MATH 7243: Homework 1

Sai Nikhil

NUID: 001564864

T. Sai Nikhil

(1) $\vec{x} \in \mathbb{R}^n, A \in \mathbb{R}^{m \times n}, \vec{b} \in \mathbb{R}^m$

$$f(\vec{x}) = \vec{b}^T A \vec{x}$$

Let, $\vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$ $A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$

$$\vec{b}^T A = \begin{bmatrix} b_1 & b_2 & \cdots & b_m \end{bmatrix} \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

$$= [b_1 a_{11} + \cdots b_m a_{m1} \quad b_1 a_{12} + \cdots b_m a_{m2} \cdots \quad b_1 a_{1n} + \cdots b_m a_{mn}]$$

$$f(\vec{x}) = \vec{b}^T A \vec{x} = (b_1 a_{11} + \cdots b_m a_{m1}) x_1$$

$$+ (b_1 a_{12} + \cdots b_m a_{m2}) x_2$$

$$\vdots$$

$$+ (b_1 a_{1n} + \cdots b_m a_{mn}) x_n$$

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = \begin{bmatrix} b_1 a_{11} + \cdots b_m a_{m1} \\ b_1 a_{12} + \cdots b_m a_{m2} \\ \vdots \\ b_1 a_{1n} + \cdots b_m a_{mn} \end{bmatrix}$$

$$= (\vec{b}^T A)^T$$

$$= A^T \vec{b}$$

② Let,

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$\vec{x}^T \vec{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$= x_1^2 + x_2^2 + \cdots x_n^2$$

$$\frac{\partial (\vec{x}^T \vec{x})}{\partial \vec{x}} = \begin{bmatrix} \frac{\partial (x_1^2 + \cdots x_n^2)}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_n}(x_1^2 + \cdots x_n^2) \end{bmatrix}$$

$$= \begin{bmatrix} 2x_1 \\ \vdots \\ 2x_n \end{bmatrix} = 2\vec{x}$$

(3) Let,

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \vec{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

$$\vec{x}^T \vec{a} = x_1 a_1 + \cdots x_n a_n$$

$$(\vec{x}^T \vec{a})^2 = (x_1 a_1 + \cdots x_n a_n)^2$$

$$\frac{\partial}{\partial \vec{x}} (\vec{x}^T \vec{a})^2 = \begin{bmatrix} \frac{\partial}{\partial x_1} (\vec{x}^T \vec{a})^2 \\ \vdots \\ \frac{\partial}{\partial x_n} (\vec{x}^T \vec{a})^2 \end{bmatrix}$$

$$= \begin{bmatrix} 2(x_1 a_1 + \cdots x_n a_n)\, a_1 \\ \vdots \\ 2(x_1 a_1 + \cdots x_n a_n)\, a_n \end{bmatrix}$$

$$= (2 \vec{x}^T \vec{a}) \vec{a}$$

④

$$\vec{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}$$

$$\vec{x}(\vec{z}) = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \qquad \vec{y}(\vec{z}) = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mm} \end{bmatrix}$$

Let,

$$G = A^T \vec{y}, \quad F = \vec{x}$$

$$\therefore F : \mathbb{R}^n \to \mathbb{R}^m, \quad G : \mathbb{R}^n \to \mathbb{R}^m$$

Assume, $H = G^T F = \vec{y}^T A \vec{x}$

According to lecture notes,

$$\frac{\partial H}{\partial \vec{z}} = \frac{\partial G}{\partial \vec{z}} F + \frac{\partial F}{\partial \vec{z}} G \quad \text{— (1)}$$

$$\therefore \frac{\partial F}{\partial \vec{z}} = \frac{\partial \vec{x}}{\partial \vec{z}} \quad \text{— (2)}$$

$$\frac{\partial G}{\partial \vec{z}} = \frac{\partial (A^T \vec{y})}{\partial \vec{z}}$$

$$= \frac{\partial}{\partial \vec{z}} \left(\begin{bmatrix} a_{11}y_1 + a_{21}y_2 + \cdots a_{m1}y_m \\ \vdots \\ a_{1m}y_1 + a_{2m}y_2 + \cdots a_{mm}y_m \end{bmatrix} \right)$$

$$= \frac{\partial}{\partial \vec{z}} \left(\begin{bmatrix} \sum_{i=1}^{m} a_{i1} y_i \\ \vdots \\ \sum_{i=1}^{m} a_{im} y_i \end{bmatrix} \right)$$

$$
= \begin{bmatrix} \dfrac{\partial}{\partial z_1}\left(\sum_{i=1}^{m} a_{i1} y_i\right) & \cdots & \dfrac{\partial}{\partial z_1}\left(\sum_{i=1}^{m} a_{im} y_i\right) \\ \vdots & \ddots & \vdots \\ \dfrac{\partial}{\partial z_n}\left(\sum_{i=1}^{m} a_{i1} y_i\right) & \cdots & \dfrac{\partial}{\partial z_n}\left(\sum_{i=1}^{m} a_{im} y_i\right) \end{bmatrix}
$$

denominator
layout
notation

$$
= \begin{bmatrix} \sum_{i=1}^{m} a_{i1} \dfrac{\partial}{\partial z_1}(y_i) & \cdots & \sum_{i=1}^{m} a_{im} \dfrac{\partial}{\partial z_1}(y_i) \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{m} a_{i1} \dfrac{\partial}{\partial z_n}(y_i) & \cdots & \sum_{i=1}^{m} a_{im} \dfrac{\partial}{\partial z_n}(y_i) \end{bmatrix}
$$

$$
= \begin{bmatrix} \dfrac{\partial y_1}{\partial z_1} & \cdots & \dfrac{\partial y_m}{\partial z_1} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial y_1}{\partial z_n} & \cdots & \dfrac{\partial y_m}{\partial z_n} \end{bmatrix} \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mm} \end{bmatrix}
$$

$$= \frac{\partial \vec{y}}{\partial \vec{z}} A$$

$$\therefore \quad \frac{\partial G}{\partial \vec{z}} = \frac{\partial \vec{y}}{\partial \vec{z}} A \quad \text{--- (3)}$$

Substituting (2) & (3) in (1),

$$\frac{\partial (\vec{y}^T A \vec{x})}{\partial \vec{z}} = \frac{\partial H}{\partial \vec{z}} = \frac{\partial G}{\partial \vec{z}} F + \frac{\partial F}{\partial \vec{z}} G$$

$$= \frac{\partial \vec{y}}{\partial \vec{z}} A \vec{x} + \frac{\partial \vec{x}}{\partial \vec{z}} A^T \vec{y}$$

Hence, Proved

(5) $A(x) : \mathbb{R} \rightarrow \mathbb{R}^{n \times n}$

We know that,

$$AA^{-1} = I_n$$

Differentiating both sides w.r.t. "x",

$$\frac{d}{dx}\left(AA^{-1}\right) = \frac{d}{dx}\left(I_n\right) = 0$$

$(\because I_n \rightarrow \text{constant})$

$$\therefore \frac{dA}{dx} A^{-1} + A \frac{d\left(A^{-1}\right)}{dx} = 0$$

Chain rule

$$A \frac{d}{dx}\left(A^{-1}\right) = - \frac{dA}{dx} A^{-1}$$

$$\Rightarrow \boxed{\frac{d}{dx}\left(A^{-1}\right) = -A^{-1} \frac{dA}{dx} A^{-1}}$$

(left multiply with A^{-1})

⑥ $\vec{x}, \vec{\beta} \in \mathbb{R}^p$

$$\frac{\partial(\vec{x}^T \beta)}{\partial \vec{x}} = \frac{\partial}{\partial \vec{x}}\left(x_1\beta_1 + x_2\beta_2 + \cdots x_p\beta_p\right)$$

$$= \frac{\partial}{\partial \vec{x}}\left(\sum_{i=1}^{p} x_i\beta_i\right)$$

$$= \begin{bmatrix} \frac{\partial}{\partial x_1}\left(\sum_{i=1}^{p} x_i\beta_i\right) \\ \vdots \\ \frac{\partial}{\partial x_p}\left(\sum_{i=1}^{p} x_i\beta_i\right) \end{bmatrix}$$

$$= \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

$$= \beta$$

(7) $X: \mathbb{R}^q \to \mathbb{R}^p \quad Y: \mathbb{R}^p \to \mathbb{R}^n$

$Y(X) \in \mathbb{R}^n$

$Z \in \mathbb{R}^q$

$$\frac{\partial Y}{\partial Z} = \begin{bmatrix} \frac{\partial y_1(x_1,\dots,x_p)}{\partial z_1} & \cdots & \frac{\partial y_m(x_1,\dots,x_p)}{\partial z_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_1(x_1,\dots,x_p)}{\partial z_n} & \cdots & \frac{\partial y_m(x_1,\dots,x_p)}{\partial z_n} \end{bmatrix}$$

denominator layout notation

$$= \begin{bmatrix} \sum_{i=1}^{p} \frac{\partial x_i}{\partial z_1} \times \frac{\partial y_1}{\partial x_i} & \cdots & \sum_{i=1}^{p} \frac{\partial x_i}{\partial z_1} \times \frac{\partial y_m}{\partial x_i} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{p} \frac{\partial x_i}{\partial z_n} \times \frac{\partial y_1}{\partial x_i} & \cdots & \sum_{i=1}^{p} \frac{\partial x_i}{\partial z_n} \times \frac{\partial y_m}{\partial x_i} \end{bmatrix}$$

(1)

$$\frac{\partial X}{\partial Z} \times \frac{\partial Y}{\partial X} = \begin{bmatrix} \frac{\partial x_1}{\partial z_1} & \cdots & \frac{\partial x_p}{\partial z_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_1}{\partial z_n} & \cdots & \frac{\partial x_p}{\partial z_n} \end{bmatrix} \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial x_p} & \cdots & \frac{\partial y_m}{\partial x_p} \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} \sum_{i=1}^{p} \frac{\partial x_i}{\partial z_1} \times \frac{\partial y_1}{\partial x_i} & \cdots & \sum_{i=1}^{p} \frac{\partial x_i}{\partial z_1} \times \frac{\partial y_m}{\partial x_i} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{p} \frac{\partial x_i}{\partial z_n} \times \frac{\partial y_1}{\partial x_i} & \cdots & \sum_{i=1}^{p} \frac{\partial x_i}{\partial z_n} \times \frac{\partial y_m}{\partial x_i} \end{bmatrix}}_{(2)}$$

From (1) & (2),

$$\underbrace{\frac{\partial Y}{\partial Z}}_{n \times m} = \underbrace{\frac{\partial X}{\partial Z}}_{n \times p} \times \underbrace{\frac{\partial Y}{\partial X}}_{p \times m}$$

As you can see from dimensions above, order definitely matters. However, it matters on whether we use denominator/numerator layout notation.

(8) Given, $\vec{x} \in \mathbb{R}^p$

$z(\vec{x}) \in \mathbb{R}$

Let, $Y(\vec{x}) \in \mathbb{R}^n$

$$\therefore \frac{\partial}{\partial \vec{x}}(zY) = \frac{\partial}{\partial \vec{x}}\left(\begin{bmatrix} zy_1 \\ \vdots \\ zy_n \end{bmatrix}\right)$$

$$= \begin{bmatrix} \frac{\partial(zy_1)}{\partial x_1} & \cdots & \frac{\partial(zy_n)}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial(zy_1)}{\partial x_p} & \cdots & \frac{\partial(zy_n)}{\partial x_p} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\partial z}{\partial x_1} y_1 + z\frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial z}{\partial x_1} y_n + z\frac{\partial y_n}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial z}{\partial x_p} y_1 + z\frac{\partial y_1}{\partial x_p} & \cdots & \frac{\partial z}{\partial x_p} y_n + z\frac{\partial y_n}{\partial x_p} \end{bmatrix}$$

$$= z \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_1} \\ \vdots & & \vdots \\ \frac{\partial y_1}{\partial x_p} & \cdots & \frac{\partial y_n}{\partial x_p} \end{bmatrix} + \begin{bmatrix} \frac{\partial z}{\partial x_1} y_1 & \cdots & \frac{\partial z}{\partial x_1} y_n \\ \vdots & \ddots & \vdots \\ \frac{\partial z}{\partial x_p} y_1 & \cdots & \frac{\partial z}{\partial x_p} y_n \end{bmatrix}$$

$$= z \frac{\partial Y}{\partial \vec{x}} + \begin{bmatrix} \frac{\partial z}{\partial x_1} \\ \vdots \\ \frac{\partial z}{\partial x_p} \end{bmatrix} \begin{bmatrix} y_1 & \cdots & y_n \end{bmatrix}$$

$$= z \frac{\partial Y}{\partial \vec{x}} + \frac{\partial z}{\partial \vec{x}} Y^T$$

Hence, proved