Indian Statistical Institute

Bachelor of Statistical Data Science

Mathematics - III

Course Instructor: Prof. Depdip Ganguly

Fall 2026

Preface

These notes are designed as a comprehensive study guide for the *Mathematics - III* course, instructed by **Prof. Debdip Ganguly** at the Indian Statistical Institute, Delhi Center during the Fall 2026 semester. However, please keep in mind that they are most effective when paired with regular attendance and active listening during the actual lectures.

While I have made every effort to ensure this document is completely accurate, it is entirely possible that a few typos or conceptual mistakes remain. I would be incredibly grateful if you could point out any errors you spot while studying. Feel free to contact me if you find any! ;)

— *Utkarsh Bharadwaj*

Contents

PART I

LINEAR ALGEBRA

1 Linear Transformations

Linear Transformation

Let V and W be two vector spaces over a field $\mathbb{F}$. Thus, a linear transformation $T : V \rightarrow W$ is defined as

1. $T(v_1 + v_2) = T(v_1) + T(v_2) \quad \forall v_1, v_2 \in V$
2. $T(\alpha v) = \alpha T(v) \quad \forall \alpha \in \mathbb{F}$

Important

1. Both V and W must come for the same field $\mathbb{F}$
2. A linear transform takes $0 \in V \rightarrow 0 \in W$

Example: Some examples of a vector space over $\mathbb{R}^n$

1. $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ such that $T(x_1, x_2) = (x_1, -x_2)$

   `# Using definition to prove it`

   Proof. Let $u = (x_1, x_2)$ and $v = (y_1, y_2)$ be in $\mathbb{R}^2$, and let $\alpha \in \mathbb{R}$.
   1. Additivity:

$$\begin{aligned} T(u+v) &= T(x_1 + y_1, x_2 + y_2) \\ &= (x_1 + y_1, -(x_2 + y_2)) \\ &= (x_1, -x_2) + (y_1, -y_2) \\ &= T(u) + T(v) \end{aligned}$$

   2. Scalar Multiplication:

$$\begin{aligned} T(\alpha u) &= T(\alpha x_1, \alpha x_2) \\ &= (\alpha x_1, -\alpha x_2) \\ &= \alpha(x_1, -x_2) \\ &= \alpha T(u) \end{aligned}$$

   Since both properties hold, T is a linear transformation. $\square$

2. $T : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ such that $T(x_1, x_2, x_3) = (x_1, x_2, 0)$. Its a projection over the $x - y$ plane
3. $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ such that $T(x_1, x_2) = (x_1 + 1, x_2)$. This is NOT a linear transform as $T(0, 0) \neq (0, 0)$
4. $T : \mathbb{M}_2(\mathbb{C}) \rightarrow \mathbb{C}$ such that $T(A) = tr(A)$

Matrix representation

Let V b a vector space of dimension k such that $\beta = \{v_1, v_2, \ldots, v_k\}$ are the basis vectors of V. Thus, any $v \in V$ can be written as $v = \sum_{i=1}^{k} \alpha_i v_i \ \forall \alpha_i \in \mathbb{F}$. Alternatively, we can also write

$$v = \underbrace{\begin{bmatrix} v_1 & v_2 & \cdots & v_k \end{bmatrix}}_{B} \underbrace{\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{bmatrix}}_{[v]_\beta}$$

Thus, we have

$$v = B[v]_\beta \iff [v]_\beta = B^{-1}v$$

> **Note**
>
> B is a non-singular $k \times k$ matrix

Matrix Form of a Linear Transformation

Consider $T : V \to W$ such that V and W are n and m dimensions respectively. Let

$$\beta = \{v_1, v_2, \ldots, v_n\} \text{ be basis of } V$$
$$\gamma = \{w_1, w_2, \ldots, w_n\} \text{ be basis of } W$$

Thus, $\forall v \in V$, we have $v = \sum_{i=1}^{n} \alpha_i v_i$ and also $T(v_i) \in W \ \forall 1 \leq i \leq n$. Thus, we have the following

$$\begin{aligned} T(v) &= T(\alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_n v_n) \\ &= \alpha_1 T(v_1) + \alpha_2 T(v_2) + \ldots + \alpha_n T(V_n) \\ &= \underbrace{\begin{bmatrix} T(v_1) & T(v_2) & \ldots & T(v_n) \end{bmatrix}}_{\text{Matrix of order } m \times n} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} \end{aligned}$$

Thus, we have the following

$$T(v) = [T]_\beta^\gamma [v]_\beta$$

1.1 Some exercises on Linear Transformation and Matrix Notation

Example: 1

Consider the following linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ such that $T(x_1, x_2) = (x_1, -x_2)$. Formulate the transformation matrix

Solution

First, we need to fix a basis for V. Here V is $\mathbb{R}^2$. Thus, consider the standard basis:

$$e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Now, we apply the linear transform on these basis vectors. Thus, we have:

$$T(e_1) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad T(e_2) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

The transformation matrix A is formed by taking these vectors as its columns. Therefore, the transformation matrix is:

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

Example: 2

Let $T : P_2(\mathbb{R}) \to P_1(\mathbb{R})$ such that $T(p) = D(p)$ where $P_n(\mathbb{R})$ is the space of all polynomials of degree *atmost* n and D is the derivation of the polynomial. Formulate the transformation matrix

Solution

First, we fix the standard bases for both the domain $P_2(\mathbb{R})$ and the codomain $P_1(\mathbb{R})$. Let the standard basis for $P_2(\mathbb{R})$ be $B_1 = \{1, x, x^2\}$ and the standard basis for $P_1(\mathbb{R})$ be $B_2 = \{1, x\}$.
Next, we apply the linear transformation T (differentiation) to each basis vector in B_1:

$$\begin{aligned} T(1) &= \frac{d}{dx}(1) = 0 \\ T(x) &= \frac{d}{dx}(x) = 1 \\ T(x^2) &= \frac{d}{dx}(x^2) = 2x \end{aligned}$$

Now, we express these results as coordinate column vectors with respect to the codomain basis $B_2 = \{1, x\}$:

$$[T(1)]_{B_2} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \qquad [T(x)]_{B_2} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad [T(x^2)]_{B_2} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$$

Finally, we construct the transformation matrix by taking these coordinate vectors as its columns. Thus, the standard transformation matrix A is:

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

Example: 3

Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the reflection with respect to the straight line $y = 2x$. Find the analytic expression for T.

Solution

To find the transformation matrix for this reflection, we can use a basis composed of vectors whose behavior under the reflection is easy to determine.
First, we identify a direction vector on the line $y = 2x$. If we let $x = 1$, then $y = 2$. Thus, the vector $\vec{v}_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ lies on the line. Since points on the mirror line do not move during a reflection, $T(\vec{v}_1) = \vec{v}_1$.
Next, we identify a vector orthogonal (perpendicular) to the line. The slope of our line is 2, so the perpendicular slope is $-1/2$. A vector pointing in this direction is $\vec{v}_2 = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$. Under reflection, a vector completely orthogonal to the mirror line is flipped to its exact negative. Thus, $T(\vec{v}_2) = -\vec{v}_2 = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$.

Now, we express the standard basis vectors $e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ as linear combinations of $\vec{v}_1$ and $\vec{v}_2$:

$$e_1 = \frac{1}{5}\vec{v}_1 - \frac{2}{5}\vec{v}_2$$
$$e_2 = \frac{2}{5}\vec{v}_1 + \frac{1}{5}\vec{v}_2$$

We apply the linear transformation T to the standard basis vectors using linearity:

$$T(e_1) = \frac{1}{5}T(\vec{v}_1) - \frac{2}{5}T(\vec{v}_2) = \frac{1}{5}\begin{pmatrix} 1 \\ 2 \end{pmatrix} - \frac{2}{5}\begin{pmatrix} 2 \\ -1 \end{pmatrix} = \begin{pmatrix} -3/5 \\ 4/5 \end{pmatrix}$$
$$T(e_2) = \frac{2}{5}T(\vec{v}_1) + \frac{1}{5}T(\vec{v}_2) = \frac{2}{5}\begin{pmatrix} 1 \\ 2 \end{pmatrix} + \frac{1}{5}\begin{pmatrix} 2 \\ -1 \end{pmatrix} = \begin{pmatrix} 4/5 \\ 3/5 \end{pmatrix}$$

The transformation matrix A is formed by taking these transformed standard basis vectors as its columns:

$$A = \begin{pmatrix} -3/5 & 4/5 \\ 4/5 & 3/5 \end{pmatrix} = \frac{1}{5}\begin{pmatrix} -3 & 4 \\ 4 & 3 \end{pmatrix}$$

Finally, the analytic expression for T applied to an arbitrary vector $\begin{pmatrix} x \\ y \end{pmatrix}$ is given by multiplying the matrix A by the vector:

$$T(x,y) = \left(-\frac{3}{5}x + \frac{4}{5}y, \frac{4}{5}x + \frac{3}{5}y\right)$$

Example: 4

Let $T : P_2(\mathbb{R}) \to P_3(\mathbb{R})$ defined by

$$T(f)(x) = 2f'(x) + \int_0^x 3f(t)dt$$

Find the Null space of T.

Solution

First, we establish the standard bases for our domain and codomain. The domain is $P_2(\mathbb{R})$ with standard basis $B = \{1, x, x^2\}$. Note that $\dim(P_2(\mathbb{R})) = 3$. The codomain is $P_3(\mathbb{R})$ with standard basis $C = \{1, x, x^2, x^3\}$.
Next, we find the standard transformation matrix A by applying T to each basis vector in B:

$$\begin{aligned}
T(1) &= 2\frac{d}{dx}(1) + \int_0^x 3(1)dt = 0 + [3t]_0^x = 3x \\
T(x) &= 2\frac{d}{dx}(x) + \int_0^x 3tdt = 2(1) + \left[\frac{3}{2}t^2\right]_0^x = 2 + \frac{3}{2}x^2 \\
T(x^2) &= 2\frac{d}{dx}(x^2) + \int_0^x 3t^2dt = 2(2x) + [t^3]_0^x = 4x + x^3
\end{aligned}$$

Writing these results as coordinate vectors relative to $C = \{1, x, x^2, x^3\}$, we form the columns of matrix A:

$$A = \begin{pmatrix} 0 & 2 & 0 \\ 3 & 0 & 4 \\ 0 & 3/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

To use the Rank-Nullity theorem, we must first find the rank of T, which is the rank of matrix A. We can determine this by checking the linear independence of the column vectors. Row reducing matrix A yields:

$$\text{RREF}(A) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

Since there are 3 pivot columns, the columns are linearly independent, meaning the rank of the transformation is $\text{Rank}(T) = 3$.
Alternatively, we can also find the range of T as explicitly writing out the span of the transformed basis vectors:

$$\text{Range}(T) = \text{span}\left\{3x, 2 + \frac{3}{2}x^2, 4x + x^3\right\}$$

Because these polynomials are linearly independent, the dimension of this range is 3, which confirms that $\text{Rank}(T) = 3$.
Now, we apply the Rank-Nullity Theorem, which states:

$$\dim(\text{Domain}) = \text{Rank}(T) + \text{Nullity}(T)$$

Substituting our known values:

$$3 = 3 + \text{Nullity}(T) \implies \text{Nullity}(T) = 0$$

The nullity is the dimension of the null space. Since the dimension is 0, the null space must consist only of the zero vector (the zero polynomial).
Therefore, the Null space of T is:

$$N(T) = \{0\}$$

Example: 5

Let $T : P_3(\mathbb{R}) \to P_2(\mathbb{R})$ such that $T(f(x)) = f'(x)$. Formulate the transformation matrix.

`# Solution`

Let the standard basis for $P_3(\mathbb{R})$ be $B_1 = \{1, x, x^2, x^3\}$ and for $P_2(\mathbb{R})$ be $B_2 = \{1, x, x^2\}$.
Next, we apply the linear transformation T (differentiation) to each basis vector in B_1:

$$\begin{aligned} T(1) &= 0 = 0 \cdot 1 + 0 \cdot x + 0 \cdot x^2 \\ T(x) &= 1 = 1 \cdot 1 + 0 \cdot x + 0 \cdot x^2 \\ T(x^2) &= 2x = 0 \cdot 1 + 2 \cdot x + 0 \cdot x^2 \\ T(x^3) &= 3x^2 = 0 \cdot 1 + 0 \cdot x + 3 \cdot x^2 \end{aligned}$$

Constructing the transformation matrix A by taking these coordinate vectors as its columns relative to B_2:

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$

2 Similar Matrix

Similar Matrix

For any $A, B \in \mathbb{M}_n(\mathbb{C})$, A and B are called similar matrices if there exists a non-singular matrix P such that

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

This is denoted as $A \sim B$

Note

1. $A \sim A$: This is trivially true
2. $A \sim B \equiv B \sim A$
3. If $A \sim B$ and $B \sim C$, this implies $A \sim C$

```
# Proof of Transitivity
```

Proof. Let $A \sim B$ and $B \sim C$. By definition, there exist non-singular matrices P and Q such that:

$$A = PBP^{-1} \quad \text{and} \quad B = QCQ^{-1}$$

Substituting B into the equation for A:

$$\begin{aligned} A &= P(QCQ^{-1})P^{-1} \\ &= (PQ)C(Q^{-1}P^{-1}) \\ &= (PQ)C(PQ)^{-1} \end{aligned}$$

Let $L = PQ$. Since the product of non-singular matrices is non-singular, L is non-singular. Thus, $A = LCL^{-1}$, which means $A \sim C$. $\square$

Important

$(\mathbb{M}_n(\mathbb{C}), \sim)$ is an *equivalent relation* on $\mathbb{M}_n(\mathbb{C})$

3 Vector Space

Definition of a vector space

A **vector space** over a field $\mathbb{F}$ (such as the real numbers $\mathbb{R}$ or complex numbers $\mathbb{C}$) is a non-empty set V of objects, called *vectors*, equipped with two operations:

- **Vector Addition:** A rule that assigns to each pair of vectors $\mathbf{u}, \mathbf{v} \in V$ a unique vector $\mathbf{u} + \mathbf{v} \in V$.
- **Scalar Multiplication:** A rule that assigns to each scalar $c \in \mathbb{F}$ and each vector $\mathbf{v} \in V$ a unique vector $c\mathbf{v} \in V$.

Example: Vector spaces

1. $\mathbb{R}^n, n \geq 1$ and $\mathbb{C}^n, n \geq 1$ are the most intuitive and basic examples of a vector space
2. Let V be a finite dimensional vector space of dimension n. Then forall $\mathbf{v} \in V$, we can write $\mathbf{v} = \sum_{i=1}^{n} \alpha_i v_i$ such that $\mathcal{B} = \{v_1, v_2, \ldots v_n\}$ are the basis vectors over field $\mathbb{F}_n$

Example: Space of matrices

Let $\mathbb{M}_{m\times n}(\mathbb{F})$ be the set of all matrices of order $m \times n$ over the field $\mathbb{F}$. For $A = (a_{ij})_{m\times n}$ and $B = (b_{ij})_{m\times n}$ consider the following

$$(A \oplus B)_{ij} := a_{ij} + b_{ij}$$

$$(c \cdot A)_{ij} := ca_{ij}$$

This implies that $(M_{m\times n}, \oplus, \cdot)$ defines a vector space over a field F

Example: Functions Mapping

Let S be any arbitrary set. Define $\mathcal{F}^S$ as the set of all mappings from S to some F. Thus, we have

$$\mathcal{F}^S := \{f : S \to \mathbb{F}\}$$

Thus, for any $f, g \in \mathcal{F}^S$ and $s \in S$, we have the following

$$(f \oplus g)(s) := f(s) + g(s)$$

$$(cf)(s) := cf(s) \quad c \in \mathbb{F}$$

In particular, if $S = \mathbb{N}$ and $\mathbb{F} = \mathbb{R}$, then $\mathcal{F}^\infty := F^{\mathbb{N}}$ is the space of all real sequences. Thus,

$$\mathcal{F}^\infty = \{\{x_n\} : x_n \in \mathbb{R}\}$$

The standard basis vectors for this space are:

$$\begin{aligned} e_1 &= \{1, 0, 0, 0, \dots\} \\ e_2 &= \{0, 1, 0, 0, \dots\} \\ &\vdots \\ e_i &= \{0, 0, \dots, 1, \dots\} \quad \text{(1 at the } i\text{-th position)} \end{aligned}$$

Example: l^2 space

Consider the set

$$l^2 = \left\{ \{x_n\} : \sum |x_n|^2 < \infty \right\}$$

This form a vector space. The proof of this is left as an exercise to the reader
Similarly, consider the following set

$$l^p = \left\{ \{x_n\} : \sum |x_n|^p < \infty \right\}$$

This can be proved using a basic property of a convex function. We'll arrive at the following results

$$|x_n + y_n|^p \leq 2^{p-1} \left(|x_n|^p + |y_n|^p\right)$$

Readers are adviced to go through the quick proof themselves. To brush up your concepts of convex optimization, you can click here.

Example: Function Space

Consider the set

$$V = \{f : \mathbb{R} \to \mathbb{R}, f \text{ is continuous}\}$$

This is also a vector space. For example, a linear transformation $T : V \to V$ on this space can be defined by:

$$T(f')(x) = \int_0^x f(t)dt$$

4 Real Inner Product Space

Geometric Motivation

In $\mathbb{R}^2$ and $\mathbb{R}^3$, we can measure the length of a vector x geometrically:

$$|x| = (x_1^2 + x_2^2)^{1/2} \quad \text{or} \quad |x| = (x_1^2 + x_2^2 + x_3^2)^{1/2}$$

The angle θ between two vectors x and y is given by:

$$\cos\theta = \frac{x \cdot y}{|x| \cdot |y|}$$

The abstract definition of the inner product generalizes these concepts of length and angle to broader vector spaces.

Definition of inner product

Let X be a real vector space. An *Inner Product* on X is a function $< \cdot, \cdot >$ from $X \times X \to \mathbb{R}$ such hat it satisfies the following properties

1. $\langle x, x \rangle \geq 0$ and $\langle x, x \rangle = 0 \iff x = 0$
2. $\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$
3. $\langle x, y \rangle = \langle y, x \rangle$

A vector space $(X, \langle \cdot, \cdot \rangle)$ with all elements following the above properties is called an *Inner Product Space*

4.1 Some examples of Inner Product Space

Example: $\mathbb{R}^n$

Consider $x, y \in \mathbb{R}^n$ and define $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$. This forms an inner product space

Proof

We have defined an inner product on this space. Now, we'll verify that the given inner product satisfies those 3 properties mentioned earlier. This will prove that R^n with the given inner product is an inner product space

Proof. 1. $\langle x, x \rangle = \sum x_i^2 \geq 0$ and $\sum x_i^2 = 0 \iff x_i = 0 \ \forall\ i$: This is trivially true

2. For $x, y, z \in V$ and $\alpha, \beta \in \mathbb{R}$, consider the following

$$\begin{aligned} \langle \alpha x + \beta y, z \rangle &= \sum_{i=1}^n (\alpha x_i + \beta y_i) z_i \\ &= \alpha \sum x_i z_i + \beta \sum y_i z_i \\ &= \alpha \langle x, z \rangle + \beta \langle y, z \rangle \end{aligned}$$

3. Consider the following

$$\langle x, y \rangle = \sum x_i y_i = \sum y_i x_i = \langle y, x \rangle$$

$\square$

Thus, this defines an inner product space

Example: Finite dimensional vector space

Let V be any finite dimensional vector space with dimension n. Then forall $v, w \in V$ and $\alpha, \beta \in \mathbb{R}^n$ we can write

$$v = \sum_{i=1}^{n} \alpha_i v_i \qquad w = \sum_{i=1}^{n} \beta_i v_i$$

such that $\mathcal{B} = \{v_1, v_2, \ldots, v_n\}$ are the basis vectors for V. Further define

$$\langle v, w \rangle := \sum_{i=1}^{n} \alpha_i \beta_i$$

This defines an inner product space. The proof is trivial

Important

Every vector space with a *finite dimension* forms an inner product space

Cauchy-Schwartz Inequality

Consider a real inner product space V. Consider $x, y \in V$. Thus, *Cauchy-Schwartz Inequality* states that

$$|\langle x, y \rangle| \leq (\langle x, x \rangle)^{1/2} (\langle y, y \rangle)^{1/2}$$

5 Complex Inner Product Space

Definition

Let X be a complex vector space. An inner product on X is a function $\langle \cdot, \cdot \rangle : X \times X \to \mathbb{C}$ such that it satisfies the following properties

1. $\langle x, x \rangle \in \mathbb{R}$, $\langle x, x \rangle \geq 0$
2. $\langle x, x \rangle = 0 \iff x = 0$

5.1 Some examples of complex inner product space

We'll see a couple of examples of a complex inner product space here

Example: $\mathbb{C}^n$

Consider $x, y \in \mathbb{C}^n$. Define the complex inner product as

$$\langle x, y \rangle := \sum_{i=1}^{n} x_i \bar{y}_i \in \mathbb{C}$$

This defines a complex inner product space

Proof

Consider the following

$$\langle x, x\rangle = \sum_{i=1}^{n} x_i \bar{x}_i = \sum_{i=1}^{n} |x_i|^2$$

Thus, $\langle x, x\rangle$ clearly belongs to $\mathbb{R}$. Also, we have

$$\sum_{i=1}^{n} |x_i|^2 \geq 0 \; \forall x_i \in \mathbb{C}$$

and

$$\sum_{i=1}^{n} |x_i|^2 = 0 \iff x_i = 0 \; \forall \; i$$

Example: Finite dimensional complex vector space

Consider a finite dimensional complex vector V such that $\mathcal{B} = \{v_1, \ldots, v_n\}$ form the basis vector for V where it is obvious that $v_i \in \mathbb{C} \; \forall \; i$. Thus, we can write for any $v, w \in V$

$$v = \sum \alpha_i v_i \qquad w = \sum \beta_i v_i$$

such that $\alpha_i, \beta_i \in \mathbb{C}$. Define the following

$$\langle v, w\rangle = \sum_{i=1}^{n} \alpha_i \bar{\beta}_i$$

This defines a complex vector space. Proof is left to the reader as an exercise

Example: l^2

Consider the following set

$$l^2 = \left\{ \{x_n\} : \sum |x_n|^2 < \infty \right\}$$

Define the complex inner product as

$$\langle \{x_n\}, \{y_n\}\rangle := \sum_{i=1}^{n} x_i \bar{y}_i$$

This defines a complex inner product space

Important

We *cannot* define a complex inner product space on the l^p for $p \neq 2$ which is defined as

$$l^p = \left\{ \{x_n\} : \sum |x_n|^p < \infty \right\}$$

Lemma

Let X be an inner product space. Consider $x, y, z \in X$ and $\alpha, \beta \in \mathbb{C}$. Then, we have the following properties

1. $\langle \vec{0}, y\rangle = \langle x, \vec{0}\rangle = 0$
2. $\langle x, \alpha y + \beta z\rangle = \bar{\alpha}\langle x, y\rangle + \bar{\beta}\langle x, z\rangle$
3. $\langle \alpha x + \beta y, \alpha x + \beta y\rangle = |\alpha|^2\langle x, x\rangle + \alpha\bar{\beta}\langle x, y\rangle + \beta\bar{\alpha}\langle y, x\rangle + |\beta|^2\langle y, y\rangle$

Proof of Property 2

Proof. Using the conjugate symmetry and linearity in the first argument:

$$\begin{aligned}
\langle x, \alpha y + \beta z\rangle &= \overline{\langle \alpha y + \beta z, x\rangle} \\
&= \overline{\alpha\langle y, x\rangle + \beta\langle z, x\rangle} \\
&= \bar{\alpha}\overline{\langle y, x\rangle} + \bar{\beta}\overline{\langle z, x\rangle} \\
&= \bar{\alpha}\langle x, y\rangle + \bar{\beta}\langle x, z\rangle
\end{aligned}$$

□

Note

$\langle x, y\rangle = \overline{\langle y, x\rangle} \in \mathbb{C}$

6 Gram Matrix

Definition

Let X be a real inner product space of finite dimension such that $\mathcal{B} = \{v_1, v_2, \ldots, v_n\}$ form the basis vector of X. Consider $x, y \in X$. Thus, we can write

$$x = \beta_1 v_1 + \ldots + \beta_n v_n$$

$$y = \alpha_1 v_1 + \ldots + \alpha_n v_n$$

Thus, we have the following inner product

$$\langle x, y\rangle = \langle \beta_1 v_1 + \ldots + \beta_n v_n, \alpha_1 v_1 + \ldots + \alpha_n v_n\rangle$$

$$= \alpha_1\langle v_1, v_1\rangle\beta_1 + \beta_1\langle v_2, v_1\rangle\alpha_2 + \ldots$$

This can also be written as

$$\langle x, y\rangle = \begin{bmatrix}\beta_1 & \beta_2 & \dots & \beta_n\end{bmatrix} \underbrace{\begin{bmatrix}\langle v_1, v_1\rangle & \langle v_1, v_2\rangle & \dots & \langle v_1, v_n\rangle \\ \langle v_2, v_1\rangle & \langle v_2, v_2\rangle & \dots & \dots \\ \vdots & & & \\ \langle v_n, v_1\rangle & \dots & \dots & \langle v_n, v_n\rangle\end{bmatrix}}_{\text{Symmetric Matrix}} \begin{bmatrix}\alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n\end{bmatrix}$$

Thus, we have the following

$$\langle x, y\rangle = x_{\mathcal{B}}^T G y_{\beta}$$

where G is the *Gram Matrix* with respect to $\mathcal{B}$

7 Symmetric Matrix

Definition and basic properties

A matrix $A = (a_{ij})_{n\times n}$ is called a *Symmetric Matrix* if $A^T = A$, i.e., $a_{ij} = a_{ji}$. It also has the following two basic properties among others

1. All the *Eigenvalues* of A are in $\mathbb{R}$. Thus $\sigma(A) \subseteq \mathbb{R}$
2. A is diagonalizable, i.e., there exists a non-singular matrix P such that $A = PDP^T$

7.1 Orthogonal Diagonalizable

Note

A real symmetric matrix of order n is said to be positive definite if forall non-zero vector $x \in \mathbb{R}^n$

$$x^T Ax > 0$$

Also, recall that for $x, y \in \mathbb{R}^n$

$$\langle x, y\rangle = \sum_{i=1}^{n} x_i y_i = y^T x$$

Proposition

Let A be real matrix of order n. The, the following are equivalent

1. The expression $\langle x, y\rangle = y^T Ax$ defines an inner product on $\mathbb{R}^n$
2. A is a real positive definite matrix

Proof

Proof. $\langle x, x\rangle = y^T Ax = \langle y, x\rangle = x^T Ay$ Since $\langle x, y\rangle$ defines an inner product space, we can write the following

$$0 \leq \langle x, x\rangle = x^T Ax$$

This implies $x^T Ax \geq 0$. Also since $\langle x, x\rangle = 0 \iff x = 0$, this implies that $x^T Ax = 0 \iff x = 0$. Hence, we have the following result that for any non-zero vector $x \in \mathbb{R}^n$

$$x^T Ax > 0$$

□

Definition of Orthogonal Matrix

A matrix $S \in \mathbb{M}_n(\mathbb{R})$ is said to be orthogonal matrix if $SA^t = I_n$ where I_n is the identity matrix of order n

Proposition

Let S be a matrix in $\mathbb{M}_n(\mathbb{R})$. Then, the following are equivalent

1. S is an orthogonal matrix
2. $SS^T = I_n$
3. $S^T S = I_n$
4. The column of S are an orthogonal basis for $\mathbb{R}^n$
5. The Row of S are an orthogonal basis for $\mathbb{R}^n$

Proof of Equivalence (3) and (4)

Proof. Let $S = \begin{bmatrix} c_1 & c_2 & \dots & c_n \end{bmatrix}$ where c_i represents the i-th column vector of S. Consider the product $S^T S$:

$$S^T S = \begin{bmatrix} c_1^T \\ c_2^T \\ \vdots \\ c_n^T \end{bmatrix} \begin{bmatrix} c_1 & c_2 & \dots & c_n \end{bmatrix}$$

The element in the i-th row and j-th column of $S^T S$ is given by $c_i^T c_j = \langle c_i, c_j\rangle$. Thus, $S^T S = I_n$ if and only if:

$$\langle c_i, c_j\rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

This means $S^T S = I_n$ is true if and only if the columns of S form an orthogonal basis for $\mathbb{R}^n$. □

Definition of orthogonally diagonalizable matrix

A real matrix is said to be orthogonally diagonalizable if there exists a diagonal matrix D and an orthogonal matrix S such that

$$D = S^T AS$$

Equivalently, we can also write

$$A = SDS^T$$

It is easy to see that A is a symmetric matrix

Lemma

Let A be a real symmetric matrix and let x_1, x_2 be eigen-vectors of A associated with distinct eigenvalues λ_1, λ_2. Then, $\langle x_1, x_2\rangle = 0$

Proof

Proof. Consider the following

$$\begin{aligned}\langle x_1, Ax_2\rangle = (Ax_2)^T x_1 = x_2^T A^T x_1 = x_2^T Ax_1 \\ = \lambda_1 x_2^T x_1 = \lambda_1\langle x_1, x_2\rangle\end{aligned}$$

Also, the same expression can be written in the following way

$$\langle x_1, Ax_2\rangle = \langle x_1, \lambda_2 x_2\rangle = \lambda_2\langle x_1, x_2\rangle$$

Thus, we have $\lambda_1\langle x_1, x_2\rangle = \lambda_2\langle x_1, x_2\rangle$. This implies

$$(\lambda_1 - \lambda_2)\langle x_1, x_2\rangle = 0$$

Thus, $\langle x_1, x_2\rangle = 0$ since $\lambda_1 \neq \lambda_2$ □

Theorem

A real square matrix is orthogonally diagonalizable iff it is symmetric

Proof

Proof. If A is orthogonally diagonalizable then A is symmetric. This has already been shown in a previous result. Thus, we'll only prove the "if" part here.
Let A be a symmetric matrix. To prove A is orthogonally diagonalizable we'll use

the method of induction. The result is trivially true for $n = 1$, since any 1×1 matrix is diagonal.
Suppose A is an $n \times n$ matrix with $n \geq 2$ and that the assertion holds for all square symmetric matrices of order $(n-1)$.
Let λ be a real eigenvalue of A with eigenvector x such that $||x|| = 1$. (Every real symmetric matrix has at least one real eigenvalue). Consider the following basis for $\mathbb{R}^n$:

$$\begin{aligned}\mathcal{B} &= \{x\} \cup (\text{span}\{x\})^{\perp} \\ &= \{x\} \cup \beta_1^{\perp}\end{aligned}$$

where $\beta_1^{\perp} = \{v_2, v_3, \ldots, v_n\}$ is an orthonormal basis for $(\text{span}\{x\})^{\perp}$, formally defined as:

$$\text{span}\{x\}^{\perp} = \{v \in \mathbb{R}^n : \langle x, v \rangle = 0\}$$

Because x is orthogonal to every vector in $\beta_1^{\perp}$, the set $\mathcal{B}$ forms an orthonormal basis for $\mathbb{R}^n$. Let P_1 be the $n \times n$ matrix whose columns are the vectors of $\mathcal{B}$:

$$P_1 = \begin{bmatrix} x & v_2 & \ldots & v_n \end{bmatrix}$$

Since the columns of P_1 are orthonormal, P_1 is an orthogonal matrix. Let us examine the matrix $P_1^T A P_1$.
The first column of $P_1^T A P_1$ is evaluated as follows:

$$P_1^T A x = P_1^T (\lambda x) = \lambda P_1^T x = \lambda \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Furthermore, since A is symmetric, the matrix $P_1^T A P_1$ is also symmetric:

$$(P_1^T A P_1)^T = P_1^T A^T (P_1^T)^T = P_1^T A P_1$$

Because $P_1^T A P_1$ is symmetric and its first column is zero everywhere except for the first entry, its first row must also be zero everywhere except for the first entry. Thus, $P_1^T A P_1$ takes the block matrix form:

$$P_1^T A P_1 = \begin{bmatrix} \lambda & \mathbf{0}^T \\ \mathbf{0} & A_1 \end{bmatrix}$$

where A_1 is an $(n-1) \times (n-1)$ symmetric matrix.
By our inductive hypothesis, since A_1 is an $(n-1) \times (n-1)$ symmetric matrix, it is orthogonally diagonalizable. Therefore, there exists an $(n-1) \times (n-1)$ orthogonal matrix Q and a diagonal matrix D_1 such that:

$$Q^T A_1 Q = D_1$$

We now construct an $n \times n$ block matrix P_2 defined by:

$$P_2 = \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & Q \end{bmatrix}$$

Since Q is orthogonal, P_2 is also an orthogonal matrix. Now, let $P = P_1 P_2$. The product of two orthogonal matrices is orthogonal, so P is an orthogonal matrix. Finally, we compute $P^T AP$:

$$\begin{aligned} P^T AP &= (P_1 P_2)^T A (P_1 P_2) \\ &= P_2^T (P_1^T A P_1) P_2 \\ &= \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & Q^T \end{bmatrix} \begin{bmatrix} \lambda & \mathbf{0}^T \\ \mathbf{0} & A_1 \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & Q \end{bmatrix} \\ &= \begin{bmatrix} \lambda & \mathbf{0}^T \\ \mathbf{0} & Q^T A_1 Q \end{bmatrix} \\ &= \begin{bmatrix} \lambda & \mathbf{0}^T \\ \mathbf{0} & D_1 \end{bmatrix} \end{aligned}$$

Since D_1 is a diagonal matrix, $P^T AP$ is a completely diagonal matrix. Thus, A is orthogonally diagonalizable, completing the induction step and the proof. □

8 Spectral Theorem

Recall: Orthogonally Diagonalizable

A real matrix A is orthogonally diagonalizable if $A = SDS^T$, where S is an orthogonal matrix and D is a diagonal matrix.

- **Theorem:** A real matrix is orthogonally diagonalizable if and only if A is symmetric.
- Given A is a real symmetric matrix, we follow two steps to diagonalize it:
  1. Find eigenvalues of A, denoted as $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$. If all eigenvalues are distinct, the corresponding eigenvectors are mutually orthogonal (i.e., $\langle v_i, v_j \rangle = 0$ for $i \neq j$).
  2. Find eigenvectors $\{v_1, v_2, \ldots, v_n\}$ corresponding to the eigenvalues, forming the matrix $S = [v_1, v_2, \ldots, v_n]$.

Spectral Theorem

Let A be an $n \times n$ real symmetric matrix. Then A has a spectral decomposition of the form:

$$A = \lambda_1 u_1 u_1^T + \lambda_2 u_2 u_2^T + \cdots + \lambda_n u_n u_n^T$$

Proof / Derivation

Proof. From orthogonal diagonalization, we can write $A = SDS^T$ where $S = [u_1, u_2, \dots, u_n]$ is an orthogonal matrix composed of orthonormal eigenvectors, and D is the diagonal matrix of eigenvalues. Expanding the matrices gives:

$$A = \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix} \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix} \begin{pmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_n^T \end{pmatrix}$$

Multiplying the first two matrices:

$$A = \begin{bmatrix} \lambda_1 u_1 & \lambda_2 u_2 & \dots & \lambda_n u_n \end{bmatrix} \begin{pmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_n^T \end{pmatrix}$$

Expanding this final product yields the spectral decomposition:

$$A = \lambda_1 u_1 u_1^T + \lambda_2 u_2 u_2^T + \dots + \lambda_n u_n u_n^T$$

□

Example: Spectral Decomposition

Find the spectral decomposition of the following matrix:

$$A = \begin{pmatrix} 5 & 1 & 1 \\ 1 & 5 & 1 \\ 1 & 1 & 5 \end{pmatrix}$$

Solution

First, we find the eigenvalues by solving the characteristic equation $\det(A - \lambda I) = 0$:

$$\det \begin{pmatrix} 5-\lambda & 1 & 1 \\ 1 & 5-\lambda & 1 \\ 1 & 1 & 5-\lambda \end{pmatrix} = 0$$

Observing that the row sums are all 7, $\lambda_1 = 7$ is an eigenvalue. Factoring out $(7 - \lambda)$ leaves a repeated root of $\lambda_2 = \lambda_3 = 4$.
Next, we find the corresponding orthonormal eigenvectors:

- For $\lambda_1 = 7$: Solving $(A - 7I)v = 0$ gives $v_1 = (1, 1, 1)^T$. Normalizing it gives $u_1 = \frac{1}{\sqrt{3}}(1, 1, 1)^T$.
- For $\lambda = 4$: Solving $(A - 4I)v = 0$ yields the plane $x + y + z = 0$. We pick two orthogonal vectors in this plane. Let $v_2 = (1, -1, 0)^T$, yielding normalized $u_2 = \frac{1}{\sqrt{2}}(1, -1, 0)^T$.

- For the third vector, we take the cross product $v_3 = v_1 \times v_2 = (1, 1, -2)^T$, yielding normalized $u_3 = \frac{1}{\sqrt{6}}(1, 1, -2)^T$.

Applying the Spectral Theorem formula, the decomposition is:

$$A = 7\left(\frac{1}{3}\begin{pmatrix}1\\1\\1\end{pmatrix}\begin{pmatrix}1 & 1 & 1\end{pmatrix}\right) + 4\left(\frac{1}{2}\begin{pmatrix}1\\-1\\0\end{pmatrix}\begin{pmatrix}1 & -1 & 0\end{pmatrix}\right) + 4\left(\frac{1}{6}\begin{pmatrix}1\\1\\-2\end{pmatrix}\begin{pmatrix}1 & 1 & -2\end{pmatrix}\right)$$

Corollary 1

Let A be a real positive definite symmetric matrix. Then there exists a non-singular matrix B such that $A = BB^T$.

Proof of Corollary 1

Proof. Since A is symmetric, we can orthogonally diagonalize it as $A = SDS^T$. Because it is positive definite, all its eigenvalues are positive, allowing us to split the diagonal matrix D into $D = D^{1/2}D^{1/2}$, where:

$$D^{1/2} = \begin{pmatrix}\sqrt{\lambda_1} & 0 & \dots & 0\\ 0 & \sqrt{\lambda_2} & \dots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \sqrt{\lambda_n}\end{pmatrix}$$

Letting $B = SD^{1/2}$, we can rewrite the decomposition as:

$$A = (SD^{1/2})(D^{1/2}S^T) = (SD^{1/2})(SD^{1/2})^T = BB^T$$

□

Corollary 2

A real symmetric matrix is positive definite if and only if its eigenvalues are strictly positive numbers.

Proof of Corollary 2

Proof. **Forward Direction ($\Longrightarrow$):** Assume A is positive definite. For any non-zero vector $x \neq 0$, we have $0 < x^TAx$. Let $v \neq 0$ be an eigenvector such that $Av = \lambda v$. By definition of the inner product, $\langle v, v\rangle = \|v\|^2 > 0$. Substituting v into the positive definite condition gives:

$$0 < v^TAv = v^T(\lambda v) = \lambda\|v\|^2 \implies \lambda > 0$$

Reverse Direction ($\Longleftarrow$): Conversely, assume all eigenvalues of A are positive

($\lambda_i > 0$). We must claim $x^T Ax > 0$ for all $x \neq 0$. Using the orthogonal decomposition $A = SDS^T$, let $x \neq 0$:

$$x^T Ax = x^T (SDS^T)x = (S^T x)^T D(S^T x)$$

Let $y = S^T x$. Since $x \neq 0$ and S^T is an invertible orthogonal matrix, it guarantees $y \neq 0$. Thus, the expression becomes $y^T Dy = \sum_{i=1}^{n} \lambda_i y_i^2$. Since at least one component of y is non-zero (meaning $y_i^2 > 0$) and all $\lambda_i > 0$, the sum is strictly greater than zero, proving $x^T Ax > 0$. □

9 Singular Value Decomposition (SVD)

Foundations of SVD

Let A be a $k \times n$ real matrix, defined as $A = (a_{ij})_{k\times n}$.

- Define $B = A^T A$. Since $B^T = (A^T A)^T = A^T (A^T)^T = A^T A = B$, it follows that $A^T A$ is a real symmetric matrix of size $n \times n$.
- The eigenvalues of $A^T A$, denoted $\sigma(A^T A)$, are a subset of $\mathbb{R}$. Let them be $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ where $\lambda_i \in \mathbb{R}$.
- Let λ be an eigenvalue of $A^T A$ corresponding to an eigenvector $v \neq 0$. We can show $\lambda \geq 0$ using the inner product:

$$0 \leq \langle Av, Av\rangle = (Av)^T (Av) = v^T (A^T A)v = v^T (\lambda v) = \lambda \|v\|^2 \implies \lambda \geq 0$$

- We define the *singular values* of matrix A as $\sigma_i = \sqrt{\lambda_i}$.

```
# Constructing the Orthogonal Bases
```

Let $\{v_1, v_2, \dots, v_n\}$ be an orthogonal set of eigenvectors of $A^T A$ which forms a basis for $\mathbb{R}^n$. Let $\{\lambda_1, \lambda_2, \dots, \lambda_r\}$ be the non-zero eigenvalues of $A^T A$, corresponding to singular values $\{\sigma_1, \sigma_2, \dots, \sigma_r\}$. We conventionally order them as:

$$\sigma_1 \geq \sigma_2 \geq \sigma_3 \dots \geq \sigma_r > 0$$

We calculate the norm of Av_i (assuming the eigenvectors are normalized so $\|v_i\| = 1$):

$$\|Av_i\| = (\langle Av_i, Av_i\rangle)^{1/2} = \left[(Av_i)^T Av_i\right]^{1/2} = \left[v_i^T (A^T A)v_i\right]^{1/2} = \sqrt{\lambda_i \|v_i\|^2} = \sqrt{\lambda_i} = \sigma_i$$

Because the eigenvectors of a symmetric matrix are orthogonal, $\langle Av_i, Av_j\rangle = 0$ for $i \neq j$. We can define normalized vectors u_i in the column space:

$$u_i = \frac{1}{\|Av_i\|} Av_i$$

This ensures $\|u_i\| = 1$ and $\langle u_i, u_j\rangle = 0$ for $i \neq j$. The set $\{u_1, u_2, \dots, u_r\}$ (for $r \leq n$) forms a basis for the column space of A, denoted col(A).

$$\text{col}(A) = \text{span}\{Av_1, \dots, Av_r\} \cup \text{span}\{Av_{r+1}, \dots\} = \text{span}\{Av_1, \dots, Av_r\}$$

Thus $\{u_1, u_2, \ldots, u_r\}$ forms an orthonormal basis for $\text{col}(A)$. We can use the extension of basis theorem to add vectors $u_{r+1}, \ldots, u_k$ to form a complete orthonormal basis for $\mathbb{R}^k$.

SVD Theorem

Let A be a $k \times n$ real matrix. Then there exist a $k \times k$ orthogonal matrix U and an $n \times n$ orthogonal matrix V such that:

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

where Σ is a $k \times n$ diagonal matrix of singular values.

Matrix Form Derivation

Since $u_i = \frac{1}{\|Av_i\|} Av_i$, rearranging gives $Av_i = \sigma_i u_i$. In block matrix form, we can write:

$$A_{k\times n} \begin{bmatrix} v_1 & v_2 & \ldots & v_n \end{bmatrix} = \begin{bmatrix} u_1 & \ldots & u_r & u_{r+1} & \ldots & u_k \end{bmatrix} \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}$$

where D is an $r \times r$ diagonal matrix containing the strictly positive singular values $\sigma_1, \ldots, \sigma_r$. Letting $U = [u_1, \ldots, u_k]$, $\Sigma = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}$, and $V = [v_1, \ldots, v_n]$, the equation becomes $AV = U\Sigma$. Right-multiplying by V^T (since V is orthogonal and $V^{-1} = V^T$) yields the final decomposition:

$$A_{k\times n} = \underbrace{[u_1 \ldots u_k]}_{k\times k} \underbrace{\Sigma}_{k\times n} \underbrace{[v_1 \ldots v_n]^T}_{n\times n}$$

Example: Example 1: SVD Calculation

Find the Singular Value Decomposition (SVD) for the following matrix:

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

Solution

First, we construct the symmetric matrix $A^T A$:

$$A^T A = \begin{pmatrix} 1 & 0 & 1 \\ -1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$$

The eigenvalues of A^TA are evidently $\sigma(A^TA) = \{3, 2\}$. The singular values are the square roots of the eigenvalues, ordered such that $\sigma_1 > \sigma_2$:

$$\sigma_1 = \sqrt{3}, \qquad \sigma_2 = \sqrt{2}$$

The corresponding normalized eigenvectors for A^TA are:

$$v_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (\text{for } \lambda = 3), \qquad v_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad (\text{for } \lambda = 2)$$

Next, we calculate the orthogonal basis vectors for the column space using $u_i = \frac{1}{\|Av_i\|}Av_i$:

$$u_1 = \frac{1}{\sqrt{3}} A \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{3}} \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$$
$$u_2 = \frac{1}{\sqrt{2}} A \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

To complete the orthonormal basis for $\mathbb{R}^3$, we find a vector u_3 that is orthogonal to both u_1 and u_2 by taking their cross product:

$$u_3' = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} \times \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$$

Normalizing u_3' yields $u_3 = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$.

The final SVD form $A = U\Sigma V^T$ is written as:

$$A = \underbrace{\begin{pmatrix} -1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & 0 & 2/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{2} & -1/\sqrt{6} \end{pmatrix}}_{U_{3\times 3}} \underbrace{\begin{pmatrix} \sqrt{3} & 0 \\ 0 & \sqrt{2} \\ 0 & 0 \end{pmatrix}}_{\Sigma_{3\times 2}} \underbrace{\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}}_{V^T_{2\times 2}}$$

10 Low Rank Approximation

This section defines a new concept of *Matrix Norms* and an important application of SVD

Finite Vector Norm

For a finite vector $x \in \mathbb{R}^n$ where $x = (x_1, x_2, \ldots, x_n)$, the general p-norm is defined as:

$$||x||_p := \left(\sum_i |x_i|^p \right)^{\frac{1}{p}}$$

Matrix Rank and Approximation Goal

- The rank of a matrix is the number of linearly independent columns or rows, which is equal to the number of non-zero singular values.
- **Aim:** In low rank approximation, the goal is to approximate a given matrix M with a lower-rank matrix M_k (where $\text{rank}(M_k) \leq k$) such that the distance $||M - M_k||$ is minimized.

Operator Norm (2-Norm)

The operator norm of a matrix M is defined as:

$$||M||_2 := \max_{x \neq 0} \frac{||Mx||_2}{||x||_2} = \max_{x \neq 0} \frac{(\langle Mx, Mx \rangle)^{1/2}}{(\langle x, x \rangle)^{1/2}}$$

Claim: $||M||_2 = \sigma_{\max}(M)$.

```
# Proof of Operator Norm Claim
```

Proof. We want to maximize the ratio $\frac{\langle Mx, Mx \rangle}{\langle x, x \rangle}$. Notice that $\langle Mx, Mx \rangle = x^T M^T M x$. By evaluating this for an eigenvector x of $M^T M$, the maximum value corresponds to the largest eigenvalue of $M^T M$, denoted as $\lambda_{\max}$. Thus, the maximum of the squared ratio is $\lambda_{\max}$. Taking the square root gives $\sqrt{\lambda_{\max}} = \sigma_{\max}(M)$. □

Courant-Fischer Characterization

Let $A = A^T \in \mathbb{M}_{n \times n}(\mathbb{R})$. Let $\{\lambda_1, \ldots, \lambda_n\}$ be the eigenvalues of A sorted in descending order. The k-th eigenvalue can be characterized using the Rayleigh Quotient $R(x) = \frac{x^T A x}{x^T x}$ as:

$$\lambda_k = \min_{\dim(V)=k} \max_{x \in V, x \neq 0} \frac{x^T A x}{x^T x}$$

Proof of Courant-Fischer Characterization

Proof. Let $A = A^T \in \mathbb{M}_{n\times n}(\mathbb{R})$. Consider the Rayleigh quotient $R(x) = \frac{x^TAx}{x^Tx}$. Let $\{v_1, v_2, \ldots, v_n\}$ be an orthonormal basis of eigenvectors for $\mathbb{R}^n$. By expanding $x = \sum_{i=1}^{k} c_i v_i$ in a fixed subspace of dimension k, we observe that $R(x) \geq \lambda_k$. Let V be any k-dimensional subspace and define $\bar{E}_k = \text{span}\{v_k, v_{k+1}, \ldots, v_n\}$. Since $\dim(\bar{E}_k) = n - k + 1$, the intersection $V \cap \bar{E}_k$ contains a non-zero vector x_0. For this vector, $R(x_0) = \frac{\sum_{i=k}^{n} \lambda_i c_i^2}{\sum c_i^2} \leq \lambda_k$. Thus, $\min_{\dim(V)=k} \max_{x\in V} R(x) = \lambda_k$. □

Low Rank Approximation in Operator Norm

Let $M \in \mathbb{M}_{m\times n}(\mathbb{R})$ with singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$. For any integer $k \geq 1$, the truncated SVD matrix $\tilde{M}_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T$ satisfies:

$$||M - \tilde{M}_k||_2 = \sigma_{k+1}$$

Furthermore, for *any* matrix M_k of rank at most k, we have $||M - M_k||_2 \geq \sigma_{k+1}$.

Proof of Approximation Error

Proof. Using SVD, the residual matrix is $M - \tilde{M}_k = \sum_{i=k+1}^{n} \sigma_i u_i v_i^T$. The operator norm of this residual matrix is its largest singular value. Since the singular values of the residual are $\sigma_{k+1}, \sigma_{k+2}, \ldots$, the maximum is σ_{k+1}. Thus, $||M - \tilde{M}_k||_2 = \sigma_{k+1}$. □

11 Frobenius Norm

Definition of Frobenius Norm

For any matrix $M \in \mathbb{M}_{m\times n}(\mathbb{R})$, the Frobenius norm is defined as:

$$||M||_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |m_{ij}|^2 \right)^{1/2}$$

This is equivalent to treating the matrix as a vector in $\mathbb{R}^{mn}$.

Properties of Frobenius Norm

1. $||M||_F^2 = \text{tr}(M^TM) = \text{tr}(MM^T)$.
2. **Non-negativity:** $||M||_F \geq 0$, and $||M||_F = 0 \implies m_{ij} = 0\ \forall i, j$.
3. **Homogeneity:** $||\alpha M||_F = |\alpha| ||M||_F$.

4. **Triangle Inequality:** $||M+N||_F \leq ||M||_F+||N||_F$. This is proven by treating the matrices as vectors in $\mathbb{R}^{mn}$ and applying the standard Cauchy-Schwarz inequality.
5. **Orthogonal Invariance:** The Frobenius norm is invariant under orthogonal transformations. If U and V are orthogonal matrices, then $||UAV||_F = ||A||_F$.

```
# Proofs for Frobenius Norm Properties
```

Proof. **Non-negativity:** The definition states $||M||_F^2 = \sum_i \sum_j |m_{ij}|^2 \geq 0$. If $||M||_F^2 = 0$, then $\sum_i \sum_j |m_{ij}|^2 = 0$, which implies $m_{ij} = 0 \quad \forall i = 1 \ldots m, j = 1 \ldots n$.
Homogeneity: Evaluating the scalar multiple gives $||\alpha M||_F^2 = \sum_i \sum_j |\alpha m_{ij}|^2 = |\alpha|^2 \sum_i \sum_j |m_{ij}|^2 = |\alpha|^2 ||M||_F^2$, meaning $||\alpha M||_F = |\alpha| ||M||_F$.
Triangle Inequality: Using the Cauchy-Schwartz inequality, we can bound the norm of $M+N$:

$$\begin{aligned}||M+N||_F^2 &= \sum_i \sum_j (m_{ij}+n_{ij})^2 \\ &= \sum_i \sum_j \left[(m_{ij})^2 + (n_{ij})^2 + 2m_{ij}n_{ij}\right] \\ &= ||M||_F^2 + ||N||_F^2 + 2\sum_i \sum_j m_{ij}n_{ij}\end{aligned}$$

By applying Cauchy-Schwartz ($\sum m_{ij}n_{ij} \leq (\sum m_{ij}^2)^{1/2}(\sum n_{ij}^2)^{1/2}$), we get:

$$||M+N||_F^2 \leq ||M||_F^2 + ||N||_F^2 + 2||M||_F||N||_F = (||M||_F + ||N||_F)^2$$

This directly simplifies to $||M+N||_F \leq ||M||_F + ||N||_F$. □

```
# Proof of Orthogonal Invariance
```

Proof. By the trace property of the Frobenius norm:

$$||UAV||_F^2 = \text{tr}((UAV)^T(UAV)) = \text{tr}(V^TA^TU^TUAV)$$

Since U is orthogonal, $U^TU = I$. Thus:

$$||UAV||_F^2 = \text{tr}(V^TA^TAV)$$

Using the cyclic property of the trace ($\text{tr}(AB) = \text{tr}(BA)$), we get:

$$\text{tr}(VV^TA^TA) = \text{tr}(IA^TA) = \text{tr}(A^TA) = ||A||_F^2$$

□

Eckart-Young-Mirsky Theorem

For any matrix $A \in \mathbb{M}_{m\times n}(\mathbb{R})$, the best rank-$k$ approximation under the Frobenius norm is given by the truncated SVD matrix A_k:

$$\min_{\text{rank}(B)\leq k} ||A - B||_F = ||A - A_k||_F = \left(\sum_{i=k+1}^{n} \sigma_i^2\right)^{1/2}$$

This theorem serves as the basis for greedy algorithms in low-rank approximations.

```
# Proof Sketch of Eckart-Young-Mirsky
```

Proof. Let $A = U\Sigma V^T$ and let B be an arbitrary matrix with rank at most k. Because the Frobenius norm is invariant under orthogonal transformations, we have:

$$||A - B||_F = ||U^T(A - B)V||_F = ||\Sigma - U^T BV||_F$$

Let $C = U^T BV$. Since U and V are full rank, $\text{rank}(C) = \text{rank}(B) \leq k$. Thus, the problem reduces to minimizing $||\Sigma - C||_F$ where C is a matrix of rank at most k. Expanding this difference shows that the minimum is achieved when C matches the first k diagonal entries of Σ (the largest singular values). The remaining error is the sum of the squares of the discarded singular values $\sum_{i=k+1}^{n} \sigma_i^2$. $\square$

Example: Real-World Application

Consider a Consumer Matrix $M \in \mathbb{M}_{m\times n}(\mathbb{R})$ where each row corresponds to a consumer and each column corresponds to a product. The entry (i, j) represents the probability that consumer i purchases product j. Low rank approximation can be used to extract the most significant purchasing trends. We can hypothesize that there are k hidden features in consumers, such as age, gender, and annual income. This allows us to represent the consumer matrix algebraically as $M = AB$, where A is a factor weight matrix and B is a features matrix. In this context, the inner product between two such matrices A and B can be defined as $\langle A, B\rangle := \text{tr}(B^T A)$, which satisfies the Cauchy-Schwarz inequality $\text{tr}(B^T A) \leq (\text{tr}(A^T A))^{1/2}(\text{tr}(B^T B))^{1/2}$.