Домашнее задание. Матричное дифференцирование

Владимиров Эдуард, группа Б05-928

8 сентября 2022 г.

Задача 1. Найти градиент $\nabla_X f$ для функции $f(X) = \det(X^{-1} + A)$,

<u>*Решение.*</u>

$$df(X) = \det(X^{-1} + A) \cdot \operatorname{tr}\Big[(X^{-1} + A)^{-1} \cdot d(X^{-1} + A)\Big] =$$

$$= \det(X^{-1} + A) \cdot \operatorname{tr}\Big[-(X^{-1} + A)^{-1} X^{-1} dX \cdot X^{-1}\Big] =$$

$$= -\det(X^{-1} + A) \cdot \Big\langle X^{-T}(X^{-1} + A)^{-T}, dX \cdot X^{-1}\Big\rangle =$$

$$= -\det(X^{-1} + A) \cdot \Big\langle X^{-T}(X^{-1} + A)^{-T} X^{-T}, dX\Big\rangle =$$

Ответ. $\nabla_X f = -\det(X^{-1} + A) \cdot X^{-T}(X^{-1} + A)^{-T} X^{-T}$

Задача 2. Для каждой из следующих функций найти градиент ∇f и гессиан $\nabla^2 f$:

(a) $f(t) = \|(A + tI_n)^{-1} b\|, \quad A \in \mathbb{S}^n_+$

(b) $f(t) = \frac{1}{2}\left\langle xx^T - A,\, xx^T - A\right\rangle, \quad A \in \mathbb{S}^n$

<u>*Решение.*</u> (a) Обозначим $Z := (A + tI_n)^{-1} b$

$$df(t) = \frac{1}{2\|Z\|} \cdot d\langle Z, Z\rangle = \frac{2\langle Z, dZ\rangle}{2\|Z\|}$$

$$dZ = -(A + tI_n)^{-1} d(A + tI_n)(A + tI_n)^{-1} b = -(A + tI_n)^{-2} b\, dt$$

$$df(t) = \frac{\langle Z, -(A + tI_n)^{-2} b\, dt\rangle}{\|Z\|}$$

(b)

$$0)\ (x+\varepsilon e_i)(x+\varepsilon e_i)^T = xx^T + \begin{pmatrix} 0 & \dots & \varepsilon x_1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \varepsilon x_1 & \dots & \varepsilon^2 + 2\varepsilon x_i & \dots & \varepsilon x_n \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \varepsilon x_n & \dots & 0 \end{pmatrix} =: xx^T + X_i$$

$$1)\ f(x+\varepsilon e_i) - f(x) = \frac{1}{2}\big\langle xx^T + X_i - A, xx^T + X_i - A\big\rangle - \frac{1}{2}\big\langle xx^T - A, xx^T - A\big\rangle =$$

$$= \frac{1}{2}\big\langle X_i,\ X_i\big\rangle + \big\langle xx^T - A,\ X_i\big\rangle$$

$$1.1)\ \frac{1}{2}\big\langle X_i, X_i\big\rangle = \frac{1}{2}\boldsymbol{\varepsilon^2}\big(2x_1^2 + \ldots + 2x_{i-1}^2 + (\varepsilon+2)^2 x_i^2 + 2x_{i+1}^2 + \ldots + 2x_n^2\big)$$

$$1.2)\ xx^T - A = \begin{pmatrix} x_1^2 - a_{11} & \dots & x_1 x_n - a_{1n} \\ \vdots & \ddots & \vdots \\ x_n x_1 - a_{n1} & \dots & x_n^2 - a_{nn} \end{pmatrix}$$

$$\big\langle xx^T - A,\ X_i\big\rangle = \sum_{j=1, j\neq i}^{n} (x_i x_j - a_{ij}) \cdot \varepsilon x_j + \sum_{j=1, j\neq i}^{n} (x_j x_i - a_{ji}) \cdot \varepsilon x_j +$$

$$+ (x_i^2 - a_{ii})(\varepsilon^2 + 2\varepsilon x_i)$$

$$2)\ \frac{\partial f}{\partial x_i}(x) = \lim_{\varepsilon\to 0} \frac{f(x+\varepsilon e_i) - f(x)}{\varepsilon} = \lim_{\varepsilon\to 0}\left[\frac{\langle X_i,\ X_i\rangle}{2\varepsilon} + \frac{\langle xx^T - A,\ X_i\rangle}{\varepsilon}\right]$$

$$\frac{\partial f}{\partial x_i}(x) = \lim_{\varepsilon\to 0} \frac{\langle xx^T - A,\ X_i\rangle}{\varepsilon} = \sum_{j=1, j\neq i}^{n} (2x_i x_j^2 - x_j(a_{ij} + a_{ji})) + 2x_i(x_i^2 - a_{ii}) =$$

$$= \Big| A \text{— симм.} \Big| = \sum_{j=1}^{n} 2x_j(x_i x_j - a_{ij})$$

Ответ. (a) $\nabla f(t) = \dfrac{1}{\|(A + tI_n)^{-1} b\|} b^T \left(A + tI_n\right)^{-3} b$

(b) $\nabla f(x) = 2\,(xx^T - A)x$

Задача 3. Задача состоит в вычислении градиентов для прохода назад по этому графу, т.е. при известном $\nabla_x L$ найти $\nabla_b L$ и $\nabla_A L$.

<u>*Решение.*</u>

$$Ax = b \Rightarrow x = A^{-1}b \Rightarrow dx = A^{-1}db + dA^{-1} \cdot b$$

$$dL = \nabla_x L^T dx = \nabla_x L^T (A^{-1}db + dA^{-1} \cdot b)$$

$$dL = -\nabla_x L^T A^{-1} dA \cdot A^{-1} b + \nabla_x L^T A^{-1} db$$

$$dL = -\nabla_x L^T A^{-1} dA \cdot x + \nabla_x L^T A^{-1} db$$

$$dL = \operatorname{tr}(-x \nabla_x L^T A^{-1} dA) + \nabla_x L^T A^{-1} db$$

Ответ.

$$\nabla_A L = -A^{-T} \nabla_x L x^T$$

$$\nabla_b L = A^{-T} \nabla_x L$$

Задача 4. Для каждой из следующих функций найти все точки стационарности и указать значения параметров, при которых они существуют:

(a) $f(x) = \langle a, x \rangle - \ln(1 - \langle b, x \rangle)$

(b) $f(x) = \langle c, x \rangle \exp(-\langle Ax, x \rangle), \quad A \in \mathbb{S}^n_{++}$

(c) $f(X) = \langle X^{-1}, I_n \rangle - \langle A, X \rangle, \quad A \in \mathbb{S}^n$

<u>*Решение.*</u> (a)

$$df(x) = \big\langle a,\, dx \big\rangle + \frac{\big\langle b,\, dx \big\rangle}{1 - \big\langle b,\, x \big\rangle} = \Big\langle a + \frac{b}{1 - \big\langle b,\, x \big\rangle},\, dx \Big\rangle$$

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

Стац. точки: $\dfrac{b}{1 - \big\langle b,\, x \big\rangle} = a.$

Если a и b неколлинеарны или противоположно направлены, то таких точек нет ($1 - \big\langle b,\, x \big\rangle > 0$).

Иначе $1 - \big\langle b,\, x \big\rangle = \dfrac{b}{a} \Leftrightarrow \big\langle b,\, x \big\rangle = 1 - \dfrac{b}{a}$.

Таким образом, множество стационарных точек образует $(n-1)$-мерную гиперплоскость.

(b)

$$df(x) = \big(-2\big\langle Ax,\, dx\big\rangle\big\langle c,\, x\big\rangle + \big\langle c,\, dx\big\rangle\big)\exp(-\big\langle Ax,\, x\big\rangle) =$$
$$= \left\langle e^{-\left\langle Ax, x\right\rangle}(c - 2Ax\big\langle c,\, x\big\rangle),\, dx\right\rangle$$

$\nabla f(x) = e^{-\left\langle Ax, x\right\rangle}(c - 2Ax\big\langle c,\, x\big\rangle)$

$\nabla f(x) = 0 :\ 2Ax\big\langle c,\, x\big\rangle = c$

Ищем решение вида $Ax = \lambda c \Leftrightarrow x = \lambda A^{-1}c :$

$$2\lambda c\big\langle c,\, \lambda A^{-1}c\big\rangle = c$$
$$2\lambda^2\big\langle c,\, A^{-1}c\big\rangle = 1$$
$$\lambda^2 = \frac{1}{2\big\langle c,\, A^{-1}c\big\rangle} > 0, \text{так как} A \succ 0 \Rightarrow A^{-1} \succ 0$$
$$\lambda = \pm\frac{1}{\sqrt{2\big\langle c,\, A^{-1}c\big\rangle}}$$
$$x = \pm\frac{A^{-1}c}{\sqrt{2\big\langle c,\, A^{-1}c\big\rangle}}$$

(c)

$$df(X) = \big\langle -X^{-1}dX{\cdot}X^{-1},\, I_n\big\rangle - \big\langle A,\, dX\big\rangle = -\big\langle dX,\, X^{-2}\big\rangle - \big\langle A,\, dX\big\rangle =$$

$df(X) = -\big\langle X^{-2} + A,\, dX\big\rangle$

$\nabla f(X) = -X^{-2} - A$

$\nabla f(X) = 0 : X^{-2} = -A$

Для существования решения необходимо и достаточно, чтобы матрица A была отрицательно определённой ($\Rightarrow$ невырожденной)

$$X^{-1} = \sqrt{-A}$$
$$X = \sqrt{-A}^{-1}$$

Ответ. (a) $\left\{x : \big\langle b,\, x\big\rangle = 1 - \frac{b}{a}\right\}$

Условия: сонаправленность векторов a и b.

(b) $\left\{ \pm \dfrac{A^{-1}c}{\sqrt{2\langle c,\, A^{-1}c\rangle}} \right\}$

Условия: отсутствуют.

(c) $X = \sqrt{-A}^{-1}$

Условия: $A \prec 0$.

Задача 5. $X \in \mathbb{S}^n,\ X = Q^T \Lambda Q,\ |\sigma(X)| = n$

f — некоторая скалярная функция потерь, которая каким-то образом зависит от собственных векторов и собственных значений матрицы X. Требуется по известным $\nabla_\Lambda f$ (диагональная матрица) и $\nabla_Q f$ найти $\nabla_X f$.

Решение. $dX = dQ^T \cdot \Lambda \cdot dQ + Q^T \cdot d\Lambda \cdot Q + Q^T \cdot \Lambda \cdot dQ$

$$df = \left\langle \nabla_X f,\, dX \right\rangle = \left\langle dQ,\, \Lambda \cdot Q \cdot \nabla_X f^T \right\rangle + \left\langle Q \cdot \nabla_X f \cdot Q^T,\, d\Lambda \right\rangle + \left\langle \Lambda \cdot Q \cdot \nabla_X f,\, dQ \right\rangle =$$

$$= \left\langle \Lambda \cdot Q \cdot (\nabla_X f + \nabla_X f^T),\, dQ \right\rangle + \left\langle Q \cdot \nabla_X f \cdot Q^T,\, d\Lambda \right\rangle =$$

$$\left\langle \nabla_Q f,\, dQ \right\rangle + \left\langle \nabla_\Lambda f,\, d\Lambda \right\rangle$$

Из последнего равенства получаем, что $\nabla_X f = Q^T \cdot \nabla_\Lambda f \cdot Q = \frac{1}{2}\, Q^T \cdot \Lambda \cdot \nabla_Q f$

Ответ. $\nabla_X f = Q^T \cdot \nabla_\Lambda f \cdot Q$