Komentář k úlohám probíraným na cvičeních dle plánu 04

Petr Olšák

5.1 Má řešení ve skriptech.

5.2 Toto je typický příklad přeurčené soustavy bez řešení (zkuste eliminovat na papíře, zjistíte, že to nemá řešení). Vložte matici $\mathbf{A}$ a vektor pravých stran $\mathbf{b}$ do Matlabu a zkuste tři metody výpočtu řešení ve smyslu nejmenších čtverců. a) přímo: `A\b`, b) podle přednášky, tj. podle vzorce (5.3): `(A'*A)^-1 * A' * b` a konečně c) QR rozkladem, tj. podle přednášky a vzorce (5.9): `[Q R] = rq(A,0); R^-1 * Q' * b`. Ve všech případech dostaneme stejné řešení. Komická hodnota `-0.0000` ve výsledku naznačuje, že Matlab má smysl pro humor.

5.3 a) Minimalizujeme $\sum \|\mathbf{x}-\mathbf{a}_i\|^2$ při daných $\mathbf{a}_i$. To odpovídá přeurčené soustavě nehomogenních rovnic

$$\begin{matrix} \mathbf{x}=\mathbf{a}_1 \\ \mathbf{x}=\mathbf{a}_2 \\ \vdots \\ \mathbf{x}=\mathbf{a}_m \end{matrix} \qquad \text{tedy} \qquad \begin{bmatrix} \mathbf{I} \\ \mathbf{I} \\ \vdots \\ \mathbf{I} \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_m \end{bmatrix}$$

a $\mathbf{x}$ je řešení dle metody nejmenších čverců. Vynásobením zleva transponovanou maticí soustavy $[\,\mathbf{I} \quad \mathbf{I} \quad \cdots \quad \mathbf{I}\,]$ (tj. převedením na normální rovnice) máme $m\,\mathbf{I}\mathbf{x} = \mathbf{a}_1+\mathbf{a}_2+\cdots+\mathbf{a}_m$ a odtud vidíme, že $\mathbf{x}$ musí být těžiště zadaných bodů $\mathbf{a}_i$.

5.3 g) Všechny naměřené vzorky (G_i, S_i, P_i) pro $i \in \{1,\dots,n\}$ mají vyhovovat rovnostem

$$P_i = G_i(a_1 + a_2 S_i + a_3 10^{G_i+S_i} + a_4 10^{-S_i})$$

což je pro $n > 4$ přeurčená nehomogenní soustava s neznámými a_1, a_2, a_3, a_4 , kterou chceme řešit ve smyslu nejmenších čtverců. Jednotlivé řádky matice soustavy včetně odpovídajícího údaje ze sloupce pravých stran jsou:

$$[\; G_i \quad G_i S_i \quad G_i\, 10^{G_i+S_i} \quad G_i\, 10{-}S_i \quad | \quad P_i \;]$$

Matice soustavy (bez sloupce pravých stran) má n řádků a čtyři sloupce. Úkolem pak je z naměřených dat vytvořit tuto matici a sloupec pravých stran, vložit údaje do počítače a provést `A\b`. Tím dostaneme odhad hodnot a_1, a_2, a_3, a_4.

5.3 h) Máme konečnou posloupnost známých údajů $p_1, p_2, \dots, p_n$ a chceme predikovat údaj p_{n+1} pomocí vzorce $p_{n+1} = \beta_1 + \beta_2 p_n + \beta_3 p_{n-1}$. K tomu účelu musíme nejprve odhadnout $\beta_1, \beta_2, \beta_3$ ze známých hodnot $p_1, p_2, \dots, p_n$ jako řešení přeurčené soustavy s maticí a sloupcem pravých stran:

$$\mathbf{A} = \begin{bmatrix} 1 & p_2 & p_1 \\ 1 & p_3 & p_2 \\ 1 & p_4 & p_3 \\ & \vdots & \\ 1 & p_{n-1} & p_{n-2} \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} p_3 \\ p_4 \\ p_5 \\ \vdots \\ p_n \end{bmatrix}$$

Vyřešením této přeurčené soustavy (pro $n > 5$) ve smyslu nejmenších čtverců dostaneme odhady $\beta_1, \beta_2, \beta_3$, které požijeme ve vzorci pro predikci.

5.4 Je potřeba si rozmyslet, že maticové násobení funguje takto:

$$\mathbf{A}\mathbf{x}-\mathbf{b} = \begin{bmatrix} \sum a_{1j}x_j - b_1 \\ \sum a_{2j}x_j - b_2 \\ \vdots \\ \sum a_{mj}x_j - b_m \end{bmatrix}, \qquad \text{takže} \qquad \sqrt{\mathbf{W}}\,(\mathbf{A}-\mathbf{b}) = \begin{bmatrix} \sqrt{w_1}\,(\sum a_{1j}x_j - b_1) \\ \sqrt{w_2}\,(\sum a_{2j}x_j - b_2) \\ \vdots \\ \sqrt{w_m}\,(\sum a_{mj}x_j - b_m) \end{bmatrix}$$

Promyslete si to, projděte si to podrobně nad rozepsanými maticemi. Symbolem $\sqrt{\mathbf{W}}$ je označena matice $\operatorname{diag}(\sqrt{w_1}, \ldots, \sqrt{w_n})$, což je čtvercová matice rozměru $n \times n$, která obsahuje na diagonále uvedené prvky a mimo ní jsou nuly. Platí $\sqrt{\mathbf{W}}\sqrt{\mathbf{W}} = \mathbf{W}$.

Dále zjevně platí, že $\|\sqrt{\mathbf{W}}(\mathbf{A}\mathbf{x} - \mathbf{b})\|^2 = f(\mathbf{x})$ ze zadání příkladu. Minimalizuje se tedy $\|\sqrt{\mathbf{W}}\,\mathbf{A}\mathbf{x} - \sqrt{\mathbf{W}}\,\mathbf{b}\|^2$. Optimální řešení je řešení soustavy $\sqrt{\mathbf{W}}\,\mathbf{A}\,\mathbf{x} = \sqrt{\mathbf{W}}\,\mathbf{b}$ ve smyslu nejmenších čtverců. Normálový tvar je podle (5.3) v naší úloze $\mathbf{A}^T\sqrt{\mathbf{W}}^T\sqrt{\mathbf{W}}\,\mathbf{A}\,\mathbf{x} = \mathbf{A}^T\sqrt{\mathbf{W}}^T\sqrt{\mathbf{W}}\,\mathbf{b}$, stručněji $\mathbf{A}^T\mathbf{W}\,\mathbf{A}\,\mathbf{x} = \mathbf{A}^T\mathbf{W}\,\mathbf{b}$. Pseudoinverzi si zapište sami.

5.8 c,d) Na to se dá jít různými cestami, my využijeme vzoreček o projektoru. Kdyby matice $\mathbf{A}$ měla ve sloupcích zadané vektory $\mathbf{u}, \mathbf{v}$, a vektor $(2, 0, 1)$ označíme $\mathbf{z}$, pak projekce $\mathbf{z}$ na $\operatorname{rng}\mathbf{A}$ (neboli odpověď na otázku c) je $\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{z}$ a projekce na kolmý doplněk (odpověď na otázku d) je $(\mathbf{I} - \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T)\mathbf{z}$. Jenže, to je zbytečně moc počítání.

Najdeme raději bázi podprostoru $\operatorname{span}\{\mathbf{u}, \mathbf{v}\}^\perp$ vyřešením soustavy $\mathbf{A}\mathbf{x} = \mathbf{0}$. Ta báze obsahuje jediný vektor, například $\mathbf{b} = (2, 5, 3)$. Nyní projekce na kolmý doplněk (odpověď d) je

$$\mathbf{d} = \mathbf{b}\,(\mathbf{b}^T\mathbf{b})^{-1}\mathbf{b}^T\mathbf{z} = \frac{1}{\|\mathbf{b}\|^2}\mathbf{b}(\mathbf{b}^T\mathbf{z}) = \frac{\mathbf{b}^T\mathbf{z}}{\|\mathbf{b}\|^2}\,\mathbf{b} = \frac{7}{38}(2, 5, 3).$$

Projekci na $\operatorname{span}\{\mathbf{u}, \mathbf{v}\}$, (tedy řešení d) spočítáme jako $\mathbf{z} - \mathbf{d} = \frac{1}{38}(62, -35, 17)$.

5.16 Vyjdeme z řešení nejobecnější úlohy, ozna4me ji jako 5.15 x). Ostatní úlohy pak vyřešíme tak, že obecné řešení upravíme na konkrétní případ. To matematici rádi dělají.

5.16 x) Najdeme vzdálenost bodu $\mathbf{v}$ od afinního prostoru $\{\mathbf{x};\ \mathbf{A}^T\mathbf{x} = \mathbf{b}\}$. Budeme postupovat analogicky jako v řešení cvičení 4.21. V tomto cvičení jsme hledali vzdálenost vektoru $\mathbf{v}$ od množiny $\{\mathbf{x}; \mathbf{U}^T\mathbf{x} = \mathbf{b}\}$, použili jsme projekce partikulárního řešení $\mathbf{p}$ a zadaného $\mathbf{v}$ na $\operatorname{rng}\mathbf{U}$ a vzdálenost těch dvou projekcí jsme spočítali. Nyní máme místo $\mathbf{U}^T$ ($\mathbf{U}^T$ měla ortonormální řádky) matici $\mathbf{A}$ (jen s lin. nezávislými řádky). Potřebujeme ty dva vektory kolmo promítnout do $\operatorname{rng}\mathbf{A}$. Projektor na tento podprostor je $\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$, jak víme z teorie k této kapitole. Hledáme tedy vzálenost projekce zadaného vektoru $\mathbf{v}$ od projekce partikulárního řešení $\mathbf{p}$ v $\operatorname{rng}\mathbf{A}$:

$$\|\text{proj.}\mathbf{v} - \text{proj.}\mathbf{p}\| = \|\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{v} - \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{p}\| = \|\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\,(\mathbf{A}^T\mathbf{v} - \mathbf{b})\| =$$

$$= \sqrt{(\mathbf{A}^T\mathbf{v} - \mathbf{b})^T(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\,(\mathbf{A}^T\mathbf{v} - \mathbf{b})} = \sqrt{(\mathbf{A}^T\mathbf{v} - \mathbf{b})^T(\mathbf{A}^T\mathbf{A})^{-1}\,(\mathbf{A}^T\mathbf{v} - \mathbf{b})}$$

5.16 b) V terminologii úlohy 5.16 x) je $\mathbf{a} = \mathbf{A}$ a $b = \mathbf{b}$, takže druhá mocnina hledané vzdálenosti je $\mathbf{b}^T(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{b} = b(\mathbf{a}^T\mathbf{a})^{-1}b = b^2/\|\mathbf{a}\|^2$ a vzdálenost je $\sqrt{b^2/\|\mathbf{a}\|^2} = |b|/\|\mathbf{a}\|$.

15.15 c) Vzdálenost množiny $\{\mathbf{x}; \mathbf{A}\mathbf{x} = \mathbf{b}\}$ od počátku získáme z obecného vzorce x) nahrazením $\mathbf{v}$ nulovým vektorem a přeznačkováním $\mathbf{A}^T \leftarrow \mathbf{A}$, $\mathbf{A} \leftarrow \mathbf{A}^T$. Protože $\mathbf{v} = \mathbf{0}$, je také $\mathbf{A}\mathbf{v} = \mathbf{0}$. Dostáváme $\sqrt{\mathbf{b}^T(\mathbf{A}\mathbf{A}^T)^{-1}\,\mathbf{b}}$.

5.16 d) Vzdálenost bodu $\mathbf{v}$ od nadroviny $\{\mathbf{x}; \mathbf{a}^T\mathbf{x} = b\}$ získáme použitím obecného vzorce z úlohy x), kde $\mathbf{A} = \mathbf{a}^T$, tedy

$$\sqrt{(\mathbf{a}^T\mathbf{v} - b)^T(\mathbf{a}^T\mathbf{a})^{-1}\,(\mathbf{a}^T\mathbf{v} - b)} = \sqrt{\frac{(\mathbf{a}^T\mathbf{v} - b)^2}{\|\mathbf{a}\|^2}} = \frac{|\mathbf{a}^T\mathbf{v} - b|}{\|a\|}$$

Poznámka Ještě připojím důležitou poznámku o projekcích na lineární podprostor daný jako $\operatorname{rng}\mathbf{A}$. Tématicky to souvisí trochu s příklady 4.17 a 5.14. V kapitole 4 jsme se naučili sestavovat projektor ve tvaru $\mathbf{U}\mathbf{U}^T$, což ale funguje jen za předpokladu, že $\mathbf{U}$ má ortonormální sloupce. Co když ale $\mathbf{A}$ jejíž rng je daný podprostor, nemá ortogonální sloupce? Přirozeně lze použít QR rozklad a v roli matice $\mathbf{U}$ použít odpovídající matici $\mathbf{Q}$, kde máme ve sloupcích ortonormální bázi daného podprostoru. Druhá možnost je podívat se do sekce 5.1.1. Nepřehlédněte to, prosím, v písemce se to může s vysokou pravděpodobností vyskytnout.

5.17 $\left\|\begin{bmatrix}\mathbf{a}\\ \mathbf{b}\end{bmatrix}\right\|^2 = \begin{bmatrix}\mathbf{a}^T & \mathbf{b}^T\end{bmatrix}\begin{bmatrix}\mathbf{a}\\ \mathbf{b}\end{bmatrix} = \mathbf{a}^T\mathbf{a} + \mathbf{b}^T\mathbf{b} = \|\mathbf{a}\|^2 + \|\mathbf{b}\|^2$.