

DS4400 HW1

Xin Guan

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

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

Solution:

Proof. Let $\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$

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

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

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

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

Solution:

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

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

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

Since \mathbf{U} is orthonormal, we have:

$$\forall i \neq k, v_i v_k^T = \mathbf{0}$$

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

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

$$\begin{aligned} \text{Therefore, } \|\mathbf{U}\mathbf{a}\|_2^2 &= (v_1 a_1)^2 + (v_2 a_2)^2 + \cdots + (v_n a_n)^2 \\ &= \sum_{i=1}^n (v_i^T a_i)^2 = \sum_{i=1}^n (v_i a_i)^T (v_i a_i) = \sum_{i=1}^n (a_i^T v_i)(v_i^T a_i) \\ &= \sum_{i=1}^n (a_i^T (v_i v_i^T) a_i) \end{aligned}$$

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

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

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

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

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

Solution:

Proof. By the definition of inverse, we have:

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

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

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

Then we have:

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

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

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

Then we have:

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

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

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

Solution:

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

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

Then we have:

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

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

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

Then we have:

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

□

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

Solution:

Proof. Let $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \ddots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$

Then $\alpha\mathbf{A} = \begin{bmatrix} \alpha a_{11} & \alpha a_{12} & \dots & \alpha a_{1n} \\ \alpha a_{21} & \alpha a_{22} & \dots & \alpha a_{2n} \\ \vdots & \ddots & & \vdots \\ \alpha a_{n1} & \alpha a_{n2} & \dots & \alpha a_{nn} \end{bmatrix}$

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

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

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

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

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

Solution:

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

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

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

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

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

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

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

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

Solution:

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

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

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

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

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

(c) $\frac{\partial \|\mathbf{A} \mathbf{x}\|_2^2}{\partial \mathbf{x}} = 2 \mathbf{A}^T \mathbf{A} \mathbf{x}$

Solution:

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

$$x \text{ as } \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix}$$

$$\text{Then } \|A\mathbf{x}\| = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \\ \dots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n \end{bmatrix}$$

$$\begin{aligned} \text{Then } \|A\mathbf{x}\|_2^2 &= (a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n)^2 + (a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n)^2 + \dots + (a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n)^2 \\ &= (\sum_{i=1}^n a_{1i}x_i)^2 + (\sum_{i=1}^n a_{2i}x_i)^2 \dots + (\sum_{i=1}^n a_{ni}x_i)^2 \end{aligned}$$

$$\begin{aligned} \text{Then } \frac{\partial \|A\mathbf{x}\|_2^2}{\partial x} &= \frac{\partial [(\sum_{i=1}^n a_{1i}x_i)^2 + (\sum_{i=1}^n a_{2i}x_i)^2 \dots + (\sum_{i=1}^n a_{ni}x_i)^2]}{\partial x} \\ &= \frac{\partial (\sum_{i=1}^n a_{1i}x_i)^2}{\partial x} + \frac{\partial (\sum_{i=1}^n a_{2i}x_i)^2}{\partial x} + \dots + \frac{\partial (\sum_{i=1}^n a_{ni}x_i)^2}{\partial x} \\ &= 2(\sum_{i=1}^n a_{1i}x_i) \frac{\partial (\sum_{i=1}^n a_{1i}x_i)}{\partial x} + 2(\sum_{i=1}^n a_{2i}x_i) \frac{\partial (\sum_{i=1}^n a_{2i}x_i)}{\partial x} + \dots + 2(\sum_{i=1}^n a_{ni}x_i) \frac{\partial (\sum_{i=1}^n a_{ni}x_i)}{\partial x} \\ &= 2(\sum_{i=1}^n a_{1i}x_i) \begin{bmatrix} a_{11} \\ a_{12} \\ \dots \\ a_{1n} \end{bmatrix} + 2(\sum_{i=1}^n a_{2i}x_i) \begin{bmatrix} a_{21} \\ a_{22} \\ \dots \\ a_{2n} \end{bmatrix} + \dots + 2(\sum_{i=1}^n a_{ni}x_i) \begin{bmatrix} a_{n1} \\ a_{n2} \\ \dots \\ a_{nn} \end{bmatrix} \\ &= 2 \begin{bmatrix} (\sum_{i=1}^n a_{1i}x_i)a_{11} + (\sum_{i=1}^n a_{2i}x_i)a_{21} \dots + (\sum_{i=1}^n a_{ni}x_i)a_{n1} \\ (\sum_{i=1}^n a_{1i}x_i)a_{12} + (\sum_{i=1}^n a_{2i}x_i)a_{22} \dots + (\sum_{i=1}^n a_{ni}x_i)a_{n2} \\ \dots \\ (\sum_{i=1}^n a_{1i}x_i)a_{1n} + (\sum_{i=1}^n a_{2i}x_i)a_{2n} \dots + (\sum_{i=1}^n a_{ni}x_i)a_{nn} \end{bmatrix} \\ &= 2 \begin{bmatrix} (\sum_{i=1}^n a_{i1}a_{i1})x_1 + (\sum_{i=1}^n a_{i2}a_{i1})x_2 + \dots + (\sum_{i=1}^n a_{in}a_{i1})x_n \\ (\sum_{i=1}^n a_{i1}a_{i2})x_1 + (\sum_{i=1}^n a_{i2}a_{i2})x_2 + \dots + (\sum_{i=1}^n a_{in}a_{i2})x_n \\ \dots \\ (\sum_{i=1}^n a_{i1}a_{in})x_1 + (\sum_{i=1}^n a_{i2}a_{in})x_2 + \dots + (\sum_{i=1}^n a_{in}a_{in})x_n \end{bmatrix} \end{aligned}$$

If we write A as $\begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix}$

$$\text{where } v_i = \begin{bmatrix} a_{1i} \\ a_{2i} \\ a_{3i} \\ \dots \\ a_{ni} \end{bmatrix}$$

$$\text{Then we can write the expression above as } 2 \begin{bmatrix} v_1^T v_1 x_1 + v_2^T v_1 x_2 + \dots + v_n^T v_1 x_n \\ v_1^T v_2 x_1 + v_2^T v_2 x_2 + \dots + v_n^T v_2 x_n \\ \dots \\ v_1^T v_n x_1 + v_2^T v_n x_2 + \dots + v_n^T v_n x_n \end{bmatrix}$$

$$\text{Therefore, } \frac{\partial \|A\mathbf{x}\|_2^2}{\partial x} = 2 \begin{bmatrix} v_1^T v_1 x_1 + v_2^T v_1 x_2 + \dots + v_n^T v_1 x_n \\ v_1^T v_2 x_1 + v_2^T v_2 x_2 + \dots + v_n^T v_2 x_n \\ \dots \\ v_1^T v_n x_1 + v_2^T v_n x_2 + \dots + v_n^T v_n x_n \end{bmatrix}$$

$$\text{on the right hand side: } 2A^T A\mathbf{x} = 2 \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix}^T \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix}$$

$$\begin{aligned}
&= 2 \begin{bmatrix} v_1^T \\ v_2^T \\ \vdots \\ v_n^T \end{bmatrix} \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\
&= 2 \begin{bmatrix} v_1^T v_1 & v_1^T v_2 & \dots & v_1^T v_n \\ v_2^T v_1 & v_2^T v_2 & \dots & v_2^T v_n \\ \vdots & \dots & \dots & \vdots \\ v_n^T v_1 & v_n^T v_2 & \dots & v_n^T v_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\
&= 2 \begin{bmatrix} v_1^T v_1 x_1 & v_1^T v_2 x_2 & \dots & v_1^T v_n x_n \\ v_2^T v_1 x_1 & v_2^T v_2 x_2 & \dots & v_2^T v_n x_n \\ \vdots & \dots & \dots & \vdots \\ v_n^T v_1 x_1 & v_n^T v_2 x_2 & \dots & v_n^T v_n x_n \end{bmatrix} \\
&= 2 \begin{bmatrix} v_1^T v_1 x_1 & v_2^T v_1 x_2 & \dots & v_n^T v_1 x_n \\ v_1^T v_2 x_1 & v_2^T v_2 x_2 & \dots & v_n^T v_2 x_n \\ \vdots & \dots & \dots & \vdots \\ v_1^T v_n x_1 & v_2^T v_n x_2 & \dots & v_n^T v_n x_n \end{bmatrix}
\end{aligned}$$

Thus, left part equals right part.

Therefore, $\frac{\partial \|\mathbf{Ax}\|_2^2}{\partial \mathbf{x}} = 2\mathbf{A}^T \mathbf{Ax}$

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, $g''(x) \geq 0$.

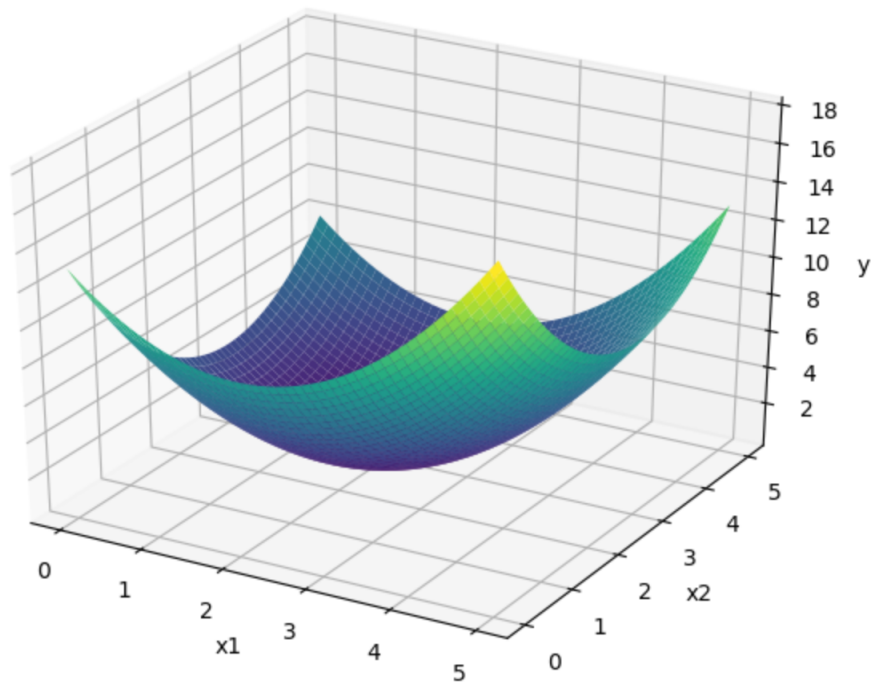
And $e^{g(x)} [g'(x)]^2 \geq 0$, $e^{g(x)} \geq 0$.

Then, $e^{g(x)} [g'(x)]^2 + e^{g(x)} g''(x) \geq 0$

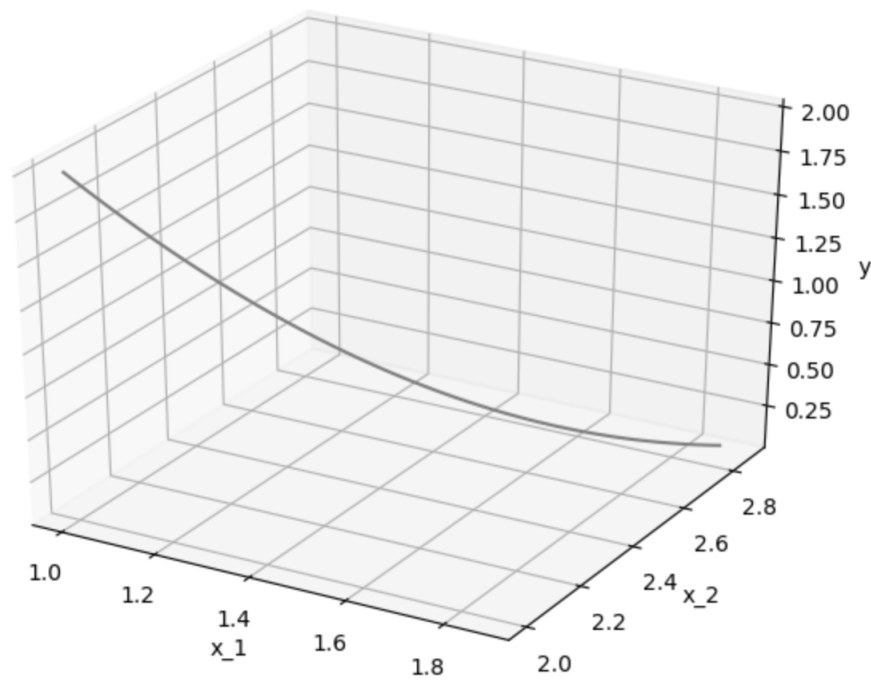
Therefore, $f(x) = e^{g(x)}$ is convex.

5. question 5

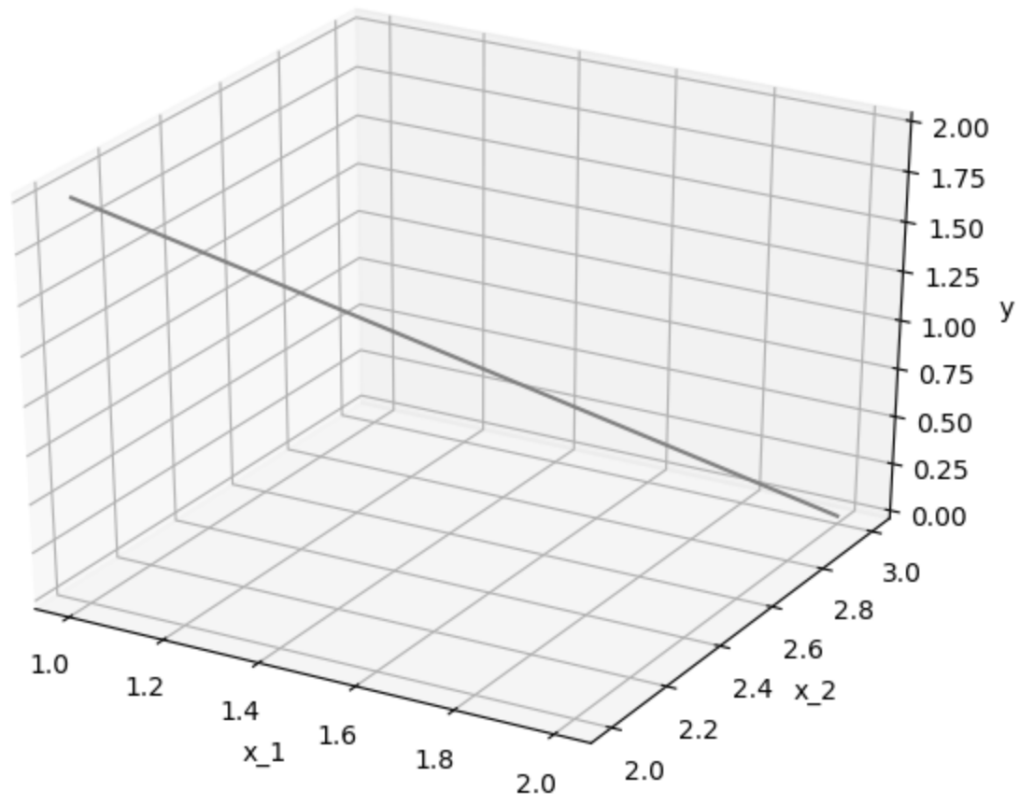
(a) Solution:



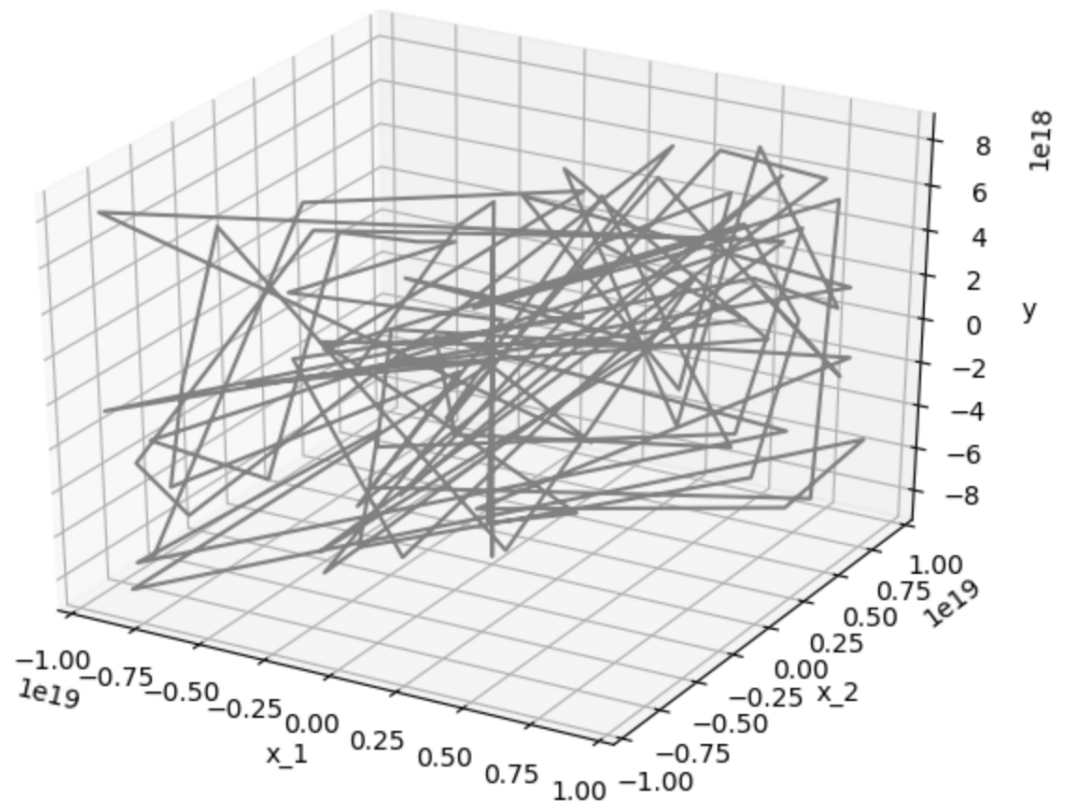
(b) Solution:



(c) Solution:



(d) [Solution:](#)



6. question 6

Solution:

