Linear Algebra Fundamentals

Vectors, Matrices, Factorization, and Tensor Operations.

Jeremie Nlandu Mabiala

AMMI BootCamp 2025

African Masters of Machine Intelligence

August 4, 2025

AIMS | African Institute for Mathematical Sciences SENEGAL

Overview I

Overview II

- Matrices Definitions
- Matrix Operations

4. Linear transformations: Rotations and Projections
   - Linear Transformations
   - Rotations Transformations
   - Orthogonal Projections

5. Determinants, Matrix inversion, and Trace
   - Determinants and Traces

6. Eigenvalues and Eigenvectors
   - Eigenvalues and Eigenvectors Definitions

7. Eigenvalues and eigenvectors

AIMS | African Institute for Mathematical Sciences SENEGAL

Overview III

AIMS | African Institute for Mathematical Sciences SENEGAL

The Mathematical Language of AI

- **Foundation:** Linear algebra is the math behind all machine learning
- **Data Structure:** Data becomes matrices and vectors we can work with
- **Speed:** Matrix operations run fast on modern computers (GPUs)
- **Pattern Finding:** Tools like PCA help find hidden patterns in data

AIMS African Institute for Mathematical Sciences SENEGAL

Real AI Applications

- **Neural Networks:** Each layer multiplies matrices to learn
- **Training Models:** Uses vectors to improve AI performance
- **Decision Making:** Helps AI draw boundaries between different categories
- **Everyday Use:** Powers Netflix recommendations, self-driving cars, ChatGPT

AIMS | African Institute for Mathematical Sciences SENEGAL

Overview

AIMS | African Institute for Mathematical Sciences SENEGAL

Vectors - Definition and Notation

- A *vector* is an ordered list of numbers.
- Written in **row representation** as:

$$[x_1, x_2, \cdots, x_n, \ldots] \text{ or } (x_1, x_2, \cdots, x_n, \ldots) \tag{2.1}$$

- Or in **column representation** as:

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \\ \vdots \end{bmatrix} \text{ or } \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \\ \vdots \end{pmatrix} \tag{2.2}$$

Vector Components and Dimensions

- Quantities $x_1, x_2, \ldots, x_n$ are called **components** (elements, coefficients, entries).
- The number of elements is the **dimension** (or length) of the vector.
- Examples of finite vectors:

$$\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} \text{ and } \begin{bmatrix} -2 \\ 1.1 \\ 0 \\ 0 \end{bmatrix} \tag{2.3}$$

are vectors of size 3 and 4, respectively.

AIMS African Institute for Mathematical Sciences SENEGAL

Vector Notation Conventions

- **Scalars:** Numbers are vectors of dimension 0.
- **Notation:** Lowercase letters denote vectors: a, x, p, r or $\vec{x}$.
- **Indexing:** The i-th element of vector x is denoted x_i.
- **Equality:** Vectors x and y are equal ($x = y$) if and only if $x_i = y_i$ for all i.
- **Dimension notation:**
    - n-vector: vector of dimension n
    - $(n, 1)$-column vector or $(1, n)$-row vector

Block Vectors and Concatenation

- **Stacked vector:** For vectors b, c, d of sizes m, n, p:

$$a = \begin{bmatrix} b \\ c \\ d \end{bmatrix} \tag{2.4}$$

- Also called **block vector** with block entries b, c, d.

- Result: a is an $(m + n + p)$-vector.

- **Expanded form:**

$$a = (b_1, b_2, \ldots, b_m, c_1, c_2, \ldots, c_n, d_1, d_2, \ldots, d_p)$$

Special Vectors

- **Zero vector:** All entries are zero, denoted $\mathbf{0}$.

- **Ones vector:** All entries are 1, denoted $\mathbf{1}$.

- **Unit vectors:** One entry is 1, all others are 0, denoted e_i.

- **Example** - Unit vectors in dimension 3:

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{2.5}$$

- **Sparse vectors:** Many entries are zero. We denote $\text{nnz}(x)$ as the number of nonzero entries.

AIMS African Institute for Mathematical Sciences SENEGAL

Geometric Representation

- A 2D vector (x_1, x_2) represents a location or displacement in a plane:
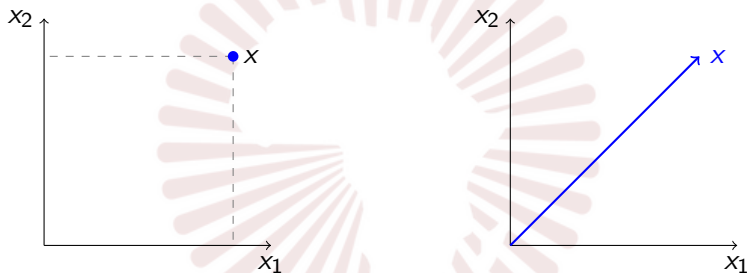
Figure: Point vs. Vector representation in 2D

Real-World Vector Examples

- **Economics:** Quantities of n different commodities

- **Education:** Student grades per subject/year

- **Finance:** Cash flow where x_i is payment in period i

- **Audio:** Acoustic pressure at sample time i

- **NLP:** Word count where x_i is frequency of word i in a document
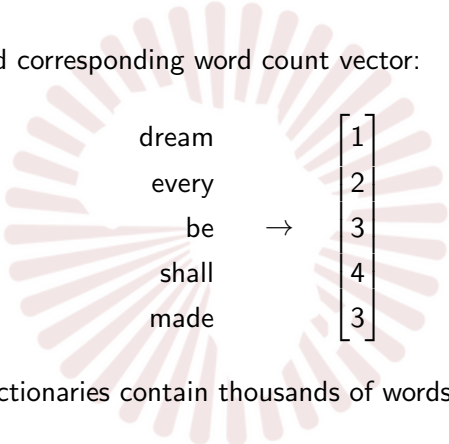
Text Processing Example

"I have a dream *that one day* every *valley* shall *be* exalted, *every* hill *and* mountain shall *be made* low, the rough places will *be made* plain, *and* the crooked places will *be made* straight; and the glory of the Lord shall *be* revealed, and all flesh shall *see it together."*

— Martin Luther King Jr., **"I have a dream"** (excerpt)

Word Count Vector Example

- Dictionary and corresponding word count vector:

$$
\begin{matrix}
\text{dream} \\
\text{every} \\
\text{be} \\
\text{shall} \\
\text{made}
\end{matrix}
\quad \rightarrow \quad
\begin{bmatrix}
1 \\
2 \\
3 \\
4 \\
3
\end{bmatrix}
$$

- In practice, dictionaries contain thousands of words.

Overview

AIMS | African Institute for Mathematical Sciences SENEGAL

Vector Operations

- We can add and subtract two vectors a and b of the same size, with the sum denoted $a \pm b$

- To get the sum, we add/subtract corresponding entries:

$$\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \pm \begin{bmatrix} 9 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \pm 9 \\ 1 \pm 2 \\ 1 \pm (-1) \end{bmatrix}$$

- Addition of vectors is commutative, associative, and has the zero vector as the neutral element

AIMS | African Institute for Mathematical Sciences SENEGAL

Geometric Interpretation: Addition

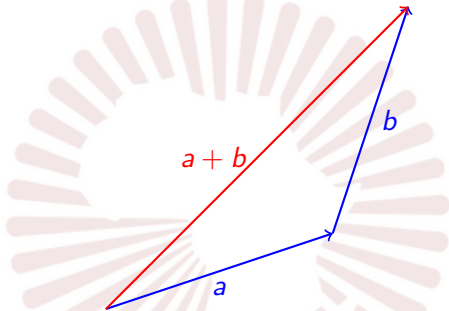- Geometrically, if a and b are vectors, their sum $a + b$ is represented as:

Figure: Addition of two vectors a and b

- The opposite of a vector a, denoted $-a$, is a vector in the opposite direction of a

Geometric Interpretation: Subtraction

- The representation of $a - b$:
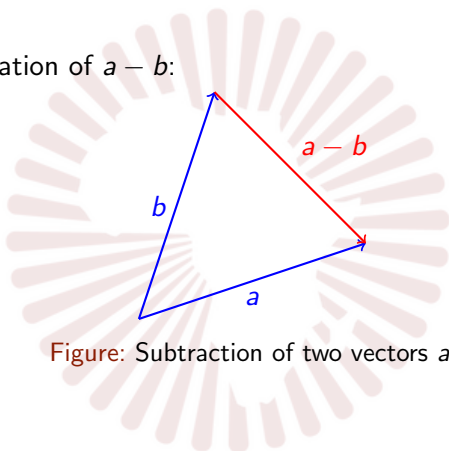
Figure: Subtraction of two vectors a and b

Scalar Multiplication

- We can multiply a vector a by a scalar β, denoted βa

- $\beta a = (\beta a_1, \beta a_2, \ldots, \beta a_n)$

- It is a scaling in the magnitude of the vector a:
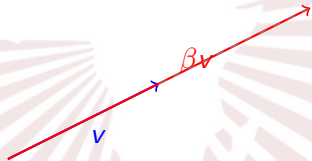
- Scalar multiplication is associative and distributive

Element-wise Multiplication

- For two *n*-vectors *a* and *b*, the element-wise multiplication $a \circ b$ produces a vector where each component is the product of corresponding elements

- Formally:

$$a \circ b = (a_1 b_1, a_2 b_2, \ldots, a_n b_n)$$

- The symbol $\circ$ or $\star$ denotes *element-wise* or *Hadamard* multiplication

- Example:

$$(0, -1, 2) \circ (1, -1, \frac{1}{2}) = (0 \cdot 1, (-1) \cdot (-1), 2 \cdot \frac{1}{2}) = (0, 1, 1)$$

- Element-wise multiplication is not commutative, i.e., $a \circ b \neq b \circ a$ in general

- However, it is associative, i.e., $(a \circ b) \circ c = a \circ (b \circ c)$

- It is distributive over vector addition, i.e., $a \circ (b + c) = a \circ b + a \circ c$

- **Note:** This operation is intensively used in NumPy and machine learning for vectorization.

Transpose and Linear Combinations

- **Transpose:** For an $(n, 1)$-vector x, its transpose x^T is a $(1, n)$-vector
- Properties of transpose:
  - $(x^T)^T = x$ (Transpose of transpose is the original vector)
  - $(\beta x)^T = \beta x^T$ for a scalar β
  - $(x + y)^T = x^T + y^T$ for vectors x and y

- **Linear combination:** A linear combination of vectors $x_1, x_2, \ldots, x_k$ with scalars $\alpha_1, \alpha_2, \ldots, \alpha_k$ is:

$$y = \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_k x_k$$

- This is a vector formed by scaling and adding the vectors x_i.

AIMS | African Institute for Mathematical Sciences SENEGAL

- Example: For $x_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $x_2 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$, and $\alpha_1 = 2$, $\alpha_2 = -1$:

$$y = 2 \begin{bmatrix} 1 \\ 2 \end{bmatrix} - 1 \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix} - \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

- The result y is a vector in the same dimension as x_1 and x_2.

- Linear combinations are fundamental in linear algebra, forming the basis for **vector spaces**.

Overview

AIMS African Institute for Mathematical Sciences SENEGAL

Vector Spaces

- A **vector space** is a set of vectors of the same dimension, equipped with vector addition and scalar multiplication operations, and satisfies specific axioms.

- **Fundamental axioms** of a vector space V (over the field $\mathbb{R}$) are:

  1. **Closure under addition**: $\mathbf{u} + \mathbf{v} \in V$ for all $\mathbf{u}, \mathbf{v} \in V$
  2. **Closure under scalar multiplication**: $c\mathbf{v} \in V$ for all $c \in \mathbb{R}$ and $\mathbf{v} \in V$
  3. **Associativity**: $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
  4. **Commutativity**: $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
  5. **Zero vector**: There exists $\mathbf{0} \in V$ such that $\mathbf{v} + \mathbf{0} = \mathbf{v}$
  6. **Additive inverse**: For each $\mathbf{v} \in V$, there exists $-\mathbf{v} \in V$ such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$
  7. **Distributivity**: $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$ and $(c + d)\mathbf{v} = c\mathbf{v} + d\mathbf{v}$
  8. **Scalar multiplication identity**: $1 \cdot \mathbf{v} = \mathbf{v}$

AIMS | African Institute for Mathematical Sciences SENEGAL

- To show that a set of vectors forms a vector space, we need to verify these axioms.
- A vector space can be defined over any field, such as $\mathbb{R}$ or $\mathbb{C}$, or even finite fields, such as $\mathbb{F}_p$.
- **Note:** We consider vector spaces over the field of real numbers $\mathbb{R}$, meaning that all the scalars are real numbers.

- Examples of vector spaces:
  - $\mathbb{R}^n$: The set of all n- vectors with real components forms a vector space.
  - The set of all polynomials of degree at most n forms a vector space, i.e. $P_n = \{p(x) = a_0 + a_1 x + \cdots + a_n x^n \mid a_i \in \mathbb{R}\}$ is a vector space.
  - The set of all continuous functions on a closed interval, i.e. $C[a, b]$, forms a vector space.
- We can easily show the examples above saisfy the axioms of a vector space.

- **Dimension:** The **dimension** of a vector space equals the number of components in each vector

- Example: $\mathbb{R}$ represents the real numbers and forms a 1-dimensional vector space

- Example: $\mathbb{R}^2$ consists of all ordered pairs:

$$\mathbb{R}^2 = \left\{ x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \mid x_1, x_2 \in \mathbb{R} \right\}$$

forming a 2-dimensional vector space that can be visualized as the Cartesian plane
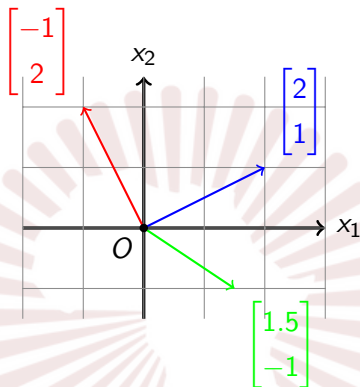
Figure: Visualization of $\mathbb{R}^2$ as a 2-dimensional vector space

- In general, $\mathbb{R}^n$ constitutes an n-dimensional vector space

Subspaces. Spannning set

- **Subspace:** A **subspace** is a subset of a vector space that is also a vector space itself, with respect to the same vector addition and scalar multiplication.

- A subspace must satisfy the same axioms as the parent vector space.

- Equivalently, a subspace is closed under vector addition and scalar multiplication, i.e. $\emptyset \neq S \subseteq V$ a subspace of V must satisfy

$$(i)\ \ \mathbf{x}, \mathbf{y} \in S \implies \mathbf{x} + \mathbf{y} \in S, \quad (ii)\ \ \lambda \in \mathbb{R}, \mathbf{x} \in S \implies \lambda \mathbf{x} \in S.$$

AIMS | African Institute for Mathematical Sciences SENEGAL

- Example: The set of all vectors in $\mathbb{R}^3$ that lie on the xy-plane is a subspace of $\mathbb{R}^3$.

- Example: The set of all vectors in $\mathbb{R}^3$ that lie on the line $x = y = z$ is a subspace of $\mathbb{R}^3$.

- Example: The set of all polynomials of degree at most n is a subspace of the vector space of all polynomials.

- Example: The set of all continuous functions on a closed interval $[a, b]$ is a subspace of the vector space of all functions defined on $[a, b]$.

- **Spanning Set:** A set of vectors $\{v_1, v_2, \ldots, v_k\}$ **spans** a vector space V if every vector in V can be expressed as a linear combination of the vectors in the set.

- Formally, $V = \mathrm{span}\{v_1, v_2, \ldots, v_k\}$ means that any vector $v \in V$ can be written as:

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_k v_k$$

where α_i are scalars.

- The set $\{v_1, v_2, \ldots, v_k\}$ is called a **spanning set** for V.

AIMS | African Institute for Mathematical Sciences SENEGAL

- Example: The set $\{e_1, e_2, e_3\}$ spans $\mathbb{R}^3$, where:

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

- Indeed, any vector in $\mathbb{R}^3$ can be expressed as a linear combination of these vectors:

$$v = x_1 e_1 + x_2 e_2 + x_3 e_3 = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Other examples of spanning sets include **lines** and **planes** in vector spaces:

- **Line:** Given a vector u, the line spanned by u is the set of all scalar multiples of u, i.e.,

$$l_u = \{v \mid v = \lambda u \text{ for some } \lambda \in \mathbb{R}\} = \text{span}\{u\}$$

- **Plane:** Given two vectors u and v, the plane spanned by u and v is the set of all linear combinations of u and v, i.e.,

$$\pi_{uv} = \{w \mid w = \lambda u + \beta v \text{ for some } \lambda, \beta \in \mathbb{R}\} = \text{span}\{u, v\}$$

- Example:

$$\mathbb{R}^2 = \{\lambda e_1 + \beta e_2 \mid \lambda, \beta \in \mathbb{R}\}$$

AIMS | African Institute for Mathematical Sciences SENEGAL

Dependence, Independence,and Basis

- **Dependence:** A set of vectors $\{a_1, \ldots, a_k\}$ (with $k \geq 1$) is **linearly dependent** if there exist coefficients $\beta_1, \ldots, \beta_k$, not all zero, such that

$$\beta_1 a_1 + \beta_2 a_2 + \cdots + \beta_k a_k = 0 \qquad (2.6)$$

- Equivalently, at least one vector a_i can be expressed as a linear combination of the others.

- We say the vectors $a_1, \ldots, a_k$ are **linearly dependent**.

AIMS | African Institute for Mathematical Sciences SENEGAL

- A single vector $\{a_1\}$ is linearly dependent if and only if $a_1 = 0$.
- Two vectors $\{a_1, a_2\}$ are linearly dependent if and only if one is a scalar multiple of the other, i,.e., $a_1 = \lambda a_2$ for some scalar λ.
- For more than two vectors, it is not easy to state their dependence using the definition. The notion **determinant** will make it "easy ".

- Example: Consider the vectors

$$a_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \ a_2 = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}, \ a_3 = \begin{bmatrix} -2 \\ 4 \\ 6 \end{bmatrix} \tag{2.7}$$

These vectors are linearly dependent because we can find coefficients $\beta_1 = 1$, $\beta_2 = 2$, $\beta_3 = -3$ (not all zero) such that $\beta_1 a_1 + \beta_2 a_2 + \beta_3 a_3 = 0$.

- **Independence:** A set of $k \geq 1$ vectors $\{a_1, \ldots, a_k\}$ is **linearly independent** if the equation

$$\beta_1 a_1 + \beta_2 a_2 + \cdots + \beta_k a_k = 0 \qquad (2.8)$$

has only the trivial solution $\beta_1 = \beta_2 = \cdots = \beta_k = 0$.

- Equivalently, none of the vectors can be written as a linear combination of the others.

- We say the vectors $a_1, \ldots, a_k$ are **linearly independent**.

- The standard unit vectors $e_1, \ldots, e_n$ form a linearly independent set.

AIMS | African Institute for Mathematical Sciences SENEGAL

- Example: Consider the vectors

$$a_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \; a_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \; a_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{2.9}$$

These vectors are linearly independent because the only solution to the equation $\beta_1 a_1 + \beta_2 a_2 + \beta_3 a_3 = 0$ is $\beta_1 = \beta_2 = \beta_3 = 0$.

AIMS African Institute for Mathematical Sciences SENEGAL

- In general, a set of n vectors in $\mathbb{R}^n$ is linearly independent if and only if the determinant of the matrix formed by these vectors is non-zero.

- We will see later how to make use of the assertion above.

- The dimension of a vector space is the number of vectors in any basis for that space.

- **Basis:** A basis for a vector space V is a set of vectors $\{b_1, b_2, \ldots, b_k\}$ such that:
  1. The vectors are linearly independent.
  2. The vectors span the vector space V, that is any vector in V can be expressed as a linear combination of the basis vectors.

- **Dimension:** The number of vectors in a basis is the **dimension** of the vector space.

- Example: In $\mathbb{R}^3$, the set $\{e_1, e_2, e_3\}$ is a basis where:

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{2.10}$$

- Any vector in $\mathbb{R}^3$ can be uniquely expressed as a linear combination of these basis vectors:

AIMS African Institute for Mathematical Sciences SENEGAL

- The dimension of $\mathbb{R}^3$ is 3, as there are three basis vectors.
- The concept of basis extends to any vector space, not just $\mathbb{R}^n$.
- For example, the set of all polynomials of degree at most n has a basis consisting of the monomials $\{1, x, x^2, \ldots, x^n\}$.
- The dimension of this polynomial space is $n + 1$.

Overview

AIMS | African Institute for Mathematical Sciences | SENEGAL

Inner product

- The dot product (also known as inner product) between two n-vectors a and b is defined and denoted as the scalar:

$$a^T b = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n \qquad (2.12)$$

- Alternative notations include: (a, b) and $a \cdot b$. And we write it as: $(a, b) = a^T b$ or $a \cdot b = a^T b$.

- Illustration:

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}^T \begin{bmatrix} 1 \\ \frac{1}{2} \\ -2 \end{bmatrix} = (1 \times 1) + (2 \times \frac{1}{2}) + (3 \times -2) = -4 \qquad (2.13)$$

Key Properties

- Fundamental properties:
  - Commutativity: $a^T b = b^T a$
  - Scalar multiplication: $(\lambda a)^T b = \lambda(a^T b)$
  - Distributivity: $(a + b)^T c = a^T c + b^T c$
  - Zero vector: $0^T a = a^T 0 = 0$ for any vector a
- Practical applications:
  - $e_i^T a = a_i$ (extracts the i-th component)
  - $\mathbf{1}^T a = a_1 + a_2 + \cdots + a_n$ (computes sum of entries)
  - $a^T a = a_1^2 + \cdots + a_n^2$ (sum of squared entries)
- **Note :** An Euclidean space is a vector space with inner product.

AIMS | African Institute for Mathematical Sciences SENEGAL

Vector Norms

- The Euclidean norm of a vector x is defined as:

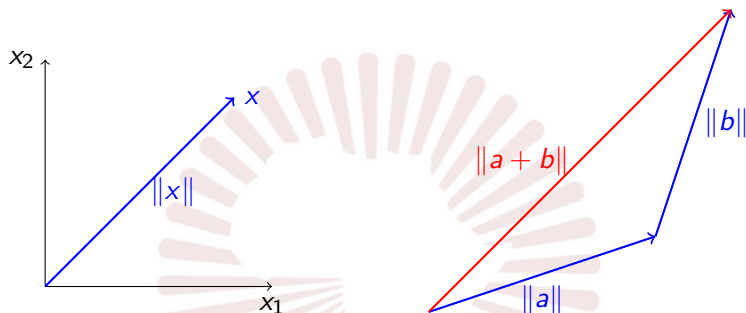$$\|x\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \sqrt{x^T x} \qquad (2.14)$$

- For $n = 1$, this simplifies to the absolute value.

- **Example:** For $x = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$, the norm is:

$$\|x\| = \sqrt{3^2 + 4^2} = \sqrt{9 + 16} = \sqrt{25} = 5$$

- The norm is a measure of the distance from the origin to the point represented by the vector in a vector space.

AIMS | African Institute for Mathematical Sciences SENEGAL

Figure: (Left) Geometric representation of 2-vector norms. (Right) Triangle inequality visualization

- The norm is also known as the **length** or **magnitude** of the vector.

- **Note :** There are other types of norms, such as the L_1 and L_∞ norms, which are defined as:

$$L_1 \text{ norm}: \ \|x\|_1 = |x_1| + |x_2| + \cdots + |x_n| \qquad (2.15)$$

$$L_\infty \text{ norm}: \ \|x\|_\infty = \max(|x_1|, |x_2|, \ldots, |x_n|) \qquad (2.16)$$

- The Euclidean(also known as L_2) norm is the most commonly used norm.

 AIMS | African Institute for Mathematical Sciences SENEGAL

- Visual comparison of different norms for the same vector:
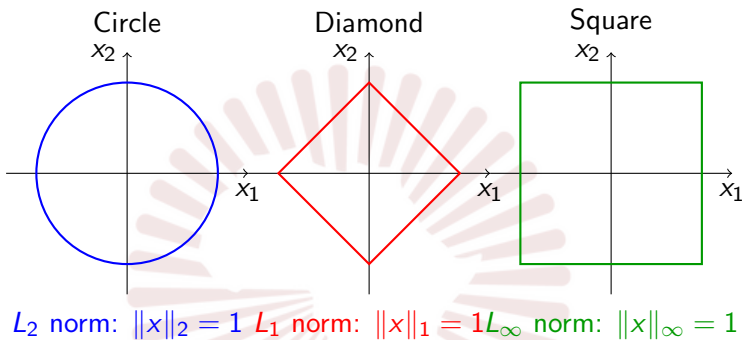
Figure: Unit balls for different norms in 2D space

L_2 norm: $\|x\|_2 = 1$ L_1 norm: $\|x\|_1 = 1$ L_∞ norm: $\|x\|_\infty = 1$

- Each shape represents all vectors with norm equal to 1 under the respective norm definition.

Norm Properties

All norms for n-vectors x and y and scalar λ must satisfy these axioms:

- **Scaling property**: $\|\lambda x\| = |\lambda| \|x\|$
- **Trinagual inequality**: $\|x + y\| \leq \|x\| + \|y\|$
- **Non-negativity**: $\|x\| \geq 0$
- **Zero condition**: $\|x\| = 0$ if and only if $x = 0$

Note: That is, for a function to be a norm, it must satisfy these properties.

AIMS African Institute for Mathematical Sciences SENEGAL

Exercise: Verify these properties hold for the L_1, L_2, and L_∞ norms.

Exercice: Does the function $\|x\| = \sqrt{x_1^2 + x_2^2 + 1}$ satisfy the norm properties? Why or why not?

Norm of Block Vectors

- Consider block vector formed by concatenating vectors a, b, and c:

- The squared norm satisfies:

$$\|(a, b, c)\|^2 = a^T a + b^T b + c^T c = \|a\|^2 + \|b\|^2 + \|c\|^2 \quad (2.17)$$

- Therefore, the norm of the block vector is:

$$\|(a, b, c)\| = \sqrt{\|a\|^2 + \|b\|^2 + \|c\|^2} \quad (2.18)$$

- This property extends to any number of block components.

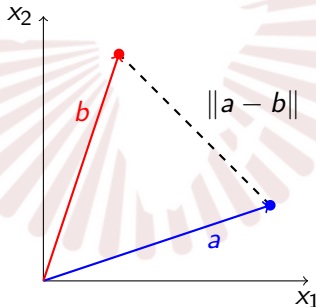AIMS | African Institute for Mathematical Sciences SENEGAL

Euclidean Distance Between Vectors

- The Euclidean distance between two n-vectors a and b is defined as:

$$\text{dist}(a, b) = \|a - b\| = \sqrt{(a_1 - b_1)^2 + \cdots + (a_n - b_n)^2} \qquad (2.19)$$

- Geometric interpretation in 2D space:

- The distance is zero if and only if $a = b$.

- The distance satisfies the **triangle inequality**:

$$\text{dist}(a, b) + \text{dist}(b, c) \geq \text{dist}(a, c) \tag{2.20}$$

- The distance can also be generalized to other norms, such as L_1 and L_∞:

$$\text{dist}_1(a, b) = \|a - b\|_1 = |a_1 - b_1| + |a_2 - b_2| + \cdots + |a_n - b_n| \tag{2.21}$$

$$\text{dist}_\infty(a, b) = \|a - b\|_\infty = \max(|a_1 - b_1|, |a_2 - b_2|, \ldots, |a_n - b_n|) \tag{2.22}$$

- Example: For vectors $a = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$ and $b = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$:

$$\text{dist}(a, b) = \sqrt{(3-1)^2 + (4-2)^2} = \sqrt{2^2 + 2^2} = \sqrt{4 + 4} = \sqrt{8} = 2\sqrt{2}$$

- The distance can be interpreted as the length of the line segment connecting points a and b in 2D space.

Statistical Properties Using Inner Products and Norms

We can compute statstic measures of vector x in terms of its norm:

- For an n-vector x, the mean can be expressed using inner products:

$$\overline{x} = \frac{1}{n}\mathbf{1}^T x \tag{2.23}$$

- The centered (de-meaned) vector is obtained by:

$$\tilde{x} = x - \overline{x}\mathbf{1}$$

- The standard deviation can be computed using the norm:

$$\text{std}(x) = \frac{\|\tilde{x}\|}{\sqrt{n}} = \frac{\|x - \overline{x}\mathbf{1}\|}{\sqrt{n}}$$

AIMS | African Institute for Mathematical Sciences SENEGAL

- The mean square value of vector x is defined as:

$$\text{MSV}(x) = \frac{x_1^2 + \cdots + x_n^2}{n} = \frac{\|x\|^2}{n} \tag{2.24}$$

- The root mean square (RMS) value of vector x is:

$$\text{RMS}(x) = \sqrt{\frac{x_1^2 + \cdots + x_n^2}{n}} = \frac{\|x\|}{\sqrt{n}} \tag{2.25}$$

- The variance of vector x is defined as:

$$\text{Var}(x) = \frac{1}{n}\|x - \overline{x}\mathbf{1}\|^2 = \frac{1}{n}(x - \overline{x}\mathbf{1})^T(x - \overline{x}\mathbf{1}) \tag{2.26}$$

- The covariance matrix of vector x is given by:

$$C(x) = \frac{1}{n}(x - \overline{x}\mathbf{1})(x - \overline{x}\mathbf{1})^T \tag{2.27}$$

- The covariance matrix captures the variance and covariance of the components of vector x.

Some Norms Inequalities

Cauchy-Schwarz Inequality

For any *n*-vectors *a* and *b*, the following inequality holds:

$$|a^T b| \leq \|a\|\|b\| \tag{2.28}$$

This fundamental result is known as the **Cauchy-Schwarz inequality**.

Proof:

- Consider the vector $c = a - \lambda b$ for some scalar λ.

- The norm of c must be non-negative:

$$\|c\|^2 = (a - \lambda b)^T (a - \lambda b) \geq 0$$

- Expanding this gives:

$$a^T a - 2\lambda a^T b + \lambda^2 b^T b \geq 0$$

- This is a quadratic in λ, which must have non-negative discriminant:

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

- Rearranging yields the Cauchy-Schwarz inequality:

$$|a^T b|^2 \leq (a^T a)(b^T b)$$

- Taking square roots gives:

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

AIMS | African Institute for Mathematical Sciences SENEGAL

Exercise: Use the Cauchy-Schwarz inequality to demonstrate that the triangle inequality holds for vector norms, i.e., show that:

$$\|a + b\| \leq \|a\| + \|b\|$$

Minkowski Inequality

For any n-vectors a and b, and for $p \geq 1$, the following holds:

$$\|a + b\|_p \leq \|a\|_p + \|b\|_p \tag{2.29}$$

This is known as the **Minkowski inequality**.

Proof:

- Consider the p-norm defined as:

$$\|x\|_p = (|x_1|^p + |x_2|^p + \cdots + |x_n|^p)^{1/p}$$

- The Minkowski inequality states that the p-norm satisfies the triangle inequality.

- This can be shown using the Cauchy-Schwarz inequality for p-norms.

- An other way to show the Minkowsky's inequality is to use the convexity of the function $f(x) = |x|^p$ for $p \geq 1$.

- The function $f(x)$ is convex, because its second derivative is non-negative:

$$f''(x) = p(p-1)|x|^{p-2}$$

- Convex functions satisfy the Jensen's inequality:

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$$

for all $\lambda \in [0,1]$.

AIMS African Institute for Mathematical Sciences SENEGAL

- Applying this to the p-norm gives:

$$\|a + b\|_p^p = (|a_1 + b_1|^p + |a_2 + b_2|^p + \cdots + |a_n + b_n|^p) \leq (|a_1|^p + |b_1|^p$$

- Taking the p-th root yields the Minkowski inequality:

$$\|a + b\|_p \leq \|a\|_p + \|b\|_p$$

Holder's Inequality

For any *n*-vectors a and b and for any $p, q > 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$, the following holds:
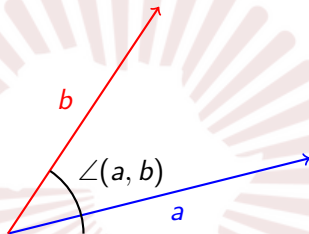
$$|a^T b| \leq \|a\|_p \|b\|_q \tag{2.30}$$

This is known as **Hölder's inequality**.

Proof: Exercice

Angle Between Vectors

- Two vectors sharing a common origin define an angle between them.

- The angle between two vectors a and b is given by:

$$\angle(a, b) = \arccos\left(\frac{\langle a, b \rangle}{\|a\| \times \|b\|}\right) \tag{2.31}$$

- The existence of the angle is guaranteed by the Cauchy-Schwarz inequality:

$$|\langle a, b \rangle| \leq \|a\| \|b\| \tag{2.32}$$

- For non-zero vectors a and b, we can normalize the inner product:

$$-1 \leq \frac{\langle a, b \rangle}{\|a\| \|b\|} \leq 1$$

- Since arccos is defined on the interval $[-1, 1]$ and maps bijectively to $[0, \pi]$, the angle $\angle(a, b)$ is well-defined.

- For more details, refer to the blog post *continuous inverse theorem*.

AIMS | African Institute for Mathematical Sciences SENEGAL

Properties of Angles

- The angle is always between 0 and π radians (or 0 and $180°$).

- If the angle is 0, the vectors are in the same direction.

- If the angle is π, the vectors are in opposite directions.

- If the angle is $\frac{\pi}{2}$, the vectors are orthogonal (perpendicular).
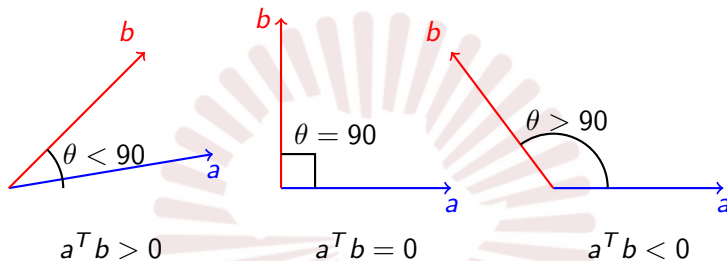
- The angle can be computed using the inner product:

- From equation 2.31, we can express the dot product of vectors a and b as:

$$a^T b = \|a\|\|b\| \cos(\angle(a, b)) \tag{2.33}$$

- This gives us insight into the sign of the dot product:
  - $a^T b > 0$ when $\angle(a, b) < \frac{\pi}{2}$ (acute angle)
  - $a^T b = 0$ when $\angle(a, b) = \frac{\pi}{2}$ (orthogonal vectors)
  - $a^T b < 0$ when $\angle(a, b) > \frac{\pi}{2}$ (obtuse angle)

Figure: Relationship between dot product sign and angle between vectors

Example

- Consider vectors $a = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $b = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$.

- Compute the angle between them:

$$a^T b = 1 \cdot 3 + 2 \cdot 4 = 3 + 8 = 11$$
$$\|a\| = \sqrt{1^2 + 2^2} = \sqrt{5}$$
$$\|b\| = \sqrt{3^2 + 4^2} = \sqrt{25} = 5$$

- The angle is:

$$\angle(a, b) = \arccos\left(\frac{11}{\sqrt{5} \cdot 5}\right) = \arccos\left(\frac{11}{5\sqrt{5}}\right)$$

AIMS | African Institute for Mathematical Sciences SENEGAL

Example: Consider vectors $a = \begin{bmatrix} 3 \\ 4 \\ 0 \end{bmatrix}$ and $b = \begin{bmatrix} 4 \\ -3 \\ 5 \end{bmatrix}$.

- Compute the angle between them:

$$a^T b = 3 \cdot 4 + 4 \cdot (-3) + 0 \cdot 5 = 12 - 12 + 0 = 0$$
$$\|a\| = \sqrt{3^2 + 4^2 + 0^2} = \sqrt{9 + 16} = 5$$
$$\|b\| = \sqrt{4^2 + (-3)^2 + 5^2} = \sqrt{16 + 9 + 25} = \sqrt{50} = 5\sqrt{2}$$

- The angle is:

$$\angle(a, b) = \arccos\left(\frac{0}{5 \cdot 5\sqrt{2}}\right) = \arccos(0) = \frac{\pi}{2}$$

AIMS | African Institute for Mathematical Sciences SENEGAL

- **Exercise** (Pythagorean theorem): Prove that for any *n*-vectors *a* and *b*:

$$\|a + b\|^2 = \|a\|^2 + \|b\|^2 \quad \text{if and only if} \quad \angle(a, b) = \frac{\pi}{2} \qquad (2.34)$$

- **Exercise**: Determine the angles between the standard unit vectors e_1, e_2, and e_3 in $\mathbb{R}^3$.

Orthogonal Vectors and Orthogonality

- **Orthonanility:** Two n-vectors a and b are orthogonal, denoted $a \perp b$, if their inner product is zero:

$$a^T b = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n = 0 \qquad (2.35)$$

- **Example:** The standard unit vectors e_1 and e_2 are orthogonal. For $n = 3$:

$$e_1^T e_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}^T \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = (1)(0) + (0)(1) + (0)(0) = 0$$

AIMS | African Institute for Mathematical Sciences SENEGAL

- **Mutual orthogonaity:** A set of vectors $a_1, \ldots, a_k$ are called **mutually orthogonal** if $a_i \perp a_j$ for all $i \neq j$.

- The standard unit vectors $e_1, \ldots, e_n$ form a mutually orthogonal set, satisfying:

$$e_i{}^T e_j = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{2.36}$$

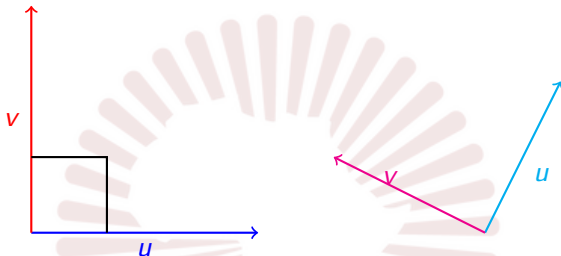- The quantity δ_{ij} is called the **Kronecker delta**.

Figure: Examples of orthogonal vector pairs in $\mathbb{R}^2$

- **Orthonormality:** A set of vectors $a_1, \ldots, a_k$ are **orthonormal** if they are mutually orthogonal and each vector has unit norm:

$$\|a_i\| = 1 \quad \text{for all } i \tag{2.37}$$

- The standard unit vectors $e_1, \ldots, e_n$ form an orthonormal set:

$$\|e_i\| = 1 \quad \text{and} \quad e_i^T e_j = \delta_{ij} \quad \text{for all } i, j \tag{2.38}$$

Exercice: Show that orthogonal vectors are linearly independent.

Parseval's Identity

For any *n*-vector $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ and any orthonormal basis $\{u_1, u_2, \ldots, u_n\}$, where $\|u_i\| = 1$ and $u_i^T u_j = \delta_{ij}$, the squared Euclidean norm satisfies:

$$\|x\|_2^2 = \langle x, x \rangle = \sum_{i=1}^{n} |(x, u_i)|^2 = \sum_{i=1}^{n} (x^T u_i)^2 \qquad (2.39)$$

where $(\cdot, \cdot)$ denotes the standard inner product on $\mathbb{R}^n$.

Note: This shows that the squared norm of a vector is preserved under any orthonormal basis representation.

AIMS | African Institute for Mathematical Sciences SENEGAL

Orthonormal Expansion

- If $\{a_1, a_2, \ldots, a_n\}$ forms an orthonormal basis for $\mathbb{R}^n$, then any vector $x \in \mathbb{R}^n$ can be uniquely expressed as:

$$x = (a_1^T x)a_1 + (a_2^T x)a_2 + \cdots + (a_n^T x)a_n = \sum_{i=1}^{n} (a_i^T x)a_i \qquad (2.40)$$

- This representation is called the **orthonormal expansion** of x with respect to the orthonormal basis $\{a_1, a_2, \ldots, a_n\}$.
- The coefficients $(a_i^T x)$ represent the projections of x onto each basis vector a_i.

AIMS African Institute for Mathematical Sciences SENEGAL

Exercice: Prove Parseval's identity by showing that the inner product of x with each basis vector u_i gives the projection of x onto u_i, and that the sum of the squares of these projections equals the squared norm of x.

Exercise: Prove this expansion 2.40 by showing that the right-hand side equals x using the orthonormality properties.

Gram-Schmidt (orthogonalization) algorithm

- An algorithm to construct orthogonal vectors from any vectors.
- **Given k vectors of dimension n, $a_1, \ldots, a_k$**
- The $G - S$ algorithm generates orthonormal vectors $q_1, q_2, \ldots, q_k$ as follows:

$$q_1 = a_1$$

$$q_2 = a_2 - \frac{a_2^T q_1}{\|q_1\|^2} q_1$$

$$\vdots = \vdots$$

$$q_k = a_k - \sum_{i=1}^{k-1} \frac{a_k^T q_i}{\|q_i\|^2} q_i$$

- **Gram-Schmidt Algorithm:**

  1. **Input:** Linearly independent vectors $a_1, a_2, \ldots, a_k \in \mathbb{R}^n$
  2. **Output:** Orthonormal vectors $q_1, q_2, \ldots, q_k \in \mathbb{R}^n$
  3. **For** $i = 1$ to k:
     - Set $\tilde{q}_i \leftarrow a_i$
     - **For** $j = 1$ to $i - 1$: $\tilde{q}_i \leftarrow \tilde{q}_i - (a_i^T q_j) q_j$
     - $q_i \leftarrow \frac{\tilde{q}_i}{\|\tilde{q}_i\|}$

- The algorithm produces orthonormal vectors where $q_i^T q_j = \delta_{ij}$.

- The newly obtained vectors $q_1, \ldots, q_k$ are orthonormal. (Proof: exercise. Hint: induction).

- **Complexity:** $O(kn^2)$ operations for k vectors of dimension n.

AIMS African Institute for Mathematical Sciences SENEGAL

Example: Gram-Schmidt Algorithm

Example: Let's consider the vectors

$$a_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad a_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad a_3 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

- These vectors are not orthogonal (verify by computing dot products).

- **Step 1:** $q_1 = \frac{a_1}{\|a_1\|} = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

AIMS African Institute for Mathematical Sciences SENEGAL

- **Step 2:** Orthogonalize a_2 against q_1:

$$\tilde{q}_2 = a_2 - (a_2^T q_1)q_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \frac{2}{\sqrt{3}} \cdot \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \frac{2}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/3 \\ -2/3 \\ 1/3 \end{bmatrix}$$

Then normalize: $q_2 = \frac{\tilde{q}_2}{\|\tilde{q}_2\|}$

- **Step 3:** Orthogonalize a_3 against q_1 and q_2:

$$\tilde{q}_3 = a_3 - (a_3^T q_1)q_1 - (a_3^T q_2)q_2$$

  Compute the projections and normalize to get q_3.

- The final orthonormal vectors are:

$$q_1, q_2, q_3$$

- **Verification:** Let us check that $q_i^T q_j = \delta_{ij}$ for all i, j.

AIMS African Institute for Mathematical Sciences SENEGAL

- Compute the inner products:

$$q_1^T q_2 = \frac{1}{\sqrt{3}} \cdot \frac{1}{\sqrt{3}} \left(1 \cdot 1 + 1 \cdot 0 + 1 \cdot 1\right) = \frac{1}{3} \cdot 2 = \frac{2}{3}$$

$$q_1^T q_3 = \frac{1}{\sqrt{3}} \cdot \frac{1}{\sqrt{3}} \left(1 \cdot 1 + 1 \cdot 1 + 1 \cdot 0\right) = \frac{1}{3} \cdot 2 = \frac{2}{3}$$

$$q_2^T q_3 = \frac{1}{\sqrt{3}} \cdot \frac{1}{\sqrt{3}} \left(\frac{1}{3} \cdot 1 + \left(-\frac{2}{3}\right) \cdot 1 + \frac{1}{3} \cdot 0\right) = \frac{1}{3} \cdot 0 = 0$$

- Since $q_1^T q_2 = 0$, $q_1^T q_3 = 0$, and $q_2^T q_3 = 0$, the vectors are orthogonal.
- Normalize each vector to ensure they have unit norm:

$$\|q_1\| = \|q_2\| = \|q_3\| = 1$$

- Therefore, the vectors q_1, q_2, q_3 form an orthonormal set.

AIMS | African Institute for Mathematical Sciences SENEGAL

Overview

Matrix Definitions and Notations

- A matrix is a rectangular arrangement of numbers organized in rows and columns, for example:

$$\begin{bmatrix} 1 & 2 & 4 & 0 \\ -2 & 0 & 1 & \pi \\ 2.1 & 10 & -9 & e \end{bmatrix}$$

- Matrix dimensions are denoted as (rows) $\times$ (columns)
  - The example above has dimensions 3×4

- Each value in a matrix is called an element, entry, or component.

- Convention: matrices are denoted with uppercase letters ($A, B, C, M,$ etc.)

- B_{ij} represents the element at row i, column j of matrix B

- For example, in matrix B:

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

- $B_{12} = 2$, $B_{21} = -1$, and $B_{22} = 1$

- The element B_{ij} is located at the intersection of row i and column j.

- The element B_{ij} is often referred to as the (i, j)-th entry of matrix B.

AIMS African Institute for Mathematical Sciences SENEGAL

Tall, Wide, Square Matrices

- An $m \times n$ matrix A can be characterized as:
    - *tall* when $m > n$ (more rows than columns)
    - *wide* when $m < n$ (more columns than rows)
    - *square* when $m = n$ (equal rows and columns)
- Example:

$$(a) \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}, \quad (b) \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{bmatrix}, \quad (c) \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

- Special matrix categories:

  - An $m \times 1$ matrix forms a column vector. For instance: $\begin{bmatrix} 1.2 \\ 0 \end{bmatrix}$.

  - A $1 \times n$ matrix forms a *row* vector, for instance: $[-1, \; 2, \; 0, \; 3]$.

  - A 1×1 matrix represents a scalar value. For instance: $\begin{bmatrix} 3.14 \end{bmatrix}$.

Block Matrix Structure

- The i-th row of A forms an n-dimensional row vector:

$$[A_{i1}, \cdots, A_{in}]$$

- A block matrix contains matrices as elements, structured as:

$$A = \begin{bmatrix} B & C \\ D & E \end{bmatrix}$$

where B, C, D, and E are submatrices (blocks) within A.

- blocks within the same row must share identical row dimensions
- blocks within the same column must share identical column dimensions
- Example: given

$$B = [0\ 2\ 3],\ C = [-1]\ D = \begin{bmatrix} 2 & 2 & 1 \\ 1 & 3 & 3 \end{bmatrix},\ E = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

This yields

$$\begin{bmatrix} B & C \\ D & E \end{bmatrix} = \begin{bmatrix} 0 & 2 & 3 & -1 \\ 2 & 2 & 1 & 3 \\ 1 & 3 & 3 & 2 \end{bmatrix}$$

AIMS | African Institute for Mathematical Sciences SENEGAL

Diagonal and Triangular Matrices

- diagonal matrix: square matrix with $A_{ij} = 0$ when $i \neq j$

- $\text{diag}(a_1, \ldots, a_n)$ denotes a diagonal matrix with $A_{ii} = a_i$ for $i = 1, \ldots, n$

- Examples of diagonal matrices:

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}, \ B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \ C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad (3.1)$$

- lower triangular matrix: $A_{ij} = 0$ when $i < j$

- upper triangular matrix: $A_{ij} = 0$ when $i > j$

- Examples:

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ (upper triangular)}, \quad \begin{bmatrix} 1 & 0 & 0 \\ -1 & 2 & 0 \\ .3 & 1 & 1 \end{bmatrix} \text{ (lower triangular)},$$

Matrix Transpose

- Given matrix A, its **transpose** is the matrix A^T where the rows of A become columns, defined as

$$(A^T)_{ij} = A_{ji}, \ i = 1, \ldots, n, j = 1, \ldots, m$$

- For instance, the transpose of $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ yields the lower

triangular matrix $A^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix}$

AIMS African Institute for Mathematical Sciences SENEGAL

Transpose Properties

- For matrices A, B, and scalar λ:

$$(A + B)^T = A^T + B^T, \ (\lambda A)^T = \lambda A^T, \ (A^T)^T = A \qquad (3.2)$$

- $(A^T)^T = A$

- Matrix addition, subtraction, and scalar multiplication are well-defined operations.

- Specifically

$$(A \pm B)_{ij} = A_{ij} \pm B_{ij}, \ i = 1, \ldots, m, j = 1, \ldots, n$$
$$(\alpha A)_{ij} = \alpha A_{ij}, \ i = 1, \ldots, m, j = 1, \ldots, n$$

- These operations follow standard algebraic properties:

$$A + B = B + A, \ \alpha(A + B) = \alpha A + \alpha B, \ (A + B)^T = A^T + B^T$$
$$(3.3)$$

Overview

Matrix-Vector Multiplication

- The product of an $m \times n$ matrix A and an n-dimensional vector x yields an m-dimensional vector $y = Ax$, where each entry is computed as:

$$y_i = A_{i1}x_1 + \cdots + A_{in}x_n, \ i = 1, \ldots, m$$

$$
\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}
$$

Figure: Computing the first element $y_1 = a_{11}x_1 + a_{12}x_2 + a_{13}x_3$

- This operation can be viewed as computing dot products between matrix rows and the vector:

$$
y_i = a_i^T x, \ i = 1, \ldots, m \tag{3.4}
$$

where $a_1^T, \ldots, a_m^T$ denote the rows of matrix A, i.e.,
$a_i^T = [A_{i1}, A_{i2}, \ldots, A_{in}]$.

- Example:

$$
\begin{bmatrix} 0 & 2 & -1 \\ -1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \cdot 1 + 2 \cdot 2 + (-1) \cdot (-1) \\ (-1) \cdot 1 + 1 \cdot 2 + 0 \cdot (-1) \end{bmatrix}
$$

$$
= \begin{bmatrix} 0 + 4 + 1 \\ -1 + 2 + 0 \end{bmatrix}
$$

$$
= \begin{bmatrix} 5 \\ 1 \end{bmatrix}
$$

Special Cases and Alternative Interpretation

- When multiplying matrix A by the vector of ones $\mathbf{1}$, the result is a vector containing the row sums of A:

$$\begin{bmatrix} 0 & 2 & -1 \\ -1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 + 2 + (-1) \\ -1 + 1 + 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

- The matrix-vector product $y = Ax$ can alternatively be expressed as:

$$y = x_1 a_1 + \cdots + x_n a_n$$

where $a_1, \ldots, a_n$ represent the columns of matrix A.

- This interpretation shows that $y = Ax$ is a linear combination of the columns of A, with weights given by the components $x_1, \ldots, x_n$.

Matrix Multiplication

- For matrices A (dimensions $n \times m$) and B (dimensions $p \times q$)

- The matrix product $C = AB$ is defined if and only if the number of columns in A equals the number of rows in B, i.e., $m = p$

- When this condition is satisfied, the resulting matrix C has dimensions $n \times q$, with elements computed as:

$$C_{ij} = \sum_{k=1}^{m} A_{ik} B_{kj} \tag{3.5}$$

- The product AB is obtained by computing the dot product of each row of A with each column of B
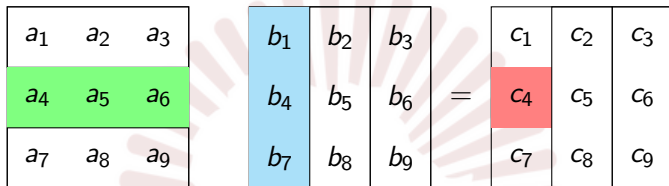
- Equivalently, the matrix product can be expressed as:

$$
C = \begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_n^T \end{bmatrix} \begin{bmatrix} b_1 & b_2 & \cdots & b_q \end{bmatrix} = \begin{bmatrix} a_1^T b_1 & a_1^T b_2 & \cdots & a_1^T b_q \\ a_2^T b_1 & a_2^T b_2 & \cdots & a_2^T b_q \\ \vdots & \vdots & \ddots & \vdots \\ a_n^T b_1 & a_n^T b_2 & \cdots & a_n^T b_q \end{bmatrix} \tag{3.6}
$$

where a_i^T denotes the i-th row of matrix A and b_j represents the j-th column of matrix B.

Figure: Matrix multiplication visualization

Matrix Multiplication Example

- Consider matrices A and B:

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

- The product $C = AB$ is computed as follows:

$$C_{11} = 1 \cdot 7 + 2 \cdot 9 + 3 \cdot 11 = 58, \quad C_{12} = 1 \cdot 8 + 2 \cdot 10 + 3 \cdot 12 = 64$$

$$C_{21} = 4 \cdot 7 + 5 \cdot 9 + 6 \cdot 11 = 139, \quad C_{22} = 4 \cdot 8 + 5 \cdot 10 + 6 \cdot 12 = 154$$

- Thus, the resulting matrix C is:

$$C = \begin{bmatrix} 58 & 64 \\ 139 & 154 \end{bmatrix}$$

- Note that matrix multiplication is not commutative, i.e., $AB \neq BA$ in general

- However, it is associative, i.e., $(AB)C = A(BC)$

- It is also distributive over addition, i.e., $A(B + C) = AB + AC$

- The identity matrix I_n serves as the multiplicative identity for matrices, i.e., $AI_n = A$ and $I_nA = A$ for any compatible matrix A

- The zero matrix O acts as the additive identity, satisfying $A + O = A$ and $O + A = A$ for any matrix A

AIMS | African Institute for Mathematical Sciences SENEGAL

Overview

AIMS African Institute for Mathematical Sciences SENEGAL

Linear Transformations

- A mapping $T : \mathbb{R}^m \to \mathbb{R}^n$ is called **linear** if it satisfies

$$T(\alpha u + \beta v) = \alpha T(u) + \beta T(v), \quad \text{for all } u, v \in \mathbb{R}^m, \ \alpha, \beta \in \mathbb{R} \quad (4.1)$$

- Consider an $m \times n$ matrix A. It defines a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$ since

$$A(\lambda x + \beta y) = A(\lambda x) + A(\beta y) = \lambda Ax + \beta Ay \quad (4.2)$$

- Another example: given $a \in \mathbb{R}^n$, the inner product $T(u) = \langle a, u \rangle$ for all $u \in \mathbb{R}^n$ is linear. Indeed, $T(\lambda u) = \langle a, \lambda u \rangle = \lambda \langle a, u \rangle = \lambda T(u)$ and $T(u + v) = \langle a, u + v \rangle = \langle a, u \rangle + \langle a, v \rangle = T(u) + T(v)$.

Exercise

- Determine which of the following functions $T : \mathbb{R}^2 \to \mathbb{R}^2$ represent linear transformations:

  1. $T(x_1, x_2) = (1 + x_1, x_2)$
  2. $T(x_1, x_2) = (x_2, x_1)$
  3. $T(x_1, x_2) = (x_1 - x_2, 0)$
  4. $T(x_1, x_2) = (\sin x_1, x_2)$
  5. $T(x_1, x_2) = (x_1^2, x_2)$

- Given a linear transformation $T : \mathbb{R}^m \to \mathbb{R}^n$, prove that $T(\mathbf{0}) = \mathbf{0}$. Does the property $T(\mathbf{1}) = \mathbf{1}$ hold for the vector of ones?

Matrix Representation of Linear Transformations

- A linear transformation $T : \mathbb{R}^m \rightarrow \mathbb{R}^n$ can be represented by an $n \times m$ matrix A such that for any vector $x \in \mathbb{R}^m$:

$$T(x) = Ax$$

- The matrix A is constructed from the images of the standard basis vectors:

$$A = \begin{bmatrix} T(e_1) & T(e_2) & \cdots & T(e_m) \end{bmatrix}$$

where e_i are the standard basis vectors in $\mathbb{R}^m$.

AIMS African Institute for Mathematical Sciences SENEGAL

- The columns of A are the transformed basis vectors, i.e.,
  $A = [T(e_1), T(e_2), \ldots, T(e_m)]$.
- For example, if $T(e_1) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $T(e_2) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, then the matrix representation is:

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

- This matrix A represents the identity transformation in $\mathbb{R}^2$.
- The linear transformation can be visualized as a mapping of vectors in $\mathbb{R}^m$ to $\mathbb{R}^n$ through the matrix multiplication Ax.

AIMS African Institute for Mathematical Sciences SENEGAL

- Example: For a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T(x_1, x_2) = (2x_1 + x_2, -x_1 + 3x_2)$, the matrix representation is:

$$A = \begin{bmatrix} T(e_1) & T(e_2) \\ \downarrow & \downarrow \\ 2 & 1 \\ -1 & 3 \end{bmatrix}$$

where $T(e_1) = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$ and $T(e_2) = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$.

- The matrix A can be used to compute the transformation of any vector $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ in $\mathbb{R}^2$ as:

$$T(x) = Ax = \begin{bmatrix} 2 & 1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

- This results in a new vector in $\mathbb{R}^2$, which is the image of x under the transformation T.
- The matrix representation allows for efficient computation of linear transformations, especially in higher dimensions.

AIMS | African Institute for Mathematical Sciences SENEGAL

Exercices:

- Prove that the matrix representation of a linear transformation is unique.

- Show that if T is a linear transformation, then $T(\mathbf{0}) = \mathbf{0}$.

- Given a linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^2$ defined by $T(x_1, x_2, x_3) = (x_1 + 2x_2, 3x_3)$, find its matrix representation.

- For the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T(x_1, x_2) = (x_1 - x_2, 2x_1 + x_2)$, compute the image of the vector $x = (1, 2)$.

- If $T : \mathbb{R}^2 \to \mathbb{R}^2$ is defined by $T(x_1, x_2) = (3x_1 + x_2, -x_1 + 4x_2)$, find the matrix representation and compute $T(e_1)$ and $T(e_2)$, where $e_1 = (1, 0)$ and $e_2 = (0, 1)$.

AIMS African Institute for Mathematical Sciences SENEGAL

Image, Kernel, and Rank

Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation. The following concepts are essential in understanding the properties of T:

- The **image** (or **range**) of T is the set of all vectors that can be obtained as outputs, i.e.

$$\text{Im}(T) = \{ T(x) \in \mathbb{R}^n : x \in \mathbb{R}^m \}$$

- The **kernel** (or **null space**) of T consists of all input vectors that are mapped to the zero vector, i.e.

$$\text{Ker}(T) = \{ x \in \mathbb{R}^m : T(x) = \mathbf{0} \}$$

- The **rank** of a linear transformation equals the dimension of its image, written as rank(T).

- In other words, the rank is the number of linearly independent vectors in the image of T.

- The **nullity** of a linear transformation equals the dimension of its kernel, written as nullity(T).

Example: Consider the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^2$ defined by:

$$T(x_1, x_2, x_3) = (x_1 + 2x_2, 3x_3)$$

- The image of T is the set of all vectors in $\mathbb{R}^2$ that can be expressed as $(x_1 + 2x_2, 3x_3)$ for some $x_1, x_2, x_3 \in \mathbb{R}$.
- That is, the image consists of all vectors of the form (a, b) where a can take any value in $\mathbb{R}$ and b can take any value in $\mathbb{R}$, leading to $\text{Im}(T) = \mathbb{R}^2$.

AIMS

- The kernel of T consists of all vectors (x_1, x_2, x_3) such that $T(x_1, x_2, x_3) = (0, 0)$, which leads to the equations $x_1 + 2x_2 = 0$ and $3x_3 = 0$.

- This implies that x_3 must be zero, and x_1 can be expressed in terms of x_2 as $x_1 = -2x_2$. Thus, the kernel can be described by the equation $x_1 + 2x_2 = 0$ and $x_3 = 0$, which is a line in $\mathbb{R}^3$.

- That is, $\text{Ker}(T) = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1 = -2x_2, x_3 = 0\}$.

- Or, the kernel can be expressed as:

$$\text{Ker}(T) = \left\{ \begin{bmatrix} -2t \\ t \\ 0 \end{bmatrix} \in \mathbb{R}^3 : t \in \mathbb{R} \right\} = \text{span} \left(\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \right)$$

AIMS | African Institute for Mathematical Sciences SENEGAL

- The nullity of T is the dimension of the kernel, which is 1, as it is spanned by one vector only.

- The rank of T is the dimension of the image, which in this case is 2, since the image spans a plane in $\mathbb{R}^2$.

Rank-Nullity Theorem

For a linear transformation $T : \mathbb{R}^m \to \mathbb{R}^n$, the following relationship holds:

$$\text{rank}(T) + \text{nullity}(T) = m$$

where rank(T) is the dimension of the image of T and nullity(T) is the dimension of the kernel of T.

Note: This theorem connects the dimensions of the image and kernel of a linear transformation to the dimension of the domain.

Example: For the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^2$ defined by
$T(x_1, x_2, x_3) = (x_1 + 2x_2, 3x_3)$:

- The rank of T is 2 (dimension of the image $\mathbb{R}^2$).

- The nullity of T is 1 (dimension of the kernel, which is a line in $\mathbb{R}^3$).

- Thus, by the rank-nullity theorem:

$$\text{rank}(T) + \text{nullity}(T) = 2 + 1 = 3$$

which matches the dimension of the domain $\mathbb{R}^3$.

Exercise: Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation.

- Prove that $Ker(T)$ is a subspace of $\mathbb{R}^m$.

- Show that $Im(T)$ is a subspace of $\mathbb{R}^n$.

Exercice:

- Prove the rank-nullity theorem for a linear transformation
  $T : \mathbb{R}^m \to \mathbb{R}^n$.

- Given a linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^2$ defined by
  $T(x_1, x_2, x_3) = (x_1 + 2x_2, 3x_3)$, find the image and kernel of T.

- Determine the rank and nullity of the transformation T.

- Verify that the rank-nullity theorem holds for this transformation.

AIMS | African Institute for Mathematical Sciences SENEGAL

Exercise: For the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^2$ defined by
$T(x_1, x_2, x_3) = (x_1 + 2x_2, 3x_3)$:

- Find the image and kernel of T.

- Determine the rank and nullity of T.

- Verify that the rank-nullity theorem holds, i.e.,
  $\text{rank}(T) + \text{nullity}(T) = m$, where m is the dimension of the domain
  $\mathbb{R}^3$.

Overview

1. Motivation
2. Vectors
   - Definitions and Notations
   - Vectors Operations
   - Vector Spaces
   - Inner Products and Norms
     - Inner product
     - Norms, Distance and Angle
     - Some Norms Inequalities
     - Angles Between Vectors
     - Orthogonality
3. Matrices
   - Matrices Definitions
   - Matrix Operations

Rotation Transformations in $\mathbb{R}^2$

- **Two-dimensional rotation:**

$\Phi(e_2) := [-\sin\theta, \cos\theta]^T$
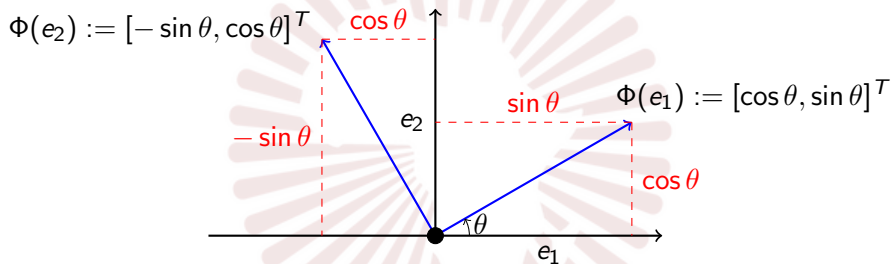
Figure: Standard basis rotation in $\mathbb{R}^2$ through angle θ

$\Phi(e_1) := [\cos\theta, \sin\theta]^T$

Matrix Representation of 2D Rotation

- **Derivation:** From the transformed basis vectors in the previous slide:

$$\Phi(e_1) = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}, \quad \Phi(e_2) = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix} \tag{4.3}$$

- The rotation transformation matrix R_θ has these as columns:

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{4.4}$$

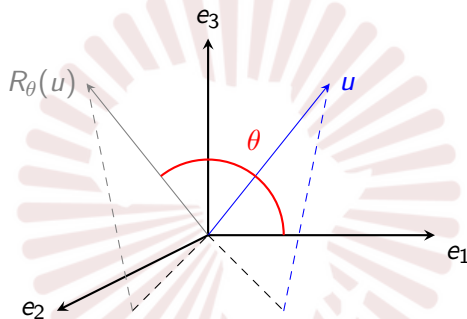**AIMS** African Institute for Mathematical Sciences SENEGAL

- Given a vector $\mathbf{x}$, its rotation by angle θ yields $\mathbf{y}$:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{4.5}$$

Rotation Transformations in $\mathbb{R}^3$

Figure: Vector rotation in $\mathbb{R}^3$ around the e_3-axis by angle θ

Rotation Matrices in Three Dimensions

- **Rotation around the e_1-axis:**

$$R_1(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \tag{4.6}$$

- **Rotation around the e_2-axis:**

$$R_2(\theta) = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} \tag{4.7}$$

Additional 3D Rotation Matrix and Key Properties

- **Rotation around the e_3-axis:**

$$R_3(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4.8}$$

- In each rotation, the corresponding axis e_i remains invariant.

Exercises on Rotation Matrices

Problem: Demonstrate the following fundamental properties of rotation matrices:

1. Distance preservation: For any angle θ, verify that:

$$\|\mathbf{x} - \mathbf{y}\| = \|R_\theta(\mathbf{x}) - R_\theta(\mathbf{y})\|$$

   holds for all vectors $\mathbf{x}, \mathbf{y}$.

2. Rotation composition: Prove that $R_{\theta+\phi} = R_\theta \cdot R_\phi$ for all $\theta, \phi \in [0, 2\pi)$.

- Identity and inverse relations: Establish that $R_0 = I$ and $R_{-\theta} = R_\theta^{-1}$.

- Angular preservation: Demonstrate that rotations maintain inner products:

$$\langle R_\theta \mathbf{x}, R_\theta \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle$$

for all vectors $\mathbf{x}, \mathbf{y}$.

Exercise: Consider the rotation matrix R_θ defined as:

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{4.9}$$

- Show that R_θ is orthogonal, i.e., $R_\theta^T R_\theta = I$.
- Verify that the determinant of R_θ is 1, confirming that it represents a rotation.
- Prove that the inverse of R_θ is its transpose, i.e., $R_\theta^{-1} = R_\theta^T$.

 AIMS African Institute for Mathematical Sciences SENEGAL

Exercise: Given a vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, compute its rotation by angle θ using the matrix R_θ:

$$\mathbf{y} = R_\theta \mathbf{x} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{4.10}$$

- Calculate the resulting vector $\mathbf{y}$.
- Discuss how the rotation affects the coordinates of $\mathbf{x}$.

AIMS | African Institute for Mathematical Sciences SENEGAL

Exercise: Consider the rotation matrix R_θ and a vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

- Compute the rotated vector $\mathbf{y} = R_\theta \mathbf{x}$.

- Verify that the angle between $\mathbf{x}$ and $\mathbf{y}$ is θ.

- Discuss the geometric interpretation of this rotation in the context of the standard basis vectors.
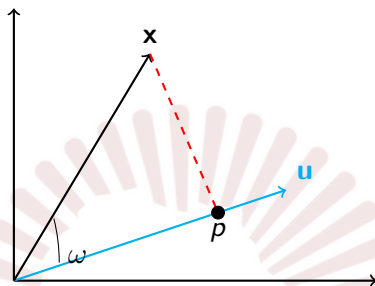
Overview

Projection onto a line

- **Vector projection onto a line**: Given a vector x, its projection p onto a line spanned by vector u is defined as

$$p = \frac{x^T u}{\|u\|^2} u \qquad (4.11)$$

- The residual $e = x - p$ is orthogonal to u, satisfying $u^T e = 0$.
- **Projection matrix**: The projection can be expressed as $p = Px$ where P is the projection matrix:

$$P = \frac{uu^T}{\|u\|^2} \qquad (4.12)$$

- **Example**: Consider vectors x and u:

$$x = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \quad u = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

- Since $\|u\|^2 = 1$ and $x^T u = 2 \cdot 1 + 3 \cdot 0 = 2$

- The projection is $p = 2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$

- **Matrix example**: For projection onto the line spanned by $b = [1, 2, 2]^T$:

$$P = \frac{bb^T}{\|b\|^2} = \frac{1}{9} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 2 \end{bmatrix} = \frac{1}{9} \begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{bmatrix} \tag{4.13}$$

- For any vector x, its projection onto this line is:

$$p = Px = \frac{1}{9} \begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \frac{1}{9} \begin{bmatrix} x_1 + 2x_2 + 2x_3 \\ 2x_1 + 4x_2 + 4x_3 \\ 2x_1 + 4x_2 + 4x_3 \end{bmatrix} \tag{4.14}$$

AIMS | African Institute for Mathematical Sciences SENEGAL

Exercise

1. Calculate the projection vector p and determine the projection matrix P for the given vectors:

$$a) \ x = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \ u = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \quad b) \ x = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, u = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} \quad (4.15)$$

2. For both cases above, evaluate P^2 and P^3. What pattern do you notice?

3. Verify that the residual $e = x - p$ is orthogonal to u, i.e., $u^T e = 0$

4. Demonstrate that $\|p\| = \|x\| \cos \theta$, where θ represents the angle between x and u.

Projection onto a Plane

- **Projection of a vector onto a plane**: Given a vector x and a 2D plane π, the projection of x onto π is a vector x' given by

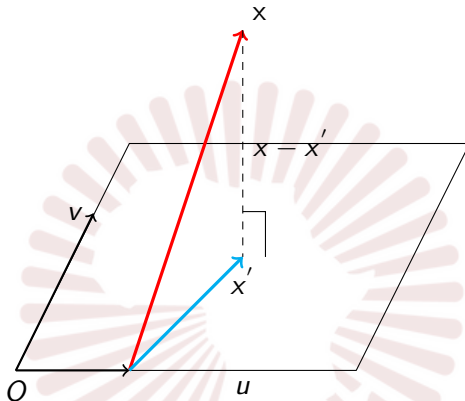$$x' = \frac{x^T u}{\|u\|^2} u + \frac{x^T v}{\|v\|^2} v \qquad (4.16)$$

where u and v are orthogonal basis vectors spanning the plane.

- In matrix notation, the projection x' of x is:

$$x' = \left(I - nn^T\right) x, \qquad (4.17)$$

where n is the unit normal vector to the plane, and $P = I - nn^T$ is the orthogonal projection matrix onto the plane. **AIMS** | African Institute for Mathematical Sciences SENEGAL

Figure: Projection onto a plane spanned by u and v. The projection x' can be expressed as a linear combination of u and v.

- **Example**: Consider a plane with normal vector *n* and a vector *x* to be projected:

$$n = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad x = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

- First, verify that *n* is a unit vector: $\|n\| = 1$ (which is already the case).

- Calculate the outer product nn^T:

$$nn^T = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- Determine the projection matrix *P*:

- The projection matrix P is:

$$P = I - nn^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

- The projection of x onto the plane is:

$$x' = Px = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$$

Projection onto a Subspace

- **General subspace projection**: Given n linearly independent vectors $a_1, \ldots, a_n \in \mathbb{R}^m$, we seek the projection p of vector b onto the subspace spanned by these vectors:

$$p = \hat{p}_1 a_1 + \cdots + \hat{p}_n a_n \qquad (4.18)$$

where the coefficients $\hat{p}_i$ are determined by the orthogonality condition.

- Let $A = [a_1 \ a_2 \ \cdots \ a_n]$ be the matrix with columns a_i. The projection coefficients satisfy:

$$A^T(b - A\hat{p}) = 0 \quad \Rightarrow \quad A^T A \hat{p} = A^T b \qquad (4.19)$$

Projection onto General Subspaces

- **Setting**: Let $x \in \mathbb{R}^n$ and $U \subset \mathbb{R}^n$ be a subspace with $\dim(U) = m \geq 1$.

- Let $\{b_1, b_2, \ldots, b_m\}$ be a basis for U, and define $B = [b_1 \; b_2 \; \cdots \; b_m]$.

- Any vector in U can be written as:

$$x' = \sum_{i=1}^{m} \lambda_i b_i = B\lambda$$

  for some coefficient vector $\lambda = [\lambda_1, \ldots, \lambda_m]^T$.

- The orthogonal projection of x onto U is:

$$x' = B(B^T B)^{-1} B^T x \tag{4.21}$$

where $P = B(B^T B)^{-1} B^T$ is the projection matrix onto U.

Exercices

1. Given a vector $x = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and a subspace spanned by $b_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ and

   $b_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, compute the projection of x onto the subspace.

2. For the same vectors, calculate the projection matrix P and verify that $P^2 = P$.

3. Show that the error vector $e = x - x'$ is orthogonal to both b_1 and b_2.

4. Prove that the projection x' satisfies $\|x'\| = \|b_1\| \cos \theta_1 + |b_2| \cos \theta_2$, where θ_1 and θ_2 are the angles between x' and b_1, and x' and b_2, respectively.

AIMS | African Institute for Mathematical Sciences SENEGAL

Overview

Determinants

- For any square matrix $A \in \mathbb{R}^{n \times n}$, the determinant $\det(A)$ (also written as $|A|$) is a scalar value given by

$$
\det(A) = \begin{vmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
a_{n1} & a_{n2} & \cdots & a_{nn}
\end{vmatrix} \tag{5.1}
$$

- Determinants are only defined for square matrices.

- For any $n \times n$ matrix A, we can compute $\det A$ using cofactor expansion:

$$\det A = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det A_{ij} \text{ or } \det A = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det A_{ij} \quad (5.2)$$

where a_{ij} is the element at row i and column j, and A_{ij} is the (i, j)-minor of matrix A.

- For a 2×2 matrix A, we have:

$$\det A = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11} a_{22} - a_{12} a_{21} \quad (5.3)$$

AIMS | African Institute for Mathematical Sciences SENEGAL

- For a 3×3 matrix A, applying Eq. 5.2 gives:

$$
\det A = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} -
$$

$$
a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}
$$

- The determinant serves as a test for matrix invertibility.

Determinant and Matrix Invertibility

- A square matrix $A \in \mathbb{R}^{n \times n}$ is invertible if and only if $\det(A) \neq 0$.

- Proof: Left as an exercise.

- Example: Consider the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 5 & 6 & 0 \end{bmatrix} \tag{5.4}$$

- We calculate:
  $\det A = 1(1 \cdot 0 - 4 \cdot 6) - 2(0 \cdot 0 - 4 \cdot 5) + 3(0 \cdot 6 - 1 \cdot 5) = -24 + 40 - 15 = 1.$

- Since $\det A = 1 \neq 0$, the matrix A is invertible.

AIMS | African Institute for Mathematical Sciences SENEGAL

Properties

- Matrix determinants satisfy the following properties:
  1. $\det(AB) = \det(A) \times \det(B)$
  2. $\det(A) = \det(A^T)$
  3. If A is invertible, then $\det(A^{-1}) = \frac{1}{\det(A)}$.
  4. $\det(\lambda A) = \lambda^n \det(A)$
  5. Exchanging two rows or columns negates $\det(A)$.

Trace

- The trace of $A \in \mathbb{R}^{n \times n}$ is the scalar defined as

$$\text{tr}(A) := \sum_{i=1}^{n} a_{ii} \tag{5.5}$$

- In other words, the trace equals the sum of all diagonal entries of A.
- It has the following properties: (Exercise)
    1. $\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$ for $A, B \in \mathbb{R}^{n \times n}$
    2. $\text{tr}(\lambda A) = \lambda \text{tr}(A)$
    3. $\text{tr}(I_n) = n$
    4. $\text{tr}(AB) = \text{tr}(BA)$ for all $A \in \mathbb{R}^{n \times k}$, $B \in \mathbb{R}^{k \times n}$.
- The trace remains unchanged under cyclic permutations:

$$\text{tr}(AKL) = \text{tr}(KLA) \tag{5.6}$$

Overview

Eigenvalues and eigenvectors

- In this section, we work exclusively with square matrices.

- A scalar $\lambda \in \mathbb{R}$ is an **eigenvalue of A** if there exists a non-zero vector x such that

$$Ax = \lambda x \tag{7.1}$$

- For example, $\lambda = 1$ is an eigenvalue of the identity matrix $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Also, $\lambda = 5$ is an eigenvalue of $A = \begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix}$ since $A \begin{bmatrix} 2 \\ 1 \end{bmatrix} = 5 \begin{bmatrix} 2 \\ 1 \end{bmatrix}$.

AIMS African Institute for Mathematical Sciences SENEGAL

Eigenvectors

- Any vector x that satisfies $Ax = \lambda x$ for some scalar λ is called an **eigenvector of** A associated with eigenvalue λ.

- We refer to equation 7.1 as **the eigenvalue equation**.

- As an example, $x = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ is an eigenvector of $A = \begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix}$ corresponding to eigenvalue $\lambda = 5$.

- Finding eigenvalues and eigenvectors for a given matrix A requires systematic computation.

AIMS | African Institute for Mathematical Sciences SENEGAL

Computing eigenvalues

- For any matrix A, these statements are equivalent:
    1. λ is an eigenvalue of A
    2. There exists a non-zero vector $x \in \mathbb{R}^n$ such that $Ax = \lambda x$
    3. $\det(A - \lambda I) = 0$

- Define $p_A(\lambda) = \det(A - \lambda I)$. This function p_A is a polynomial in λ.

- **The eigenvalues of matrix A correspond exactly to the roots of polynomial p_A.**

AIMS African Institute for Mathematical Sciences SENEGAL

Exercise

- Determine the eigenvalues of these matrices:

$$a)\ A = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} \quad b)\ I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad c)\ C = \begin{bmatrix} 4 & 1 & 0 \\ 1 & 4 & 0 \\ 0 & 0 & 2 \end{bmatrix} \quad (7.2)$$

$$d)\ U = \begin{bmatrix} 3 & 5 & -1 \\ 0 & 2 & 4 \\ 0 & 0 & 7 \end{bmatrix} \quad e)\ L = \begin{bmatrix} 1 & 0 & 0 \\ 4 & -3 & 0 \\ 2 & 5 & 6 \end{bmatrix} \quad (7.3)$$

What common property do you notice for matrices d) and e)?

AIMS | African Institute for Mathematical Sciences SENEGAL

- For any matrix A:

  1. Show that if x is an eigenvector of A with eigenvalue λ, then any non-zero scalar multiple cx is also an eigenvector of A with the same eigenvalue λ.

  2. Show that if λ is an eigenvalue of A, then λ^k is an eigenvalue of A^k for any positive integer k.

  3. For a 2×2 triangular matrix A, show that the eigenvalues are the diagonal elements.

  4. For any 2×2 matrix A, find the characteristic polynomial p_A and verify that:

  $$p_A(\lambda) = \lambda^2 - \lambda \text{tr}(A) + \det(A) \tag{7.4}$$

  5. Show that if A is symmetric, then all eigenvalues of A are real.

Finding Eigenvectors

- For a matrix A, these statements are equivalent:
  1. x is an eigenvector of A with eigenvalue λ
  2. x satisfies the equation $Ax = \lambda x$
  3. x is a non-zero solution to the homogeneous system $(A - \lambda I)x = 0$
  4. x lies in the nullspace of $A - \lambda I$, i.e., $x \in \ker(A - \lambda I)$
  5. The matrix $A - \lambda I$ is singular (non-invertible), which means $\det(A - \lambda I) = 0$

- **Eigenvectors corresponding to λ are non-zero solutions to** $(A - \lambda I)x = 0$.

AIMS African Institute for Mathematical Sciences SENEGAL

Example: Finding Eigenvectors

1. Consider $A = \begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix}$. We will find its eigenvectors.

2. First, we compute the eigenvalues using the characteristic equation:

$$\det(A - \lambda I) = \begin{vmatrix} 4 - \lambda & 2 \\ 1 & 3 - \lambda \end{vmatrix} = 0 \tag{7.5}$$

$$= (4 - \lambda)(3 - \lambda) - 2 \tag{7.6}$$

$$= \lambda^2 - 7\lambda + 10 = 0 \Longrightarrow \lambda_1 = 5, \ \lambda_2 = 2 \tag{7.7}$$

3. Next, we find eigenvectors by solving $(A - \lambda I)x = 0$.

AIMS | African Institute for Mathematical Sciences SENEGAL

- **For eigenvalue** $\lambda = 5$:

$$(A - 5I) \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1 & 2 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0 \implies x = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

- **For eigenvalue** $\lambda = 2$:

$$(A - 2I) \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0 \implies x = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Exercise

1. Compute the eigenvectors corresponding to the eigenvalues of matrices $(a) - (e)$ from the previous slide.

2. If x and y are eigenvectors of matrix A with the same eigenvalue λ, prove that any linear combination $\alpha x + \beta y$ is also an eigenvector for λ.

3. Can the zero vector serve as an eigenvector? Explain your reasoning.

4. Is it possible for a single eigenvector to correspond to two distinct eigenvalues? Provide justification.

5. If matrix A has eigenvalue $\lambda = 0$, what can you conclude about the properties of A, particularly its invertibility?

AIMS | African Institute for Mathematical Sciences SENEGAL

Algebraic Multiplicity

- An eigenvalue λ of matrix A corresponds to a root of the characteristic polynomial $p_A(\lambda)$.

- The algebraic multiplicity of eigenvalue λ_i refers to how many times λ_i appears as a root in p_A.

- Example: For a 3×3 matrix A with characteristic polynomial

$$p_A(\lambda) = (\lambda - 1)^2(\lambda - 3) \tag{7.8}$$

eigenvalue $\lambda = 1$ has algebraic multiplicity 2, while $\lambda = 3$ has multiplicity 1.

Exercise

- Determine the algebraic multiplicity of each eigenvalue for the following matrices:

$$a)\ A = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix} \quad b)\ B = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \quad c)\ C = \begin{bmatrix} 4 & 1 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix} \quad (7.9)$$

Geometric Multiplicity

- For an eigenvalue λ of matrix A, the **geometric multiplicity** equals $\dim(\ker(A - \lambda I))$.

- Equivalently, it represents the **maximum number of linearly independent eigenvectors associated with** λ.

- Example: For matrix $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$, eigenvalue $\lambda = 2$ has geometric multiplicity 2.

Eigenspace and Spectrum

- Given matrix A and eigenvalue λ, the **eigenspace** E_λ contains all corresponding eigenvectors:

$$E_\lambda = \{x \in \mathbb{R}^n \mid (A - \lambda I)x = 0\} = \ker(A - \lambda I) \qquad (7.10)$$

- The eigenspace E_λ forms a subspace spanned by eigenvectors of λ.

- From our earlier example with $\lambda = 5$, the eigenspace is:

$$E_5 = \text{span} \left\{ \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right\} = \left\{ \begin{bmatrix} 2\alpha \\ \alpha \end{bmatrix} \mid \alpha \in \mathbb{R} \right\} \qquad (7.11)$$

- Note: **geometric multiplicity = dimension of eigenspace**.
- The **spectrum** (or **eigenspectrum**) of A, denoted $\sigma(A)$, is the **complete set of eigenvalues**:

$$\sigma(A) = \{\lambda \in \mathbb{R} \mid \det(A - \lambda I) = 0\} \qquad (7.12)$$

- Example: For $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ with eigenvalues $\lambda = 1$ and $\lambda = 3$, we have $\sigma(A) = \{1, 3\}$.

Computing Eigenvalues, Eigenvectors, and Eigenspaces

- Consider the 2×2 matrix:

$$A = \begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix} \tag{7.13}$$

- **Characteristic polynomial**:

$$p_A(\lambda) = \det(A - \lambda I) = \begin{vmatrix} 4 - \lambda & 2 \\ 1 & 3 - \lambda \end{vmatrix} = (4 - \lambda)(3 - \lambda) - 2 \tag{7.14}$$

$$= \lambda^2 - 7\lambda + 10 \tag{7.15}$$

$$= (\lambda - 2)(\lambda - 5) \tag{7.16}$$

- **Eigenvalues**: The roots are $\lambda_1 = 2$ and $\lambda_2 = 5$, each with algebraic

- **Eigenvectors and Eigenspaces**: For eigenvalue $\lambda = 5$, we find eigenvectors by solving:

$$(A - \lambda I)x = 0 \implies \begin{bmatrix} 4 - 5 & 2 \\ 1 & 3 - 5 \end{bmatrix} x = 0 \qquad (7.17)$$

- This gives us the system:

$$\begin{bmatrix} -1 & 2 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0 \qquad (7.18)$$

- From the first row: $-x_1 + 2x_2 = 0 \Rightarrow x_1 = 2x_2$
- Setting $x_2 = t$, the eigenspace is:

$$E_5 = \text{span} \left\{ \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right\} = \left\{ t \begin{bmatrix} 2 \\ 1 \end{bmatrix} \mid t \in \mathbb{R}, t \neq 0 \right\} \qquad (7.19)$$

- The eigenvector and eigenspace associated to $\lambda = 5$ are

$$v_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, E_5 = \text{span}\{v_1\} := \{cv_1 \mid c \in \mathbb{R}\}$$

- For $\lambda = 2$, we find analogously the equations

$$\begin{bmatrix} 4-2 & 2 \\ 1 & 3-2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0 \iff \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$

which gives $x_2 = -x_1$, such as $v_2 = \begin{bmatrix} 1 & -1 \end{bmatrix}^T$, is an eigenvector. The corresponding eigenspace is $E_2 = \text{span}\{v_2\}$.

AIMS | African Institute for Mathematical Sciences SENEGAL

Back to Matrix Norms: Spectral Norm

- For any matrix A (not necessarily square), the **spectral norm** $\|A\|_2$ is defined as the square root of the largest eigenvalue of $A^T A$:

$$\|A\|_2 = \sqrt{\lambda_{\max}(A^T A)} \tag{7.20}$$

where $\lambda_{\max}$ denotes the maximum eigenvalue.

- Example: Let $A = \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}$. First, compute $A^T A$:

$$A^T A = \begin{bmatrix} 1 & 0 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 8 \end{bmatrix} \tag{7.21}$$

AIMS | African Institute for Mathematical Sciences SENEGAL

- To find the eigenvalues of $A^T A$, we solve the characteristic equation:

$$\det(A^T A - \lambda I) = \begin{vmatrix} 1 - \lambda & 2 \\ 2 & 8 - \lambda \end{vmatrix} = (1 - \lambda)(8 - \lambda) - 4 = 0 \quad (7.22)$$

- Solving yields eigenvalues $\lambda_1 \approx 8.531$ and $\lambda_2 \approx 0.469$.

- Therefore, the spectral norm is:

$$\|A\|_2 = \sqrt{\lambda_{\max}} = \sqrt{8.531} \approx 2.921 \quad (7.23)$$

- Compare this with other norms: What are the Frobenius norm, 1-norm, and ∞-norm of A? What patterns do you observe?

Spectral theorem

- Racall that a symmetric matrix A is such that $A^T = A$.
- For example, the following matrices are symmetric

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} = A^T \text{ and } D = \begin{bmatrix} -2 & 0 & 1 \\ 0 & 3 & 2 \\ 1 & 2 & 0 \end{bmatrix} = D^T \qquad (7.24)$$

- What are the eigenvalues of A and D and the eigenvectors for each? What do you observe?
- Let's consider A. A quick computation yields

$$\lambda_1 = 1, \quad \lambda_2 = 3, \; v_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \text{ and } v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad (7.25)$$

- Property 1: All eigenvalues of A are real numbers.
- Property 2: The eigenvectors of A are orthogonal to each other, as shown by:

$$v_1^T v_2 = \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 1 \cdot 1 + (-1) \cdot 1 = 0 \qquad (7.26)$$

- By normalizing v_1 and v_2 to unit vectors $x_1 = \frac{v_1}{\|v_1\|}$ and $x_2 = \frac{v_2}{\|v_2\|}$, we obtain the orthogonal matrix Q satisfying $Q^T Q = I$:

$$Q = \begin{bmatrix} x_1 & x_2 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \qquad (7.27)$$

- Property 3: Matrix A can be decomposed as:

$$A = Q \Lambda Q^T = \frac{1}{\sqrt{}} \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} \frac{1}{\sqrt{}} \begin{bmatrix} 1 & -1 \end{bmatrix} \qquad (7.28)$$

Spectral Theorem for Symmetric Matrices

- **Spectral Theorem**: If A is a symmetric matrix (i.e., $A = A^\top$), then:
  1. Every eigenvalue of A is real
  2. $\mathbb{R}^n$ has an orthonormal basis formed entirely by eigenvectors of A

- From property 2, we can construct an orthogonal matrix Q such that

$$A = QDQ^T \tag{7.29}$$

where D contains the eigenvalues of A on its diagonal, and the columns of Q are the corresponding normalized eigenvectors.

- **Eigenvalues are real**. For any complex eigenpair (λ, x):

$$Ax = \lambda x \Rightarrow x^T A x = \lambda x^T x \Rightarrow \lambda = \overline{\lambda} \Rightarrow \lambda \in \mathbb{R}$$

- **Orthonormal eigenvector basis exists**:

  1. Choose an eigenvalue λ of A with corresponding unit eigenvector $v \in \mathbb{R}^n$. Let $V_1 = \text{Span}(v)$ and consider its orthogonal complement $V_1^{\perp}$

  2. **Key insight:** $V_1^{\perp}$ is A-invariant, meaning $A(V_1^{\perp}) \subseteq V_1^{\perp}$

  3. **Induction:** Apply the theorem to the restriction of A on $V_1^{\perp}$ (dimension $n - 1$), which remains symmetric

Proof, continued

5 **Step 5:** By the induction hypothesis, there exists an orthonormal basis of eigenvectors for the subspace $V_1^{\perp}$

6 **Conclusion:** By combining the unit eigenvector v with the orthonormal eigenvectors from $V_1^{\perp}$, we construct a complete orthonormal basis of $\mathbb{R}^n$ consisting entirely of eigenvectors of A

Example

- Consider the 2×2 symmetric matrix:

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

This matrix admits the following spectral decomposition:

$$A = QDQ^T \iff \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}^T$$

- Here, $\lambda_1 = 4$ and $\lambda_2 = 2$ are the eigenvalues of A, while the columns of Q contain the corresponding normalized eigenvectors.

- What is the determinant of A?

AIMS | African Institute for Mathematical Sciences SENEGAL

Relationship Between Determinants, Traces, and Eigenvalues

- **Determinant formula:** For any matrix A, its determinant equals the product of all eigenvalues:

$$\det(A) = \prod_{i=1}^{n} \lambda_i \qquad (7.30)$$

where λ_i denote the eigenvalues of A.

- **Invertibility criterion:** Matrix A is invertible if and only if all eigenvalues are non-zero.

- **Trace formula:** The trace of matrix A equals the sum of all eigenvalues:

Eigenvalue Localization: Gershgorin Circle Theorem

- **Theorem:** For any matrix A, all eigenvalues lie within the union of Gershgorin discs:

$$\sigma(A) \subseteq \bigcup_{i=1}^{n} \left\{ \lambda \in \mathbb{C} \mid |\lambda - a_{ii}| \leq \sum_{j=1,\, j \neq i}^{n} |a_{ij}| \right\} \tag{7.32}$$

- Each disc D_i is centered at diagonal element a_{ii} with radius equal to the sum of absolute values of off-diagonal elements in row i.

- This theorem provides bounds for eigenvalue locations without computing them explicitly.

Example: Applying Gershgorin's Theorem

- Consider the matrix:

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

- We construct three Gershgorin discs, one for each row.

- **Disc 1 (Row 1):**

  Center: $a_{11} = 5$, Radius: $R_1 = |a_{12}| + |a_{13}| = |1| + |0| = 1$

  $$D_1 = \{\lambda \in \mathbb{C} : |\lambda - 5| \leq 1\}$$

- **Disc 2 (Row 2):**

  Center: $a_{22} = 4$, Radius: $R_2 = |a_{21}| + |a_{23}| = |2| + |1| = 3$

  $$D_2 = \{\lambda \in \mathbb{C} : |\lambda - 4| \leq 3\}$$

AIMS | African Institute for Mathematical Sciences SENEGAL

- **Disc 3 (Row 3):**

  Center: $a_{33} = 6$, Radius: $R_3 = |a_{31}| + |a_{32}| = |1| + |2| = 3$

  $$D_3 = \{\lambda \in \mathbb{C} : |\lambda - 6| \leq 3\}$$

- **Conclusion:** All eigenvalues of A must lie within $D_1 \cup D_2 \cup D_3$.

- For real eigenvalues, this corresponds to the interval $[1, 9]$.

Exercise

1. Prove or explain why the geometric multiplicity of an eigenvalue is always less than or equal to its algebraic multiplicity.

2. Give an example of a matrix where the algebraic multiplicity is greater than the geometric multiplicity.

3. Can a matrix have an eigenvalue with algebraic multiplicity 2 and geometric multiplicity 1? What does this imply about the matrix?

4. Is it possible for a matrix to have an eigenvalue whose geometric multiplicity is greater than its algebraic multiplicity? Why or why not?

5. Let $A \in \mathbb{R}^{n \times n}$. If an eigenvalue λ has algebraic multiplicity n, what does this imply about the characteristic polynomial?

6. If A is diagonalizable, how are the algebraic and geometric multiplicities of its eigenvalues related?

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

AIMS | African Institute for Mathematical Sciences SENEGAL

Overview

Cholesky decomposition

- **Any symmetric, positive definite matrix** A can be uniquely factorized as $A = LL^T$, where L is a **lower triangular matrix with positive diagonal elements:**

$$
\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} l_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ l_{n1} & \cdots & l_{nn} \end{bmatrix} \begin{bmatrix} l_{11} & \cdots & l_{n1} \\ 0 & \ddots & \vdots \\ 0 & \cdots & l_{nn} \end{bmatrix} \tag{8.1}
$$

- The lower triangular matrix L is called the Cholesky factor of A.

AIMS | African Institute for Mathematical Sciences SENEGAL

Cholesky Decomposition: Example

- **Example:** Consider the symmetric, positive-definite matrix A:

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

- Let $L = \begin{bmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{bmatrix}$. Then:

$$LL^T = \begin{bmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} \\ 0 & l_{22} \end{bmatrix} = \begin{bmatrix} l_{11}^2 & l_{11}l_{21} \\ l_{21}l_{11} & l_{21}^2 + l_{22}^2 \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 2 & 5 \end{bmatrix}$$

- **Solving for L:**

1. From $a_{11} = l_{11}^2$:

$$l_{11}^2 = 4 \implies l_{11} = 2 \quad (\text{since } l_{11} > 0)$$

2. From $a_{12} = l_{11}l_{21}$:

$$2l_{21} = 2 \implies l_{21} = 1$$

3. From $a_{22} = l_{21}^2 + l_{22}^2$:

$$l_{21}^2 + l_{22}^2 = 5 \implies 1^2 + l_{22}^2 = 5 \implies l_{22}^2 = 4 \implies l_{22} = 2 \quad (\text{since } l_{22} > 0)$$

Thus:

$$L = \begin{bmatrix} 2 & 0 \\ 1 & 2 \end{bmatrix}$$

Overview

Eigendecomposition/Diagonalization

- Consider a matrix A possessing n eigenvalues $\lambda_1, \ldots, \lambda_n$ and n linearly independent eigenvectors $x_1, \ldots, x_n$. We can construct:
  - Diagonal matrix: $D = \text{diag}(\lambda_1, \ldots, \lambda_n)$
  - Matrix of eigenvectors: $P = [x_1 \ x_2 \ \cdots \ x_n]$

  This allows us to represent A as:

$$
A = PDP^{-1} = P \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} P^{-1} \tag{8.2}
$$

- A matrix that allows such a decomposition is termed "diagonalizable". **A matrix is diagonalizable if and only if it can**

Example

- **All symmetric matrices are diagonalizable.** The spectral theorem guarantees the existence of a matrix P such that:

$$A = PDP^{-1} \tag{8.3}$$

- Consider the matrix A:

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

- We aim to find the eigendecomposition $A = PDP^{-1}$, where P has eigenvectors as columns and D is the diagonal matrix of corresponding eigenvalues.

- **Step 1: Find the Eigenvalues**. Solve $\det(A - \lambda I) = 0$:

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

$$= \lambda^2 - 6\lambda + 8 = (\lambda - 4)(\lambda$$

- The eigenvalues are: $\sigma(A) = \{\lambda_1 = 4, \lambda_2 = 2\}$.
- **Step 2: Find the Eigenvectors**

  For $\lambda_1 = 4$:

  Solve $(A - 4I) \begin{bmatrix} x \\ y \end{bmatrix} = 0 \Longleftrightarrow \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0$. This yields:

$$-x + y = 0 \implies y = x. \text{ Thus: } v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

AIMS African Institute for Mathematical Sciences SENEGAL

- For $\lambda_2 = 2$:

  Solve $(A - 2I) \begin{bmatrix} x \\ y \end{bmatrix} = 0 \iff \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0$. This gives:

  $$x + y = 0 \implies y = -x. \text{ Thus: } v_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

- **Step 3: Form D and P:**

  $$P = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad D = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}$$

Conditions for Diagonalization

- Not all matrices are symmetric. We must therefore establish conditions for a matrix to be diagonalizable. Let A be an $n \times n$ matrix.

- **Condition 1:** A is diagonalizable if there exists a basis of $\mathbb{R}^n$ composed of eigenvectors of A, i.e., A **possesses n linearly independent eigenvectors.**

- **Condition 2:** If A **has n distinct eigenvalues**, then it is diagonalizable (since eigenvectors corresponding to distinct eigenvalues are linearly independent).

- **Condition 3:** A is diagonalizable if A **satisfies $GM = AM$.**

AIMS | African Institute for Mathematical Sciences SENEGAL

Exercises: Diagonalizability

- Determine which of the following matrices are diagonalizable over $\mathbb{R}$. Provide justification for your answer.

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

b) $B = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$

c) $C = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$

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

e) $E = \begin{bmatrix} 4 & 1 & 1 \\ 1 & 4 & 1 \\ 1 & 1 & 4 \end{bmatrix}$

f) $F = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$

Overview

SVD Theorem

- Let $A \in \mathbb{R}^{m \times n}$ be a matrix of rank $r \leq \min(m, n)$. The Singular Value Decomposition (SVD) of A expresses it as:

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

where:

  - $U \in \mathbb{R}^{m \times m}$ is an orthogonal matrix with columns u_i (left singular vectors)
  - $V \in \mathbb{R}^{n \times n}$ is an orthogonal matrix with columns v_j (right singular vectors)
  - $\Sigma \in \mathbb{R}^{m \times n}$ has diagonal entries $\sigma_i \geq 0$ and zeros elsewhere

AIMS | African Institute for Mathematical Sciences SENEGAL

Properties of SVD

- The diagonal entries $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$ are called **singular values** of A.
- The columns u_i are **left singular vectors**, and columns v_j are **right singular vectors**.
- The singular values are conventionally ordered in decreasing order.
- The matrix Σ is uniquely determined. For $m > n$:

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \end{bmatrix} \tag{8.4}$$

- For $m < n$, Σ has the structure:

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_m & 0 & \cdots & 0 \end{bmatrix} \tag{8.5}$$

- **Key fact:** The SVD exists for *any* matrix $A \in \mathbb{R}^{m \times n}$, regardless of rank or shape.

Computing the SVD

- The SVD can be constructed using the eigendecomposition of $A^T A$:

$$A^T A = Q \Lambda Q^T = Q \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} Q^T \qquad (8.6)$$

where Q contains orthonormal eigenvectors and $\lambda_i \geq 0$ are eigenvalues of $A^T A$.

- Starting from the SVD form $A = U\Sigma V^T$, we compute:

$$A^T A = (U\Sigma V^T)^T (U\Sigma V^T) = V\Sigma^T U^T U\Sigma V^T \qquad (8.7)$$

- Since U is orthogonal ($U^T U = I$), we get:

$$A^T A = V \Sigma^T \Sigma V^T = V \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{bmatrix} V^T \qquad (8.8)$$

- This shows that the singular values are $\sigma_i = \sqrt{\lambda_i}$, where λ_i are eigenvalues of $A^T A$.

References

📖 Deisenroth, Marc Peter and Faisal, A Aldo and Ong, Cheng Soon.
Mathematics for machine learning.
Cambridge University Press, 2020 .

📖 Coursera: David Dye et al.
Mathematics for machine learning: Linear Algebra.
Imperial College London, 2025 .

📖 Rachel Thomas
Computational Linear Algebra.
FastAI, 2017 .

📄 B. Author.
On this and that.
Journal on This and That. 2(1):50–100, 2000.