Poznámky k předmětu Numerická lineární algebra I. (přednáška 12)

Michal Merta*

*Katedra aplikované matematiky, VŠB-Technická univerzita Ostrava, e-mail: michal.merta@vsb.cz

1 Aplikace QR rozkladu - řešení problému nejmenších čtverců

Motivační příklad Tlak v automatickém espresso kávovaru závisí mj. na hrubosti mletí kávových zrn. Závislost jsme zjistili experimentálně a zakreslili do grafu na Obrázku 1.1.

Obrázek 1.1: Tlak v kávovaru v závislosti na hrubosti mletí kávy.

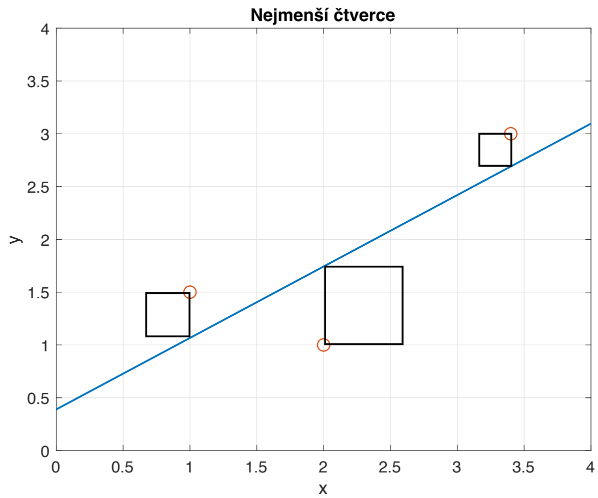
Graf nám naznačuje lineární závislost. Měření je však zatížené chybou. Již na první pohled je proto zřejmé, že nedokážeme nalézt lineární funkci, která prochází všemi naměřenými hodnotami. Taková funkce by měla obecně tvar $p(x) = c_1 + c_2 x$. Kdybychom provedli pouze dvě měření $\{x_1, y_1\}, \{x_2, y_2\}$, stačilo by dosadit a získali bychom soustavu dvou lineárních rovnic o dvou neznámých

$$\begin{aligned} y_1 &= c_1 + c_2 x_1 \\ y_2 &= c_1 + c_2 x_2, \end{aligned}$$

ze které bychom snadno dopočetli koeficienty c_1, c_2. V případě měření z Obrázku 1.1 však dostaneme pět rovnic o dvou neznámých, tedy přeurčený systém, který obecně nemusí mít žádné řešení.

Můžeme se ovšem pokusit nalézt přímku p, která sice nebude procházet všemi body, ale chyba, které se takto dopustíme, bude v jistém smyslu minimální (hledáme tzv. lineární regresi). Přesněji řečeno, budeme chtít, aby součet druhých mocnin (čtverců) odchylek lineární funce p od naměřených hodnot byl co nejmenší:

$$\min_{p \text{ linearni}} \sum_{j=1}^{m} (p(x_j) - y_j)^2 \tag{1.1}$$



Obrázek 1.2: Lineární funkce (modře) minimalizující součet druhých mocnin odchylek od daných bodů (červeně).

Geometrická interpretace tohoto přístupu (metody nejmenších čtverců) je znázorněna na Obrázku 1.2.

Dosadíme-li do (1.1) $p(x_j) = c_1 + c_2 x_j$, můžeme minimalizační problém přepsat na:

$$\min_{c_1, c_2 \in \mathbb{R}} \left\| \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ & \vdots \\ 1 & x_{m-1} \\ 1 & x_m \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} - \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{m-1} \\ y_m \end{bmatrix} \right\|^2 . \tag{1.2}$$

Druhou mocninu můžeme zanedbat, aniž bychom změnili řešení.

Problém nejmenších čtverců nemusí vznikat pouze při aproximaci funkce polynomem. V obecném případě, máme-li $\mathsf{A} \in \mathbb{R}^{m \times n}, \boldsymbol{c} \in \mathbb{R}^n$ a $\boldsymbol{y} \in \mathbb{R}^m, n < m$ a chceme-li aproximovat řešení přeurčeného systému $\mathsf{A}\boldsymbol{c} = \boldsymbol{y}$, nabyde (1.2) tvar

$$\min_{\boldsymbol{c} \in \mathbb{R}^m} \|\mathsf{A}\boldsymbol{c} - \boldsymbol{y}\|.$$

Jednou z možností, jak tento problém řešit je pomocí QR rozkladu. Předpokládejme tedy, že známe úplný QR rozklad

$$\mathsf{A} = \mathsf{Q}\hat{\mathsf{R}},$$

kde

$$\hat{\mathsf{R}} = \begin{bmatrix} \mathsf{R} \\ \mathsf{O} \end{bmatrix}$$

a $\mathsf{Q} \in \mathbb{R}^{m\times m}, \mathsf{R} \in \mathbb{R}^{n\times n}, \mathsf{O} \in \mathbb{R}^{(m-n)\times n}$. Pak

$$\begin{aligned}\min_{\boldsymbol{c}\in\mathbb{R}^m} \|\mathsf{A}\boldsymbol{c} - \boldsymbol{y}\| &= \min_{\boldsymbol{c}\in\mathbb{R}^m} \|\mathsf{Q}\hat{\mathsf{R}}\boldsymbol{c} - \boldsymbol{y}\| = \min_{\boldsymbol{c}\in\mathbb{R}^m} \|\underbrace{\mathsf{Q}^T\mathsf{Q}}_{=\mathsf{I}}\hat{\mathsf{R}}\boldsymbol{c} - \mathsf{Q}^T\boldsymbol{y}\| \\ &= \min_{\boldsymbol{c}\in\mathbb{R}^m} \|\hat{\mathsf{R}}\boldsymbol{c} - \underbrace{\mathsf{Q}^T\boldsymbol{y}}_{\overset{\text{ozn.}}{=}\hat{\boldsymbol{y}}}\| = \min_{\boldsymbol{c}\in\mathbb{R}^m} \|\hat{\mathsf{R}}\boldsymbol{c} - \hat{\boldsymbol{y}}\|\end{aligned}$$

Poznamenejme, že druhá rovnost vychází z toho, že přenásobíme-li vektor ortogonální maticí, nezměníme jeho normu ($\|\mathsf{Q}\boldsymbol{v}\| = \sqrt{(\mathsf{Q}\boldsymbol{v})^T\mathsf{Q}\boldsymbol{v}} = \sqrt{\boldsymbol{v}^T\mathsf{Q}^T\mathsf{Q}\boldsymbol{v}} = \sqrt{\boldsymbol{v}^T\boldsymbol{v}} = \|\boldsymbol{v}\|$). Náš minimalizační problém jsme tedy převedli na

$$\min_{\boldsymbol{c}\in\mathbb{R}^n} \left\| \begin{bmatrix} \mathsf{R} \\ \mathsf{O} \end{bmatrix} \boldsymbol{c} - \begin{bmatrix} \hat{\boldsymbol{y}}(1:n) \\ \hat{\boldsymbol{y}}(n+1:m) \end{bmatrix} \right\|$$

Je jasné, že prvky na pozicích $n+1:m$ výrazu uvnitř normy žádnou volbou $\boldsymbol{c}$ neovlivníme (vektor $\boldsymbol{c}$ je na těchto pozicích přenásoben nulovou maticí). Prvních n prvků ale můžeme vynulovat, položíme-li

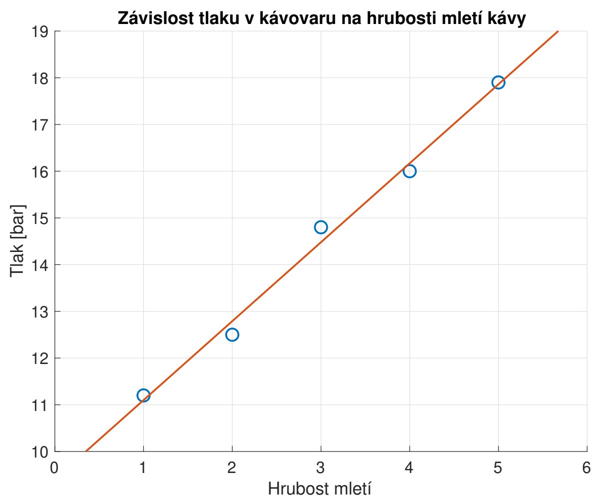
$$\mathsf{R}\boldsymbol{c} = \hat{\boldsymbol{y}}(1:n),$$

čímž najdeme hledané minimum. Poznamenejme, že se jedná o systém s horní trojúhelníkovou maticí R a neznámým vektorem $\boldsymbol{c}$, který můžeme řešit efektivně zpětnou substitucí.

Pokud tímto způsobem vyřešíme problém z Obrázku 1.1, získáme lineární funkci znázorněnou na Obrázku 1.3.

References

[1] Reichel, L. Least Square Approximation by QR Decomposition. Dostupné z `http://www.math.kent.edu/~reichel/courses/intr.num.comp.1/fall09/lecture4/lecture4.pdf`.



Obrázek 1.3: Lineární regrese nalezená pomocí metody nejmenších čtverců.