Optimalizace

3. Metoda nejmenších čtverců

Tom Werner

FEL ČVUT

Řešitelnost lineárních soustav

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

$$\mathbf{Ax} = \mathbf{b}$$

má řešení, právě když $\mathbf{b} \in \operatorname{rng} \mathbf{A}$, tj. $\operatorname{rank}[\mathbf{A} \;\; \mathbf{b}] = \operatorname{rank} \mathbf{A}$ (Frobeniova věta).

Tři případy:

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

Lineární úloha nejmenších čtverců

Přibližné řešení přeurčené soustavy $\mathbf{Ax} = \mathbf{b}$ ve smyslu nejmenších čtverců:

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

Také se nazývá **(lineární) úloha nejmenších čtverců**.

Příklad: Pro soustavu

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

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

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

Řešení pomocí derivací

Minimalizujeme funkci

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

- funkce f je konvexní, kvadratická, zdola omezená nulou
- Nutná podmínka na lokální extrémy (= stacionární podmínka):

$$\nabla f(\mathbf{x}) = 2\mathbf{A}^T(\mathbf{A}\mathbf{x} - \mathbf{b}) = \mathbf{0}$$

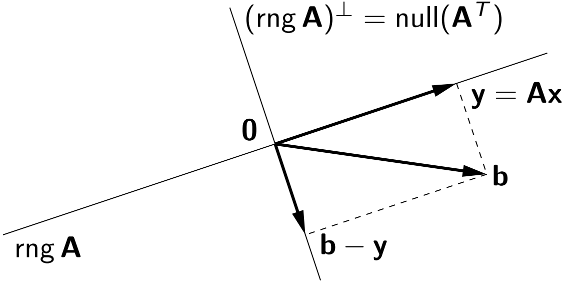
- f je konvexní $\implies$ stacionární body jsou globální minima

Věta

Vektor $\mathbf{x}$ je optimální řešení úlohy právě tehdy, když splňuje soustavu **normálních rovnic**

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

To je lineární soustava n rovnic o n neznámých.



Řešíme úlohu

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

Věta o kolmici: $\mathbf{y}=\mathbf{A}\mathbf{x}$ je optimální právě tehdy, když

$$\begin{aligned}(\mathbf{b}-\mathbf{A}\mathbf{x}) \perp \operatorname{rng}\mathbf{A} &\iff \mathbf{b}-\mathbf{A}\mathbf{x} \in (\operatorname{rng}\mathbf{A})^\perp = \operatorname{null}(\mathbf{A}^T)\\ &\iff \mathbf{A}^T(\mathbf{b}-\mathbf{A}\mathbf{x}) = \mathbf{0}\\ &\iff \mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}\end{aligned}$$

Řešení normálních rovnic pomocí (pseudo)inverze

Věta

- $\operatorname{rng}(\mathbf{A}^T\mathbf{A}) = \operatorname{rng}(\mathbf{A}^T)$
- $\operatorname{null}(\mathbf{A}^T\mathbf{A}) = \operatorname{null}\mathbf{A}$

Důkaz druhé rovnosti: Zřejmě $\mathbf{Ax} = \mathbf{0} \implies \mathbf{A}^T\mathbf{Ax} = \mathbf{0}$.
Ale také $\mathbf{A}^T\mathbf{Ax} = \mathbf{0} \implies \mathbf{x}^T\mathbf{A}^T\mathbf{Ax} = (\mathbf{Ax})^T(\mathbf{Ax}) = \|\mathbf{Ax}\|^2 = 0 \implies \mathbf{Ax} = \mathbf{0}$.
První rovnost pak plyne z druhé rovnosti a z věty o hodnosti a nulitě.

Důsledky pro soustavu $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$:

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

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

  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 normálních rovnic pomocí QR rozkladu

Řešíme soustavu normálních rovnic

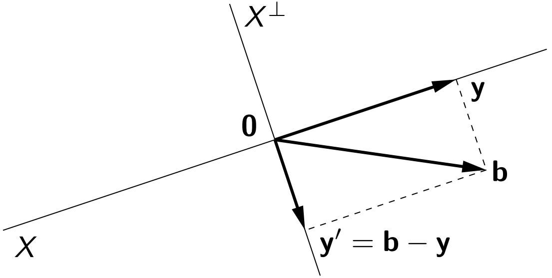
$$\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$$

kde sloupce $\mathbf{A}$ jsou l.n.

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

$$\begin{aligned} \mathbf{A}^T\mathbf{A}\mathbf{x} &= \mathbf{A}^T\mathbf{b} \\ \mathbf{R}^T\mathbf{Q}^T\mathbf{Q}\mathbf{R}\mathbf{x} &= \mathbf{R}^T\mathbf{Q}^T\mathbf{b} \\ \mathbf{R}^T\mathbf{R}\mathbf{x} &= \mathbf{R}^T\mathbf{Q}^T\mathbf{b} \\ \mathbf{R}\mathbf{x} &= \mathbf{Q}^T\mathbf{b} \end{aligned}$$

V Matlabu: `x=A\b`



Ortogonální projekce vektoru $\mathbf{b} \in \mathbb{R}^m$ na podprostor $X \subseteq \mathbb{R}^m$ je vektor $\mathbf{y} \in X$ takový, že $(\mathbf{b} - \mathbf{y}) \perp X$.

Jestliže X má bázi tvořenou sloupci matice $\mathbf{A}$ (tedy $X = \operatorname{rng} \mathbf{A}$), pak

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

Matice $\mathbf{P} \in \mathbb{R}^{m\times m}$ je **ortogonální projektor** na podprostor X.

Tvoří-li sloupce $\mathbf{A}$ ortonormální bázi podprostoru X (tj. $\mathbf{A}^T\mathbf{A} = \mathbf{I}$), máme $\mathbf{P} = \mathbf{A}\mathbf{A}^T$.

Tvrzení: $\mathbf{P}$ nezávisí na volbě báze podprostoru X. Tedy ortog. projekce existuje a je jediná.

Důsledek

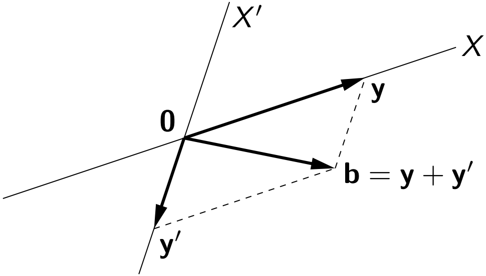
Každé $\mathbf{b} \in \mathbb{R}^m$ lze jednoznačně rozložit jako $\mathbf{b} = \mathbf{y} + \mathbf{y}'$ kde $\mathbf{y} \in X$ a $\mathbf{y}' \in X^\perp$.

- $\mathbf{y}' = \mathbf{b} - \mathbf{y} = \mathbf{b} - \mathbf{Pb} = (\mathbf{I} - \mathbf{P})\mathbf{b}$ je ortog. projekce vektoru $\mathbf{b}$ na podprostor $X^\perp$.
- $\mathbf{I} - \mathbf{P}$ je ortog. projektor na $X^\perp$.
- Zřejmě $\operatorname{rng} \mathbf{P} = X$ a $\operatorname{null} \mathbf{P} = X^\perp$.

Věta (charakterizace ortog. projektorů)

Matice $\mathbf{P}$ je ortog. projektor (na nějaký podprostor) právě když $\mathbf{P}^2 = \mathbf{P} = \mathbf{P}^T$.

Nepovinné: Obecná projekce



Podprostory $X, X' \subseteq \mathbb{R}^m$ nazveme **komplementární**, jestliže

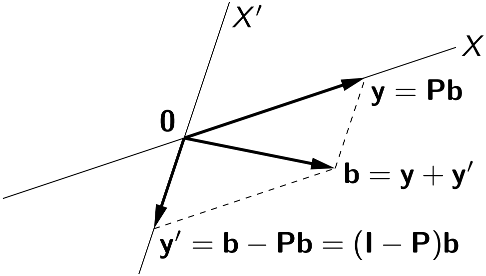
$$X \cap X' = \{\mathbf{0}\}$$

$$\dim X + \dim X' = m$$

Věta

Každé $\mathbf{b} \in \mathbb{R}^m$ lze jednoznačně rozložit jako $\mathbf{b} = \mathbf{y} + \mathbf{y}'$ kde $\mathbf{y} \in X$ a $\mathbf{y}' = X'$.

Vektor $\mathbf{y}$ se nazývá **projekce** vektoru $\mathbf{b}$ na podprostor X ve směru podprostoru X'.



Definice: Matice $\mathbf{P}$ se nazývá (obecný) **projektor**, jestliže $\mathbf{P}^2 = \mathbf{P}$.

- Podprostory $\operatorname{rng} \mathbf{P}$ a $\operatorname{null} \mathbf{P}$ jsou komplementární a platí $\operatorname{null} \mathbf{P} = \operatorname{rng}(\mathbf{I} - \mathbf{P})$.
- Tedy $\mathbf{y} = \mathbf{Pb}$ je projekce vektoru $\mathbf{b}$ na podprostor $\operatorname{rng} \mathbf{P}$ ve směru podprostoru $\operatorname{null} \mathbf{P}$.
- Pro každou komplementární dvojici podprostorů X, X' existuje právě jeden projektor tak, že $\operatorname{rng} \mathbf{P} = X$ a $\operatorname{null} \mathbf{P} = X'$.
- Matice $\mathbf{P}$ je projektor, právě když $\mathbf{P} = \mathbf{B}(\mathbf{AB})^{-1}\mathbf{A}$ pro nějaké matice $\mathbf{A}, \mathbf{B}$ takové, že $\mathbf{AB}$ je regulární.
- Speciálně, když $\mathbf{AB} = \mathbf{I}$ tak $\mathbf{P} = \mathbf{BA}$.

Ortogonální projekce/projektory jsou spec. případ, kdy $X' = X^\perp$ a $\mathbf{P} = \mathbf{P}^T$.

Řešení nedourčené lineární soustavy s nejmenší normou

Mezi řešeními soustavy $\mathbf{A}\mathbf{x} = \mathbf{b}$ najdeme to s nejmenší normou:

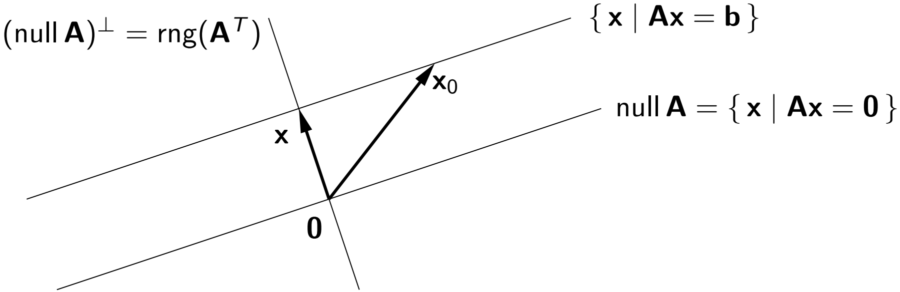
$$\min\{\, \|\mathbf{x}\|^2 \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{A}\mathbf{x} = \mathbf{b} \,\}$$

Příklad: Soustava

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

má nekonečně mnoho řešení. Hledáme řešení které minimalizuje číslo $x^2 + y^2 + z^2$.

Řešení úlohy



$\mathbf{x}$ má nejmenší normu $\iff \mathbf{x} \perp \operatorname{null}\mathbf{A} \iff \mathbf{x} \in (\operatorname{null}\mathbf{A})^\perp = \operatorname{rng}(\mathbf{A}^T) \iff \mathbf{x} = \mathbf{A}^T\mathbf{y}$

Věta

$\mathbf{x}$ je optimální řešení úlohy, právě když existuje $\mathbf{y}$ tak, že

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

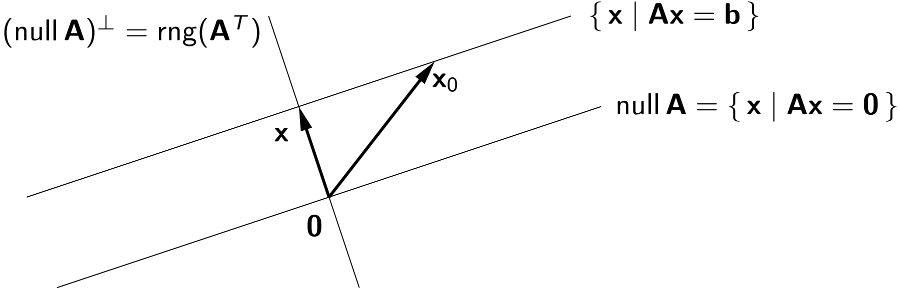
Řešme soustavu

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

- Dosazením dostaneme $\mathbf{A}\mathbf{A}^T\mathbf{y} = \mathbf{b}$.
- Má-li $\mathbf{A}$ l.n. řádky, je $\mathbf{y} = (\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b}$. Tedy

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

  kde $\mathbf{A}^+$ je **pseudoinverze** matice $\mathbf{A}$ (s l.n. řádky).
  Je to jedna z pravých inverzí matice $\mathbf{A}$, neboť $\mathbf{A}\mathbf{A}^+ = \mathbf{A}\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1} = \mathbf{I}$.



- Nechť $\mathbf{x}_0$ je libovolné splňující $\mathbf{A}\mathbf{x}_0 = \mathbf{b}$.
- Optimální $\mathbf{x}$ je ortog. projekce vektoru $\mathbf{x}_0$ na podprostor $(\operatorname{null} \mathbf{A})^\perp = \operatorname{rng}(\mathbf{A}^T)$, tedy

$$\mathbf{x} = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{x}_0 = \underbrace{\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}}_{\mathbf{A}^+}\mathbf{b}$$

(nezávisí na tom, jaké partikulární řešení $\mathbf{x}_0$ jsme vybrali!)

Důsledek: Má-li soustava $\mathbf{A}\mathbf{x} = \mathbf{b}$ řešení, má právě jedno řešení s nejmenší normou.

Pseudoinverze obecné matice

Hledejme optimální řešení úlohy

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

s nejmenší normou.

To vede na úlohu

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

Nechť $\mathbf{x}^*$ je (jediné) řešení úlohy (LSN).

- Má-li soustava $\mathbf{A}\mathbf{x} = \mathbf{b}$ jediné řešení, pak $\mathbf{x}^*$ je toto řešení.
- Má-li soustava $\mathbf{A}\mathbf{x} = \mathbf{b}$ více řešení, pak $\mathbf{x}^*$ je její řešení s nejmenší normou.
- Nemá-li soustava $\mathbf{A}\mathbf{x} = \mathbf{b}$ řešení, pak $\mathbf{x}^*$ je optimální řešení úlohy (LS) s nejmenší normou.

Fakt: Existuje matice $\mathbf{A}^+$ (**Moore-Penroesova pseudoinverze** matice $\mathbf{A}$) tak, že $\mathbf{x}^* = \mathbf{A}^+\mathbf{b}$.

- Má-li $\mathbf{A}$ l.n. sloupce, pak $\mathbf{A}^+ = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$.
- Má-li $\mathbf{A}$ l.n. řádky, pak $\mathbf{A}^+ = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}$.
- Nemá-li $\mathbf{A}$ plnou hodnost, pak $\mathbf{A}^+$ lze spočítat pomocí SVD.