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) Отражение Хаусхолдера: $H(V) = I - 2VV^*$

Примерим преобразования Хаусхолдера

$A = \begin{pmatrix} Q_1 & Q_2 \end{pmatrix} \begin{pmatrix} R_1 \\ 0 \end{pmatrix} = Q_1 R_1$

С лекции знаем, что $H_n \cdot H_{n-1} \cdot \cdot \ldots \cdot H_2 \cdot H_1 A = R$

Мы хотим посчитать эту сложность:

$H_i \in Mat_{m\times m}$ $\quad v \in \mathbb{C}^{m\times 1}$

Подсчет векторов v из ф-ли $H(V) = I - 2vv^*$ займет

$4 \cdot m^2$ (выведен на семе)

$\Rightarrow$ нам нужно посчитать только сложность $H_{n-1} \cdot H_{n-2} \cdot \ldots \cdot H_2 \cdot H_1 A = R$

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

Умножение $I \cdot A$ займет 0 операций

Умножение $(2v)(\underbrace{v^* A}_{1\times m})$ займет $\underset{\text{вектор на 2}}{m} + \underset{\text{умножить } v^* \text{ на } A}{(2m-1)n} + \underset{\text{умножить } 2v \text{ на } v^*A}{mn} = 3mn - n + m$

$2V \cdot V^*$

Вычислить $A - 2vv^*A$ займет mn операций

Получаем, что умножить одну матрицу Хаусхолдера на A займет $3mn+m-n+mn=4mn+m-n =$
$=4mn+O(m)$

Дальше матрицы Хаусхолдера мы будем применять уже в виде $\begin{pmatrix} I & \\ & H_{(2)} \end{pmatrix}$ и т.д.

т.е. итоговая сложность:

$$4mn + 4(m-1)(n-1) + \dots + 4 = 4\sum_{k=0}^{n-1}(m-k)(n-k) = 4\sum_{k=0}^{n-1} mn - km - kn + k^2 =$$

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

b) $A = (Q_1\ Q_2)\begin{pmatrix} R_1 \\ 0 \end{pmatrix} = Q_1 R_1$ – thin QR

$$Q_1 = (Q_1\ Q_2)\begin{pmatrix} I \\ 0 \end{pmatrix} = \Big(H_1\Big(H_2 \dots \big(\underbrace{H_n}_{m\times m}\begin{pmatrix} I \\ 0 \end{pmatrix}\big)\Big)\Big)^{n\times n}_{m-n\times n}$$

При умножении H_n на $\begin{pmatrix} I \\ 0 \end{pmatrix}$ мы просто обрезаем H_n за $O(1)$. Затем на k-ой итерации $H_{n-k} = \begin{pmatrix} I_{n-k} & 0 \\ 0 & H(v_{n-k}) \end{pmatrix}$ и $4(m-n+k)\cdot k$ итераций (пользуемся пунктом a здесь), для того чтобы умножить такую матрицу на все предыдущие справа. Т.е. делаем то же самое, что и в пункте a, но в обратном порядке. Итог: $\sum_{k=1}^{n} 4mk - 4nk + 4k^2 = 2mn^2 - \frac{2}{3}n^3 + O(mn)$

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:

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

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

$$A = U\Sigma V \qquad \Sigma = \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} \qquad \Sigma^+ = \begin{pmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{pmatrix}$$

$$A^+ = V\Sigma^+ U^*$$

$$B(\mu) = (A^TA + \mu I)^{-1} A^T$$

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

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

$$= \|V((\Sigma^2 + \mu I)^{-1}\Sigma U^T - \Sigma^+ U^T)\|_2 = \|(\Sigma^2 + \mu I)^{-1}\Sigma - \Sigma^+\|_2 \leftarrow$$ это старшее сингулярное значение

$\det \Sigma_r \neq 0 \Rightarrow \det \Sigma_r^2 \neq 0 \Rightarrow \det(\Sigma^2 + \mu I) \neq 0$, $\mu > 0 \Rightarrow$ обратима

$$\Sigma^2 + \mu I = \operatorname{diag}(\sigma_1^2 + \mu, \dots, \sigma_r^2 + \mu, \mu \dots \mu)$$

$$(\Sigma^2 + \mu I)^{-1} = \operatorname{diag}\left(\frac{1}{\sigma_1^2 + \mu}, \dots, \frac{1}{\sigma_r^2 + \mu}, \frac{1}{\mu} \dots, \frac{1}{\mu}\right)$$

$$(\Sigma^2 + \mu I)^{-1} \cdot \Sigma = \operatorname{diag}\left(\frac{\sigma_1}{\sigma_1^2 + \mu}, \dots, \frac{\sigma_r}{\sigma_r^2 + \mu}, 0, \dots, 0\right)$$

$$(\Sigma^2 + \mu I)^{-1}\Sigma - \Sigma^+ = \operatorname{diag}\left(\frac{\sigma_1}{\sigma_1^2 + \mu} - \frac{1}{\sigma_1}, \dots, \frac{\sigma_r}{\sigma_r^2 + \mu} - \frac{1}{\sigma_r}, 0 \dots 0\right)$$

Посмотрим SVD разложение $(\Sigma^2 + \mu I)^{-1}\Sigma - \Sigma^+ = I\,(\Sigma^2 + \mu I)^{-1}\Sigma - \Sigma^+ \cdot (-I) \Rightarrow$

$$\Rightarrow \|(\Sigma^2 + \mu I)^{-1}\Sigma - \Sigma^+\|_2 = \left|\frac{\sigma_r}{\sigma_r^2 + \mu} - \frac{1}{\sigma_r}\right| = \left|\frac{\sigma_r^2 - \sigma_r^2 - \mu}{(\sigma_r^2 + \mu)\,\sigma_r}\right| = \frac{\mu}{(\sigma_r^2 + \mu)\,\sigma_r}$$

в) $B(\mu) \to A^+$, $x_\mu \to A^+ b$, $\mu \to +0$

$$B(\mu) \to A^+ \Leftrightarrow \lim_{\mu \to +0} \|B(\mu) - A^+\|_2 = 0 \Leftrightarrow \lim_{\mu \to +0} \frac{\mu}{(\sigma_r^2 + \mu)\sigma_r} = 0$$

$$x_\mu = B(\mu) b$$

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

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 $\|x\|_2$.

Все решения имеют вид $x = A^+b + (I - A^+A)y \ \forall y \in \mathbb{R}^n$ и при этом

$x = A^+b$ – минимален по второй норме

$\|x\|_2^2 = \|A^+b\|_2^2 + \|(I-A^+A)y\|_2^2 \Rightarrow \exists z, w \in \mathbb{R}^n: x = z+w$ и $\langle z, w\rangle = 0$

Пусть $w = (I - A^+A)y \in \ker(A)$

Мы знаем, что $\ker(A) \perp \operatorname{Im}(A^T)$

$Ay = (A_{(1)}w, \dots A_{(m)}w)^T = 0$

$A^Tx = V(\Sigma U^T x) \Rightarrow \operatorname{Im}(A^T) = \operatorname{Im}(V)$

$A^+b = V\Sigma^+U^Tb = V(\Sigma^+U^Tb) \in \operatorname{Im}(V) \Rightarrow \in \operatorname{Im}(A^T) \Rightarrow$

$\perp \ker(A) \Rightarrow \langle A^+b, (I - A^+A)y\rangle = 0$ и выполнено р-во из условия

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

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

$V = [a \quad b]$

$W = [a \quad a]$

b) $G = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ при этом $W = [a]$

$V = [a \quad b]$

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

c) $A_{(1)} = U G_{(1)} (\overbrace{W \otimes V}^{n^2\times 2})^T$ $\qquad G_{(1)} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

$A_{(2)} = V G_{(2)} (W \otimes U)^T$ $\qquad G_{(2)} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

$A_{(3)} = W G_{(3)} (V \otimes U)^T$ $\qquad G_{(3)} = (1 \ 0 \ 0 \ 1)$

$$A_{(3)} = \underset{n\times 1}{[a]} \ \underset{1\times 4}{(1 \ 0 \ 0 \ 1)} \ \underset{4\times n^2}{\begin{pmatrix} a_1 [a \ \ 2a+3b] & b_1 [a \ \ 2a+3b] \\ a_2 [a \ \ 2a+3b] & b_2 [a \ \ 2a+3b] \\ \vdots & \vdots \\ a_n [a \ \ 2a+3b] & b_2 [a \ \ 2a+3b] \end{pmatrix}^T}$$

Понятно, что здесь ранг будет равен 1 т.к у $G_{(3)}$ rnk 1 $\Rightarrow$ т.к

$rk(AB) \leq \min(rkA, rkB) \Rightarrow rk(A_{(3)}) = 1$

$$A_{(1)} = [a \quad 2a+3b] \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a_1[a \; b] \\ \vdots \\ a_n[a \; b] \end{pmatrix}^T = [a \quad 2a+3b] \begin{pmatrix} a_1 a & a_1 b \\ \vdots & \\ a_n a & a_n b \end{pmatrix}^T$$

$n \times 2$ $\quad 2\times 2 \quad 2 \times n^2$

Т.к a и b ортогональны $\Rightarrow$ лин. незав rank каждой из этих матриц равен 2 $\Rightarrow rk\, A_{(1)} = 2$

$$A_{(2)} = [a \quad b] \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a_1 a & a_1(2a+3b) \\ \vdots & \vdots \\ a_n a & a_n(2a+3b) \end{pmatrix}^T$$

Здесь те же самые утверждения об ортогональности и линейной независимости как для первой развертки $\Rightarrow rk\, A_{(2)} = 2$

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) Сложность $(A \otimes B)$ $\quad n^2 \times n^2$

Сложность $(A \otimes B)\,x \quad n^2 \cdot n^2 = n^4$

Итоговая сложность: $2n^2 + n^4 = O(n^4)$

б) $(A \otimes B)x = vec(BXA^T)$, где $X = x.\text{reshape}((n,n))$

$$(A\otimes B)x = \begin{pmatrix} a_{11}BX^{(1)} + a_{12}BX^{(2)} + \dots + a_{1n}BX^{(n)} \\ \vdots \\ a_{n1}BX^{(1)} + a_{n2}BX^{(2)} + \dots + a_{nn}BX^{(n)} \end{pmatrix} \rightsquigarrow \left(\sum a_{1i}BX^{(i)} \dots \sum a_{ni}BX^{(i)}\right) =$$

если применим reshape((n,n))

$$= \begin{pmatrix} \sum a_{1i} B_{(1)}X^{(i)} & \dots & \sum a_{ni} BX^{(i)} \\ \vdots & & \vdots \\ \sum a_{1i} B_{(n)} X^{(i)} & \dots & \sum a_{ni} B_{(n)} X^{(i)} \end{pmatrix} = \begin{pmatrix} B_{(1)}X^{(1)} & \dots & B_{(1)}X^{(n)} \\ \vdots & & \vdots \\ B_{(n)}X^{(1)} & & B_{n}X^{(n)} \end{pmatrix} \begin{pmatrix} a_{11} & \dots & a_{n1} \\ \vdots & & \\ a_{1n} & \dots & a_{nn} \end{pmatrix} =$$

$$= \begin{pmatrix} b_{11} & \dots & b_{1n} \\ \vdots & & \vdots \\ b_{n1} & & b_{nn} \end{pmatrix} \begin{pmatrix} x_{11} & \dots & x_{n1} \\ \vdots & & \vdots \\ x_{1n} & \dots & x_{nn} \end{pmatrix} \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \\ a_{n1} & \dots & a_{nn} \end{pmatrix} = BXA^T$$

Чтобы найти BXA^T нужно потратить:

$B_{(i)}X^{(j)}$ – $2n-1$ операций $\Rightarrow (2n-1)\cdot n^2 = 2n^3 - n^2$ т.к. $\forall i,j$

BXA^T тоже $2n^3 - n^2$

Итоговая сложность: $2n^3 - n^2 + 2n^3 - n^2 = O(n^3)$

c) Если C – циркулянт $\Rightarrow C = F_n^{-1}\,\text{diag}(F_n c)\,F_n$

Если C – циркулянт $\Rightarrow C^T$ – тоже циркулянт

BX: $BX^{(i)}$ – умножение циркулянта на вектор $O(n\log n)$

Таких n столбцов $\Rightarrow BX$: $nO(n\log n) = O(n^2\log n)$

$BXA^T = (AX^TB^T)^T$ $A(X^TB^T)^{(i)}$ – за $O(n \log n)$. Всего за $O(n^2 \log n)$

Итог: $O(n^2 \log n)$