Úloha nejmenších čtverců

Petr Olšák

petr@olsak.net

http://petr.olsak.net/

Vlastnosti Gramovy matice $\mathbf{A}^T\mathbf{A}$

Nejprve dokážeme tvrzení, které budeme potřebovat:

Nechť $\mathbf{A} \in \mathbb{R}^{m \times n}$ je libovolná matice. Pak

$$\operatorname{Null} \mathbf{A}^T\mathbf{A} = \operatorname{Null} \mathbf{A}$$
$$\operatorname{rng} \mathbf{A}^T\mathbf{A} = \operatorname{rng} \mathbf{A}^T$$

Důsledek: $\mathbf{A}$ má LN sloupce, právě když $\mathbf{A}^T\mathbf{A}$ je regulární.
$\mathbf{A}$ má LN řádky, právě když $\mathbf{A}\mathbf{A}^T$ je regulární.

Soustavy nehomogenních rovnic $\mathbf{A}\mathbf{x} = \mathbf{b}$

Z hlediska množiny řešení rozlišujeme tři případy:

- Soustava nemá řešení:
  nastane právě když $\mathbf{b} \notin \operatorname{rng} \mathbf{A}$.
- Soustava má jedniné řešení:
  nastane právě když $\mathbf{b} \in \operatorname{rng} \mathbf{A}$ a matice $\mathbf{A}$ má LN sloupce.
- Soustava má více řešení:
  nastane právě když $\mathbf{b} \in \operatorname{rng} \mathbf{A}$ a matice $\mathbf{A}$ má LZ sloupce.

Prvnímu případu říkáme přeurčená soustava, třetímu případu nedourčená soustava.

Úloha nejmenších čtverců najde „optimální řešení“ přeurčené soustavy.

Motivační příklad

Máme soustavu nehomogenních rovnic

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

Je to přeurčená soustava, $\mathbf{b} \notin \operatorname{rng} \mathbf{A}$, konkrétně $\operatorname{rank} \mathbf{A} = 2$, $\operatorname{rank}[\mathbf{A}\ \mathbf{b}] = 3$.

Například každý řádek soustavy vyjadřuje lineární závislost nějakých parametrů danou třeba z fyzikální podstaty měřeného jevu. Jednotlivým měřením získáme koeficienty jedné rovnice. Těch měření provedeme třeba desítky či stovky. Kvůli zaokrouhlovacím chybám a chybám měření skoro jistě budeme mít přeurčenou soustavu s úzkou maticí.

Úloha nejmenších čtverců

Hledá se $\mathbf{x}$ takové, že po dosazení do soustavy má vektor levých stran nejmenší vzdálenost od vektoru pravých stran. Tedy:

$$\min\{\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2;\ \mathbf{x} \in \mathbb{R}^n\} \tag{1}$$

Mezinárodní název: *least squares method* (LS).

- Pozorování: Má-li soustava $\mathbf{A}\,\mathbf{x} = \mathbf{b}$ řešení, pak optimalizační úloha (1) najde toto řešení (vzdálenost je nulová), jinak optimalizační úloha nenajde skutečné řešení, ale vektor, který intuitivně nejlépe vyhovuje zadané soustavě.
- Předchozí příklad zformulovaný jako úloha nejmenších čtverců:

$$\min\ (x + 2y - 6)^2 + (-x + y - 3)^2 + (x + y - 4)^2, \quad x, y \in \mathbb{R}$$

Řešení bychom mohli najít pomocí parciálních derivací.
Ale dnes na to půjdeme jinak...

Algebraické řešení úlohy LS

- $\min \|\mathbf{Ax} - \mathbf{b}\|^2 = \min \|\mathbf{y} - \mathbf{b}\|^2, \ \mathbf{y} \in \operatorname{rng} \mathbf{A}$.
- Geometricky: $\mathbf{y}$ musí být kolmý průmět vektoru $\mathbf{b}$ do podprostoru $\operatorname{rng} \mathbf{A}$.
- Fakt, že vektor $\mathbf{y}$ (tj. vektor $\mathbf{Ax}$) musí být kolmý na $\operatorname{rng} \mathbf{A}$ vyjádříme skalárními součiny: $\mathbf{A}^T(\mathbf{Ax} - \mathbf{b}) = \mathbf{0}$.
- To vede na normální soustavu rovnic úlohy:

$$\mathbf{A}^T\mathbf{A}\,\mathbf{x} = \mathbf{A}^T\mathbf{b}. \tag{2}$$

- Soustava (2) má vždy řešení díky tvrzení ze slide 2.
- Řešení soustavy (2) je řešením optimalizační úlohy LS (1).
- Má-li $\mathbf{A}$ LZ sloupce, má úloha (1) více řešení.

Úloha s maticí s LN sloupci

- Gramova matice $\mathbf{A}^T\mathbf{A}$ je v takovém případě regulární.
- Normální rovnici lze vyřešit násobením inverzní maticí zleva:

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

- Takto spočítaný vektor $\mathbf{x}$ je tedy řešením úlohy LS (1).
- Matici $(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ nazýváme levou pseudoinverzí k matici $\mathbf{A}$ (značíme ji $\mathbf{A}^+$). Je to jedna z levých inverzí matice $\mathbf{A}$, protože:

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

- Pro srovnání: Má-li matice $\mathbf{A}$ LN řádky, definujeme pravou pseudoinverzi k matici $\mathbf{A}$ jako $\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}$ (značíme ji taky $\mathbf{A}^+$). Platí totiž:

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

- Úloha $\mathbf{A}\mathbf{x} = \mathbf{b}$ má optimální řešení $\mathbf{x} = \mathbf{A}^+\mathbf{b}$.
- Poznámka: existuje definice pseudoinverze k libovolné matici, ale tomu se zatím vyhneme.

Řešení úlohy QR rozkladem

- Protože výpočet $\mathbf{A}^T\mathbf{A}$ může být drahý a může značně snížit přesnost, je řešení úlohy nejmenších čtverců postaveno na QR rozkladu matice $\mathbf{A}$.
- Předpokládejme $\mathbf{A}$ s LN sloupci. Pak úlohu LS (1) řešíme použitím redukovaného rozkladu matice $\mathbf{A} = \mathbf{QR}$ takto:

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

- Úlohu tedy vyřešíme z rovnice $\mathbf{R}\,\mathbf{x} = \mathbf{Q}^T\mathbf{b}$ zpětným dosazením.
- Obdobně pracuje algoritmus `A\b`.

Ortogonální projektor podruhé

- Z minula víme, že ortogonální projektor na $\operatorname{rng} \mathbf{U}$ je matice $\mathbf{U}\mathbf{U}^T$, má-li matice $\mathbf{U}$ ortonormální sloupce.
- Nyní předpokládejme podprostor zadaný jako $\operatorname{rng} \mathbf{A}$, kde $\mathbf{A}$ má LN sloupce, ale ne nutně ortonormální.
- Nechť $\mathbf{x}$ je řešení úlohy LS (1), pak $\mathbf{A}\mathbf{x}$ je kolmá projekce vektoru $\mathbf{b}$ na $\operatorname{rng} \mathbf{A}$. Konkrétně:

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

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

- Je tedy $\mathbf{P} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ ortogonální projektor na $\operatorname{rng} \mathbf{A}$.
- V případě $\mathbf{A} = \mathbf{U}$ matice s ortonormálními sloupci tento obecnější vzorec pro $\mathbf{P}$ přechází na $\mathbf{U}\mathbf{U}^T$.

Úloha nejm. čtverců s maticí s LZ sloupci

- Tato úloha má více řešení.
- I normální rovnice $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$ má více řešení.
- Najdeme jedno partikulární řešení normální rovnice a přidáme k němu podprostor $\operatorname{Null}\mathbf{A}^T\mathbf{A} = \operatorname{Null}\mathbf{A}$.
- K tomu potřebujeme umět řešit nedourčené nehomogenní soustavy...

Nedourčená soustava $\mathbf{A}\mathbf{x} = \mathbf{b}$ s LN řádky

- Předpokládáme $\mathbf{b} \in \operatorname{rng} \mathbf{A}$ a $\mathbf{A}$ má LN řádky, soustava má tedy řešení.
- **Úloha:** Mezi řešeními najdeme řešení $\mathbf{x}$ s nejmenší normou.
- Musí být splněno $\mathbf{x} \perp \operatorname{Null} \mathbf{A}$ (přednášející načrtne obrázek).
- Když $\mathbf{x} \perp \operatorname{Null} \mathbf{A}$, pak je $\mathbf{x} \in \operatorname{rng} \mathbf{A}^T$, neboli existuje $\mathbf{y}$ tak, že $\mathbf{x} = \mathbf{A}^T\mathbf{y}$.
- Dosazením do rovnice $\mathbf{A}\mathbf{x} = \mathbf{b}$ máme $\mathbf{A}\mathbf{A}^T\mathbf{y} = \mathbf{b}$.
- Protože $\mathbf{A}$ má LN řádky, je $\mathbf{A}\mathbf{A}^T$ regulární.
- Vyřešíme: $\mathbf{y} = (\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b}$ tj. $\mathbf{x} = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b}$. Takže $\mathbf{x} = \mathbf{A}^+\mathbf{b}$.
- $\mathbf{x} = \mathbf{A}^+\mathbf{b}$ je tedy řešení optimalizační úlohy

$$\min\{\|\mathbf{x}\|^2;\ \mathbf{A}\mathbf{x} = \mathbf{b}\}.$$

Sjednocující formulace pro řešení $\mathbf{Ax} = \mathbf{b}$

- Lze zformulovat jedinou úlohu, která najde v případě soustavy $\mathbf{Ax} = \mathbf{b}$:
  - řešení s nejmenší normou, má-li soustava více řešení,
  - řešení, má-li soustava jediné řešení,
  - optimální řešení ve smyslu nejmenších čtverců, nemá-li soustava řešení; je-li takových optimálních řešení více, najde úloha mezi nimi optimální řešení s nejmenší normou.

  Taková sjednocující optimalizační úloha zní:

$$\min\{\|\mathbf{x}\|^2;\ \mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}\}.$$

- Operátor `A\b` z Matlabu najde vždy řešení s vlastnostmi popsanými zde.

Vícekriteriální nejmenší čtverce

Myšlenka na tomto slide se opírá o následující jednoduché tvrzení:

- Je-li $\mathbf{a} \in \mathbb{R}^m$, $\mathbf{b} \in \mathbb{R}^n$, pak $\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \in \mathbb{R}^{m+n}$ a platí:

$$\left\| \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} \right\|^2 = \|\mathbf{a}\|^2 + \|\mathbf{b}\|^2.$$

- Nechť $\mathbf{A}_i \in \mathbb{R}^{m_i \times n}$, $\mathbf{b}_i \in \mathbb{R}^{m_i}$, $\mu_i > 0$. Úloha

$$\min\{\mu_1 \|\mathbf{A}_1\mathbf{x} - \mathbf{b}_1\|^2 + \cdots + \mu_k \|\mathbf{A}_k\mathbf{x} - \mathbf{b}_k\|^2, \ \mathbf{x} \in \mathbb{R}^n\}$$

je vícekriteriální úloha minimalizace nejmenších čtverců a lze ji převést na standardní úlohu $\min\{\|\mathbf{A}\mathbf{x} - \mathbf{b}\}$, kde

$$\mathbf{A} = \begin{bmatrix} \sqrt{\mu_1}\mathbf{A}_1 \\ \vdots \\ \sqrt{\mu_k}\mathbf{A}_k \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} \sqrt{\mu_1}\mathbf{b}_1 \\ \vdots \\ \sqrt{\mu_k}\mathbf{b}_k \end{bmatrix}.$$

Normální soustava rovnic pro tuto úlohu zní:

$$(\mu_1 \mathbf{A}_1^T\mathbf{A}_1 + \cdots + \mu_k \mathbf{A}_k^T\mathbf{A}_k)\mathbf{x} = \mu_1 \mathbf{A}_1^T\mathbf{b}_1 + \cdots + \mu_k \mathbf{A}_k^T\mathbf{b}_k.$$

Příklad dvou-kriteriální úlohy

Chceme najít řešení soustavy $\mathbf{Ax} = \mathbf{b}$ ve smyslu nejmenších čtverců, ale rádi bychom s určitou vahou přihlédli také k požadavku, aby norma řešení byla pokud možno malá.

- Pro zadanou váhu $\mu > 0$ hledáme

$$\min\{\|\mathbf{Ax} - \mathbf{b}\|^2 + \mu\,\|\mathbf{x}\|^2;\ \mathbf{x} \in \mathbb{R}^n\}.$$

Toto lze převést na jedinou úlohu nejmenších čtverců $\min\{\|\mathbf{Bx} - \mathbf{c}\|^2;\ \mathbf{x} \in \mathbb{R}^n\}$, kde

$$\mathbf{B} = \begin{bmatrix} \mathbf{A} \\ \sqrt{\mu}\,\mathbf{I} \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix}$$

s normální soustavou $(\mathbf{A}^T\mathbf{A} + \mu\,\mathbf{I})\,\mathbf{x} = \mathbf{A}^T\mathbf{b}$. Matice této soustavy je regulární, takže máme řešení

$$\mathbf{x} = (\mathbf{A}^T\mathbf{A} + \mu\,\mathbf{I})^{-1}\mathbf{A}^T\mathbf{b}.$$