DS4400 HW1

Xin Guan

1. Let $\boldsymbol{a} \in \mathbb{R}^n$ be an n-dimensional vector and let $\boldsymbol{U} \in \mathbb{R}^{n\times n}$ be an orthonormal matrix, i.e., $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{U}\boldsymbol{U}^T = \mathbf{I}_n$. show the following:

   (a) $\text{trace}(\boldsymbol{a}\boldsymbol{a}^T) = \|\boldsymbol{a}\|_2^2$

   Solution:

   Proof. Let $\boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_n \end{bmatrix}$

   Then, $\boldsymbol{a}\boldsymbol{a}^T = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \cdot \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} = \begin{bmatrix} a_1 \cdot a_1 & a_1 \cdot a_2 & \dots & a_1 \cdot a_n \\ a_2 \cdot a_1 & a_2 \cdot a_2 & \dots & a_2 \cdot a_n \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ a_n \cdot a_1 & a_n \cdot a_2 & \dots & a_n \cdot a_n \end{bmatrix}$

   Therefore, $\text{trace}(\boldsymbol{a}\boldsymbol{a}^T) = a_1 \cdot a_1 + a_2 \cdot a_2 + \cdots + a_n \cdot a_n = \sum_{i=1}^n a_i^2 = \|\boldsymbol{a}\|_2^2$

   Thus, $\text{trace}(\boldsymbol{a}\boldsymbol{a}^T) = \|\boldsymbol{a}\|_2^2$ □

   (b) $\|\boldsymbol{U}\boldsymbol{a}\|_2^2 = \|\boldsymbol{a}\|_2^2$

   Solution:

   Proof. We can write $\boldsymbol{U}$ as follow:

   $\boldsymbol{U} = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix}$, where $v_i \in \mathbb{R}^n, i \in \mathbf{Z}, 1 \le i \le n$.

   We can write $\boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$, where $a_i \in \mathbb{R}, 1 \le i \le n$

   Since $\boldsymbol{U}$ is orthonormal, we have:

   $\forall i \ne k, v_i v_k = \overrightarrow{\mathbf{0}}$

   $\forall i \in \mathbf{Z}, 1 \le i \le n, v_i^T v_i = 1$

   Then, we can write $\boldsymbol{U}\boldsymbol{a} = \begin{bmatrix} v_1 a_1 + v_2 a_2 + \cdots + v_n a_n \end{bmatrix}$

   Therefore, $\|\boldsymbol{U}\boldsymbol{a}\|_2^2 = (v_1 a_1)^2 + (v_2 a_2)^2 + \cdots + (v_n a_n)^2$

   $= \sum_{i=1}^n (v_i^T a_i)^2 = \sum_{i=1}^n (v_i a_i)^T (v_i a_i) = \sum_{i=1}^n (a_i^T v_i)(v_i^T a_i)$

   $= \sum_{i=1}^n (a_i^T (v_i v_i^T) a_i)$

   Since $\forall i \in \mathbf{Z}, 1 \le i \le n, v_i^T v_i = 1$, we have:

   $\sum_{i=1}^n (a_i^T (v_i v_i^T) a_i) = \sum_{i=1}^n (a_i^T a_i) = \|\boldsymbol{a}\|_2^2$

   Thus, $\|\boldsymbol{U}\boldsymbol{a}\|_2^2 = \|\boldsymbol{a}\|_2^2$. □

2. Let $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{B} \in \mathbb{R}^{n\times n}$ be arbitrary but invertible matrices and let α be a scalar. Show the following:

(a) $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$

Solution:

Proof. By the definition of inverse, we have:

$$(\boldsymbol{AB})(\boldsymbol{AB})^{-1} = \boldsymbol{I}_n$$

Multiply both sides by $\boldsymbol{A}^{-1}$:

$$\boldsymbol{A}^{-1}\boldsymbol{AB}(\boldsymbol{AB})^{-1} = \boldsymbol{A}^{-1}\boldsymbol{I}_n$$

Then we have:

$$\boldsymbol{B}(\boldsymbol{AB})^{-1} = \boldsymbol{A}^{-1}$$

Multiply both sides by $\boldsymbol{B}^{-1}$:

$$\boldsymbol{B}^{-1}\boldsymbol{B}(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$$

Then we have:

$$(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$$

Therefore, $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$ □

(b) $(\boldsymbol{A}^T)^{-1} = (\boldsymbol{A}^{-1})^T$

Solution:

Proof. By the definition of inverse, we have $\boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}_n$. Then we transpose both sides:

$$(\boldsymbol{A}^{-1}\boldsymbol{A})^T = (\boldsymbol{I}_n)^T$$

Then we have:

$$\boldsymbol{A}^T(\boldsymbol{A}^{-1})^T = \boldsymbol{I}_n$$

Multiply both sides by $(\boldsymbol{A}^T)^{-1}$:

$$(\boldsymbol{A}^T)^{-1}\boldsymbol{A}^T(\boldsymbol{A}^{-1})^T = (\boldsymbol{A}^T)^{-1}\boldsymbol{I}_n$$

Then we have:

$$(\boldsymbol{A}^{-1})^T = (\boldsymbol{A}^T)^{-1}$$

□

(c) $\text{trace}(\alpha\boldsymbol{A}) = \alpha\ \text{trace}(\boldsymbol{A})$

Solution:

Proof. Let $\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \ddots & & \vdots \\ a_{11} & a_{12} & \dots & a_{1n} \end{bmatrix}$

Then $\alpha\boldsymbol{A} = \begin{bmatrix} \alpha a_{11} & \alpha a_{12} & \dots & \alpha a_{1n} \\ \alpha a_{21} & \alpha a_{22} & \dots & \alpha a_{2n} \\ \vdots & \ddots & & \vdots \\ \alpha a_{11} & \alpha a_{12} & \dots & \alpha a_{nn} \end{bmatrix}$

Then, $\text{trace}(\alpha\boldsymbol{A}) = \alpha a_{11} + \alpha a_{22} + \dots \alpha a_{nn} = \sum_{i=1}^{n} \alpha a_{ii} = \alpha \sum_{i=1}^{n} a_{ii}$

On the other hand, $\alpha\ \text{trace}(\boldsymbol{A}) = \alpha \cdot (a_{11} + a_{22} + \dots a_{nn}) = \alpha \sum_{i=1}^{n} a_{ii}$

Therefore, $\text{trace}(\alpha\boldsymbol{A}) = \alpha\ \text{trace}(\boldsymbol{A})$ □

3. For vectors $\boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{a} \in \mathbb{R}^n$ and matrices $\boldsymbol{X} \in \mathbb{R}^{n\times n}, \boldsymbol{A} \in \mathbb{R}^{n\times n}$, show the following:

   (a) $\frac{\partial \boldsymbol{a}^T \boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{A}^T\boldsymbol{a}$

   Solution:

   We write $\boldsymbol{A}$ as follow: $\boldsymbol{A} = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix}$, where $v_i \in \mathbb{R}^n, \forall i \in \mathbf{Z}, 1 \le i \le n$

   Then $\boldsymbol{a}^T\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}^T v_1 & \boldsymbol{a}^T v_2 & \dots & \boldsymbol{a}^T v_n \end{bmatrix}$

   We write $\boldsymbol{x}$ as follow: $\begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix}$, where $x_i \in \mathbb{R}, \forall i \in \mathbf{Z}, 1 \le i \le n$

   Then $\boldsymbol{a}^T\boldsymbol{A}\boldsymbol{x} = \boldsymbol{a}^T v_1 x_1 + \boldsymbol{a}^T v_2 x_2 + \dots + \boldsymbol{a}^T v_n x_n$

   Then $\frac{\partial \boldsymbol{a}^T \boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = \begin{bmatrix} \boldsymbol{a}^T v_1 \\ \boldsymbol{a}^T v_2 \\ \dots \\ \boldsymbol{a}^T v_n \end{bmatrix}$

   $\boldsymbol{A}^T\boldsymbol{a} = \begin{bmatrix} v_1^T \\ v_2^T \\ \dots \\ v_n^T \end{bmatrix} \boldsymbol{a} = \begin{bmatrix} v_1^T\boldsymbol{a} \\ v_2^T\boldsymbol{a} \\ \dots \\ v_n^T\boldsymbol{a} \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}^T v_1 \\ \boldsymbol{a}^T v_2 \\ \dots \\ \boldsymbol{a}^T v_n \end{bmatrix}$

   Therefore, $\frac{\partial \boldsymbol{a}^T \boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{A}^T\boldsymbol{a}$

   (b) $\frac{\partial trace(\boldsymbol{A}^T\boldsymbol{X})}{\partial \boldsymbol{X}} = \boldsymbol{A}$

   Solution:

   we write $\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \ddots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}, \boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \ddots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nn} \end{bmatrix}$

   Then, $\boldsymbol{A}^T\boldsymbol{X} = \begin{bmatrix} \sum_{i=1}^n a_{i1}x_{i1} & \sum_{i=1}^n a_{i1}x_{i2} & \dots & \sum_{i=1}^n a_{i1}x_{in} \\ \sum_{i=1}^n a_{i2}x_{i1} & \sum_{i=1}^n a_{i2}x_{i2} & \dots & \sum_{i=1}^n a_{i2}x_{in} \\ \vdots & \ddots & & \vdots \\ \sum_{i=1}^n a_{in}x_{i1} & \sum_{i=1}^n a_{in}x_{i2} & \dots & \sum_{i=1}^n a_{in}x_{in} \end{bmatrix}$

   $\text{trace}(\boldsymbol{A}^T\boldsymbol{X}) = \sum_{i=1}^n a_{i1}x_{i1} + \sum_{i=1}^n a_{i2}x_{i2} + \dots \sum_{i=1}^n a_{in}x_{in} = \sum_{j=1}^n \sum_{i=1}^n a_{ij}x_{ij}$

   Then, $\frac{\partial trace(\boldsymbol{A}^T\boldsymbol{X})}{\partial \boldsymbol{X}} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \ddots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} = \boldsymbol{A}$

   Therefore, $\frac{\partial trace(\boldsymbol{A}^T\boldsymbol{X})}{\partial \boldsymbol{X}} = \boldsymbol{A}$

   (c) $\frac{\partial \|\boldsymbol{A}\boldsymbol{x}\|^2}{\partial x} = 2\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x}$

4. Determine whether each of the following functions is convex or not.

   (a) $f(x) = (x-a)^2$, for any real number a.

   Solution:

Since this function is $\mathbb{R} \Rightarrow \mathbb{R}$, we can calculate its second order derivative to determine whether it is convex.

$f'(x) = 2(x-a)$
$f''(x) = 2 \geq 0$
Therefore, $f(x) = (x-a)^2$ is convex.

(b) $f(x) = -log(2x)$, with the domain $x \in (0, +\infty)$.

Solution:

Since this function is $\mathbb{R} \Rightarrow \mathbb{R}$, we can calculate its second order derivative to determine whether it is convex.

$f'(x) = -\frac{1}{x}$
$f''(x) = \frac{1}{x^2} \geq 0$
Therefore, $f(x) = -log(2x)$ is convex.

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

Solution:

Since this function is $\mathbb{R} \Rightarrow \mathbb{R}$, we can calculate its second order derivative to determine whether it is convex.

$f'(x) = e^{g(x)}g'(x)$
$f''(x) = e^{g(x)}[g'(x)]^2 + e^{g(x)}g''(x)$
Since $g(x)$ is convex, $g''(x) \geq 0$.
And $e^{g(x)}[g'(x)]^2 \geq 0$, $e^{g(x)} \geq 0$.
Then, $e^{g(x)}[g'(x)]^2 + e^{g(x)}g''(x) \geq 0$
Therefore, $f(x) = e^{g(x)}$ is convex.