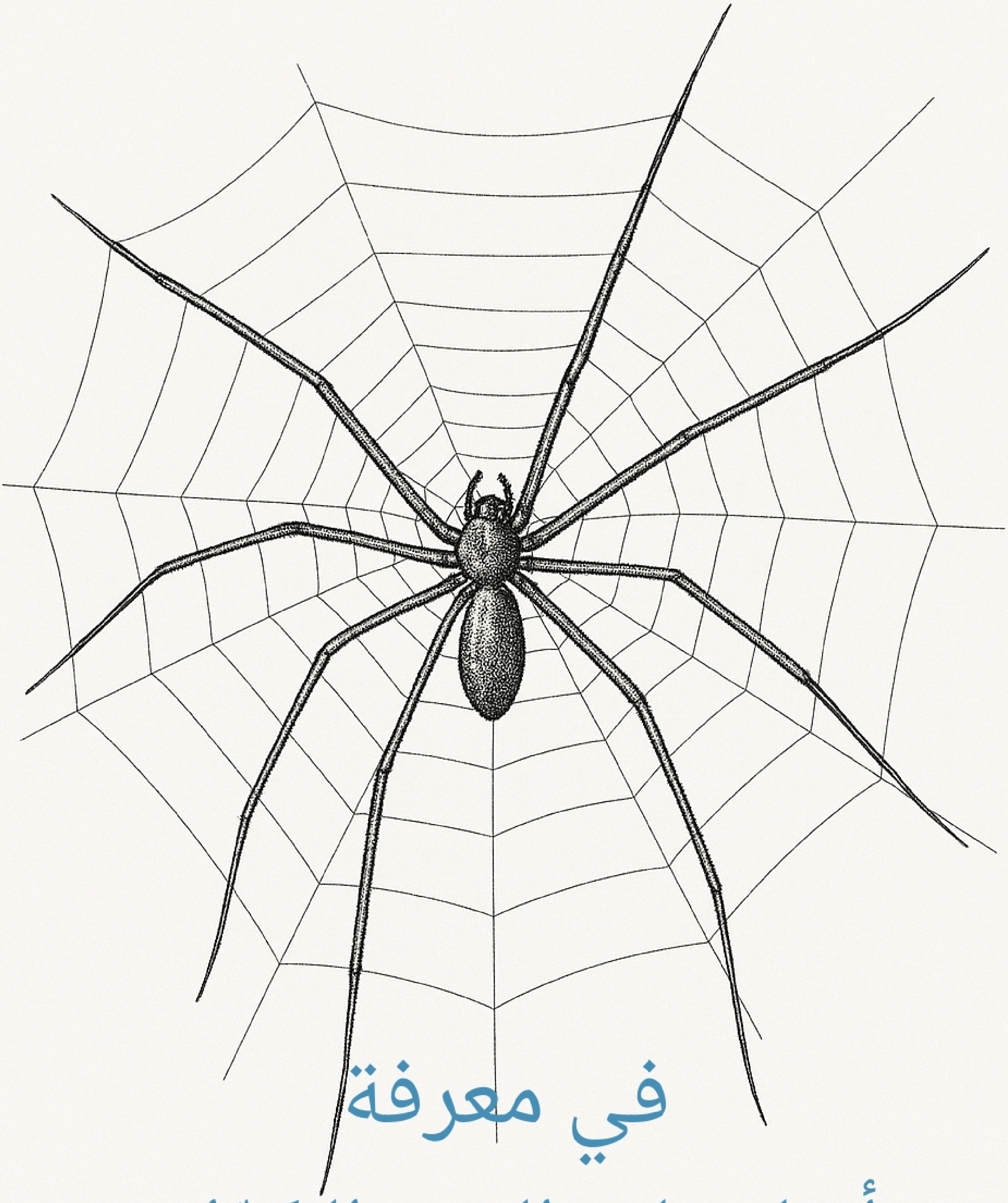
الجامع المُغطِّي



في معرفة
أساسيات الجبر الخَطِّي

📚 Linear Algebra — Table of Contents

- Hadamard Product
    - Elementwise Operation
    - Example with 3×3 Matrices
    - Applications

---

6. Linear Transformations

- 2D Linear Transformations:
    - Rotations
    - Reflections
    - Scaling
    - Shearing
    - Orthogonal Projections

---

7. Projections

- Projection onto a line
- Projection onto a subspace
- Properties of projections
- Worked numeric example

---

8. Solving Linear Systems

- Row Reduction (Gaussian Elimination)
- Row Echelon Form (REF) and Reduced Row Echelon Form (RREF)
- Using Row Reduction for:
    - Solving Linear Systems
    - Checking Linear Independence of Vectors
    - Determining Full Rank

---

9. Matrix Factorizations

- LU Decomposition
    - Step-by-Step Guide
    - Permutation Matrix and Why It's Needed
- QR Decomposition
    - Relation to Gram–Schmidt Process
- Gram–Schmidt Process

---

10. Determinants and Inverses

- Determinant
  - Intuitive and Mathematical Definitions
  - Calculation in 2D and 3D
  - Properties (5 detailed, rest listed)
- Invertibility vs Singularity
- Inverse of a 2×2 Matrix (by hand)

---

11. Eigenvalues & Eigenvectors

- Conceptual & intuitive explanation
- General steps to compute them
- Worked numerical example
- Relation to diagonalization

---

12. Diagonalization

- General steps with a numeric example
- Conditions for a matrix to be diagonalizable
- Applications of diagonalization

---

13. Positive (Semi)Definite Matrices

- Positive definite vs positive semidefinite
- Properties and how to distinguish them

---

14. Singular Value Decomposition (SVD)

- Statement: every matrix can be decomposed into SVD
- Proof walkthrough (step by step with formulas)
- Which matrices are PSD in the proof
- Worked numeric example

---

15. Covariance Matrix & PCA

- Proof that covariance matrix is PSD
- Relation to PCA
- Eigenvalues as principal components
- Steps of the PCA algorithm
- PCA for dimensionality reduction
- Why dimension reduction is needed

Scalars, Vectors, Matrices, and Tensors — Definitions and Examples

1. Scalar

Definition:
A scalar is a single number — an element of a field (usually real numbers $\mathbb{R}$ or complex numbers $\mathbb{C}$) that represents magnitude or value without direction.
Scalars are used to scale (stretch, shrink, or flip) vectors and matrices.

Notation:

- Lowercase italic letters: a, b, c, α, β
- $a \in \mathbb{R}$ means “a is a real number.”

Examples:

- $a = 5$ (real scalar)
- $\alpha = -2.7$ (real scalar)
- $\lambda = 3 + 4i$ (complex scalar)

2. Vector

Definition:
A vector is an ordered list (tuple) of scalars, representing a point or direction in space.
Vectors can be **column** or **row** vectors, and live in a vector space $\mathbb{R}^n$ or $\mathbb{C}^n$.

Notation:

- Bold lowercase: $\mathbf{v}, \mathbf{x}$
- Coordinates with indices: $\mathbf{v} = (v_1, v_2, \ldots, v_n)$
- Column vector form:

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

Examples:

- $\mathbf{v} = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} \in \mathbb{R}^3$

- $\mathbf{x} = (0.5, 4.2) \in \mathbb{R}^2$

---

3. Matrix

Definition:
A matrix is a rectangular array of scalars arranged in rows and columns.
A matrix represents a **linear transformation** from one vector space to another, or a dataset of numbers in tabular form.

Notation:

- Bold uppercase: $\mathbf{A}, \mathbf{M}$
- An $m \times n$ matrix has m rows and n columns.
- Element a_{ij} is in the i-th row and j-th column.

Example:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \in \mathbb{R}^{2\times 3}$$

Here:

- 2 rows, 3 columns.
- $a_{21} = 4$.

---

4. Tensor

Definition:
A tensor is a generalization of scalars (0D), vectors (1D), and matrices (2D) to higher dimensions.
Formally, it is a **multidimensional array** of numbers that transforms according to certain rules under a change of coordinates.
In machine learning, “tensor” usually just means “N-dimensional array of numbers.”

Notation:

- Script or calligraphic letters: $\mathcal{T}, \mathcal{X}$
- Indices for multiple dimensions: $T_{ijk\ell}$

Examples:

- **Scalar:** 7 → 0D tensor
- **Vector:** $[1, 2, 3]$ → 1D tensor
- **Matrix:** $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ → 2D tensor
- **3D Tensor:**

$$\mathcal{T}_{ijk} \in \mathbb{R}^{2\times 3\times 4}$$

could be thought of as a stack of 2×3 matrices in 4 layers.

Vector Addition and Subtraction

---

1. Mathematical Definition

Let:

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

Addition

$$\mathbf{a} + \mathbf{b} = \begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \end{bmatrix}$$

Subtraction

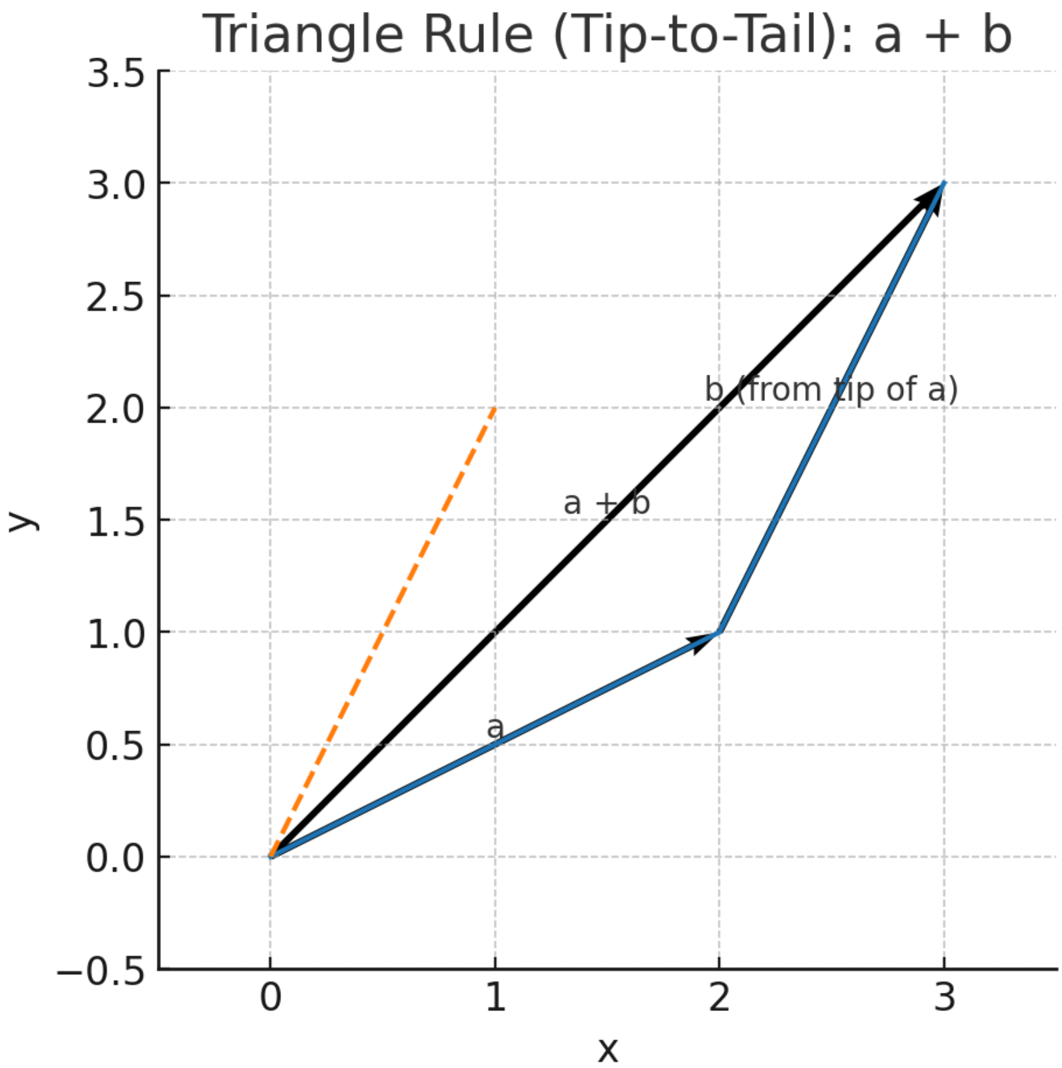
$$\mathbf{a} - \mathbf{b} = \begin{bmatrix} a_1 - b_1 \\ a_2 - b_2 \end{bmatrix}$$

---

2. Geometric Interpretation

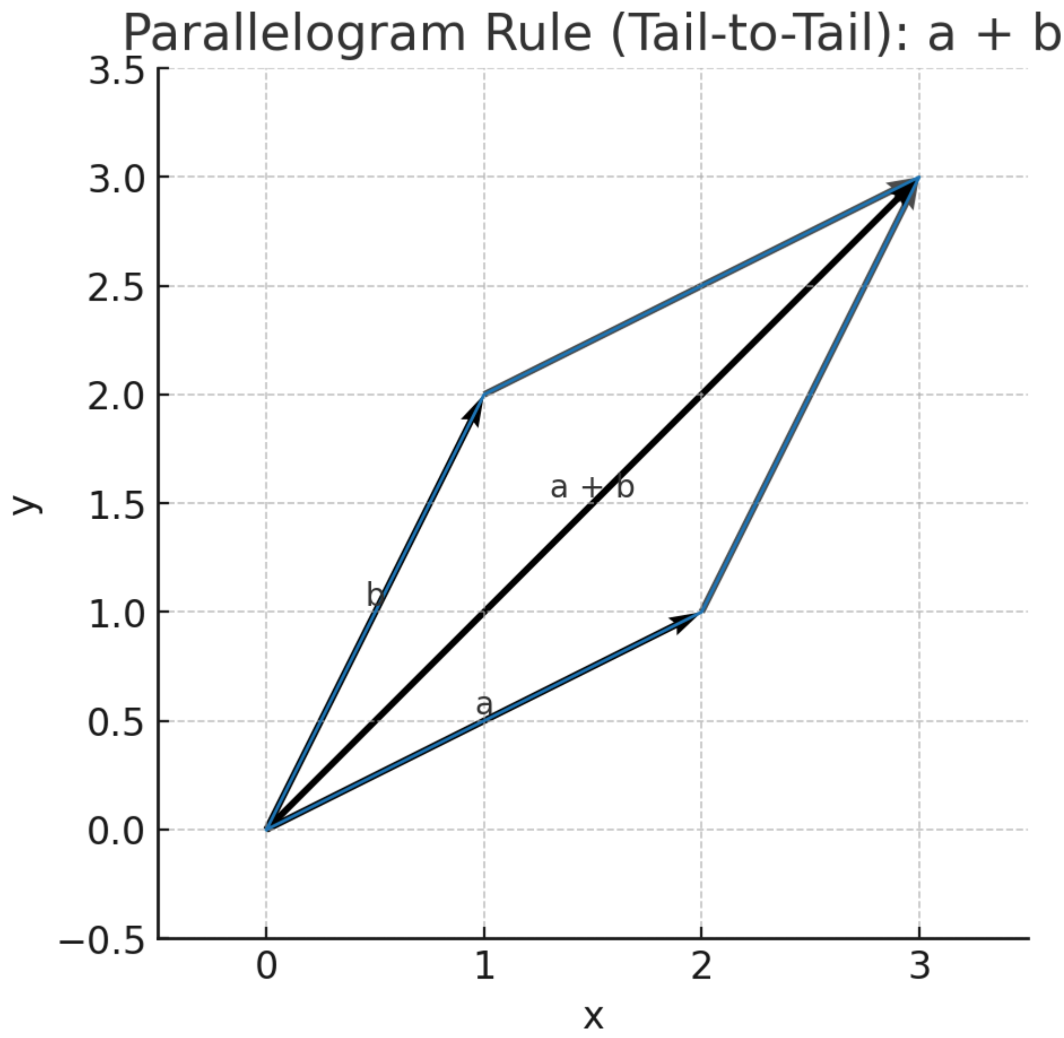
Triangle Rule (Tip-to-Tail Method)

- Place the **tail** of vector $\mathbf{b}$ at the **tip** of vector $\mathbf{a}$.
- The resultant vector $\mathbf{a} + \mathbf{b}$ is drawn from the **tail** of $\mathbf{a}$ to the **tip** of $\mathbf{b}$.



Parallelogram Rule (Tail-to-Tail Method)

- Place both vectors **tail-to-tail**.
- Draw a parallelogram where $\mathbf{a}$ and $\mathbf{b}$ are adjacent sides.
- The diagonal from the common tail is the resultant $\mathbf{a} + \mathbf{b}$.



3. Example

Let:

$$\mathbf{a} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

Addition:

$$\mathbf{a} + \mathbf{b} = \begin{bmatrix} 2+1 \\ 1+3 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

Subtraction:

$$\mathbf{a} - \mathbf{b} = \begin{bmatrix} 2-1 \\ 1-3 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

4. Graphical Meaning of Subtraction

- Subtraction $\mathbf{a} - \mathbf{b}$ is equivalent to **adding** $\mathbf{a}$ to **the negative** of $\mathbf{b}$:

$$\mathbf{a} - \mathbf{b} = \mathbf{a} + (-\mathbf{b})$$

- Geometrically, reverse $\mathbf{b}$, then apply the triangle rule.

---

5. Notes

- Vector addition is **commutative**:

$$\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$$

- Vector subtraction is **not** commutative:

$$\mathbf{a} - \mathbf{b} \neq \mathbf{b} - \mathbf{a}$$

Graphical Representation:

- *Triangle Rule*: move one vector so its tail meets the other's tip.
- *Parallelogram Rule*: keep both tails together and draw a parallelogram.

Algebraic Properties of Vector Addition and Scalar Multiplication in a Plane

1. For all vectors $\mathbf{x}$ and $\mathbf{y}$,

$$\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$$

2. For all vectors $\mathbf{x}, \mathbf{y}, \mathbf{z}$,

$$(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$$

3. There exists a vector denoted $\mathbf{0}$ such that

$$\mathbf{x} + \mathbf{0} = \mathbf{x} \quad \text{for each vector } \mathbf{x}.$$

4. For each vector $\mathbf{x}$, there is a vector $\mathbf{y}$ such that

$$\mathbf{x} + \mathbf{y} = \mathbf{0}.$$

5. For each vector $\mathbf{x}$,

$$1\mathbf{x} = \mathbf{x}.$$

6. For each pair of real numbers a and b and each vector $\mathbf{x}$,

$$(ab)\mathbf{x} = a(b\mathbf{x}).$$

7. For each real number a and each pair of vectors $\mathbf{x}$ and $\mathbf{y}$,

$$a(\mathbf{x} + \mathbf{y}) = a\mathbf{x} + a\mathbf{y}.$$

8. For each pair of real numbers a and b and each vector $\mathbf{x}$,

$$(a + b)\mathbf{x} = a\mathbf{x} + b\mathbf{x}.$$

The Dot Product in Linear Algebra

---

1. Definition

The **dot product** (also called **scalar product**) between two vectors $x, y \in \mathbb{R}^n$ is defined as:

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

It can also be written in matrix form as:

$$x \cdot y = x^T y$$

where x^T is the **transpose** of vector x.

---

2. Geometric Interpretation

The dot product measures how much two vectors point in the same direction.
It is given by:

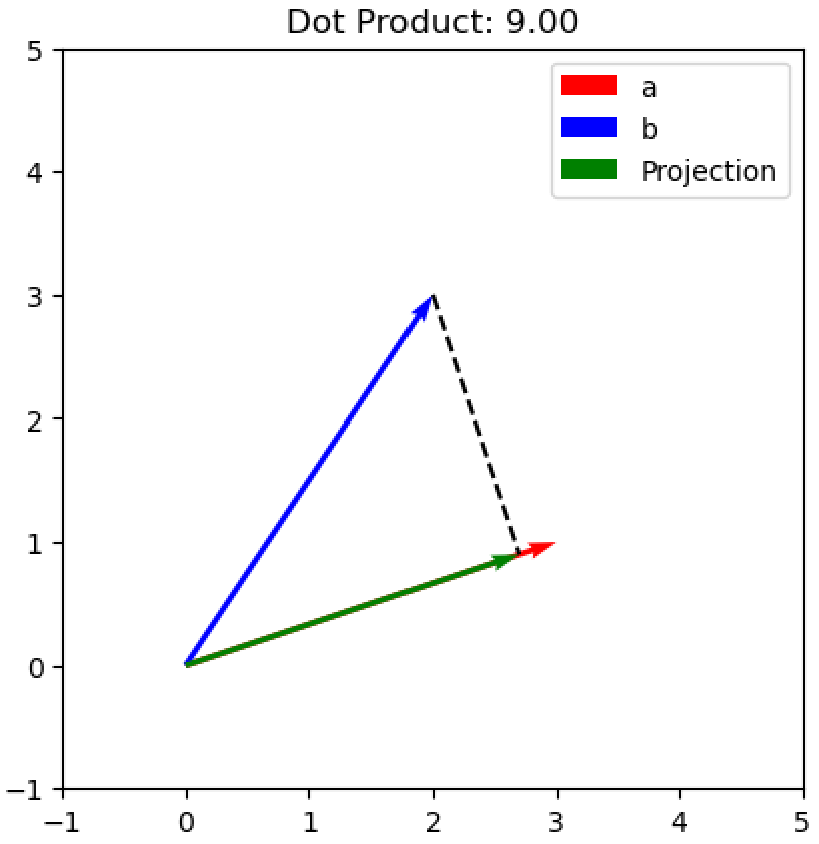
$$x \cdot y = ||x|| \, ||y|| \cos \theta$$

where:

- $||x||$ and $||y||$ are the magnitudes (lengths) of x and y
- θ is the angle between x and y

Projection form:

$$x \cdot y = ||x|| \cdot (\text{projection of } y \text{ onto } x)$$



3. Example

Let:

$$x = \begin{bmatrix} 3 \\ 4 \end{bmatrix}, \quad y = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

Algebraic form:

$$x \cdot y = (3)(2) + (4)(-1) = 6 - 4 = 2$$

Geometric form:

- Magnitudes:

$$\|x\| = \sqrt{3^2 + 4^2} = 5, \quad \|y\| = \sqrt{2^2 + (-1)^2} = \sqrt{5}$$

- Using the cosine formula:

$$x \cdot y = \|x\|\|y\| \cos\theta$$

Substitute:

$$2 = (5)(\sqrt{5}) \cos\theta$$

$$\cos\theta = \frac{2}{5\sqrt{5}}$$

4. Interpretation of Results

- **Positive** ($x \cdot y > 0$): Vectors point in a generally **same** direction ($\theta < 90^\circ$)
- **Zero** ($x \cdot y = 0$): Vectors are **orthogonal** ($\theta = 90^\circ$)
- **Negative** ($x \cdot y < 0$): Vectors point in **opposite** directions ($\theta > 90^\circ$)

---

5. Self Dot Product

If $x \in \mathbb{R}^n$:

$$x^T x = \sum_{i=1}^{n} x_i^2 = \|x\|^2$$

So the dot product of a vector with itself equals its **magnitude squared**.

Example:

$$x = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

$$x^T x = 3^2 + 4^2 = 25 = \|x\|^2$$

---

6. Row vs Column Multiplication

- **Row vector × Column vector** (inner product):

$$[1 \quad 2 \quad 3] \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} = 1 \cdot 4 + 2 \cdot 5 + 3 \cdot 6 = 32$$

Result: Scalar

- **Column vector × Row vector** (outer product):

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} [4 \quad 5 \quad 6] = \begin{bmatrix} 4 & 5 & 6 \\ 8 & 10 & 12 \\ 12 & 15 & 18 \end{bmatrix}$$

Result: Matrix (**outer product**).

---

7. Applications of Dot Product in Machine Learning

- **Cosine similarity** for document similarity in NLP:

$$\text{similarity}(x, y) = \frac{x \cdot y}{\|x\| \|y\|}$$

- **Word embeddings**: Measuring semantic similarity (e.g., Word2Vec, GloVe)
- **Attention mechanisms** in Transformers (query-key dot products)
- **Projection operations** in dimensionality reduction (PCA)
- **Computing distances** in k-Nearest Neighbors (via norms)
- **Neural networks**: Weighted sums in perceptrons and dense layers
- **Recommendation systems**: Matching user and item feature vectors
- **Computer vision**: Template matching via cross-correlation
- **Gradient computations**: Backpropagation dot products

Cross Product (Vector Product)

Definition

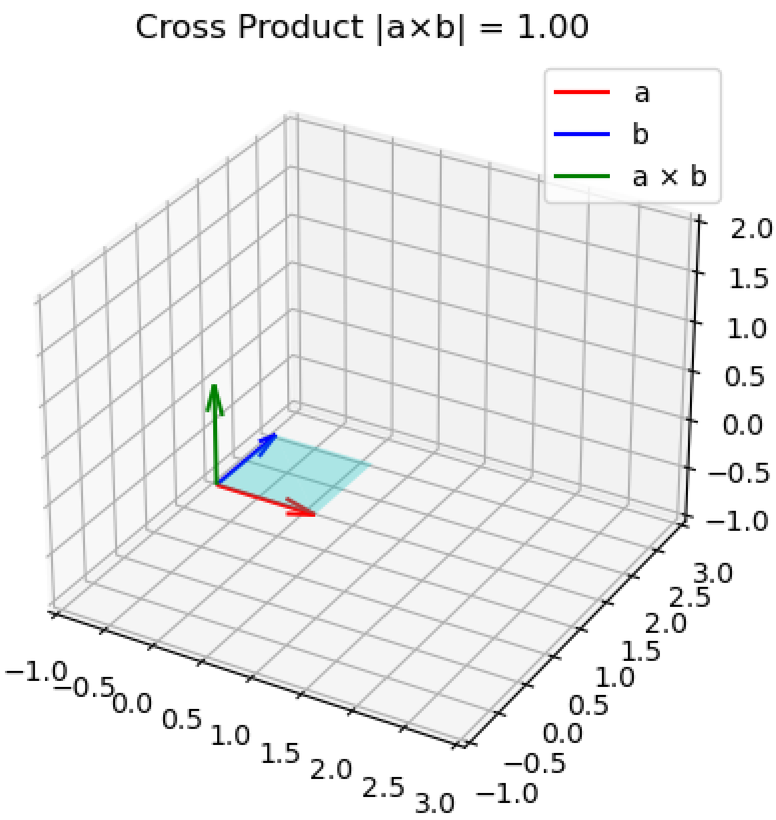
The **cross product** is an operation between two **3D vectors** $a, b \in \mathbb{R}^3$ that results in another **3D vector** perpendicular to both a and b.

If:

$$a = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

then:

$$a \times b = \begin{bmatrix} a_2b_3 - a_3b_2 \\ a_3b_1 - a_1b_3 \\ a_1b_2 - a_2b_1 \end{bmatrix}$$



Magnitude and Direction

- **Magnitude**:

$$\|a \times b\| = \|a\| \, \|b\| \, \sin\theta$$

where θ is the angle between a and b.

- **Direction**: Perpendicular to both a and b, following the **right-hand rule**.

Example

Let:

$$a = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad b = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$$

Step-by-step:

$$a \times b = \begin{bmatrix} (2)(6) - (3)(5) \\ (3)(4) - (1)(6) \\ (1)(5) - (2)(4) \end{bmatrix} = \begin{bmatrix} 12 - 15 \\ 12 - 6 \\ 5 - 8 \end{bmatrix} = \begin{bmatrix} -3 \\ 6 \\ -3 \end{bmatrix}$$

Interpretation

- $a \times b$ is **orthogonal** to both a and b.

- The **magnitude** of $a \times b$ equals the **area of the parallelogram** spanned by a and b.

Applications

- **Computer graphics**: finding surface normals for lighting calculations
- **Physics**: torque $\tau = r \times F$ and angular momentum
- **Robotics**: computing rotational effects of forces
- **3D geometry**: finding perpendicular vectors to planes
- **Navigation**: determining orientation using gyroscopic data

---

Linear Algebra — Key Terminology and Examples (Interview Style)

Q1: What is linear dependence and independence?

- **Linear dependence:** A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ is **linearly dependent** if at least one vector can be written as a linear combination of the others.
  Example: In $\mathbb{R}^3$,
  $\mathbf{v}_1 = (1, 2, 3)$, $\mathbf{v}_2 = (2, 4, 6)$, $\mathbf{v}_3 = (0, 1, 1)$
  Here, $\mathbf{v}_2 = 2\mathbf{v}_1$, so the set is dependent.

- **Linear independence:** A set of vectors is **linearly independent** if the only solution to
  $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_k\mathbf{v}_k = \mathbf{0}$
  is $c_1 = c_2 = \dots = c_k = 0$.
  Example: In $\mathbb{R}^2$, $(1, 0)$ and $(0, 1)$ are independent.

Q2: What are linear combinations?

A **linear combination** of vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ is any vector that can be formed as:

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \dots + a_k\mathbf{v}_k$$

where $a_1, a_2, \dots, a_k$ are scalars.
Example: If $\mathbf{v}_1 = (1, 0)$, $\mathbf{v}_2 = (0, 1)$, then $(3, 4) = 3\mathbf{v}_1 + 4\mathbf{v}_2$.

Q3: What are the row space, column space, and null space?

- **Row space**: The span of the row vectors of a matrix.

- **Column space**: The span of the column vectors of a matrix (also the image of the linear transformation).
- **Null space**: The set of all vectors $\mathbf{x}$ such that $\mathbf{A}\mathbf{x} = \mathbf{0}$.

Example:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

- Row space: span of $(1, 2, 3)$ and $(4, 5, 6)$ in $\mathbb{R}^3$
- Column space: span of $(1, 4)$, $(2, 5)$, $(3, 6)$ in $\mathbb{R}^2$
- Null space: all $\mathbf{x} \in \mathbb{R}^3$ satisfying $\mathbf{A}\mathbf{x} = \mathbf{0}$

Q4: What is the span?

The **span** of a set of vectors is the set of all possible linear combinations of those vectors.
Example: In $\mathbb{R}^2$, the span of $(1, 0)$ and $(0, 1)$ is all of $\mathbb{R}^2$.

Q5: What is a basis?

A **basis** is a set of **linearly independent vectors** that spans a vector space.
Example: In $\mathbb{R}^3$, the standard basis is:

$$\{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$$

Q6: What is the rank of a matrix and what makes it full rank?

- **Rank:** The dimension of the column space (or row space) of the matrix.
- **Full rank:** A matrix is full rank if its rank equals the smallest of its dimensions ($\min(m, n)$ for an $m \times n$ matrix).

Example:
For $\mathbf{A} \in \mathbb{R}^{3\times3}$, full rank means rank $= 3$.

Q7: What is the Rank–Nullity Theorem?

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

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

where:

- **rank** = dimension of the column space
- **nullity** = dimension of the null space

Q8: What is a subspace and what are the conditions for a valid subspace?

A **subspace** of a vector space V is a subset $W \subseteq V$ that is also a vector space under the same operations.

Conditions:

1. The zero vector $\mathbf{0}$ is in W
2. If $\mathbf{u}, \mathbf{v} \in W$, then $\mathbf{u} + \mathbf{v} \in W$ (closed under addition)
3. If $\mathbf{u} \in W$ and c is a scalar, then $c\mathbf{u} \in W$ (closed under scalar multiplication)

Example: The set of all vectors in $\mathbb{R}^3$ with $z = 0$ is a subspace (the xy-plane).

Q9: What are normal transformations?

A linear transformation represented by a matrix $\mathbf{A}$ is **normal** if:

$$\mathbf{A}\mathbf{A}^* = \mathbf{A}^*\mathbf{A}$$

where $\mathbf{A}^*$ is the conjugate transpose of $\mathbf{A}$.

- For real matrices, $\mathbf{A}^T\mathbf{A} = \mathbf{A}\mathbf{A}^T$

Example: All symmetric matrices are normal.

Q10: How to check if a transformation is linear?

A transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is **linear** if for all vectors $\mathbf{u}, \mathbf{v}$ and scalar c:

1. **Additivity:** $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$
2. **Homogeneity:** $T(c\mathbf{u}) = cT(\mathbf{u})$

Intuitive check: The transformation preserves straight lines and the origin.

Example:
$T(x, y) = (2x, 3y)$ is linear.
$T(x, y) = (x + 1, y)$ is **not** linear (fails origin preservation).

Special Matrices in Linear Algebra

---

1. Diagonal Matrix

A **diagonal matrix** has all non-diagonal entries equal to zero:

$$A_{ij} = 0 \quad \text{for} \quad i \neq j$$

Example:

$$D = \begin{bmatrix} 4 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

---

2. Anti-Diagonal Matrix

An **anti-diagonal matrix** has all elements zero except those on the anti-diagonal (from top-right to bottom-left):

$$A_{ij} \neq 0 \quad \text{only if} \quad i + j = n + 1$$

Example:

$$M = \begin{bmatrix} 0 & 0 & 7 \\ 0 & 5 & 0 \\ 3 & 0 & 0 \end{bmatrix}$$

---

3. Identity Matrix

An **identity matrix** is a special diagonal matrix with all diagonal entries equal to 1:

$$I_n = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

It satisfies:

$$I_n x = x, \quad I_n A = A, \quad A I_n = A$$

---

4. Triangular Matrices

- **Upper Triangular**: All elements below the main diagonal are zero.

$$U = \begin{bmatrix} 2 & -1 & 3 \\ 0 & 5 & 4 \\ 0 & 0 & 6 \end{bmatrix}$$

- **Lower Triangular**: All elements above the main diagonal are zero.

$$L = \begin{bmatrix} 2 & 0 & 0 \\ -3 & 4 & 0 \\ 1 & 2 & 5 \end{bmatrix}$$

---

5. Symmetric Matrix

A **symmetric matrix** satisfies:

$$A^T = A$$

Example:

$$S = \begin{bmatrix} 2 & 3 & 4 \\ 3 & 5 & -1 \\ 4 & -1 & 0 \end{bmatrix}$$

---

6. Skew-Symmetric Matrix

A **skew-symmetric matrix** satisfies:

$$A^T = -A$$

Note: All diagonal elements must be zero. Example:

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

---

7. Orthogonal Matrix

A **square matrix** Q is orthogonal if:

$$Q^TQ = QQ^T = I$$

This means columns (and rows) are orthonormal vectors. Example:

$$Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

Here:

$$Q^TQ = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

---

8. Permutation Matrix

A **permutation matrix** is obtained by permuting the rows (or columns) of an identity matrix.
It represents a reordering of vector components when multiplied.

If P is a permutation matrix and x a column vector, then Px reorders the elements of x.

Example (swap row 1 and row 3 of I_3):

$$
P = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}
$$

Applied to:

$$
x = \begin{bmatrix} a \\ b \\ c \end{bmatrix}
$$

We get:

$$
Px = \begin{bmatrix} c \\ b \\ a \end{bmatrix}
$$

Uses:

- Rearranging equations in a system for numerical stability.
- Row swapping in **LU decomposition with partial pivoting**.
- Representing discrete permutations in combinatorics.

Matrix Dimensions and Multiplication Rules

---

1. Dimensions of a Matrix

The **dimension** (or size) of a matrix is given as:

$$
m \times n
$$

where:

- m = number of **rows**
- n = number of **columns**

Example:

$$
\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \in \mathbb{R}^{2\times3}
$$

This is a 2×3 matrix (2 rows, 3 columns).

---

2. Condition for Matrix Multiplication

For two matrices $\mathbf{A}$ and $\mathbf{B}$:

- If $\mathbf{A}$ has dimensions $m \times n$
- and $\mathbf{B}$ has dimensions $n \times p$

then the multiplication:

$$\mathbf{C} = \mathbf{A} \cdot \mathbf{B}$$

is **defined** only when the **number of columns of $\mathbf{A}$ equals the number of rows of $\mathbf{B}$**.

The **result $\mathbf{C}$** will have dimensions:

$$m \times p$$

---

3. Examples of Matrix Multiplication

Example 1 — 2D Square Matrices

$\mathbf{A} \in \mathbb{R}^{2\times 2}$ and $\mathbf{B} \in \mathbb{R}^{2\times 2}$

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}$$

Multiplication:

$$\mathbf{AB} = \begin{bmatrix} 1\cdot 5 + 2\cdot 7 & 1\cdot 6 + 2\cdot 8 \\ 3\cdot 5 + 4\cdot 7 & 3\cdot 6 + 4\cdot 8 \end{bmatrix} = \begin{bmatrix} 19 & 22 \\ 43 & 50 \end{bmatrix}$$

Dimensions: 2×2

Example 2 — 2D Rectangular Matrices

$\mathbf{A} \in \mathbb{R}^{2\times 3}$ and $\mathbf{B} \in \mathbb{R}^{3\times 2}$

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 2 \\ -1 & 3 & 1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 3 & 1 \\ 2 & 1 \\ 1 & 0 \end{bmatrix}$$

Multiplication:

$$\mathbf{AB} = \begin{bmatrix} 1\cdot 3 + 0\cdot 2 + 2\cdot 1 & 1\cdot 1 + 0\cdot 1 + 2\cdot 0 \\ (-1)\cdot 3 + 3\cdot 2 + 1\cdot 1 & (-1)\cdot 1 + 3\cdot 1 + 1\cdot 0 \end{bmatrix} = \begin{bmatrix} 5 & 1 \\ 4 & 2 \end{bmatrix}$$

Dimensions: 2×2

Example 3 — 3D Square Matrices

$\mathbf{A} \in \mathbb{R}^{3\times 3}$ and $\mathbf{B} \in \mathbb{R}^{3\times 3}$

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 5 & 6 & 0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} -2 & 1 & 0 \\ 3 & 0 & 1 \\ 4 & -1 & 2 \end{bmatrix}$$

Multiplication:

$$\mathbf{AB} = \begin{bmatrix} 1(-2)+2(3)+3(4) & 1(1)+2(0)+3(-1) & 1(0)+2(1)+3(2) \\ 0(-2)+1(3)+4(4) & 0(1)+1(0)+4(-1) & 0(0)+1(1)+4(2) \\ 5(-2)+6(3)+0(4) & 5(1)+6(0)+0(-1) & 5(0)+6(1)+0(2) \end{bmatrix} = \begin{bmatrix} 16 & - \\ 19 & - \\ 8 & \end{bmatrix}$$

Dimensions: 3×3

Example 4 — 3D Rectangular Matrices

$\mathbf{A} \in \mathbb{R}^{3\times 2}$ and $\mathbf{B} \in \mathbb{R}^{2\times 4}$

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 \end{bmatrix}$$

Multiplication:

$$\mathbf{AB} = \begin{bmatrix} 1(7)+2(11) & 1(8)+2(12) & 1(9)+2(13) & 1(10)+2(14) \\ 3(7)+4(11) & 3(8)+4(12) & 3(9)+4(13) & 3(10)+4(14) \\ 5(7)+6(11) & 5(8)+6(12) & 5(9)+6(13) & 5(10)+6(14) \end{bmatrix} = \begin{bmatrix} 29 & 32 \\ 65 & 72 \\ 101 & 112 \end{bmatrix}$$

Dimensions: 3×4

Hadamard Product

---

1. Hadamard Product (Elementwise Multiplication)

Definition

The **Hadamard product** between two matrices $A, B \in \mathbb{R}^{m\times n}$ is defined as:

$$A \circ B = [a_{ij} \cdot b_{ij}]$$

This means **multiply each element in A with the corresponding element in B**.

Example

Let:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}, \quad B = \begin{bmatrix} 9 & 8 & 7 \\ 6 & 5 & 4 \\ 3 & 2 & 1 \end{bmatrix}$$

The Hadamard product is:

$$
A \circ B = \begin{bmatrix} 1 \cdot 9 & 2 \cdot 8 & 3 \cdot 7 \\ 4 \cdot 6 & 5 \cdot 5 & 6 \cdot 4 \\ 7 \cdot 3 & 8 \cdot 2 & 9 \cdot 1 \end{bmatrix} = \begin{bmatrix} 9 & 16 & 21 \\ 24 & 25 & 24 \\ 21 & 16 & 9 \end{bmatrix}
$$

Applications

- **Elementwise weighting** in image processing (e.g., applying masks/filters)
- **Feature-wise multiplication** in machine learning (e.g., attention masks)
- **Neural networks**: gating mechanisms in LSTMs/GRUs
- **Financial modeling**: multiplying returns or rates by risk factors per element
- **Matrix blending** in graphics and simulations

Linear Transformations in Linear Algebra 2D version

---

1. Rotation

A **rotation** transformation rotates vectors in a plane around the origin by an angle θ.

Matrix (2D):

$$
R(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}
$$

Example (rotation by 90° counterclockwise):

$$
R\left(\frac{\pi}{2}\right) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}
$$

Applied to $\mathbf{v} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$:

$$
R\mathbf{v} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}
$$

---

2. Reflection

A **reflection** flips vectors over a line (in 2D) or a plane (in 3D).

Matrix (2D reflection over x-axis):

$$
M_x = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}
$$

Example:

$$M_x \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 3 \\ -4 \end{bmatrix}$$

---

3. Scaling

A **scaling** transformation stretches or shrinks vectors along coordinate axes.

Matrix (2D scaling by s_x in x and s_y in y):

$$S = \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix}$$

Example (double x, halve y):

$$S = \begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix}$$

Applied to $\begin{bmatrix} 3 \\ 4 \end{bmatrix}$:

$$\begin{bmatrix} 6 \\ 2 \end{bmatrix}$$

---

4. Shearing

A **shearing** transformation slants the shape of an object.

Matrix (shear in x direction by k):

$$H_x = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$$

Matrix (shear in y direction by k):

$$H_y = \begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}$$

Example (shear x by $k = 1.5$):

$$H_x = \begin{bmatrix} 1 & 1.5 \\ 0 & 1 \end{bmatrix}$$

Applied to $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$:

$$\begin{bmatrix} 2 + 1.5 \cdot 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 6.5 \\ 3 \end{bmatrix}$$

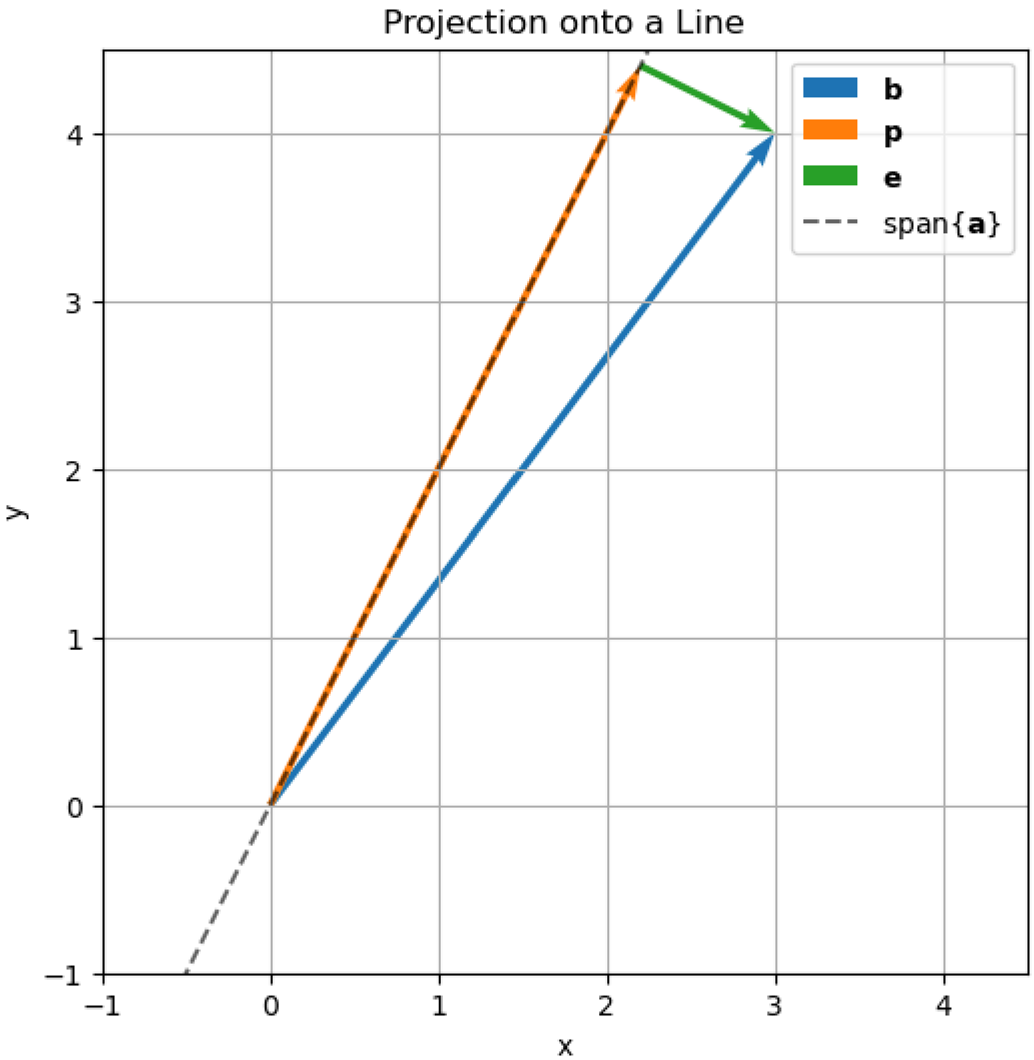
Projections onto Lines and Subspaces

Big picture (what projection does)

Given a vector $\mathbf{b}$ and a subspace S (a line, plane, or column space), the **orthogonal projection** finds the **closest** point to b, vector $\mathbf{p} \in S$ and decomposes

$$\mathbf{b} = \mathbf{p} + \mathbf{e}, \qquad \mathbf{p} \in S,\ \mathbf{e} \perp S.$$

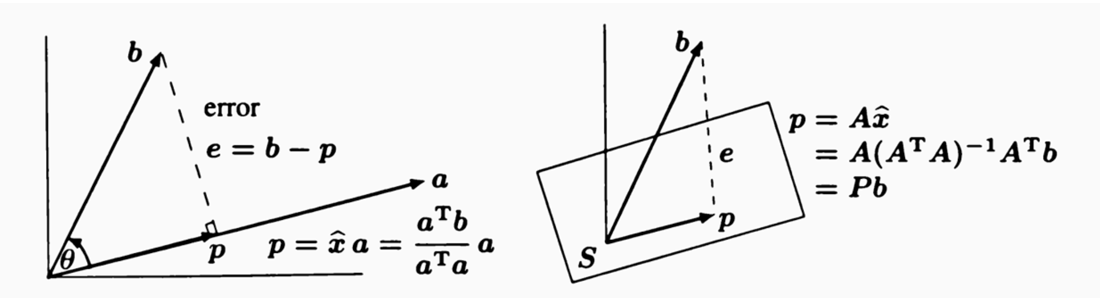
This makes a right triangle with legs $\mathbf{p}$ and $\mathbf{e}$ and hypotenuse $\mathbf{b}$.



---

1) Projection onto a line through a

Claim (formula): The projection of $\mathbf{b}$ onto the line through a nonzero vector $\mathbf{a}$ is the closest point to $\mathbf{b}$ on that line and equals



$$\boxed{\mathbf{p} = \mathbf{a}\,\frac{\mathbf{a}^\top \mathbf{b}}{\mathbf{a}^\top \mathbf{a}}}.$$

Why (derivation sketch): Points on the line through $\mathbf{a}$ have the form $x\,\mathbf{a}$. Minimize the distance

$$\min_x \|\mathbf{b} - x\mathbf{a}\|^2.$$

Differentiate w.r.t. x (or use normal equations) to get

$$\mathbf{a}^\top(\mathbf{b} - x\mathbf{a}) = 0 \;\Rightarrow\; x^{*} = \frac{\mathbf{a}^\top \mathbf{b}}{\mathbf{a}^\top \mathbf{a}}, \quad \mathbf{p} = x^{*}\,\mathbf{a}.$$

---

2) Error is perpendicular; Pythagorean identity

Let $\mathbf{e} = \mathbf{b} - \mathbf{p}$.

- **Perpendicularity:**

  $$\mathbf{a}^\top \mathbf{e} = \mathbf{a}^\top(\mathbf{b} - \mathbf{p}) = \mathbf{a}^\top \mathbf{b} - \mathbf{a}^\top \mathbf{a}\,\frac{\mathbf{a}^\top \mathbf{b}}{\mathbf{a}^\top \mathbf{a}} = 0.$$

  So $\mathbf{e} \perp \mathbf{a}$ (i.e., to the line).

- **Right triangle and Pythagorean identity:** Because $\mathbf{p} \perp \mathbf{e}$,

  $$\boxed{\|\mathbf{b}\|^2 = \|\mathbf{p}\|^2 + \|\mathbf{e}\|^2}.$$

---

3) Projection onto the column space $\mathcal{C}(A)$

If the columns of A are independent (so $A^\top A$ is invertible), the projection of $\mathbf{b}$ onto $S = \mathcal{C}(A)$ is

$$\boxed{\mathbf{p} = A(A^\top A)^{-1} A^\top \mathbf{b}}.$$

Reason: Write $\mathbf{p} = A\mathbf{x}$ for some coefficients $\mathbf{x}$. The orthogonality condition $A^\top(\mathbf{b} - A\mathbf{x}) = \mathbf{0}$ gives

$$A^\top A\,\mathbf{x} = A^\top \mathbf{b} \quad\Rightarrow\quad \mathbf{x} = (A^\top A)^{-1} A^\top \mathbf{b},$$

and hence $\mathbf{p} = A\mathbf{x} = A(A^\top A)^{-1} A^\top \mathbf{b}$.

---

6) The projection matrix onto $\mathcal{C}(A)$

Define

$$\boxed{P = A(A^\top A)^{-1} A^\top}.$$

Then

$$\boxed{\mathbf{p} = P\mathbf{b}}, \qquad \boxed{P^2 = P}, \qquad \boxed{P^\top = P}.$$

- $P^2 = P$ (**idempotent**) and $P^\top = P$ (**symmetric**) are the hallmark properties of an orthogonal projector.
- Consequences: $\mathrm{im}(P) = \mathcal{C}(A)$, $\mathcal{N}(P) = \mathcal{C}(A)^\perp$, and the eigenvalues of P are only 0 and 1.

Worked Numeric Examples

A) Projection onto a **line**

Let

$$\mathbf{a} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}.$$

Projection:

$$\mathbf{a}^\top \mathbf{b} = 1 \cdot 3 + 2 \cdot 4 = 11, \qquad \mathbf{a}^\top \mathbf{a} = 1^2 + 2^2 = 5.$$

$$\boxed{\mathbf{p} = \mathbf{a}\,\frac{11}{5} = \begin{bmatrix} \frac{11}{5} \\ \frac{22}{5} \end{bmatrix}}, \qquad \mathbf{e} = \mathbf{b} - \mathbf{p} = \begin{bmatrix} \frac{4}{5} \\ -\frac{2}{5} \end{bmatrix}.$$

Checks:

- Orthogonality:

$$\mathbf{a}^\top \mathbf{e} = 1 \cdot \tfrac{4}{5} + 2 \cdot \left(-\tfrac{2}{5} \right) = 0.$$

- Pythagorean:

$$\|\mathbf{p}\|^2 = \frac{121}{25} + \frac{484}{25} = \frac{121}{5}, \quad \|\mathbf{e}\|^2 = \frac{16}{25} + \frac{4}{25} = \frac{4}{5},$$

$$\|\mathbf{p}\|^2 + \|\mathbf{e}\|^2 = \frac{121}{5} + \frac{4}{5} = \frac{125}{5} = 25 = \|\mathbf{b}\|^2.$$

B) Projection onto a **subspace** $\mathcal{C}(A)$

Let

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \in \mathbb{R}^{3\times 2}, \qquad \mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}.$$

The columns of A are independent, so $A^\top A$ is invertible.

Compute the ingredients:

$$A^\top A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \qquad (A^\top A)^{-1} = \frac{1}{3}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

$$A^\top \mathbf{b} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}.$$

Projection vector:

$$\mathbf{p} = A(A^\top A)^{-1}A^\top \mathbf{b} = A\left(\frac{1}{3}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} 3 \\ 4 \end{bmatrix}\right) = A\left(\frac{1}{3}\begin{bmatrix} 2 \\ 5 \end{bmatrix}\right) = A\begin{bmatrix} \frac{2}{3} \\ \frac{5}{3} \end{bmatrix} = \begin{bmatrix} \frac{2}{3} \\ \frac{7}{3} \\ \frac{5}{3} \end{bmatrix}.$$

Error and orthogonality:

$$\mathbf{e} = \mathbf{b} - \mathbf{p} = \begin{bmatrix} \frac{1}{3} \\ -\frac{1}{3} \\ \frac{1}{3} \end{bmatrix}, \qquad A^\top \mathbf{e} = \mathbf{0} \ \ (\text{so } \mathbf{e} \perp \mathcal{C}(A)).$$

Projection matrix P (onto $\mathcal{C}(A)$):

$$P = A(A^\top A)^{-1}A^\top = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} & -\frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} & \frac{1}{3} \\ -\frac{1}{3} & \frac{1}{3} & \frac{2}{3} \end{bmatrix}.$$

- Apply to $\mathbf{b}$:

$$P\mathbf{b} = \mathbf{p}.$$

- Properties:

$$P^2 = P, \qquad P^\top = P.$$

- Pythagorean check:

$$\|\mathbf{p}\|^2 = \frac{4+49+25}{9} = \frac{78}{9} = \frac{26}{3}, \quad \|\mathbf{e}\|^2 = \frac{1+1+1}{9} = \frac{1}{3},$$

$$\|\mathbf{p}\|^2 + \|\mathbf{e}\|^2 = \frac{27}{3} = 9 = \|\mathbf{b}\|^2.$$

Quick How-To (recipes)

- **Onto a line $\operatorname{span}\{\mathbf{a}\}$:**

$$\mathbf{p} = \mathbf{a}\,\frac{\mathbf{a}^\top \mathbf{b}}{\mathbf{a}^\top \mathbf{a}}, \quad \mathbf{e} = \mathbf{b} - \mathbf{p}.$$

- **Onto a column space $S = \mathcal{C}(A)$ (full column rank):**

$$\mathbf{p} = A(A^\top A)^{-1}A^\top \mathbf{b} = P\mathbf{b}, \quad P = A(A^\top A)^{-1}A^\top, \quad P^2 = P = P^\top.$$

- **Geometric facts (always):** $\mathbf{b} = \mathbf{p} + \mathbf{e}$ with $\mathbf{p} \in S$, $\mathbf{e} \perp S$, and $\|\mathbf{b}\|^2 = \|\mathbf{p}\|^2 + \|\mathbf{e}\|^2$.

Determinants — Intuition, Definition, and Properties

---

1. Intuitive Meaning

The determinant of a square matrix measures the **scaling factor** of the linear transformation represented by the matrix.

- In **2D**: It represents the **signed area** of the parallelogram spanned by the columns (or rows).
- In **3D**: It represents the **signed volume** of the parallelepiped spanned by the columns (or rows).
- **Sign** indicates orientation:
  - Positive → orientation preserved
  - Negative → orientation reversed
  - Zero → transformation squashes space into a lower dimension (matrix is **singular**, not invertible)

---

2. Mathematical Definition

For a **2×2 matrix**:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad \det(A) = ad - bc$$

For a **3×3 matrix**:

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

Using cofactor expansion along the first row:

$$\det(A) = a(ei - fh) - b(di - fg) + c(dh - eg)$$

---

3. Example Calculations

Example (2D):

$$A = \begin{bmatrix} 3 & 4 \\ 2 & 5 \end{bmatrix}, \quad \det(A) = (3)(5) - (4)(2) = 15 - 8 = 7$$

Interpretation: Area scaled by factor 7, orientation preserved.

Example (3D):

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 5 & 6 & 0 \end{bmatrix}$$

$$\det(A) = 1(1 \cdot 0 - 4 \cdot 6) - 2(0 \cdot 0 - 4 \cdot 5) + 3(0 \cdot 6 - 1 \cdot 5)$$

$$\det(A) = 1(0 - 24) - 2(0 - 20) + 3(0 - 5)$$

$$\det(A) = -24 + 40 - 15 = 1$$

Interpretation: Volume preserved (factor 1), orientation preserved.

---

4. Five Key Properties of Determinants

1. **Effect of Row Swap**: Swapping two rows (or columns) changes the sign of the determinant.
2. **Row/Column Multiplication**: Multiplying a row (or column) by k multiplies the determinant by k.
3. **Row Addition**: Adding a multiple of one row to another does not change the determinant.
4. **Triangular Matrices**: The determinant of a triangular matrix is the product of its diagonal entries.
5. **Invertibility**: A matrix is invertible **iff** $\det(A) \neq 0$.

---

5. Other Properties (List Only)

- $\det(AB) = \det(A) \cdot \det(B)$
- $\det(A^T) = \det(A)$
- $\det(A^{-1}) = \frac{1}{\det(A)}$ (if invertible)
- $\det(kA) = k^n \det(A)$ for $n \times n$ matrix
- Determinant of orthogonal matrix is ± 1
- Determinant of projection matrix ≤ 1
- Determinant is multilinear and alternating in its rows/columns

---

Invertibility of a Matrix

---

1. Definition

A square matrix A is **invertible** (or **nonsingular**) if there exists another matrix A^{-1} such that:

$$A \cdot A^{-1} = A^{-1} \cdot A = I$$

where I is the identity matrix of the same size.

---

2. Conditions for Invertibility

A square matrix A is invertible **iff** any (and hence all) of the following equivalent conditions hold:

1. $\det(A) \neq 0$
2. A has **full rank** ($\mathrm{rank}(A) = n$ for an $n \times n$ matrix)
3. The **columns** (and rows) of A are **linearly independent**
4. The linear transformation $T(\mathbf{x}) = A\mathbf{x}$ is **bijective** (both one-to-one and onto)
5. The **null space** of A contains **only** the zero vector:

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

---

3. What Makes a Matrix Singular

A matrix is **singular** (non-invertible) if **any** of the following are true:

- $\det(A) = 0$
- $\mathrm{rank}(A) < n$ (columns/rows are linearly dependent)
- At least one row (or column) is a linear combination of the others
- The transformation $A\mathbf{x}$ squashes space into a lower dimension (e.g., a 3D object becomes flat, or a 2D shape becomes a line/point)

---

4. Geometric Intuition

- In 2D: An invertible matrix maps a unit square to a parallelogram with **nonzero area**.
- In 3D: An invertible matrix maps a unit cube to a parallelepiped with **nonzero volume**.
- A singular matrix collapses the shape into a lower dimension (area/volume = 0).

---

5. Quick Checks for Invertibility

- Check determinant: if $\det(A) \neq 0$, invertible.
- Check rank: if $\mathrm{rank}(A) = n$, invertible.
- Try solving $A\mathbf{x} = \mathbf{b}$: if every $\mathbf{b}$ has a unique solution, invertible.

Example

Invertible Matrix:

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}, \quad \det(A) = 2(1) - 1(1) = 1 \neq 0$$

Singular Matrix:

$$B = \begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix}, \quad \det(B) = 2(2) - 4(1) = 0$$

(Second column is $2\times$ the first column → columns are linearly dependent)

Properties of Inverse and Transpose of a Product

---

1) Inverse of a Product

For two invertible square matrices A and B:

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

⚠️ The order is reversed.

---

2) Transpose of a Product

For any conformable matrices A and B:

$$(AB)^T = B^T A^T$$

⚠️ The order is reversed.

Inverse of 2 by 2 Matrix — By Hand

---

1. Formula

For a matrix:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

The inverse exists **iff**:

$$\det(A) = ad - bc \neq 0$$

The formula is:

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

---

2. Steps to Calculate

1. **Compute the determinant**:

$$\det(A) = ad - bc$$

   If $\det(A) = 0$, the matrix is singular and has **no inverse**.

2. **Swap the diagonal elements** (a \leftrightarrow d).

3. **Change the sign** of the off-diagonal elements (b) and (c).

4. **Multiply** the resulting matrix by (\frac{1}{\det(A)}).

---

3. Example

Let:

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

Step 1 — Determinant:

$$\det(A) = (3)(5) - (4)(2) = 15 - 8 = 7$$

Step 2 — Swap diagonals:

$$\begin{bmatrix} 5 & 4 \\ 2 & 3 \end{bmatrix} \quad \longrightarrow \quad \begin{bmatrix} 5 & 4 \\ 2 & 3 \end{bmatrix} \text{(swap } 3 \leftrightarrow 5\text{)}$$

Step 3 — Change signs of off-diagonals:

$$\begin{bmatrix} 5 & -4 \\ -2 & 3 \end{bmatrix}$$

Step 4 — Multiply by (1/7):

$$A^{-1} = \frac{1}{7} \begin{bmatrix} 5 & -4 \\ -2 & 3 \end{bmatrix}$$

Final Result:

$$A^{-1} = \begin{bmatrix} \frac{5}{7} & -\frac{4}{7} \\ -\frac{2}{7} & \frac{3}{7} \end{bmatrix}$$

Verification:

$$A \cdot A^{-1} = I$$

where (I) is the (2\times 2) identity matrix.

Solving Linear Systems

Row Reduction, REF, and RREF — Definitions and Uses

Q1: What is Row Reduction (Gaussian Elimination)?

Definition:
Row reduction (Gaussian elimination) is a sequence of operations applied to the rows of a matrix to simplify it.
The goal is to transform the matrix into a **Row Echelon Form (REF)** or **Reduced Row Echelon Form (RREF)** using **elementary row operations**:

1. Swap two rows
2. Multiply a row by a nonzero scalar
3. Add (or subtract) a multiple of one row to another

Q2: What is Row Echelon Form (REF)?

A matrix is in **Row Echelon Form** if:

1. All nonzero rows are above any rows of all zeros.
2. The leading entry (pivot) of each nonzero row is to the right of the leading entry in the row above it.
3. All entries below each pivot are zero.

Example (REF):

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 5 \end{bmatrix}$$

Q3: What is Reduced Row Echelon Form (RREF)?

A matrix is in **RREF** if:

1. It is in **REF**.
2. Each pivot is equal to 1.
3. Each pivot is the only nonzero entry in its column.

Example (RREF):

$$\begin{bmatrix} 1 & 0 & -5 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix}$$

Q4: How to use REF/RREF for solving linear systems?

Given a system $\mathbf{A}\mathbf{x} = \mathbf{b}$:

1. Form the **augmented matrix** $[\mathbf{A} \mid \mathbf{b}]$.
2. Apply **row reduction** to get REF or RREF.
3. Use:
   - **REF** → solve by back-substitution.
   - **RREF** → read the solution directly (pivots give fixed variables, free columns give parameters).

Example: System:

$$\begin{cases} x + y + z = 6 \\ 2y + 5z = -4 \\ 2x + 5y - z = 27 \end{cases}$$

Augmented matrix:

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 6 \\ 0 & 2 & 5 & -4 \\ 2 & 5 & -1 & 27 \end{array}\right]$$

Row reduce → RREF → read solution. **see the full steps in the section below**

Q5: How to use RREF to check linear independence?

For vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$:

1. Place them as **columns** in a matrix.
2. Row reduce to RREF.

3. If **every column has a pivot** → vectors are **linearly independent**.
   If **any column lacks a pivot** → vectors are **dependent**.

Example:

$$\mathbf{V} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 5 \end{bmatrix}$$

RREF shows 3 pivots → independent. **see the full steps in the section below**

Q6: How to check if a matrix is full rank using RREF?

- Let $\mathbf{A}$ be $m \times n$.
- Row reduce $\mathbf{A}$ to RREF.
- **Full rank** means:

$$\operatorname{rank}(\mathbf{A}) = \min(m, n)$$

  i.e., the number of pivots equals the smallest dimension.

Example:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

- $m = 2, n = 3, \min(m, n) = 2$
- RREF has 2 pivots → full rank.

see the full steps in the section below

Row Reduction, REF, and RREF — Step-by-Step Examples

---

1. Solving a Linear System using Row Reduction

We solve:

$$\begin{cases} x + y + z = 6 \\ 2y + 5z = -4 \\ 2x + 5y - z = 27 \end{cases}$$

Step 1 — Write the Augmented Matrix

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 6 \\ 0 & 2 & 5 & -4 \\ 2 & 5 & -1 & 27 \end{array}\right]$$

Step 2 — Eliminate below the first pivot (a_{11})

$R_3 \leftarrow R_3 - 2R_1$

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 6 \\ 0 & 2 & 5 & -4 \\ 0 & 3 & -3 & 15 \end{array}\right]$$

Step 3 — Make pivot at a_{22} equal to 1

$R_2 \leftarrow R_2/2$

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 6 \\ 0 & 1 & 2.5 & -2 \\ 0 & 3 & -3 & 15 \end{array}\right]$$

Step 4 — Eliminate below a_{22}

$R_3 \leftarrow R_3 - 3R_2$

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 6 \\ 0 & 1 & 2.5 & -2 \\ 0 & 0 & -10.5 & 21 \end{array}\right]$$

Step 5 — Make pivot at a_{33} equal to 1

$R_3 \leftarrow R_3/(-10.5)$

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 6 \\ 0 & 1 & 2.5 & -2 \\ 0 & 0 & 1 & -2 \end{array}\right]$$

Step 6 — Back-substitution

From last row: $z = -2$
From second row: $y + 2.5(-2) = -2 \implies y = 3$
From first row: $x + 3 + (-2) = 6 \implies x = 5$

Final solution:

$$x = 5, \quad y = 3, \quad z = -2$$

---

2. Checking Linear Independence

We check if:

$$\mathbf{v}_1 = (1, 2, 3), \quad \mathbf{v}_2 = (4, 5, 6), \quad \mathbf{v}_3 = (7, 8, 9)$$

are independent.

Step 1 — Form matrix with vectors as columns

$$\mathbf{V} = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}$$

Step 2 — Row reduction to RREF

$R_2 \leftarrow R_2 - 2R_1$, $R_3 \leftarrow R_3 - 3R_1$

$$\begin{bmatrix} 1 & 4 & 7 \\ 0 & -3 & -6 \\ 0 & -6 & -12 \end{bmatrix}$$

$R_3 \leftarrow R_3 - 2R_2$

$$\begin{bmatrix} 1 & 4 & 7 \\ 0 & -3 & -6 \\ 0 & 0 & 0 \end{bmatrix}$$

$R_2 \leftarrow R_2/(-3)$

$$\begin{bmatrix} 1 & 4 & 7 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

Eliminate above a_{22}: $R_1 \leftarrow R_1 - 4R_2$

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

Step 3 — Count pivots

We have only 2 pivots (in col 1 and col 2) → **not full pivot coverage** → **vectors are linearly dependent**.

---

3. Checking if a Matrix is Full Rank

Matrix:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

Step 1 — Row reduction

$R_2 \leftarrow R_2 - 4R_1$

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \end{bmatrix}$$

$R_2 \leftarrow R_2/(-3)$

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

Eliminate above a_{22}: $R_1 \leftarrow R_1 - 2R_2$

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{bmatrix}$$

Step 2 — Count pivots

We have 2 pivots.
Since $\min(m, n) = \min(2, 3) = 2$, rank = 2 = full rank.

LU Decomposition

---

1) What is LU Decomposition?

LU decomposition factors a square matrix A into

$$A = LU$$

where L is **lower triangular** (with 1's on the diagonal) and U is **upper triangular**.
This is useful because solving $Ax = b$ becomes two easier steps: forward substitution with L then back substitution with U.

---

2) General Steps (without pivoting)

1. Start with A.
2. Perform Gaussian elimination to create zeros **below** each pivot.
3. The matrix you end with is U.
4. The elimination **multipliers** you used are the subdiagonal entries of L (with L having 1's on the diagonal).

---

3) Example

We decompose

$$A = \begin{bmatrix} 2 & 4 & 3 & 5 \\ -4 & -7 & -5 & -8 \\ 6 & 8 & 2 & 9 \\ 4 & 9 & -2 & 14 \end{bmatrix}.$$

Step 1: First pivot ($u_{11} = 2$)

Multipliers:

$$\ell_{21} = -2, \quad \ell_{31} = 3, \quad \ell_{41} = 2.$$

Step 2: Second pivot ($u_{22} = 1$)

Multipliers:

$$\ell_{32} = -4, \quad \ell_{42} = 1.$$

Step 3: Third pivot ($u_{33} = -3$)

Multiplier:

$$\ell_{43} = 3.$$

Final result

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ 3 & -4 & 1 & 0 \\ 2 & 1 & 3 & 1 \end{bmatrix}, \qquad U = \begin{bmatrix} 2 & 4 & 3 & 5 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & -3 & 2 \\ 0 & 0 & 0 & -4 \end{bmatrix}, \qquad A = LU.$$

---

4) Pivoting note

Sometimes a pivot is 0 or very small. In that case we first **permute** rows or columns using a **permutation matrix** P, then factor:

$$PA = LU.$$

This makes the process feasible and more numerically stable.

QR Decomposition — Step-by-Step

---

1. Definition

The **QR decomposition** factors a matrix $\mathbf{A}$ into:

$$\mathbf{A} = \mathbf{QR}$$

where:

- $\mathbf{Q}$ is an **orthogonal** (or unitary in complex case) matrix:

$$\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$$

  Columns of $\mathbf{Q}$ are orthonormal basis vectors.
- $\mathbf{R}$ is an **upper triangular** matrix.

---

2. Why is QR useful?

- **Solving linear systems** more stably than Gaussian elimination.
- **Least squares problems**: $\min \|Ax - b\|$.
- **Eigenvalue algorithms** (QR algorithm).
- Useful in **orthogonalization** and **numerical stability**.

---

3. Manual QR decomposition via Gram–Schmidt Process

We decompose:

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Step 1 — Take the first column as a_1

$$a_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

Normalize to get q_1:

$$\|a_1\| = \sqrt{1^2 + 1^2 + 0^2} = \sqrt{2}$$

$$q_1 = \frac{a_1}{\|a_1\|} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

Step 2 — Take the second column as a_2

$$a_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

Project a_2 onto q_1:

$$\text{proj}_{q_1}(a_2) = (q_1^T a_2) q_1$$

$$q_1^T a_2 = \frac{1}{\sqrt{2}}(1 \cdot 1 + 1 \cdot 0 + 0 \cdot 1) = \frac{1}{\sqrt{2}}$$

$$\text{proj}_{q_1}(a_2) = \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

Step 3 — Subtract projection to orthogonalize

$$u_2 = a_2 - \text{proj}_{q_1}(a_2) = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 0.5 \\ 0.5 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.5 \\ -0.5 \\ 1 \end{bmatrix}$$

Step 4 — Normalize u_2 to get q_2

$$\|u_2\| = \sqrt{0.5^2 + (-0.5)^2 + 1^2} = \sqrt{0.25 + 0.25 + 1} = \sqrt{1.5}$$

$$q_2 = \frac{u_2}{\|u_2\|} = \frac{1}{\sqrt{1.5}} \begin{bmatrix} 0.5 \\ -0.5 \\ 1 \end{bmatrix}$$

Step 5 — Build Q

$$\mathbf{Q} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{0.5}{\sqrt{1.5}} \\ \frac{1}{\sqrt{2}} & \frac{-0.5}{\sqrt{1.5}} \\ 0 & \frac{1}{\sqrt{1.5}} \end{bmatrix}$$

Step 6 — Build R

We have:

$$R = Q^T A$$

Compute:

$$R = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{0.5}{\sqrt{1.5}} & \frac{-0.5}{\sqrt{1.5}} & \frac{1}{\sqrt{1.5}} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} \\ 0 & \sqrt{1.5} \end{bmatrix}$$

Final QR decomposition

$$\mathbf{A} = \mathbf{QR}$$

where:

$$\mathbf{Q} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{0.5}{\sqrt{1.5}} \\ \frac{1}{\sqrt{2}} & \frac{-0.5}{\sqrt{1.5}} \\ 0 & \frac{1}{\sqrt{1.5}} \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} \\ 0 & \sqrt{1.5} \end{bmatrix}$$

---

4. Notes

- In numerical computing, **modified Gram–Schmidt** is used for stability.
- Alternatively, **Householder reflections** can be used for QR in fewer operations.

Relationship Between Gram–Schmidt and QR Decomposition

---

1. QR Decomposition Recap

For a matrix:

$$A \in \mathbb{R}^{m \times n}, \quad m \geq n$$

The QR decomposition factors (A) as:

$$A = QR$$

where:

- (Q) is an (m \times n) matrix with **orthonormal columns** ((Q^T Q = I))
- (R) is an (n \times n) **upper triangular** matrix

---

2. Gram–Schmidt Process

The **Gram–Schmidt orthogonalization** takes a set of linearly independent vectors:

$$\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$$

and constructs an orthonormal set:

$$\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n\}$$

such that:

- Each (\mathbf{q}_i) is **orthogonal** to all previous (\mathbf{q}_j) ((j < i))
- Each (\mathbf{q}_i) has unit length

Classical Gram–Schmidt formulas:

1. Start:

$$\mathbf{u}_1 = \mathbf{a}_1$$

$$\mathbf{q}_1 = \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|}$$

2. For (k = 2, \dots, n):

$$\mathbf{u}_k = \mathbf{a}_k - \sum_{j=1}^{k-1} \mathrm{proj}_{\mathbf{q}_j}(\mathbf{a}_k)$$

where:

$$\mathrm{proj}_{\mathbf{q}_j}(\mathbf{a}_k) = \left(\mathbf{q}_j^T \mathbf{a}_k\right) \mathbf{q}_j$$

$$\mathbf{q}_k = \frac{\mathbf{u}_k}{\|\mathbf{u}_k\|}$$

---

3. How Gram–Schmidt Produces QR

- The orthonormal vectors (\mathbf{q}_1, \dots, \mathbf{q}_n) form the **columns of (Q)**.
- The coefficients from the projection steps form the **entries of (R)**.

Specifically:

$$R_{jk} = \mathbf{q}_j^T \mathbf{a}_k, \quad j \le k$$

$$Q = [\mathbf{q}_1 \; \mathbf{q}_2 \; \dots \; \mathbf{q}_n]$$

$$A = QR$$

---

4. Example

Let:

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Step 1 — Gram–Schmidt:

1. (\mathbf{a}_1 = (1,1,0)^T)

$$\mathbf{q}_1 = \frac{(1,1,0)}{\sqrt{2}}$$

2. (\mathbf{a}_2 = (1,0,1)^T)

Remove projection on (\mathbf{q}_1):

$$\mathbf{u}_2 = (1,0,1) - \frac{\sqrt{2}}{2}(1,1,0) = (0.5,-0.5,1)$$

Normalize:

$$\mathbf{q}_2 = \frac{(0.5,-0.5,1)}{\sqrt{1.5}}$$

Step 2 — Build (R):

$$R = \begin{bmatrix} \mathbf{q}_1^T\mathbf{a}_1 & \mathbf{q}_1^T\mathbf{a}_2 \\ 0 & \mathbf{q}_2^T\mathbf{a}_2 \end{bmatrix}$$

Step 3 — Verify:

$$A = QR$$

---

5. Summary

- Gram–Schmidt **orthonormalizes** the columns of (A) → these become (Q).
- The **projection coefficients** fill in the upper triangular matrix (R).
- Therefore, **Gram–Schmidt is a constructive algorithm for QR decomposition**.

Eigenvalues, Eigenvectors, and Matrix Diagonalization

1. Concept & Intuition

- **Eigenvector** of a square matrix A: a non-zero vector $\mathbf{v}$ whose direction does not change when A is applied — it only scales.
- **Eigenvalue** λ: the scalar factor by which the eigenvector is stretched or shrunk.

Mathematically:

$$A\mathbf{v} = \lambda\mathbf{v}, \quad \mathbf{v} \neq \mathbf{0}$$

Analogy:
Imagine a rubber sheet with arrows (vectors) drawn on it. When the sheet is stretched by A, most arrows change **both** length and direction — except a few that change **only** in length.
These special arrows are **eigenvectors**, and the scale factor is the **eigenvalue**.

---

2. Finding Eigenvalues & Eigenvectors — General Steps

Given an $n \times n$ matrix A:

Step 1 — Eigenvalues

1. Start with:

$$A\mathbf{v} = \lambda\mathbf{v}$$

2. Rewrite:

$$(A - \lambda I)\mathbf{v} = 0$$

3. For a **non-zero $\mathbf{v}$**, the determinant must vanish:

$$\det(A - \lambda I) = 0$$

4. Solve this **characteristic polynomial** for λ.

Step 2 — Eigenvectors

For each eigenvalue λ:

1. Solve:

$$(A - \lambda I)\mathbf{v} = 0$$

2. This gives the eigenvector(s) (up to scalar multiples).

Example

Let:

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

Step 1 — Eigenvalues:

$$\det(A - \lambda I) = \det \begin{bmatrix} 4-\lambda & 1 \\ 2 & 3-\lambda \end{bmatrix} = (4-\lambda)(3-\lambda) - 2$$

$$= (12 - 4\lambda - 3\lambda + \lambda^2) - 2$$

$$= \lambda^2 - 7\lambda + 10 = 0$$

Factor:

$$(\lambda - 5)(\lambda - 2) = 0$$

Thus:

$$\lambda_1 = 5, \quad \lambda_2 = 2$$

Step 2 — Eigenvectors:

- For $\lambda = 5$:

$$A - 5I = \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix}$$

From the first row:

$$-x + y = 0 \quad \Rightarrow \quad y = x$$

Eigenvector:

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

- For $\lambda = 2$:

$$A - 2I = \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix}$$

From the first row:

$$2x + y = 0 \quad \Rightarrow \quad y = -2x$$

Eigenvector:

$$\mathbf{v}_2 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

---

3. Connection to Diagonalization

Diagonalization rewrites A as:

$$A = PDP^{-1}$$

where:

- D is a diagonal matrix with the **eigenvalues** on the diagonal.
- P is a matrix whose columns are the **eigenvectors**.

Interpretation:
In the eigenvector basis, the transformation A looks like **pure scaling**.

---

4. Steps to Diagonalize a Matrix

1. Find all eigenvalues λ_i.
2. Find corresponding linearly independent eigenvectors $\mathbf{v}_i$.
3. Form:

$$P = [\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_n]$$

4. Form:

$$D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$$

5. Verify:

$$A = PDP^{-1}$$

Example (Diagonalizing our A)

We have:

$$\lambda_1 = 5, \quad \lambda_2 = 2$$

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

Step 1:

$$P = \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix}$$

Step 2:

$$D = \begin{bmatrix} 5 & 0 \\ 0 & 2 \end{bmatrix}$$

Step 3: Inverse of P:

$$P^{-1} = \frac{1}{(1)(-2) - (1)(1)} \begin{bmatrix} -2 & -1 \\ -1 & 1 \end{bmatrix} = \frac{1}{-3} \begin{bmatrix} -2 & -1 \\ -1 & 1 \end{bmatrix}$$

$$P^{-1} = \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{1}{3} \end{bmatrix}$$

Step 4: Check:

$$PDP^{-1} = \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{1}{3} \end{bmatrix} = A$$

---

5. Conditions for Diagonalizability

A matrix A is diagonalizable if:

1. It has n **linearly independent eigenvectors** (for an $n \times n$ matrix).

This is always true if:

- A has **distinct eigenvalues** (sufficient but not necessary).
- A is **symmetric** ($A = A^T$) — in this case, A is always diagonalizable with **orthogonal** P.

---

6. Applications of Diagonalization

- **Solving systems of differential equations**: $\mathbf{x}' = A\mathbf{x}$
- **Computing powers of a matrix**:

$$A^k = PD^kP^{-1}$$

Useful for Markov chains.

- **Quantum mechanics**: eigenvalues = energy levels, eigenvectors = states.
- **Principal Component Analysis (PCA)**: diagonalizing covariance matrices to find principal directions.
- **Vibration analysis** in engineering.
- **Graph theory**: using eigenvalues of adjacency matrices for network analysis.

Positive Definite vs Positive Semidefinite Matrices

1. Definitions

Let $A \in \mathbb{R}^{n \times n}$ be a **symmetric** matrix ($A = A^\top$).
We test positivity by the quadratic form:

$$q(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x}, \quad \mathbf{x} \in \mathbb{R}^n.$$

- **Positive Definite (PD):**

$$\mathbf{x}^\top A\mathbf{x} > 0 \quad \forall\, \mathbf{x} \neq \mathbf{0}.$$

- **Positive Semidefinite (PSD):**

$$\mathbf{x}^\top A\mathbf{x} \geq 0 \quad \forall\, \mathbf{x}.$$

---

2. Key Properties

Positive Definite (PD)

- All eigenvalues of A are **strictly positive**.

- A is **invertible** (since $\det(A) \neq 0$).
- Cholesky decomposition exists: $A = LL^\top$ with L invertible.
- Quadratic form is **strictly convex**.

Positive Semidefinite (PSD)

- All eigenvalues of A are **nonnegative** (some may be 0).
- A may be **singular** (not invertible).
- Cholesky decomposition may exist but L may be non-invertible.
- Quadratic form is **convex but not strictly convex** (can be flat in some directions).

---

3. How to Tell Them Apart

Methods

1. **Eigenvalue test:**
   - PD $\iff$ all eigenvalues $\lambda_i > 0$.
   - PSD $\iff$ all eigenvalues $\lambda_i \geq 0$.
2. **Principal minors test (Sylvester's criterion):**
   - PD $\iff$ all leading principal minors > 0.
   - PSD $\iff$ all principal minors ≥ 0 (but need to check all, not only leading).
3. **Quadratic form check:**
   - Compute $\mathbf{x}^\top A\mathbf{x}$ for various $\mathbf{x}$.
   - If always > 0, PD; if always ≥ 0, PSD.

---

4. Comparison Table

Property	Positive Definite (PD)	Positive Semidefinite (PSD)
Quadratic form	$\mathbf{x}^\top A\mathbf{x} > 0$ for all $\mathbf{x} \neq 0$	$\mathbf{x}^\top A\mathbf{x} \geq 0$ for all $\mathbf{x}$
Eigenvalues	All > 0	All ≥ 0
Invertibility	Always invertible	May be singular
Determinant	> 0	≥ 0 (possibly 0)
Cholesky	Exists, with nonsingular L	May exist, but L can be singular
Convexity	Strictly convex quadratic form	Convex (flat directions possible)

---

5. Examples

Example 1 — PD

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}, \qquad \lambda_1 = 2 > 0,\ \lambda_2 = 3 > 0.$$

So A is **positive definite**.

Quadratic form:

$$\mathbf{x}^\top A\mathbf{x} = 2x_1^2 + 3x_2^2 > 0 \quad \forall (x_1, x_2) \neq (0, 0).$$

Example 2 — PSD

$$B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \qquad \lambda_1 = 1 > 0,\ \lambda_2 = 0.$$

So B is **positive semidefinite** (not definite).

Quadratic form:

$$\mathbf{x}^\top B\mathbf{x} = x_1^2 \geq 0,$$

but for $\mathbf{x} = (0, 1)$, $\mathbf{x}^\top B\mathbf{x} = 0$.

---

6. Summary

- **PD $\Rightarrow$ strictly positive eigenvalues, invertible, strictly convex.**
- **PSD $\Rightarrow$ nonnegative eigenvalues, possibly singular, convex but not strict.**
- Always check **eigenvalues** or **quadratic form** to decide.

Singular Value Decomposition (SVD): Definition, Existence Proof, PSD links, and Example

1) What is the SVD?

For any real matrix $A \in \mathbb{R}^{m\times n}$, the **Singular Value Decomposition (SVD)** factors A as

$$\boxed{A = U\,\Sigma\,V^\top}$$

where

- $U \in \mathbb{R}^{m\times m}$ is **orthogonal** ($U^\top U = I_m$). Its columns $u_1, \ldots, u_m$ are **left singular vectors**.
- $V \in \mathbb{R}^{n\times n}$ is **orthogonal** ($V^\top V = I_n$). Its columns $v_1, \ldots, v_n$ are **right singular vectors**.
- $\Sigma \in \mathbb{R}^{m\times n}$ is **diagonal (rectangular)** with nonnegative entries:

$$\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0), \quad \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0, \quad r = \operatorname{rank}(A).$$

The σ_i's are the **singular values** of A.

Thin (economy) SVD. Often we use $U_r \in \mathbb{R}^{m\times r},\ \Sigma_r \in \mathbb{R}^{r\times r},\ V_r \in \mathbb{R}^{n\times r}$ with

$$A = U_r\,\Sigma_r\,V_r^{\top}, \qquad U_r^{\top}U_r = I_r, \quad V_r^{\top}V_r = I_r.$$

---

2) Why does every matrix have an SVD? (Existence)

Key PSD fact (used repeatedly)

Both $A^{\top}A \in \mathbb{R}^{n\times n}$ and $AA^{\top} \in \mathbb{R}^{m\times m}$ are **symmetric positive semidefinite (PSD)** because for all vectors x and y,

$$x^{\top}(A^{\top}A)\,x = \|Ax\|^2 \geq 0, \qquad y^{\top}(AA^{\top})\,y = \|A^{\top}y\|^2 \geq 0.$$

Therefore, by the **spectral theorem**, each admits an orthonormal eigenbasis and a diagonalization with **nonnegative** eigenvalues.

Constructive proof

1. **Diagonalize $A^{\top}A$ (symmetric PSD):**

   $$A^{\top}A = V\,\Lambda\,V^{\top},$$

   where V is orthogonal and $\Lambda = \operatorname{diag}(\lambda_1,\ldots,\lambda_n)$ with $\lambda_i \geq 0$.

2. **Define singular values:**

   $$\sigma_i = \sqrt{\lambda_i} \geq 0.$$

   Let $r = \#\{i : \sigma_i > 0\} = \operatorname{rank}(A)$.

3. **Define right singular vectors:** columns of V are eigenvectors of $A^{\top}A$. For each $i \leq r$ (nonzero σ_i), set v_i = i-th column of V.

4. **Define left singular vectors for nonzero σ_i:**

   $$u_i \;=\; \frac{Av_i}{\sigma_i}, \qquad i = 1,\ldots,r.$$

   These satisfy $\|u_i\| = 1$ and are orthonormal.

5. **Complete to orthonormal bases:** Extend $\{u_1,\ldots,u_r\}$ to an orthonormal basis of $\mathbb{R}^m$ to form U, and $\{v_1,\ldots,v_r\}$ to an orthonormal basis of $\mathbb{R}^n$ to form V (these extra vectors correspond to $\sigma = 0$).

6. **Form Σ and conclude:** With Σ placing $\sigma_1,\ldots,\sigma_r$ on the diagonal (and zeros elsewhere),

   $$\boxed{A = U\,\Sigma\,V^{\top}}.$$

(You can verify $Av_i = \sigma_i u_i$ and $A^\top u_i = \sigma_i v_i$; columns along the zero singular values map to 0.)

This proves **existence** for **every** real matrix A.

---

3) The "Proof of the SVD" in your image — step-by-step (with equations)

The image argues via the connections among A, $A^\top A$, and $AA^\top$ **assuming** an SVD and then identifying what U, Σ, V must be.

Goal: $A = U\Sigma V^\top$.

1. **Form symmetric matrices $A^\top A$ and $AA^\top$:**

$$A^\top A = (V\Sigma^\top U^\top)(U\Sigma V^\top) = V\,\Sigma^\top\Sigma\,V^\top \quad (\text{because } U^\top U = I) \tag{8}$$

$$AA^\top = (U\Sigma V^\top)(V\Sigma^\top U^\top) = U\,\Sigma\Sigma^\top\,U^\top \quad (\text{because } V^\top V = I) \tag{9}$$

2. **Both right-hand sides are of the spectral form $Q\Lambda Q^\top$:**
   - $A^\top A = V(\Sigma^\top\Sigma)V^\top$ is symmetric.
   - $AA^\top = U(\Sigma\Sigma^\top)U^\top$ is symmetric.
3. **Identify eigenvectors and eigenvalues:**
   - Columns of V are **orthonormal eigenvectors of $A^\top A$**.
   - Columns of U are **orthonormal eigenvectors of $AA^\top$**.
   - The **nonzero eigenvalues** of both $A^\top A$ and $AA^\top$ are the **same** and equal to σ_i^2 (the squares of singular values).

   In short (as boxed in the image):

   - V contains orthonormal eigenvectors of $A^\top A$.
   - U contains orthonormal eigenvectors of $AA^\top$.
   - $\sigma_1^2, \dots, \sigma_r^2$ are the **nonzero** eigenvalues of both $A^\top A$ and $AA^\top$.
4. **Which matrices are PSD here?**
   - $\Sigma^\top\Sigma$ and $\Sigma\Sigma^\top$ are **diagonal with nonnegative entries σ_i^2**, hence **PSD**.
   - Congruences by orthogonal matrices preserve PSD, so $A^\top A = V(\Sigma^\top\Sigma)V^\top$ and $AA^\top = U(\Sigma\Sigma^\top)U^\top$ are also **PSD**.

This exactly matches the structure and statements in your picture.

---

4) Useful identities that follow from SVD

For $i \le r$:

$$Av_i = \sigma_i u_i, \qquad A^\top u_i = \sigma_i v_i,$$

and

$$A^\top A\, v_i = \sigma_i^2 v_i, \qquad AA^\top\, u_i = \sigma_i^2 u_i.$$

So v_i's are eigenvectors of $A^\top A$ and u_i's are eigenvectors of $AA^\top$.

---

5) Fully worked numeric example (with a zero singular value)

Let

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{2\times 2}.$$

Step 1 — Compute $A^\top A$ and its eigendecomposition

$$A^\top A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}.$$

- Characteristic polynomial: $\lambda(\lambda - 5) = 0 \;\Rightarrow\; \lambda_1 = 5,\ \lambda_2 = 0.$
- Unit eigenvector for $\lambda_1 = 5$: solve $(A^\top A - 5I)v = 0$ gives $v_1 \propto (1, 2)$. Normalize:

$$v_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

- Unit eigenvector for $\lambda_2 = 0$: solve $A^\top A\, v = 0$ gives $v_2 \propto (-2, 1)$. Normalize:

$$v_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} -2 \\ 1 \end{bmatrix}.$$

Set

$$V = \begin{bmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix}, \qquad \Lambda = \operatorname{diag}(5, 0).$$

Step 2 — Singular values

$$\sigma_1 = \sqrt{5}, \qquad \sigma_2 = 0.$$

Thus

$$\Sigma = \operatorname{diag}(\sqrt{5},\, 0).$$

Step 3 — Left singular vectors

For the nonzero singular value,

$$u_1 = \frac{Av_1}{\sigma_1} = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix} = \frac{1}{\sqrt{5}} \begin{bmatrix} \frac{5}{\sqrt{5}} \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Choose any unit vector orthogonal to u_1 as u_2, e.g.

$$u_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Hence

$$U = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2.$$

Step 4 — Verify $A = U\Sigma V^\top$

Compute

$$V^\top = \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix}, \qquad \Sigma V^\top = \begin{bmatrix} \sqrt{5} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix} = A.$$

So indeed $A = U\Sigma V^\top$.

Step 5 — Check the symmetric/PSD relationships (from the image)

- Using Σ and V:

$$V\,\Sigma^\top \Sigma\, V^\top = V \begin{bmatrix} 5 & 0 \\ 0 & 0 \end{bmatrix} V^\top = A^\top A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}.$$

- Using Σ and U:

$$U\,\Sigma\Sigma^\top\, U^\top = \begin{bmatrix} 5 & 0 \\ 0 & 0 \end{bmatrix} = AA^\top.$$

Both $\Sigma^\top\Sigma$ and $\Sigma\Sigma^\top$ are diagonal with nonnegative entries, hence **PSD**, and therefore $A^\top A$ and $AA^\top$ are **PSD** as well.

---

6) Takeaways

- **Every** real matrix A admits $A = U\Sigma V^\top$.
- $A^\top A$ and $AA^\top$ are **symmetric PSD**; their **nonzero eigenvalues** are σ_i^2.
- Columns of V (resp. U) are eigenvectors of $A^\top A$ (resp. $AA^\top$).
- Identities: $Av_i = \sigma_i u_i$ and $A^\top u_i = \sigma_i v_i$.
- Zero singular values correspond to directions mapped to 0; extend U, V orthonormally to complete the SVD.

Covariance Matrix, PSD Property, and PCA (with algorithm + intuition)

1) Covariance matrices are symmetric positive semidefinite (PSD)

Population covariance

For a random vector $X \in \mathbb{R}^d$ with mean $\mu = \mathbb{E}[X]$, the covariance matrix is

$$\Sigma \;=\; \mathbb{E}\big[(X-\mu)(X-\mu)^\top\big] \in \mathbb{R}^{d\times d}.$$

- **Symmetric:** $\Sigma^\top = \mathbb{E}[(X-\mu)(X-\mu)^\top]^\top = \mathbb{E}[(X-\mu)(X-\mu)^\top] = \Sigma$.
- **PSD:** For any $\mathbf{w} \in \mathbb{R}^d$,

$$\mathbf{w}^\top \Sigma\, \mathbf{w} = \mathbf{w}^\top \mathbb{E}\big[(X-\mu)(X-\mu)^\top\big]\mathbf{w} = \mathbb{E}\big[\mathbf{w}^\top (X-\mu)(X-\mu)^\top \mathbf{w}\big] = \mathbb{E}\big[(\mathbf{w}^\top (X-\mu))^2$$

Therefore Σ is **symmetric PSD**.

Sample covariance

Given data matrix $X \in \mathbb{R}^{n\times d}$ (rows are samples), let X_c be **column-centered** (subtract each column mean). The unbiased sample covariance is

$$S \;=\; \frac{1}{n-1}\, X_c^\top X_c.$$

- **Symmetric** (obvious).
- **PSD:** For any $\mathbf{w}$,

$$\mathbf{w}^\top S\, \mathbf{w} = \frac{1}{n-1}\, \mathbf{w}^\top X_c^\top X_c\, \mathbf{w} = \frac{1}{n-1}\, \|X_c \mathbf{w}\|^2 \;\ge\; 0.$$

So S is also **symmetric PSD**.
Because covariance is PSD, all its eigenvalues are **nonnegative**.

---

2) PCA as variance maximization → eigenvalue problem on the covariance

Goal of PCA: Find orthonormal directions (principal components) that **maximize variance** of the projected data.

Let Σ be the (population or sample) covariance. The **first principal component** direction $\mathbf{w}_1$ solves

$$\max_{\|\mathbf{w}\|=1}\; \mathbf{w}^\top \Sigma\, \mathbf{w}.$$

Form the Lagrangian $L(\mathbf{w}, \lambda) = \mathbf{w}^\top \Sigma \mathbf{w} - \lambda(\mathbf{w}^\top \mathbf{w} - 1)$. Stationarity gives

$$\nabla_{\mathbf{w}} L \;=\; 2\Sigma\mathbf{w} - 2\lambda\mathbf{w} \;=\; 0 \quad\Longrightarrow\quad \Sigma\,\mathbf{w} = \lambda\,\mathbf{w}.$$

Thus **optimal directions are eigenvectors of** Σ, with the objective value equal to the eigenvalue.

- The maximizer is the eigenvector for the **largest eigenvalue** λ_1.
- The next components $\mathbf{w}_2, \mathbf{w}_3, \ldots$ solve the same problem with **orthogonality constraints** to the previous ones, yielding the remaining eigenvectors of Σ in **descending** eigenvalue order.

Therefore:

- **1st principal component $\mathbf{w}_1$** = eigenvector of Σ with **largest eigenvalue** λ_1.
- **k-th principal component $\mathbf{w}_k$** = eigenvector with the k-th largest eigenvalue λ_k.
- **Explained variance** of PC k = λ_k; **explained variance ratio** = $\lambda_k / \sum_{j=1}^{d} \lambda_j$.

---

3) PCA algorithm (two equivalent views)

(A) Eigen-decomposition of covariance

1. **Standardize (optional)**: often scale each feature to zero mean and unit variance **if units differ**.
2. **Center** the data: $X \mapsto X_c$ (subtract column means).
3. **Covariance:** $S = \frac{1}{n-1} X_c^\top X_c \in \mathbb{R}^{d \times d}$.
4. **Eigen-decompose:** $S = V \Lambda V^\top$ with $\Lambda = \mathrm{diag}(\lambda_1 \geq \cdots \geq \lambda_d \geq 0)$, $V = [\mathbf{v}_1, \ldots, \mathbf{v}_d]$.
5. **Choose** k (e.g., the smallest k with $\sum_{i=1}^{k} \lambda_i / \sum_{j=1}^{d} \lambda_j \geq$ threshold like 90%).
6. **Project:** low-dimensional representation

$$Z = X_c V_k \in \mathbb{R}^{n \times k}, \quad V_k = [\mathbf{v}_1, \ldots, \mathbf{v}_k].$$

(Rows of Z are the **PC scores**.)

(B) SVD of centered data (numerically preferred)

- Compute the SVD of the centered data:

$$X_c = U \Sigma_{\text{svd}} V^\top, \qquad U^\top U = I, \; V^\top V = I.$$

- Then the covariance is $S = \frac{1}{n-1} X_c^\top X_c = V \left(\frac{\Sigma_{\text{svd}}^2}{n-1} \right) V^\top$.
  - **Right singular vectors** = principal directions (V).
  - **Eigenvalues** of S = $\sigma_{\text{svd},i}^2 / (n-1)$ (squares of singular values scaled by $1/(n-1)$).
- **Scores:** $Z = X_c V_k = U_k \Sigma_{\text{svd},k}$.

---

4) PCA is dimensionality reduction

PCA replaces the original d features with $k \ll d$ **orthogonal** features (PCs) that keep the **largest variance**:

- Approximate reconstruction (rank-k):

$$X_c \;\approx\; Z V_k^\top \;=\; (X_c V_k) V_k^\top .$$

- The best rank-k approximation (in Frobenius norm) is given by **top k singular values/vectors** (Eckart–Young–Mirsky theorem).
  The **reconstruction error** equals $\sum_{i=k+1}^{d} \lambda_i$.

---

5) Why reduce dimensionality?

- **Noise reduction / denoising:** Small-variance directions often capture noise; removing them improves signal-to-noise.
- **Mitigate multicollinearity:** PCs are uncorrelated; models can be more stable.
- **Generalization:** Reduces risk of overfitting in high-d with limited n (curse of dimensionality).
- **Computation & storage:** Fewer features → faster training/inference and smaller models.
- **Visualization:** Project to 2D/3D for inspection and exploratory analysis.
- **Downstream algorithms:** Many methods perform better with compact, informative features.

---

6) Key summary links

- Covariance matrices (population Σ and sample S) are **symmetric PSD**:

$$\mathbf{w}^\top \Sigma \mathbf{w} = \mathrm{Var}(\mathbf{w}^\top X) \ge 0, \qquad \mathbf{w}^\top S \mathbf{w} = \frac{1}{n-1} \|X_c \mathbf{w}\|^2 \ge 0.$$

- PCA chooses eigenvectors of Σ with **largest eigenvalues** to maximize projected variance:

$$\max_{\|\mathbf{w}\|=1} \; \mathbf{w}^\top \Sigma \mathbf{w} \;\Rightarrow\; \Sigma \mathbf{w} = \lambda \mathbf{w}, \quad \mathbf{w} = \text{principal direction},\; \lambda = \text{variance along it.}$$

- **First PC $\mathbf{w}_1$** corresponds to the **largest eigenvalue** λ_1, second PC to λ_2, etc.
- Use **eigendecomposition** of S or **SVD** of X_c to compute PCA in practice.