Homework 1 – Convex Terminologies

[Pathri Vidya Praveen, CS24BTECH11047]

1. ⇒ Let $A \in R^{n \times n}$: A is diagonalizable if there exists an invertible matrix P such that

$$P^{-1}AP = D \longrightarrow A = PDP^{-1}$$

where D is a diagonal matrix.

⇒ Consider a 2×2 symmetric matrix $(A = A^T)$

$$A = \begin{bmatrix} a & c \\ c & b \end{bmatrix}, \quad a, b, c \in R$$

⇒ A is diagonalizable iff there are 2 linearly independent eigen vectors.

⇒ Let 2 eigenvalues be λ_1, λ_2 which satisfy $|A - \lambda I| = 0$.

$$\Rightarrow \begin{vmatrix} a-\lambda & c \\ c & b-\lambda \end{vmatrix} = 0 \quad \Rightarrow (a-\lambda)(b-\lambda) = c^2$$

$$\Rightarrow \lambda^2 - (a+b)\lambda + (ab - c^2) = 0.$$

$$\lambda = \frac{(a+b) \pm \sqrt{(a+b)^2 - 4(ab-c^2)}}{2}.$$

$$\lambda = \frac{a+b \pm \sqrt{(a-b)^2 + 4c^2}}{2} \qquad (\lambda_1 > \lambda_2 \text{ WLOG})$$

$$\lambda_1 = \frac{a+b+\sqrt{(a-b)^2+4c^2}}{2}, \quad \lambda_2 = \frac{a+b-\sqrt{(a-b)^2+4c^2}}{2}$$

$\Rightarrow$ We know that $(a-b)^2 + 4c^2 \geq 0 \ \forall \ a,b,c \in R$

$\Rightarrow \lambda_1 \in R, \ \lambda_2 \in R \longrightarrow$ $\boxed{\text{eigen values are always real}}$

<u>Case 1: $\lambda_1 = \lambda_2$:</u>

$\Rightarrow (a-b)^2 + 4c^2 = 0 \longrightarrow c = 0, \ a = b = \lambda = \lambda_1 = \lambda_2$

$\Rightarrow A = \lambda I$. So, now every non zero vector is eigen vector. We can pick 2 independent eigenvectors so A is diagonalizable.

<u>Case 2: $\lambda_1 \neq \lambda_2$:</u>

$\Rightarrow x_1, x_2$ are eigenvectors of λ_1, λ_2.

$\Rightarrow x_1^T A x_2 = (Ax_1)^T x_2$ $\quad \left[\begin{array}{l} Ax_1 = \lambda x_1 \\ Ax_2 = \lambda x_2 \end{array} \right.$

$x_1^T \lambda_2 x_2 = \lambda_1 x_1^T x_2$

$(\lambda_1 - \lambda_2) x_1^T x_2 = 0 \longrightarrow \lambda_1 = \lambda_2$ ✗

x_1, x_2 are orthogonal $\longleftarrow \boxed{x_1^T x_2 = 0}$

x_1, x_2 are 2 linearly independent eigenvectors

$\Rightarrow$ $\boxed{\text{A is diagonalizable}}$

2. $\Rightarrow$ Given an arbitrary non empty set $S \subseteq \mathbb{R}^n$. Define set of vectors in S and $\forall \underline{y} \in S$, $C = \underline{x} : \underline{x}^T \underline{y} \geq 0$.

$\Rightarrow C = \{\underline{x} \in \mathbb{R}^n \mid \underline{x}^T \underline{y} \geq 0 \quad \forall \underline{y} \in S\}$

$$= \bigcap_{\underline{y} \in S} \{\underline{x} \in \mathbb{R}^n \mid \underline{x}^T \underline{y} \geq 0\} \longrightarrow \text{halfspace}$$

(Intersection of halfspaces)

Cone:

$\Rightarrow$ Set C is a cone if $\forall \underline{x} \in C$, $\theta \geq 0$, $\theta \underline{x} \in C$.

$\Rightarrow$ ~~let~~ Take $\underline{x} \in C$, $\underline{x}^T \underline{y} \geq 0 \quad \forall \underline{y} \in S$.

Take $\theta \geq 0$. $\longrightarrow (\theta \underline{x})^T \underline{y} \geq 0 \quad \forall \underline{y} \in S$

$\theta \underline{x} \in C$.

$\Rightarrow \boxed{C \text{ is a cone}}$ — independent on structure of S

Convex set:

$\Rightarrow$ Set C is convex if for any distinct $\underline{x}_1, \underline{x}_2 \in C$, the line segment joining $\underline{x}_1, \underline{x}_2 \in C$ i.e $\theta \underline{x}_1 + (1-\theta) \underline{x}_2 \in C$ where $\theta \in [0,1]$

$\Rightarrow$ Take $\underline{x}_1 \in C$, $\underline{x}_2 \in C$, $\theta \in [0,1]$.

$\Rightarrow \underline{x}_1^T \underline{y} \geq 0$ and $\underline{x}_2^T \underline{y} \geq 0 \quad \forall \underline{y} \in S$

$\Rightarrow \theta \geq 0$, $\underline{x}_1^T \underline{y} \geq 0 \longrightarrow \theta \underline{x}_1^T \underline{y} \geq 0 \quad \forall \underline{y} \in S$ — ①

$\Rightarrow (1-\theta) \geq 0$, $\underline{x}_2^T \underline{y} \geq 0$

$\longrightarrow (1-\theta) \underline{x}_2^T \underline{y} \geq 0 \quad \forall \underline{y} \in S$ — ②

$\Rightarrow$ By eq. ① and ②, $\theta \underline{x}_1^T \underline{y} + (1-\theta) \underline{x}_2^T \underline{y} = (\theta \underline{x}_1 + (1-\theta) \underline{x}_2)^T \underline{y} \geq 0$

$\forall \underline{y} \in S$.

$\Rightarrow \theta \underline{x_1} + (1-\theta)\, \underline{x_2} \in C.$

$\Rightarrow$ $\boxed{C \text{ is convex set}}$ — independent of specific structure of set S

Affine set:

$\Rightarrow$ Set C is affine if $\forall\, \underline{x_1}, \underline{x_2} \in C$, $\theta \in \mathbb{R}$,

$\theta \underline{x_1} + (1-\theta)\underline{x_2} \in C.$

$\Rightarrow$ Take $\underline{x_1}, \underline{x_2} \in C$. $\underline{x_1}^T \underline{y} \geq 0$, $\underline{x_2}^T \underline{y} \geq 0 \;\; \forall \underline{y} \in S$

$\Rightarrow$ ~~Take~~ We proved above that

if $\theta \in [0,1]$, $\theta \underline{x_1} + (1-\theta)\underline{x_2} \in C$.

$\Rightarrow$ Take $\theta < 0$. $\theta \underline{x_1}^T \underline{y} \leq 0 \;\; \forall \underline{y} \in S$.

$(1-\theta)\, \underline{x_2}^T \underline{y} \geq 0 \quad \forall \underline{y} \in S.$

Take $\underline{x_2} = \underline{0} \in C$ $(\because \underline{0}^T \underline{y} = 0 \geq 0 \;\; \forall \underline{y} \in S)$

$(\theta \underline{x_1} + (1-\theta)\underline{x_2})^T \underline{y} = \theta \underline{x_1}^T \underline{y} \leq 0 \;\; \forall \underline{y} \in S.$

$\theta \underline{x_1} + (1-\theta)\underline{x_2} \notin C$ in general

$\Rightarrow$ $\boxed{C \text{ is not affine set in general}}$

$\Rightarrow$ Depends on specific structure of set S.

① $S = \{\underline{0}\}$, then $C = \mathbb{R}^n$ is affine

② If S is symmetric i.e. if $\underline{y} \in S$ then

$-\underline{y} \in S.$

$\underline{x}^T \underline{y} \geq 0$, $\underline{x}^T(-\underline{y}) \geq 0 \longrightarrow \underline{x}^T \underline{y} = 0$

$C = \{\underline{x} : \underline{x}^T \underline{y} = 0 \;\; \forall \underline{y} \in S\}$

This is affine.

Take $\theta \in \mathbb{R}$, $\underline{x_1}, \underline{x_2} \in C$.

$\underline{x_1}^T \underline{y} = 0$, $\underline{x_2}^T \underline{y} = 0$.

$(\theta \underline{x_1} + (1-\theta)\underline{x_2})^T \underline{y}$

$= \theta \underline{x_1}^T \underline{y} + (1-\theta)\underline{x_2}^T \underline{y} = 0 \geq 0$

$= \theta(\underline{0}) + (1-\theta)(\underline{0})$

$\theta \underline{x_1} + (1-\theta)\underline{x_2} \in C \longrightarrow C$ is affine

Subspace:

$\Rightarrow C \subseteq R^n$ is a subspace iff

① $\underline{0} \in C$ ② $\underline{x_1}, \underline{x_2} \in C \longrightarrow \underline{x_1} + \underline{x_2} \in C$

③ $\underline{x} \in C \Rightarrow \alpha \underline{x} \in C \;\forall\; \alpha \in R$

(or)

$\underline{x_1}, \underline{x_2} \in C \longrightarrow \alpha \underline{x_1} + \beta \underline{x_2} \in C,$

$\alpha, \beta \in R$

$\Rightarrow \underline{0} \in C$ because $\underline{0}^T \underline{y} = 0 \;\forall\; \underline{y} \in S$

$\Rightarrow \underline{x_1}, \underline{x_2} \in C \longrightarrow \underline{x_1}^T \underline{y} \geq 0, \underline{x_2}^T \underline{y} \geq 0$

$(\underline{x_1} + \underline{x_2})^T \underline{y} = \underline{x_1}^T \underline{y} + \underline{x_2}^T \underline{y} \geq 0 \;\forall\; \underline{y} \in S$

$\underline{x_1} + \underline{x_2} \in C$

$\Rightarrow \underline{x} \in C, \alpha \in R : \underline{x}^T \underline{y} \geq 0 \;\forall\; \underline{y} \in S$

But $\alpha < 0$ then $(\alpha \underline{x})^T \underline{y} \leq 0 \;\forall\; \underline{y} \in S$

$\alpha \underline{x} \notin C$

$\Rightarrow$ $\boxed{C \text{ is not subspace in general}}$

$\Rightarrow$ Depends on specific structure of set S

$\Rightarrow$ If S is symmetric,

$C = \{\underline{x} : \underline{x}^T \underline{y} = 0 \;\forall\; \underline{y} \in S\}$

(check affine case)

then if $\underline{x} \in C$, $\alpha \underline{x} \in C \;\forall\; \alpha \in R$

So, C is subspace

---

3. $\Rightarrow A \in R^{n \times n}$ is a $n \times n$ dimensional matrix

$\Rightarrow$ Vector $\underline{y} \in R^n$

Function f_1 convexity:

$\Rightarrow f_1(\underline{x}) = \| \underline{y} - A \underline{x} \|_2$

$\Rightarrow$ For f_1 to be convex,

$f_1(\theta \underline{x_1} + (1-\theta) \underline{x_2}) \leq \theta f_1(\underline{x_1}) + (1-\theta) f_1(\underline{x_2})$

where $\theta \in [0, 1]$

$\Rightarrow f_1(\theta \underline{x_1} + (1-\theta)\underline{x_2})$

$= \| \underline{y} - A(\theta \underline{x_1} + (1-\theta)\underline{x_2}) \|_2$

$= \| \theta \underline{y} + (1-\theta)\underline{y} - \theta A \underline{x_1} - (1-\theta) A \underline{x_2} \|_2$

$= \| \theta(\underline{y} - A\underline{x_1}) + (1-\theta)(\underline{y} - A\underline{x_2}) \|_2$

$\leq \| \theta(\underline{y} - A\underline{x_1}) \|_2 + \| (1-\theta)(\underline{y} - A\underline{x_2}) \|_2$

(By definition of norm, Δ^{le} inequality satisfied)

$\leq \theta \| \underline{y} - A\underline{x_1} \|_2 + (1-\theta) \| \underline{y} - A\underline{x_2} \|_2$

$\Rightarrow f_1(\theta \underline{x_1} + (1-\theta)\underline{x_2}) \leq \theta f_1(\underline{x_1}) + (1-\theta) f_1(\underline{x_2})$

$\Rightarrow \boxed{f_1 \text{ is convex}}$

(Here θ and $1-\theta$ are non-ve, so we can write $|\theta| = \theta$, $|1-\theta| = 1-\theta$)

<u>Function f_2 convexity:</u>

$\Rightarrow f_2(\underline{x}) = \| \underline{y} - A\underline{x} \|_2^2$

$$\Rightarrow f_2(\underline{x}) = \|\underline{y} - A\underline{x}\|_2^2$$

$$= \langle \underline{y} - A\underline{x}, \underline{y} - A\underline{x} \rangle$$

$$= \langle \underline{y}, \underline{y} \rangle - \cancel{A}\langle \underline{y}, A\underline{x} \rangle - \cancel{A}\langle A\underline{x}, \underline{y} \rangle + \langle A\underline{x}, A\underline{x} \rangle$$

$$= \underline{y}^T\underline{y} - \cancel{A}\,\underline{y}^T\cancel{x}A\underline{x} - (A\underline{x})^T\underline{y} + (A\underline{x})^T A\underline{x}$$

$(\underline{x}, \underline{y} \in \mathbb{R}^n$, so $\underline{x}^T\underline{y} = \underline{y}^T\underline{x})$

$$= \underline{y}^T\underline{y} - 2\cancel{A}\,\underline{y}^T A\underline{x} + \underline{x}^T A^T A\underline{x}$$

$$\Rightarrow f_2(\underline{x}) = \underline{y}^T\underline{y} - 2\underline{y}^T A\underline{x} + \underline{x}^T A^T A\underline{x} \quad — (1)$$

$$\Rightarrow \left(\nabla^2 f_2(\underline{x})\right)_{ij} = \frac{\partial}{\partial x_i}\left(\frac{\partial}{\partial x_j} f_2(\underline{x})\right)$$

↳ Hessian matrix compute!

$\Rightarrow \nabla^2(\underline{y}^T\underline{y}) = 0$ because it is constant

$$\Rightarrow \nabla^2(-2\underline{y}^T A\underline{x}) = \nabla^2\left(-2(A^T\underline{y})^T\underline{x}\right)$$

$$= \nabla\left(\nabla\left(-2(A^T\underline{y})^T\underline{x}\right)\right)$$

$$= \nabla(-2A^T\underline{y}) = 0 \quad (\because -2A^T\underline{y} \text{ is constant})$$

$$\Rightarrow \nabla^2(\underline{x}^T A^T A\underline{x}) = \nabla\left(\nabla(\underline{x}^T A^T A\underline{x})\right)$$

$$\cancel{= \nabla\left(\nabla\left(\sum_{i,j}(A^TA)_{ij}\, x_i x_j\right)\right)}$$

$$= \frac{\partial}{\partial x_i}\left(\frac{\partial}{\partial x_j}\left(\sum_{i,j}(A^TA)_{ij}\, x_i x_j\right)\right)$$

$\Rightarrow$ First we compute gradient.

$$\cancel{\frac{\partial}{\partial x_k}\sum_{i,j}(A^TA)_{ij}\, x_i x_j = \sum_j}$$

$$\frac{\partial}{\partial x_k}\sum_{i,j}(A^TA)_{ij}\, x_i x_j = \sum_j (A^TA)_{kj}\, x_j + \sum_i (A^TA)_{ik}\, x_i$$

$$\frac{\partial}{\partial x_k} \sum_{i,j} (A^T A)_{ij}\, x_i x_j = 2 \sum_j (A^T A)_{kj}\, x_j$$

$$\frac{\partial f_2}{\partial x_k} = 2 \sum_j (A^T A)_{kj}\, x_j = 2(A^T A \underline{x})_k$$

$$\nabla (\underline{x}^T A^T A \underline{x}) = 2 A^T A \underline{x}$$

$$\Rightarrow \nabla(\nabla(\underline{x}^T A^T A \underline{x})) = \nabla(2 A^T A \underline{x})$$

$$= \nabla(2(A^T A)^T \underline{x}) = 2 A^T A$$

$$\Rightarrow \boxed{\nabla^2 (\underline{x}^T A^T A \underline{x}) = 2 A^T A}$$

⇒ Now substitute all these Hessians of individual terms into final hessian.

⇒ By ①,

$$\boxed{\nabla^2 f_2(\underline{x}) = 2 A^T A} \qquad 2A^T A \succeq 0$$

⇒ $2A^T A$ is positive semi definite

⇒ $\boxed{f_2(\underline{x}) \text{ is convex}}$ because Hessian of f_2 is positive semi definite

---

4. ⇒ $f : \mathbb{R}^n \to \mathbb{R}$ is a differentiable function. with well-defined gradient at all points.

⇒ For any point $\underline{x} \in \mathbb{R}^n$, any direction $\underline{u} \in \mathbb{R}^n$ with $\|\underline{u}\|_2 = 1$.

⇒ Directional derivative of f at $\underline{x}$ in the direction $\underline{u}$ by

$$f'_{\underline{u}}(\underline{x}) = \lim_{\varepsilon \to 0} \frac{f(\underline{x} + \varepsilon \underline{u}) - f(\underline{x})}{\varepsilon}$$

(1.) $\Rightarrow f$ is differentiable at x.

$$\Rightarrow f(\underline{x}+\underline{h}) = f(\underline{x}) + \nabla f(\underline{x})^T \underline{h} + o(\|h\|)$$

o little-o as $h \to 0$

$\Rightarrow$ Take $\underline{h} = \varepsilon \underline{u}$

$$\Rightarrow f(\underline{x}+\varepsilon\underline{u}) = f(\underline{x}) + \nabla f(\underline{x})^T \varepsilon \underline{u} + o(\|\varepsilon \underline{u}\|)$$

$$= f(\underline{x}) + \nabla f(\underline{x})^T \varepsilon \underline{u} + o(\varepsilon)$$

$$\Rightarrow \frac{f(\underline{x}+\varepsilon\underline{u}) - f(\underline{x})}{\varepsilon} = \nabla f(\underline{x})^T \underline{u} + \frac{o(\varepsilon)}{\varepsilon}$$

$$\Rightarrow \lim_{\varepsilon\to 0} \frac{f(\underline{x}+\varepsilon\underline{u}) - f(\underline{x})}{\varepsilon} = \lim_{\varepsilon\to 0}\left(\nabla f(\underline{x})^T \underline{u} + \frac{o(\varepsilon)}{\varepsilon}\right)$$

$$\Rightarrow \boxed{f_{\underline{u}}'(\underline{x}) = \nabla f(\underline{x})^T \underline{u}} \quad \left(\because \lim_{\varepsilon\to 0}\frac{o(\varepsilon)}{\varepsilon} = 0\right)$$

(2.) $\Rightarrow \nabla f(\underline{x})^T \underline{u} = \langle \nabla f(\underline{x}), \underline{u} \rangle$

$\Rightarrow$ euclidean inner product

$$\Rightarrow \boxed{f_{\underline{u}}'(\underline{x}) = \langle \nabla f(\underline{x}), \underline{u} \rangle}$$

(3.) $\Rightarrow \|\underline{u}\|_2 = 1$, direction $\underline{u}$ for which $f_{\underline{u}}'(\underline{x})$ is maximized

$\Rightarrow f_{\underline{u}}'(\underline{x}) = \langle \nabla f(\underline{x}), \underline{u} \rangle$ is maximized when $\boxed{\underline{u} \text{ is in direction of gradient } \nabla f(\underline{x})}$

$\Rightarrow f_{\underline{u}}'(\underline{x}) = \langle \nabla f(\underline{x}), \underline{u} \rangle \leq \|\nabla f(\underline{x})\|_2 \|\underline{u}\|_2$

(Cauchy Schwartz inequality)

$\leq \|\nabla f(\underline{x})\|_2 \cdot 1$ $\Rightarrow$ ~~$f_{\underline{u}}'(\underline{x}) \leq \|\nabla f(\underline{x})\|_2$~~

$\Rightarrow f_{\underline{u}}'(\underline{x}) \leq \|\nabla f(\underline{x})\|_2$ and $f_{\underline{u}}'(\underline{x}) = \|\nabla f(\underline{x})\|$

if $\underline{u}$ is in direction of $\nabla f(\underline{x})$

$$\Rightarrow \boxed{\underline{u} = \frac{\nabla f(\underline{x})}{\|\nabla f(\underline{x})\|_2}} \quad \text{as } \|\underline{u}\|_2 = 1$$

(4.) ⇒ direction $\underline{u}$ for which directional derivative is minimized.

⇒ $f_{\underline{u}}'(\underline{x}) = \langle \nabla f(\underline{x}), \underline{u} \rangle$ is minimum when $\boxed{\underline{u} \text{ is opposite direction to gradient } \nabla f(\underline{x})}$

⇒ $f_{\underline{u}}'(\underline{x}) = \langle \nabla f(\underline{x}), \underline{u} \rangle \geq -\|\nabla f(\underline{x})\|_2 \|\underline{u}\|_2$

$\geq -\|\nabla f(\underline{x})\|_2$

⇒ $f_{\underline{u}}'(\underline{x}) \geq -\|\nabla f(\underline{x})\|_2$ and equality if $\underline{u}$ is opposite to ~~φ~~ $\nabla f(\underline{x})$ i.e

(5) $\boxed{\underline{u} = -\dfrac{\nabla f(\underline{x})}{\|\nabla f(\underline{x})\|_2}}$

$\|\underline{u}\|_2 = 1$

(5.) ⇒ Direction of steepest ascent of f at $\underline{x}$ is direction in which function ↑ at greatest rate.
direction in which directional derivative is maximized.
i.e $\underline{u}$ where directional derivative

$$\text{Direction of steepest ascent} = \frac{\nabla f(\underline{x})}{\|\nabla f(\underline{x})\|_2}$$

$$\text{Rate of increase} = \left\langle \nabla f(\underline{x}), \frac{\nabla f(\underline{x})}{\|\nabla f(\underline{x})\|_2} \right\rangle$$

$$= \frac{\|\nabla f(\underline{x})\|_2^2}{\|\nabla f(\underline{x})\|_2} = \|\nabla f(\underline{x})\|_2$$

⇒ Direction of steepest descent of f at $\underline{x}$ is direction in which function ↓ at ~~least~~ greatest rate.
direction in which directional derivative is minimized.
i.e $\underline{u}$ in which directional derivative is minimized.

$$\text{Direction of steepest descent} = -\frac{\nabla f(\underline{x})}{\|\nabla f(\underline{x})\|_2}$$

$$\text{Rate of decrease} = \left\langle \nabla f(\underline{x}), -\frac{\nabla f(\underline{x})}{\|\nabla f(\underline{x})\|_2} \right\rangle$$

$$= \frac{-\|\nabla f(\underline{x})\|_2^2}{\|\nabla f(\underline{x})\|_2} = -\|\nabla f(\underline{x})\|_2$$