EECS 545 Homework 1. Yuan Yin

1) (a) U is orthogonal $\Rightarrow UU^T = U^TU = I$.

Thus, $\|U\vec{x}\| = \sqrt{\langle U\vec{x}, U\vec{x}\rangle} = \sqrt{(U\vec{x})^T(U\vec{x})} = \sqrt{\vec{x}^TU^TU\vec{x}}$

$= \sqrt{\vec{x}^TI\vec{x}} = \sqrt{\vec{x}^T\vec{x}} = \|\vec{x}\|$

(b) Suppose an orthogonal matrix U has the form

$$U = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Thus, we know that

$$UU^T = \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} a & c \\ b & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_{2\times 2}.$$

$$\Leftrightarrow \begin{cases} a^2+b^2=1 & ① \\ c^2+d^2=1 & ② \\ ac+bd=0 & ③ \end{cases}$$

There is 3 equations but with four unknown variables.

Thus the degree of freedom is 1.

According to ①, we can assume $a=\cos\theta$, thus b have 2 solutions, which is $b=\sin\theta$ or $b=-\sin\theta$.

Since θ is an unknow variable. we can know that c,d can also be represented by θ. According to ② and ③

$$\begin{cases} c^2+d^2=1 \\ \cos\theta\cdot c+\sin\theta\cdot d=0 \end{cases} \quad \text{or} \quad \begin{cases} d^2+c^2=1 \\ \cos\theta\cdot c+(-\sin\theta)\cdot d=0. \end{cases}$$

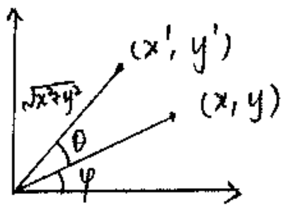
$$\Rightarrow \begin{cases} c=\sin\theta \\ d=-\cos\theta \end{cases} \quad \text{or} \quad \begin{cases} c=\sin\theta \\ d=\cos\theta. \end{cases} \quad (\theta\in\mathbb{R})$$

Thus, we solve U.

$$U = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \text{ or } \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix} \quad (\theta \in \mathbb{R})$$

Geometric interpretation:

① for $A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$



suppose there is a point (x, y) having an angle of φ with the positive x-axis.

Thus we know $x = \sqrt{x^2+y^2}\cos\varphi$.
$y = \sqrt{x^2+y^2}\sin\varphi$.

Transformation by A:

$$A\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x\cos\theta - y\sin\theta \\ x\sin\theta + y\cos\theta \end{bmatrix}$$

$$= \begin{bmatrix} \sqrt{x^2+y^2}\cos\theta\cos\varphi - \sqrt{x^2+y^2}\sin\theta\sin\varphi \\ \sqrt{x^2+y^2}\sin\theta\cos\varphi + \sqrt{x^2+y^2}\cos\theta\sin\varphi \end{bmatrix}$$

$$= \begin{bmatrix} \sqrt{x^2+y^2}\cos(\theta+\varphi) \\ \sqrt{x^2+y^2}\sin(\theta+\varphi) \end{bmatrix}$$

$$= \begin{bmatrix} x' \\ y' \end{bmatrix}$$

Thus the effect of A is a rotation, it rotates the points counterclockwise through an angle θ about the origin

② for $B = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$

we can decompose it by:

$$B = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = A \times \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

Thus it has ~~simen~~ similar part of effect with A.

for the other part:

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ -y \end{bmatrix}$$

Thus we can see the effect of this part is a reflection.

In conclusion, the effect of B has two steps, first it reflect the point with respect to the x-axis, and second it rotates the points counterclockwise through an angle θ about the origin.

(c) Let's first look at an original ellipse which is centralized at origin point, and major axis is on x-axis ~~HX~~, minor axis is on y-axis, it has form: $\frac{x^2}{a^2} + \frac{y^2}{b} = 1$.

According to the question, we choose $a = \frac{3}{2}$, $b = \frac{1}{2}$ so that we get an appropriate ellipse without movement or rotation.

Now we can write it as matrix form: let $\vec{x} = (x, y)^T$.

Then $\vec{x}^T \begin{bmatrix} \frac{4}{9} & 0 \\ 0 & 4 \end{bmatrix} \vec{x} = 1$ is just the same ellipse as above

Now, let's move it to center $[3, -1]^T$, then new ellipse:

$$(\vec{x} - \vec{c})^T \Lambda (\vec{x} - \vec{c}) = 1 \quad \text{where } \Lambda = \begin{bmatrix} \frac{4}{9} & 0 \\ 0 & 4 \end{bmatrix} \quad \vec{c} = [3\ -1]^T$$

At last, let's rotate so that the major axis makes an angle of $+\frac{\pi}{6}$ radians with positive x-axis. According to conclusion in (b), we know $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ has effect of rotation.

Thus we take $\theta = +\frac{\pi}{6}$ and $U' = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$, the new ellipse will be like:

$$[U'(\vec{x} - \vec{c})]^T \Lambda [U'(\vec{x} - \vec{c})] = 1.$$

$$\Rightarrow (\vec{x} - \vec{c})^T U'^T \Lambda U' (\vec{x} - \vec{c}) = 1.$$

2).

(A) (i) We know that $P_{X|Y}(x|y) = \frac{P_{X,Y}(x,y)}{P_Y(y)}$

Thus, $\mathbb{E}[X] = \int_X x \cdot P_X(x)\, dx$

$= \int_X x \cdot \left[\int_Y P_{X,Y}(x,y)\, dy\right] dx$

$= \int_X \int_Y x \cdot P_{X,Y}(x,y)\, dy\, dx.$

$= \int_Y \int_X x \cdot \frac{P_{X,Y}(x,y)}{P_Y(y)}\, dx \cdot P_Y(y)\, dy$

$= \int_Y \int_X x \cdot P_{X|Y}(x|y)\, dx \cdot P_Y(y)\, dy$

$= \int_Y \mathbb{E}_X[X|Y] \cdot P_Y(y)\, dy$

$= \mathbb{E}_Y[\mathbb{E}_X[X|Y]]$ ∎

(ii) $\mathbb{E}[\mathbb{I}[X \in \mathcal{C}]]$

$= 1 * \mathbb{P}[X \in C] + 0 * (1 - \mathbb{P}[X \in \mathcal{C}])$

$= \mathbb{P}(X \in \mathcal{C})$ ∎

(iii) $\mathbb{E}[XY] = \iint x \cdot y \cdot P(x,y)\, dx\, dy$

$= \iint x \cdot y \cdot P(x) \cdot P(y)\, dx\, dy$ (since X, Y are indep)

$= \int x \cdot P(x)\, dx \cdot \int y\, P(y)\, dy$

$= \mathbb{E}[X] \cdot \mathbb{E}[Y]$ ∎

Thus, in order to get the ellipse in the question,

we can choose $\vec{c} = [3\ -1]^T$, $r = 1$, $U = U'^T = \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix}$

$$\Lambda = \begin{bmatrix} \frac{4}{9} & 0 \\ 0 & 4 \end{bmatrix}$$

Note that this is only one of the solutions. we can choose different r and Λ actually

3.

(iv) since $X, Y \in \{0,1\}$ $X * Y = \begin{cases} 0 & X=0 \text{ or } Y=0. \\ 1 & X=1 \text{ and } Y=1 \end{cases}$

We know $\mathbb{E}[XY] = 0 * P(X=0 \text{ or } Y=0) + 1 * \mathbb{P}(X=1 \text{ and } Y=1)$

$= \mathbb{P}(X=1, Y=1)$

$\mathbb{E}[X]\mathbb{E}[Y] = (0 * \mathbb{P}(X=0) + 1 * \mathbb{P}(X=1)) * (0 \cdot P(Y=0) + 1 * P(Y=1))$

$= P(X=1) P(Y=1)$

According to the question, we have $P(X=1, Y=1) = P(X=1) P(Y=1)$ ⑤

Thus we also have

$P(X=1, Y=1) = (1 - P(X=0)) P(Y=1)$

$= P(Y=1) - P(X=0) P(Y=1)$. ①

since we know $P(Y=1) = P(X=0, Y=1) + P(X=1, Y=1)$. ②

$\Rightarrow$ ①+② : $P(X=0) P(Y=1) = P(X=0, Y=1)$. ③

with similar reason, we have $P(X=1) P(Y=0) = P(X=1, Y=0)$. ④

Thus ③+④+⑤ $\Rightarrow$ $P(X=0, Y=0) = P(X=0) P(Y=0)$.

$\Rightarrow$. Thus we can say that ~~$P(X) P(Y) = P(X=X)$~~

$P(X=x) P(Y=y) = P(X=x, Y=y)$

X, Y are independent

(b) (i) $P(H=h, D=d) \le P(H=h)$

(ii) $P(H=h \mid D=d)$ and $P(H=h)$ depend on the relation between H and D.

(iii) $P(H=h \mid D=d) \geq P(D=d \mid H=h) P(H=h)$

For (i)., since set $\{H=h, D=d\} \subseteq \{H=h\}$

the probability of left term will be alway equal or less than

the right term.

more briefly, we can know that

$$P(H=h) = P(H=h, D=d) + P(H=h, D \neq d)$$

And we also know $P(H=h, D\neq d) \geqslant 0$.

Thus $P(H=h, D=d) \leq P(H=h)$

3) (a) & (b).

" $\Rightarrow$ " If A is PSD (PD), then since $A = U \Lambda U^T$ (by Spectral theorem), where U is orthogonal matrix, we have

$UU^T = U^TU = I$. , $\vec{u}_i$ denotes the i-th column of U.

$$\Rightarrow \vec{u}_i \vec{u}_i^T = 1. = \vec{u}_i^T \vec{u}_i$$

Also since we have $A\vec{u}_i = \lambda_i \vec{u}_i$.

$$\Rightarrow \vec{u}_i^T A \vec{u}_i = \lambda_i$$

by A is PSD (PD) definition,

$$\Rightarrow \lambda_i = \vec{u}_i^T A \vec{u}_i \geqslant 0 \quad (>0).$$

" $\Leftarrow$ " if $\lambda_i \geqslant 0$ ($\lambda_i > 0$)., then for $\forall \vec{z} \in \mathbb{R}^d$ ($\vec{z} \neq \vec{0}$ when $\lambda_i > 0$)

we have
$$\vec{z}^T A \vec{z} = \vec{z}^T U \Lambda U^T \vec{z}$$
$$= \vec{z}^T \left(\sum_{i=1}^{d} \lambda_i \vec{u}_i \vec{u}_i^T\right) \vec{z}$$
$$= \sum_{i=1}^{d} \lambda_i (\langle \vec{z}, \vec{u}_i \rangle)^2 \geqslant 0 \quad (>0 \text{ when } \lambda_i > 0).$$

This is because $\vec{z} \neq 0$ and $\vec{u}_i$ $(i=1,\cdots,d) \in \mathbb{R}^d$ and orthogonal so $\langle \vec{z} \cdot \vec{u}_i \rangle^2 > 0$ must exist

Thus, A is PSD (PD)

4.

4) a) Assume there is at least two global minimizer $\vec{x}_1, \vec{x}_2$, then since f is strictly convex, we have

$$f(t\vec{x}_1+(1-t)\vec{x}_2) < tf(\vec{x}_1) + (1-t) f(\vec{x}_2) \qquad (\vec{x}_1 \neq \vec{x}_2).$$

Also since they are both global minimizer. $\Rightarrow f(\vec{x}_1) = f(\vec{x}_2)$
$= \min f(\vec{x})$

Thus $f(t\vec{x}_1+(1-t)\vec{x}_2) < tf(\vec{x}_1)+(1-t)f(\vec{x}_2) = f(\vec{x}_1) = f(\vec{x}_2)$

which is contradict with $\vec{x}_1, \vec{x}_2$ are global minimizer

Thus f has at most one global minimizer

(b) Assume $f(\vec{x})$ and $g(\vec{x})$ are two convex functions.

Suppose $\nabla^2 f(\vec{x}) = A$ and $\nabla^2 g(\vec{x}) = B$.

$\Rightarrow$ A and B are PSD (will be shown in (e))

Then the Hessian matrix of new function $h(\vec{x}) = f(\vec{x})+g(\vec{x})$ is

$$\nabla^2 h(\vec{x}) = \nabla^2 f(\vec{x}) + \nabla^2 g(\vec{x}) = A+B =: C$$

for $\forall \vec{z} \in \mathbb{R}^d$. $\vec{z}^T C \vec{z} = \vec{z}^T (A+B) \vec{z}$

$$= \underbrace{\vec{z}^T A \vec{z}}_{\geq 0} + \underbrace{\vec{z}^T B \vec{z}}_{\geq 0} \quad \text{(since } A, B \text{ are PSD).}$$

Thus we have $\vec{z}^T C \vec{z} \geq 0$.

$\Rightarrow$ C is a PSD matrix

$\Rightarrow$ $h(\vec{x})$ is a convex function (according to (e))

(c) Suppose $A = \begin{bmatrix} a_{11} & \cdots & a_{1d} \\ \vdots & & \\ a_{d1} & \cdots & a_{dd} \end{bmatrix}$ $\quad \vec{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_d \end{bmatrix}$

$$\Rightarrow f(\vec{x}) = \frac{1}{2} \begin{bmatrix} \sum_1^d a_{i_1 1} x_{i_1} \\ \sum_1^d a_{i_2 1} x_{i_2} \\ \vdots \\ \sum_1^d a_{i_d 1} x_{i_d} \end{bmatrix}^T \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix} + \sum_1^d b_i x_i + C$$

$$= \frac{1}{2} \sum_{j=1}^d \left(\sum_{i=1}^d a_{ij} x_i \right) \cdot x_j + \sum_{i=1}^d b_i x_i + C.$$

$$\Rightarrow \nabla^2 f(\vec{x}) = \begin{bmatrix} \frac{\partial^2 f(\vec{x})}{\partial x_1^2} & \cdots & \frac{\partial^2 f(\vec{x})}{\partial x_1 \partial x_d} \\ \vdots & & \\ \frac{\partial^2 f(\vec{x})}{\partial x_d^2} & \cdots & \frac{\partial^2 f(\vec{x})}{\partial x_d \partial x_d} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1d} \\ \vdots & & & \\ a_{d1} & \cdots & \cdots & a_{dd} \end{bmatrix} = A.$$

According conclusion in (e).

when A is PSD, f is convex

when A is PD, f is strictly convex

5.

(d) since $\vec{x}^*$ is a local minimizer

we have $\nabla f(\vec{x}^*) = 0$

Using second-order expansion:

$$f(\vec{x}^*+d) = f(\vec{x}^*) + \frac{1}{2} d^T \nabla^2 f(\vec{x}^*) d + o(\|d\|^2)$$

since $\vec{x}^*$ is local minimimizer

$$0 \le f(\vec{x}^*+d) - f(\vec{x}^*) \underset{\forall \|d\| \to 0}{\sim} \frac{1}{2} d^T \nabla^2 f(\vec{x}^*) d$$

Thus $d^T \nabla^2 f(\vec{x}^*) d$ is non-negative for any d (and $\|d\| \to 0$)

Thus at this point, $\nabla^2 (f(\vec{x}^*))$ is positive semi-definite

(e) Lemma: f is a convex function $\Leftrightarrow$ $\forall \vec{x}, \vec{y} \in S \subseteq \mathbb{R}^n$
$f(\vec{y}) \ge f(\vec{x}) + \nabla f(\vec{x})^T (\vec{y} - \vec{x})$

Pf of lemma: "$\Rightarrow$" f is convex, suppose $\vec{z} = \lambda \vec{y} + (1-\lambda)\vec{x}$ $\lambda \in [0,1]$

Then $f(\vec{z}) = f(\lambda \vec{y} + (1-\lambda)\vec{x}) \le \lambda f(\vec{y}) + (1-\lambda) f(\vec{x})$

Then $f(\vec{z}) - f(\vec{x}) \le \lambda f(\vec{y}) - \lambda f(\vec{x})$

From the lecture we know $\nabla f(\vec{x})^T \vec{y} = \lim\limits_{\lambda \to 0^+} \dfrac{f(\vec{x} + \lambda \vec{y}) - f(\vec{x})}{\lambda}$

$$\Rightarrow \nabla f(\vec{x})^T (\vec{y} - \vec{x}) = \lim_{\lambda \to 0^+} \frac{f(\vec{x} + \lambda(\vec{y} - \vec{x})) - f(\vec{x})}{\lambda} = \lim_{\lambda \to 0^+} \frac{f(\vec{z}) - f(\vec{x})}{\lambda}$$

$$\le f(\vec{y}) - f(\vec{x})$$

"$\Leftarrow$" still, suppose $\vec{z} = \lambda \vec{y} + (1-\lambda)\vec{x}$

$f(\vec{y}) \ge f(\vec{z}) + \nabla f(\vec{z})^T (\vec{y} - \vec{z})$ ①

$f(\vec{x}) \ge f(\vec{z}) + \nabla f(\vec{z})^T (\vec{x} - \vec{z})$ ②

λ① $+ (1-\lambda)$② : $\lambda f(\vec{y}) + (1-\lambda) f(\vec{x}) \ge f(\vec{z}) + \nabla f(\vec{z})^T (\lambda \vec{y} + (1-\lambda)\vec{x} - \vec{z})$

Thus $\lambda f(\vec{y}) + (1-\lambda) f(\vec{x}) \geq f(\lambda \vec{y} + (1-\lambda)\vec{x}) + 0$

Thus f is convex

Now we want to prove $\nabla^2(f(\vec{x}))$ is PSD $\Leftrightarrow$ f is convex.

"$\Rightarrow$" For some $\vec{z} \in [\vec{x}, \vec{y}]$, we have

$$f(\vec{y}) = f(\vec{x}) + \nabla f(\vec{x})^T (\vec{y} - \vec{x}) + \frac{1}{2}((\vec{y}-\vec{x})^T H_f(\vec{z})(\vec{y}-\vec{x}))$$

$H_f(\vec{z})$ is PSD. then: $\frac{1}{2}((\vec{y}-\vec{x})^T H_f(\vec{z})(\vec{y}-\vec{x})) \geq 0$

Thus $f(\vec{y}) \geq f(\vec{x}) + \nabla f(\vec{x})^T (\vec{y} - \vec{x})$

by lemma, f is convex

"$\Leftarrow$" for small $\lambda > 0$, $\vec{d} \in \mathbb{R}^n$, we have

$$f(\vec{x} + \lambda \vec{d}) = f(\vec{x}) + \lambda \nabla f(\vec{x})^T \vec{d} + \frac{1}{2}\lambda^2 \vec{d}^T H_f(\vec{x}) \vec{d} + o(\|\lambda \vec{d}\|^2)$$

by lemma, we have

$$f(\vec{x} + \lambda \vec{d}) \geq f(\vec{x}) + \lambda \nabla f(\vec{x})^T \vec{d}$$

Then we have for $\forall \vec{d} \in \mathbb{R}^n$,

$$\frac{1}{2}\lambda^2 \vec{d}^T H_f(\vec{x}) \vec{d} + o(\|\lambda \vec{d}\|^2) \geq 0.$$

$$\Rightarrow \quad \vec{d}^T H_f(\vec{x}) \vec{d} + \frac{2\, o(\|\lambda \vec{d}\|^2)}{\lambda^2} \geq 0.$$

take ~~lim~~ $\lambda \to 0^+$. we have $\lim_{\lambda \to 0} \frac{2\, o(\|\lambda \vec{d}\|^2)}{\lambda^2} = 0$

Thus $\vec{d}^T H_f(\vec{x}) \vec{d} \geq 0$, $H_f(\vec{x})$ is PSD ∎