Contents

Vector Spaces

What is a Vector Space?

A **vector space** is a set where you can add things together and multiply them by numbers ("scalars"), and the results stay within the set. The rules are:

- **Addition**: If you have two vectors $\vec{u}, \vec{v}$, then $\vec{u} + \vec{v}$ is also a vector in that space.
- **Scalar multiplication**: Multiply a vector $\vec{v}$ by a number a to get another vector: $a\vec{v}$.
- There are other properties (like distributivity, existence of a zero vector, etc.), but these are the core.

Examples

- The set of **all ordered pairs of real numbers** (x, y) is a vector space: $\mathbb{R}^2$.
- The **Euclidean space** $\mathbb{R}^n$ with usual addition and scalar multiplication.
- The set of **all polynomials** with real coefficients.
- The set of **matrices** of fixed size over $\mathbb{R}$.

Axioms

These operations satisfy the following axioms for all $\vec{u}, \vec{v}, \vec{w} \in V$ and scalars $a, b \in \mathbb{F}$:

1. **Associativity of addition:** $(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$.
2. **Commutativity of addition:** $\vec{u} + \vec{v} = \vec{v} + \vec{u}$.
3. **Identity element of addition:** There is a zero vector $\vec{0} \in V$ such that $\vec{v} + \vec{0} = \vec{v}$ for all $\vec{v} \in V$.
4. **Inverse elements of addition:** For each $\vec{v} \in V$, there exists a vector $-\vec{v} \in V$ such that $\vec{v} + (-\vec{v}) = \vec{0}$.
5. **Compatibility of scalar multiplication with field multiplication:** $a(b\vec{v}) = (ab)\vec{v}$.
6. **Identity element of scalar multiplication:** $1\vec{v} = \vec{v}$, where 1 is the multiplicative identity in $\mathbb{F}$.
7. **Distributivity of scalar multiplication with respect to vector addition:** $a(\vec{u} + \vec{v}) = a\vec{u} + a\vec{v}$.
8. **Distributivity of scalar multiplication with respect to field addition:** $(a + b)\vec{v} = a\vec{v} + b\vec{v}$.

Linear Combination

A **linear combination** in a vector space is an expression where you take some vectors and multiply each of them by a scalar (a number), then add all those results together.

Formally, given vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ in a vector space V, and scalars $c_1, c_2, \ldots, c_n$ from the underlying field (like real numbers), the **linear combination** is:

$$\vec{w} = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n$$

This creates another vector $\vec{w}$ in the vector space.

Linear independence

In a vector space, a set of vectors is called **linearly independent** if no vector in the set can be written as a linear combination of the others.

More formally, given vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ in a vector space V, they are linearly independent if the only scalars $c_1, c_2, \ldots, c_n$ satisfying

$$c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n = \vec{0}$$

is the trivial solution where every scalar is zero:

$$c_1 = c_2 = \cdots = c_n = 0$$

Basis and Dimension

A **basis** of a vector space is a set of vectors in V that are **linearly independent** and **span** V.

- **Linear independence** means no vector in the set can be written as a linear combination of the others.

- **Spanning** means every vector in V can be written as a linear combination of basis vectors.

The **dimension** of a vector space is the number of vectors in any (and hence every) basis.

Hierarchy of Spaces

- **Vector Space**: Most general setting for linear algebra.
- **Normed Space**: A vector space with a way to measure "length" (norm).
- **Inner Product Space**: A normed space where you can take "dot products" (measuring angles).
- **Euclidean Space**: The example you know best—$\mathbb{R}^n$ with the usual dot product, forming our usual geometry.

A simple mnemonic: "All Euclidean spaces are inner product spaces, all inner product spaces are normed spaces, all normed spaces are vector spaces—but not vice versa."

Normed Vector Space

- A **normed space** is a vector space equipped with a **norm** $\|\cdot\|$, which measures the "length" or "size" of a vector.
- The norm satisfies:
    - Positivity: $\|\vec{v}\| \geq 0$, and $\|\vec{v}\| = 0 \iff \vec{v} = \vec{0}$
    - Scalar multiplication: $\|a\vec{v}\| = |a|\|\vec{v}\|$
    - Triangle inequality: $\|\vec{u} + \vec{v}\| \leq \|\vec{u}\| + \|\vec{v}\|$
- Norm provides a way to measure distances by: $d(\vec{u}, \vec{v}) = \|\vec{u} - \vec{v}\|$
- Example: Euclidean length in $\mathbb{R}^n$.

Inner Product Space

- An **inner product space** is a normed space with an additional structure called an **inner product** $\langle \cdot, \cdot \rangle$.
- The inner product:
  - Gives a way to define length and angle,
  - Satisfies linearity, symmetry (or conjugate symmetry in complex spaces), and positivity.
- The norm is induced by the inner product: $\|\vec{v}\| = \sqrt{\langle \vec{v}, \vec{v} \rangle}$
- The scalar (dot) product in $\mathbb{R}^n$ is a standard example.
- Allows definitions of **orthogonality** (perpendicularity).

Euclidean Space

- A **Euclidean space** is a finite-dimensional inner product space over $\mathbb{R}$ with the standard dot product.
- Examples: $\mathbb{R}^2, \mathbb{R}^3, \ldots$
- Has geometric notions like angles, lengths, orthogonality, projections which you use in physical intuition.

Summary Table

Space Type	Structure Added	Key Features	Examples
Vector Space	Addition, scalar multiplication	Linear structure only	$\mathbb{R}^n$, polynomials
Normed Space	Norm $\|\cdot\|$	Length and distance	$\mathbb{R}^n$ with Euclidean norm, function spaces with L^p norms
Inner Product Space	Inner product $\langle \cdot, \cdot \rangle$	Angles, orthogonal projections, induced norm	$\mathbb{R}^n$ with dot product
Euclidean Space	Standard inner product in $\mathbb{R}^n$	Familiar geometric space	$\mathbb{R}^2, \mathbb{R}^3$

Norm and Inner Product

Norm

A **norm** is a function that assigns a *length* or *size* to each vector in a vector space, generalizing the familiar idea of length from geometry to more abstract spaces.

Formally, a norm on a vector space V over a field $\mathbb{R}$ or $\mathbb{C}$ is a function $\|\cdot\| : V \to \mathbb{R}$, that satisfies the following properties for all vectors $\vec{x}, \vec{y} \in V$ and scalars a:

1. **Non-negativity:** $\|\vec{x}\| \geq 0$ and $\|\vec{x}\| = 0 \iff \vec{x} = \vec{0}$

2. **Absolute homogeneity (scaling):** $\|a\vec{x}\| = |a| \cdot \|\vec{x}\|$, where $|a|$ is the absolute value or modulus of a.
3. **Triangle inequality:** $\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|$

Manhattan Norm (also called L^1 norm or Taxicab norm)

- **Definition:** The Manhattan norm of a vector $\vec{x} = (x_1, x_2, \ldots, x_n)$ in $\mathbb{R}^n$ is the sum of the absolute values of its components: $\|\vec{x}\|_1 = |x_1| + |x_2| + \cdots + |x_n|$
- **Intuition:** It measures distance as if you could only travel along grid lines, like counting city blocks in Manhattan (hence the name).
- **Use cases:** Often used in optimization, sparse data settings, and in machine learning when encouraging sparsity.

Euclidean Norm (also called L^2 norm or magnitude)

- **Definition:** The Euclidean norm of $\vec{x} = (x_1, \ldots, x_n)$ is the square root of the sum of the squares of the components: $\|\vec{x}\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$
- **Intuition:** This is the familiar geometric length ("straight-line distance") of the vector.
- **Use cases:** Fundamental in physics and geometry; widely used in statistics, data science, and optimization tasks due to its smoothness and nice mathematical properties.

Infinity Norm (also called L^∞ norm or Max norm)

- **Definition:** The infinity norm of $\vec{x} = (x_1, \ldots, x_n)$ is the largest absolute value among the components: $\|\vec{x}\|_\infty = \max(|x_1|, |x_2|, \ldots, |x_n|)$
- **Intuition:** Measures the "maximum stretch" of a vector along any coordinate direction.
- **Use cases:** Used in optimization formulations focusing on worst-case or maximum deviation scenarios, uniform convergence analysis.

Inner Product

An **inner product** is a generalization of the familiar dot product that applies to abstract vector spaces, allowing you to define and measure angles, lengths, and orthogonality within the space.

An inner product on a vector space V (over real or complex numbers) is a function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{F}$, that takes two vectors $\vec{u}, \vec{v} \in V$ and returns a scalar, satisfying the following properties for all vectors $\vec{u}, \vec{v}, \vec{w} \in V$ and scalars a:

1. **Linearity in the first argument:** $\langle a\vec{u} + \vec{v}, \vec{w} \rangle = a\langle \vec{u}, \vec{w} \rangle + \langle \vec{v}, \vec{w} \rangle$
2. **Conjugate symmetry (or symmetry in real spaces):** $\langle \vec{u}, \vec{v} \rangle = \overline{\langle \vec{v}, \vec{u} \rangle}$. For real vector spaces, this reduces to $\langle \vec{u}, \vec{v} \rangle = \langle \vec{v}, \vec{u} \rangle$.
3. **Positive-definiteness:** $\langle \vec{v}, \vec{v} \rangle \geq 0,$ and $\langle \vec{v}, \vec{v} \rangle = 0 \iff \vec{v} = \vec{0}$

What Does an Inner Product Do?

- **Generalizes the dot product:** It allows multiplication of vectors resulting in a scalar, capturing geometric intuitions of length and angle.
- **Defines length (norm):** $\|\vec{v}\| = \sqrt{\langle \vec{v}, \vec{v} \rangle}$
- **Defines angle between vectors:** $\cos \theta = \frac{\langle \vec{u}, \vec{v} \rangle}{\|\vec{u}\| \|\vec{v}\|}$
- **Determines orthogonality:** Two vectors $\vec{u}$ and $\vec{v}$ are orthogonal if: $\langle \vec{u}, \vec{v} \rangle = 0$

Examples

- In $\mathbb{R}^n$ with real numbers, the inner product is the usual **dot product**: $\langle \vec{u}, \vec{v} \rangle = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$
- In complex vector spaces, the inner product generalizes to the **Hermitian inner product**, where conjugation is applied to ensure positive-definiteness.

Euclidean Spaces

Definition of Euclidean Space

- A Euclidean space $\mathbb{E}^n$ (or $\mathbf{E}^n$) is a **finite-dimensional vector space over the real numbers equipped with a standard inner product (dot product)**.
- The elements of $\mathbb{E}^n$ are called **points**.
- The associated vector space $\overrightarrow{\mathbb{E}^n}$ consists of vectors called **Euclidean vectors**.
- The inner product allows defining lengths, angles, and distances consistent with classical geometry.

Properties of Euclidean Space

- **Dimension:** The integer n defines the dimension (e.g., line is 1D, plane is 2D, space is 3D).
- **Distance:** The distance between two points $P, Q \in \mathbb{E}^n$ is given by the norm induced by the inner product:

$$d(P, Q) = \|\overrightarrow{PQ}\| = \sqrt{(Q_1 - P_1)^2 + (Q_2 - P_2)^2 + \cdots + (Q_n - P_n)^2}$$

- **No preferred origin or direction:** Any point can serve as origin, and the space is uniform in all directions.
- **Affine structure:** You can add vectors to points and subtract points to get vectors; this follows rules making the space behave nicely geometrically.

Examples

- The set of tuples $(x_1, x_2, \ldots, x_n)$ of real numbers with usual vector addition and dot product is the standard Euclidean space.

- The classical 2D and 3D geometry spaces are special cases ($\mathbb{E}^2$, $\mathbb{E}^3$).

Subspaces

What is a Subspace?

A **subspace** is a "smaller" vector space inside a bigger one. For example, a line or a plane through the origin inside $\mathbb{R}^3$.

Three Key Criteria

A subset W of a vector space V is a **subspace** if:

1. **Zero Vector:** $\vec{0} \in W$.
2. **Closed Under Addition:** For any $\vec{u}, \vec{v} \in W$, the sum $\vec{u} + \vec{v} \in W$.
3. **Closed Under Scalar Multiplication:** For any $\vec{v} \in W$ and any scalar a, $a\vec{v} \in W$.

If you see a set that doesn't satisfy these, it's not a subspace.

Quick Examples

- **Entire vector space** itself: Always a subspace.
- **Line through origin:** E.g., all multiples of $(1, 2)$ inside $\mathbb{R}^2$.
- **Plane through origin:** E.g., all $(x, y, 0)$ inside $\mathbb{R}^3$.
- **Set of solutions to a homogeneous linear equation:** Like $ax + by = 0$, forms a subspace.

Not Subspaces (Counterexamples)

- **Line not through origin:** E.g., $(1, 2) + t(3, 4)$ for $t \in \mathbb{R}$, where $(1, 2)$ isn't the origin — this fails the zero vector criterion.
- **Subset not closed under addition/scalar multiplication:** E.g., only vectors of length 1.

Why Are Subspaces Important?

- They often describe **solution sets** to linear systems.
- They're the backbone of **span**, **basis**, and **dimension** concepts.

Vector Geometry

Angles Between Vectors

The angle θ between two vectors $\vec{a}$ and $\vec{b}$ in $\mathbb{R}^n$ uses their dot product: $\cos\theta = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| \, |\vec{b}|}$

To find θ: $\theta = \cos^{-1}\left(\frac{\vec{a} \cdot \vec{b}}{|\vec{a}| \, |\vec{b}|}\right)$

Here:

- $\mathbf{a} \cdot \mathbf{b}$ is the dot product of vectors $\mathbf{a}$ and $\mathbf{b}$.
- $|\mathbf{a}|$ and $|\mathbf{b}|$ are the magnitudes (lengths) of the vectors.
- $\cos^{-1}$ denotes the inverse cosine function (also called arccosine).

This formula computes the angle in radians (or degrees if converted) between the two vectors by dividing their dot product by the product of their magnitudes and then taking the inverse cosine of the result.

Vector & Scalar Projections

Scalar projection (also called "component") of $\vec{a}$ onto $\vec{b}$: $\text{comp}_{\vec{b}}\vec{a} = \frac{\vec{a}\cdot\vec{b}}{|\vec{b}|}$. This is a scalar value that gives the magnitude of this projection (tells you how long the shadow of $\mathbf{a}$ is when cast on $\mathbf{b}$, without the direction.

Vector projection puts the scalar projection in the direction of $\vec{b}$: $\text{proj}_{\vec{b}}\vec{a} = \frac{\vec{a}\cdot\vec{b}}{|\vec{b}|^2}\vec{b}$.

Here, the dot product $\mathbf{a} \cdot \mathbf{b}$ measures how much $\mathbf{a}$ aligns with $\mathbf{b}$, and dividing by $|\mathbf{b}|^2$ normalizes by the length of $\mathbf{b}$, so the result scales vector $\mathbf{b}$ accordingly.

Think of it as the **"shadow"** of $\vec{a}$ on $\vec{b}$. The vector projection of a vector $\mathbf{a}$ onto another vector $\mathbf{b}$ is a vector that represents the component of $\mathbf{a}$ that points in the direction of $\mathbf{b}$. It essentially gives the "shadow" or "footprint" of $\mathbf{a}$ along $\mathbf{b}$.

In summary:

- Vector projection gives a vector along $\mathbf{b}$ showing how much of $\mathbf{a}$ lies in the direction of $\mathbf{b}$.
- Scalar projection gives the length (magnitude) of the vector projection, a scalar value.

Orthogonal Projections

***Orthogonalization** is the process of transforming a set of vectors or features into a new set where each pair is independent or "perpendicular" to each other. This is achieved by creating vectors with a dot product of zero, making them mutually orthogonal, and is often done using methods like the Gram-Schmidt process.*

Orthogonal projection in linear algebra is the process of projecting a vector onto a subspace such that the difference between the original vector and the projection is orthogonal (perpendicular) to that subspace.

- Given a vector $\mathbf{x}$ and a subspace W, the **orthogonal projection** of $\mathbf{x}$ onto W is the vector $\mathbf{x}_W \in W$ closest to $\mathbf{x}$.
- The vector difference $\mathbf{x} - \mathbf{x}_W$ is orthogonal to every vector in W, meaning $\mathbf{x} - \mathbf{x}_W$ is perpendicular to the subspace.

- For projection onto a line spanned by a unit vector $\mathbf{u}$, the orthogonal projection of $\mathbf{x}$ is given by: $P_{\mathbf{u}}(\mathbf{x}) = (\mathbf{x} \cdot \mathbf{u})\mathbf{u}$, where $\mathbf{u}$ is a unit vector ($|\mathbf{u}| = 1$).
- More generally, if $\mathbf{u}$ is not a unit vector, the projection is: $P_{\mathbf{u}}(\mathbf{x}) = \frac{\mathbf{x} \cdot \mathbf{u}}{|\mathbf{u}|^2}\mathbf{u}$
- Orthogonal projections are **idempotent** operators: applying the projection operator twice is the same as applying it once.
- Orthogonal projections are also **self-adjoint (symmetric)** linear operators, ensuring geometric and algebraic properties that make them fundamental in decomposing vectors and spaces.
- Applications include minimizing the distance between vectors and subspaces, least squares problems, and decomposing signals or data along orthogonal components.

Suppose you want the projection onto a subspace (like a plane) that is defined by some vector $\vec{b}$. The vector projection $\mathrm{proj}_{\vec{b}}\vec{a}$ is always **orthogonal** to the difference $\vec{a} - \mathrm{proj}_{\vec{b}}\vec{a}$ (the part left over is perpendicular to $\vec{b}$).

In a nutshell:

- The **orthogonal projection** of $\vec{a}$ onto a line (or vector) is its vector projection.
- The part perpendicular to $\vec{b}$ is called the "orthogonal complement".

Normals

A **normal** to a surface (like a plane) is a vector perpendicular to that surface.

- For a plane defined by $Ax + By + Cz = D$, the vector (A, B, C) is normal to the plane.
- Normals are important for defining hyperplanes, reflecting vectors, optimization, and computer graphics.

Normal Vector

A **normal vector** to a surface or a plane is a vector that is **perpendicular** (orthogonal) to that surface or plane at a given point. For example:

- In 2D or 3D, a vector normal to a line or a plane is perpendicular to every vector lying along that line or plane.
- Normal vectors are fundamental for defining orientations of surfaces, calculating angles between surfaces, and for various applications such as physics (e.g., normal force) and computer graphics (lighting, shading).

Unit Normal Vector

A **unit normal vector** is a normal vector that has been normalized to have a magnitude (length) of exactly 1. It is obtained by dividing the normal vector by its magnitude: $\hat{\mathbf{n}} = \frac{\mathbf{n}}{|\mathbf{n}|}$

Unit normal vectors are used when only the direction of the normal matters, independent of its length.

Examples and Properties

- For a plane defined by the equation $ax + by + cz = d$, the vector $\mathbf{n} = (a, b, c)$ is a normal vector to the plane.
- Normal vectors are used to compute projections, reflections, and determine orientations.
- The normal vector is often denoted by $\mathbf{n}$ or $\hat{\mathbf{n}}$ when unit normalized.

Hyperplanes

A **hyperplane** is a fundamental concept in linear algebra and geometry that generalizes the idea of a plane to higher dimensions.

Definition and properties

- In an n-dimensional space, a hyperplane is an $(n-1)$-dimensional flat affine subspace. This means it has one dimension less than the ambient space.
- For example:
    - In 2D, a hyperplane is a line (1D).
    - In 3D, a hyperplane is a plane (2D).
    - In 4D, it is a 3-dimensional "slice" or subspace.
- A hyperplane divides the space into two half-spaces, which are the sets of points lying on either side of the hyperplane.

Equation of a Hyperplanes

- The equation of a hyperplane is a linear equation of the form: $a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = b$ where $a_1, a_2, \ldots, a_n$ are real coefficients and b is a constant.
- The vector $\mathbf{w} = (a_1, a_2, \ldots, a_n)$ is **normal** (perpendicular) to the hyperplane.
- A point $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ lies on the hyperplane if it satisfies the equation.
- If we write $\mathbf{w} \cdot \mathbf{x} + b = 0$, then:
    - $\mathbf{w} \cdot \mathbf{x} + b > 0$ means the point is on one side of the hyperplane.
    - $\mathbf{w} \cdot \mathbf{x} + b < 0$ means the point is on the other side.

Applications

- Hyperplanes are widely used in machine learning for classification (e.g., Support Vector Machines) where they separate data into classes.
- They serve in optimization problems and geometric computations to define boundaries and constraints.

- In vector spaces, hyperplanes help decompose the space into meaningful regions by "cutting" it into parts.

Vector Arithmetic

Vector arithmetic is fundamental in linear algebra and involves operations with vectors such as addition, subtraction, and scalar multiplication.

- **Vector addition**: For vectors $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ and $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ in $\mathbb{R}^n$, their sum is $\mathbf{u} + \mathbf{v} = (u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n)$
- **Vector subtraction**: Vector subtraction $\mathbf{a} - \mathbf{b}$ can be defined as the addition of the first vector $\mathbf{a}$ with the negative of the second vector $-\mathbf{b}$: $\mathbf{a} - \mathbf{b} = \mathbf{a} + (-\mathbf{b})$
- **Scalar multiplication**: For a scalar a and vector $\mathbf{v}$, $a\mathbf{v} = (av_1, av_2, \ldots, av_n)$

These operations follow commutative, associative, and distributive laws analogous to ordinary algebra.

Vector algebra: Span, Basis, Generating Set, Coordinates

Span

- The **span** of a set of vectors $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ in a vector space V is **the set of all possible linear combinations** of those vectors:

$$\text{span}(S) = \{c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n \mid c_i \in \mathbb{F}\}$$

- It forms a **subspace** of V.
- Intuitively, span is all the points you can "reach" by combining vectors in S with scalar multiplications and additions.

Generating Set

- A set of vectors S is called a **generating set** or **spanning set** if: $\text{span}(S) = V$
- This means every vector in the vector space V can be expressed as a linear combination of vectors in S.
- Generating sets are a starting point for building or describing the entire vector space.

Basis

- A **basis** of a vector space V is a generating set that is also **linearly independent**.
- Properties of a basis $B = \{\vec{b}_1, \ldots, \vec{b}_n\}$:
  - **Spanning:** $V = \text{span}(B)$
  - **Linear independence:** No vector in B can be written as a linear combination of the others.

- Every vector in V can be **uniquely expressed** as a linear combination of basis vectors.
- The number of basis vectors is called the **dimension** of V.

Example:
In $\mathbb{R}^2$, the canonical basis is: $\{(1,0),(0,1)\}$, which spans the plane, and these two vectors are linearly independent.

Coordinates

- Given a basis $B = \{\vec{b}_1, \ldots, \vec{b}_n\}$, every vector $\vec{v} \in V$ can be uniquely written as:

$$\vec{v} = c_1\vec{b}_1 + c_2\vec{b}_2 + \cdots + c_n\vec{b}_n$$

- The scalars $(c_1, c_2, \ldots, c_n)$ are called the **coordinates** of $\vec{v}$ with respect to the basis B.
- Coordinates provide a way to represent vectors as tuples of numbers relative to the basis.

Summary Table

Concept	Definition	Key Property
Span	All linear combinations of a vector set	Forms a subspace
Generating Set	A set whose span equals the entire space	Spans whole vector space
Basis	Generating set that is linearly independent	Unique representation of vectors
Coordinates	Scalars expressing vector concerning a basis	Coordinates uniquely specify a vector

Linear Combination and Linear Dependence

- A **linear combination** of vectors is an expression of the form $a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \ldots + a_k\mathbf{v}_k$, where a_i are scalars.
- Vectors are **linearly dependent** if there exist some scalars, not all zero, such that the linear combination equals the zero vector.
- If the only scalars that satisfy this are all zero, the vectors are **linearly independent**.

Orthonormal Sets and Basis

Defintion

An **orthonormal set** is a set of vectors in a vector space that satisfy two properties simultaneously:

- The vectors are **orthogonal** to each other, meaning the inner product of any two distinct vectors is zero: $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = 0 \quad \text{for } i \neq j$
- Each vector is a **unit vector** with length 1: $\|\mathbf{u}_i\| = 1$

An **orthonormal basis** is an orthonormal set of vectors that also forms a **basis** of a vector space. This means:

- The orthonormal vectors span the entire vector space.
- They are linearly independent.
- Every vector in the space can be uniquely expressed as a linear combination of these orthonormal basis vectors.

Why orthonormal bases are useful:

- Simplify calculations of projections and inner products.
- Representation of vectors in this basis involves simply taking dot products with basis vectors.
- The Gram-Schmidt process is often used to convert any basis into an orthonormal basis.
- In Euclidean space $\mathbb{R}^n$, the standard basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is an orthonormal basis.

Gram-Schmidt Process

Purpose of Gram-Schmidt Process

- Turns a non-orthogonal basis into an orthonormal basis.
- Preserves the span of the original vectors.
- Useful in simplifying many linear algebra problems, including least squares and QR factorization.

The Process Step-by-Step

Suppose you have a set of linearly independent vectors in a vector space: $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$. The goal is to produce an orthonormal set $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$, that spans the same subspace.

Step 1: Initialize the first orthogonal vector

Set: $\mathbf{u}_1 = \mathbf{v}_1$

Normalize it: $\mathbf{e}_1 = \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|}$, where $\|\mathbf{u}_1\|$ is the norm (length) of $\mathbf{u}_1$.

Step 2: Orthogonalize the second vector

Subtract from $\mathbf{v}_2$ its projection onto $\mathbf{e}_1$:

$$\mathbf{u}_2 = \mathbf{v}_2 - \text{proj}_{\mathbf{e}_1}(\mathbf{v}_2) = \mathbf{v}_2 - \langle \mathbf{v}_2, \mathbf{e}_1 \rangle \mathbf{e}_1$$

Normalize: $\mathbf{e}_2 = \frac{\mathbf{u}_2}{\|\mathbf{u}_2\|}$

Step 3: Orthogonalize the k-th vector (for $k = 3$ to n)

Remove projections onto all previous orthonormal vectors:

$$\mathbf{u}_k = \mathbf{v}_k - \sum_{j=1}^{k-1} \text{proj}_{\mathbf{e}_j}(\mathbf{v}_k) = \mathbf{v}_k - \sum_{j=1}^{k-1} \langle \mathbf{v}_k, \mathbf{e}_j \rangle \mathbf{e}_j$$

Normalize: $\mathbf{e}_k = \frac{\mathbf{u}_k}{\|\mathbf{u}_k\|}$

- $\langle \mathbf{v}, \mathbf{e} \rangle$ is the inner product (dot product in Euclidean space).
- Projection formula: $\text{proj}_{\mathbf{e}}(\mathbf{v}) = \langle \mathbf{v}, \mathbf{e} \rangle \mathbf{e}$
- Each $\mathbf{u}_k$ is orthogonal to all previous $\mathbf{e}_j$ for $j < k$.
- The final set $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is orthonormal.

Key Properties

- The set $\{\mathbf{e}_1, ..., \mathbf{e}_n\}$ is **orthonormal**: vectors are mutually perpendicular and have unit length.
- The vectors $\mathbf{e}_1, ..., \mathbf{e}_n$ span the same subspace as the original $\mathbf{v}_1, ..., \mathbf{v}_n$.
- The process is **sequential**: each new vector is orthogonalized relative to all previously computed orthonormal vectors.

Applications

- Simplifying computation in solving linear systems.
- Basis construction in vector spaces.
- QR factorization where a matrix A is decomposed as $A = QR$, with Q orthogonal and R upper-triangular.
- Numerical stability improvements in computations.

Linear Maps (Transformations)

Linear maps (also called linear transformations) are functions between vector spaces that preserve the structure of vector addition and scalar multiplication. They are fundamental in linear algebra, connecting abstract vector spaces through operations that respect their linear structure.

Definition and Properties

A function $T : V \to W$ between vector spaces V and W (over the same field) is a **linear map** if for all vectors $\vec{u}, \vec{v} \in V$ and scalars c, the following hold:

- **Additivity:** $T(\vec{u} + \vec{v}) = T(\vec{u}) + T(\vec{v})$
- **Homogeneity (scalar multiplication preservation):** $T(c\vec{v}) = cT(\vec{v})$
- From these, it follows that $T(\vec{0}) = \vec{0}$ (zero vector maps to zero vector).

Range and Kernel

Kernel of a Linear Map

Given a linear map $T : V \to W$ between vector spaces V and W:

- The **kernel** (or null space) of T is the set of all vectors in V that map to the zero vector in W: $\ker(T) = \{\vec{v} \in V : T(\vec{v}) = \vec{0}\}$
- **Properties:**
  - The kernel is a **subspace** of the domain V.
  - It captures vectors "collapsed" to zero by T.
  - The dimension of the kernel is called the **nullity** of T.
  - If the kernel contains only the zero vector, T is **injective** (one-to-one).

Range of a Linear Map

- The **range** (also called image) of T is the set of all vectors in W that are images of vectors in V: $\mathrm{Range}(T) = \{T(\vec{v}) : \vec{v} \in V\}$
- **Properties:**
  - The range is a **subspace** of the codomain W.
  - It describes all possible outputs of the linear transformation.
  - The dimension of the range is called the **rank** of T.
  - If the range equals W, then T is **surjective** (onto).

Fundamental Theorem of Linear Maps (Rank-Nullity Theorem)

For a linear map $T : V \to W$ where V is finite-dimensional, the following holds: $\dim(V) = \dim(\mathrm{Ker}(T)) + \dim(\mathrm{Range}(T))$. Meaning the dimension of V splits into the dimension of the kernel plus the dimension of the range.

Statement of the Theorem

Let $T : V \to W$ be a linear map between finite-dimensional vector spaces. Then, **rank(T) + nullity(T) = dim(V)**, where

- **rank(T) = dim(Range(T))**, the dimension of the image (range) of T,
- **nullity(T) = dim(Ker(T))**, the dimension of the kernel (null space) of T,
- **dim(V)** is the dimension of the domain vector space V.

Interpretation

- The theorem says the dimension of the domain V splits into two parts:
  1. The dimension of all vectors sent to zero by T (the nullity),
  2. The dimension of the subspace "covered" by T in W (the rank).
- Intuitively, the linear transformation T "collapses" the kernel to the zero vector and "injects" the rest into the image.

Sketch of the Proof

1. **Basis for Kernel:**
   Start by finding a basis $\{\vec{u}_1, \ldots, \vec{u}_k\}$ of $\ker(T)$.
2. **Extend to Basis of Domain:**
   Extend this basis to a basis of the whole space V by adding vectors $\vec{v}_1, \ldots, \vec{v}_m$ so that $\{\vec{u}_1, \ldots, \vec{u}_k, \vec{v}_1, \ldots, \vec{v}_m\}$ forms a basis of V. This means $\dim(V) = k + m$, where $k = \text{nullity}(T)$.
3. **Image of Extended Basis:**
   Consider the images under T of the additional basis vectors: $T(\vec{v}_1), \ldots, T(\vec{v}_m)$. These vectors span $\text{Range}(T)$.
4. **Linear Independence of Images:**
   The set $\{T(\vec{v}_1), \ldots, T(\vec{v}_m)\}$ is linearly independent (otherwise, one could reduce m).
5. **Conclusion:**
   Hence, $\dim(\text{Range}(T)) = m$ and $\dim(V) = k + m = \text{nullity}(T) + \text{rank}(T)$

Composition and Sum of Linear Maps

- **Composition:** If $S : U \to V$ and $T : V \to W$ are linear maps, their composition $T \circ S : U \to W$ defined by $(T \circ S)(\vec{u}) = T(S(\vec{u}))$ is also a linear map.
- **Sum:** If $T_1, T_2 : V \to W$ are linear maps, their sum $T_1 + T_2 : V \to W$ defined by $(T_1 + T_2)(\vec{v}) = T_1(\vec{v}) + T_2(\vec{v})$ is a linear map.

Fundamental Subspaces Associated With a Linear Map

The **fundamental subspaces** of a linear map $T : V \to W$ are:

- The **kernel** subspace in V.
- The **range** subspace in W.

These subspaces capture the essence of T's behavior in terms of which vectors get mapped to zero and which vectors in W are reachable as images.

Matrices

Matrix Algebra

Basic Operations on Matrices

Given matrices A and B of the same dimension (same number of rows and columns), and a scalar k:

- **Addition:** $(A + B)_{ij} = A_{ij} + B_{ij}$
- **Subtraction:** $(A - B)_{ij} = A_{ij} - B_{ij}$
- **Scalar Multiplication:** $(kA)_{ij} = k \times A_{ij}$

Matrix Multiplication

- If A is an $m \times n$ matrix and B is an $n \times p$ matrix, their product $C = AB$ is an $m \times p$ matrix with entries: $C_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}$
- Note: Matrix multiplication is **associative** but generally **not commutative**.
- The product represents composition of linear transformations.

Identity Matrix and Inverse

- The $n \times n$ **identity matrix** I_n has 1's on the diagonal and 0's elsewhere.
- For any $n \times n$ matrix A, if an inverse A^{-1} exists: $AA^{-1} = A^{-1}A = I_n$
- Inverse exists only if A is **non-singular** (full rank).

Transpose of a Matrix

- The **transpose** of an $m \times n$ matrix A, denoted A^T, is the $n \times m$ matrix where rows and columns are swapped: $(A^T)_{ij} = A_{ji}$
- Important property: $(AB)^T = B^T A^T$.

Determinant

- Defined for **square matrices**, the determinant is a scalar that encodes important properties:
    - If the determinant is 0, the matrix is singular (non-invertible).
    - The determinant is used in solving linear systems, finding eigenvalues, and more.

Matrix Rank

- The **rank** is the dimension of the column space (maximum number of linearly independent columns).
- Determines the solvability of systems and matrix invertibility.

Special Types of Matrices

- **Square:** Equal number of rows and columns.
- **Diagonal:** Non-zero only on the diagonal.
- **Symmetric:** $A = A^T$.
- **Orthogonal:** $A^T A = I$.
- **Zero matrix:** All entries zero.

Practical Example

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

- **Addition:**
$$A + B = \begin{bmatrix} 1+5 & 2+6 \\ 3+7 & 4+8 \end{bmatrix} = \begin{bmatrix} 6 & 8 \\ 10 & 12 \end{bmatrix}$$

- **Multiplication:**

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

Cofactor Expansion for Determinants — Detailed Explanation

Minor and Cofactor To understand cofactor expansion, first define:

- The **minor** M_{ij} of an element a_{ij} in an $n \times n$ matrix A is the determinant of the $(n-1) \times (n-1)$ matrix obtained by deleting the i-th row and j-th column from A.
- The **cofactor** C_{ij} of a_{ij} is: $C_{ij} = (-1)^{i+j} M_{ij}$. The sign alternates according to the position to account for orientation in determinant calculation.

Cofactor Expansion Formula The **cofactor expansion** (or Laplace expansion) allows the determinant of A to be computed by expanding along any row i or column j:

- Expansion along row i: $\det(A) = \sum_{j=1}^{n} a_{ij} C_{ij} = \sum_{j=1}^{n} a_{ij} (-1)^{i+j} M_{ij}$
- Expansion along column j: $\det(A) = \sum_{i=1}^{n} a_{ij} C_{ij} = \sum_{i=1}^{n} a_{ij} (-1)^{i+j} M_{ij}$

Choosing a row or column with many zeros simplifies calculation because terms with zero coefficients vanish.

Step-by-Step Example (3x3 Matrix) Let

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

Expand along the first row:

- For $a_{11} = 1$, minor M_{11} is determinant of

$$\begin{bmatrix} 0 & 4 \\ 6 & 1 \end{bmatrix} = (0)(1) - (4)(6) = -24$$

- Cofactor $C_{11} = (-1)^{1+1} \times -24 = +(-24) = -24$.

- For $a_{12} = 2$, minor M_{12} is determinant of

$$\begin{bmatrix} 3 & 4 \\ 5 & 1 \end{bmatrix} = (3)(1) - (4)(5) = 3 - 20 = -17$$

- Cofactor $C_{12} = (-1)^{1+2} \times -17 = -(-17) = +17$.

- For $a_{13} = 0$, term contributes zero.

Calculate determinant: $\det(A) = 1 \times (-24) + 2 \times 17 + 0 = -24 + 34 = 10$

Properties of Cofactor Expansion

- Expanding along any row or column gives the same determinant.
- Recursively reduces determinant calculation to smaller matrices (minors).
- The method becomes inefficient for very large matrices; other numerical methods preferred.

Standard Matrix of a Linear Map

What is the Standard Matrix?

For a linear transformation (linear map) $T : \mathbb{R}^n \to \mathbb{R}^m$, there exists a unique matrix $A \in \mathbb{R}^{m \times n}$ such that for every vector $\mathbf{x} \in \mathbb{R}^n$, $T(\mathbf{x}) = A\mathbf{x}$.

This matrix A is called the **standard matrix** (or matrix representation) of the linear map T.

How is the Standard Matrix Constructed?

1. **Start with the standard basis of the domain:**
   The standard basis vectors in $\mathbb{R}^n$ are

   $$\mathbf{e}_1 = (1, 0, \ldots, 0), \quad \mathbf{e}_2 = (0, 1, \ldots, 0), \quad \ldots, \quad \mathbf{e}_n = (0, 0, \ldots, 1)$$

2. **Apply the transformation T to each basis vector:**
   Calculate $T(\mathbf{e}_1), \quad T(\mathbf{e}_2), \quad \ldots \quad T(\mathbf{e}_n)$ Each $T(\mathbf{e}_j)$ is a vector in $\mathbb{R}^m$.

3. **Form the matrix using these images as columns:**
   The matrix A is formed by taking the transformed basis vectors as columns:

   $$A = \begin{bmatrix} | & | & & | \\ T(\mathbf{e}_1) & T(\mathbf{e}_2) & \cdots & T(\mathbf{e}_n) \\ | & | & & | \end{bmatrix}$$

4. **Verification:**
   For any vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, since $\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n$, by linearity

   $$T(\mathbf{x}) = x_1 T(\mathbf{e}_1) + \cdots + x_n T(\mathbf{e}_n) = A\mathbf{x}$$

   which confirms the matrix represents the linear map.

Example

Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be defined by:

$$T \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 + x_2 \\ -x_3 \end{pmatrix}$$

- Compute images of standard basis vectors of $\mathbb{R}^3$:

$$T(\mathbf{e}_1) = T \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$T(\mathbf{e}_2) = T \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$T(\mathbf{e}_3) = T \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

- Form the matrix:

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

Why Is This Useful?

- It translates abstract linear maps into concrete matrix multiplication.
- Enables computational techniques to analyze linear maps.
- Essential for changing between bases and understanding the structure of transformations.
- Every linear transformation corresponds uniquely to a standard matrix once bases are fixed.

Transition Matrix

What is a Transition Matrix?

A **transition matrix** (also called a **change-of-basis matrix**) is a square matrix that describes how to convert (or translate) coordinate vectors of a vector from one basis to another basis within the same vector space.

Setting

- Suppose V is an n-dimensional vector space with two ordered bases:

$$\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}, \quad \mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n\}$$

- Any vector $\mathbf{v} \in V$ has different coordinate representations with respect to bases $\mathcal{B}$ and $\mathcal{C}$:

$$[\mathbf{v}]_{\mathcal{B}}, \quad [\mathbf{v}]_{\mathcal{C}} \in \mathbb{R}^n$$

Purpose of Transition Matrix

- The transition matrix $P_{\mathcal{C} \leftarrow \mathcal{B}}$ converts the coordinates of $\mathbf{v}$ from basis $\mathcal{B}$ to basis $\mathcal{C}$: $[\mathbf{v}]_{\mathcal{C}} = P_{\mathcal{C} \leftarrow \mathcal{B}}[\mathbf{v}]_{\mathcal{B}}$
- It encapsulates the relationship between two bases.

How to Construct the Transition Matrix

1. Express each vector of the basis $\mathcal{B}$ in terms of the basis $\mathcal{C}$: $\mathbf{b}_j = a_{1j}\mathbf{c}_1 + a_{2j}\mathbf{c}_2 + \cdots + a_{nj}\mathbf{c}_n$
2. The coordinates $(a_{1j}, a_{2j}, \ldots, a_{nj})^{\mathrm{T}}$ form the j-th column of the matrix $P_{\mathcal{C} \leftarrow \mathcal{B}}$.

In matrix form: $P_{\mathcal{C} \leftarrow \mathcal{B}} = \begin{bmatrix} [\mathbf{b}_1]_{\mathcal{C}} & [\mathbf{b}_2]_{\mathcal{C}} & \cdots & [\mathbf{b}_n]_{\mathcal{C}} \end{bmatrix}$

Properties

- $P_{\mathcal{C} \leftarrow \mathcal{B}}$ is **invertible**.
- Its inverse is the transition matrix from $\mathcal{C}$ to $\mathcal{B}$: $(P_{\mathcal{C} \leftarrow \mathcal{B}})^{-1} = P_{\mathcal{B} \leftarrow \mathcal{C}}$
- Transition matrices allow switching between coordinate representations and analyzing linear transformations relative to different bases.

Matrix Rank

What is Matrix Rank?

The **rank** of a matrix A is the dimension of the vector space spanned by its columns, equivalently the maximum number of linearly independent columns in A.

By a fundamental theorem in linear algebra, the **column rank** and the **row rank** of a matrix are always equal; thus the rank can also be described as the maximum number of linearly independent rows.

Why is Rank Important?

- Rank measures the **"nondegenerateness"** of the matrix, i.e., how much information it conveys.
- It helps determine whether a system of linear equations has solutions:
  - Full rank (equal to the smaller of number of rows or columns) often indicates unique or full solutions.
  - Rank deficiency means dependencies exist between rows or columns.
- Rank guides matrix invertibility:

      – A square matrix is invertible if and only if its rank equals its size.
- Rank is used in matrix factorizations, dimensionality reduction, and more.

How to Find the Rank?

1. **Minor Method:**
   The rank is the size of the largest non-zero minor (determinant of a square submatrix) of A.

2. **Row Echelon Form:**
   Perform row operations to get the matrix into echelon form. The number of non-zero rows after reduction equals the rank.

3. **Normal Form:**
   Convert A to a block matrix with an identity matrix I_r in the top-left:

$$\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$$

Here, r is the rank.

Properties of Rank

- $\text{rank}(A) \leq \min(\text{number of rows}, \text{number of columns})$
- $\text{rank}(A) = \text{rank}(A^T)$
- If A is invertible (square), $\text{rank}(A) = \text{size of } A$
- Rank is invariant under elementary row operations.

Example

For the matrix

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

- Perform Gaussian elimination to echelon form:

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

- Number of non-zero rows is 2, so $\text{rank}(A) = 2$.

Null Space and Column Space

Definiton

- **Null space** (kernel) ($\mathbb{R}^n$): Set of vectors $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{0}$.

- Null space is a subspace of the domain; dimension called **nullity**.

- Geometrically, it is the set of vectors mapped to the zero vector by the linear transformation associated with A.

- It represents the freedom or degrees of freedom in the solution of the homogeneous system.

- If the null space consists only of the zero vector, the columns of AA are linearly independent.

- **Column space**($\mathbb{R}^m$): Span of the columns of A; a subspace of the codomain.

- It contains all possible vectors b for which the system $Ax = b$ has at least one solution.

- Dimensions (rank of A) satisfy Rank-Nullity Theorem: rank(A) + nullity(A) = number of columns

Relationship Between Null Space and Column Space

- Null space and column space live in **different spaces**:
    - Null space is in the domain space $\mathbb{R}^n$.
    - Column space is in the codomain space $\mathbb{R}^m$.
- The **rank-nullity theorem** relates their dimensions: rank(A) + nullity$(A) = n$, where n is the number of columns of A.
- This theorem captures how much of the input space is "collapsed" (null space) and how much is "preserved" (column space).

Visualization and Example

- Consider $A = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & -1 \end{bmatrix}$
    - Its column space is a subspace of $\mathbb{R}^2$.
    - Its null space may be illustrated as lines/planes in $\mathbb{R}^3$ orthogonal to row space.
- The null space vectors correspond to solutions of the homogeneous equation $A\mathbf{x} = \mathbf{0}$.
- The column space vectors correspond to all vectors that can be expressed as linear combinations of the columns.

Solving Systems of Linear Equations (SLE)

System of Linear Equations in Matrix Form

A system of linear equations with m equations and n unknowns can be written as: $A\mathbf{x} = \mathbf{b}$, where

- A is an $m \times n$ coefficient matrix,
- $\mathbf{x}$ is an $n \times 1$ column vector of unknowns,
- $\mathbf{b}$ is an $m \times 1$ column vector of constants.

Types of Solutions

- **Unique solution:** Exists if A is square and has full rank (rank $= n$).
- **No solution:** The system is inconsistent if $\mathbf{b}$ lies outside the column space of A.
- **Infinite solutions:** When the system is underdetermined or rank-deficient (rank $< n$), solutions depend on free parameters.

Methods of Solving SLE

Direct Methods

1. **Gaussian Elimination:**
   - Use row operations to reduce A to echelon or reduced row echelon form.
   - Solve the resulting triangular system by back-substitution.
   - Efficient for moderate-sized systems.
2. **Matrix Inversion:**
   - If A is square and invertible, solution is: $\mathbf{x} = A^{-1}\mathbf{b}$
   - Generally not preferred for large systems as inversion is costly.
3. **Cramer's Rule:**
   - For square $n \times n$ systems, uses determinants: $x_i = \frac{\det(A_i)}{\det(A)}$
   - A_i is the matrix A with the ith column replaced by $\mathbf{b}$.
   - Computationally expensive for large n.
4. **LU Decomposition:**
   - Factorize $A = LU$ where L is lower-triangular and U is upper-triangular.
   - Solve via forward and backward substitution.
   - Efficient for repeated solves with same A.

Gaussian elimination   Gaussian elimination is a systematic method for solving systems of linear equations. It transforms the system's augmented matrix into an upper triangular (or row echelon) form from which solutions can be found by back substitution.

Step 1: Form the Augmented Matrix

Convert the system of linear equations into an augmented matrix combining coefficients with constants on the right. For example, the system:

$$\begin{cases} 2x + y - z = 8 \\ -3x - y + 2z = -11 \\ -2x + y + 2z = -3 \end{cases}$$

becomes

$$\begin{bmatrix} 2 & 1 & -1 & | & 8 \\ -3 & -1 & 2 & | & -11 \\ -2 & 1 & 2 & | & -3 \end{bmatrix}$$

Step 2: Forward Elimination to Row Echelon Form

Use elementary row operations to manipulate the matrix:

- **Swap rows**
- **Multiply a row by a nonzero scalar**
- **Add/subtract multiples of one row to/from another**

The goal is to create zeros under the leading coefficients (pivots) in each column.

Example steps:

- Use row 1 to eliminate x from rows 2 and 3:

    - Multiply row 1 by $\frac{3}{2}$ and add to row 2.
    - Multiply row 1 by 1 and add to row 3.

- This yields zeros in first column below pivot.

- Move to second column and eliminate entries below pivot in row 2 similarly.

At the end, matrix looks like:

$$\begin{bmatrix} 2 & 1 & -1 & | & 8 \\ 0 & \text{value} & \text{value} & | & \text{value} \\ 0 & 0 & \text{value} & | & \text{value} \end{bmatrix}$$

Step 3: Back Substitution

Once the matrix is in upper triangular form, solve for variables starting from the last row upward:

- Solve the last equation for the last variable.
- Substitute back into above equations to find other variables stepwise.

For the example, after elimination:

- Solve for z from row 3.

- Substitute z into row 2 to find y.
- Substitute y, z into row 1 to find x.

Key Points

- Gaussian elimination works with any system size.
- Produces either a unique solution, infinitely many solutions (when free variables appear), or no solution (when contradiction occurs).
- Efficient and standard algorithm in numerical linear algebra.

Classification of Matrices

Based on Shape and Size

- **Row Matrix:** Has only one row and multiple columns.
  Example:
  $$\begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$$

- **Column Matrix:** Has one column and multiple rows.
  Example:
  $$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

- **Rectangular Matrix:** Number of rows different from number of columns ($m \times n$ where $m \neq n$).
  Example: 2×3 matrix.

- **Square Matrix:** Number of rows equals number of columns (n × n).
  Example: 3×3 matrix.

Based on Element Values and Patterns

- **Zero Matrix:** All elements are zero.
- **Diagonal Matrix:** Square matrix with non-zero entries only on the main diagonal.
- **Scalar Matrix:** Diagonal matrix with all diagonal entries equal.
- **Identity Matrix:** Scalar matrix with diagonal entries equal to 1; acts as multiplicative identity.
- **Upper Triangular Matrix:** All entries below the main diagonal are zero.
- **Lower Triangular Matrix:** All entries above the main diagonal are zero.
- **Symmetric Matrix:** $A = A^T$ (equal to its transpose).
- **Skew-Symmetric Matrix:** $A = -A^T$.
- **Orthogonal Matrix:** Square matrix with orthonormal rows and columns; satisfies $A^T A = I$.

Based on Matrix Properties

- **Singular Matrix:** Square matrix with determinant zero; not invertible.
- **Non-Singular Matrix:** Square matrix with non-zero determinant; invertible.
- **Hermitian Matrix:** Complex square matrix equal to its conjugate transpose.
- **Skew-Hermitian Matrix:** Complex matrix equal to negative of its conjugate transpose.

Eigenvalues and Eigenvectors

Definiton

An **eigenvector** of a square matrix A is a non-zero vector x such that $Ax = \lambda x$, where λ is a scalar known as the **eigenvalue**. In essence, eigenvectors represent directions in which the action of the matrix is a simple stretching or compression by a factor of λ.

Key Properties

- A matrix may have real or complex eigenvalues.
- The collection of all eigenvectors corresponding to distinct eigenvalues are linearly independent.
- The **characteristic equation** is given by $\det(A - \lambda I) = 0$.

Existence of an Eigenbasis and Diagonalization

A matrix A is **diagonalizable** if it has a basis of eigenvectors called an **eigenbasis**. This means that $A = PDP^{-1}$, where D is a diagonal matrix with eigenvalues of A on the diagonal, and P contains the corresponding eigenvectors as its columns.

Eigenspaces and Geometric Multiplicity

Given a square matrix A acting on a vector space V, an **eigenspace** corresponding to an eigenvalue λ is defined as: $E_\lambda = \{v \in V \mid Av = \lambda v\}$.

This is equivalent to the kernel (null space) of $A - \lambda I$, meaning: $E_\lambda = \ker(A - \lambda I)$.

The **geometric multiplicity** of an eigenvalue λ is the dimension of its eigenspace; it tells us how many linearly independent eigenvectors correspond to λ.

Each eigenspace forms a subspace, and eigenvectors associated with **distinct eigenvalues** are always linearly independent. Hence, the sum of the dimensions of all eigenspaces cannot exceed the size n of the space.

Existence of an Eigenbasis

An **eigenbasis** is a basis for the entire space formed by linearly independent eigenvectors of A. This happens *if and only if* the sum of the geometric multiplicities of all eigenvalues equals n, the dimension of the space: $\sum_{i=1}^{k} \dim(E_{\lambda_i}) = n$, where $\lambda_1, \ldots, \lambda_k$ are the distinct eigenvalues of A.

If this condition is satisfied, the matrix is said to be **diagonalizable**, and an eigenbasis exists. Conversely, if the sum is less than n, no eigenbasis can be formed, and the matrix cannot be diagonalized.

Connection Between Eigenbasis and Diagonalization

A matrix A is **diagonalizable** if it has an eigenbasis. This means there exists an invertible matrix P whose columns are the eigenvectors of A, and a diagonal matrix D such that: $A = PDP^{-1}$.

Here, D has the eigenvalues of A on its diagonal. The diagonalization theorem thus states:

Theorem: A matrix A is diagonalizable if and only if it has n linearly independent eigenvectors — equivalently, if it admits an eigenbasis.

Examples and Observations

- **Diagonal matrices** are trivially diagonalizable; their standard basis vectors form an eigenbasis with the diagonal entries as eigenvalues.
- **Symmetric matrices** over real numbers always have an eigenbasis due to the spectral theorem — their eigenvectors can be chosen orthogonal.
- **Defective matrices** (e.g., having repeated eigenvalues but insufficient independent eigenvectors) **do not** have an eigenbasis. For example, a matrix with a double eigenvalue but only one eigenvector cannot be diagonalized.

Geometric Interpretation

When an eigenbasis exists, the linear transformation represented by A acts simply as a **stretching** operation along each eigendirection. Transforming to the eigenbasis "aligns" the coordinate system with these natural directions, and in this basis, the action of A becomes purely scalar multiplication along each coordinate axis.

Real Spectral Theorem

The **spectral theorem** provides a crucial result for symmetric matrices:

- Every real symmetric matrix can be orthogonally diagonalized.
- All its eigenvalues are real, and there exists an orthonormal basis of eigenvectors.

Formally, if A is symmetric, then: $A = PDP^T$, where P is an orthogonal matrix ($P^T P = I$) and D is a diagonal matrix of eigenvalues.

This theorem has fundamental implications in physics and machine learning (e.g., Principal Component Analysis).

Singular Value Decomposition (SVD)

The **singular value decomposition** generalizes diagonalization for any $m \times n$ matrix A: $A = U\Sigma V^T$, where:

- U is an $m \times m$ orthogonal matrix,
- V is an $n \times n$ orthogonal matrix,
- Σ is an $m \times n$ diagonal matrix with nonnegative **singular values** on the diagonal.

Each singular value corresponds to the length of the semiaxes of the ellipsoid obtained by transforming the unit sphere with A. This decomposition is vital in data compression and numerical optimization.

Step-by-Step Instruction

Step 1: Compute $A^T A$ and AA^T

Start with your matrix A of size $m \times n$:

1. Compute $A^T A$ (an $n \times n$ matrix).

2. Compute AA^T (an $m \times m$ matrix).

These two matrices are **symmetric** and **positive semidefinite**, which guarantees real, nonnegative eigenvalues.

Step 2: Find Eigenvalues and Singular Values

1. Solve the **eigenvalue problem** for $A^T A$: $A^T A v_i = \lambda_i v_i$
2. The **singular values** of A are the square roots of these eigenvalues: $\sigma_i = \sqrt{\lambda_i}$. The singular values are always nonnegative and are arranged in **descending order** in Σ.

Step 3: Compute the Right Singular Vectors (V)

The normalized eigenvectors v_i from $A^T A$ form the columns of V: $V = [v_1, v_2, \ldots, v_n]$. Since eigenvectors corresponding to distinct eigenvalues are orthogonal, V is an **orthogonal matrix** ($V^T V = I$).

Step 4: Compute the Left Singular Vectors (U)

For each singular value σ_i and corresponding right singular vector v_i, compute: $u_i = \frac{1}{\sigma_i} A v_i$. These u_i become the columns of the matrix U, and they form an orthonormal basis for the column space of A.

If A has rank $r < \min(m, n)$, additional orthonormal columns are added to U to make it square.

Step 5: Form the Diagonal Matrix Σ

Construct the $m \times n$ diagonal matrix Σ such that: $\Sigma_{ii} = \sigma_i$. All off-diagonal entries are 0. If $m > n$, there will be zero rows at the bottom; if $n > m$, there will be zero columns on the right.

Step 6: Verify the Decomposition

Check that: $A = U\Sigma V^T$.

This verification ensures the correctness of your SVD. The columns of U represent orthonormal directions in the input space, and those of V represent orthonormal directions in the output space.

Illustration Example (Conceptual)

Let:

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

Following the above steps yields:

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

where:

$$U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \end{bmatrix}, \quad V = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{18}} & -\frac{1}{\sqrt{18}} & \frac{4}{\sqrt{18}} \\ \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \end{bmatrix}$$

Geometric Meaning

- V rotates the coordinate system to align with principal axes.
- Σ scales space along those axes by factors σ_i.
- U rotates or flips the space to match the transformation output.

Thus, SVD describes **how any matrix transforms space**: as a rotation, followed by stretching, then another rotation.

Positive Definite Matrices

A symmetric matrix A is **positive definite** if: $x^T A x > 0$ for all nonzero vectors x.

Equivalently, all its eigenvalues are positive. Properties include:

- The determinant of a positive definite matrix is always positive.
- It admits a unique positive definite matrix square root.
- The inverse of a positive definite matrix is itself positive definite.

Such matrices model energy, covariance, and metric structures in applied contexts.

Determinant: Geometric Meaning and Invariance

The **determinant** of a square matrix measures how volume or area changes under the associated linear transformation. Specifically:

- In 2D, $|\det(A)|$ gives the area scaling factor of the parallelogram formed by the transformation.
- In 3D, $|\det(A)|$ gives the volume scaling factor of the transformed parallelepiped.

Invariance

The determinant's value changes by the same scalar multiple as the product of eigenvalues. It remains invariant under:

- row/column operations that preserve linear independence,
- similarity transformations $(\det(A) = \det(P^{-1}AP))$.

A zero determinant indicates that the transformation collapses space, meaning the matrix is singular.

Decompositions

LU (Lower-Upper factorization)

Definition

The **LU decomposition** (Lower–Upper factorization) expresses a square matrix A as the product of two triangular matrices: $A = L \times U$, where:

- L is a **lower triangular matrix** with 1s on the diagonal,

- U is an **upper triangular matrix**.

This technique is primarily used to solve systems of linear equations, calculate determinants, and find matrix inverses efficiently.

Step-by-Step Construction

1. **Start with a square matrix A:** $A = [a_{ij}]_{n \times n}$

2. **Apply Gaussian elimination** to transform A into an upper-triangular form U.

Record the multipliers used to eliminate elements below the diagonal — these form the off-diagonal entries in L.

3. **Form matrices:**

   - U is the resulting upper-triangular matrix.

   - L contains the elimination multipliers with 1s on the diagonal.

4. **Verify:**
   Check $A = LU$ to ensure correctness.

Example:
$$A = \begin{bmatrix} 4 & 3 \\ 6 & 3 \end{bmatrix}, \quad L = \begin{bmatrix} 1 & 0 \\ 1.5 & 1 \end{bmatrix}, \quad U = \begin{bmatrix} 4 & 3 \\ 0 & -1.5 \end{bmatrix}.$$

Then $A = LU$.

Applications

- Fast solution of $Ax = b$ using two triangular solves: $Ly = b$, then $Ux = y$.
- Efficient calculation of determinants via $\det(A) = \det(L)\det(U)$.
- Foundation for the **Cholesky** and **QR** methods.

QR Decomposition

Definition

The **QR decomposition** expresses a matrix A (not necessarily square) as: $A = QR$, where:

- Q is an orthogonal (or unitary) matrix ($Q^T Q = I$),

- R is an upper triangular matrix.

Method

1. **Use the Gram–Schmidt process** on the columns of A to create orthonormal vectors forming Q.
2. **Compute** R as: $R = Q^T A$

This decomposition is especially suited to solving least squares problems and is numerically stable.

Example

Let

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

Applying Gram–Schmidt:

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

Applications

- Solving $Ax = b$ when A is tall (least squares).
- Numerical stability in regression and PCA.
- Finding eigenvalues via iterative algorithms.

Cholesky Decomposition

Definition

The **Cholesky decomposition** is a special case of LU factorization for **symmetric positive definite** matrices: $A = LL^T$, where L is a lower triangular matrix with positive diagonal entries.

Steps

1. Ensure A is symmetric and positive definite ($x^T A x > 0$ for all nonzero x).

2. Compute entries of L as:

$$L_{ii} = \sqrt{A_{ii} - \sum_{k=1}^{i-1} L_{ik}^2}$$

$$L_{ij} = \frac{1}{L_{jj}} \left(A_{ij} - \sum_{k=1}^{j-1} L_{ik} L_{jk} \right)$$

3. Return $A = LL^T$.

Example

$$A = \begin{bmatrix} 4 & 12 & -16 \\ 12 & 37 & -43 \\ -16 & -43 & 98 \end{bmatrix} \Rightarrow L = \begin{bmatrix} 2 & 0 & 0 \\ 6 & 1 & 0 \\ -8 & 5 & 3 \end{bmatrix}.$$

Applications

- Solving symmetric systems more efficiently than LU.
- Covariance modeling in statistics.
- Basis of optimization methods in machine learning (e.g., Gaussian processes).

Eigendecomposition

Definition

The **eigendecomposition** expresses a square matrix A as: $A = P\Lambda P^{-1}$, where:

- Λ is diagonal containing the **eigenvalues** of A,

- P contains the corresponding **eigenvectors** as columns.

This decomposition exists when A has n linearly independent eigenvectors (i.e., is diagonalizable).

Steps

1. Compute the **characteristic polynomial**: $\det(A - \lambda I) = 0$.

2. Find eigenvalues λ_i and corresponding eigenvectors v_i.

3. Construct: $P = [v_1, v_2, \ldots, v_n], \quad \Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$.

   Then $A = P\Lambda P^{-1}$.

Example

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \Rightarrow \Lambda = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}, \quad P = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Applications

- Fundamental to the **spectral theorem** for symmetric matrices.
- Used in **Principal Component Analysis (PCA)**.
- Forms the basis of the **Singular Value Decomposition (SVD)**.

Comparative Table

Decomposition	Form	Matrix Type	Purpose	Notes
LU	$A = LU$	Any square A	Solving $Ax = b$	Based on Gaussian elimination
QR	$A = QR$	Any real A	Least-squares, PCA	Uses orthogonalization
Cholesky	$A = LL^T$	Symmetric positive definite	Fast & stable for symmetric systems	Special LU case

Decomposition	Form	Matrix Type	Purpose	Notes
Eigen	$A = P\Lambda P^{-1}$	Diagonalizable A	Spectral analysis, PCA	Relates to SVD

These decompositions provide complementary ways to simplify and analyze matrices—each suited for different matrix structures and computational goals in applied linear algebra and data analysis.

Core Analysis

Open and Closed Sets

Definition

In a metric or topological space (X, d):

- A set $U \subseteq X$ is **open** if, for every point $x \in U$, there exists a radius $r > 0$ such that the entire open ball $B_r(x) = \{y \in X : d(x, y) < r\} \subseteq U$.
- A set $F \subseteq X$ is **closed** if it contains all its limit points; equivalently, its complement $X \setminus F$ is open.

Some sets can be both open and closed; these are called **clopen sets**.

Examples (in $\mathbb{R}$)

- $(0, 1)$ is open.
- $(0, 1]$ is neither open nor closed.
- $\mathbb{R}$ and $\emptyset$ are both clopen.

Properties

- The union of any collection of open sets is open.
- The intersection of finitely many open sets is open.
- The complement of an open set is closed (and vice versa).
- A set is closed iff every convergent sequence in it has its limit also in it.

Continuity

Definition

A function $f : X \to Y$ between metric spaces is **continuous at a point** $a \in X$ if:

$$\forall \varepsilon > 0, \exists \delta > 0 \text{ such that } d_X(x, a) < \delta \Rightarrow d_Y(f(x), f(a)) < \varepsilon.$$

Equivalently, f is continuous if and only if the preimage of every open set in Y is open in X.

Key Theorems

- **Composition Theorem:** The composition of continuous functions is continuous.
- **Extreme Value Theorem:** A continuous function on a closed and bounded interval $[a, b]$ achieves maximum and minimum values.
- **Intermediate Value Theorem:** If f is continuous on $[a, b]$, it takes every value between $f(a)$ and $f(b)$.

Sequences and Convergence

Definition of a Sequence

A **sequence** is a function $a : \mathbb{N} \to \mathbb{R}$, denoted $(a_n)_{n=1}^{\infty}$.

It is a fundamental building block for calculus and real analysis, used to define limits, continuity, and compactness.

Convergent Sequences

A sequence (a_n) **converges** to a limit $L \in \mathbb{R}$ if:

$$\forall \varepsilon > 0, \exists N \in \mathbb{N} \text{ such that } |a_n - L| < \varepsilon \text{ for all } n > N.$$

We write $\lim_{n \to \infty} a_n = L$.

Divergent and Bounded Sequences

- A sequence is **divergent** if it does not converge to any finite limit.
- A sequence is **bounded** if there exists $M > 0$ such that $|a_n| \leq M$ for all n.

Examples:

- $a_n = \frac{1}{n} \to 0$ (convergent).
- $a_n = (-1)^n$ (divergent).
- $a_n = n$ (diverges to infinity).

Limits and Their Properties

Definition ($\varepsilon - \delta$ Form)

For a real function $f : \mathbb{R} \to \mathbb{R}$, the **limit** of $f(x)$ as $x \to a$ equals L if:

$$\forall \varepsilon > 0, \exists \delta > 0 \text{ such that } |x - a| < \delta \Rightarrow |f(x) - L| < \varepsilon.$$

Denoted: $\lim_{x \to a} f(x) = L$.

One-Sided Limits

- **Right-hand limit:** $\lim_{x \to a^+} f(x)$
- **Left-hand limit:** $\lim_{x \to a^-} f(x)$

If both exist and are equal, the two-sided limit exists.

Limit Laws

If $\lim_{x \to a} f(x) = L_1$ and $\lim_{x \to a} g(x) = L_2$, then:

- $\lim_{x \to a}[f(x) + g(x)] = L_1 + L_2$
- $\lim_{x \to a}[cf(x)] = cL_1$
- $\lim_{x \to a}[f(x)g(x)] = L_1 L_2$
- $\lim_{x \to a}[f(x)/g(x)] = L_1/L_2$, provided $L_2 \neq 0$.

Sequential Criterion for Limits

A limit $\lim_{x \to a} f(x) = L$ exists **iff** for every sequence (x_n) converging to a (excluding $x_n = a$), the sequence $f(x_n)$ converges to L. This connects **sequences** and **function limits**.

Notations

Notation	Description	Example/Meaning
$A, B, C, \ldots$	Matrices	$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$
$\mathbf{v}, \mathbf{w}$	Vectors	$\mathbf{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$
$a \cdot b$	Dot product of vectors	$a \cdot b = \sum_i a_i b_i$
$A \times B$	Cross product (in $\mathbb{R}^3$)	$\mathbf{a} \times \mathbf{b}$
A^T	Transpose of matrix A	$A^T = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$
A^{-1}	Inverse of matrix A	$AA^{-1} = I$
$\|\mathbf{v}\|$	Norm (length) of vector $\mathbf{v}$	$\|\mathbf{v}\| = \sqrt{\sum_i v_i^2}$
$\mathbf{0}$	Zero vector or zero matrix	$\mathbf{0} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$
$\mathbb{R}^n$	n-dimensional real vector space	Space of all vectors with n real components
$\mathbb{F}$	Field (general notation, e.g., $\mathbb{R}$, $\mathbb{C}$)	Field over which vector space is defined
$\mathbf{I}$	Identity matrix	$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
$A \circ B$	Hadamard (element-wise) product of matrices	$(A \circ B)_{ij} = A_{ij} B_{ij}$

Notation	Description	Example/Meaning
$A \otimes B$	Kronecker product of matrices	Larger matrix formed by multiplying elements
$\det(A)$ or A	Determinant of matrix A	Scalar value representing volume scaling
$\langle \mathbf{u}, \mathbf{v} \rangle$	Inner product of vectors	Generalized dot product in inner product spaces
$\operatorname{span}(\{\mathbf{v}_i\})$	Span of vectors	All linear combinations of $\mathbf{v}_i$
$\mathbf{u} \perp \mathbf{v}$	Orthogonality of vectors	$\mathbf{u} \cdot \mathbf{v} = 0$
$\mathcal{N}(A)$	Null space (kernel) of matrix A	Set of vectors $\mathbf{x}$ with $A\mathbf{x} = \mathbf{0}$
$\mathcal{R}(A)$	Range (image) of matrix A	Set of all vectors $A\mathbf{x}$ for $\mathbf{x}$ varying
$\mathbf{e}_i$	Standard basis vector with 1 in the i-th position	$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \end{bmatrix}$