МАТРИЧНЫЕ ВЫЧИСЛЕНИЯ И МАТРИЧНОЕ ДИФФЕРЕНЦИРОВАНИЕ

Теоретическое задание 1

Каратыщев Дмитрий Иванович

Октябрь, 2023

1. Доказать тождество Вудбери

$$(\mathbf{A}+\mathbf{UCV})^{\mathbf{-1}} = \mathbf{A^{-1}} - \mathbf{A^{-1}U}(\mathbf{C^{-1}}+\mathbf{VA^{-1}U})^{\mathbf{-1}}\mathbf{VA^{-1}}, \tag{1.1}$$

Где $A \in \mathbb{R}^{n\times n}$, $C \in \mathbb{R}^{m\times m}$, $U \in \mathbb{R}^{n\times m}$ и $V \in \mathbb{R}^{m\times n}$, $\det(A) \neq 0$, $\det(C) \neq 0$

Доказательство:

Преобразуем правую часть равенства, пользуясь известным свойством операции обращения матриц: $(AB)^{-1} = B^{-1}A^{-1}$

Из данного свойства тривиально выводится следующий факт: $V^{-1}C^{-1}U^{-1} = (UCV)^{-1}$

► $A^{-1}-A^{-1}U(C^{-1}+VA^{-1}U)^{-1}VA^{-1} = A^{-1}-A^{-1}U[(C^{-1}U^{-1}+VA^{-1})U]^{-1}VA^{-1} = A^{-1} - A^{-1}(C^{-1}U^{-1}+VA^{-1})^{-1}VA^{-1} = A^{-1} - A^{-1}[(C^{-1}U^{-1}A+V)A^{-1}]^{-1}VA^{-1} = A^{-1} - (C^{-1}U^{-1}A+V)^{-1}VA^{-1} = A^{-1} - [V(V^{-1}C^{-1}U^{-1}A+I)]^{-1}VA^{-1} = A^{-1} - (V^{-1}C^{-1}U^{-1}A+I)^{-1}A^{-1} = A^{-1} - [(UCV)^{-1}A+I]^{-1}A^{-1} = A^{-1} - [(UCV)^{-1}(A+UCV)]^{-1}A^{-1} = A^{-1} - (A+UCV)^{-1}UCVA^{-1} = (A+UCV)^{-1}[(A+UCV)A^{-1} - UCVA^{-1}] = (A+UCV)^{-1}(I+UCVA^{-1}-UCVA^{-1}) = (A+UCV)^{-1}I = (A+UCV)^{-1}$

◄

2. Упростить выражения

$$\left\|\mathbf{uv^T} - \mathbf{A}\right\|_{\mathbf{F}}^{\mathbf{2}} - \|\mathbf{A}\|_{\mathbf{F}}^{\mathbf{2}}, \tag{2.1}$$

Где $u \in \mathbb{R}^m$, $v \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$

Решение:

$\left\|uv^T - A\right\|_F^2 - \|A\|_F^2 = \langle uv^T - A\ ,\ uv^T - A\rangle - \langle A\ ,\ A\rangle = \langle uv^T\ ,\ uv^T\rangle - 2\langle uv^T\ ,\ A\rangle + \langle A\ ,\ A\rangle - \langle A\ ,\ A\rangle = \langle uv^T\ ,\ uv^T\rangle\ - 2\langle uv^T\ ,\ A\rangle$

1. $\langle uv^T\ ,\ uv^T\rangle\ - 2\langle uv^T\ ,\ A\rangle = \left\|uv^T\right\|_F^2 - 2\langle uv^T\ ,\ A\rangle = \|u\|_F^2\|v\|_F^2 - 2\langle uv^T\ ,\ A\rangle$
2. $\langle uv^T\ ,\ uv^T\rangle\ - 2\langle uv^T\ ,\ A\rangle = \operatorname{tr}\big((uv^T)^T\cdot uv^T\big) - 2\operatorname{tr}\big((uv^T)^T\cdot A\big) =$
   $= \operatorname{tr}\big(vu^Tuv^T\big) - 2\operatorname{tr}\big(vu^T\cdot A\big) = \operatorname{tr}\big((uv^T)^T(uv^T-2A)\big)$

$$\operatorname{tr}\big((\mathbf{2I_n}+\mathbf{aa^T})^{\mathbf{-1}}(\mathbf{uv^T}+\mathbf{vu^T})\big), \tag{2.2}$$

Где $a,\ u,\ v \in \mathbb{R}^n$. *Подсказка: воспользоваться тождеством Вудбери [1.1].*

Решение:

Положим в [1.1] $A = 2I_n,\ U = a,\ C = [1],\ V = a^T$

Получаем, что $(2I_n + aa^T)^{-1} = \frac{1}{2}I_n - \frac{1}{2}I_n a([1] + a^T\frac{1}{2}I_n a)^{-1}a^T\frac{1}{2}I_n = \frac{1}{2}I_n - \frac{1}{4}a([1]+a^T\frac{1}{2}a)^{-1}a^T = \frac{1}{2}\left[I_n - a([2]+a^Ta)^{-1}a^T\right] = \frac{1}{2}\left(I_n - \frac{aa^T}{2+a^Ta}\right)$

Получаем исходный след: $\frac{1}{2}\left\langle uv^T + vu^T,\ \left(I_n - \frac{aa^T}{2+a^Ta}\right)^T\right\rangle = (*)$, записанный в виде скалярного произведения. Пользуясь тем, что $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ и $\mathrm{tr}\big(A^T\big) = \mathrm{tr}(A)$, получим:

$(*) = \frac{1}{2}\langle uv^T,\ I_n\rangle + \frac{1}{2}\langle vu^T,\ I_n\rangle - \frac{1}{2+a^Ta}\,\mathrm{tr}\big(uv^Taa^T\big) = \frac{1}{2}\langle uv^T,\ I_n\rangle + \frac{1}{2}\langle vu^T,\ I_n\rangle - \frac{1}{2+a^Ta}\,\mathrm{tr}\big(a^Tuv^Ta\big) = \langle u,\ v\rangle + \frac{1}{2+a^Ta}\langle v^Ta,\ u^Ta\rangle = \langle u,\ v\rangle + \frac{v^Ta\cdot u^Ta}{2+a^Ta}$

$$\sum_{\mathbf{i=1}}^{\mathbf{n}}\langle \mathbf{S^{-1}a_i},\ \mathbf{a_i}\rangle, \tag{2.3}$$

Где $a_1, \ldots, a_n \in \mathbb{R}^d,\ S = \sum_{i=1}^n a_ia_i^T, \det(S) \neq 0$.

Решение:

$\sum_{i=1}^n \langle S^{-1}a_i,\ a_i\rangle = \sum_{i=1}^n \mathrm{tr}\left(a_i^T\cdot\left(\sum_{j=1}^n a_ja_j^T\right)^{-1}\cdot a_i\right) = $ [По свойству следа] $= \sum_{i=1}^n \mathrm{tr}\left(\left(\sum_{j=1}^n a_ja_j^T\right)^{-1}\cdot a_i\cdot a_i^T\right) = \mathrm{tr}\left(\sum_{i=1}^n\left[\left(\sum_{j=1}^n a_ja_j^T\right)^{-1}\cdot a_i\cdot a_i^T\right]\right) = \mathrm{tr}\big(S^{-1}S\big) = \mathrm{tr}(I_d) = d$

3. Найти первую и вторую производную

$$\mathbf{f}:\mathbf{E}\longrightarrow\mathbb{R}, \mathbf{f(t)} = \det(\mathbf{A} - \mathbf{tI_n}), \text{где } \mathbf{A}\in\mathbb{R}^{n\times n}, E = \{t\in\mathbb{R}\colon\ \det(A - tI_n)\neq 0\} \tag{3.1}$$

Решение:

$f(t) = (g\circ h)(t))$, где $g(Y) = \det(Y),\ h(t) = A - tI_n$

Правило дифференцирования композиции: $d(g\circ h)_{t_0}[dt] = dg_{h(t_0)}[dh_{t_0}(dt)]$

$dg_Y[dY] = \det(Y)\left\langle Y^{-T},\ dY\right\rangle,\ dh_{t_0}[dt] = -I_n\,dt \Rightarrow df_{t_0}[dt] = d(g\circ h)_{t_0}[dt] =$

$= \det(A - t_0I_n)\left\langle (A-t_0I_n)^{-T},\ -I_n\,dt\right\rangle = \det(A - t_0I_n)\left\langle (A-t_0I_n)^{-T},\ -I_n\right\rangle dt =$

$= -\det(A-t_0I_n)\,\mathrm{tr}\big((A-t_0I_n)^{-1}\big)\,dt$ - искомая производная (дифференциал)

Дополнительно:

Получили первый градиент функции $\nabla f(t) = -\det(A - tI_n)\,\mathrm{tr}\big((A-tI_n)^{-1}\big)$

Найдём вторую производную (дифференциал второго порядка) функции f: $d^2(f)_t[dt, \hat{dt}] = d\,(df)_t\,[dt, \hat{dt}] = d\left[-\det(A-tI_n)\cdot\mathrm{tr}\big((A-tI_n)^{-1}\big)\cdot dt\right]_t[dt, \hat{dt}] =$ *[фиксируем dt]* $=$

$-\,dt\cdot d\left[\det(A-tI_n)\right]_t[\hat{dt}]\cdot\mathrm{tr}\big((A-tI_n)^{-1}\big)\ -\ dt\cdot\det(A-tI_n)\cdot d\left[\mathrm{tr}\big((A-tI_n)^{-1}\big)\right]_t[\hat{dt}] =$

$= dt\cdot\det(A-tI_n)\cdot\mathrm{tr}^2((A-tI_n)^{-1})\cdot\hat{dt}\ -\ dt\cdot\det(A-tI_n)\cdot\mathrm{tr}\left(d\left[(A-tI_n)^{-1}\right]_t[\hat{dt}]\right) =$

$= dt\cdot\det(A-tI_n)\cdot\mathrm{tr}^2((A-tI_n)^{-1})\cdot\hat{dt}\ -\ dt\cdot\det(A-tI_n)\times$

$\times\,\mathrm{tr}\left(-(A-tI_n)^{-1}\cdot(-\hat{dt}I_n)\cdot(A-tI_n)^{-1}\right) =$

$= dt\cdot\left[\det(A-tI_n)\cdot\left(\mathrm{tr}^2((A-tI_n)^{-1}) - \mathrm{tr}\big((A-tI_n)^{-2}\big)\right)\right]\cdot\hat{dt}$

Дополнительно:

Получили следующий гессиан $\nabla^2 f(t) = \det(A-tI_n)\cdot\left(\mathrm{tr}^2((A-tI_n)^{-1}) - \mathrm{tr}\big((A-tI_n)^{-2}\big)\right)$

$$\mathbf{f}:\mathbb{R}_{++}\longrightarrow\mathbb{R},\ \mathbf{f(t)}=\left\|(\mathbf{A}+\mathbf{tI_n})^{-1}\mathbf{b}\right\|,\ \textbf{где } \mathbf{A}\in\mathbb{S}^n_+, \mathbf{b}\in\mathbb{R}^n \tag{3.2}$$

Решение:

$f(t)=(g\circ h)(t))$, где $g(Y)=\|Y\|,\ h(t)=(A+tI_n)^{-1}b$

$dh_t[dt]=d\left((A+tI_n)^{-1}\right)_t[dt]\cdot b=\left[-(A+tI_n)^{-1}(dtI_n)(A+tI_n)^{-1}\right]b$

$dg_Y[dY]=d(\|Y\|)_Y[dY]=d(\langle Y,\ Y\rangle^{\frac{1}{2}})_Y[dY]=\frac{1}{2}\langle Y,\ Y\rangle^{-\frac{1}{2}}\cdot d(\langle Y,\ Y\rangle)_Y[dY]=$

$=\frac{1}{2}\langle Y,\ Y\rangle^{-\frac{1}{2}}\cdot d(Y^TY)_Y[dY]=\frac{1}{2}\langle Y,\ Y\rangle^{-\frac{1}{2}}\cdot\left(d(Y^T)_Y[dY]\cdot Y+Y^TdY\right)=$

$=\frac{1}{2}\langle Y,\ Y\rangle^{-\frac{1}{2}}\cdot\left((dY)^TY+Y^TdY\right)=\frac{1}{2}\langle Y,\ Y\rangle^{-\frac{1}{2}}\cdot\left((Y^TdY)^T+Y^TdY\right)=$

$=[$ *учитываем, что* $(Y^TdY)^T=Y^TdY]=\langle Y,\ Y\rangle^{-\frac{1}{2}}\cdot Y^TdY=\frac{\langle dY,\ Y\rangle}{\|Y\|}$

Получаем, что $d(g\circ h)_t[dt]=\dfrac{\langle -(A+tI_n)^{-1}(dtI_n)(A+tI_n)^{-1}b,\ (A+tI_n)^{-1}b\rangle}{\|(A+tI_n)^{-1}b\|}=$

$=[$ *учитываем, что матрица* A *симметричная:* $(A+tI_n)^{-T}=(A+tI_n)^{-1}]=$

$$=-\frac{\operatorname{tr}\left(b^T(A+tI_n)^{-T}(A+tI_n)^{-2}b\right)}{\|(A+tI_n)^{-1}b\|}\cdot dt=-\frac{\operatorname{tr}\left(b^T(A+tI_n)^{-3}b\right)}{\|(A+tI_n)^{-1}b\|}\cdot dt=$$

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

Дополнительно:

Получили первый градиент функции $\nabla f(t)=-\dfrac{\langle(A+tI_n)^{-3}b,\ b\rangle}{\|(A+tI_n)^{-1}b\|}$

Найдём вторую производную (дифференциал второго порядка) функции f: $d^2(f)_t[dt,\hat{dt}]=$

$d\left(df\right)_t[dt,\hat{dt}]=d\left[-\dfrac{\langle(A+tI_n)^{-3}b,\ b\rangle}{\|(A+tI_n)^{-1}b\|}\cdot dt\right]_t[dt,\hat{dt}]=[$*фиксируем* $dt]=$

$$-dt\cdot\frac{d\left[\langle(A+tI_n)^{-3}b,\ b\rangle\right]_t[\hat{dt}]\cdot\left\|(A+tI_n)^{-1}b\right\|-d\left[\left\|(A+tI_n)^{-1}b\right\|\right]_t[\hat{dt}]\cdot\langle(A+tI_n)^{-3}b,\ b\rangle}{\|(A+tI_n)^{-1}b\|^2}=$$

$$=-dt\cdot\frac{\langle -3(A+tI_n)^{-4}b\hat{dt},\ b\rangle\cdot\left\|(A+tI_n)^{-1}b\right\|+\dfrac{\langle(A+tI_n)^{-3}b,\ b\rangle^2}{\|(A+tI_n)^{-1}b\|}\cdot\hat{dt}}{\|(A+tI_n)^{-1}b\|^2}=$$

$$=dt\cdot\left[\frac{3\langle(A+tI_n)^{-4}b,\ b\rangle}{\|(A+tI_n)^{-1}b\|}-\frac{\langle(A+tI_n)^{-3}b,\ b\rangle^2}{\|(A+tI_n)^{-1}b\|^3}\right]\cdot\hat{dt}$$

Дополнительно:

Получили следующий гессиан $\nabla^2 f(t)=\dfrac{3\langle(A+tI_n)^{-4}b,\ b\rangle}{\|(A+tI_n)^{-1}b\|}-\dfrac{\langle(A+tI_n)^{-3}b,\ b\rangle^2}{\|(A+tI_n)^{-1}b\|^3}$

4. Найти градиент ∇f и гессиан $\nabla^2 f$

$$\mathbf{f}:\mathbb{R}^n \longrightarrow \mathbb{R},\ \mathbf{f(x)} = \frac{\mathbf{1}}{\mathbf{2}}\left\|\mathbf{xx^T} - \mathbf{A}\right\|^{\mathbf{2}}_{\mathbf{F}},\ \text{где } \mathbf{A} \in \mathbb{S}^n \tag{4.1}$$

Решение:

$df_x[dx] = d\left(\frac{1}{2}\left\|xx^T - A\right\|_F^2\right)_x[dx] = \frac{1}{2}d\left(\langle xx^T - A,\ xx^T - A\rangle\right)_x[dx] =$

$= \frac{1}{2}d\left[\mathrm{tr}\big((xx^T - A)^T(xx^T - A)\big)\right]_x[dx] =$ *[учтём, что матрица A симметрична]*

$= \frac{1}{2}d\left[\mathrm{tr}\big((xx^T - A)^2\big)\right]_x[dx] = \frac{1}{2}\,\mathrm{tr}\big(d\left[(xx^T - A)^2\right]_x[dx]\big) =$

$\frac{1}{2}\,\mathrm{tr}\big(2(xx^T - A)\cdot d(xx^T - A)_x[dx]\big) = \mathrm{tr}\big((xx^T - A)\cdot(dx\cdot x^T + x(dx)^T)\big) =$

$= \mathrm{tr}\big((xx^T - A)\cdot dx\cdot x^T\big) + \mathrm{tr}\big((xx^T - A)\cdot x(dx)^T\big) = \mathrm{tr}\big(x^T\cdot(xx^T - A)\cdot dx\big) +$

$+\,\mathrm{tr}\big(x^T\cdot(xx^T - A)\cdot dx\big) = 2\,\mathrm{tr}\big(x^T\cdot(xx^T - A)\cdot dx\big) = \langle 2(xx^T - A)x,\ dx\rangle$

Получили первый градиент функции $\nabla f(x) = 2(xx^T - A)x$

$d(df)_x[dx, \hat{dx}] = d\left[2\,\mathrm{tr}\big(x^T\cdot(xx^T - A)\cdot dx\big)\right]_x[dx, \hat{dx}] = 2\,\mathrm{tr}\Big(d\left[x^T\cdot(xx^T - A)\cdot dx\right]_x[dx, \hat{dx}]\Big) =$

$= 2\,\mathrm{tr}\Big(\Big[(\hat{dx})^T\cdot(xx^T - A) + x^T\cdot\Big(\hat{dx}x^T + (\hat{dx}x^T)^T\Big)\Big]\cdot dx\Big) = 2\,\mathrm{tr}\Big((\hat{dx})^T\cdot(xx^T - A)\cdot dx\Big) +$

$+2\,\mathrm{tr}\Big(x^T\cdot\Big(\hat{dx}x^T + (\hat{dx}x^T)^T\Big)\cdot dx\Big) = \langle 2(xx^T - A)dx,\ \hat{dx}\rangle + 2\,\mathrm{tr}\Big((\hat{dx})^T xx^T dx\Big) +$

$+2\,\mathrm{tr}\Big(x^T x(\hat{dx})^T dx\Big) = \langle 2(xx^T - A)dx,\ \hat{dx}\rangle + \langle 2xx^T dx,\ \hat{dx}\rangle + \langle 2x^T x dx,\ \hat{dx}\rangle =$

$\langle\big(2(xx^T - A) + 2xx^T + 2x^T xI_n\big)\,dx,\ \hat{dx}\rangle = \langle 2\,\big(\langle x,\ x\rangle I_n + 2xx^T - A\big)\,dx,\ \hat{dx}\rangle$

Получили следующий гессиан $\nabla^2 f(x) = 2\,\big(\langle x,\ x\rangle I_n + 2xx^T - A\big)$

$$\mathbf{f}:\mathbb{R}^n\backslash\{0\} \longrightarrow \mathbb{R},\ \mathbf{f(x)} = \langle x,\ x\rangle^{\langle x,\ x\rangle} \tag{4.2}$$

Решение:

$df_x[dx] = d\left[e^{\langle x,\ x\rangle\ln\langle x,\ x\rangle}\right]_x[dx] = e^{\langle x,\ x\rangle\ln\langle x,\ x\rangle}\cdot(d(\langle x,\ x\rangle)_x[dx]\cdot\ln\langle x,\ x\rangle + \langle x,\ x\rangle\cdot d(\ln\langle x,\ x\rangle)_x[dx])$

Имеем $d(\langle x,\ x\rangle)_x[dx] = d(x^T x)_x[dx] = 2x^T dx$

Тогда $df_x[dx] = e^{\langle x,\ x\rangle\ln\langle x,\ x\rangle}\cdot(2x^T dx\cdot(\ln\langle x,\ x\rangle + 1)) = \langle 2\langle x,\ x\rangle^{\langle x,\ x\rangle}\cdot(\ln\langle x,\ x\rangle + 1)x,\ dx\rangle$

Получили первый градиент функции $\nabla f(x) = 2\langle x,\ x\rangle^{\langle x,\ x\rangle}\cdot(\ln\langle x,\ x\rangle + 1)x$

$d(df)_x[dx, \hat{dx}] = d\left(\langle 2\langle x,\ x\rangle^{\langle x,\ x\rangle}\cdot(\ln\langle x,\ x\rangle + 1)x,\ dx\rangle\right)_x[dx, \hat{dx}] =$

$= d\left(2\langle x,\ x\rangle^{\langle x,\ x\rangle}\cdot(\ln\langle x,\ x\rangle + 1)\cdot x^T\cdot dx\right)_x[dx, \hat{dx}] = 2\langle d\left(\langle x,\ x\rangle^{\langle x,\ x\rangle}\right)_x[\hat{dx}]\cdot(\ln\langle x,\ x\rangle + 1)\cdot x +$

$+\langle x,\ x\rangle^{\langle x,\ x\rangle}\cdot d(\ln\langle x,\ x\rangle + 1)_x[\hat{dx}]\cdot x + \langle x,\ x\rangle^{\langle x,\ x\rangle}\cdot(\ln\langle x,\ x\rangle + 1)\hat{dx},\ dx\rangle =$

$= 2\langle 2\langle x,\ x\rangle^{\langle x,\ x\rangle}\cdot(\ln\langle x,\ x\rangle + 1)^2 x\cdot x^T\hat{dx} + 2\langle x,\ x\rangle^{\langle x,\ x\rangle - 1}\cdot x\cdot x^T\hat{dx} + \langle x,\ x\rangle^{\langle x,\ x\rangle}\cdot(\ln\langle x,\ x\rangle + 1)\hat{dx},\ dx\rangle =$

$= \langle\big[4\langle x,\ x\rangle^{\langle x,\ x\rangle}\cdot(\ln\langle x,\ x\rangle + 1)^2 x\cdot x^T + 4\langle x,\ x\rangle^{\langle x,\ x\rangle - 1}\cdot x\cdot x^T + 2\langle x,\ x\rangle^{\langle x,\ x\rangle}\cdot(\ln\langle x,\ x\rangle + 1)I_n\big]\,\hat{dx},\ dx\rangle$

Итоговый гессиан:

$\nabla^2 f(x) = 4\langle x,\ x\rangle^{\langle x,\ x\rangle}\cdot\left[(\ln\langle x,\ x\rangle + 1)^2 + \frac{1}{\langle x,\ x\rangle}\right]\cdot xx^T + 2\langle x,\ x\rangle^{\langle x,\ x\rangle}\cdot(\ln\langle x,\ x\rangle + 1)I_n$

$$\mathbf{f}:\mathbb{R}^n\longrightarrow\mathbb{R},\ \mathbf{f(x)}=\|\mathbf{Ax}-\mathbf{b}\|^{\mathbf{p}},\ \textbf{где}\ \mathbf{A}\in\mathbb{R}^{m\times n},\ \mathbf{b}\in\mathbb{R}^m,\ \mathbf{p}\geq\mathbf{2} \tag{4.3}$$

Решение:

$df_x[dx]=d\left(\|Ax-b\|^p\right)_x[dx]=d\left(\langle Ax-b,\ Ax-b\rangle^{\frac{p}{2}}\right)_x[dx]=d\left[\left((Ax-b)^T(Ax-b)\right)^{\frac{p}{2}}\right]_x[dx]=$

$=\frac{p}{2}\left((Ax-b)^T(Ax-b)\right)^{\frac{p}{2}-1}d\left((Ax-b)^T(Ax-b)\right)_x[dx]$

Распишем дифференциал: $d\left((Ax-b)^T(Ax-b)\right)_x[dx]=d\left[(Ax-b)^T\right]_x[dx](Ax-b)+$

$+(Ax-b)^Td\left[(Ax-b)\right]_x[dx]=(Adx)^T(Ax-b)+(Ax-b)^TAdx=2\langle Ax-b,\ Adx\rangle=2(Ax-b)^TAdx$

В итоге получаем $df_x[dx]=p\cdot\left((Ax-b)^T(Ax-b)\right)^{\frac{p}{2}-1}(Ax-b)^TAdx=$

$=\langle p\cdot\left((Ax-b)^T(Ax-b)\right)^{\frac{p}{2}-1}A^T(Ax-b),\ dx\rangle$

Получили первый градиент функции $\nabla f(x)=p\cdot\left((Ax-b)^T(Ax-b)\right)^{\frac{p}{2}-1}A^T(Ax-b)$

$d(df)_x[dx,\hat{dx}]=p\langle d\left[\left((Ax-b)^T(Ax-b)\right)^{\frac{p-2}{2}}A^T(Ax-b)\right]_x[\hat{dx}],\ dx\rangle=$

$=p\langle d\left[\left((Ax-b)^T(Ax-b)\right)^{\frac{p-2}{2}}\right]_x[\hat{dx}]\cdot A^T(Ax-b)+$

$+\ \left((Ax-b)^T(Ax-b)\right)^{\frac{p-2}{2}}\cdot d\left[A^T(Ax-b)\right]_x[\hat{dx}],\ dx\rangle$

Подставляем значения для уже посчитанных дифференциалов и получаем следующее выражение: $\Big\langle\Big[p\cdot(p-2)\cdot\left((Ax-b)^T(Ax-b)\right)^{\frac{p-4}{2}}\cdot A^T(Ax-b)\cdot(Ax-b)^TA+$

$+\left((Ax-b)^T(Ax-b)\right)^{\frac{p-2}{2}}\cdot A^TA\Big]\hat{dx},\ dx\Big\rangle=$

$=\left\langle\left[p(p-2)\|Ax-b\|^{p-4}\cdot A^T(Ax-b)\cdot(Ax-b)^TA+\|Ax-b\|^{p-2}\cdot A^TA\right]\hat{dx},dx\right\rangle$

Результирующий гессиан:

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

5. Знакоопределённость второго дифференциала

$$\mathbf{f}:\mathbb{S}^n_{++}\longrightarrow\mathbb{R},\ \mathbf{f(X)}=\operatorname{tr}\big(\mathbf{X^{-1}}\big) \tag{5.1}$$

Решение:

$df_X[dX]=-\operatorname{tr}\big(X^{-1}dXX^{-1}\big)=-\operatorname{tr}\big(X^{-2}dX\big)$

$d^2f_X[dX,dX]=-tr(d(X^{-2})_X[dX]dX)=\operatorname{tr}\big((X^{-1}dXX^{-2}+X^{-2}dXX^{-1})dX\big)=$

$=2\operatorname{tr}\big(X^{-1}dXX^{-2}dX\big)=2\operatorname{tr}\big(X^{-1}dXX^{-1}dXX^{-1}\big)=$ /*Квадратный корень для положительно определённых симметричных матриц однозначно определён*/ $=$

$=2\operatorname{tr}\Big(X^{-1}dXX^{-\frac{1}{2}}X^{-\frac{1}{2}}dXX^{-1}\Big)=2\operatorname{tr}\Big(\Big(X^{-\frac{1}{2}}dXX^{-1}\Big)^TX^{-\frac{1}{2}}dXX^{-1}\Big)=$

$=2\langle X^{-\frac{1}{2}}dXX^{-1},\ X^{-\frac{1}{2}}dXX^{-1}\rangle>0$ по определению скалярного произведения

Получаем, что второй дифференциал является **отрицательно определённым**

$$\mathbf{f}:\mathbb{S}^n_{++}\longrightarrow\mathbb{R},\ \mathbf{f}(\mathbf{X})=(\det(\mathbf{X}))^{\frac{1}{\mathbf{n}}} \tag{5.2}$$

Решение:

$df_x[dx]=\frac{1}{n}(\det(X))^{\frac{1}{n}-1}\cdot\det(X)\langle X^{-T},\ dX\rangle=\langle\frac{1}{n}(\det(X))^{\frac{1}{n}}X^{-T},\ dX\rangle$

$df_x[dx,dx]=\frac{1}{n}\left\langle d\left((\det(X))^{\frac{1}{n}}X^{-T}\right)_X[dX],\ dX\right\rangle=$

$=\frac{1}{n}\left\langle\langle\frac{1}{n}(\det(X))^{\frac{1}{n}}X^{-1},\ dX\rangle X^{-1}-(\det(X))^{\frac{1}{n}}\cdot X^{-1}dXX^{-1},\ dX\right\rangle=$

$=\frac{1}{n^2}(\det(X))^{\frac{1}{n}}\langle X^{-1},\ dX\rangle^2-\frac{1}{n}(\det(X))^{\frac{1}{n}}\langle X^{-1}dXX^{-1},\ dX\rangle=(\star)$

Заметим, что $\langle X^{-1},\ dX\rangle=\mathrm{tr}\big(X^{-1}dX\big)=\mathrm{tr}\Big(X^{-\frac{1}{2}}X^{-\frac{1}{2}}dX\Big)=\mathrm{tr}\Big(X^{-\frac{1}{2}}dXX^{-\frac{1}{2}}\cdot I_n\Big)=$

$=\langle X^{-\frac{1}{2}}dXX^{-\frac{1}{2}},\ I_n\rangle$ и при этом $\langle X^{-1}dXX^{-1},\ dX\rangle=\langle X^{-\frac{1}{2}}dXX^{-\frac{1}{2}},\ X^{-\frac{1}{2}}dXX^{-\frac{1}{2}}\rangle$

$(\star)=\frac{1}{n^2}(\det(X))^{\frac{1}{n}}\langle X^{-\frac{1}{2}}dXX^{-\frac{1}{2}},\ I_n\rangle^2-\frac{1}{n}(\det(X))^{\frac{1}{n}}\langle X^{-\frac{1}{2}}dXX^{-\frac{1}{2}},\ X^{-\frac{1}{2}}dXX^{-\frac{1}{2}}\rangle$

Воспользуемся неравенством Коши-Буняковского-Шварца

$(\star)\leq\frac{1}{n^2}(\det(X))^{\frac{1}{n}}\left\|X^{-\frac{1}{2}}dXX^{-\frac{1}{2}}\right\|^2\|I_n\|^2-\frac{1}{n}(\det(X))^{\frac{1}{n}}\left\|X^{-\frac{1}{2}}dXX^{-\frac{1}{2}}\right\|^2=$

$=\frac{1}{n^2}(\det(X))^{\frac{1}{n}}\left\|X^{-\frac{1}{2}}dXX^{-\frac{1}{2}}\right\|^2\mathrm{tr}^2(I_n)-\frac{1}{n}(\det(X))^{\frac{1}{n}}\left\|X^{-\frac{1}{2}}dXX^{-\frac{1}{2}}\right\|^2=0$

Равенство второй производной нулю достигается при $dX\propto X$, что следует из условий неравенства Коши-Буняковского-Шварца

6. Найти точки стационарности

$$\mathbf{f}:\mathbb{R}^n\longrightarrow\mathbb{R},\ \mathbf{f}(\mathbf{x})=\langle\mathbf{c},\ \mathbf{x}\rangle+\frac{\sigma}{\mathbf{3}}\|\mathbf{x}\|^{\mathbf{3}},\ \textit{где}\ \mathbf{c}\in\mathbb{R}^{\mathbf{n}},\ \mathbf{c}\neq\mathbf{0},\ \sigma>\mathbf{0} \tag{6.1}$$

Решение:

$df_x[dx]=\langle c,\ dx\rangle+\sigma\|x\|\langle x,\ dx\rangle\Rightarrow$ *Точки стационарности находятся при равенстве нулю первого дифференциала* $\Rightarrow c+\sigma\|x\|x=0$

Получаем, что $\|x\|x=-\frac{c}{\sigma}\Rightarrow\|x\|^2=\frac{\|c\|}{\sigma}\Rightarrow\|x\|=\sqrt{\frac{\|c\|}{\sigma}}\Rightarrow x=\frac{c}{\sqrt{\sigma\|c\|}}$

$$\mathbf{f}:E\longrightarrow\mathbb{R},\mathbf{f}(\mathbf{x})=\langle\mathbf{a},\ \mathbf{x}\rangle-\ln(\mathbf{1}-\langle\mathbf{b},\ \mathbf{x}\rangle),\textit{где}\,\mathbf{a},\ \mathbf{b}\in\mathbb{R}^{\mathbf{n}},\ \mathbf{a},\ \mathbf{b}\neq\mathbf{0},\ E=\{\mathbf{x}\in\mathbb{R}^{\mathbf{n}}\mid\langle\mathbf{b},\ \mathbf{x}\rangle<\mathbf{1}\}\Rightarrow \tag{6.2}$$

Решение:

$df_x[dx]=\langle a,\ dx\rangle-\dfrac{-\langle b,\ dx\rangle}{1-\langle b,\ x\rangle}=\left\langle a+\dfrac{b}{1-\langle b,\ x\rangle},\ dx\right\rangle\Rightarrow a+\dfrac{b}{1-\langle b,\ x\rangle}=0\Rightarrow$

$\Rightarrow\|a\|=\dfrac{\|b\|}{1-\langle b,\ x\rangle}\Rightarrow\langle b,\ x\rangle=1-\dfrac{\|b\|}{\|a\|}\ (\star)$

Ответом является гиперплоскость $(\star)$

$$\mathbf{f}:\mathbb{R}^n \longrightarrow \mathbb{R},\ \mathbf{f}(\mathbf{x}) = \langle \mathbf{c},\ \mathbf{x}\rangle \cdot \mathbf{e}(-\langle \mathbf{Ax},\ \mathbf{x}\rangle),\ \textit{где}\ \mathbf{c} \in \mathbb{R}^{\mathbf{n}},\ \mathbf{c} \neq \mathbf{0},\ A \in \mathbb{S}^{\mathbf{n}}_{++} \tag{6.3}$$

Решение:

$df_x[dx] = \langle c,\ dx\rangle e^{-\langle Ax,\ x\rangle} - 2\langle Ax,\ dx\rangle\langle c,\ x\rangle e^{-\langle Ax,\ x\rangle}$, при этом учтён факт, что матрица A симметрична. В дальнейшем положительная определённость будет использоваться для обращения матрицы

Получаем, что $(c - 2Ax\langle c,\ x\rangle)e^{-\langle Ax,\ x\rangle} = 0 \Rightarrow c - 2Ax\langle c,\ x\rangle = 0 \Rightarrow$

$\Rightarrow\ \langle x,\ c\rangle x = \dfrac{A^{-1}c}{2} = \text{[\textit{Умножим скалярно на c}]} = \langle x,\ c\rangle^2 = \dfrac{\langle A^{-1}c,\ c\rangle}{2}$

Извлечём корень у каждой части равенства. Данная операция корректна ввиду положительной определённости матрицы

$\langle x,\ c\rangle = \pm\sqrt{\dfrac{\langle A^{-1}c,\ c\rangle}{2}} \Rightarrow \text{[\textit{Подставим скалярное произведение в равенство выше}]} \Rightarrow$

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

7. Найти предел

$$\lim_{\mathbf{k}\to+\infty} \operatorname{tr}\big(\mathbf{X}^{-\mathbf{k}} - (\mathbf{X}^{\mathbf{k}} + \mathbf{X}^{\mathbf{2k}})^{-\mathbf{1}}\big),\ \textit{где}\ \mathbf{X} \in \mathbb{S}^n_{++} \tag{7.1}$$

Решение:

Воспользуемся неравенством [1.1], положив $A = U = V = X^k,\ C = I_n$

Получим, что $(X^k + X^k I_n X^k)^{-1} = X^{-k} - X^{-k}X^{-k}(I_n + X^k X^{-k} X^k)^{-1} X^k X^{-k} = X^{-k} - (I_n + X^k)^{-1}$

Тогда [7.1] преобразовывается в $\lim_{k\to+\infty} \operatorname{tr}\big((I + X^k)^{-1}\big) = (\star)$

$X \in \mathbb{S}^n_{++} \Rightarrow \exists Q : QQ^T = Q^T Q = I_n,\ X = Q\Lambda Q^T,\ \Lambda = diag(\lambda_1, ..., \lambda_n)$, где $\lambda_1, ..., \lambda_n -$ собственные значения матрицы X

Получаем: $(\star) = \lim_{k\to+\infty} \operatorname{tr}\big((I + Q\Lambda^k Q^T)^{-1}\big) = \lim_{k\to+\infty} \operatorname{tr}\big((QIQ^T + Q\Lambda^k Q^T)^{-1}\big) =$

$= \lim_{k\to+\infty} \operatorname{tr}\big(Q^{-T}(I + \Lambda^k)^{-1}Q^{-1}\big) = \lim_{k\to+\infty} \operatorname{tr}\big(I + \Lambda^k\big)^{-1} = \lim_{k\to+\infty} \sum_{i=1}^n \dfrac{1}{1 + \lambda_i^k} =$

$= \sum_{i=1}^n [\lambda_i < 1] + \frac{1}{2}\sum_{i=1}^n [\lambda_i = 1]$

8. Метод главных компонент

$$\mathbf{F}(\mathbf{P}) = \sum_{\mathbf{i=1}}^{\mathbf{N}} \left\|\mathbf{x_i} - \mathbf{P}(\mathbf{P^T P})^{-\mathbf{1}}\mathbf{P^T x_i}\right\|^{\mathbf{2}} = \mathbf{N}\operatorname{tr}\big((\mathbf{I_D} - \mathbf{P}(\mathbf{P^T P})^{-\mathbf{1}}\mathbf{P^T})^{\mathbf{2}}\ \mathbf{S}\big) \to \min_{\mathbf{P}} \tag{8.1}$$

При этом $\mathbf{S} = \frac{\mathbf{1}}{\mathbf{N}}\sum_{\mathbf{i=1}}^{\mathbf{N}} \mathbf{x_i x_i^T},\ \mathbf{P} \in \mathbb{R}^{\mathbf{D\times d}},\ \mathbf{x_i} \in \mathbb{R}^{\mathbf{D}}$

Задание 1 Найти градиент $\nabla_P F(P)$ *при* $P^T P = I_D$

$dF_P[dP] = N\operatorname{tr}\left(d\left[(I_D - P(P^T P)^{-1}P^T)^2\right]_P [dP]S\right)$

Положим $g := P(P^T P)^{-1}P^T,\ \hat{g} = I_D - P(P^T P)^{-1}P^T,\ \tilde{g} = \hat{g}^2$

Тогда $d\tilde{g}_P[dP] = d\hat{g}_P[dP]\cdot\hat{g} + \hat{g}\cdot d\hat{g}_P[dP] = -dg_P[dP]\cdot\hat{g} - \hat{g}\cdot dg_P[dP]$

$dg_P[dP] = dP(P^TP)^{-1}P^T + P(P^TP)^{-1}(dP)^T - P(P^TP)^{-1}(d(P^TP)_P[dP])(P^TP)^{-1}P^T =$

$= dP(P^TP)^{-1}P^T + P(P^TP)^{-1}(dP)^T - P(P^TP)^{-1}(dP)^TP(P^TP)^{-1}P^T -$

$-P(P^TP)^{-1}P^TdP(P^TP)^{-1}P^T = \hat{g}(dP(P^TP)^{-1}P^T) + (P(P^TP)^{-1}(dP)^T)\hat{g}$

$d\tilde{g}_P[dP] = -\hat{g}(dP(P^TP)^{-1}P^T)\hat{g} - (P(P^TP)^{-1}(dP)^T)\hat{g}^2 - \hat{g}^2(dP(P^TP)^{-1}P^T) -$

$-\hat{g}(P(P^TP)^{-1}(dP)^T)\hat{g} = -(I_D - P(P^TP)^{-1}P^T)(dP(P^TP)^{-1}P^T)(I_D - P(P^TP)^{-1}P^T) -$

$-(P(P^TP)^{-1}(dP)^T)(I_D - P(P^TP)^{-1}P^T)^2 - (I_D - P(P^TP)^{-1}P^T)^2(dP(P^TP)^{-1}P^T) -$

$-(I_D - P(P^TP)^{-1}P^T)(P(P^TP)^{-1}(dP)^T)(I_D - P(P^TP)^{-1}P^T) =$

$= -(P(P^TP)^{-1}(dP)^T)(I_D - P(P^TP)^{-1}P^T)^2 - (I_D - P(P^TP)^{-1}P^T)^2(dP(P^TP)^{-1}P^T) =$

$= -(P(P^TP)^{-1}(dP)^T)(I_D - P(P^TP)^{-1}P^T) - (I_D - P(P^TP)^{-1}P^T)(dP(P^TP)^{-1}P^T) =$

$= -P(P^TP)^{-1}(dP)^T - dP(P^TP)^{-1}P^T + P(P^TP)^{-1}(dP)^TP(P^TP)^{-1}P^T +$

$+P(P^TP)^{-1}P^TdP(P^TP)^{-1}P^T$

$dF_P[dP] = N\,\mathrm{tr}(d\tilde{g}_P[dP]\cdot S) = N\Big\langle P(P^TP)^{-1}P^TSP(P^TP)^{-1} - SP(P^TP)^{-1} - S^TP(P^TP)^{-1} +$

$+P(P^TP)^{-1}P^TS^TP(P^TP)^{-1},\ dP\Big\rangle = N\Big\langle P(P^TP)^{-1}P^T(S+S^T)P(P^TP)^{-1} -$

$-(S+S^T)P(P^TP)^{-1},\ dP\Big\rangle = [\text{Учтём, что } S = S^T] = \Big\langle 2N(P(P^TP)^{-1}P^T - I)SP(P^TP)^{-1},\ dP\Big\rangle$

В случае, когда $P^TP = I_D$ получим $dF_P[dP] = \Big\langle 2N(PP^T - I)SP,\ dP\Big\rangle$

Итоговый градиент $\nabla_P F(P) = 2N(PP^T - I)SP$

Задание 2.1 Доказать, что градиент равен нулю при заданных условиях

Необходимо показать, что градиент равен нулю для матрицы P следующего вида: $P = [p_{k_1}, ..., p_{k_d}]$, где $p_{k_1}, ..., p_{k_d}$ - различные собственные векторы из матрицы $Q : S = Q\Lambda Q^T$ Условие означает, что $P^TP = I_D$, и, следовательно, $\nabla_P F(P) = 2N(PP^T - I)Q\Lambda Q^TP$

$$\hat{P} = Q^TP = \underbrace{\begin{array}{c} k_m \\ k_l \\ \\ k_i \\ k_j \end{array}\begin{bmatrix} 0 & 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 1 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \end{bmatrix}}_{D\times d}$$

Единицы стоят на пересечении k_i-й строки и i-го столбца

$$\tilde{P} = \Lambda\hat{P} = \underbrace{\begin{array}{c} k_m \\ k_l \\ \\ k_i \\ k_j \end{array}\begin{bmatrix} 0 & 0 & \lambda_{k_m} & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & \lambda_{k_l} & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \lambda_{k_i} & 0 & 0 & \dots & 0 & 0 \\ 0 & \lambda_{k_j} & 0 & \dots & 0 & 0 \end{bmatrix}}_{D\times d}$$

Собственные значения стоят на пересечении k_i-й строки и i-го столбца

$$P^{'} = Q\tilde{P} = \underbrace{\begin{bmatrix} \lambda_{k_1} q_{k_1} & \lambda_{k_2} q_{k_2} & \lambda_{k_3} q_{k_3} & \dots & \lambda_{k_d} q_{k_d} \end{bmatrix}}_{D\times d}$$

Матрица $\tilde{P}$, умноженная на Q слева

Легко видеть, что

$$P^{'} = P\Lambda^{'} = P \cdot \underbrace{\begin{bmatrix} \lambda_{k_1} & 0 & 0 & \dots & 0 \\ 0 & \lambda_{k_2} & 0 & \dots & 0 \\ 0 & 0 & \lambda_{k_3} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_{k_d} \end{bmatrix}}_{d\times d}$$

Получаем, что $\nabla_P F(P) = 2N(PP^T - I_D)P\Lambda^{'} = 2N(P\underbrace{P^T P}_{I_d} - P)\Lambda^{'} = 2N(P - P)\Lambda^{'} = \theta$

Задание 2.2 Доказать, что значение минимума функции $F(P)$ достигается при матрице P, столбцы которой - собственные векторы, отвечающие наибольшим собственным значениям матрицы S

Доказательство:

$F(P) = N\,\mathrm{tr}\big((I_D - P(P^T P)^{-1} P^T)^2\, S\big) = N\,\mathrm{tr}\big((I_D - PP^T)^2\; Q\Lambda Q^T\big) = N\,\mathrm{tr}\big((I_D - PP^T)\; Q\Lambda Q^T\big) =$

$= N\,\mathrm{tr}\big(Q\Lambda Q^T\big) - N\,\mathrm{tr}\big(PP^T Q\Lambda Q^T\big) = N\,\mathrm{tr}(\Lambda) - N\,\mathrm{tr}\left(P^T \underbrace{Q\Lambda Q^T P}_{2.1 \Rightarrow P\Lambda'}\right) = N\,\mathrm{tr}(\Lambda) - N\,\mathrm{tr}\Big(P^T P\Lambda^{'}\Big) =$

$= N\sum_{i=1}^{D} \lambda_i - N\sum_{i=1}^{d} \lambda_{k_i}$

Видно, что чем большему собственному значению соответствует столбец матрицы P, тем больше вторая сумма в полученной разности $\Rightarrow$ тем меньше итоговая сумма. Следовательно, при выборе собственных векторов, отвечающих максимальных собственным значениям, мы получим минимальное значение функции P $\blacksquare$