بنام خدا

98102595 تمرین اول یادگیری عمیق فربد یداللهی

۱ـ آ)

$$\|A\|_2 \le \|A\|_F \le \sqrt{\operatorname{rank}(A)}\,\|A\|_2$$

$$\|A\|_2 = \max_{\|x\|_2=1} \|Ax\|_2$$

if P is hermitian then we prove that:

$$\lambda_{max} = \max_{\|x\|_2=1} x^H P x$$

because P is hermitian there exist only a unitary matrix U such that $U^H P U = D$ which diagonalize P ($P = U D U^H$)

D is the diagonal matrix with P's eigenvalues corresponding to the eigenvectors placed in the columns of U

$$y = U^H x \Rightarrow \max_{\|x\|_2=1} x^H P x = \max_{\|y\|_2=1} y^H D y$$

$$= \max_{\|y\|_2=1} \sum_{i=1}^{n} \lambda_i |y_i|^2 \le \lambda_{max} \underbrace{\max_{\|y\|_2=1} \sum_{i=1}^{n} |y_i|^2}_{1} = \lambda_{max}$$

– for x such that

$$\max_{\|x\|_2=1} x^H P x = \lambda_{max} \Rightarrow \|A\|_2 = \max_{\|x\|_2=1} \|Ax\|_2$$

$$= \sqrt{\lambda_{max}(A^H A)}$$

As $A^H A$ is positive semidefinite then its eigenvalues are greater than or equal to 0 we have rank(A) eigenvalues so in a decreasing order:

$$\lambda_1 > \lambda_2 > \lambda_3 > \dots > \lambda_r = \dots = \lambda_n = 0$$

for a matrix $X^{n \times n}$:

$$\operatorname{trace}(X) = \sum_{i=1}^{n} \lambda_i$$

we also have $\|A\|_F = \sqrt{\text{trace}(A^H A)}$

So:

$$\sqrt{\lambda_1} \le \underbrace{\sqrt{\sum_{i=1}^{n} \lambda_i}}_{\|A\|_F} \le \sqrt{r.\lambda_1}$$

and $\lambda_1 = \lambda_{max} = \|A\|_2^2$

So:

$$\|A\|_2 \le \|A\|_F \le \sqrt{r}.\ \|A\|_2$$

ب) i)

$$P(X \geq a) \le \frac{EX}{a}$$

$$EX = \int_{-\infty}^{\infty} x f_X(x)\, dx$$

$$EX = \int_{0}^{\infty} x f_X(x)\, dx$$ x is non negative valued

$$\geq \int_{a}^{\infty} x f_X(x)\, dx$$ for $a > 0$

$$\geq \int_{a}^{\infty} a f_X(x)\, dx$$

$$\geq a \int_{a}^{\infty} f_X(x)\, dx = a P(X \geq a)$$

$$\frac{EX}{a} \geq P(X \geq a)$$

ii)

$$P(|Z-\mu| > \varepsilon) \le \frac{\sigma^2}{\varepsilon^2}$$

using markov inequality:

$$Y = (Z-\mu)^2 \Rightarrow P(Y \geq \varepsilon^2) \le \frac{EY}{\varepsilon^2}$$

$$EY = E[(Z-\mu)^2] = \sigma^2$$

$$P(Y \geq \varepsilon^2) = P((Z-\mu)^2 \geq \varepsilon^2) = P(|Z-\mu| \geq \varepsilon) \le \frac{EY}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2} \Rightarrow P(|Z-\mu| \geq \varepsilon) \le \frac{\sigma^2}{\varepsilon^2}$$

ادامه (۱۷

$P(|Z-\mu|<0.01)\geq 0.95$ $\qquad \mu = \pi$

$1-P(|Z-\mu|>0.01)\geq 0.95$

$P(|Z-\mu|>0.01)\leq 0.05$

با توجه به مشاهدات و صورت سوال تعریف می‌کنیم:

$X_i \sim \text{uniform}(-1,1)$

$Y_i = g(x_i) = 4\sqrt{1+x_i^2} \rightarrow g' = 4\cdot\frac{-2x_i}{2\sqrt{1-x_i^2}} = -\frac{4x_i}{\sqrt{1-x_i^2}}$

$\Rightarrow \mu_{Y_i} = \int_{-1}^{1} 4\sqrt{1-x_i^2}\cdot\frac{1}{2}\,dx = \pi$

$\sigma_{Y_i} = \int_{-1}^{1} 16(1-x_i^2)\cdot\frac{1}{2}\,dx - \mu_{Y_i}^2 = \frac{32}{3}-\pi^2$

by the use of the law of large number we use $n\ (n\rightarrow\infty)$ number of Random variables Y_i so we have:

$Z = Y_1+Y_2+Y_3+\dots+Y_n$

$\overset{LLN}{\Longrightarrow} \mu_Z = \mu_{Y_i} = \pi,\quad \sigma_Z^2 = \frac{\sigma_{Y_i}^2}{n}$

So using the chebyshev inequality:

$P(|Z-\mu_Z|>\varepsilon)\leq\frac{\sigma_Z^2}{\varepsilon^2}$ and we need $\varepsilon=0.01$ and $\frac{\sigma_Z^2}{\varepsilon^2}=0.05$

and:

$\mu_Z=\mu_{Y_i}=\pi,\quad \sigma_Z^2=\frac{\sigma_{Y_i}^2}{n}=\frac{32}{3n}-\frac{\pi^2}{n}$

So:

$P(|Z-\pi|>0.01)\leq\frac{32}{3\times10^{-4}n}-\frac{\pi^2}{10^{-4}n}=0.05$

$\Rightarrow 32-3\pi^2 = 15\times10^{-6}n \Rightarrow n = \frac{32-3\pi^2}{15\times10^{-6}} = 159412$

$$\frac{\partial a^T x}{\partial x} = a^T$$

$$a^T x = \sum_{ij} a_{ij} x_j \Rightarrow \frac{\partial}{\partial x_p} a^T x = \frac{\partial}{\partial x_p} \sum_{ij} a_{ji} x_j = \sum_i a_{pi} = (a^T)_p$$

for $p=1,\dots,n$

$$\Longrightarrow \frac{\partial a^T x}{\partial x} = a^T$$

$$\frac{\partial x^T A x}{\partial x} = x^T(A + A^T)$$

$$x^T A x = \sum_{ij} x_i A_{ij} x_j \Rightarrow \frac{\partial}{\partial x_p} x^T A x = \frac{\partial}{\partial x_p} \sum_{ij} x_i A_{ij} x_j$$

$$= \sum_j A_{pj} x_j + \sum_i x_i A_{ip} = \sum_j x_j A^T_{jp} + \sum_i x_i A_{ip}$$

$$= (x^T A^T)_p + (x^T A)_p = (x^T (A + A^T))_p$$

For all $p = 1, 2, \dots, n$

$$\Longrightarrow \frac{\partial x^T A x}{\partial x} = x^T(A + A^T)$$

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

$$0 = \frac{\partial I}{\partial \beta} = \frac{\partial A A^{-1}}{\partial \beta} = A \frac{\partial A^{-1}}{\partial \beta} + \frac{\partial A}{\partial \beta} A^{-1}$$

$$\Rightarrow A \frac{\partial A^{-1}}{\partial \beta} = -\frac{\partial A}{\partial \beta} A^{-1} \Rightarrow \frac{\partial A^{-1}}{\partial \beta} = -A^{-1} \frac{\partial A}{\partial \beta} A^{-1}$$

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

$$|A| = \sum_j A_{1j} \underbrace{adj(A)_{j1}}_{\text{Cofactor}}$$

$$= \sum_j A_{2j}\, adj(A)_{j2}$$

$$\vdots$$

$$= \sum_j A_{nj}\, adj(A)_{jn}$$

$$\Rightarrow \nabla_A |A| = \begin{bmatrix} \frac{\partial |A|}{\partial A_{11}} & \frac{\partial |A|}{\partial A_{12}} & \cdots & \frac{\partial |A|}{\partial A_{1n}} \\ \vdots & & & \vdots \\ \frac{\partial |A|}{\partial A_{n1}} & \cdots & & \frac{\partial |A|}{\partial A_{nn}} \end{bmatrix} =$$

$$\begin{bmatrix} \frac{\partial \sum_j A_{1j} adj(A)_{j1}}{\partial A_{11}} & \cdots & \frac{\partial \sum_j A_{nj}\, adj(A)_{jn}}{} \\ & & \\ \frac{\partial \sum_j A_{1j} adj(A)_{j1}}{\partial A_{n1}} & \cdots & \frac{\partial \sum_j A_{nj} adj(A)_{jn}}{\partial A_{nn}} \end{bmatrix}$$

$$= \begin{bmatrix} adj(A)_{11} & \cdots & adj(A)_{1n} \\ \vdots & & \vdots \\ adj(A)_{n1} & \cdots & adj(A)_{nn} \end{bmatrix} = adj(A)$$

$$\Rightarrow \nabla_A |A| = adj(A) = |A| \cdot \underbrace{\frac{adj(A)}{|A|}}_{A^{-T}} = |A| A^{-T}$$

$\nabla_A \log|A| = A^{-1}$

$\nabla_A \log|A| = \frac{\partial}{\partial A_{ij}} \log|A| = \frac{1}{|A|} \frac{\partial}{\partial A_{ij}} |A| = \frac{1}{|A|} \operatorname{adj}(A)_{ji} = (A^{-1})_{ji}$

$\xRightarrow[\text{i and j}]{\text{for all}} \nabla_A \log|A| = A^{-1}$

-3

$\lambda_1 + \lambda_2 + \lambda_3 + \cdots + \lambda_n = \operatorname{trace}(A)$

$\lambda_1 \lambda_2 \cdots\cdots \lambda_n = |A|$

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

we calculate the characteristic polynomial of A:

$P(x) = |xI - A| = x^n + C_{n-1} x^{n-1} + \cdots + C_1 x + C_0$

we also have:

$P(x) = (x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_n)$

to calculate C_0 we need to compute $P(x)$ for $x=0$ i.e. $P(0)$

① $P(0) = (0-\lambda_1)(0-\lambda_2)\cdots(0-\lambda_n) = (-1)^n \lambda_1 \cdots\cdots \lambda_n$

② $P(0) = |0I - A| = |-A| = (-1)^n |A|$

$\Rightarrow C_0 = (-1)^n \lambda_1 \cdots \lambda_n = (-1)^n |A| \Rightarrow \lambda_1 \lambda_2 \cdots \lambda_n = |A|$

now we calculate C_{n-1}

① $C_{n-1} = -\lambda_1 x^{n-1} - \cdots - \lambda_n x^{n-1} = -(\lambda_1 + \lambda_2 + \cdots + \lambda_n) x^{n-1}$

choosing x from $n-1$ of the $(x-\lambda_i)$ factors and the constant from the remaining.

② $$|xI - A| = \begin{vmatrix} x-a_{11} & -a_{12} & \cdots & -a_{1n} \\ \vdots & x-a_{22} & & \vdots \\ & & \ddots & \\ -a_{n1} & \cdots & \cdots & x-a_{nn} \end{vmatrix}$$

$P(x) = (x-a_{11})\cdots(x-a_{nn}) + q(x)$ with q with the degree atmost $n-2$.

$P(x) = -(a_{11} + \cdots + a_{nn}) x^{n-1} \Rightarrow a_{11} + a_{22} + \cdots + a_{nn} = \lambda_1 + \cdots + \lambda_n$

$\Rightarrow \lambda_1 + \cdots + \lambda_n = \operatorname{trace}(A)$

$$A^+ = (A^TA)^{-1}A^T$$

we now that A^+ meets these conditions:

1) $AA^+A = A$
2) $A^+AA^+ = A^+$
3) $(AA^+)^T = AA^+$
4) $(A^+A)^T = A^+A$

اگر ماتریس A رتبه کامل ستونی داشته باشد. rank

رنک کامل ستونی

(I) if $\rho(A) = n \Rightarrow A^TA$ is invertible.

(II) if $\rho(A) = m \Rightarrow AA^T$ is invertible

رنک کامل سطری

$$① \Rightarrow A = AA^+A \Rightarrow A^T = (AA^+A)^T = A^T(AA^+)^T$$

$$\xrightarrow{③} A^T = A^TAA^+ \xrightarrow{\mathrm{I}} A^+ = (A^TA)^{-1}A^T$$

$$A^+ = A^T(AA^T)^{-1}$$

$$① \Rightarrow A = AA^+A \Rightarrow A^T = (AA^+A)^T = (A^+A)^TA^T$$

$$\xrightarrow{④} A^T = A^+AA^T \xrightarrow{\mathrm{II}} A^+ = A^T(AA^T)^{-1}$$

5

$$M = \begin{bmatrix} A_{n\times n} & B_{n\times k} \\ C_{k\times n} & D_{k\times k} \end{bmatrix}$$

$$\det\left(\begin{bmatrix} A_{n\times n} & B_{n\times k} \\ C_{k\times n} & D_{k\times k} \end{bmatrix}\right) = \det(D)\det(A - BD^{-1}C)$$

if D is invertible;

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} c \\ d \end{pmatrix}$$

$$\begin{aligned} Ax + By &= c \\ Cx + Dy &= d \end{aligned} \Rightarrow y = D^{-1}(d - Cx)$$

$$Ax + B(D^{-1}(d - Cx)) = c$$

$$(A - BD^{-1}C)x = c - BD^{-1}d$$

if $A - BD^{-1}C$ is invertible then,

$$x = (A - BD^{-1}C)^{-1}(c - BD^{-1}d)$$

$$y = D^{-1}(d - C(A - BD^{-1}C)^{-1}(c - BD^{-1}d))$$

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

$$\det M = \det\left(\begin{bmatrix} I & BD^{-1} \\ 0 & I \end{bmatrix}\begin{bmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{bmatrix}\begin{bmatrix} I & 0 \\ D^{-1}C & I \end{bmatrix}\right) =$$

$$\det\begin{pmatrix} I & BD^{-1} \\ 0 & I \end{pmatrix}\det\begin{bmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{bmatrix}\det\begin{bmatrix} I & 0 \\ D^{-1}C & I \end{bmatrix} = 1 \times \det D \cdot \det(A - BD^{-1}C) \times 1$$

$$\det\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det(A)\det(D - CA^{-1}B)$$

like before if A is invertible then we can have:

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

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

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

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

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

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

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

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} (A-BD^{-1}C)^{-1} & -A^{-1}B(D-CA^{-1}B)^{-1} \\ -(D-CA^{-1}B)^{-1}CA^{-1} & (D-CA^{-1}B)^{-1} \end{pmatrix} \quad (1)$$

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} + A^{-1}B(D-CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D-CA^{-1}B)^{-1} \\ -(D-CA^{-1}B)^{-1}CA^{-1} & (D-CA^{-1}B)^{-1} \end{pmatrix} \quad (2)$$

$$\overset{(1)(2)}{\Longrightarrow} (A - BD^{-1}C)^{-1} = A^{-1} + A^{-1}B(D-CA^{-1}B)^{-1}CA^{-1}$$

if we replace B with $-B$ then we will have:

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

$$\det(A - uv^T) = \det(A)(1 - v^T A^{-1} u)$$

$$\det\left(\begin{bmatrix} A & \underset{n\times1}{u} \\ \underset{1\times n}{v^T} & \underset{1\times1}{I} \end{bmatrix}\right) = \det(A - uv^T)$$

$$\det\left(\begin{bmatrix} A & u \\ v^T & I \end{bmatrix}\right) = \det(A)\det(\underset{1\times1}{I} - \underbrace{\overset{1\times n}{v^T}\,\overset{n\times n}{A^{-1}}\,\overset{n\times1}{u}}_{1\times1}\,T)$$

$$= \det(A)(1 - v^T A^{-1} u)$$

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

-7

$x_1, x_2, \ldots, x_n$

$$f(x) = \frac{1}{\theta^2} x e^{\frac{-x}{\theta}}, \quad 0 < \theta < \infty$$

نمونه ها iid هستند

$$L(x_1, \ldots, x_n) = \frac{\prod_{i=1}^{n} x_i}{\theta^{2n}} e^{\frac{-\sum_{i=1}^{n} x_i}{\theta}}$$

It is easier to work with log likelihood so:

$$\ln L = \sum_{i=1}^{n} \ln x_i - 2n \ln\theta - \frac{1}{\theta}\sum_{i=1}^{n} x_i$$

$$\Rightarrow \frac{\partial \ln L}{\partial\theta} = \frac{-2n}{\theta} + \frac{\sum_{i=1}^{n} x_i}{\theta^2} = 0 \Rightarrow \theta = \frac{\sum_{i=1}^{n} x_i}{2n} = \frac{1}{2}\bar{x}$$

infact θ is half of the sample mean $\frac{1}{2}\bar{X}$ so it is a estinator of the mean.

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

MLE:

$$L(x_1, \dots, x_n) = \frac{1}{\sigma^n (\sqrt{2\pi})^n} e^{\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

$$\ln L = -n \ln \sigma - \frac{n}{2}\ln 2\pi + \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{2\sigma^2}\sum_{i=1}^{n} -2(x_i-\mu) = 0$$

$$\frac{-1}{\sigma^2}\left(\sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\mu\right) = 0$$

$$\frac{n\mu}{\sigma^2} = \frac{\sum_{i=1}^{n} x_i}{\sigma^2} \Rightarrow \mu = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{X} = \hat{\mu}_{ML}$$

MAP:

$$f_{prior} = \frac{1}{\beta\sqrt{2\pi}} e^{\frac{(x-\gamma)^2}{2\beta^2}}$$

we must maximize:

$$f_{prior}(\mu)\, f(x|\mu) = \frac{1}{\sqrt{2\pi}\,\beta} \exp\left(\frac{-(\mu-\gamma)^2}{2\beta^2}\right) \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(\frac{-(x_i-\mu)^2}{2\sigma^2}\right)$$

Similarly maximizing above equation is equal to minimize:

$$\left(\frac{\mu-\gamma}{\beta}\right)^2 + \sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2$$

$$\hat{\mu}_{MAP} = \frac{\beta n}{\beta^2 n+\sigma^2}\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) + \frac{\sigma^2}{\beta^2 n+\sigma^2}\gamma = \frac{\beta^2 \sum_{i=1}^{n} x_i + \sigma^2\gamma}{\beta^2 n + \sigma^2}$$

As the sample size tends to ∞, MLE and MAP become equal. The prior is important if we don't have much data, but as data increases, evidence overwhelms the prior.

$N(X \mid \mu, \Sigma)$

$$X = \begin{pmatrix} x_a \\ x_b \end{pmatrix} \quad \mu = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix} \quad \Sigma = \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix}$$

$$x_b = \sqrt{\Sigma_{bb}}\, z_1 + \mu_b \qquad z_1, z_2 \sim N(0,1)$$

$$x_a = \Sigma_{ab} \Sigma_{bb}^{-1} (x_b - \mu_b) + \sqrt{\Sigma_{aa} - \frac{\Sigma_{ab}\Sigma_{ba}}{\Sigma_{bb}}}\, z_2 + \mu_{xa}$$

$$E[x_a \mid x_b] = \Sigma_{ab} \Sigma_{bb}^{-1} (x_b - \mu_b) + \sqrt{\frac{\det \Sigma}{\Sigma_{bb}}}\, \underbrace{E[z_2]}_{0} + \mu_a$$

$$E[x_a \mid x_b] = \mu_{xa} + \Sigma_{ab} \Sigma_{bb}^{-1} (x_b - \mu_b)$$

$$\Sigma_{a|b} = \left(\sqrt{\frac{\det \Sigma}{\Sigma_{bb}}}\right)^2 \operatorname{var}(z_2) = \left(\sqrt{\det \Sigma}\right)^2 = \det \Sigma$$

$$= \frac{\Sigma_{aa}\Sigma_{bb} - \Sigma_{ab}\Sigma_{ba}}{\Sigma_{bb}} = \Sigma_{aa} - \frac{\Sigma_{ab}\Sigma_{ba}}{\Sigma_{bb}}$$

$$= \Sigma_{aa} - \Sigma_{ab} \Sigma_{bb}^{-1} \Sigma_{ba}$$

$$E[x_a] = \mu_a$$

$$E[x_a] = E[E[x_a \mid x_b]] \qquad \text{law of iterated expectations}$$

$$E[x_a] = E[E[x_a \mid x_b]] = E\left[\Sigma_{ab} \Sigma_{bb}^{-1} (x_b - \mu_b) + \sqrt{\frac{\det \Sigma}{\Sigma_{bb}}}\, E[z_2] + \mu_a\right]$$

$$= \Sigma_{ab}\Sigma_{bb}^{-1} \underbrace{(\underbrace{E[x_b]}_{\mu_b} - \mu_b)}_{0} + \sqrt{\frac{\det \Sigma}{\Sigma_{bb}}}\, \underbrace{E[z_2]}_{0} + \mu_a$$

$$E[x_a] = \mu_a$$

$$E[\Sigma_{a|b}] = E[x_a^2] - E[E[x_a \mid x_b]^2]$$

$$\Sigma_{aa} - \Sigma_{ab} \Sigma_{bb}^{-1} \Sigma_{ba} = E[x_a^2] - E\left[\mu_a^2 + (x_b - \mu_b)^2 \Sigma_{ab} \Sigma_{bb}^{-1} + \mu_a \Sigma_{ab} \Sigma_{bb}^{-1} (x_b - \mu_b)\right]$$

$$= E[x_a^2] - \mu_a^2 + \Sigma_{ab}$$

$$f_{x_a,x_b}(x_a,x_b) = \frac{1}{2\pi\, \Sigma_{aa}\, \Sigma_{bb} \sqrt{1-\rho^2}} \exp\left(-\frac{1}{2} Q(x_a,x_b)\right)$$

$$Q(x_a,x_b) = \frac{1}{1-\rho^2}\left[\left(\frac{x_a-\mu_a}{\sqrt{\Sigma_{aa}}} - \rho \frac{x_b-\mu_b}{\sqrt{\Sigma_{bb}}}\right)^2 + (1-\rho^2)\left(\frac{x_b-\mu_b}{\sqrt{\Sigma_{bb}}}\right)^2\right]$$

$$Q(x_a,x_b) = \frac{(x_a - A(x_b))^2}{(1-\rho^2)\,\Sigma_{aa}} + \frac{(x_b-\mu_b)^2}{\Sigma_{bb}}$$

and $A = \mu_a + \rho\sqrt{\frac{\Sigma_{aa}}{\Sigma_{bb}}}\,(x_b-\mu_b)$

with integration over x_b:

$$f(x_a) = \frac{1}{\sqrt{2\pi}\,\Sigma_{aa}} \exp\left(\frac{-(x_a-\mu_a)^2}{2\Sigma_{aa}}\right)$$

$$\Rightarrow x_a \sim (\mu_a, \Sigma_{aa})$$

$x^* = A^+ b \rightarrow$ min norm 2

$$x = (A^+A)x + (I - A^+A)x = A^+b + (I - A^+A)x$$

$$\Rightarrow \|x\|^2 = \|A^+b\|^2 + \|(I - A^+A)x\|^2 \geq \|A^+b\|^2$$

$$\Rightarrow \|x\| \geq \|A^+b\|$$

(b

$$Ax = b \xrightarrow{SVD} A = U\Sigma V^T = U\begin{bmatrix} \Sigma_1 & 0 \end{bmatrix}\begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} = U\Sigma^{-1} V_1^T$$

least Norm:

$$x_{LN} = A^T(AA^T)^{-1}b = V_1 \Sigma_1^{-1} U^T b$$

$$f(x) = \|Ax - b\|_2^2$$

$$\nabla f(x) = 2A^T(Ax - b) \Rightarrow$$ ۲ را در داخل ν در نظر می‌گیریم.

$$x^{(t+1)} = x^{(t)} - \nu A^T(Ax^{(t)} - b)$$

$$= (I - \nu A^TA)x^{(t)} + \nu A^T b$$

$$\Rightarrow x^{(t)} = (I - \nu A^TA)^k x_0 + \nu \sum_{l=0}^{k-1} (I - \nu A^TA)^l A^T b$$

$$y := V^T x$$

$$y^{(t)} = (I - \nu\Sigma^T\Sigma)^k y_0 + \nu \sum_{l=0}^{k-1} (I - \nu\Sigma^T\Sigma)^l \Sigma^T U^T b$$

$$= \begin{bmatrix} (I - \nu\Sigma_1^2)^k & 0 \\ 0 & I \end{bmatrix} y_0 + \nu \sum_{l=0}^{k-1} \begin{bmatrix} (I - \nu\Sigma_1^2)^l & 0 \\ 0 & I \end{bmatrix}\begin{bmatrix} \Sigma_1 \\ 0 \end{bmatrix} U^T b$$

$$= \begin{bmatrix} (I - \nu\Sigma_1^2)^k & 0 \\ 0 & I \end{bmatrix} y_0 + \nu \sum_{l=0}^{k-1} \begin{bmatrix} (I - \nu\Sigma_1^2)^l \Sigma_1 \\ 0 \end{bmatrix} U^T b$$

$$y^\infty = \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix} y_0 + \nu \sum_{l=0}^{\infty} \begin{bmatrix} (I - \nu\Sigma_1^2)^l \Sigma_1 \\ 0 \end{bmatrix} U^T b$$

$$y^\infty = \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix} y_0 + \begin{bmatrix} \Sigma_1^{-1} \\ 0 \end{bmatrix} U^T b \xrightarrow{x = Vy} x = V_2 V_2^T x_0 + \underbrace{V_1 \Sigma_1^{-1} U^T b}_{x_{LN}}$$

c) با توجه به عبارت نوشته شده در بخش قبل برای کاهش در هر مرحله باید داشته باشیم:

$$1 - \nu \Sigma_1^2 \geq 0 \Rightarrow 1 \geq \nu \sigma^2_{max}(A)$$

$$\Rightarrow \nu \leq \frac{2}{\sigma^2_{max}(A)}$$

we know that:

$$\det\begin{bmatrix}\tilde{A} & x \\ y^T & a\end{bmatrix} = a\det\tilde{A} - y^T(\operatorname{adj}\tilde{A})x$$

now we have:

$$P_A(t) = \det(tI-A) = \det\begin{bmatrix} tI-B & x \\ y^* & t-a\end{bmatrix} =$$

$$(t-a)\det(tI-B) - y^*\operatorname{adj}(tI-B)x$$

$$P_A(t) = (t-a)P_B(t) - y^*(\operatorname{adj}(tI-B))x$$

1b

by the ~~uscoft~~ courant fisher theorem:

$\exists\, S_A, S_B \in \mathbb{F}^n$, $\dim S_A = k+j$, $\dim S_B = n-j$

$$\lambda_{k+j}(A) = \max_{\substack{x\in S_A\\ \|x\|=1}} \langle Ax, x\rangle$$

$$\lambda_{n-j}(B) = \max_{\substack{x\in S_B\\ \|x\|=1}} \langle Bx, x\rangle$$

$$\dim(S_A\cap S_B) \geq k \Rightarrow \lambda_k(A+B) = \min_{\dim S=k} \max_{\substack{x\in S\\ \|x\|=1}} \langle (A+B)x, x\rangle$$

$$\leq \min_{\substack{\dim S=k\\ S\subseteq S_A\cap S_B}} \left(\max_{\substack{x\in S\\ \|x\|=1}} (\langle Ax,x\rangle + \langle Bx,x\rangle)\right)$$

$$\leq \min_{\substack{\dim S=k\\ S\in S_A\cap S_B}} \left(\max_{\substack{x\in S\\ \|x\|=1}} \langle Ax,x\rangle + \max_{\substack{x\in S\\ \|x\|=1}} \langle Bx,x\rangle\right)$$

$$= \lambda_{k+j}(A) + \lambda_{n-j}(B)$$

if $y \in \mathbb{F}^m$ and $x = (y_1, \dots, y_m, 0, \dots, 0) \in \mathbb{F}^n$ (C

$\Rightarrow \langle Ax, x \rangle = \langle By, y \rangle$, $\|x\| = \|y\|$

$$\lambda_k(A) = \min_{\substack{S \subseteq \mathbb{F}^n \\ \dim S = k}} \max_{\substack{x \in S \\ \|x\| = 1}} \langle Ax, x \rangle \leq \min_{\substack{S \subseteq \mathrm{span}(e_1, \dots, e_m) \in \mathbb{F}^n \\ \dim S = m-k+1}} \min_{\substack{x \in S \\ \|x\| = 1}} \langle Ax, x \rangle$$

$$= \max_{\substack{S \subseteq \mathbb{F}^m \\ \dim S = n-k+1}} \min_{\substack{y \in S \\ \|y\| = 1}} \langle By, y \rangle = \lambda_k(B)$$

and:

$$\lambda_{k+n-m}(A) = \max_{\substack{S \subseteq \mathbb{F}^n \\ \dim S = m-k+1}} \min_{\substack{x \in S \\ \|x\| = 1}} \langle Ax, x \rangle \geq \max_{\substack{S \subseteq \mathrm{span}(e_1, \dots, e_m) \in \mathbb{F}^n \\ \dim S = m-k+1}} \min_{\substack{x \in S \\ \|x\| = 1}} \langle Ax, x \rangle$$

$$= \max_{\substack{S \subseteq \mathbb{F}^m \\ \dim S = n-k+1}} \min_{\substack{y \in S \\ \|y\| = 1}} \langle By, y \rangle = \lambda_k(B)$$

when $m = n\cancel{-1}$:

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