Komentář k úlohám probíraným na cvičeních dle plánu 06

Petr Olšák

Shrneme nejprve nejdůležitější poznatky. Předpokládáme $\mathbf{C}$ symetrickou matici z $\mathbb{R}^{n\times n}$ s vlastními čísly $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$. Úloha

$$\min\{\mathbf{x}^T\mathbf{C}\mathbf{x};\ \mathbf{x}^T\mathbf{x} = 1\} \tag{1}$$

má argument minima $\mathbf{v}_1$ (první vlastní vektor) a hodnotu λ_1. Důkaz plyne ze spektrálního rozkladu $\mathbf{C} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ a z vyjádření kvadratické formy $\mathbf{x}^T\mathbf{C}\mathbf{x}$ v nové bázi. Vektor $\mathbf{v}_1$ je první sloupec matice $\mathbf{V}$. Úloha (1) se dá zobecnit pro „vícesloupcové“ $\mathbf{x}$, tedy matici $\mathbf{X}$ s k sloupci, $k < n$:

$$\min\{\operatorname{tr}(\mathbf{X}^T\mathbf{C}\mathbf{X});\ \mathbf{X} \in \mathbb{R}^{n\times k}, \mathbf{X}^T\mathbf{X} = \mathbf{I}\} \tag{2}$$

Rozdělíme-li $\mathbf{V}$ ze spektrálního rozkladu $\mathbf{C}$ do dvou bloků, z nichž první má k sloupců a druhý $n-k$ sloupců, tedy $\mathbf{V} = [\mathbf{X}\ \mathbf{Y}]$, pak argumentem minima je první blok a hodnotou minima je $\lambda_1 + \cdots + \lambda_k$. Toto je *úloha nejmenší stopy*.

Další úlohou je proložení m bodů $\mathbf{a}_i \in \mathbb{R}^n$ podprostorem Y stanovené dimenze d tak, aby součet čtverců vzdálensotí těch bodů od podprostoru Y byl nejmenší. Tedy:

$$\min\Big\{\sum_{i=1}^{m}\Big(\operatorname{dist}(\mathbf{a}_i, Y)\Big)^2;\ \dim Y = d\Big\} \tag{3}$$

Naplníme-li matici $\mathbf{A}$ ve sloupcích souřadnicemi daných bodů $\mathbf{a}_i$, pak tuto úlohu lze převést na úlohu (2) pro $\mathbf{C} = \mathbf{A}\mathbf{A}^T$, $k = n - d$. Úloha (2) pak vrátí ve sloupcích $\mathbf{X}$ bázi kolmého doplňku X k hledanému prostoru Y. Argument, proč tomu tak je, se opírá o představu projekce daných bodů $\mathbf{a}_i$ do prostoru X kolmého na hledaný podprostor Y. Hledá se tedy X, aby součet čtverců velikostí těchto projekcí byl co nejmenší. Ten součet vede (po chvíli výpočtu) na výraz $\operatorname{tr}(\mathbf{X}^T\mathbf{A}\mathbf{A}^T\mathbf{X})$, kde ve sloupcích $\mathbf{X}$ je báze prostoru X. Můžeme tedy matici $\mathbf{V}$ ze spektrálního rozkladu matice $\mathbf{A}\mathbf{A}^T$ rozdělit do dvou bloků $\mathbf{V} = [\mathbf{X}\ \mathbf{Y}]$, v prvním je báze X a ve druhém bloku s d sloupci (tj. s d vlastními vektory příslušejícími největším vlastním číslům matice $\mathbf{A}\mathbf{A}^T$) je báze hledaného podprostoru Y. Úloha se jmenuje PCA (principal component analyzis), protože umožňuje analýzu nikoli daných m bodů, ale jen jejich projekcí do podrostoru Y, což je sice *ztrátová komprese*, ale podprostor Y stanovené dimenze d je volen tak, aby ztráta byla co nejmenší. Projekce výchozích bodů do Y pak můžeme popsat jen souřadnicemi v bázi podprostoru Y, což je typicky daleko méně čísel než původní souřadnice bodů v $\mathbb{R}^n$. Zabýváme se tedy jen těmi podstatnými (principal) souřadnicemi v nové bázi.

Poslední příbuznou úlohou je hledání matice stanovené hodnosti d, která se od zadané matice $\mathbf{A} \in \mathbb{R}^{n\times m}$ liší co nejméně, tedy

$$\min\{\|\mathbf{A} - \mathbf{B}\|;\ \operatorname{rank}\mathbf{B} \le d\} \tag{4}$$

Předpokládáme $d < \operatorname{rank}\mathbf{A}$, protože jinak lze volit $\mathbf{B} = \mathbf{A}$ a úloha je o ničem. Názorné řešení této úlohy plyne ze SVD rozkladu matice $\mathbf{A}$, tedy $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$, kde nyní v matici $\mathbf{S}$ odstraníme (pronulujeme) nejmenší singulární čísla, aby tam zbylo jen d největších čísel. Vzniklou matici $\mathbf{S}_1$ použijeme na sestavení hledané $\mathbf{B} = \mathbf{U}\mathbf{S}_1\mathbf{V}^T$. Je pak jistě $\operatorname{rank}\mathbf{B} = d$ a dá se ukázat, že kvadrát Frobeniovy vzdálenosti $\|\mathbf{A} - \mathbf{B}\|^2$ je opravdu nejmenší možný a je roven sumě kvadrátů odstraněných singulárních čísel.

Úlohu (4) lze řešit i převodem na úlohu (3): ve sloupcích dané matice $\mathbf{A}$ máme body $\mathbf{a}_i$ a provedeme jejich projekci na podrostor Y, který je řešením úlohy (3). Souřadnice těchto projekcí zapíšeme do sloupců matice $\mathbf{B}$. Udělali jsme tedy nejmenší možné změny sloupců se smyslu součtu čtverců velikostí těch změn. A protože sloupce $\mathbf{B}$ leží v prostoru dimenze d, je $\operatorname{rank}\mathbf{B} \le d$. Máme-li $\mathbf{Y}$ druhý blok matice $\mathbf{V}$ ze spektrálního rozkladu matice $\mathbf{A}\mathbf{A}^T$, pak $\mathbf{B} = \mathbf{Y}\mathbf{Y}^T\mathbf{A}$, protože $\mathbf{Y}\mathbf{Y}^T$ je matice kolmého projektoru na podprostor Y.

7.1 Úloha (1) najde minimum ve vlastním vektoru příslušejícímu nejmenšímu vlastnímu číslu. Důkaz se skoro nezmění, když místo minima hledáme maximum. To je nabyto ve vlastním vektoru příslušejícím největšímu vlastnímu číslu λ_n a hodnota maxima je přímo rovna vlastnímu číslu λ_n.

7.2 Matice $\mathbf{C}$ na vstupu do spektrálního rozkladu musí být symetrická. Ovšem každá kvadratická forma má symetrickou matici, tj. stačí takovou matici zvolit. Pokud to neuděláte, dostanete chybný výsledek nebo dokonce vlastní čísla a vektory budou komplexní a matice ani nemusí mít diagonální rozklad.

7.3 Hledá se minimum $\lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$ za podmínky $y_1^2 + y_2^2 + \cdots + y_n^2 = 1$. Eliminujeme y_1^2 z podmínky, tedy $y_1^2 = 1 - y_2^2 - \cdots - y_n^2$ a dosadíme do účelové funkce:

$$\lambda_1(1 - y_2^2 - \cdots - y_n^2) + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2 \;=\; \lambda_1 + (\lambda_2 - \lambda_1)\, y_2^2 + \cdots + (\lambda_n - \lambda_1)\, y_n^2$$

Nejmenší hodnota tohoto výrazu je λ_1, protože dále je ve výrazu nezáporný součet: kvadráty jsou vždy nezáporné a jsou násobené závorkami, které jsou taky nezáporné, protože λ_1 je nejmenší. Nejmenší hodnotu zajistíme při volbě $y_2 = y_3 = \cdots = y_n = 0$, pak tedy musí být $y_1 = \pm 1$.

7.4 Dokážeme, že množiny $P = \{\mathbf{x}^T\mathbf{A}\,\mathbf{x};\ \mathbf{x}^T\mathbf{x} = 1\}$ a $Q = \{\mathbf{x}^T\mathbf{A}\,\mathbf{x}/\mathbf{x}^T\mathbf{x};\ \mathbf{x} \neq \mathbf{o}\}$ se rovnají. Tím pádem se rovnají i jejich minima. Vezmeme nějaké $u \in P$, tj. existuje $\mathbf{x}$ tak, že $u = \mathbf{x}^T\mathbf{A}\,\mathbf{x}$ a navíc $\mathbf{x}^T\mathbf{x} = 1$. Ovšem pak $u = \mathbf{x}^T\mathbf{A}\,\mathbf{x}/\mathbf{x}^T\mathbf{x}$, protože jmenovatel v tomto výrazu je roven jedné. Takže $u \in Q$. Vezmeme nyní $v \in Q$, tedy $v = \mathbf{x}^T\mathbf{A}\,\mathbf{x}/\mathbf{x}^T\mathbf{x}$ pro nějaký nenulový vektor $\mathbf{x}$. Označme $\mathbf{y} = \mathbf{x}/\|\mathbf{x}\|$, tj. $\mathbf{x} = \|\mathbf{x}\|\,\mathbf{y}$. Je zřejmé, že $\mathbf{y}^T\mathbf{y} = 1$. A dosazením za $\mathbf{x}$ do vzorce vyjadřující v máme $v = \|\mathbf{x}\|\,\mathbf{y}^T\mathbf{A}\,\|\mathbf{x}\|\,\mathbf{y}/\mathbf{x}^T\mathbf{x} = \mathbf{y}^T\mathbf{A}\,\mathbf{y}$. Reálné číslo v tedy leží v množině P. Rovnost $P = Q$ je dokázána. Tudíž také $\min P = \min Q$.

7.5 Máme úlohu $\min\{\mathbf{x}^T\mathbf{C}\mathbf{x};\ \mathbf{x}^T\mathbf{x} = 1, \mathbf{v}_i^T\mathbf{x} = 0 \text{ pro } i = 1,\ldots,k\}$. Provedeme $\mathbf{C} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ a úloha po změně souřadnic $\mathbf{y} = \mathbf{V}^T\mathbf{x}$ přechází na $\min\{\mathbf{y}^T\mathbf{\Lambda}\mathbf{y};\ \mathbf{y}^T\mathbf{y} = 1, y_i = 0 \text{ pro } i = 1,\ldots,k\}$, protože požadavek $\mathbf{v}_i^T\mathbf{x} = 0$ pro $i = 1,\ldots,k$ je požadavek na nulový výsledek prvních k levých stran soustavy rovností $\mathbf{V}^T\mathbf{x} = \mathbf{y}$, takže $y_i = 0$ pro $i = 1,\ldots,k$.

7.6 Připomínáme definici, $\langle\mathbf{A},\mathbf{B}\rangle = \sum_i\sum_j a_{ij}b_{ij}$. Platí $\langle\mathbf{A},\mathbf{B}\rangle = \operatorname{tr}(\mathbf{A}^T\mathbf{B})$, protože pro $\mathbf{C} = \mathbf{A}^T\mathbf{B}$ je $c_{ij} = \sum_k a_{ki}b_{kj}$ a dále $\operatorname{tr}(\mathbf{A}^T\mathbf{B}) = \operatorname{tr}(\mathbf{C}) = \sum_j c_{jj} = \sum_j\sum_k a_{kj}b_{kj}$. Zřejmě také $\langle\mathbf{A},\mathbf{B}\rangle = \langle\mathbf{B},\mathbf{A}\rangle$.

Takže pro čtvercovou $\mathbf{A}$ je $\langle\mathbf{AX},\mathbf{X}\rangle = \langle\mathbf{X},\mathbf{AX}\rangle = \operatorname{tr}(\mathbf{X}^T\mathbf{AX})$.

Rovněž $\langle\mathbf{A},\mathbf{XX}^T\rangle = \langle\mathbf{XX}^T,\mathbf{A}\rangle = \operatorname{tr}(\mathbf{XX}^T\mathbf{A})$. Zde jsme využili faktu, že $\mathbf{XX}^T$ je symetrická.

K ověření poslední rovnosti je třeba si uvědomit, že při značení $\mathbf{C} = \mathbf{X}^T\mathbf{AX}$ je $c_{ij} = \mathbf{x}_i^T\mathbf{A}\mathbf{x}_j$ a pro tuto matici počítáme stopu, tedy $\sum_j c_{jj}$.

7.8 Je to úloha (3). Je třeba provést spektrální rozklad matice $\mathbf{AA}^T$, kde souřadnice zadaných bodů jsou ve sloupcích matice $\mathbf{A}$:

```
octave:1> A = [3 -3 4; -2 -3 -2; 1 0 -1; 3 1 0]'
A =

   3  -2   1   3
  -3  -3   0   1
   4  -2  -1   0

octave:2> [V D] = eig(A*A')
V =

  -0.616418   0.377131   0.691232
   0.319558   0.922118  -0.218130
   0.719661  -0.086429   0.688925

D =

Diagonal Matrix

    5.4877         0         0
         0   19.5624         0
         0         0   37.9500

octave:3>
```

báze přímky

báze roviny

největší vlastní číslo

Matlab srovná vlastní čísla od nejmenšího, výsledek tedy čteme ve sloupcích matice $\mathbf{V}$ zprava. Potřebujeme totiž vlastní vektory příslušející d největším vlastním číslům (d je dimenze hledaného podprostoru).

Bázi přímky vložme do matice `Y=V(:,3)`. Pak projekce daných bodů na přímku jsou `Y*Y'*A` a souřadnice těchto projekcí vzhledem k bázi přímky jsou `Y'*A`. Nyní vložme do matice `Y` bázi roviny, tedy `Y=V(:,2:3)`. Pak projekce daných bodů na rovinu jsou `Y*Y'*A` a souřadnice těchto projekcí vzhledem k bázi roviny jsou `Y'*A`.

Hledáme-li přímku neprocházející počátkem, je třeba od všech sloupců matice $\mathbf{A}$ odečíst těžiště $\mathbf{T}$ všech zadaných bodů, tedy sumu sloupců matice $\mathbf{A}$ dělenou počtem sloupců:

```
octave:3> T = A*[1 1 1 1]'*1/4
T =

   1.25000
  -1.25000
   0.25000

octave:4> B = A - T*[1 1 1 1]
B =

   1.75000  -3.25000  -0.25000   1.75000
  -1.75000  -1.75000   1.25000   2.25000
   3.75000  -2.25000  -1.25000  -0.25000

octave:5> [V D] = eig(B*B')
V =

   0.649075   0.378395  -0.659939
  -0.524571   0.850905  -0.028045
  -0.550933  -0.364388  -0.750796

D =

Diagonal Matrix

    0.027894           0           0
           0   17.563479           0
           0           0   32.658627

octave:6>
```

Báze přímky rovnoběžné s hledanou přímkou

Hledanou přímku parametricky zapíšeme jako $\mathbf{T} + \mathrm{Span}(\text{výše vyznačená báze})$.

V zadání byl ještě úkol dospět ke stejným výsledkům pomocí SVD. V tomto případě počítáme přímo SVD rozklad matice $\mathbf{A}$ resp. $\mathbf{B}$. (Nikoli $\mathbf{AA}^T$ ani $\mathbf{BB}^T$.) Použijeme `[U S V] = svd(A)`. Výsledná matice $\mathbf{U}$ má ve sloupcích levé singulární vektory a to jsou vlastní vektory matice $\mathbf{AA}^T$. Kvadráty singulárních čísel matice $\mathbf{A}$ jsou vlastní čísla matice $\mathbf{AA}^T$, jen je stroj seřadil v opačném pořadí, od největšího singulárního čísla k nejmenšímu. V opačném pořadí jsou tedy i vlastní vektory v matici $\mathbf{U}$, takže výsledek pro přímku čteme v levém sloupci matice $\mathbf{U}$.

7.9 Protože je matice $\mathbf{A}$ symetrická, vlastní čísla jsou reálná. Jsou-li nezáporná, pak SVD rozklad $\mathbf{USV}^T$ je shodný se spekrálním rozkladem $\mathbf{U\Lambda U}^T$, kde volíme $\mathbf{S} = \mathbf{\Lambda}$ a $\mathbf{V} = \mathbf{U}$. Vyjdeme z této volby obecně. Je-li nějaké vlastní číslo λ záporné, pak na jeho místo do $\mathbf{S}$ napíšeme $-\lambda$ a odpovídající řádek matice $\mathbf{V}^T$ vynásobíme -1. Tím se celkový součin nezmění. Toto provedeme se všemi zápornými vlastními čísly. Získali jsme $\mathbf{S} \geq \mathbf{O}$ a máme SVD rozklad $\mathbf{A} = \mathbf{USV}^T$.

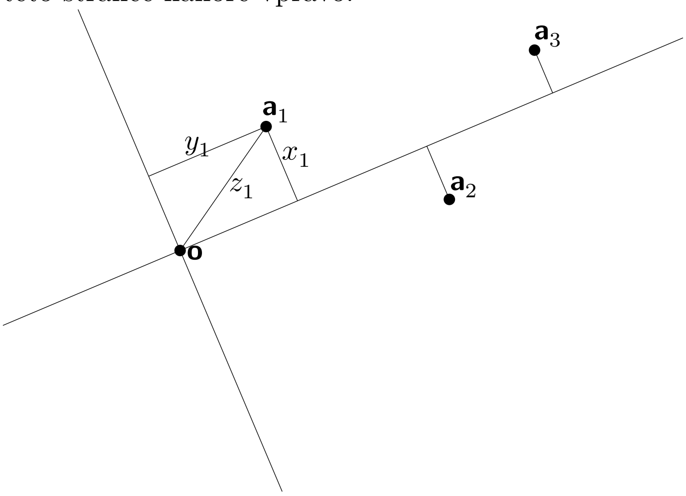
Pozor: tento příklad nemá nic splečného se známým vztahem $\sigma_i^2 = \lambda_i$, protože tento vztah platí pro singulární čísla matice $\mathbf{A}$ a vlastní čísla matice $\mathbf{AA}^T$, což jsou různé matice.

7.10 Je to úloha (4). Je třeba najít SVD rozklad, pronulovat nejmenší diagonální prvek matice S (v tomto případě to je jednička) a najít novou matici s nejbližší hodnotou jako součin $\mathbf{US}_1\mathbf{V}^T$. Výpočet vidíme při pravém okraji.

```
>> A

A =

   -0.8667    0.1333   -1.4667
   -1.0667    0.9333   -0.2667

>> [U S V] = svd(A);
>> S

S =

    2.0000         0         0
         0    1.0000         0

>> S1 = [2 0 0; 0 0 0];
>> U*S1*V'

ans =

   -1.0667    0.5333   -1.0667
   -0.8000    0.4000   -0.8000

>> ans*15

ans =

  -16.0000    8.0000  -16.0000
  -12.0000    6.0000  -12.0000

>>
```

7.12 Problém je, že $\mathbf{AA}^T$ má rozměr $10^6 \times 10^6$, takže raději spočítáme $\mathbf{A}^T\mathbf{A}$, což je rozměru 100×100 a do počítače se nám vejde. Obě matice mají stejná nenulová vlastní čísla. Je-li $\mathbf{v}$ vlastní vektor matice $\mathbf{A}^T\mathbf{A}$, pak $\mathbf{u} = \mathbf{Av}$ je vlastní vektor matice $\mathbf{AA}^T$. Ta se nám sice nevejde do počítače, ale její vlastní vektory příslušející nenulovým vlastním číslům umíme tedy spočítat z vlastních vektorů matice $\mathbf{A}^T\mathbf{A}$. Vzorec pro $\mathbf{u}$ lze odvodit z definice vlastních čísel a vektorů: Je-li $\mathbf{v}$ vlastní vektor $\mathbf{A}^T\mathbf{A}$, pak je $\mathbf{A}^T\mathbf{Av} = \lambda\mathbf{v}$. Vynásobením zleva maticí $\mathbf{A}$ a přezávorkováním máme $(\mathbf{AA}^T)(\mathbf{Av}) = \lambda(\mathbf{Av})$. Z toho jednak vidíme, že λ je také vlastní číslo matice $\mathbf{AA}^T$, a dále ve druhé a třetí závorce vidíme její vlastní vektor.

7.13 Klíčové heslo v odpovědi na toto cvičení zní: „Pythagorova věta“. Body v prostoru i jeho nulový vektor (počátek) jsou pevně dány, tj. velikosti z_i jsou dány. Hledáme x_i (vzdálenosti od podprostoru) aby součet jejich kvadrátů byl minimální, nebo hledáme y_i (délky jejich ortogonálních projekcí) aby součet těchto kvadrátů byl maximální. Přitom je $x_i^2 + y_i^2 = z_i^2 = \text{const.}$, takže když minimalizujeme sumu x_i^2, je to totéž, jako když maximalizujeme sumu y_i^2. Ilustrace je na této stránce nahoře vpravo.



7.14 Zahrňme souřadnice všech m bodů do sloupců matice $\mathbf{A}$. Jejich projekce $P(\mathbf{A})$ získáme maticovým názobením $\mathbf{PA}$. Jejich těžiště najdeme maticovým násobením $\frac{1}{m}\mathbf{A}\,\mathbf{1}$. Tyto dvě operace skutečně komutují: $\mathbf{P}(\frac{1}{m}\mathbf{A}\,\mathbf{1}) = \frac{1}{m}(\mathbf{PA})\mathbf{1}$.

7.15 a) Vzdálenost bodu $\mathbf{z}$ od přímky (nadroviny v $\mathbb{R}^2$) zadané vzorcem $\mathbf{a}^T\mathbf{x} = \beta$ jsme spočítali v příkladu 5.16 (d), tato vzdálenost je $|\mathbf{a}^T\mathbf{z} - \beta|/\|\mathbf{a}\|$. V terminologii našeho příkladu je $\mathbf{a} = (a, b)$, $\beta = -c$, takže vzálenost bodu (x_i, y_i) od naší přímky je $|ax_i + by_i + c|/\sqrt{a^2 + b^2}$. Suma čtverců těchto vzdáleností tedy je $1/(a^2 + b^2)\sum_i(ax_i + by_i + c)^2$, což je něco jiného než $\sum_i(ax_i + bx_i + c)^2$. Takže úloha neminimalizuje součet čverců vzdáleností.

b) Platí-li $a^2 + b^2 = 1$, je suma čtverců vzdáleností přímo rovna $\sum_i(ax_i + bx_i + c)^2$, což je účelová funkce naší úlohy. Prokládáme tedy zadanými body přímku, abychom minimalizovali sumu čtverců vzdáleností, což je úloha (3) při volbě $d = 1$. Není to přímo úloha (3): nehledá se lineární podprostor ale afinní, je tedy třeba zadané body nejprve posunout proti těžišti, najít podprostor a nakonec posunout podprotor zpět o těžiště, viz třeba cvičení 7.8 c.

c) Vzdálenost bodu (x_i, y_i) od kuželosečky se rozhodně ne vždy rovná vzorci $|Q(x_i, y_i)|$, takže úloha s účelovou funkcí $\sum_i Q^2(x_i, y_i)$ neminimalizuje součet čtverců vzdáleností. Například parabola $x^2 - y = 0$ (tedy $a = 1, e = -1$, ostatní jsou nuly) vyhovuje podmínce $a^2 + b^2 + c^2 = 1$ a ze cvičení 1.1 (i) o ni víme, že je od bodu $(3, 0)$ vzdálená $\sqrt{5}$, zatímco $Q(3, 0) = 9$.