Теоретическая домашняя работа «Матричные вычисления и матричное дифференцирование»

ММРО 317 группы, ММП ВМК МГУ

Борисов Иван Максимович

21.10.2023

1 Тождество Вудбери

$$\forall A \in \mathbb{R}^{n\times n}, C \in \mathbb{R}^{m\times m}, U \in \mathbb{R}^{n\times m}, V \in \mathbb{R}^{m\times n}, |A| \neq 0, |C| \neq 0$$
$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}$$

►
1) Рассмотрим представление I:
Пусть $P \in \mathbb{R}^{n\times n} : |I + P| \neq 0$
Тогда $I = (I + P)^{-1}(I + P) = (I + P)^{-1} + (I + P)^{-1}P$

$$(I + P)^{-1} = I - (I + P)^{-1}P, \tag{1}$$

2) Пусть $X, Y \in \mathbb{R}^{n\times n} : |I + XY| \neq 0, |I + YX| \neq 0$
Рассмотрим $X(I + YX) = (X + XYX) = (I + XY)X$

$$(I + XY)^{-1}X = X(I + YX)^{-1}, \tag{2}$$

3) Положим в (1) $P = A^{-1}UCV$
$\Rightarrow (I + A^{-1}UCV) = I - (I + \underline{A^{-1}U}\underline{CV})^{-1}\underline{A^{-1}U}\underline{CV}$
Пользуясь (2), получаем $(A + UCV)^{-1}A = I - A^{-1}U(I + CVA{-}1U)^{-1}CV$
Так как $|A| \neq 0$: $(A + UCV)^{-1} = A^{-1} - A^{-1}U(I + CVA{-}1U)^{-1}CVA^{-1}$
Так как $|C| \neq 0$: $(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA{-}1U)^{-1}VA^{-1}$

■

2 Упростить выражения

а) $||uv^T - A||_F^2 - ||A||_F^2$, где $u \in \mathbb{R}^m, v \in \mathbb{R}^n, A \in \mathbb{R}^{m\times n}$

$$\begin{aligned}
||uv^T - A||_F^2 - ||A||_F^2 &= \langle uv^T - A, uv^T - A\rangle - \langle A, A\rangle = (uv^T - A)^T(uv^T - A) - A^TA = \\
&= (vu^T - A^T)(uv^T - A) - A^TA = vu^Tuv^T - vu^TA - A^Tuv^T + \cancel{A^TA} - \cancel{A^TA} \\
&= (uv^T)^Tuv^T - vu^TA - A^Tuv^T = ||uv^T||_F^2 - vu^TA - (A^Tuv^T)^T = \\
&= ||uv^T||_F^2 - 2vu^TA
\end{aligned}$$

b) $tr((2I + aa^T)^{-1}(uv^T + vu^T))$, где $a, u, v \in \mathbb{R}^n$

$$
\begin{aligned}
tr((2I + aa^T)^{-1}(uv^T + vu^T)) &= tr((\frac{1}{2}I - \frac{1}{4}aI(\frac{1}{2}I + \frac{1}{2}a^T aI)^{-1}a^T I)(uv^T + vu^T)) = \\
&= \frac{1}{2}tr((I - \frac{aa^T}{2 + ||a||^2}I)(uv^T + vu^T)) = \\
&= \frac{1}{2}tr(uv^T) + \frac{1}{2}tr(vu^T) - \frac{1}{2(2 + ||a||^2)}tr(aa^T(uv^T + vu^T)) = \langle u, v\rangle - \frac{\langle a, v\rangle\langle a, u\rangle}{2 + ||a||^2}
\end{aligned}
$$

При первом переходе использовалось тождество Вудбери при $A = 2I, U = a, C = I, V = a^T$

c) $\sum_{i=1}^n \langle S^{-1}a_i, a_i\rangle$ где $a_1, a_2, \dots, a_n \in \mathbb{R}^d, S = \sum_{i=1}^n a_i a_i^T, |S| \neq 0$

$$
\begin{aligned}
\sum_{i=1}^n \langle S^{-1}a_i, a_i\rangle &= tr(\sum_{i=1}^n \langle S^{-1}a_i, a_i\rangle) = \sum_{i=1}^n tr(\langle S^{-1}a_i, a_i\rangle) = \sum_{i=1}^n tr(a_i^T S^{-1} a_i) = \sum_{i=1}^n tr(S^{-1}a_i^T a_i) = \\
&= tr(S^{-1}\sum_{i=1}^n a_i^T a_i) = tr(S^{-1}S) = tr(I_{d\times d}) = d
\end{aligned}
$$

3 Поиск производных 1 и 2 порядков

a) $f : E \to \mathbb{R}, f(t) = det(A - tI)$, где $A \in \mathbb{R}^{n\times n}, E = \{t \in \mathbb{R} : |A - tI| \neq 0\}$

$$
d(|A - tI|) = |A - tI|\langle (A - tI)^{-T}, d(A - tI)\rangle = -|A - tI|\langle (A - tI)^{-T}, I\rangle dt \Rightarrow \boxed{f'(t) = -|A - tI|\langle (A - tI)^{-T}, I\rangle}
$$

$$
\begin{aligned}
-d(|A - tI|\langle (A - tI)^{-T}, I\rangle) &= -(-|A - tI|\langle (A - tI)^{-T}, I\rangle\langle (A - tI)^{-T}, I\rangle dt + d((A - tI)^{-T})I + 0) = \\
&= |A - tI|\langle (A - tI)^{T}, I\rangle^2 dt - (d(A - tI)^{-1})^T = \\
&= |A - tI|\langle (A - tI)^{-T}, I\rangle^2 dt - (-(A - tI)^{-1}dt(A - tI)^{-1})^T = \\
&= |A - tI|\langle (A - tI)^{-T}, I\rangle^2 dt + (A - tI)^{-T}dt(A - tI)^{-T} \\
&= (|A - tI|\langle (A - tI)^{-T}, I\rangle^2 + (A - tI)^{-2T})dt
\end{aligned}
$$

$$
\Rightarrow \boxed{f''(t) = |A - tI|\langle (A - tI)^{-T}, I\rangle^2 + (A - tI)^{-2T}}
$$

b) $f : \mathbb{R}_{++} \to \mathbb{R}, f(t) = ||(A + tI)^{-1}b||$, где A - неотр. определенная, $b \in \mathbb{R}^n$

$$
\begin{aligned}
d(||(A + tI)^{-1}b||) &= d(\langle (A + tI)^{-1}b, (A + tI)^{-1}b\rangle^{\frac{1}{2}}) = \frac{1}{2||(A + tI)^{-1}b||}\cdot d(\langle (A + tI)^{-1}b, (A + tI)^{-1}b\rangle) = \\
&= \frac{1}{\cancel{2}||(A + tI)^{-1}b||}\cdot \cancel{2}\langle (A + tI)^{-1}b, d((A + tI)^{-1}b)\rangle \\
&= \frac{1}{||(A + tI)^{-1}b||}\langle (A + tI)^{-1}b, b(-(A + tI)^{-1}dt(A + tI)^{-1})\rangle \\
&= \frac{-\langle (A + tI)^{-1}b, (A + tI)^{-2}b\rangle}{||(A + tI)^{-1}b||}dt \Rightarrow \boxed{f'(t) = \frac{-\langle (A + tI)^{-1}b, (A + tI)^{-2}b\rangle}{||(A + tI)^{-1}b||}}
\end{aligned}
$$

$-d(\frac{\langle (A+tI)^{-1}b, (A+tI)^{-2}b\rangle}{||(A+tI)^{-1}b||})-?$

Рассм. числитель:

$$
\begin{aligned}
d(\langle (A + tI)^{-1}b, (A + tI)^{-2}b\rangle) &= \langle d((A + tI)^{-1}b), (A + tI)^{-2}b\rangle + \langle (A + tI)^{-1}b, d((A + tI)^{-2}b)\rangle = \\
&= \langle -(A + tI)^{-1}dt(A + tI)^{-1}b, (A + tI)^{-2}b\rangle dt + \langle (A + tI)^{-1}b, -2(A + tI)^{-3}b\rangle dt = \\
&= -||(A + tI)^{-2}b||^2 dt - 2\langle (A + tI)^{-1}b, (A + tI)^{-3}b\rangle dt
\end{aligned}
$$

Тогда: $-d(\frac{\langle (A+tI)^{-1}b,(A+tI)^{-2}b\rangle}{||(A+tI)^{-1}b||}) =$

$$= -\frac{(-||(A+tI)^{-2}b||^2 - 2\langle (A+tI)^{-1}b,(A+tI)^{-3}b\rangle)\cdot||(A+tI)^{-1}b||dt - \langle (A+tI)^{-1}b,(A+tI)^{-2}b\rangle \cdot \frac{-\langle (A+tI)^{-1}b,(A+tI)^{-2}b\rangle}{||(A+tI)^{-1}b||}dt}{||(A+tI)^{-1}b||^2}$$

$$\Rightarrow f''(t) = \frac{(||(A+tI)^{-2}b||^2 + 2\langle (A+tI)^{-1}b,(A+tI)^{-3}b\rangle)\cdot||(A+tI)^{-1}b|| - \langle (A+tI)^{-1}b,(A+tI)^{-2}b\rangle \cdot \frac{\langle (A+tI)^{-1}b,(A+tI)^{-2}b\rangle}{||(A+tI)^{-1}b||}}{||(A+tI)^{-1}b||^2}$$

4 Поиск производных градиента и гессиана

a) $f : \mathbb{R}^n \to \mathbb{R}, f(x) = \frac{1}{2}||xx^T - A||_F^2$, где A - симметрическая

$$\begin{aligned} d(\frac{1}{2}||xx^T - A||_F^2) &= \frac{1}{2}d\langle xx^T - A, xx^T - A\rangle = \langle xx^T - A, d(xx^T)\rangle = \\ &= \langle xx^T - A, dx\cdot x^T\rangle + \langle xx^T - A, dx^T\cdot x\rangle = \langle (xx^T - A)x, dx\rangle + \langle (xx^T - A)^T x, dx\rangle = \\ &= 2\langle (xx^T - A)x, dx\rangle \Rightarrow \boxed{\nabla f = 2(xx^T - A)x} \end{aligned}$$

$$\begin{aligned} 2d(\langle (xx^T - A)x, dx_1\rangle) &= 2\langle d((xx^T - A)x), dx_1\rangle = 2\langle d(xx^T - A)\cdot x + (xx^T - A)dx_2, dx_1\rangle = \\ &= 2\langle (dx_2\cdot x^T + (x\cdot dx_2^T))\cdot x + (xx^T - A)dx_2, dx_1\rangle = \\ &= 2\langle dx_2\cdot x^T x, dx_1\rangle + 2\langle x\cdot dx_2^T\cdot x, dx_1\rangle + 2\langle (xx^T - A)dx_2, dx_1\rangle = \\ &= 2x^T x\langle dx_2, dx_1\rangle + 2\langle xx^T dx_2, dx_1\rangle + 2\langle (xx^T - A)dx_2, dx_1\rangle = \\ &= \langle 2(x^T xI + xx^T + (xx^T - A))dx_2, dx_1\rangle \Rightarrow \boxed{\nabla^2 f = 2(x^T xI + 2xx^T - A)} \end{aligned}$$

b) $f : \mathbb{R}^n\backslash\{\theta\} \to \mathbb{R}, f(x) = \langle x, x\rangle^{\langle x,x\rangle}$

$$\begin{aligned} d(e^{\langle x,x\rangle ln(\langle x,x\rangle)}) &= \langle x,x\rangle^{\langle x,x\rangle} d(\langle x,x\rangle ln(\langle x,x\rangle) = \langle x,x\rangle^{\langle x,x\rangle}(d(\langle x,x\rangle)ln\langle x,x\rangle + \langle x,x\rangle d(ln\langle x,x\rangle)) = \\ &= \langle x,x\rangle^{\langle x,x\rangle}(\langle 2xln\langle x,x\rangle, dx\rangle + \langle 2x, dx\rangle) = \langle 2\langle x,x\rangle^{\langle x,x\rangle}x(1 + ln\langle x,x\rangle), dx\rangle \end{aligned}$$

$$\Rightarrow \boxed{\nabla f(x) = \langle 2\langle x,x\rangle^{\langle x,x\rangle}x(1 + ln\langle x,x\rangle)}$$

$$\begin{aligned} 2\cdot d\Big(\langle\langle x,x\rangle^{\langle x,x\rangle}x(1 + ln\langle x,x\rangle), dx_1\rangle\Big) &= 2\Big(\langle x(1 + ln\langle x,x\rangle)d\big(\langle x,x\rangle^{\langle x,x\rangle}\big) + \langle x,x\rangle^{\langle x,x\rangle}d\big(x(1 + ln\langle x,x\rangle)\big), dx_1\rangle\Big) = \\ &= 2\Big(\big\langle x(1 + ln\langle x,x\rangle)\langle 2\langle x,x\rangle^{\langle x,x\rangle}x(1 + ln\langle x,x\rangle), dx_2\rangle + \langle x,x\rangle^{\langle x,x\rangle}\big((1 + ln\langle x,x\rangle)dx_2 + xd(ln\langle x,x\rangle)\big), dx_1\big\rangle\Big) = \\ &= 2\Big(\big\langle 2\langle x,x\rangle^{\langle x,x\rangle}(1 + ln\langle x,x\rangle)^2 x\langle x, dx_2\rangle + \langle x,x\rangle^{\langle x,x\rangle}(1 + ln\langle x,x\rangle)Idx_2 + 2\frac{x}{\langle x,x\rangle}\langle x, dx_2\rangle, dx_1\big\rangle\Big) = \\ &= 2\Big(\big\langle 2\langle x,x\rangle^{\langle x,x\rangle}(1 + ln\langle x,x\rangle)^2 x dx_2^T x + 2\langle x,x\rangle^{\langle x,x\rangle}(1 + ln\langle x,x\rangle)Idx_2 + 4\frac{x(dx_2)^T x}{\langle x,x\rangle}, dx_1\big\rangle = \\ &= \big\langle 4\langle x,x\rangle^{\langle x,x\rangle}(1 + ln\langle x,x\rangle)^2 xx^T dx_2 + 2\langle x,x\rangle^{\langle x,x\rangle}(1 + ln\langle x,x\rangle)Idx_2 + 4\frac{xx^T}{\langle x,x\rangle}dx_2, dx_1\big\rangle \end{aligned}$$

$$\Rightarrow \boxed{\nabla^2 f(x) = 4\langle x,x\rangle^{\langle x,x\rangle - 1}\big(\langle x,x\rangle(1 + ln\langle x,x\rangle)^2 + 1\big)xx^T + 2\langle x,x\rangle^{\langle x,x\rangle}(1 + ln\langle x,x\rangle)I}$$

c) $f : \mathbb{R}^n \to \mathbb{R}, f(x) = ||Ax - b||^p$, где $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m, p \geq 2$

$$\begin{aligned} d(||Ax - b||^p) &= d(\langle Ax - b, Ax - b\rangle^{\frac{p}{2}}) = \frac{p}{2}||Ax - b||^{p-2}d(\langle Ax - b, Ax - b\rangle) = \\ &= \frac{p}{\cancel{2}}||Ax - b||^{p-2}\cancel{2}\langle Ax - b, Adx\rangle = \langle p||Ax - b||^{p-2}A^T(Ax - b), dx\rangle \end{aligned}$$

$$\Rightarrow \boxed{\nabla f(x) = p||Ax-b||^{p-2}A^T(Ax-b)}$$

$$\begin{aligned}
p\cdot d(\langle ||Ax-b||^{p-2}A^T(Ax-b), dx_1\rangle) &= p\cdot \langle A^T(Ax-b)d(||Ax-b||^{p-2}) + ||Ax-b||^{p-2}A^T d(Ax-b), dx_1\rangle = \\
&= p\cdot \big\langle A^T(Ax-b)\big\langle (p-2)||Ax-b||^{p-4}A^T(Ax-b), dx_2\big\rangle + ||Ax-b||^{p-2}A^T A dx_2, dx_1\rangle) = \\
&= p\cdot \big\langle (p-2)||Ax-b||^{p-4}A^T(Ax-b)\langle dx_2, A^T(Ax-b)\rangle + ||Ax-b||^{p-2}A^T A dx_2, dx_1\rangle) = \\
&= p\cdot \big\langle (p-2)||Ax-b||^{p-4}A^T(Ax-b)(A^T(Ax-b))^T dx_2 + ||Ax-b||^{p-2}A^T A dx_2, dx_1\rangle) = \\
&= p\cdot \big\langle (p-2)||Ax-b||^{p-4}A^T(Ax-b)(Ax-b)^T A dx_2 + ||Ax-b||^{p-2}A^T A dx_2, dx_1\rangle) =
\end{aligned}$$

$$\Rightarrow \boxed{\nabla^2 f(x) = p(p-2)||Ax-b||^{p-4}A^T(Ax-b)(Ax-b)^T A + ||Ax-b||^{p-2}A^T A}$$

5 Знакоопределенность второго дифференциала

a) $f : \mathbb{S}^n_{++} \to \mathbb{R}, f(X) = tr(X^{-1})$, где S^n_{++} — множество положительно определенных матриц.

$$d(tr(X^{-1})) = d(\langle X^{-1}, I\rangle) = \langle -X^{-1}dXX^{-1}, I\rangle = -\langle X^{-2}, dX\rangle \Rightarrow \boxed{\nabla f(X) = -X^{-2}}$$

$$\begin{aligned}
d(-\langle X^{-2}, dX_1\rangle) &= -\langle d(X^{-1}X^{-1}), dX_1\rangle = -\langle -X^{-1}dX_2X^{-1}X^{-1} - X^{-1}X^{-1}dX_2X^{-1}, dX_1\rangle = \\
&= \langle X^{-1}dX_2X^{-2} + X^2dX_2X^{-1}, dX_1\rangle
\end{aligned}$$

Фиксируем $\forall H$ — симметрическую матрицу:

$$\begin{aligned}
D^2f(X)[H,H] &= \langle X^{-1}HX^{-2} + X^{-2}HX^{-1}, H\rangle = \langle X^{-1}HX^{-2}, H\rangle + \langle X^{-2}HX^{-1}, H\rangle = \\
&= \langle X^{-\frac{1}{2}}HX^{-1}, X^{-\frac{1}{2}}HX^{-1}\rangle + \langle X^{-1}HX^{-\frac{1}{2}}, X^{-1}HX^{-\frac{1}{2}}\rangle = \\
&= ||X^{-\frac{1}{2}}HX^{-1}||_F^2 + ||X^{-1}HX^{-\frac{1}{2}}||_F^2 > 0, \forall X \in \mathbb{S}^n_{++},
\end{aligned}$$

b) $f : \mathbb{S}^n_{++} \to \mathbb{R}, f(X) = |X|^{\frac{1}{n}}$, где S^n_{++} — множество положительно определенных матриц.

$$d(|X|^{\frac{1}{n}}) = \tfrac{1}{n}|X|^{\frac{1-n}{n}}d(|X|) = \tfrac{1}{n}|X|^{\frac{1-n}{n}}|X|\langle X^{-T}, dX\rangle = \tfrac{1}{n}|X|^{\frac{1}{n}}\langle X^{-1}, dX\rangle \Rightarrow \boxed{\nabla f(X) = \frac{1}{n}|X|^{\frac{1}{n}}X^{-1}}$$

$$\begin{aligned}
d(\langle \frac{1}{n}|X|^{\frac{1}{n}}X^{-1}, dX_1\rangle) &= \frac{1}{n}\langle d(|X|^{\frac{1}{n}}X^{-1}), dX_1\rangle = \frac{1}{n}\big\langle \langle \frac{1}{n}|X|^{\frac{1}{n}}X^{-1}, dX_2\rangle X^{-1} - |X|^{\frac{1}{n}}X^{-1}dX_2X^{-1}, dX_1\big\rangle = \\
&= \frac{|X|^{\frac{1}{n}}}{n^2}\langle X^{-T}dX_2X^{-1} - nX^{-1}dX_2X^{-1}, dX_1\rangle = \\
&= \frac{|X|^{\frac{1}{n}}}{n^2}\langle X^{-1}dX_2X^{-1} - nX^{-1}dX_2X^{-1}, dX_1\rangle
\end{aligned}$$

Фиксируем $\forall H$ — симметрическую матрицу:

$$\begin{aligned}
D^2f(X)[H,H] &= \frac{|X|^{\frac{1}{n}}}{n^2}\langle X^{-1}HX^{-1} - nX^{-1}HX^{-1}, H\rangle = \frac{|X|^{\frac{1}{n}}}{n^2}\langle X^{-1}HX^{-1}, H\rangle - \frac{|X|^{\frac{1}{n}}}{n}\langle X^{-1}HX^{-1}, H\rangle = \\
&= \frac{|X|^{\frac{1}{n}}}{n^2}\langle X^{-1}HX^{-1}H, I\rangle - \frac{|X|^{\frac{1}{n}}}{n}\langle X^{-1}HX^{-1}, H\rangle \le \\
&\le \frac{|X|^{\frac{1}{n}}}{n^2}||X^{-1}HX^{-1}H||_F||I||_F - \frac{|X|^{\frac{1}{n}}}{n}\langle X^{-1}HX^{-1}, H\rangle = \\
&= \frac{|X|^{\frac{1}{n}}}{n^{\frac{3}{2}}}||X^{-1}HX^{-1}H||_F - \frac{|X|^{\frac{1}{n}}}{n}||X^{-\frac{1}{2}}HX^{-\frac{1}{2}}||_F^2 \le 0
\end{aligned}$$

6 Точки стационарности

a)$f : \mathbb{R}^n \to \mathbb{R}, f(x) = \langle c, x\rangle + \frac{\sigma}{3}||x||^3$, где $c \in \mathbb{R}^n, c \neq \theta, \sigma > 0$

$$d(\langle c,x\rangle + \frac{\sigma}{3}||x||^3) = \langle c, dx\rangle + \frac{\sigma}{\cancel{3}}\cdot \cancel{3}||x||^2 d(||x||) = \langle c, dx\rangle + \sigma||x||^2 d\langle x,x\rangle^{\frac{1}{2}} = \langle c, dx\rangle + \sigma||x||^{\cancel{2}}\frac{\cancel{2}}{2\cancel{||x||}}\langle x, dx\rangle =$$
$$= \langle c, dx\rangle + \sigma||x||\langle x, dx\rangle$$

$\Rightarrow \nabla f(x) = c + \sigma||x||x = 0$

$c + \sigma||x||x = 0 \Rightarrow \sigma||x||x = -c$

Рассмотрим уравнения по норме: $\sigma||x||^2 = ||c|| \Rightarrow ||x|| = \sqrt{\frac{||c||}{\sigma}}$

$c + \sigma\sqrt{\frac{||c||}{\sigma}}x = 0 \Rightarrow \boxed{x = -\frac{c}{\sigma}\sqrt{\frac{\sigma}{||c||}}}$

b)$f : E \to \mathbb{R}, f(x) = \langle a, x\rangle - ln(1 - \langle b, x\rangle)$, где $a, b \in \mathbb{R}^n, a, b \neq \theta, E = \{x \in \mathbb{R}^n | \langle b, x\rangle < 1\}$

$$d(\langle a,x\rangle - ln(1-\langle b,x\rangle)) = \langle a, dx\rangle - \frac{1}{1-\langle b,x\rangle}d(-\langle b,x\rangle) = \langle a, dx\rangle + \frac{1}{1-\langle b,x\rangle}\langle b, dx\rangle$$

$\Rightarrow \nabla f(x) = a + \frac{b}{1-\langle b,x\rangle} = 0$

$$a + \frac{b}{1-\langle b,x\rangle} = 0 \Rightarrow (1-\langle b,x\rangle)^{-1}b = -a$$

Таким образом b и a - коллинеарные векторы, т.е. $a = \lambda^{-1}b, \lambda \in \mathbb{R}^-$

$$(1-\langle b,x\rangle)^{-1}b = -\lambda^{-1}b \Rightarrow 1 - \langle b,x\rangle = -\lambda \Rightarrow \langle b,x\rangle = 1 + \lambda$$

$$\boxed{\forall x \in E : \langle b,x\rangle = 1+\lambda, \lambda \in \mathbb{R}^- : a = \lambda^{-1}b}$$

c)$f : \mathbb{R}^n \to \mathbb{R}, f(x) = \langle c, x\rangle e^{-\langle Ax,x\rangle}$, где $c \in \mathbb{R}^n, c \neq \theta, A$ — положительно определенная матрица

$$d(\langle c,x\rangle e^{-\langle Ax,x\rangle}) = e^{-\langle Ax,x\rangle}\langle c, dx\rangle - \langle c,x\rangle e^{-\langle Ax,x\rangle}d(\langle Ax,x\rangle) = e^{-\langle Ax,x\rangle}\big(\langle c,dx\rangle - 2\langle c,x\rangle\langle Ax,dx\rangle\big)$$

$\Rightarrow \nabla f(x) = e^{-\langle Ax,x\rangle}\big(c - 2\langle c,x\rangle Ax\big) = 0 \Rightarrow c - 2Ax\langle c,x\rangle = 0 \Rightarrow x = \frac{A^{-1}c}{2\langle c,x\rangle}$

Скалярно умножим на c:

$\langle x,c\rangle = \frac{\langle A^{-1}c,c\rangle}{2\langle c,x\rangle} \Rightarrow \langle x,c\rangle^2 = \frac{\langle A^{-1}c,c\rangle}{2} \Rightarrow \langle x,c\rangle = \sqrt{\frac{\langle A^{-1}c,c\rangle}{2}}$

Таким образом:

$$\boxed{x = \frac{A^{-1}c}{\sqrt{2\langle A^{-1}c,c\rangle}}}$$

7 Предел матриц

$$X \in \mathbb{S}^n_{++} \Rightarrow \lim_{k\to\infty} tr(X^{-k} - (X^k + X^{2k})^{-1}) - ?$$

$$S^n_{++} = \{A \in \mathbb{R}^{n\times n} | \langle Ax, x\rangle > 0, \forall x \in \mathbb{R}^n \backslash \{\theta\}\}$$

▶

$$X^{-k} - (X^k + X^{2k})^{-1} = X^{-k} - (I + X^k)^{-1}X^{-k} = X^{-k} - (\sum_{i=0}^{\infty}(-1)^i X^{ik})X^{-k} = X^{-k} - (\sum_{i=0}^{\infty}(-1)^i X^{k(i-1)}) =$$

$$= \cancel{X^{-k}} - \cancel{X^{-k}} - \sum_{i=1}^{\infty}(-1)^i X^{k(i-1)} = \sum_{i=1}^{\infty}(-1)^{i-1} X^{k(i-1)} = \sum_{j=0}^{\infty}(-1)^j X^{kj} = (I + X^k)^{-1}$$

$$\Rightarrow \lim_{k\to\infty} tr(X^{-k} - (X^k + X^{2k})^{-1}) = \lim_{k\to\infty} tr(I + X^k)^{-1}$$

Так как $X \in S^n_{++}$, то применяя спектральное разложение, имеем:

$$X = \sum_i \lambda_i P_i$$

λ_i - собств. зн. X, P_i - орт. проекторы на собственные подпространства W_i и $trP_i = |W_i| \overset{\text{def}}{=\!=} d_i$
Тогда:

$$X^k = \sum_i \lambda_i^k P_i$$

$$tr(I + X^k) = \sum_i d_i(1 + \lambda_i^k)$$

$$\lim_{k\to\infty} tr(I + X^k)^{-1} = \lim_{k\to\infty} \sum_i \frac{d_i}{(1 + \lambda_i^k)}$$

$$\lim_{k\to\infty} tr(I + X^k)^{-1} = \lim_{k\to\infty} \sum_i \frac{d_i}{(1 + \lambda_i^k)}$$

Если $\lambda_i > 1 \Rightarrow \lim_{k\to\infty} \frac{d_i}{1+\lambda_i^k} = 0$
Если $\lambda_i = 1 \Rightarrow \lim_{k\to\infty} \frac{d_i}{1+\lambda_i^k} = \frac{d_i}{2}$
Если $\lambda_i < 1 \Rightarrow \lim_{k\to\infty} \frac{d_i}{1+\lambda_i^k} = d_i$
Таким образом:

$$\boxed{\lim_{k\to\infty} tr(I + X^k)^{-1} = \sum_{\lambda_i:\lambda_i<1} d_i + \sum_{\lambda_i:\lambda_i=1} \frac{d_i}{2}}$$

■

8 Метод наименьших компонент

Пусть $\{x_i\}_{i=1}^n, x_i \in \mathbb{R}^D$ - выборка, размерность которой необходимо понизить до d, с помощью проектора, задаваемого матрицей $P \in \mathbb{R}^{D\times d}$.
В таком случае ортогональная проекция $x_{pr} = P(P^TP)^{-1}P^Tx$.
Тогда для нахождения P рассмотрим оптимизационную задачу:

$$F(P) = \sum_{i=1}^{n} ||x_i - P(P^TP)^{-1}P^Tx_i||^2 = Ntr((I - P(P^TP)^{-1}P^T)^2S) \to \min_P$$

где $S = \frac{1}{N}\sum_{i=1}^n x_i x_i^T$ - выборочная матрица ковариации для нормированной выборки

8.1 Поиск градиента

$$
\begin{aligned}
d(F(P)) = Nd(tr((I - P(P^TP)^{-1}P^T)^2S)) = Nd\langle(I - P(P^TP)^{-1}P^T)^2S, I\rangle = N\langle d\big((I - P(P^TP)^{-1}P^T)^2S\big), I\rangle = \\
= N\langle 2(I - P(P^TP)^{-1}P^T)d\big((I - P(P^TP)^{-1}P^T)\big)S, I\rangle \\
= -2N\langle(I - P(P^TP)^{-1}P^T)d\big((P(P^TP)^{-1}P^T)\big)S, I\rangle
\end{aligned}
$$

Рассмотрим $d(P(P^TP)^{-1}P^T)$:

$$
\begin{aligned}
d(P(P^TP)^{-1}P^T) &= d(P(P^TP)^{-1})P^T + P(P^TP)^{-1}dP^T = \\
&= \big(dP(P^TP)^{-1} + Pd((P^TP)^{-1}\big)P^T + P(P^TP)^{-1}dP^T = \\
&= \big(dP(P^TP)^{-1} - P(P^TP)^{-1}d(P^TP)(P^TP)^{-1}\big)P^T + P(P^TP)^{-1}dP^T = \\
&= \big(dP(P^TP)^{-1} - P(P^TP)^{-1}(dP^TP + P^TdP)(P^TP)^{-1}\big)P^T + P(P^TP)^{-1}dP^T = \\
&= \big(dP - P(dP^TP + P^TdP)\big)P^T + PdP^T = \\
&= dPP^T - P(dP)^TPP^T - PP^TdPP^T + PdP^T
\end{aligned}
$$

Рассмотрим $(I - P(P^TP)^{-1}P^T)d\big((P(P^TP)^{-1}P^T)\big)$:

$$
\begin{aligned}
&(I - P(P^TP)^{-1}P^T)d(P(P^TP)^{-1}P^T) = (I - PP^T)(dPP^T - P(dP)^TPP^T - PP^TdPP^T + PdP^T) = \\
&= dPP^T - P(dP)^TPP^T - PP^TdPP^T + PdP^T - PP^TdPP^T + PP^TP(dP)^TPP^T + PP^TPP^TdPP^T - PP^TPdP^T = \\
&= dPP^T - \cancel{P(dP)^TPP^T} - PP^TdPP^T + \cancel{PdP^T} - \cancel{PP^TdPP^T} + \cancel{P(dP)^TPP^T} + \cancel{PP^TdPP^T} - \cancel{PdP^T} = \\
&= (I - PP^T)dPP^T
\end{aligned}
$$

Таким образом:

$$
d(F(P)) = -2N\langle(I - PP^T)dPP^TS, I\rangle = -2N\langle(I - PP^T)^T(P^TS)^T, dP\rangle = 2N\langle(PP^T - I)SP, dP\rangle
$$

$$
\boxed{\nabla F(P) = 2N(PP^T - I)SP}
$$

8.2 Теорема

Рассмотри $S = Q\Lambda Q^T$, где $\Lambda - diag\{\lambda_1, \dots, \lambda_D\}, \lambda_i$ - собств. зн. $S, 1 \leq i \leq D$
$Q = [q_1|q_2|...|q_D] \in \mathbb{R}^{D\times D}$ - орт. матрица из собственных векторов q_i

Тогда $\forall\ 1 \leq d \leq D:$

1. $\nabla F(P) = 0$, где $P = [q_{i_1}|\dots|q_{i_d}], q_i \neq q_j$, при $i \neq j$.
2. $min_P F(P) = F(\hat{P})$, где $\hat{P} = [q_{i_1}|\dots|q_{i_d}]$, $q_{i_1}, \dots, q_{i_l}$ - собственные вектора, отвечающие максимальным собственным значениям.

►

Так как Q - ортогональная матрица, то $\forall\ 1 \leq d \leq D \Rightarrow P = [q_{i_1}, \dots, q_{i_d}]$ — матрица с орт. столбцами.

Тогда:

$$
\begin{aligned}
\nabla F(P) = 2N(PP^T - I)SP = 2N(PP^T - I)Q\Lambda Q^TP = 2N(PP^T - I)P\hat{\Lambda} = 2N(PP^TP - P)\hat{\Lambda} = \\
= 2N(P - P)\hat{\Lambda} = 0
\end{aligned}
$$

$$
\hat{\Lambda} = diag\{\lambda_{i_1}, \dots, \lambda_{i_d}\}
$$

Рассмотрим $F(P)$:

$$
\begin{aligned}
F(P) &= Ntr((I - P(P^TP)^{-1}P^T)(I - P(P^TP)^{-1}P^T)Q\Lambda Q^T) = Ntr((I - PP^T)(S - PP^TQ\Lambda Q^T)) = \\
&= Ntr((I - PP^T)(S - P\hat{\Lambda}P^T) = N(tr(S) - tr(P\hat{\Lambda}P^T) - tr(PP^TS) + tr(PP^TP\hat{\Lambda}P^T)) = \\
&= N(tr(S) - \cancel{tr(P\hat{\Lambda}P^T)} - tr(PP^TQ\Lambda Q^T) + \cancel{tr(P\hat{\Lambda}P^T)}) = N(tr(S) - tr(P\hat{\Lambda}P^T)) = \\
&= N(tr(S) - \sum_{j=1}^{d} q_{i_j})
\end{aligned}
$$

То есть минимум при $\sum_{j=1}^{d} q_{i_j} \to max$, что и завершает доказательство теоремы.

■