Lecture 3: Fundamental Theorem of Linear Algebra

1. Orthogonal Decomposition of Space

Any finite-dimensional inner product space decomposes into two orthogonal pieces. **Why care?** This decomposition underpins least squares, projections, and the structure of linear systems. When $Ax = b$ has no exact solution, orthogonal decomposition reveals the best approximation.

We first make precise what "orthogonal pieces" means.

Definition 1 (Orthogonal Complement)**.** For a subspace S of an inner product space X, the **orthogonal complement** is

$$S^\perp = \{x \in X : \langle x, s\rangle = 0 \text{ for all } s \in S\}$$

Lemma 1 (Basis Test for Orthogonality)**.** *Let $S \subseteq X$ have basis $\{s_1, \ldots, s_k\}$. Then $x \in S^\perp$ iff $\langle x, s_i\rangle = 0$ for $i = 1, \ldots, k$.*

Proof. ($\Rightarrow$) If $x \in S^\perp$, then $\langle x, s\rangle = 0$ for all $s \in S$; in particular, $\langle x, s_i\rangle = 0$ for each basis vector.

($\Leftarrow$) Suppose $\langle x, s_i\rangle = 0$ for all i. Any $s \in S$ can be written $s = \sum_{i=1}^k \alpha_i s_i$. By linearity:

$$\langle x, s\rangle = \sum_{i=1}^k \alpha_i \langle x, s_i\rangle = 0$$

Thus $x \in S^\perp$. □

Practical Implication

To verify $x \in S^\perp$, check orthogonality to a *basis* of S, not every vector. Infinitely many conditions reduce to finitely many.

Definition 2 (Direct Sum)**.** The **direct sum** $X = S \oplus S^\perp$ means:

(i) **Existence:** Every $x \in X$ can be written as $x = s + s^\perp$ with $s \in S$, $s^\perp \in S^\perp$.

(ii) **Uniqueness:** The decomposition is unique.

Equivalently: $S + S^\perp = X$ and $S \cap S^\perp = \{0\}$.

Why Uniqueness Follows from $S \cap S^\perp = \{0\}$

If $s_1 + s_1^\perp = s_2 + s_2^\perp$, then $s_1 - s_2 = s_2^\perp - s_1^\perp$. The LHS is in S; the RHS is in $S^\perp$. Thus both sides lie in $S \cap S^\perp = \{0\}$, so $s_1 = s_2$.

Theorem 1 (Orthogonal Decomposition)**.** *For any subspace S of a finite-dimensional inner product space X:*

$$X = S \oplus S^\perp$$

Moreover, $\dim(S) + \dim(S^\perp) = \dim(X)$.

Proof. **Step 1: $S \cap S^\perp = \{0\}$.** If $v \in S \cap S^\perp$, then $\langle v, v\rangle = 0$. By positive definiteness, $v = 0$.

Step 2: Existence. Let $\{s_1, \ldots, s_k\}$ be an orthonormal basis for S (obtained via Gram–Schmidt). For any $x \in X$, define:

$$s = \sum_{i=1}^k \langle x, s_i\rangle s_i \in S, \qquad s^\perp = x - s$$

Then $\langle s^\perp, s_j\rangle = \langle x, s_j\rangle - \langle x, s_j\rangle = 0$ using orthonormality $\langle s_i, s_j\rangle = \delta_{ij}$. By Lemma 1, $s^\perp \in S^\perp$.

Step 3: Dimension. From Step 2, $X = S + S^\perp$. Combined with Step 1 ($S \cap S^\perp = \{0\}$), we have $X = S \oplus S^\perp$. Thus:

$$\dim(X) = \dim(S) + \dim(S^\perp)$$

□

Geometric Intuition

In $\mathbb{R}^3$: if S is a plane through the origin, then $S^\perp$ is the perpendicular line. Every vector splits into a plane component plus a line component—orthogonal pieces summing to the original.

Example 1 (Orthogonal Decomposition in $\mathbb{R}^2$)**.** Let $S = \text{span}\{(1, 0)\}$ be the x-axis in $\mathbb{R}^2$. Then:

$$S^\perp = \text{span}\{(0, 1)\} \quad \text{(the } y\text{-axis)}$$

Any vector $(a, b) \in \mathbb{R}^2$ decomposes as:

$$\underbrace{(a, b)}_{\in \mathbb{R}^2} = \underbrace{(a, 0)}_{\in S} + \underbrace{(0, b)}_{\in S^\perp}$$

Notice that $(a, 0) \cdot (0, b) = 0$, confirming orthogonality.

We now turn to how matrices decompose $\mathbb{R}^n$ and $\mathbb{R}^m$ into orthogonal pieces.

2. Four Fundamental Subspaces

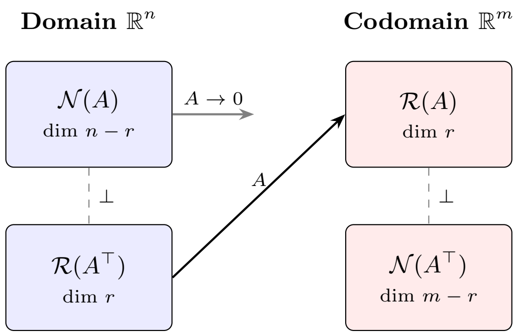
A matrix $A \in \mathbb{R}^{m\times n}$ has four fundamental subspaces. **Why four?** A maps $\mathbb{R}^n \to \mathbb{R}^m$, creating two complementary subspaces in each space.

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

Here $r = \text{rank}(A)$.

Dimension Check

In $\mathbb{R}^n$: $(n-r)+r=n$. In $\mathbb{R}^m$: $r+(m-r)=m$. (Rank-Nullity Theorem.)



2.1 Definitions and Significance

Null space (kernel): $\mathcal{N}(A) = \{x \in \mathbb{R}^n : Ax = 0\}$. Captures information loss: if $x_1 - x_2 \in \mathcal{N}(A)$, then $Ax_1 = Ax_2$.

Column space (range): $\mathcal{R}(A) = \{Ax : x \in \mathbb{R}^n\}$. All possible outputs. $Ax = b$ is solvable iff $b \in \mathcal{R}(A)$.

Row space: $\mathcal{R}(A^\top) = \{A^\top y : y \in \mathbb{R}^m\}$. Orthogonal to the null space; represents directions A can detect.

Left null space: $\mathcal{N}(A^\top) = \{y \in \mathbb{R}^m : A^\top y = 0\}$. When $Ax = b$ has no solution, the component of b in $\mathcal{N}(A^\top)$ measures the residual.

2.2 Example: A 2×3 Matrix

Example 2. Consider $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{pmatrix} \in \mathbb{R}^{2\times 3}$. Row 2 equals $2\times$ row 1, so $\text{rank}(A) = 1$.

Step 1: Null space $\mathcal{N}(A) \subseteq \mathbb{R}^3$. Solve $Ax = 0$: both rows give $x_1 + 2x_2 + 3x_3 = 0$, so $x_1 = -2x_2 - 3x_3$:

$$\mathcal{N}(A) = \text{span}\left\{ \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -3 \\ 0 \\ 1 \end{pmatrix} \right\} \quad (\dim = 2 = n - r)$$

Step 2: Row space $\mathcal{R}(A^\top) \subseteq \mathbb{R}^3$. Spanned by rows of A:

$$\mathcal{R}(A^\top) = \text{span}\left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \right\} \quad (\dim = 1 = r)$$

Step 3: Verify orthogonality. $(1,2,3)\cdot(-2,1,0) = 0$ ✓ and $(1,2,3)\cdot(-3,0,1) = 0$ ✓

Step 4: Column space $\mathcal{R}(A) \subseteq \mathbb{R}^2$. All columns are multiples of $(1,2)^\top$:

$$\mathcal{R}(A) = \text{span}\left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\} \quad (\dim = 1 = r)$$

Step 5: Left null space $\mathcal{N}(A^\top) \subseteq \mathbb{R}^2$. Solve $A^\top y = 0$: all rows give $y_1 + 2y_2 = 0$:

$$\mathcal{N}(A^\top) = \text{span}\left\{ \begin{pmatrix} -2 \\ 1 \end{pmatrix} \right\} \quad (\dim = 1 = m - r)$$

Step 6: Verify. $(1,2)\cdot(-2,1) = 0$ ✓ Column space $\perp$ left null space.

Summary

For $A \in \mathbb{R}^{2\times 3}$ with rank 1: In $\mathbb{R}^3$, $\mathcal{N}(A)$ (dim 2) $\perp$ $\mathcal{R}(A^\top)$ (dim 1). In $\mathbb{R}^2$, $\mathcal{N}(A^\top)$ (dim 1) $\perp$ $\mathcal{R}(A)$ (dim 1). Dimensions always sum to ambient space dimension.

We now prove *why* these orthogonality relationships hold.

3. Fundamental Orthogonal Decompositions

The four subspaces form orthogonal pairs: a consequence of the adjoint relationship between A and $A^\top$.

Lemma 2 (Adjoint Identity)**.** *For $A \in \mathbb{R}^{m\times n}$, $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$: $\langle A^\top y, x\rangle = \langle y, Ax\rangle$.*

Proof. $\langle A^\top y, x\rangle = (A^\top y)^\top x = y^\top A x = \langle y, Ax\rangle$. □

Remark: This identity underlies both orthogonality relations. A and $A^\top$ are adjoint operators.

Fundamental Theorem of Linear Algebra:

$$\mathcal{N}(A) \oplus \mathcal{R}(A^\top) = \mathbb{R}^n$$

$$\mathcal{N}(A^\top) \oplus \mathcal{R}(A) = \mathbb{R}^m$$

Every $x \in \mathbb{R}^n$ uniquely decomposes into null space and row space components; every $y \in \mathbb{R}^m$ splits into column space and left null space components. Equivalently:

$$\mathcal{N}(A) = \mathcal{R}(A^\top)^\perp \quad \text{and} \quad \mathcal{N}(A^\top) = \mathcal{R}(A)^\perp$$

Applications

Least Squares: When $b \notin \mathcal{R}(A)$, decompose $b = b_\parallel + b_\perp$ with $b_\parallel \in \mathcal{R}(A)$ and $b_\perp \in \mathcal{N}(A^\top)$. The least squares solution solves $Ax = b_\parallel$.
Rank-Nullity: The decomposition implies $\dim(\mathcal{N}(A)) + \text{rank}(A) = n$.

4. Proof: $\mathcal{N}(A) = \mathcal{R}(A^\top)^\perp$

We prove one orthogonality relation; the other follows by applying the same argument to $A^\top$.

Proof Strategy: Double Inclusion

To prove $X = Y$: show (a) $X \subseteq Y$ and (b) $Y \subseteq X$. Here $X = \mathcal{N}(A)$, $Y = \mathcal{R}(A^\top)^\perp$.

4.1 Part (a): $\mathcal{N}(A) \subseteq \mathcal{R}(A^\top)^\perp$

Goal: Every null space vector is orthogonal to every row space vector.

Step 1: Take $x \in \mathcal{N}(A)$, so $Ax = 0$. For any $w \in \mathcal{R}(A^\top)$, write $w = A^\top z$ for some $z \in \mathbb{R}^m$.

Step 2: Apply Lemma 2:

$$\langle w, x\rangle = \langle A^\top z, x\rangle = \langle z, Ax\rangle = \langle z, 0\rangle = 0$$

Step 3: Since w was arbitrary, $x \in \mathcal{R}(A^\top)^\perp$.

4.2 Part (b): $\mathcal{R}(A^\top)^\perp \subseteq \mathcal{N}(A)$

Goal: Any vector orthogonal to the row space lies in the null space.

Step 1: Take $x \in \mathcal{R}(A^\top)^\perp$, so $\langle x, w\rangle = 0$ for all $w \in \mathcal{R}(A^\top)$.

Step 2: In particular, for all $z \in \mathbb{R}^m$:

$$0 = \langle x, A^\top z\rangle = \langle A^\top z, x\rangle = \langle z, Ax\rangle$$

where the last step uses Lemma 2.

Step 3: Since $\langle Ax, z\rangle = 0$ for all z, choosing $z = Ax$ gives $\|Ax\|^2 = 0$, hence $Ax = 0$.

The Finishing Move: Why This Works

If $\langle v, z\rangle = 0$ for *all* $z \in \mathbb{R}^m$, then $v = 0$. Why?
Method 1: Choose $z = e_i$ (standard basis): $\langle v, e_i\rangle = v_i = 0$ for each i. Every component vanishes, so $v = 0$.
Method 2: Choose $z = v$ to get $\|v\|^2 = 0$, hence $v = 0$.

4.3 Completing the Proof

Both inclusions give $\mathcal{N}(A) = \mathcal{R}(A^\top)^\perp$, equivalent to $\mathcal{N}(A) \oplus \mathcal{R}(A^\top) = \mathbb{R}^n$. $\square$

Corollary: Rank-Nullity

From $\mathcal{N}(A) = \mathcal{R}(A^\top)^\perp$: $\dim(\mathcal{N}(A)) = n - \dim(\mathcal{R}(A^\top)) = n - r$. The Fundamental Theorem implies Rank-Nullity.

Intuition

Each row $a_i^\top$ of A gives an equation $a_i^\top x = 0$: "x is orthogonal to a_i." To satisfy $Ax = 0$, x must be orthogonal to every row, hence to their span (the row space).

5. The Generalized Proof Method

Key Insight

The proof of $\mathcal{N}(A) = \mathcal{R}(A^\top)^\perp$ is not just a result—it's a **reusable template** for proving orthogonality relationships throughout linear algebra.

5.1 The Six-Step Template

1. **Reduce to two inclusions.** To prove $S = T$, show $S \subseteq T$ and $T \subseteq S$ separately.
2. **Start with arbitrary element.** Begin each direction with "Let $x \in S$" (or T). This is the only starting point.
3. **Translate orthogonal complement.** Use the "dictionary": $x \in S^\perp$ means $\langle x, w\rangle = 0$ for all $w \in S$.
4. **Replace subspace elements.** If $w \in \mathcal{R}(A^\top)$, write $w = A^\top z$ for some z. If $w \in \mathcal{R}(A)$, write $w = Ax$.
5. **Apply the adjoint identity.** The "bridge" between null space and range: $\langle A^\top y, x\rangle = \langle y, Ax\rangle$.
6. **Use the finishing move.** "Orthogonal to everything $\Rightarrow$ zero": if $\langle v, z\rangle = 0$ for all z, then $v = 0$ (choose $z = v$).

One-Line Summary

Unpack $\rightarrow$ Rewrite $\rightarrow$ Transpose $\rightarrow$ Zero $\rightarrow$ Conclude

Where This Method Applies

- Proving $\mathcal{N}(A^\top) = \mathcal{R}(A)^\perp$ (Section 5)
- Deriving the normal equations for least squares
- Showing projection matrices satisfy $P^2 = P$
- Proving optimality conditions in constrained optimization

6. Sketch: $\mathcal{N}(A^\top) = \mathcal{R}(A)^\perp$

For completeness, we outline the analogous proof.

Part (a): $\mathcal{N}(A^\top) \subseteq \mathcal{R}(A)^\perp$. Let $y \in \mathcal{N}(A^\top)$. For any $w = Ax \in \mathcal{R}(A)$:

$$\langle y, w\rangle = \langle y, Ax\rangle = (A^\top y)^\top x = 0$$

Part (b): $\mathcal{R}(A)^\perp \subseteq \mathcal{N}(A^\top)$. Let $y \in \mathcal{R}(A)^\perp$. Then $\langle y, Ax\rangle = 0$ for all x, so $\langle A^\top y, x\rangle = 0$ for all x. Choosing $x = A^\top y$ gives $A^\top y = 0$. $\square$

7. Application: Minimum-Norm Solutions

The FTLA has immediate applications to optimization. Consider *underdetermined* systems where infinitely many solutions exist.

Problem: Minimum-Norm Solution

Given $A \in \mathbb{R}^{m\times n}$ with $m < n$ and full row rank, and $b \in \mathbb{R}^m$:

$$\min_{x\in\mathbb{R}^n} \|x\|^2 \quad \text{subject to} \quad Ax = b$$

Since A has full row rank, $\mathcal{R}(A) = \mathbb{R}^m$, so $Ax = b$ is feasible for any $b \in \mathbb{R}^m$.

Key insight: If x_p is any particular solution ($Ax_p = b$), then the full solution set is $\{x_p + z : z \in \mathcal{N}(A)\}$. Adding null space directions doesn't change Ax. Which choice minimizes $\|x\|$?

7.1 Method 1: Geometric Derivation (via FTLA)

By the FTLA, $\mathbb{R}^n = \mathcal{R}(A^\top) \oplus \mathcal{N}(A)$. Any x decomposes uniquely:

$$x = x_r + x_n, \quad x_r \in \mathcal{R}(A^\top), \quad x_n \in \mathcal{N}(A)$$

Since $x_r \perp x_n$, the Pythagorean theorem gives:

$$\|x\|^2 = \|x_r\|^2 + \|x_n\|^2$$

Now consider the constraint. If $Ax = b$, then $Ax_r + Ax_n = b$. Since $x_n \in \mathcal{N}(A)$, we have $Ax_n = 0$, so $Ax_r = b$. The row space component x_r alone satisfies the constraint!

Conclusion: Among all solutions, the one with $x_n = 0$ (i.e., $x^\star = x_r \in \mathcal{R}(A^\top)$) minimizes the norm. The minimum-norm solution lies entirely in the row space.

7.2 Method 2: Lagrange Multipliers

Form the Lagrangian with multiplier $\lambda \in \mathbb{R}^m$:

$$\mathcal{L}(x, \lambda) = \tfrac{1}{2}\|x\|^2 + \lambda^\top(b - Ax)$$

Stationarity: $\nabla_x \mathcal{L} = x - A^\top\lambda = 0 \implies x = A^\top\lambda$

This confirms $x^\star \in \mathcal{R}(A^\top)$ automatically!

Constraint: Substituting into $Ax = b$:

$$AA^\top\lambda = b$$

Since A has full row rank, $AA^\top$ is invertible, so $\lambda = (AA^\top)^{-1}b$.

Solution: $x^\star = A^\top(AA^\top)^{-1}b$.

Minimum-Norm Solution: For $A \in \mathbb{R}^{m\times n}$ with full row rank ($m < n$):

$$x^\star = A^\top(AA^\top)^{-1}b = A^\dagger b$$

where $A^\dagger = A^\top(AA^\top)^{-1}$ is the **Moore–Penrose pseudoinverse**. Since A has full row rank, $AA^\dagger = I_m$, making it a right inverse.

Example 3 (Minimum-Norm for a 1×2 System)**.** Let $A = \begin{pmatrix} 1 & 1 \end{pmatrix}$ and $b = 2$. The constraint $x_1 + x_2 = 2$ defines a line of solutions.

General solution: $x = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + t\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ for $t \in \mathbb{R}$.

Note: $(1, -1)^\top \in \mathcal{N}(A)$ and $(1, 1)^\top \in \mathcal{R}(A^\top)$.

Using the formula: $AA^\top = 2$, so $(AA^\top)^{-1} = \frac{1}{2}$.

$$x^\star = A^\top(AA^\top)^{-1}b = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \cdot \frac{1}{2} \cdot 2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

This is the point on the solution line closest to the origin—the null space component has been "killed."

Key Takeaways

1. **Decomposition ⇒ Projection.** $X = S \oplus S^\perp$ means every vector splits uniquely into orthogonal components. *Use this to project onto subspaces.*
2. **Four subspaces, two orthogonal pairs.** Memorize: $\mathcal{N}(A) \perp \mathcal{R}(A^\top)$ in $\mathbb{R}^n$; $\mathcal{N}(A^\top) \perp \mathcal{R}(A)$ in $\mathbb{R}^m$. Dimensions sum to ambient space.
3. **Adjoint identity.** $\langle A^\top y, x\rangle = \langle y, Ax\rangle$ is the engine behind both orthogonality proofs.
4. **Solvability test.** $Ax = b$ solvable $\Leftrightarrow b \in \mathcal{R}(A) \Leftrightarrow b \perp \mathcal{N}(A^\top)$.
5. **Minimum-norm solutions.** For underdetermined systems: the FTLA shows the optimal $x^\star$ lies in $\mathcal{R}(A^\top)$. Formula: $x^\star = A^\top(AA^\top)^{-1}b$.