# 24300 HW2

Bin Yu

October 18, 2024

## Question 1: Linear Dependence and Independence

Given matrices:

$$A = \begin{bmatrix} -2 & 3 & -3 \\ -6 & 9 & -11 \\ -4 & 6 & -8 \end{bmatrix}$$

and

$$B = \begin{bmatrix} 0 & 4 & 2 \\ 1 & 0 & 3 \\ -1 & 0 & 0 \end{bmatrix}$$

To determine the linear dependence or independence of the columns of matrices $A$ and $B$, we can perform row reduction and analyze the rank of each matrix.

The rank of a matrix is the maximum number of linearly independent rows or columns it contains.

For an $n \times n$ matrix:

- If the rank equals $n$, the columns are **linearly independent**.
- If the rank is less than $n$, the columns are **linearly dependent**.

### Matrix A

$$A = \begin{bmatrix} -2 & 3 & -3 \\ -6 & 9 & -11 \\ -4 & 6 & -8 \end{bmatrix}$$

Perform row reduction:

$$\text{Row 2} = \text{Row 2} - 3 \cdot \text{Row 1}$$
$$\text{Row 3} = \text{Row 3} - 2 \cdot \text{Row 1}$$

$$A = \begin{bmatrix} -2 & 3 & -3 \\ 0 & 0 & -2 \\ 0 & 0 & -2 \end{bmatrix}$$

$$\text{Row 3} = \text{Row 3} - \text{Row 2}$$

$$A = \begin{bmatrix} -2 & 3 & -3 \\ 0 & 0 & -2 \\ 0 & 0 & 0 \end{bmatrix}$$

In the row-echelon form of matrix $A$, there are two non-zero rows. Therefore, the rank of $A$ is:

$$\text{rank}(A) = 2$$

Since the rank of matrix $A$ is 2, which is less than the number of columns (3), the columns of matrix $A$ are **linearly dependent**.

## Matrix B

$$B = \begin{bmatrix} 0 & 4 & 2 \\ 1 & 0 & 3 \\ -1 & 0 & 0 \end{bmatrix}$$

Perform row reduction:

$$\text{Row 3} = \text{Row 3} + \text{Row 2} \quad \text{and} \quad \text{Swap Row 1 and Row 2}$$

$$B = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 4 & 2 \\ 0 & 0 & 3 \end{bmatrix}$$

In the row-echelon form of matrix $B$, all three rows are non-zero. Therefore, the rank of $B$ is:

$$\text{rank}(B) = 3$$

Since the rank of matrix $B$ is 3, which equals the number of columns (3), the columns of matrix $B$ are **linearly independent**.

Therefore,

1. **Matrix $A$:**
   - $\text{rank}(A) = 2 < 3$ (number of columns)
   - $\Rightarrow$ Columns of $A$ are **linearly dependent**.
2. **Matrix $B$:**
   - $\text{rank}(B) = 3 = 3$ (number of columns)
   - $\Rightarrow$ Columns of $B$ are **linearly independent**.

## Question 2: Row Rank and Column Rank

### Matrix A

Given the matrix:

$$A = \begin{bmatrix} 6 & 4 & 2 \\ -3 & -2 & -1 \\ 9 & 6 & q \end{bmatrix}$$

To find the rank for different values of $q$, perform row reduction.

Row 2 = 2 Row 2 + Row 1
Row 3 = Row 3 - $\frac{3}{2}$ Row 1

We have:

$$A = \begin{bmatrix} 6 & 4 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & q-3 \end{bmatrix}$$

**Case 1:** $q = 3$

When $q = 3$, the matrix becomes:

$$A = \begin{bmatrix} 6 & 4 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Since there are two rows of all zeros, the rank of $A$ is 1.

**Case 2:** $q \neq 3$

When $q \neq 3$, we have:

$$A = \begin{bmatrix} 6 & 4 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & q-3 \end{bmatrix}$$

In this case, there is one row of all zeros, so the rank of $A$ is 2.

$$\text{Therefore, } \begin{cases} q = 3 \implies \text{rank}(A) = 1, \\ q \neq 3 \implies \text{rank}(A) = 2, \\ \text{rank}(A) = 3 \text{ is not achievable for any value of } q, \text{since A at least has one row of all zeros} \end{cases}$$

### Matrix B

Given the matrix:

$$B = \begin{bmatrix} 3 & 1 & 3 \\ q & 2 & q \end{bmatrix}$$

Perform row reduction to find the rank.

Row 2 = Row 2 - 2 Row 1

$$B = \begin{bmatrix} 3 & 1 & 3 \\ q-6 & 0 & q-6 \end{bmatrix}$$

**Case 1:** $q = 6$

When $q = 6$, the matrix becomes:

$$B = \begin{bmatrix} 3 & 1 & 3 \\ 0 & 0 & 0 \end{bmatrix}$$

Thus, the rank of $B$ is 1.

**Case 2:** $q \neq 6$

When $q \neq 6$, after performing row operations, we get:

$$B = \begin{bmatrix} 3 & 1 & 3 \\ q-6 & 0 & q-6 \end{bmatrix}$$

Therefore, the rank of $B$ is 2.

$$\text{Therefore, } \begin{cases} q = 6 \implies \text{rank}(B) = 1, \\ q \neq 6 \implies \text{rank}(B) = 2, \\ \text{rank}(B) = 3 \text{ is not achievable for any value of } q, \text{ since } B \text{ only has 2 rows, so } \text{rank}(B) \leq \min(m, n). \end{cases}$$

## Question 3: Fundamental Subspaces

### (1)

Given the matrix:

$$A = \begin{bmatrix} 0 & 3 & -6 & 6 \\ 3 & -9 & 12 & -9 \\ 3 & -7 & 8 & -5 \\ -1 & 3 & -4 & 3 \end{bmatrix}$$

we need to find the span of the null space of $A$ and use this to determine the rank of $A$.

Let:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

solving:

$$A\mathbf{x} = \begin{bmatrix} 0 & 3 & -6 & 6 \\ 3 & -9 & 12 & -9 \\ 3 & -7 & 8 & -5 \\ -1 & 3 & -4 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Augmented matrix:

$$\left[\begin{array}{cccc|c} 0 & 3 & -6 & 6 & 0 \\ 3 & -9 & 12 & -9 & 0 \\ 3 & -7 & 8 & -5 & 0 \\ -1 & 3 & -4 & 3 & 0 \end{array}\right]$$

Row Operations:
Row 2 = Row 2 + 3 Row 1

$$\left[\begin{array}{cccc|c} 0 & 3 & -6 & 6 & 0 \\ 3 & 0 & -6 & 9 & 0 \\ 3 & -7 & 8 & -5 & 0 \\ -1 & 3 & -4 & 3 & 0 \end{array}\right]$$

Row 3 = Row 3 - Row 2
Row 4 = Row 4 - Row 1

$$\left[\begin{array}{cccc|c} 0 & 3 & -6 & 6 & 0 \\ 3 & 0 & -6 & 9 & 0 \\ 0 & -7 & 14 & -14 & 0 \\ -1 & 0 & 2 & -3 & 0 \end{array}\right]$$

Row 4 = Row 2 + 3 Row 4
Row 3 = $\frac{7}{3}$ Row 1 + Row 3

$$\left[\begin{array}{cccc|c} 0 & 3 & -6 & 6 & 0 \\ 3 & 0 & -6 & 9 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Swapping Row 1 and Row 2 Normalize Row 1 and Row 2 by dividing by 3.

Final row-echelon form:

$$\left[\begin{array}{cccc|c} 1 & 0 & -2 & 3 & 0 \\ 0 & 1 & -2 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

the system of equations is:

$$\begin{aligned} x_1 - 2x_3 + 3x_4 &= 0 \quad \text{(Equation 1)} \\ x_2 - 2x_3 + 2x_4 &= 0 \quad \text{(Equation 2)} \end{aligned}$$

Let $x_3 = s$ and $x_4 = t$, where $s, t \in R$.

$$\text{From Equation 2:} \quad x_2 = 2x_3 - 2x_4 = 2s - 2t$$

$$\text{From Equation 1:} \quad x_1 = 2x_3 - 3x_4 = 2s - 3t$$

Therefore,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 2s - 3t \\ 2s - 2t \\ s \\ t \end{bmatrix} = s \begin{bmatrix} 2 \\ 2 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -3 \\ -2 \\ 0 \\ 1 \end{bmatrix}, \quad s, t \in R$$

Thus:

$$\text{Null}(A) = \text{Span} \left\{ \begin{bmatrix} 2 \\ 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ -2 \\ 0 \\ 1 \end{bmatrix} \right\}$$

Thus,

$$\text{Nullity}(A) = 2$$

From Rank-Nullity Theorem:

$$\text{Rank}(A) + \text{Nullity}(A) = \text{Number of Columns of } A$$

$$\text{Rank}(A) + 2 = 4 \quad \Rightarrow \quad \text{Rank}(A) = 2$$

## (2)

Given:

$$\mathbf{b} = \begin{bmatrix} 3 \\ 7 \\ 9 \\ -\frac{7}{3} \end{bmatrix}$$

To find a solution to the system:

$$A\mathbf{x} = \mathbf{b}$$

Augmented matrix:

$$\left[\begin{array}{cccc|c} 0 & 3 & -6 & 6 & 3 \\ 3 & -9 & 12 & -9 & 7 \\ 3 & -7 & 8 & -5 & 9 \\ -1 & 3 & -4 & 3 & -\frac{7}{3} \end{array}\right]$$

Row Operations: Row 2 = Row 2 + 3 Row 1

$$\left[\begin{array}{cccc|c} 0 & 3 & -6 & 6 & 3 \\ 3 & 0 & -6 & 9 & 16 \\ 3 & -7 & 8 & -5 & 9 \\ -1 & 3 & -4 & 3 & -\frac{7}{3} \end{array}\right]$$

Row 3 = Row 3 - Row 2
Row 4 = Row 4 - Row 1

$$\left[\begin{array}{cccc|c} 0 & 3 & -6 & 6 & 3 \\ 3 & 0 & -6 & 9 & 16 \\ 0 & -7 & 14 & -14 & -7 \\ -1 & 0 & 2 & -3 & -\frac{16}{3} \end{array}\right]$$

**Row 4** = Row 2 + 3 Row 4
**Row 3** = $\frac{7}{3}$ Row 1 + Row 3

$$\left[\begin{array}{cccc|c} 0 & 3 & -6 & 6 & 3 \\ 3 & 0 & -6 & 9 & 16 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Swap Row 1 and Row 2
Normalize Row 1 and Row 2 by dividing by 3.

Final row-echelon form:

$$\left[\begin{array}{cccc|c} 1 & 0 & -2 & 3 & \frac{16}{3} \\ 0 & 1 & -2 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

From the reduced row-echelon form, the system of equations is:

$$x_1 - 2x_3 + 3x_4 = \frac{16}{3} \quad \text{(Equation 1)}$$
$$x_2 - 2x_3 + 2x_4 = 1 \quad \text{(Equation 2)}$$

Let $x_3 = s$ and $x_4 = t$, where $s, t \in R$.

$$\text{From Equation 2:} \quad x_2 = 2x_3 - 2x_4 + 1 = 2s - 2t + 1$$

$$\text{From Equation 1:} \quad x_1 = 2x_3 - 3x_4 + \frac{16}{3} = 2s - 3t + \frac{16}{3}$$

Therefore,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 2s - 3t + \frac{16}{3} \\ 2s - 2t + 1 \\ s \\ t \end{bmatrix} = s \begin{bmatrix} 2 \\ 2 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -3 \\ -2 \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} \frac{16}{3} \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad s, t \in R$$

Solution to $A\mathbf{x} = \mathbf{b}$:

$$\mathbf{x} = \begin{bmatrix} \frac{16}{3} \\ 1 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} 2 \\ 2 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -3 \\ -2 \\ 0 \\ 1 \end{bmatrix}, \quad s, t \in R$$

Span of the Solution Set:

$$\mathbf{x} = \begin{bmatrix} \frac{16}{3} \\ 1 \\ 0 \\ 0 \end{bmatrix} + \text{Span} \left\{ \begin{bmatrix} 2 \\ 2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ -2 \\ 0 \\ 1 \end{bmatrix} \right\}$$

**Therefore, dimension of the Solution Space:** 2

## (3)

Suppose:

$$\mathbf{b} = \begin{bmatrix} 3 \\ 7 \\ 0 \\ 0 \end{bmatrix}$$

To find a solution to the linear system $A\mathbf{x} = \mathbf{b}$:

Augmented matrix:

$$\left[\begin{array}{cccc|c} 0 & 3 & -6 & 6 & 3 \\ 3 & -9 & 12 & -9 & 7 \\ 3 & -7 & 8 & -5 & 0 \\ -1 & 3 & -4 & 3 & 0 \end{array}\right]$$

Row Operations:
Row 2 = Row2 + 3 Row1

$$\left[\begin{array}{cccc|c} 0 & 3 & -6 & 6 & 3 \\ 3 & 0 & -6 & 9 & 16 \\ 3 & -7 & 8 & -5 & 0 \\ -1 & 3 & -4 & 3 & 0 \end{array}\right]$$

Row 3 = Row3 - Row2
Row 4 = Row4 - Row1

$$\left[\begin{array}{cccc|c} 0 & 3 & -6 & 6 & 3 \\ 3 & 0 & -6 & 9 & 16 \\ 0 & -7 & 14 & -14 & -16 \\ -1 & 0 & 2 & -3 & -3 \end{array}\right]$$

Row 4 = Row2 + 3 Row4
Row 3 = $\frac{7}{3}$ Row1 + Row3

$$\left[\begin{array}{cccc|c} 0 & 3 & -6 & 6 & 3 \\ 3 & 0 & -6 & 9 & 16 \\ 0 & 0 & 0 & 0 & -9 \\ 0 & 0 & 0 & 0 & 7 \end{array}\right]$$

Swap Row 1 and Row 2
Normalize Row 1 and Row 2 by dividing by 3.

Final row-echelon form:

$$\left[\begin{array}{cccc|c} 1 & 0 & -2 & 3 & \frac{16}{3} \\ 0 & 1 & -2 & 2 & 1 \\ 0 & 0 & 0 & 0 & -9 \\ 0 & 0 & 0 & 0 & 7 \end{array}\right]$$

From the reduced augmented matrix, observe the following rows:

$$0 = -9 \quad \text{(Row 3)}$$
$$0 = 7 \quad \text{(Row 4)}$$

These are contradictory equations, which indicate that the system is inconsistent. Therefore, there is no solution to the linear system $A\mathbf{x} = \mathbf{b}$

**Explanation in Terms of Fundamental Subspaces:**

The system $A\mathbf{x} = \mathbf{b}$ has a solution if and only if $\mathbf{b}$ is in the column space of $A$, denoted Col($A$). From Part 1, we have that Rank($A$) = 2, meaning the column space of $A$ is a 2-dimensional subspace of $R^4$, which implies that only vectors $\mathbf{b} \in R^4$ that lie within this 2-dimensional plane can make the system $A\mathbf{x} = \mathbf{b}$solvable.

However, the augmented matrix's inconsistency (i.e., contradictory equations) shows that $\mathbf{b}$ is not in Col($A$), that is, $\mathbf{b}$ can not be the linear combination of Col($A$). Therefore, the linear system $A\mathbf{x} = \mathbf{b}$ does not have a solution.

## Question 4: Least Squares with Calculus

### (1)

Let $f(\mathbf{x}) = \|A\mathbf{x} - \mathbf{b}\|^2$. Show that:

$$f(\mathbf{x}) = \mathbf{x}^\top A^\top A\mathbf{x} - 2\mathbf{x}^\top A^\top \mathbf{b} + \mathbf{b}^\top \mathbf{b}.$$

Since we have the definition:

$$\|\mathbf{y}\|^2 = \mathbf{y}^\top \mathbf{y} = \sum_{i=1}^{n} y_i^2$$

Therefore,

$$f(\mathbf{x}) = \|A\mathbf{x} - \mathbf{b}\|^2 = (A\mathbf{x} - \mathbf{b})^\top (A\mathbf{x} - \mathbf{b})$$

Using the distributive property of the transpose and inner product:

$$(A\mathbf{x} - \mathbf{b})^\top (A\mathbf{x} - \mathbf{b}) = (A\mathbf{x})^\top (A\mathbf{x}) - (A\mathbf{x})^\top \mathbf{b} - \mathbf{b}^\top (A\mathbf{x}) + \mathbf{b}^\top \mathbf{b}$$

Since we have:

$$(A\mathbf{x})^\top = \mathbf{x}^\top A^\top.$$

Thus,

$$(A\mathbf{x})^\top (A\mathbf{x}) = \mathbf{x}^\top A^\top A\mathbf{x}$$
$$(A\mathbf{x})^\top \mathbf{b} = \mathbf{x}^\top A^\top \mathbf{b}$$

Since that $\mathbf{b}^\top (A\mathbf{x})$ is a scalar and equals its transpose, thus,

$$\mathbf{b}^\top (A\mathbf{x}) = (A\mathbf{x})^\top \mathbf{b} = \mathbf{x}^\top A^\top \mathbf{b}.$$

Substituting the simplified terms back into the expanded expression:

$$f(\mathbf{x}) = \mathbf{x}^\top A^\top A\mathbf{x} - \mathbf{x}^\top A^\top \mathbf{b} - \mathbf{x}^\top A^\top \mathbf{b} + \mathbf{b}^\top \mathbf{b}$$
$$f(\mathbf{x}) = \mathbf{x}^\top A^\top A\mathbf{x} - 2\mathbf{x}^\top A^\top \mathbf{b} + \mathbf{b}^\top \mathbf{b}$$

Therefore,

$$f(\mathbf{x}) = \|A\mathbf{x} - \mathbf{b}\|^2 = \mathbf{x}^\top A^\top A\mathbf{x} - 2\mathbf{x}^\top A^\top \mathbf{b} + \mathbf{b}^\top \mathbf{b}$$

### (2)

To show that:

$$\nabla \left(\mathbf{v}^\top \mathbf{x}\right) = \mathbf{v}$$

We need to prove:

$$\frac{\partial}{\partial x_k} \left(\mathbf{v}^\top \mathbf{x}\right) = v_k \quad \text{for all } k = 1, 2, \ldots, n.$$

Let $\mathbf{v}$ and $\mathbf{x}$ be $n$-dimensional column vectors:

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Then, the inner product is:

$$\mathbf{v}^\top \mathbf{x} = v_1 x_1 + v_2 x_2 + \cdots + v_n x_n = \sum_{i=1}^{n} v_i x_i$$

To compute the partial derivative of $\mathbf{v}^\top \mathbf{x}$ with respect to each component $x_k$ of $\mathbf{x}$:

$$\frac{\partial}{\partial x_k} \left( \mathbf{v}^\top \mathbf{x} \right) = \frac{\partial}{\partial x_k} \left( \sum_{i=1}^{n} v_i x_i \right)$$

For all $i \neq k$, the term is a constant and the derivative is zero. Therefore, the partial derivative with respect to $x_k$ affects only the term where $i = k$:

$$\frac{\partial}{\partial x_k} \left( \sum_{i=1}^{n} v_i x_i \right) = \frac{\partial}{\partial x_k} (v_k x_k) = v_k$$

Thus, the gradient $\nabla \left( \mathbf{v}^\top \mathbf{x} \right)$ is a vector composed of all these partial derivatives:

$$\nabla \left( \mathbf{v}^\top \mathbf{x} \right) = \begin{bmatrix} \frac{\partial}{\partial x_1} \left( \mathbf{v}^\top \mathbf{x} \right) \\ \frac{\partial}{\partial x_2} \left( \mathbf{v}^\top \mathbf{x} \right) \\ \vdots \\ \frac{\partial}{\partial x_n} \left( \mathbf{v}^\top \mathbf{x} \right) \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \mathbf{v}$$

Therefore,

$$\nabla \left( \mathbf{v}^\top \mathbf{x} \right) = \mathbf{v}$$

## (3)


## (3)


To show that:

$$\mathbf{x}^\top M \mathbf{x} = \sum_{i,j=1}^{n} m_{ij} x_i x_j$$

Let $M$ be an $n \times n$ symmetric matrix and $\mathbf{x}$ be an $n \times 1$ vector. The product $M\mathbf{x}$ can be expressed component-wise as:

$$(M\mathbf{x})_i = \sum_{j=1}^{n} m_{ij} x_j \quad \text{for } i = 1, 2, \ldots, n.$$

Thus, the vector $M\mathbf{x}$ is:

$$M\mathbf{x} = \begin{bmatrix} \sum_{j=1}^{n} m_{1j} x_j \\ \sum_{j=1}^{n} m_{2j} x_j \\ \vdots \\ \sum_{j=1}^{n} m_{nj} x_j \end{bmatrix}.$$

$\mathbf{x}^\top M\mathbf{x}$ is the inner product of $\mathbf{x}$ with $M\mathbf{x}$:

$$\mathbf{x}^\top M\mathbf{x} = \sum_{i=1}^{n} x_i (M\mathbf{x})_i = \sum_{i=1}^{n} x_i \left( \sum_{j=1}^{n} m_{ij} x_j \right).$$

Rearranging the summations:

$$\mathbf{x}^\top M\mathbf{x} = \sum_{i=1}^{n} \sum_{j=1}^{n} m_{ij} x_i x_j.$$

Since $M$ is symmetric ($m_{ij} = m_{ji}$), this can also be written as:

$$\mathbf{x}^\top M\mathbf{x} = \sum_{i,j=1}^{n} m_{ij} x_i x_j.$$

Therefore,

$$\mathbf{x}^\top M\mathbf{x} = \sum_{i,j=1}^{n} m_{ij} x_i x_j.$$

Given the quadratic form:

$$f(\mathbf{x}) = \mathbf{x}^\top M\mathbf{x} = \sum_{i,j=1}^{n} m_{ij} x_i x_j,$$

compute the partial derivative of $f$ with respect to $x_k$:

$$\frac{\partial f}{\partial x_k} = \frac{\partial}{\partial x_k} \left( \sum_{i=1}^{n} \sum_{j=1}^{n} m_{ij} x_i x_j \right).$$

Since $m_{ij}$ are constants, the derivative acts only on the terms involving $x_i$ and $x_j$.

Similar to section 2, we have:

$$\frac{\partial}{\partial x_k}(x_i x_j) = \begin{cases} x_j & \text{if } i = k, \\ x_i & \text{if } j = k, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, the partial derivative becomes:

$$\frac{\partial f}{\partial x_k} = \sum_{i=1}^{n} \sum_{j=1}^{n} m_{ij} \frac{\partial}{\partial x_k}(x_i x_j) = \sum_{i=1}^{n} \sum_{j=1}^{n} m_{ij} \left( \frac{\partial}{\partial x_k} x_i \cdot x_j + x_i \cdot \frac{\partial}{\partial x_k} x_j \right).$$

This can be split into two sums:

$$\frac{\partial f}{\partial x_k} = \sum_{i=1}^{n} \sum_{j=1}^{n} m_{ij} \delta_{ik} x_j + \sum_{i=1}^{n} \sum_{j=1}^{n} m_{ij} x_i \delta_{jk},$$

where $\delta_{ik}$ is 1 if $i = k$ and 0 otherwise, and similarly for $\delta_{jk}$.

For the first sum, since $\delta_{ik}$ is only non-zero when $i = k$:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} m_{ij} \delta_{ik} x_j = \sum_{j=1}^{n} m_{kj} x_j.$$

For the second sum, $\delta_{jk}$ is only non-zero when $j = k$:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} m_{ij} x_i \delta_{jk} = \sum_{i=1}^{n} m_{ik} x_i.$$

Thus,

$$\frac{\partial f}{\partial x_k} = \sum_{j=1}^{n} m_{kj} x_j + \sum_{i=1}^{n} m_{ik} x_i.$$

Since $M$ is symmetric ($M^\top = M$), we have $m_{kj} = m_{jk}$ and $m_{ik} = m_{ki}$. Therefore:

$$\frac{\partial f}{\partial x_k} = \sum_{j=1}^{n} m_{jk} x_j + \sum_{i=1}^{n} m_{ki} x_i.$$

Now, observe that both sums involve the same terms due to the symmetry $m_{ki} = m_{ik} = m_{jk}$ when $i = j$. Thus, we can combine them as follows:

$$\frac{\partial f}{\partial x_k} = \sum_{j=1}^{n} m_{jk} x_j + \sum_{j=1}^{n} m_{jk} x_j = \sum_{j=1}^{n} (m_{jk} + m_{jk}) x_j = 2 \sum_{j=1}^{n} m_{jk} x_j = 2(M\mathbf{x})_k.$$

The gradient vector $\nabla f(\mathbf{x})$ is composed of all these partial derivatives:

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = \begin{bmatrix} 2(M\mathbf{x})_1 \\ 2(M\mathbf{x})_2 \\ \vdots \\ 2(M\mathbf{x})_n \end{bmatrix} = 2M\mathbf{x}.$$

Therefore, the gradient of the quadratic form $\mathbf{x}^\top M \mathbf{x}$ with respect to $\mathbf{x}$ is:

$$\nabla \left( \mathbf{x}^\top M \mathbf{x} \right) = 2M\mathbf{x}.$$

## (4)

First, we need to show that:

$$\nabla f(\mathbf{x}) = 2(A^\top A\mathbf{x} - A^\top \mathbf{b}).$$

From Part 1 we have:

$$\begin{aligned} f(\mathbf{x}) &= \|A\mathbf{x} - \mathbf{b}\|^2 = (A\mathbf{x} - \mathbf{b})^\top (A\mathbf{x} - \mathbf{b}). \\ &= \mathbf{x}^\top A^\top A\mathbf{x} - 2\mathbf{x}^\top A^\top \mathbf{b} + \mathbf{b}^\top \mathbf{b}. \end{aligned}$$

From Part 3, we know that for a quadratic form $\mathbf{x}^\top M \mathbf{x}$, the gradient is $\nabla(\mathbf{x}^\top M \mathbf{x}) = 2M\mathbf{x}$.

Here, let $M = A^\top A$. Then, the gradient of the first term $\mathbf{x}^\top A^\top A\mathbf{x}$ is:

$$\nabla(\mathbf{x}^\top A^\top A\mathbf{x}) = 2A^\top A\mathbf{x}.$$

From Part 2, we have established that for vector $\mathbf{v}$:

$$\nabla \left( \mathbf{v}^\top \mathbf{x} \right) = \mathbf{v}.$$

Let:

$$\mathbf{v} = -2A^\top \mathbf{b}.$$

Therefore,

$$-2\mathbf{x}^\top A^\top \mathbf{b} = \mathbf{v}^\top \mathbf{x}.$$

Therefore, the gradient of the linear term is:

$$\nabla(-2\mathbf{x}^\top A^\top \mathbf{b}) = \nabla\left(\mathbf{v}^\top \mathbf{x}\right) = \mathbf{v} = -2A^\top \mathbf{b}.$$

The constant term $\mathbf{b}^\top \mathbf{b}$ has a gradient of zero.

The gradient of the objective function $f(\mathbf{x})$ is:

$$\nabla f(\mathbf{x}) = 2A^\top A\mathbf{x} - 2A^\top \mathbf{b} = 2(A^\top A\mathbf{x} - A^\top \mathbf{b}).$$

Therefore,

$$\nabla f(\mathbf{x}) = 2(A^\top A\mathbf{x} - A^\top \mathbf{b}).$$

To find the minimizer $\mathbf{x}^*$ of the objective function $f(\mathbf{x})$, we set the gradient of $f(\mathbf{x})$ to zero:

$$\nabla f(\mathbf{x}^*) = 0.$$

$$2(A^\top A\mathbf{x}^* - A^\top \mathbf{b}) = 0.$$
$$A^\top A\mathbf{x}^* - A^\top \mathbf{b} = 0.$$
$$A^\top A\mathbf{x}^* = A^\top \mathbf{b}.$$

Setting the gradient of the objective function to zero leads directly to the normal equations:

$$A^\top A\mathbf{x}^* = A^\top \mathbf{b}.$$

Therefore, $f(\mathbf{x})$ is minimized when $\mathbf{x}^*$ satisfies the normal equations.