МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В.Ломоносова
Факультет вычислительной математики и кибернетики

Задание 3. Практическое задание по матричному дифференцированию

Подготовил:
студент 317 группы ВМК МГУ
Долгушев Глеб Дмитриевич

October 2024

Содержание

Задание 1

Для проверки тождества воспользуемся простым перемножением.

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

$$=\mathbf{I}-\mathbf{U}(\mathbf{C}^{-1}+\mathbf{VA}^{-1}\mathbf{U})^{-1}\mathbf{VA}^{-1}+\mathbf{UCVA}^{-1}-\mathbf{UCVA}^{-1}\mathbf{U}(\mathbf{C}^{-1}+\mathbf{VA}^{-1}\mathbf{U})^{-1}\mathbf{VA}^{-1}$$

Нужно доказать, что это равно I, что равносильно

$$-\mathbf{U}(\mathbf{C}^{-1}+\mathbf{VA}^{-1}\mathbf{U})^{-1}\mathbf{VA}^{-1}+\mathbf{UCVA}^{-1}-\mathbf{UCVA}^{-1}\mathbf{U}(\mathbf{C}^{-1}+\mathbf{VA}^{-1}\mathbf{U})^{-1}\mathbf{VA}^{-1}=\mathbf{0}$$

Домножив на A справа получим

$$-\mathbf{U}(\mathbf{C}^{-1}+\mathbf{VA}^{-1}\mathbf{U})^{-1}\mathbf{V}+\mathbf{UCV}-\mathbf{UCVA}^{-1}\mathbf{U}(\mathbf{C}^{-1}+\mathbf{VA}^{-1}\mathbf{U})^{-1}\mathbf{V}=$$

$$=\mathbf{U}(-(\mathbf{C}^{-1}+\mathbf{VA}^{-1}\mathbf{U})^{-1}+\mathbf{C}-\mathbf{CVA}^{-1}\mathbf{U}(\mathbf{C}^{-1}+\mathbf{VA}^{-1}\mathbf{U})^{-1})\mathbf{V}$$

Домножим каждое слагаемое на $\mathbf{I}=\mathbf{CC}^{-1}$ слева и вынесем $\mathbf{C}$ за скобки

$$\mathbf{UC}(-\mathbf{C}^{-1}(\mathbf{C}^{-1}+\mathbf{VA}^{-1}\mathbf{U})^{-1}+\mathbf{I}-\mathbf{VA}^{-1}\mathbf{U}(\mathbf{C}^{-1}+\mathbf{VA}^{-1}\mathbf{U})^{-1})\mathbf{V}=$$

$$=\mathbf{UC}(-(\mathbf{C}^{-1}+\mathbf{VA}^{-1}\mathbf{U})(\mathbf{C}^{-1}+\mathbf{VA}^{-1}\mathbf{U})^{-1}+\mathbf{I})\mathbf{V}=$$

$$=\mathbf{UC}(-\mathbf{I}+\mathbf{I})\mathbf{V}=\mathbf{0}, \text{ что и требовалось показать}$$

Задание 2

(a)

Преобразуйте выражение: $||\mathbf{uv}^T-\mathbf{A}||_F^2-||\mathbf{A}||_F^2$

$$||\mathbf{uv}^T-\mathbf{A}||_F^2-||\mathbf{A}||_F^2=tr((\mathbf{uv}^T-\mathbf{A})^T(\mathbf{uv}^T-\mathbf{A}))-||\mathbf{A}||_F=$$

$$=tr(||\mathbf{u}||^2\mathbf{vv}^T)-tr(\mathbf{vu}^T\mathbf{A})-tr(\mathbf{A}^T\mathbf{uv}^T)+tr(\mathbf{A}^T\mathbf{A})-tr(\mathbf{A}^T\mathbf{A})=$$

$$=tr(||\mathbf{u}||^2\mathbf{vv}^T)-tr(\mathbf{vu}^T\mathbf{A})-tr(\mathbf{A}^T\mathbf{uv}^T)=$$

Используя св-во следа $tr(AB)=tr(BA)$ получим

$$||\mathbf{u}||^2||\mathbf{v}||^2-2tr(\mathbf{vu}^T\mathbf{A})$$

(b)

Преобразуйте выражение: $tr((2\mathbf{I}_n+\mathbf{aa}^T)^{-1}(\mathbf{uv}^T+\mathbf{vu}^T))$

Применяя тождество Вудбери, в котором положим $\mathbf{U}=\mathbf{a},\mathbf{V}=\mathbf{a}^T,\mathbf{A}=2\mathbf{I}_n,\mathbf{C}=1$ получим

$$(2I_n+aa^T)^{-1}=\frac{1}{2}I_n-\frac{1}{2}\mathbf{a}(1+\frac{1}{2}\mathbf{a}^T\mathbf{a})^{-1}\mathbf{a}^T\frac{1}{2}=$$

$$=\frac{1}{2}\left[\mathbf{I}_n-\frac{1}{2+||\mathbf{a}||^2}\mathbf{aa}^T\right]$$

Тогда исходное выражение примет вид

$$\frac{1}{2}tr(\left[\mathbf{I}_n - \frac{1}{2+||\mathbf{a}||^2}\mathbf{a}\mathbf{a}^T\right](\mathbf{u}\mathbf{v}^T + \mathbf{v}\mathbf{u}^T)) =$$

$$\frac{1}{2}(\left[tr(\mathbf{u}\mathbf{v}^T) + tr(\mathbf{v}\mathbf{u}^T) - \frac{1}{2+||\mathbf{a}||^2}tr(\mathbf{a}\mathbf{a}^T(\mathbf{u}\mathbf{v}^T + \mathbf{v}\mathbf{u}^T))\right]$$

Заметим, что $tr(X) = tr(X^T)$ и $(\mathbf{v}\mathbf{u}^T)^T = \mathbf{u}\mathbf{v}^T, (\mathbf{a}\mathbf{a}^T\mathbf{u}\mathbf{v}^T)^T = \mathbf{v}\mathbf{u}^T\mathbf{a}\mathbf{a}^T$, кроме того учтем, что $tr(\mathbf{a}\mathbf{a}^T\mathbf{v}\mathbf{u}^T) = tr(\mathbf{v}\mathbf{u}^T\mathbf{a}\mathbf{a}^T)$. Перепишем начальное выражение

$$\begin{aligned} tr(\mathbf{u}\mathbf{v}^T) - \frac{1}{2+||\mathbf{a}||^2}tr(\mathbf{a}\mathbf{a}^T\mathbf{u}\mathbf{v}^T) = \\ = \langle \mathbf{u}, \mathbf{v}\rangle - \frac{1}{2+||\mathbf{a}||^2}\langle \mathbf{a}, \mathbf{u}\rangle tr(\mathbf{a}\mathbf{v}^T) = \\ = \langle \mathbf{u}, \mathbf{v}\rangle - \frac{1}{2+||\mathbf{a}||^2}\langle \mathbf{a}, \mathbf{u}\rangle \langle \mathbf{a}, \mathbf{v}\rangle \end{aligned}$$

В последнем выражении также учли свойство следа $tr(ABC) = tr(CAB)$.

(c)

Преобразуйте выражение $\sum\limits_{i=1}^{n}(\mathbf{S}^{-1}\mathbf{a}_i, \mathbf{a}_i)$, где $\mathbf{a}_i \in \mathbb{R}^d$ и $\mathbf{S} = \sum\limits_{i=1}^{n}\mathbf{a}_i\mathbf{a}_i^T$ Рассмотрим, что такое $\mathbf{S}^{-1}$

$$\mathbf{S}^{-1}\mathbf{S} = \mathbf{I}_n \Rightarrow \mathbf{S}^{-1}\sum_{i=1}^{n}\mathbf{a}_i\mathbf{a}_i^T = \mathbf{I}_d$$

Обозначим $\mathbf{b}_i = \mathbf{S}^{-1}\mathbf{a}_i$, тогда $\sum\limits_{i=1}^{n}\mathbf{b}_i\mathbf{a}_i^T = \mathbf{I}_d$

$$\Rightarrow \sum_{i=1}^{n}\mathbf{b}_{ik_1}\mathbf{a}_{ik_2} = \begin{cases} 1, k_1 = k_2 \\ 0, k_1 \neq k_2 \end{cases}, k_1, k_2 = \overline{1, d}$$

Тогда

$$\sum_{i=1}^{n}\langle \mathbf{b}_i, \mathbf{a}_i\rangle = \sum_{i=1}^{n}\sum_{j=1}^{d}\mathbf{b}_{ij}\mathbf{a}_{ij} = \sum_{i=1}^{d}\sum_{j=1}^{n}\mathbf{b}_{ij}\mathbf{a}_{ij} = d$$

Задание 3

(a)

Учитывая полученные на семинаре результаты, а именно $d(det\mathbf{X})[\mathbf{H}] = det\mathbf{X}\left\langle \mathbf{X}^{-T}, \mathbf{H}\right\rangle$, а также формулу для дифференцирования сложной функции получим

$$\begin{aligned} df(t)[h_1] = d(det(\mathbf{A} - t\mathbf{I}_n)) = det(\mathbf{A} - t\mathbf{I}_n)\left\langle (\mathbf{A} - t\mathbf{I}_n)^{-T}, d(\mathbf{A} - t\mathbf{I}_n)[h_1]\right\rangle = \\ = det(\mathbf{A} - t\mathbf{I}_n)\left\langle (\mathbf{A} - t\mathbf{I}_n)^{-T}, -\mathbf{I}_n h_1\right\rangle = \\ = -det(\mathbf{A} - t\mathbf{I}_n)\left\langle (\mathbf{A} - t\mathbf{I}_n)^{-T}, \mathbf{I}_n h_1\right\rangle \end{aligned}$$

Тогда

$$d(df(t)[h_1])[h_2] = d(-det(\mathbf{A} - t\mathbf{I}_n)\left\langle (\mathbf{A} - t\mathbf{I}_n)^{-T}, \mathbf{I}_n h_1 \right\rangle)[h_2] =$$
$$= -d(det(\mathbf{A} - t\mathbf{I}_n))[h_2]\left\langle (\mathbf{A} - t\mathbf{I}_n)^{-T}, \mathbf{I}_n h_1 \right\rangle -$$
$$-det(\mathbf{A} - t\mathbf{I}_n) d\left\langle (\mathbf{A} - t\mathbf{I}_n)^{-T}, \mathbf{I}_n h_1 \right\rangle [h_2] =$$
$$= det(\mathbf{A} - t\mathbf{I}_n)\left\langle (\mathbf{A} - t\mathbf{I}_n)^{-T}, \mathbf{I}_n h_2 \right\rangle \left\langle (\mathbf{A} - t\mathbf{I}_n)^{-T}, \mathbf{I}_n h_1 \right\rangle -$$
$$-det(\mathbf{A} - t\mathbf{I}_n)\left\langle d((\mathbf{A} - t\mathbf{I}_n)^{-T})[h_2], \mathbf{I}_n h_1 \right\rangle$$

Пользуясь фактом с семинара, который легко вывести $d(X^{-1}) = -X^{-1}dXX^{-1}$ преобразуем полученное выражение к виду

$$det(\mathbf{A} - t\mathbf{I}_n)\left\langle (\mathbf{A} - t\mathbf{I}_n)^{-T}, \mathbf{I}_n h_2 \right\rangle \left\langle (\mathbf{A} - t\mathbf{I}_n)^{-T}, \mathbf{I}_n h_1 \right\rangle +$$
$$+det(\mathbf{A} - t\mathbf{I}_n)\left\langle (\mathbf{A} - t\mathbf{I}_n)^{-T} d(\mathbf{A} - t\mathbf{I}_n)^T[h_2](\mathbf{A} - t\mathbf{I}_n)^{-T}, \mathbf{I}_n h_1 \right\rangle =$$
$$= det(\mathbf{A} - t\mathbf{I}_n)\left\langle (\mathbf{A} - t\mathbf{I}_n)^{-T}, \mathbf{I}_n h_2 \right\rangle \left\langle (\mathbf{A} - t\mathbf{I}_n)^{-T}, \mathbf{I}_n h_1 \right\rangle - det(\mathbf{A} - t\mathbf{I}_n)\left\langle (\mathbf{A} - t\mathbf{I}_n)^{-2T} h_2, \mathbf{I}_n h_1 \right\rangle$$

Далее вспомним, что приращения h_1, h_2 - числа и приведем полученный результат к удобному виду

$$d(df(t)[h_1])[h_2] = det(\mathbf{A} - t\mathbf{I}_n)\left[tr((\mathbf{A} - t\mathbf{I}_n)^{-1})^2 - tr((\mathbf{A} - t\mathbf{I}_n)^{-2})\right] h_2 h_1$$

(b)

Исходная функция $||(\mathbf{A} + t\mathbf{I}_n)^{-1}\mathbf{b}||$

$$d(||(\mathbf{A} + t\mathbf{I}_n)^{-1}\mathbf{b}||)[h_1] = d\sqrt{\langle (\mathbf{A} + t\mathbf{I}_n)^{-1}\mathbf{b}, (\mathbf{A} + t\mathbf{I}_n)^{-1}\mathbf{b} \rangle} =$$
$$= \frac{d\langle (\mathbf{A} + t\mathbf{I}_n)^{-1}\mathbf{b}, (\mathbf{A} + t\mathbf{I}_n)^{-1}\mathbf{b} \rangle [h_1]}{2||(\mathbf{A} + t\mathbf{I}_n)^{-1}\mathbf{b}||} =$$
$$= \frac{\langle (\mathbf{A} + t\mathbf{I}_n)^{-1}\mathbf{b}, d\left[(\mathbf{A} + t\mathbf{I}_n)^{-1}\mathbf{b}\right] \rangle [h_1]}{||(\mathbf{A} + t\mathbf{I}_n)^{-1}\mathbf{b}||} =$$
$$= \frac{\langle (\mathbf{A} + t\mathbf{I}_n)^{-1}\mathbf{b}, d(\mathbf{A} + t\mathbf{I}_n)^{-1}[h_1]\mathbf{b} \rangle}{||(\mathbf{A} + t\mathbf{I}_n)^{-1}\mathbf{b}||} =$$
$$= -\frac{\langle (\mathbf{A} + t\mathbf{I}_n)^{-1}\mathbf{b}, (\mathbf{A} + t\mathbf{I}_n)^{-2}\mathbf{b}h_1 \rangle}{||(\mathbf{A} + t\mathbf{I}_n)^{-1}\mathbf{b}||}$$

Так как $\mathbf{A} \in \mathbb{S}^n_+$, то $(\mathbf{A} + t\mathbf{I}_n)^{-T} = (\mathbf{A} + t\mathbf{I}_n)^{-1}$, тогда преобразованное выражение примет вид

$$-\frac{b^T(\mathbf{A} + t\mathbf{I}_n)^{-3}\mathbf{b}}{||(\mathbf{A} + t\mathbf{I}_n)^{-1}\mathbf{b}||} h_1$$

Тогда вторая производная

$$d(df(t)[h_1])[h_2] = d\left(-\frac{\mathbf{b}^T(\mathbf{A} + t\mathbf{I}_n)^{-3}\mathbf{b}}{||(\mathbf{A} + t\mathbf{I}_n)^{-1}\mathbf{b}||} h_1\right)[h_2] =$$
$$= -\frac{d(\mathbf{b}^T(\mathbf{A} + t\mathbf{I}_n)^{-3}\mathbf{b})[h_2]}{||(\mathbf{A} + t\mathbf{I}_n)^{-1}\mathbf{b}||} h_1 - d\left(\frac{1}{||(\mathbf{A} + t\mathbf{I}_n)^{-1}\mathbf{b}||}\right)[h_2]\mathbf{b}^T(\mathbf{A} + t\mathbf{I}_n)^{-3}\mathbf{b})h_1 =$$
$$= -\frac{\mathbf{b}^T d\left((\mathbf{A} + t\mathbf{I}_n)^{-3}\right)[h_2]\mathbf{b}}{||(\mathbf{A} + t\mathbf{I}_n)^{-1}\mathbf{b}||} h_1 + \frac{1}{||(\mathbf{A} + t\mathbf{I}_n)^{-1}b||^2} d\left(||(\mathbf{A} + t\mathbf{I}_n)^{-1}\mathbf{b}||\right)[h_2]\mathbf{b}^T(\mathbf{A} + t\mathbf{I}_n)^{-3}\mathbf{b})h_1 =$$

Производная $d\left(||(\mathbf{A}+t\mathbf{I}_n)^{-1}\mathbf{b}||\right)[h_2]$ уже посчитана и равна $-\frac{\mathbf{b}^T(\mathbf{A}+t\mathbf{I}_n)^{-3}\mathbf{b}}{||(\mathbf{A}+t\mathbf{I}_n)^{-1}\mathbf{b}||}h_2$, посчитаем выражение $d(\mathbf{X}^{-3})$

$$d(\mathbf{X}^{-3}) = d((\mathbf{X}^{-1})^3) = \mathbf{X}^{-2}d(\mathbf{X}^{-1}) + \mathbf{X}^{-1}d(\mathbf{X}^{-1})\mathbf{X}^{-1} + d(\mathbf{X}^{-1})\mathbf{X}^{-2} =$$
$$-\mathbf{X}^{-3}\mathbf{X}\mathbf{X}^{-1} - \mathbf{X}^{-2}\mathbf{X}\mathbf{X}^{-2} - \mathbf{X}^{-1}\mathbf{X}\mathbf{X}^{-3})$$

тогда, т.к. $d(\mathbf{A}+t\mathbf{I}_n) = \mathbf{I}$

$$d\left((\mathbf{A}+t\mathbf{I}_n)^{-3}\right)[h_2] = -3(\mathbf{A}+t\mathbf{I}_n)^{-4}h_2$$

и искомое выражение равно

$$\frac{3\mathbf{b}^T(\mathbf{A}+t\mathbf{I}_n)^{-4}\mathbf{b}}{||(\mathbf{A}+t\mathbf{I}_n)^{-1}\mathbf{b}||}h_2h_1 - \frac{(\mathbf{b}^T(\mathbf{A}+t\mathbf{I}_n)^{-3}\mathbf{b})^2}{||(\mathbf{A}+t\mathbf{I}_n)^{-1}\mathbf{b}||^3}h_2h_1$$

Задание 4

(a)

$$f(x) = \frac{1}{2}||xx^T - A||_F^2$$

$$df(x)[h_1] = d\left(\frac{1}{2}||xx^T - A||_F^2\right) = d\left(\frac{1}{2}\left\langle xx^T - A, xx^T - A\right\rangle\right) =$$
$$= \left\langle xx^T - A, d\left(xx^T - A\right)\right\rangle =$$
$$= \left\langle xx^T - A, dxx^T + xdx^T\right\rangle$$

Покажем что $\left\langle xx^T - A, dxx^T\right\rangle = \left\langle xx^T - A, xdx^T\right\rangle$ (заметим, что $(dx)^T = d(x^T) = dx^T$)

$$\left\langle xx^T - A, dxx^T\right\rangle = tr(xdx^T(xx^T - A)) = tr((xx^T - A)^T dxx^T) = tr(dxx^T(xx^T - A)^T) =$$
$$= \left\langle xx^T - A, xdx^T\right\rangle$$

Тогда искомое

$$2\left\langle xx^T - A, dxx^T\right\rangle = 2\left\langle xx^Tx - Ax, dx\right\rangle = 2\left\langle (||x||^2 I - A)x, dx\right\rangle$$
$$\Rightarrow \nabla f = 2(||x||^2 I - A)x$$

$$d(df(x)[h1])[h_2] = d\left(2\left\langle (||x||^2 I - A)x, h_1\right\rangle\right)[h_2] =$$
$$= 2\left\langle (d(||x||^2)[h_2]Ix, h_1\right\rangle + 2\left\langle (||x||^2 I - A)dx[h_2], h_1\right\rangle$$

Из предыдущих заданий $d(||x||^2) = 2\langle x, dx\rangle$, тогда

$$2\left\langle 2\left\langle x, h_2\right\rangle x, h_1\right\rangle + 2\left\langle (||x||^2 I - A)h_2, h_1\right\rangle = 2\left\langle \left(2xx^t + ||x||^2 I - A\right)h_2, h_1\right\rangle$$
$$\Rightarrow \nabla^2 f = 4xx^t + 2||x||^2 I - 2A$$

(b)

$$f(x) = \langle x, x\rangle^{\langle x,x\rangle}$$

$$df(x)[h_1] = d\left(\langle x, x\rangle^{\langle x,x\rangle}\right)[h_1] = d\left(e^{ln(\langle x,x\rangle)\langle x,x\rangle}\right)[h_1] =$$

$$= \langle x, x\rangle^{\langle x,x\rangle}\left[\frac{2\langle x, dx\rangle}{\langle x, x\rangle}\langle x, x\rangle + 2ln\langle x, x\rangle\langle x, dx\rangle\right] =$$

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

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

$$d(df(x)[h_1])[h_2] = d\left(2\langle x, x\rangle^{\langle x,x\rangle}\left(1 + ln\langle x, x\rangle\right)\langle x, h_1\rangle\right)[h_2]$$

Данная производная будет представляться как сумма трех, которые мы посчитаем далее

1.
$$d\left(2\langle x, x\rangle^{\langle x,x\rangle}\right)[h_2]\left(1 + ln\langle x, x\rangle\right)\langle x, h_1\rangle = 4\langle x, x\rangle^{\langle x,x\rangle}\left(1 + ln\langle x, x\rangle\right)^2\langle x, h_1\rangle\langle x, h_2\rangle =$$
$$= h_1^T 4x\langle x, x\rangle^{\langle x,x\rangle}\left(1 + ln\langle x, x\rangle\right)^2 x^T h_2$$

2.
$$2\langle x, x\rangle^{\langle x,x\rangle} d\left(ln\langle x, x\rangle\right)[h_2]\langle x, h_1\rangle = 4\langle x, x\rangle^{\langle x,x\rangle - 1}\langle x, h_2\rangle\langle x, h_1\rangle$$
$$= h_1^T 4x\langle x, x\rangle^{\langle x,x\rangle - 1} x^T h_2$$

3.
$$2\langle x, x\rangle^{\langle x,x\rangle}\left(1 + ln\langle x, x\rangle\right)\langle h_2, h_1\rangle =$$
$$= h_1^T 2\langle x, x\rangle^{\langle x,x\rangle}\left(1 + ln\langle x, x\rangle\right)h_2$$

Тогда вторая производная примет вид

$$h_1^T 2\langle x, x\rangle^{\langle x,x\rangle}\left[2\left(1 + ln\langle x, x\rangle\right)^2 xx^T + 2\langle x, x\rangle^{-1} xx^T + \left(1 + ln\langle x, x\rangle\right)I\right]h_2$$

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

(c)

$$f(x) = ||Ax-b||^p$$

$$df(x)[h_1] = d\left\langle Ax-b, Ax-b\right\rangle^{\frac{p}{2}}[h_1] = \frac{p}{2}\left\langle Ax-b, Ax-b\right\rangle^{\frac{p-2}{2}} 2\left\langle Ax-b, Ah_1\right\rangle =$$

$$= p||Ax-b||^{p-2}\left\langle A^T(Ax-b), h_1\right\rangle$$

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

Вторая производная

$$d(f(x)[h_1])[h_2] = d\left(p||Ax-b||^{p-2}\left\langle A^T(Ax-b), h_1\right\rangle\right) =$$

$$= pd\left(||Ax-b||^{p-2}\right)\left\langle A^T(Ax-b), h_1\right\rangle + p||Ax-b||^{p-2}\left\langle d\left(A^T(Ax-b)\right), h_1\right\rangle =$$

$$= p(p-2)||Ax-b||^{p-4}\left\langle A^T(Ax-b), h_2\right\rangle\left\langle A^T(Ax-b), h_1\right\rangle + p||Ax-b||^{p-2}\left\langle A^TAh_2, h_1\right\rangle =$$

$$= h_1^T\left[p(p-2)||Ax-b||^{p-4}A^T(Ax-b)(Ax-b)^TA + p||Ax-b||^{p-2}A^TA\right]h_2 =$$

$$= \left\langle\left[p(p-2)||Ax-b||^{p-4}A^T(Ax-b)(Ax-b)^TA + p||Ax-b||^{p-2}A^TA\right]h_2, h_1\right\rangle =$$

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

Задание 5

(a)

Показать знакоопределенность второго дифференциала $f(\mathbf{X}) = tr(\mathbf{X}^{-1})$

$$df(\mathbf{X})[\mathbf{H}_1] = d\left\langle \mathbf{X}^{-1}, \mathbf{I}_n\right\rangle[\mathbf{H}_1] = -\left\langle \mathbf{X}^{-1}\mathbf{H}\mathbf{X}^{-1}, \mathbf{I}_n\right\rangle$$

$$d(df(\mathbf{X})[\mathbf{H}_1])[\mathbf{H}_2] = -\left\langle d\mathbf{X}^{-1}\mathbf{H}\mathbf{X}^{-1}, \mathbf{I}_n\right\rangle - \left\langle \mathbf{X}^{-1}\mathbf{H}d\mathbf{X}^{-1}, \mathbf{I}_n\right\rangle =$$

$$= \left\langle \mathbf{X}^{-1}\mathbf{H}_1\mathbf{X}^{-1}\mathbf{H}_2\mathbf{X}^{-1}, \mathbf{I}_n\right\rangle + \left\langle \mathbf{X}^{-1}\mathbf{H}_2\mathbf{X}^{-1}\mathbf{H}_1\mathbf{X}^{-1}, \mathbf{I}_n\right\rangle$$

В силу положительной определенности $\mathbf{X}$ положительно определена и $\mathbf{X}^{-1}$ и, кроме того, из нее можно извлечь положительно определенный корень $\mathbf{X}^{-\frac{1}{2}}$.

$$d^2f(\mathbf{X})[\mathbf{H},\mathbf{H}] = 2\left\langle \mathbf{X}^{-1}\mathbf{H}\mathbf{X}^{-1}\mathbf{H}\mathbf{X}^{-1}, \mathbf{I}_n\right\rangle = 2\left\langle \mathbf{X}^{-1}\mathbf{H}\mathbf{X}^{-\frac{1}{2}}\mathbf{X}^{-\frac{1}{2}}\mathbf{H}\mathbf{X}^{-1}, \mathbf{I}_n\right\rangle =$$

$$= 2\left\langle \mathbf{X}^{-\frac{1}{2}}\mathbf{H}\mathbf{X}^{-1}, \mathbf{X}^{-\frac{1}{2}}\mathbf{H}\mathbf{X}^{-1}\right\rangle > 0, \forall \mathbf{H} \in \mathbb{S}^n_{++}$$

$\Rightarrow$ Вторая производная является положительно-определенной

(b)

Показать знакоопределенность второго дифференциала $det(\mathbf{X})^{\frac{1}{n}}$
Учитывая $\mathbf{X} \in \mathbb{S}^n_{++}$ и, следовательно, $\mathbf{X}^{-T} = \mathbf{X}^{-1}$

$$df(\mathbf{X})[\mathbf{H}_1] = \frac{1}{n} det(\mathbf{X})^{\frac{1}{n}-1} det(\mathbf{X}) \left\langle \mathbf{X}^{-1}, \mathbf{H}_1 \right\rangle = \frac{1}{n} det(\mathbf{X})^{\frac{1}{n}} \left\langle \mathbf{X}^{-1}, \mathbf{H_1} \right\rangle$$

$$d^2 f(\mathbf{X})[\mathbf{H}_1, \mathbf{H}_2] = \frac{1}{n} d\left(det(\mathbf{X})^{\frac{1}{n}}\right) \left\langle \mathbf{X}^{-1}, \mathbf{H}_1 \right\rangle + \frac{1}{n} det(\mathbf{X})^{\frac{1}{n}} \left\langle d\mathbf{X}^{-1}, \mathbf{H}_1 \right\rangle =$$

$$= \frac{1}{n}^2 det(\mathbf{X})^{\frac{1}{n}} \left\langle \mathbf{X}^{-1}, \mathbf{H}_1 \right\rangle \left\langle \mathbf{X}^{-1}, \mathbf{H}_2 \right\rangle - \frac{1}{n} det(\mathbf{X})^{\frac{1}{n}} \left\langle \mathbf{X}^{-1}\mathbf{H}_2\mathbf{X}^{-1}, \mathbf{H}_1 \right\rangle$$

$$d^2 f(\mathbf{X})[\mathbf{H}, \mathbf{H}] = \frac{1}{n} det(\mathbf{X})^{\frac{1}{n}} \left[\frac{1}{n} \left\langle \mathbf{X}^{-1}, \mathbf{H} \right\rangle^2 - \left\langle \mathbf{X}^{-1}\mathbf{H}\mathbf{X}^{-1}, \mathbf{H} \right\rangle\right] =$$

$$= \frac{1}{n} det(\mathbf{X})^{\frac{1}{n}} \left[\frac{1}{n} tr\left(\mathbf{H}^T\mathbf{X}^{-1}\right)^2 - tr\left(\mathbf{H}^T\mathbf{X}^{-1}\mathbf{H}\mathbf{X}^{-1}\right)\right] =$$

$$= \frac{1}{n} det(\mathbf{X})^{\frac{1}{n}} \left[\frac{1}{n} tr\left(\mathbf{H}^T\mathbf{X}^{-\frac{1}{2}}\mathbf{X}^{-\frac{1}{2}}\right)^2 - tr\left(\mathbf{H}^T\mathbf{X}^{-\frac{1}{2}}\mathbf{X}^{-\frac{1}{2}}\mathbf{H}\mathbf{X}^{-\frac{1}{2}}\mathbf{X}^{-\frac{1}{2}}\right)\right] =$$

$$= \frac{1}{n} det(\mathbf{X})^{\frac{1}{n}} \left[\frac{1}{n} tr\left(\mathbf{X}^{-\frac{1}{2}}\mathbf{H}^T\mathbf{X}^{-\frac{1}{2}}\right)^2 - tr\left(\mathbf{X}^{-\frac{1}{2}}\mathbf{H}^T\mathbf{X}^{-\frac{1}{2}}\mathbf{X}^{-\frac{1}{2}}\mathbf{H}\mathbf{X}^{-\frac{1}{2}}\right)\right]$$

Обозначим $\widetilde{\mathbf{H}} = \mathbf{X}^{-\frac{1}{2}}\mathbf{H}\mathbf{X}^{-\frac{1}{2}}$

$$= \frac{1}{n} det(\mathbf{X})^{\frac{1}{n}} \left[\frac{1}{n} \left\langle \widetilde{\mathbf{H}}, \mathbf{I}_n \right\rangle^2 - \left\langle \widetilde{\mathbf{H}}, \widetilde{\mathbf{H}} \right\rangle\right]$$

Из неравенства КБШ

$$\left\langle \widetilde{\mathbf{H}}, \mathbf{I}_n \right\rangle^2 \leq \left\langle \widetilde{\mathbf{H}}, \widetilde{\mathbf{H}} \right\rangle \langle \mathbf{I}_n, \mathbf{I}_n \rangle = n \left\langle \widetilde{\mathbf{H}}, \widetilde{\mathbf{H}} \right\rangle$$

Следовательно

$$\frac{1}{n} \left\langle \widetilde{\mathbf{H}}, \mathbf{I}_n \right\rangle^2 - \left\langle \widetilde{\mathbf{H}}, \widetilde{\mathbf{H}} \right\rangle \leq 0$$

И найденная вторая производная неположительно-определенная.

Задание 6

(a)

$$f(\mathbf{x}) = \langle \mathbf{c}, \mathbf{x} \rangle + \frac{\sigma}{3} ||\mathbf{x}||^3$$

$$df(\mathbf{x})[\mathbf{h}] = \langle \mathbf{c}, \mathbf{h} \rangle + \frac{\sigma}{3} 3||\mathbf{x}||^2 \frac{2 \langle \mathbf{x}, \mathbf{h} \rangle}{2||\mathbf{x}||} = \langle \mathbf{c}, \mathbf{h} \rangle + \sigma ||\mathbf{x}|| \langle \mathbf{x}, \mathbf{h} \rangle$$

Тогда $\nabla f = \mathbf{c} + \sigma ||\mathbf{x}|| \mathbf{x}$ Точки стационарности

$$\mathbf{c} + \sigma ||\mathbf{x}|| \mathbf{x} = 0$$

$$||\mathbf{x}||\mathbf{x} = -\frac{\mathbf{c}}{\sigma}$$

Тогда

$$\mathbf{x} = -\frac{\sqrt{||\mathbf{c}||}\mathbf{c}}{\sqrt{\sigma}||\mathbf{c}||} = -\frac{\mathbf{c}}{\sqrt{\sigma||\mathbf{c}||}}$$

Для любых значений параметров

(b)

$$f(\mathbf{x}) = \langle \mathbf{a}, \mathbf{x} \rangle - ln\left(1 - \langle \mathbf{b}, \mathbf{x} \rangle\right)$$

$$df(\mathbf{x})[\mathbf{h}] = \langle \mathbf{a}, \mathbf{h} \rangle + \frac{\langle \mathbf{b}, \mathbf{h} \rangle}{1 - \langle \mathbf{b}, \mathbf{x} \rangle}$$

Тогда $\nabla f = \mathbf{a} + \frac{\mathbf{b}}{1-\langle \mathbf{b}, \mathbf{x} \rangle}$ Точки стационарности

$$\mathbf{a} + \frac{\mathbf{b}}{1 - \langle \mathbf{b}, \mathbf{x} \rangle} = 0$$

$$\frac{\mathbf{b}}{1 - \langle \mathbf{b}, \mathbf{x} \rangle} = -\mathbf{a}$$

Данное равенство имеет решения, если вектора $\mathbf{a}$ и $\mathbf{b}$ коллинеарны, тогда

$$\mathbf{b} = \pm\frac{||\mathbf{b}||}{||\mathbf{a}||}\mathbf{a}, \text{ и}$$

$$1 - \langle \mathbf{b}, \mathbf{x} \rangle = \mp\frac{||\mathbf{b}||}{||\mathbf{a}||}$$

Тогда множество стационарных точек задается как гиперплоскость

$$\langle \mathbf{b}, \mathbf{x} \rangle = 1 \pm \frac{||\mathbf{b}||}{||\mathbf{a}||}$$

(c)

$$f(\mathbf{x}) = \langle \mathbf{c}, \mathbf{x} \rangle\, exp\left(-\langle \mathbf{A}\mathbf{x}, \mathbf{x} \rangle\right)$$

$$df(\mathbf{x})[\mathbf{h}] = \langle \mathbf{c}, \mathbf{h} \rangle\, exp\left(-\langle \mathbf{A}\mathbf{x}, \mathbf{x} \rangle\right) - \langle \mathbf{c}, \mathbf{x} \rangle\, exp\left(-\langle \mathbf{A}\mathbf{x}, \mathbf{x} \rangle\right) \left\langle (\mathbf{A} + \mathbf{A}^{\mathbf{T}})\mathbf{x}, \mathbf{h} \right\rangle$$

Так как $\mathbf{A} \in \mathbb{S}^n_{++}$

$$df(\mathbf{x})[\mathbf{h}] = \langle \mathbf{c}, \mathbf{h} \rangle\, exp\left(-\langle \mathbf{A}\mathbf{x}, \mathbf{x} \rangle\right) - \langle \mathbf{c}, \mathbf{x} \rangle\, exp\left(-\langle \mathbf{A}\mathbf{x}, \mathbf{x} \rangle\right) \langle 2\mathbf{A}\mathbf{x}, \mathbf{h} \rangle$$

Тогда

$$\nabla f = exp\left(-\langle \mathbf{A}\mathbf{x}, \mathbf{x} \rangle\right)\left[\mathbf{c} - \langle \mathbf{c}, \mathbf{x} \rangle\, 2\mathbf{A}\mathbf{x}\right]$$

Точки стационарности

$$\mathbf{c} - 2\langle \mathbf{c}, \mathbf{x} \rangle\, \mathbf{A}\mathbf{x} = 0$$

$$\mathbf{A}\mathbf{x} = \frac{\mathbf{c}}{2\left\langle \mathbf{c}, \mathbf{x} \right\rangle}$$

Так как $\mathbf{A} \in \mathbb{S}^n_{++}$, то имеет обратную и

$$\mathbf{x} = \frac{\mathbf{A}^{-1}\mathbf{c}}{2\left\langle \mathbf{c}, \mathbf{x} \right\rangle} \tag{1}$$

Чтобы преобразовать это выражение домножим обе части на $\mathbf{c}^T$ слева

$$\mathbf{c}^T\mathbf{x} = \frac{\mathbf{c}^T\mathbf{A}^{-1}\mathbf{c}}{2\left\langle \mathbf{c}, \mathbf{x} \right\rangle}$$

$$\left\langle \mathbf{c}, \mathbf{x} \right\rangle = \frac{\mathbf{c}^T\mathbf{A}^{-1}\mathbf{c}}{2\left\langle \mathbf{c}, \mathbf{x} \right\rangle}$$

Отсюда

$$\left\langle \mathbf{c}, \mathbf{x} \right\rangle = \frac{1}{\sqrt{2}}\sqrt{\mathbf{c}^T\mathbf{A}^{-1}\mathbf{c}}$$

И выражение (1) для $\mathbf{x}$ принимает вид

$$\mathbf{x} = \frac{\mathbf{A}^{-1}\mathbf{c}}{\sqrt{2\mathbf{c}^T\mathbf{A}^{-1}\mathbf{c}}} \tag{2}$$

То есть для функции $f(\mathbf{x}) = \left\langle \mathbf{c}, \mathbf{x} \right\rangle exp\left(-\left\langle \mathbf{A}\mathbf{x}, \mathbf{x} \right\rangle\right)$ существует единственная точка стационарности при любых значениях параметров, которая определена выражением (2).

Задание 7

Воспользуемся утверждением о том, что любую матрицу можно свести преобразованиями подобия в верхнетреугольному виду, то есть $\mathbf{X} = \mathbf{P}^{-1}\mathbf{U}\mathbf{P}$, тогда

$$tr\left(\mathbf{X}^{-k} - \left(\mathbf{X}^k + \mathbf{X}^{2k}\right)^{-1}\right) = tr\left(\mathbf{P}^{-1}\mathbf{U}^{-k}\mathbf{P} - \left(\mathbf{P}^{-1}\mathbf{U}^k\mathbf{P} + \mathbf{P}^{-1}\mathbf{U}^{2k}\mathbf{P}\right)^{-1}\right) =$$

$$= tr\left(\mathbf{P}^{-1}\mathbf{U}^{-k}\mathbf{P} - \left(\mathbf{P}^{-1}\left(\mathbf{U}^k + \mathbf{U}^{2k}\right)\mathbf{P}\right)^{-1}\right) = tr\left(\mathbf{P}^{-1}\mathbf{U}^{-k}\mathbf{P} - \mathbf{P}^{-1}\left(\mathbf{U}^k + \mathbf{U}^{2k}\right)^{-1}\mathbf{P}\right) =$$

$$tr\left(\mathbf{U}^{-k} - \left(\mathbf{U}^k + \mathbf{U}^{2k}\right)^{-1}\right)$$

При этом матрица $\mathbf{U}^k + \mathbf{U}^{2k}$ - верхнетреугольная, с собственными значениями $\lambda_i^k + \lambda_i^{2k}$ на диагонали. Так как обратная к верхнетреугольной - верхнетреугольная, а собственные значения матрицы и обратной ей связаны следующим соотношением

$$\widetilde{\lambda}_i = \frac{1}{\lambda_i}, \text{ где } \lambda_i \text{ - с.з. матрицы, а } \widetilde{\lambda}_i \text{ - с.з. обратной,}$$

то получим, что $\left(\mathbf{U}^k + \mathbf{U}^{2k}\right)^{-1}$ - верхнетреугольная, с собственными значениями $\frac{1}{\lambda_i^k + \lambda_i^{2k}}$ на диагонали.

Аналогично $\mathbf{U}^{-k}$ - верхнетреугольная, с собственными значениями $\frac{1}{\lambda_i^k}$ на диагонали. Тогда ясно, что

$$tr\left(\mathbf{U}^{-k}+\left(\mathbf{U}^{k}-\mathbf{U}^{2k}\right)^{-1}\right)=\sum_{i=1}^{n}\frac{1}{\lambda_i^k}-\frac{1}{\lambda_i^k+\lambda_i^{2k}}=\sum_{i=1}^{n}\frac{\lambda_i^{2k}}{\lambda_i^k(\lambda_i^k+\lambda_i^{2k})}=$$

$$=\sum_{i=1}^{n}\frac{1}{1+\lambda_i^k}$$

$$\lim_{k\to\inf}\frac{1}{1+\lambda_i^k}=\begin{cases}1,\lambda_i<1\\ \frac{1}{2},\lambda_i=1\\ 0,\lambda_i>1\end{cases}\quad\Rightarrow\lim_{k\to\inf}\sum_{i=1}^{n}\frac{1}{1+\lambda_i^k}=\frac{1}{2}N_1+N_{(0,1)},\text{ где}$$

N_1- кол-во с.з. равных 1, $N_{(0,1)}$ - кол-во с.з. принадлежащих интервалу $(0,1)$.

Задание 8

$$F(\mathbf{P})=Ntr\left(\left(\mathbf{I}-\mathbf{P}(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\right)^2\mathbf{S}\right),\text{ где}$$

$\mathbf{S}=\frac{1}{N}\sum\limits_{i=1}^{N}\mathbf{x_i}\mathbf{x_i^T}\Rightarrow\mathbf{S}^T=\mathbf{S}$

(a)

$$dF(\mathbf{P})=Nd\left\langle\left(\mathbf{I}-\mathbf{P}(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\right)^2,\mathbf{S}\right\rangle=$$

$$=N\left\langle\left(\mathbf{I}-\mathbf{P}(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\right)d\left(\mathbf{I}-\mathbf{P}(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\right),\mathbf{S}\right\rangle+$$
$$+N\left\langle d\left(\mathbf{I}-\mathbf{P}(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\right)\left(\mathbf{I}-\mathbf{P}(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\right),\mathbf{S}\right\rangle=$$

$$=N\left\langle\left(\mathbf{I}-\mathbf{P}\mathbf{P}^T\right)d\left(\mathbf{I}-\mathbf{P}(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\right),\mathbf{S}\right\rangle+$$
$$+N\left\langle d\left(\mathbf{I}-\mathbf{P}(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\right)\left(\mathbf{I}-\mathbf{P}\mathbf{P}^T\right),\mathbf{S}\right\rangle=$$

$$=N\big[2\left\langle d\left(\mathbf{I}-\mathbf{P}(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\right),\mathbf{S}\right\rangle-$$
$$-\left\langle\mathbf{P}\mathbf{P}^Td\left(\mathbf{I}-\mathbf{P}(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\right),\mathbf{S}\right\rangle-$$
$$-\left\langle d\left(\mathbf{I}-\mathbf{P}(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\right)\mathbf{P}\mathbf{P}^T,\mathbf{S}\right\rangle\big]$$

Далее отдельно вычислим следующие производные

$$d((\mathbf{P}^T\mathbf{P})^{-1})=-\left(\mathbf{P}^T\mathbf{P}\right)^{-1}\left(d\mathbf{P}^T\mathbf{P}+\mathbf{P}^Td\mathbf{P}\right)\left(\mathbf{P}^T\mathbf{P}\right)^{-1}=-d\mathbf{P}^T\mathbf{P}-\mathbf{P}^Td\mathbf{P}$$

$$d\left(\mathbf{I}-\mathbf{P}(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\right)=-d\mathbf{P}(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T-\mathbf{P}(\mathbf{P}^T\mathbf{P})^{-1}d\mathbf{P}^T-\mathbf{P}d\left(\mathbf{P}^T\mathbf{P}\right)^{-1}\mathbf{P}^T=$$
$$=-d\mathbf{P}\mathbf{P}^T-\mathbf{P}d\mathbf{P}^T+\mathbf{P}\left(d\mathbf{P}^T\mathbf{P}+\mathbf{P}^Td\mathbf{P}\right)\mathbf{P}^T$$

Тогда

$$
\begin{gathered}
dF(\mathbf{P}) = -N[2\left\langle d\mathbf{P}\mathbf{P}^T + \mathbf{P}d\mathbf{P}^T - \mathbf{P}\left(d\mathbf{P}^T\mathbf{P} + \mathbf{P}^T d\mathbf{P}\right)\mathbf{P}^T, \mathbf{S}\right\rangle - \\
- \left\langle \mathbf{P}\mathbf{P}^T\left(d\mathbf{P}\mathbf{P}^T + \mathbf{P}d\mathbf{P}^T - \mathbf{P}\left(d\mathbf{P}^T\mathbf{P} + \mathbf{P}^T d\mathbf{P}\right)\mathbf{P}^T\right), \mathbf{S}\right\rangle - \\
- \left\langle\left(d\mathbf{P}\mathbf{P}^T + \mathbf{P}d\mathbf{P}^T - \mathbf{P}\left(d\mathbf{P}^T\mathbf{P} + \mathbf{P}^T d\mathbf{P}\right)\mathbf{P}^T\right)\mathbf{P}\mathbf{P}^T, \mathbf{S}\right\rangle]
\end{gathered}
$$

Раскрывая скобки и учитывая $\mathbf{P}^T\mathbf{P} = \mathbf{I}$

$$
\begin{aligned}
dF(\mathbf{P}) = -N[2&\left\langle d\mathbf{P}\mathbf{P}^T + \mathbf{P}d\mathbf{P}^T - \mathbf{P}d\mathbf{P}^T\mathbf{P}\mathbf{P}^T - \mathbf{P}\mathbf{P}^T d\mathbf{P}\mathbf{P}^T, \mathbf{S}\right\rangle - \\
&- \left\langle \mathbf{P}\mathbf{P}^T d\mathbf{P}\mathbf{P}^T + \mathbf{P}d\mathbf{P}^T - \mathbf{P}d\mathbf{P}^T\mathbf{P}\mathbf{P}^T - \mathbf{P}\mathbf{P}^T d\mathbf{P}\mathbf{P}^T, \mathbf{S}\right\rangle - \\
&- \left\langle d\mathbf{P}\mathbf{P}^T + \mathbf{P}d\mathbf{P}^T\mathbf{P}\mathbf{P}^T - \mathbf{P}d\mathbf{P}^T\mathbf{P}\mathbf{P}^T - \mathbf{P}\mathbf{P}^T d\mathbf{P}\mathbf{P}^T, \mathbf{S}\right\rangle] =
\end{aligned}
$$

$$
\begin{aligned}
= -Ntr\big(\mathbf{S}^T\big[&2d\mathbf{P}\mathbf{P}^T + 2\mathbf{P}d\mathbf{P}^T - 2\mathbf{P}d\mathbf{P}^T\mathbf{P}\mathbf{P}^T - 2\mathbf{P}\mathbf{P}^T d\mathbf{P}\mathbf{P}^T - \\
&-\mathbf{P}\mathbf{P}^T d\mathbf{P}\mathbf{P}^T - \mathbf{P}d\mathbf{P}^T + \mathbf{P}d\mathbf{P}^T\mathbf{P}\mathbf{P}^T + \mathbf{P}\mathbf{P}^T d\mathbf{P}\mathbf{P}^T - \\
&-d\mathbf{P}\mathbf{P}^T - \mathbf{P}d\mathbf{P}^T\mathbf{P}\mathbf{P}^T + \mathbf{P}d\mathbf{P}^T\mathbf{P}\mathbf{P}^T + \mathbf{P}\mathbf{P}^T d\mathbf{P}\mathbf{P}^T\big]\big) =
\end{aligned}
$$

$$
\begin{gathered}
= -Ntr\big(\mathbf{S}^T\big[d\mathbf{P}\mathbf{P}^T + \mathbf{P}d\mathbf{P}^T - \mathbf{P}\mathbf{P}^T d\mathbf{P}\mathbf{P}^T - \mathbf{P}d\mathbf{P}^T\mathbf{P}\mathbf{P}^T\big]\big) = \\
= -Ntr\big(\mathbf{S}^T\big((\mathbf{I} - \mathbf{P}\mathbf{P}^T)d\mathbf{P}\mathbf{P}^T + \mathbf{P}d\mathbf{P}^T(\mathbf{I} - \mathbf{P}\mathbf{P}^T)\big)\big)
\end{gathered}
$$

Воспользовавшись свойствами следа:

1. $tr(ABC) = tr(CAB)$
2. $tr(A + B) = tr(A) + tr(B)$
3. $tr(A^T) = tr(A)$

Получим

$$
dF(\mathbf{P}) = Ntr\big(\mathbf{S}^T(\mathbf{P}\mathbf{P}^T - \mathbf{I})d\mathbf{P}\mathbf{P}^T\big) + Ntr\big(\mathbf{S}^T\mathbf{P}d\mathbf{P}^T(\mathbf{P}\mathbf{P}^T - \mathbf{I})\big)
$$

Отдельно для каждого слагаемого, учитывая что $(\mathbf{P}\mathbf{P}^T - \mathbf{I})$ - симметрическая

$$
tr\big(\mathbf{S}^T(\mathbf{P}\mathbf{P}^T - \mathbf{I})d\mathbf{P}\mathbf{P}^T\big) = tr\big(d\mathbf{P}\mathbf{P}^T\mathbf{S}^T(\mathbf{P}\mathbf{P}^T - \mathbf{I})\big) = N\left\langle(\mathbf{P}\mathbf{P}^T - \mathbf{I})\mathbf{S}\mathbf{P}, d\mathbf{P}\right\rangle
$$

$$
Ntr\big(\mathbf{S}^T\mathbf{P}d\mathbf{P}^T(\mathbf{P}\mathbf{P}^T - \mathbf{I})\big) = Ntr\big((\mathbf{P}\mathbf{P}^T - \mathbf{I})\mathbf{S}^T\mathbf{P}d\mathbf{P}^T\big) = N\left\langle(\mathbf{P}\mathbf{P}^T - \mathbf{I})\mathbf{S}\mathbf{P}, d\mathbf{P}\right\rangle
$$

И наконец, производная функции

$$
dF(\mathbf{P}) = 2N\left\langle(\mathbf{P}\mathbf{P}^T - \mathbf{I})\mathbf{S}\mathbf{P}, d\mathbf{P}\right\rangle
$$

$$
\nabla_{\mathbf{P}}F = 2N(\mathbf{P}\mathbf{P}^T - \mathbf{I})\mathbf{S}\mathbf{P}
$$

(b)

Пусть $\mathbf{P}$, состоит из с.в $\mathbf{q_{i_k}}, k = \overline{1, d}$ матрицы $\mathbf{S}$ и $\mathbf{S} = \mathbf{Q\Lambda Q}^T$, тогда

$$\mathbf{SP} = [\lambda_{i_1}\mathbf{q_{i_1}}, \dots, \lambda_{i_d}\mathbf{q_{i_d}}], \text{ где } \lambda_{i_k}\text{- с.з., соответствующее с.в.}\mathbf{q_{i_k}}$$

Тогда $\mathbf{SP} = \mathbf{P}\widetilde{\mathbf{\Lambda}}$, где

$$\widetilde{\mathbf{\Lambda}} = \begin{pmatrix} \lambda_{i_1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \lambda_{i_d} \end{pmatrix}$$

Раскрывая скобки и применяя полученный результат

$$\nabla_{\mathbf{P}} F = 2N(\mathbf{PP}^T - \mathbf{I})\mathbf{SP} = 2N\left(\mathbf{PP}^T\mathbf{SP} - \mathbf{SP}\right) = 2N\left(\mathbf{PP}^T\mathbf{P}\widetilde{\mathbf{\Lambda}} - \mathbf{P}\widetilde{\mathbf{\Lambda}}\right)$$

Принимая во внимание ортогональность столбцов $\mathbf{P}$

$$\nabla_{\mathbf{P}} F = 2N\left(\mathbf{P}\widetilde{\mathbf{\Lambda}} - \mathbf{P}\widetilde{\mathbf{\Lambda}}\right) = \mathbf{0}$$

ч.т.д

Рассмотрим $\mathbf{P}$, состоящие из с.в $\mathbf{q_{i_k}}, k = \overline{1, d}$ матрицы $\mathbf{S}$, $\widetilde{\mathbf{\Lambda}}$ такая же как выше, тогда $\mathbf{P}^T\mathbf{P} = \mathbf{I}$ и

$$F(\mathbf{P}) = Ntr\left(\left(\mathbf{I} - \mathbf{P}(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\right)^2\mathbf{S}\right) = Ntr\left(\left(\mathbf{I} - \mathbf{PP}^T\right)^2\mathbf{S}\right) =$$
$$Ntr\left((\mathbf{I} - \mathbf{PP}^T)\mathbf{S}\right) = Ntr\left(\mathbf{S} - \mathbf{SPP}^T\right) = Ntr\left(\mathbf{S} - \mathbf{P}\widetilde{\mathbf{\Lambda}}\mathbf{P}^T\right)$$

Так как след - это инвариант подобия, то домножим выражение в скобках на Q^T слева и Q - справа

$$F(\mathbf{P}) = Ntr\left(\mathbf{\Lambda} - \mathbf{Q^T}\mathbf{P}\widetilde{\mathbf{\Lambda}}\mathbf{P}^T\mathbf{Q}\right) = N\left[tr(\mathbf{\Lambda}) - tr(\mathbf{Q^T}\mathbf{P}\widetilde{\mathbf{\Lambda}}\mathbf{P}^T\mathbf{Q})\right] =$$
$$= N\left[tr(\mathbf{\Lambda}) - tr(\widetilde{\mathbf{\Lambda}})\right] = N\left[\sum_{i=1}^{D}\lambda_i - \sum_{k=1}^{d}\lambda_{i_k}\right]$$

Такое выражение, очевидно, минимально, когда λ_{i_k} - максимальные с.з. матрицы $\mathbf{S}$, вспоминая как введена $\widetilde{\mathbf{\Lambda}}$ получим, что для рассмотренного класса матриц минимум $F(\mathbf{P})$ действительно достигается, когда матрица $\mathbf{P}$ состоит из с.в., отвечающих наибольшим с.з. $\mathbf{S}$.