DS4400 HW1

Xin Guan

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

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

   Solution:

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

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

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

   Thus, $\text{trace}(aa^T) = \|a\|_2^2$ $\qquad\square$

   (b) $\|Ua\|_2^2 = \|a\|_2^2$

   Solution:

   Proof. We can write U as follow:

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

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

   Since U is orthonormal, we have:

   $\forall i \neq k, v_i v_k = \vec{0}$

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

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

   Therefore, $\|Ua\|_2^2 = (v_1 a_1)^2 + (v_2 a_2)^2 + \cdots + (v_n a_n)^2$

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

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

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

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

1

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

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

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

$= \|a\|_2^2$

Thus, $\|Ua\|_2^2 = \|a\|_2^2$. $\qquad\square$

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

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

Solution:

Proof. By the definition of inverse, we have:

$$(AB)(AB)^{-1} = I_n$$

Multiply both sides by A^{-1}:
$$A^{-1}AB(AB)^{-1} = A^{-1}I_n$$

Then we have:
$$B(AB)^{-1} = A^{-1}$$

Multiply both sides by B^{-1}:
$$B^{-1}B(AB)^{-1} = B^{-1}A^{-1}$$

Then we have:
$$(AB)^{-1} = B^{-1}A^{-1}$$

Therefore, $(AB)^{-1} = B^{-1}A^{-1}$ $\qquad\square$

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

Solution:

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

$$(A^{-1}A)^T = (I_n)^T$$

Then we have:
$$A^T(A^{-1})^T = I_n$$

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

$$(A^T)^{-1}A^T(A^{-1})^T = (A^T)^{-1}I_n$$

Then we have:

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

$\qquad\square$

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

Solution:

2

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

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

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

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

Therefore, $\text{trace}(\alpha A) = \alpha \, \text{trace}(A)$ $\qquad\qquad\qquad \square$

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

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

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

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

Then, on the left hand side: $\frac{\partial \boldsymbol{a}^T \boldsymbol{A} \boldsymbol{x}}{\partial \boldsymbol{x}} = \begin{bmatrix} \boldsymbol{a}^T v_1 \\ \boldsymbol{a}^T v_2 \\ \cdots \\ \boldsymbol{a}^T v_n \end{bmatrix}$

On the right hand side: $\boldsymbol{A}^T \boldsymbol{a} = \begin{bmatrix} v_1^T \\ v_2^T \\ \cdots \\ v_n^T \end{bmatrix} \boldsymbol{a} = \begin{bmatrix} v_1^T \boldsymbol{a} \\ v_2^T \boldsymbol{a} \\ \cdots \\ v_n^T \boldsymbol{a} \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}^T v_1 \\ \boldsymbol{a}^T v_2 \\ \cdots \\ \boldsymbol{a}^T v_n \end{bmatrix}$

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

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

Solution:

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

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

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

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

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) $\frac{\partial \|\boldsymbol{Ax}\|_2^2}{\partial x} = 2\boldsymbol{A}^T\boldsymbol{Ax}$

Solution:

we write $\boldsymbol{A}$ 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 as $\begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix}$

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

Then $\|Ax\|_2^2$
$= (a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n)^2 + (a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n)^2 + \dots + (a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n)^2$
$= (\sum_{i=1}^{n} a_{1i}x_i)^2 + (\sum_{i=1}^{n} a_{2i}x_i)^2 \dots + (\sum_{i=1}^{n} a_{ni}x_i)^2$

Then on the left hand side:

$$\frac{\partial \|\boldsymbol{Ax}\|^2}{\partial x}$$

$$= \frac{\partial[(\sum_{i=1}^{n} a_{1i}x_i)^2 + (\sum_{i=1}^{n} a_{2i}x_i)^2 \dots + (\sum_{i=1}^{n} a_{ni}x_i)^2]}{\partial x}$$

$$= \frac{\partial(\sum_{i=1}^{n} a_{1i}x_i)^2}{\partial x} + \frac{\partial(\sum_{i=1}^{n} a_{2i}x_i)^2}{\partial x} + \dots + \frac{\partial(\sum_{i=1}^{n} a_{ni}x_i)^2}{\partial x}$$

$$= 2(\sum_{i=1}^{n} a_{1i}x_i)\frac{\partial(\sum_{i=1}^{n} a_{1i}x_i)}{\partial x} + 2(\sum_{i=1}^{n} a_{2i}x_i)\frac{\partial(\sum_{i=1}^{n} a_{2i}x_i)}{\partial x} + \dots + 2(\sum_{i=1}^{n} a_{ni}x_i)\frac{\partial(\sum_{i=1}^{n} a_{ni}x_i)}{\partial x}$$

$$= 2(\sum_{i=1}^{n} a_{1i}x_i)\begin{bmatrix} a_{11} \\ a_{12} \\ \dots \\ a_{1n} \end{bmatrix} + 2(\sum_{i=1}^{n} a_{2i}x_i)\begin{bmatrix} a_{21} \\ a_{22} \\ \dots \\ a_{2n} \end{bmatrix} + \dots + 2(\sum_{i=1}^{n} a_{ni}x_i)\begin{bmatrix} a_{n1} \\ a_{n2} \\ \dots \\ a_{nn} \end{bmatrix}$$

$$= 2\begin{bmatrix} (\sum_{i=1}^{n} a_{1i}x_i)a_{11} + (\sum_{i=1}^{n} a_{2i}x_i)a_{21} + \dots + (\sum_{i=1}^{n} a_{ni}x_i)a_{n1} \\ (\sum_{i=1}^{n} a_{1i}x_i)a_{12} + (\sum_{i=1}^{n} a_{2i}x_i)a_{22} + \dots + (\sum_{i=1}^{n} a_{ni}x_i)a_{n2} \\ \vdots \\ (\sum_{i=1}^{n} a_{1i}x_i)a_{1n} + (\sum_{i=1}^{n} a_{2i}x_i)a_{2n} + \dots + (\sum_{i=1}^{n} a_{ni}x_i)a_{nn} \end{bmatrix}$$

On right hand side:

$$2A^T Ax = 2A^T(Ax)$$

$$= 2 \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \ddots & & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \cdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n \end{bmatrix}$$

$$= 2 \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \ddots & & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{n} a_{1i}x_i \\ \sum_{i=1}^{n} a_{2i}x_i \\ \cdots \\ \sum_{i=1}^{n} a_{ni}x_i \end{bmatrix}$$

$$= 2 \begin{bmatrix} (\sum_{i=1}^{n} a_{1i}x_i) \cdot a_{11} + (\sum_{i=1}^{n} a_{2i}x_i) \cdot a_{21} + \cdots + (\sum_{i=1}^{n} a_{ni}x_i) \cdot a_{n1} \\ (\sum_{i=1}^{n} a_{1i}x_i) \cdot a_{12} + (\sum_{i=1}^{n} a_{2i}x_i) \cdot a_{22} + \cdots + (\sum_{i=1}^{n} a_{ni}x_i) \cdot a_{n2} \\ \vdots \\ (\sum_{i=1}^{n} a_{1i}x_i) \cdot a_{1n} + (\sum_{i=1}^{n} a_{2i}x_i) \cdot a_{2n} + \cdots + (\sum_{i=1}^{n} a_{ni}x_i) \cdot a_{nn} \end{bmatrix}$$

Thus, left part equals right part.

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

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.

$f'(x) = -\frac{1}{x}$
$f''(x) = \frac{1}{x^2}$
since $x \in (0, +\infty), f''(x) \geq 0$

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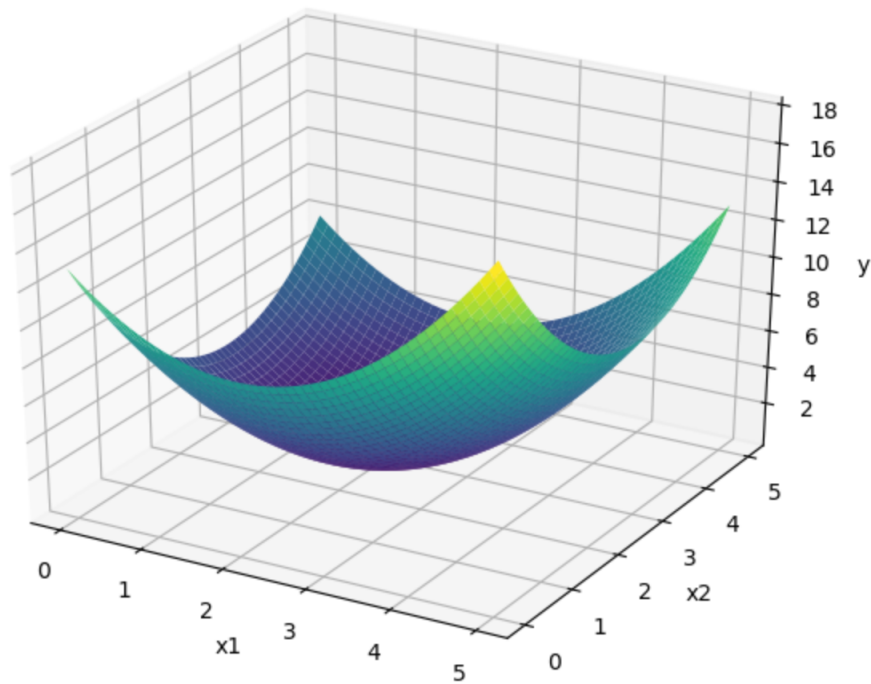
$f'(x) = e^{g(x)}g'(x)$
$f''(x) = e^{g(x)}[g'(x)]^2 + e^{g(x)}g''(x)$
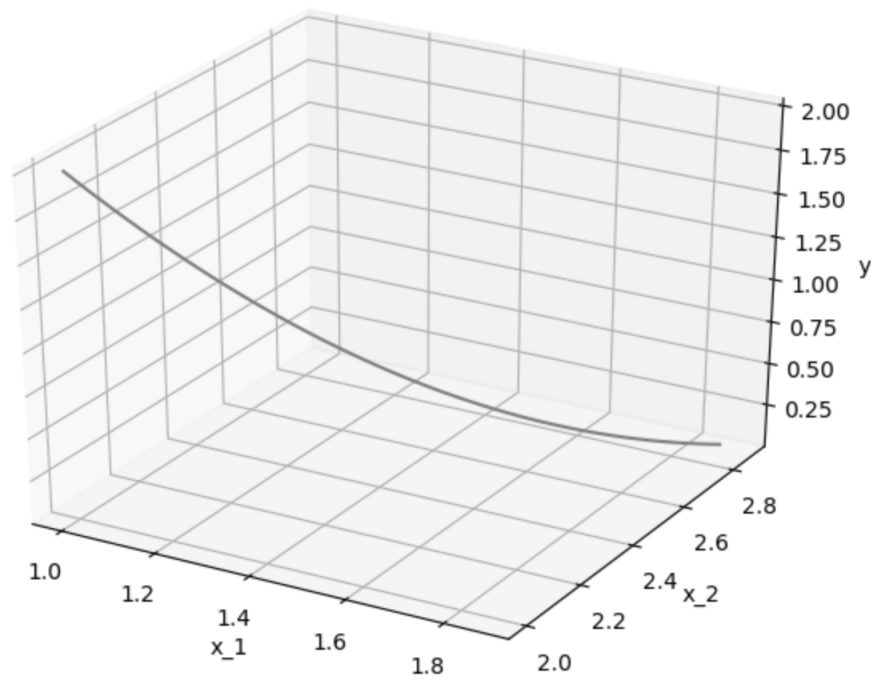Since $g(x)$ is convex, $g''(x) \geq 0$.

5

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:
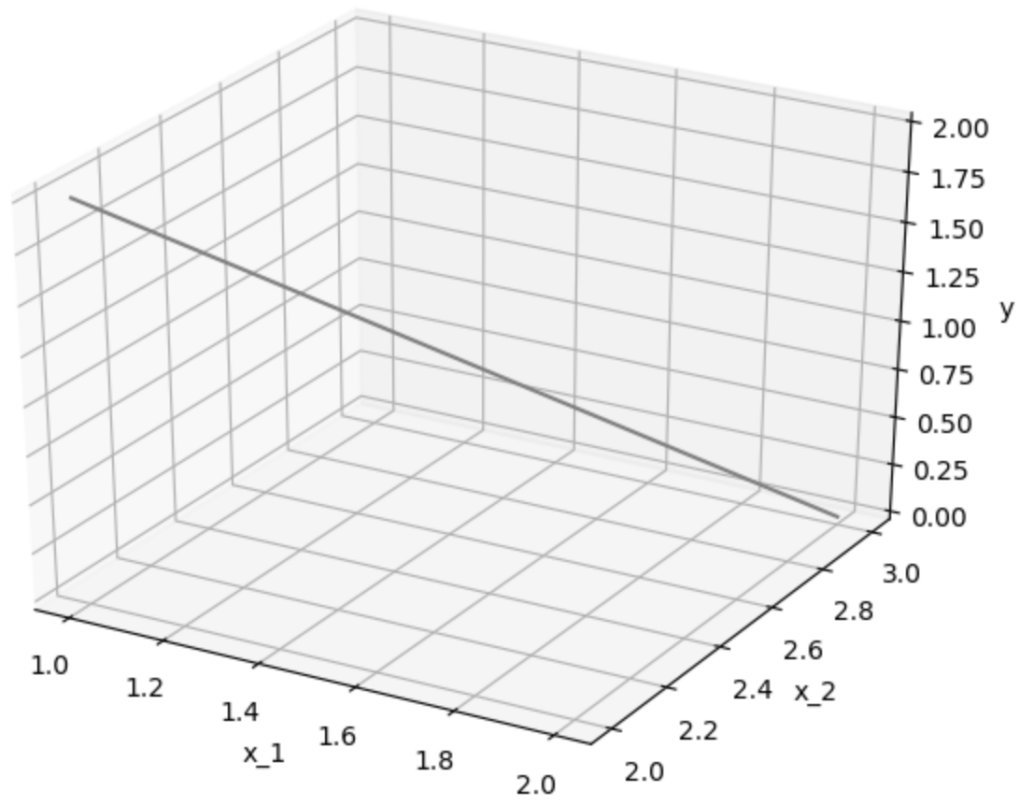
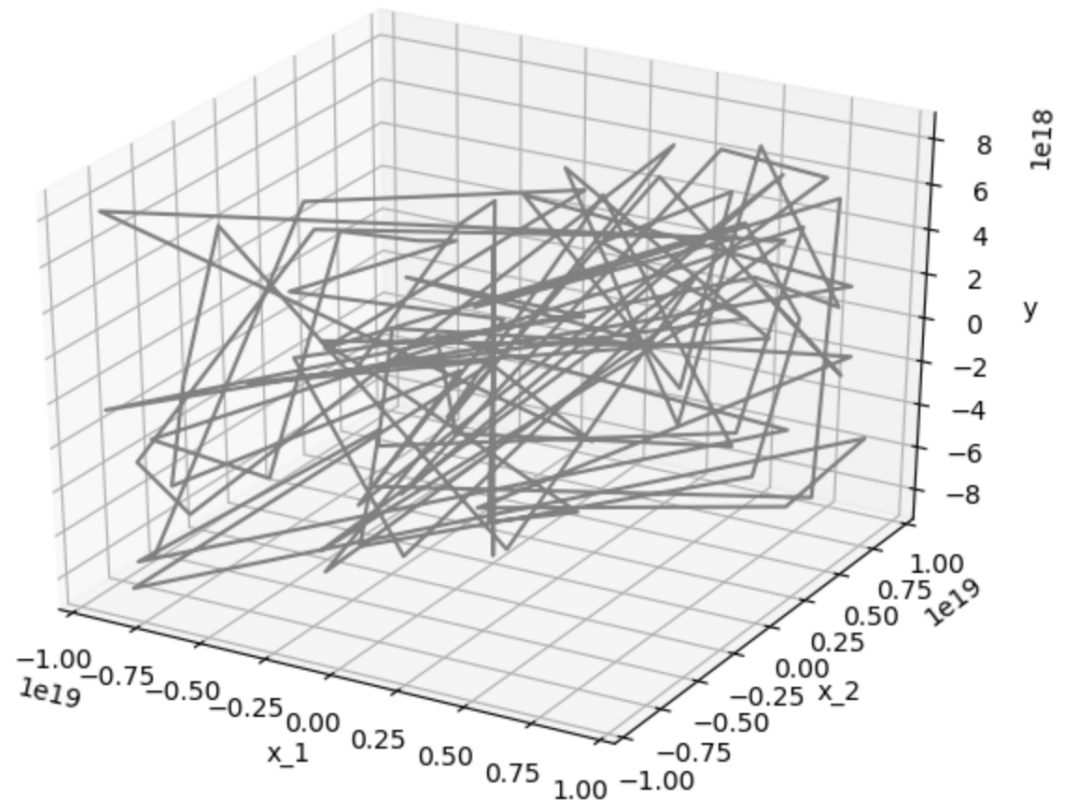(c) Solution:

(d) Solution:

6. question 6

Solution: