۹۸۱۰۰۸۷۹ - علی عباسی — تمرین اول یادگیری عمیق

۱) آ)

$$\|A\|_F^2 = tr(A^HA) = \sum_i \lambda_i(A^HA) \geqslant \lambda_{max}(A^HA) = \|A\|_2^2 \quad (1)$$

(جمع قطر = جمع مقادیر ویژه)
اثبات پس سوال ۳

می‌دانیم rank یک ماتریس تعداد مقادیر ویژه ناصفر آن ماتریس را نشان می‌دهد. بنابراین:

$$\|A\|_F^2 = \sum_i \lambda_i(A^HA) = \sum_{i:\lambda_i\neq 0} \lambda_i(A^HA) \leqslant \underbrace{rank(A^HA)}_{= \text{تعداد } \lambda_i \text{ های ناصفر}} \times \lambda_{max}(A^HA)$$

همچنین $rank(A^HA) \overset{(*)}{=} rank(A)$ (اثبات در پایین). پس:

$$\|A\|_F^2 \leqslant rank(A)\,\lambda_{max}(A^HA) = rank(A)\,\|A\|_2^2 \quad (2)$$

$$(1,2) \rightarrow \boxed{\|A\|_2 \leqslant \|A\|_F \leqslant \sqrt{rank(A)}\,\|A\|_2}$$

اثبات (*):

$$x\in N(A) \Rightarrow Ax=0 \xRightarrow{\times A^H} A^HAx=0 \Rightarrow x\in N(A^HA) \Rightarrow N(A)\subseteq N(A^HA) \quad (1)$$

$$x\in N(A^HA) \Rightarrow A^HAx=0 \xrightarrow{\times x^H} x^HA^HAx=0 \Rightarrow \|Ax\|^2=0 \Rightarrow \|Ax\|=0 \Rightarrow Ax=0$$

$$\Rightarrow x\in N(A) \Rightarrow N(A^HA)\subseteq N(A) \quad (2)$$

$$(1,2) \rightarrow N(A)=N(A^HA) \Rightarrow \boxed{rank(A)=rank(A^HA)}$$

ب) i) فرض کنید X یک متغیر گسسته باشد
(در صورت پیوسته بودن کافی است به‌جای Σ، انتگرال گرفت)

$$E(X) = \sum_x x\,p(x)$$

$$\xrightarrow{xp(x)\geq 0} \geqslant \sum_{x\geqslant\alpha} x\,p(x) \geqslant \sum_{x\geqslant\alpha} \alpha\,p(x)$$

$$= \alpha\sum_{x\geqslant\alpha} p(x) = \alpha\,P(X\geqslant\alpha) \Rightarrow \boxed{P(X\geqslant\alpha) \leqslant \frac{E(X)}{\alpha}}$$

ii)

$$\boxed{P(|Z-\mu|\geqslant\varepsilon)} = \underbrace{P(|Z-\mu|^2\geqslant\varepsilon^2)}_{\text{هردو مثبت}} \geqslant \frac{E[|Z-\mu|^2]}{\varepsilon^2} = \frac{Var(Z)}{\varepsilon^2} = \boxed{\frac{\sigma^2}{\varepsilon^2}}$$

iii) متغیر تصادفی I را به این صورت تعریف می‌کنیم:

$$I = \begin{cases} 1 & (\text{نقطه در دایره}) \quad p=\frac{\pi}{4} \\ 0 & (\text{نقطه بیرون دایره}) \quad p=1-\frac{\pi}{4} \end{cases}$$

حال n بار نقطه‌ی تصادفی ایجاد می‌کنیم. $\frac{1}{n}$ حاصل جمع I ها، تخمین $\frac{\pi}{4}$ خواهد بود. پس Z که خود π را تخمین می‌زند به این صورت تعریف می‌کنیم:

$$Z = \frac{4}{n}\sum_{i=1}^n I_i$$

میانگین Z، خود π خواهد بود ($\mu = E[Z] = \frac{4}{n}\sum_{i=1}^n E[I_i] = \pi$)

واریانس Z را نیز برای استفاده در چبیشف محاسبه می‌کنیم:

$$Var(Z) = Var\left(\frac{4}{n}\sum_{i=1}^n I_i\right) = \frac{16}{n^2}\,Var\left(\sum_i I_i\right)$$

$$\overset{(\text{مستقل هستند } I_i)}{=} \frac{16}{n^2}\sum_{i=1}^n Var(I_i) = \frac{16}{n^2}\times n\times\left(\frac{\pi}{4}\right)\left(1-\frac{\pi}{4}\right) = \boxed{\frac{1}{n}\times\pi(4-\pi)} = \sigma^2$$

Scanned with CamScanner

۱۷) می‌خواهیم با قطعیت ۹۵ درصد خطای تخمین کمتر از یک درصد باشد:

$$\text{خطا} = \frac{|Z-\pi|}{\pi} \leq 0/01 \Rightarrow |Z-\pi| \leq \frac{\pi}{100}$$

$$P\left(|Z-\pi| \leq \frac{\pi}{100}\right) \geq 0/95 \Rightarrow P\left(|Z-\pi| \geq \frac{\pi}{100}\right) \leq 0/05$$

جایگذاری در از بخش قبل

$$\Rightarrow \varepsilon^* = \frac{\pi}{100}, \quad \frac{\sigma^2}{\varepsilon^2} = \frac{5}{100} \Rightarrow \sigma^2 = \frac{5}{100}\varepsilon^2 = \frac{5\pi^2}{10^6}$$

حداقل تعداد نقاط مورد نیاز

$$\sigma^2 = \frac{1}{n}\pi(4-\pi) = \frac{5\pi^2}{10^6} \Rightarrow \boxed{n_{min} = \frac{10^6(4-\pi)}{5\pi}}$$

می‌توانیم یک حد بالا نیز برای آن به دست بیاوریم (با توجه به این که می‌دانیم π بین ۳ و ۴ است ولی مقدار دقیق آن را در این سوال نمی‌دانیم)

$$\pi > 3, \quad 4-\pi < 1 \Rightarrow n_{min} < \frac{10^6}{5 \times 3} = \frac{2}{3}\times 10^5$$

۲)

- $$\left[\frac{\partial a^T x}{\partial x}\right]_k = \frac{\partial a^T x}{\partial x_k} = \frac{\partial}{\partial x_k}\sum_{i=1}^{n} a_i x_i = a_k \xrightarrow{\text{برداری می‌کنیم}} \frac{\partial a^T x}{\partial x} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \boxed{a}$$

- $$\left[\frac{\partial x^T A x}{\partial x}\right]_k = \frac{\partial x^T A x}{\partial x_k} = \frac{\partial}{\partial x_k}\sum_{i=1}^{n} x_i (Ax)_i = \frac{\partial}{\partial x_k}\sum_{i=1}^{n}\sum_{j=1}^{n} x_i A_{ij} x_j$$

$$= \sum_{i\neq k} x_i A_{ik} + \sum_{j \neq k} x_j A_{kj} + 2A_{kk}x_k$$

یک جمله نسبت به x_k درجه ۲ است. بقیه درجه ۱ هستند. جداگانه محاسبه می‌شوند.

$$= \sum_i A_{ik}x_i + \sum_j A_{kj}x_j \overset{B \triangleq A^T}{=} \langle B_k, x\rangle + \langle A_k, x\rangle$$

سطر kام

$$\xrightarrow{\text{برداری}} \frac{\partial x^T A x}{\partial x} = \begin{bmatrix} B_1^T x \\ \vdots \\ B_n^T x \end{bmatrix} + \begin{bmatrix} A_1^T x \\ \vdots \\ A_n^T x \end{bmatrix} = Bx + Ax = \boxed{(A^T + A)x}$$

- $$AA^{-1} = I \xrightarrow{\partial/\partial\beta} \frac{\partial A}{\partial \beta}A^{-1} + A\frac{\partial A^{-1}}{\partial\beta} = 0 \to A\frac{\partial A^{-1}}{\partial \beta} = -\frac{\partial A}{\partial\beta}A^{-1}$$

$$\xrightarrow{\times A^{-1}} \boxed{\frac{\partial A^{-1}}{\partial\beta} = -A^{-1}\frac{\partial A}{\partial\beta}A^{-1}}$$

- $\nabla_A |A| \overset{?}{=} |A| A^{-T}$

ابتدا ماتریس cofactor ماتریس A را به این صورت تعریف می‌کنیم:

$$C = [c_{ij}], \quad c_{ij} = (-1)^{i+j}\det A(i|j)$$

که منظور از $A(i|j)$ ماتریسی حاصل از در انداختن سطر i و ستون j ماتریس A است. طبق تعریف، ضرب داخلی ستون j ام ماتریس A و ماتریس C برابر با $|A|$ است. نشان می‌دهیم ضرب داخلی ستون k ماتریس A و ستون j ماتریس C برابر با صفر است (زمانی که $k \neq j$)

ماتریس B را به این صورت تعریف می‌کنیم:

$$B = \begin{bmatrix} a_{11} & \cdots & a_{1k} & \cdots & a_{1k} & \cdots & a_{1n} \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{n1} & \cdots & a_{nk} & \cdots & a_{nk} & \cdots & a_{nn} \end{bmatrix}$$

ستون j ام

یعنی ماتریس B برابر با ماتریس A است که ستون j ام آن را با ستون k ام ماتریس A جایگذاری کرده ایم.

با توجه به اینکه B دو ستون تکراری دارد، پس $\det B = 0$. از طرف دیگر دترمینان B را می‌توان به کمک گسترش برحسب ~~سطر~~ ستون j اُم اش نوشت:

$$\det B = \sum_i B_{ij} C_{ij} = \sum_i A_{ik} C_{ij} = 0 \quad (k \neq j)$$

و همانطور که گفته شد $\sum_i A_{ij} C_{ij} = |A|$ ، پس:

ستون j اُم cofactor(A) با ستون j اُم C یکی است

$$C^T A = |A| I \Rightarrow \frac{C^T}{|A|} A = I \Rightarrow A^{-1} = \frac{C^T}{|A|} \Rightarrow C^T = |A| A^{-1} \Rightarrow \boxed{C = |A| A^{-T}} \; (*)$$

حال به کمک این ماتریس، مشتق خواسته شده را به دست می‌آوریم:

$$\left[\frac{\partial |A|}{\partial A}\right]_{ij} = \frac{\partial |A|}{\partial A_{ij}} = C_{ij}$$

دقت شود چون در محاسبه C_{ij}، ستون j و سطر i ماتریس A حذف شده اند، C_{ij} تابعی از A_{ij} نیست.

$$\xrightarrow{\text{پس}} \frac{\partial |A|}{\partial A} = C \xrightarrow{(*)} \boxed{\frac{\partial |A|}{\partial A} = |A| A^{-T}}$$

(منظور از $\frac{\partial}{\partial A}$ همان ∇_A است)

$$\boxed{\nabla_A \log |A|} = \nabla_A |A| \times \frac{\partial \log z}{\partial z}\Big|_{z=|A|} = |A| A^{-T} \times \frac{1}{|A|} = \boxed{A^{-T}}$$

(الف) (چون مشتق نسبت به ماتریس است، ترانهاده اهمیت دارد!)

۴) فرض کنید ماتریس A ابعاد $m \times n$ داشته باشد و رنک آن r باشد. ← سوال ۳ صفحه بعد

(الف)

$$r = n \rightarrow m \geq n \rightarrow \Sigma = \begin{bmatrix} \sigma_1 & & \\ & \sigma_2 \ddots & \\ & & \sigma_n \\ \hline & 0 & \end{bmatrix}_{m \times n}, \quad A = U \Sigma V^T$$

حال سعی می‌کنیم از $A^+ = (A^T A)^{-1} A$ برسیم:

$$A^T A = V \Sigma^T \underbrace{U^T U}_{I} \Sigma V^T = V \Sigma^T \Sigma V^T = V \begin{bmatrix} \sigma_1^2 & & \\ & \sigma_2^2 \ddots & \\ & & \sigma_n^2 \end{bmatrix}_{n \times n} V^T$$

$$\rightarrow (A^T A)^{-1} = (V^T)^{-1} (\Sigma^T \Sigma)^{-1} V^{-1} = V \begin{bmatrix} \frac{1}{\sigma_1^2} & & \\ & \ddots & \\ & & \frac{1}{\sigma_n^2} \end{bmatrix}_{n \times n} V^T$$

$$\rightarrow \boxed{(A^T A)^{-1} A^T} = V \begin{bmatrix} \frac{1}{\sigma_1^2} & & \\ & \ddots & \\ & & \frac{1}{\sigma_n^2} \end{bmatrix} \underbrace{V^T V}_{I} \Sigma^T U^T = V \begin{bmatrix} \frac{1}{\sigma_1^2} & & \\ & \ddots & \\ & & \frac{1}{\sigma_n^2} \end{bmatrix} \left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & 0 \\ & & \sigma_n & \end{array}\right] U^T$$

$$= V \left[\begin{array}{ccc|c} \frac{1}{\sigma_1} & & & \\ & \frac{1}{\sigma_2} \ddots & & 0 \\ & & \frac{1}{\sigma_n} & \end{array}\right]_{n \times m} U^T = V \Sigma^+ U^T = \boxed{A^+}$$

منظور از Σ^+ همان Σ^{-1} در صورت سوال است.

ب) به طور مشابه ثابت می‌شود.

$$\Sigma = \left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \sigma_2 \ddots & & 0 \\ & & \sigma_m & \end{array}\right]_{m \times n} \Leftarrow (r = m ,\; n > m)$$

$$A A^T = U \Sigma \underbrace{V^T V}_{I} \Sigma^T U^T = U \left[\begin{array}{ccc|c} \sigma_1 & & & \\ & \ddots & & 0 \\ & & \sigma_m & \end{array}\right]_{m \times n} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_m \\ \hline & 0 & \end{bmatrix}_{n \times m} U^T$$

$$\Rightarrow A A^T = U \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_m^2 \end{bmatrix} U^T \Rightarrow (A A^T)^{-1} = U \begin{bmatrix} \frac{1}{\sigma_1^2} & & \\ & \ddots & \\ & & \frac{1}{\sigma_m^2} \end{bmatrix} U^T$$

$$\Rightarrow \boxed{A^T (A A^T)^{-1}} = V \Sigma^T \underbrace{U^T U}_{I} \begin{bmatrix} \frac{1}{\sigma_1^2} & & \\ & \ddots & \\ & & \frac{1}{\sigma_m^2} \end{bmatrix} U^T = V \begin{bmatrix} \frac{1}{\sigma_1} & & \\ & \ddots & \\ & & \frac{1}{\sigma_m} \\ \hline & 0 & \end{bmatrix} U^T = V \Sigma^+ U^T = \boxed{A^+}$$

سوال ۲) الف)

$$\lambda_1+\lambda_2+\dots+\lambda_n=\operatorname{trace}(A)$$

می خواهیم ثابت کنیم جمع عناصر روی قطر ماتریس A با جمع مقدار ویژه‌هایش برابر است.

$$A=\begin{bmatrix} a_{11} & & & \\ & a_{22} & & \\ & & \ddots & \\ & & & a_{nn}\end{bmatrix} \rightarrow \operatorname{trace}(A)=\sum_{i=1}^{n} a_{ii}$$

می دانیم مقادیر ویژه‌ی A یعنی $\lambda_1, \dots, \lambda_n$، ریشه‌های $\det(A-\lambda I)$ هستند. (که چندجمله‌ای حاصل از $\det(A-\lambda I)$ برحسب λ را چندجمله‌ای مشخصه می نامیم).

که این چندجمله‌ای به صورت روبه‌رو خواهد بود:

$$\det(A-\lambda I)=(-1)^n\lambda^n+c_{n-1}\lambda^{n-1}+\dots+c_1\lambda+c_0$$

این چندجمله‌ای از درجه‌ی n خواهد بود و ضریب λ^n برابر با $(-1)^n$ می باشد؛

(زیرا در
$$\begin{vmatrix} a_{11}-\lambda & & * \\ & a_{22}-\lambda & \\ * & & a_{nn}-\lambda \end{vmatrix}$$
تنها در صورتی که قطر اصلی را به عنوان یک قطر پراکنده در نظر بگیریم، λ^n خواهیم داشت و با توجه به منفی بودن ضریب λ در قطر اصلی، ضریب λ^n برابر با $(-1)^n$ خواهد بود).

با توجه به اینکه $\lambda_1, \dots, \lambda_n$ مقادیر ویژه‌ی A و ریشه‌های $\det(A-\lambda I)$ هستند، پس می توان این چندجمله‌ای را به صورت زیر تجزیه کرد:

$$\det(A-\lambda I)=(\lambda_1-\lambda)\dots(\lambda_n-\lambda)=(-1)^n\lambda^n+c_{n-1}\lambda^{n-1}+\dots+c_1\lambda+c_0$$

حال به کمک این تجزیه، $\boxed{\text{ضریب } \lambda^{n-1}}$ ← کلید اثبات، یعنی c_{n-1} را می یابیم:

$$c_{n-1}\lambda^{n-1}=\lambda_1(-1)^{n-1}\lambda^{n-1}+\dots+\lambda_n(-1)^{n-1}\lambda^{n-1}=(-1)^{n-1}\lambda^{n-1}\sum_{i=1}^{n}\lambda_i$$

$$\rightarrow \boxed{c_{n-1}=(-1)^{n-1}\sum_{i=1}^{n}\lambda_i} \quad ①$$

Scanned with CamScanner

از طرفی در $\det(A-\lambda I) = \begin{vmatrix} a_{11}-\lambda & & * \\ & a_{22}-\lambda & \\ * & & \ddots \\ & & & a_{nn}-\lambda \end{vmatrix}$ تنها در حالتی λ^{n-1} خواهیم داشت که قطر برگزیده را قطر اصلی در نظر بگیریم. بنابراین باید ضریب λ^{n-1} را در حاصل ضرب زیر بیابیم:

$$(a_{11}-\lambda)(a_{22}-\lambda)\cdots(a_{nn}-\lambda)$$

$$c_{n-1}\lambda^{n-1} = a_{11}(-1)^{n-1}\lambda^{n-1} + a_{22}(-1)^{n-1}\lambda^{n-1} + \cdots + a_{nn}(-1)^{n-1}\lambda^{n-1}$$

ضریب λ^{n-1}

$$= (-1)^{n-1}\lambda^{n-1}\sum_{i=1}^{n} a_{ii}$$

$$\rightarrow \boxed{c_{n-1} = (-1)^{n-1}\sum_{i=1}^{n} a_{ii}} \quad (2)$$

$$\xrightarrow{(1\text{ و }2)} (-1)^{n-1}\sum_{i=1}^{n}\lambda_i = (-1)^{n-1}\sum_{i=1}^{n} a_{ii} \Rightarrow \boxed{\sum_{i=1}^{n}\lambda_i = \sum_{i=1}^{n} a_{ii} = \mathrm{trace}(A)}$$

حکم ثابت شد.

ب)

$$\lambda_1\lambda_2\cdots\lambda_n = \det(A)$$

اثبات این مورد به کمک مواردی که در بخش الف گفته شد ساده است.
می دانیم مقادیر ویژه‌ی A ریشه‌های چندجمله‌ای $\det(A-\lambda I)$ هستند که این چندجمله‌ای را می توان به صورت زیر تجزیه کرد:

$$\det(A-\lambda I) = (\lambda_1-\lambda)(\lambda_2-\lambda)\cdots(\lambda_n-\lambda)$$

λ را برابر با صفر قرار می دهیم (λ متغیر چندجمله‌ای است و مجاز به این کار هستیم):

$$\xrightarrow{\lambda=0} \det(A-0\times I) = (\lambda_1-0)\cdots(\lambda_n-0)$$

$$\rightarrow \boxed{\det(A) = \lambda_1\lambda_2\cdots\lambda_n}$$

ثابت شد.

سوال ۵. الف)

$$M = \begin{bmatrix} A_{n\times n} & B_{n\times k} \\ C_{k\times n} & D_{k\times k} \end{bmatrix}, \quad A \text{ is invertible}$$

با ابعاد k در n و n ابعادند

$$\begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix}\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} A & B \\ 0 & D-CA^{-1}B \end{bmatrix} \xrightarrow{\times \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix}} = \begin{bmatrix} A & 0 \\ 0 & D-CA^{-1}B \end{bmatrix}$$

تا همینجا برای حل سوال کافی است. اما برای بخش‌های بعدی نیاز است کمی بیشتر پیش برویم.

وارون‌ها در ج) ثابت می‌شوند $\Longrightarrow$

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix}\begin{bmatrix} A & 0 \\ 0 & D-CA^{-1}B \end{bmatrix}\begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}$$

بنابراین به تجزیه‌ی LDU ماتریس M دست یافتیم. اکنون می‌توانیم خواسته‌ی سوال را به دست بیاوریم.
همانطور که از جبرخطی می‌دانیم، دترمینان ماتریس‌های سمت راست به راحتی برحسب دترمینان بلوک‌های آن‌ها به دست می‌آید (اثبات آن‌ها به کمک تعریف عملگر n خطی و مفهوم قطر پرکننده انجام می‌شود ولی به دلیل طولانی بودن از آوردن آن خودداری می‌کنیم.)

$$\det M = \det\begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix} \det\begin{bmatrix} A & 0 \\ 0 & D-CA^{-1}B \end{bmatrix} \det\begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}$$

$$\boxed{\det M = 1 \times \det A \times \det(D-CA^{-1}B) \times 1}$$

ب) مشابه بخش الف ثابت می‌کنیم:

$$\begin{bmatrix} I & -BD^{-1} \\ 0 & I \end{bmatrix}\begin{bmatrix} A & B \\ C & D \end{bmatrix}\begin{bmatrix} I & 0 \\ -D^{-1}C & I \end{bmatrix} = \begin{bmatrix} I & -BD^{-1} \\ 0 & I \end{bmatrix}\begin{bmatrix} A-BD^{-1}C & B \\ 0 & D \end{bmatrix} = \begin{bmatrix} A-BD^{-1}C & 0 \\ 0 & D \end{bmatrix}$$

$$\xrightarrow{\text{ضرب طرفین}} M = \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & BD^{-1} \\ 0 & I \end{bmatrix}\begin{bmatrix} A-BD^{-1}C & 0 \\ 0 & D \end{bmatrix}\begin{bmatrix} I & 0 \\ D^{-1}C & I \end{bmatrix}$$

بنابراین:

$$\boxed{\det M = 1 \times \det(A-BD^{-1}C)\det(D) \times 1}$$

ج) ابتدا روش وارون کردن در این ماتریس‌ها را نشان می‌دهیم.

$$\begin{bmatrix} I & 0 \\ X & I \end{bmatrix}^{-1} = \begin{bmatrix} I & 0 \\ -X & I \end{bmatrix} \xrightarrow{\text{Proof}} \begin{bmatrix} I & 0 \\ X & I \end{bmatrix}\begin{bmatrix} I & 0 \\ -X & I \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} = I \quad \text{و} \quad \begin{bmatrix} I & X \\ 0 & I \end{bmatrix}^{-1} = \begin{bmatrix} I & -X \\ 0 & I \end{bmatrix}$$ به طور مشابه

$$\begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix}^{-1} = \begin{bmatrix} X^{-1} & 0 \\ 0 & Y^{-1} \end{bmatrix} \xrightarrow{\text{Proof}} \begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix}\begin{bmatrix} X^{-1} & 0 \\ 0 & Y^{-1} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} = I$$

حال از الف و ب داریم:

$$M = \begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix}\begin{bmatrix} A & 0 \\ 0 & \underbrace{D-CA^{-1}B}_{\triangleq S_A \text{ به اختصار}} \end{bmatrix}\begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix} = \begin{bmatrix} I & BD^{-1} \\ 0 & I \end{bmatrix}\begin{bmatrix} \overbrace{A-BD^{-1}C}^{S_D \text{ به اختصار}} & 0 \\ 0 & D \end{bmatrix}\begin{bmatrix} I & 0 \\ D^{-1}C & I \end{bmatrix}$$

$$\xrightarrow{\text{وارون}} M^{-1} = \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix}\begin{bmatrix} A^{-1} & 0 \\ 0 & S_A^{-1} \end{bmatrix}\begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix} = \begin{bmatrix} I & 0 \\ -D^{-1}C & I \end{bmatrix}\begin{bmatrix} S_D^{-1} & 0 \\ 0 & D^{-1} \end{bmatrix}\begin{bmatrix} I & -BD^{-1} \\ 0 & I \end{bmatrix}$$

حال دو طرف را محاسبه می‌کنیم و برابر قرار می‌دهیم:

$$\Rightarrow M^{-1} = \begin{bmatrix} A^{-1}+A^{-1}BS_A^{-1}CA^{-1} & -A^{-1}BS_A^{-1} \\ -S_A^{-1}CA^{-1} & S_A^{-1} \end{bmatrix} = \begin{bmatrix} S_D^{-1} & -S_D^{-1}BD^{-1} \\ -D^{-1}CS_D^{-1} & D^{-1}+D^{-1}CS_D^{-1}BD^{-1} \end{bmatrix}$$

بنابراین،

$$S_D^{-1} = (A-BD^{-1}C)^{-1} = A^{-1}+A^{-1}BS_ACA^{-1} = A^{-1}+A^{-1}B(D-CA^{-1}B)CA^{-1}$$

با تبدیل $C \to -C$ نیز تساوی برقرار است:

$$\boxed{(A+BD^{-1}C)^{-1} = A^{-1}-A^{-1}B(D+CA^{-1}B)CA^{-1}}$$

د) اگر ماتریس A، $n\times n$ باشد، ماتریس $(n+1)\times(n+1)$ M را به صورت زیر تعریف می‌کنیم:

$$M = \left[\begin{array}{c|c} A & u \\ \hline v^T & 1 \end{array}\right] \Rightarrow \det M = \det(A)\times\det(1-v^TA^{-1}u) \quad \text{(بخش الف)}$$
$$= \det[1]\times\det(A-u\times 1\times v^T) \quad \text{(بخش ب)}$$

$$\Rightarrow \boxed{\det(A-uv^T) = \det(A)\det(1-v^Tu)}$$

سوال ۶) الف:

$$A = \left[\begin{array}{c|c} B & x \\ \hline y^H & a \end{array}\right] \quad \boxed{P_A(t) = \det(tI-A)} = \det\begin{bmatrix} tI-B & -x \\ -y^H & t-a \end{bmatrix}$$

det برابر با خودش که عدد است

سوال ۵ ←

$$= \det(tI-B)\times\det\left((t-a)-y^H(tI-B)^{-1}x\right)$$

$\det(A)\times A^{-1} = adj(A)$

$$= (t-a)\det(tI-B) - y^H\times\det(tI-B)\times(tI-B)^{-1}\times x$$
$$\boxed{= (t-a)P_B(t) - y^H\times adj(tI-B)\times x}$$

ب) ابتدا نامساوی Rayleigh را ثابت می‌کنیم:

$$\lambda_{min} \le x^HAx \le \lambda_{max}$$
$$s.t.\; x^Hx=1$$

از روش ضرایب لاگرانژ استفاده می‌کنیم:

$$\mathcal{L}(x,\lambda) = x^HAx + (1-x^Hx) \longrightarrow \min_{x,\lambda}\mathcal{L}(x,\lambda)=?$$

$$\xrightarrow{\partial/\partial x} 2Ax - 2\lambda x = 0 \to Ax = \lambda x \to$$ x باید بردار ویژه A باشد.

اگر x بردار ویژه A و λ مقدار ویژه مربوط به آن باشد، داریم،

$$x^HAx = x^H\lambda x = \lambda\times 1 = \lambda$$ که A هرمیتی است

بنابراین به ازای λ_{min} و بردار ویژه مربوط به آن کمینه می‌شود. به طور مشابه می‌توان نشان داد بیشینه مقدار x^HAx، λ_{max} است.

حال به حل سوال می‌پردازیم:

$$\lambda_i(A+B) \le \lambda_{i+j}(A)+\lambda_{n-j}(B), \quad i\in\{1,\dots,n\},\; j\in\{0,\dots,n-i\}$$

فرض کنید $\{x_1,\dots,x_n\}$، $\{y_1,\dots,y_n\}$ و $\{z_1,\dots,z_n\}$ مجموعه بردار ویژه‌های A، B، و $A+B$ باشند (به ترتیب)

به طوری که:

$$Ax_i = \lambda_i(A)x_i, \quad By_i = \lambda_i(B)y_i, \quad (A+B)z_i = \lambda_i(A+B)z_i$$

به ازای i و j داده شده، S_1 و S_2 و S_3 را به این صورت در نظر می‌گیریم:

$$S_1 = span\{x_1,\dots,x_{i+j}\}$$
$$S_2 = span\{y_1,\dots,y_{n-j}\}$$
$$S_3 = span\{z_i,\dots,z_n\}$$

در این صورت داریم:

$$\dim S_1 + \dim S_2 + \dim S_3 = (i+j)+(n-j)+(n-i+1) = 2n+1$$

با توجه به این که بعد فضا n است، این موضوع نشان می‌دهد که یک بردار مشترک در هر سه زیر فضا وجود دارد. یعنی:

$$\exists x\in S_1\cap S_2\cap S_3$$

و فرض می‌کنیم x یکه است (با توجه به اینکه x و kx در زیرفضا وجود دارند، x را می‌توان یکه کرد.)

حال از نامساوی ریلی ثابت شده استفاده می‌کنیم:

$$\lambda_i(A+B) \leq x^H(A+B)x = x^H A x + x^H B x$$
$$\leq \lambda_{i+j}(A) + \lambda_{n-j}(B)$$

دقت کنید $\lambda_i(A+B)$، کوچک‌ترین مقدار ویژه برای بردار ویژه‌های موجود در زیرفضای S_3 است و به طور مشابه، $\lambda_{i+j}(A)$ و $\lambda_{n-j}(B)$ نیز برای زیرفضاهای S_1 و S_2 بزرگترین مقدار ویژه‌ها هستند و در شرایط نامساوی صدق می‌کنند. بنابراین حکم سوال ثابت شد.

ج) اثبات این بخش به کمک بخش الف انجام می‌شود و نیازی به بخش ب نیست!

$$A = \begin{bmatrix} B & y \\ y^H & a \end{bmatrix} \longrightarrow P_A(t) = \det(tI - A) = (t-a)P_B(t) - y^H \operatorname{adj}(tI-B)y$$
$$= (t-a)P_B(t) - y^H \underbrace{\det(tI-B)}_{P_B(t)}(tI-B)^{-1}y$$

$$\Rightarrow P_A(t) = P_B(t)\left[t-a - y^H(tI-B)^{-1}y\right]$$

اگر Λ_A را مجموعه مقدار ویژه‌های A تعریف کنیم.

$$\lambda \in \Lambda_B \rightarrow P_B(\lambda) = 0 \Rightarrow P_A(\lambda) = 0 \rightarrow \lambda \in \Lambda_A \Rightarrow \Lambda_B \subseteq \Lambda_A$$

این به این معنی است که همه‌ی مقادیر ویژه‌ی B، مقادیر ویژه‌ی A نیز هستند. با توجه به این که $P_B(t)$ به صورت یک فاکتور در عبارت $P_A(t)$ در آمده است نیز می‌توان گفت تعداد تکرار آن‌ها نیز یکسان است (یا در A یکی‌شان تکرارش یک واحد بیشتر است) و از $n+1$ مقدار ویژه‌ی A، n تای آن‌ها در B نیز هستند و تنها یک مقدار ویژه دیگر در فاکتور دیگر حضور دارد. اگر این بردار ویژه $\lambda_k(A)$ باشد خواهیم داشت:

$$\lambda_1(A) = \lambda_1(B) \leq \lambda_2(A) = \lambda_2(B) \leq \cdots \leq \lambda_{k-1}(A) = \lambda_{k-1}(B) \leq \lambda_k(A) \leq \lambda_k(B) = \lambda_{k+1}(A)$$
$$\leq \cdots \leq \lambda_n(B) = \lambda_{n+1}(A)$$

یعنی فرقی نمی‌کند مقدار ویژه‌ی جدید A در ابتدا، انتها یا وسط باشد. در هر صورت عبارت سوال صحیح است.

~~$p(x) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp($~~

Scanned with CamScanner

سوال ۷:

$D = \{x_1, \dots, x_n\}$

$$\mathcal{L} = p(D|\theta) \Rightarrow \max_\theta \mathcal{L} = \max_\theta p(D|\theta) \equiv \max_\theta \log p(D|\theta)$$

$$= \max_\theta \sum_i -2\log\theta + \log x_i - \frac{x_i}{\theta}$$

$$\rightarrow \frac{\partial \mathcal{L}}{\partial \theta} = 0 \Rightarrow \sum_i \frac{-2}{\theta} + \frac{+x_i}{\theta^2} = \sum_i \frac{x_i - 2\theta}{\theta^2} = 0 \rightarrow \sum_i x_i = \sum_i 2\theta = 2n\theta$$

$$\Rightarrow \boxed{\hat{\theta} = \frac{1}{2n}\sum_i x_i}$$ تخمین بیشینه درستنمایی

سوال ۸:

$\cancel{MAP}$, MLE: $\max_\mu \mathcal{L} = \max_\mu p(D|\mu) \equiv \max_\mu \log p(D|\mu)$

$$= \max_\mu \sum_i \frac{-1}{2}\log(2\pi\sigma^2) + \frac{-(x_i-\mu)^2}{2\sigma^2}$$

$$\xrightarrow{\partial/\partial\mu = 0} \sum_i 0 + \frac{x_i - \mu}{\sigma^2} = 0 \Rightarrow \sum x_i - \mu = 0 \Rightarrow \hat{\mu}_{MLE} = \frac{1}{n}\sum_{i=1}^n x_i = \bar{x}$$

با میل دادن n به بی‌نهایت، طبق قانون اعداد بزرگ، تخمین‌گر برابر با میانگین واقعی توزیع می‌شود.

$$\lim_{n\to\infty} \hat{\mu}_{MLE} = \mu$$

MAP: $\max_\mu p(\mu)\, p(D|\mu) \equiv \max_\mu \log p(\mu) + \log p(D|\mu)$

$$= \max_\mu \frac{-1}{2}\log(2\pi\beta^2) + \frac{-(\mu-\gamma)^2}{2\beta^2} + \sum_i \frac{-1}{2}\log(2\pi\sigma^2) + \frac{-(x_i-\mu)^2}{2\sigma^2}$$

$$\xrightarrow{\partial/\partial\mu = 0} 0 + \frac{-(\mu-\gamma)}{\beta^2} + \sum_i 0 + \frac{-(\mu - x_i)}{\sigma^2} = 0$$

$$\Rightarrow \mu\left(\frac{1}{\beta^2} + \frac{n}{\sigma^2}\right) = \frac{\gamma}{\beta^2} + \frac{1}{\sigma^2}\sum_i x_i = \frac{\gamma}{\beta^2} + \frac{n\bar{x}}{\sigma^2}$$

$$\Rightarrow \hat{\mu}_{MAP} = \frac{\frac{\gamma}{\beta^2} + \frac{n\bar{x}}{\sigma^2}}{\frac{1}{\beta^2} + \frac{n}{\sigma^2}} = \boxed{\frac{\gamma\sigma^2 + \beta^2 n\bar{x}}{\sigma^2 + n\beta^2}}$$

$$\lim_{n\to\infty} \hat{\mu}_{MAP} = \lim_{n\to\infty} \frac{\beta^2 n\bar{x}}{n\beta^2} = \bar{x}$$

پس در بی‌نهایت $\hat{\mu}_{MAP} = \bar{x}$ می‌شود

و مشابه قبل، طبق قانون اعداد بزرگ برابر با میانگین واقعی توزیع می‌شود (همین طور هم انتظار می‌رفت که به ازای $n \to \infty$، داده‌ها بسیار مهم‌تر از prior شوند و $\hat{\mu}_{MLE} = \hat{\mu}_{MAP}$ شود)

سوال ۹) ابتدا ثابت می‌کنیم اگر x نرمال باشد $x \sim \mathcal{N}(\mu, \Sigma)$ آن‌گاه x_b و x_a نیز توزیع نرمال‌هایی به صورت زیر هستند:

$$x_a \sim \mathcal{N}(\mu_a, \Sigma_{aa}), \quad x_b \sim \mathcal{N}(\mu_b, \Sigma_{bb}) \qquad x = \begin{bmatrix} x_a \\ x_b \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix}$$

قضیه تبدیل خطی توزیع نرمال: اگر $x \sim \mathcal{N}(\mu, \Sigma)$ از توزیع نرمال پیروی کند، آن‌گاه هر تبدیل خطی آن هم یک توزیع نرمال است:

$$y = Ax + b \sim \mathcal{N}(A\mu + b, A\Sigma A^T)$$

اثبات: تابع مولد گشتاور یک متغیر تصادفی به این صورت تعریف می‌شود:

$$M_x(t) = \mathbb{E}(\exp[t^T x])$$

تابع مولد گشتاور برای متغیر تصادفی نرمال $x \sim \mathcal{N}(\mu, \Sigma)$ به این صورت است:

$$M_x(t) = \exp\left[t^T\mu + \frac{1}{2} t^T \Sigma t\right]$$

بنابراین تابع مولد گشتاور متغیر y برابر است با:

$$M_y(t) = \mathbb{E}\left[\exp\left[t^T(Ax+b)\right]\right]$$

$$= \mathbb{E}\left(\exp(t^T A x) \cdot \exp(t^T b)\right) = \exp(t^T b)\, M_x(A^T t)$$

$$= \exp(t^T b)\, \exp\left(t^T A\mu + \frac{1}{2} t^T A \Sigma A^T t\right)$$

$$= \exp\left(t^T(A\mu + b) + \frac{1}{2} t^T A\Sigma A^T t\right)$$

بنابراین y یک متغیر تصادفی گاوسی با توزیع $\mathcal{N}(A\mu + b, A\Sigma A^T)$ است! Q.E.D

پس گاوسی بودن و پارامترهای توزیع‌های x_a و x_b به راحتی به دست می‌آید:

$$x_a = [I_a \;\; 0_b]\, x$$

$$x_b = [0_a \;\; I_b]\, x \implies \mathbb{E}[x_a] = [I \;\; 0]\mu = \mu_a, \quad \mathrm{Cov}[x_a] = [I \;\; 0]\begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix}\begin{bmatrix} I \\ 0 \end{bmatrix} = \Sigma_{aa}$$

به طور مشابه: $\mathbb{E}[x_b] = [0 \;\; I]\mu = \mu_b$, $\mathrm{Cov}[x_b] = [0 \;\; I]\,\Sigma \begin{bmatrix} 0 \\ I \end{bmatrix} = \Sigma_{bb}$

$$\implies x_a \sim \mathcal{N}(\mu_a, \Sigma_{aa}) \quad \text{و} \quad x_b \sim \mathcal{N}(\mu_b, \Sigma_{bb})$$

(البته برای x_a و x_b تقریباً بدیهی بود و نیازی به اثبات قضیه نبود! اما چون درباره x_a در بخش الف بررسی شده اثبات کردم.)

الان به اثبات توزیع شرطی می‌پردازیم. در این سوال از سوال ۵ کمک می‌گیریم:

$$\Sigma = \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix}$$

به آن مکمل شور می‌گوییم

$$(\Sigma/\Sigma_{bb}) \triangleq \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba} \implies \boxed{\det(\Sigma) = \det(\Sigma_{bb})\, \det(\Sigma/\Sigma_{bb})}$$ سوال ۵

$$p(x) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right) \quad \text{و} \quad \boxed{d = a + b}$$

برای راحتی، z و z_a و z_b را به این صورت تعریف می‌کنیم،
و الان عبارت توان $p(x)$ را محاسبه می‌کنیم (بدون $-\frac{1}{2}$)

$$z \triangleq x - \mu = \begin{bmatrix} x_a - \mu_a \\ x_b - \mu_b \end{bmatrix} = \begin{bmatrix} z_a \\ z_b \end{bmatrix}$$

$$z^T \Sigma^{-1} z = [z_a^T \;\; z_b^T] \underbrace{\begin{bmatrix} I & 0 \\ -\Sigma_{bb}^{-1}\Sigma_{ba} & I \end{bmatrix}\begin{bmatrix} (\Sigma/\Sigma_{bb})^{-1} & 0 \\ 0 & \Sigma_{bb}^{-1} \end{bmatrix}\begin{bmatrix} I & -\Sigma_{ab}\Sigma_{bb}^{-1} \\ 0 & I \end{bmatrix}}_{\Sigma^{-1}} \begin{bmatrix} z_a \\ z_b \end{bmatrix}$$

که این تجزیه از اثبات سوال ۵ آمده است.

$$= [z_a^T - z_b^T \Sigma_{bb}^{-1}\Sigma_{ba} \quad z_b^T] \begin{bmatrix} (\Sigma/\Sigma_{bb})^{-1} & 0 \\ 0 & \Sigma_{bb}^{-1} \end{bmatrix} \begin{bmatrix} z_a - \Sigma_{ab}\Sigma_{bb}^{-1} z_b \\ z_b \end{bmatrix}$$

$$\Rightarrow z^T \Sigma^{-1} z = \begin{bmatrix} z_a^T (\Sigma/\Sigma_{bb})^{-1} - z_b^T \Sigma_{bb}^{-1} \Sigma_{ba} (\Sigma/\Sigma_{bb})^{-1} & z_b^T \Sigma_{bb}^{-1} \end{bmatrix} \begin{bmatrix} z_a - \Sigma_{ab} \Sigma_{bb}^{-1} z_b \\ z_b \end{bmatrix}$$

$$= z_b^T \Sigma_{bb}^{-1} z_b + \left(z_a^T (\Sigma/\Sigma_{bb})^{-1} - z_b^T \Sigma_{bb}^{-1} \Sigma_{ba} (\Sigma/\Sigma_{bb})^{-1} \right) \left(z_a - \Sigma_{ab} \Sigma_{bb}^{-1} z_b \right)$$

$$= z_b^T \Sigma_{bb}^{-1} z_b + \left(z_a^T - z_b^T \Sigma_{bb}^{-1} \Sigma_{ba} \right) (\Sigma/\Sigma_{bb})^{-1} \left(z_a - \Sigma_{ab} \Sigma_{bb}^{-1} z_b \right)$$

$$= z_b^T \Sigma_{bb}^{-1} z_b + \left(z_a - \Sigma_{ab} \Sigma_{bb}^{-1} z_b \right)^T (\Sigma/\Sigma_{bb})^{-1} \left(z_a - \Sigma_{ab} \Sigma_{bb}^{-1} z_b \right)$$

Σ is symmetric $\rightarrow \Sigma_{ba}^T = \Sigma_{ab}$

حال $p(x_a|x_b)$ را محاسبه می‌کنیم.

$$P(x_a|x_b) = \frac{P(x_a, x_b)}{P(x_b)} = \frac{P(x)}{P(x_b)}$$

$$= \frac{1}{\sqrt{|2\pi|^{d-b} \frac{|\Sigma|}{|\Sigma_{bb}|}}} \frac{\exp\left(\frac{-1}{2} \overbrace{(x-\mu)^T \Sigma^{-1} (x-\mu)}^{z^T \Sigma^{-1} z}\right)}{\exp\left(\frac{-1}{2} \underbrace{(x_b-\mu_b)^T \Sigma_{bb}^{-1} (x_b-\mu_b)}_{z_b^T \Sigma_{bb}^{-1} z_b}\right)}$$

$$= \frac{1}{\sqrt{|2\pi|^a \det(\Sigma/\Sigma_{bb})}} \times \exp\left(\frac{-1}{2} \left(z_a - \Sigma_{ab} \Sigma_{bb}^{-1} z_b\right)^T (\Sigma/\Sigma_{bb})^{-1} \left(z_a - \Sigma_{ab} \Sigma_{bb}^{-1} z_b\right)\right)$$

$$\overset{z = x-\mu}{=} \frac{1}{\sqrt{(2\pi)^a \det(\Sigma/\Sigma_{ab})}} \exp\left(\frac{-1}{2} \left(x_a - \mu_a - \Sigma_{ab} \Sigma_{bb}^{-1} (x_b - \mu_b)\right)^T (\Sigma/\Sigma_{bb})^{-1} \left(x_a - \mu_a - \Sigma_{ab} \Sigma_{bb}^{-1} (x_b - \mu_b)\right)\right)$$

بنابراین:

$$\boxed{x_a|x_b \sim \mathcal{N}(\mu_{a|b}, \Sigma_{a|b}) \quad , \quad \mu_{a|b} = \mu_a + \Sigma_{ab} \Sigma_{bb}^{-1} (x_b - \mu_b) \\ \Sigma_{a|b} = \Sigma/\Sigma_{bb} = \Sigma_{aa} - \Sigma_{ab} \Sigma_{bb}^{-1} \Sigma_{ba}}$$

بخش بعد در ابتدای سوال اثبات شد. اما یک راه حل دیگر نیز برای به دست آوردن $\mathbb{E}[x_a]$ و $Cov(x_a)$ نیز به کمک $\mu_{a|b}$ و $Cov_{a|b}$ می‌توان ارائه داد. اگر x_a و x_b مستقل باشند، $x_a|x_b = x_a$ می‌شود. بنابراین اگر Σ_{ab} و Σ_{ba} را در $\mu_{a|b}$ و $\Sigma_{a|b}$ صفر قرار دهیم، به میانگین و کواریانس x_a می‌رسیم که همان μ_a و Σ_{aa} هستند!

سوال ۱۰) الف) مسأله را در دو حالت حل می‌کنیم. حالتی که $Ax=b$ پاسخ داشته باشد (یعنی b در فضای ستونی A باشد) و حالتی که $Ax=b$ پاسخ نداشته باشد ($b\notin C(A)$)

(۱) $b\in C(A)$: در این حالت کافی است ثابت کنیم در بین xهایی که $Ax=b$، $x^*=A^+b$ کمینه نرم را دارد.

فرض کنید x یک پاسخ دلخواه از معادله باشد. نشان می‌دهیم نرم آن بزرگ‌تر (یا مساوی) نرم x^* خواهد بود:

$$\left.\begin{array}{l} Ax=b \\ Ax^*=b \end{array}\right\} \Rightarrow A(x-x^*)=0 \Rightarrow x-x^*\in N(A)$$

حال نشان می‌دهیم به ازای هر $z\in N(A)$، $\langle z,x^*\rangle$ برابر با صفر است.

$$\langle z,x^*\rangle = z^TA^+b$$

$$A^+=V\Sigma^+U^T=V\left[\begin{array}{ccc|c} \frac{1}{\sigma_1} & & & \\ & \ddots & & 0 \\ & & \frac{1}{\sigma_r} & \\ \hline & 0 & & 0 \end{array}\right]U^T=\frac{1}{\sigma_1}v_1u_1^T+\frac{1}{\sigma_2}v_2u_2^T+\dots+\frac{1}{\sigma_r}v_ru_r^T$$

رنک A

$$\Rightarrow z^TA^+b=\frac{1}{\sigma_1}\langle z,v_1\rangle\langle u_1,b\rangle+\frac{1}{\sigma_2}\langle z,v_2\rangle\langle u_2,b\rangle+\dots+\frac{1}{\sigma_r}\langle z,v_r\rangle\langle u_r,b\rangle$$

دقت کنید که v_1 تا v_r، بردار ویژه‌های (مربوط به مقدار ویژه‌های ناصفر) A^TA هستند؛ بنابراین یک پایه‌ی متعامد برای فضای سطری A^TA هستند و چون فضای سطری A^TA و A یکسان است، پس پایه‌ای برای فضای سطری A نیز هستند. همچنین $\underbrace{C(A^T)}_{\text{فضای سطری } A}=N(A)^\perp$. بنابراین v_1 تا v_r همگی بر z که از $N(A)$ است عمودند.

پس:

$$\langle z,x^*\rangle=z^TA^+b=0 \qquad (\forall z\in N(A))$$

حال حکم نهایی به راحتی ثابت می‌شود:

$$x-x^*\in N(A)\Rightarrow \langle x-x^*,x^*\rangle=0$$

$$\|x\|^2=\|x-x^*+x^*\|^2=\underbrace{\|x-x^*\|^2}_{\geq 0}+\|x^*\|^2+2\overbrace{\langle x-x^*,x^*\rangle}^{\text{صفر}}$$

$$\Rightarrow \|x^*\|^2\leq\|x\|^2\Rightarrow\|x^*\|\leq\|x\|\Rightarrow$$

پس $\|x^*\|$ کمینه نرم را در بین تمام پاسخ‌ها دارد

(۲) $b\notin C(A)$ و در نتیجه $Ax=b$ پاسخ ندارد.

در این صورت کمینه مقدار $\|Ax-b\|$ در حالتی خواهد بود که Ax برابر با projection بردار b در فضای ستونی A باشد. چنین برداری را b^* می‌نامیم و ثابت می‌کنیم $b^*=AA^+b$ و x کمینه، پاسخ بهینه معادله $Ax=b^*$ خواهد بود.

$$AA^+b=U\Sigma\underbrace{V^TV}_{I}\Sigma^+U^Tb=U\left[\begin{array}{c|c} \overbrace{\begin{matrix}1 & & \\ & \ddots & \\ & & 1\end{matrix}}^{r} & 0 \\ \hline 0 & 0 \end{array}\right]U^Tb=(u_1u_1^T+u_2u_2^T+\dots+u_ru_r^T)b$$

$$=\langle u_1,b\rangle u_1+\langle u_2,b\rangle u_2+\dots+\langle u_r,b\rangle u_r$$

که این فرمول projection بردار b در فضای حاصل از $\text{span}\{u_1,\dots,u_r\}$ است و همانطور که می‌دانیم، این بردارها فضای ستونی A را span می‌کنند (مشابه v_iها که در بخش قبل نشان دادیم فضای سطری A را span می‌کنند، می‌توان نشان داد). بنابراین:

$$AA^+b=b^*$$

حال برای معادله جدید یعنی $Ax=b^*$، با توجه به این که $b^*\in C(A)$ است، از (۱) می‌توانیم x کمینه را بیابیم:

$$x^*=A^+b^*=A^+AA^+b$$

طبق تعریف، شبه وارون دو ویژگی هم دارد $\begin{cases}A^+AA^+=A^+\\ AA^+A=A\end{cases}$. بنابراین:

$$\boxed{x^*=A^+b}$$

پس با کنار هم قرار گرفتن (۱) و (۲)، حکم کلی ثابت می‌شود.

سوال ۱۰ : (ب)

$$x^{(t+1)} = x^{(t)} - vA^T(Ax^{(t)}-b) = \underbrace{(I - vA^TA)}_{M}x^t + \underbrace{vA^Tb}_{B}$$

$$\Rightarrow x^{(1)} = B \,,\quad x^{(2)} = Mx^{(1)} + B = MB + B \,,\quad x^{(3)} = M(MB+B) + B = (M^2 + M + I)B$$

$$\cdots \Rightarrow x^{(t+1)} = \left(\sum_{i=0}^{t} M^i\right)B \qquad [M^0 = I]$$

ما می‌خواهیم نشان دهیم x به A^+b همگرا می‌شود. پس $\lim_{t\to\infty} x^t$ را باید محاسبه کنیم. ماتریس M متقارن است. پس قطری می‌شود و

$$M = S^T\Lambda S \qquad \boxed{M^i = S^T\Lambda^i S}$$

$$\Rightarrow x^{(t+1)} = \left(\sum_{i=0}^{t} S^T\Lambda^i S\right)B = S^T\left(\sum_{i=0}^{t}\begin{bmatrix}\lambda_1^i & & & \\ & \ddots & & \\ & & \lambda_r^i & \\ & & & 0\end{bmatrix}\right)SB$$

r صفر

شبه‌وارون $I-\Lambda$ است

$$x^{(\infty)} = \lim_{t\to\infty} x^{(t+1)} = S^T\left(\sum_{i=0}^{\infty}\begin{bmatrix}\lambda_1^i & & & \\ & \ddots & & \\ & & \lambda_r^i & \\ & & & 0\end{bmatrix}\right)SB = S^T\begin{bmatrix}\frac{1}{1-\lambda_1} & & & \\ & \ddots & & \\ & & \frac{1}{1-\lambda_r} & \\ & & & 0\end{bmatrix}SB$$

$$= S^T(I-\Lambda)^+SB$$

$$\left.[S^T(I-\Lambda)^+S]^+ = S^T(I-\Lambda)S = I - S^T\Lambda S = I - M \Rightarrow S^T(I-\Lambda)^+S = (I-M)^+\right\} \Rightarrow$$

$$x^{(\infty)} = (I-M)^+B = (vA^TA)^+vA^Tb = (A^TA)^+A^Tb$$

چون درون شبه وارون است ساده می‌شوند

حال ثابت می‌کنیم $\boxed{(A^TA)^+A^T = A^+}$

$$A = U\Sigma V^T$$

$$A^TA = V\Sigma^TU^TU\Sigma V^T = V\Sigma^T\Sigma V^T \Rightarrow (A^TA)^+ = V(\Sigma^T\Sigma)^+V^T$$

$$\Rightarrow (A^TA)^+A^T = V(\Sigma^T\Sigma)^+V^TV\Sigma^TU^T = V(\Sigma^T\Sigma)^+\Sigma^TU^T$$

می‌توان نشان داد: $(\Sigma^T\Sigma)^+\Sigma^T = \Sigma^+$

$$\Sigma^T\Sigma = \begin{bmatrix}\sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots\ \sigma_r^2 & \\ & & & 0\end{bmatrix} \Rightarrow (\Sigma^T\Sigma)^+ = \begin{bmatrix}\frac{1}{\sigma_1^2} & & & \\ & \frac{1}{\sigma_2^2} & & \\ & & \ddots\ \frac{1}{\sigma_r^2} & \\ & & & 0\end{bmatrix} \Rightarrow (\Sigma^T\Sigma)^+\Sigma^T = \begin{bmatrix}\frac{1}{\sigma_1} & & & \\ & \frac{1}{\sigma_2} & & \\ & & \ddots\ \frac{1}{\sigma_r} & \\ & & & 0\end{bmatrix} = \Sigma^+$$

$$\Rightarrow (A^TA)^+A^T = V\Sigma^+U^T = A^+ \Longrightarrow \boxed{x^{(\infty)} = A^+b}$$

x به A^+b میل می‌کند.

(ج)

$$Ax^{(t+1)} = Ax^{(t)} - vAA^T(Ax^{(t)}-b) \Rightarrow Ax^{(t+1)} - b = Ax^{(t)} - b - vAA^T(Ax^{(t)}-b)$$

$$\Rightarrow (Ax^{(t+1)}-b) = \underbrace{(I - vAA^T)}_{B}\underbrace{(Ax^{(t)}-b)}_{z} \Rightarrow \|Ax^{(t+1)}-b\|^2 = \|(I-vAA^T)(Ax^{(t)}-b)\|^2$$

از خارج قسمت رایلی استفاده می‌کنیم:

$$r(z) = \frac{\|Bz\|^2}{\|z\|^2} = \frac{z^TB^TBz}{z^Tz} \leq \lambda_{max}(B^TB)$$

با توجه به این که B متقارن است، مقدار ویژه‌های $B^TB = B^2$ مجذور مقدار ویژه‌های B هستند. بنابراین،

(1) $\lambda_{max}(B^TB) = \lambda_{max}(B)^2$ or (2) $\lambda_{max}(B^TB) = \lambda_{min}(B)^2$

(درصورتی که $\lambda_{min}(B)$ عدد منفی کوچکی باشد مجذورش از $\lambda_{max}(B)$ بیشتر است)

$$\Rightarrow \|Bz\|^2 \leq \lambda_{max}(B^TB)\,\|z\|^2$$

حکم سوال

$$(1):\ \|Ax^{(t+1)}-b\|^2 \leq \lambda_{max}(I-vAA^T)^2\|Ax^{(t)}-b\|^2 \Rightarrow \lambda_{max}(I-vAA^T)^2 \leq 1$$

$$\Rightarrow -1 \leq \lambda_{max}(I - vAA^T) \leq 1 \Rightarrow -1 \leq 1 - v\lambda_{min}(AA^T) \leq 1 \Rightarrow$$

$$v\lambda_{min}(AA^T) \leq 2 \Rightarrow v\sigma^2_{min}(A) \leq 2 \Rightarrow \boxed{v \leq \frac{2}{\sigma^2_{min}(A)}}$$

(۲) $\|Ax^{(t+1)}-b\|^2 \leq \lambda_{min}(I-vAA^T)^2\|Ax^{(t)}-b\|^2 \xrightarrow{\text{حکم سوال}} \lambda_{min}(I-vAA^T)^2 \leq 1$

$\Rightarrow -1 \leq \lambda_{min}(I-vAA^T) \leq 1 \Rightarrow -1 \leq 1 - v\lambda_{max}(AA^T) \leq 1$

$\Rightarrow v\lambda_{max}(AA^T) \leq 2 \Rightarrow v\sigma^2_{max}(A) \leq 2 \Rightarrow \boxed{v \leq \frac{2}{\sigma^2_{max}(A)}}$

اشتراک نتیجه‌های (۱) و (۲) $\Rightarrow \boxed{v \leq \frac{2}{\sigma^2_{max}(A)}}$ Q.E.D.