Linear Algebra for Deep Learning

Praveen Kumar Chandaliya, DoAI

Department of Artificial Intelligence, SVNIT Surat

January 7, 2026

What is a Vector?

Definition

A **vector** is an ordered collection of numbers representing magnitude and direction in space, or features in a dataset.

Applications in Data Science, ML & DL:

- Representing **feature vectors** in supervised learning.
- Encoding **text, images, and audio** as embeddings.
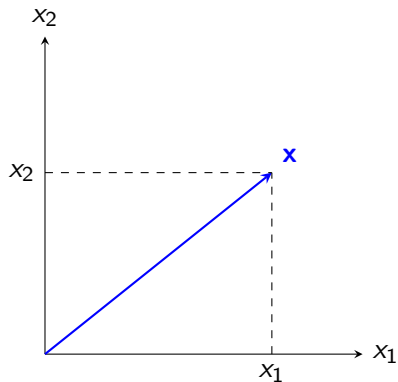- Input to **Neural Networks**.

Vector Notation

Mathematical Representation:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{or} \quad \mathbf{x} = (x_1, x_2, \ldots, x_n)$$

Example (3D Feature Vector):

$$\mathbf{x} = \begin{bmatrix} \text{Height} \\ \text{Weight} \\ \text{Age} \end{bmatrix} = \begin{bmatrix} 170 \\ 65 \\ 25 \end{bmatrix}$$

Vector Visualization

A 2D feature vector (x_1, x_2) in space.

Vectors in $\mathbb{R}^n$

Definition: A vector in $\mathbb{R}^n$ is an ordered *n*-tuple of real numbers:

$$\mathbf{x} = (x_1, x_2, \ldots, x_n), \quad x_i \in \mathbb{R}.$$

Column Vector:

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

- Default form in mathematics and ML for data samples. - Each row is a **feature**.

Row Vector:

$$\mathbf{x}^\top = \begin{bmatrix} x_1 & x_2 & \ldots & x_n \end{bmatrix} \in \mathbb{R}^{1 \times n}$$

- Transpose of the column vector. - Often used for **dataset rows** in tabular form. **ML Context:**

- Column vector: single sample's feature representation.
- Row vector: same sample in dataset matrix format.

Vector: Direction and Magnitude

Definition: A vector has:

- **Magnitude (Length)**: how long the vector is.
- **Direction**: the orientation from the origin.

Magnitude Formula: For $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ in $\mathbb{R}^2$:

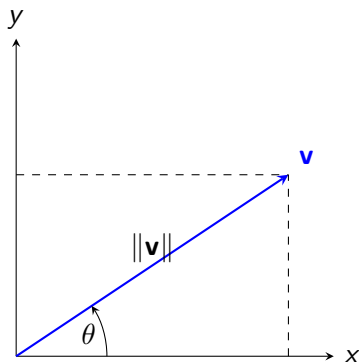$$\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2}$$

In $\mathbb{R}^n$:

$$\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$$

Direction: The unit vector (direction) is:

$$\hat{\mathbf{v}} = \frac{\mathbf{v}}{\|\mathbf{v}\|}$$

Direction and Magnitude
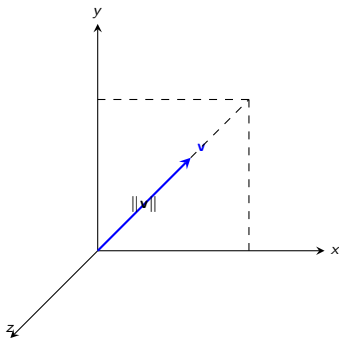
3D Vector: Direction and Magnitude

Definition: A 3D vector $\mathbf{v} = \begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix}$ has:

- **Magnitude (Length):**

$$\|\mathbf{v}\| = \sqrt{v_x^2 + v_y^2 + v_z^2}$$

- **Direction:** represented by the **unit vector**:

$$\hat{\mathbf{v}} = \frac{\mathbf{v}}{\|\mathbf{v}\|}$$

Interpretation: Magnitude = distance from origin; Direction = orientation in 3D space.

Vector Algebra

Given: Two vectors in $\mathbb{R}^n$

$$\mathbf{u} = (u_1, u_2, \ldots, u_n), \quad \mathbf{v} = (v_1, v_2, \ldots, v_n)$$

1. Addition & Subtraction

$$\mathbf{u} + \mathbf{v} = (u_1 + v_1, \ldots, u_n + v_n)$$
$$\mathbf{u} - \mathbf{v} = (u_1 - v_1, \ldots, u_n - v_n)$$

2. Scalar Multiplication

$$c\mathbf{u} = (cu_1, cu_2, \ldots, cu_n)$$

3. Dot Product (Inner Product)

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$$
$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos\theta$$

4. Cross Product (in $\mathbb{R}^3$)

$$\mathbf{u} \times \mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_x & u_y & u_z \\ v_x & v_y & v_z \end{vmatrix}$$

4. Norm (Length)

$$\|\mathbf{u}\| = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2}$$

Angle Between Two Vectors

Definition: The angle θ between two non-zero vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ is given by:

$$\cos \theta = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \, \|\mathbf{v}\|}$$

where:

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$$
$$\|\mathbf{u}\| = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2}$$
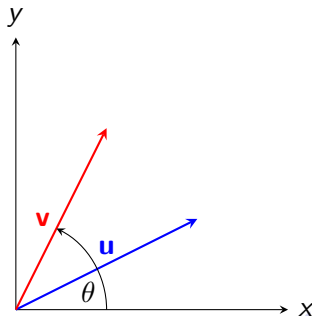
Angle Between Two Vectors

Definition: The angle θ between two non-zero vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ is given by:

$$\cos\theta = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \, \|\mathbf{v}\|}$$

where:

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$$

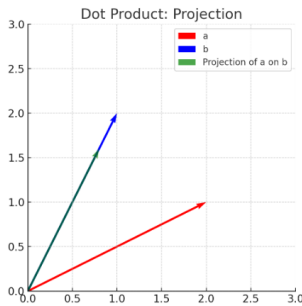$$\|\mathbf{u}\| = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2}$$
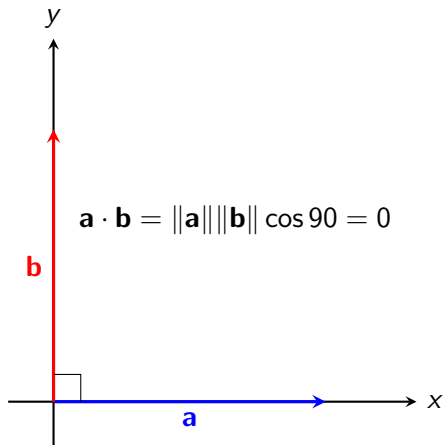
Dot Product (Scalar Product)

Definition

$$\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\|\|\mathbf{b}\| \cos\theta = a_1 b_1 + a_2 b_2 + a_3 b_3$$

- Result: **Scalar**
- Measures projection of one vector onto another
- $\mathbf{a} \cdot \mathbf{b} = 0 \Rightarrow$ Vectors are perpendicular
- Applications: angle between vectors, work ($W = \mathbf{F} \cdot \mathbf{d}$)

Perpendicular Vectors (Dot Product = 0)

$$\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\| \|\mathbf{b}\| \cos 90 = 0$$

Neural-Network Intuition: Orthogonal Weight Vectors

In a neural network, each neuron computes a weighted sum of the input features:

$$z = \mathbf{w}^\top \mathbf{x} + b, \tag{1}$$

where $\mathbf{w} \in \mathbb{R}^n$ denotes the weight vector, $\mathbf{x} \in \mathbb{R}^n$ represents the input feature vector, and b is the bias term.
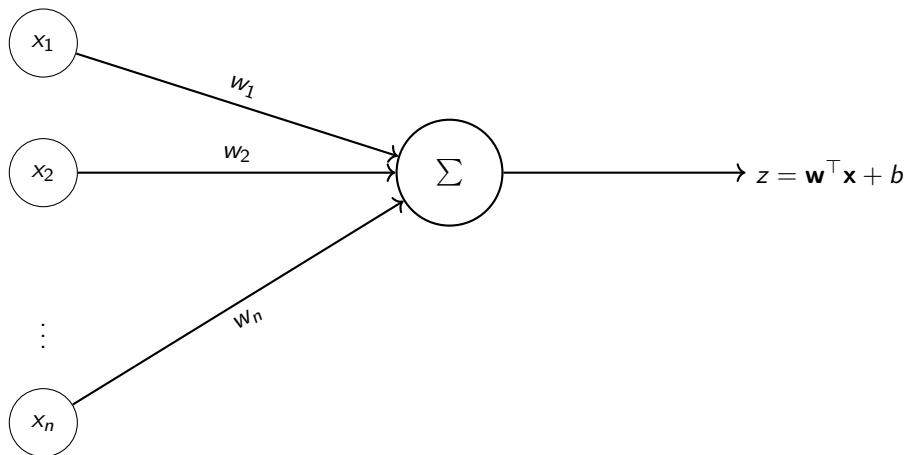
Consider two neurons with weight vectors $\mathbf{w}_1$ and $\mathbf{w}_2$. These weight vectors are said to be *orthogonal* if they satisfy the following condition:

$$\mathbf{w}_1^\top \mathbf{w}_2 = 0. \tag{2}$$

This orthogonality implies that the corresponding neurons respond to **Independent directions** in the feature space, thereby encouraging feature **Decorrelation**.

Orthogonal weight vectors enable neurons to learn independent features, improving representation diversity, optimization stability, and gradient propagation in deep neural networks.

Single Neuron Architecture

- $\mathbf{x} = [x_1, x_2, \ldots, x_n]^\top$ – input vector
- $\mathbf{w} = [w_1, w_2, \ldots, w_n]^\top$ – weight vector
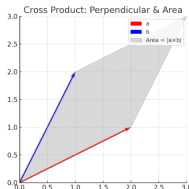- b – bias term

Cross Product (Vector Product)

The *cross product* (or *vector product*) of two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$ is defined

$$\mathbf{a} \times \mathbf{b} = \|\mathbf{a}\|\|\mathbf{b}\| \sin \theta \, \mathbf{n}$$

$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} = \begin{bmatrix} a_2 b_3 - a_3 b_2 \\ a_3 b_1 - a_1 b_3 \\ a_1 b_2 - a_2 b_1 \end{bmatrix}.$$

- Result: **Vector** (perpendicular to $\mathbf{a}$ and $\mathbf{b}$)
- Magnitude = area of parallelogram formed by $\mathbf{a}$ and $\mathbf{b}$
- $\mathbf{a} \times \mathbf{b} = \mathbf{0} \Rightarrow$ Vectors are parallel

Comparison: Dot vs Cross Product

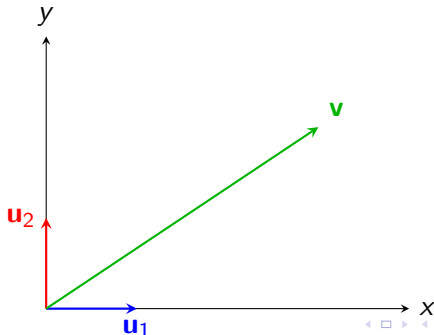| Aspect	Dot Product	Cross Product
Result	Scalar	Vector
Dimension	Any $\mathbb{R}^n$	Only $\mathbb{R}^3$
Geometric Meaning	Projection	Perpendicular vector + Area
Zero Condition	Vectors $\perp$	Vectors $\parallel$
Applications	Angle, gradients	Torque, Normal Vector
Used in	Attention, embedding	3D vision, robotics

Linear Combination of Vectors

Definition: A vector $\mathbf{v}$ is a **linear combination** of vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ if:

$$\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k$$

where $c_i \in \mathbb{R}$ are scalars.

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

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \mathbf{v} = 3\mathbf{u}_1 + 2\mathbf{u}_2 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

Linear Independence

Definition: Vectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ are **linearly independent** if:

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k = \mathbf{0}$$

has only the trivial solution (no solution exists)

$$c_1 = c_2 = \cdots = c_k = 0$$

Intuition:

- No vector in the set can be written as a linear combination of the others.
- Geometrically: - In $\mathbb{R}^2$: independent $\rightarrow$ not on the same line. - In $\mathbb{R}^3$: independent $\rightarrow$ not in the same plane.

Example:

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \text{are independent in } \mathbb{R}^2$$

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \end{bmatrix} \quad \text{are dependent (collinear).}$$

Linearly Independent vs Dependent Vectors

- **Linearly Independent:** $c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k = \mathbf{0}$ only if $c_1 = c_2 = \cdots = c_k = 0$.

- **Linearly Dependent:** At least one vector can be written as a linear combination of the others.

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

$$\underbrace{\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}}_{\text{Independent}}$$

Step 1: Check orthogonality v1.v2=(1)(0)+(0)(2)=0.
Step 2: Test linear independence

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = \mathbf{0}.$$

$$c_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

$$\begin{cases} c_1 = 0, \\ 2c_2 = 0 \implies c_2 = 0. \end{cases}$$

$$\underbrace{\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 2 \\ 4 \end{bmatrix}}_{\text{Dependent: } \mathbf{v}_2 = 2\mathbf{v}_1}$$

Criteria for Linear Independence

Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a set of vectors in $\mathbb{R}^n$. Construct a matrix M by placing these vectors as its columns:

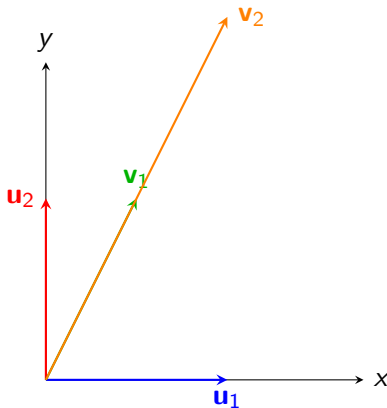$$M = \begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \\ | & | & & | \end{bmatrix}.$$

Then compute the determinant $\det(M)$.

$$\det(M) = \begin{cases} \neq 0, & \text{the vectors are linearly independent,} \\ = 0, & \text{the vectors are linearly dependent.} \end{cases}$$

Note: If the set of vectors forms an orthogonal set (i.e., each pair of vectors in the set is orthogonal to each other), then they are linearly independent.

Visual Representation

Independent vectors point in different directions.
Dependent vectors lie on the same line (collinear).

Orthogonal Vectors

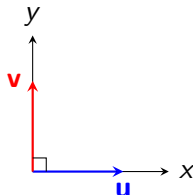**Definition:** Two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are **orthogonal** if:

$$\mathbf{u} \cdot \mathbf{v} = 0$$

which means the angle between them is $\theta = 90°$.

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

$$\mathbf{u} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

$$\mathbf{u} \cdot \mathbf{v} = 1 \cdot 0 + 0 \cdot 3 = 0 \quad \Rightarrow \quad \text{Orthogonal}$$

- Orthogonal feature vectors $\rightarrow$ no linear correlation.
- Feature Decorrelation, Stable Gradient Flow, Deep networks and RNNs, Attention & Subspace Projections

Inverse of a Matrix

Definition: For a square matrix $A \in \mathbb{R}^{n \times n}$, the **inverse** A^{-1} is the matrix such that:

$$AA^{-1} = A^{-1}A = I_n$$

where I_n is the $n \times n$ identity matrix.

Conditions for Invertibility:

- A must be a square matrix ($n \times n$).
- $\det(A) \neq 0$.
- Columns (or rows) of A are linearly independent.

Inverse of a Matrix

Definition: For a square matrix $A \in \mathbb{R}^{n \times n}$, the **inverse** A^{-1} is the matrix such that:

$$AA^{-1} = A^{-1}A = I_n$$

where I_n is the $n \times n$ identity matrix.

Conditions for Invertibility:

- A must be a square matrix ($n \times n$).
- $\det(A) \neq 0$.
- Columns (or rows) of A are linearly independent.

Formula (for 2×2 matrix):

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

Example:

$$A = \begin{bmatrix} 4 & 7 \\ 2 & 6 \end{bmatrix}, \quad \det(A) = 4(6) - 7(2) = 10$$

$$A^{-1} = \frac{1}{10} \begin{bmatrix} 6 & -7 \\ -2 & 4 \end{bmatrix}$$

Singular and Non-Singular Matrices

1. Singular Matrix: A square matrix A is called **singular** if:

$$\det(A) = 0$$

- No inverse exists.
- Columns (or rows) are linearly dependent.
- $Ax = b$ may have no solution or infinitely many solutions.

Singular and Non-Singular Matrices

1. Singular Matrix: A square matrix A is called **singular** if:

$$\det(A) = 0$$

- No inverse exists.
- Columns (or rows) are linearly dependent.
- $Ax = b$ may have no solution or infinitely many solutions.

2. Non-Singular Matrix: A square matrix A is called **non-singular** if:

$$\det(A) \neq 0$$

- Inverse A^{-1} exists.
- Columns (or rows) are linearly independent.
- $Ax = b$ has a unique solution.

Singular and Non-Singular Matrices

Example:

$$A_1 = \begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix}, \quad \det(A_1) = 0 \quad \text{(Singular)}$$

$$A_2 = \begin{bmatrix} 3 & 2 \\ 5 & 4 \end{bmatrix}, \quad \det(A_2) = 3(4) - 2(5) = 2 \neq 0 \quad \text{(Non-Singular)}$$

In Data Science, ML, and DL:

- In deep learning, non-singular matrices preserve information and ensure stable gradient flow (Core layers), while singular matrices collapse dimensions and are useful only when dimensionality reduction is intended (Bottlenecks, pooling)
- Singular matrices appear when features are highly correlated (multicollinearity).
- Non-singular matrices are required for certain matrix inversion operations.

What is a Vector Space?

- A **vector space** V is a set of vectors defined over a field $\mathbb{R}$ (usually real numbers).
- Two operations:
  - **Vector addition** $(+)$
  - **Scalar multiplication** $(\cdot)$
- Must satisfy 6 properties (axioms) to be a vector space.

Property 1 – Addition = Abelian Group

$(V, +)$ must form an **Abelian group**:

- **Closure:** $v, w \in V \Rightarrow v + w \in V$
- **Commutativity:** $v + w = w + v$
- **Associativity:** $(u + v) + w = u + (v + w)$
- **Additive identity:** $\exists\, 0 \in V$ s.t. $v + 0 = v$
- **Additive inverse:** $\forall v, \exists (-v)$ s.t. $v + (-v) = 0$

Properties 2 & 3 – Scalar Multiplication

- **Closure under scalar multiplication:** $a \in \mathbb{R}, v \in V \Rightarrow a \cdot v \in V$
- **Distributivity over vector addition:** $a \cdot (v + w) = a \cdot v + a \cdot w$

- **Distributivity over scalar addition:** $(a + b) \cdot v = a \cdot v + b \cdot v$
- **Associativity of scalar multiplication:** $a \cdot (b \cdot v) = (ab) \cdot v$

Property 6 – Unitary Law

- **Multiplicative identity:** $1 \cdot v = v \quad \forall v \in V$

Summary: If all 6 properties hold, V is a vector space.

Example: $\mathbb{R}^2$ as a Vector Space

$\mathbf{V} = \mathbb{R}^2$ is a vector space over the field of real numbers.

- Let $(a, b), (c, d) \in \mathbb{R}^2$. Then:

$$(a, b) + (c, d) = (a + c, \ b + d) \in \mathbb{R}^2$$

  $\therefore$ Closure property holds in $\mathbb{R}^2$.

- Let $(a, b), (c, d) \in \mathbb{R}^2$. Then:

$$(a, b) + (c, d) = (a + c, \ b + d)$$
$$= (c + a, \ d + b)$$
$$= (c, d) + (a, b)$$

  $\therefore$ Commutativity holds in $\mathbb{R}^2$.

- Let $(a, b), (c, d), (e, f) \in \mathbb{R}^2$. Then:

$$(a, b) + \big((c, d) + (e, f)\big) = (a, b) + (c + e, \ d + f)$$
$$= (a + (c + e), \ b + (d + f))$$
$$= ((a + c) + e, \ (b + d) + f)$$

  $\therefore$ Associativity holds in $\mathbb{R}^2$

- Let $(a, b) \in \mathbb{R}^2$. Then:

$$(0, 0) + (a, b) = (0 + a, \ 0 + b) \quad \text{where } (0, 0) \in \mathbb{R}^2$$

$$= (a, b)$$

$\therefore (0, 0)$ is the **identity element** of $\mathbb{R}^2$.

- Let $(a, b) \in \mathbb{R}^2$. Also $(-a, -b) \in \mathbb{R}^2$. Now:

$$(a, b) + (-a, -b) = (a + (-a), \ b + (-b))$$

$$= (0, 0)$$

$\therefore (-a, -b)$ is the **additive inverse** of (a, b) in $\mathbb{R}^2$.

$\therefore$ The above five properties show that $(\mathbb{R}^2, +)$ is an **Abelian group**.

Example: $\mathbb{R}^2$ – Scalar Multiplication Closure

P2: Let $k \in \mathbb{R}$, $(a, b) \in \mathbb{R}^2$. Then:

$$k(a, b) = (ka, \ kb) \in \mathbb{R}^2$$

$\therefore$ Closure property holds for scalar multiplication in $\mathbb{R}^2$.

P3: Let $k \in \mathbb{R}$, $(a, b), (c, d) \in \mathbb{R}^2$. Then:

$$\begin{aligned}
k\big((a, b) + (c, d)\big) &= k(a + c, \ b + d) \\
&= \big(k(a + c), \ k(b + d)\big) \\
&= (ka + kc, \ kb + kd) \\
&= (ka, kb) + (kc, kd) \\
&= k(a, b) + k(c, d)
\end{aligned}$$

P4: Let $k, m \in \mathbb{R}$, $(a, b) \in \mathbb{R}^2$. Then:

$$\begin{aligned}
(k + m)(a, b) &= ((k + m)a, \ (k + m)b) \\
&= (ka + ma, \ kb + mb) \\
&= (ka, kb) + (ma, mb)
\end{aligned}$$

Continued...

P5: Let $k, m \in \mathbb{R}$, $(a, b) \in \mathbb{R}^2$. Then:

$$k\big(m(a, b)\big) = k(ma, mb)$$

$$= \big(k(ma), \ k(mb)\big)$$

$$= \big((km)a, \ (km)b\big)$$
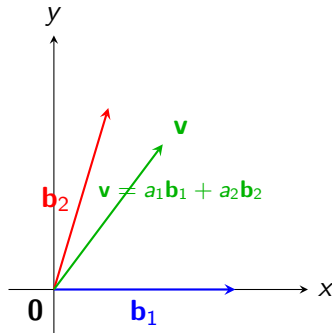
$$= (km)(a, b)$$

P6: Let $(a, b) \in \mathbb{R}^2$. Then:

$$1(a, b) = (1a, \ 1b)$$

$$= (a, b)$$

Since P1–P6 are satisfied, therefore $(\mathbb{R}^2, +, \cdot)$ is a vector space.

Basis in Linear Algebra

A **basis** of a vector space V is a set of vectors $\{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ such that:

- The vectors are **linearly independent**.
- They **span** V (every vector in V can be expressed as a linear combination of them).

Caption: A basis $\{\mathbf{b}_1, \mathbf{b}_2\}$ spans $\mathbb{R}^2$. Every vector $\mathbf{v}$ in the space is a unique linear combination of basis vectors.

Basis Example

If V has a basis of n vectors, then $\dim(V) = n$.
Example: Consider $\mathbb{R}^3$ with:

$$B = \{\mathbf{e}_1 = (1, 0, 0), \ \mathbf{e}_2 = (0, 1, 0), \ \mathbf{e}_3 = (0, 0, 1)\}.$$

- B is **linearly independent** — the only solution to
  $c_1\mathbf{e}_1 + c_2\mathbf{e}_2 + c_3\mathbf{e}_3 = \mathbf{0}$ is $c_1 = c_2 = c_3 = 0$.
- B **spans** $\mathbb{R}^3$ — any $(x, y, z) \in \mathbb{R}^3$ can be written as:
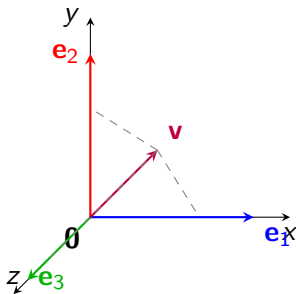
$$(x, y, z) = x\mathbf{e}_1 + y\mathbf{e}_2 + z\mathbf{e}_3.$$

B is a basis for $\mathbb{R}^3$ and $\dim(\mathbb{R}^3) = 3$.

Basis of $\mathbb{R}^3$

Example: The standard basis of $\mathbb{R}^3$ is

$$\mathcal{B} = \{\mathbf{e}_1 = (1,0,0),\ \mathbf{e}_2 = (0,1,0),\ \mathbf{e}_3 = (0,0,1)\}.$$

Caption: $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ form a basis of $\mathbb{R}^3$. Any vector $\mathbf{v} \in \mathbb{R}^3$ can be written as $\mathbf{v} = a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + a_3\mathbf{e}_3$.

Examples

Examples of standard basis

1. If $V = \mathbb{R}^2(\mathbb{R})$, then $\{(1,0),(0,1)\}$ For $(\alpha, \beta) \in \mathbb{R}^2$,

$$(\alpha, \beta) = \alpha(1,0) + \beta(0,1)$$

2. $V = \mathbb{R}^3(\mathbb{R}), \quad \{(1,0,0),(0,1,0),(0,0,1)\}$

3. $V = \mathbb{R}^n(\mathbb{R}), \quad \{(1,0,\ldots,0),(0,1,\ldots,0),\ldots,(0,0,\ldots,1)\}$
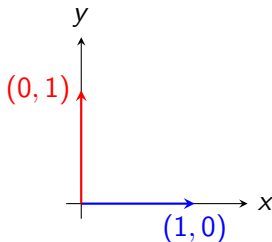
When a Set is Not a Basis

Recall: A basis of V must be a set of vectors that are
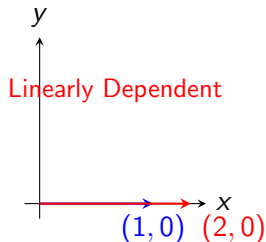
(1) Linearly Independent   and   (2) Span the space.

Valid Basis of $\mathbb{R}^2$

Not a Basis of $\mathbb{R}^2$

$\mathcal{B} = \{(1,0), (0,1)\}$

$S = \{(1,0), (2,0)\}$

A basis requires independence and spanning. Here S fails: both vectors lie on the x-axis, so they cannot span $\mathbb{R}^2$.

Dimension

The number of vectors in a basis of $V(F)$ (V over a field F (denoted VF)) is called the **dimension** of the vector space $V(F)$. **Examples:**

1. Dimensions of vector space $\mathbb{R}^2(\mathbb{R})$ is 2.
2. Dimensions of vector space $\mathbb{R}^n(\mathbb{R})$ is n.
3. Dimensions of vector space $M_{m \times n}(\mathbb{R})$ is $m \times n$.
4. Dimensions of vector space $P_n(\mathbb{R})$ is $n + 1$.

1. $\mathbb{R}^2(\mathbb{R})$: Basis $= \{(1,0),(0,1)\}$, dim $= 2$
2. $\mathbb{R}^n(\mathbb{R})$: Basis $= \{e_1, e_2, \ldots, e_n\}$, dim $= n$
3. $M_{m \times n}(\mathbb{R})$: Space of $m \times n$ real matrices, dim $= m \times n$
4. $P_n(\mathbb{R})$: Polynomials of degree $\leq n$, Basis $= \{1, x, x^2, \ldots, x^n\}$, dim $= n + 1$

What is a Linear Transformation?

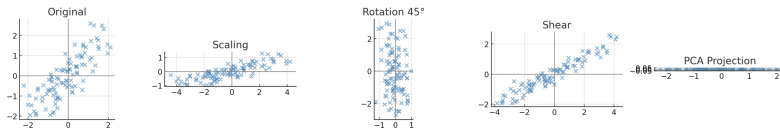**Definition:** A *linear transformation* $T : \mathbb{R}^n \to \mathbb{R}^m$ satisfies:

$$T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y}), \quad T(c\mathbf{x}) = c \cdot T(\mathbf{x})$$

and can be written as:

$$T(\mathbf{x}) = A\mathbf{x}$$

- Preserves the origin ($0 \to 0$).
- Maps lines to lines, planes to planes.
- Can rotate, scale, shear, or project data.

Example of Linear Transformer

- Projects data onto the *x*-axis.
- Example of dimensionality reduction.
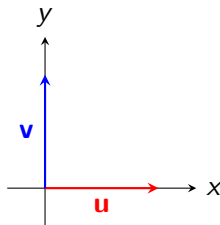
Orthogonal Vectors

Definition: Two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are **orthogonal** if

$$\mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^{n} u_i v_i = 0$$

Example:

$$\mathbf{u} = (1, 0), \quad \mathbf{v} = (0, 1)$$

$$\mathbf{u} \cdot \mathbf{v} = 0$$
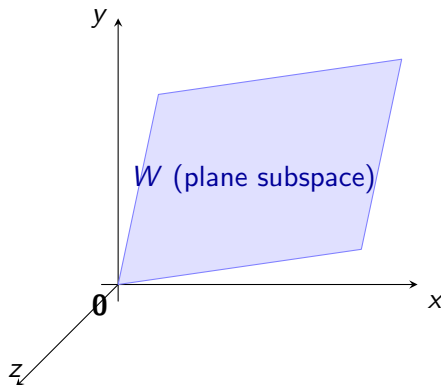
Vector Space and Subspace

Vector Space V over a field $\mathbb{F}$: A set of vectors with two operations (addition and scalar multiplication) satisfying:

- Closure under addition and scalar multiplication

- Associativity, commutativity of addition

- Existence of additive identity ($\mathbf{0}$) and additive inverses

- Distributive and compatibility properties

Subspace $W \subseteq V$: A subset W is a *subspace* of V if:
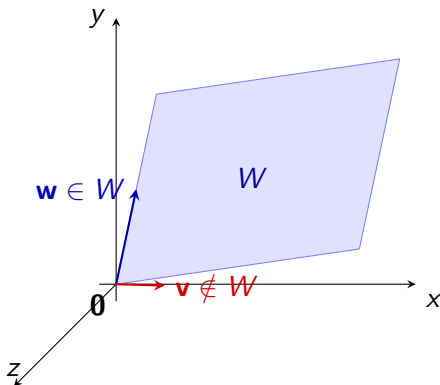
- $\mathbf{0} \in W$ (contains zero vector)

- Closed under addition: $\mathbf{u}, \mathbf{v} \in W \implies \mathbf{u} + \mathbf{v} \in W$

- Closed under scalar multiplication: $c \in \mathbb{F}$, $\mathbf{u} \in W \implies c\mathbf{u} \in W$

Vector Space and Subspace: $W \subseteq V$

Caption: The vector space $V = \mathbb{R}^3$ (entire space with axes) contains a subspace W (a plane through the origin).

Subspace Illustration: $W \subseteq V$

Caption: Vectors inside W remain within the subspace, whereas **v** (red) lies outside W.

Subspace Hierarchy: $U \subseteq W \subseteq V$

$$U \subseteq W \subseteq V$$

Caption: Subspaces form a nested structure: the line U lies inside the plane W, and both lie inside the space V.

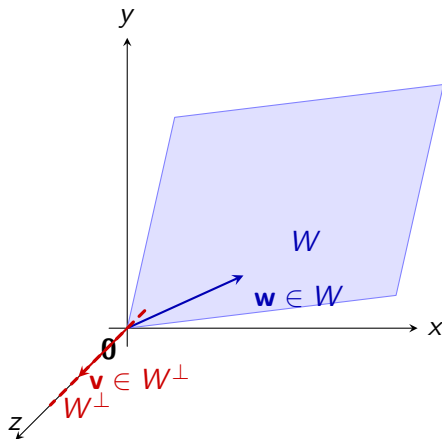# Orthogonal Complement

Definition

For a subspace $W \subseteq V$, the **orthogonal complement** of W is

$$W^\perp = \{\mathbf{v} \in V \mid \mathbf{v} \cdot \mathbf{w} = 0, \ \forall \mathbf{w} \in W\}.$$

- $W^\perp$ is also a subspace of V.
- Geometrically: all vectors perpendicular to every vector in W.

The orthogonal complement of a subspace W in a vector space V is defined as the set of all vectors in V that are orthogonal to every vector in W. This means that the dot product of any vector in the orthogonal complement with any vector in the subspace W is zero. The orthogonal complement itself is also a subspace of V.

Caption: $W^\perp$ is the set of vectors orthogonal to every vector in W. In $\mathbb{R}^3$, if W is a plane, then $W^\perp$ is a line perpendicular to it.

Example in $\mathbb{R}^3$

Let
$$W = \text{span}\{(1,1,0),\ (0,1,1)\}.$$

Step 1: A general vector in W is

$$\mathbf{w} = a(1,1,0) + b(0,1,1) = (a,\ a+b,\ b).$$

Step 2: A vector $\mathbf{v} = (x,y,z)$ is in $W^{\perp}$ if

$$\mathbf{v} \cdot (1,1,0) = x + y = 0, \qquad \mathbf{v} \cdot (0,1,1) = y + z = 0.$$

Solving the Conditions

$$\begin{cases} x + y = 0 \implies x = -y, \\ y + z = 0 \implies z = -y. \end{cases}$$
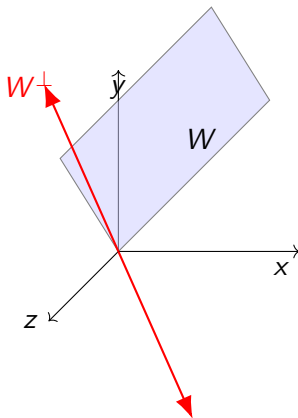
So

$$\mathbf{v} = (-y,\, y,\, -y) = y(-1, 1, -1).$$

Result

$$W^{\perp} = \text{span}\{(-1, 1, -1)\}.$$

- W is a **plane** in $\mathbb{R}^3$.
- $W^{\perp}$ is a **line** perpendicular to that plane.

Geometric Illustration

Orthogonal Projection

V be an inner product space and W be subspace of V. Also let $\{w_1, w_2 .. w_k\}$ be orthogonal basic of W. The linear transformation

$$P : V \longrightarrow W$$

is called the **orthogonal projection** onto W, where for any $v \in V$, we have

$$P_W(v) = \frac{\langle v, w_1 \rangle}{\|w_1\|^2} w_1 + \frac{\langle v, w_2 \rangle}{\|w_2\|^2} w_2 + \cdots + \frac{\langle v, w_k \rangle}{\|w_k\|^2} w_k$$

subspace W (the x-axis), Vector v, Projection ($PW(v)$) in green, Residual r in purple (orthogonal).

Let $w_1 = (3, 0, 1)$ and $w_2 = (0, 1, 0)$. Then, $W = L(w_1, w_2)$.
Take a vector $v = (0, 3, 10) \in \mathbb{R}^3$.

$$P_W(v) = \frac{\langle v, w_1 \rangle}{\langle w_1, w_1 \rangle} w_1 + \frac{\langle v, w_2 \rangle}{\langle w_2, w_2 \rangle} w_2$$

$$= \frac{10}{10}(3, 0, 1) + \frac{3}{1}(0, 1, 0)$$

$$= (3, 0, 1) + (0, 3, 0) = (3, 3, 1)$$

Now, the residual is:

$$v - P_W(v) = (0, 3, 10) - (3, 3, 1) = (-3, 0, 9) \in W^\perp$$

plane W spanned by w_1, w_2 (light blue shaded). The vector v in red. The projection $P_W(v)$ in blue. The residual $v - P_W(v)$ in green, which is orthogonal to W.

Motivation for Eigenvalues and Eigenvectors

Why do we need them?

- Linear transformations *stretch*, *shrink*, or *rotate* vectors.
- Most vectors change both **direction** and **magnitude**.
- **Eigenvectors** are the *special directions* that remain unchanged in direction.
- The associated **eigenvalue** tells how much scaling occurs.

Definition: For a square matrix $A \in \mathbb{R}^{n \times n}$, a non-zero vector $v \in \mathbb{R}^n$ is an eigenvector and $\lambda \in \mathbb{R}$ is the eigenvalue.

$$Av = \lambda v$$

Example

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

Step 1: Characteristic polynomial

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

$$\implies (\lambda - 1)(\lambda - 3) = 0$$

$$\therefore \lambda_1 = 3, \quad \lambda_2 = 1$$

Step 2: Eigenvectors For $\lambda = 3$:

$$(A - 3I)x = 0 \quad \Rightarrow \quad \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$

$$\Rightarrow x_1 = x_2 \quad \Rightarrow v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

For $\lambda = 1$:

$$(A - I)x = 0 \quad \Rightarrow \quad \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$

$$\Rightarrow x_1 = -x_2 \quad \Rightarrow v_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Final Answer:

$$\lambda_1 = 3, \ v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad \lambda_2 = 1, \ v_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Orthonormal basis of eigenvectors:

$$\hat{v}_1 = \tfrac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad \hat{v}_2 = \tfrac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Eigenvalues and Eigenvectors

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

$$\lambda_1 = 3, \ v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \lambda_2 = 1, \ v_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

- Eigenvectors = invariant directions
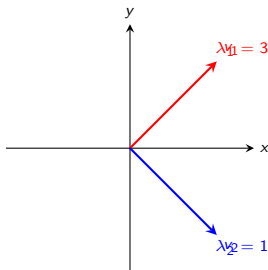- Eigenvalues = scaling factors

Quadratic Form in Linear Algebra

A **quadratic form** in n variables is defined as:

$$f(x) = x^T A x, \quad x \in \mathbb{R}^n, \ A^T = A.$$

where

- $A \in \mathbb{R}^{n \times n}$ is a **symmetric matrix** $(A^T = A)$,
- x is a column vector

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Deep Learning

Quadratic forms provide the mathematical backbone for loss functions, regularization, distance metrics, and curvature analysis in deep learning

Example of Quadratic Form

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

$$f(x) = x^T A x = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 2x_1^2 + 2x_1 x_2 + 3x_2^2$$

$$\Rightarrow f(x) \text{ is a quadratic form in } x_1, x_2.$$

(A) Mean Squared Error (MSE) Loss For linear prediction,

$$\hat{\mathbf{y}} = W\mathbf{x},$$

the loss function is defined as

$$\mathcal{L} = \|\mathbf{y} - W\mathbf{x}\|^2 = (\mathbf{y} - W\mathbf{x})^\top (\mathbf{y} - W\mathbf{x}),$$

which is a quadratic form in $\mathbf{x}$ (and in the model parameters when expanded).
Insight: Since the loss is a convex quadratic function, it admits a unique global minimum.
(B) Weight Decay / L_2 Regularization The regularization term is given by

$$\mathcal{L}_{\text{reg}} = \lambda \|\mathbf{w}\|^2 = \lambda \, \mathbf{w}^\top I \mathbf{w}.$$

Interpretation: Penalizes large weights and improves generalization.
Matrix view: The identity matrix I defines a simple quadratic form.

Positive Semi-Definite Matrix

A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is called **positive semi-definite (PSD)** if

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

Note

If $x^T A x > 0$ for all non-zero x, then A is called **positive definite (PD)**. x is vector.

Example of PSD Matrix

Consider

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

- A is symmetric since $A^T = A$.
- For any $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$,

$$x^T A x = 2x_1^2 - 2x_1 x_2 + 2x_2^2 = (x_1 - x_2)^2 + x_1^2 + x_2^2 \geq 0.$$

$$\Rightarrow \quad A \text{ is positive semi-definite.}$$

Positive Definite Matrix

A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is called **positive definite (PD)** if

$$x^T A x > 0, \quad \forall x \in \mathbb{R}^n, \ x \neq 0.$$

Key Property

All eigenvalues of a positive definite matrix are strictly positive.

Example of Positive Definite Matrix

Consider

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

- A is symmetric since $A^T = A$.
- For any $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$,

$$x^T A x = 2x_1^2 - 2x_1 x_2 + 2x_2^2.$$

- Rewrite:

$$x^T A x = (x_1 - x_2)^2 + x_1^2 + x_2^2.$$

- Since $(x_1 - x_2)^2 + x_1^2 + x_2^2 > 0$ for all non-zero x, A is **positive definite**.

Norm

Definition: A *norm* on a vector space V is a function $\| \cdot \| : V \to \mathbb{R}$ such that for all $\mathbf{x}, \mathbf{y} \in V$ and scalar α:

1. **Non-negativity:** $\|\mathbf{x}\| \geq 0$, $\|\mathbf{x}\| = 0 \iff \mathbf{x} = \mathbf{0}$
2. **Homogeneity:** $\|\alpha\mathbf{x}\| = |\alpha|\, \|\mathbf{x}\|$
3. **Triangle inequality:** $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$

Common Norms: Manhattan distance (L_1), Euclidean length (L_2), and Max norm (L_∞)

$$\|\mathbf{x}\|_1 = \sum |x_i|, \quad \|\mathbf{x}\|_2 = \sqrt{\sum x_i^2}, \quad \|\mathbf{x}\|_\infty = \max |x_i|$$
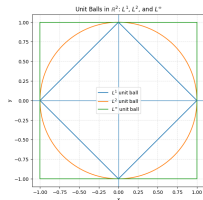
Norm Examples

Given: $\mathbf{x} = (3, 4)$

$$\|\mathbf{x}\|_1 = |3| + |4| = 7$$
$$\|\mathbf{x}\|_2 = \sqrt{3^2 + 4^2} = 5$$
$$\|\mathbf{x}\|_\infty = \max(3, 4) = 4$$

Unit balls for L^1, L^2, L^∞ norms

Metric Space

Definition: A metric space (M, d) consists of a set M and a function $d : M \times M \to \mathbb{R}$ satisfying:

1. **Non-negativity:** $d(x, y) \geq 0$, and $d(x, y) = 0 \iff x = y$
2. **Symmetry:** $d(x, y) = d(y, x)$
3. **Triangle inequality:** $d(x, z) \leq d(x, y) + d(y, z)$

From Norm to Metric: Given norm $\| \cdot \|$, define:

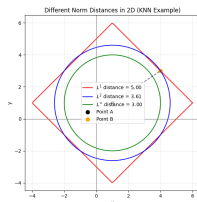$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$$

Metric Examples

Given: $\mathbf{x} = (1, 2)$, $\mathbf{y} = (4, 6)$

$$d_1(\mathbf{x}, \mathbf{y}) = |1 - 4| + |2 - 6| = 7$$

$$d_2(\mathbf{x}, \mathbf{y}) = \sqrt{(1 - 4)^2 + (2 - 6)^2} = 5$$

$$d_\infty(\mathbf{x}, \mathbf{y}) = \max(|1 - 4|, |2 - 6|) = 4$$

Visualizing distances in KNN using different norms

Orthogonal Vectors

Definition

A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is **orthogonal** if

$$\mathbf{v}_i^\top \mathbf{v}_j = 0 \quad \text{for } i \neq j.$$

Key Properties

- Vectors are mutually perpendicular
- Vector magnitudes are arbitrary (non-zero)
- Orthogonal sets are linearly independent

Example

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 3 \end{bmatrix}, \quad \mathbf{v}_1^\top \mathbf{v}_2 = 0$$

Orthonormal Vectors

Definition

A set of vectors is **orthonormal** if

- Vectors are orthogonal
- Each vector has unit norm

$$\|\mathbf{v}_i\| = 1$$

Equivalent Condition

$$\mathbf{v}_i^\top \mathbf{v}_j = \begin{cases} 1, & i = j, \\ 0, & i \neq j \end{cases}$$

Example

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Orthogonal vs Orthonormal

Property	Orthogonal	Orthonormal
Dot Product	$= 0$	$= 0$
Vector Length	Arbitrary	1
Normalized	No	Yes
Ease of Projection	Medium	High
Common Usage	Geometry	PCA, SVD, DL

Relationship Between Them

Normalization

Any orthogonal set can be converted into an orthonormal set by:

$$\mathbf{u}_i = \frac{\mathbf{v}_i}{\|\mathbf{v}_i\|}$$

Key Takeaway

Orthogonal = Perpendicular
Orthonormal = Perpendicular + Unit Length

Explore New Loss function

[a] ArcFace, CosFace, SphereFace

---

[a]https://www.researchgate.net/publication/369911128_Machine_Learning_
Deep_Learning_and_Face_Recognition_Loss_Functions_Cross_Entropy_KL_
Softmax_Regression_Triplet_Center_Constructive_Sphere_and_ArcFace_Deep_
Face_Recognition

The eigenvalues of

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

are found by solving $\det(A - \lambda I) = 0$:

$$\det \begin{bmatrix} 2 - \lambda & -1 \\ -1 & 2 - \lambda \end{bmatrix} = (2 - \lambda)^2 - 1 = 0.$$

$$\Rightarrow \lambda^2 - 4\lambda + 3 = 0 \quad \Rightarrow \quad \lambda_1 = 1, \ \lambda_2 = 3.$$

Both eigenvalues are positive $\Rightarrow$ A is positive definite.

Linearly Independent Eigenvector

- If $A \in \mathbb{R}^{n \times n}$ has distinct eigenvalues $\lambda_1, \ldots, \lambda_k$, then the corresponding eigenvectors $v_1, \ldots, v_k$ are **linearly independent**.

- If A has n distinct eigenvalues, then

$$A = V \Lambda V^{-1}, \quad V = [v_1 \ \cdots \ v_n], \ \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n).$$

- This property is very useful because it ensures we can form a basis of eigenvectors and use it for diagonalization of A

$$A = V \Lambda V^{-1},$$

where V contains linearly independent eigenvectors and Λ is the diagonal matrix of eigenvalues.

Example Matrix

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

Characteristic polynomial:

$$\det(A - \lambda I) = \begin{vmatrix} 4 - \lambda & 1 \\ 2 & 3 - \lambda \end{vmatrix} = (4 - \lambda)(3 - \lambda) - 2 = \lambda^2 - 7\lambda + 10.$$

Distinct eigenvalues:

$$\lambda_1 = 5, \qquad \lambda_2 = 2.$$

Eigenvectors

For $\lambda_1 = 5$:

$$(A - 5I)v = 0 \;\Rightarrow\; \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix} v = 0 \;\Rightarrow\; v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

For $\lambda_2 = 2$:

$$(A - 2I)v = 0 \;\Rightarrow\; \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix} v = 0 \;\Rightarrow\; v_2 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}.$$

Eigenvectors

For $\lambda_1 = 5$:

$$(A - 5I)v = 0 \Rightarrow \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix} v = 0 \Rightarrow v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

For $\lambda_2 = 2$:

$$(A - 2I)v = 0 \Rightarrow \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix} v = 0 \Rightarrow v_2 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}.$$

Form

$$V = \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix}, \qquad \Lambda = \begin{bmatrix} 5 & 0 \\ 0 & 2 \end{bmatrix}.$$

Check Linear Independence

$$\det(V) = \det \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix} = 1 \cdot 2 - (-1) \cdot 1 = 3 \neq 0.$$

Hence v_1 and v_2 are **linearly independent**.

(Equivalently: eigenvectors for distinct eigenvalues are independent by the theorem.)

Diagonalization (Verification)

Compute

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

Then

$$V \Lambda V^{-1} = \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 2 \end{bmatrix} \frac{1}{3} \begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix} = A.$$

Diagonalization (Verification)

Compute

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

Then

$$V\Lambda V^{-1} = \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 2 \end{bmatrix} \frac{1}{3} \begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix} = A.$$

Thus $A = V\Lambda V^{-1}$ with V built from **linearly independent eigenvectors**.
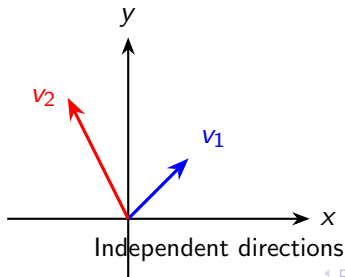
Linearly Independent Eigenvectors

- If a matrix has distinct eigenvalues, their eigenvectors are linearly independent.
- Example:

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

Eigenvalues: $\lambda_1 = 5$, $\lambda_2 = 2$

$$v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad v_2 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}.$$

Independent directions

Eigen Decomposition

For a square matrix $A \in \mathbb{R}^{n \times n}$, If A has n linearly independent eigenvectors, it can be decomposed (factorized) as:

$$A = V \Lambda V^{-1}$$

where

- $V = [v_1 \ v_2 \ \cdots \ v_n]$ is the eigenvector matrix,
- $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is the eigenvalue matrix.

Special Case (Symmetric A)

$$A = V \Lambda V^T, \quad V^T V = I$$

Decomposition involves breaking down this matrix into several component matrices, revealing essential structure within the data.

Example: Eigen Decomposition Step 1 - Eigenvalues

Consider

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

Characteristic polynomial:

$$\det(A - \lambda I) = (4 - \lambda)(3 - \lambda) - 2$$

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

Eigenvalues:

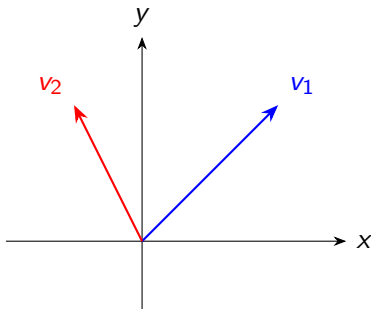$$\lambda_1 = 5, \quad \lambda_2 = 2$$

Example: Step 2 - Eigenvectors

For $\lambda = 5$:

$$(A - 5I)v = 0 \quad \Rightarrow \quad v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

For $\lambda = 2$:

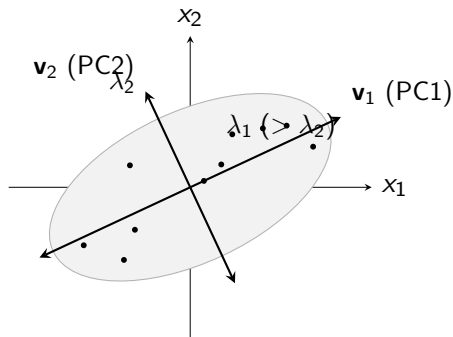$$(A - 2I)v = 0 \quad \Rightarrow \quad v_2 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

Construct matrices:

$$V = \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix}, \quad \Lambda = \begin{bmatrix} 5 & 0 \\ 0 & 2 \end{bmatrix}$$

Hence:

$$A = V\Lambda V^{-1}$$

PCA: Orthogonal Components in 2D

Setup

$X \in \mathbb{R}^{n \times 2}$ (centered), $\quad C = \frac{1}{n-1} X^{\top}$

Eigen-decomposition

$$C \mathbf{v}_i = \lambda_i \mathbf{v}_i, \quad \mathbf{v}_1 \perp \mathbf{v}_2, \ \|\mathbf{v}_i\| = 1$$

Interpretation

- $\mathbf{v}_1$ (PC1): direction of *max* variance λ_1.

- $\mathbf{v}_2$ (PC2): orthogonal, variance λ_2.

- Project data: $Z = XV$ with $V = [\mathbf{v}_1 \ \mathbf{v}_2]$.

Applications of Eigen Decomposition

- Principal Component Analysis (PCA)
- Solving systems of differential equations
- Computing matrix powers A^k
- Quantum mechanics: operators and observables
- Google PageRank algorithm

Spectral Decomposition: Definition

Spectral Theorem: If $A \in \mathbb{R}^{n \times n}$ is symmetric, then

$$A = Q \Lambda Q^T$$

where

- Q is an orthogonal matrix ($Q^T Q = I$),
- Λ is a diagonal matrix of eigenvalues.

Key facts:

- All eigenvalues are real.
- Eigenvectors can be chosen orthonormal.

Matrix for Spectral Decomposition

We want to decompose

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

into the form

$$A = Q\Lambda Q^T$$

where Q is orthogonal and Λ is diagonal.

Step 1: Eigenvalues

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

$$\det \begin{bmatrix} 4 - \lambda & 1 \\ 1 & 4 - \lambda \end{bmatrix} = (4 - \lambda)^2 - 1 = 0$$

$$\lambda_1 = 5, \quad \lambda_2 = 3$$

Step 2: Eigenvectors

For $\lambda_1 = 5$:

$$v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad q_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

For $\lambda_2 = 3$:

$$v_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad q_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$
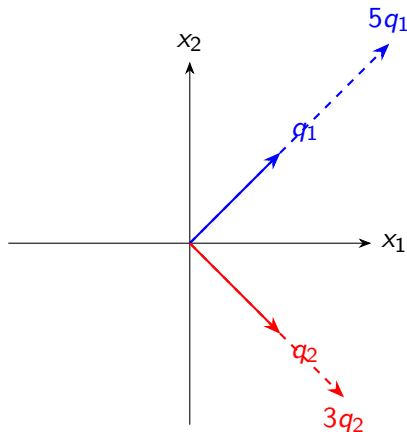
$$Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}, \quad \Lambda = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}$$

$$A = Q\Lambda Q^T$$

$$\begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}^T$$
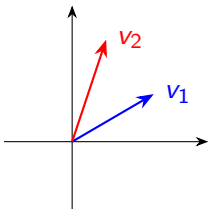
A scales vectors along q_1 by 5 and along q_2 by 3. Also shows eigenvectors q_2 and q_1 how they are scaled by Λ.

Eigenvalue Decomposition vs. Spectral Decomposition
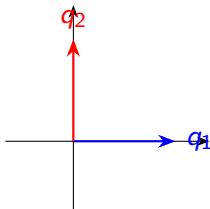
EVD

$$A = V \Lambda V^{-1}$$

- Works for any diagonalizable A
- V: eigenvectors (not necessarily orthogonal)
- Λ: eigenvalues (real or complex)

Spectral Decomposition (SD)

$$A = Q \Lambda Q^T$$

- Only for symmetric/Hermitian A
- Q: orthogonal (or unitary) eigenvectors
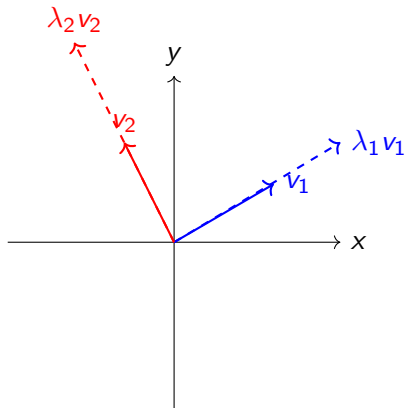- Λ: real eigenvalues

Eigenvalue Decomposition and Spectral Decomposition

Aspect	Eigendecomposition ($A = PDP^{-1}$)	Singular Value Decomposition (SVD) ($A = U\Sigma V^*$)
Orthonormality of Vectors	Columns of P are eigenvectors, not necessarily orthogonal.	U and V are orthonormal matrices (pure rotations/reflections).
Relation between Basis Matrices	P and P^{-1} are exact inverses.	U and V are not inverses and are usually unrelated.
Diagonal Entries	D contains eigenvalues, which may be real or complex (positive/negative/imaginary).	Σ contains singular values, always real and nonnegative.
Existence	Only for square matrices, and not guaranteed (requires diagonalizability).	Always exists for any rectangular or square matrix.

Geometric Interpretation

Eigendecomposition
($A = PDP^{-1}$)

Singular Value Decomposition
($A = U\Sigma V^*$)

Vectors are stretched along eigenvector directions.

Rotation $\rightarrow$ Scaling $\rightarrow$ Rotation. (unit circle transformed into an ellipse via SVD)

PCA vs. SVD: Definitions

Singular Value Decomposition (SVD)

$$A = U \Sigma V^T$$

- Works for any real/complex matrix. - U, V are orthogonal, Σ has singular values.

Principal Component Analysis (PCA)

$$C = \frac{1}{n} A^T A$$

- Finds orthogonal directions maximizing variance. - Based on eigendecomposition of covariance matrix C.

Drawbacks of PCA compared to SVD

- **Covariance dependency:** PCA works on $A^T A$ instead of $A \rightarrow$ squares condition number, may reduce stability.

- **Loss of sign/structural info:** PCA gives only principal axes, SVD retains both left and right singular vectors.

- **Restriction:** PCA requires a symmetric, positive semidefinite covariance matrix; SVD works for any rectangular matrix.

- **Scaling sensitivity:** PCA depends on feature variance, requiring normalization.

- **Numerical stability:** PCA via eigendecomposition can be unstable; modern PCA implementations rely on SVD.

Quick Comparison

Aspect	PCA	SVD
Input	Covariance ($A^T A$)	Original matrix (A)
Matrix type	Symmetric, PSD	Any matrix
Output	Principal components	Singular vectors + values
Stability	Less stable (eig.)	More stable
Scaling issue	Sensitive	Not sensitive
Use case	Dimensionality reduction	General factorization

SVD

PCA

PC_1 (max variance)

PC_2

$\sigma_2 u_2$

A

$\sigma_1 u_1$

Unit circle

Ellipse

SVD: stretching of axes

PCA: directions of max variance

Summary: PCA vs. SVD

- **Use PCA** when the goal is **dimensionality reduction**, feature extraction, or finding **directions of maximum variance** in data.

- **Use SVD** when you need a **general matrix factorization**, numerical stability, or when working with **any rectangular matrix**.

- In practice: **PCA is often computed using SVD**, since it is more stable and efficient.

Inner Product: Definition

Definition: An *inner product* on a vector space V is a function

$$\langle \mathbf{u}, \mathbf{v} \rangle : V \times V \to \mathbb{R}$$

that satisfies:

1. **Symmetry:** $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$
2. **Linearity:** $\langle a\mathbf{u} + b\mathbf{w}, \mathbf{v} \rangle = a\langle \mathbf{u}, \mathbf{v} \rangle + b\langle \mathbf{w}, \mathbf{v} \rangle$
3. **Positive-definiteness:** $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ and $= 0$ iff $\mathbf{u} = \mathbf{0}$
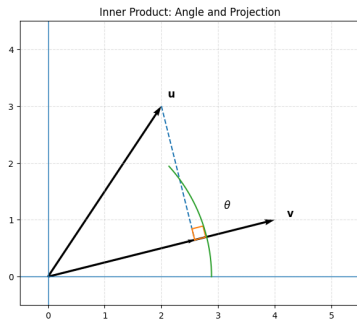
Euclidean inner product:

$$\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{i=1}^{n} u_i v_i$$

Geometric Interpretation

For vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$:

$$\langle \mathbf{u}, \mathbf{v} \rangle = \|\mathbf{u}\|\|\mathbf{v}\|\cos\theta$$

where θ is the angle between them.

Inner Product: Angle and Projection

Key idea: - If $\langle \mathbf{u}, \mathbf{v} \rangle > 0 \to$ acute angle - If $\langle \mathbf{u}, \mathbf{v} \rangle = 0 \to$ orthogonal - If $\langle \mathbf{u}, \mathbf{v} \rangle < 0 \to$ obtuse angle

Example: Numeric Computation

$$\mathbf{u} = [2,\ 3], \quad \mathbf{v} = [4,\ 1]$$

$$\langle \mathbf{u}, \mathbf{v} \rangle = (2)(4) + (3)(1) = 8 + 3 = 11$$

Geometric check:

$$\|\mathbf{u}\| = \sqrt{2^2 + 3^2} = \sqrt{13}, \quad \|\mathbf{v}\| = \sqrt{4^2 + 1^2} = \sqrt{17}$$

$$\cos\theta = \frac{11}{\sqrt{13} \cdot \sqrt{17}} \approx 0.741$$

$$\theta \approx 42.0^\circ$$

Inner Product in Data Science & ML

Applications:

- **Cosine Similarity:**

$$\text{sim}(\mathbf{u}, \mathbf{v}) = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|\|\mathbf{v}\|}$$

  Used in NLP, recommendation systems, and document retrieval.

- **Kernel Methods:** In SVM, the kernel $k(\mathbf{x}, \mathbf{y}) = \langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle$ is an inner product in feature space.

- **Projection:** Project $\mathbf{u}$ onto $\mathbf{v}$:

$$\text{proj}_{\mathbf{v}}(\mathbf{u}) = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle}\mathbf{v}$$

Applications of Spectral Decomposition

- Analyzing quadratic forms in optimization.
- Principal Component Analysis (PCA).
- Computing matrix functions:

$$f(A) = Q f(\Lambda) Q^T$$

  (e.g., exponential, square root).
- Signal processing and statistics.

Orthogonal Complements in ML:

- **Linear Regression:** Residual $r = y - \hat{y}$ lies in $(\text{Col}(X))^{\perp}$.

$$X^T r = 0$$

- **PCA:** Dropped components lie in the orthogonal complement of chosen principal subspace.
- **Signal Processing:** Separates signal subspace from noise subspace.
- **Word Embeddings:** Bias removal by projecting into $W^{\perp}$ of bias direction.

Applications in Data Science & Machine Learning:

- **PCA:** Principal components are orthogonal $\Rightarrow$ uncorrelated features.
- **SVD:** $A = U\Sigma V^T$, where U, V are orthogonal matrices.
- **Gradient Descent:** Orthogonal initialization improves stability.
- **Feature Independence:** Orthogonal features avoid redundancy.
- **Word Embeddings:** Nearly orthogonal vectors represent unrelated concepts.

Conclusion

- Vector spaces underpin linear algebra and machine learning.
- Next: **Subspaces** – subsets that are themselves vector spaces.

Vector Space

Definition: A **vector space** (or linear space) over a field $\mathbb{F}$ (e.g., $\mathbb{R}$ or $\mathbb{C}$) is a set V together with two operations:

1. **Vector addition:** $u + v \in V$
2. **Scalar multiplication:** $c \cdot v \in V$ for $c \in \mathbb{F}$

that satisfy the **vector space axioms** (closure, associativity, commutativity, identity, inverse, distributive properties, etc.).

Vector Space

Definition: A **vector space** (or linear space) over a field $\mathbb{F}$ (e.g., $\mathbb{R}$ or $\mathbb{C}$) is a set V together with two operations:

1. **Vector addition:** $u + v \in V$
2. **Scalar multiplication:** $c \cdot v \in V$ for $c \in \mathbb{F}$

that satisfy the **vector space axioms** (closure, associativity, commutativity, identity, inverse, distributive properties, etc.).

Example:

$$\mathbb{R}^n = \{\mathbf{x} = (x_1, x_2, \ldots, x_n) \mid x_i \in \mathbb{R}\}$$

- Vectors can be added component-wise. - Vectors can be multiplied by scalars.

Vector Space

Definition: A **vector space** (or linear space) over a field $\mathbb{F}$ (e.g., $\mathbb{R}$ or $\mathbb{C}$) is a set V together with two operations:

1. **Vector addition:** $u + v \in V$
2. **Scalar multiplication:** $c \cdot v \in V$ for $c \in \mathbb{F}$

that satisfy the **vector space axioms** (closure, associativity, commutativity, identity, inverse, distributive properties, etc.).

Example:

$$\mathbb{R}^n = \{\mathbf{x} = (x_1, x_2, \ldots, x_n) \mid x_i \in \mathbb{R}\}$$

- Vectors can be added component-wise. - Vectors can be multiplied by scalars.

Non-example: Set of all positive real numbers under normal addition is **not** a vector space (no additive inverse for $x > 0$).

Vector Space

Definition: A **vector space** (or linear space) over a field $\mathbb{F}$ (e.g., $\mathbb{R}$ or $\mathbb{C}$) is a set V together with two operations:

1. **Vector addition:** $u + v \in V$
2. **Scalar multiplication:** $c \cdot v \in V$ for $c \in \mathbb{F}$

that satisfy the **vector space axioms** (closure, associativity, commutativity, identity, inverse, distributive properties, etc.).

Example:

$$\mathbb{R}^n = \{\mathbf{x} = (x_1, x_2, \ldots, x_n) \mid x_i \in \mathbb{R}\}$$

- Vectors can be added component-wise. - Vectors can be multiplied by scalars.

Non-example: Set of all positive real numbers under normal addition is **not** a vector space (no additive inverse for $x > 0$).

In Data Science:

- Feature vectors form vector spaces.
- Embedding spaces (e.g., word2vec, face embeddings) are vector spaces.
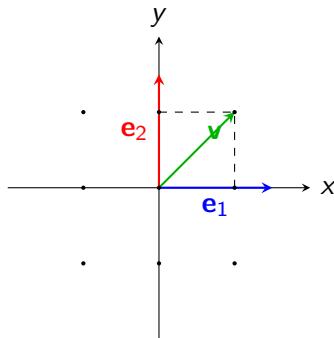- Linear models operate on vector spaces.

Visualizing a 2D Vector Space

- A 2D vector space over $\mathbb{R}$ contains all vectors of the form:

$$\mathbf{v} = a\mathbf{e}_1 + b\mathbf{e}_2$$

where $\mathbf{e}_1, \mathbf{e}_2$ are basis vectors.

- Every point in the plane (origin included) is in the space.

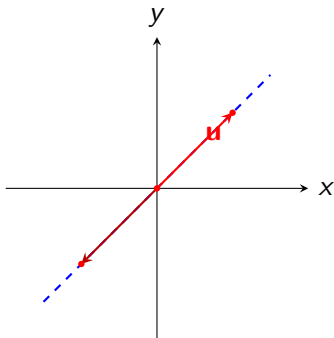- Operations: vector addition, scalar multiplication.

Visualizing a 1D Subspace of $\mathbb{R}^2$

- A **subspace** is a subset of a vector space that is itself a vector space.
- Example: All scalar multiples of a nonzero vector **u** in $\mathbb{R}^2$.

$$W = \{c\mathbf{u} \mid c \in \mathbb{R}\}$$

- Geometrically, in $\mathbb{R}^2$, this forms a line through the origin.
- Closed under vector addition and scalar multiplication.

Basis, Dimension, and Dimensionality Reduction

Basis:

- A **basis** of a vector space V is a set of linearly independent vectors that span V.
- Every vector in V can be uniquely expressed as a linear combination of basis vectors.
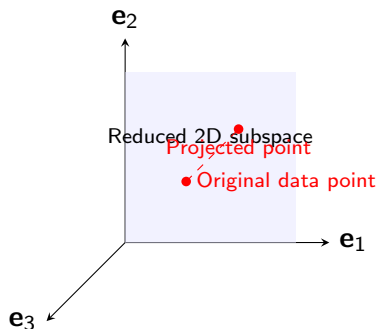- Example in $\mathbb{R}^3$:

$$\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$$

Dimension

Dimension:

- The number of vectors in any basis of V.
- $\dim(\mathbb{R}^n) = n$.

Dimensionality Reduction

- Goal: Reduce the number of basis vectors while preserving key information.
- Methods: PCA, t-SNE, UMAP.
- Example: Reduce $\mathbb{R}^{100}$ data to $\mathbb{R}^2$ for visualization.

Why It Matters in ML

- **KNN, K-Means, Clustering:** Distance metric defines "nearest".
- **Different norms** change cluster shapes:
  - L^1: Diamond-shaped boundaries
  - L^2: Circular boundaries
  - L^∞: Square boundaries
- **Regularization:**
  - $L^1 \rightarrow$ Sparsity (Lasso)
  - $L^2 \rightarrow$ Smoothness (Ridge)

Norms and Metrics in DS and ML

Norms in DS & ML:

- L^2 norm: Feature scaling, Ridge regression regularization.
- L^1 norm: Sparsity promotion in Lasso regression.

Metrics in DS & ML:

- Euclidean distance in KNN, clustering.
- Cosine distance for text and embeddings.
- Mahalanobis distance for correlated features.

Why Vectors Matter in DS and ML?

- **Data Representation:** Every sample in your dataset can be seen as a vector in $\mathbb{R}^n$.
- **Distance & Similarity:** Euclidean distance, cosine similarity.
- **Linear Algebra Operations:** Dot product, matrix multiplication for model computations.
- **Transformations:** Scaling, normalization, and dimensionality reduction (PCA).
- $\cos\theta$ is the **cosine similarity** between two feature vectors.
- Used in **text embeddings**, **face recognition**, and **recommendation systems**.

Resources

- Essential Mathematics for Machine (Prof Sanjeev Kumar): `https://www.youtube.com/watch?v=JO9jNe6BemE&list=PLLy_2iUCG87D1CXFxE-SxCFZUiJzQ3IvE&index=1`
- Data Science for Biology (Prof Biplab Bose): `https://www.youtube.com/watch?v=ZU-N7dmmEqI&list=PLWdnzlV3ogoXmoCXczKiu6WGWOr05Zw02`
- 3DBrown1Blue (Linear Algebra) `https://www.youtube.com/watch?v=fNk_zzaMoSs&list=PLZHQObOWTQDPD3MizzM2xVFitgF8hE_ab&index=2`
- 3DBrown1Blue (Calculus) `https://www.youtube.com/watch?v=TrcCbdWwCBc&list=PLSQl0a2vh4HC5feHa6Rc5c0wbRTx56nF7&index=1`
- Issues with PCA `https://www.youtube.com/watch?v=sgU4zbO-W4M`
- PCA and SVD `https://www.youtube.com/watch?v=DQ_BkPHIl-g`
- SVD User Movie: `https://www.youtube.com/watch?v=P5mlg91as1c`