Math 7243

Homework 1. Matrix calculus:

Using the denominator layout notation conventions.
(One point each question except 4, 7,8 with 2 points each.)
Problems 4,7,8 have longer calculations.

Correct methods are not unique. Just make sure it is correct.

Problem 1. Assume $\vec{x} \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$ and $\vec{b} \in \mathbb{R}^m$. Let $f(\vec{x}) = \vec{b}^T A\vec{x}$. Find ∇f.

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

Problem 2. Assume $\vec{x} \in \mathbb{R}^n$. Find $\dfrac{\partial \vec{x}^T \vec{x}}{\partial \vec{x}}$.

We know that $\dfrac{\partial \vec{x}^T A\vec{x}}{\partial \vec{x}} = (A^T + A)\vec{x}$ for any $n \times n$ matrix A.

In this question, $A = I$ identity matrix. So, $\dfrac{\partial \vec{x}^T \vec{x}}{\partial \vec{x}} = 2\vec{x}$.

Problem 3. Assume $\vec{x}$ and $\vec{a} \in \mathbb{R}^n$. Find $\dfrac{\partial(\vec{x}^T \vec{a})^2}{\partial \vec{x}}$

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

Remarks: You can also use product rule.
The results can be in different correct format: $2\vec{a}(\vec{x}^T \vec{a})$ or $2\vec{x}^T \vec{a}\vec{a}$, etc. Each one is a vector.

Problem 4. Suppose $\vec{x} : \mathbb{R}^n \to \mathbb{R}^m$ is a map sending $\vec{z} \in \mathbb{R}^n$ to $\vec{x}(\vec{z}) \in \mathbb{R}^m$. Similarly, suppose $\vec{y} : \mathbb{R}^n \to \mathbb{R}^m$ and A is an $m \times m$ constant matrix. Prove that $\dfrac{\partial(\vec{y}^T A\vec{x})}{\partial \vec{z}} = \dfrac{\partial \vec{y}}{\partial \vec{z}} A\vec{x} + \dfrac{\partial \vec{x}}{\partial \vec{z}} A^T \vec{y}$

Using product rule: $\dfrac{\partial(\vec{y}^T A\vec{x})}{\partial \vec{z}} = \dfrac{\partial((A^T \vec{y})^T \vec{x})}{\partial \vec{z}} = \dfrac{\partial(A^T \vec{y})}{\partial \vec{z}}\vec{x} + \dfrac{\partial \vec{x}}{\partial \vec{z}}(A^T \vec{y})$

Compare to the right side, we need to show that $\dfrac{\partial(A^T \vec{y})}{\partial \vec{z}} = \dfrac{\partial \vec{y}}{\partial \vec{z}} A$. We prove it by entry:

$\dfrac{\partial(A^T \vec{y})_i}{\partial \vec{z}_j} = \dfrac{\partial(\vec{a}_i^T \vec{y})}{\partial \vec{z}_j} = \dfrac{\partial(a_{1i}y_1 + \cdots + a_{mi}y_m)}{\partial \vec{z}_j} = \begin{bmatrix} \dfrac{\partial y_1}{\partial \vec{z}_j} \cdots \dfrac{\partial y_m}{\partial \vec{z}_j} \end{bmatrix} \begin{bmatrix} a_{1i} \\ \vdots \\ a_{mi} \end{bmatrix}$ which equal to

(the j-th row of $\dfrac{\partial \vec{y}}{\partial \vec{z}}$)(the i-th column of A). So $\dfrac{\partial(A^T \vec{y})}{\partial \vec{z}} = \dfrac{\partial \vec{y}}{\partial \vec{z}} A$.

Remark: for the product rule.

Problem 5. Suppose $A(x) : \mathbb{R} \to \mathbb{R}^{n\times n}$ is a map from $\mathbb{R}$ to $\mathbb{R}^{n\times n}$.

Show that if $A(x)$ is invertible, then $\frac{dA^{-1}}{dx} = -A^{-1}\frac{dA}{dx}A^{-1}$

Suppose $A = [a_{i,j}]$ such each a_{ij} is a function from $\mathbb{R}$ to $\mathbb{R}$.
Suppose $A^{-1} = [b_{i,j}]$ such each b_{ij} is a function from $\mathbb{R}$ to $\mathbb{R}$.
From $A^{-1}A = I_n$, we have the (i, j)−th entry

$$\sum_{k=1}^{n} b_{ik}a_{kj} = \delta_{ij}$$

Take derivatives of both sides

$$\frac{d}{dx}\left(\sum_{k=1}^{n} b_{ik}a_{kj}\right) = 0$$

That is

$$\sum_{k=1}^{n} \frac{d}{dx}\left(b_{ik}a_{kj}\right) = 0$$

Using product rule for derivative of functions:

$$\sum_{k=1}^{n} \frac{d\, b_{ik}}{dx}a_{kj} + b_{ik}\frac{d\, a_{kj}}{dx} = 0$$

That is

$$\frac{dA^{-1}}{dx}A + A^{-1}\frac{dA}{dx} = 0$$

The order of the product $A^{-1}\frac{dA}{dx}$ is important.

Problem 6. Let $\vec{x}$ and $\beta \in \mathbb{R}^p$. Prove that $\frac{\partial \vec{x}^T\beta}{\partial \vec{x}} = \beta$

Write $\vec{x}^T\beta = b_1x_1 + \cdots b_px_p$. This is a map from $\mathbb{R}^p$ to $\mathbb{R}$.

Using the denominator layout notation conventions. $\frac{\partial \vec{x}^T\beta}{\partial \vec{x}} = \begin{bmatrix} \frac{\partial \vec{x}^T\beta}{\partial \vec{x}_1} \\ \vdots \\ \frac{\partial \vec{x}^T\beta}{\partial \vec{x}_p} \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_p \end{bmatrix} = \beta$

Problem 7. Chain Rule. Assume that Y is an n vector but assume that Y depends on X and X depends on some $Z \in \mathbb{R}^q$. Show that

$$\frac{\partial Y}{\partial Z} = \frac{\partial X}{\partial Z}\frac{\partial Y}{\partial X}$$

Does the order matter?

Hint: This means that $X : \mathbb{R}^q \to \mathbb{R}^p$ and $Y : \mathbb{R}^p \to \mathbb{R}^n$.

Let $Z = \begin{bmatrix} z_1 \\ \vdots \\ z_q \end{bmatrix}$ Then $X = \begin{bmatrix} x_1(\vec{z}) \\ \vdots \\ x_q(\vec{z}) \end{bmatrix}$ and $Y = \begin{bmatrix} y_1(\vec{x}) \\ \vdots \\ y_n(\vec{x}) \end{bmatrix}$

Write done both sides of the equation explicitly and compare the $q \times n$ matrix.

$$\frac{\partial Y}{\partial Z} = \begin{bmatrix} \frac{\partial y_1}{\partial z_1} & \cdots & \frac{\partial y_n}{\partial z_1} \\ \vdots & & \vdots \\ \frac{\partial y_1}{\partial z_q} & \cdots & \frac{\partial y_n}{\partial z_q} \end{bmatrix} =$$

In particular, the (i,j)-entry

$$\frac{\partial y_j}{\partial z_i} = \frac{\partial y_j}{\partial x_1}\frac{\partial x_1}{\partial z_i} + \frac{\partial y_j}{\partial x_2}\frac{\partial x_2}{\partial z_i} + \cdots + \frac{\partial y_j}{\partial x_p}\frac{\partial x_p}{\partial z_i}$$

and compare the (i,j)-entry of $\frac{\partial X}{\partial Z}\frac{\partial Y}{\partial X}$

$$\frac{\partial X}{\partial Z}\frac{\partial Y}{\partial X} = \begin{bmatrix} \frac{\partial x_1}{\partial z_1} & \cdots & \frac{\partial x_p}{\partial z_1} \\ \vdots & & \vdots \\ \frac{\partial x_1}{\partial z_q} & \cdots & \frac{\partial x_p}{\partial z_q} \end{bmatrix} \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_1} \\ \vdots & & \vdots \\ \frac{\partial y_1}{\partial x_p} & \cdots & \frac{\partial y_n}{\partial x_p} \end{bmatrix}$$

which is $[\frac{\partial x_1}{\partial z_i} \cdots \frac{\partial x_p}{\partial z_i}] \begin{bmatrix} \frac{\partial y_j}{\partial x_1} \\ \vdots \\ \frac{\partial y_j}{\partial x_p} \end{bmatrix} = \frac{\partial y_j}{\partial x_1}\frac{\partial x_1}{\partial z_i} + \frac{\partial y_j}{\partial x_2}\frac{\partial x_2}{\partial z_i} + \cdots + \frac{\partial y_j}{\partial x_p}\frac{\partial x_p}{\partial z_i}$ So, $\frac{\partial Y}{\partial Z} = \frac{\partial X}{\partial Z}\frac{\partial Y}{\partial X}$

The order matters. $\frac{\partial X}{\partial Z}\frac{\partial Y}{\partial X} \neq \frac{\partial Y}{\partial X}\frac{\partial X}{\partial Z}$. Even the size of the right side does not compatible.

Problem 8. Let $z : \mathbb{R}^p \to \mathbb{R}$ be a function that depends on $\vec{x} \in \mathbb{R}^p$ and let Y be a n-vector that depends on $\vec{x} \in \mathbb{R}^p$. Prove that

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

Here $Y = \begin{bmatrix} y_1(\vec{x}) \\ \vdots \\ y_n(\vec{x}) \end{bmatrix}$ is a map $\mathbb{R}^p$ to $\mathbb{R}^n$.

Let $H := zY = \begin{bmatrix} z(\vec{x})y_1(\vec{x}) \\ \vdots \\ z(\vec{x})y_n(\vec{x}) \end{bmatrix}$ which is a function from $\mathbb{R}^p to \mathbb{R}^n$.

Then $\dfrac{\partial H}{\partial \vec{x}} = \begin{bmatrix} \dfrac{\partial H_1}{\partial \vec{x}_1} & \cdots & \dfrac{\partial H_n}{\partial \vec{x}_1} \\ \vdots & & \vdots \\ \dfrac{\partial H_1}{\partial \vec{x}_p} & \cdots & \dfrac{\partial H_n}{\partial \vec{x}_p} \end{bmatrix}$

The i, j entry of $\dfrac{\partial H_j}{\partial \vec{x}_i}$ is calculated by product rule for functions $\dfrac{\partial z y_j}{\partial \vec{x}_i} = z\dfrac{\partial y_j}{\partial \vec{x}_i} + y_i \dfrac{\partial z}{\partial \vec{x}_i}$.

Compare the (i,j)-entry of

$$z\frac{\partial Y}{\partial \vec{x}} + \frac{\partial z}{\partial \vec{x}} Y^T = z\begin{bmatrix} \dfrac{\partial y_1}{\partial \vec{x}_1} & \cdots & \dfrac{\partial y_n}{\partial \vec{x}_1} \\ \vdots & & \vdots \\ \dfrac{\partial y_1}{\partial \vec{x}_p} & \cdots & \dfrac{\partial y_n}{\partial \vec{x}_p} \end{bmatrix} + \begin{bmatrix} \dfrac{\partial z}{\partial \vec{x}_1} \\ \vdots \\ \dfrac{\partial z}{\partial \vec{x}_p} \end{bmatrix} \begin{bmatrix} y_1 \ldots y_n \end{bmatrix}$$

which is also $z\dfrac{\partial y_j}{\partial \vec{x}_i} + y_i \dfrac{\partial z}{\partial \vec{x}_i}$.

So, $\dfrac{\partial}{\partial \vec{x}}(zY) = z\dfrac{\partial Y}{\partial \vec{x}} + \dfrac{\partial z}{\partial \vec{x}} Y^T$

Remark: This formula is different from the product formula in class. (Find the reason.)