سوال ۱)

الف)

$$\nabla f(w) = \frac{\partial}{\partial w}\Big(\underbrace{\frac{1}{2n}(X^T w - y)^T(X^T w - y)}_{B} + \underbrace{\lambda/2\, w^T w}_{A}\Big)$$

$$= \frac{\partial A}{\partial w} + \frac{\partial B}{\partial w}$$

$$= 2w\,\lambda/2 + \frac{\partial\left(1/2n\left(2XX^T w - (y^T X^T)^T - Xy\right)\right)}{\partial w}$$

$$= \lambda w + \frac{XX^T w}{n} - \frac{Xy}{n}$$

$$= \boxed{Aw - \frac{Xy}{n}}$$

$$\rightarrow \nabla f(w^*) = 0 \longrightarrow Aw^* - \frac{xy}{n} = 0 \rightarrow \frac{xy}{n} = Aw^*$$

$$\longrightarrow \nabla f(w) = Aw - Aw^* = A(w - w^*)$$

ب)

$$x^T A x = 1/n\, x^T X X^T x + \lambda\, x^T x$$

$$= 1/n\, \|X^T x\|_2^2 + \lambda \|x\|_2^2$$

$\lambda \geq 0 \longleftarrow$ با توجه به منظم سازی داریم

$$\longrightarrow x^T A x = \underbrace{\frac{\|X^T x\|^2}{n}}_{\geq 0} + \underbrace{\lambda \|x\|^2}_{\geq 0} \geq 0 \rightarrow A: \text{semidefinite pos}$$

$$\sigma_{max}(I-\alpha A) = \max_{\|x\|_2=1} \|(I-\alpha A)x\|_2 = \qquad (ج$$

$$= -\alpha \min_{\|x\|_2=1} \|Ax - x/\alpha\|$$

$$\|Ax - \frac{x}{\alpha}\|_2 \geq \|Ax\|_2 - 1/\alpha \, \|x\|_2$$ (باتوجه به خاصیت نرم)

$$\longrightarrow \sigma_{max} = (I-\alpha A) = -\alpha \min_{\|x\|_2=1} \|Ax - x/\alpha\|_2$$

$$= -\alpha \min \left(\|Ax\|_2 - \frac{\|x\|_2}{\alpha} \right) = -\alpha \min \|Ax\|_2 + 1$$

$$= -\alpha \min_{\|x\|_2=1} \|U\Sigma \underbrace{V^T x}_{y}\|_2 + 1 = -\alpha \min \|\Sigma y\|_2 + 1$$

$$= -\alpha \sigma_{min}(A) + 1 = 1 - \alpha \sigma_{min}(A)$$

$$\boxed{= 1 - \frac{\sigma_{min}(A)}{\sigma_{max}(A)}}$$

$$\|w^{t+1} - w^*\| = \|w^t - \alpha \nabla f(w^t) - w^*\| = \|w^t - \alpha A(w^t - w^*) - w^*\|$$

$$= \|(I-\alpha A)(w^t - w^*)\| = \frac{\|(I-\alpha A)(w^t - w^*)\|}{\|w^t - w^*\|} \|w^t - w^*\| \qquad (د$$

$$\leq \max_{x \neq 0} \frac{\|(I-\alpha A)x\|}{\|x\|} \|w^t - w^*\| = \sigma_{max}(I-\alpha A) \|w^t - w^*\|$$

$$\boxed{= \left(1 - \frac{\sigma_{min}(A)}{\sigma_{max}(A)}\right) \|w^t - w^*\|}$$

سوال ۲)

$$M = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$$

operation:

$$\text{row1} = A^{-1} \times \text{row1} \qquad U: \begin{bmatrix} I & A^{-1}B \\ B^T & C \end{bmatrix} \qquad L: \begin{bmatrix} A & 0 \\ 0 & I \end{bmatrix}$$

$$\text{row2} = \text{row2} - B^T \text{row1} \qquad \begin{bmatrix} I & A^{-1}B \\ 0 & C - B^TA^{-1}B \end{bmatrix} \qquad \begin{bmatrix} A & 0 \\ B^T & I \end{bmatrix}$$

$$\longrightarrow M = \underbrace{\begin{bmatrix} A & 0 \\ B^T & I \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} I & A^{-1}B \\ 0 & C - B^TA^{-1}B \end{bmatrix}}_{U}$$

برای تجزیه قطری M ← تلاش بر تجزیه قطری L و U به صورت ψD و $D\psi$

$$L = \begin{bmatrix} I & 0 \\ B^TA^{-1} & I \end{bmatrix}\begin{bmatrix} A & 0 \\ 0 & I \end{bmatrix} \qquad U = \begin{bmatrix} I & 0 \\ 0 & C - B^TA^{-1}B \end{bmatrix}\begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}$$

$$\longrightarrow M = LU = \psi^T \begin{pmatrix} A & 0 \\ 0 & C - B^TA^{-1}B \end{pmatrix} \psi = \psi^T D \psi$$

اول: $C - B^TA^{-1}B$, A psd $\rightarrow$ M psd

$$x^T M x = x^T P^T D P x = y^T D y \rightarrow y = Px$$

P فول رنک است پس $y=0 \longleftrightarrow x=0$

پس D psd است $\longleftrightarrow$ M psd است.

حالا باید اثبات کنیم D، psd است.

$$X^T D X = (x_1^T \; x_2^T)\begin{pmatrix} A & 0 \\ 0 & C - B^T A^{-1} B \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= \underbrace{x_1^T A x_1}_{\substack{\geq 0 \\ psd}} + \underbrace{x_2^T (C - B^T A^{-1} B) x_2}_{\substack{\geq 0 \\ psd}} \geq 0 \longrightarrow D \text{، psd است}$$

حالا:

$$M \; psd \longrightarrow A \text{ و } C - B^T A^{-1} B \; psd$$

for $x^T A x$: $\tilde{x} = \begin{bmatrix} x \\ 0 \end{bmatrix}$ و $\tilde{x}^T D \tilde{x} = (x^T \; 0)\begin{pmatrix} A & 0 \\ 0 & C - B^T A^{-1} B \end{pmatrix}\begin{pmatrix} x \\ 0 \end{pmatrix}$

$$= x^T A x \geq 0$$

$$\longrightarrow \forall x : x^T A x \geq 0 \longrightarrow A \; psd$$

همانند بالا:

$x^T (C - B^T A^{-1} B) x$: $\tilde{x} = \begin{bmatrix} 0 \\ x \end{bmatrix}$

$$\longrightarrow \tilde{x}^T D \tilde{x} = (0 \; x^T)\begin{pmatrix} A & 0 \\ 0 & C - B^T A^{-1} B \end{pmatrix}\begin{pmatrix} 0 \\ x \end{pmatrix} = x^T (C - B^T A^{-1} B) x \geq 0$$

$$\longrightarrow (C - B^T A^{-1} B) \text{ psd است.}$$

(سوال ۳)

$$C = \begin{bmatrix} c_1 & 0 & \cdots & 0 \\ 0 & c_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & c_n \end{bmatrix} , \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} , \quad X = \begin{bmatrix} | & & | \\ x_1 & \cdots & x_n \\ | & & | \end{bmatrix}$$

$$\longrightarrow \min \sum c_i^2 (w \cdot x_i - y_i)^2 = \min \| \overbrace{\underbrace{CX^T}_{A} \underbrace{w}_{x}}^{Ax} - \overbrace{Cy}^{B} \|^2$$

$$\longrightarrow Ax = b \rightsquigarrow A^T A x = A^T b$$

$$\longrightarrow \underline{\hat{x} = (A^T A)^{-1} A^T b}$$

$$\longrightarrow \hat{w} = \left((CX^T)^T (CX^T) \right)^{-1} (CX^T)^T (Cy)$$

$$= (X C^T C X^T)^{-1} (X C^T C y)$$

$$\longrightarrow \boxed{\hat{w} = (X C^T C X^T)^{-1} (X C^T C y)}$$

سوال ٤)

الف ← ب)

$Av_i = \lambda_i v_i \rightarrow v_i^T A v_i \geq 0$

$\longrightarrow v_i^T \lambda_i v_i \geq 0 \rightarrow \lambda_i \|v_i\|^2 \geq 0$ و $\|v_i\|^2 > 0 \leftarrow v_i \neq 0$

$\longrightarrow \boxed{\lambda_i \geq 0}$

ب ← ج)

$A^T = A \rightarrow A = UDU^T$

$\longrightarrow \begin{cases} D: & \text{ماتریس قطری با قطر درایه‌های } \lambda_i \\ U: & \text{Unitary} \end{cases}$

$\Sigma = \sqrt{\lambda_i}$ ماتریس قطری با درایه‌های $\rightarrow D = \Sigma \Sigma^T$

$\longrightarrow A = U\Sigma\Sigma^T U^T = (U\Sigma)(U\Sigma)^T \rightarrow B = U\Sigma$

$\longrightarrow \boxed{A = BB^T}$

ج ← الف)

$A = BB^T \longrightarrow x^T A x = x^T B B^T x = \|B^T x\|^2 \geq 0$

$\longrightarrow A$ مثبت semidefinite

از هر یک می‌توان به دیگری رسید ←

هر ۳ مورد معادل اند.

سوال ۱۵ الف)

$$\mathrm{Tr}(A^TA) \longrightarrow A = \begin{bmatrix} | & & | \\ a_1 & \text{—} & a_n \\ | & & | \end{bmatrix}$$

$$\longrightarrow \mathrm{Tr}(A^TA) = \sum_{i=1}^{n} a_i^T a_i = \sum_i \sum_j A_{ji}^2 = \sum_{i,j} |A_{ij}|^2$$

$$\longrightarrow \boxed{\|A\|_F = \sqrt{\mathrm{Tr}(A^TA)} = \Big(\sum_{i,j} |A_{ij}|^2\Big)^{1/2}}$$

ب)

$$\|UA\|_F = \sqrt{\mathrm{Tr}((UA)^T(UA))} = \sqrt{\mathrm{Tr}(A^TA)} = \|A\|_F \quad (1)$$

$$\|AV\|_F = \sqrt{\mathrm{Tr}((AV)^T(AV))} = \sqrt{\mathrm{Tr}(V^TA^TAV)} = \sqrt{\mathrm{Tr}(VV^TA^TA)}$$

$$= \sqrt{\mathrm{Tr}(A^TA)} = \|A\|_F \quad (2)$$

$$\xrightarrow{1,2} \boxed{\|UA\|_F = \|AV\|_F = \|A\|_F}$$

ج)

$$A = U\Sigma V^* \longrightarrow \|A\|_F = \|U\Sigma V^*\|_F = \|\Sigma V^*\|_F = \|\Sigma\|_F$$

$$= \sqrt{\mathrm{Tr}(\Sigma^T\Sigma)} = \sqrt{\sum_{i=1}^{n} \sigma_i^2} = \sqrt{\sum_{i=1}^{r} \sigma_i^2}$$

$$* \quad \Sigma^T\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & \cdots \\ 0 & \ddots & \\ \vdots & & \sigma_n^2 \end{pmatrix} \text{ و } \sigma_{r+1} = \cdots = \sigma_n = 0$$

د)

$$\sigma_{max}(A) \leq \sqrt{\sigma_1^2 + \cdots + \sigma_r^2} = \|A\|_F \leq \sqrt{r\,\sigma_{max}^2(A)}$$

$$= \sqrt{r}\,(\sigma_{max}(A))$$

طل ۹)

$$\frac{\partial(|A|A^TxA)}{\partial A} = \frac{\partial|A|}{\partial A}A^TxA + |A|\frac{\partial(A^TxA)}{\partial A}$$

$$= |A|A^{-T}A^TxA + |A|\left(\frac{\partial A^T}{\partial A}xA + A^T\frac{\partial x}{\partial A}A + A^Tx\frac{\partial A}{\partial A}\right)$$

$$= |A|xA + |A|xA + |A|A^T\frac{\partial x}{A}A + A^Tx$$

$$= \boxed{|A|\left(2xA + A^T\frac{\partial x}{A}A + A^Tx\right)}$$

بن :

$$\frac{\partial|A|}{\partial A} = |A|A^{-T}$$

$$\frac{\partial|A|}{\partial A} = \begin{bmatrix} \frac{\partial|A|}{\partial A_{11}} & \cdots & \frac{\partial|A|}{\partial A_{1n}} \\ \vdots & & \\ \frac{\partial|A|}{\partial A_{n1}} & \cdots & \frac{\partial|A|}{\partial A_{nn}} \end{bmatrix} = \begin{bmatrix} C_{11} & \cdots & C_{1n} \\ \vdots & & \\ C_{n1} & & C_{nn} \end{bmatrix} = Adj(A)^T$$

$$adj(A) \times |A|^{-1} = A^{-1} \longrightarrow \left(\frac{\partial|A|}{\partial A}\right)^T|A|^{-1} = A^{-1}$$

$$\longrightarrow |A|A^{-1} = \left(\frac{\partial|A|}{\partial A}\right)^T$$

$$\longrightarrow \frac{\partial(|A|)}{\partial A} = |A|A^{-T}$$

سوال ۷) الف)

$$\frac{\partial y}{\partial x} = \begin{bmatrix} \frac{\partial f(x_1)}{\partial x_1} & \cdots & \frac{\partial f(x_1)}{\partial (x_n)} \\ \vdots & & \vdots \\ \frac{\partial f(x_n)}{\partial (x_1)} & \cdots & \frac{\partial f(x_n)}{\partial (x_n)} \end{bmatrix}$$

$$= \mathrm{diag}\left(f'(x_1), f'(x_2), \dots, f'(x_n)\right)$$

ب)

$$\frac{\partial z}{\partial x} = \begin{bmatrix} \frac{\partial x_1 y_1}{\partial x_1} & \cdots & \frac{\partial x_1 y_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial x_n y_n}{\partial x_1} & \cdots & \frac{\partial x_n y_n}{\partial x_n} \end{bmatrix} = \mathrm{diag}(y_1, y_2, \dots, y_n)$$

پ)

$$\frac{\partial z}{\partial x} = \begin{bmatrix} \frac{\partial (f(x_1, y_1))}{\partial x_1} & \cdots & \frac{\partial (f(x_1, y_1))}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial (f(x_n, y_n))}{\partial x_1} & \cdots & \frac{\partial (f(x_n, y_n))}{\partial x_n} \end{bmatrix}$$

$$= \mathrm{diag}\left(\frac{\partial (f(x_1, y_1))}{\partial x_1}, \dots, \frac{\partial (f(x_n, y_n))}{\partial x_n}\right)$$