A0 ~~Assig~~ Problem Set 0 [Time given : around 10 days]
software used to grade assignments : Gradescope.
Marks : 0 points.
It is just to make students familiar with Gradescope.

Problem Set #0 : Linear Algebra and Multivariable Calculus.

1. If $f : \mathbb{R}^n \to \mathbb{R}$
then gradient $\nabla f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{bmatrix}$ where $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

The hessian $\nabla^2 f(x)$ is given as:-

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

9/01

PAGE NO. :
DATE : / /

(a) Let $f(x) = \frac{1}{2}x^TAx + b^Tx$ where A is symmetric matrix and $b \in \mathbb{R}^n$ is a vector. What is $\nabla f(x)$?

→ clearly $f: \mathbb{R}^n \longrightarrow \mathbb{R}$.

$$\begin{array}{cc} & \begin{array}{ccc} x_1 & \cdots & x_n \end{array} \\ x^TAx := \begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_n \end{array} & \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \end{array}$$

$$x^TAx = x_1^2 a_{11} + x_1x_2(a_{12}+a_{21}) + \cdots + x_1x_n(a_{1n}+a_{n1}) + x_2^2 a_{22} + x_{23}(a_{23}+a_{32}) + \cdots + x_2x_n(a_{2n}+a_{n2}) + \cdots + x_n^2 a_{nn}.$$

$$\nabla x^TAx = \begin{bmatrix} 2x_1a_{11} + x_2(a_{12}+a_{21}) + \cdots + x_n(a_{1n}+a_{n1}) \\ x_1(a_{12}+a_{21}) + 2x_2a_{22} + \cdots + x_n(a_{2n}+a_{n2}) \\ \vdots \\ x_1(a_{1n}+a_{n1}) + x_2(a_{2n}+a_{n2}) + \cdots + 2x_n(a_{nn}) \end{bmatrix}$$

Since A is a symmetric matrix $\Rightarrow A = A^T$

Thus

$$\nabla x^TAx = \begin{bmatrix} 2a_{11} & 2a_{12} & \cdots & 2a_{1n} \\ 2a_{21} & 2a_{22} & \cdots & 2a_{2n} \\ \vdots & \vdots & & \vdots \\ 2a_{n1} & 2a_{n2} & \cdots & 2a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$= 2Ax$$

$\nabla b^Tx = ?$

$$b^Tx = \begin{bmatrix} b_1 & b_2 & \cdots & b_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = b_1x_1 + b_2x_2 + \cdots + b_nx_n$$

$$\nabla b^T x = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = b.$$

Thus.

$$\nabla f(x) = \frac{1}{2}(2Ax) + b. = Ax + b.$$

(b) $f(x) = g(h(x)) = g \circ h(x)$

$g: \mathbb{R} \to \mathbb{R}$ is differentiable and.

$h: \mathbb{R}^n \to \mathbb{R}$ is differentiable.

→ What is $\nabla f(x)$?

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

• $\nabla f(x) = \nabla g(h(x)) \, \nabla h(x)$

$\Rightarrow \nabla f(x) = g'(h(x)) \, \nabla h(x)$

simply a fxn of $h(x)$

(c) Let $f(x) = \frac{1}{2} x^T A x + b^T x$ where A is symmetric and $b \in \mathbb{R}^n$ is a vector. What is $\nabla^2 f(x)$?

$$\rightarrow \nabla^2 \frac{1}{2}(x^T A x) = \frac{1}{2} \begin{bmatrix} 2a_{11} & 2a_{12} & \cdots & 2a_{1n} \\ 2a_{21} & 2a_{22} & \cdots & 2a_{2n} \\ \vdots & \vdots & & \\ 2a_{n1} & 2a_{n2} & \cdots & 2a_{nn} \end{bmatrix} = A.$$

$$\nabla^2 b^T x = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} = 0$$

$$\Rightarrow \nabla^2 f(x) = A.$$

(d) $f(x) = g(a^T x)$

$g: \mathbb{R} \to \mathbb{R}$. is continuously differentiable and $a \in \mathbb{R}^n$ is a vector. What are $\nabla f(x)$ and $\nabla^2 f(x)$?

$$\nabla f(x) = g'(a^T x) \; \nabla a^T x.$$
$$= g'(a^T x)\, a$$

$$\cancel{\nabla^2 f(x) = g'(f(x)) \nabla^2 f(x) + g''(f(x)) \nabla f(x) \nabla f(x)^T}$$

$$\cancel{\nabla^2 f(x) = g'(a^T x) \nabla^2 (a^T x) + g''(a^T x)}$$

let $h(x) = g(f(x))$ $\quad f: \mathbb{R}^n \to \mathbb{R}, \; g: \mathbb{R} \to \mathbb{R}$

$$\nabla^2 h(x) = g'(f(x)) \nabla^2 f(x) + g''(f(x)) \nabla f(x) \nabla f(x)^T$$

$$\nabla^2 h(x) = \begin{bmatrix} \frac{\partial^2}{\partial x_1^2} h(x) & \frac{\partial^2}{\partial x_1 \partial x_2} h(x) & \cdots & \frac{\partial^2}{\partial x_1 \partial x_n} h(x) \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2}{\partial x_n^2} h(x) & \frac{\partial^2}{\partial x_n \partial x_2} h(x) & \cdots & \frac{\partial^2}{\partial x_n^2} h(x) \end{bmatrix}$$

$$\frac{\partial}{\partial x_1^2} h(x) = \frac{\partial}{\partial x_1^2} g(f(x)) = \frac{\partial}{\partial x_1}\left(\frac{\partial}{\partial x_1} g(f(x))\right)$$

$$= \frac{\partial}{\partial x_1}\left(g'(f(x)) \frac{\partial}{\partial x_1} f(x)\right).$$

$$= g''(f(x)) \frac{\partial}{\partial x_1} f(x) \frac{\partial}{\partial x_1} f(x) + g'(f(x)) \frac{\partial^2}{\partial x_1^2} f(x)$$

PAGE NO.:
DATE: / /

$$\frac{\partial^2}{\partial x_1 \partial x_2} h(x) = \frac{\partial}{\partial x_1}\left(\frac{\partial}{\partial x_2} h(x)\right)$$

$$= \frac{\partial}{\partial x_1}\left(\frac{\partial}{\partial x_2} g(f(x))\right)$$

$$= \frac{\partial}{\partial x_1}\left(g'(f(x)) \frac{\partial}{\partial x_2} f(x)\right)$$

$$= g''(f(x)) \frac{\partial}{\partial x_1} \cancel{g'(f(x))} \frac{\partial}{\partial x_2} f(x) +$$

$$g'(f(x)) \frac{\partial^2}{\partial x_1 \partial x_2} f(x).$$

Thus generalizing we get

$i=j$

$$\frac{\partial}{\partial x_i^2} h(x) = g''(f(x)) \frac{\partial}{\partial x_j} f(x) \frac{\partial}{\partial x_i} f(x) + g'(f(x)) \frac{\partial^2}{\partial x_i^2} f(x)$$

$i \neq j$

$$\frac{\partial^2}{\partial x_i \partial x_j} h(x) = g''(f(x)) \frac{\partial}{\partial x_i} f(x) \frac{\partial}{\partial x_j} f(x) + g'(f(x)) \frac{\partial^2 f(x)}{\partial x_i \partial x_j}$$

Thus

$$\nabla^2 h(x) = g'(f(x)) \nabla^2 f(x) + g''(f(x)) \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{bmatrix} \begin{bmatrix} \frac{\partial}{\partial x_1} f(x), \dots \frac{\partial f(x)}{\partial x_n} \end{bmatrix}$$

$$\nabla^2 h(x) = g'(f(x)) \nabla^2 f(x) + g''(f(x)) \nabla f(x) \nabla f(x)^T$$

Thus for our question :-

$$\nabla^2 (g(a^T x)) = g'(a^T x) \nabla^2 (a^T x) + g''(a^T x) \nabla(a^T x) (\nabla(a^T x))^T$$

$$= g'(a^T x)\, 0 + g''(a^T x)\, a a^T$$

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

Q2 Positive definite matrices

(a) let $z \in \mathbb{R}^n$ be an n-vector. Show that $A = zz^T$ is positive semidefinite.

$\rightarrow$ $A = A^T \quad (zz^T = (zz^T)^T)$

$A = zz^T$

$A^T = (zz^T)^T = (z^T)^T z^T = zz^T$

$x^T A x = x^T (zz^T) x$

$= (z^T x)^T (z^T x)$

$= \|z^T x\|_2^2 \geq 0 \quad \forall \; x \in \mathbb{R}^n$.

(b) let $z \in \mathbb{R}^n$ be a non-zero n-vector. Let $A = zz^T$. What is the null-space of A? What is rank of A?

$\rightarrow$ Null space of A $= \{x \in \mathbb{R}^n \mid Ax = 0\}$.

$\Rightarrow zz^T x = 0$

$\Rightarrow (z^T x) z = 0$

since z is a non-zero n-vector

$\Rightarrow z^T x = 0$

Thus Null space of $A = \{x \in \mathbb{R}^n \mid z^T x = 0\}$.

$$[z_1 \; z_2 \; \cdots \; z_n] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = 0$$

$z_1 x_1 + z_2 x_2 + \cdots + z_n x_n = 0$

~~Thus~~.

Since z_i's $\neq 0$.

Thus. $x_1 = \frac{z_2}{z_1} x_2 + \dots + \frac{z_n}{z_1} x_n$.

~~Hence Null space =~~

(b) What is the rank of A.

$$n = \text{rank} + \text{Nullity}.$$

$$\text{nullity} = n-1$$

$\Rightarrow$ rank $= 1$.

Simply also $A = ZZ^T$

$$= \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} \begin{bmatrix} z_1 & \dots & z_n \end{bmatrix}$$

$$= \begin{bmatrix} z_1 \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} & \dots & z_n \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} \end{bmatrix}.$$

Thus rank $= 1$.

(c) Let $A \in \mathbb{R}^{n \times n}$ be positive semidefinite and $B \in \mathbb{R}^{m \times n}$ be arbitrary, where $m, n \in \mathbb{N}$. Is BAB^T PSD? If so prove it. If not, give a counter example with explicit A and B.

$\rightarrow$ $(BAB^T)^T = (B^T)^T A^T B^T = BA^T B^T = BAB^T$

$x^T (BAB^T)^T x = x^T (BAB^T) x$.

$= (B^T x) \, A \, (B^T x) \geq 0$

for ~~any~~ all $B^T x \in \mathbb{R}^n$

can any $y \in \mathbb{R}^n$ be written as $B^T x$. ie does ~~for~~ every $y \in \mathbb{R}^n$, $\exists x \in \mathbb{R}^n$ s.t $y = B^T x$. This is not always true!

PAGE NO.:
DATE: / /

Q3 Eigenvectors, Eigenvalues and Spectral theorem.

$$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n) = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

(a) $A = T\Lambda T^{-1}$; A is diagonalizable ie.

T is investible

$T = \begin{bmatrix} t^{(1)} & \cdots & t^{(n)} \end{bmatrix}$

$t^{(i)} \in \mathbb{R}^n$.

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

$A, B \subseteq R$

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

$$AT = T\Lambda$$

$$A\begin{bmatrix} t^{(1)} & \cdots & t^{(n)} \end{bmatrix} = \begin{bmatrix} t^{(1)} & \cdots & t^{(n)} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} At^{(1)} & \cdots & At^{(n)} \end{bmatrix} = \begin{bmatrix} \lambda_1 t^{(1)} & \lambda_2 t^{(2)} & \cdots & \lambda_n t^{(n)} \end{bmatrix}$$

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

(b) $U^T = U^{-1}$

$\Rightarrow U^T U = \underline{I}$

$\Rightarrow U$ is orthogonal.

(c) If A is PSD then $\lambda_i(A) \geq 0$ for each i

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

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

$$u^{(i)T} A u^{(i)} = \lambda_i u^{(i)T} u^{(i)} = \lambda_i$$

Since A is Positive semi-definite

$\Rightarrow u^{(i)T} A u^{(i)} \geq 0$

$\Rightarrow \lambda_i \geq 0$ for each i.