4) a) Singular values of A are $\sqrt{\lambda}$ of $A^T A$.

$A^T A$ is written as PDP^T, where D is diagonal and P, P^T are orthonormal inverses and as such

$\left(\frac{1}{\sqrt{2}}\right)^2 + \left(\frac{1}{\sqrt{2}}\right)^2 = 1$ → orthonormal

$\frac{1/\sqrt{2}}{1/\sqrt{2}}$ L.I w/ $\frac{-1/\sqrt{2}}{1/\sqrt{2}}$

$A^T A$ is given in spectral decomposition form

↳ $\sigma_1(A) = \sqrt{5}$ $\sigma_2(A) = \sqrt{3}$

b) $CC^T = \begin{bmatrix} B & -B & 3I_3 \end{bmatrix}^{3\times 7} \cdot \begin{bmatrix} B^T \\ -B^T \\ 3I_3^T \end{bmatrix}^{7\times 3} = \begin{bmatrix} BB^T + BB^T + 9I_3^T \end{bmatrix}$

$= \begin{bmatrix} 2BB^T + 9I_3 \end{bmatrix}^{3\times 3} = V(2\sigma_B^2 + 9I_3)V^T$, where σ_B denotes diag σ_i of B.

Hence $\sigma_C = \sqrt{\sigma_{CC^T}} = \sqrt{2\sigma_B^2 + 9I_3}$

so $\sigma_1 = \sqrt{2 \cdot 7 + 9} = \sqrt{23}$ $\sigma_2 = \sqrt{13}$ $\sigma_3 = 3$

5) a) $Df(\vec{x}) = 2\vec{x}$, $D^2 f(\vec{x}) = 2I$

$\vec{x}\, 2I \vec{x} = 2\|\vec{x}\|_2^2 \geq 0 \;\forall \vec{x} \in \mathbb{R}^n \Rightarrow$ PSD hessian

so f is convex.

Scanned with CamScanner

6) $\nabla g(\vec{x}) = e^{\|\vec{x}\|_2^2} \cdot 2\vec{x}$

$$\nabla^2 g(\vec{x}) = e^{\|\vec{x}\|_2^2} \cdot 2\nabla^2 f(\vec{x}) + e^{\|\vec{x}\|_2^2} \cdot 4x^T x$$

$$= \underbrace{e^{\|\vec{x}\|_2^2}}_{>0\ \forall x} \cdot \underbrace{2I}_{PSD} + \underbrace{e^{\|\vec{x}\|_2^2}}_{>0\ \forall x} \cdot \underbrace{4\|\vec{x}\|_2^2}_{\geq 0\ \forall x}$$

$\succeq 0$ everywhere so convex

6) ~~Show~~ S is convex set iff convex combination of any $S_1, S_2 \in$ is also in S

Let $S_3 = \theta S_1 + (1-\theta) S_2$, where

$S_1 = \sum x_i F_i$ for some $\vec{x} \in \mathbb{R}^n$

$S_2 = \sum y_i F_i$ for some $\vec{y} \in \mathbb{R}^n$

$$S_3 = \theta \sum_1^n x_i F_i + \sum_1^n (1-\theta) y_i F_i = \sum_1^n (\theta x_i + (1-\theta) y_i) F_i$$

so $S_3 = \sum_1^n z_i F_i$ for a $\vec{z}$ in $\mathbb{R}^n$ where $\vec{z} = \theta\vec{x} + (1-\theta)\vec{y}$

we know S_3 is PSD from hint as well so $S_3 \in S$

7) a) $f'(x) = 2ax$

i) $x_{t+1} = x_t - \eta f'(x_t) = x_t - \eta \cdot 2a x_t$

c) $x_{t+1} = (1-\eta a) x_t \Rightarrow x_{t+1} = (1-\eta a)^2 x_{t-1}$

recursive so $x_{t} = (1-\eta a)^{t-1} x_0$

6) $f(\vec{x}) = \vec{v}(\vec{v}^T\vec{x}) = (\vec{v}\cdot\vec{v}^T)\vec{x}$

$$Df(\vec{x}) = \begin{bmatrix} \partial f_1/\partial x_1 & \partial f_1/\partial x_2 & \cdots & \partial f_1/\partial x_n \\ \partial f_2 & \cdots & & \\ \partial f_n/\partial x_1 & \cdots & & \partial f_n/\partial x_n \end{bmatrix}$$

Here $f_i(\vec{x}) = U_i\vec{x}$, where $U = \vec{v}\vec{v}^T$ and U_i is i^{th} row

$\partial f_i/\partial x_j = U_{i,j}$ (entry at i^{th} row, j^{th} col of U)

$$Df(\vec{x}) = \begin{bmatrix} U_{1,1} & U_{1,2} & \cdots & U_{1,n} \\ U_{2,1} & \cdots & & \\ & & & \\ U_{n,1} & \cdots & & U_{n,n} \end{bmatrix} = U = \vec{v}\vec{v}^T \checkmark$$

9) $A = \cancel{U^{n\times k}\Sigma^{k\times k}V^{T\,k\times n}}$ By spectral decomposition
$U\Lambda U^T$

$A = U\Lambda U^T = \sum_1^k \sigma_i u_i u_i^T$ ($rk = k \Rightarrow k$ σ_i)

$= U_k \cdot \Sigma_r U_k^T$, U^k has dims $n\times k$, Σ has $k\times k$ (full diag of $\lambda_i \neq 0$)

Since A is PSD, its eigenvalues ~~are sing~~ are non-negative

$\cancel{\Sigma^{1/2}}$ ~~Define~~ so we can define

~~$P = U_k\Sigma^{1/2}$ U_k has dim $n\times k$ and rank k~~

~~as can also be written as~~

$P = U_k^{n\times k}\Sigma^{1/2\,k\times k}V^{T\,k\times k}$ where V^T has right singular vectors of $\Sigma^{1/2}$ and as such P has rank k and is dim $n\times k$

$\cancel{A =} PP^T = U\Sigma^{1/2}\underbrace{V^TV}_{I}\Sigma^{1/2}U^T = U\Sigma U^T = A \checkmark$

Scanned with CamScanner

$(rk(A) = rk(A^T))$ $(N(A) + rk(A) = n \; \forall A^{m \times n})$

6) $rk(P) = k \Rightarrow rk(P^T) = k \Rightarrow N(P^T) = n-k$

$N(P^T) = N(PP^T) = N(A) = n-k$

$\uparrow$ hint $\hookrightarrow rk(A) = k \quad (N(A) + rk(A) = n)$

10) $A\vec{v} = \vec{r}_1^T \vec{v}_1 + \vec{r}_2^T \vec{v}_2 \ldots \vec{r}_n^T \vec{v}_n$

$\|A\vec{v}\|_\infty = \cancel{max} \max |\vec{r}_i^T \vec{v}|$ $\downarrow$ choose $\vec{r}_i$ w/ largest l-2 norm

$\max_{\substack{\vec{v} \in \mathbb{R}^n \\ \|\vec{v}\|_2 = 1}} \|A\vec{v}\|_\infty \leq \max |\vec{r}_i^T \vec{v}| \overset{C.S}{=} \max \|\vec{r}_i\|_2 \cdot \|\vec{v}\|_2$

$\leq \max \|\vec{r}_i\|_2 \cdot 1$

Equality achieved when $\vec{v}$ points in same direction as $\vec{r}_i$ $(\cos\theta = 0)$

So $\max_{\substack{\vec{v} \in \mathbb{R}^n \\ \|\vec{v}\|_2 = 1}} \|A\vec{v}\|_\infty = \max_i \|\vec{r}_i\|_2$

11) a) $\vec{v}_2, \vec{v}_3$ since explain most variance

b) top k principal components are eigenvalues of $X^T X$, which are singular values of $X \Rightarrow k_i = \sigma_i(X)$

c) $XP = \begin{bmatrix} \vec{x}_1^T \vec{p}_1 & \vec{x}_1^T \vec{p}_2 & \vec{x}_1^T \vec{p}_3 & \cdots & \vec{x}_1^T \vec{p}_k \\ \vec{x}_2^T \vec{p}_1 & \cdots & & & \\ \vdots & & & & \\ \vec{x}_n^T \vec{p}_1 & \cdots & & & \vec{x}_n^T \vec{p}_k \end{bmatrix} = Z$

$\hookrightarrow Z_{ij} = \vec{x}_i^T \vec{p}_j$

Scanned with CamScanner

Solution to ridge regression $\|A\vec{x}-\vec{y}\|_2^2+\lambda\|\vec{x}\|_2^2$

is $x^* = (A^T A + \lambda I)^{-1} A^T \vec{y}$

Adapting to this problem:

$\beta^* = (Z^T Z + \lambda I_k)^{-1} Z^T \vec{y}$ ~~of~~

$Z^T Z = (XV_k)^T (XV_k) = V_k^T X^T X V_k$

using $X = U_d \Sigma_d V_d^T$ ~~$\Rightarrow Z^T Z = V_k^T V_d \Sigma_d^T U_d^T U_d \Sigma_d V_d^T V_k$~~

$\hookrightarrow$ $Z^T Z = V_k^T V_d \Sigma_d^T \underbrace{U_d^T U_d}_{I} \Sigma_d V_d^T V_k$

~~$V_k^T \cdot V_d = \underbrace{V_k^T \cdot V_k}_{I} + \underbrace{V_k^T \cdot V_{d-k}}_{0 \text{ as columns of } V \text{ are orthogonal}} = I_k$~~

$$\begin{bmatrix} \underbrace{V_k^T \cdot V_k}_{I} & \underbrace{V_k^T \cdot V_{d-k}}_{0 \text{ as } V \text{ has } \perp \text{ cols}} \end{bmatrix}_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ & \ddots & 1 & 0 \end{bmatrix}^{k \times d}$$

so $Z^T Z = \begin{bmatrix} 1 & & 0 & \\ & \ddots & & 0 \\ 0 & & 1 & \end{bmatrix}^{k \times d} \Sigma^d \Sigma^d \begin{bmatrix} 1 & & 0 \\ & \ddots & 1 \\ 0 & & 0 \end{bmatrix}^{d \times k}$ (erases $d-k$ λ of Σ)

$= \Sigma^k \Sigma^k = \Sigma_k^2$

Similarly

$Z^T = V_k^T V_d \Sigma_d U_d^T = \begin{bmatrix} 1 & & & 0 \\ & 1 & 1 & \\ 0 & & 1 & 0 \end{bmatrix} \Sigma_d U_d^T = \Sigma_k U_k^T$

$k \times d$ $d \times d$ This gives us $(\Sigma_k^2 + \lambda I_k)^{-1} \Sigma_k U^{k\,T} \vec{y}$ as solution

e) First simplify $X\vec{a}^* - Z\beta^*$

$$X\vec{a}^* = U_d \Sigma_d \underbrace{V_d^T V_d}_{I} (\Sigma_d^2 + \lambda I)^{-1} \Sigma_d U_d^T \vec{y}$$

$$= U_d \underbrace{\begin{bmatrix} \sigma_1 \cdot \sigma_1 / (\sigma_1^2 + \lambda) & & \\ & \sigma_2 \cdot \sigma_2 / (\sigma_2^2 + \lambda) & \\ & & \ddots \end{bmatrix}}_{\text{call this } D_d} U_d^T y$$

$$Z\vec{b} = \underbrace{U_d \Sigma_d V_d^T V_k}_{U_k \Sigma_k \text{ from a)}} (\Sigma_k^2 + \lambda I_k)^{-1} \Sigma_k U_k^T \vec{y} = U_k \begin{bmatrix} \sigma_1^2 / (\sigma_1^2 + \lambda) & & \\ & \ddots & \\ & & \sigma_k^2 / (\sigma_k^2 + \lambda) \end{bmatrix}$$

$$= U_k \underbrace{\begin{bmatrix} \sigma_1^2 / (\sigma_1^2 + \lambda) & & \\ & \ddots & \\ & & \sigma_k^2 / (\sigma_k^2 + \lambda) \end{bmatrix}}_{D_k} U_k^T \vec{y}$$

$$X\vec{a}^* - Z\vec{b}^* = U_d D_d U_d^T \vec{y} - U_k D_k U_k^T \vec{y}$$

$$= (U_d D_d U_d^T - U_k D_k U_k^T) \vec{y}$$

$$= \left(\sum_{1}^{d} \frac{\sigma_i^2}{\sigma_i^2 + \lambda} \vec{u}_i \vec{u}_i^T - \sum_{1}^{k} \frac{\sigma_i^2}{\sigma_i^2 + \lambda} \vec{u}_i \vec{u}_i^T \right) \vec{y}$$

$$\| X\vec{a} - Z\vec{b}^* \|_2^2 = \left\| \left(\sum_{i=k+1}^{d} \frac{\sigma_i^2}{\sigma_i^2 + \lambda} u_i u_i^T \right) \vec{y} \right\|_2^2$$

d) ~~x_t~~ since $x_0 = 1$, $\lim_{t\to\infty} x_t = (1-\eta a)^{\infty}$ ~~t~~

Since $\eta > 0$ and $0 < a < 1/\eta \Rightarrow 0 < \eta a < 1$,

function converges: $x_t = 0$ as $t \to \infty$

8) a) $\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \end{bmatrix} = \cancel{f(\vec{x})} \cdot \begin{bmatrix} \cos(\vec{a}^T\vec{x}-6) \cdot \frac{\partial \vec{a}^T\vec{x}}{\partial x_1} \\ \cos(\vec{a}^T\vec{x}-6) \cdot \frac{\partial \vec{a}^T\vec{x}}{\partial x_2} \\ \vdots \end{bmatrix}$ (chain rule)

$= \begin{bmatrix} \cos(\vec{a}^T\vec{x}-6) \cdot a_1 \\ \cos(\vec{a}^T\vec{x}-6) \cdot a_2 \\ \vdots \\ \cos(\vec{a}^T\vec{x}-6) \cdot a_n \end{bmatrix} = \cos(\vec{a}^T\vec{x}-6) \cdot \vec{a}$

$D^2 f(\vec{x})$ is symmetric by Clairaut's Thm.

$D^2 f(\vec{x}) = \begin{bmatrix} \partial^2 f/\partial x_1^2 & \partial f/\partial x_1 \partial x_2 & \cdots & \partial f/\partial x_1 \partial x_n \\ \partial f/\partial x_2 \partial x_1 & & & \\ \vdots & & & \\ \partial f/\partial x_n \partial x_1 & & & \end{bmatrix}$

~~For $\partial f/\partial x_i^2$:~~ ~~$\partial f$~~ For $i = j$:

$\frac{\partial f}{\partial x_i} \cos(\vec{a}^T\vec{x}-6) \cdot a_i = -\sin(\vec{a}^T\vec{x}-6) \cdot a_i \cdot a_i$

for $i \neq j$: $-\sin(\vec{a}^T\vec{x}-6) \cdot a_i \cdot a_j$

$\hookrightarrow D^2 f(\vec{x}) = -\sin(\vec{a}^T\vec{x}-6) \cdot \begin{bmatrix} a_1^2 & a_1 a_2 & \cdots & a_1 a_n \\ a_2 a_1 & a_2^2 & & \\ \vdots & & \ddots & \\ & & & a_n^2 \end{bmatrix}$

$= -\sin(\vec{a}^T\vec{x}-6) \cdot \vec{a} \cdot \vec{a}^T$

Scanned with CamScanner

I will follow the rules and do this exam on my own

Percy Jatison Joe Rogan

245065599