Домашняя работа 1
Богданов Александр Б05-003

1 Задача

Дано:

$$f(X) = \det(X^{-1} + A)$$

Известно:

$$d(\det X) = \langle \det(X)X^{-T}, dX \rangle$$
$$d(X^{-1}) = -X^{-1}d(X)X^{-1}$$

Решение:

$$df(X) = \langle \det(X^{-1}+A)(X^{-1}+A)^{-T}, d(X^{-1}+A) \rangle = -\det(X^{-1}+A)\langle (X^{-1}+A)^{-T}, X^{-1}d(X)X^{-1} \rangle =$$
$$= -\det(X^{-1}+A)\langle X^{-T}(X^{-1}+A)^{-T}X^{-T}, dX \rangle = -\det(X^{-1}+A)\langle (X(X^{-1}+A)X)^{-T}, dX \rangle =$$
$$= -\det(X^{-1}+A)\langle (X+XAX)^{-T}, dX \rangle$$

Тогда:

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

Ответ:

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

2 Задача

2.1 Пункт а)

Дано:

$$f(t) = \|(A+tI_n)^{-1}b\|_2$$

Известно:

$$d(\|x\|_2) = \frac{\langle x, dx \rangle}{\|x\|_2}$$
$$d(X^{-1}) = -X^{-1}d(X)X^{-1}$$

Решение (первая часть):

$$df(t) = \frac{\langle (A+tI_n)^{-1}b, d((A+tI_n)^{-1}b) \rangle}{\|(A+tI_n)^{-1}b\|_2} = -\frac{\langle (A+tI_n)^{-1}b, (A+tI_n)^{-1}dt(A+tI_n)^{-1}b \rangle}{\|(A+tI_n)^{-1}b\|_2} =$$
$$= -\frac{b^T(A+tI_n)^{-T}(A+tI_n)^{-2}b}{\|(A+tI_n)^{-1}b\|_2}dt = -\frac{b^T(A+tI_n)^{-3}b}{\|(A+tI_n)^{-1}b\|_2}dt$$

Тогда:

$$\nabla f = -\frac{b^T(A+tI_n)^{-3}b}{\|(A+tI_n)^{-1}b\|_2}$$

Решение (вторая часть):

$$d^2f(t) = -\frac{\|(A+tI_n)^{-1}b\|_2 d(b^T(A+tI_n)^{-3}b) - d(\|(A+tI_n)^{-1}b\|_2)(b^T(A+tI_n)^{-3}b)}{\|(A+tI_n)^{-1}b\|_2^2}dt =$$

$$= -\frac{b^T d(A+tI_n)^{-3}b}{\|(A+tI_n)^{-1}b\|_2}dt + \frac{d(\|(A+tI_n)^{-1}b\|_2)(b^T(A+tI_n)^{-3}b)}{\|(A+tI_n)^{-1}b\|_2^2}dt =$$

$$= 3\frac{b^T(A+tI_n)^{-4}b}{\|(A+tI_n)^{-1}b\|_2}(dt)^2 - \frac{(b^T(A+tI_n)^{-3}b)^2}{\|(A+tI_n)^{-1}b\|_2^3}(dt)^2$$

Тогда:

$$\nabla^2 f = -\frac{(b^T(A+tI_n)^{-3}b)^2}{\|(A+tI_n)^{-1}b\|_2^3} + 3\frac{b^T(A+tI_n)^{-4}b}{\|(A+tI_n)^{-1}b\|_2}$$

Ответ:

$$\nabla f = -\frac{b^T(A+tI_n)^{-3}b}{\|(A+tI_n)^{-1}b\|_2}$$

$$\nabla^2 f = -\frac{(b^T(A+tI_n)^{-3}b)^2}{\|(A+tI_n)^{-1}b\|_2^3} + 3\frac{b^T(A+tI_n)^{-4}b}{\|(A+tI_n)^{-1}b\|_2}$$

2.2 Пункт b)

Дано:

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

Известно:

$$d\langle Ax, x\rangle = \langle (A+A^T)x, dx\rangle$$

Решение (первая часть):

$$df(t) = \frac{1}{2}d\langle xx^T - A, xx^T - A\rangle = \langle xx^T - A, d(xx^T - A)\rangle = \langle xx^T - A, d(x)x^T + xd(x)^T\rangle =$$

$$= \langle (xx^T - A)x, dx\rangle + \langle (xx^T - A)x, dx\rangle = \langle 2(xx^T - A)x, dx\rangle$$

Тогда:

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

Решение (вторая часть):

$$d^2 f(t) = \langle d(2(xx^T - A)x), dx_1\rangle = \langle 2(d(xx^Tx) - Adx_2), dx_1\rangle = \langle 2(dx_2(x^Tx) + xd\langle x, x\rangle - Adx_2), dx_1\rangle =$$

$$= \langle 2(dx_2(x^Tx) + 2xx^Tdx_2 - Adx_2), dx_1\rangle = \langle 2x^TxI + 4xx^T - 2A)dx_2, dx_1\rangle = \langle (2x^TxI + 4xx^T - 2A)dx_1, dx_2\rangle$$

Тогда:

$$\nabla^2 f = 2x^TxI + 4xx^T - 2A$$

Ответ:

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

$$\nabla^2 f = 2x^TxI + 4xx^T - 2A$$

3 Задача

Воспользуемся алгоритмом backpropagation:

$$dL(x) = \langle \nabla_x L, dx\rangle$$

$$dx(A, b) = d(A^{-1}b) = -A^{-1}dAA^{-1}b + A^{-1}db$$

$$dL(A, b) = -\langle \nabla_x L, A^{-1}dAA^{-1}b\rangle + \langle \nabla_x L, A^{-1}db\rangle = -\langle A^{-T}\nabla_x L b^T A^{-T}, dA\rangle + \langle A^{-T}\nabla_x L, db\rangle$$

Так как в стандартном виде $dL = \langle \nabla_A L, dA \rangle + \langle \nabla_b L, db \rangle$ получаем:

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

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

Ответ:

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

4 Задача

4.1 Пункт a)

Дано:

$$f(x) = \langle a, x \rangle - \ln(1 - \langle b, x \rangle),\ a, b \neq 0,\ E = \{x \in \mathbb{R}^n | \langle x, b \rangle < 1\}$$

Решение:

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

$$\nabla f = a + \frac{b}{1 - \langle b, x \rangle} = 0 \Rightarrow a \text{ и } b \text{ должны быть линейно зависимыми, так как } a, b \neq 0$$

Тогда $b = \lambda a \Rightarrow 1 - \langle b, x \rangle + \lambda = 0 \Rightarrow \langle b, x \rangle = \lambda + 1 \Rightarrow$ для того, чтобы выполнялось условие принадлежности множеству λ должна быть меньше нуля.
Ответ:

1. Если a и b линейно независимы, то точек стационарности нет,
2. Если $b = \lambda a$ и $\lambda \geq 0$, то точек стационарности нет на E,
3. Если $b = \lambda a$ и $\lambda < 0$, то точки стационарности на E это точки гиперплоскости $\langle b, x \rangle = \lambda + 1$.

4.2 Пункт b)

Дано:

$$f(x) = \langle c, x \rangle \exp\left(-\langle Ax, x \rangle\right),\ c \neq 0$$

Известно:

$$d\langle Ax, x \rangle = \langle 2Ax, dx \rangle \text{ для симметричной матрицы}$$

Решение:

$$df(x) = \langle c, dx \rangle \exp\left(-\langle Ax, x \rangle\right) - \langle c, x \rangle \langle 2Ax, dx \rangle \exp\left(-\langle Ax, x \rangle\right) =$$

$$= \langle (c - 2\langle c, x \rangle Ax) \exp\left(-\langle Ax, x \rangle\right), dx \rangle$$

$$\nabla f = (c - 2\langle c, x \rangle Ax) \exp\left(-\langle Ax, x \rangle\right) = 0$$

$$2\langle c, x \rangle Ax = c$$

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

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

$$\langle c, x \rangle = \pm \sqrt{\frac{1}{2} c^T A^{-1} c},\ \text{можно, так как матрица положительно определенная}$$

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

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

Ответ: При любых параметрах ($c \neq 0$) существуют стационарные точки: $x = \pm\frac{A^{-1}c}{\sqrt{2c^T A^{-1}c}}$

4.3 Пункт с)

Дано:

$$f(X) = \langle X^{-1}, I_n\rangle - \langle A, X\rangle$$

Известно:

$$d(X^{-1}) = -X^{-1}d(X)X^{-1}$$

Решение:

$$df(X) = f(X) = -\langle X^{-1}d(X)X^{-1}, I_n\rangle - \langle A, dX\rangle = -\langle dX, X^{-T}X^{-T}\rangle - \langle A, dX\rangle =$$

$$= \langle -X^{-T}X^{-T} - A, dX\rangle$$

$$\nabla f = -(XX)^{-T} - A = 0$$

$$X^{-2} = -A$$

Если A не отрицательно определенная матрица, то точек стационарности нет, так как слева положительно определенная матрица. Если A отрицательно определенная матрица, то получаем:

$$X^2 = -A^{-1}$$

Так как A симметричная, то ее можно представить в виде: $A = Q^T\Lambda Q$, где Q - ортогональная матрица, Λ - диагональная матрица из собственных чисел. Тогда $-A^{-1} = Q^T(-\Lambda^{-1})Q$ и следовательно $X = Q^T\sqrt{-\Lambda^{-1}}Q$ **Ответ:**

1. Если A не отрицательно определенная, то точек стационарности нет,
2. Если A отрицательно определенная, $X = Q^T\sqrt{-\Lambda^{-1}}Q$, где $A = Q^T\Lambda Q$, Q - ортогональная матрица, Λ - диагональная матрица из собственных чисел.

5 Задача

Воспользуемся алгоритмом backpropagation:

$$df(X) = \langle \nabla_X f, dX\rangle$$

$$dX(\Lambda, Q) = d(Q^T\Lambda Q) = (dQ)^T\Lambda Q + Q^T d\Lambda Q + Q^T\Lambda dQ$$

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

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

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

Так как в стандартном виде $dL = \langle \nabla_Q f, dQ\rangle + \langle \nabla_\Lambda f, d\Lambda\rangle$ получаем:

$$\nabla_Q f = \Lambda Q((\nabla_X f)^T + \nabla_X f)$$

$$\nabla_\Lambda f = Q\nabla_X f Q^T$$

$$\nabla_Q f = 2\Lambda Q \nabla_\Lambda f Q$$
$$\nabla_X f = Q^T \nabla_\Lambda f Q$$

Ответ:

$$\nabla_X f = Q^T \nabla_\Lambda f Q$$

При условии: $\nabla_Q f = 2\Lambda Q \nabla_\Lambda f Q$