Optimalizace

3. Metoda nejmenších čtverců

Tom Werner

FEL ČVUT

Řešitelnost lineárních soustav

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

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

- 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. $\operatorname{rank}[\mathbf{A} \;\; \mathbf{b}] = \operatorname{rank} \mathbf{A}$ (Frobeniova věta)
- množina řešení je afinní podprostor $\mathbb{R}^n$

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

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**.

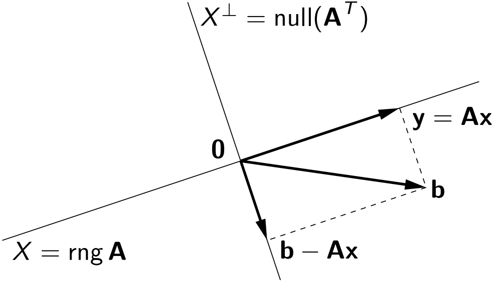
Příklad: Pro soustavu

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

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

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

Řešení úlohy



Věta

Vektor $\mathbf{x}$ 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**

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

Důkaz: Věta o kolmici: $\mathbf{x}$ je optimální řešení právě když $(\mathbf{b} - \mathbf{A}\mathbf{x}) \perp \operatorname{rng}\mathbf{A}$, tj. $\mathbf{A}^T(\mathbf{b} - \mathbf{A}\mathbf{x}) = \mathbf{0}$.

Řešení pomocí derivací

Stejnou podmínku dostaneme z analýzy: minimalizujeme funkci

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

- 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

Řešitelnost soustavy normálních rovnic

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 $\mathbf{Ax} = \mathbf{0} \;\Rightarrow\; \mathbf{A}^T\mathbf{Ax} = \mathbf{0}$. Ale také
$\mathbf{A}^T\mathbf{Ax} = \mathbf{0} \;\Rightarrow\; \mathbf{x}^T\mathbf{A}^T\mathbf{Ax} = (\mathbf{Ax})^T(\mathbf{Ax}) = \|\mathbf{Ax}\|^2 = 0 \;\Rightarrow\; \mathbf{Ax} = \mathbf{0}$.
První rovnost pak plyne z druhé rovnosti a rank-nullity theoremu.

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

- 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 pomocí QR rozkladu

Řešíme soustavu normálních rovnic

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

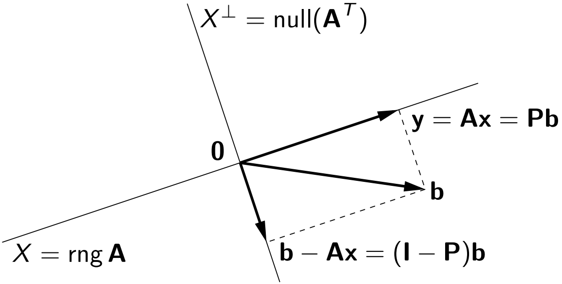
kde sloupce $\mathbf{A}$ jsou l.n.

- 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 na podprostor zadaný obecnou bází



Zopakujme úlohu nejmenších čtverců: $\min\limits_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{Ax} - \mathbf{b}\|^2 = \min\limits_{\mathbf{y}\in\mathrm{rng}\,\mathbf{A}} \|\mathbf{y} - \mathbf{b}\|^2$

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

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

  je ortogonální projekce vektoru $\mathbf{b}$ na podprostor X a $\mathbf{P}$ je ortog. projektor.
- Když $\mathbf{A}$ má ortonormální sloupce ($\mathbf{A}^T\mathbf{A} = \mathbf{I}$), tak $\mathbf{P} = \mathbf{AA}^T$. To už známe!

Řešení nedourčené lineární soustavy s nejmenší normou

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

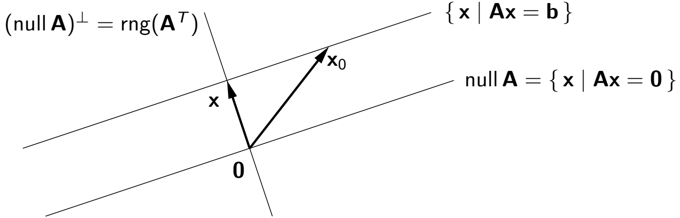
$$\min\{\, \|\mathbf{x}\|^2 \mid \mathbf{x} \in \mathbb{R}^n,\ \mathbf{A}\mathbf{x} = \mathbf{b}\,\}$$

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 $\;\Leftrightarrow\; \mathbf{x} \perp \text{null}\,\mathbf{A} \;\Leftrightarrow\; \mathbf{x} \in (\text{null}\,\mathbf{A})^\perp = \text{rng}(\mathbf{A}^T) \;\Leftrightarrow\; \mathbf{x} = \mathbf{A}^T\mathbf{y}$

Věta

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

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

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

Dosazení dá $\mathbf{A}\mathbf{A}^T\mathbf{y} = \mathbf{b}$. 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 $\mathbf{A}\mathbf{A}^+ = \mathbf{A}\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1} = \mathbf{I}$.

Použití metody nejmenších čtverců

Lineární regrese

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

$$y = f(x, \boldsymbol{\theta}).$$

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

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

- Regrese je **lineární**, jestliže pro každé $x \in X$ je f lineární funkcí $\boldsymbol{\theta}$. Tedy

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

  kde $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_n)\colon X \to \mathbb{R}$ 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 $\mathbf{y} = (y_1, \ldots, y_m)$ a prvky matice $\mathbf{A}$ jsou $a_{ij} = \varphi_j(x_i)$.

Příklad: Aproximace polynomiální funkcí

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

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

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: $\theta^* = \frac{1}{m} \sum_{i=1}^m x_i$.
- 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.

Mnoho dalších použití!

Mnoho dalších použití úlohy

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

je v této knize (zdarma ke stažení i se slajdy, viz doplňující literatura na CW):