3.2) skalárně vektorem $\mathbf{u}_i$, což dá

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

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

Věta 4.4. *Nechť jsou vektory* $\mathbf{u}_1, \dots, \mathbf{u}_n$ *ortonormální. Nechť*

$$\mathbf{x} = \alpha_1\mathbf{u}_1 + \cdots + \alpha_n\mathbf{u}_n. \tag{4.9}$$

Pak $\alpha_i = \mathbf{u}_i^T\mathbf{x}$.

Důkaz. Vynásobením rovnice (4.9) zleva skalárně vektorem $\mathbf{u}_i$ máme $\mathbf{u}_i^T\mathbf{x} = \mathbf{u}_i^T\mathbf{u}_i\alpha_i = \alpha_i$. □

Připomeňme (§3.1), že skaláry $\alpha_1, \dots, \alpha_n$ jsou souřadnice vektoru $\mathbf{x}$ v ortonormální bázi $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ podprostoru $\operatorname{span}\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$. Ze střední školy víte (zopakujte a odvoďte!), že $\mathbf{u}_i^T\mathbf{x}$ je délka ortogonální projekce vektoru $\mathbf{x}$ do (normalizovaného) vektoru $\mathbf{u}_i$. Uvědomte si rozdíl oproti situaci, kdy vektory $\mathbf{u}_1, \dots, \mathbf{u}_n$ jsou lineárně nezávislé, ale ne ortonormální (Věta 3.1): v tom případě koeficienty α_i musíme počítat řešením lineární soustavy (4.9).

Dosazení do (4.9) ukáže, že pro $\mathbf{u}_1, \dots, \mathbf{u}_n$ ortonormální a $\mathbf{x} \in \operatorname{span}\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ platí

$$\mathbf{x} = (\mathbf{u}_1^T\mathbf{x})\mathbf{u}_1 + \cdots + (\mathbf{u}_n^T\mathbf{x})\mathbf{u}_n. \tag{4.10}$$

Podotkněme, že ortonormální báze lineárního (pod)prostoru odpovídá tomu, co ze základní školy znáte pod pojmem ‘kartézský souřadnicový systém’. Souřadnice bodu (pod)prostoru vůči jeho ortonormální bázi (viz §3.1) pak známe jako jeho ‘kartézské souřadnice’.

[2] Pro vektory z ortonormální množiny je obvyklé používat písmena $\mathbf{u}, \mathbf{v}, \mathbf{q}$ a ne $\mathbf{x}, \mathbf{y}, \mathbf{a}$, jak jsme zvyklí. Podobně, matice s ortonormálními sloupci (§4.4) je obvyklé značit $\mathbf{U}, \mathbf{V}, \mathbf{Q}$ místo $\mathbf{A}, \mathbf{B}$.

[3] Nepřesně proto, že ortonormalita není vlastnost jednoho vektoru, ale množiny vektorů.

4.4 Matice s ortonormálními sloupci

Nechť sloupce matice $\mathbf{U} \in \mathbb{R}^{m\times n}$ tvoří ortonormální množinu vektorů. Dle Věty 4.3 jsou sloupce $\mathbf{U}$ lineárně nezávislé, tedy nutně $m \geq n$ (tj. $\mathbf{U}$ je čtvercová nebo úzká). Podmínku ortonormality (4.8) sloupců matice $\mathbf{U}$ lze psát stručně jako

$$\mathbf{U}^T\mathbf{U} = \mathbf{I}. \tag{4.11}$$

To také znamená, že $\mathbf{U}^T$ je levá inverze matice $\mathbf{U}$ a $\mathbf{U}$ je pravá inverze matice $\mathbf{U}^T$ (viz §2.4). Dle Věty 4.4 pro každé $\mathbf{x} \in \operatorname{rng}\mathbf{U}$ platí $\mathbf{U}\mathbf{U}^T\mathbf{x} = \mathbf{x}$, což je rovnost (4.10) v maticovém tvaru.

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

$$\mathbf{f}(\mathbf{x})^T\mathbf{f}(\mathbf{y}) = (\mathbf{U}\mathbf{x})^T(\mathbf{U}\mathbf{y}) = \mathbf{x}^T\mathbf{U}^T\mathbf{U}\mathbf{y} = \mathbf{x}^T\mathbf{y}. \tag{4.12}$$

Pro $\mathbf{x} = \mathbf{y}$ dostaneme $\|\mathbf{f}(\mathbf{x})\| = \|\mathbf{U}\mathbf{x}\| = \|\mathbf{x}\|$, neboli zobrazení zachovává také eukleidovskou normu. Zobrazení $\mathbf{f}$ tedy zachovává délky a úhly. Taková zobrazení se nazývají **isometrie**[4].

Věta 4.5. *Pro každou čtvercovou matici* $\mathbf{U}$ *platí*

$$\mathbf{U}^T\mathbf{U} = \mathbf{I} \quad \Longleftrightarrow \quad \mathbf{U}^T = \mathbf{U}^{-1} \quad \Longleftrightarrow \quad \mathbf{U}\mathbf{U}^T = \mathbf{I}. \tag{4.13}$$

Důkaz. Nechť platí levá rovnost. Pak jsou sloupce $\mathbf{U}$ ortonormální a tedy lineárně nezávislé. To ale znamená, že $\mathbf{U}$ je regulární, protože je čtvercová. Vynásobením levé rovnice maticí $\mathbf{U}^{-1}$ zprava získáme prostřední rovnici. Vynásobením prostřední rovnice maticí $\mathbf{U}$ zleva získáme pravou rovnici. 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 (4.13) se říká **ortogonální matice**.

Zdůrazněme, že pokud $\mathbf{U}$ je obdélníková s ortonormálními sloupci, neplatí $\mathbf{U}\mathbf{U}^T = \mathbf{I}$ (platí pouze $\mathbf{U}\mathbf{U}^T\mathbf{x} = \mathbf{x}$ pro každé $\mathbf{x} \in \operatorname{rng}\mathbf{U}$). Dále, pokud má $\mathbf{U}$ ortogonální (ne však ortonormální) sloupce, nemusí mít ortogonální řádky[5].

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 *zrcadlení* (reflexe) kolem nadroviny procházející počátkem.

Příklad 4.4. 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 $\det\mathbf{U} = 1$. □

[4]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.

[5]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 4.5. Zrcadlení (neboli reflexe) v $\mathbb{R}^2$ kolem přímky procházející počátkem a směrnicí $\tan(\varphi/2)$ je reprezentováno ortogonální maticí

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

□

Příklad 4.6. 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.□

4.5 QR rozklad

Věta 4.6. *Každou matici $\mathbf{A} \in \mathbb{R}^{m\times n}$ kde $m \geq n$ lze rozložit na součin*

$$\mathbf{A} = \mathbf{QR}, \tag{4.14}$$

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

Rozkladu matice (4.14) se říká **QR rozklad** matice $\mathbf{A}$. QR rozklad se počítá (což zároveň dokáže Větu 4.6) pomocí algoritmů založených na *Gramm-Schmidtově ortogonalizaci*, *Householderových reflexích* nebo *Givensových rotacích* (neuvádíme). V Matlabu jej spočítáte příkazem `[Q,R]=qr(A,0)`. Podotkněme, že Věta 4.6 popisuje tzv. *redukovanou* verzi QR rozkladu. Příkaz `[Q,R]=qr(A)` počítá *plnou* verzi, která doplní matici $\mathbf{Q}$ přidáním sloupců na ortogonální matici $m \times m$ a matici $\mathbf{R}$ nulovými řádky na rozměr $m \times n$. Napište si v Matlabu příkaz `help qr`!

Vztah (4.14) říká, že každý sloupec matice $\mathbf{A}$ je lineární kombinací sloupců matice $\mathbf{Q}$ (příslušný sloupec matice $\mathbf{R}$ jsou koeficienty této lineární kombinace). Jelikož rozměry matic $\mathbf{A}$ a $\mathbf{Q}$ jsou stejné, znamená to, že sloupce matice $\mathbf{Q}$ tvoří ortonormální bázi podprostoru rng $\mathbf{A}$. Z Věty 4.6 tedy plyne, že *každý podprostor má ortonormální bázi* (což je netriviální tvrzení!).

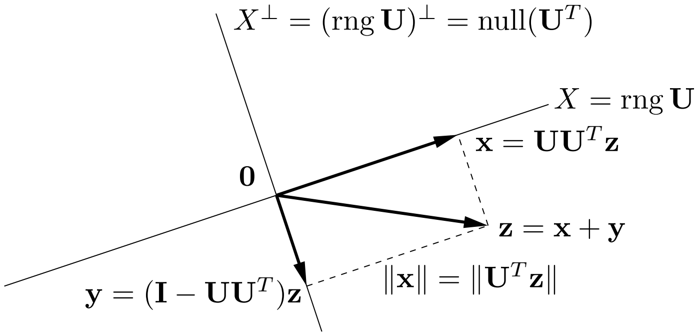
QR rozklad je velmi užitečný. Typické je jeho užití na řešení lineárních soustav. Řešíme-li soustavu $\mathbf{Ax} = \mathbf{b}$, rozložíme $\mathbf{A} = \mathbf{QR}$ a vynásobíme soustavu zleva $\mathbf{Q}^T$, což dá

$$\mathbf{Rx} = \mathbf{Q}^T\mathbf{b}. \tag{4.15}$$

Toto je 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 (4.15) lze levně najít zpětnou substitucí.

4.6 Ortogonální projekce

Ortogonální projekce vektoru $\mathbf{z} \in \mathbb{R}^m$ na podprostor $X \subseteq \mathbb{R}^m$ je vektor $\mathbf{x} \in X$ takový, že $(\mathbf{z} - \mathbf{x}) \perp X$. Viz obrázek:



Ukážeme, jak ortogonální projekci spočítat, je-li dána ortonormální báze podprostoru X, tedy je-li dána matice $\mathbf{U} \in \mathbb{R}^{m\times n}$ taková, že $X = \operatorname{rng}\mathbf{U}$ a $\mathbf{U}^T\mathbf{U} = \mathbf{I}$. Protože (jak ukážeme v §4.5) každý podprostor má ortonormální bázi, dokážeme tím zároveň i existenci a jednoznačnost ortogonální projekce. Případem, kdy podprostor X je zadán libovolnou (ne nutně ortonormální) bází, se budeme zabývat později v §5.1.1.

Věta 4.7. *Nechť matice $\mathbf{U} \in \mathbb{R}^{m\times n}$ se sloupci $\mathbf{u}_1, \ldots, \mathbf{u}_n \in \mathbb{R}^m$ splňuje $\mathbf{U}^T\mathbf{U} = \mathbf{I}$. Ortogonální projekce vektoru $\mathbf{z} \in \mathbb{R}^m$ na podprostor $\operatorname{rng}\mathbf{U}$ je vektor*

$$\mathbf{x} = \mathbf{U}\mathbf{U}^T\mathbf{z} = (\mathbf{u}_1^T\mathbf{z})\mathbf{u}_1 + \cdots + (\mathbf{u}_n^T\mathbf{z})\mathbf{u}_n. \tag{4.16}$$

Důkaz. Hledáme vektor $\mathbf{x} \in \operatorname{rng}\mathbf{U}$ splňující $(\mathbf{z} - \mathbf{x}) \perp \operatorname{rng}\mathbf{U}$. Dle (3.10) je $\mathbf{x} \in \operatorname{rng}\mathbf{U}$ právě když $\mathbf{x} = \mathbf{U}\boldsymbol{\alpha}$ pro nějaké $\boldsymbol{\alpha}$. Podmínka $(\mathbf{z} - \mathbf{x}) \perp \operatorname{rng}\mathbf{U}$ říká, že vektor $\mathbf{z} - \mathbf{x} = \mathbf{z} - \mathbf{U}\boldsymbol{\alpha}$ je kolmý na sloupce matice $\mathbf{U}$. To lze psát jako $\mathbf{U}^T(\mathbf{z} - \mathbf{U}\boldsymbol{\alpha}) = \mathbf{0}$ neboli $\mathbf{U}^T\mathbf{z} = \mathbf{U}^T\mathbf{U}\boldsymbol{\alpha}$. Protože $\mathbf{U}^T\mathbf{U} = \mathbf{I}$, tato rovnice má právě jedno řešení $\boldsymbol{\alpha} = \mathbf{U}^T\mathbf{z}$. Po dosazení $\mathbf{x} = \mathbf{U}\boldsymbol{\alpha} = \mathbf{U}\mathbf{U}^T\mathbf{z}$.

Druhá rovnost v (4.16) je jen otázka značení. Uvědomíme si[6], že $\mathbf{u}\mathbf{u}^T\mathbf{z} = (\mathbf{u}^T\mathbf{z})\mathbf{u}$, a pak

$$\mathbf{U}\mathbf{U}^T\mathbf{z} = \mathbf{u}_1\mathbf{u}_1^T\mathbf{z} + \cdots + \mathbf{u}_n\mathbf{u}_n^T\mathbf{z} = (\mathbf{u}_1^T\mathbf{z})\mathbf{u}_1 + \cdots + (\mathbf{u}_n^T\mathbf{z})\mathbf{u}_n. \qquad \square$$

Poznámky k větě:

- Skaláry $\alpha_i = \mathbf{u}_i^T\mathbf{z}$ v (4.16) jsou souřadnice vektoru $\mathbf{x}$ v ortonormální bázi $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ podprostoru X (srov. Věta 4.4). Řečeno maticově, $\boldsymbol{\alpha} = \mathbf{U}^T\mathbf{z} \in \mathbb{R}^n$ jsou souřadnice vektoru $\mathbf{x} = \mathbf{U}\boldsymbol{\alpha}$ v ortonormální bázi tvořené sloupci $\mathbf{U}$.
- Přidáme-li k vektorům $\mathbf{u}_1, \ldots, \mathbf{u}_n$ vektory $\mathbf{u}_{n+1}, \ldots, \mathbf{u}_m$ tak, aby množina $\mathbf{u}_1, \ldots, \mathbf{u}_m$ byla ortonormální, dostaneme ortonormální bázi prostoru $\mathbb{R}^m$. Dle (4.10) tedy pro každé $\mathbf{z} \in \mathbb{R}^m$ máme
$$\mathbf{z} = (\mathbf{u}_1^T\mathbf{z})\mathbf{u}_1 + \cdots + (\mathbf{u}_m^T\mathbf{z})\mathbf{u}_m. \tag{4.17}$$
Vidíme, že (4.16) je vlastně 'zkrácená' suma (4.17), ze které jsme vzali jen prvních n členů.
- Má-li matice $\mathbf{U}$ jediný sloupec (označme ho $\mathbf{U} = \mathbf{u} \in \mathbb{R}^m$), pak
$$\mathbf{x} = \mathbf{u}\mathbf{u}^T\mathbf{z} = (\mathbf{u}^T\mathbf{z})\mathbf{u}. \tag{4.18}$$
To je tedy vzoreček pro projekci vektoru $\mathbf{z}$ na přímku $X = \operatorname{rng}\mathbf{u} = \operatorname{span}\{\mathbf{u}\}$ procházející počátkem (kde $\|\mathbf{u}\| = 1$), který máte znát ze střední školy. Středoškolské odvození: skalární součin $\mathbf{u}^T\mathbf{z}$ je délka (se znaménkem) průmětu a (4.18) je vektor této délky ve směru $\mathbf{u}$.

[6]Pozor:, závorka ve výrazu $(\mathbf{u}^T\mathbf{z})\mathbf{u}$ je nutná, protože součin výrazů $\mathbf{u}^T\mathbf{z}$ a $\mathbf{u}$ není maticový součin, ale násobení vektoru skalárem (viz poznámka v §2.1). Oproti tomu, výraz $\mathbf{u}\mathbf{u}^T\mathbf{z}$ je maticový součin tří matic $\mathbf{u}, \mathbf{u}^T, \mathbf{z}$, takže závorky v něm být nemusí.

Zobrazení dané vzorcem (4.16) (za předpokladu $\mathbf{U}^T\mathbf{U} = \mathbf{I}$) je lineární zobrazení reprezentované maticí

$$\mathbf{P} = \mathbf{U}\mathbf{U}^T, \tag{4.19}$$

které se proto říká **ortogonální projektor** (na podprostor $X = \operatorname{rng}\mathbf{U}$). Z (4.19) ihned plynou (ověřte!) dvě vlastnosti ortogonálního projektoru:

$$\mathbf{P}^2 = \mathbf{P} = \mathbf{P}^T \tag{4.20}$$

(kde $\mathbf{P}^2$ je zkratka pro $\mathbf{P}\mathbf{P}$). Vlastnost $\mathbf{P}^2 = \mathbf{P}$ (říká se jí *idempotence* matice $\mathbf{P}$) říká očividnou věc: když vektor jednou promítneme na nějaký podprostor, tak opětovné promítnutí na tentýž podprostor ho už nezmění. Dá se ukázat (důkaz neuvádíme), že každá čtvercová matice s vlastnostmi (4.20) je ortogonálni projektor (na nějaký podprostor).

4.6.1 Projekce na ortogonální doplněk

Pohled na ortogonální projekci se stane úplnější, zahrneme-li do něj kromě podprostoru X i jeho ortogonální doplněk $X^\perp$ (viz obrázek výše). Vlastně jsme ukázali, že pro každý vektor $\mathbf{z} \in \mathbb{R}^m$ existuje právě jedna dvojice vektorů $\mathbf{x} \in X$ a $\mathbf{y} \in X^\perp$ tak, že $\mathbf{x} + \mathbf{y} = \mathbf{z}$. Opravdu, podmínku $\mathbf{y} \in X^\perp$ lze psát jako $\mathbf{y} \perp X$ neboli $(\mathbf{z} - \mathbf{x}) \perp X$ (domyslete do konce!).

Co je prostorem obrazů a nulovým prostorem ortogonálního projektoru? Úvahou (Kam se promítne každý vektor? Které vektory se promítnou do počátku?) snadno vidíme, že

$$\operatorname{rng}\mathbf{P} = X, \tag{4.21a}$$
$$\operatorname{null}\mathbf{P} = X^\perp. \tag{4.21b}$$

Algebraicky to také plyne z Vět 3.6 a 3.8.

Dále z obrázku vidíme, že je-li $\mathbf{P}$ projektor na X, tak $\mathbf{I} - \mathbf{P}$ je projektor na $X^\perp$. Vskutku:

$$\mathbf{y} = \mathbf{z} - \mathbf{x} = \mathbf{z} - \mathbf{U}\mathbf{U}^T\mathbf{z} = \mathbf{z} - \mathbf{P}\mathbf{z} = (\mathbf{I} - \mathbf{P})\mathbf{z}. \tag{4.22}$$

Rozveďme tuto myšlenku: zkoumejme ortogonální matici $\begin{bmatrix}\mathbf{U} & \mathbf{V}\end{bmatrix} \in \mathbb{R}^{m\times m}$, která je složená ze dvou bloků $\mathbf{U} \in \mathbb{R}^{m\times n}$ a $\mathbf{V} \in \mathbb{R}^{m\times(m-n)}$. Z ortonormality sloupců máme $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ a $\mathbf{V}^T\mathbf{V} = \mathbf{I}$, a dále $\mathbf{U}^T\mathbf{V} = \mathbf{0}$ neboli $\operatorname{rng}\mathbf{U} \perp \operatorname{rng}\mathbf{V}$. Platí ovšem dokonce

$$(\operatorname{rng}\mathbf{U})^\perp = \operatorname{rng}\mathbf{V}, \tag{4.23}$$

což plyne z Věty 4.1, neboť $\dim\operatorname{rng}\mathbf{U} + \dim\operatorname{rng}\mathbf{V} = n + (m-n) = m$. Vidíme, že rozdělení sloupců ortogonální matice do dvou bloků generuje rozklad prostoru na podprostor a jeho ortogonální doplněk. Navíc z Věty 4.5 máme

$$\begin{bmatrix}\mathbf{U} & \mathbf{V}\end{bmatrix}\begin{bmatrix}\mathbf{U} & \mathbf{V}\end{bmatrix}^T = \mathbf{U}\mathbf{U}^T + \mathbf{V}\mathbf{V}^T = \mathbf{I}, \tag{4.24}$$

jak ověříme roznásobením blokových matic (proveďte!). To souhlasí, protože projektor na podprostor X je $\mathbf{P} = \mathbf{U}\mathbf{U}^T$ a tedy projektor na podprostor $X^\perp$ je $\mathbf{V}\mathbf{V}^T = \mathbf{I} - \mathbf{U}\mathbf{U}^T = \mathbf{I} - \mathbf{P}$.

4.6.2 Vzdálenost bodu od podprostoru

Z obrázku na začátku §4.6 je intuitivně jasné, že projekce $\mathbf{x} = \mathbf{P}\mathbf{z}$ je bod podprostoru X nejbližší bodu $\mathbf{z}$. Tato vlastnost ortogonální projekce je důležitá pro optimalizaci, využijeme ji např. v §5. Dokažme ji přesně:

Věta 4.8. *Je-li $\mathbf{x} \in X$ ortogonální projekce vektoru $\mathbf{z}$ na podprostor X, pak pro každé $\mathbf{x}' \in X$ platí $\|\mathbf{z} - \mathbf{x}\| \leq \|\mathbf{z} - \mathbf{x}'\|$.*

Důkaz. Protože X je podprostor a $\mathbf{x}, \mathbf{x}' \in X$, je $\mathbf{x} - \mathbf{x}' \in X$. Protože $(\mathbf{z} - \mathbf{x}) \perp X$, platí $(\mathbf{z} - \mathbf{x}) \perp (\mathbf{x} - \mathbf{x}')$. Z Pythagorovy věty tedy $\|\mathbf{z} - \mathbf{x}'\|^2 = \|\mathbf{z} - \mathbf{x}\|^2 + \|\mathbf{x} - \mathbf{x}'\|^2 \geq \|\mathbf{z} - \mathbf{x}\|^2$. □

Z toho plyne, že *vzdálenost* bodu $\mathbf{z}$ od podprostoru $X^\perp = (\operatorname{rng} \mathbf{U})^\perp = \operatorname{null}(\mathbf{U}^T)$ je rovna délce projekce bodu $\mathbf{z}$ na podprostor $X = \operatorname{rng} \mathbf{U}$, tedy

$$\|\mathbf{x}\| = \|\mathbf{U}\mathbf{U}^T\mathbf{z}\| = \|\mathbf{U}^T\mathbf{z}\|, \tag{4.25}$$

kde jsme použili, že zobrazení reprezentované maticí $\mathbf{U}$ zachovává eukleidovskou normu (viz §4.4). Samozřejmě stále předpokládáme, že $\mathbf{U}$ má ortonormální sloupce.

Jako v §4.6.1, doplňme matici $\mathbf{U}$ na ortogonální matici $\begin{bmatrix}\mathbf{U} & \mathbf{V}\end{bmatrix}$, tedy $X^\perp = \operatorname{rng} \mathbf{V}$. Pak vzdálenost bodu $\mathbf{z}$ od podprostoru X je $\|\mathbf{y}\| = \|\mathbf{V}^T\mathbf{z}\|$. Z (4.24) máme

$$\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 = \|\mathbf{U}^T\mathbf{z}\|^2 + \|\mathbf{V}^T\mathbf{z}\|^2 = \mathbf{z}^T\mathbf{U}\mathbf{U}^T\mathbf{z} + \mathbf{z}^T\mathbf{V}\mathbf{V}^T\mathbf{z} = \mathbf{z}^T(\mathbf{U}\mathbf{U}^T + \mathbf{V}\mathbf{V}^T)\mathbf{z} = \mathbf{z}^T\mathbf{z} = \|\mathbf{z}\|^2. \tag{4.26}$$

To není nic jiného než Pythagorova věta.

Vzorec (4.25) počítá vzdálenost bodu od lineárního podprostoru $X^\perp = \operatorname{null}(\mathbf{U}^T)$, tj. tento podprostor je tvořen řešeními homogenní soustavy $\mathbf{U}^T\mathbf{x} = \mathbf{0}$. Často potřebujeme spočítat vzdálenost bodu of afinního podprostoru. Jak to uděláme? Reprezentujme tento afinní podprostor, $A \subseteq \mathbb{R}^m$, množinou řešení lineární soustavy $\mathbf{U}^T\mathbf{x} = \mathbf{b}$ (viz Věta 3.13), tedy

$$A = \{\, \mathbf{x} \in \mathbb{R}^m \mid \mathbf{U}^T\mathbf{x} = \mathbf{b} \,\} = \{\, \mathbf{x} \in \mathbb{R}^m \mid \mathbf{U}^T(\mathbf{x} - \mathbf{x}_0) = \mathbf{0} \,\} = X^\perp + \mathbf{x}_0, \tag{4.27}$$

kde $\mathbf{x}_0$ je libovolný bod splňující $\mathbf{U}^T\mathbf{x}_0 = \mathbf{b}$. Vidíme, že vzdálenost bodu $\mathbf{z}$ od afinního podprostoru A je rovna vzdálenosti bodu $\mathbf{z} - \mathbf{x}_0$ od lineárního podprostoru $X^\perp$, což je

$$\|\mathbf{U}^T(\mathbf{z} - \mathbf{x}_0)\| = \|\mathbf{U}^T\mathbf{z} - \mathbf{b}\|. \tag{4.28}$$

Speciálně, necht' náš afinní podprostor je nadrovina $\{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}^T\mathbf{x} = b \,\}$, kde předpokládáme $\mathbf{a} \neq \mathbf{0}$ (jinak by podprostor nebyl nadrovinou, tj. neměl by dimenzi $n - 1$). Jestliže $\|\mathbf{a}\| = 1$, vzdálenost bodu $\mathbf{z}$ od nadroviny je rovna $|\mathbf{a}^T\mathbf{z} - b\|$. Obecně je tato vzdálenost rovna $|\mathbf{a}^T\mathbf{z} - b\|/\|\mathbf{a}\|$. Vidíme zde geometrický význam vektoru $\mathbf{a}$ a čísla b: $\mathbf{a}$ je normálový vektor nadroviny a $|b|/\|\mathbf{a}\|$ je vzdálenost nadroviny od počátku.

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

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

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

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

4.6. Počet nezávislých parametrů ("stupňů volnosti") ortogonální matice $n \times n$ se získá jako rozdíl počtu prvků matice (n^2) 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 n-rozměrném prostoru. Jaké je toto číslo pro $n = 2, 3, 4$? Najděte vzorec pro obecné n.

4.7. Pokud $\|\mathbf{x}\| = \|\mathbf{y}\|$, dokažte, že $(\mathbf{x} + \mathbf{y}) \perp (\mathbf{x} - \mathbf{y})$. V prostoru $\mathbb{R}^2$ nakreslete vektory $\mathbf{x} + \mathbf{y}$ a $\mathbf{x} - \mathbf{y}$.

4.8. 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 $\operatorname{span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}^\perp$.
c) Najděte ortogonální projektory na podprostory $\operatorname{span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ a $\operatorname{span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}^\perp$.

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

4.10. Spočtěte co nejjednodušším způsobem inverzi matice $\frac{1}{3}\begin{bmatrix} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{bmatrix}$. Je tato matice ortogonální projektor?

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

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

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

4.13. Pro $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ reprezentuje matice $\mathbf{H} = \mathbf{I} - 2\mathbf{U}\mathbf{U}^T$ zrcadlení (reflexi) vzhledem k podprostoru $X = (\operatorname{rng}\mathbf{U})^\perp$, proto této matici budeme říkat *reflektor*.

a) Odvoďte vzorec pro reflektor geometrickou úvahou, podobnou jako pro projektor.
b) Dokažte, že $\mathbf{H}^{-1} = \mathbf{H} = \mathbf{H}^T$ (srov. s (4.20)).
c) Co je $\mathbf{H}\mathbf{x}$ když $\mathbf{x} \in X$? Ukažte algebraicky a odůvodněte geometricky.
d) Co je $\mathbf{H}\mathbf{x}$ když $\mathbf{x} \perp X$? Ukažte algebraicky a odůvodněte geometricky.

Poznamenejme, že známější je případ, kdy $\mathbf{U}$ má jediný sloupec, tedy $\mathbf{H} = \mathbf{I} - 2\mathbf{u}\mathbf{u}^T$ kde $\mathbf{u}^T\mathbf{u} = 1$. Tato matice reprezentuje zrcadlení kolem nadroviny s normálovým vektorem $\mathbf{u}$ a je známa jako *elementární reflektor* nebo *Householderova matice*.

4.14. ($\star$) *RQ rozklad* rozloží matici $\mathbf{A} = \mathbf{R}\mathbf{Q}$, kde $\mathbf{R}$ je horní trojúhelníková a $\mathbf{Q}$ je ortogonální. Jak byste spočítali RQ rozklad z QR rozkladu?

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

4.16. Dokažte: Jestliže sloupce matice $\mathbf{U}$ tvoří ortonormální bázi nějakého podprostoru, pak sloupce matice $\mathbf{V} = \mathbf{UC}$ tvoří ortonormální bázi téhož podprostoru pro každou ortogonální matici $\mathbf{C}$.

4.17. Nechť $\mathbf{U}^T\mathbf{U} = \mathbf{I}$. Najděte co nejjednodušší vzorec pro

a) vzdálenost bodu $\mathbf{x}$ od podprostoru $\operatorname{rng}\mathbf{U}$,
b) vzdálenost bodu $\mathbf{x}$ od podprostoru $\operatorname{null}(\mathbf{U}^T) = \{\, \mathbf{x} \mid \mathbf{U}^T\mathbf{x} = \mathbf{0} \,\}$,
c) vzdálenost počátku $\mathbf{0}$ od afinního podprostoru $\{\, \mathbf{x} \mid \mathbf{U}^T\mathbf{x} = \mathbf{b} \,\}$,
d) vzdálenost bodu $\mathbf{x}$ od afinního podprostoru $\{\, \mathbf{x} \mid \mathbf{U}^T\mathbf{x} = \mathbf{b} \,\}$.

4.18. Nechť $X, Y \subseteq \mathbb{R}^n$ jsou podprostory takové, že $X \subseteq Y$. Tvrdíme, že když vektor ortogonálně promítneme nejdříve na Y a potom na X, dostaneme stejný výsledek, jako když ho ortogonálně promítneme rovnou na X. Dokažte toto tvrzení algebraicky nebo ho vyvraťte.

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

a) Prvky matice $\mathbf{U}$ splňují $|u_{ij}| \le 1$.
b) Pro každý řádek $\mathbf{u}^T$ matice $\mathbf{U}$ platí $\|\mathbf{u}\| \le 1$.

4.20. Nechť $\mathbf{P}$ je ortogonální projektor (na nějaký podprostor). Dokažte, že

a) Prvky matice $\mathbf{P}$ splňují $|p_{ij}| \le 1$.
b) Diagonální prvky matice $\mathbf{P}$ jsou nezáporné.

4.21. Obecnou *projekcí* se v lineární algebře rozumí každé lineární zobrazení $\mathbf{f}(\mathbf{z}) = \mathbf{Pz}$, které je idempotentní, tedy $\mathbf{f}(\mathbf{f}(\mathbf{z})) = \mathbf{f}(\mathbf{z})$ neboli $\mathbf{P}^2 = \mathbf{P}$. Projekce nemusí být ortogonální, může být šikmá – pak promítáme ve směru podprostoru $\operatorname{null}\mathbf{P}$ na podprostor $\operatorname{rng}\mathbf{P}$.

a) Dokažte, že pro každý vektor $\mathbf{z}$ platí $\mathbf{z} - \mathbf{Pz} \in \operatorname{null}\mathbf{P}$.
b) Projekce je ortogonální, právě když $\mathbf{P}^T = \mathbf{P}$. Dokažte, že potom $(\operatorname{rng}\mathbf{P})^\perp = \operatorname{null}\mathbf{P}$.

4.22. Nechť $\mathbf{A}, \mathbf{B}$ jsou libovolné matice splňující $\mathbf{AB} = \mathbf{I}$ (tedy $\mathbf{A}$ je levá inverze $\mathbf{B}$ a $\mathbf{B}$ je pravá inverze $\mathbf{A}$, viz §2.4). Dokažte, že:

a) $\mathbf{P} = \mathbf{BA}$ je (obecný) projektor dle Cvičení 4.21, tedy platí $\mathbf{P}^2 = \mathbf{P}$.
b) $\operatorname{null}\mathbf{P} = \operatorname{null}\mathbf{A}$ a $\operatorname{rng}\mathbf{P} = \operatorname{rng}\mathbf{B}$.
c) Co jsou matice $\mathbf{A}, \mathbf{B}$ v případě ortogonální projekce?

4.23. Pro jaké vektory $\mathbf{x}, \mathbf{y}$ platí trojúhelníková nerovnost $\|\mathbf{x}+\mathbf{y}\| \le \|\mathbf{x}\| + \|\mathbf{y}\|$ s rovností?

4.24. Dokažte Pythagorovu větu: $\mathbf{x} \perp \mathbf{y} \implies \|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$. Dokažte její zobecnění: jsou-li vektory $\mathbf{x}_1, \ldots, \mathbf{x}_k$ po dvojicích ortogonální, pak $\|\mathbf{x}_1 + \cdots + \mathbf{x}_k\|^2 = \|\mathbf{x}_1\|^2 + \cdots + \|\mathbf{x}_k\|^2$.

4.25. Zobrazení $\mathbf{F}\colon \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ je dané vzorcem $\mathbf{F}(\mathbf{A}) = (\mathbf{I}-\mathbf{A})(\mathbf{I}+\mathbf{A})^{-1}$. Předpokládejte, že $\mathbf{A}$ je taková, že $\mathbf{I}+\mathbf{A}$ je regulární. Dokažte, že:

a) Pro každou $\mathbf{A}$ je $(\mathbf{I}-\mathbf{A})(\mathbf{I}+\mathbf{A}) = (\mathbf{I}+\mathbf{A})(\mathbf{I}-\mathbf{A})$.
b) Pro každou $\mathbf{A}$ je $(\mathbf{I}-\mathbf{A})(\mathbf{I}+\mathbf{A})^{-1} = (\mathbf{I}+\mathbf{A})^{-1}(\mathbf{I}-\mathbf{A})$.
c) Pro každou antisymetrickou $\mathbf{A}$ je $\mathbf{F}(\mathbf{A})$ ortogonální.
d) Pro každou ortogonální $\mathbf{A}$ je $\mathbf{F}(\mathbf{A})$ antisymetrická.

e) Zobrazení $\mathbf{F}$ je inverzí sama sebe, tedy $\mathbf{F}(\mathbf{F}(\mathbf{A})) = \mathbf{A}$ pro každé $\mathbf{A}$.

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

Nápověda a řešení

4.2. Např. $(1, 1, -1)$

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

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

4.6. Podmínka $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ je soustava rovnic $\mathbf{u}_i^T\mathbf{u}_j = \delta_{ij}$ pro $i, j = 1, \ldots, n$, kde $\mathbf{u}_1, \ldots, \mathbf{u}_n$ 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 $n^2 - \binom{n}{2} - n = \binom{n}{2} = n(n-1)/2$.

4.9. 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 Gram-Smidtovy ortogonalizace, jestli ji znáte.)

4.10. Není matice náhodou ortogonální? Ort. projektor to není, protože nesplňuje $\mathbf{P}^2 = \mathbf{P}$.

4.11.c) Plyne z (b).

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

4.13.c) Je $\mathbf{x} \in X = (\operatorname{rng} \mathbf{U})^\perp = \operatorname{null}(\mathbf{U}^T)$, tedy $\mathbf{U}^T\mathbf{x} = \mathbf{0}$, tedy $\mathbf{H}\mathbf{x} = \mathbf{x}$. Geometricky to znamená, že když vektor leží v rovině zrcadlení, tak ho zrcadlení nezmění.

4.13.d) $\mathbf{x} \perp X$ je to samé jako $\mathbf{x} \in X^\perp = ((\operatorname{rng} \mathbf{U})^\perp)^\perp = \operatorname{rng} \mathbf{U}$, tedy $\mathbf{x} = \mathbf{U}\boldsymbol{\alpha}$ pro nějaké $\boldsymbol{\alpha}$. Tedy $\mathbf{U}\mathbf{U}^T\mathbf{x} = \mathbf{U}\mathbf{U}^T\mathbf{U}\boldsymbol{\alpha} = \mathbf{U}\boldsymbol{\alpha} = \mathbf{x}$ (taky to lze vidět z (4.21a), protože $\mathbf{U}\mathbf{U}^T$ je projektor na X). Tedy $\mathbf{H}\mathbf{x} = -\mathbf{x}$. Geometricky: zrcadlení vektoru kolmého k rovině zrcadlení vektoru obrátí orientaci.

4.15. Ne, protože isometrie zachovává eukleidovskou normu, ale $\|(1, -1, 2)\| \neq \|(1, 2, -1, 1)\|$.

4.16. Víme, že $\mathbf{U} \in \mathbb{R}^{m\times n}$ a $\mathbf{C} \in \mathbb{R}^{n\times n}$, kde $m \geq n$, a $\mathbf{U}^T\mathbf{U} = \mathbf{I}$, $\mathbf{C}^T\mathbf{C} = \mathbf{I} = \mathbf{C}\mathbf{C}^T$. Sloupce matic $\mathbf{U}$ a $\mathbf{V}$ tvoří báze stejného podprostoru, nebot' dle Věty 3.6 je $\operatorname{rng} \mathbf{U} = \operatorname{rng} \mathbf{V}$ ($\mathbf{C}$ je matice přechodu k jiné bázi). Sloupce matice $\mathbf{V}$ jsou ortonormální, nebot' $\mathbf{V}^T\mathbf{V} = \mathbf{C}^T\mathbf{U}^T\mathbf{U}\mathbf{C} = \mathbf{C}^T\mathbf{C} = \mathbf{I}$.

4.17.a) $\|(\mathbf{I} - \mathbf{U}\mathbf{U}^T)\mathbf{x}\|$, víc to už zjednodušit nejde.

4.17.b) $\|\mathbf{U}^T\mathbf{x}\|$, viz §4.6.2.

4.17.c) $\|\mathbf{b}\|$, viz §4.6.2.

4.17.d) $\|\mathbf{U}^T\mathbf{x} - \mathbf{b}\|$, viz §4.6.2.

4.18. Je pravdivé. Důkaz: Necht' $\begin{bmatrix}\mathbf{U} & \mathbf{V}\end{bmatrix}$ je matice s ortonormálními sloupci taková, že $X = \operatorname{rng} \mathbf{U}$ a $Y = \operatorname{rng} \begin{bmatrix}\mathbf{U} & \mathbf{V}\end{bmatrix}$. Z toho plyne $X \subseteq Y$. Všimněte si (použijeme později), že $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ a $\mathbf{U}^T\mathbf{V} = \mathbf{0}$. Označme $\mathbf{P} = \mathbf{U}\mathbf{U}^T$ projektor na X a $\mathbf{Q} = \begin{bmatrix}\mathbf{U} & \mathbf{V}\end{bmatrix}\begin{bmatrix}\mathbf{U} & \mathbf{V}\end{bmatrix}^T$ projektor na Y. Máme dokázat, že pro každé $\mathbf{z}$ platí $\mathbf{P}\mathbf{Q}\mathbf{z} = \mathbf{P}\mathbf{z}$, neboli $\mathbf{P}\mathbf{Q} = \mathbf{P}$. Platí $\mathbf{P}\mathbf{Q} = \mathbf{U}\mathbf{U}^T\begin{bmatrix}\mathbf{U} & \mathbf{V}\end{bmatrix}\begin{bmatrix}\mathbf{U} & \mathbf{V}\end{bmatrix}^T = \mathbf{U}\begin{bmatrix}\mathbf{U}^T\mathbf{U} & \mathbf{U}^T\mathbf{V}\end{bmatrix}\begin{bmatrix}\mathbf{U} & \mathbf{V}\end{bmatrix}^T = \mathbf{U}\begin{bmatrix}\mathbf{I} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{U} & \mathbf{V}\end{bmatrix}^T = \mathbf{U}\mathbf{U}^T = \mathbf{P}$.

4.19.b) Doplňme matici $\mathbf{U}$ na ortogonální matici $\mathbf{W} = \begin{bmatrix}\mathbf{U} & \mathbf{V}\end{bmatrix}$. Řádek této matice je $\mathbf{w}^T = \begin{bmatrix}\mathbf{u}^T & \mathbf{v}^T\end{bmatrix}$. Z ortogonality $\mathbf{W}$ je však $\mathbf{w}^T\mathbf{w} = \mathbf{u}^T\mathbf{u} + \mathbf{v}^T\mathbf{v} = 1$. Z toho $\mathbf{u}^T\mathbf{u} \leq 1$.

4.20.a) Z $\mathbf{P} = \mathbf{U}\mathbf{U}^T$ máme $p_{ij} = \mathbf{u}_i^T\mathbf{u}_j$, kde $\mathbf{u}_i$ je i-tý řádek $\mathbf{U}$. Protože $\|\mathbf{u}_i\| \leq 1$, musí být $|\mathbf{u}_i^T\mathbf{u}_j| \leq 1$.

4.20.b) Máme $p_{ii} = \mathbf{u}_i^T\mathbf{u}_i \geq 0$

4.21.a) $\mathbf{z} - \mathbf{P}\mathbf{z} \in \operatorname{null} \mathbf{P}$ znamená $\mathbf{P}(\mathbf{z} - \mathbf{P}\mathbf{z}) = \mathbf{P}(\mathbf{I} - \mathbf{P})\mathbf{z} = \mathbf{0}$. Ale $\mathbf{P}(\mathbf{I} - \mathbf{P}) = \mathbf{P} - \mathbf{P}^2 = \mathbf{P} - \mathbf{P} = \mathbf{0}$.

4.21.b) Dle (4.7a) je $(\operatorname{rng} \mathbf{P})^\perp = \operatorname{null}(\mathbf{P}^T) = \operatorname{null} \mathbf{P}$. Srov. s Cvičením 5.16.

4.22.a) $\mathbf{P}^2 = \mathbf{P}\mathbf{P} = \mathbf{B}\mathbf{A}\mathbf{B}\mathbf{A} = \mathbf{B}\mathbf{A} = \mathbf{P}$

4.22.b) $\mathbf{B}$ má levou inverzi a tedy dle Věty 3.7 má l.n. sloupce. Dle Věty 3.8 tedy $\operatorname{null}\mathbf{P} = \operatorname{null}(\mathbf{BA}) = \operatorname{null}\mathbf{A}$. $\operatorname{rng}\mathbf{P} = \operatorname{rng}\mathbf{B}$ dokážeme podobně z Vět 3.5 a 3.6.

4.22.c) $\mathbf{A} = \mathbf{U}$ a $\mathbf{B} = \mathbf{U}^T$, kde $\mathbf{U}$ má ortonormální sloupce.

Kapitola 5

Nehomogenní lineární soustavy

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 $\operatorname{rank}[\mathbf{A}\ \mathbf{b}] = \operatorname{rank} \mathbf{A}$ (Frobeniova věta). Pokud je množina řešení soustavy neprázdná, je to 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 §5.2.

5.1 Přibližné řešení ve smyslu nejmenších čtverců

Pokud soustava (5.1) nemá řešení, řešme ji přibližně (což můžeme značit $\mathbf{A}\mathbf{x} \approx \mathbf{b}$). Hledejme takové $\mathbf{x}$, aby eukleidovská norma vektoru $\mathbf{r} = \mathbf{b} - \mathbf{A}\mathbf{x}$ zbytků (neboli *reziduí*) byla co nejmenší. Úloha se nezmění (proč?), když místo eukleidovské normy budeme minimalizovat její čtverec $\|\mathbf{r}\|^2 = \mathbf{r}^T\mathbf{r} = r_1^2 + \cdots + r_m^2$. Tedy řešíme úlohu

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2. \tag{5.2}$$

Protože minimalizujeme součet čtverců reziduí, mluvíme o přibližném řešení soustavy **ve smyslu nejmenších čtverců** (*least squares solution*)[1].

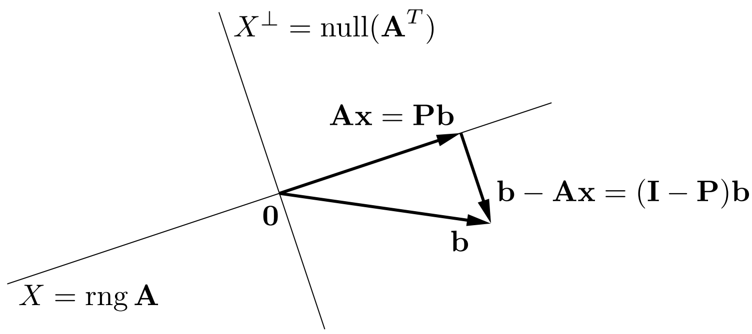
Příklad 5.1. Soustava třech rovnic o dvou neznámých

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

[1] Přesně by se mělo říkat *least sum of squares*, protože je i metoda založená na *least median of squares*.

je přeurčená. Její přibližné řešení ve smyslu nejmenších čtverců znamená najít taková čísla x, y, která minimalizují číslo $(x+2y-6)^2 + (-x+y-3)^2 + (x+y-4)^2$. $\square$

Úlohu (5.2) vyřešíme následující úvahou, patrnou z obrázku:



Pokud vzdálenost $\|\mathbf{Ax}-\mathbf{b}\|$ bodů $\mathbf{Ax}$ a $\mathbf{b}$ má být minimální, musí být vektor $\mathbf{b}-\mathbf{Ax}$ kolmý na podprostor $\text{rng}\,\mathbf{A}$. To je intuitivně jasné a přesně jsme to dokázali ve Větě 4.8. To ale znamená (viz (4.4)), že vektor $\mathbf{b} - \mathbf{Ax}$ musí být kolmý na každý sloupec matice $\mathbf{A}$. Tuto podmínku lze zapsat jako $\mathbf{A}^T(\mathbf{Ax} - \mathbf{b}) = \mathbf{0}$, tedy

$$\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}. \tag{5.3}$$

Tedy máme tento výsledek: $\mathbf{x}$ je optimální řešení optimalizační úlohy (5.2) právě tehdy, když $\mathbf{x}$ je řešením soustavy (5.3). Soustava (5.3) se nazývá **normální rovnice** (protože normála = kolmice). Je to soustava n rovnic o n neznámých. Abychom mohli zkoumat její řešitelnost, uvedeme následující větu.

Věta 5.1. *Pro každou matici* $\mathbf{A}$ *platí*[2]

$$\text{rng}(\mathbf{A}^T\mathbf{A}) = \text{rng}(\mathbf{A}^T), \tag{5.4a}$$
$$\text{null}(\mathbf{A}^T\mathbf{A}) = \text{null}\,\mathbf{A}. \tag{5.4b}$$

Důkaz. Dokažme nejprve rovnost (5.4b), tj. $\mathbf{A}^T\mathbf{Ax} = \mathbf{0} \iff \mathbf{Ax} = \mathbf{0}$. Implikace $\Leftarrow$ je jasná, vynásobením $\mathbf{Ax} = \mathbf{0}$ zleva maticí $\mathbf{A}^T$. Implikace $\Rightarrow$ se dokáže takto:

$$\mathbf{A}^T\mathbf{Ax} = \mathbf{0} \quad \Longrightarrow \quad \mathbf{x}^T\mathbf{A}^T\mathbf{Ax} = (\mathbf{Ax})^T(\mathbf{Ax}) = 0 \quad \Longrightarrow \quad \mathbf{Ax} = \mathbf{0},$$

neboť pro libovolný vektor $\mathbf{y}$ platí $\mathbf{y}^T\mathbf{y} = 0 \implies \mathbf{y} = \mathbf{0}$ (proč?).

Dokažme (5.4a). Z definice (3.10) je jasné (viz Věta 3.6), že $\text{rng}(\mathbf{A}^T\mathbf{A}) \subseteq \text{rng}(\mathbf{A}^T)$. Nyní použijeme Větu 3.11 jednou na matici $\mathbf{A}$ a jednou na matici $\mathbf{A}^T\mathbf{A}$:

$$\dim\text{rng}\,\mathbf{A} + \dim\text{null}\,\mathbf{A} = n = \dim\text{rng}(\mathbf{A}^T\mathbf{A}) + \dim\text{null}(\mathbf{A}^T\mathbf{A}).$$

Díky (5.4b) z toho máme $\dim\text{rng}(\mathbf{A}^T\mathbf{A}) = \dim\text{rng}\,\mathbf{A} = \dim\text{rng}(\mathbf{A}^T)$, kde druhá rovnost je (2.2). Ale pokud je podprostor podmnožinou jiného podprostoru a oba mají stejnou dimenzi, jsou stejné (dle Věty 3.3). $\square$

[2] Matice tvaru $\mathbf{A}^T\mathbf{A}$ či $\mathbf{AA}^T$ se objevují v různých situacích. Označme jako $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^m$ sloupce matice $\mathbf{A}$. Matici $\mathbf{A}^T\mathbf{A} \in \mathbb{R}^{n\times n}$ se říká *Gramova matice* vektorů $\mathbf{a}_1, \dots, \mathbf{a}_n$ a její prvky jsou skalární součiny $\mathbf{a}_i^T\mathbf{a}_j$ (viz (2.8)). Matici $\mathbf{AA}^T \in \mathbb{R}^{m\times m}$ lze zase vidět (až na skalární násobek) jako empirickou kovarianční matici n pozorování $\mathbf{a}_1, \dots, \mathbf{a}_n$ m-tice náhodných proměnných.

Soustava (5.3) má řešení, právě když $\mathbf{A}^T\mathbf{b} \in \operatorname{rng}(\mathbf{A}^T\mathbf{A})$. Ale z (5.4a) je $\operatorname{rng}(\mathbf{A}^T) = \operatorname{rng}(\mathbf{A}^T\mathbf{A})$. Protože $\mathbf{A}^T\mathbf{b} \in \operatorname{rng}(\mathbf{A}^T)$ pro libovolné $\mathbf{A}, \mathbf{b}$, vidíme, že soustava má *vždy* řešení. Je zajímavé, že pro důkaz tohoto tvrzení jsme potřebovali rovnost (5.4a), jejíž důkaz není snadný (používá Větu 3.11). Zdá-li se vám, že musí existovat elementárnější důkaz, schválně ho zkuste najít!

Zkombinujeme-li (5.4a) (použité jednou na matici $\mathbf{A}$ a jednou na matici $\mathbf{A}^T$) a (2.2), máme

$$\operatorname{rank}(\mathbf{A}\mathbf{A}^T) = \operatorname{rank}\mathbf{A} = \operatorname{rank}(\mathbf{A}^T) = \operatorname{rank}(\mathbf{A}^T\mathbf{A}). \tag{5.5}$$

Dle (5.5) je matice $\mathbf{A}^T\mathbf{A}$ regulární, právě když $\mathbf{A}$ má lineárně nezávislé sloupce. V tom případě můžeme soustavu (5.3) řešit pomocí inverze. Řešením je vektor $\mathbf{x} = \mathbf{A}^+\mathbf{b}$, kde

$$\mathbf{A}^+ = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T. \tag{5.6}$$

Matice (5.6) se nazývá **pseudoinverze** matice $\mathbf{A}$ s lineárně nezávislými sloupci. Je to jedna z levých inverzí matice $\mathbf{A}$, neboť $\mathbf{A}^+\mathbf{A} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{A} = \mathbf{I}$.

Má-li matice $\mathbf{A}$ lineárně závislé sloupce, vzorec (5.6) nelze použít (zkrátka proto, že matice $\mathbf{A}^T\mathbf{A}$ nemá inverzi). V tom případě soustava (5.3), a tedy i úloha (5.2), mají nekonečně mnoho (afinní podprostor) řešení (pozor, to je něco jiného, než že soustava (5.1) má nekonečně mnoho řešení!).

5.1.1 Ortogonální projekce na podprostor daný obecnou bází

Pokud $\mathbf{x}$ je řešení normální rovnice, vektor $\mathbf{A}\mathbf{x}$ je ortogonální projekcí vektoru $\mathbf{b}$ na podprostor $X = \operatorname{rng}\mathbf{A}$ (viz obrázek výše). Pokud $\mathbf{A}$ má lineárně nezávislé sloupce (tj. tyto sloupce tvoří bázi podprostoru X), z (5.6) máme $\mathbf{A}\mathbf{x} = \mathbf{P}\mathbf{b}$, kde

$$\mathbf{P} = \mathbf{A}\mathbf{A}^+ = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T. \tag{5.7}$$

Toto je tedy projektor na podprostor X s bází (ne nutně ortonormální) tvořenou sloupci matice $\mathbf{A}$. Zdůrazněme, že projektor (5.7) vyjde stejný pro libovolnou bázi podprostoru X (viz Cvičení 5.11). Pokud je báze ortonormální, je $\mathbf{A}^T\mathbf{A} = \mathbf{I}$ a (5.7) se redukuje na (4.19).

Projekce na $X^\perp$ má přirozenou úlohu v problému (5.2): hodnota jeho minima je $\|\mathbf{b}-\mathbf{A}\mathbf{x}\|^2 = \|\mathbf{b}-\mathbf{P}\mathbf{b}\|^2 = \|(\mathbf{I}-\mathbf{P})\mathbf{b}\|^2$.

5.1.2 Řešení pomocí QR rozkladu

I když má matice $\mathbf{A}$ lineárně nezávislé sloupce, řešení pomocí pseudoinverze (5.6) nemusí být vhodné pro numerické výpočty, kdy nezbytně používáme aritmetiku s konečnou přesností.

Příklad 5.2. Řešme soustavu $\mathbf{A}\mathbf{x} = \mathbf{b}$ 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 $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$, 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 $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$ 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 $\mathbf{R}^T\mathbf{Q}^T\mathbf{Q}\mathbf{R}\mathbf{x} = \mathbf{R}^T\mathbf{Q}^T\mathbf{b}$ neboli $\mathbf{R}^T\mathbf{R}\mathbf{x} = \mathbf{R}^T\mathbf{Q}^T\mathbf{b}$. Jestliže $\mathbf{A}$ má lineárně nezávislé sloupce, matice $\mathbf{R}$ je regulární. Vynásobením maticí $\mathbf{R}^{-T}$ zleva (což je tedy ekvivalentní úprava) máme

$$\mathbf{R}\mathbf{x} = \mathbf{Q}^T\mathbf{b}. \tag{5.8}$$

Zdůraněme, že pokud $\mathbf{A}$ není čtvercová, pak soustava (5.8) není ekvivalentní původní soustavě $\mathbf{A}\mathbf{x} = \mathbf{b}$.

Jestliže sloupce $\mathbf{A}$ jsou lineárně závislé, postup je trochu složitější, ale také stojí na QR rozkladu. V Matlabu je řešení nehomogenní lineární soustavy 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`!

5.1.3 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), $i = 1, \dots, m$, 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

$$\min_{\boldsymbol{\theta} \in \mathbb{R}^n} \sum_{i=1}^m (y_i - f(x_i, \boldsymbol{\theta}))^2. \tag{5.9}$$

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

$$f(x, \boldsymbol{\theta}) = \theta_1 \varphi_1(x) + \cdots + \theta_n \varphi_n(x) = \boldsymbol{\varphi}(x)^T \boldsymbol{\theta} \tag{5.10}$$

nějakých daných funkcí[3] $\varphi_j \colon X \to \mathbb{R}$. Pak

$$\sum_{i=1}^m (y_i - f(x_i, \boldsymbol{\theta}))^2 = \|\mathbf{y} - \mathbf{A}\boldsymbol{\theta}\|^2,$$

kde $\mathbf{y} = (y_1, \dots, y_m)$ a prvky matice $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{m \times n}$ jsou $a_{ij} = \varphi_j(x_i)$ (odvoďte!). Tedy vyjádřili jsme úlohu (5.9) ve tvaru (5.2).

Příklad 5.3. Nejjednodušší případ je pro $n = 1$ a konstantní funkci $\varphi_i(x) = 1$. Funkce (5.10) je tedy $f(x, \theta) = \theta$. Úloha (5.9) 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 $y_1, \dots, y_m$. □

[3] Funkce φ_j se často nazývají *bázové funkce* (pokud jsou ovšem lineárně nezávislé).

Příklad 5.4. *Proložení bodů polynomem*[4]. 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 $T = \mathbb{R}^d$ a bázové funkce jsou monomy proměnných $x_1, \ldots, x_d$ (viz §6) až do nějakého stupně. □

5.1.4 Statistické odůvodnění kritéria nejmenších čtverců

Možná se ptáte, proč se má nalezení přibližného rešení přeurčené soustavy formulovat zrovna jako (5.2). Uvedeme statistický důvod, odkud se kritérium nejmenšího součtu čtverců vzalo.

Odhadujme skryté parametry $\mathbf{x}$ nějakého systému z měření $\mathbf{b}$ na systému. Budiž vázány známou lineární závislostí $\mathbf{b} = \mathbf{A}\mathbf{x}$. Měření jsou zatížena chybami, které jsou způsobeny šumem senzorů, nepřesnostmi měření, nedokonalou znalostí modelu, apod. Tedy

$$\mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{r}, \tag{5.11}$$

kde $\mathbf{r} = (r_1, \ldots, r_m)$ jsou náhodné proměnné modelující chyby měření $\mathbf{b} = (b_1, \ldots, b_m)$. Metoda nejmenších čtverců říká, že máme minimalizovat $\|\mathbf{r}\|_2^2 = \sum_{i=1}^m r_i^2$, ale neříká proč.

Důvod odvodíme statistickou úvahou. Metoda činí dva předpoklady:

- Náhodné proměnné r_i mají normální (neboli Gaussovo) rozdělení s nulovou střední hodnotou a směrodatnou odchylkou σ, s hustotou pravděpodobnosti

  $$p(r_i) = c\, \mathrm{e}^{-r_i^2/(2\sigma^2)},$$

  kde $c = \left(\sigma\sqrt{2\pi}\right)^{-1}$ je normalizační konstanta.
- Náhodné proměnné $r_1, \ldots, r_m$ jsou na sobě nezávislé. Tedy sdružená hustota pravděpodobnosti je rovna součinu

  $$p(\mathbf{r}) = p(r_1, \ldots, r_m) = \prod_{i=1}^m p(r_i) = \prod_{i=1}^m c\, \mathrm{e}^{-r_i^2/(2\sigma^2)}. \tag{5.12}$$

Dále použijeme *princip maxima věrohodnosti*. Ten říká, že parametry $\mathbf{x}$ se mají najít tak, aby $p(\mathbf{r}) = p(\mathbf{b} - \mathbf{A}\mathbf{x})$ bylo maximální. Je pohodlnější minimalizovat záporný logaritmus

$$-\log p(r_1, \ldots, r_m) = -\sum_{i=1}^m \log p(r_i) = \sum_{i=1}^m \left(\frac{r_i^2}{2\sigma^2} - \log c \right).$$

Jelikož σ je konstanta, je to totéž jako minimalizovat $\sum_i r_i^2$.

[4] Nedejte se zmást tím, že polynom není lineární funkce a přesto jde o lineární regresi. Důležité je, že regresní funkce (5.10) je lineární v parametrech $\boldsymbol{\theta}$.

5.1.5 Vícekriteriální nejmenší čtverce, regularizace

V některých úlohách se hodí minimalizovat více kritérií tvaru $\|\mathbf{Ax}-\mathbf{b}\|$ 'současně'. K tomu se dá přistoupit tak, že minimalizujeme (nezáporně) vážený součet kritérií[5], tedy funkci[6]

$$\mu_1\|\mathbf{A}_1\mathbf{x}-\mathbf{b}_1\|^2+\cdots+\mu_k\|\mathbf{A}_k\mathbf{x}-\mathbf{b}_k\|^2 \tag{5.13}$$

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 (5.13) je roven (viz Cvičení 5.18)

$$\left\|\begin{bmatrix}\sqrt{\mu_1}(\mathbf{A}_1\mathbf{x}-\mathbf{b}_1)\\ \vdots\\ \sqrt{\mu_k}(\mathbf{A}_k\mathbf{x}-\mathbf{b}_k)\end{bmatrix}\right\|^2=\left\|\begin{bmatrix}\sqrt{\mu_1}\mathbf{A}_1\\ \vdots\\ \sqrt{\mu_k}\mathbf{A}_k\end{bmatrix}\mathbf{x}-\begin{bmatrix}\sqrt{\mu_1}\mathbf{b}_1\\ \vdots\\ \sqrt{\mu_k}\mathbf{b}_k\end{bmatrix}\right\|^2=\|\mathbf{A}'\mathbf{x}-\mathbf{b}'\|^2, \tag{5.14}$$

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!)

$$\mathbf{x}=(\mathbf{A}'^T\mathbf{A}')^{-1}\mathbf{A}'^T\mathbf{b}'=(\mu_1\mathbf{A}_1^T\mathbf{A}_1+\cdots+\mu_k\mathbf{A}_k^T\mathbf{A}_k)^{-1}(\mu_1\mathbf{A}_1^T\mathbf{b}_1+\cdots+\mu_k\mathbf{A}_k^T\mathbf{b}_k). \tag{5.15}$$

Speciálně, někdy chceme přibližně řešit soustavu $\mathbf{Ax}=\mathbf{b}$ a zároveň chceme, aby norma řešení $\mathbf{x}$ nebyla moc velká. To lze formulovat jako

$$\min_{\mathbf{x}\in\mathbb{R}^n}\big(\|\mathbf{Ax}-\mathbf{b}\|^2+\mu\|\mathbf{x}\|^2\big). \tag{5.16}$$

pro zvolenou váhu $\mu>0$. Přidání členu $\mu\|\mathbf{x}\|^2$ se říká (Tichonovova) **regularizace** úlohy (5.2). Dosazením do vzorečku (5.15) je optimální řešení rovno $\mathbf{x}=\mathbf{A}_\mu^+\mathbf{b}$ kde

$$\mathbf{A}_\mu^+=(\mathbf{A}^T\mathbf{A}+\mu\mathbf{I})^{-1}\mathbf{A}^T \tag{5.17}$$

je 'regularizovaná pseudoinverze'[7] 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 5.19), tedy $\mathbf{A}_\mu^+$ je vždy definována. Viz také Cvičení 5.20.

5.2 Řešení s nejmenší normou

Předpokládejme nyní, že soustava (5.1) je nedourčená, neboli má nekonečně mnoho řešení. Je často užitečné z této množiny řešení vybrat jediné podle nějakého kritéria. Přirozeným kritériem je minimalizovat eukleidovskou normu (tedy vzdálenost od počátku) řešení, což vede na úlohu

$$\min\{\,\|\mathbf{x}\|^2 \mid \mathbf{x}\in\mathbb{R}^n,\ \mathbf{Ax}=\mathbf{b}\,\}. \tag{5.18}$$

Místo normy $\|\mathbf{x}\|$ opět minimalizujeme její čtverec. Tato úloha je známa jako řešení nehomogenní lineární soustavy **s nejmenší normou** (*least norm solution*). Podotkněme, že někdy je vhodné použít jiná kritéria než nejmenší eukleidovskou normu, viz např. Cvičení **??**.

[5]Minimalizací více kritérií současně se zabývá obor *vícekriteriální optimalizace (multiobjective optimization)*.

[6]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 (5.13) by pak bylo kratší.

[7]Symbol $\mathbf{A}_\mu^+$ zde neoznačuje pseudoinverzi nějaké matice $\mathbf{A}_\mu$, taková matice $\mathbf{A}_\mu$ totiž neexistuje.

Příklad 5.5. Soustava dvou rovnic o třech neznámých

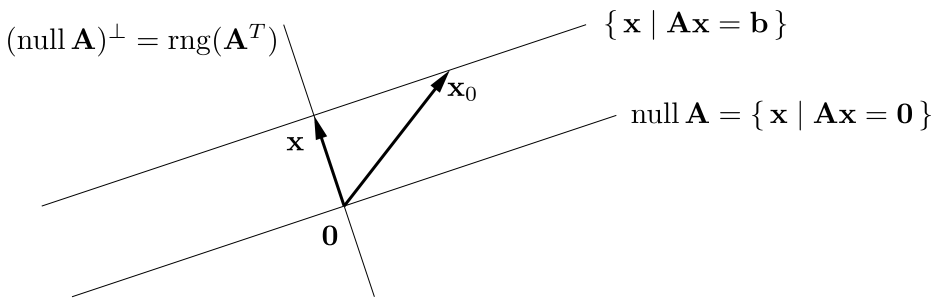
$$\begin{aligned} x + 2y + z &= 1 \\ -x + y + 2z &= 2 \end{aligned}$$

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

$$\{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} = \mathbf{b} \,\} = \operatorname{null} \mathbf{A} + \mathbf{x}_0, \tag{5.19}$$

kde $\mathbf{x}_0$ je libovolné (partikulární) řešení soustavy, tedy $\mathbf{A}\mathbf{x}_0 = \mathbf{b}$. Množina (5.19) 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 $\mathbf{x} \perp \operatorname{null} \mathbf{A}$, neboli $\mathbf{x} \in (\operatorname{null} \mathbf{A})^\perp = \operatorname{rng}(\mathbf{A}^T)$, kde poslední rovnost je (4.7b). Neboli musí existovat vektor $\mathbf{y} \in \mathbb{R}^m$ tak, že $\mathbf{x} = \mathbf{A}^T\mathbf{y}$. Pro vyřešení úlohy (5.18) tedy musíme vyřešit soustavu rovnic

$$\mathbf{A}^T\mathbf{y} = \mathbf{x}, \tag{5.20a}$$
$$\mathbf{A}\mathbf{x} = \mathbf{b}. \tag{5.20b}$$

To je soustava $m + n$ rovnic o $m + n$ neznámých $(\mathbf{x}, \mathbf{y})$.

Vyřešme tuto soustavu. Dosazením $\mathbf{x}$ do druhé rovnice obdržíme $\mathbf{A}\mathbf{A}^T\mathbf{y} = \mathbf{b}$. Tato soustava má vždy řešení, protože $\mathbf{b} \in \operatorname{rng} \mathbf{A} = \operatorname{rng}(\mathbf{A}\mathbf{A}^T)$. Předpokládejme, že matice $\mathbf{A}\mathbf{A}^T$ má plnou hodnost, což dle (5.5) nastane právě když $\mathbf{A}$ má lineárně nezávislé řádky. Potom $\mathbf{y} = (\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b}$. Dosazením do první rovnice dostaneme $\mathbf{x} = \mathbf{A}^+\mathbf{b}$, kde

$$\mathbf{A}^+ = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1} \tag{5.21}$$

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 jinou úvahou. 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 (5.7), zde ovšem promítáme na $\operatorname{rng}(\mathbf{A}^T)$ a tedy musíme vzorec použít s $\mathbf{A}^T$ místo s $\mathbf{A}$. Tedy

$$\mathbf{x} = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{x}_0 = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b} = \mathbf{A}^+\mathbf{b}. \tag{5.22}$$

Vzorce (5.6) a (5.21) dohromady definují pseudoinverzi libovolné matice (čtvercové, úzké nebo široké) s plnou hodností (tedy $\operatorname{rank} \mathbf{A} = \min\{m, n\}$).

5.3 ($\star$) 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 (5.3). Co když je ale sama soustava (5.3) nedourčená? Pak můžeme hledat její řešení s nejmenší normou, tj. řešit úlohu

$$\min\{\, \|\mathbf{x}\|^2 \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}\,\}. \tag{5.23}$$

Protože úloha (5.18) má pro každou matici $\mathbf{A}$ právě jedno optimální řešení, má i úloha (5.23) pro každou $\mathbf{A}$ právě jedno optimální řešení. Toto řešení, $\mathbf{x}^*$, má tedy následující vlastnosti:

- Pokud soustava $\mathbf{A}\mathbf{x} = \mathbf{b}$ má jediné řešení, $\mathbf{x}^*$ je toto řešení.
- Pokud soustava $\mathbf{A}\mathbf{x} = \mathbf{b}$ nemá řešení, $\mathbf{x}^*$ je její přibližné řešení ve smyslu nejmenších čtverců, tj. řešení problému (5.2). Pokud ovšem problém (5.2) má více než jedno (tedy nekonečně mnoho) řešení, $\mathbf{x}^*$ je řešení problému (5.2) s nejmenší normou.
- Pokud soustava $\mathbf{A}\mathbf{x} = \mathbf{b}$ má nekonečně mnoho řešení, $\mathbf{x}^*$ je řešení této soustavy s nejmenší normou, tj. řešení problému (5.18).

Lze ukázat, že řešení úlohy (5.23) lze opět psát jako $\mathbf{x} = \mathbf{A}^+\mathbf{b}$ kde matici $\mathbf{A}^+$ nazýváme **pseudoinverze** (přesněji *Moore-Penroseova pseudoinverze*) matice $\mathbf{A}$. Když $\mathbf{A}$ má lineárně nezávislé sloupce, pseudoinverze je rovna (5.6). Když $\mathbf{A}$ má lineárně nezávislé řádky, pseudoinverze je rovna (5.21). Když ovšem $\mathbf{A}$ nemá plnou hodnost, pseudoinverzi je nutno počítat jinak. Elegantně se to udělá pomocí SVD, což neuvádíme.

Je několik jiných způsobů, jak definovat pseudoinverzi obecné matice. Zmíníme ještě jeden zajímavý. Všiměte si, že úloha (5.16) je 'něco mezi' úlohami (5.2) a (5.18). Neformálně, (5.18) je minimalizace $\|\mathbf{x}\|^2 + \mu\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2$ pro velmi velké μ. To je ale totéž jako (5.16) pro velmi malé (kladné) μ. Lze ukázat, že pseudoinverze obecné matice je rovna

$$\mathbf{A}^+ = \lim_{\mu\to 0+} \mathbf{A}^+_\mu. \tag{5.24}$$

5.4 Cvičení

5.1. Máme soustavu $\mathbf{A}\mathbf{x} = \mathbf{b}$, kde $\mathbf{A} \in \mathbb{R}^{m\times n}$ a $\mathbf{b} \neq \mathbf{0}$. Jsou tyto výroky pravdivé? Odpovědi dokažte.

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 1.14.

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 1.11.

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 k přímkám 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

$$f(\mathbf{x}) = \sum_{i=1}^{m} w_i \left(\sum_{j=1}^{n} a_{ij} x_j - b_i \right)^2$$

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. Máme vektory $\mathbf{u} = (2, 1, -3)$ a $\mathbf{v} = (1, -1, 1)$. Najdi 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) $(\operatorname{span}\{\mathbf{u}, \mathbf{v}\})^\perp$.

5.6. Nechť $X = \operatorname{span}\{\, (-\frac{3}{5}, 0, \frac{4}{5}, 0),\ (0, 0, 0, 1),\ (\frac{4}{5}, 0, \frac{3}{5}, 0) \,\}$. Najdi projektory na podprostor X a podprostor $X^\perp$.

5.7. Máme $\mathbf{A} = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 4 & 1 \\ 1 & 2 & 0 \end{bmatrix}$. Najdi ortogonální projekci vektoru $(1, 1, 1)$ na podprostory (a) $\operatorname{rng} \mathbf{A}$, (b) $\operatorname{null} \mathbf{A}$, (c) $\operatorname{rng}(\mathbf{A}^T)$, (d) $\operatorname{null}(\mathbf{A}^T)$.

5.8. 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 (5.7) a podprostor $X = \operatorname{rng} \mathbf{A}$? Jaký je geometrický význam této situace?

5.9. Dokažte následující vlastnosti pseudoinverze ze vztahů (5.6) a (5.21) pro libovolné (úzké, široké nebo čtvercové) matice plné hodnosti:

a) $\mathbf{A}^+ = \mathbf{A}^{-1}$ 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) $(\mathbf{A}^T\mathbf{A})^+ = \mathbf{A}^+(\mathbf{A}^T)^+$, $(\mathbf{A}\mathbf{A}^T)^+ = (\mathbf{A}^T)^+\mathbf{A}^+$

5.10. Spočítejte pseudoinverzi nenulového skaláru (tj. matice s jedním řádkem a jedním sloupcem), nenulového sloupcového vektoru (tj. matice s jednim sloupcem) a nenulového řádkového vektoru (tj. matice s jedním řádkem).

5.11. Uvažujme projektor (5.7). Báze podprostoru X, na který promítáme, je tvořena sloupci matice $\mathbf{A}$. Projektor $\mathbf{P}$ se nesmí změnit, vezmeme-li jinou bázi podprostoru. Různé báze podprostoru jsou dány sloupci matice $\mathbf{AC}$ pro různé regulární matice $\mathbf{C} \in \mathbb{R}^{n\times n}$ (tedy $\mathbf{C}$ je matice přechodu k jiné bázi), viz Cvičení 3.12.

5.12. Jaké bude řešení normálních rovnic (5.3) v případě, že $\mathbf{A}$ má lineárně nezávislé sloupce a $\mathbf{b} \perp \operatorname{rng} \mathbf{A}$? Vyřešte geometrickou úvahou (vzpomeňte si na orgogonální projekci a koukejte na obrázek v §5.1!) a pak zkuste dokázat algebraicky.

5.13. Dokažte, že pokud $\mathbf{A}^T\mathbf{A} = \mathbf{B}^T\mathbf{B}$, pak existuje $\mathbf{C}$ tak, že $\mathbf{A} = \mathbf{CB}$.

5.14. Najděte co nejjednodušší vzorec pro

a) vzdálenost počátku $\mathbf{0}$ od nadroviny $\{\, \mathbf{x} \mid \mathbf{a}^T\mathbf{x} = b \,\}$,
b) vzdálenost počátku $\mathbf{0}$ od afinního podprostoru $\{\, \mathbf{x} \mid \mathbf{A}\mathbf{x} = \mathbf{b} \,\}$ (kde $\mathbf{A}$ má lineárně nezávislé řádky),
c) vzdálenost bodu $\mathbf{x}$ od nadroviny $\{\, \mathbf{x} \mid \mathbf{a}^T\mathbf{x} = b \,\}$.

5.15. Neformálně je jasné, že nedourčená soustava bude mít vždy jen jedno řešení s nejmenší normou. Dokažte toto formálně, tj. dokažte, že jestliže soustava $\mathbf{A}\mathbf{x} = \mathbf{b}$ má alespoň jedno řešení pak soustava (5.20) má právě jedno řešení.

5.16. 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 $(\operatorname{rng} \mathbf{A})^\perp = \operatorname{null} \mathbf{A}$.

5.17. Ukázali jsme (viz §4.6), že ortogonální projektor (4.19) na podprostor reprezentovaný ortonormální bází splňuje $\mathbf{P}^2 = \mathbf{P} = \mathbf{P}^T$. Dokažte tyto rovnosti i pro ortogonální projektor (5.7) na podprostor reprezentovaný obecnou bází.

5.18. Pro vektory $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ ukažte, že $\left\| \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \right\|^2 = \|\mathbf{a}\|^2 + \|\mathbf{b}\|^2$.

5.19. Dokažte, že matice $\mathbf{A}^T\mathbf{A} + \mu\mathbf{I}$ je regulární pro každou matici $\mathbf{A}$ a každé $\mu > 0$.

5.20. Uvažujme 'regularizovanou pseudoinverzi' (5.24). Dokažte, že pro každou matici $\mathbf{A}$ a každé $\mu > 0$ platí $(\mathbf{A}^T\mathbf{A} + \mu\mathbf{I})^{-1}\mathbf{A}^T = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T + \mu\mathbf{I})^{-1}$. Dumejte nad významem této rovnosti.

5.21. ($\star$) Dokažte, že pro každé matice $\mathbf{A}, \mathbf{B}$ takové, že matice $\begin{bmatrix}\mathbf{A} & \mathbf{B}\end{bmatrix}$ je čtvercová regulární a $\mathbf{A}^T\mathbf{B} = \mathbf{0}$, platí $\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T + \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T = \mathbf{I}$. Vysvětlete rozdíl oproti (4.24).

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) Zbývá napsat funkci (1.14) ve tvaru $f(\mathbf{u}) = \|\mathbf{P}\mathbf{u} - \mathbf{q}\|^2$. To je snadné (srov. s (5.14)): máme $\sum_{i=1}^m \|\mathbf{x} - \mathbf{a}_i\|^2 = \left\|\begin{bmatrix}\mathbf{x} - \mathbf{a}_1\\ \vdots \\ \mathbf{x} - \mathbf{a}_m\end{bmatrix}\right\|^2 = \left\|\begin{bmatrix}\mathbf{I}_n\\ \vdots\\ \mathbf{I}_n\end{bmatrix}\mathbf{x} - \begin{bmatrix}\mathbf{a}_1\\ \vdots\\ \mathbf{a}_m\end{bmatrix}\right\|^2$, tedy $\mathbf{u} = \mathbf{x}$, $\mathbf{P} = \begin{bmatrix}\mathbf{I}_n\\ \vdots\\ \mathbf{I}_n\end{bmatrix}$, $\mathbf{q} = \begin{bmatrix}\mathbf{a}_1\\ \vdots\\ \mathbf{a}_m\end{bmatrix}$

5.3.c) Zbývá napsat funkci (1.13) ve tvaru $f(\mathbf{u}) = \|\mathbf{P}\mathbf{u} - \mathbf{q}\|^2$. 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 $\mathbf{y}, t_1, \ldots, t_m$.

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. $f(\mathbf{x}) = (\mathbf{A}\mathbf{x} - \mathbf{b})^T\mathbf{W}(\mathbf{A}\mathbf{x} - \mathbf{b}) = \|\mathbf{W}^{1/2}\mathbf{A}\mathbf{x} - \mathbf{W}^{1/2}\mathbf{b}\|^2$.

5.5. (a) $(2, 1, -3)/14$, (b) $(26, -1, 17)/14$, (c) $(62, -35, 17)/38$, (d) $(14, 35, 21)/38$

5.6. Nejsou náhodou vektory ortonormální?

5.6. Projektor na X je $\mathbf{P} = \operatorname{diag}(1, 0, 1, 1)$. Projektor na $X^\perp$ je $\mathbf{P} = \operatorname{diag}(0, 1, 0, 0)$.

5.7. (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.

5.8. $\mathbf{A}$ je regulární, tedy $X = \mathbb{R}^m$. Projektor je identita.

5.9.b) Když $\mathbf{A}$ má l.n. sloupce, dle (5.6) 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 (5.21) 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.

5.11. $\tilde{\mathbf{A}}(\tilde{\mathbf{A}}^T\tilde{\mathbf{A}})^{-1}\tilde{\mathbf{A}}^T = \mathbf{A}\mathbf{C}(\mathbf{C}^T\mathbf{A}^T\mathbf{A}\mathbf{C})^{-1}\mathbf{C}^T\mathbf{A}^T = \mathbf{A}\mathbf{C}\mathbf{C}^{-1}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{C}^{-T}\mathbf{C}^T\mathbf{A}^T = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$

5.12. $\mathbf{x} = \mathbf{0}$

5.13. Dle (5.4a) je $\operatorname{rng}(\mathbf{A}^T) = \operatorname{rng}(\mathbf{B}^T)$. Tedy každý řádek $\mathbf{A}$ je lineární kombinací řádků $\mathbf{B}$. Dle Cvičení 3.14 to jde napsat jako $\mathbf{A} = \mathbf{C}\mathbf{B}$ pro nějaké $\mathbf{C}$.

5.14.a) $|b|/\|\mathbf{a}\|$

5.14.b) Čtverec vzdálenosti je roven optimální hodnotě úlohy (5.18), tedy $(\mathbf{A}^+\mathbf{b})^T(\mathbf{A}^+\mathbf{b}) = \mathbf{b}^T(\mathbf{A}\mathbf{A}^T)^{-T}\mathbf{A}\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b} = \mathbf{b}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b}$.

5.14.c) $|\mathbf{a}^T\mathbf{x} - b|/\|\mathbf{a}\|$

5.16. Dle (4.7a) a (5.4b) je $(\operatorname{rng}\mathbf{A})^\perp = \operatorname{null}(\mathbf{A}^T) = \operatorname{null}(\mathbf{A}\mathbf{A}^T) = \operatorname{null}(\mathbf{A}^T\mathbf{A}) = \operatorname{null}\mathbf{A}$.

5.17. Ihned plyne z toho, že každý podprostor má ortonormální bázi (což jsme řekli v §4.5), tedy stačí rovnosti dokázat pro (4.19). Ale můžeme také dokázat přímým dosazením:
$\mathbf{P}^2 = \mathbf{P}\mathbf{P} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$,
$\mathbf{P}^T = [\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T]^T = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-T}\mathbf{A}^T = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$.

5.19. 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 (5.5) má $\mathbf{B}^T\mathbf{B}$ plnou hodnost, tedy je regulární.

5.20. Rovnici vynásobte zleva maticí $\mathbf{A}^T\mathbf{A} + \mu\mathbf{I}$ a zprava maticí $\mathbf{A}\mathbf{A}^T + \mu\mathbf{I}$ a pak roznásobte závorky.

5.21. $\mathbf{I} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{B} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{A}^T \\ \mathbf{B}^T \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{A}^T \\ \mathbf{B}^T \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \end{bmatrix} \left(\begin{bmatrix} \mathbf{A}^T \\ \mathbf{B}^T \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{B} \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{A}^T \\ \mathbf{B}^T \end{bmatrix}$
$= \begin{bmatrix} \mathbf{A} & \mathbf{B} \end{bmatrix} \begin{bmatrix} \mathbf{A}^T\mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}^T\mathbf{B} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{A}^T \\ \mathbf{B}^T \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \end{bmatrix} \begin{bmatrix} (\mathbf{A}^T\mathbf{A})^{-1} & \mathbf{0} \\ \mathbf{0} & (\mathbf{B}^T\mathbf{B})^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{A}^T \\ \mathbf{B}^T \end{bmatrix}$
$= \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T + \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T$.

Význam: $\begin{bmatrix} \mathbf{A} & \mathbf{B} \end{bmatrix}$ regulární a $\mathbf{A}^T\mathbf{B} = \mathbf{0}$ implikuje $\operatorname{rng}\mathbf{B} = (\operatorname{rng}\mathbf{A})^\perp$. To souhlasí s tím, že $\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ je ort. projektor na $\operatorname{rng}\mathbf{A}$ a $\mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T$ je ort. projektor na $\operatorname{rng}\mathbf{B} = (\operatorname{rng}\mathbf{A})^\perp$.

Kapitola 6

Spektrální rozklad a kvadratické funkce

Ze základní školy znáte polynomy jedné proměnné, co jsou ale polynomy více proměnných? **Monom** k-tého stupně n proměnných je výraz

$$x_1^{k_1} \cdots x_n^{k_n},$$

kde $k_1, \dots, k_n \in \{0, \dots, k\}$ splňují $k_1 + \cdots + k_n = k$. **Polynom** 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. $f(x, y) = x^2 y - 5y^3$ je homogenní stupně tři.

Vidíme, že afinní funkce (3.23) je jen jiný název pro polynom prvního stupně a lineární funkce (3.6) (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 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

$$\mathbf{A}\mathbf{v} = \lambda \mathbf{v}. \tag{6.2}$$

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.

Rovnici (6.2) lze přepsat jako

$$(\mathbf{A} - \lambda \mathbf{I})\mathbf{v} = \mathbf{0}. \tag{6.3}$$

[1] Tohle samozřejmě platí jen pro funkce na vektorovém prostoru $\mathbb{R}^n$. Pokud uvažujeme funkce na abstraktním (tedy definovaným axiomy) vektorovém prostoru, který znáte z lineární algebry, pak např. lineární forma není polynom jednoduše proto, že 'polynom' na tomto prostoru není definován.

[2] Názvosloví není zcela konzistentní, což je opět dáno tím, že některá jména pocházejí z lineární algebry a některá z matematické analýzy.

To je soustava homogenních lineárních rovnic pro $\mathbf{v}$, která má netriviální řešení právě tehdy, když matice $\mathbf{A} - \lambda\mathbf{I}$ je singulární. Tedy vlastní čísla splňují

$$p_{\mathbf{A}}(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}) = 0, \tag{6.4}$$

kde funkce $p_A\colon \mathbb{R} \to \mathbb{R}$ se nazývá **charakteristický polynom** matice $\mathbf{A}$.

Vlastní vektory příslušné vlastnímu číslu λ pak spočítáme ze soustavy (6.3). Vlastní vektor není svým vlastním číslem určen jednoznačně, vlastní vektory příslušné vlastnímu číslu λ tvoří celý podprostor $\operatorname{null}(\mathbf{A} - \lambda\mathbf{I})$ (kromě počátku $\mathbf{0}$). Speciálně, velikost vlastních vektorů nehraje roli a je proto zvykem je normalizovat, $\|\mathbf{v}\| = 1$.

Příklad 6.1. Vlastní čísla matice $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ jsou řešeními rovnice

$$\det(\mathbf{A} - \lambda\mathbf{I}) = \det \begin{bmatrix} 1-\lambda & 2 \\ 3 & 4-\lambda \end{bmatrix} = (1-\lambda)(4-\lambda) - 3 \cdot 2 = \lambda^2 - 5\lambda - 2 = 0.$$

Tato kvadratická rovnice má dva kořeny $\lambda = (5 \pm \sqrt{33})/2$, což jsou tedy vlastní čísla matice $\mathbf{A}$. Vlastní vektory příslušné každému λ najdeme řešením homogenní lineární soustavy

$$\begin{bmatrix} 1-\lambda & 2 \\ 3 & 4-\lambda \end{bmatrix} \mathbf{v} = \mathbf{0}.$$

□

Z definice determinantu (2.6) plyne (promyslete!), že charakteristický polynom má stupeň n. Podle *základní věty algebry* má tedy právě n komplexních kořenů, počítáme-li k-násobný kořen k-krát. Označíme-li kořeny $\lambda_1, \ldots, \lambda_n$, platí

$$p_{\mathbf{A}}(\lambda) = \prod_{i=1}^{n} (\lambda_i - \lambda).$$

V tomto smyslu má matice právě n vlastních čísel, z nichž některá mohou být stejná kvůli násobnosti. Tomuto seznamu vlastních čísel se říká **spektrum** matice.

Nechť $\lambda_1, \ldots, \lambda_n$ jsou vlastní čísla matice a $\mathbf{v}_1, \ldots, \mathbf{v}_n$ jsou jim příslušné vlastní vektory. Rovnice (6.2) lze pro ně napsat jako jedinou maticovou rovnici (rozmyslete!)

$$\mathbf{AV} = \mathbf{V\Lambda}, \tag{6.5}$$

kde diagonální matice $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^{n\times n}$ má na diagonále vlastní čísla a sloupce čtvercové matice $\mathbf{V} = [\mathbf{v}_1 \ \cdots \ \mathbf{v}_n] \in \mathbb{R}^{n\times n}$ jsou vlastní vektory.

Vlastní vektory mohou být lineárně závislé. Otázka, kdy se tak stane a co to znamená, není jednoduchá a podrobně ji zde diskutovat nepotřebujeme. Řekneme jen, že existuje dobrý důvod vlastní vektory vybrat tak, aby hodnost matice $\mathbf{V}$ byla co možná největší.

Jak se počítají vlastní čísla a vektory? Charakteristický polynom je hlavně teoeretický nástroj a přímé hledání jeho kořenů není vhodné pro numerický výpočet[3]. Pro větší matice se používají numerické iterační algoritmy, přičemž pro matice různého typu jsou vhodné různé algoritmy. Zásadní rozdíl oproti např. QR rozkladu je v tom, že vlastní čísla obecně nejdou přesně spočítat konečným počtem elementárních operací (+, −, ×, /), tedy vždy je spočítáme jen s konečnou přesností. Matlabská funkce `[V,D]=eig(A)` spočítá matice $\mathbf{V}$ a $\mathbf{\Lambda}$ splňující (6.5).

[3]Naopak, hledání kořenů libovolného polynomu lze převést na hledání vlastních čísel matice, která se nazývá *doprovodná matice (companion matrix)* polynomu.

6.1.1 Spektrální rozklad

Pokud je $\mathbf{V}$ regulární (tj. matice $\mathbf{A}$ má n lineárně nezávislých vlastních vektorů), je invertovatelná a (6.5) lze psát jako

$$\mathbf{A} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1}. \tag{6.6}$$

Vztahu (6.6) se pak říká **rozklad matice podle vlastních čísel** nebo **spektrální rozklad**. V tom případě je matice $\mathbf{A}$ podobná diagonální matici (neboli **diagonalizovatelná**), protože z (6.6) plyne $\mathbf{V}^{-1}\mathbf{A}\mathbf{V} = \boldsymbol{\Lambda}$.

Věta 6.1. *Nechť $\mathbf{A} \in \mathbb{R}^{n\times n}$. Následující dvě tvrzení jsou ekvivalentní:*

- *Matice $\mathbf{A}$ je symetrická.*
- *Všechna vlastní čísla matice $\mathbf{A}$ jsou reálná a $\mathbf{A}$ má n vlastních vektorů které tvoří ortonormální množinu.*

Důkaz této věty neuvádíme. Větě 6.1 se někdy říká **spektrální věta**. Podle ní pro každou symetrickou $\mathbf{A}$ je v (6.5) matice $\boldsymbol{\Lambda}$ reálná a $\mathbf{V}$ může být zvolena ortogonální ($\mathbf{V}^{-1} = \mathbf{V}^T$). Tedy

$$\mathbf{A} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T = \lambda_1\mathbf{v}_1\mathbf{v}_1^T + \cdots + \lambda_n\mathbf{v}_n\mathbf{v}_n^T. \tag{6.7}$$

Zároveň jsme vpravo uvedli i druhou formu rozkladu jako součet dyád (viz §2.6) (přesvědčte se, že druhá rovnost v (6.7) platí!).

Všimněme si, že (např. dle Věty 3.6, použité dvakrát na výraz $\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T$) máme

$$\operatorname{rank}\boldsymbol{\Lambda} = \operatorname{rank}\mathbf{A}. \tag{6.8}$$

Ale hodnost diagonální matice Λ je jednoduše počet jejích nenulových prvků, tedy počet nenulových vlastních čísel. Když $\operatorname{rank}\mathbf{A} < n$, některá vlastní čísla jsou nulová a můžeme vynechat jim odpovídající sloupce+řádky matice $\boldsymbol{\Lambda}$ a sloupce matice $\mathbf{V}$. To znamená vynechání nulových sčítanců v sumě dyád v (6.7).

Příklad 6.2. Zde je spektrální rozklad matice 3×3 hodnosti 2:

$$\begin{aligned}
\begin{bmatrix} 2 & 3 & -2 \\ 3 & 2 & 2 \\ -2 & 2 & -8 \end{bmatrix} &= \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{18} & -2/3 \\ 1/\sqrt{2} & -1/\sqrt{18} & 2/3 \\ 0 & 4/\sqrt{18} & 1/3 \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 \\ 0 & -9 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 1/\sqrt{18} & -1/\sqrt{18} & 4/\sqrt{18} \\ -2/3 & 2/3 & 1/3 \end{bmatrix} \\
&= \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{18} \\ 1/\sqrt{2} & -1/\sqrt{18} \\ 0 & 4/\sqrt{18} \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & -9 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 1/\sqrt{18} & -1/\sqrt{18} & 4/\sqrt{18} \end{bmatrix} \\
&= 5 \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \end{bmatrix} - 9 \begin{bmatrix} 1/\sqrt{18} \\ -1/\sqrt{18} \\ 4/\sqrt{18} \end{bmatrix} \begin{bmatrix} 1/\sqrt{18} & -1/\sqrt{18} & 4/\sqrt{18} \end{bmatrix} \qquad \square
\end{aligned}$$

Vlastní čísla a vektory jsou rozsáhlé téma, které jsme zde zdaleka nevyčerpali. To ale není ani třeba, protože dále budeme pořebovat jen spektrální rozklad symetrické matice.

6.2 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

$$f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} = \sum_{j=1}^n \sum_{i=1}^n a_{ij}x_ix_j \tag{6.9}$$

pro nějakou matici $\mathbf{A} \in \mathbb{R}^{n\times n}$. Protože $x_ix_j = x_jx_i$ (násobení čísel je komutativní), máme

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \sum_{j=1}^n \sum_{i=1}^n a_{ij}x_ix_j = \frac{1}{2}\sum_{j=1}^n \sum_{i=1}^n (a_{ij}+a_{ji})x_ix_j = \frac{1}{2}\mathbf{x}^T(\mathbf{A}+\mathbf{A}^T)\mathbf{x}. \tag{6.10}$$

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 $\frac{1}{2}(\mathbf{A}+\mathbf{A}^T) = \mathbf{A}$.

Abychom si procvičili maticovou algebru, dokážeme rovnost (6.10) i jinak. Každou čtvercovou matici lze jednoznačně napsat jako součet symetrické a antisymetrické části (viz Cvičení 2.11):

$$\mathbf{A} = \underbrace{\frac{1}{2}(\mathbf{A}+\mathbf{A}^T)}_{\text{symetrická}} + \underbrace{\frac{1}{2}(\mathbf{A}-\mathbf{A}^T)}_{\text{antisymetrická}}.$$

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.

Příklad 6.3. 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. □

6.2.1 Definitnost kvadratické formy / její matice

Čtvercovou matici $\mathbf{A}$ nazýváme

- **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 existuje $\mathbf{x}$ a $\mathbf{y}$ tak, že $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ a $\mathbf{y}^T\mathbf{A}\mathbf{y} < 0$.

Matice může mít i několik těchto vlastností najednou. Např. positivně definitní matice je zároveň positivně semidefinitní. Nulová matice je zároveň positivně i negativně semidefinitní.

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 definice je jasné, má-li kvadratická forma extrém a případně jaký:

Věta 6.2. *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.*

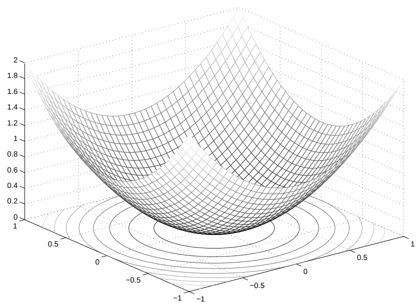
Důkaz. Je-li $\mathbf{A}$ positivně semidefinitní, kvadratická forma nemůže být záporná a zároveň pro $\mathbf{x} = \mathbf{0}$ je nulová, proto v bodě $\mathbf{x} = \mathbf{0}$ (i když možná i jinde) nabývá svého minima. Je-li $\mathbf{A}$ indefinitní a např. $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$, 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 $\mathbf{y}^T\mathbf{A}\mathbf{y} < 0$. □

Ze spektrálního rozkladu (6.7) máme

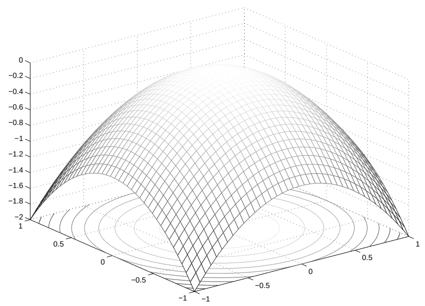
$$f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{x}^T\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T\mathbf{x} = \mathbf{y}^T\mathbf{\Lambda}\mathbf{y} = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2 = g(\mathbf{y}) = g(\mathbf{v}^T\mathbf{x}). \tag{6.11}$$

Vidíme, že substituce $\mathbf{x} = \mathbf{V}\mathbf{y}$ (tj. $\mathbf{y} = \mathbf{V}^T\mathbf{x}$) diagonalizovala matici kvadratické formy. Protože matice $\mathbf{V}$ je ortogonální (a proto regulární), transformace $\mathbf{x} = \mathbf{V}\mathbf{y}$ je isometrie a navíc vzájemně jednoznačná (neboli bijekce, viz §1.1.2). Tedy funkce f a g se liší jen otočením příp. zrcadlením. Totéž můžeme říct i v jazyce bází: zatímco ve standardní (ortonormální) bázi (tvořené sloupečky jednotkové matice $\mathbf{I}$) má kvadratická forma matici $\mathbf{A}$, v ortonormální bázi tvořené sloupci matice $\mathbf{V}$ má forma diagonální matici $\mathbf{\Lambda}$.

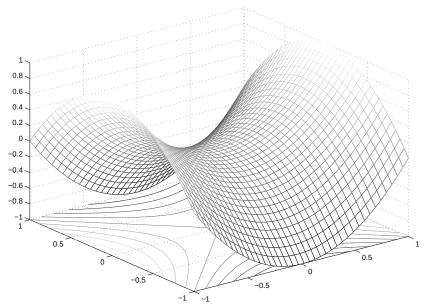
Vlastnosti kvadratické formy jsou mnohem lépe patrny z jejího diagonálního tvaru g než z původního f. Pro případ dvou proměnných ($n = 2$) si grafy diagonální formy g snadno představíme. Jsou-li obě vlastní čísla kladná, funkce g vypadá jako 'dolík'. Jsou-li obě vlastní čísla záporná, funkce g vypadá jako 'kopec'. Mají-li vlastní čísla opačná znaménka, tvarem je 'sedlo':



$g(y_1, y_2) = y_1^2 + y_2^2$

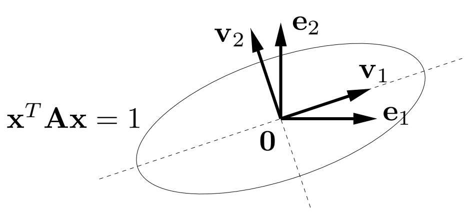


$g(y_1, y_2) = -y_1^2 - y_2^2$



$g(y_1, y_2) = y_1^2 - y_2^2$

Vrstevnice kvadratické formy (tj. množina řešení rovnice $\mathbf{x}^T\mathbf{A}\mathbf{x} = y$, kde ovšem bez ztráty obecnosti volíme $y = 1$) s oběma vlastními čísly kladnými je elipsa se středem v počátku, směry jejíchž hlavních os jsou vlastní vektory:



Věta 6.3. *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.*

Důkaz. To je snadný důsledek diagonalizace (6.11). Protože transformace $\mathbf{x} = \mathbf{V}\mathbf{y}$ je bijekce, platí např. (promyslete!)

$$(\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0 \ \forall \mathbf{x} \in \mathbb{R}^n) \quad \iff \quad (\mathbf{y}^T\mathbf{\Lambda}\mathbf{y} \geq 0 \ \forall \mathbf{y} \in \mathbb{R}^n).$$

Tedy definitnost matice $\mathbf{A}$ je stejná jako definitnost matice $\mathbf{\Lambda}$. Ale definitnost diagonální matice $\mathbf{\Lambda}$ je okamžitě patrná ze znamének čísel λ_i. Např. výraz (6.11) je nezáporný pro každé $\mathbf{y}$, právě když všechna λ_i jsou nezáporná. □

6.3 Kvadratická funkce

Kvadratická funkce je polynom (ne nutně homogenní) druhého stupně. Lze jej psát v maticovém tvaru[4]

$$f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} + c, \tag{6.12}$$

kde $\mathbf{A}^T = \mathbf{A} \in \mathbb{R}^{n\times n}$, $\mathbf{b} \in \mathbb{R}^n$ a $c \in \mathbb{R}$. Oproti[5] kvadratické formě tedy přibyl lineární a konstantní člen. Všimněte si, že pro $n = 1$ je (6.12) 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.

Někdy lze najít $\mathbf{x}_0 \in \mathbb{R}^n$ a $y_0 \in \mathbb{R}$ takové, že

$$\mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} + c = (\mathbf{x} - \mathbf{x}_0)^T\mathbf{A}(\mathbf{x} - \mathbf{x}_0) + y_0. \tag{6.13}$$

Výraz na pravé straně je kvadratická forma s počátkem posunutým do bodu $\mathbf{x}_0$, plus konstanta. Této úpravě se říká **doplnění na čtverec**. Znáte ji pro případ $n = 1$, neboť tak se na základní škole odvozuje vzorec pro kořeny kvadratické rovnice jedné proměnné. Zkusme spočíst $\mathbf{x}_0, y_0$ z daných $\mathbf{A}, \mathbf{b}, c$. Roznásobením pravé strany dostaneme

$$\begin{aligned}(\mathbf{x} - \mathbf{x}_0)^T\mathbf{A}(\mathbf{x} - \mathbf{x}_0) + y_0 &= \mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{A}\mathbf{x}_0 - \mathbf{x}_0^T\mathbf{A}\mathbf{x} + \mathbf{x}_0^T\mathbf{A}\mathbf{x}_0 + y_0 \\ &= \mathbf{x}^T\mathbf{A}\mathbf{x} - 2\mathbf{x}_0^T\mathbf{A}\mathbf{x} + \mathbf{x}_0^T\mathbf{A}\mathbf{x}_0 + y_0.\end{aligned}$$

Porovnáním členů stejného stupně máme

$$\mathbf{b} = -2\mathbf{A}\mathbf{x}_0, \tag{6.14a}$$

$$c = \mathbf{x}_0^T\mathbf{A}\mathbf{x}_0 + y_0, \tag{6.14b}$$

[4] Jak už jsme zmínili v §2.8 pro afinní zobrazení, kvadratickou funkci nemusíme vždy dostat zadanou přímo ve tvaru (6.12). 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 (6.12) může stát dost práce a umu.

[5] Pro $\mathbf{A} = \mathbf{0}$ bude f pouhá afinní funkce. Je věcí konvence, zda afinní funkci máme nazývat kvadratickou či nikoliv, tedy zda máme zakázat případ $\mathbf{A} = \mathbf{0}$.

Pokud soustava (6.14a) má řešení, spočítáme z ní $\mathbf{x}_0$ a pak z druhé rovnice y_0. Pokud soustava (6.14a) 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 hledat pomocí derivací, ale to si ukážeme až později.

Příklad 6.4. 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 6.5. 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. □

Příklad 6.6. Řešme znovu úlohu (5.2). Účelová funkce této úlohy je kvadratická, je totiž

$$\begin{aligned} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 &= (\mathbf{A}\mathbf{x} - \mathbf{b})^T (\mathbf{A}\mathbf{x} - \mathbf{b}) \\ &= (\mathbf{x}^T\mathbf{A}^T - \mathbf{b}^T)(\mathbf{A}\mathbf{x} - \mathbf{b}) \\ &= \mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{A}^T\mathbf{b} - \mathbf{b}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{b} \\ &= \mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x} - 2\mathbf{b}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{b}, \end{aligned} \tag{6.15}$$

kde jsme použili rovnost $\mathbf{b}^T\mathbf{A}\mathbf{x} = (\mathbf{b}^T\mathbf{A}\mathbf{x})^T = \mathbf{x}^T\mathbf{A}^T\mathbf{b}$ (neboť skalár je roven své transpozici). Extrém této kvadratické funkce můžeme najít doplněním na čtverec (viz §6.3). Soustava (6.14a) 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 (6.14a)), tedy dostali jsme normální rovnici (5.3). Zároveň je jasné, že matice $\mathbf{A}^T\mathbf{A}$ je positivně semidefinitní, neboť pro každé $\mathbf{x} \in \mathbb{R}^n$ máme

$$\mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x} = (\mathbf{A}\mathbf{x})^T\mathbf{A}\mathbf{x} = \|\mathbf{A}\mathbf{x}\|^2 \geq 0. \tag{6.16}$$

Tedy v bodě $\mathbf{x}_0$ bude minimum. □

6.3.1 Kvadrika

Vrstevnice kvadratické funkce se nazývá **kvadrika**. Tedy kvadrika je množina[6]

$$\{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} + c = 0 \,\} \tag{6.17}$$

všech řešení kvadratické rovnice, neboli množina všech kořenů kvadratické funkce.

Když $\mathbf{b} = \mathbf{0}$, množina (6.17) 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 (6.17) je pouhá nadrovina. Když $\mathbf{A}$ nemá plnou hodnost, kvadrika je degenerovaná. Množina (6.17) může být i prázdná.

Jestliže kvadratická funkce dovoluje doplnění na čtverec, můžeme množinu (6.17) psát jako

$$\{\, \mathbf{x} \in \mathbb{R}^n \mid (\mathbf{x} - \mathbf{x}_0)^T\mathbf{A}(\mathbf{x} - \mathbf{x}_0) + y_0 = 0 \,\}, \tag{6.18}$$

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**[7]. 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.

Pro $n = 2$ se kvadrika nazývá **kuželosečka** (angl. *conic*).

6.4 Cvičení

6.1. Spočítejte vlastní čísla a vlastní vektory matic $\begin{bmatrix} 1 & 2 \\ -1 & -3 \end{bmatrix}$, $\begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix}$.

6.2. Napište rovnici, jejímiž kořeny jsou vlastní čísla matice $\begin{bmatrix} 2 & 0 & 3 \\ 0 & -2 & -1 \\ 3 & -1 & 2 \end{bmatrix}$.

6.3. Jaká jsou vlastní čísla a vlastní vektory (a) nulové, (b) jednotkové, (c) diagonální matice? Jaká jsou vlastní čísla trojúhelníkové matice?

6.4. Známe vlastní čísla a vektory matice $\mathbf{A}$. Jaká jsou vlastní čísla a vektory matice $\mathbf{A} + \alpha\mathbf{I}$?

6.5. 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 jejich vlastních vektorů?

6.6. ($\star$) Ukažte, že dvě čtvercové matice komutují, právě když mají stejné vlastní vektory.

6.7. Pro dané matice určete, zda jsou positivně/negativně (semi)definitní nebo indefinitní:

$$\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}, \quad \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

6.8. Mějme matici $\mathbf{A} = \begin{bmatrix} 1 & -3 \\ 2 & -4 \end{bmatrix}$. Která z následujících tvrzení jsou pravdivá?

[6]Bez ztráty obecnosti uvažujeme vrstevnici výšky nula (tj. množina kořenů) kvadratické funkce (6.12), neboť c můžeme zvolit libovolně.

[7]Někteří autoři myslí elipsoidem množinu i s vnitřkem, někteří jen její hranici. Rozdíl je stejný jako mezi sférou a koulí.

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.

6.9. 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.

6.10. Je množina $\{\, (x, y) \in \mathbb{R}^2 \mid x^2 - 3xy + y^2 = 1 \,\}$ elipsa nebo hyperbola? Odůvodněte.

6.11. ($\star$) 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).

6.12. Ukažte, že je-li $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ spektrální rozklad symetrické matice $\mathbf{A}$, platí $\mathbf{A}^n = \mathbf{V}\mathbf{\Lambda}^n\mathbf{V}^T$.

6.13. Následující kvadratické funkce napište ve tvaru (6.12) se symetrickou $\mathbf{A}$. Pak najděte jejich extrémy a určete typ každého extrému (použijte doplnění na čtverec).

a) $f(x, y) = x^2 + 4xy - 2y^2 + 3x - 6y + 5$
b) $f(x, y) = x^2 + 2y^2 - 2xy + 2x - y$

6.14. Máme neorientovaný graf (V, E) s množinou vrcholů $V = \{1, \ldots, n\}$ a množinou hran $E \subseteq \binom{V}{2} = \{\, \{i, j\} \mid i, j \in V,\ i \neq j \,\}$. 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 $f\colon \mathbb{R}^n \to \mathbb{R}$ jako $f(\mathbf{x}) = \frac{1}{2}\sum_{\{i,j\} \in E}(x_i - x_j)^2$.

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 je graf zadán maticí sousednosti $\mathbf{A} \in \{0, 1\}^{n \times n}$ tak, že $a_{ij} = 1$ právě když $\{i, j\} \in E$. Najděte matici $\mathbf{L} \in \mathbb{R}^{n \times n}$ tak, že $f(\mathbf{x}) = \mathbf{x}^T\mathbf{L}\mathbf{x}$. Hledejte co nejjednodušší vztah pro $\mathbf{L}$. Použijte přitom kromě matice $\mathbf{A}$ také diagonální matici $\mathbf{D} = \operatorname{diag}(\mathbf{A}\mathbf{1})$; jaký je význam matice $\mathbf{D}$?

Poznamenejme, že funkci f (příp. matici $\mathbf{L}$) se říká *Laplacián* grafu (V, E).

6.15. Musí mít positivně semidefinitní matice na diagonále nezáporné prvky? Odpověď dokažte.

6.16. Dokažte, že matice $\mathbf{A}^T\mathbf{A} + \mu\mathbf{I}$ je positivně definitní pro každou matici $\mathbf{A}$ a každé $\mu > 0$.

6.17. Dokažte, že symetrická matice je positivně definitní, právě když je positivně semidefinitní a invertovatelná.

6.18. Dokažte, že každá symetrická positivně definitní matice je invertovatelná a její inverze je také positivně definitní. Dokažte

a) s použitím spektrálního rozkladu,
b) bez použití spektrálního rozkladu.

6.19. 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 matice $\mathbf{A}$ má lineárně nezávislé sloupce,
c) Pro každou positivně semidefinitní matici $\mathbf{B}$ existuje matice $\mathbf{A}$ tak, že $\mathbf{B} = \mathbf{A}^T\mathbf{A}$.
d) Pro každou positivně definitní matici $\mathbf{B}$ existuje regulární matice $\mathbf{A}$ tak, že $\mathbf{B} = \mathbf{A}^T\mathbf{A}$.

6.20. ($\star$) Positivně semidefinitní symetrické matice lze vnímat jako zobecnění nezáporných čísel. Proto se někdy positivní semidefinitnost značí $\mathbf{A} \succeq \mathbf{0}$. Zápis $\mathbf{A} \preceq \mathbf{B}$ je pak zkratkou $\mathbf{B} - \mathbf{A} \succeq \mathbf{0}$. Dokažte, že relace $\preceq$ je částečné uspořádání (tj. reflexivní, tranzitivní a antisymetrická) na množině symetrických matic $n \times n$.

6.21. ($\star$) Na základě podobnosti relace $\preceq$ z předchozího cvičení a relace $\leq$ na množině $\mathbb{R}$ bychom očekávali, že:

a) Pokud $\mathbf{A} \preceq \mathbf{B}$ a $\mathbf{C} \preceq \mathbf{D}$, potom $\mathbf{A} + \mathbf{C} \preceq \mathbf{B} + \mathbf{D}$.
b) Pokud $\mathbf{A} \succeq \mathbf{0}$ a $\alpha \geq 0$, potom $\alpha\mathbf{A} \succeq \mathbf{0}$.
c) Pokud $\mathbf{A} \succeq \mathbf{0}$, potom $\mathbf{A}^2 \succeq \mathbf{0}$.
d) Pokud $\mathbf{A} \succeq \mathbf{0}$ a $\mathbf{B} \succeq \mathbf{0}$, potom $\mathbf{AB} \succeq \mathbf{0}$.
e) 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.

6.22. Najděte všechna vlastní čísla projektoru. Odpověď nalezněte algebraicky z idempotence projektoru a poté odůvodněte geometrickou úvahou.

6.23. Geometrickou úvahou najděte aspoň dva vlastní vektory a příslušná vlastní čísla Householderovy matice ze Cvičení 4.13.

6.24. 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 ortogonální projektor (tj. matice reprezentující ortogonální projekci) je positivně semidefinitní. Jaký to má geometrický význam?

Nápověda a řešení

6.4. Vlastní čísla se zvětší o α. Vlastní vektory jsou stejné.

6.5. Nechť $\mathbf{ABv} = \lambda\mathbf{v} \neq \mathbf{0}$. Z toho plyne $\mathbf{BABv} = \lambda\mathbf{Bv}$, tedy $\mathbf{BAu} = \lambda\mathbf{u}$ kde $\mathbf{u} = \mathbf{Bv}$. Zároveň $\mathbf{u} \neq \mathbf{0}$ protože jinak by bylo $\lambda\mathbf{v} = \mathbf{0}$.

6.7. indefinitní, positivně definitní, indefinitní, positivně semidefinitní

6.8. 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.

6.9.a) $\mathbf{A} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$

6.9.b) $a = 2$, $b = 4$, $\mathbf{U} = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$

6.10. Hyperbola, neboť $\mathbf{A}$ má nenulová vlastní čísla opačných znamének.

6.13.a) Převod na tvar (6.12): $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$, $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 3 \\ -6 \end{bmatrix}$, $c = 5$. Doplnění na čtverec existuje (už proto, že $\mathbf{A}$ je regulární). Funkce nemá extrém (protože $\mathbf{A}$ je indefinitní), má sedlo v bodě $(\frac{1}{2}, -1)$.

6.13.b) Má minimum v bodě $-(3,1)/2$.

6.14.a) Je $(x_i - x_j)^2 = x_i^2 - 2x_ix_j + x_j^2$. Tedy f je homogenní polynom stupně dva, což je kvadratická forma.

6.14.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.

6.14.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 když všechna x_i jsou stejná. Jinak to nastane právě když jsou x_i stejná v každé komponentě grafu.

6.14.d) Prvek d_{ii} matice $\mathbf{D}$ je stupeň (tedy počet incidentních hran) vrcholu i. Je $\mathbf{L} = \mathbf{D} - \mathbf{A}$, neboť

$$\frac{1}{2}\sum_{\{i,j\}}(x_i - x_j)^2 = \frac{1}{2}\sum_{\{i,j\}} x_i^2 - \sum_{\{i,j\}} x_ix_j + \frac{1}{2}\sum_{\{i,j\}} x_j^2 = \sum_{\{i,j\}} x_i^2 - \sum_{\{i,j\}} x_ix_j = \mathbf{x}^T\mathbf{D}\mathbf{x} - \mathbf{x}^T\mathbf{A}\mathbf{x}.$$

6.15. Musí. Stačí vzít $\mathbf{x} = \mathbf{e}_i$ vektory standardní báze.

6.16. $\mathbf{x}^T(\mathbf{A}^T\mathbf{A} + \mu\mathbf{I})\mathbf{x} = \mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x} + \mu\mathbf{x}^T\mathbf{I}\mathbf{x} = \|\mathbf{A}\mathbf{x}\|^2 + \mu\|\mathbf{x}\|^2 > 0$ pro každé $\mathbf{x} \neq \mathbf{0}$.

6.17. Necht' $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$. Matice $\mathbf{A}$ je positivně semidefinitní, právě když jsou všechny diagonální prvky matice $\mathbf{\Lambda}$ (což jsou vlastní čísla matice $\mathbf{A}$) nezáporné. Dle (6.8) je $\mathbf{A}$ 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.

6.18.a) Je $\mathbf{A}^{-1} = (\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T)^{-1} = \mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{V}^T$. Ale definitnost matice $\mathbf{A}$ je stejná jako definitnost matice $\mathbf{\Lambda}$. Je jasné, že pokud diagonální prvky λ_i matice $\mathbf{\Lambda}$ jsou kladné, pak jsou kladné i diagonální prvky $1/\lambda_i$ matice $\mathbf{\Lambda}^{-1}$.

6.18.b) Necht' $\mathbf{A}$ 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 3.7 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 $\mathbf{x}^T\mathbf{A}\mathbf{x} = 0$.

Necht' $\mathbf{A}$ 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 $\mathbf{y} \neq \mathbf{0}$.

6.19.a) $\mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x} = (\mathbf{A}\mathbf{x})^T\mathbf{A}\mathbf{x} = \|\mathbf{A}\mathbf{x}\|^2 \geq 0$

6.19.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 (5.5).

6.19.c) Ve spektrálním rozkladu $\mathbf{B} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ má $\mathbf{\Lambda}$ nezáporné diagonální prvky. Položme $\mathbf{A} = \mathbf{\Lambda}^{1/2}\mathbf{V}^T$.

6.19.d) Ve spektrálním rozkladu $\mathbf{B} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ má $\mathbf{\Lambda}$ kladné diagonální prvky. Položme $\mathbf{A} = \mathbf{\Lambda}^{1/2}\mathbf{V}^T$.

6.20. 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 $\mathbf{A} \preceq \mathbf{B} \preceq \mathbf{A} \implies \mathbf{A} = \mathbf{B}$. 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 $\mathbf{A} \preceq \mathbf{B} \preceq \mathbf{C} \implies \mathbf{A} \preceq \mathbf{C}$. To platí, neboť $\mathbf{x}^T\mathbf{A}\mathbf{x} \leq \mathbf{x}^T\mathbf{B}\mathbf{x} \leq \mathbf{x}^T\mathbf{C}\mathbf{x} \implies \mathbf{x}^T\mathbf{A}\mathbf{x} \leq \mathbf{x}^T\mathbf{C}\mathbf{x}$.

6.22. Algebraicky: Projektor $\mathbf{P}$ splňuje $\mathbf{PP} = \mathbf{P}$ (idempotence). Tedy z $\mathbf{Pv} = \lambda\mathbf{v}$ plyne (násobením maticí $\mathbf{P}$) $\mathbf{Pv} = \lambda\mathbf{Pv}$. To může platit jen pro $\lambda = 1$ nebo pro $\mathbf{Pv} = \mathbf{0}$. Ale když $\mathbf{Pv} = \mathbf{0}$, tak z rovnice $\mathbf{Pv} = \lambda\mathbf{v}$ musí být $\lambda = 0$. Tedy vlastní čísla projektoru mohou být buď 1 nebo 0.

Úvaha: Projekce na podprostor $X = \operatorname{rng}\mathbf{P}$ promítne každý vektor $\mathbf{x} \in X$ na sebe, tedy $\mathbf{Px} = \mathbf{x}$. Každý vektor $\mathbf{x} \in X^\perp$ se promítne do $\mathbf{0}$, tedy $\mathbf{Px} = \mathbf{0}$. Každý vektor $\mathbf{x} \notin X \cup X^\perp$ se promítne do vektoru, který není rovnoběžný s $\mathbf{x}$. Shrnuto, $\mathbf{P}$ má jen dvě vlastní čísla 1 a 0 (předpokládáme nyní, že $X^\perp \neq \{\mathbf{0}\}$, neboli $\mathbf{P}$ není regulární). Vlastní vektory příslušné číslu 1 tvoří podprostor X, vlastní vektory příslušné číslu 0 tvoří podprostor $X^\perp$.

6.23. Transformace $\mathbf{f}(\mathbf{x}) = \mathbf{Hx}$ je zrcadlení okolo nadroviny s normálovým vektorem $\mathbf{v}$. Tedy $\mathbf{Hx} = \mathbf{x}$ pro všechna $\mathbf{x}$ ležící v této nadrovině, neboli $\mathbf{v}^T\mathbf{x} = 0$. Dále je $\mathbf{Hv} = -\mathbf{v}$, neboť normála nadroviny zrcadlením změní orientaci.

6.25. Projektor $\mathbf{P}$ splňuje $\mathbf{P}^2 = \mathbf{P}$ a $\mathbf{P}^T = \mathbf{P}$ (viz §4.6 a Cvičení 5.17), z čehož plyne $\mathbf{P} = \mathbf{P}^T\mathbf{P}$. Nyní $\mathbf{z}^T\mathbf{Pz} = \mathbf{z}^T\mathbf{P}^T\mathbf{Pz} = \|\mathbf{Pz}\|^2 \geq 0$ (viz Cvičení 6.19). Geometrický význam je takový, že libovolný vektor $\mathbf{z}$ svírá se svou ortogonální projekcí $\mathbf{x} = \mathbf{Pz}$ ostrý úhel nebo jsou na sebe kolmé (když $\mathbf{z} = \mathbf{0}$ nebo $\mathbf{x} = \mathbf{0}$), tedy $\mathbf{z}^T\mathbf{x} = \mathbf{z}^T\mathbf{Pz} \geq 0$.