نیکو مرادی - 400101934 - تمرین دوم تانسوری

۱) الف)

$\|A\|_* = \sum_{i=1}^{n} \sigma_i(A) \qquad A = U\Sigma V^T \rightsquigarrow A^T = V\Sigma U^T$

$$A^TA = V\Sigma \overbrace{U^TU}^{I} \Sigma V^T = V\Sigma^2 V^T \Rightarrow \sqrt{A^TA} = V(\Sigma^2)^{1/2}V^T = V\Sigma V^T$$

$$\Rightarrow tr(\sqrt{A^TA}) = tr(V\Sigma V^T) = tr(\overbrace{VV^T}^{I}\Sigma) = tr(\Sigma) = \sum_{i=1}^{n}\sigma_i(A) = \|A\|_*$$

$$\Rightarrow \boxed{\|A\|_* = tr(\sqrt{A^TA}) = \sum_{i=1}^{n}\sigma_i(A)}$$

---

۱) ب)

$A, B \in R^{n\times n}$

$$AB = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}\begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nn} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} a_{1i}b_{i1} & \cdots & \square \\ \square & \cdots & \sum_{i=1}^{n} a_{ni}b_{in} \end{bmatrix} \Rightarrow tr(AB) = \sum_{j=1}^{n}\sum_{i=1}^{n} a_{ji}b_{ij} \quad \text{I}$$

$$BA = \begin{bmatrix} \sum_{i=1}^{n} b_{1i}a_{i1} & \cdots & \square \\ \square & \ddots & \sum_{i=1}^{n} b_{ni}a_{ni} \end{bmatrix} \Rightarrow tr(BA) = \sum_{j=1}^{n}\sum_{i=1}^{n} b_{ji}a_{ij} \xrightarrow{j\leftrightarrow i} tr(BA) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ji}b_{ij} \quad \text{II}$$

$$\text{I}, \text{II} \Rightarrow \boxed{tr(AB) = tr(BA)}$$

---

۱) پ)

$\|A\|_* = \max_{C^TC=I} tr(AC) \qquad A = U\Sigma V^T \qquad C^TC = I \rightarrow C$ is orthogonal

$$AC = U\Sigma V^TC \Rightarrow tr(U\Sigma V^TC) = tr(\overbrace{UCV^T}^{B}\Sigma) = tr(B\Sigma) = \sum_{i=1}^{n} b_{ii}\sigma_i(A)$$

$$\max_{C^TC=I} tr(AC) = \max tr(B\Sigma) = \max \sum_{i=1}^{n} b_{ii}\sigma_i(A) = \sum_{i=1}^{n}(\max b_{ii})\sigma_i(A)$$

product of 3 orthogonal matrices $(U, V^T, C) \Rightarrow B$ is orthogonal $\Rightarrow |b_{ii}| \leqslant 1 \Rightarrow \max|b_{ii}| = 1$

$$\Rightarrow \boxed{\max_{C^TC=I} tr(AC) = \sum_{i=1}^{n}\sigma_i(A) = \|A\|_*}$$

---

۱) ت)

می‌دانیم: $\sigma_i(A+B) \leqslant \sigma_i(A) + \sigma_i(B) \Rightarrow \sum_i \sigma_i(A+B) \leqslant \sum_i \sigma_i(A) + \sum_i \sigma_i(B)$

$$\Rightarrow \boxed{\|A+B\|_* \leqslant \|A\|_* + \|B\|_*}$$

Householder : $H = I - \frac{2}{v^Tv} vv^T$ 2) الف)

$H^T = (I - \frac{2}{v^Tv} vv^T)^T = I^T - \frac{2}{v^Tv}(vv^T)^T = I - \frac{2}{v^Tv} vv^T \Rightarrow H^T = H \Rightarrow H$ is symmetric

$HH^T = H^TH = (I - \frac{2}{v^Tv}vv^T)(I - \frac{2}{v^Tv}vv^T) = I - \frac{4}{v^Tv}vv^T + \overbrace{\frac{4}{(v^Tv)^2}(vv^T)(vv^T)}^{\frac{4}{v^Tv}vv^T} = I \Rightarrow HH^T = H^TH = I \Rightarrow H$ is orthogonal

---

From part A : H is orthogonal! 2) ب)

$\forall x : \|Hx\|^2 \overset{!}{=} \|x\|^2$

$Hx = \lambda x \rightsquigarrow \|Hx\|^2 = |\lambda|^2 \cdot \|x\|^2 \Rightarrow \|x\|^2 = |\lambda|^2 \cdot \|x\|^2 \Rightarrow |\lambda|^2 = 1 \Rightarrow \lambda = \pm 1$

Householder: $H = I - \frac{2}{v^Tv} vv^T$ 3)

A) $x = \alpha v \Rightarrow Hx = (I - \frac{2}{v^Tv}vv^T)x = x - \frac{2}{v^Tv}vv^Tx = x - \frac{2\alpha}{\cancel{v^Tv}}v\cancel{v^Tv} = x - \overbrace{2\alpha v}^{x} = -x$

$\Rightarrow x = \alpha v,\ \alpha \neq 0 \Rightarrow Hx = -x$

B) $v \cdot x = 0 \Rightarrow v^Tx = 0 \Rightarrow Hx = (I - \frac{2}{v^Tv}vv^T)x = x - \frac{2}{v^Tv}v\overbrace{v^Tx}^{0} = x$

$\Rightarrow v^Tx = 0 \Rightarrow Hx = x$

A, B $\Rightarrow$ Householder transformation is indeed a reflector.

$A \in \mathbb{R}^{m \times n}$ rank(A) = n $\rightsquigarrow$ A is full column rank $V^Tx = y \xrightarrow{V^TV = I} \|x\| = \|y\| = 1$ 4)

$A = U\Sigma V^T \rightsquigarrow \|Ax\|_2 = [(Ax)^TAx]^{1/2} = [(U\Sigma V^Tx)^T(U\Sigma V^Tx)]^{1/2} \xrightarrow[x = Vy]{V^Tx = y} [(U\Sigma y)^T(U\Sigma y)]^{1/2} = [y^T\Sigma^T\overbrace{U^TU}^{I}\Sigma y]^{1/2}$

$\Rightarrow \|Ax\|_2 = [y^T\Sigma^T\Sigma y]^{1/2} = [y^Ty\sum_{i=1}^{n}\sigma_i^2(A)]^{1/2} = \sqrt{\sum_{i=1}^{n}\sigma_i^2 y_i^2}$

$\|Ax\|_2 = \sqrt{\sum_{i=1}^{n}\sigma_i^2 y_i^2} \geq \sigma_{min}\sqrt{\sum_{i=1}^{n} y_i^2} = \sigma_{min}\overbrace{\|y\|}^{1} \Rightarrow \|Ax\|_2 \geq \sigma_{min}(A) \Rightarrow$

$\min \|Ax\|_2 = [x^T\overbrace{A^TAx}^{\sigma_{min}^2 x}]^{1/2} = \sigma_{min}(A)\|x\| \xrightarrow{\|x\|=1} x$ is the right singular vector corresponding to σ_{min}

$A \in \mathbb{R}^{m \times n} \rightsquigarrow$ full column rank $\Rightarrow rank(A) = n$ $\quad A^TA \in \mathbb{R}^{n \times n}$ 5) الف

only vector with norm 0, is $\mathbf{0}$ vector

$A^TAx = 0 \Rightarrow x^TA^TAx = 0 \Rightarrow (Ax)^T(Ax) = 0 \Rightarrow \|Ax\|^2 = 0 \Rightarrow Ax = 0 \rightsquigarrow$

A is a full column rank, meaning that $N(A) = \emptyset$ & the columns of A are linearly independent $\Rightarrow x = 0$

Conclusion : $A^TAx = 0 \Rightarrow x = 0 \rightsquigarrow$ It means that $N(A^TA) = \emptyset$ and that $(A^TA)_{n \times n}$ is also full rank $\Rightarrow rank(A^TA) = n \Rightarrow det(A^TA) \neq 0 \Rightarrow$ A^TA is non_singular

---

$A \in \mathbb{R}^{m \times n} \rightsquigarrow$ full row rank $\Rightarrow rank(A) = m$ $\quad AA^T \in \mathbb{R}^{m \times m}$ 5) ب

only vector with norm 0, is $\mathbf{0}$ vector

$AA^Tx = 0 \Rightarrow x^TAA^Tx = 0 \Rightarrow (A^Tx)^T(A^Tx) = 0 \Rightarrow \|A^Tx\|^2 = 0 \Rightarrow A^Tx = 0 \rightsquigarrow$ A is a full row rank, So A^T is a full column rank, meaning that the null space of A^T is empty & A^T's columns are linearly independent so the only possible answer for $A^Tx = 0$ is $x = 0$

Conclusion : $AA^Tx = 0 \Rightarrow x = 0 \Rightarrow$ It means that the null space of AA^T is empty $N(AA^T) = \emptyset$ & $(AA^T)_{m \times m}$ is full rank $\Rightarrow rank(AA^T) = m \Rightarrow det(AA^T) \neq 0 \Rightarrow$ A^TA is non-Singular

---

* $rank(A) = n$: From part A : if $rank(A) = n \Rightarrow A^TA$ is non-Singular $\Rightarrow (A^TA)^{-1}(A^TA) = I$ 5) ج

$A^TA = V\Sigma^TU^TU\Sigma V^T = V\Sigma^T\Sigma V^T \to diag(\sigma_1^2, \dots, \sigma_n^2) \Rightarrow \Sigma^T\Sigma$ is non-singular $\Rightarrow (A^TA)^{-1} = V(\Sigma^T\Sigma)^{-1}V^T$

!

$(A^TA)^{-1}A^T = V(\Sigma^T\Sigma)^{-1}V^T(V\Sigma^TU^T) = V(\Sigma^T\Sigma)^{-1}\Sigma^TU^T = V\Sigma^+U^T \Rightarrow$ $A^+ = (A^TA)^{-1}A^T = V\Sigma^+U^T$

* $rank(A) = m$: From part B : if $rank(A) = m \Rightarrow AA^T$ is non-Singular

$AA^T = U\Sigma V^TV\Sigma^TU^T = U\Sigma\Sigma^TU^T \to diag(\sigma_1^2, \dots, \sigma_m^2) \Rightarrow \Sigma\Sigma^T$ is non_singular $\Rightarrow (AA^T)^{-1} = U(\Sigma\Sigma^T)^{-1}U^T$

!!

$A^T(AA^T)^{-1} = V\Sigma^TU^TU(\Sigma\Sigma^T)^{-1}U^T = V\Sigma^T(\Sigma\Sigma^T)^{-1}U^T = V\Sigma^+U^T \Rightarrow$ $A^+ = A^T(AA^T)^{-1} = V\Sigma^+U^T$

! $\Sigma^T\Sigma = diag(\sigma_1^2, \dots, \sigma_n^2) \Rightarrow (\Sigma^T\Sigma)^{-1} = diag\left(\frac{1}{\sigma_1^2}, \dots, \frac{1}{\sigma_n^2}\right) \Rightarrow (\Sigma^T\Sigma)^{-1}\Sigma^T = \begin{bmatrix} \frac{1}{\sigma_1} & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \\ 0 & \cdots & \frac{1}{\sigma_n} & \cdots & 0 \end{bmatrix}_{n \times m}$, $[(\Sigma^T\Sigma)^{-1}\Sigma^T]\Sigma = I \Rightarrow (\Sigma^T\Sigma)^{-1}\Sigma^T = \Sigma^+$

!! $\Sigma\Sigma^T = diag(\sigma_1^2, \dots, \sigma_m^2) \Rightarrow (\Sigma\Sigma^T)^{-1} = diag\left(\frac{1}{\sigma_1^2}, \dots, \frac{1}{\sigma_n^2}\right) \Rightarrow \Sigma^T(\Sigma\Sigma^T)^{-1} = \begin{bmatrix} \frac{1}{\sigma_1} & \cdots & 0 \\ \vdots & \ddots & \frac{1}{\sigma_m} \\ 0 & \cdots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}_{n \times m}$, $\Sigma[\Sigma^T(\Sigma\Sigma^T)^{-1}] = I \Rightarrow \Sigma^T(\Sigma\Sigma^T)^{-1} = \Sigma^+$

$\forall x \in$ singular vectors of A : $\|Ax\|_2 = [(Ax)^T(Ax)]^{1/2} = [x^T \underbrace{A^T A x}_{\sigma_i^2 x}]^{1/2} = \sigma_i (x^T x)^{1/2} = \sigma_i \|x\|_2$ (6

For any non-zero x : $\|Ax\|_2 \leqslant \sigma_{max} \|x\|_2$ I

This comes from the definition of singular values & the fact that the action of A on any x can't stretch x by more than σ_{max} in any direction.

$\forall x \in$ eigenvalues of A : $\|Ax\|_2 = \|\lambda x\|_2 = |\lambda| \, \|x\|_2$ II

I, II $\Rightarrow$ $|\lambda| \, \|x\|_2 = \|Ax\|_2 \leqslant \sigma_{max} \|x\|_2 \Rightarrow |\lambda| \leqslant \sigma_{max}$ $\quad \sigma_{max} = \sigma_1$

$\|A\|_F = \sqrt{\sum_{i=1}^{r} \sigma_i^2}$ $\quad$ $\text{rank}(A) = r$ $\quad$ $A = U \Sigma V^T$ (7 الف)

A : $\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2} = \sqrt{tr(AA^T)}$

B : $AA^T = U\Sigma \underbrace{V^T V}_{I} \Sigma^T U^T = U \Sigma \Sigma^T U^T \Rightarrow tr(AA^T) = tr(U\Sigma\Sigma^T U^T) = tr(\underbrace{U U^T}_{I} \Sigma \Sigma^T) = tr(\Sigma\Sigma^T)$

C : $\Sigma = \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \sigma_r \\ 0 & & 0 \end{bmatrix} \rightsquigarrow \Sigma\Sigma^T = \underset{m \times n}{\begin{bmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_r \\ 0 & \cdots & 0 \end{bmatrix}} \underset{n \times m}{\begin{bmatrix} \sigma_1 & \cdots 0 \cdots & 0 \\ \vdots & \ddots \;\; \sigma_r & 0 \end{bmatrix}} = \underset{m \times m}{\begin{bmatrix} \sigma_1^2 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & & & \vdots \\ 0 & \cdots & \sigma_r^2 & \cdots & 0 \\ 0 & \cdots & 0 & \cdots & 0 \end{bmatrix}} \Rightarrow tr(\Sigma\Sigma^T) = \sum_{i=1}^{r} \sigma_i^2$

A, B, C $\Rightarrow$ $\|A\|_F = \sqrt{tr(AA^T)} = \sqrt{tr(\Sigma\Sigma^T)} = \sqrt{\sum_{i=1}^{r} \sigma_i^2}$

A : $\|A\|_F = \sqrt{\sum_{i=1}^{r} \sigma_i^2} \leqslant \sqrt{\sum_{i=1}^{r} \sigma_{max}^2} = \sqrt{r \sigma_{max}^2} = \sqrt{r}\, \sigma_{max}(A)$ $\quad \sigma_1 = \sigma_{max}$ (7 ب)

B : $\sigma_{max}^2(A) \leqslant \sigma_{max}^2 + \cdots + \sigma_r^2 = \sum_{i=1}^{r} \sigma_i^2 \Rightarrow \sigma_{max}(A) \leqslant \sqrt{\sum_{i=1}^{r} \sigma_i^2} = \|A\|_F$

A, B $\Rightarrow$ $\sigma_{max}(A) \leqslant \|A\|_F \leqslant \sqrt{r}\, \sigma_{max}(A)$

$H = I - \frac{2}{v^T v} v v^T$, $x \neq 0$ (8

$x^T H x = x^T (I - \frac{2}{v^T v} v v^T) x = x^T x - \frac{2}{v^T v} x^T v v^T x = \|x\|^2 - \frac{2}{v^T v} (v^T x)^T (v^T x) = \|x\|^2 - \frac{2}{\|v\|^2} \|v^T x\|$

$\longrightarrow$ 1. $x = \alpha v$, $\alpha \neq 0$ $\Rightarrow x^T H x = |\alpha|^2 \|v\|^2 - 2|\alpha| \|v\|^2 = \underbrace{|\alpha|}_{+} \underbrace{\|v\|^2}_{+} \underbrace{(|\alpha| - 2)}_{+ \text{ or } -} \Rightarrow$ H is indefinite

$\longrightarrow$ 2. $v^T x = 0 \Rightarrow x^T H x = \|x\|^2 - 0 = \|x\|^2 > 0$

Most of the algorithms that we present in this book have complex versions that are fairly straightforward to derive from their real counterparts. (This is *not* to say that everything is easy and obvious at the implementation level.) As an illustration we briefly discuss complex Householder and complex Givens transformations. (9

Recall that if $A = (a_{ij}) \in \mathbb{C}^{m \times n}$, then $B = A^H \in \mathbb{C}^{n \times m}$ is its conjugate transpose. The 2-norm of a vector $x \in \mathbb{C}^n$ is defined by

$$\| x \|_2^2 = x^H x = |x_1|^2 + \cdots + |x_n|^2$$

and $Q \in \mathbb{C}^{n \times n}$ is *unitary* if $Q^H Q = I_n$. Unitary matrices preserve the 2-norm.

A complex Householder transformation is a unitary matrix of the form

$$P = I_m - \beta v v^H, \qquad 0 \neq v \in \mathbb{C}^m,$$

where $\beta = 2/v^H v$. Given a nonzero vector $x \in \mathbb{C}^m$, it is easy to determine v so that if $y = Px$, then $y(2{:}m) = 0$. Indeed, if

$$x_1 = re^{i\theta}$$

where $r, \theta \in \mathbb{R}$ and

$$v = x \pm e^{i\theta} \| x \|_2 e_1, \qquad e_1 = I_m(:,1),$$

then $Px = \mp e^{i\theta} \| x \|_2 e_1$. The sign can be determined to maximize $\| v \|_2$ for the sake of stability.

Regarding complex Givens rotations, it is easy to verify that a 2-by-2 matrix of the form

$$Q = \begin{bmatrix} \cos(\theta) & \sin(\theta)e^{i\phi} \\ -\sin(\theta)e^{-i\phi} & \cos(\theta) \end{bmatrix}$$

where $\theta, \phi \in \mathbb{R}$ is unitary. We show how to compute $c = \cos(\theta)$ and $s = \sin(\theta)e^{i\phi}$ so that

$$\begin{bmatrix} c & s \\ -\bar{s} & c \end{bmatrix}^H \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} r \\ 0 \end{bmatrix} \tag{5.1.12}$$

where $u = u_1 + iu_2$ and $v = v_1 + iv_2$ are given complex numbers. First, givens is applied to compute real cosine-sine pairs $\{c_\alpha, s_\alpha\}$, $\{c_\beta, s_\beta\}$, and $\{c_\theta, s_\theta\}$ so that

$$\begin{bmatrix} c_\alpha & s_\alpha \\ -s_\alpha & c_\alpha \end{bmatrix}^T \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} r_u \\ 0 \end{bmatrix},$$

$$\begin{bmatrix} c_\beta & s_\beta \\ -s_\beta & c_\beta \end{bmatrix}^T \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} r_v \\ 0 \end{bmatrix},$$

and

$$\begin{bmatrix} c_\theta & s_\theta \\ -s_\theta & c_\theta \end{bmatrix}^T \begin{bmatrix} r_u \\ r_v \end{bmatrix} = \begin{bmatrix} r \\ 0 \end{bmatrix}.$$

Note that $u = r_u e^{-i\alpha}$ and $v = r_v e^{-i\beta}$. If we set

$$e^{i\phi} = e^{i(\beta - \alpha)} = (c_\alpha c_\beta + s_\alpha s_\beta) + i(c_\alpha s_\beta - c_\beta s_\alpha),$$

$c = c_\theta$, and $s = s_\theta e^{i\phi}$, then

$$\bar{s}u + cv = s_\theta e^{-i\phi} r_u e^{-i\alpha} + c_\theta r_v e^{-i\beta} = e^{-i\beta}(s_\theta r_u + c_\theta r_v) = 0$$

which confirms (5.1.12).