15 - OPT – Optimalizace

Použití lineární algebry v optimalizaci: lineární podprostory a zobrazení, ortogonální projekce na podprostor, metoda nejmenších čtverců, spektrální rozklad a kvadratické funkce, prokládání bodů podprostorem. https://cw.fel.cvut.cz/b201/courses/b0b33opt/start

Lineární podprostory a zobrazení:

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

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

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

$$\alpha_1 \mathbf{x}_1 + \cdots + \alpha_k \mathbf{x}_k = \mathbf{0} \implies \alpha_1 = \cdots = \alpha_k = 0.$$
 (3.2)

V opačném případě jsou vektory $\mathbf{x}_1, \dots, \mathbf{x}_k$ lineárně závislé. Pokud jsou vektory lineárně závislé, tak je aspoň jeden z nich lineární kombinací ostatních. To je jasné: je-li např. $\alpha_1 \neq 0$, pak $\alpha_1 \mathbf{x}_1 + \dots + \alpha_k \mathbf{x}_k = \mathbf{0}$ lze napsat jako $\mathbf{x}_1 = \alpha_2' \mathbf{x}_2 + \dots + \alpha_k' \mathbf{x}_k$ kde $\alpha_i' = \alpha_i/\alpha_1$.

Lineární obal vektorů $x_1, ..., x_k$ je množina

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

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):

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

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

Báze lineárního podprostoru¹ $X \subseteq \mathbb{R}^n$ je lineárně nezávislá množina vektorů, jejíž lineární obal je X. Platí následující netriviální tvrzení (důkazy najdete v učebnicích lineární algebry):

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 bázi².

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 $\{(1,1,1),(1,-1,0),(2,0,0)\}$.

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

Lineární zobrazení

Zobrazení $\mathbf{f} \colon \mathbb{R}^n \to \mathbb{R}^m$ je lineární, jestliže

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

pro všechna $\mathbf{x}_1, \dots, \mathbf{x}_k \in \mathbb{R}^n$ a $\alpha_1, \dots, \alpha_k \in \mathbb{R}$.

Věta

Zobrazení $\mathbf{f} \colon \mathbb{R}^n \to \mathbb{R}^m$ je lineární, právě když

$$f(x) = Ax$$

pro nějakou matici $\mathbf{A} \in \mathbb{R}^{m \times n}$. Matice \mathbf{A} je zobrazením \mathbf{f} určena jednoznačně.

Tvrzení: Matice složeného zobrazení je součinem matic jednotlivých zobrazení:

Pro
$$f(x) = Ax a g(y) = By je$$

$$(g \circ f)(x) = g(f(x)) = B(Ax) = (BA)x = BAx$$

Pomocí sloupců a_1, \ldots, a_n matice **A**:

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

Vektor $\mathbf{A}\mathbf{x}$ je lineární kombinace sloupců \mathbf{A} s koeficienty x_i .

Pomocí řádků $\mathbf{a}_1^T, \dots, \mathbf{a}_m^T$ matice \mathbf{A} :

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

Složky vektoru **Ax** jsou skalární součiny řádků **A** a vektoru **x**.

Prostor obrazů matice

$$\operatorname{\mathsf{rng}} \mathbf{A} = \{ \ \mathbf{A}\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n \ \} \quad \subseteq \mathbb{R}^m$$

Interpretace:

- Obor hodnot (range, image) lineárního zobrazení $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x}$ (rng $\mathbf{A} = \mathbf{f}(\mathbb{R}^n)$)
- Množina všech vektorů y takových, že Ax = y má řešení
- Lineární obal sloupců matice A

Hodnost matice A je číslo

$$rank \mathbf{A} = dim rng \mathbf{A}$$
.

Věta

 $\operatorname{rank} \mathbf{A} = \operatorname{rank} \mathbf{A}^T$

Nulový prostor matice

$$\mathsf{null}\,\mathbf{A} = \{\; \mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} = \mathbf{0} \;\} \quad \subseteq \mathbb{R}^n$$

Interpretace:

- Množina vektorů, které se zobrazí na nulový vektor
- Množina řešení soustavy lineárních rovnic $\mathbf{A}\mathbf{x}=\mathbf{0}$
- Množina všech vektorů kolmých na každý řádek matice A

Pro matici $\mathbf{A} \in \mathbb{R}^{m \times n}$ jsou tvrzení pod sebou ekvivalentní:

Prostor obrazů

- $\operatorname{rng} \mathbf{A} = \mathbb{R}^m$
- Ax = y má řešení pro každé y
- rank **A** = *m*
- zobrazení f(x) = Ax je surjektivní
- A má lin. nezávislé řádky
- A má pravou inverzi
- AA^T je regulární

Nulový prostor

- null $\mathbf{A} = \{\mathbf{0}\}$
- Ax = 0 má jediné řešení x = 0
- rank $\mathbf{A} = n$
- zobrazení f(x) = Ax je injektivní
- A má lin. nezávislé sloupce
- A má levou inverzi
- A^TA je regulární

Pro každou matici
$$\mathbf{A} \in \mathbb{R}^{m \times n}$$
 platí
$$\underbrace{\dim \operatorname{rng} \mathbf{A}}_{\operatorname{rank} \mathbf{A}} + \dim \operatorname{null} \mathbf{A} = n.$$

Afinní kombinace a podprostory

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

$$\alpha_1 \mathbf{x}_1 + \dots + \alpha_k \mathbf{x}_k$$
 kde $\alpha_1 + \dots + \alpha_k = 1$

Afinní obal

$$\mathsf{aff}\{\mathbf{x}_1,\ldots,\mathbf{x}_k\}$$

vektorů $\mathbf{x}_1,\ldots,\mathbf{x}_k\in\mathbb{R}^n$ je množina všech jejich afinních kombinací.

• Množina $A \subseteq \mathbb{R}^n$ je afinní podprostor, jestliže je uzavřená na afinní kombinace.

Pro $X \subseteq \mathbb{R}^n$ a $\mathbf{x} \in \mathbb{R}^n$ označíme $X + \mathbf{x} = \mathbf{x} + X = \{ \mathbf{x} + \mathbf{y} \mid \mathbf{y} \in X \}.$

Věta

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

Dimenze afinního podprostoru A definována jako dim X.

Věta

Množina $A \subseteq \mathbb{R}^n$ je afinní podprostor právě tehdy, když je množinou řešení nějaké lineární soustavy, tj. $A = \{ \mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} = \mathbf{b} \}$.

Příklady afinních podprostorů:

- bod (dimenze 0) $\mathbf{x} \in \mathbb{R}^n$
- přímka (dimenze 1) $\{ \mathbf{x} + \alpha \mathbf{s} \mid \alpha \in \mathbb{R} \}$
- rovina (dimenze 2) { $\mathbf{x} + \alpha \mathbf{s} + \beta \mathbf{t} \mid \alpha, \beta \in \mathbb{R}$ }
- nadrovina (dimenze n-1) { $\mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}^T \mathbf{x} = b$ }

Lineární	Afinní	
lineární kombinace	afinní kombinace	
lineární obal	afinní obal	
lineární podprostor X	afinní podprostor $X + \mathbf{x}_0$	
řešení soustavy $\mathbf{A}\mathbf{x}=0$	řešení soustavy $\mathbf{A}\mathbf{x} = \mathbf{b}$	
null A	$\mathbf{x}_0 + null\mathbf{A}$	
lineární zobrazení A x	afinní zobrazení $\mathbf{A}\mathbf{x} + \mathbf{b}$	

Pro vektory $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$:

- Standardní skalární součin: $\mathbf{x}^T \mathbf{y} = x_1 y_1 + \cdots + x_n y_n$
- Eukleidovská norma (délka vektoru): $\|\mathbf{x}\| = (\mathbf{x}^T\mathbf{x})^{1/2} = (x_1^2 + \dots + x_n^2)^{1/2}$
- Úhel φ mezi vektory: $\mathbf{x}^T \mathbf{y} = \|\mathbf{x}\| \|\mathbf{y}\| \cos \varphi$

Jestliže $\mathbf{x}^T \mathbf{y} = 0$, vektory \mathbf{x}, \mathbf{y} jsou ortogonální, což značíme také $\mathbf{x} \perp \mathbf{y}$ (nevylučuje možnost $\mathbf{x} = \mathbf{0}$ nebo $\mathbf{y} = \mathbf{0}$).

- Eukleidovská metrika (vzdálenost bodů): $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} \mathbf{y}\|$
- Vektor $\mathbf{u} \in \mathbb{R}^m$ je normalizovaný, jestliže $\|\mathbf{u}\| = 1 = \mathbf{u}^T \mathbf{u}$.
- Množina vektorů $\{\mathbf{u}_1,\ldots,\mathbf{u}_n\}$ je ortonormální, jestliže

$$\mathbf{u}_i^T \mathbf{u}_j = \begin{cases} 0 & \text{pro } i \neq j \\ 1 & \text{pro } i = j \end{cases}$$

Věta

Jestliže matice **U** je čtvercová, pak tato tvrzení jsou ekvivalentní:

- $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ (tj. \mathbf{U} má ortonormální sloupce)
- $\mathbf{U}\mathbf{U}^T = \mathbf{I}$ (tj. \mathbf{U} má ortonormální řádky)
- \mathbf{U} je regulární a splňuje $\mathbf{U}^T = \mathbf{U}^{-1}$

Čtvercová matice s ortonormálními sloupci se nazývá ortogonální matice.

Ortogonální podprostory

- Vektor $\mathbf{y} \in \mathbb{R}^n$ je ortogonální na podprostor $X \subseteq \mathbb{R}^n$ (značíme $\mathbf{y} \perp X$), jestliže $\mathbf{y} \perp \mathbf{x}$ pro všechna $\mathbf{x} \in X$.
- Podprostory $X, Y \subseteq \mathbb{R}^n$ jsou ortogonální (značíme $X \perp Y$), jestliže $\mathbf{x} \perp \mathbf{y}$ pro všechna $\mathbf{x} \in X$ a $\mathbf{y} \in Y$.

Tvrzení: $X \perp Y \Rightarrow X \cap Y = \{\mathbf{0}\}\$

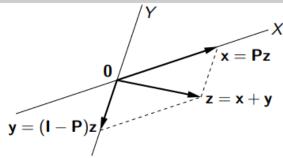
• Ortogonální doplněk podprostoru $X \subseteq \mathbb{R}^n$ je množina

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

všech vektorů ortogonálních na X.

Tvrzení: X^{\perp} je podprostor.

Obecná projekce



Vektor \mathbf{x} se nazývá projekce vektoru \mathbf{z} na podprostor X podél podprostoru Y. Projekce je tedy transformace $\mathbf{f} \colon \mathbb{R}^n \to \mathbb{R}^n$ taková, že $\mathbf{x} = \mathbf{f}(\mathbf{z})$.

Věta

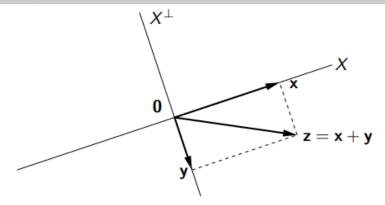
Projekce je lineární transformace.

Důkaz: Nechť $\mathbf{x} \in X$, $\mathbf{y} \in Y$ a $\mathbf{z} = \mathbf{x} + \mathbf{y}$. Pak $\alpha \mathbf{x} \in X$, $\alpha \mathbf{y} \in Y$ a $\alpha \mathbf{z} = \alpha \mathbf{x} + \alpha \mathbf{y}$. Díky jednoznačnosti jiný rozklad není. Proto $\mathbf{f}(\alpha \mathbf{z}) = \alpha \mathbf{x}$. Podobně dokážeme $\mathbf{f}(\mathbf{z} + \mathbf{z}') = \mathbf{f}(\mathbf{z}) + \mathbf{f}(\mathbf{z}')$.

Tedy existuje projektor $\mathbf{P} \in \mathbb{R}^{n \times n}$ tak, že $\mathbf{f}(\mathbf{z}) = \mathbf{P}\mathbf{z}$. Vlastnosti projektoru:

- $\operatorname{rng} \mathbf{P} = X$
- $\text{null } \mathbf{P} = Y$
- $\mathbf{P}^2 = \mathbf{P}$ (charakterizuje projektory)

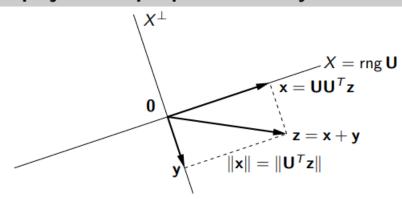
Ortogonální projekce



Je-li X podprostor, pak X^{\perp} je k němu komplementární (viz slide 26).

- Tedy pro každé **z** jednoznačně existují $\mathbf{x} \in X$ a $\mathbf{y} \in X^{\perp}$ tak, že $\mathbf{z} = \mathbf{x} + \mathbf{y}$.
- Projekce na X podél X^{\perp} se nazývá ortogonální projekce na X.
- Tedy $x \in X$ je ortogonální projekce vektoru z na X, jestliže $(z x) \perp X$.

Ortogonální projekce na podprostor zadaný ortonormální bází



Věta

Nechť $\mathbf{U}^T\mathbf{U} = \mathbf{I}$. Ortogonální projekce vektoru \mathbf{z} na podprostor $X = \operatorname{rng} \mathbf{U}$ je $\mathbf{x} = \mathbf{U}\mathbf{U}^T\mathbf{z}$.

Důkaz: Máme dokázat, že $x \in rng U$ a $(z - x) \perp rng U$.

- $UU^Tz \in rng U$ platí z definice rng.
- $(\mathbf{z} \mathbf{U}\mathbf{U}^T\mathbf{z}) \perp \operatorname{rng} \mathbf{U}$ znamená $\mathbf{U}^T(\mathbf{z} \mathbf{U}\mathbf{U}^T\mathbf{z}) = \mathbf{0}$, což platí neboť $\mathbf{U}^T\mathbf{U} = \mathbf{I}$.

Matice $P = UU^T$ je ortogonální projektor na podprostor X.

Tvrzení: Splňuje navíc (kromě vlastností obecného projektoru) $\mathbf{P} = \mathbf{P}^T$.

Nejbližší bod k podprostoru

Hledáme bod x na podprostoru X, který je nejbližší danému bodu z. Tedy

$$\mathbf{x} \in \operatorname*{argmin}_{\mathbf{x}' \in X} \|\mathbf{z} - \mathbf{x}'\|$$

Věta

Bod **x** je nejbližší bodu **z** na podprostoru X, právě když $(\mathbf{z} - \mathbf{x}) \perp X$.

Důkaz směru ←: z Pythagorovy věty.

Délka ort. projekce **z** na X je rovna vzdálenosti bodu **z** od X^{\perp} :

$$\|\mathbf{U}\mathbf{U}^T\mathbf{z}\| = \|\mathbf{U}^T\mathbf{z}\|$$

Metoda nejmenších Čtverců:

Na tohle téma je to celá přednáška důležit. Info: odkaz

Řešitelnost lineárních soustav

Soustava m lineárních rovnic o n neznámých ($\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$)

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

- pro $\mathbf{b} = \mathbf{0}$ je homogenní, pro $\mathbf{b} \neq \mathbf{0}$ je nehomogenní.
- má řešení, právě když $\mathbf{b} \in \operatorname{rng} \mathbf{A}$, tj. rank $[\mathbf{A} \ \mathbf{b}] = \operatorname{rank} \mathbf{A}$ (Frobeniova věta)
- ullet množina řešení je afinní podprostor \mathbb{R}^n

Tři případy:

- nemá řešení (b ∉ rng A): přeurčená soustava
- právě jedno řešení $(\mathbf{b} \in \operatorname{rng} \mathbf{A}, \operatorname{rank} \mathbf{A} = n)$
- nekonečně mnoho řešení ($\mathbf{b} \in \operatorname{rng} \mathbf{A}$, rank $\mathbf{A} < n$): nedourčená soustava

Úloha nejmenších čtverců

Přibližné řešení přeurčené soustavy $\mathbf{A}\mathbf{x} = \mathbf{b}$: minimalizuj součet

$$\|\mathbf{r}\|^2 = \mathbf{r}^T \mathbf{r} = r_1^2 + \dots + r_m^2$$

čtverců reziduí $(r_1,\ldots,r_m)=\mathbf{r}=\mathbf{b}-\mathbf{A}\mathbf{x}$. Tedy

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

Příklad: Pro soustavu

$$x + 2y = 6$$

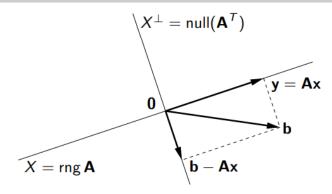
$$-x + y = 3$$

$$x + y = 4$$

hledáme přibližné řešení $x,y\in\mathbb{R}$, které minimalizuje

$$(x+2y-6)^2+(-x+y-3)^2+(x+y-4)^2$$
.

Řešení úlohy nejmenších čtverců



Věta

Vektor x je optimální řešení úlohy

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 = \min_{\mathbf{y} \in \operatorname{rng} \mathbf{A}} \|\mathbf{y} - \mathbf{b}\|^2$$

právě když splňuje soustavu normálních rovnic

$$\mathbf{A}^T \mathbf{A} \mathbf{x} = \mathbf{A}^T \mathbf{b}$$
.

Řešení pomocí derivací

Stejnou podmínku dostaneme z analýzy: minimalizujeme funkci

$$f(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 = (\mathbf{A}\mathbf{x} - \mathbf{b})^T (\mathbf{A}\mathbf{x} - \mathbf{b})$$
$$= \mathbf{x}^T \mathbf{A}^T \mathbf{A}\mathbf{x} - \mathbf{x}^T \mathbf{A}^T \mathbf{b} - \mathbf{b}^T \mathbf{A}\mathbf{x} + \mathbf{b}^T \mathbf{b}$$
$$= \mathbf{x}^T \mathbf{A}^T \mathbf{A}\mathbf{x} - 2\mathbf{b}^T \mathbf{A}\mathbf{x} + \mathbf{b}^T \mathbf{b}$$

- f je konvexní, kvadratická, zdola omezená funkce n proměnných
- Gradient f (vrátíme se k němu později):

$$\nabla f(\mathbf{x}) = 2\mathbf{A}^T \mathbf{A} \mathbf{x} - 2\mathbf{A}^T \mathbf{b}$$

• stacionární podmínka $\nabla f(\mathbf{x}) = \mathbf{0}$ jsou normální rovnice

Důsledky pro řešitelnost soustavy $\mathbf{A}^T \mathbf{A} \mathbf{x} = \mathbf{A}^T \mathbf{b}$:

- Soustava má řešení pro každé \mathbf{A}, \mathbf{b} (neboť vždy $\mathbf{A}^T \mathbf{b} \in \operatorname{rng}(\mathbf{A}^T) = \operatorname{rng}(\mathbf{A}^T \mathbf{A})$).
- Matice A^TA je regulární, právě když A má lin. nezávislé sloupce.
 Pak řešení soustavy lze napsat pomocí inverze:

$$\mathbf{x} = \underbrace{(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T}_{\mathbf{A}^+} \mathbf{b}$$

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

Řešení pomocí QR rozkladu

Řešíme soustavu normálních rovnic

$$\mathbf{A}^T \mathbf{A} \mathbf{x} = \mathbf{A}^T \mathbf{b}$$

kde sloupce A jsou l.n.

- řešení pomocí inverze matice A^TA zanáší zaokrouhlovací chyby
- lépe pomocí (redukovaného) QR rozkladu $\mathbf{A} = \mathbf{QR}$:

$$\mathbf{A}^T \mathbf{A} \mathbf{x} = \mathbf{A}^T \mathbf{b}$$
 $\mathbf{R}^T \mathbf{Q}^T \mathbf{Q} \mathbf{R} \mathbf{x} = \mathbf{R}^T \mathbf{Q}^T \mathbf{b}$
 $\mathbf{R}^T \mathbf{R} \mathbf{x} = \mathbf{R}^T \mathbf{Q}^T \mathbf{b}$
 $\mathbf{R} \mathbf{x} = \mathbf{Q}^T \mathbf{b}$

V Matlabu: x=A\b

Ortogonální projekce na podprostor zadaný obecnou bází

Zopakujme úlohu nejmenších čtverců

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 = \min_{\mathbf{y} \in \operatorname{rng} \mathbf{A}} \|\mathbf{y} - \mathbf{b}\|^2$$

- Když jsou sloupce **A** lin. nezávislé, tvoří bázi podprostoru $X = \operatorname{rng} \mathbf{A}$.
- Pak optimální řešení

$$\mathbf{y} = \mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{A}^{+}\mathbf{b} = \underbrace{\mathbf{A}(\mathbf{A}^{T}\mathbf{A})^{-1}\mathbf{A}^{T}}_{\mathbf{P}}\mathbf{b}$$

je ortogonální projekce vektoru \mathbf{b} na podprostor X (a matice \mathbf{P} je ortogonální projektor).

• Když **A** má ortonormální sloupce ($\mathbf{A}^T \mathbf{A} = \mathbf{I}$), tak $\mathbf{P} = \mathbf{A} \mathbf{A}^T$. To už známe!

Lineární regrese

Regrese je modelování funkční závislosti proměnné $y \in \mathbb{R}$ na proměnné $x \in X$ (kde X je libovolná množina) regresní funkcí

$$y = f(x, \theta).$$

Odhad parametrů: Hledáme parametry $\theta \in \mathbb{R}^n$ tak, aby funkce f dobře modelovala naměřená data $(x_1, y_1), \ldots, (x_m, y_m)$:

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

• Lineární regrese: pro každé $x \in X$ je f lineární funkcí θ . Tedy

$$f(x, \theta) = \theta_1 \varphi_1(x) + \cdots + \theta_n \varphi_n(x) = \varphi(x)^T \theta$$

kde $\varphi = (\varphi_1, \dots, \varphi_n): X \to \mathbb{R}$ jsou dané "bázové funkce".

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 $\mathbf{A} = [a_{ii}] = [\varphi_i(x_i)].$

Příklad: Aproximace polynomiální funkcí

Nechť $X = \mathbb{R}$ a $\varphi_i(x) = x^{j-1}$. Pak

$$f(x, \theta) = \theta_1 + \theta_2 x + \theta_3 x^2 + \dots + \theta_n x^{n-1}$$

a

$$\mathbf{A} = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ & & \vdots & \\ 1 & x_m & x_m^2 & \cdots & x_m^{n-1} \end{bmatrix}$$

(Vandermondova matice).

- Pro n=1 je $f(x,\theta)=\theta$ a $f(x,\theta)=\sum_{i=1}^m(x_i-\theta)^2$. Řešení: $\theta^*=\frac{1}{m}\sum_{i=1}^mx_i$.
- Zobecnění na polynomy více proměnných: $X=\mathbb{R}^d$ a φ_j jsou monomy proměnných x_1,\ldots,x_d

Pozor! Pořád jde o lineární regresi, protože f je lineární v parametrech θ .

Spektrální rozklad a kvadratické funkce:

Vlastní čísla a vektory

Nechť pro matici $\mathbf{A} \in \mathbb{R}^{n \times n}$, nenulový vektor $\mathbf{v} \in \mathbb{C}^n$ a $\lambda \in \mathbb{C}$ platí

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

Pak λ je vlastní číslo matice **A** a v je vlastní vektor příslušný λ .

Můžeme přepsat jako

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

λ je vlastní číslo, právě když je to kořen charakteristického polynomu:

$$p_{\mathbf{A}}(\lambda) = \det(\mathbf{A} - \lambda \mathbf{I}) = \prod_{i=1}^{n} (\lambda_i - \lambda) = 0$$

Je to polynom stupně n, tedy \mathbf{A} má n vlastních čísel (počítaje násobnost).

• Množina vlastních vektorů příslušných jednomu vlastnímu číslu λ je

$$null(\mathbf{A} - \lambda \mathbf{I}) \setminus \{\mathbf{0}\}\$$

Diagonalizovatelné matice

Nechť $\lambda_1,\ldots,\lambda_n$ jsou vlastní čísla matice a $\mathbf{v}_1,\ldots,\mathbf{v}_n$ jim příslušné vlastní vektory. Soustavu

$$\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i, \quad i = 1, \dots, n$$

lze napsat jako

$$AV = V\Lambda$$

kde

$$\mathbf{V} = [\mathbf{v}_1 \cdots \mathbf{v}_n], \qquad \mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_n \end{bmatrix}$$

- V Matlabu: [V,D]=eig(A) (kde D je Λ)
- Jestliže V je regulární, pak

$$A = V\Lambda V^{-1}$$

$$\mathbf{\Lambda} = \mathbf{V}^{-1}\mathbf{A}\mathbf{V}$$

Tedy A je diagonalizovatelná (= podobná diagonální matici).

Věta (Spektrální věta)

Nechť matice $\mathbf{A} \in \mathbb{R}^{n \times n}$ je symetrická $(\mathbf{A}^T = \mathbf{A})$. Pak:

- Všechna vlastní čísla matice A jsou reálná.
- Existuje ortonormální množina n vlastních vektorů matice A.

Důsledek (spektrální rozklad symetrické matice)

Je-li $\mathbf{A} \in \mathbb{R}^{n \times n}$ symetrická, ve spektrálním rozkladu lze \mathbf{V} zvolit ortogonální a

$$\mathbf{A} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^{-1} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^{T} = \lambda_{1} \mathbf{v}_{1} \mathbf{v}_{1}^{T} + \cdots + \lambda_{n} \mathbf{v}_{n} \mathbf{v}_{n}^{T}$$

Kvadratická forma

Funkce $f: \mathbb{R}^n \to \mathbb{R}$ je kvadratická forma, je-li to je homogenní polynom stupně 2. V maticové formě:

$$f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j$$
 pro nějakou $\mathbf{A} \in \mathbb{R}^{n \times n}$

Tvrzení

Pro každou kvadratickou formu $f: \mathbb{R}^n \to \mathbb{R}$ existuje symetrická matice $\mathbf{A} \in \mathbb{R}^{n \times n}$ taková, že $f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$.

Příklad:

$$f(x,y) = 2x^2 - 2xy + y^2 = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 2 & -2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Definitnost kvadratické formy / její matice

Kvadratická forma f je

- positivně semidefinitní, když f(x) ≥ 0 pro každé x
- negativně semidefinitní, když f(x) ≤ 0 pro každé x
- positivně definitní, když f(x) > 0 pro každé x ≠ 0
- negativně definitní, když f(x) < 0 pro každé $x \neq 0$
- indefinitní, když f(x) > 0 a f(y) < 0 pro nějaká x, y

Definitností čtvercové matice **A** rozumíme definitnost kvadr. formy $f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x}$.

Kvadratická funkce

Kvadratická funkce je polynom $f: \mathbb{R}^n \to \mathbb{R}$ druhého stupně, tj.

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

pro nějaké $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{b} \in \mathbb{R}^n$ a $c \in \mathbb{R}$.

Příklad pro n = 2:

$$f(x,y) = 2x^2 - 2xy + y^2 - 2y + 3 = \begin{bmatrix} x \\ y \end{bmatrix}^T \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 0 \\ -2 \end{bmatrix}^T \begin{bmatrix} x \\ y \end{bmatrix} + 3$$

Doplnění kvadratické funkce na čtverec

Pro některé kvadratické lze nalézt $\mathbf{x}_0 \in \mathbb{R}^n$ a $y_0 \in \mathbb{R}$ tak, že

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

Porovnáme koeficienty u stejných monomů:

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

$$\mathbf{c} = \mathbf{x}_0^T \mathbf{A} \mathbf{x}_0 + y_0$$
(*)

- Rovnice (*) je vlastně stacionární podmínka ∇f(x) = 0.
- Doplnění na čtverec je možné, právě když rovnice (*) má řešení (tj. f má aspoň jeden stacionární bod).
- Druhá forma f je posunutá kvadratická forma.
 Z ní ihned zjistíme, zda f má extrém a jaký.

Příklad: Tato kvadr. funkce doplnit na čtverec jde:

$$f(x,y) = 2x^{2} - 2xy + y^{2} - 2y + 3 = \begin{bmatrix} x \\ y \end{bmatrix}^{T} \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 0 \\ -2 \end{bmatrix}^{T} \begin{bmatrix} x \\ y \end{bmatrix} + 3$$
$$= \begin{bmatrix} x - 1 \\ y - 2 \end{bmatrix}^{T} \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x - 1 \\ y - 2 \end{bmatrix} + 1$$
$$= 2(x - 1)^{2} - 2(x - 1)(y - 2) + (y - 2)^{2} + 1$$

tedy $\mathbf{x}_0 = (1, 2), y_0 = 1.$

Příklad: A tato nejde:

$$f(x,y) = x^2 + y = \begin{bmatrix} x \\ y \end{bmatrix}^T \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}^T \begin{bmatrix} x \\ y \end{bmatrix}$$

Soustava $\mathbf{b} = -(\mathbf{A} + \mathbf{A}^T)\mathbf{x}_0$, tj.

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} = -2 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$$

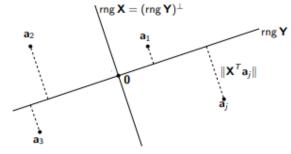
nemá řešení.

Prokládání bodů podprostorem:

Optimální proložení bodů podprostorem

Úloha: Proložení bodů podprostorem

Nechť $\mathbf{a}_1,\ldots,\mathbf{a}_n\in\mathbb{R}^m$ a $k\leq m$. Najdi podprostor dimenze k, který minimalizuje součet čtverců vzdáleností k bodům $\mathbf{a}_1,\ldots,\mathbf{a}_n$.



- Hledaný podprostor je rng \mathbf{Y} kde $\mathbf{Y} \in \mathbb{R}^{m \times k}$.
- Jeho ortogonální doplněk je $(\operatorname{rng} \mathbf{Y})^{\perp} = \operatorname{rng} \mathbf{X}$ kde $\mathbf{X} \in \mathbb{R}^{m \times (m-k)}$ a $\mathbf{X}^T \mathbf{X} = \mathbf{I}$.
- Označíme-li $\mathbf{A} = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_n] \in \mathbb{R}^{m \times n}$, součet čtverců vzdáleností v bodům je

$$\|\mathbf{X}^T \mathbf{a}_1\|^2 + \dots + \|\mathbf{X}^T \mathbf{a}_n\|^2 = \|\mathbf{X}^T \mathbf{A}\|^2 = \operatorname{tr}(\mathbf{X}^T \mathbf{A} \mathbf{A}^T \mathbf{X})$$

min
$$\operatorname{tr}(\mathbf{X}^T\mathbf{A}\mathbf{A}^T\mathbf{X})$$

za podmínek $\mathbf{X}^T\mathbf{X} = \mathbf{I}$
 $\mathbf{X} \in \mathbb{R}^{m \times (m-k)}$

- Spočítej spektrální rozklad AA^T = VAV^T.
- Rozděl matici $\mathbf{V} = \begin{bmatrix} \mathbf{X} & \mathbf{Y} \end{bmatrix} \in \mathbb{R}^{m \times m}$ na bloky $\mathbf{X} \in \mathbb{R}^{m \times (m-k)}$ a $\mathbf{Y} \in \mathbb{R}^{m \times k}$.
- Sloupce Y tvoří ortonormální bázi hledaného podprostoru.
- Sloupce X tvoří ortonormální bázi jeho ortogonálního doplňku.
- Optimální hodnota úlohy je λ₁ + · · · + λ_{m-k}.

Prokládáme afinním podprostorem

Tvrzení

Nechť $\mathbf{a}_1, \dots, \mathbf{a}_n \in \mathbb{R}^m$ a $k \leq m$. Afinní podprostor dimenze k, který minimalizuje součet čtverců vzáleností k bodům $\mathbf{a}_1, \dots, \mathbf{a}_n$, prochází jejich těžištěm

$$\overline{\mathbf{a}} = \frac{1}{n}(\mathbf{a}_1 + \cdots + \mathbf{a}_n).$$

Proložení bodů afinním podprostorem dimenze $k \le n$:

- 1. Zadané body posuneme tak, aby měly těžiště v počátku: $\mathbf{a}_i := \mathbf{a}_i \overline{\mathbf{a}}$.
- 2. Posunuté body proložíme (lineárním) podprostorem dimenze k
- 3. Nalezený podprostor posuneme o a.

Opakování

Spektrální rozklad $\mathbf{A}\mathbf{A}^T = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ pro $\mathbf{A} = [\mathbf{a}_1 \cdots \mathbf{a}_n] \in \mathbb{R}^{m \times n}$.

- Sloupce Y ∈ R^{m×k} matice V odpovídající k největším vlastním číslům tvoří ortonormální bázi podprostoru dimenze k, který minimalizuje součet čtverců vzdáleností k bodům A
- Chyba proložení je součet $\lambda_1 + \cdots + \lambda_{k-1}$ nejmenších vlastních čísel.
- Ortog. projekce bodů a_i na optimální podprostor je b_i = YY^Ta_i. Pišme

$$\mathbf{b}_j = \mathbf{Y}\mathbf{c}_j, \qquad \mathbf{c}_j = \mathbf{Y}^T\mathbf{a}_j$$

neboli

$$B = YC$$
, $C = Y^TA$

Tedy sloupce matice $\mathbf{C} \in \mathbb{R}^{k \times n}$ jsou souřadnice bodů \mathbf{B} v ortonormální bázi \mathbf{Y} .

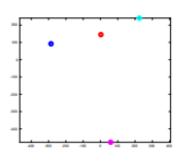
Použití:

- Komprese: Matice A má mn prvků, matice Y, C dohromady (m + n)k prvků.
- Redukce dimenze: Body C mají menší dimenzi než body A.
- Visualizace: Pro k ≤ 3 si body C můžeme prohlédnout.
- Rozpoznávání: Body C často vhodnější pro klasifikaci/shlukování/mining.

Příklad: Visualizace + rozpoznávání dat

Average consumption of food types in grams per person per week in the UK

	England	N.Ireland	Scotland	Wales
Alcoholic drinks	375	135	458	475
Beverages	57	47	53	73
Carcase meat	245	267	242	227
Cereals	1472	1494	1462	1582
Cheese	105	66	103	103
Confectionery	54	41	62	64
Fats and oils	193	209	184	235
Fish	147	93	122	160
Fresh fruit	1102	674	957	1137
Fresh potatoes	720	1033	566	874
Fresh Veg	253	143	171	265
Other meat	685	586	750	803
Other Veg	488	355	418	570
Processed potatoes	198	187	220	203
Processed Veg	360	334	337	365
Soft drinks	1374	1506	1572	1256
Sugars	156	139	147	175

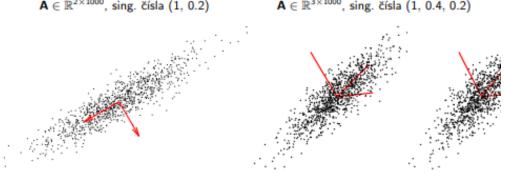


- 4 body v 17-rozměrném prostoru
- Odečteme těžiště.
- Proložíme podprostorem dimenze 2 a promítneme body do něj.
- Zobrazíme souřadnice bodů v ortonormální bázi v₁₆, v₁₇

Příklad: Vlastní vektory matice AA^T (= levé sing. vektory matice A)

$$\boldsymbol{A} \in \mathbb{R}^{2 \times 1000}$$
, sing. čísla (1, 0.2)

$$\boldsymbol{A} \in \mathbb{R}^{3 \times 1000}$$
 , sing. čísla (1, 0.4, 0.2)



Příklad: Přibližné proložení bodů kuželosečkou

Úloha: Najdi kuželosečku, která minimalizuje součet čtverců vzdáleností k daným bodům $(x_1, y_1), \ldots, (x_m, y_m) \in \mathbb{R}^2$.

- Kuželosečka: $Q(x, y) = ax^2 + bxy + cy^2 + dx + ey + f = 0$
- Přesné řešení velmi těžké.
- Přibližná formulace:

 $a, b, c, d, e, f \in \mathbb{R}$

Tedy: minimalizuj $\|\mathbf{A}\boldsymbol{\theta}\|$ za podmínky $\boldsymbol{\theta}^T\boldsymbol{\theta}=1$ kde

$$\mathbf{A} = \begin{bmatrix} x_1^2 & x_1 y_1 & y_1^2 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_m^2 & x_m y_m & y_m^2 & x_m & y_m & 1 \end{bmatrix}$$