EECS 545 F17 HW1

Name: Po-Jen Lai   Members: Ming-Ko Cho, Yi-Lun Wu, Chun-Yu Hsiung

1) (a)

(i) True. We know $I = I^T$, since $AA^{-1} = I \Rightarrow AA^{-1} = (AA^{-1})^T$, since $(AB)^T = B^T A^T \Rightarrow AA^{-1} = (A^{-1})^T A^T$

$\Rightarrow AA^{-1} = A^{-1}A = (A^{-1})^T A^T$. since $A = A^T \Rightarrow A^{-1}A = (A^{-1})^T A$. $\Rightarrow A^{-1}AA^{-1} = (A^{-1})^T A A^{-1} \Rightarrow \underline{A^{-1} = (A^{-1})^T}$ #

(ii) True. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. if $\det(A) = 1$, $\begin{bmatrix} a & c \\ b & d \end{bmatrix} = A^T = A^{-1} = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.

$\Rightarrow d = a,\ b = -c \Rightarrow A$ has the form $\begin{bmatrix} a & -c \\ c & a \end{bmatrix}$, and $a^2 + c^2 = 1$

$\Rightarrow$ We can write A as $\underline{\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}}$ # if $\det(A) = -1$, $\begin{bmatrix} a & c \\ b & d \end{bmatrix} = \begin{bmatrix} -d & b \\ c & -a \end{bmatrix}$

$\Rightarrow a = -d,\ b = c \Rightarrow A$ has the form $\begin{bmatrix} a & c \\ c & -a \end{bmatrix}$, and $-(a^2 + c^2) = -1 \Rightarrow \underline{A = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}}$ #

(iii) eigenvalues of A are $-15, -4.899, 4.899 \Rightarrow A$ is not positive-definite

$\Rightarrow A$ cannot be decomposed as CC^T. <u>the answer is false.</u> #

2)

(a)

(i) $E[X] = \int_{-\infty}^{\infty} x f_X(x)\,dx$, since $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy \Rightarrow E[X] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x,y)\,dx\,dy$

since $f_{X|Y}(x|y) = \dfrac{f_{X,Y}(x,y)}{f_Y(y)} \Rightarrow \underline{E[X] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x \cdot f_{X|Y}(x|y) \cdot f_Y(y)\,dx\,dy = E_Y[E_X[X|Y]]}$ #

(ii) $E[I[X \in C]] = \int_X I[X \in C]\,dP = \int_X 1 \cdot dP = \underline{P(X \in C)}$ #

(iii) $Var[X] = E[X^2] - (E[X])^2 = E_Y[E_X[X^2|Y]] - (E_Y[E_X[X|Y]])^2$

$= E_Y[E_X[X^2|Y]] - E_Y[(E_X[X|Y])^2] + E_Y[(E_X[X|Y])^2] - (E_Y[E_X[X|Y]])^2$

$\underline{= E_Y[Var_X[X|Y]] + Var_Y[E_X[X|Y]]}$ #

(iv) $E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dx\,dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_X(x) f_Y(y)\,dx\,dy = \int_{-\infty}^{\infty} f_X(x)\,dx \int_{-\infty}^{\infty} f_Y(y)\,dy = \underline{E[X]E[Y]}$ #

2) (a) (i) $E[XY] = \Sigma\, xy P_r(X=x) P_r(Y=y) = P_r(X=1, Y=1)$.

$E[X] = \Sigma\, x P_r(X=x, Y=y) = P_r(X=1, Y=0) + P_r(X=1, Y=1) = P_r(X=1)$

$E[Y] = \Sigma\, y P_r(X=x, Y=y) = P_r(X=1, Y=1) + P_r(X=0, Y=1) = P_r(Y=1)$

$\Rightarrow P_r(X=1, Y=1) = E[XY] = E[X]E[Y] = \underline{P_r(X=1) \cdot P_r(Y=1)} \rightarrow$ independent

and for $X = X,\ Y = 1-y$. $E[XY] = \Sigma\, x(1-y) P_r(X=x) P_r(Y=y) = P_r(X=1, Y=0)$

$E[X] = \Sigma\, x P_r(X=x, Y=y) = P_r(X=1)$, $E[Y] = \Sigma (1-y) P_r(X=x, Y=y) = P_r(Y=0)$

$\Rightarrow P_r(X=1, Y=0) = E[XY] = E[X]E[Y] = \underline{P_r(X=1) P_r(Y=0)} \rightarrow$ independent.

for $X = 1-X,\ Y = 1-y$ and $X = 1-X,\ Y = y$, the same as above.

$\Rightarrow E[XY] = E[X]E[Y] \rightarrow \underline{X, Y \text{ are independent.}}$ #

(b)

(i) $P(H=h) = \sum_{i} P(H=h, D=d_i) \Rightarrow P(H=h, D=d) \leq P(H=h)$ #

(ii) $P(H=h \mid D=d) = \dfrac{P(H=h, D=d)}{P(D=d)} = a \quad 0 \leq a \leq 1 \Rightarrow \underline{\text{depends}}$ #

(iii) $\dfrac{P(H=h, D=d)}{P(D=d)} \geq P(H=h, D=d) = P(D=d \mid H=h) \cdot P(H=h) \Rightarrow \underline{\geq}$ #

3)

(a) $A = U \Lambda U^T$. $X^T \cdot A \cdot X = \sum_{i=1}^{d} X^T \cdot \lambda_i u_i u_i^T \cdot X = \sum_{i=1}^{d} \lambda_i (u_i^T X)^T u_i X$.

$\because (u_i^T X)^T \cdot u_i X > 0 \Rightarrow \underline{\lambda_i \geq 0 \Rightarrow A = PSD.}$

If A is PSD, then $X^T A \cdot X = \sum_{i=1}^{d} \lambda_i (u_i^T X)^T \cdot u_i X \geq 0$ $\underline{\text{for } \lambda_i \geq 0}$ #

(b) Same as (a) $\underline{\text{with } \lambda_i > 0.}$ #

4)

(a) $f(x) = a^T x + b \Rightarrow f(tx + (1-t)y) = a^T t x + a^T (1-t) y + b$

$= t a^T x + (1-t) a^T y + tb + (1-t) b \Rightarrow f(x)$ is convex and concave.

$f(x)$ is not strickly convex.

4)

(b) Let x_1, x_2 be global minimizer. $x' = \frac{x_1+x_2}{2}$. $\Rightarrow f(x') > f(x_1) = f(x_2)$

but $f\left(\frac{x_1+x_2}{2}\right) < \frac{1}{2} f(x_1) + \frac{1}{2} f(x_2)$ $\Rightarrow$ Contradiction. f has at most one global minimizer.

(c) $\nabla f(x^*) = 0$. Let $x = x^* + th$, where $t \in \mathbb{R}$, $h \in \mathbb{R}^n$

$$f(x^*+th) = f(x^*) + \langle \underbrace{\nabla f(x^*)}_{\to 0}, th \rangle + \frac{1}{2} \langle th, \nabla^2 f(x^*) th \rangle + o(\|th\|^2)$$

$$\Rightarrow 0 < \frac{f(x^*+th) - f(x^*)}{\|th\|^2} = \frac{(th)^T \cdot \nabla^2 f(x^*)(th)}{2\|th\|^2} + \underbrace{\frac{o(\|th\|^2)}{\|th\|^2}}_{\to 0}$$

$\Rightarrow 0 < (th)^T \cdot \nabla^2 f(x^*)(th)$, where $th \in \mathbb{R}^n$. $\Rightarrow$ $t^2 h^T \cdot \nabla^2 f(x^*) \cdot h > 0$ #

(d) f is convex $\Rightarrow f(x) \geq f(y) + \langle \nabla f(y), x-y \rangle \Rightarrow \langle x-y, \nabla^2 f(y)(x-y) \rangle \geq 0$

$\Rightarrow (x-y)^T \cdot \nabla^2 f(y) \cdot (x-y) \geq 0 \Rightarrow$ $\nabla^2 f(y)$ is PSD #

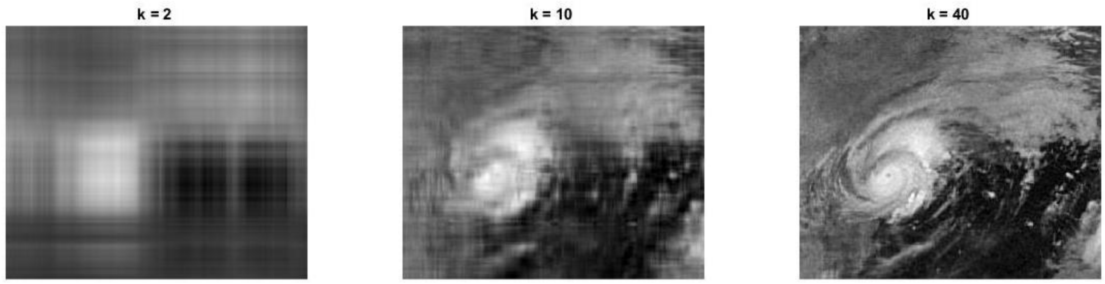
$\nabla^2 f(y)$ is PSD. $\langle x-y, \nabla^2 f(y+t(x-y))(x-y) \rangle \geq 0$.

$\Rightarrow f(x) \leq f(y) + \langle \nabla f(y), x-y \rangle \Rightarrow$ f is convex #.

(e) $f(x) = \frac{1}{2} x^T A x + b^T x + c$. $\nabla^2 f(x) = A$. # . f is $\begin{cases} \text{convex} & \text{if } A \text{ is PSD} \\ \text{strictly convex} & \text{PD.} \end{cases}$ #

5.

(a)



For $k \in \{2, 10, 40\}$, $$\frac{\|X-\bar{X}\|_F}{\|X\|_F} = [0.2815, 0.1587, 0.0837]$$

- Matlab Code

```
X = double(rgb2gray(imread('harvey-saturday-goes7am.jpg')));
[U,S,V] = svd(X);
k = [2 10 40];
X_approximate = zeros(size(X,1), size(X,2), size(k,2));
error = [];
for i = 1 : size(k,2)
        X_approximate(:, :, i) = U(:, 1:k(i))*S(1:k(i), 1:k(i))*transpose(V(:, 1:k(i)));
        error = [error norm(X - X_approximate(:,:,i),'fro')/norm(X,'fro')];
        subplot(1, size(k,2), i);
        imshow(uint8(X_approximate(:, :, i)));
        txt = sprintf('k = %d ', k(i));
        title(txt);
end
disp(error);
```

(b) The numbers (n) we need to describe the approximation

For k = 2, n = 1296*2 + 2 + 2*1548

For k = 10, n = 1296*10 + 10 + 10*1548

For k = 40, n = 1296*40 + 40 + 40*1548