

# Kapitola 8

## Skalární součin

Vektorové prostory byly definovány velice obecně, takže pokryjí velkou třídu objektů. Na druhou stranu, jestliže přidáme další požadavky na to co mají prostory splňovat, tak budeme moci odvodit hlubší výsledky. Konkrétně, skalární součin dává možnost přirozeně zavést pojem kolmosti, velikost a vzdálenost vektorů (a tím i limity) atd.

### 8.1 Skalární součin a norma

Skalární součin (stejně jako grupu, vektorové prostory aj.) zavádíme obecně výčtem vlastností, které má splňovat.

**Definice 8.1** (Skalární součin nad  $\mathbb{R}$ ). Buď  $V$  vektorový prostor nad  $\mathbb{R}$ . Pak *skalární součin* je zobrazení  $\langle \cdot, \cdot \rangle: V^2 \rightarrow \mathbb{R}$ , splňující:

- (1)  $\langle x, x \rangle \geq 0 \ \forall x \in V$ , a rovnost nastane pouze pro  $x = 0$ ,
- (2)  $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle \ \forall x, y, z \in V$ ,
- (3)  $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle \ \forall x, y \in V, \forall \alpha \in \mathbb{R}$ ,
- (4)  $\langle x, y \rangle = \langle y, x \rangle \ \forall x, y \in V$ .

Zobecnění na těleso komplexních čísel je následující. Připomeňme ještě, že komplexně sdružené číslo k  $a + bi \in \mathbb{C}$  je definované jako  $\overline{a + bi} = a - bi$ .

**Definice 8.2** (Skalární součin nad  $\mathbb{C}$ ). Buď  $V$  vektorový prostor nad  $\mathbb{C}$ . Pak *skalární součin* je zobrazení  $\langle \cdot, \cdot \rangle: V^2 \rightarrow \mathbb{C}$ , splňující:

- (1)  $\langle x, x \rangle \geq 0 \ \forall x \in V$ , a rovnost nastane pouze pro  $x = 0$ ,
- (2)  $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle \ \forall x, y, z \in V$ ,
- (3)  $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle \ \forall x, y \in V, \forall \alpha \in \mathbb{C}$ ,
- (4)  $\langle x, y \rangle = \overline{\langle y, x \rangle} \ \forall x, y \in V$ .

Kvůli vlastnosti (1), kde je potřeba uspořádání, zavádíme skalární součin pouze nad tělesy  $\mathbb{R}$  a  $\mathbb{C}$ .

Čtvrtá vlastnost u komplexního skalárního součinu zařídí, že  $\langle x, x \rangle = \overline{\langle x, x \rangle} \in \mathbb{R}$ , tedy  $\langle x, x \rangle$  je vždy reálné číslo a lze jej porovnávat s nulou u první vlastnosti.

Vlastnosti (2) a (3) říkají, že skalární součin je lineární funkcí v první složce. Jak je to s druhou?

$$\begin{aligned}\langle x, y + z \rangle &= \overline{\langle y + z, x \rangle} = \overline{\langle y, x \rangle + \langle z, x \rangle} = \overline{\langle y, x \rangle} + \overline{\langle z, x \rangle} = \langle x, y \rangle + \langle x, z \rangle, \\ \langle x, \alpha y \rangle &= \overline{\langle \alpha y, x \rangle} = \overline{\alpha \langle y, x \rangle} = \overline{\alpha} \overline{\langle y, x \rangle} = \overline{\alpha} \langle x, y \rangle.\end{aligned}$$

Ve druhé složce tedy komplexní skalární součin není lineární, ačkoli reálný skalární součin jest.

Pokud dosadíme  $\alpha = 0$ , dostáváme  $\langle o, x \rangle = \langle x, o \rangle = 0$ , tedy násobení jakéhokoli vektoru s nulovým dá nulu.

**Příklad 8.3** (Příklady standardních skalárních součinů).

- V prostoru  $\mathbb{R}^n$ : standardní skalární součin  $\langle x, y \rangle = x^T y = \sum_{i=1}^n x_i y_i$ .
- V prostoru  $\mathbb{C}^n$ : standardní skalární součin  $\langle x, y \rangle = x^T \bar{y} = \sum_{i=1}^n x_i \bar{y}_i$ .
- V prostoru  $\mathbb{R}^{m \times n}$ : standardní skalární součin  $\langle A, B \rangle = \sum_{i=1}^m \sum_{j=1}^n a_{ij} b_{ij}$ .
- V  $\mathcal{C}_{[a,b]}$ , prostoru spojitých funkcí na intervalu  $[a, b]$ : standardní skalární součin  $\langle f, g \rangle = \int_a^b f(x)g(x)dx$ .

Výše zmíněné skalární součiny jsou pouze příklady možných zavedení součinů na daných prostorech; jako skalární součin mohou fungovat i jiné operace. Později, ve větě 11.17, popíšeme všechny skalární součiny v prostoru  $\mathbb{R}^n$ .  $\square$

Nadále uvažujeme vektorový prostor  $V$  nad  $\mathbb{R}$  či  $\mathbb{C}$  se skalárním součinem. Nejprve ukážeme, že skalární součin umožňuje zavést normu, neboli velikost vektoru.

**Definice 8.4** (Norma indukovaná skalárním součinem). *Norma indukovaná skalárním součinem* je definovaná  $\|x\| := \sqrt{\langle x, x \rangle}$ , kde  $x \in V$ .

Norma je dobře definovaná díky první vlastnosti z definice skalárního součinu, a je to vždy nezáporná hodnota.

Pro standardní skalární součin v  $\mathbb{R}^n$  dostáváme známou eukleidovskou normu  $\|x\| = \sqrt{x^T x} = \sqrt{\sum_{i=1}^n x_i^2}$ .

Geometrická interpretace standardního skalárního součinu v  $\mathbb{R}^n$  je  $\langle x, y \rangle = \|x\| \cdot \|y\| \cos(\varphi)$ , kde  $\varphi$  je úhel mezi vektory  $x, y$ . Speciálně,  $x, y$  jsou kolmé právě tehdy, když  $\langle x, y \rangle = 0$ . V jiných prostorech takováto geometrie chybí, proto kolmost zavedeme právě pomocí vztahu  $\langle x, y \rangle = 0$ .

**Definice 8.5** (Kolmost). Vektory  $x, y \in V$  jsou *kolmé*, pokud  $\langle x, y \rangle = 0$ . Značení:  $x \perp y$ .

**Příklad 8.6** (Příklady kolmých vektorů pro standardní skalární součiny).

- V prostoru  $\mathbb{R}^3$ :  $(1, 2, 3) \perp (1, 1, -1)$ .
- V prostoru  $\mathcal{C}_{[-\pi, \pi]}$ :  $\sin x \perp \cos x \perp 1$ .  $\square$

**Věta 8.7** (Pythagorova). *Pokud  $x, y \in V$  jsou kolmé, tak  $\|x + y\|^2 = \|x\|^2 + \|y\|^2$ .*

*Důkaz.*  $\|x + y\|^2 = \langle x + y, x + y \rangle = \langle x, x \rangle + \underbrace{\langle x, y \rangle}_{=0} + \underbrace{\langle y, x \rangle}_{=0} + \langle y, y \rangle = \langle x, x \rangle + \langle y, y \rangle = \|x\|^2 + \|y\|^2$ .  $\square$

Poznamenejme, že nad  $\mathbb{R}$  platí i opačná implikace, ale nad  $\mathbb{C}$  obecně nikoli (viz problém 8.2).

**Věta 8.8** (Cauchyho–Schwarzova nerovnost<sup>1)</sup>). *Pro každé  $x, y \in V$  platí  $|\langle x, y \rangle| \leq \|x\| \cdot \|y\|$ .*

*Důkaz.* (Reálná verze) Nejprve ukážeme reálnou verzi, protože má elegantní důkaz. Pro  $y = o$  platí nerovnost triviálně, tak předpokládejme  $y \neq o$ . Uvažujme reálnou funkci  $f(t) = \langle x + ty, x + ty \rangle \geq 0$  proměnné  $t \in \mathbb{R}$ . Pak

$$f(t) = \langle x, x \rangle + t\langle x, y \rangle + t\langle y, x \rangle + t^2\langle y, y \rangle = \langle x, x \rangle + 2t\langle x, y \rangle + t^2\langle y, y \rangle.$$

Jedná se o kvadratickou funkci, která je všude nezáporná, nemůže mít tedy dva různé kořeny. Proto je příslušný diskriminant nekladný:

$$4\langle x, y \rangle^2 - 4\langle x, x \rangle\langle y, y \rangle \leq 0.$$

Z toho dostáváme  $\langle x, y \rangle^2 \leq \langle x, x \rangle\langle y, y \rangle$ , odmocněním  $|\langle x, y \rangle| \leq \|x\| \cdot \|y\|$ .  $\square$

*Důkaz.* (Komplexní verze) Je-li  $\langle x, y \rangle = 0$ , pak tvrzení platí triviálně. Budeme tedy předpokládat, že  $\langle x, y \rangle \neq 0$ , a zavedeme vektor

$$z := \frac{\langle y, y \rangle}{\langle x, y \rangle} x - y.$$

<sup>1)</sup>Nerovnost se také někdy nazývá jen Schwarzova, nebo Cauchyho–Bunjakovského, popř. Cauchyho–Schwarzova–Bunjakovského. Augustin-Louis Cauchy ji dokázal r. 1821 pro prostor  $\mathbb{R}^n$  a později ji nezávisle na sobě zobecnili Hermann Amandus Schwarz (1880) a Viktor Jakovlevič Bunjakovskij (1859).

Potom platí

$$\langle z, y \rangle = \left\langle \frac{\langle y, y \rangle}{\langle x, y \rangle} x - y, y \right\rangle = \frac{\langle y, y \rangle}{\langle x, y \rangle} \langle x, y \rangle - \langle y, y \rangle = 0.$$

Vektory  $z, y$  jsou tudíž na sebe kolmé a podle Pythagorovy věty

$$\|z + y\|^2 = \|z\|^2 + \|y\|^2,$$

neboli

$$\frac{\langle y, y \rangle^2}{|\langle x, y \rangle|^2} \|x\|^2 = \|z\|^2 + \|y\|^2 \geq \|y\|^2.$$

Zkrácením  $\|y\|^2$  a vynásobením  $|\langle x, y \rangle|^2$  dostaneme hledanou nerovnost  $\|x\|^2 \|y\|^2 \geq |\langle x, y \rangle|^2$  v druhých mocninách.  $\square$

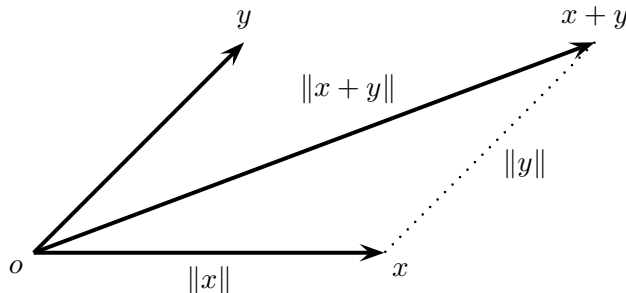
Občas se Cauchyho–Schwarzova nerovnost uvádí v ekvivalentní podobě

$$|\langle x, y \rangle|^2 \leq \langle x, x \rangle \langle y, y \rangle.$$

Cauchyho–Schwarzova nerovnost je užitečná a často používaná pro odvozování dalších výsledků na obecné bázi, nebo i pro konkrétní algebraické výrazy. Např. pro standardní skalární součin v  $\mathbb{R}^n$  dostaneme nerovnost

$$\left( \sum_{i=1}^n x_i y_i \right)^2 \leq \left( \sum_{i=1}^n x_i^2 \right) \left( \sum_{i=1}^n y_i^2 \right).$$

Další využití viz např. [Krisl, 2008]. My použijeme Cauchyho–Schwarzovu nerovnost hned v následujícím k odvození trojúhelníkové nerovnosti.



**Důsledek 8.9** (Trojúhelníková nerovnost). *Pro každé  $x, y \in V$  platí  $\|x + y\| \leq \|x\| + \|y\|$ .*

*Důkaz.* Nejprve připomeňme, že pro každé komplexní číslo  $z = a + bi$  platí:  $z + \bar{z} = 2a = 2 \operatorname{Re}(z)$ , a dále  $a \leq |z|$ . Nyní můžeme odvodit:

$$\begin{aligned} \|x + y\|^2 &= \langle x + y, x + y \rangle = \langle x, x \rangle + \langle y, y \rangle + \langle x, y \rangle + \langle y, x \rangle = \\ &= \langle x, x \rangle + \langle y, y \rangle + 2 \operatorname{Re}(\langle x, y \rangle) \leq \langle x, x \rangle + \langle y, y \rangle + 2 |\langle x, y \rangle| \leq \\ &\leq \|x\|^2 + \|y\|^2 + 2 \|x\| \cdot \|y\| = (\|x\| + \|y\|)^2, \end{aligned}$$

kde poslední nerovnost plyne z Cauchyho–Schwarzovy nerovnosti. Tedy máme  $\|x + y\|^2 \leq (\|x\| + \|y\|)^2$  a odmocněním získáme hledaný vztah.  $\square$

Norma indukovaná skalárním součinem je jen jedním typem normy, pojem normy je ale definován obecněji. My budeme vesměs pracovat s normou indukovanou skalárním součinem, takže následující je pouze malou odbočkou.

**Definice 8.10** (Norma). Buď  $V$  vektorový prostor nad  $\mathbb{R}$  nebo  $\mathbb{C}$ . Pak *norma* je zobrazení  $\|\cdot\|: V \rightarrow \mathbb{R}$ , splňující:

- (1)  $\|x\| \geq 0$  pro všechna  $x \in V$ , a rovnost nastane pouze pro  $x = 0$ ,
- (2)  $\|\alpha x\| = |\alpha| \cdot \|x\|$  pro všechna  $x \in V$ , a pro všechna  $\alpha \in \mathbb{R}$  resp.  $\alpha \in \mathbb{C}$ ,

$$(3) \|x + y\| \leq \|x\| + \|y\|.$$

**Tvrzení 8.11.** Norma indukovaná skalárním součinem je normou.

*Důkaz.* Vlastnost (1) je splněna díky definici normy indukované skalárním součinem. Vlastnost (3) je ukázána v důsledku 8.9. Zbývá vlastnost (2):

$$\|\alpha x\| = \sqrt{\langle \alpha x, \alpha x \rangle} = \sqrt{\alpha \bar{\alpha} \langle x, x \rangle} = \sqrt{\alpha \bar{\alpha}} \sqrt{\langle x, x \rangle} = |\alpha| \cdot \|x\|. \quad \square$$

**Příklad 8.12** (Příklady norem v  $\mathbb{R}^n$ ). Užitečná třída norem jsou tzv.  $p$ -normy. Pro  $p = 1, 2, \dots$  definujeme  $p$ -normu vektoru  $x \in \mathbb{R}^n$  jako

$$\|x\|_p = \left( \sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}.$$

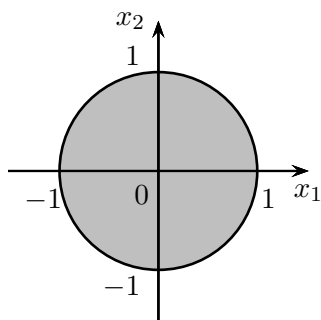
Speciální volby  $p$  vedou ke známým normám:

- pro  $p = 2$ : eukleidovská norma  $\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$ , což je norma indukovaná standardním skalárním součinem,
- pro  $p = 1$ : součtová norma  $\|x\|_1 = \sum_{i=1}^n |x_i|$ ; nazývá se manhattanská norma, protože odpovídá reálným vzdálenostem při procházení pravoúhlé sítě ulic v městě,
- pro  $p = \infty$  (limitním přechodem): maximová (Čebyševova) norma  $\|x\|_\infty = \max_{i=1, \dots, n} |x_i|$ .  $\square$

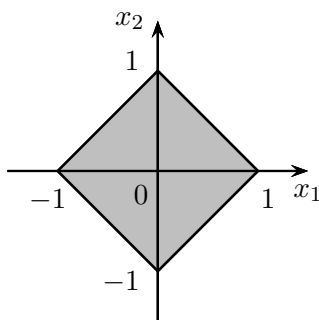
**Příklad 8.13** (Jednotková koule). Jednotková koule je množina vektorů, které mají normu nanejvýš 1 a tedy jsou od počátku vzdáleny maximálně 1. Formálně definujeme jednotkovou kouli jako

$$\{x \in V; \|x\| \leq 1\}.$$

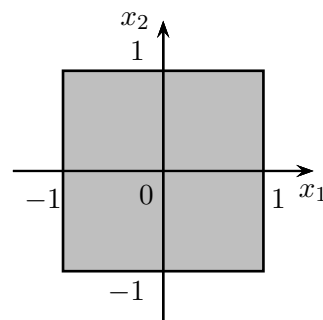
Tento pojem je nejenom užitečný nástroj, ale zároveň umožňuje normu geometricky vizualizovat. Uvažujme pro konkrétnost rovinu  $\mathbb{R}^2$  a tři základní normy z předchozího příkladu 8.12. Jednotková koule má tvar:



eukleidovská norma



součtová norma



maximová norma

Jiné normy mají za jednotkovou kouli jiný geometrický objekt. Každá jednotková koule v  $\mathbb{R}^n$  ale musí být množina, která je uzavřená, omezená, symetrická dle počátku, konvexní (tj., s každými dvěma body obsahuje i jejich spojnici) a počátek leží v jejím vnitřku. Platí i opačné tvrzení – každá množina splňující tyto vlastnosti představuje jednotkovou kouli nějaké normy.  $\square$

**Příklad 8.14** (Příklady norem v  $\mathcal{C}_{[a,b]}$ ). Normu spojitě funkce  $f: [a, b] \rightarrow \mathbb{R}$  lze zavést analogicky jako pro Eukleidovský prostor:

- analogie eukleidovské normy:  $\|f\|_2 = \sqrt{\int_a^b f(x)^2 dx}$ ,
- analogie součtové normy:  $\|f\|_1 = \int_a^b |f(x)| dx$ ,
- analogie maximové normy:  $\|f\|_\infty = \max_{x \in [a,b]} |f(x)|$ ,
- analogie  $p$ -normy:  $\|f\|_p = \left( \int_a^b |f(x)|^p dx \right)^{\frac{1}{p}}$ .  $\square$

**Poznámka 8.15** (Rovnoběžníkové pravidlo). Pro normu indukovanou skalárním součinem platí tzv. *rovnoběžníkové pravidlo*:

$$\|x - y\|^2 + \|x + y\|^2 = 2\|x\|^2 + 2\|y\|^2.$$

*Důkaz.*  $\|x - y\|^2 + \|x + y\|^2 = \langle x - y, x - y \rangle + \langle x + y, x + y \rangle = 2\langle x, x \rangle + 2\langle y, y \rangle = 2\|x\|^2 + 2\|y\|^2.$   $\square$

Díky tomu snadno nahlédneme, že součtová a maximová norma nejsou indukované žádným skalárním součinem. Např. pro  $x = (1, 0)$  a  $y = (0, 1)$  nespĺňují rovnoběžníkové pravidlo.

Platí dokonce silnější tvrzení: Pokud pro normu platí rovnoběžníkové pravidlo, pak je indukovaná nějakým skalárním součinem; viz Horn and Johnson [1985].

Norma umožňuje zavést vzdálenost (neboli metriku) mezi vektory  $x, y$  jako  $\|x - y\|$ . A pokud máme vzdálenost, můžeme zavést limity, etc. Čtenáře by již nemělo překvapit, že i metriku lze zavést axiomatically. Navíc k definici metriky nepotřebujeme ani vektorový prostor, stačí libovolná množina.

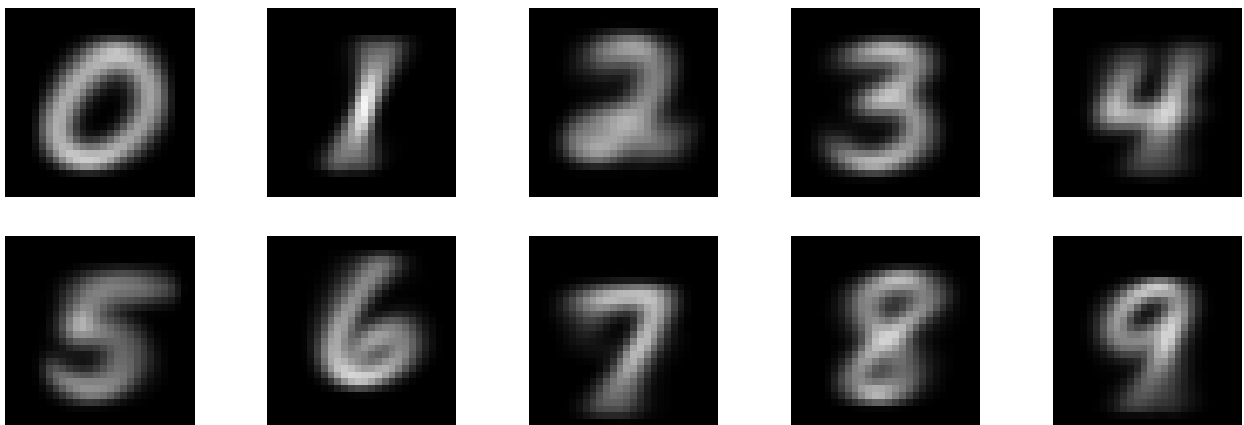
**Poznámka 8.16** (Metrika). Metriku na množině  $M$  definujeme jako je zobrazení  $d: M^2 \rightarrow \mathbb{R}$ , splňující:

- (1)  $d(x, y) \geq 0$  pro všechna  $x, y \in M$ , a rovnost nastane pouze pro  $x = y$ ,
- (2)  $d(x, y) = d(y, x)$  pro všechna  $x, y \in M$ ,
- (3)  $d(x, z) \leq d(x, y) + d(y, z)$  pro všechna  $x, y, z \in M$ .

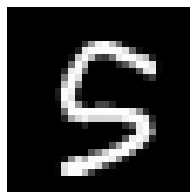
Každá norma určuje metriku předpisem  $d(x, y) := \|x - y\|$ , tedy vzdálenost vektorů  $x, y$  zavádí jako velikost jejich rozdílu. Na druhou stranu to ale obecně neplatí. Existují prostory s metriku, která není indukována žádnou normou, např. diskretní metrika  $d(x, y) := \lceil \|x - y\|_2 \rceil$ , nebo diskretní metrika definovaná  $d(x, y) := 1$  pro  $x \neq y$  a  $d(x, y) := 0$  pro  $x = y$ .

**Příklad 8.17** (Klasifikace psaných číslic). Ukážeme použití vzdálenosti na vytvoření jednoduchého klasifikátoru pro automatickou identifikaci ručně psaných číslic. Tento problém je velmi důležitý mj. pro efektivitu poštovní služby. Uvádí se [Chartier, 2015], že jen v roce 2012 bylo v USA doručeno 160 miliard dopisů. Automatizace pak znamená zrychlení a zlevnění celého procesu doručování dopisů. První program pro detekci číslic byl spuštěn v roce 1997 a přestože měl úspěšnost pouze 10%, znamenal výrazný posun kupředu.

Předpokládáme, že každá číslice je zadána jako obrázek, reprezentovaný maticí  $A \in \mathbb{R}^{m \times n}$ , tedy pixel obrázku na pozici  $(i, j)$  má barvu s číslem  $a_{ij}$ . Jako vzory použijeme zprůměrované hodnoty z databáze MNIST (<http://yann.lecun.com/exdb/mnist/>):



Nyní uvažujme následující obrázek, který chceme klasifikovat jako číslici:



Na prostoru matic proto musíme zavést metriku. K tomuto účelu adaptujeme klasickou eukleidovskou vzdálenost a vzdálenost matic  $A, B \in \mathbb{R}^{m \times n}$  definujeme jako

$$\|A - B\| := \sqrt{\sum_{i=1}^m \sum_{j=1}^n (a_{ij} - b_{ij})^2}.$$

Pokud spočítáme vzdálenost mezi maticí reprezentující klasifikovaný obrázek a jednotlivými vzory, pak dostaneme hodnoty:

0: 1957.44    1: 2237.30    2: 2015.79    3: 1816.23    4: 1868.78  
 5: 1771.64    6: 2038.57    7: 2090.51    8: 1843.22    9: 1900.81

Vidíme, že nejmenší vzdálenost je k obrázku číslo 5, proto klasifikujeme daný obrázek správně jako číslo 5. Takovýto jednoduchý klasifikátor se ale snadno může splést. Náš klasifikátor nedokáže rozpoznat tvary čar (znaménko křivosti atp.), a tudíž je malá vzdálenost je i třeba k obrázku číslo 3.  $\square$

## 8.2 Ortonormální báze, Gramova–Schmidtova ortogonalizace

Každý vektorový prostor má bázi. U prostoru se skalárním součinem je přirozené se ptát, zda existuje báze složená z navzájem kolmých vektorů. V této sekci ukážeme, že je to pravda, že taková báze má řadu pozoruhodných vlastností a také odvodíme algoritmus na její nalezení.

**Definice 8.18** (Ortogonalní a ortonormální systém). Systém vektorů  $z_1, \dots, z_n$  je *ortogonální*, pokud  $\langle z_i, z_j \rangle = 0$  pro všechna  $i \neq j$ . Systém vektorů  $z_1, \dots, z_n$  je *ortonormální*, pokud je ortogonální a  $\|z_i\| = 1$  pro všechna  $i = 1, \dots, n$ .

Je-li systém  $z_1, \dots, z_n$  ortonormální, pak je také ortogonální. Naopak to obecně neplatí, ale není problém ortogonální systém zortonormalizovat. Jsou-li  $z_1, \dots, z_n$  nenulové a ortogonální, pak  $\frac{1}{\|z_1\|}z_1, \dots, \frac{1}{\|z_n\|}z_n$  je ortonormální. Důkaz:  $\|\frac{1}{\|z_i\|}z_i\| = \frac{1}{\|z_i\|}\|z_i\| = 1$ .

**Příklad 8.19.** V prostoru  $\mathbb{R}^n$  se standardním skalárním součinem je ortonormálním systémem například kanonická báze  $e_1, \dots, e_n$ . Speciálně v rovině  $\mathbb{R}^2$  tvoří normální bázi vektory  $(1, 0)^T$ ,  $(0, 1)^T$ . Jiný příklad normální báze v  $\mathbb{R}^2$  je například:  $\frac{\sqrt{2}}{2}(1, 1)^T$ ,  $\frac{\sqrt{2}}{2}(1, -1)^T$ .  $\square$

**Tvrzení 8.20.** Je-li systém vektorů  $z_1, \dots, z_n$  ortonormální, pak je lineárně nezávislý.

*Důkaz.* Uvažujme lineární kombinaci  $\sum_{i=1}^n \alpha_i z_i = o$ . Pak pro každé  $k = 1, \dots, n$  platí:

$$0 = \langle o, z_k \rangle = \left\langle \sum_{i=1}^n \alpha_i z_i, z_k \right\rangle = \sum_{i=1}^n \alpha_i \langle z_i, z_k \rangle = \alpha_k \langle z_k, z_k \rangle = \alpha_k. \quad \square$$

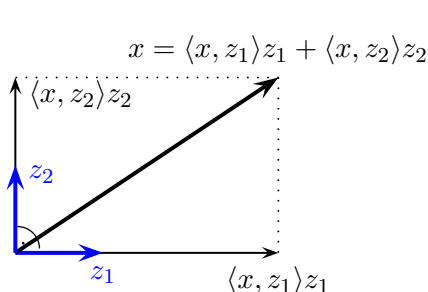
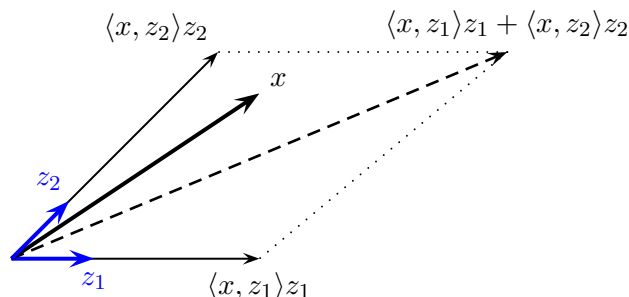
Ortonormalita vektorů tedy znamená jejich lineární nezávislost plus něco navíc – jejich kolmost. A právě tato vlastnost umožní některé problémy řešit efektivně. Zrovna následující věta říká, jak jednoduše spočítat souřadnice vůči bázi, která je ortonormální.

**Věta 8.21** (Fourierovy koeficienty). Buď  $z_1, \dots, z_n$  ortonormální báze prostoru  $V$ . Pak pro každé  $x \in V$  platí  $x = \sum_{i=1}^n \langle x, z_i \rangle z_i$ .

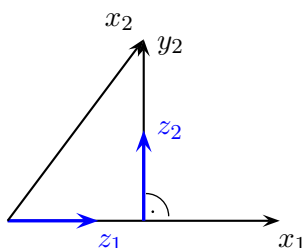
*Důkaz.* Víme, že  $x = \sum_{i=1}^n \alpha_i z_i$  a souřadnice  $\alpha_1, \dots, \alpha_n$  jsou jednoznačné (věta 5.31). Nyní pro každé  $k = 1, \dots, n$  platí:

$$\langle x, z_k \rangle = \left\langle \sum_{i=1}^n \alpha_i z_i, z_k \right\rangle = \sum_{i=1}^n \alpha_i \langle z_i, z_k \rangle = \alpha_k \langle z_k, z_k \rangle = \alpha_k. \quad \square$$

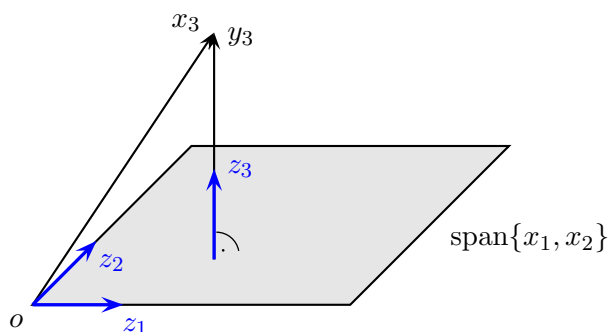
Vyjádření  $x = \sum_{i=1}^n \langle x, z_i \rangle z_i$  se nazývá *Fourierův rozvoj*, a skaláry  $\langle x, z_i \rangle$ ,  $i = 1, \dots, n$  se nazývají *Fourierovy koeficienty*<sup>2)</sup>. Geometrický význam Fourierova koeficientu  $\langle x, z_i \rangle$  je ten, že  $\langle x, z_i \rangle z_i$  udává projekci vektoru  $x$  na přímku  $\text{span}\{z_i\}$ . Jinými slovy,  $\langle x, z_i \rangle z_i$  je vektor na přímce se směrnici  $z_i$ , který je nejbližší vektoru  $x$ . Potom vektor  $x$  lze složit z těchto dílčích projekcí jednoduchým součtem  $x = \sum_{i=1}^n \langle x, z_i \rangle z_i$  (více o projekcích budeme hovořit v sekci 8.3). Jak ilustruje obrázek dole, pokud by báze  $z_1, \dots, z_n$  nebyla ortonormální, pak by tato vlastnost už obecně neplatila.

Vektory  $z_1, z_2$  ortonormální.Vektory  $z_1, z_2$  délky 1, ale ne ortogonální.

Jak sestavit ortonormální bázi nějakého prostoru? Následující procedura, Gramova–Schmidtova ortogonalizační metoda, začne s libovolnou bází a postupným nakolmíváním vektorů vytvoří bázi, která je ortonormální. Nakolmívání v kroku 2 funguje tak, že od vektoru  $x_k$  odečteme jeho projekci do prostoru generovaného vektory  $x_1, \dots, x_{k-1}$ ; tak bude kolmý na všechny předchozí. O projekci budeme pojednávat více v sekci 8.3.



Nakolmení druhého vektoru.



Nakolmení třetího vektoru.

**Algoritmus 8.22** (Gramova–Schmidtova ortogonalizace<sup>3)</sup>). Buďte  $x_1, \dots, x_n \in V$  lineárně nezávislé.

- 1: **for**  $k := 1$  **to**  $n$  **do**
- 2:      $y_k := x_k - \sum_{j=1}^{k-1} \langle x_k, z_j \rangle z_j$ , //vypočítáme kolmici
- 3:      $z_k := \frac{1}{\|y_k\|} y_k$ , //normalizujeme délku na 1
- 4: **end for**

Výstup:  $z_1, \dots, z_n$  ortonormální báze prostoru  $\text{span}\{x_1, \dots, x_n\}$ .

*Důkaz.* (Správnost Gramovy–Schmidtovy ortogonalizace.) Matematickou indukcí podle  $n$  dokážeme, že  $z_1, \dots, z_n$  je ortonormální báze prostoru  $\text{span}\{x_1, \dots, x_n\}$ . Pro  $n = 1$  je  $y_1 = x_1 \neq o$  a  $z_1 = \frac{1}{\|x_1\|} x_1$  je dobře definované a  $\text{span}\{x_1\} = \text{span}\{z_1\}$ .

<sup>2)</sup>Jean Baptiste Joseph Fourier (1768–1830), francouzský matematik a fyzik. Rozvoj použil kolem r. 1807 pro řešení problému vedení tepla v pevných látkách.

<sup>3)</sup>Metoda pochází od dánského finančního matematika Jørgen Pedersen Grama z r. 1883, explicitní vzorec publikoval r. 1907 německý matematik Erhard Schmidt. Jak už to bývá, nezávisle na nich a dříve objevili postup také P.S. Laplace (1816) nebo A.L. Cauchy (1836).

Indukční krok  $n \leftarrow n-1$ . Předpokládejme, že  $z_1, \dots, z_{n-1}$  je ortonormální báze prostoru  $\text{span}\{x_1, \dots, x_{n-1}\}$ . Kdyby bylo  $y_n = 0$ , tak  $x_n = \sum_{j=1}^{n-1} \langle x_n, z_j \rangle z_j$  a  $x_n \in \text{span}\{z_1, \dots, z_{n-1}\} = \text{span}\{x_1, \dots, x_{n-1}\}$ , což by byl spor s lineární nezávislostí vektorů  $x_1, \dots, x_n$ . Proto  $y_n \neq 0$  a  $z_n = \frac{1}{\|y_n\|} y_n$  je dobře definovaný a má jednotkovou normu.

Nyní dokážeme, že  $z_1, \dots, z_n$  je ortonormální systém. Z indukčního předpokladu je  $z_1, \dots, z_{n-1}$  ortonormální systém a proto  $\langle z_i, z_j \rangle$  je rovno 0 pro  $i \neq j$  a rovno 1 pro  $i = j$ . Stačí ukázat, že  $z_n$  je kolmé na ostatní  $z_i$  pro  $i < n$ :

$$\begin{aligned} \langle z_n, z_i \rangle &= \frac{1}{\|y_n\|} \langle y_n, z_i \rangle = \frac{1}{\|y_n\|} \left\langle x_n - \sum_{j=1}^{n-1} \langle x_n, z_j \rangle z_j, z_i \right\rangle = \\ &= \frac{1}{\|y_n\|} \langle x_n, z_i \rangle - \frac{1}{\|y_n\|} \sum_{j=1}^{n-1} \langle x_n, z_j \rangle \langle z_j, z_i \rangle = \frac{1}{\|y_n\|} \langle x_n, z_i \rangle - \frac{1}{\|y_n\|} \langle x_n, z_i \rangle = 0. \end{aligned}$$

Zbývá ověřit  $\text{span}\{z_1, \dots, z_n\} = \text{span}\{x_1, \dots, x_n\}$ . Z algoritmu je vidět, že  $z_n \in \text{span}\{z_1, \dots, z_{n-1}, x_n\} \subseteq \text{span}\{x_1, \dots, x_n\}$ , a tedy  $\text{span}\{z_1, \dots, z_n\} \subseteq \text{span}\{x_1, \dots, x_n\}$ . Protože oba prostory mají stejnou dimenzi, nastane rovnost (věta 5.46).  $\square$

Gramova–Schmidtova ortogonalizace má tu přednost, že je použitelná v každém prostoru se skalárním součinem. Speciálně při standardním skalárním součinu v  $\mathbb{R}^n$  můžeme ortogonalizaci vyjádřit maticově (viz poznámka 13.9), ale na druhou stranu v tomto případě existují i jiné metody, které mají lepší numerické vlastnosti; srov. sekce 13.3.

**Důsledek 8.23** (Existence ortonormální báze). *Každý konečně generovaný prostor (se skalárním součinem) má ortonormální bázi.*

*Důkaz.* Víme (věta 5.38), že každý konečně generovaný prostor má bázi, a tu můžeme Gramovou–Schmidtovou metodou zortogonalizovat.  $\square$

Poznamenejme, že pro nekonečně-dimenzionální prostory tvrzení věty neplatí – existují prostory se skalárním součinem, které nemají ortonormální bázi; viz Bečvář [2005].

**Důsledek 8.24** (Rozšíření ortonormálního systému na ortonormální bázi). *Každý ortonormální systém vektorů v konečně generovaném prostoru lze rozšířit na ortonormální bázi.*

*Důkaz.* Víme (věta 5.45), že každý ortonormální systém vektorů  $z_1, \dots, z_m$  lze rozšířit na bázi  $z_1, \dots, z_m, x_{m+1}, \dots, x_n$ , a tu můžeme Gramovou–Schmidtovou metodou zortogonalizovat na  $z_1, \dots, z_m, z_{m+1}, \dots, z_n$ . Pověšme si, že ortogonalizací se prvních  $m$  vektorů nezmění.  $\square$

Další užitečný vztah je Besselova nerovnost a Parsevalova rovnost.

**Věta 8.25.** *Bud'  $z_1, \dots, z_n$  ortonormální systém ve  $V$  a bud'  $x \in V$ . Pak platí:*

- (1) *Besselova nerovnost:  $\|x\|^2 \geq \sum_{j=1}^n |\langle x, z_j \rangle|^2$ ,*
- (2) *Parsevalova rovnost:  $\|x\|^2 = \sum_{j=1}^n |\langle x, z_j \rangle|^2$  právě tehdy, když  $x \in \text{span}\{z_1, \dots, z_n\}$ .*

*Důkaz.*

- (1) Vyplývá z úpravy

$$\begin{aligned} 0 &\leq \left\langle x - \sum_{j=1}^n \langle x, z_j \rangle z_j, x - \sum_{j=1}^n \langle x, z_j \rangle z_j \right\rangle = \\ &= \langle x, x \rangle - \sum_{j=1}^n \overline{\langle x, z_j \rangle} \langle x, z_j \rangle - \sum_{j=1}^n \langle x, z_j \rangle \langle z_j, x \rangle + \sum_{j=1}^n \langle x, z_j \rangle \overline{\langle x, z_j \rangle} = \\ &= \|x\|^2 - \sum_{j=1}^n |\langle x, z_j \rangle|^2. \end{aligned}$$



(2) Vyplývá z předchozího, neboť rovnost nastane právě tehdy, když  $x = \sum_{j=1}^n \langle x, z_j \rangle z_j$ .  $\square$

Parsevalova rovnost ukazuje, že pro vektory blízké počátku musí i jejich souřadnice být dostatečně malé. Dále, rovnost se dá zobecnit i pro nekonečně-dimenzionální prostory jako je  $\mathcal{C}_{[-\pi, \pi]}$ , což mj. znamená, že Fourierovy koeficienty v nekonečném rozvoji musí konvergovat k nule.

Parsevalova rovnost také jinými slovy říká, že v jakémkoli konečně generovaném prostoru  $V$  se norma libovolného  $x \in V$  dá vyjádřit jako standardní eukleidovská norma jeho vektoru souřadnic:  $\|x\| = \sqrt{[x]_B^T [x]_B}$ , kde  $B$  je ortonormální báze  $V$ . Jak ukážeme dole v tvrzení 8.26, tato vlastnost analogicky platí i pro skalární součin:  $\langle x, y \rangle = [x]_B^T [y]_B$  pro reálný a  $\langle x, y \rangle = [x]_B^T \overline{[y]_B}$  pro komplexní prostor.

**Tvrzení 8.26.** *Bud'  $B$  ortonormální báze prostoru  $V$  a bud'  $x, y \in V$ . Pak  $\langle x, y \rangle = [x]_B^T \overline{[y]_B}$ .*

*Důkaz.* Bud'  $B = \{z_1, \dots, z_n\}$ . Podle věty 8.21 je  $[x]_B = (\langle x, z_1 \rangle, \dots, \langle x, z_n \rangle)^T$ . Nyní

$$\begin{aligned} \langle x, y \rangle &= \langle \sum_{j=1}^n \langle x, z_j \rangle z_j, y \rangle = \sum_{j=1}^n \langle x, z_j \rangle \langle z_j, y \rangle \\ &= \sum_{j=1}^n \langle x, z_j \rangle \overline{\langle y, z_j \rangle} = [x]_B^T \overline{[y]_B}. \end{aligned} \quad \square$$

Není těžké nahlédnout, že tato věta platí i naopak. Čili dostáváme, že zobrazení  $\langle \cdot, \cdot \rangle$  je skalárním součinem na prostoru  $V$  právě tehdy, když se dá vyjádřit jako  $\langle x, y \rangle = [x]_B^T \overline{[y]_B}$  pro nějakou (či pro libovolnou) ortonormální bázi  $B$ . Každý skalární součin je tedy standardním skalárním součinem při pohledu z libovolné ortonormální báze.

### 8.3 Ortogonální doplněk a projekce

Ortogonální doplněk je užitečný pojem s názornou geometrickou interpretací. Ortogonální projekce je pak navíc ohromně důležitý nástroj, jehož použití v mnoha různých oborech překonává jeho základní geometrický význam.

**Definice 8.27** (Ortogonální doplněk). Bud'  $V$  vektorový prostor a  $M \subseteq V$ . Pak *ortogonální doplněk* množiny  $M$  je  $M^\perp := \{x \in V; \langle x, y \rangle = 0 \ \forall y \in M\}$ .

Ortogonální doplněk  $M^\perp$  tedy obsahuje takové vektory  $x$ , které jsou kolmé na všechny vektory z  $M$  (někdy zkráceně říkáme, že  $x$  je kolmé na  $M$ ).

**Příklad 8.28.** Ortogonální doplněk k vektoru  $(2, 5)^T$  je přímka  $\text{span}\{(5, -2)^T\}$ . Ortogonální doplněk k celé přímce  $\text{span}\{(2, 5)^T\}$  je rovněž přímka  $\text{span}\{(5, -2)^T\}$ .  $\square$

**Tvrzení 8.29** (Vlastnosti ortogonálního doplňku množiny). Bud'  $V$  vektorový prostor a  $M, N \subseteq V$ . Pak

- (1)  $M^\perp$  je podprostor  $V$ ,
- (2) je-li  $M \subseteq N$  pak  $M^\perp \supseteq N^\perp$ ,
- (3)  $M^\perp = \text{span}(M)^\perp$ .

*Důkaz.*

- (1) Ověříme vlastnosti podprostoru:  $0 \in M^\perp$  triviálně. Nyní buďte  $x_1, x_2 \in M^\perp$ . Pak  $\langle x_1, y \rangle = \langle x_2, y \rangle = 0 \ \forall y \in M$ , tedy i  $\langle x_1 + x_2, y \rangle = \langle x_1, y \rangle + \langle x_2, y \rangle = 0$ . Nakonec, buď  $x \in M^\perp$ , tedy  $\langle x, y \rangle = 0 \ \forall y \in M$ . Pak pro každý skalár  $\alpha$  je  $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle = 0$ .
- (2) Bud'  $x \in N^\perp$ , tedy  $\langle x, y \rangle = 0 \ \forall y \in N$ . Tím spíš  $\langle x, y \rangle = 0 \ \forall y \in M \subseteq N$ , a proto  $x \in M^\perp$ .
- (3)  $M \subseteq \text{span}(M)$ , tedy dle předchozího je  $M^\perp \supseteq \text{span}(M)^\perp$ . Důkaz druhé inkluze spočívá v tom, že je-li vektor  $x$  kolmý na určité vektory, pak je kolmý na jejich lineární kombinace, a tím pádem na jejich lineární obal. Formálně: buď  $x \in M^\perp$  tedy  $\langle x, y \rangle = 0 \ \forall y \in M$ . Speciálně,  $\langle x, y_i \rangle = 0$ , kde  $y_1, \dots, y_n \in M$  je báze  $\text{span}(M)$ . Pak pro libovolné  $y = \sum_{i=1}^n \alpha_i y_i \in \text{span}(M)$  jest  $\langle x, y \rangle = \langle x, \sum_{i=1}^n \alpha_i y_i \rangle = \sum_{i=1}^n \overline{\alpha_i} \langle x, y_i \rangle = 0$ .  $\square$

Vlastnost (3) říká, že ortogonální doplněk prostoru nebo jeho báze je ten samý. To ulehčí práci pro praktické hledání ortogonálního doplňku, protože stačí ověřit kolmost jen na bázecké vektory.

Zatímco předchozí věta se týkala ortogonálního doplňku libovolné množiny vektorů, nyní se zaměříme na ortogonální doplněk podprostoru. Povšimněme si, že důkaz první části je poměrně konstruktivní a dává návod jak ortogonální doplněk (resp. jeho bázi) spočítat.

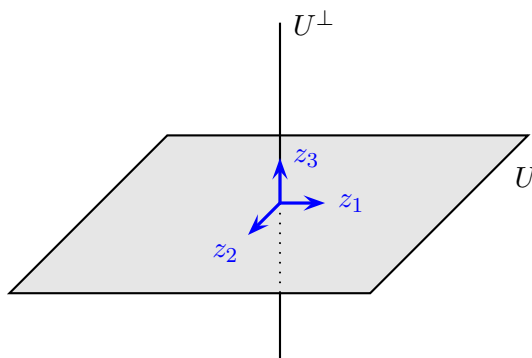
**Věta 8.30** (Vlastnosti ortogonálního doplňku podprostoru). *Bud'  $U$  podprostor vektorového prostoru  $V$ . Potom platí:*

- (1) *Je-li  $z_1, \dots, z_m$  ortonormální báze  $U$ , a je-li  $z_1, \dots, z_m, z_{m+1}, \dots, z_n$  její rozšíření na ortonormální bázi  $V$ , pak  $z_{m+1}, \dots, z_n$  je ortonormální báze  $U^\perp$ .*
- (2)  $\dim V = \dim U + \dim U^\perp$ ,
- (3)  $V = U + U^\perp$ ,
- (4)  $(U^\perp)^\perp = U$ ,
- (5)  $U \cap U^\perp = \{0\}$ .

*Důkaz.*

- (1) Zřejmě  $z_{m+1}, \dots, z_n$  je ortonormální systém v  $V$ , a tudíž stačí dokázat  $\text{span}\{z_{m+1}, \dots, z_n\} = U^\perp$ .  
Inkluze „ $\supseteq$ “. Každý  $x \in U^\perp$  má Fourierův rozvoj  $x = \sum_{i=1}^n \langle x, z_i \rangle z_i$ . Je-li  $x \in U^\perp$ , pak  $\langle x, z_i \rangle = 0$ ,  $i = 1, \dots, m$ , a tudíž  $x = \sum_{i=m+1}^n \langle x, z_i \rangle z_i \in \text{span}\{z_{m+1}, \dots, z_n\}$ .  
Inkluze „ $\subseteq$ “. Bud'  $x \in \text{span}\{z_{m+1}, \dots, z_n\}$ , pak  $x = \sum_{i=m+1}^n \langle x, z_i \rangle z_i = \sum_{i=1}^m 0 z_i + \sum_{i=m+1}^n \langle x, z_i \rangle z_i$ . Z jednoznačnosti souřadnic dostáváme  $\langle x, z_i \rangle = 0$ ,  $i = 1, \dots, m$ , a tím  $x \in U^\perp$ .
- (2) Z první vlastnosti máme  $\dim V = n$ ,  $\dim U = m$ ,  $\dim U^\perp = n - m$ .
- (3) Z první vlastnosti máme  $x = \underbrace{\sum_{i=1}^m \langle x, z_i \rangle z_i}_{\in U} + \underbrace{\sum_{i=m+1}^n \langle x, z_i \rangle z_i}_{\in U^\perp} \in U + U^\perp$ .
- (4) Z první vlastnosti je  $z_{m+1}, \dots, z_n$  ortonormální báze  $U^\perp$ , tedy  $z_1, \dots, z_m$  je ortonormální báze  $(U^\perp)^\perp$ .
- (5) Z předchozího a podle věty 5.52 o dimenzi spojení a průniku je  $\dim(U \cap U^\perp) = \dim V - \dim U - \dim U^\perp = 0$ .  $\square$

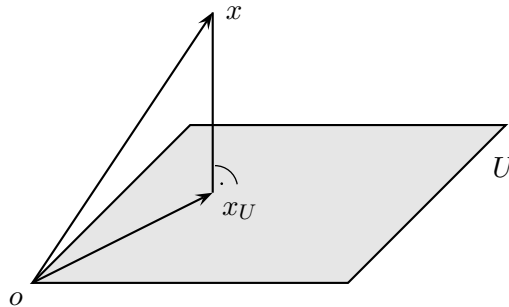
**Příklad 8.31.** Ilustrace podprostoru  $U$  a jeho ortogonálního doplňku  $U^\perp$ :



$\square$

Další z pěkných vlastností ortonormálních systémů je, že umožňují jednoduše spočítat projekci  $x_U$  vektoru  $x$  do podprostoru  $U$ , což je ten vektor z  $U$ , který je nejbližší k  $x$ .<sup>4)</sup> Jak ilustruje obrázek dole, a jak ukážeme formálně v následujících větách, projekce  $x_U$  je jednoznačná a je to takový vektor z  $U$ , pro který  $x - x_U$  je kolmé na  $U$ , tedy  $x - x_U \in U^\perp$ . Díky této kolmosti se mluví o *ortogonální projekci*. Vzhledem k tomu, že jiné typy projekcí uvažovat nebudeme, nadále používáme pouze pojem *projekce*.

<sup>4)</sup>Pro lepší názornost projekce v  $\mathbb{R}^n$  je možná lepší interpretovat vektory z  $\mathbb{R}^n$  jako body. Potom představa projekce bodu  $x \in \mathbb{R}^n$  do podprostoru  $U$  jako nejbližšího bodu  $x_U \in U$  k bodu  $x$  je intuitivně pochopitelnější.



**Definice 8.32** (Ortogonální projekce). Buď  $V$  vektorový prostor a  $U$  jeho podprostor. Pak *projekcí* vektoru  $x \in V$  rozumíme takový vektor  $x_U \in U$ , který splňuje

$$\|x - x_U\| = \min_{y \in U} \|x - y\|.$$

Následující věta ukazuje, že každý vektor má jednoznačnou projekci, a tím věta také opravňuje k zavedení projekce jakožto zobrazení  $V \rightarrow U$  definované  $x \mapsto x_U$ . Pojem projekce tedy budeme používat jak pro zobrazení  $x \mapsto x_U$ , tak pro obraz  $x_U$  konkrétního vektoru  $x$ .

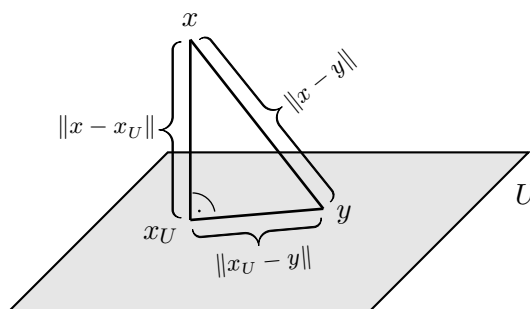
**Věta 8.33** (O ortogonální projekci). Buď  $U$  podprostor vektorového prostoru  $V$ . Pak pro každé  $x \in V$  existuje právě jedna projekce  $x_U \in U$  do podprostoru  $U$ . Navíc, je-li  $z_1, \dots, z_m$  ortonormální báze  $U$ , pak

$$x_U = \sum_{i=1}^m \langle x, z_i \rangle z_i. \quad (8.1)$$

*Důkaz.* Buď  $z_1, \dots, z_m, z_{m+1}, \dots, z_n$  rozšíření na ortonormální bázi  $V$ . Zdefinujme  $x_U := \sum_{i=1}^m \langle x, z_i \rangle z_i \in U$  a ukážeme, že je to hledaný vektor. Nyní

$$x - x_U = \sum_{i=1}^n \langle x, z_i \rangle z_i - \sum_{i=1}^m \langle x, z_i \rangle z_i = \sum_{i=m+1}^n \langle x, z_i \rangle z_i \in U^\perp. \quad (8.2)$$

Buď  $y \in U$  libovolné. Nyní máme  $x - x_U \in U^\perp$  a  $x_U - y \in U$ , viz obrázek:



Tudíž  $(x - x_U) \perp (x_U - y)$  a můžeme použít Pythagorovu větu, která dává

$$\|x - y\|^2 = \|(x - x_U) + (x_U - y)\|^2 = \|x - x_U\|^2 + \|x_U - y\|^2 \geq \|x - x_U\|^2,$$

neboli  $\|x - y\| \geq \|x - x_U\|$ , což dokazuje minimalitu. Abychom dokázali jednoznačnost, uvědomíme si, že rovnost nastane pouze tehdy, když  $\|x_U - y\|^2 = 0$ , čili když  $x_U = y$ .  $\square$

Pokud vektor  $x$  náleží do podprostoru  $U$ , pak jeho projekcí do  $U$  je on sám, a vzoreček (8.1) odpovídá Fourierově rozvoji z věty 8.21. Rovněž je snadné nahlédnout, že náleží-li vektor  $x$  do podprostoru  $U^\perp$ , pak jeho projekcí do  $U$  je  $o$ . Předpis (8.1) navíc ukazuje, že zobrazení  $x \mapsto x_U$ , které vektor  $x \in V$  zobrazuje na jeho projekci do podprostoru  $U$ , je lineárním zobrazením.

**Příklad 8.34** (Projekce na přímku). Buď  $a \in \mathbb{R}^n$  nenulový vektor a uvažujme projekci vektoru  $x \in \mathbb{R}^n$  na přímku se směrnici  $a$ , čili projekci do podprostoru  $U = \text{span}\{a\}$ . Ortonormální báze prostoru  $U$  je vektor  $z = \frac{1}{\|a\|}a$  a podle vzorce (8.1) má projekce vektoru  $x$  tvar

$$x_U = \langle x, z \rangle z = \frac{1}{\|a\|^2} \langle x, a \rangle a = \frac{x^T a}{a^T a} a. \quad \square$$

**Poznámka 8.35.** Vzhledem k vlastnostem (3) a (5) věty 8.30 se dá prostor  $V$  vyjádřit jako direktní součet podprostorů  $U$  a  $U^\perp$  (poznámka 5.54). To mj. znamená, že každý vektor  $v \in V$  má jednoznačné vyjádření  $v = u + u'$ , kde  $u \in U$  a  $u' \in U^\perp$ . Podle věty 8.33 je navíc vektor  $u$  projekcí vektoru  $v$  do  $U$ , a vektor  $u'$  projekcí  $v$  do  $U^\perp$ .

Víme ze vztahu (8.2) důkazu věty 8.33, že  $x - x_U \in U^\perp$ . Nyní nahlédneme, že tato vlastnost je nejenom nutnou, ale i postačující podmínkou pro to, aby  $x_U$  byla projekce vektoru  $x$ .

**Tvrzení 8.36.** Při značení z věty 8.33, pokud nějaké  $y \in U$  splňuje  $x - y \in U^\perp$ , pak  $y = x_U$ .

*Důkaz.* Protože  $(x - y) \perp (y - x_U)$ , použijeme Pythagorovu větu, která říká

$$\|x - x_U\|^2 = \|x - y\|^2 + \|y - x_U\|^2 \geq \|x - y\|^2.$$

Dostáváme  $\|x - x_U\| \geq \|x - y\|$ , tedy z vlastnosti a jednoznačnosti projekce musí  $y = x_U$ .  $\square$

**Příklad 8.37** (Legendreovy polynomy). Uvažujme prostor polynomů  $\mathcal{P}^n$ . Jaký pro něj zavést skalární součin? První nápad je využít isomorfismu s  $\mathbb{R}^{n+1}$  a pro polynomy  $p(x) = a_n x^n + \dots + a_0$ ,  $q(x) = b_n x^n + \dots + b_0$  zavést skalární součin jako

$$\langle p, q \rangle = \sum_{i=1}^n a_i b_i$$

a vektory  $1, x, x^2, \dots, x^n$  pak budou tvořit ortonormální systém. To je sice v pořádku, ale není to jediný možný způsob. Pokud si uvědomíme, že  $\mathcal{P}^n$  je podprostorem prostoru spojitých funkcí  $\mathcal{C}_{[a,b]}$ , tak můžeme na  $\mathcal{P}^n$  použít standardní skalární součin prostoru  $\mathcal{C}_{[a,b]}$ . Pokud zortogonalizujeme Gramovou–Schmidtovou metodou vektory  $1, x, x^2, \dots$  speciálně na  $\mathcal{C}_{[-1,1]}$ , pak dostaneme tzv. *Legendreovy polynomy*

$$p_0(x) = 1, \quad p_1(x) = x, \quad p_2(x) = \frac{1}{2}(3x^2 - 1), \quad p_3(x) = \frac{1}{5}(5x^3 - 3x), \quad \dots$$

Technické detaily výpočtu přeskočíme. Ani explicitní vyjádření není příliš jednoduché, neboť  $n$ -tý člen má tvar

$$p_n(x) = 2^{-n} \sum_{k=0}^n \binom{n}{k}^2 (x-1)^{n-k} (x+1)^k.$$

Tyto polynomy jsou na sebe kolmé, ale z důvodu určitých aplikací jsou znormovány tak, že  $n$ -tý polynom má normu  $2/(2n+1)$ .

Legendreovy polynomy můžeme použít třeba k aproximaci funkce polynomem, srov. metodu v sekci 3.6. Pokud funkci  $f$  chceme aproximovat polynomem  $n$ -tého stupně, tak spočítáme projekci  $f$  do podprostoru  $\mathcal{P}^n$  v tomto skalárním součinu. Projekci spočítáme podle věty 8.33 a za ortonormální bázi  $\mathcal{P}^n$  použijeme Legendreovy polynomy. Výsledná projekce má třeba tu vlastnost, že ze všech polynomů stupně  $n$  je nejbližší k  $f$  v normě indukované daným skalárním součinem, což zhruba odpovídá snaze minimalizovat plochu mezi  $f$  a polynomem. Historicky byly Legendreovy polynomy poprvé použity ve fyzice k vyjádření určitých operátorů a řešení diferenciálních rovnic.  $\square$

**Příklad 8.38** (Ortonormální systém v prostoru funkcí). V prostoru  $\mathcal{C}_{[-\pi,\pi]}$  existuje spočetný ortonormální systém  $z_1, z_2, \dots$  sestávající z vektorů

$$\frac{1}{\sqrt{2\pi}}, \quad \frac{1}{\sqrt{\pi}} \cos x, \quad \frac{1}{\sqrt{\pi}} \sin x, \quad \frac{1}{\sqrt{\pi}} \cos 2x, \quad \frac{1}{\sqrt{\pi}} \sin 2x, \quad \frac{1}{\sqrt{\pi}} \cos 3x, \quad \frac{1}{\sqrt{\pi}} \sin 3x, \quad \dots$$

I když to není báze v pravém slova smyslu, každou funkci  $f \in \mathcal{C}_{[-\pi, \pi]}$  lze vyjádřit jako nekonečnou řadu  $f(x) = \sum_{i=1}^{\infty} \langle f, z_i \rangle z_i$ . Poznámka: zde trochu zjednodušujeme a odbýváme pojem nekonečného součtu, ale pro intuitivní pochopení to snad postačuje.

Vyjádření několika prvních členů  $f(x) \approx \sum_{i=1}^k \langle f, z_i \rangle z_i$ , což je vlastně projekce do prostoru  $\text{span}\{z_1, \dots, z_k\}$  dimenze  $k$ , dává dobrou aproximaci funkce  $f(x)$ . Takovéto aproximace se používá hojně v teorii zpracování signálů (např. zvuku).

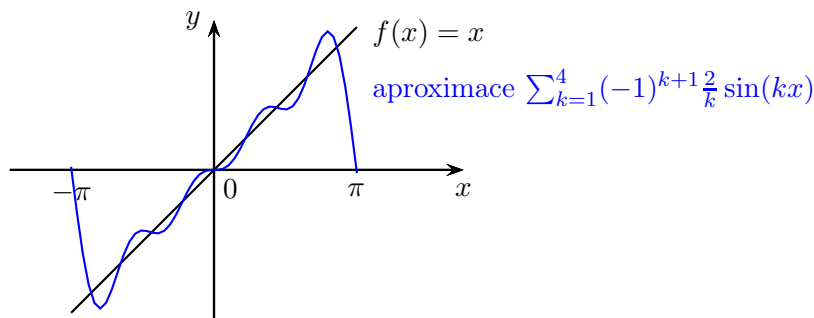
Konkrétně, spočítejme Fourierův rozvoj funkce  $f(x) = x$  na intervalu  $[-\pi, \pi]$

$$x = a_0 + \sum_{k=1}^{\infty} (a_k \sin(kx) + b_k \cos(kx)),$$

kde

$$\begin{aligned} a_0 &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) \cdot 1 \, dx = \frac{1}{2\pi} \int_{-\pi}^{\pi} x \, dx = 0, \\ a_k &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(kx) \, dx = \frac{1}{\pi} \int_{-\pi}^{\pi} x \sin(kx) \, dx = (-1)^{k+1} \frac{2}{k}, \\ b_k &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(kx) \, dx = \frac{1}{\pi} \int_{-\pi}^{\pi} x \cos(kx) \, dx = 0. \end{aligned}$$

Tedy  $x = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{2}{k} \sin(kx)$ .



□

Je možné spočítat projekci bez nutnosti mít ortonormální bázi podprostoru? Ano, pomocí řešení soustavy s tzv. Gramovou maticí.

**Věta 8.39** (Gramova matice). *Bud'  $U$  podprostor reálného vektorového prostoru  $V$ . Nechť  $U$  má bázi  $B = \{w_1, \dots, w_m\}$ . Označme jako Gramovu matici  $G \in \mathbb{R}^{m \times m}$  matici s prvky  $G_{ij} = \langle w_i, w_j \rangle$ . Pak  $G$  je regulární matice a vektor souřadnic  $s = [x_U]_B$  projekce  $x_U$  libovolného vektoru  $x \in V$  do podprostoru  $U$  je řešením soustavy*

$$Gs = (\langle w_1, x \rangle, \dots, \langle w_m, x \rangle)^T. \quad (8.3)$$

*Důkaz.* Pro důkaz regularity  $G$  bud'  $s \in \mathbb{R}^m$  řešením soustavy  $Gs = o$ . Pak  $i$ -tý řádek soustavy rovnic má tvar  $\sum_{j=1}^m G_{ij} s_j = \langle w_i, \sum_{j=1}^m s_j w_j \rangle = 0$ , čili  $\sum_{j=1}^m s_j w_j \in U^\perp \cap U = \{o\}$ . Z lineární nezávislosti  $w_1, \dots, w_m$  nutně  $s = o$ .

Víme, že  $x_U$  existuje a je jednoznačná a lze psát ve tvaru  $x_U = \sum_{j=1}^m \alpha_j w_j$  pro vhodné skaláry  $\alpha_j$ . Protože  $x - x_U \in U^\perp$ , dostáváme speciálně  $\langle w_i, x - x_U \rangle = 0$ , pro všechna  $i = 1, \dots, m$ . Dosazením za  $x_U$  získáme  $\langle w_i, x - \sum_{j=1}^m \alpha_j w_j \rangle = 0$ , neboli

$$\sum_{j=1}^m \alpha_j \langle w_i, w_j \rangle = \langle w_i, x \rangle, \quad i = 1, \dots, m.$$

Tedy  $s := [x_U]_B = (\alpha_1, \dots, \alpha_m)^T$  řeší soustavu (8.3). Z regularity  $G$  pak existuje pouze jediné řešení soustavy a odpovídá dané projekci. □

V důkazu jsme nahlédli, že Gramova matice  $G$  je regulární. Není těžké nahlédnout, že pokud by generátory  $w_1, \dots, w_m$  podprostoru  $U$  byly lineárně závislé, pak by matice  $G$  byla singulární. Tudíž  $G$  je regulární právě tehdy, když vektory  $w_1, \dots, w_m$  jsou lineárně nezávislé.

## 8.4 Ortogonální doplněk a projekce v $\mathbb{R}^n$

Z minulé sekce víme, jak počítat ortogonální doplněk a projekci pro libovolný konečně generovaný vektorový prostor se skalárním součinem, a to pomocí ortonormální báze. Nyní ukážeme, že v  $\mathbb{R}^n$  pro standardní skalární součin tyto transformace lze vyjádřit explicitně a přímo bez počítání ortonormální báze.

Následující věta říká, jak spočítat ortogonální doplněk libovolného podprostoru  $\mathbb{R}^n$ , známe-li jeho bázi nebo konečný systém generátorů (představují řádky matice  $A$ ).

**Věta 8.40** (Ortogonální doplněk v  $\mathbb{R}^n$ ). *Bud'  $A \in \mathbb{R}^{m \times n}$ . Pak  $\mathcal{R}(A)^\perp = \text{Ker}(A)$ .*

*Důkaz.* Z vlastností ortogonálního doplňku (tvrzení 8.29(3)) víme  $\mathcal{R}(A)^\perp = \{A_{1*}, \dots, A_{m*}\}^\perp$ . Tedy  $x \in \mathcal{R}(A)^\perp$  právě tehdy, když  $x$  je kolmé na řádky matice  $A$ , neboli  $A_{i*}x = 0$  pro všechna  $i = 1, \dots, m$ . Ekvivalentně,  $Ax = o$ , to jest  $x \in \text{Ker}(A)$ .  $\square$

**Příklad 8.41.** Bud'  $V$  prostor generovaný vektory  $(1, 2, 3)^T$  a  $(1, -1, 0)^T$ . Chceme-li určit  $V^\perp$ , tak sestavíme matici

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & -1 & 0 \end{pmatrix},$$

protože  $V = \mathcal{R}(A)$ . Nyní již stačí nalézt bázi  $V^\perp = \text{Ker}(A)$ , kterou tvoří např. vektor  $(1, 1, -1)^T$ .  $\square$

Charakterizace ortogonálního doplňku má i teoretické důsledky, např. vztah matice  $A$  a matice  $A^T A$ . Pozor, pro sloupcové prostory analogie neplatí!

**Důsledek 8.42.** *Bud'  $A \in \mathbb{R}^{m \times n}$ . Pak*

- (1)  $\text{Ker}(A^T A) = \text{Ker}(A)$ ,
- (2)  $\mathcal{R}(A^T A) = \mathcal{R}(A)$ ,
- (3)  $\text{rank}(A^T A) = \text{rank}(A)$ .

*Důkaz.*

- (1) Je-li  $x \in \text{Ker}(A)$ , pak  $Ax = o$ , a tedy také  $A^T Ax = A^T o = o$ , čímž  $x \in \text{Ker}(A^T A)$ . Naopak, je-li  $x \in \text{Ker}(A^T A)$ , pak  $A^T Ax = o$ . Pronásobením  $x^T$  dostaneme  $x^T A^T Ax = o$ , neboli  $\|Ax\|^2 = o$ . Z vlastnosti normy musí  $Ax = o$  a tudíž  $x \in \text{Ker}(A)$ .
- (2)  $\mathcal{R}(A^T A) = \text{Ker}(A^T A)^\perp = \text{Ker}(A)^\perp = \mathcal{R}(A)$ .
- (3) Triviálně z předchozího bodu.  $\square$

Nyní se podíváme na projekci vektoru  $x$  do podprostoru  $V$ , pro kterou odvodíme explicitní vzorec. Pokud vektory báze prostoru  $V$  dáme do sloupců matice  $A$ , tak projekci vektoru  $x$  do  $V$  lze formulovat jako projekci  $x$  do  $\mathcal{S}(A)$ .

**Věta 8.43** (Ortogonální projekce v  $\mathbb{R}^m$ ). *Bud'  $A \in \mathbb{R}^{m \times n}$  hodnosti  $n$ . Pak projekce vektoru  $x \in \mathbb{R}^m$  do sloupcového prostoru  $\mathcal{S}(A)$  je  $x' = A(A^T A)^{-1} A^T x$ .*

*Důkaz.* Nejprve si uvědomíme, že  $x'$  je dobře definované. Matice  $A^T A$  má dimenzi  $n$  (důsledek 8.42(3)), tedy je regulární a má inverzi. Podle tvrzení 8.36 stačí nyní ukázat, že  $x' \in \mathcal{S}(A)$  a  $x - x' \in \mathcal{S}(A)^\perp$ . První vlastnost platí, neboť  $x' = Az$  pro  $z = (A^T A)^{-1} A^T x$ . Pro druhou vlastnost stačí ověřit, že  $x - x' \in \mathcal{S}(A)^\perp = \mathcal{R}(A^T)^\perp = \text{Ker}(A^T)$ , a to plyne z vyjádření

$$A^T(x - x') = A^T(x - A(A^T A)^{-1} A^T x) = A^T x - A^T A(A^T A)^{-1} A^T x = A^T x - A^T x = o. \quad \square$$

Poznamenejme, že projekce je lineární zobrazení a podle předchozí věty je  $P := A(A^T A)^{-1} A^T$  jeho matice (vzhledem ke kanonické bázi). Navíc tato matice má pozoruhodné vlastnosti. Např. je symetrická,  $P^2 = P$  a regulární pouze tehdy, když  $m = n$ .

Speciálně, matice projekce na jednodimenzionální podprostor (přímku) má tvar  $P = a(a^T a)^{-1} a^T$ , kde  $a \in \mathbb{R}^n$  je směrnice přímky. Projekce vektoru  $x$  na přímku je pak vektor  $Px = a(a^T a)^{-1} a^T x = \frac{a^T x}{a^T a} a$  (srov. příklad 8.34). Pokud navíc směrnici normujeme tak, aby  $\|a\|_2 = 1$ , potom  $a^T a = 1$  a tudíž matice projekce získá jednoduchý tvar  $P = aa^T$ .

**Věta 8.44** (Ortogonalní projekce do doplňku). *Bud'  $P \in \mathbb{R}^{n \times n}$  matice projekce do podprostoru  $V \subseteq \mathbb{R}^n$ . Pak  $I - P$  je maticí projekce do  $V^\perp$ .*

*Důkaz.* Podle věty 8.30 lze každý vektor  $x \in \mathbb{R}^n$  jednoznačně rozložit na součet  $x = y + z$ , kde  $y \in V$  a  $z \in V^\perp$ . Z pohledu věty 8.33 je  $y$  projekce  $x$  do  $V$  a  $z$  projekce  $x$  do  $V^\perp$ . Tedy  $z = x - y = x - Px = (I - P)x$ .  $\square$

**Příklad 8.45** (Matice projekce do  $\text{Ker}(A)$ ). Věta 8.44 umožňuje elegantně vyjádřit projekci do jádra matice  $A \in \mathbb{R}^{m \times n}$ . Předpokládejme, že  $\text{rank}(A) = m$ . Protože  $\text{Ker}(A)^\perp = \mathcal{R}(A) = \mathcal{S}(A^T)$ , tak matice projekce do  $\text{Ker}(A)$  je dána předpisem  $I - A^T(AA^T)^{-1}A$ , kde  $A^T(AA^T)^{-1}A$  je matice projekce do  $\mathcal{S}(A^T)$ .  $\square$

**Poznámka 8.46** (Vzdálenosti podprostorů). Jedním z elegantních využití projekcí v geometrii je určení vzdálenosti afinních podprostorů – vzdálenost bodu od přímky, vzdálenost dvou přímek, vzdálenost bodu od roviny atp. Vzdáleností dvou afinních podprostorů  $U + a$ ,  $V + b$  pak rozumíme nejmenší vzdálenost  $\|x - y\|$ , kde  $x \in U + a$ ,  $y \in V + b$ . Bez důkazu poznamenáváme, že nejmenší vzdálenost se vždy nabyde.

Univerzální postup je následující. Mějme  $U + a$ ,  $V + b$  dva afinní podprostory prostoru  $\mathbb{R}^n$ , kde  $U = \text{span}\{u_1, \dots, u_m\}$  a  $V = \text{span}\{v_1, \dots, v_n\}$ . Necht' nejmenší vzdálenost se nabyde pro body  $x \in U + a$ ,  $y \in V + b$ ; tyto body jdou vyjádřit jako  $x = a + \sum_{i=1}^m \alpha_i u_i$ ,  $y = b + \sum_{j=1}^n \beta_j v_j$ . Vzdálenost těchto dvou bodů je stejná jako vzdálenost bodu  $a$  od bodu  $b + \sum_{j=1}^n \beta_j v_j - \sum_{i=1}^m \alpha_i u_i$ . Čili hledanou vzdálenost můžeme ekvivalentně vyjádřit jako vzdálenost bodu  $a$  od afinního podprostoru  $U + V + b$ . Posunutím ve směru  $-b$  pak vzdálenost spočítáme jako vzdálenost bodu  $a - b$  od podprostoru  $U + V = \text{span}\{u_1, \dots, u_m, v_1, \dots, v_n\}$ . To už je standardní úloha, kterou vyřešíme pomocí věty 8.33 resp. věty 8.43 jakožto vzdálenost bodu  $a - b$  od své projekce do prostoru  $U + V$ .

Uvažujme pro konkrétnost dvě přímky  $\text{span}\{u_1\} + a$ ,  $\text{span}\{v_1\} + b$ , kde  $u_1 = (1, 1, 4)^T$ ,  $a = (3, 3, 3)^T$ ,  $v_1 = (1, 0, 2)^T$ ,  $b = (1, 2, 6)^T$ . Jejich vzdálenost je tedy stejná jako vzdálenost bodu  $a - b = (2, 1, -3)^T$  od  $\text{span}\{u_1, v_1\}$ . Podle věty 8.43 je projekce bodu  $a - b$  na podprostor  $\text{span}\{u_1, v_1\}$  rovna  $(0, -1, -2)^T$ . Tudíž hledaná vzdálenost je  $\|(2, 1, -3)^T - (0, -1, -2)^T\| = \|(2, 2, -1)^T\| = 3$ .

## 8.5 Metoda nejmenších čtverců

Věta o projekci má široké použití nejenom v geometrii. Uvažujme soustavu  $Ax = b$ , která nemá řešení (typicky, když  $m$  je mnohem větší než  $n$ ). V tom případě bychom chtěli nějakou dobrou aproximaci, tj. takový vektor  $x$ , že levá a pravá strana jsou si co nejblíže. Formálně,

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

Tento přístup se studuje pro různé normy, ale pro eukleidovskou dostáváme

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

což je vzhledem k monotonii druhé mocniny ekvivalentní s úlohou

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2 = \min_{x \in \mathbb{R}^n} \sum_{j=1}^n (A_{j*}x - b_j)^2.$$

Odtud název *metoda nejmenších čtverců*. S využitím věty o projekci najdeme řešení snadno. Následující věta říká, že řešení metodou nejmenších čtverců jsou zároveň řešeními soustavy rovnic

$$A^T Ax = A^T b. \quad (8.4)$$

Tato soustava se nazývá *soustava normálních rovnic*. Zajímavé je, že tuto soustavu dostaneme z původní soustavy  $Ax = b$  pouhým přenásobením maticí  $A^T$ .

**Věta 8.47** (Množina řešení metodou nejmenších čtverců). *Bud'  $A \in \mathbb{R}^{m \times n}$ . Pak množina přibližných řešení soustavy  $Ax = b$  metodou nejmenších čtverců je neprázdná a rovna množině řešení normálních rovnic (8.4).*

*Důkaz.* Hledáme vlastně projekci vektoru  $b$  do podprostoru  $\mathcal{S}(A)$ , a tato projekce je vektor tvaru  $Ax$ , kde  $x \in \mathbb{R}^n$ . Podle tvrzení 8.36 je  $Ax$  projekcí právě tehdy, když  $Ax - b \in \mathcal{S}(A)^\perp = \text{Ker}(A^T)$ . Jinými slovy, musí platit  $A^T(Ax - b) = 0$ , neboli  $A^T Ax = A^T b$ . Tato soustava má řešení, protože projekce musí existovat.  $\square$

Poznamenejme, že pokud soustava  $Ax = b$  je řešitelná, potom každé její řešení je zároveň řešením metodou nejmenších čtverců, a naopak, každé řešení metodou nejmenších čtverců je skutečným řešením soustavy.

Jednoznačnost řešení nejmenších čtverců nastane, má-li matice  $A$  lineárně nezávislé sloupce. Pak totiž podle důsledku 8.42(3)) je matice  $A^T A$  regulární a soustava normálních rovnic tak má právě jedno řešení. To je typický případ.

**Důsledek 8.48.** *Bud'  $A \in \mathbb{R}^{m \times n}$  hodnosti  $n$ . Pak přibližné řešení soustavy  $Ax = b$  metodou nejmenších čtverců je  $x = (A^T A)^{-1} A^T b$ , a je jednoznačné.*

Metoda nejmenších čtverců<sup>5)</sup> má uplatnění v řadě oborů, zejména ve statistice při lineární regresi. Ta studuje chování a odhaduje budoucí vývoj různých veličin, např. globální teploty, HDP, ceny akcií či ropy v čase. Setkáme se s ní ale skoro ve všech vědních oborech. Např. ve fyzice tzv. Hookeův zákon říká, že natažení materiálu je přímo úměrné působící síle. Chceme-li odhadnout konstantu úměrnosti, provedeme velké množství pozorování a z nich vypočítáme hodnotu metodou nejmenších čtverců.

**Příklad 8.49** (Lineární regrese: vývoj světové populace). Data vývoje světové populace jsou následující:

rok	1950	1960	1970	1980	1990	2000
populace (mld.)	2,519	2,982	3,692	4,435	5,263	6,070

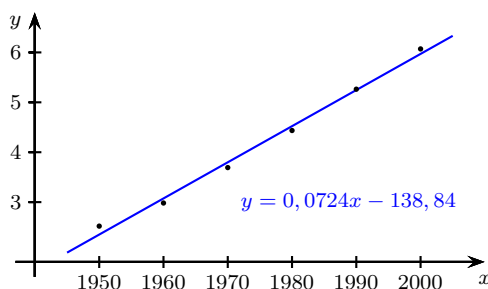
Chceme najít závislost velikosti populace na čase. Předpokládejme, že závislost je lineární. (To není vůbec samozřejmé, třeba Fibonacciho zjednodušený model růstu populace králíků byl exponenciální.) Obrázek dole, ilustrující data, v zásadě lineární závislost naznačuje. Lineární vztah popíšeme přímkou  $y = px + q$ , kde  $x$  je čas a  $y$  velikost populace. Neznámé parametry  $p, q$  vypočítáme. Po dosazení dat do rovnic by parametry  $p, q$  měly splňovat podmínky

$$2,519 = p \cdot 1950 + q$$

$$\vdots$$

$$6,070 = p \cdot 2000 + q$$

Přesné řešení neexistuje ale řešení metodou nejmenších čtverců je  $p = 0,0724$ ,  $q = -138,84$ . Grafické znázornění závislosti a zdrojový kód pro Matlab / Octave:



```
A = [1950 1960 1970
      1980 1990 2000;
      1 1 1 1 1 1]';
b = [2.519 2.982 3.692
      4.435 5.263 6.070]';
x=(A'*A)\(A'*b),
x'*[2009; 1]
```

Výslednou závislost lze využít pro predikce na další rok. Odhad pro rok 2010 je 6,6943 mld. obyvatel, ve skutečnost jich bylo 6,853 mld. Ovšem pozor, má smysl vytvářet pouze krátkodobé odhady – v roce 1900 určitě nebyla velikost populace záporná.  $\square$

<sup>5)</sup>Metoda nejmenších čtverců byla vyvinuta Gaussem kolem roku 1801 pro astronomická pozorování. Tehdy se objevil asteroid Ceres, aby vzápětí zase zmizel. Gauss metodou popsal jeho dráhu a předpověděl kdy se znovu objeví.



## 8.6 Ortogonální matice

Uvažujme lineární zobrazení v prostoru  $\mathbb{R}^n$ . Jaké toto zobrazení (potažmo jeho matice) musí být, aby nijak nedeformovalo geometrické objekty? Otočení kolem počátku či překlopení podle osy jsou příklady takových zobrazení, ale chtěli bychom je analyzovat podrobně. Ukážeme, že tato vlastnost souvisí s tzv. ortogonálními maticemi. Ty ale mají dalekosáhlejší význam. Protože mají dobré numerické vlastnosti (viz sekce 1.3 a 3.5.1), setkáváme se s nimi často v nejrůznějších numerických algoritmech.

I v této sekci uvažujeme standardní skalární součin v  $\mathbb{R}^n$  a eukleidovskou normu.

**Definice 8.50** (Ortogonální a unitární matice). Matice  $Q \in \mathbb{R}^{n \times n}$  je *ortogonální*, pokud  $Q^T Q = I_n$ . Matice  $Q \in \mathbb{C}^{n \times n}$  je *unitární*, pokud  $\overline{Q}^T Q = I_n$ .

Pojem unitární matice je zobecnění ortogonálních matic pro komplexní čísla. Nadále ale budeme vesměs pracovat jen s ortogonálními maticemi.

**Věta 8.51** (Charakterizace ortogonálních matic). *Bud'  $Q \in \mathbb{R}^{n \times n}$ . Pak následující jsou ekvivalentní:*

- (1)  $Q$  je ortogonální,
- (2)  $Q$  je regulární a  $Q^{-1} = Q^T$ ,
- (3)  $QQ^T = I_n$ ,
- (4)  $Q^T$  je ortogonální,
- (5)  $Q^{-1}$  existuje a je ortogonální,
- (6) sloupce  $Q$  tvoří ortonormální bázi  $\mathbb{R}^n$ ,
- (7) řádky  $Q$  tvoří ortonormální bázi  $\mathbb{R}^n$ .

*Důkaz.* Stručně. (1)–(5): Je-li  $Q$  ortogonální, pak  $Q^T Q = I$  a tedy  $Q^{-1} = Q^T$ ; podobně naopak. Dle vlastnosti inverze máme i  $QQ^T = I$ , neboli  $(Q^T)^T Q^T = I$ , tedy  $Q^T$  je ortogonální.

(6): Z rovnosti  $Q^T Q = I$  dostáváme porovnáním prvků na pozici  $i, j$ , že  $\langle Q_{*i}, Q_{*j} \rangle = 1$ , pokud  $i = j$ , a  $\langle Q_{*i}, Q_{*j} \rangle = 0$ , pokud  $i \neq j$ . Tedy sloupce  $Q$  tvoří ortonormální systém. Analogicky naopak.  $\square$

Vzhledem k vlastnosti (6) by se spíš slušelo říkat „ortonormální matice“, ale termín ortogonální matice je již zažitý.

**Tvrzení 8.52** (Součin ortogonálních matic). *Jsou-li  $Q_1, Q_2 \in \mathbb{R}^{n \times n}$  ortogonální, pak  $Q_1 Q_2$  je ortogonální.*

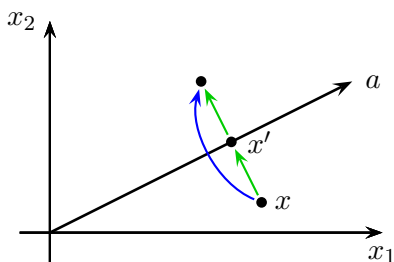
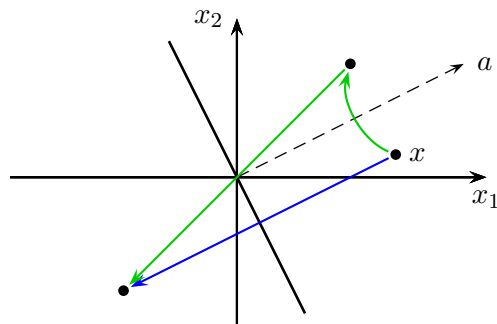
*Důkaz.*  $(Q_1 Q_2)^T Q_1 Q_2 = Q_2^T Q_1^T Q_1 Q_2 = Q_2^T Q_2 = I_n$ .  $\square$

**Příklad 8.53** (Příklady ortogonálních matic).

- Jednotková matice  $I_n$ , nebo k ní opačná  $-I_n$ .
- *Householderova matice*:  $H(a) := I_n - \frac{2}{a^T a} a a^T$ , kde  $o \neq a \in \mathbb{R}^n$ . Její geometrický význam je následující. Nechť  $x'$  je projekce bodu  $x$  na přímku  $\text{span}\{a\}$ , a uvažujme lineární zobrazení otočení bodu  $x$  dle přímky  $\text{span}\{a\}$  o úhel  $180^\circ$ . Pomocí věty 8.43 o projekci dostáváme, že bod  $x$  se zobrazí na vektor

$$x + 2(x' - x) = 2x' - x = 2a(a^T a)^{-1} a^T x - x = \left( 2 \frac{a a^T}{a^T a} - I \right) x.$$

Tedy matice otočení je  $\frac{2}{a^T a} a a^T - I$ . Uvažujme nyní zrcadlení dle nadroviny s normálou  $a$ . To můžeme reprezentovat jako otočení o  $180^\circ$  dle  $a$ , a pak překlopení dle počátku. Tedy matice tohoto zobrazení je  $I - 2 \frac{a a^T}{a^T a} = H(a)$ .

Otočení kolem přímky  $a$  o  $180^\circ$ .Zrcadlení dle nadroviny s normálou  $a$ .

Navíc se dá ukázat, že každou ortogonální matici řádu  $n$  lze rozložit jako součin nanejvýš  $n$  vhodných Householderových matic. Tudíž lineární zobrazení s ortogonální maticí geometricky reprezentuje složení nanejvýš  $n$  zrcadlení. Další vlastnosti Householderových matic podkryjeme v sekci 13.1.

- *Givensova matice*<sup>6)</sup>: Pro  $n = 2$  je to matice otočení o úhel  $\alpha$  proti směru hodinových ručiček

$$\begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix}.$$

Je to tedy matice tvaru  $\begin{pmatrix} c & -s \\ s & c \end{pmatrix}$ , kde  $c^2 + s^2 = 1$  a každá takováto matice odpovídá nějaké matici otočení. Obecně pro dimenzi  $n$  je to matice reprezentující otočení o úhel  $\alpha$  v rovině os  $x_i, x_j$ , tedy schematicky

$$G_{i,j}(c, s) = \begin{pmatrix} I & & & \\ & c & & -s \\ & & I & \\ & s & & c \\ & & & & I \end{pmatrix}$$

Také z Givensových matic lze složit každou ortogonální matici, ale je jich potřeba v součinu až  $\binom{n}{2}$  a případně navíc jedna diagonální matice s  $\pm 1$  na diagonále. Geometricky to znamená, že každé lineární zobrazení s ortogonální maticí reprezentuje složení nanejvýš  $\binom{n}{2}$  jednoduchých otočení a případně jedno zrcadlení ve směru souřadných os.  $\square$

**Věta 8.54** (Vlastnosti ortogonálních matic). *Bud'  $Q \in \mathbb{R}^{n \times n}$  ortogonální. Pak:*

- (1)  $\langle Qx, Qy \rangle = \langle x, y \rangle$  pro každé  $x, y \in \mathbb{R}^n$ ,
- (2)  $\|Qx\| = \|x\|$  pro každé  $x \in \mathbb{R}^n$ ,
- (3)  $|Q_{ij}| \leq 1$  a  $|Q_{ij}^{-1}| \leq 1$  pro každé  $i, j = 1, \dots, n$ ,
- (4)  $\begin{pmatrix} 1 & o^T \\ o & Q \end{pmatrix}$  je ortogonální matice.

*Důkaz.*

- (1)  $\langle Qx, Qy \rangle = (Qx)^T Qy = x^T Q^T Qy = x^T Iy = x^T y = \langle x, y \rangle$ .
- (2)  $\|Qx\| = \sqrt{\langle Qx, Qx \rangle} = \sqrt{\langle x, x \rangle} = \|x\|$ .
- (3) Vzhledem k vlastnosti (6) z věty 8.51 je  $\|Q_{*j}\| = 1$  pro každé  $j = 1, \dots, n$ . Tedy  $1 = \|Q_{*j}\|^2 = \sum_{i=1}^n q_{ij}^2$ , z čehož  $q_{ij}^2 \leq 1$ , a proto  $|q_{ij}| \leq 1$ . Matice  $Q^{-1}$  je ortogonální, takže pro ni tvrzení platí také.
- (4) Z definice  $\begin{pmatrix} 1 & o^T \\ o & Q \end{pmatrix}^T \begin{pmatrix} 1 & o^T \\ o & Q \end{pmatrix} = \begin{pmatrix} 1 & o^T \\ o & Q^T Q \end{pmatrix} = I_{n+1}$ .  $\square$

<sup>6)</sup>James Wallace Givens (1910–1993), Jr., americký matematik.

Díváme-li se na  $Q$  jako na matici příslušného lineárního zobrazení  $x \mapsto Qx$ , pak vlastnost (1) věty 8.54 říká, že při tomto zobrazení se zachovávají úhly, a vlastnost (2) zase říká, že se zachovávají délky. Tvzení platí i naopak: matice zobrazení zachovávající skalární součin musí být nutně ortogonální (srov. věta 8.55) a dokonce matice zobrazení zachovávající eukleidovskou normu musí být ortogonální [Horn and Johnson, 1985]. Vlastnost (3) je zase ceněná v numerické matematice, protože  $Q$  a  $Q^{-1}$  mají omezené velikosti složek. Nejdůležitější vlastností pro numerické počítání je (2), protože při násobení s ortogonální maticí prvky (a tedy i zaokrouhlovací chyby) nemají tendenci se zvětšovat.

Na závěr ukažme některá zobecnění výše zmíněných vlastností na libovolný skalární součin a libovolné lineární zobrazení.

**Věta 8.55** (Ortogonalní matice a lineární zobrazení). *Budť  $U, V$  prostory nad  $\mathbb{R}$  s libovolným skalárním součinem a  $f: U \rightarrow V$  lineární zobrazení. Necht'  $B_U$  resp.  $B_V$  je ortonormální báze  $U$  resp.  $V$ . Pak matice zobrazení  ${}_{B_V}[f]_{B_U}$  je ortogonální právě tehdy, když  $\langle f(x), f(y) \rangle = \langle x, y \rangle$  pro každé  $x, y \in U$ .*

*Důkaz.* Podle tvrzení 8.26 a vlastností matice zobrazení je

$$\begin{aligned}\langle x, y \rangle &= [x]_{B_U}^T \cdot [y]_{B_U}, \\ \langle f(x), f(y) \rangle &= [f(x)]_{B_V}^T \cdot [f(y)]_{B_V} = ({}_{B_V}[f]_{B_U} \cdot [x]_{B_U})^T \cdot {}_{B_V}[f]_{B_U} \cdot [y]_{B_U} = \\ &= [x]_{B_U}^T \cdot {}_{B_V}[f]_{B_U}^T \cdot {}_{B_V}[f]_{B_U} \cdot [y]_{B_U}.\end{aligned}$$

Tudíž, je-li  ${}_{B_V}[f]_{B_U}$  ortogonální, pak  $\langle f(x), f(y) \rangle = \langle x, y \rangle$ . Naopak, pokud rovnost  $\langle f(x), f(y) \rangle = \langle x, y \rangle$  platí pro každé  $x, y \in U$ , platí rovnost speciálně pro vektory, jejichž souřadnice jsou jednotkové vektory. Dosadíme-li za  $x$  a  $y$  konkrétně  $i$ -tý a  $j$ -tý vektor báze  $B_U$ , máme  $[x]_{B_U} = e_i$ ,  $[y]_{B_U} = e_j$ , a proto

$$\begin{aligned}(I_n)_{ij} &= e_i^T e_j = [x]_{B_U}^T \cdot [y]_{B_U} = \langle x, y \rangle = \langle f(x), f(y) \rangle = [x]_{B_U}^T \cdot {}_{B_V}[f]_{B_U}^T \cdot {}_{B_V}[f]_{B_U} \cdot [y]_{B_U} = \\ &= e_i^T \cdot {}_{B_V}[f]_{B_U}^T \cdot {}_{B_V}[f]_{B_U} \cdot e_j = ({}_{B_V}[f]_{B_U}^T \cdot {}_{B_V}[f]_{B_U})_{ij}.\end{aligned}$$

Tímto po složkách dostáváme rovnost  $I_n = {}_{B_V}[f]_{B_U}^T \cdot {}_{B_V}[f]_{B_U}$ . □

**Tvrzení 8.56** (Ortogonalní matice a matice přechodu). *Bud'  $V$  prostor nad  $\mathbb{R}$  s libovolným skalárním součinem a  $B_1, B_2$  dvě jeho báze. Jakékoli dvě z následujících vlastností implikují tu třetí:*

- (1)  $B_1$  je ortonormální báze,
- (2)  $B_2$  je ortonormální báze,
- (3)  ${}_{B_2}[id]_{B_1}$  je ortogonální matice.

*Důkaz.*

Implikace „(1), (2)  $\Rightarrow$  (3)“. Plyne z věty 8.55, neboť identita zachovává skalární součin.

Implikace „(2), (3)  $\Rightarrow$  (1)“. Bud'  $B_1 = \{x_1, \dots, x_n\}$ . Z definice pak sloupce  ${}_{B_2}[id]_{B_1}$  tvoří vektory  $[x_i]_{B_2}$ , které jsou (díky ortogonalitě matice přechodu) ortonormální při standardním skalárním součinu v  $\mathbb{R}^n$ . Podle tvrzení 8.26 pak  $\langle x_i, x_j \rangle = [x_i]_{B_2}^T [x_j]_{B_2}$ , což je 0 pro  $i \neq j$  a 1 jinak.

Implikace „(3), (1)  $\Rightarrow$  (2)“. Platí z předchozího ze symetrie, neboť  ${}_{B_1}[id]_{B_2} = {}_{B_2}[id]_{B_1}^{-1}$ . □

## Problémy

- 8.1. Stopa matice  $A \in \mathbb{R}^{n \times n}$  je číslo  $\text{trace}(A) = \sum_{i=1}^n a_{ii}$ . Ukažte, že  $\langle A, B \rangle := \text{trace}(A^T B)$  je skalární součin na prostoru matic  $\mathbb{R}^{m \times n}$ .
- 8.2. Dokažte, že následující tvrzení platí pro reálný vektorový prostor, ale pro komplexní už obecně ne: Vektory  $x, y \in V$  jsou kolmé právě tehdy, když  $\|x + y\|^2 = \|x\|^2 + \|y\|^2$ .
- 8.3. Rozhodněte, pro které vektory se Cauchyho–Schwarzova nerovnost nabyde jako rovnost a pro které jako ostrá nerovnost.
- 8.4. Porovnejte velikost (normu) vektoru  $x \in V$  s velikostí jeho projekce do podprostoru  $U \subseteq V$ .
- 8.5. Určete vzdálenost  $c \in \mathbb{R}^n$  od

- (a) nadroviny  $a^T x = b$ , kde  $o \neq a \in \mathbb{R}^n$  a  $b \in \mathbb{R}$ .
  - (b) přímky  $p = b + \text{span}\{a\}$ ,  $a \neq o$ .
- 8.6. Ukažte, že projekce vektoru  $x$  do podprostoru  $U$  se dá vyjádřit jako jednoznačný bod v průniku podprostoru  $U$  a afinního podprostoru  $U^\perp + x$ .
- 8.7. Buď  $P$  matice projekce do  $U \subseteq \mathbb{R}^n$ . Dokažte, že  $\text{rank}(P) = \text{trace}(P)$ .
- 8.8. Odvoďte vzorečky pro ortogonální projekci v  $\mathbb{R}^n$  (věta 8.43) a pro metodu nejmenších čtverců (důsledek 8.48) pomocí Gramovy matice (věta 8.39).
- 8.9. Může být součet ortogonálních matic zase ortogonální matice?
- 8.10. Buďte  $U, V$  prostory nad  $\mathbb{R}$  se skalárním součinem a  $f: U \rightarrow V$  zobrazení zachovávající skalární součin, tj.  $\langle f(x), f(y) \rangle = \langle x, y \rangle$  pro všechna  $x, y \in U$ . Ukažte, že  $f$  je lineární a prosté.

### Shrnutí ke kapitole 8. Skalární součin

Skalární součin zavádí speciální součin dvou vektorů, kdy výsledkem je skalár. Máme-li vektorový prostor vybaven skalárním součinem, pak tento skalární součin přirozeně na prostoru definuje také normu, tedy velikost vektoru. A norma pak definuje vzdálenost vektorů jako normu jejich rozdílu. Oba pojmy potřebujeme k tomu, abychom byli schopni měřit v prostoru, ale taky třeba vyjádřit, že posloupnost vektorů konverguje.

Skalární součin dále přirozeně zavádí kolmost vektorů. Ortonormální báze je báze složená z vektorů velikosti 1 a navzájem kolmých. S takovou bází se pak jednoduše počítají souřadnice, projekce aj. Ortonormální bázi umíme sestavit Gramovou–Schmidtovou ortogonalizační metodou. Ačkoli jsme pojem skalárního součinu definovali abstraktně, ukázalo se, že každý skalární součin má podobu standardního skalárního součinu v souřadném systému (libovolné) ortonormální báze.

Ortogonalní projekce je zobrazení, které vektor zobrazí na jemu nejbližší v daném podprostoru. Přímka, vedená od vektoru k jeho projekci, musí být kolmá na podprostoru (odtud „ortogonalní“ projekce). Projekce se snadno spočítá, pokud známe ortonormální bázi podprostoru. V opačném případě použijeme maticový vzoreček. Projekce je velmi užitečný nástroj nejen v geometrii, kde nám umožňuje elegantně vyjádřit vzdálenosti různých objektů. Jako negeometrickou aplikaci jsme uvedli metodu nejmenších čtverců, počítající nejlepší přibližné řešení přeuročené soustavy rovnic.

Algebraicky jsou ortogonální matice takové matice, jejichž inverzní matice se jednoduše vyjádří jako transpozice. Ortogonální matice pak geometricky reprezentují lineární zobrazení, které nedeformují objekty – zachovávají úhly i vzdálenosti. Jedná se tedy o rotace a zrcadlení a jejich složení. Tato geometrická podstata se odráží i v numerických vlastnostech – počítání s ortogonálními maticemi je výhodné, protože zaokrouhlovací chyby se tolik neamplifikují.