

**Домашнее задание. Матричные вычисления и матричное дифференцирование.**

Мягкий дедлайн: 5 ноября, 23:59 (за каждый день просрочки снимается 1 балл)

Жёсткий дедлайн: 12 ноября, 23:59

Формат сдачи: pdf-файл, подготовленный с помощью  $\text{\LaTeX}$  или скан рукописных записей (скан должен быть собран в один pdf-файл с порядком страниц, соответствующим задачам в задании).

Обозначения:

- $\langle x, y \rangle$  – Евклидово скалярное произведение;
- $\|x\| = \langle x, x \rangle^{1/2} = (x^T x)^{1/2}$  – Евклидова норма вектора;
- $\|A\|_F = \langle A, A \rangle^{1/2} = \text{tr}(A^T A)^{1/2}$  – матричная норма Фробениуса;
- $I_n$  – единичная матрица размера  $n \times n$ ;
- $\mathbb{R}_+^n = \{x \in \mathbb{R}^n \mid x_i \geq 0 \ \forall i\}$ ,  $\mathbb{R}_{++}^n = \{x \in \mathbb{R}^n \mid x_i > 0 \ \forall i\}$ ;
- $\mathbb{S}^n = \{A \in \mathbb{R}^{n \times n} \mid A = A^T\}$ ;
- $\mathbb{S}_+^n = \{A \in \mathbb{S}^n \mid A \text{ – неотр. определённая}\}$ ,  $\mathbb{S}_{++}^n = \{A \in \mathbb{S}^n \mid A \text{ – полож. определённая}\}$ .

Все задачи (каждый подпункт) оцениваются одинаково из общей суммы в 10 баллов.

1. Докажите тождество Вудбери:

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

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

2. Упростите каждое из следующих выражений:

- $\|uv^T - A\|_F^2 - \|A\|_F^2$ , где  $u \in \mathbb{R}^m$ ,  $v \in \mathbb{R}^n$ ,  $A \in \mathbb{R}^{m \times n}$ .
- $\text{tr}((2I_n + aa^T)^{-1}(uv^T + vu^T))$ , где  $a, u, v \in \mathbb{R}^n$ . Подсказка: воспользуйтесь тождеством Вудбери.
- $\sum_{i=1}^n \langle S^{-1}a_i, a_i \rangle$ , где  $a_1, \dots, a_n \in \mathbb{R}^d$ ,  $S = \sum_{i=1}^n a_i a_i^T$ ,  $\det(S) \neq 0$ .

3. Для каждой из следующих функций найдите первую и вторую производную:

- $f : E \rightarrow \mathbb{R}$ ,  $f(t) = \det(A - tI_n)$ , где  $A \in \mathbb{R}^{n \times n}$ ,  $E = \{t \in \mathbb{R} : \det(A - tI_n) \neq 0\}$ .
- $f : \mathbb{R}_{++} \rightarrow \mathbb{R}$ ,  $f(t) = \|(A + tI_n)^{-1}b\|$ , где  $A \in \mathbb{S}_+^n$ ,  $b \in \mathbb{R}^n$ .

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

- $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f(x) = \frac{1}{2}\|xx^T - A\|_F^2$ , где  $A \in \mathbb{S}^n$ .
- $f : \mathbb{R} \setminus \{0\} \rightarrow \mathbb{R}$ ,  $f(x) = \langle x, x \rangle^{\langle x, x \rangle}$ .
- $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f(x) = \|Ax - b\|^p$ , где  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ ,  $p \geq 2$ .

5. Для каждой из следующих функций  $f$  покажите, что второй дифференциал является знакоопределённым (т.е.  $d^2 f(x)[dx, dx]$  имеет постоянный знак) и установите этот знак:

- $f : \mathbb{S}_{++}^n \rightarrow \mathbb{R}$ ,  $f(X) = \text{tr}(X^{-1})$ .
- $f : \mathbb{S}_{++}^n \rightarrow \mathbb{R}$ ,  $f(X) = (\det(X))^{1/n}$ .

Подсказка: использовать неравенство Коши-Буняковского.

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

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

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

(c)  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $f(x) = \langle c, x \rangle \exp(-\langle Ax, x \rangle)$ , где  $c \in \mathbb{R}^n$ ,  $c \neq 0$ ,  $A \in \mathbb{S}_{++}^n$ .

7. Пусть  $X \in \mathbb{S}_{++}^n$ . Вычислите значение следующего выражения:

$$\lim_{k \rightarrow +\infty} \operatorname{tr}(X^{-k} - (X^k + X^{2k})^{-1}).$$

8. Рассмотрим метод главных компонент. Пусть имеется выборка  $\{x_i\}_{i=1}^N$ ,  $x_i \in \mathbb{R}^D$ , которую мы хотим перевести в выборку меньшей размерности  $d$  с помощью проектирования на линейное пространство, задаваемое матрицей  $P \in \mathbb{R}^{D \times d}$ . Ортогональная проекция вектора  $x$  на это пространство может быть вычислена как  $P(P^T P)^{-1} P^T x$ . Тогда для поиска наилучшей матрицы  $P$  рассмотрим следующую задачу оптимизации:

$$F(P) = \sum_{i=1}^N \|x_i - P(P^T P)^{-1} P^T x_i\|^2 = N \operatorname{tr}((I - P(P^T P)^{-1} P^T)^2 S) \rightarrow \min_P.$$

Здесь  $S = \frac{1}{N} \sum_{i=1}^N x_i x_i^T$  – выборочная матрица ковариации для нормированной выборки.

(a) Найти градиент  $\nabla_P F(P)$ , вычисленный для произвольной матрицы с ортогональными столбцами, т.е.  $P : P^T P = I$  (При вычислении дифференциала  $dF(P)$  нужно сначала действовать так, как если бы матрица  $P$  была произвольной, а потом в полученном выражении пользоваться свойством ортогональности столбцов  $P$ ).

(b) Рассмотрим собственное разложение матрицы  $S$ :  $S = Q \Lambda Q^T$ , где  $\Lambda$  – диагональная матрица с собственными значениями на диагонали,  $Q = [q_1 | q_2 | \dots | q_D] \in \mathbb{R}^{D \times D}$  – ортогональная матрица, состоящая из собственных векторов  $q_i$  по столбцам. Требуется доказать, что градиент  $\nabla_P F(P)$  равен нулю для матрицы  $P$ , состоящей из любых  $d$  различных собственных векторов  $q_i$  по столбцам. Требуется также доказать, что значение минимума  $F(P)$  достигается для матрицы  $P$ , состоящей из собственных векторов  $q_i$ , отвечающих наибольшим собственным значениям матрицы  $S$ .