Stanford. CS229. 2018. autumn. ps0.

1. $f: \mathbb{R}^n \to \mathbb{R}$

(a). $f(x) = \frac{1}{2}x^T A x + b^T x$., $A \in \mathbb{S}$. $b \in \mathbb{R}^n$.

$$\nabla f(x) = \nabla_x\left(\frac{1}{2}x^T A x + b^T x\right).$$

$$= \begin{bmatrix} \frac{\partial}{\partial x_1}\left(\frac{1}{2}x^T A x + b^T x\right) \\ \vdots \\ \frac{\partial}{\partial x_n}\left(\frac{1}{2}x^T A x + b^T x\right) \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\partial}{\partial x_1}\left(\frac{1}{2}\sum_i^n\sum_j^n A_{ij}x_i x_j + \sum_i^n b_i x_i\right). \\ \vdots \\ \frac{\partial}{\partial x_n}\left(\frac{1}{2}\sum_i^n\sum_j^n A_{ij}x_i x_j + \sum_i^n b_i x_i\right) \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\partial}{\partial x_1}\left(\frac{1}{2}\left(\sum_j^n A_{1j}x_1 x_j + \sum_i^n A_{i1}x_i x_1 + \sum\sum A_{ij}x_i x_j\right) + \sum_i^n b_i x_i\right) \\ \vdots \\ \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{2}\left(\sum_j^n A_{1j}x_j + \sum_i^n A_{i1}x_i\right) + b_1 \\ \vdots \\ \frac{1}{2}\sum_j \quad \vdots \end{bmatrix} \qquad \frac{\partial}{\partial x} A_{11}x_1^2 = 2A_{11}x_1$$

$$= \begin{bmatrix} \sum_i^n A_{i1}x_i + b_1 \\ \vdots \\ \sum_i^n A_{in}x_i + b_n \end{bmatrix}$$

$$= \begin{bmatrix} \sum_i^n A_{i1}x_i \\ \vdots \\ \sum_i^n A_{in}x_i \end{bmatrix} + \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

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

$$= A^T x + b$$

$$= Ax + b.$$

$$[x_1, x_2, \cdots, x_n] \cdot \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & & \vdots \\ A_{n1} & \cdots & A_{nn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$= \left[\sum_i^n A_{i1} \cdot x_i \,,\ \sum_i^n A_{i2} \cdot x_i \,,\ \cdots,\ \sum_i^n A_{in} x_i \right] \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$= \sum_i^n \sum_j^n A_{ij} x_i x_j$$

(b). $f(x) = g(h(x))$, $g: \mathbb{R} \to \mathbb{R}$, $h: \mathbb{R}^n \to \mathbb{R}$

$$\nabla f(x) = \nabla_x g(h(x))$$

$$= \begin{bmatrix} \frac{\partial}{\partial x_1} g(h(x)) \\ \vdots \\ \frac{\partial}{\partial x_n} g(h(x)) \end{bmatrix}$$

$$\frac{\partial}{\partial x_i} g(h(x)) = \frac{\partial g(h(x))}{\partial h(x)} \cdot \frac{\partial h(x)}{\partial x_i}$$

$$= \begin{bmatrix} \frac{\partial}{\partial h(x)} g(h(x)) \cdot \frac{\partial}{\partial x_1} h(x) \\ \vdots \\ \vdots \end{bmatrix}$$

$$= \frac{\partial}{\partial h(x)} g(h(x)) \cdot \begin{bmatrix} \frac{\partial}{\partial x_1} h(x) \\ \vdots \\ \frac{\partial}{\partial x_n} h(x) \end{bmatrix}$$

$$= \frac{\partial}{\partial h(x)} g(h(x)) \nabla h(x)$$

(c)

$f(x) = \frac{1}{2} x^T A x + b^T x$, $A \in \mathbb{S}^{n \times n}$, $b \in \mathbb{R}^n$

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

$$= \begin{bmatrix} \frac{\partial}{\partial x_1} \frac{\partial}{\partial x_1} f(x) & \cdots & \frac{\partial}{\partial x_1} \frac{\partial}{\partial x_n} f(x) \\ \vdots & & \vdots \\ \frac{\partial}{\partial x_n} \frac{\partial}{\partial x_1} f(x) & \cdots & \frac{\partial}{\partial x_n} \frac{\partial}{\partial x_n} f(x) \end{bmatrix}$$

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

$$= A.$$

(d)

$f(x) = g(a^T x)$, $g : \mathbb{R} \to \mathbb{R}$, $a \in \mathbb{R}^n$

$$\nabla f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} g(a^T x) \\ \vdots \\ \frac{\partial}{\partial x_n} g(a^T x) \end{bmatrix}$$

$$\frac{\partial}{\partial x_i} g(a^T x) = \frac{\partial g(a^T x)}{\partial a^T x} \cdot \frac{\partial a^T x}{\partial x_i}$$

$$= \frac{\partial g(a^T x)}{\partial a^T x} \cdot a_i$$

$$= \begin{bmatrix} g'(a^T x) \cdot a_1 \\ \vdots \\ g'(a^T x) \cdot a_n \end{bmatrix}$$

$$= g'(a^T x) \cdot \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$$

$$= g'(a^T x) \cdot a$$

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1 \partial x_1} g(a^T x) & \cdots & \frac{\partial}{\partial x_1 \partial x_n} g(a^T x) \\ \vdots & & \vdots \\ \frac{\partial}{\partial x_n \partial x_1} g(a^T x) & \cdots & \frac{\partial}{\partial x_n \partial x_n} g(a^T x) \end{bmatrix}$$

$$\begin{aligned} \frac{\partial^2}{\partial x_i \partial x_j} g(a^T x) &= \frac{\partial}{\partial x_i} \cdot \frac{\partial}{\partial x_j} g(a^T x) \\ &= \frac{\partial}{\partial x_i} \cdot g'(a^T x) \cdot a_j \\ &= a_j \frac{\partial}{\partial x_i} g'(a^T x) \\ &= a_j \frac{\partial}{\partial a^T x} \cdot g'(a^T x) \cdot \frac{\partial}{\partial x_i} a^T x \\ &= a_j \cdot g''(a^T x) \cdot a_i \\ &= a_i \cdot a_j \cdot g''(a^T x) \end{aligned}$$

$$= \begin{bmatrix} a_1 \cdot a_1 \cdot g''(a^T x) & \cdots & a_1 a_n g''(a^T x) \\ \vdots & & \vdots \\ a_n a_1 g''(a^T x) & \cdots & a_n a_n g''(a^T x) \end{bmatrix}$$

$$= g''(a^T x) \begin{bmatrix} a_1 a_1 & \cdots & a_1 a_n \\ \vdots & & \vdots \\ a_n a_1 & \cdots & a_n a_n \end{bmatrix}$$

$$= g''(a^T x)\, a\, a^T$$

2.

(a).

$z \in \mathbb{R}^n$

$A = zz^T \succeq 0$?

let $x \in \mathbb{R}^n$.

$$
\begin{aligned}
x^T A x &= x^T z z^T x \\
&= \sum_i^n x_i z_i \cdot \sum_j^n x_j z_j \\
&= \left(\sum_i^n x_i z_i\right)^2 \geq 0.
\end{aligned}
$$

so, $A \succeq 0$.

(b)

~~$z \in \mathbb{R}^n$, $z_i \neq 0$ for i in n. $A = zz^T$~~

~~let $x \in \mathbb{R}^n$, $x_i \neq 0$ for i in n.~~

~~$x^T A x = x^T z z^T x = (\sum_i^n x_i z_i)^2 > 0$~~

~~so, $A \succ 0$, A is positive.~~

nullspace :

let $x \in \mathbb{R}^n$, $Ax = z z^T x$

$$= z \cdot \sum_i^n x_i z_i$$

if $Ax = 0$. $\longrightarrow$ $\cancel{x}$ $\sum_i^n x_i z_i = 0$

$z^T x = 0$

$\mathcal{N}(A) = \{ x \in \mathbb{R}^n : z^T x = 0 \}$.

rank :

~~because A is positive~~

~~so A is full rank~~

$$
\begin{aligned}
\text{rank}(A) &= \text{rank}(z z^T) \\
&\leq \min(\text{rank}(z), \text{rank}(z^T)) \\
&= 1
\end{aligned}
$$

(C).

$A \in \mathbb{R}^{n\times n}$, $A \succeq 0$, $B \in \mathbb{R}^{m\times n}$

BAB^T : $\cancel{\mathbb{R}^{m\times m}}$ PSD?

let $x \in \mathbb{R}^n$,

$x^T A x \geq 0$.

if BAB^T is PSD:

$x \in \mathbb{R}^m$. $x^T B A B^T x$

$(x^T B) A (B^T x)$

$(B^T x)^T A (B^T x)$

$B^T x = y \in \mathbb{R}^n$. $y^T A y \geq 0$.

So BAB^T is PSD.

3.

(a). $A \in \mathbb{R}^{n \times n}$, $A = T \Lambda T^{-1}$

$T \in \mathbb{R}^{n \times n}$, $\Lambda = diag(\lambda_1, \dots, \lambda_n)$.

$T = [t^{(1)}, \dots, t^{(n)}]$, $t^{(i)} = \mathbb{R}^n$

show: $At^{(i)} = \lambda_i t^{(i)}$

$$A = T \Lambda T^{-1}$$

$$AT = T \Lambda.$$

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \cdot [t^{(1)} \dots t^{(n)}] = [t^{(1)} \dots t^{(n)}] \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}$$

$$At^{(i)} = t^{(i)} \lambda_i \quad \text{(for each i in n.)}$$

$$At^{(i)} = \lambda_i t^{(i)}$$

(b).

$A^T = A$, $U = [u^{(1)} \dots u^{(n)}]$, $U^T U = I$.

$A = U \Lambda U^T$ $\quad u^{(i)} \in \mathbb{R}^n$

$\Lambda = diag(\lambda, \dots, \lambda_n)$

show: $Au^{(i)} = \lambda_i u^{(i)}$

$$A = U \Lambda U^T$$

$$AU = U \Lambda$$

$$A[u^{(1)} \dots u^{(n)}] = [u^{(1)} \dots u^{(n)}] \cdot diag(\lambda_1, \dots, \lambda_n)$$

$$[Au^{(1)} \dots Au^{(n)}] = [\lambda_1 u^{(1)} \dots \lambda_n u^{(n)}].$$

$$Au^{(i)} = \lambda_i u^{(i)}$$

so $u^{(i)}$ is an eigen vector of A.

(c) A is PSD, $A \succeq 0$.

$x \in \mathbb{R}^n$, $x^T A x \geq 0$.

$\lambda_i(A)$ means ith eigenvalue of A.

show: $\lambda_i(A) \geq 0$ for each i, $\lambda_i(A) = \lambda_i$

$$A t^{(i)} = \lambda_i t^{(i)}$$

$$(t^{(i)})^T A t^{(i)} \geq 0.$$

$$(t^{(i)})^T \lambda_i t^{(i)} \geq 0$$

$$\lambda (t^{(i)})^T t^{(i)} \geq 0.$$

$$\therefore \lambda \|t^{(i)}\|_2^2 \geq 0.$$

$$\lambda \geq 0$$