Homework-1

Arushi Sharma

NUID - 001306699

Email - sharma.aru@northeastern.edu

Question-1 :

Given : Square matrix $A \in R^{n \times n}$

To prove : A can be written as sum of B & C where B is symmetric & C is anti-symmetric.

Solⁿ 1 :

$$A = B + C \qquad -①$$

If, B is symmetric, then by defⁿ:

$$B = B^T \qquad -②$$

If C is anti-symmetric, then by defn:

$$C = -C^T \quad -(3)$$

By Transposition property, we can write

$$A^T = (B+C)^T \quad [\text{from } (1)]$$

$$= B^T + C^T \quad [\because (A+B)^T = A^T + B^T]$$

$$= B + (-C) \quad [\text{from } (2) \,\&\, (3)]$$

$$A^T = B - C \quad -(4)$$

We have:

$$A = B + C$$

$$A^T = B - C$$

Adding above 2 equations, we get

$$A + A^T = 2B$$

$$\Rightarrow \boxed{B = \frac{A + A^T}{2}}$$

Now, subtracting A^T from A, we get

$$A - A^T = 2C$$

$$\Rightarrow \boxed{C = \frac{A - A^T}{2}}$$

proved

Question 2:

Given: For a matrix $A \in R^{m \times n}$,

Frobenius norm $\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2}$

where a_{ij} is $(i,j)^{th}$ entry of A.

To prove: $\|A\|_F^2 = \text{trace}(A^T A)$

Solution 2 : $\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2}$

$$\Rightarrow \|A\|_F^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2 \quad \text{— (1)}$$

Now, checking RHS,

$$\text{trace}(A^T A) = \sum_{i=1}^{m} [A^T A]_{ii}$$

We can also write it as:

$$\sum_{i=1}^{m} \left(\sum_{j=1}^{n} A_{ij}^T A_{ji} \right)$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2 \quad \text{— (2)}$$

So, we see $\|A\|_F^2 = \text{trace}(A^T A)$

$$= \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2$$

(proved)

Question-3:

Given: $U \in R^{n \times n}$ is an orthonormal matrix ie $U^T U = U U^T = I_n$.

To prove: $||Ux||_2^2 = ||x||_2^2$ for $x \in R^n$

Sol^n- $||Ux||_2^2 = Ux \cdot Ux$

$= (Ux)^T \cdot Ux$ [$\because x \cdot x = x^T \cdot x$]

$= x^T \underbrace{U^T \cdot U}x$ [$\because (AB)^T = B^T \cdot A^T$]

$= x^T I x$ [$\because U^T U = U U^T = I$]

$= x^T x$

$$= x^2$$
$$= \|x\|_2^2$$

Hence, we saw that transformation of a vector by an orthonormal matrix preserves the magnitude of the vector. (proved)

Question 4:

Given: $x \in R^n$, $a \in R^n$ & matrices
$X \in R^{n \times n}$, $A \in R^{n \times n}$

To prove:

4.1) $\partial a^T x = a$

∂x

To prove this, let's take as example.

Let $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ & $a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$

$$a^T x = [a_1 \; a_2 \; --- \; a_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$= [a_1 x_1 + a_2 x_2 + \cdots + a_n x_n]$$

$$= \begin{bmatrix} \dfrac{\partial}{\partial x_1}(a_1 x_1 + a_2 x_2 + \cdots + a_n x_n) \\ \dfrac{\partial}{\partial x_2}(a_1 x_1 + a_2 x_2 + \cdots + a_n x_n) \\ \vdots \qquad\qquad \vdots \\ \dfrac{\partial}{\partial x_n}(a_1 x_1 + a_2 x_2 + \cdots + a_n x_n) \end{bmatrix}$$

$$= \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = a$$

(proved)

4.2) $$\frac{\partial x^T A x}{\partial x} = (A + A^T)x$$

Proof:

$$x^T A x = [x_1 \; x_2 \; \cdots \; x_n] \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & & a_{2n} \\ \vdots & & \\ a_{n1} & & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$= \big[(a_{11}x_1 + \cdots + a_{n1}x_n) \; \cdots \; (a_{1n}x_1 + \cdots + a_{nn}x_n)\big] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$= \left[\sum_{i=1}^{n} a_{i1} x_i + \cdots + \sum_{i=1}^{n} a_{in} x_i\right] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Multiplying both, we get,

$$= x_1 \sum_{i=1}^{n} a_{i1} x_i + \cdots + x_n \sum_{i=1}^{n} a_{in} x_i$$

$$= \sum_{j=1}^{n} x_j \cdot \sum_{i=1}^{n} a_{ij} x_i$$

$$= \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij} x_i x_j \qquad — (1)$$

Now, we will differentiate (1) w.r.t x_k

$$\frac{\partial x^T A x}{\partial x_k} = \frac{\partial}{\partial x_k} \left(\sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij} x_i x_j \right)$$

$$= \frac{\partial}{\partial x_k}\left(x_1 \sum_{i=1}^{n} a_{i1} x_i + \dots + x_k \sum_{i=1}^{n} a_{ik} x_i + \dots \right.$$

$$\left. + x_n \sum_{i=1}^{n} a_{in} x_i \right)$$

$$= x_1 a_{k1} + \dots + \sum_{i=1}^{n} a_{ik} x_i + x_k \cdot a_{kk} + \dots$$

$$\dots + x_n a_{kn}$$

$$= \sum_{j=1}^{n} a_{kj} x_j + \sum_{i=1}^{n} a_{ik} x_i$$

This can now be written as:

$$= (Ax)_k + (Ax^T)^T_k$$

$$= (Ax)_k + (xA^T)_k$$

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

proved.

4.3) $$\frac{\partial\, \mathrm{trace}(A^T X)}{\partial X} = A$$

Proof:

$$\mathrm{trace}(A^T X) = tr \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} \leftarrow \vec{x}_1 \rightarrow \\ \leftarrow \vec{x}_2 \rightarrow \\ \vdots \\ \leftarrow \vec{x}_n \rightarrow \end{bmatrix}$$

$$= tr \begin{bmatrix} \vec{a}_1 \vec{x}_1 & \vec{a}_2 \vec{x}_2 & \cdots & \vec{a}_n \vec{x}_n \\ \vdots & & & \\ \vec{a}_1 \vec{x}_1 & \vec{a}_2 \vec{x}_2 & \cdots & \vec{a}_n \vec{x}_n \end{bmatrix}$$

$$= \sum_{i=1}^{n} a_{i1} x_{i1} + \sum_{i=1}^{n} a_{i2} x_{i2} + \cdots + \sum_{i=1}^{n} a_{in} x_{in}$$

This can be generalised as $\sum_{j=1}^{n}\sum_{i=1}^{n} a_{ij} \cdot x_{ij}$

$$\frac{\partial \operatorname{trace}(A^T X)}{\partial X} = a_{ij} = A$$

(proved)

4.4) $$\frac{\partial \|y - Ax\|_2^2}{\partial x} = 2A^T(Ax - y)$$

Proof: $\|y - Ax\|_2^2 = (y - Ax)^T (y - Ax)$

$= (y^T - x^T A^T)(y - Ax)$

$= y^T y - y^T A x - x^T A^T y + x^T A^T A x$

$= y^2 - y^T A x - (Ax)^T y + x^T A^T A x$

$= y^2 - y^T A x - y^T A x + x^T A^T A x$

$[\because A^T B = B^T A]$

$$= y^2 - 2y^TAx + x^TA^TAx$$

$$\frac{\partial}{\partial x}\left(y^2 - 2y^TAx + x^TA^TAx\right)$$

$$= \frac{\partial}{\partial x}\left(-2y^TAx + (Ax)^TAx\right)$$

$$= \frac{\partial}{\partial x}\left(-2y^TAx + (Ax)^2\right) \quad [\because B^TB = B^2]$$

$$= -2y^TA + 2Ax \cdot A$$

$$= -2y^TA + 2A^2x$$

$$= -2y^TA + 2A^T \cdot Ax$$

$$= 2A^T(Ax - y) \quad \text{proved!}$$

Q-5- Determine whether each of the following functions is convex or not:

5.1) $f(x) = e^{ax}$, for a fixed $a \in \mathbb{R}$.

Sol^n:

If the second derivative of $f(x)$ is positive, then we can prove that $f(x)$ is convex.

$$\frac{\partial}{\partial x} e^{ax} = e^{ax} \cdot \frac{\partial}{\partial x}(ax) = a \cdot e^{ax}$$

$$\frac{\partial^2 (a \cdot e^{ax})}{\partial x^2} = a \cdot \frac{\partial}{\partial x}(e^{ax}) + e^{ax} \cdot \frac{\partial}{\partial x}(a)$$

$$= a \cdot (a \cdot e^{ax}) = a^2 \cdot e^{ax}$$

2nd derivative is positive.

Hence, $f(x)$ is convex. (proved)

5.2) $f(x) = -\log(x)$ with the domain $x \in (0, +\infty)$

Sol^n:

$$\frac{\partial}{\partial x}(-\log(x)) = -\frac{1}{x}$$

$$\frac{\partial^2}{\partial x^2}(-\log(x)) = +\frac{1}{x^2}$$

The 2nd derivative is positive.
Hence, $f(x)$ is a convex function.

5.3) $f(x) = e^{g(x)}$, where $g(x)$ is convex.

Sol^n - $f'(x) = e^{g(x)} \cdot g'(x)$

$$f''(x) = e^{g(x)} \cdot g''(x) + g'(x) \cdot e^{g(x)} \cdot g'(x)$$

$$= e^{g(x)} \cdot g''(x) + e^{g(x)} \cdot (g'(x))^2$$

We know $g(x)$ is convex. So, its double derivative will be positive. And single derivative is in square. Thus, the function $f(x)$ is convex.

(proved)

Q.6 - Given: Training dataset $\{(x_1, y_1), \dots (x_N, y_N)\}$ where $x_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}$. Consider the modified regression problem:

$$\min_{\theta \in \mathbb{R}^d} \sum_{i=1}^{N} s_i (y_i - \theta^T x_i)^2,$$

where each $s_i \in [0, 1]$ is given & indicates importance score of (x_i, y_i) for regression.

6.1) Write closed form solution of the above regression problem.

Sol^n - $$\frac{\partial}{\partial \theta}\left(\sum_{i=1}^{N} s_i (y_i - \theta^T x_i)^2 \right)$$

$$\Rightarrow \sum_{i=1}^{N} s_i \frac{\partial}{\partial \theta} (y_i - \theta^T x_i)^2 = 0$$

Now, $\sum_{i=1}^{n} (y_i - \theta^T x_i)^2$ can be written as:

$$\| Y - \theta^T X \|_2^2 \quad \text{where}$$

$$\| Y - \theta^T X \| = \left\| \underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_n \end{pmatrix}}_{n \times 1} - \underbrace{\begin{pmatrix} - x_1^T - \\ - x_2^T - \\ \vdots \\ \vdots \\ - x_n^T - \end{pmatrix}}_{n \times (d+1)} \theta \right\|$$

(with Y above the first vector and X above the matrix)

So, $\sum_{i=1}^{n} s_i \times \frac{\partial}{\partial \theta} \left(\| Y - \theta^T X \|_2^2 \right) = 0$

$\Rightarrow \sum_{i=1}^{n} s_i \times -2 X^T (Y - \theta^T X) = 0$

$\Rightarrow -X^T Y + X^T \theta^T X = 0$

$\Rightarrow \theta = (X^T X)^{-1} X^T Y$ Ans.

6.2) Gradient Decent Steps:

Sol^n -> $$\min_{\theta \in R^d} \sum_{i=1}^{N} s_i (y_i - \theta^T x_i)^2$$

1) Initialize $\theta^0 \in R^{d+1}$ with random vector.

where $\theta = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_d \end{bmatrix}$ & is gaussian distributed.

2) Define learning rate(ρ)

3) Loop until convergence & keep updating the cost func with each step, & update the step size.

while $\left\|\frac{\partial \text{cost}}{\partial \theta}\Big|_{\theta^t}\right\|_2 \leq \epsilon$ or $(t > T_{max})$

$$\theta^t = \theta^{t-1} - \rho \frac{\partial \text{cost}}{\partial \theta}\Big|_{\theta^{t-1}}$$

$$= \theta^{t-1} + \rho S_i \, X^T (Y - \theta^T X)$$

$$= \theta^{t-1} - \rho S_i \, X^T (X\theta^{t-1} - Y)$$

$t = t+1$

Q.7 - $f(x_1, x_2) = x_1^2 + (x_2 - 2)^2$

Expression of gradient of f:

Soln- $\frac{\partial}{\partial x_1 \partial x_2}(f(x_1, x_2))$

$$= \begin{bmatrix} \dfrac{\partial}{\partial x_1}\left(x_1^2 + (x_2-2)^2\right) \\ \dfrac{\partial}{\partial x_2}\left(x_1^2 + (x_2-2)^2\right) \end{bmatrix}$$

$$= \begin{bmatrix} 2x_1 \\ 2(x_2-2) \end{bmatrix} = 2\begin{bmatrix} x_1 \\ x_2-2 \end{bmatrix} \text{ Ans}$$