Week 3

Lecture 1

Four Fundamental Subspaces

1. Overview

In this lecture, we study the **four fundamental subspaces** associated with a matrix A.
Given a matrix $A \in \mathbb{R}^{m \times n}$, we define:

1. Column Space $\mathcal{C}(A)$
2. Null Space $\mathcal{N}(A)$
3. Row Space $\mathcal{R}(A)$
4. Left Null Space $\mathcal{N}(A^\top)$
   We analyze their definitions, geometric meaning, computation, and dimensional relationships.

---

2. Column Space

2.1 Definition

Let

$$A = [u_1 \; u_2 \; \dots \; u_n]$$

where u_i are the columns of A.
We define the **column space** as

$$\mathcal{C}(A) = \operatorname{span}(u_1, \dots, u_n)$$

That is,

$$\mathcal{C}(A) = \left\{ \sum_{i=1}^{n} \alpha_i u_i \;\middle|\; \alpha_i \in \mathbb{R} \right\}$$

Thus, $\mathcal{C}(A)$ consists of all linear combinations of the columns of A.

2.2 Connection to Solving $Ax = b$

We observe that:

$$Ax = b$$

is solvable **if and only if**

$$b \in \mathcal{C}(A)$$

Therefore, the column space characterizes exactly which right-hand sides b admit solutions.

2.3 Example

Consider

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{bmatrix}$$

We observe:

$$\mathrm{col}_3 = \mathrm{col}_1 + \mathrm{col}_2$$

Hence,

$$\mathcal{C}(A) = \mathrm{span}(\mathrm{col}_1, \mathrm{col}_2)$$

Thus,

$$\dim(\mathcal{C}(A)) = 2$$

Even though A has 3 columns, its column space is 2-dimensional inside $\mathbb{R}^4$.

---

3. Null Space

3.1 Definition

We define the **null space** as

$$\mathcal{N}(A) = \{x \in \mathbb{R}^n \mid Ax = 0\}$$

3.2 Subspace Property

We verify subspace conditions:
If $x_1, x_2 \in \mathcal{N}(A)$:

$$Ax_1 = 0, \quad Ax_2 = 0$$

Then

$$A(x_1 + x_2) = Ax_1 + Ax_2 = 0$$

Also, for scalar α:

$$A(\alpha x) = \alpha Ax = 0$$

Hence, $\mathcal{N}(A)$ is a subspace of $\mathbb{R}^n$.

3.3 Example (Same Matrix)

Since

$$\mathrm{col}_1 + \mathrm{col}_2 - \mathrm{col}_3 = 0$$

we obtain

$$x = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \in \mathcal{N}(A)$$

Thus, $\mathcal{N}(A)$ is a line in $\mathbb{R}^3$.

3.4 Invertible Case

If A is invertible:

$$\mathcal{N}(A) = \{0\}$$

and

$$\mathcal{C}(A) = \mathbb{R}^n$$

In this case, $Ax = b$ has a unique solution.
If A is not invertible:

$$x = x_p + x_n$$

where

- $Ax_p = b$
- $Ax_n = 0$

3.5 Gaussian Elimination and Null Space

We compute null space by solving:

$$Ax = 0$$

After row reduction:

- Pivot columns correspond to basic variables
- Non-pivot columns correspond to free variables
  Each free variable generates one basis vector of the null space.

---

4. Rank and Nullity

We define:

- **Rank** = number of pivot columns
- **Nullity** = number of free variables
  We have:

  $$\operatorname{rank}(A) = \dim(\mathcal{C}(A))$$

  $$\operatorname{nullity}(A) = \dim(\mathcal{N}(A))$$

  If A has n columns:

  $$\operatorname{rank}(A) + \operatorname{nullity}(A) = n$$

  This is the **Rank–Nullity Theorem**.

---

5. Row Space

5.1 Definition

We define the row space as:

$$\mathcal{R}(A) = \mathcal{C}(A^\top)$$

Equivalently:

$$\mathcal{R}(A) = \text{span of rows of } A$$

5.2 Important Fact

$$\text{column rank} = \text{row rank}$$

Thus:

$$\dim(\mathcal{R}(A)) = \dim(\mathcal{C}(A)) = r$$

---

6. Left Null Space

6.1 Definition

We define:

$$\mathcal{N}(A^\top) = \{y \in \mathbb{R}^m \mid A^\top y = 0\}$$

Equivalently:

$$y^\top A = 0$$

Thus, the left null space consists of all linear combinations of rows that produce the zero vector.

6.2 Dimension Relationship

For $A \in \mathbb{R}^{m \times n}$:
We know:

$$\dim(\mathcal{C}(A)) + \dim(\mathcal{N}(A)) = n$$

Applying the same to $A^\top$:

$$\dim(\mathcal{C}(A^\top)) + \dim(\mathcal{N}(A^\top)) = m$$

Since:

$$\dim(\mathcal{C}(A^\top)) = r$$

we obtain:

$$\dim(\mathcal{N}(A^\top)) = m - r$$

7. Summary of Dimensions

For $A \in \mathbb{R}^{m \times n}$ with rank r:

Subspace	Dimension
Column space $\mathcal{C}(A)$	r
Null space $\mathcal{N}(A)$	$n - r$
Row space $\mathcal{R}(A)$	r
Left null space $\mathcal{N}(A^\top)$	$m - r$

---

8. Final Perspective

We conclude:

- Column space controls solvability of $Ax = b$
- Null space controls non-uniqueness of solutions
- Row space equals column space of $A^\top$
- Left null space captures row dependencies
- Rank determines dimensions of all four subspaces

These four spaces completely describe the structure of a matrix and form the backbone of linear algebra used in machine learning.

```
*************************************************************************************
```

Lecture 2

Orthogonality and Orthogonal Subspaces

1. Motivation

In this lecture, we build the geometric foundation required to understand orthogonal projections and subsequently least squares problems.

We study:

1. Length of a vector
2. Orthogonal vectors
3. Orthonormal vectors
4. Orthogonal subspaces
5. Orthogonality among the four fundamental subspaces

---

2. Length of a Vector

Let

$$x = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

Then

$$\|x\|^2 = 1^2 + 2^2 = 5$$

In general, for

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$$

the Euclidean norm is

$$\|x\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

Equivalently,

$$\|x\|^2 = x^\top x$$

---

3. Orthogonal Vectors

Two vectors $x, y \in \mathbb{R}^n$ are **orthogonal** if

$$x^\top y = 0$$

If

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

then

$$x^\top y = \sum_{i=1}^{n} x_i y_i$$

3.1 Connection with Pythagoras

Orthogonality is equivalent to

$$\|x\|^2 + \|y\|^2 = \|x+y\|^2$$

Expanding,

$$\|x+y\|^2 = (x+y)^\top (x+y) = x^\top x + y^\top y + x^\top y + y^\top x$$

Since $y^\top x = x^\top y$, we obtain

$$\|x+y\|^2 = x^\top x + y^\top y + 2x^\top y$$

Thus orthogonality requires

$$x^\top y = 0$$

3.2 Remarks

1. The zero vector is orthogonal to every vector:

$$0^\top x = 0$$

2. A mutually orthogonal nonzero set of vectors is linearly independent.
   If

$$c_1 v_1 + \cdots + c_k v_k = 0$$

   Taking inner product with v_1,

$$v_1^\top (c_1 v_1 + \cdots + c_k v_k) = c_1 v_1^\top v_1$$

   Since $\|v_1\|^2 \neq 0$, we conclude $c_1 = 0$.
   Repeating yields $c_i = 0$ for all i.

---

4. Orthonormal Vectors

Vectors u and v are **orthonormal** if

1. $u^\top v = 0$
2. $\|u\| = \|v\| = 1$

Examples in $\mathbb{R}^2$:

$$\begin{bmatrix}1\\0\end{bmatrix}, \quad \begin{bmatrix}0\\1\end{bmatrix}$$

Another example:

$$\begin{bmatrix}\cos\theta\\ \sin\theta\end{bmatrix}, \quad \begin{bmatrix}-\sin\theta\\ \cos\theta\end{bmatrix}$$

---

5. Orthogonal Subspaces

Let $U, V \subseteq \mathbb{R}^n$ be subspaces.
We say U and V are orthogonal if

$$x^\top y = 0 \quad \text{for all } x \in U,\ y \in V$$

---

6. Orthogonality and the Four Fundamental Subspaces

Let $A \in \mathbb{R}^{m\times n}$.

6.1 Row Space and Null Space

Claim:

$$\mathcal{R}(A) \perp \mathcal{N}(A)$$

If $x \in \mathcal{N}(A)$, then

$$Ax = 0$$

Writing rows explicitly,

$$A = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{bmatrix}$$

Then

$$Ax = \begin{bmatrix} r_1^\top x \\ r_2^\top x \\ \vdots \\ r_m^\top x \end{bmatrix} = 0$$

Thus

$$r_i^\top x = 0 \quad \text{for all } i$$

So every row is orthogonal to x, and therefore every linear combination of rows is orthogonal to x. Hence,

$$\mathcal{R}(A) \perp \mathcal{N}(A)$$

Equivalently,

$$\mathcal{C}(A^\top) \perp \mathcal{N}(A)$$

6.2 Column Space and Left Null Space

Similarly,

$$\mathcal{C}(A) \perp \mathcal{N}(A^\top)$$

---

7. Dimension Relationships

Let $\operatorname{rank}(A) = r$.
Then:

$$\dim \mathcal{C}(A) = r$$

$$\dim \mathcal{N}(A) = n - r$$

$$\dim \mathcal{C}(A^\top) = r$$

$$\dim \mathcal{N}(A^\top) = m - r$$

These relationships reflect orthogonal decompositions of $\mathbb{R}^n$ and $\mathbb{R}^m$.

8. Summary

In this lecture, we established:

- The definition of vector length
- Orthogonality via inner product
- Orthonormal vectors
- Orthogonal subspaces
- $\mathcal{R}(A) \perp \mathcal{N}(A)$
- $\mathcal{C}(A) \perp \mathcal{N}(A^\top)$

These results form the geometric foundation for projections and least squares.

```
********************************************************************************
```

Lecture 3

Projections and Motivation from Inconsistent Systems

1. Motivation for Projections

We begin with a linear system:

$$Ax = b$$

If

$$b \notin \mathcal{C}(A)$$

then the system is **inconsistent**, meaning there exists no vector x such that $Ax = b$.
In such a situation, we seek the **best approximation** to b within the column space of A.
The geometric solution is:

> We project b onto $\mathcal{C}(A)$.

This projected vector will be the closest vector in the column space to b.

---

2. Projection Onto a Line

We first study the simplest case: projection onto a line.

Let:

$$a \in \mathbb{R}^n$$

We want to project b onto the line spanned by a.
Let the projection be:

$$p = \hat{x}a$$

Define the error vector:

$$e = b - p = b - \hat{x}a$$

For orthogonal projection, we require:

$$e \perp a$$

Thus:

$$a^T(b - \hat{x}a) = 0$$

Expanding:

$$a^T b - \hat{x}a^T a = 0$$

Solving for $\hat{x}$:

$$\hat{x} = \frac{a^T b}{a^T a}$$

Therefore, the projection is:

$$p = \frac{a^T b}{a^T a} a$$

---

3. Projection Matrix Form

We rewrite:

$$p = \left(\frac{a^T b}{a^T a}\right) a$$

Rearranging:

$$p = \left(\frac{a a^T}{a^T a}\right) b$$

Define the **projection matrix**:

$$P = \frac{aa^T}{a^T a}$$

Then the projection of any vector b onto the line spanned by a is:

$$p = Pb$$

---

4. Properties of the Projection Matrix

4.1 Symmetry

$$P^T = P$$

Projection matrices are symmetric.

4.2 Idempotence

$$P^2 = P$$

Once a vector is projected, projecting again does not change it:

$$P(Pb) = Pb$$

4.3 Column Space

$$\mathcal{C}(P) = \text{span}(a)$$

4.4 Null Space

$$\mathcal{N}(P) = \{x \in \mathbb{R}^n \mid a^T x = 0\}$$

The null space is the subspace orthogonal to a.

4.5 Rank

Since projection is onto a 1-dimensional line:

$$\text{rank}(P) = 1$$

---

5. Invariance Under Scaling

If we replace a by αa:

$$P = \frac{(\alpha a)(\alpha a)^T}{(\alpha a)^T(\alpha a)}$$

The scalar cancels.
Therefore:

> The projection matrix depends only on the line, not the specific vector chosen.

---

6. Cauchy–Schwarz Inequality via Projection

We know:

$$||e||^2 = ||b - p||^2 \geq 0$$

Expanding gives:

$$(b^T b)(a^T a) - (a^T b)^2 \geq 0$$

Thus:

$$|a^T b| \leq ||a|| \, ||b||$$

This is the **Cauchy–Schwarz inequality**.

---

7. Connection to Least Squares

When:

$$Ax = b$$

is inconsistent, meaning:

$$b \notin \mathcal{C}(A),$$

we compute the orthogonal projection of b onto $\mathcal{C}(A)$.
This geometric idea forms the foundation of:

> Least squares solutions.

```
**********************************************************************************
```

Lecture 4

Least Squares and Projection onto a Subspace

1. Revisiting Least Squares

We begin with an overdetermined system:

$$\begin{aligned} 2x &= b_1 \\ 3x &= b_2 \\ 4x &= b_3 \end{aligned}$$

We observe that this system is solvable **if and only if** the vector

$$b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

lies on the line spanned by

$$a = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$$

If b does not lie in $\text{span}(a)$, then the system is inconsistent.

---

2. Least Squares Formulation

Since exact solvability may fail, we instead minimize the squared error:

$$E(x) = (2x - b_1)^2 + (3x - b_2)^2 + (4x - b_3)^2$$

We compute the derivative:

$$\frac{dE}{dx} = 2(2x - b_1) + 3(3x - b_2) + 4(4x - b_3)$$

Setting derivative equal to zero:

$$2(2x - b_1) + 3(3x - b_2) + 4(4x - b_3) = 0$$

Solving, we obtain:

$$\hat{x} = \frac{2b_1 + 3b_2 + 4b_3}{2^2 + 3^2 + 4^2}$$

We observe that this equals:

$$\hat{x} = \frac{a^T b}{a^T a}$$

This is exactly the projection coefficient of b onto a.
Thus we conclude:

> Minimizing squared error is equivalent to performing an orthogonal projection.

---

3. Projection onto a Subspace

We now generalize.
Let A be an $m \times n$ matrix.
We want to project b onto the column space of A.
We denote the projection by:

$$p = A\hat{x}$$

The error vector is:

$$e = b - p = b - A\hat{x}$$

By definition of orthogonal projection:

$$e \perp \mathrm{Col}(A)$$

Since e is orthogonal to every column of A, we must have:

$$A^T e = 0$$

Substituting:

$$A^T(b - A\hat{x}) = 0$$

This gives the **normal equations**:

$$A^T A\hat{x} = A^T b$$

This is the key equation.
Even if $Ax = b$ is inconsistent, the normal equations always have a solution.

---

4. Projection Matrix

If the columns of A are linearly independent, then $A^T A$ is invertible.
We solve:

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

Thus the projection is:

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

We define the projection matrix:

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

Then:

$$p = Pb$$

---

5. Properties of Projection Matrix

We verify:

1. Symmetry

$$P^T = P$$

Thus the projection matrix is symmetric.

---

2. Idempotence

$$P^2 = P$$

This means projecting twice does nothing new.

---

3. Characterization

If a matrix satisfies:

$$P^T = P \quad \text{and} \quad P^2 = P$$

then P is a projection matrix.

---

6. Special Cases

Case 1: $b \in \mathrm{Col}(A)$

If $b = Ax$ for some x, then:

$$Pb = b$$

Projection does not change vectors already in the subspace.

Case 2: $b \in \mathrm{Null}(A^T)$

If:

$$A^T b = 0$$

then:

$$Pb = 0$$

Case 3: A is square and invertible

Then $\mathrm{Col}(A) = \mathbb{R}^n$.
Hence:

$$Pb = b$$

Projection becomes the identity transformation.

---

7. Final Takeaway

Minimizing:

$$||Ax - b||^2$$

is equivalent to solving:

$$A^T A\hat{x} = A^T b$$

which geometrically means:

> We orthogonally project b onto the column space of A.

This is the geometric foundation of least squares.

```
********************************************************************************
```

Lecture 5

Example of Least Squares (Ordinary Least Squares)

1. Problem Setup — Curve Fitting

We consider a simple linear regression problem.
Suppose we are given data points:

$$(x_1, b_1), (x_2, b_2), \dots, (x_m, b_m)$$

We want to fit a straight line of the form:

$$b \approx \theta_1 x + \theta_2$$

This is a linear model with slope θ_1 and intercept θ_2.

---

2. Matrix Formulation

We rewrite the system as:

$$A\theta = b$$

where

$$A = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{bmatrix}, \quad \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

In general, this system may be inconsistent.

---

3. Least Squares Objective

We minimize the squared error:

$$\min_{\theta} \|b - A\theta\|^2$$

Explicitly,

$$\sum_{i=1}^{m}(b_i - \theta_1 x_i - \theta_2)^2$$

The least squares solution satisfies the normal equations:

$$A^T A\hat{\theta} = A^T b$$

---

4. Concrete Example

Consider:

$$A = \begin{bmatrix} -1 & 1 \\ 1 & 1 \\ 2 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}$$

We first check consistency by Gaussian elimination.
After row reduction, the augmented system produces a contradiction, so the system is inconsistent.
Thus, we solve the normal equations.

---

5. Compute Normal Equations

Compute:

$$A^T A = \begin{bmatrix} 6 & 2 \\ 2 & 3 \end{bmatrix}$$

Compute:

$$A^T b = \begin{bmatrix} 6 \\ 5 \end{bmatrix}$$

Thus we solve:

$$\begin{bmatrix} 6 & 2 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 5 \end{bmatrix}$$

This gives:

$$\hat{\theta}_1 = \frac{4}{7}, \quad \hat{\theta}_2 = \frac{9}{7}$$

---

6. Final Least Squares Line

The best-fit line is:

$$b = \frac{4}{7}x + \frac{9}{7}$$

---

7. Projection Interpretation

The projection is:

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

The projected vector is:

$$p = A\hat{\theta}$$

The residual is:

$$e = b - p$$

For this example:

$$e = \begin{bmatrix} \frac{2}{7} \\ -\frac{6}{7} \\ \frac{4}{7} \end{bmatrix}$$

We verify:

$$A^Te = 0$$

Thus the residual is orthogonal to the column space of A.

---

8. Key Insight

Minimizing

$$||A\theta - b||^2$$

is equivalent to solving

$$A^TA\hat{\theta} = A^Tb$$

which geometrically corresponds to projecting b onto the column space of A.

---

9. Conceptual Summary

- If $b \in \mathrm{Col}(A)$, the system is exactly solvable.
- If not, we project b onto $\mathrm{Col}(A)$.
- The error vector e lies in $\mathrm{Null}(A^T)$.
- Least squares is orthogonal projection.

```
********************************************************************************
```
