ECE586/AI586 Applied Matrix Analysis - Homework 1

Winter 2024
Instructor: Xiao Fu

School of Electrical Engineering and Computer Science
Oregon State University

January 12, 2025

Woonki Kim
kimwoon@oregonstate.edu

Q1. Show the following facts for subspaces:

(a) $\mathcal{R}(\mathbf{A})^\perp = \mathcal{N}(\mathbf{A}^T)$ (5%)

(b) $\mathcal{N}(\mathbf{A}) = \mathcal{R}(\mathbf{A}^T)^\perp$ (5%)

A1.

(a) $\mathcal{R}(\mathbf{A})^\perp = \mathcal{N}(\mathbf{A}^T)$:

For $\mathbf{A} \in \mathbb{R}^{m\times n}$

$$\mathcal{R}(\mathbf{A}) = \{\mathbf{y} \in \mathbb{R}^m \mid \mathbf{y} = \mathbf{A}\mathbf{x},\, \mathbf{x} \in \mathbb{R}^n\}$$
$$\mathcal{R}(\mathbf{A})^\perp = \{\mathbf{y} \in \mathbb{R}^m \mid \mathbf{z}^T\mathbf{y} = 0,\, \forall \mathbf{z} \in \mathcal{R}(\mathbf{A})\}$$
$$\mathcal{N}(\mathbf{A}^T) = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}^T\mathbf{x} = \mathbf{0}\}$$

To prove $\mathcal{R}(\mathbf{A})^\perp = \mathcal{N}(\mathbf{A}^T)$, we need to show:

$$\mathcal{R}(\mathbf{A})^\perp \subseteq \mathcal{N}(\mathbf{A}^T)$$
$$\mathcal{N}(\mathbf{A}^T) \subseteq \mathcal{R}(\mathbf{A})^\perp$$

1. Prove $\mathcal{R}(\mathbf{A})^\perp \subseteq \mathcal{N}(\mathbf{A}^T)$:

Proof by construction:

Let $\mathbf{z} \in \mathcal{R}(\mathbf{A})^\perp$, then $\mathbf{z}^T\mathbf{y} = 0, \forall \mathbf{y} \in \mathcal{R}(\mathbf{A})$.

Since $\mathbf{y} \in \mathcal{R}(\mathbf{A})$, by definition of range space:

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \mathbf{x} \in \mathbb{R}^n$$

To verify orthogonality with all vectors in $\mathcal{R}(\mathbf{A})$, we note that $\mathbf{z}^T\mathbf{y} = 0$ must hold for every possible $\mathbf{y} \in \mathcal{R}(\mathbf{A})$:

Substituting $\mathbf{y} = \mathbf{A}\mathbf{x}$, we get:

$$\mathbf{z}^T\mathbf{y} = \mathbf{z}^T\mathbf{A}\mathbf{x} = \mathbf{0}, \quad \forall \mathbf{x} \in \mathbb{R}^n.$$

Since $\mathbf{x}$ can be any vector in $\mathbb{R}^n$, the condition $\mathbf{z}^T(\mathbf{A}\mathbf{x}) = 0$ must hold for all $\mathbf{x}$:

$$\mathbf{z}^T\mathbf{A}\mathbf{x} = \mathbf{0}, \quad \forall \mathbf{x} \in \mathbb{R}^n.$$

To make this equation true for all $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{z}^T\mathbf{A} = \mathbf{0}$ should hold.
Thus:

$$(\mathbf{z}^T\mathbf{A})^T = \mathbf{A}^T\mathbf{z} = \mathbf{0}$$

Which is definition of null space, proving that for any $\mathbf{z} \in \mathcal{R}(A)$, $\mathbf{z} \in \mathcal{N}(\mathbf{A}^T)$ holds.
Thus:

$$\mathcal{R}(\mathbf{A})^\perp \subseteq \mathcal{N}(\mathbf{A}^T)$$

2. Prove $\mathcal{N}(\mathbf{A}^T) \subseteq \mathcal{R}(\mathbf{A})^\perp$:

Proof by construction:
By definition of the null space, let $\mathbf{z} \in \mathcal{N}(\mathbf{A}^T)$, then:

$$\mathbf{A}^T\mathbf{z} = \mathbf{0}.$$

Taking the transpose, we have:

$$(\mathbf{A}^T\mathbf{z})^T = \mathbf{z}^T\mathbf{A} = \mathbf{0}^T.$$

Now, lets consider the range space. Let $\mathbf{y} \in \mathcal{R}(A)$
By definition of $\mathcal{R}(\mathbf{A})$, there exists some $\mathbf{x} \in \mathbb{R}^n$ such that:

$$\mathbf{y} = \mathbf{A}\mathbf{x}.$$

Now to verify orthogonality with all vectors in $\mathcal{R}(\mathbf{A})$, we need to check that $\mathbf{z}^T\mathbf{y} = 0$ holds for every possible $\mathbf{y} \in \mathcal{R}(\mathbf{A})$:

$$\mathbf{z}^T\mathbf{y} = \mathbf{z}^T(\mathbf{A}\mathbf{x}).$$

By associativity of matrix multiplication:

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

Since $\mathbf{z}^T\mathbf{A} = 0$, it follows that:

$$(\mathbf{z}^T\mathbf{A})\mathbf{x} = 0.$$

Therefore:

$$\mathbf{z}^T\mathbf{y} = (\mathbf{z}^T\mathbf{A})\mathbf{x} = 0, \quad \forall \mathbf{y} \in \mathcal{R}(\mathbf{A}).$$

Since $\mathbf{z}^T\mathbf{y} = 0$ for all $\mathbf{y} \in \mathcal{R}(\mathbf{A})$, we conclude that:

$$\mathbf{z} \in \mathcal{R}(\mathbf{A})^{\perp}.$$

Thus, for any $\mathbf{z} \in \mathcal{N}(\mathbf{A}^T)$ also belongs to $\mathcal{R}(\mathbf{A})^{\perp}$. Therefore:

$$\mathcal{N}(\mathbf{A}^T) \subseteq \mathcal{R}(\mathbf{A})^{\perp}.$$

Since $\mathcal{R}(\mathbf{A})^{\perp} \subseteq \mathcal{N}(\mathbf{A}^T)$ and $\mathcal{N}(\mathbf{A}^T) \subseteq \mathcal{R}(\mathbf{A})^{\perp}$, we have:

$$\mathcal{R}(\mathbf{A})^{\perp} = \mathcal{N}(\mathbf{A}^T).$$

(b) **Prove $\mathcal{N}(\mathbf{A}) = \mathcal{R}(\mathbf{A}^T)^{\perp}$:**

$$\mathcal{N}(\mathbf{A}) = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} = \mathbf{0}\}$$
$$\mathcal{R}(\mathbf{A}^T) = \{\mathbf{y} \in \mathbb{R}^n \mid \mathbf{y} = \mathbf{A}^T\mathbf{z},\, \mathbf{z} \in \mathbb{R}^m\}$$
$$\mathcal{R}(\mathbf{A}^T)^{\perp} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{w}^T\mathbf{x} = 0,\, \forall \mathbf{w} \in \mathcal{R}(\mathbf{A}^T)\}$$

To prove $\mathcal{N}(\mathbf{A}) = \mathcal{R}(\mathbf{A}^T)^{\perp}$, we show the following:

$$\mathcal{N}(\mathbf{A}) \subseteq \mathcal{R}(\mathbf{A}^T)^{\perp}$$
$$\mathcal{R}(\mathbf{A}^T)^{\perp} \subseteq \mathcal{N}(\mathbf{A})$$

1. Prove $\mathcal{N}(\mathbf{A}) \subseteq \mathcal{R}(\mathbf{A}^T)^{\perp}$:
Let $\mathbf{x} \in \mathcal{N}(\mathbf{A})$. Then $\mathbf{A}\mathbf{x} = \mathbf{0}$.
Let $\mathbf{w} \in \mathcal{R}(\mathbf{A}^T)$. Then $\mathbf{w} = \mathbf{A}^T\mathbf{z}$, $\mathbf{z} \in \mathbb{R}^m$.
Taking the dot product $\mathbf{w}^T\mathbf{x}$, we get:

$$\mathbf{w}^T\mathbf{x} = (\mathbf{A}^T\mathbf{z})^T\mathbf{x} = \mathbf{z}^T(\mathbf{A}\mathbf{x}) = \mathbf{z}^T\mathbf{0} = 0.$$

which holds for all $\mathbf{w}$
Thus, $\mathbf{x} \in \mathcal{R}(\mathbf{A}^T)^{\perp}$.

2. Prove $\mathcal{R}(\mathbf{A}^T)^{\perp} \subseteq \mathcal{N}(\mathbf{A})$:

The orthogonal complement $\mathcal{R}(\mathbf{A}^T)^{\perp}$ consists of all vectors $\mathbf{y}$ such that

$$\mathbf{v}^T\mathbf{y} = 0 \quad \text{for all } \mathbf{v} \in \mathcal{R}(\mathbf{A}^T).$$

Since $\mathcal{R}(\mathbf{A}^T)$ is the range space of $\mathbf{A}^T$, any vector $\mathbf{v} \in \mathcal{R}(\mathbf{A}^T)$ can be written as $\mathbf{v} = \mathbf{A}^T\mathbf{x}$ for some $\mathbf{x}$.

Thus, $\mathcal{R}(\mathbf{A}^T)^\perp$ can be expressed as:

$$\mathcal{R}(\mathbf{A}^T)^\perp = \left\{\mathbf{y} \mid (\mathbf{A}^T\mathbf{x})^T\mathbf{y} = 0 \forall \mathbf{x}\right\}.$$

Equivalently, we can write:

$$\mathbf{x}^T(\mathbf{A}\mathbf{y}) = 0 \quad \forall \mathbf{x}.$$

Since this equation should hold for all $\mathbf{x}$:

$$\mathbf{A}\mathbf{y} = \mathbf{0}.$$

Any $\mathbf{y} \in \mathcal{R}(\mathbf{A}^T)^\perp$ satisfies $\mathbf{A}\mathbf{y} = \mathbf{0}$, which means $\mathbf{y} \in \mathcal{N}(\mathbf{A})$.

Thus:

$$\mathcal{R}(\mathbf{A}^T)^\perp \subseteq \mathcal{N}(\mathbf{A})$$

Since $\mathcal{N}(\mathbf{A}) \subseteq \mathcal{R}(\mathbf{A}^T)^\perp$ and $\mathcal{R}(\mathbf{A}^T)^\perp \subseteq \mathcal{N}(\mathbf{A})$, we have:

$$\mathcal{N}(\mathbf{A}) = \mathcal{R}(\mathbf{A}^T)^\perp.$$

Q2. Show the following statements for $\mathbf{x} \in \mathbb{R}^n$:

(a) $\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1 \leq \sqrt{n}\|\mathbf{x}\|_2$ (3%)

(b) $\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_2 \leq \sqrt{n}\|\mathbf{x}\|_\infty$ (3%)

(c) $\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_1 \leq n\|\mathbf{x}\|_\infty$ (4%)

A2.

(a) Show $\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1 \leq \sqrt{n}\|\mathbf{x}\|_2$:

1. Prove $\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1$:

Since norm is non-negative, $\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1$ holds if and only if $\|\mathbf{x}\|_2^2 \leq \|\mathbf{x}\|_1^2$ holds.

Expand $\|\mathbf{x}\|_1^2$ and $\|\mathbf{x}\|_2^2$:

$$\|\mathbf{x}\|_1^2 = \left(\sum_{i=1}^{n} |x_i|\right)^2 = \sum_{i=1}^{n} |x_i|^2 + 2\sum_{i<j} |x_i x_j|$$

$$\|\mathbf{x}\|_2^2 = \sum_{i=1}^{n} |x_i|^2.$$

Since the extra cross terms $2\sum_{i<j}|x_i x_j|$ is sum of absolute value making it at most positive, and equal to zero when number of non-zero element in $\mathbf{x}$ is less than or equal to one:

$$\sum_{i=1}^{n}|x_i|^2 \leq \sum_{i=1}^{n}|x_i|^2 + 2\sum_{i<j}|x_i x_j|$$

Thus:

$$\|\mathbf{x}\|_2^2 \leq \|\mathbf{x}\|_1^2$$

Since norm is a non-negative, inequality still holds when taking square root:

$$\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1.$$

2. Prove $\|\mathbf{x}\|_1 \leq \sqrt{n}\|\mathbf{x}\|_2$:

Let:

$$\mathbf{x} = (x_1, x_2, \cdots, x_n)^T, \mathbf{y} = (1, 1, \cdots, 1)^T$$

By the Cauchy-Schwartz inequality, we have:

$$|\mathbf{x} \cdot \mathbf{y}| \leq \|\mathbf{x}\|_2\|\mathbf{y}\|_2$$

Since, left hand side is absolute value of linear combination of two vectors and right hand side is multiplication of norms, both sides are all non-negative. Thus, inequality still holds when taking square on both sides:

$$|\mathbf{x} \cdot \mathbf{y}|^2 \leq \|\mathbf{x}\|_2^2\|\mathbf{y}\|_2^2$$

Reformulating:

$$\left(\sum_{i=1}^{n}|x_i||y_i|\right)^2 \leq \left(\sum_{i=1}^{n}|x_i|^2\right)\left(\sum_{i=1}^{n}|y_i|^2\right),$$

Since, $y_i = 1, \forall y_i$

$$\left(\sum_{i=1}^{n}|x_i|\right)^2 \leq \left(\sum_{i=1}^{n}|x_i|^2\right) \cdot n$$

Taking square root on both sides:

$$\sum_{i=1}^{n}|x_i| \leq \sqrt{\left(\sum_{i=1}^{n}|x_i|^2\right)} \cdot \sqrt{n}$$

Expressing as norm:

$$\|\mathbf{x}\|_1 \leq \sqrt{n} \cdot \|\mathbf{x}\|_2$$

Combining the 1. and 2. we get:

$$\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1 \leq \sqrt{n}\|\mathbf{x}\|_2.$$

(b) Show $\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_2 \leq \sqrt{n}\|\mathbf{x}\|_\infty$

1. Prove $\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_2$:

Since $\|\mathbf{x}\|_2^2 = \sum_{i=1}^n |x_i|^2$, it is obvious that $|x_i|^2 \leq \|\mathbf{x}\|_2^2,\ \forall x_i$.

This inequality is for all x_i, including the maximum value:

$$\max |x_i|^2 \leq \|\mathbf{x}\|_2^2.$$

Reformulating using norm:

$$\|\mathbf{x}\|_\infty^2 \leq \|\mathbf{x}\|_2^2.$$

Since both $\|\mathbf{x}\|_\infty$ and $\|\mathbf{x}\|_2$ are norm they are non-negative, thus $\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_2$ holds if and only if $\|\mathbf{x}\|_\infty^2 \leq \|\mathbf{x}\|_2^2$ holds

Taking the square root:

$$\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_2.$$

2. Prove $\|\mathbf{x}\|_2 \leq \sqrt{n}\|\mathbf{x}\|_\infty$:

Since $|x_i| \leq |x|_{max}\,\forall x_i$, sum of all elements in $\mathbf{x}$ is always smaller than n times the element with maximum absolute value, and is equal when all x_i has same absolute value.

Thus expressing using summation:

$$\sum_{i=1}^n |x_i| \leq \sum_{i=1}^n |x|_{max}$$

Since in this equation both sides are sum of absolute values they are non-negative. Thus taking square does not change inequality:

$$\sum_{i=1}^n |x_i|^2 \leq \sum_{i=1}^n |x|_{max}^2$$

Taking the square root:

$$\left(\sum_{i=1}^n |x_i|^2\right)^{1/2} \leq \left(\sum_{i=1}^n |x|_{max}^2\right)^{1/2}$$

Expressing with norms:

$$\|\mathbf{x}\|_2 \leq \left(n\|\mathbf{x}\|_\infty^2\right)^{1/2} = \sqrt{n}\|\mathbf{x}\|_\infty$$

Thus:

$$\|\mathbf{x}\|_2 \leq \sqrt{n}\|\mathbf{x}\|_\infty$$

Combining 1. and 2., we get:

$$\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_2 \leq \sqrt{n}\|\mathbf{x}\|_\infty.$$

(c) Show $\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_1 \leq n\|\mathbf{x}\|_\infty$:

- $\|\mathbf{x}\|_\infty = \max_i |x_i|$
- $\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$

1. Prove $\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_1$:
Since $\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$, it is obvious that $|x_i| \leq \|\mathbf{x}\|_1,\ \forall x_i$.
This inequality is for all x_i, including the maximum value:

$$\max |x_i| \leq \|\mathbf{x}\|_1.$$

Reformulating using norm:

$$\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_1.$$

2. Prove $\|\mathbf{x}\|_1 \leq n\|\mathbf{x}\|_\infty$:
Again, since $|x_i| \leq \|\mathbf{x}\|_\infty \forall x_i$, we have:

$$\sum_{i=1}^n |x_i| \leq \sum_{i=1}^n |\mathbf{x}|_{max} = n|\mathbf{x}|_{max}$$

Expressing in norm:

$$\|\mathbf{x}\|_1 \leq n\|\mathbf{x}\|_\infty.$$

Combining the two parts, we have:

$$\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_1 \leq n\|\mathbf{x}\|_\infty.$$

Q3. Show that if $\mathbf{A} \in \mathbb{R}^{m\times m}$ is triangular, either upper or lower, then the following holds:

$$\det(\mathbf{A}) = \prod_{i=1}^m a_{ii}.$$

(10%)

A3.
A square matrix $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{m\times m}$ is triangular if all entries either above or below the diagonal are zero:

- Upper triangular: $a_{ij} = 0$ for $i > j$.
- Lower triangular: $a_{ij} = 0$ for $i < j$.

The determinant of a matrix is defined recursively using minors or can be computed directly via cofactor expansion:

$$\det(\mathbf{A}) = \sum_{j=1}^{m} a_{ij} c_{ij}, \text{ for any } i = 1, ...m$$

where $c_{ij} = (-1)^{i+j} \det(A_{ij})$.

Case 1: Upper Triangular Matrix ($a_{ij} = 0$ for $i > j$):
The determinant of an $m \times m$ upper triangular matrix can be computed using cofactor expansion along the first row. The determinant simplifies as follows:

$$\det(\mathbf{A}) = \sum_{j=1}^{m} a_{1j} c_{1j}.$$

Since $\mathbf{A}$ is upper triangular, all elements below the diagonal are zero. Thus, expanding the determinant along any row or column only retains terms involving diagonal entries.

Proof by Induction:
For $m = 1$ and $m = 2$, this result is straightforward:
For $m = 1$, the matrix is $\mathbf{A} = [a_{11}]$, and $\det(\mathbf{A}) = a_{11}$.
For $m = 2$, the matrix is:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{bmatrix},$$

and the determinant is:

$$\det(\mathbf{A}) = a_{11}a_{22} - 0 = a_{11}a_{22}.$$

These cases are too obvious, so we use base case as $m = 3$.

Base Case: $m = 3$
Now consider an upper triangular 3×3 matrix:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{bmatrix}.$$

Using cofactor expansion along the first row:

$$\det(\mathbf{A}) = a_{11} \det \begin{bmatrix} a_{22} & a_{23} \\ 0 & a_{33} \end{bmatrix} - a_{12} \cdot 0 + a_{13} \cdot 0.$$

Thus:

$$\det(\mathbf{A}) = a_{11} \det \begin{bmatrix} a_{22} & a_{23} \\ 0 & a_{33} \end{bmatrix}.$$

Now compute the determinant of the 2×2 matrix:

$$\det \begin{bmatrix} a_{22} & a_{23} \\ 0 & a_{33} \end{bmatrix} = a_{22}a_{33}.$$

Substituting back:

$$\det(\mathbf{A}) = a_{11}(a_{22}a_{33}) = a_{11}a_{22}a_{33} = \prod_{i=1}^{3} a_{ii}$$

Thus, the result holds for $m = 3$.

Induction Step:
Assume the result holds for an upper triangular matrix of size $(m-1) \times (m-1)$:

$$\det(\mathbf{A}_{m-1}) = \prod_{i=1}^{m-1} a_{ii},$$

where $\mathbf{A}_{m-1}$ is an upper triangular matrix of size $(m-1) \times (m-1)$.
Now consider an $m \times m$ upper triangular matrix $\mathbf{A}$:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ 0 & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{mm} \end{bmatrix}.$$

Expand the determinant along the first row:

$$\det(\mathbf{A}) = a_{11} \det(\mathbf{A}_{11}) - \sum_{j=2}^{m} a_{1j} \cdot 0.$$

Thus:

$$\det(\mathbf{A}) = a_{11} \det(\mathbf{A}_{11}),$$

where $\mathbf{A}_{11}$ is the $(m-1) \times (m-1)$ upper triangular submatrix obtained by removing the first row and first column of $\mathbf{A}$.
By the inductive hypothesis:

$$\det(\mathbf{A}_{11}) = \prod_{i=2}^{m} a_{ii}.$$

Substituting back:

$$\det(\mathbf{A}) = a_{11} \prod_{i=2}^{m} a_{ii} = \prod_{i=1}^{m} a_{ii}.$$

Thus it is true for upper triangular matrix $\mathbf{A} \in \mathbb{R}^{m \times m}$, that:

$$\det(\mathbf{A}) = \prod_{i=1}^{m} a_{ii}.$$

Case 2: Lower Triangular Matrix ($a_{ij} = 0$ for $i < j$):
Proof by Induction:
For $m = 1$ and $m = 2$, this result is straightforward:
For $m = 1$, the matrix is $\mathbf{A} = [a_{11}]$, and $\det(\mathbf{A}) = a_{11}$.
For $m = 2$, the matrix is:

$$\mathbf{A} = \begin{bmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{bmatrix},$$

and the determinant is:

$$\det(\mathbf{A}) = a_{11}a_{22} - 0 = a_{11}a_{22}.$$

These cases are too obvious, so we use base case as $m = 3$.

Base Case: $m = 3$
Now consider a 3×3 lower triangular matrix:

$$\mathbf{A} = \begin{bmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

Using cofactor expansion along the first row:

$$\det(\mathbf{A}) = a_{11} \det \begin{bmatrix} a_{22} & 0 \\ a_{32} & a_{33} \end{bmatrix} - 0 + 0.$$

Thus:

$$\det(\mathbf{A}) = a_{11} \det \begin{bmatrix} a_{22} & 0 \\ a_{32} & a_{33} \end{bmatrix}.$$

Now compute the determinant of the 2×2 matrix:

$$\det \begin{bmatrix} a_{22} & 0 \\ a_{32} & a_{33} \end{bmatrix} = a_{22}a_{33} - 0 = a_{22}a_{33}.$$

Substituting back:

$$\det(\mathbf{A}) = a_{11}(a_{22}a_{33}) = a_{11}a_{22}a_{33} = \prod_{i=1}^{3} a_{ii}$$

Thus, the result holds for $m = 3$.

Induction Step:
Assume the result holds for a lower triangular matrix of size $(m-1) \times (m-1)$, i.e.,

$$\det(\mathbf{A}_{m-1}) = \prod_{i=1}^{m-1} a_{ii},$$

where $\mathbf{A}_{m-1}$ is a lower triangular matrix of size $(m-1) \times (m-1)$.

Now consider an $m \times m$ lower triangular matrix $\mathbf{A}$:

$$\mathbf{A} = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{bmatrix}.$$

Expand the determinant along the first row:

$$\det(\mathbf{A}) = a_{11} \det(\mathbf{A}_{11}) + \sum_{j=2}^{m} 0.$$

Thus:

$$\det(\mathbf{A}) = a_{11} \det(\mathbf{A}_{11}),$$

where $\mathbf{A}_{11}$ is the $(m-1) \times (m-1)$ lower triangular submatrix obtained by removing the first row and first column of $\mathbf{A}$.
By the inductive hypothesis:

$$\det(\mathbf{A}_{11}) = \prod_{i=2}^{m} a_{ii}.$$

Substituting back:

$$\det(\mathbf{A}) = a_{11} \prod_{i=2}^{m} a_{ii} = \prod_{i=1}^{m} a_{ii}.$$

Thus it is true for lower triangular matrix $\mathbf{A} \in \mathbb{R}^{m \times m}$, that:

$$\det(\mathbf{A}) = \prod_{i=1}^{m} a_{ii}.$$

For both upper and lower triangular matrices, the determinant is given by:

$$\det(\mathbf{A}) = \prod_{i=1}^{m} a_{ii}.$$

Q4. Consider a square matrix $\mathbf{A} \in \mathbb{R}^{m \times m}$. The matrix is nonsingular if and only if $\mathbf{Ax} \neq 0$ for any $\mathbf{x}$. One interesting type of matrix is the so-called diagonally dominant matrices, which admit the following property:

$$|a_{ii}| > \sum_{j=1, j \neq i} |a_{ij}|.$$

Show that any diagonally dominant matrix is nonsingular. (10%)

A4.
Proof by Contradiction:

Assume A is singular:
If A is singular, then there exists a nonzero vector $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T$ such that:

$$A\mathbf{x} = 0.$$

This means:

$$\sum_{j=1}^{n} a_{ij} x_j = 0, \quad \forall i = 1, 2, \ldots, n.$$

Rewrite the equation for each i:

$$a_{ii} x_i + \sum_{j \neq i} a_{ij} x_j = 0.$$

Rearrange:

$$x_i = -\frac{\sum_{j \neq i} a_{ij} x_j}{a_{ii}}.$$

Consider the magnitude of x_i: Taking the absolute value:

$$|x_i| = \frac{\left|\sum_{j \neq i} a_{ij} x_j\right|}{|a_{ii}|}.$$

Use the triangle inequality:

$$|x_i| = \frac{\left|\sum_{j \neq i} a_{ij} x_j\right|}{|a_{ii}|} \leq \frac{\sum_{j \neq i} |a_{ij}||x_j|}{|a_{ii}|}.$$

Let $|x|_{\max} = \max_{1 \leq i \leq n} |x_i|$. Then for all i:

$$|x_i| \leq \frac{\sum_{j \neq i} |a_{ij}||x_j|}{|a_{ii}|} \leq \frac{\sum_{j \neq i} |a_{ij}||x|_{\max}}{|a_{ii}|}.$$

still holds, since $|x_i| \leq |x|_{\max}, \forall x_i$.
Divide through by $|x|_{\max}$ ($|x|_{\max} \neq 0$ since x is non-zero vector):

$$\frac{|x_i|}{|x|_{\max}} \leq \frac{\sum_{j \neq i} |a_{ij}|}{|a_{ii}|}.$$

By the definition of diagonal dominant matrix:

$$|a_{ii}| > \sum_{j \neq i} |a_{ij}|.$$

Then:

$$1 > \frac{\sum_{j \neq i} |a_{ij}|}{|a_{ii}|} \geq \frac{|x_i|}{|x|_{\max}}.$$

Thus:

$$1 > \frac{|x_i|}{|x|_{\max}}, \quad \forall i.$$

This is saying that $|x_i| < |x|_{\max}$ for all i, which is a contradiction because $|x|_{\max}$ is the maximum value of $|x_i|$, but $|x_i| < |x|_{\max}$ shows that no x_i can be $x_{\max}$ contradicting our assumption.

Thus, diagonally dominant matrices are non-singular.

Q5. Consider a transformation $\mathbf{y} = \mathbf{Q}\mathbf{x}$, where

$$\mathbf{Q} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$

Show that $\mathbf{Q}$ rotates $\mathbf{x}$ by an angle of θ. (Hint: use the definition of θ) (10%)

A5.
Let vector $\mathbf{x} = [x_1, x_2]^T$, $x \in \mathbb{R}^2$
After applying the transformation $\mathbf{y} = \mathbf{Q}\mathbf{x}$:

$$\mathbf{y} = \mathbf{Q}\mathbf{x} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

$$\mathbf{y} = \begin{bmatrix} \cos(\theta)x_1 - \sin(\theta)x_2 \\ \sin(\theta)x_1 + \cos(\theta)x_2 \end{bmatrix}.$$

To show that $\mathbf{Q}$ preserves the magnitude of $\mathbf{x}$, we need to calculate the magnitude of $\mathbf{y}$ to verify that it remains the same with the magnitude of a vector $\mathbf{x}$ which is $\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2}$. Lets calculate the magnitude of $\mathbf{y}$:

$$\|\mathbf{y}\|^2 = y_1^2 + y_2^2.$$

Substituting $y_1 = \cos(\theta)x_1 - \sin(\theta)x_2$ and $y_2 = \sin(\theta)x_1 + \cos(\theta)x_2$:

$$y_1^2 = (\cos(\theta)x_1 - \sin(\theta)x_2)^2 = \cos^2(\theta)x_1^2 - 2\cos(\theta)\sin(\theta)x_1x_2 + \sin^2(\theta)x_2^2,$$

$$y_2^2 = (\sin(\theta)x_1 + \cos(\theta)x_2)^2 = \sin^2(\theta)x_1^2 + 2\cos(\theta)\sin(\theta)x_1x_2 + \cos^2(\theta)x_2^2.$$

Adding y_1^2 and y_2^2:

$$y_1^2+y_2^2 = \cos^2(\theta)x_1^2-2\cos(\theta)\sin(\theta)x_1x_2+\sin^2(\theta)x_2^2+\sin^2(\theta)x_1^2+2\cos(\theta)\sin(\theta)x_1x_2+\cos^2(\theta)x_2^2.$$

Simplify using $\cos^2(\theta) + \sin^2(\theta) = 1$:

$$y_1^2 + y_2^2 = (\cos^2(\theta) + \sin^2(\theta))x_1^2 + (\cos^2(\theta) + \sin^2(\theta))x_2^2 = x_1^2 + x_2^2.$$

Thus:

$$\|\mathbf{y}\| = \sqrt{y_1^2 + y_2^2} = \sqrt{x_1^2 + x_2^2} = \|\mathbf{x}\|.$$

This shows that the transformation $\mathbf{Q}$ preserves the magnitude of $\mathbf{x}$.

And now we need to show that $\mathbf{Q}$ rotates $\mathbf{x}$ by angle of θ.
As learned in class we can represent angle between two vector $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ as:

$$\phi = \cos^{-1}\left(\frac{\mathbf{y}^T\mathbf{x}}{\|\mathbf{x}\|_2\|\mathbf{y}\|_2}\right),$$

Thus:

$$\cos(\phi) = \frac{\mathbf{y}^T\mathbf{x}}{\|\mathbf{x}\|_2\|\mathbf{y}\|_2}$$

To expand $\cos(\phi)$, we are going to represent $\mathbf{y}^T x$ and $||\mathbf{x}||_2, ||\mathbf{y}||_2$ with θ.
First compute the dot product $\mathbf{y}^T\mathbf{x}$:

$$\mathbf{y}^T\mathbf{x} = \begin{bmatrix}\cos(\theta)x_1 - \sin(\theta)x_2 \\ \sin(\theta)x_1 + \cos(\theta)x_2\end{bmatrix}^T \begin{bmatrix}x_1 \\ x_2\end{bmatrix}.$$

Expanding this:

$$\mathbf{y}^T\mathbf{x} = (\cos(\theta)x_1 - \sin(\theta)x_2)x_1 + (\sin(\theta)x_1 + \cos(\theta)x_2)x_2.$$

Simplify the terms:

$$\mathbf{y}^T\mathbf{x} = \cos(\theta)x_1^2 - \sin(\theta)x_1x_2 + \sin(\theta)x_1x_2 + \cos(\theta)x_2^2.$$

$$\mathbf{y}^T\mathbf{x} = \cos(\theta)(x_1^2 + x_2^2).$$

Second as shown earlier:

$$\|\mathbf{y}\|_2 = \sqrt{y_1^2 + y_2^2} = \sqrt{x_1^2 + x_2^2} = \|\mathbf{x}\|_2.$$

Thus:

$$\|\mathbf{x}\|_2 = \|\mathbf{y}\|_2.$$

Now recall the formula for $\cos(\phi)$:

$$\cos(\phi) = \frac{\mathbf{y}^T\mathbf{x}}{\|\mathbf{x}\|_2\|\mathbf{y}\|_2}.$$

Substituting $\mathbf{y}^T\mathbf{x} = \cos(\theta)(x_1^2 + x_2^2)$ and $\|\mathbf{x}\|_2 = \|\mathbf{y}\|_2 = \sqrt{x_1^2 + x_2^2}$, we get:

$$\cos(\phi) = \frac{\cos(\theta)(x_1^2 + x_2^2)}{(\sqrt{x_1^2 + x_2^2})(\sqrt{x_1^2 + x_2^2})}.$$

Simplify:

$$\cos(\phi) = \cos(\theta) \cdot \frac{x_1^2 + x_2^2}{x_1^2 + x_2^2}.$$

$$\cos(\phi) = \cos(\theta).$$

Since before the modification by Q, $\mathbf{y} = \mathbf{x}$, the angle between the two vectors was $0°$. After the modification, we found that the cosine of the angle between the two vectors ($\cos(\phi)$) is equal to $\cos(\theta)$. This proves that $\mathbf{x}$ has rotated by θ, either in the positive or negative direction.

Thus, we have demonstrated that the magnitude of $\mathbf{x}$ has not changed while its direction has been rotated by θ.

Q6. Given two matrices $\mathbf{X} \in \mathbb{R}^{m\times n}$ and $\mathbf{Y} \in \mathbb{R}^{m\times n}$, show that

$$\mathrm{Tr}(\mathbf{X}^T\mathbf{Y}) = \mathrm{vec}(\mathbf{X})^T \,\mathrm{vec}(\mathbf{Y}),$$

where $\mathrm{vec}(\cdot)$ is the vectorization operator, i.e.,

$$\mathrm{vec}(\mathbf{X}) = [\mathbf{x}_1^T, \ldots, \mathbf{x}_n^T]^T \in \mathbb{R}^{mn},$$

and here $\mathbf{x}_n \in \mathbb{R}^m$ denotes the n-th column of $\mathbf{X}$. Also use the above to show that $\mathrm{Tr}(\mathbf{AB}) = \mathrm{Tr}(\mathbf{BA})$ in our lecture notes. (10%)

A6.
The trace of a matrix $\mathbf{A}$ is defined as the sum of its diagonal elements:

$$\mathrm{Tr}(\mathbf{X}^T\mathbf{Y}) = \sum_{i=1}^{n}\sum_{j=1}^{m}(\mathbf{X}^T\mathbf{Y})_{ij}.$$

The (i, j)-th element of the matrix product $\mathbf{X}^T\mathbf{Y}$ is computed as:

$$(\mathbf{X}^T\mathbf{Y})_{ij} = \sum_{k=1}^{m}(\mathbf{X}^T)_{ik}\mathbf{Y}_{kj}.$$

By the definition of the transpose, $(\mathbf{X}^T)_{ik} = \mathbf{X}_{ki}$. Substituting this into the formula:

$$(\mathbf{X}^T\mathbf{Y})_{ij} = \sum_{k=1}^{m}\mathbf{X}_{ki}\mathbf{Y}_{kj}.$$

The diagonal elements of $\mathbf{X}^T\mathbf{Y}$ occur when $i = j$. Substituting $i = j$ into the formula:

$$(\mathbf{X}^T\mathbf{Y})_{ii} = \sum_{k=1}^{m}\mathbf{X}_{ki}\mathbf{Y}_{ki}.$$

Now, the trace is the sum of all diagonal elements:

$$\mathrm{Tr}(\mathbf{X}^T\mathbf{Y}) = \sum_{i=1}^{n}(\mathbf{X}^T\mathbf{Y})_{ii}.$$

Substitute $(\mathbf{X}^T\mathbf{Y})_{ii} = \sum_{k=1}^{m} \mathbf{X}_{ki}\mathbf{Y}_{ki}$ into the equation:

$$\mathrm{Tr}(\mathbf{X}^T\mathbf{Y}) = \sum_{i=1}^{n}\sum_{k=1}^{m} \mathbf{X}_{ki}\mathbf{Y}_{ki}.$$

By switching the order of summation :

$$\mathrm{Tr}(\mathbf{X}^T\mathbf{Y}) = \sum_{k=1}^{m}\sum_{i=1}^{n} \mathbf{X}_{ki}\mathbf{Y}_{ki}.$$

Since $\mathbf{X}_{ki}$ and $\mathbf{Y}_{ki}$ are just the corresponding elements of $\mathbf{X}$ and $\mathbf{Y}$, we can write this as:

$$\mathrm{Tr}(\mathbf{X}^T\mathbf{Y}) = \sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij}y_{ij}.$$

Now vectorize both matrices
The vectorization operator, $\mathrm{vec}(\mathbf{X})$, reshapes $\mathbf{X}$ into a column vector by stacking its columns:

$$\mathrm{vec}(\mathbf{X}) = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{bmatrix},$$

where $\mathbf{x}_j$ is the j-th column of $\mathbf{X}$, and $\mathrm{vec}(\mathbf{X}) \in \mathbb{R}^{mn}$.
Alternatively, in terms of elements:

$$\mathrm{vec}(\mathbf{X}) = \left[x_{11}, x_{12}, \ldots, x_{1n}, x_{21}, x_{22}, \ldots, x_{2n}, \ldots, x_{m1}, x_{m2}, \ldots, x_{mn}\right]^T.$$

Similarly,

$$\mathrm{vec}(\mathbf{Y}) = \left[y_{11}, y_{12}, \ldots, y_{1n}, y_{21}, y_{22}, \ldots, y_{2n}, \ldots, y_{m1}, y_{m2}, \ldots, y_{mn}\right]^T.$$

The k-th element of $\mathrm{vec}(\mathbf{X})$ corresponds to an element x_{ij} of the original matrix $\mathbf{X}$, where: i indexes the row of $\mathbf{X}$, j indexes the column of $\mathbf{X}$, and the mapping follows the column-stacking order:

$$k = (j-1)m + i,$$

where j runs from 1 to n, and for each j, i runs from 1 to m.

Substituting back into the dot product formula:

$$\mathrm{vec}(\mathbf{X})^T\mathrm{vec}(\mathbf{Y}) = \sum_{k=1}^{mn} \mathrm{vec}(\mathbf{X})_k\,\mathrm{vec}(\mathbf{Y})_k.$$

Since $\mathrm{vec}(\mathbf{X})_k = x_{ij}$ and $\mathrm{vec}(\mathbf{Y})_k = y_{ij}$, we can rewrite the summation over k as a double summation over i and j of the matrices:

$$\mathrm{vec}(\mathbf{X})^T\mathrm{vec}(\mathbf{Y}) = \sum_{j=1}^{n}\sum_{i=1}^{m} x_{ij}y_{ij} = \sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij}y_{ij}.$$

We can see that the two results are identical:

$$\mathrm{Tr}(\mathbf{X}^T\mathbf{Y}) = \sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij}y_{ij} = \mathrm{vec}(\mathbf{X})^T\mathrm{vec}(\mathbf{Y}).$$

Thus, we have shown:

$$\mathrm{Tr}(\mathbf{X}^T\mathbf{Y}) = \mathrm{vec}(\mathbf{X})^T\mathrm{vec}(\mathbf{Y}).$$

Now show that $\mathrm{Tr}(\mathbf{BA}) = \mathrm{Tr}(\mathbf{AB})$ using this property
In our lecture note it states that:

$$\mathrm{Tr}(\mathbf{BA}) = \mathrm{Tr}(\mathbf{AB})$$

for $\mathbf{A}, \mathbf{B}$ of appropriate size. Which means that $\mathbf{AB}$ should be square.
Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{B} \in \mathbb{R}^{n\times m}$, then $\mathbf{A}^T \in \mathbb{R}^{n\times m}$ and $\mathbf{B^T} \in \mathbb{R}^{m\times n}$

For $\mathbf{A}^T \in \mathbb{R}^{n\times m}$ and $\mathbf{B} \in \mathbb{R}^{n\times m}$, we know:

$$\mathrm{Tr}((\mathbf{A^T})^\mathbf{T}\mathbf{B}) = \mathrm{vec}(\mathbf{A}^T)^T\mathrm{vec}(\mathbf{B}).$$

$$\mathrm{Tr}(\mathbf{AB}) = \mathrm{vec}(\mathbf{A}^T)^T\mathrm{vec}(\mathbf{B}).$$

For $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{B}^T \in \mathbb{R}^{m\times n}$, we know:

$$\mathrm{Tr}((\mathbf{B^T})^\mathbf{T}\mathbf{A}) = \mathrm{vec}(\mathbf{B^T})^T\mathrm{vec}(\mathbf{A}).$$

$$\mathrm{Tr}(\mathbf{BA}) = \mathrm{vec}(\mathbf{B^T})^T\mathrm{vec}(\mathbf{A}).$$

As shown in the proof above for $\mathbf{X} \in \mathbb{R}^{m\times n}$ and $\mathbf{Y} \in \mathbb{R}^{m\times n}$:

$$\mathrm{vec}(\mathbf{X})^T\mathrm{vec}(\mathbf{Y}) = \sum_{j=1}^{n}\sum_{i=1}^{m} x_{ij}y_{ij} = \sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij}y_{ij}.$$

Using this to represent $\mathrm{vec}(\mathbf{A}^T)^T\mathrm{vec}(\mathbf{B})$ for $\mathbf{A}^T \in \mathbb{R}^{n\times m}$ and $\mathbf{B} \in \mathbb{R}^{n\times m}$:

$$\mathrm{vec}(\mathbf{A}^T)^T\mathrm{vec}(\mathbf{B}) = \sum_{i=1}^{n}\sum_{j=1}^{m} \mathbf{A}^T_{ij}\mathbf{B}_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ji}b_{ij}$$

Similarly for $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{B}^T \in \mathbb{R}^{m\times n}$:

$$\mathrm{vec}(\mathbf{B}^T)^T\mathrm{vec}(\mathbf{A}) = \sum_{i=1}^{m}\sum_{j=1}^{n} \mathbf{B}^T_{ij}\mathbf{A}_{ij} = \sum_{i=1}^{m}\sum_{j=1}^{n} b_{ji}a_{ij} = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}b_{ji}$$

Just writing i as j and j as i:

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

Thus :

$$\mathrm{Tr}(\mathbf{AB}) = \mathrm{Tr}(\mathbf{BA})$$

Q7. Consider a system of linear equations:

$$\mathbf{y} = \mathbf{A}\mathbf{x},$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$. Assume that $\mathrm{rank}(\mathbf{A}) = n$. Also assume that there is a solution $\mathbf{x}_0$ such that $\mathbf{y} = \mathbf{A}\mathbf{x}_0$ holds. Show that there is no other solution that can satisfy the equality—i.e., the solution to the above system is unique. (10%)

A7.
Proof by Contradiction:
Assume there is another solution $\mathbf{x}_1$ where $\mathbf{x_1} \neq \mathbf{x_0}$.
Suppose $\mathbf{x}_1$ is also a solution to $\mathbf{y} = \mathbf{A}\mathbf{x}$. Then:

$$\mathbf{y} = \mathbf{A}\mathbf{x}_1.$$

Since $\mathbf{x}_0$ is also a solution, we can write:

$$\mathbf{y} = \mathbf{A}\mathbf{x}_0.$$

Thus, for both solutions to satisfy the same equation:

$$\mathbf{A}\mathbf{x}_0 = \mathbf{A}\mathbf{x}_1.$$

By subtracting two equations:

$$\mathbf{A}\mathbf{x}_0 - \mathbf{A}\mathbf{x}_1 = 0 \implies \mathbf{A}(\mathbf{x}_0 - \mathbf{x}_1) = 0.$$

Let $\mathbf{h} = \mathbf{x}_0 - \mathbf{x}_1$. This gives:

$$\mathbf{A}\mathbf{h} = 0,$$

where $\mathbf{h}$ lies in the null space of $\mathbf{A}$.
Deriving dimension of $\mathcal{N}(\mathbf{A})$:
We have learned from class that

$$\dim\mathcal{R}(\mathbf{A}^T) + \dim\mathcal{N}(\mathbf{A}) = n, \text{ where } \mathbf{A} \in \mathbb{R}^{m\times n}$$

Then,

$$\dim\mathcal{N}(\mathbf{A}) = n - \dim\mathcal{R}(\mathbf{A}^T) = n - \mathrm{rank}(\mathbf{A}^T) = n - \mathrm{rank}(\mathbf{A})$$

Since in this problem $\mathrm{rank}(\mathbf{A}) = n$, we have:

$$\dim\mathcal{N}(\mathbf{A}) = n - n = 0.$$

This means the null space of $\mathbf{A}$ contains only the zero vector:

$$\mathbf{h} = 0 \implies \mathbf{x}_0 - \mathbf{x}_1 = 0.$$

Thus:

$$\mathbf{x}_0 = \mathbf{x}_1.$$

Contradicting our assumption that $\mathbf{x_0} \neq \mathbf{x_1}$
The solution $\mathbf{x}_0$ is unique, as there is no other $\mathbf{x}_1$ that satisfies $\mathbf{y} = \mathbf{A}\mathbf{x}$.

Q8. Given $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{B} \in \mathbb{R}^{n \times p}$. Assume that $\text{rank}(\mathbf{B}) = n$. Show that

$$\text{rank}(\mathbf{AB}) = \text{rank}(\mathbf{A}).$$

(10%)

A8.
$\text{rank}(\mathbf{AB}) = \text{rank}(\mathbf{A})$ holds if and only if $\text{rank}(\mathbf{AB}) \leq \text{rank}(\mathbf{A})$ and $\text{rank}(\mathbf{AB}) \geq \text{rank}(\mathbf{A})$ both hold.

1. Prove $\text{rank}(\mathbf{AB}) \leq \text{rank}(\mathbf{A})$
As learned in class:

$$\text{rank}(\mathbf{AB}) \leq \min(\text{rank}(\mathbf{A}), \text{rank}(\mathbf{B})).$$

Thus, $\text{rank}(\mathbf{AB})$ cannot exceed neither $\text{rank}(\mathbf{A})$ nor $\text{rank}(\mathbf{B})$:

$$\text{rank}(\mathbf{AB}) \leq \text{rank}(\mathbf{A})$$

2. Prove $\text{rank}(\mathbf{A}) \leq \text{rank}(\mathbf{AB})$

Let $\mathbf{x} \in \mathcal{N}(\mathbf{AB})$, then:

$$\mathbf{ABx} = 0 \implies \mathbf{Ax} \in \mathcal{N}(\mathbf{B}).$$

Since $\text{rank}\mathbf{B} = n$, we know that

$$\dim\mathcal{N}(\mathbf{B}) = n - \dim\mathcal{R}(\mathbf{B}^T) = n - \text{rank}(\mathbf{B}) = 0$$

Which means that $\mathcal{N}(\mathbf{B}) = \{0\}$, it follows that $\mathbf{Ax} = 0$.
Thus, any $\mathbf{x} \in \mathcal{N}(\mathbf{AB})$ must also satisfy:

$$\mathbf{x} \in \mathcal{N}(\mathbf{A}).$$

Now since $\mathbf{Ax} = 0$, any $\mathbf{x} \in \mathcal{N}(\mathbf{A})$ satisfies:

$$\mathbf{ABx} = \mathbf{A}(\mathbf{Bx}) = 0,$$

Thus:

$$\mathbf{x} \in \mathcal{N}(\mathbf{AB}).$$

Therefore, there is a one-to-one correspondence between $\mathcal{N}(\mathbf{AB})$ and $\mathcal{N}(\mathbf{A})$.

In other words:

$$\dim(\mathcal{N}(\mathbf{AB})) = \dim(\mathcal{N}(\mathbf{A})).$$

We know that:

$$\dim\mathcal{N}(\mathbf{A}) = n - \dim\mathcal{R}(\mathbf{A}^T) = n - \text{rank}(\mathbf{A})$$

and:

$$\dim(\mathcal{N}(\mathbf{AB})) = p - \text{rank}(\mathbf{AB}).$$

Since $\dim(\mathcal{N}(\mathbf{AB})) = \dim(\mathcal{N}(\mathbf{A}))$, we have:

$$p - \text{rank}(\mathbf{AB}) = n - \text{rank}(\mathbf{A}).$$

Rearranging:

$$\text{rank}(\mathbf{AB}) = \text{rank}(\mathbf{A}) + (p - n).$$

Since $\text{rank}(\mathbf{B}) = n$ meaning that $p \geq n$ (and $p - n \geq 0$), it follows that:

$$\text{rank}(\mathbf{AB}) \geq \text{rank}(\mathbf{A}).$$

Since

$$\text{rank}(\mathbf{AB}) \leq \text{rank}(\mathbf{A})$$
$$\text{rank}(\mathbf{AB}) \geq \text{rank}(\mathbf{A})$$

both holds, then:

$$\text{rank}(\mathbf{AB}) = \text{rank}(\mathbf{A})$$

Q9. Given $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{B} \in \mathbb{R}^{n \times p}$. Show that

$$\text{rank}(\mathbf{AB}) \geq \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B}) - n.$$

(10%)

A9.

- Let:

$$r_{\mathbf{A}} = \text{rank}(\mathbf{A})$$
$$r_{\mathbf{B}} = \text{rank}(\mathbf{B})$$
$$r_{\mathbf{AB}} = \text{rank}(\mathbf{AB})$$

  The rank of $\mathbf{B}$, $r_{\mathbf{B}}$, tells us that the range space of $\mathbf{B}$, denoted $\mathcal{R}(\mathbf{B})$, has dimension $r_{\mathbf{B}}$.

  This means there are $r_{\mathbf{B}}$ linearly independent columns in $\mathbf{B}$.

We can form a basis for $\mathcal{R}(\mathbf{B})$ using $r_{\mathbf{B}}$ vectors:

$$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{r_{\mathbf{B}}}\}$$

To span all of $\mathbb{R}^n$, we add $n - r_{\mathbf{B}}$ linearly independent vectors, $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{n-r_{\mathbf{B}}}\}$, which span the complementary subspace to $\mathcal{R}(\mathbf{B})$ in $\mathbb{R}^n$.

Thus, the full set of $\{\mathbf{v}_1, \ldots, \mathbf{v}_{r_{\mathbf{B}}}, \mathbf{w}_1, \ldots, \mathbf{w}_{n-r_{\mathbf{B}}}\}$ forms a basis for $\mathbb{R}^n$.

Let's organize these basis vectors into a matrix:

$$\mathbf{M} = [\mathbf{V} \mid \mathbf{W}],$$

where:

$\mathbf{V} \in \mathbb{R}^{n \times r_{\mathbf{B}}}$ contains the $r_{\mathbf{B}}$ basis vectors for $\mathcal{R}(\mathbf{B})$,

$\mathbf{W} \in \mathbb{R}^{n \times (n-r_{\mathbf{B}})}$ contains the $n - r_{\mathbf{B}}$ basis vectors for the complementary subspace.

Any vector $\mathbf{y} \in \mathbb{R}^n$ can now be written as:

$$\mathbf{y} = \mathbf{M}\boldsymbol{\alpha}, \boldsymbol{\alpha} \in \mathbb{R}^n$$

When we multiply $\mathbf{A}$ by $\mathbf{M}$, the matrix is split into two parts:

$$\mathbf{AM} = [\mathbf{AV} \mid \mathbf{AW}].$$

This separates the mapping of $\mathbf{A}$ on the subspace $\mathcal{R}(\mathbf{B})$ and the complementary subspace.

- Now considering relationship of rank($\mathbf{A}$) and rank($\mathbf{AM}$)

  Let $\mathbf{x} \in \mathcal{R}(\mathbf{A})$, then $\mathbf{Ay} = \mathbf{x}$ for $\mathbf{y} \in \mathbb{R}^n$

  As shown above, any $\mathbf{y} \in \mathbb{R}^n$ can be written as:

$$\mathbf{y} = \mathbf{M}\boldsymbol{\alpha}, \boldsymbol{\alpha} \in \mathbb{R}^n$$

  Then, substituting $\mathbf{y} = \mathbf{M}\boldsymbol{\alpha}$, we have:

$$\mathbf{Ay} = \mathbf{AM}\boldsymbol{\alpha} = \mathbf{x}$$

  Thus:

$$\mathbf{x} \in \mathcal{R}(\mathbf{AM})$$

  Showing that:

$$\mathcal{R}(\mathbf{A}) \subseteq \mathcal{R}(\mathbf{AM})$$

  Now for the opposite direction, let $\mathbf{x} \in \mathcal{R}(\mathbf{AM})$.

Then, there exists $\boldsymbol{\alpha} \in \mathbb{R}^n$ then:

$$AM\alpha = x$$

Let $\mathbf{z} = \mathbf{M}\boldsymbol{\alpha}$.

Since $\mathbf{z} \in \mathbb{R}^n$

$$\mathbf{Az} = \mathbf{x}$$

Thus,

$$\mathbf{x} \in \mathcal{R}(\mathbf{A})$$

Showing that:

$$\mathcal{R}(\mathbf{AM}) \subseteq \mathcal{R}(\mathbf{A})$$

Since

$$\mathcal{R}(\mathbf{AM}) \subseteq \mathcal{R}(\mathbf{A}), \mathcal{R}(\mathbf{A}) \subseteq \mathcal{R}(\mathbf{AM})$$

We can say that:

$$\mathcal{R}(\mathbf{AM}) = \mathcal{R}(\mathbf{A})$$

- Now considering the relationship between $\mathbf{AB}$ and $\mathbf{AV}$

  Let $\mathbf{x} \in \mathcal{R}(\mathbf{AV})$. Then $\mathbf{x} = \mathbf{AVy}$ for some $\mathbf{y} \in \mathbb{R}^{r_B}$.

  Since the columns of $\mathbf{V}$ span the same subspace as the columns of $\mathbf{B}$, we can express $\mathbf{Vy}$ as a linear combination of the columns of $\mathbf{B}$:

  $$\mathbf{Vy} = \mathbf{B}\boldsymbol{\alpha},$$

  for some $\boldsymbol{\alpha} \in \mathbb{R}^p$.

  Substituting this into $\mathbf{x}$, we have:

  $$\mathbf{x} = \mathbf{AVy} = \mathbf{A}(\mathbf{B}\boldsymbol{\alpha}) = \mathbf{AB}\boldsymbol{\alpha}.$$

  Therefore, $\mathbf{x} \in \mathcal{R}(\mathbf{AB})$:

  $$\mathcal{R}(\mathbf{AV}) \subseteq \mathcal{R}(\mathbf{AB}).$$

  Now, let $\mathbf{x} \in \mathcal{R}(\mathbf{AB})$. Then $\mathbf{x} = \mathbf{ABy}$ for some $\mathbf{y} \in \mathbb{R}^p$.

  Since the columns of $\mathbf{B}$ lie in $\mathcal{R}(\mathbf{B})$, which is spanned by the columns of $\mathbf{V}$, we can express $\mathbf{By}$ as:

  $$\mathbf{By} = \mathbf{V}\boldsymbol{\theta},$$

  for some $\boldsymbol{\theta} \in \mathbb{R}^{r_B}$.

  Substituting this into $\mathbf{x}$, we have:

  $$\mathbf{x} = \mathbf{ABy} = \mathbf{A}(\mathbf{By}) = \mathbf{A}(\mathbf{V}\boldsymbol{\theta}) = \mathbf{AV}\boldsymbol{\theta}.$$

Thus, $\mathbf{x} \in \mathcal{R}(\mathbf{AV})$:

$$\mathcal{R}(\mathbf{AB}) \subseteq \mathcal{R}(\mathbf{AV}).$$

Since both directions are proven, we have:

$$\mathcal{R}(\mathbf{AB}) = \mathcal{R}(\mathbf{AV}).$$

Thus:

$$\operatorname{rank}(\mathbf{AB}) = \operatorname{rank}(\mathbf{AV}).$$

- Now consider the full mapping of $\mathbf{A}$:

$$\mathbf{AM} = [\mathbf{AV} \mid \mathbf{AW}].$$

The rank of $\mathbf{AM}$ is at most the sum of the ranks of these two components. Equals to when all columns of $\mathbf{AV}$:

$$\operatorname{rank}(\mathbf{AM}) \leq \operatorname{rank}(\mathbf{AV}) + \operatorname{rank}(\mathbf{AW}).$$

Since we have proven that, $\mathcal{R}(\mathbf{AM}) = \mathcal{R}(\mathbf{A})$:

$$\operatorname{rank}(\mathbf{AM}) = \operatorname{rank}(\mathbf{A})$$

Substituting in $r_{\mathbf{A}}$ for rank($\mathbf{A}$):

$$r_{\mathbf{A}} \leq \operatorname{rank}(\mathbf{AV}) + \operatorname{rank}(\mathbf{AW}).$$

The matrix $\mathbf{AW}$ corresponds to the mapping of $\mathbf{A}$ on the complementary subspace to $\mathcal{R}(\mathbf{B})$.

Since $\mathbf{W}$ has $n - r_{\mathbf{B}}$ columns, the rank of $\mathbf{AW}$ is at most:

$$\operatorname{rank}(\mathbf{AW}) \leq n - r_{\mathbf{B}}.$$

Substituting back:

$$r_{\mathbf{A}} \leq \operatorname{rank}(\mathbf{AV}) + (n - r_{\mathbf{B}}).$$

Since $\operatorname{rank}(\mathbf{AV}) = \operatorname{rank}(\mathbf{AB})$, we have:

$$r_{\mathbf{A}} \leq \operatorname{rank}(\mathbf{AB}) + (n - r_{\mathbf{B}}).$$

Rearranging:

$$r_{\mathbf{A}} - (n - r_{\mathbf{B}}) \leq \operatorname{rank}(\mathbf{AB}).$$

Thus:

$$r_{\mathbf{A}} + r_{\mathbf{B}} - n \leq \text{rank}(\mathbf{AB}).$$

Finally:

$$\text{rank}(\mathbf{AB}) \geq \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B}) - n.$$

Q10. Show the Hölder's inequality in the lecture notes. (10%)

A10.

Hölder's inequality:

$$|\mathbf{x}^T\mathbf{y}| \leq \|\mathbf{x}\|_p\|\mathbf{y}\|_q, \text{ where, } 1/p + 1/q = 1, p \geq 1$$

Which can be reformulated as:

$$\sum_{i=1}^{n} |x_i y_i| \leq \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} \left(\sum_{i=1}^{n} |y_i|^q\right)^{1/q}$$

Let $S_1 = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}, S_2 = \left(\sum_{i=1}^{n} |y_i|^q\right)^{1/q}$

We need to prove:

$$\sum_{i=1}^{n} |x_i y_i| \leq S_1 S_2$$

To make the inequality simple, let $a_i = \frac{|x_i|}{S_1}$ and $b_i = \frac{|y_i|}{S_2}$,
Then:

$$\sum_{i=1}^{n} a_i^p = \sum_{i=1}^{n} \left(\frac{|x_i|}{S_1}\right)^p = \frac{\sum_{i=1}^{n} |x_i|^p}{\left(\sum_{i=1}^{n} |x_i|^p\right)} = 1$$

Similarly:

$$\sum_{i=1}^{n} b_i^q = 1.$$

Then our goal is to show:

$$\sum_{i=1}^{n} a_i b_i \leq 1.$$

We now apply the young's inequality which states that for $x, y \geq 0$, $p > 1$ and $1/p + 1/q = 1$:

$$xy \leq \frac{x^p}{p} + \frac{y^q}{q}$$

Applying this to a, b:

$$a_i b_i \leq \frac{a_i^p}{p} + \frac{b_i^q}{q},$$

for $\frac{1}{p} + \frac{1}{q} = 1$.
Summing this over all i, we have:

$$\sum_{i=1}^{n} a_i b_i \leq \sum_{i=1}^{n} \left(\frac{a_i^p}{p} + \frac{b_i^q}{q} \right).$$

Since $\sum_{i=1}^{n} a_i^p = 1$ and $\sum_{i=1}^{n} b_i^q = 1$:

$$\sum_{i=1}^{n} a_i b_i \leq \frac{1}{p} + \frac{1}{q}.$$

Finally since $\frac{1}{p} + \frac{1}{q} = 1$:

$$\sum_{i=1}^{n} a_i b_i \leq 1.$$

Substitute in original variables x_i and y_i:

$$\sum_{i=1}^{n} |x_i y_i| = S_1 S_2 \cdot \sum_{i=1}^{n} a_i b_i.$$

Since $\sum_{i=1}^{n} a_i b_i \leq 1$, it follows that:

$$\sum_{i=1}^{n} |x_i y_i| \leq S_1 S_2.$$

Thus:

$$\sum_{i=1}^{n} |x_i y_i| \leq \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} \left(\sum_{i=1}^{n} |y_i|^q \right)^{1/q}$$