HW 3

1. **(20 баллов)**. Пусть задана матрица $A \in \mathbb{R}^{m\times n}$, $m \geq n$.

(a) Покажите, что A можно привести к верхнетреугольной матрице R с помощью преобразований Хаусхолдера, используя

$$2mn^2 - \frac{2}{3}n^3 + \mathcal{O}(mn),$$

арифметических операций.

(b) Покажите, что количество арифметических операций для вычисления $Q \in \mathbb{R}^{m\times n}$ из тонкого QR будет:

$$2mn^2 - \frac{2}{3}n^3 + \mathcal{O}(mn).$$

(a) С лекции знаем, что $R = (H_n(H_{n-1}\dots(H_1A)))$

Посчитаем сложность вычисления одного умножения:

$H_1A = (I - 2vv^*)A = A - 2vv^*A$

v^*A: $2mn + O(mn)$ операций

$2vv^*A$: $3mn + O(mn)$ операций

$A - 2vv^*A$: $4mn + O(mn)$ операций

Мы знаем, что H_i выглядит как $\begin{bmatrix} I_{i-1} & 0 \\ 0 & H \end{bmatrix}$, то есть на каждой следующей итерации мы умножаем на матрицу меньшего размера. Получается, что сложность каждого умножения: $4(m-k)(n-k) + O(mn)$, $k \in \{0,\dots,n-1\}$

Сложность всех умножений: $\sum\limits_{k=0}^{n-1}[4(m-k)(n-k) + O(mn)] =$

$= \sum\limits_{k=0}^{n-1} 4(mn - mk - kn + k^2) + O(mn) = 4mn^2 - \sum\limits_{k=0}^{n-1} 4mk - \sum\limits_{k=0}^{n-1} 4kn + \sum\limits_{k=0}^{n-1} 4k^2 + O(mn) =$

$= 4mn^2 - 4m\cdot\frac{n}{2}\cdot n - 4n\cdot\frac{n}{2}\cdot n + 4\cdot\frac{(n-1)\cdot n(2n-1)}{6} + O(mn) = 4mn^2 - 2mn^2 - 2n^3 + \frac{8n^3}{6} + O(mn) =$

$= 2mn^2 - \frac{2}{3}n^3 + O(mn)$

(b) С лекции знаем, что $Q = (H_1(H_2\dots(H_n\begin{bmatrix} I \\ 0 \end{bmatrix})))$

Можем заметить, что мы считали сложность произведения $(H_n\dots(H_2H_1))$.

Сложность нового произведения - это та же сумма, только с слагаемыми в другом порядке. Сложность остается $2mn^2 - \frac{2}{3}n^3 + O(mn)$

2. **(20 баллов)**. Запишем решение x_μ задачи наименьших квадратов с ℓ_2-регуляризацией

$$\|Ax-b\|_2^2+\mu\|x\|_2^2 \to \min_x$$

для заданной матрицы $A\in\mathbb{R}^{m\times n}$ ранга r, вектора правой части $b\in\mathbb{R}^{m\times n}$ и константы $\mu\in\mathbb{R}$ в виде $x_\mu = B(\mu)b$ с матрицей $B(\mu)\in\mathbb{R}^{n\times m}$, которая выражается через A и μ (см. лекцию).

(a) Покажите, что для $\mu>0$ справедливо:

$$\|B(\mu)-A^+\|_2 = \frac{\mu}{(\mu+\sigma_r(A)^2)\,\sigma_r(A)}.$$

$n\times m$ $\quad m\times n$

(b) Покажите, что $B(\mu)\to A^+$ и что $x_\mu \to A^+b$ при $\mu\to +0$.

(a) $x_\mu = (A^TA+\mu I)^{-1}A^Tb = B(\mu)b$

$$\|B(\mu)-A^+\|_2 = \|(A^TA+\mu I)^{-1}A^T - A^+\|_2 = \|(A^TA+\mu I)^{-1}A^T - V\Sigma^+U^T\|_2 =$$

$$= \|(A^TA+\mu I)^{-1}(U\Sigma V^T)^T - V\Sigma^+U^T\|_2 = \|(A^TA+\mu I)^{-1}V\Sigma U^T - V\Sigma^+U^T\|_2 =$$

Типо квадрат. Формально: матрица n×n с квадр. сингл. чисел

$$= \|((V\Sigma^T\underbrace{U^T)(U}_{I}\Sigma V^T)+\mu I)^{-1}V\Sigma U^T - V\Sigma^+U^T\|_2 = \|(V\Sigma^2V^T+\mu I)^{-1}V\Sigma - V\Sigma^+\|_2 =$$

$$= \|(V\Sigma^2V^T+\mu I)^{-1}(V^T)^{-1}\Sigma - V\Sigma^+\|_2 = \|(V^T(V\Sigma^2V^T+\mu I))^{-1}\Sigma - V\Sigma^+\|_2 =$$

$$= \|(\Sigma^2V^T+\mu V^T)^{-1}\Sigma - V\Sigma^+\|_2 = \|((\Sigma^2+\mu I)V^T)^{-1}\Sigma - V\Sigma^+\|_2 = \|V(\Sigma^2+\mu I)^{-1}\Sigma - V\Sigma^+\|_2 =$$

$$= \|(\Sigma^2+\mu I)^{-1}\Sigma - \Sigma^+\|_2$$

$$\Vert$$

$$\left\| \begin{matrix} \frac{\sigma_1}{\sigma_1^2+\mu}-\frac{1}{\sigma_1} & & & & \\ & \ddots & & & \\ & & \frac{\sigma_r}{\sigma_r^2+\mu}-\frac{1}{\sigma_r} & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{matrix} \right\|_2 = \left\| \begin{matrix} \frac{-\mu}{(\sigma_1^2+\mu)\sigma_1} & & & & \\ & \ddots & & & \\ & & \frac{-\mu}{(\sigma_r^2+\mu)\sigma_r} & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{matrix} \right\|_2 =$$

Это же и есть Σ (не обрезанная) из SVD

$$= \left\| \begin{matrix} \frac{\mu}{(\sigma_1^2+\mu)\sigma_1} & & & & \\ & \ddots & & & \\ & & \frac{\mu}{(\sigma_r^2+\mu)\sigma_r} & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{matrix} \right\|_2 = \left[\text{Вторая норма - это старшее сингл. число}\right] = \frac{\mu}{(\sigma_r^2-\mu)\sigma_r}$$

(b) $\|B(\mu)-A^+\|_2 = \dfrac{\mu}{(\sigma_r^2-\mu)\sigma_r} \xrightarrow{\mu\to+0} \dfrac{0}{const} = 0 \Rightarrow B(\mu)\to A^+$

$$\|A^+b - x_\mu\|_F = \|A^+b - B(\mu)b\|_F = \|(A^+-B(\mu))b\|_F \le \overset{\to 0}{\|A^+-B(\mu)\|_F}\,\overset{const}{\|b\|_2} \to 0 \Rightarrow x_\mu \to A^+b$$

3. **(15 баллов)**. Покажите, что для решений $x \in \mathbb{R}^n$ задачи $\|Ax-b\| \to \min_x$, где $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ заданы, справедливо:

$$\|x\|_2^2 = \|A^+b\|_2^2 + \|(I - A^+A)y\|_2^2,$$

где y – произвольный вектор (см. обозначения в лекции). Сделайте отсюда вывод, какое решение имеет наименьшую $\|x\|_2$.

С лекции знаем, что $x = A^+b + (I - A^+A)y$

$\|x\|_2^2 = \langle x, x\rangle = \langle A^+b + (I-A^+A)y, A^+b + (I-A^+A)y\rangle = \langle A^+b, A^+b\rangle + 2\langle A^+b, (I-A^+A)y\rangle +$
$+ \langle (I-A^+A)y, (I-A^+A)y\rangle = \|A^+b\|_2^2 + 2\underbrace{\langle A^+b, (I-A^+A)y\rangle} + \|(I-A^+A)y\|_2^2 \circledcirc$

Хотим, чтобы $=0$, т.е. $A^+b \perp (I-A^+A)y$

Если $z \in (I-A^+A)y$, то $z \in Ker(A)$ или же $z \in Im(V_r^\perp)$ из SVD

Если $w \in A^+b$, то $w \in Im(A^*)$, т.е. $w \in Im(V_r)$ ортогональны

Значит $\forall z \in (I-A^+A)y$, $\forall w \in A^+b$ $z \perp w$, т.е. подпространства ортогональны, получаем, что $A^+b \perp (I-A^+A)y$.

$\circledcirc \|A^+b\|_2^2 + \|(I-A^+A)y\|_2^2$

Наименьшую норму x будет иметь решение с $y = \vec{0}$, т.к. $\|(I-A^+A)y\|_2^2$ неотрицательна $\Rightarrow$ минимум в $\vec{0}$.

4. **(25 баллов)**. Пусть ненулевые $a, b \in \mathbb{R}^n$, $n \geq 2$ ортогональны друг другу и

$$A = a \circ a \circ a + 2(a \circ b \circ a) + 3(b \circ b \circ a).$$

(a) Запишите матрицы $U, V, W \in \mathbb{R}^{n\times 2}$ из канонического разложения A.
Подсказка: используйте линейность тензорного произведения по каждому из аргументов.

(b) Запишите ядро $G \in \mathbb{R}^{2\times 2\times 1}$ и факторы U, V, W из разложения Таккера A.

(c) Докажите, что мультилинейный ранг тензора A равен $(2, 2, 1)$.

(a) $A = a \circ a \circ a + (2a) \circ b \circ a + (3b) \circ b \circ a = a \circ a \circ a + (2a + 3b) \circ b \circ a$

$U = [a \;\; 2a+3b]$

$V = [a \;\; b]$

$W = [a \;\; a]$

(b) $U = [b \ a]$ $\quad g_{221} = 1$

$V = [b \ a]$ $\quad g_{211} = 2$ $\quad G = \begin{bmatrix} 3 & 0 \\ 2 & 1 \end{bmatrix}$

$W = [a]$ $\quad g_{111} = 3$

ядро

(c) $rank = (r_1, r_2, r_3)$

$r_1 = rank(A_{(1)}) = rank(U G_{(1)} (W \otimes V)^T)$

$r_2 = rank(A_{(2)}) = rank(V G_{(2)} (W \otimes U)^T)$

$r_3 = rank(A_{(3)}) = rank(W G_{(3)} (V \otimes U)^T)$

Рассмотрим $W G_{(3)} (V \otimes U)^T$

$rk = 1 \Rightarrow rank\, A_{(3)} = 1 = r_3$

Рассмотрим $V G_{(2)} (W \otimes U)^T$. Так как $rk(V) = 2$ (a и b лин.нез.), то $rk(V G_{(2)} (W \otimes U)^T) \leq 2$. Докажем, что он ровно 2.

$rk\, G_{(2)} = 2$

$(W \otimes U)^T = W^T \otimes U^T = [a^T] \otimes \begin{bmatrix} b^T \\ a^T \end{bmatrix} = \begin{bmatrix} a^T \otimes b^T \\ a^T \otimes a^T \end{bmatrix}$

Если $rk\, (W \otimes U)^T < 2$, то $\exists \alpha, \beta \neq 0 : (\alpha b^T + \beta a^T) \otimes a^T = 0$. Но по условию a и b ортогональны $\Rightarrow$ лин. нез-мы, т.е. $\alpha, \beta = 0$. Значит $rk(W \otimes U^T) = 2$.

Так как $rk\, G_{(2)} = 2$ и $rk(W \otimes U^T) = 2$, можем сделать вывод, что $rk(V G_{(2)} (W \otimes U)^T) = 2 = rank(A_{(2)}) = r_2$. Аналогично можем сказать про $rk(A_{(1)})$.

Итого, $(2, 2, 1)$ - минимальный ранг.

5. **(20 баллов)**. Пусть $A, B \in \mathbb{R}^{n \times n}$ – некоторые заданные матрицы, и пусть стоит задача вычисления матрично-векторного произведения:

$$y = (A \otimes B)\, x, \quad x \in \mathbb{R}^{n^2}.$$

(a) Каково асимптотическое число арифметических операций для вычисления y по x без учета дополнительной структуры матрицы $(A \otimes B)$?

(b) Предложите алгоритм вычисления y, имеющий асимптотическое число операций $\mathcal{O}(n^3)$. **Подсказка:** в этой задаче может помочь операция векторизации.

(c) Как получить число операций $\mathcal{O}(n^2 \log n)$, если A и B являются циркулянтами?

(a) Посчитаем сложность вычисления $A \otimes B$. i,j-й элемент - это $a_{ij} \cdot B$. То есть каждый элемент мы вычисляем за $O(n^2)$. Всего таких элементов n^2.
i-ю координату y вычисляем за $O(n^2)$, всего координат n^2.
Значит итоговая сложность: $O(n^4)$

(b) $(A \otimes B)x = (A \otimes B)\,\underset{\substack{\uparrow \\ \text{сделаем } \mathbb{R}^{n\times n} \text{ матрицу} \\ \text{из } \mathbb{R}^{n^2} \text{ вектора}}}{\mathrm{vec}(X)} = \mathrm{vec}(BXA^T) \Rightarrow$ сложность $O(n^3)$

(c) Также посчитаем BXA^T, но B и A^T уже циркулянты.

$$A = F_n^{-1}\,\mathrm{diag}(F_n \cdot a) \cdot F_n$$
$$B = F_n^{-1}\,\mathrm{diag}(F_n \cdot b) \cdot F_n$$

новый $X = BX$ - $n \cdot n\log n = O(n^2 \log n)$

итог $=$ новый $X \cdot A^T = (\underbrace{A \cdot (\text{новый } X)^T}_{\substack{\text{умн. на циркул.} \\ \text{за } n^2 \log n}})^T \Rightarrow$ Итого: $O(n^2 \log n)$