Algèbre Linéaire

Author: CatMono

Date: July, 2025

Version: 0.1

Contents

Preface

The set of all $m \times n$ matrices over the number field P is denoted by $M_{m \times n}(P)$ or $P^{m \times n}$. Similarly, the set of all linear mappings from V to V' over the field F is represented by $\mathrm{Hom}(V, V')$ or $\mathcal{L}(V, V')$. Furthermore, when considering the set of all linear transformations, it is commonly denoted by $\mathrm{Hom}(V)$, $\mathcal{L}(V)$, or $\mathrm{End}(V)$.

Chapter 1 Preliminaries

Chapter 2 Determinants

2.1 Permutations

2.2 Determinant and Its Properties

2.3 Expanding by Rows (Columns)

¶ Expanding by One Row
¶ Cramer's Rule
¶ Expanding by k Rows

2.4 Special Determinants

Definition 2.1 (Vandermonde Determinant)

The Vandermonde determinant is defined as

$$V_n = \begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ x_1 & x_2 & x_3 & \cdots & x_n \\ x_1^2 & x_2^2 & x_3^2 & \cdots & x_n^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_1^{n-1} & x_2^{n-1} & x_3^{n-1} & \cdots & x_n^{n-1} \end{vmatrix}$$

where $x_1, x_2, \ldots, x_n$ are distinct variables. ♣

The value of the Vandermonde determinant is given by

$$V_n = \prod_{1 \leq i < j \leq n} (x_j - x_i).$$

Definition 2.2 (Arrow Determinant)

The Arrow determinant ($\searrow$) is defined as

$$A_n = \begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & 0 & \cdots & 0 \\ a_{31} & 0 & a_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & 0 & 0 & \cdots & a_{nn} \end{vmatrix}.$$

The value of the Arrow determinant is given by

$$A_n = \left(a_{11} - \sum_{k=2}^{n} \frac{a_{1k} a_{k1}}{a_{kk}} \right) \prod_{k=2}^{n} a_{kk}.$$

♣

From the first column sequentially, subtract $\frac{a_{21}}{a_{22}}$ times the second column, $\cdots$, $\frac{a_{n1}}{a_{nn}}$ times the n-th column, so that the first column becomes:

$$\begin{bmatrix} a_{11} - \sum_{k=2}^{n} \frac{a_{1k} a_{k1}}{a_{kk}} & 0 & 0 & \vdots & 0 \end{bmatrix}^{\mathrm{T}}.$$

Then expand along the first column.

> **Definition 2.3 (Two-Triangular Determinant)**
>
> If the determinant satisfies
> $$a_{ij} = \begin{cases} a, & i < j, \\ x_i, & i = j, \\ b, & i > j, \end{cases}$$
> then D_n is called a two-triangular determinant. ♣

The value of the two-triangular determinant is given by

$$\begin{vmatrix} x_1 & a & a & \dots & a \\ b & x_2 & a & \dots & a \\ b & b & x_3 & \dots & a \\ \vdots & \vdots & \vdots & & \vdots \\ b & b & b & \dots & x_n \end{vmatrix} = \begin{cases} \left[x_1 + a \sum\limits_{k=2}^{n} \frac{x_1 - a}{x_k - a} \right] \cdot \prod\limits_{k=2}^{n} (x_k - a), & a = b \\ (x_n - b)D_{n-1} + \prod\limits_{k=1}^{n-1} (x_k - a), & a \neq b \end{cases}$$

Chapter 3   System of Linear Equations

> **Definition 3.1 (System of Linear Equations)**
>
> A system of linear equations is a collection of one or more linear equations involving the same set of variables. For example, a system of m linear equations in n variables can be written as:
>
> $$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$
>
> where $x_1, x_2, \ldots, x_n$ are the variables, a_{ij} are the coefficients, and b_i are the constants.
>
> A solution to the system is an ordered set of values for the variables that satisfies all equations simultaneously. If two systems have the same solution set, they are called equivalent systems, whose relationship are equivalence. ♣

3.1 Elimination Method

3.2 Linear Space

> **Definition 3.2 (Linear Space over Field F)**
>
> Let V be a non-empty set, and F be a field. If the following two operations (binary mapping) are defined on V:
>
> **Vector addition** For any $\alpha, \beta \in V$, there exists $\gamma \in V$ such that $\gamma = \alpha + \beta$.
>
> **Scalar multiplication** For any $\alpha \in V$ and $c \in F$, there exists $\gamma \in V$ such that $\gamma = c \cdot \alpha$.
>
> and the following axioms are satisfied (for any $\alpha, \beta, \gamma \in V$ and $k, l \in F$):
>
> **A1. Commutativity of Addition** $\alpha + \beta = \beta + \alpha$.
>
> **A2. Associativity of Addition** $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$.
>
> **A3. Existence of Additive Identity** There exists an element $0 \in V$ such that $\alpha + 0 = 0 + \alpha = \alpha$.
>
> **A4. Existence of Additive Inverse** There exists an element $-\alpha \in V$ such that $\alpha + (-\alpha) = (-\alpha) + \alpha = 0$.
>
> **M1. Compatibility of Scalar Multiplication with Field Multiplication** $k(l\alpha) = (kl)\alpha$.
>
> **M2. Identity Element of Scalar Multiplication** $1 \cdot \alpha = \alpha$, where 1 is the multiplicative identity in F.
>
> **D1. Distributivity of Scalar Multiplication with respect to Vector Addition** $k(\alpha + \beta) = k\alpha + k\beta$.
>
> **D2. Distributivity of Scalar Multiplication with respect to Field Addition** $(k + l)\alpha = k\alpha + l\alpha$.
>
> Then V is called a linear space (or vector space) over the field F. ♣

⚡**Remark** In the definition of linear space, **A1. Commutativity of Addition** can be derived from the other axioms, so it is not strictly necessary to include it as an axiom.

Give some examples of linear spaces:

1. The ring of univariate polynomials $P[x]$ over the number field P, with the usual polynomial addition and scalar multiplication by elements from P, forms a linear space over the number field P. If we consider only polynomials of degree less than n, along with the zero polynomial, this also forms a linear space over P, denoted as $P[x]_n$.

2. The set of all positive real numbers $\mathbb{R}^+$, with addition and scalar multiplication defined as:

$$a \oplus b = ab, \quad k \circ a = a^k,$$

forms a linear space over the number field $\mathbb{R}$.

3. The set of all real-valued functions, under function addition and scalar multiplication by elements from $\mathbb{R}$, forms a linear space over the field $\mathbb{R}$.

4. The field P itself, under its own addition and multiplication, forms a linear space over itself.

5. The set of all n-dimensional vectors (ordered n-tuples of elements from P) over the number field P, with the usual vector addition and scalar multiplication, forms a linear space over the number field P.

In this section, we primarily examine linear spaces consisting of n-dimensional vectors over the real number field $\mathbb{R}$. As a matter of fact, all elements of a linear space can be called vectors, and the definitions and properties of vectors in $\mathbb{R}^n$ can be generalized to any linear space.

¶ **Linear Independence**

Definition 3.3 (Linear Combination)

Let V be a linear space over the field F, and let $\alpha_1, \alpha_2, \ldots, \alpha_n \in V$. For any scalars $k_1, k_2, \ldots, k_n \in F$, the vector

$$\beta = k_1\alpha_1 + k_2\alpha_2 + \cdots + k_n\alpha_n$$

is called a **linear combination** of the vectors $\alpha_1, \alpha_2, \ldots, \alpha_n$ (we also say that β is linearly expressed by $\alpha_1, \alpha_2, \ldots, \alpha_n$), where k_i are called the coefficients of the linear combination. ♣

Definition 3.4 (Linear Representation and Equivalence of Vector Groups)

If every vector α_i $(i = 1, 2, \ldots, t)$ in the vector group $\alpha_1, \alpha_2, \ldots, \alpha_t$ can be linearly expressed by the vector group $\beta_1, \beta_2, \ldots, \beta_s$, then the vector group $\alpha_1, \alpha_2, \ldots, \alpha_t$ is said to be linearly expressed by the vector group $\beta_1, \beta_2, \ldots, \beta_s$.

If two vector groups can be linearly expressed by each other, they are said to be **equivalent**. ♣

⚡**Remark** Linear representation can be used to identify redundant equations in a system of linear equations. In a system of linear equations, if one equation can be linearly expressed by other equations, then through elementary transformations, the corresponding row in the matrix can be reduced to a row of zeros.

If two vector groups are equivalent, then their corresponding systems of linear equations are equivalent.

Definition 3.5 (Linear Dependence and Independence of Vector Groups)

A vector group $\alpha_1, \alpha_2, \ldots, \alpha_t (t \geq 2)$ is said to be **linearly dependent** if there exists a vector in the group that can be linearly expressed by the other vectors.

Another equivalent definition is: A vector group $\alpha_1, \alpha_2, \ldots, \alpha_t (t \geq 1)$ is linearly dependent if there exist <u>not all zero</u> scalars $k_1, k_2, \ldots, k_t$ in the field F such that:

$$k_1\alpha_1 + k_2\alpha_2 + \cdots + k_t\alpha_t = 0.$$

Conversely, a vector group $\alpha_1, \alpha_2, \ldots, \alpha_t (t \geq 1)$ is **linearly independent** if: the fact that

$$k_1\alpha_1 + k_2\alpha_2 + \cdots + k_t\alpha_t = 0$$

implies:

$$k_1 = k_2 = \cdots = k_t = 0.$$

♣

🕏 **Note** *If a subset of a vector group is linearly dependent, then the entire vector group is linearly dependent. If a vector group is linearly independent, then any non-empty subset of the group is also linearly independent.* (**Partial dependence implies overall dependence; overall independence implies partial independence.**)

Particularly, since two proportional vectors are linearly dependent, a linearly independent vector group cannot contain two proportional vectors.

A few propositions about linear dependence and independence are given below:

Proposition 3.1

1. Generally, for a vector group $\alpha_i = (a_{i1}, a_{i2}, \ldots, a_{in})$, $i = 1, 2, \ldots, s$, the necessary and sufficient condition for linear dependence is that the equation:

$$x_1\alpha_1 + x_2\alpha_2 + \cdots + x_s\alpha_s = 0$$

or the homogeneous system of linear equations:

$$\begin{cases} a_{11}x_1 + a_{21}x_2 + \cdots + a_{s1}x_s = 0 \\ a_{12}x_1 + a_{22}x_2 + \cdots + a_{s2}x_s = 0 \\ \vdots \\ a_{1n}x_1 + a_{2n}x_2 + \cdots + a_{sn}x_s = 0 \end{cases}$$

has a non-zero solution.

2. If a vector group $\alpha_i = (a_{i1}, a_{i2}, \ldots, a_{in})$, $i = 1, 2, \ldots, s$ is linearly independent, then adding one component to each vector to form an $n+1$-dimensional vector group $\beta_i = (a_{i1}, a_{i2}, \ldots, a_{in}, a_{i,n+1})$, $i = 1, 2, \ldots, s$, will also be linearly independent.

3. For two vector groups $\alpha_1, \alpha_2, \ldots, \alpha_r$ and $\beta_1, \beta_2, \ldots, \beta_s$, if the former can be linearly expressed by the latter:

   i. if $r > s$, then the former is linearly dependent.

   ii. if $\alpha_1, \alpha_2, \ldots, \alpha_r$ is linearly independent, then $r \leqslant s$; and there exists a subset of $\beta_1, \beta_2, \ldots, \beta_s$ (without loss of generality, assume it comprises the first r vectors, $\beta_1, \beta_2, \ldots, \beta_r$), it can be replaced by $\alpha_1, \alpha_2, \ldots, \alpha_r$ to form a new vector group $\alpha_1, \alpha_2, \ldots, \alpha_r, \beta_{r+1}, \beta_{r+2}, \ldots, \beta_s$, which is equivalent to $\beta_1, \beta_2, \ldots, \beta_s$ (This is **the Steinitz replacement principle**).

4. Any $n+1$ n-dimensional vectors must be linearly dependent.

5. Two equivalent linearly independent vector groups must contain the same number of vectors.

6. Let vector group $\alpha_1, \alpha_2, \ldots, \alpha_r$ be linearly independent, and β be another vector. Then, either β can be uniquely linearly expressed by $\alpha_1, \alpha_2, \ldots, \alpha_r$, or the vector group $\alpha_1, \alpha_2, \ldots, \alpha_r, \beta$ is linearly independent.

7. If a vector β can be linearly expressed by a vector group $\alpha_1, \alpha_2, \ldots, \alpha_r$, then $\alpha_1, \alpha_2, \ldots, \alpha_r$ is linearly independent if and only if the expression for β is unique.

♠

📌 *Proof*

1.

2.

6. If $\alpha_1, \alpha_2, \ldots, \alpha_r, \beta$ is linearly independent, then conclusion holds naturally.

Now suppose $\alpha_1, \alpha_2, \ldots, \alpha_r, \beta$ is linearly dependent, that is, there exist not all zero scalars $k_1, k_2, \ldots, k_r, k$ such that:

$$k_1\alpha_1 + k_2\alpha_2 + \cdots + k_r\alpha_r + k\beta = 0.$$

If $k = 0$, then $k_1\alpha_1 + k_2\alpha_2 + \cdots + k_r\alpha_r = 0$ (not all k_i are zero), which contradicts the linear independence of $\alpha_1, \alpha_2, \ldots, \alpha_r$.

Thus, $k \neq 0$, and we have:

$$\beta = -\frac{k_1}{k}\alpha_1 - \frac{k_2}{k}\alpha_2 - \cdots - \frac{k_r}{k}\alpha_r.$$

The expression for β is obtained.

Now, suppose there exists another expression:

$$\beta = l_1\alpha_1 + l_2\alpha_2 + \cdots + l_r\alpha_r.$$

Subtracting the two expressions yields:

$$\left(l_1 + \frac{k_1}{k}\right)\alpha_1 + \left(l_2 + \frac{k_2}{k}\right)\alpha_2 + \cdots + \left(l_r + \frac{k_r}{k}\right)\alpha_r = 0.$$

Due to the linear independence of $\alpha_1, \alpha_2, \ldots, \alpha_r$, we have:

$$l_1 + \frac{k_1}{k} = 0, \quad l_2 + \frac{k_2}{k} = 0, \quad \ldots, \quad l_r + \frac{k_r}{k} = 0,$$

which implies $l_i = -\frac{k_i}{k}$ for $i = 1, 2, \ldots, r$. Thus, the expression for β is unique. ∎

¶ **Dimension, Basis, and Coordinates**

> **Definition 3.6 (Dimension)**
>
> If a linear space V over field F contains n vectors that are linearly independent, but any more than n vectors are linearly dependent, then V is called an n-dimensional linear space, noted as $\dim V_F = n$, simplified as $\dim V = n$ when the field is clear from context.
>
> If a linear space does contain as many vectors that are linearly independent as you wish, it is called an infinite-dimensional linear space. ♣

> **Definition 3.7 (Basis and Coordinates)**
>
> In an n-dimensional linear space V, if $\alpha_1, \alpha_2, \ldots, \alpha_n$ are n linearly independent vectors in V, then they form a **basis** of V.
>
> Any vector $\alpha \in V$ can be uniquely expressed as a linear combination of the basis vectors:
>
> $$\alpha = k_1\alpha_1 + k_2\alpha_2 + \cdots + k_n\alpha_n,$$
>
> where $k_1, k_2, \ldots, k_n \in F$ are called the **coordinates** of α with respect to the basis $\alpha_1, \alpha_2, \ldots, \alpha_n$. ♣

The definition leads to a theorem that is easy to see: If there are n linearly independent vectors in a linear space V, and any vector in V can be expressed as a linear combination of these n vectors, then these n vectors form a basis of V, and V is an n-dimensional linear space.

¶ **Linear Subspaces**

> **Definition 3.8 (Linear Subspace)**
>
> Let V be a linear space over field F, and let $W \subseteq V$ be a non-empty subset of V. If W itself is a linear space under the same addition and scalar multiplication as in V, then W is called a **linear subspace** of V.
>
> Zero Subspace and whole space V itself are called trivial subspaces of V. ♣

It can be easily derived that: Let W be a non-empty subset of a linear space V. Then W is a linear subspace of V if and only if:

1. For any $\alpha, \beta \in W$, $\alpha + \beta \in W$.
2. For any $\alpha \in W$ and $k \in F$, $k\alpha \in W$.

3.3 Rank of Vector Groups and Matrices

¶ **Maximal Linearly Independent Group and Rank of Vector Groups**

> **Definition 3.9 (Maximal Linearly Independent Group)**
>
> A subset of a vector group is called a **maximal linearly independent group** if:
>
> 1. The subset itself is linearly independent.
> 2. Adding any vector (if available) from the vector group to this subset makes the new subset linearly dependent. ♣

Property

1. *The maximal linearly independent group is not unique; any maximal linearly independent group is equivalent to the original vector group.*
2. *All maximal linearly independent groups of a vector group contain the same number of vectors.*

> **Definition 3.10 (Rank of Vector Groups)**
>
> The number of vectors contained in the maximal linearly independent group of a vector group is called the **rank** of the vector group. ♣

Property

1. *A vector group is linearly independent if and only if its rank is equal to the number of vectors it contains.*
2. *Equivalent vector groups have the same rank, but vector groups with the same rank are not necessarily equivalent.*
3. *If rank of a vector group is r, then any r linearly independent vectors from the group form a maximal linearly independent group.*
4. *A vector group containing non-zero vectors always has a maximal linearly independent group, and any linearly independent subset of vectors can be expanded into a maximal linearly independent group.*
5. *If vector group (I) can be linearly expressed by vector group (II), then the rank of (I) does not exceed the rank of (II).*

¶ **Rank of Matrices**

> **Definition 3.11 (Rank of Matrices)**
>
> The row rank of a matrix is defined as the rank of its row vectors; the column rank of a matrix is defined as the rank of its column vectors. The rank of a matrix is defined as the highest order of its non-zero minors. ♣

> **Theorem 3.1**
>
> The row rank and column rank of any matrix are equal. ♡

Property

1. *Elementary row and column transformations do not change the rank of a matrix.*
2. *The rank of a matrix is equal to the number of non-zero rows in the row (column) echelon form of the matrix obtained by elementary row (column) operations.*

3. *Elementary row (column) operations do not alter the linear relationships among column (row) vectors.*

4. *For a square matrix* $(a_{ij})_{n \times n}$, *its row vectors (or column vectors) are linearly independent if and only if its determinant* $\det(a_{ij}) \neq 0$.

The third property gives a method to find the maximal linearly independent group of row (or column) vectors of a matrix, take the column vectors as an example:

1. Transform the matrix into its row echelon form (REF) or reduced row echelon form (RREF) using elementary row operations. For example, for a matrix $(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)$, its REF is shown below:

$$\begin{pmatrix} a_{11} & a_{12} & 0 & 0 & 0 \\ 0 & 0 & a_{23} & 0 & 0 \\ 0 & 0 & 0 & 0 & a_{35} \end{pmatrix}.$$

2. Identify the leading 1s (pivots) in each row. The columns containing these pivots correspond to the maximal linearly independent group of row vectors. In this example, $\alpha_1, \alpha_3, \alpha_5$ form a maximal linearly independent group.

3.4 Solution to Linear Systems

¶ **Necessary and Sufficient Conditions for the Existence of Solutions**

¶ **The Structure of Solutions to Homogeneous Systems**

For a homogeneous system of linear equations, it is evident that the linear combination of solutions is still a solution to the system. Thus, the concept of a **fundamental system of solutions** is introduced.

> **Definition 3.12 (Fundamental System of Solutions)**
>
> The homogeneous system of linear equations:
>
> $$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = 0 \\ \vdots \\ a_{s1}x_1 + a_{s2}x_2 + \cdots + a_{sn}x_n = 0 \end{cases}$$
>
> (or in matrix form $AX = O$) has a set of solutions $\eta_1, \eta_2, \ldots, \eta_t$, which is called a **fundamental system of solutions**, if:
>
> 1. Any solution of the system can be expressed as a linear combination of $\eta_1, \eta_2, \ldots, \eta_t$.
> 2. $\eta_1, \eta_2, \ldots, \eta_t$ are linearly independent[a].
>
> ---
> [a] ensuring there are no redundant solutions in the fundamental system of solutions

⚡**Remark** This definition is similar to the maximal linearly independent group of a vector group. In fact, the fundamental solution set is a maximal linearly independent group of the solution vector group. Therefore, the number of solutions it contains equals its rank.

All solutions of a homogeneous system of linear equations ($AX = O$) form a linear space (solution space), whose basis is the fundamental system of solutions. Dimension of solution space is $n - r(A)$ (where n is the number of variables, and $r(A)$ is the rank of the coefficient matrix A). Thus, a fundamental system of solutions contains exactly $n - r(A)$ solutions.

¶ **The Structure of Solutions to Nonhomogeneous Linear Systems**

The homogeneous system of linear equations $AX = O$ is called the **derived system** of the nonhomogeneous system $AX = B$.

It can be easily proven that:

- The difference between two solutions of $AX = B$ is a solution of its derived system $AX = 0$.
- The sum of a solution of $AX = B$ and a solution of $AX = 0$ is still a solution of $AX = B$.

If γ_0 is a particular solution of $AX = B$, then any solution γ of $AX = B$ can be expressed as:

$$\gamma = \gamma_0 + \eta,$$

where η is a solution of $AX = 0$.

Thus, for any particular solution γ_0 of $AX = B$, when η takes all the solutions of its derived system $AX = 0$, the above formula provides all the solutions of $AX = B$.

Chapter 4 Matrices

4.1 Basic Operations

¶ `Addition`
¶ `Scalar Multiplication`
¶ `Transpose`
¶ `Matrix Multiplication`
¶ `Operations of Determinants`

> **Theorem 4.1 (Determinant of a Product)**
>
> For square matrices A and B of the same order, the determinant of their product is equal to the product of their determinants:
> $$|AB| = |A| \cdot |B|.$$
>
> It can be easily extended to multiple matrices:
> $$|A_1 A_2 \cdots A_k| = |A_1| \cdot |A_2| \cdots |A_k|.$$
> ♡

> **Theorem 4.2 (Cauchy-Binet Formula)**
>
> Let $A = (a_{ij})_{m \times n}$ and $B = (b_{ij})_{n \times m}$:
> 1. If $m > n$, then $|AB| = 0$;
> 2. If $m \leq n$, then $|AB|$ is equal to the sum of products of all m-step minors of A and the corresponding m-step minors of B, that is:
> $$|AB| = \sum_{1 \leq v_1 < v_2 < \cdots < v_m \leq n} \left| A \begin{pmatrix} 1, 2, \ldots, m \\ v_1, v_2, \ldots, v_m \end{pmatrix} \right| \cdot \left| B \begin{pmatrix} v_1, v_2, \ldots, v_m \\ 1, 2, \ldots, m \end{pmatrix} \right|.$$
> ♡

¶ `Matrix Equivalence`

> **Definition 4.1 (Matrix Equivalence)**
>
> Two matrices A and B are said to be **equivalent**, denoted as $A \cong B$, if they can be transformed into one another by a combination of elementary row and column operations.
> ♣

⚘ **Note** *Obviously, two matrices are equivalent if and only if they have the same rank.*

> **Definition 4.2 (Canonical Form of Matrix Equivalence)**
>
> For any matrix A with rank r, it can be equivalent to the matrix
> $$J_r = \begin{pmatrix} I_r & O \\ O & O \end{pmatrix}_{m \times n},$$
> which is called the **canonical form** of matrix equivalence.
> ♣

4.2 Special Square Matrices

¶ `Diagonal Matrix`

> **Definition 4.3 (Diagonal Matrix)**
>
> A matrix $A = (a_{ij})_{n \times n}$ is called a **diagonal matrix** if:
> $$a_{ij} = 0, \quad i \neq j,$$
> denoted as:
> $$A = \mathrm{diag}(a_{11}, a_{22}, \ldots, a_{nn}).$$
> If
> $$a_{i,n+1-i} = 0, \quad i = 1, 2, \ldots, n,$$
> then A is called an **skew-diagonal matrix**, denoted as:
> $$A = \mathrm{sdiag}(a_{1n}, a_{2,n-1}, \ldots, a_{n1}).$$

✒ Property

1. *The sum, difference, and product of two diagonal matrices of the same order are still diagonal matrices.*

2. *Multiplying a matrix A on the left (right) by a diagonal matrix is equivalent to multiplying the corresponding rows (columns) of A by the diagonal elements of the diagonal matrix.*

3.
$$\mathrm{diag}(a_1, a_2, \ldots, a_n)^{-1} = \mathrm{diag}\left(\frac{1}{a_1}, \frac{1}{a_2}, \ldots, \frac{1}{a_n}\right),$$
$$\mathrm{sdiag}(a_1, a_2, \ldots, a_n)^{-1} = \mathrm{sdiag}\left(\frac{1}{a_n}, \frac{1}{a_{n-1}}, \ldots, \frac{1}{a_1}\right);$$
$$\begin{bmatrix} A & O \\ O & C \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} & O \\ O & C^{-1} \end{bmatrix},$$
$$\begin{bmatrix} O & B \\ D & O \end{bmatrix}^{-1} = \begin{bmatrix} O & D^{-1} \\ B^{-1} & O \end{bmatrix}.$$

¶ Fundamental Matrix

> **Definition 4.4 (Fundamental Matrix)**
>
> E_{ij} is called a **fundamental matrix** if the element in the i-th row and j-th column is 1, and all other elements are 0.

✒ Property

1. *Multiplying a matrix A on the left by a fundamental matrix is equivalent to moving the j-th row of A to the i-th position, with all other entries in the resulting matrix being zero; multiplying A on the right by a fundamental matrix is equivalent to moving the i-th column of A to the j-th position, with all other entries in the resulting matrix being zero.*[1]

2. *If A is commutative with all fundamental matrices, then A is a scalar matrix.*

¶ Triangular Matrix

---

[1]The mnemonic is:
$$\text{moves row: } E_{i \leftarrow j}A; \qquad \text{moves column: } AE_{i \to j};$$

> **Definition 4.5 (Triangular Matrix)**
>
> A matrix $A = (a_{ij})_{n \times n}$ is called an **upper triangular matrix** if:
> $$a_{ij} = 0, \quad i > j,$$
> denoted as:
> $$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}.$$
>
> Similarly, if:
> $$a_{ij} = 0, \quad i < j,$$
> then A is called a **lower triangular matrix**, denoted as:
> $$A = \begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}.$$
> ♣

🖉 Property

1. *The sum, difference, and product of two upper (lower) triangular matrices of the same order are still upper (lower) triangular matrices.*
2. *For all square matrix A, A can be expressed as the sum of an upper triangular matrix and a lower triangular matrix:*
$$A = \frac{A + A^T}{2} + \frac{A - A^T}{2}.$$
3. *(LU **Decomposition**) If all leading principal minors of a square matrix A are non-zero, then A can be uniquely decomposed into the product of a lower triangular matrix L (with all diagonal elements equal to 1) and an invertible upper triangular matrix U:*
$$A = LU.$$

   The converse is also true.

¶ Elementary Matrix

> **Definition 4.6 (Elementary Matrix)**
>
> An **elementary matrix** is obtained by performing a single elementary row operation on an identity matrix E.
> There are three types of elementary matrices (also hold for column operations):
> **Row swapping** Interchanges rows i and j of E: $P(i, j)$.
> **Row scaling** Multiplies row i of E by a non-zero scalar c: $P(i(c))$.
> **Row addition** Adds k times row j to row i of E: $P(i, j(k))$.
> ♣

🖐 **Note** *It can be proven that $P(i, j)$ can be implemented by a sequence of $P(i(c))$ and $P(i, j(k))$ operations.*

¶ Symmetric Matrix

> **Definition 4.7 (Symmetric Matrix)**
>
> A matrix $A = (a_{ij})_{n \times n}$ is called a **symmetric matrix** if:
>
> $$a_{ij} = a_{ji}, \quad i, j = 1, 2, \ldots, n,$$
>
> or equivalently:
>
> $$A = A^T.$$
>
> Similarly, if:
>
> $$A = -A^T,$$
>
> then A is called a **skew-symmetric matrix**. ♣

Property

1. *The linear combination of two symmetric (skew-symmetric) matrices of the same order are still symmetric (skew-symmetric) matrices.*

2. *For any square matrix A, A can be expressed as the sum of a symmetric matrix and a skew-symmetric matrix:*
$$A = \frac{A + A^T}{2} + \frac{A - A^T}{2}.$$

3. *The determinant of an odd-order skew-symmetric matrix is zero; the rank of a skew-symmetric matrix is even.*

¶ **Circulant Matrix**

Let

$$C = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix},$$

it is called a **cyclic shift matrix**.

The matrix C performs specific cyclic operations when multiplied by another matrix:

Left multiplication Multiplying a matrix A from the left by C (CA) is equivalent to cyclically shifting the rows of A upwards by one position. The first row of A moves to the last row.

Right multiplication Multiplying a matrix A from the right by C (AC) is equivalent to cyclically shifting the columns of A to the right by one position. The last column of A moves to the first column.

> **Definition 4.8 (Circulant Matrix)**
>
> A matrix A is called a **circulant matrix** if it can be expressed as:
>
> $$A = \begin{bmatrix} a_0 & a_1 & a_2 & \cdots & a_{n-1} \\ a_{n-1} & a_0 & a_1 & \cdots & a_{n-2} \\ a_{n-2} & a_{n-1} & a_0 & \cdots & a_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_1 & a_2 & a_3 & \cdots & a_0 \end{bmatrix},$$
>
> which is constructed such that each row is obtained by cyclically shifting the elements of the first row one position to the right.

It can also be expressed as:

$$A = a_0 E + a_1 C + a_2 C^2 + \cdots + a_{n-1} C^{n-1} = \sum_{i=0}^{n-1} a_i C^i,$$

where E is the identity matrix and C is the cyclic shift matrix.

♣

¶ **Nilpotent Matrix**

Definition 4.9 (Nilpotent Matrix)

A square matrix A is called a **nilpotent matrix** if there exists $l \in \mathbb{N}_+$ such that:

$$A^l = O,$$

where O is the zero matrix. The smallest such l is called the **nilpotent index** of the nilpotent matrix A.

♣

Proposition 4.1

1. A triangular matrix is nilpotent if and only if all its diagonal elements are zero.
2. If a n-order triangular matrix is nilpotent, then its nilpotent index l must satisfy $l \leq n$.

♠

¶ **Full Row Rank Matrix**

Definition 4.10 (Full Row Rank Matrix)

A matrix A is called a **full row rank matrix** if its row vectors are linearly independent. Similarly, if its column vectors are linearly independent, then A is called a **full column rank matrix**.

♣

🔗 **Property**

1. $A_{m \times n}$ is a full row rank matrix if and only if there exists an invertible matrix $Q_{n \times n}$ such that
$$A = \begin{bmatrix} E_m & O \end{bmatrix} Q.$$

2. $A_{m \times n}$ is a full column rank matrix if and only if there exists an invertible matrix $P_{m \times m}$ such that
$$A = P \begin{bmatrix} E_n \\ O \end{bmatrix}.$$

4.3 Inverse Matrix

¶ **Inverse Matrix and Its Operations**
¶ **Equivalent Propositions and Method of Inversion**
¶ **Generalized Inverse**

4.4 Block Matrix

Definition 4.11 (Kronecker Product)

Let $A = (a_{ij})_{m \times n}$ and $B = (b_{ij})_{p \times q}$ be two matrices. The **Kronecker product** of A and B, denoted as $A \otimes B$,

is defined as the block matrix:

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix}_{mp \times nq}.$$

♣

¶ **Common Conclusions**

Theorem 4.3 (Determinant Reduction Formula)

Let $A_{m \times m}, B_{m \times n}, C_{n \times m}, D_{n \times n}$ be matrices. Then:

1. If A is invertible, then:
$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = |A| \cdot |D - CA^{-1}B|.$$

2. If D is invertible, then:
$$\begin{vmatrix} A & B \\ C & D \end{vmatrix} = |D| \cdot |A - BD^{-1}C|.$$

3. If both A and D are invertible, then:
$$|D| \cdot |A - BD^{-1}C| = |A| \cdot |D - CA^{-1}B|.$$

♡

⚡**Remark** The mnemonic is: For $\begin{vmatrix} A & B \\ C & D \end{vmatrix}$, for example, if A is invertible, one factor is $|A|$, and the other factor is D (the diagonal element of A) minus the product of the other three terms arranged <u>clockwise</u>, where the middle one is the inverse matrix.

4.5 Matrix Functions

4.6 Operations of Rank

Proposition 4.2

The matrices A and B in the following operations do not need to be square matrices; they only need to be compatible for multiplication or addition.

1. Addition [a]
$$|\text{rank}(A) - \text{rank}(B)| \leqslant \text{rank}(A + B) \leqslant \text{rank}(A) + \text{rank}(B).$$

2. Multiplication
$$\text{rank}(AB) \leqslant \text{rank}(A), \quad \text{rank}(AB) \leqslant \text{rank}(B).$$

2.1. Sylvester's Inequality
$$\text{rank}(AB) \geqslant \text{rank}(A) + \text{rank}(B) - n \quad (A_{s \times n}, B_{n \times m}).$$

Specially, if $AB = O$, then:
$$\text{rank}(A) + \text{rank}(B) \leqslant n.$$

2.2. Frobenius Inequality

$$\text{rank}(ABC) \geqslant \text{rank}(AB) + \text{rank}(BC) - \text{rank}(B).$$

3. Transpose

$$\text{rank}(AA^T) = \text{rank}(A^TA) = \text{rank}(A) = \text{rank}(A^T).$$

4. Inverse

$$\text{rank}(A) = \text{rank}(A^{-1}) = n.$$

5. Block Matrix

$$\text{rank}\begin{pmatrix} A & O \\ O & D \end{pmatrix} = \text{rank}(A) + \text{rank}(D).$$

---

[a]Actually, on the basis of dimension formula (Thm 5.1), let $\mathcal{R}(A)$ and $\mathcal{R}(B)$ be the column spaces of matrices A and B respectively, then:

$$\dim(\mathcal{R}(A) + \mathcal{R}(B)) = \dim(\mathcal{R}(A)) + \dim(\mathcal{R}(B)) - \dim(\mathcal{R}(A) \cap \mathcal{R}(B)),$$

i.e.,

$$\text{rank}(A + B) = \text{rank}(A) + \text{rank}(B) - \dim(\mathcal{R}(A) \cap \mathcal{R}(B)).$$

4.7 Low–Rank Corrections

Due to all the row and column vectors of a rank-1 matrix are linearly dependent, it can be expressed as the outer product of two non-zero vectors; in other words, a rank-1 matrix can be expressed as $\alpha\beta^T$, where α and β are non-zero column vectors.

Based on the decomposition $A = \alpha\beta^T$, the matrix of rank-1 has simplified calculation rules:

🔗 **Property**

Exponentiation *For any positive integer $k \geq 1$,*

$$A^k = (\beta^T\alpha)^{k-1} \cdot A,$$

where $\beta^T\alpha$ is a constant (the inner product of vectors).

Rank Transmission *If B is any matrix, then:*

$$\text{rank}(AB) \leq 1 \quad and \quad \text{rank}(BA) \leq 1,$$

(rank 1 matrices multiplied by arbitrary matrices result in ranks not exceeding 1).

> **Theorem 4.4 (Sherman-Morrison Formula)**
>
> If $A \in \mathbb{R}^{n \times n}$ is an invertible matrix, and $\alpha, \beta \in \mathbb{R}^n$ are column vectors, then $A + \alpha\beta^T$ is invertible if and only if $1 + \beta^T A^{-1}\alpha \neq 0$. In this case, the inverse of $A + \alpha\beta^T$ is given by:
>
> $$\left(A + \alpha\beta^T\right)^{-1} = A^{-1} - \frac{A^{-1}\alpha\beta^T A^{-1}}{1 + \beta^T A^{-1}\alpha},$$
>
> where $\alpha\beta^T$ is the outer product of α and β.

🕯 **Note** *Combining the properties of determinants, we can derive the determinant version of the Sherman-Morrison formula:*

$$\left|A + \alpha\beta^T\right| = |A| \cdot \left(1 + \beta^T A^{-1}\alpha\right),$$

*which is known as the **matrix determinant lemma**.*

The theorem can also be stated in terms of the adjugate matrix of A:

$$\det(A + uv^T) = \det(A) + v^T \operatorname{adj}(A)u,$$

in which case it applies whether or not the matrix A is invertible.

Replace α, β with general matrices $U, V \in \mathbb{R}^{n \times k}$, we can extend the Sherman-Morrison formula to the Woodbury matrix identity:

Theorem 4.5 (Woodbury Matrix Identity)

If $A \in \mathbb{R}^{n \times n}$ is an invertible matrix, $C \in \mathbb{R}^{k \times k}$ is also an invertible matrix, and $U, V \in \mathbb{R}^{n \times k}$ are matrices, then $A + UCV^T$ is invertible if and only if $C^{-1} + V^T A^{-1} U$ is invertible. In this case, the inverse of $A + UCV^T$ is given by:

$$\left(A + UCV^T\right)^{-1} = A^{-1} - A^{-1}U\left(C^{-1} + V^T A^{-1} U\right)^{-1} V^T A^{-1}.$$

$\heartsuit$

Without loss of generality, let A, C be conformable identity matrices, i.e., $A = E_n$ and $C = E_k$. Then the Woodbury matrix identity simplifies to:

$$\left(E_n + UV^T\right)^{-1} = E_n - U\left(E_k + V^T U\right)^{-1} V^T.$$

This form is particularly useful in applications involving low-rank updates to identity matrices.

Chapter 5 Linear Spaces

5.1 Linear Spaces and Their Bases

¶ `Basis Transformation and Coordinate Transformation`

5.2 Subspaces

¶ `Intersection and Sum of Subspaces`

> **Definition 5.1 (Intersection and Sum of Subspaces)**
>
> Let V be a linear space over field F, and let $V_1, V_2 \subseteq V$ be two subspaces of V. The **intersection** of V_1 and V_2 is defined as:
> $$V_1 \cap V_2 = \{\alpha | \alpha \in V_1 \text{ and } \alpha \in V_2\}.$$
>
> The **sum** of V_1 and V_2 is defined as:
> $$V_1 + V_2 = \{\alpha | \alpha = \alpha_1 + \alpha_2, \alpha_1 \in V_1, \alpha_2 \in V_2\}.$$
> ♣

If V_1 and V_2 are two subspaces of V, then both their intersection $V_1 \cap V_2$ and their sum $V_1 + V_2$ are also subspaces of V. Simultaneously, $V_1 + V_2$ is the smallest subspace of V that contains both V_1 and V_2 ($V_1 \cup V_2$).

¶ `Dimension Formula`

> **Proposition 5.1**
>
> In a finite-dimensional linear space V,
> $$L(\alpha_1, \alpha_2, \ldots, \alpha_s) + L(\beta_1, \beta_2, \ldots, \beta_t) = L(\alpha_1, \alpha_2, \ldots, \alpha_s, \beta_1, \beta_2, \ldots, \beta_t).$$
> ♠

> **Theorem 5.1 (Dimension Formula)**
>
> Let V be a finite-dimensional linear space over field F, and let $V_1, V_2 \subseteq V$ be two subspaces of V. Then:
> $$\dim(V_1) + \dim(V_2) = \dim(V_1 + V_2) + \dim(V_1 \cap V_2).$$
> ♡

¶ `Direct Sum of Subspaces`

> **Definition 5.2 (Direct Sum of Subspaces)**
>
> Let V be a linear space over field F, and let $V_1, V_2 \subseteq V$ be two subspaces of V. If any vector $\alpha \in V_1 + V_2$ can be uniquely expressed as:
> $$\alpha = \alpha_1 + \alpha_2, \quad \alpha_1 \in V_1, \alpha_2 \in V_2,$$
>
> then $V_1 + V_2$ is called the **direct sum** of V_1 and V_2, denoted as $V_1 \oplus V_2$.
> ♣

> **Proposition 5.2**
>
> The necessary and sufficient condition for $V_1 + V_2$ to be a direct sum is:
> 1. $0 = 0 + 0$ (the zero vector can only be expressed as the sum of two zero vectors);
> 2. $V_1 \cap V_2 = \{0\}$ (the intersection of the two subspaces is only the zero vector);
> 3. $\dim(V_1 + V_2) = \dim(V_1) + \dim(V_2)$, or equivalently, $\dim(V_1 \cap V_2) = 0$.

> 4. Any bases of V_1 and V_2 together form a basis of $V_1 + V_2$.

5.3 Isomorphisms

5.4 Quotient Spaces

5.5 Special Linear Spaces

¶ Commutative Spaces

Let A is an n-order matrix over number field P. It is obvious that the set of all matrices satisfying $AX = XA$ is a subspace of $M_{n \times n}(P)$, denoted as $C(A)$.

⌐ **Property**

1. *For any n-order matrix A, B, if A is similar to B ($A \sim B$), then $C(A)$ is isomorphic to $C(B)$ ($C(A) \cong C(B)$), further $\dim(C(A)) = \dim(C(B))$.*

Example 5.1 Find the dimension and a basis of the $C(A)$ where:

(i) A is $3-$order matrix over number field P, and A has three distinct eigenvalues.

(ii)

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

Chapter 6 Linear Mappings

6.1 Linear Mappings and Their Computation

¶ **Definition of Linear Mappings**

> **Definition 6.1 (Linear Mapping)**
>
> Let V and V' be linear spaces over field F. A mapping $\mathcal{A} : V \to V'$ is called a **linear mapping** if:
> $$\mathcal{A}(\alpha + \beta) = \mathcal{A}\alpha + \mathcal{A}\beta, \quad \forall \alpha, \beta \in V,$$
> and
> $$\mathcal{A}(k\alpha) = k\mathcal{A}\alpha, \quad \forall \alpha \in V, k \in F.$$
>
> If $V = V'$, then $\mathcal{A}$ is called a **linear transformation** of V. The linear mapping from V over field F to F is called a **linear function** on V. ♣

🔖**Remark** The field naturally constitutes a one-dimensional linear space over itself, so a linear function is a special case of a linear mapping.

¶ **Existence and Uniqueness of Linear Mappings**

> **Theorem 6.1 (Existence and Uniqueness of Linear Mappings (Linear Extension Theorem))**
>
> Let V and V' be linear spaces over field F, and let $\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\}$ be a basis of V. For any set of vectors $\{\alpha_1, \alpha_2, \ldots, \alpha_n\} \subseteq V'$, there exists a unique linear mapping
> $$\mathcal{A} : V \to V', \quad \alpha = \sum_{i=1}^{n} x_i \varepsilon_i \mapsto \mathcal{A}(\alpha) = \sum_{i=1}^{n} x_i \alpha_i$$
> such that:
> $$\mathcal{A}(\varepsilon_i) = \alpha_i, \quad i = 1, 2, \ldots, n.$$
> ♡

¶ **Operations of Linear Mappings**

In the following, all linear spaces are over the field F.

Addition Let $\mathcal{A}, \mathcal{B} \in \mathrm{Hom}(V, V')$. The sum of $\mathcal{A}$ and $\mathcal{B}$, denoted $\mathcal{A} + \mathcal{B}$, is defined by:
$$(\mathcal{A} + \mathcal{B})(\alpha) = \mathcal{A}(\alpha) + \mathcal{B}(\alpha), \quad \forall \alpha \in V.$$

The sum of linear maps is still a linear map, i.e., $\mathcal{A} + \mathcal{B} \in \mathrm{Hom}(V, V')$.

1. **Commutativity:** $\mathcal{A} + \mathcal{B} = \mathcal{B} + \mathcal{A}$,
2. **Associativity:** $(\mathcal{A} + \mathcal{B}) + \mathcal{C} = \mathcal{A} + (\mathcal{B} + \mathcal{C})$,
3. **Additive Identity:** The zero map $\prime$ satisfies $\mathcal{A} + \prime = \mathcal{A}$,
4. **Additive Inverse:** For every $\mathcal{A}$, the additive inverse $-\mathcal{A}$ satisfies $\mathcal{A} + (-\mathcal{A}) = \prime$.

Scalar Multiplication Let $k \in F$ and $\mathcal{A} \in \mathrm{Hom}(V, V')$. The scalar product $k\mathcal{A}$ is defined by:
$$(k\mathcal{A})(\alpha) = k\mathcal{A}(\alpha), \quad \forall \alpha \in V.$$

Scalar multiplication of a linear map is still a linear map, i.e., $k\mathcal{A} \in \mathrm{Hom}(V, V')$.

1. **Unit Identity:** $1 \cdot \mathcal{A} = \mathcal{A}$,
2. **Associativity of Scalars:** $(kl)\mathcal{A} = k(l\mathcal{A})$,

3. **Distributive Properties:**

$$k(\mathcal{A} + \mathcal{B}) = k\mathcal{A} + k\mathcal{B},$$

$$(k + l)\mathcal{A} = k\mathcal{A} + l\mathcal{A}.$$

Multiplication (Composition of Linear Maps) Let $\mathcal{A} \in \mathrm{Hom}(V', W)$ and $\mathcal{B} \in \mathrm{Hom}(U, V')$. The product (composition) of $\mathcal{A}$ and $\mathcal{B}$, denoted $\mathcal{A}\mathcal{B}$, is defined by:

$$(\mathcal{A}\mathcal{B})(\alpha) = \mathcal{A}(\mathcal{B}(\alpha)), \quad \forall \alpha \in U.$$

The composition of linear maps is still a linear map, i.e., $\mathcal{A}\mathcal{B} \in \mathrm{Hom}(U, W)$.

1. **Associativity of Composition:** $(\mathcal{A}\mathcal{B})\mathcal{C} = \mathcal{A}(\mathcal{B}\mathcal{C})$,

2. **Distributive Properties:**

$$\mathcal{A}(\mathcal{B} + \mathcal{C}) = \mathcal{A}\mathcal{B} + \mathcal{A}\mathcal{C},$$

$$(\mathcal{A} + \mathcal{B})\mathcal{C} = \mathcal{A}\mathcal{C} + \mathcal{B}\mathcal{C},$$

3. **Non-Commutativity:** In general, $\mathcal{A}\mathcal{B} \neq \mathcal{B}\mathcal{A}$,

4. **Identity Transformation:** The identity transformation $\mathcal{E}$ satisfies:

$$\mathcal{E}\mathcal{A} = \mathcal{A}\mathcal{E} = \mathcal{A}, \quad \forall \mathcal{A} \in \mathrm{Hom}(V).$$

Let V and V' be vector spaces over the field F.

1. The collection of all linear mappings from V to V', denoted by $\mathrm{Hom}(V, V')$, under the addition and scalar multiplication defined above, forms a linear space over the field F.

2. The collection of all linear transformations on V, denoted by $\mathrm{Hom}(V)$, under the addition and composition of maps defined above, forms a unital ring. In this case, it is referred to as the **endomorphism ring** and is denoted by $\mathrm{End}(V)$.

> **Definition 6.2 (Algebra over a Field)**
>
> Let V be a linear space over the field F equipped with an additional binary operation from $V \times V$ to V, denoted by $\cdot$. Then V is called an **algebra over the field** F if (forall $\alpha, \beta, \gamma \in V$ and $k, l \in F$):
>
> **Right Distributivity** $(\alpha + \beta) \cdot \gamma = \alpha \cdot \gamma + \beta \cdot \gamma$;
> **Left Distributivity** $\gamma \cdot (\alpha + \beta) = \gamma \cdot \alpha + \gamma \cdot \beta$;
> **Compatibility with Scalar Multiplication** $(k\alpha) \cdot (l\beta) = kl(\alpha \cdot \beta)$. ♣

⚡**Remark** The left and right distributivity conditions can be combined into a single condition: V forms a unital ring under the addition and scalar multiplication.

It is evident that $\mathrm{Hom}(V)$ is an algebra over the field F.

6.2 Kernel and Image of Linear Mappings

> **Definition 6.3 (Kernel and Image of Linear Mappings)**
>
> Let V and V' be linear spaces over field F, and let $\mathcal{A} : V \to V'$ be a linear mapping. The **kernel** of $\mathcal{A}$ is defined as:
>
> $$\mathrm{Ker}(\mathcal{A}) = \{\alpha | \alpha \in V, \mathcal{A}(\alpha) = 0\}.$$

The **image** of $\mathcal{A}$ is defined as:

$$\mathrm{Im}(\mathcal{A}) = \{\beta | \beta = \mathcal{A}(\alpha), \alpha \in V, \beta \in V'\}.$$

They can be also denoted as $\mathcal{A}^{-1}(0)$ and $\mathcal{A}(V)$ respectively.

The rank of $\mathrm{Im}(\mathcal{A})$ is called the **rank** of $\mathcal{A}$; the dimension of $\mathrm{Ker}\,\mathcal{A}$ is called the **nullity** of $\mathcal{A}$.

✎ Property

> **Theorem 6.2 (Rank-Nullity Theorem)**
>
> Let V and V' be finite-dimensional linear spaces over field F, and let $\mathcal{A} : V \to V'$ be a linear mapping. Then:
>
> $$\dim(\mathrm{Ker}\,\mathcal{A}) + \dim(\mathrm{Im}\,\mathcal{A}) = \dim(V).$$

> **Definition 6.4 (Cokernel)**
>
> Let V and V' be linear spaces over field F, and let $\mathcal{A} : V \to V'$ be a linear mapping. The **cokernel** of $\mathcal{A}$ is defined as the quotient space:
>
> $$\mathrm{Coker}(\mathcal{A}) = V'/\mathrm{Im}(\mathcal{A}),$$
>
> denoted as $\mathrm{Coker}(\mathcal{A})$.

¶ Ascending Chain of Subspaces

> **Definition 6.5 (Ascending Chain of Subspaces)**
>
> Let V be a finite-dimensional linear space over field F, and let $\mathcal{A} \in \mathrm{Hom}(V)$. For $k = 1, 2, 3, \cdots$, define:
>
> $$V_k = \mathrm{Ker}(\mathcal{A}^k) = \{\alpha \in V | \mathcal{A}^k(\alpha) = 0\},$$
>
> then the sequence of subspaces:
>
> $$V_1 \subseteq V_2 \subseteq \cdots \subseteq V_k \subseteq \cdots$$
>
> is called the **ascending chain of subspaces** induced by $\mathcal{A}$.

✎ Property

1. *The dimensions of these subspaces are non-decreasing:*

   $$\dim(V_1) \leq \dim(V_2) \leq \cdots \leq \dim(V_k) \leq \cdots$$

2. *Since V is finite-dimensional, there exists a positive integer k_0 such that:*

   $$\dim(V_{k_0}) = \dim(V_{k_0+1}) = \cdots$$

   Consequently, for all $k \geq k_0$:

   $$V_k = V_{k_0},$$

   that is, the ascending chain stabilizes at V_{k_0}.

3. *This k_0 is precisely a stability index for the corresponding linear operator (for example, an upper bound on the degree of its annihilating polynomial or a restriction on the size of its Jordan blocks).*

⚡**Remark** In some cases, the stability index k_0 can be equal to the dimension of the space V. For instance, $\mathcal{A}$ is a nilpotent or all eigenvalues of $\mathcal{A}$ are zero.

6.3 Matrix Representation of Linear Mappings

Let V and V' be finite-dimensional linear spaces over field F, $\dim(V) = n$ and $\dim(V') = m$, and let $\mathcal{A} : V \to V'$ be a linear mapping. Let $\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\}$ be a basis of V, and $\{\eta_1, \eta_2, \ldots, \eta_m\}$ be a basis of V'. Then the image of each basis vector of V under $\mathcal{A}$ can be expressed as the basis vectors of V':

$$\begin{cases} \mathcal{A}(\varepsilon_1) = a_{11}\eta_1 + a_{21}\eta_2 + \cdots + a_{m1}\eta_m, \\ \mathcal{A}(\varepsilon_i) = a_{12}\eta_1 + a_{22}\eta_2 + \cdots + a_{m2}\eta_m, \\ \qquad\qquad\qquad \vdots \\ \mathcal{A}(\varepsilon_n) = a_{1n}\eta_1 + a_{2n}\eta_2 + \cdots + a_{mn}\eta_m. \end{cases}$$

It can be expressed in matrix form as:

$$\mathcal{A}(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n) = (\mathcal{A}\varepsilon_1, \mathcal{A}\varepsilon_2, \ldots, \mathcal{A}\varepsilon_n) = (\eta_1, \eta_2, \ldots, \eta_m)A,$$

where

$$A = (a_{ij})_{m \times n} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

The matrix A is called the **matrix representation** of the linear mapping $\mathcal{A}$ under the bases $\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\}$ of V and $\{\eta_1, \eta_2, \ldots, \eta_m\}$ of V'.

Example 6.1 Let $A, B \in M_{s \times n}(F)$. Prove that homogeneous linear systems with n variables $AX = 0$ and $BX = 0$ have the same solution space if and only if there exists an invertible matrix $P \in M_{s \times s}(F)$ such that $B = PA$.

✒ *Proof*

∎

6.4 Special Linear Transformations

¶ `Common Linear Transformations`

Some common linear transformations include:

Identity Transformation The identity transformation $\mathcal{E}$ on V is defined by:

$$\mathcal{E}(\alpha) = \alpha, \quad \forall \alpha \in V.$$

Zero Transformation The zero transformation $\mathcal{O}$ from V to V' is defined by:

$$\mathcal{O}(\alpha) = 0, \quad \forall \alpha \in V.$$

Scalar Transformation For a fixed scalar $k \in F$, the scalar transformation $\mathcal{K}$ on V is defined by:

$$\mathcal{K}(\alpha) = k\alpha, \quad \forall \alpha \in V.$$

When $k = 1$, it is the identity transformation; when $k = 0$, it is the zero transformation.

Spin Axis Transformation The vectors on a plane form a two-dimensional linear space over real number field $\mathbb{R}$. Rotate the plane counterclockwise by an angle θ around the origin, then the transformation of any vector on the plane is a linear transformation, denoted by $\mathcal{P}_\theta$. Similarly, in three-dimensional space,

rotating around a fixed axis by an angle θ also defines a linear transformation.

¶ **Idempotent and Projection Transformations**

> **Definition 6.6 (Idempotent Transformation)**
>
> A linear transformation $\mathcal{A}$ on V is called an **idempotent transformation** if:
> $$\mathcal{A}^2 = \mathcal{A}.$$
> ♣

Two linear transformations $\mathcal{A}$ and $\mathcal{B}$ on V are called **orthogonal** if:
$$\mathcal{A}\mathcal{B} = \mathcal{B}\mathcal{A} = \mathcal{O}.$$

With the definitions above, we can define the projection transformation as follows:

> **Definition 6.7 (Projection Transformation)**
>
> Let V be a linear space over field F, and let $U, W \subseteq V$ be two subspaces of V such that $V = U \oplus W$. For all $\alpha \in V$, there exist unique $\alpha_1 \in U$ and $\alpha_2 \in W$ such that:
> $$\alpha = \alpha_1 + \alpha_2.$$
> Let
> $$\mathcal{P}_U : \begin{array}{c} V \to V \\ \alpha \mapsto \alpha_1 \end{array}$$
> then $\mathcal{P}_U$ is called the **projection** onto U along W, which is a linear transformation on V. Similarly, the projection onto W along U can be defined.
> ♣

$\mathcal{P}_U$ satisfies and uniquely satisfies:
$$\mathcal{P}_U(\alpha) = \begin{cases} \alpha, & \alpha \in U, \\ 0, & \alpha \in W. \end{cases}$$

In fact, the projection transformation is equivalent to the idempotent transformation.

> **Proposition 6.1**
>
> Let V be a linear space over field F.
> 1. Let $U, W \subseteq V$ be two subspaces of V such that $V = U \oplus W$. Then the projection $\mathcal{P}_U$ onto U along W and the projection $\mathcal{P}_W$ onto W along U are orthogonal idempotent transformations, and their sum is the identity transformation:
> $$\mathcal{P}_U \mathcal{P}_W = \mathcal{P}_W \mathcal{P}_U = \mathcal{O}, \quad \mathcal{P}_U^2 = \mathcal{P}_U, \quad \mathcal{P}_W^2 = \mathcal{P}_W, \quad \mathcal{P}_U + \mathcal{P}_W = \mathcal{E}.$$
> 2. Let $\mathcal{A}, \mathcal{B} \in \mathrm{Hom}(V)$ be two linear transformations. If $\mathcal{A}$ and $\mathcal{B}$ are orthogonal idempotent transformations satisfying $\mathcal{A} + \mathcal{B} = \mathcal{E}$, then $V = \mathrm{Im}(\mathcal{A}) \oplus \mathrm{Im}(\mathcal{B})$, and
> $$\mathcal{A} = \mathcal{P}_{\mathrm{Im}(\mathcal{B})}\left(\mathrm{Im}(\mathcal{A})\right), \quad \mathcal{B} = \mathcal{P}_{\mathrm{Im}(\mathcal{A})}\left(\mathrm{Im}(\mathcal{B})\right).$$
> ♠

¶ **Other Propositions about Special Linear Transformations**

> **Proposition 6.2**
>
> Let V be a linear space over field F, and let $\mathcal{A} \in \mathrm{Hom}(V)$.
> 1. $\mathcal{A}$ is an idempotent transformation if and only if
> $$\mathrm{rank}(\mathcal{A}) + \mathrm{rank}(\mathcal{E} - \mathcal{A}) = \dim(V).$$

2. $\mathcal{A}$ is an involutional transformation if and only if

$$\text{rank}(\mathcal{A}) + \text{rank}(\mathcal{E} - \mathcal{A}) = \dim(V).$$

Proposition 6.3

Let V be a linear space over field F, and let $\mathcal{A} \in \text{Hom}(V)$. Then the following statements are equivalent:

1. $\mathcal{A}$ is invertible;
2. if $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ are a basis of V, then $\mathcal{A}\varepsilon_1, \mathcal{A}\varepsilon_2, \cdots, \mathcal{A}\varepsilon_n$ are also a basis of V;
3. if $\alpha \neq 0$, then $\mathcal{A}\alpha \neq 0$;
4. if $V = V_1 \oplus V_2$, then $V = \mathcal{A}V_1 \oplus \mathcal{A}V_2$;

6.5 Linear Functions and Dual Spaces

Chapter 7 Diagonalization

7.1 Similarity of Matrices

> **Definition 7.1 (Similar Matrices)**
>
> Let $A, B \in F^{n \times n}$ be two square matrices over field F. If there exists an invertible matrix $P \in F^{n \times n}$ such that:
> $$B = P^{-1}AP,$$
> then A is said to be **similar** to B, denoted as $A \sim B$, and B is called a **similarity transformation** of A. ♣

> **Definition 7.2 (Trace)**
>
> Let $A = (a_{ij})_{n \times n} \in F^{n \times n}$ be a square matrix over field F. The sum of the diagonal elements of A:
> $$\mathrm{tr}(A) = a_{11} + a_{22} + \cdots + a_{nn},$$
> is called the **trace** of A. ♣

Obviously, similarity is an equivalence relation on the set of square matrices of the same order, i.e, it satisfies reflexivity, symmetry, and transitivity.

It is easy to verify that the arithmetic properties of traces hold, for any $A, B \in F^{n \times n}, k \in F$:

$$\mathrm{tr}(A + B) = \mathrm{tr}(A) + \mathrm{tr}(B),$$
$$\mathrm{tr}(kA) = k\,\mathrm{tr}(A),$$
$$\mathrm{tr}(AB) = \mathrm{tr}(BA).$$

From the third property, it can be derived that the trace of an $n \times n$ matrix is a commutative quantity extracted from the non-commutativity of matrix multiplication.

In the meanwhile, we have the arithmetic properties of similarity, if $B_1 = P^{-1}A_1P, B_2 = P^{-1}A_2P$:

$$B_1 + B_2 = P^{-1}(A_1 + A_2)P,$$
$$B_1 B_2 = P^{-1}(A_1 A_2)P,$$
$$B_1^k = P^{-1}A_1^k P, \quad k = 1, 2, \ldots$$

The third property is called the **similarity invariance of matrix powers**. From it, we can derive some important applications:

Eigenvalue invariance Matrix similarity guarantees the preservation of eigenvalues. Therefore, if A and B are similar matrices, then the eigenvalues of A^k and B^k are also identical. For any eigenvalue λ: If λ is an eigenvalue of A, then λ^k is an eigenvalue of A^k (the power corresponds to the operation).

The result of diagonalization is invariant If a matrix A is diagonalizable, that is, there exists an invertible matrix P and a diagonal matrix D such that $A = P^{-1}DP$, then: $A^k = P^{-1}D^kP$. In other words, calculating A^k after diagonalization becomes simple. You only need to perform the power operation on the diagonal matrix D, and then transform the result back to the original basis by a similarity transformation.

The Convenience of Jordan Form Even if a matrix A is not diagonalizable, it can still be similar to a Jordan form matrix J. In this case: $A^k = P^{-1}J^kP$, where the calculation of J^k can still be done block by block on the Jordan blocks. This property is extremely important when dealing with powers of non-

diagonalizable matrices.

Generalization of Matrix Functions The power relations of matrices can be used to derive operations of matrix functions. For example, the exponential function of a matrix is defined as $e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!}$. If $B = P^{-1}AP$, then: $e^B = P^{-1}e^A P$. Similarly, other matrix functions (such as logarithm, power functions, etc.) also possess similar properties.

In summary, the <u>determinant</u>, <u>rank</u>, and <u>trace</u> of a matrix are invariants[1] under similarity relations, and are collectively referred to as **similarity invariants**.

7.2 Eigenvectors and Eigenvalues

> **Definition 7.3 (Eigenvalues and Eigenvectors)**
>
> Let V be a linear space over field F, and let $\mathcal{A} \in \mathrm{Hom}(V)$. A non-zero vector $\xi \in V$ is called an **eigenvector** of $\mathcal{A}$ if there exists a scalar $\lambda \in F$ such that:
>
> $$\mathcal{A}(\xi) = \lambda\xi.$$
>
> The scalar λ is called the **eigenvalue** corresponding to the eigenvector ξ.
>
> From the perspective of matrices, let $A \in F^{n \times n}$ be a square matrix over field F. A non-zero vector $x \in F^n$ is called an **eigenvector** of A if there exists a scalar $\lambda \in F$ such that:
>
> $$Ax = \lambda x.$$
>
> The scalar λ is called the **eigenvalue** corresponding to the eigenvector x. ♣

According to the definition of eigenvalues and eigenvectors, if λ is an eigenvalue of A, ξ is an eigenvector corresponding to λ, then eigenvalue of some specific functions of A can be derived as follows (eigenvector remains the same):

- $A + kE$: eigenvalue is $\lambda + k$;
- kA: eigenvalue is $k\lambda$;
- $A^m, m \in \mathbb{N}^*$: eigenvalue is λ^m;
- A^{-1}: eigenvalue is $\frac{1}{\lambda}$;
- $f(A)$, where $f(x)$ is a polynomial: eigenvalue is $f(\lambda)$.

> **Definition 7.4 (Eigenpolynomial)**
>
> Let V be a linear space over field F, and let $\mathcal{A} \in \mathrm{Hom}(V)$. The polynomial
>
> $$f(\lambda) = |\lambda E - A| = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \cdots & \lambda - a_{nn} \end{vmatrix}$$
>
> is called the **eigenpolynomial** (or characteristic polynomial) of A, where $A = (a_{ij})_{n \times n}$ is the matrix representation of $\mathcal{A}$ under some basis of V, and E is the identity matrix. ♣

⚡**Remark** For example, in $P[x]_n \quad (n > 1)$, the eigenpolynomial of linear transformation $\mathcal{D}$ (derivation) is

---

[1]And also include the <u>characteristic polynomial</u>, <u>eigenvalues</u>, and <u>eigenvectors</u> in next section.

$f(\lambda) = \lambda^n$.

The roots of the eigenpolynomial $|\lambda E - A|$ correspond to the eigenvalues of A; the non-zero solutions to the homogeneous system of equations $(\lambda_0 E - A)x = 0$ correspond to the eigenvectors belonging to the eigenvalue λ_0. It is easy to verify that the eigenpolynomial as well as the eigenvalues, eigenvectors is invariant under similarity relations.

The following proposition summarizes the properties of the eigenpolynomial:

> **Proposition 7.1**
>
> Let $A \in F^{n \times n}$, the eigenpolynomial of A is $f(\lambda) = |\lambda E - A|$. Then $f(\lambda)$ is a monic polynomial of degree n:
> $$f(\lambda) = \lambda^n - \mathrm{tr}(A)\lambda^{n-1} + \cdots + (-1)^n |A|,$$
> where the coefficient of $\lambda^{n-k} \quad (1 \leqslant k < n)$ is the sum of all k-order principal minors of A multiplied by $(-1)^k$. ♠

Note *Combine Vièta's formulas with the proposition above, we can derive that:*

1. *The sum of the eigenvalues (counting multiplicities) is equal to the trace of the matrix:*
$$\sum \lambda_i = \mathrm{tr}(A).$$

2. *The product of the eigenvalues (counting multiplicities) is equal to the determinant of the matrix:*
$$\prod \lambda_i = \det A.$$

¶ **Multiplicity and Eigenspaces**

> **Definition 7.5 (Eigenspace)**
>
> Let V be a linear space over field F, and let $\mathcal{A} \in \mathrm{Hom}(V)$. For an eigenvalue λ of $\mathcal{A}$, the set of all eigenvectors corresponding to λ along with the zero vector:
> $$V_\lambda = \{\alpha | \alpha \in V, \mathcal{A}(\alpha) = \lambda\alpha\} = \mathrm{Ker}(\mathcal{A} - \lambda\mathcal{E}),$$
> is called the **eigenspace** (or eigenspace) corresponding to the eigenvalue λ, denoted as E_λ. It is evident that V_λ is a subspace of V and solution space of the equation system $(\lambda E - \mathcal{A})X = 0$. ♣

> **Definition 7.6 (Geometric and Algebraic Multiplicity of Eigenvalues)**
>
> Let V be a linear space over field F, and let $\mathcal{A} \in \mathrm{Hom}(V)$. For an eigenvalue λ of $\mathcal{A}$:
> - The dimension of the eigenspace $V_\lambda = \mathrm{Ker}(\mathcal{A} - \lambda\mathcal{E})$ is called the **geometric multiplicity** of the eigenvalue λ.
> - The multiplicity of λ as a root of the eigenpolynomial $f(\lambda) = |\lambda E - A|$ is called the **algebraic multiplicity** of the eigenvalue λ. ♣

Property *For an eigenvalue λ of $\mathcal{A}$, let $g(\lambda)$ and $a(\lambda)$ be the geometric and algebraic multiplicities of λ, respectively. Then:*
$$1 \leqslant g(\lambda) \leqslant a(\lambda) \leqslant n.$$

¶ **Estimation of Eigenvalues**

Theorem 7.1 (Gerschgorin Disks First Theorem)

Let $A = (a_{ij})_{n \times n}$ be a square matrix over complex number field C. Then all eigenvalues of A are located within the following circular disks in the complex plane:

$$|z - a_{ii}| \le R_i, \quad 1 \le i \le n,$$

where $R_i = |a_{i1}| + \cdots + |a_{i,i-1}| + |a_{i,i+1}| + \cdots + |a_{in}|$.

Theorem 7.2 (Gerschgorin Disks Second Theorem)

If the n Gerschgorin disks of the n-order matrix A form k connected regions, and one of the connected regions contains k Gerschgorin disks, then there are exactly k eigenvalues located in this connected region (if the eigenvalues are repeated, the count includes multiplicities).

7.3 Minimal Polynomial

Definition 7.7 (Minimal Polynomial)

Let V be a linear space over field F, and let $\mathcal{A} \in \mathrm{Hom}(V)$. A polynomial $f(\lambda)$ is called an **annihilating polynomial** of $\mathcal{A}$ if:

$$f(\mathcal{A}) = 0.$$

The monic polynomial $m(\lambda)$ of least degree such that:

$$m(\mathcal{A}) = 0,$$

is called the **minimal polynomial** of $\mathcal{A}$.
It is similar for a square matrix $A \in F^{n \times n}$.

7.4 Space Decomposition

Studying linear transformations through invariant subspaces is a commonly used method, as it reduces problems on the whole space to problems on lower-dimensional subspaces.

¶ **Invariant Subspace**

Definition 7.8 (Invariant Subspace)

Let V be a linear space over field F, and let $\mathcal{A} \in \mathrm{Hom}(V)$. A subspace $W \subseteq V$ is called an **invariant subspace** of V under the linear transformation $\mathcal{A}$ if:

$$\mathcal{A}(\alpha) \subseteq W, \quad \forall \alpha \in W,$$

simplified as $\mathcal{A}$-subspace.

⚡**Remark** Kernel and image of $\mathcal{A}$, as well as eigenspaces (including generalized eigenspaces), cyclic subspaces, are both invariant subspaces of V under $\mathcal{A}$.

📎 **Property**

1. *The zero subspace $\{0\}$ and the entire space V are invariant subspaces of V under any linear transformation $\mathcal{A}$, which are called trivial invariant subspaces.*

2. Let $\mathcal{A}, \mathcal{B} \in \mathrm{Hom}(V)$. If $\mathcal{A}, \mathcal{B}$ is commutative ($\mathcal{A}\mathcal{B} = \mathcal{B}\mathcal{A}$), then $\mathrm{Ker}(\mathcal{A})$, $\mathrm{Im}(\mathcal{A})$ and eigenspace of $\mathcal{A}$ are both invariant subspaces of V under $\mathcal{B}$.

3. Any subspace of scalar transformation is an invariant subspace.

4. The intersection and sum of any two invariant subspaces are also invariant subspaces.

Proposition 7.2

Let V be a linear space over field F, and let $\mathcal{A} \in \mathrm{Hom}(V)$. $W = \langle \xi \rangle$ $(\xi \neq 0)$, the subspace of V is $\mathcal{A}$-subspace if and only if ξ is an eigenvector of $\mathcal{A}$.

♠

Theorem 7.3

Let V be a n-dimensional linear space over field F, and let $\mathcal{A} \in \mathrm{Hom}(V)$.

1. Let W be a nontrivial $\mathcal{A}$-subspace, take a basis $\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_r\}$ of W, and extend it to a basis $\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\}$ of V. Then the matrix representation of $\mathcal{A}$ under this basis takes the block upper triangular form:

$$\begin{pmatrix} A & B \\ O & C \end{pmatrix}$$

where:
   - A is the matrix representation of $\mathcal{A}$ restricted to W (i.e., $\mathcal{A}|_W$) under the basis $\varepsilon_1, \ldots, \varepsilon_r$,
   - C is the matrix representation of $\widetilde{\mathcal{A}}$, the linear transformation on the quotient space V/W induced by $\mathcal{A}$, under the basis $\varepsilon_{r+1} + W, \ldots, \varepsilon_n + W$.

   Let the characteristic polynomials of $\mathcal{A}$, $\mathcal{A}|_W$, and $\widetilde{\mathcal{A}}$ be $f(\lambda)$, $f_1(\lambda)$, and $f_2(\lambda)$, respectively. Then:

$$f(\lambda) = f_1(\lambda) f_2(\lambda).$$

2. Let the matrix representation of $\mathcal{A}$ under a basis $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_r, \varepsilon_{r+1}, \ldots, \varepsilon_n$ of V take the block upper triangular form:

$$\mathcal{A} = \begin{pmatrix} A & B \\ O & C \end{pmatrix}.$$

   Define $W = \langle \varepsilon_1, \ldots, \varepsilon_r \rangle$. Then W is an $\mathcal{A}$-subspace, and the matrix representation of $\mathcal{A}|_W$ under the basis $\varepsilon_1, \ldots, \varepsilon_r$ is given by A.

3. The matrix representation of $\mathcal{A}$ under a basis of V is a block diagonal form

$$\begin{pmatrix} A_1 & O & \cdots & O \\ O & A_2 & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & A_s \end{pmatrix}$$

   if and only if V can be decomposed into a direct sum of nontrivial $\mathcal{A}$-subspaces (W_i):

$$V = W_1 \oplus W_2 \oplus \cdots \oplus W_s,$$

   then A_i represents the matrix form of $\mathcal{A}|_{W_i}$ under a basis of W_i.

♡

Proposition 7.3

Let V be a linear space over field F (finite or infinite dimensional), and let $\mathcal{A} \in \mathrm{Hom}(V)$. On $F[x]$,

$$f(x) = f_1(x) f_2(x) \cdots f_s(x),$$

where $f_i(x)$ are pairwise coprime polynomials. Then

$$\operatorname{Ker} f(\mathcal{A}) = \operatorname{Ker} f_1(\mathcal{A}) \oplus \operatorname{Ker} f_2(\mathcal{A}) \oplus \cdots \oplus \operatorname{Ker} f_s(\mathcal{A}).$$

♠

¶ **Hamilton-Cayley Theorem**

Theorem 7.4 (Hamilton-Cayley Theorem)

Let V be an n-dimensional linear space over field F, and let $\mathcal{A} \in \operatorname{Hom}(V)$. Let A be the matrix representation of $\mathcal{A}$ under some basis of V, and let $f(\lambda) = |\lambda E - A|$ be the characteristic polynomial of A. Then:

$$f(\mathcal{A}) = 0.$$

♡

🖋 **Note** *The Hamilton-Cayley theorem states that characteristic polynomial of a linear transformation annihilates the transformation itself.*

In fact, characteristic polynomial is a special case of annihilating polynomial, and it is one of the highest degree annihilating polynomials. It can be divided exactly by the minimal polynomial of $\mathcal{A}$.

Definition 7.9 (Root Subspaces)

Let V be a linear space over field F, and let $\mathcal{A} \in \operatorname{Hom}(V)$. For an eigenvalue λ of $\mathcal{A}$, the set

$$V_\lambda^{(r)} = \operatorname{Ker}(\mathcal{A} - \lambda\mathcal{E})^r = \{\alpha | \alpha \in V, (\mathcal{A} - \lambda\mathcal{E})^r(\alpha) = 0\},$$

is called the **root subspace** (or generalized eigenspace) of order r^a corresponding to the eigenvalue λ.

---

^{a}Here, r has two kinds of equivalent definitions:
1. r is the smallest positive integer such that $(\mathcal{A} - \lambda\mathcal{E})^r(\alpha) = 0$;
2. r is the algebraic multiplicity of the eigenvalue λ.

All vectors in the root subspace are annihilated by the operator raised to the r-th power, and the minimal such r is precisely the algebraic multiplicity.

♣

Similarly, generalized kernel and image can be defined.

Theorem 7.5 (Root Space Decomposition Theorem)

Let V be an n-dimensional linear space over field F, and let $\mathcal{A} \in \operatorname{Hom}(V)$. If the characteristic polynomial of $\mathcal{A}$ can be factored into linear factors over F:

$$f(\lambda) = (\lambda - \lambda_1)^{r_1}(\lambda - \lambda_2)^{r_2} \cdots (\lambda - \lambda_k)^{r_k},$$

then V can be decomposed into a direct sum of root subspaces:

$$V = V_{\lambda_1}^{(r_1)} \oplus V_{\lambda_2}^{(r_2)} \oplus \cdots \oplus V_{\lambda_k}^{(r_k)}.$$

♡

7.5 Diagonalization

Lemma 7.1

Eigenvectors corresponding to distinct eigenvalues are linearly independent.

Furthermore, if a linear transformation $\mathcal{A}$ on a linear space V has k distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$, and $\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{ir_i}$ $(i = 1, 2, \ldots, k)$ are the corresponding linearly independent eigenvectors of λ_i, then the set

of vectors

$$\{\alpha_{ij}|i = 1, 2, \ldots, k, j = 1, 2, \ldots, r_i\}$$

is linearly independent.

Theorem 7.6

Let V be an n-dimensional linear space over field F, and let $\mathcal{A} \in \mathrm{Hom}(V)$. Then the following statements are equivalent:

1. $\mathcal{A}$ is diagonalizable, i.e., there exists a basis of V such that the matrix representation of $\mathcal{A}$ under this basis is a diagonal matrix;

2. There are n linearly independent eigenvectors $(\xi_1, \xi_2, \cdots, \xi_n)$ of $\mathcal{A}$;

3. For each eigenvalue of $\mathcal{A}$, the geometric multiplicity equals the algebraic multiplicity;[a]

4. $V = V_{\lambda_1} \oplus V_{\lambda_2} \oplus \cdots \oplus V_{\lambda_k}$, where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are the <u>distinct</u> eigenvalues of $\mathcal{A}$.

5. The minimal polynomial[b] of $\mathcal{A}$ has not repeated roots, i.e. it can be factored into <u>distinct</u> linear factors over field F:

$$m(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_k).$$

---

[a] This condition is often called for any eigenvalue of $\mathcal{A}$, it has **complete set of eigenvectors**. It also can be restated as: the sum of the dimensions of all eigenspaces equals n.

[b] Since the minimal polynomial is the monic polynomial with least degree of all annihilating polynomials of $\mathcal{A}$, the condition also holds for any annihilating polynomial of $\mathcal{A}$.

7.6 Diagonalization of Special Matrices

In this section, we will discuss the diagonalization of several special types of matrices. Let V be a linear space over field F (without specific statements, it can be <u>infinite</u> dimensional), and let $\mathcal{A} \in \mathrm{Hom}(V)$. A is the matrix representation of $\mathcal{A}$ under some basis of V.

Since the isomorphism between linear transformations and square matrices, the following special matrices can also be understood as special linear transformations.

⚡**Remark** Some properties mentioned below require V to be finite dimensional or F to be a number field (commonly used are real number field $\mathbb{R}$ or complex number field $\mathbb{C}$).

In fact, in some cases, the constraints on F (being a number field) can be relaxed to the characteristic of F is not 2.

¶ **Invertible Matrix**

Let A be an invertible matrix.

1. 0 is not an eigenvalue of A; the converse proposition is also true.

2. If λ is an eigenvalue of A, then $\frac{1}{\lambda}$ is an eigenvalue of A^{-1}; and if ξ is an eigenvector of A corresponding to λ, then ξ is also an eigenvector of A^{-1} corresponding to $\frac{1}{\lambda}$.

3. (When F is number field) If A is diagonalizable, then A^{-1} is also diagonalizable.

¶ **Idempotent Matrix**

Let A be an idempotent matrix, i.e., $A^2 = A$.

1. Any eigenvalue of A is either 0 or 1.

2. (V is finite dimensional) A is diagonalizable, and

$$V = E_1 \oplus E_0 = \mathrm{Ker}(A - E) \oplus \mathrm{Ker}(A - 0 \cdot E) = \mathrm{Im}(A) \oplus \mathrm{Ker}(A)$$

3. (When F is number field, and V is finite dimensional) The rank of A equals the trace of A:

$$\mathrm{rank}(A) = \mathrm{tr}(A).$$

¶ **Nilpotent Matrix**

Let A be a nilpotent matrix, i.e., there exists a positive integer k such that $A^k = 0$. The smallest such k is called the index of nilpotency of A.

1. All eigenvalues of A are 0.
2. (V is finite dimensional) A is diagonalizable if and only if $A = 0$, i.e., $k = 1$.

¶ **Involution Matrix**

Let V be finite dimensional, F be number field, and let A be an involution matrix, i.e., $A^2 = E$. In geometry, an involution typically corresponds to a "reflection" or "symmetry" operation, since $A^2 = E$ implies that $A = A^{-1}$.

1. All eigenvalues of A are either 1 or -1.
2. A is diagonalizable, and

$$V = E_1 \oplus E_{-1} = \mathrm{Ker}(A - E) \oplus \mathrm{Ker}(A + E).$$

¶ **1 Matrix**

Let F be rational field $\mathbb{Q}$, and let A be a 1 matrix, i.e., all entries of A are 1. Denote

$$1_n = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}_{n \times 1}.$$

1. All eigenvalues of A are either n (algebraic multiplicity is 1) or 0 (algebraic multiplicity is $n - 1$).
2. The set of eigenvector corresponding to eigenvalue n is

$$\{k1_n | k \in \mathbb{Q}, k \neq 0\};$$

the set of eigenvectors corresponding to eigenvalue 0 is

$$\{k_1 \eta_1 + k_2 \eta_2 + \cdots + k_{n-1} \eta_{n-1} | k_1, k_2, \ldots, k_{n-1} \in \mathbb{Q} \text{ and not all zero }\},$$

where

$$\eta_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \eta_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots, \quad \eta_{n-1} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ -1 \end{pmatrix}.$$

3. A is diagonalizable.

It is worth noting that $\frac{1}{n} A$ is an idempotent orthonormal projection matrix.

¶ **Cyclic Shift Matrix**

Let F be complex number field $\mathbb{C}$, and let A be a cyclic shift matrix, i.e.,

$$A = \begin{pmatrix} \varepsilon_n & \varepsilon_1 & \varepsilon_2 & \cdots & \varepsilon_{n-1} \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

All of its properties are closely related to the discrete fourier transform (DFT) and the roots of unity.

1. The eigenvalues of A are the n-th roots of unity $\omega^{(k)} = \exp\left(\frac{2\pi ik}{n}\right)$ for $k = 0, 1, \ldots, n-1$.

2. The eigenvector corresponding to the eigenvalue $\omega^{(k)} = \exp\left(\frac{2\pi ik}{n}\right)$ is

$$\xi^{(k)} = \begin{pmatrix} \left(\omega^{(k)}\right)^0 \\ \left(\omega^{(k)}\right)^1 \\ \left(\omega^{(k)}\right)^2 \\ \vdots \\ \left(\omega^{(k)}\right)^{n-1} \end{pmatrix}.$$

These eigenvectors $\xi^{(0)}, \xi^{(1)}, \ldots, \xi^{(n-1)}$ are precisely the columns of the DFT matrix (or columns after a simple conjugate and scaling).

3. A is diagonalizable. Furthermore, it can be unitarily diagonalized.

¶ **Circulant Matrix**

Let F be complex number field $\mathbb{C}$, and let A be a circulant matrix, i.e.,

$$A = \begin{pmatrix} a_0 & a_1 & a_2 & \cdots & a_{n-1} \\ a_{n-1} & a_0 & a_1 & \cdots & a_{n-2} \\ a_{n-2} & a_{n-1} & a_0 & \cdots & a_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_1 & a_2 & a_3 & \cdots & a_0 \end{pmatrix} = f(C) = a_0 E + a_1 C + a_2 C^2 + \cdots + a_{n-1} C^{n-1},$$

where C is the cyclic shift matrix.

1. The eigenvalues of A are given by:

$$\lambda_k = f(\omega^{(k)}) = a_0 + a_1\omega^{(k)} + a_2(\omega^{(k)})^2 + \cdots + a_{n-1}(\omega^{(k)})^{n-1},$$

where $\omega^{(k)} = \exp\left(\frac{2\pi ik}{n}\right)$ for $k = 0, 1, \ldots, n-1$ [2].

2. The eigenvector corresponding to the eigenvalue $\lambda_k = f(\omega^{(k)})$ is

$$\xi^{(k)} = \begin{pmatrix} \left(\omega^{(k)}\right)^0 \\ \left(\omega^{(k)}\right)^1 \\ \left(\omega^{(k)}\right)^2 \\ \vdots \\ \left(\omega^{(k)}\right)^{n-1} \end{pmatrix}.$$

It means that circulant matrices share the same set of eigenvectors as the cyclic shift matrix.

3. A is diagonalizable. Furthermore, it can be unitarily diagonalized by the <u>same</u> DFT matrix.

¶ **Triangular Matrix**

---

[2] That is, the eigenvalues of a circulant matrix are the discrete Fourier transform of its first row vector $(a_0, a_1, \ldots, a_{n-1})$ (possibly differing by a constant factor and a conjugate, depending on the specific definition of the DFT).

Let F be number field, and let A be a upper triangular matrix:

$$A = (a_{ij})_{n \times n}, \quad a_{ij} = 0, \quad \forall i > j.$$

1. The eigenvalues of A are the diagonal entries a_{ii} for $i = 1, 2, \dots, n$.
2. If all diagonal entries a_{ii} are distinct, then A is diagonalizable. However, if there are repeated diagonal entries, A may be or not be diagonalizable.
3. If all diagonal entries a_{ii} are equal, and there exists some non-zero entry above the diagonal, then A is not diagonalizable.

7.7 Simultaneous Characteristics Induced by Commutativity of Multiplication

In this section, we will explore the properties of matrices and linear transformations under similarity relations induced by the commutativity of multiplication operations. We refer to these properties as **simultaneous characteristics**. The discussion will proceed in five steps.

¶ Eigenspaces are mutually Invariant Subspaces
¶ Shared Eigenvectors
¶ Simultaneous Upper Triangularization
¶ Simultaneous Diagonalization
¶ Generalization in Multiple Cases

Chapter 8   Jordan Forms

8.1 Polynomial Matrices

> **Definition 8.1 (Polynomial Matrix)**
>
> A matrix in the form:
> $$A(\lambda) = \begin{pmatrix} a_{11}(\lambda) & a_{12}(\lambda) & \cdots & a_{1n}(\lambda) \\ a_{21}(\lambda) & a_{22}(\lambda) & \cdots & a_{2n}(\lambda) \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}(\lambda) & a_{m2}(\lambda) & \cdots & a_{mn}(\lambda) \end{pmatrix},$$
> is called a **polynomial matrix** over F, where each $a_{ij}(\lambda) \in F[\lambda]$ is a polynomial in variable λ over field F. ♣

Similar to ordinary matrices, polynomial matrices can also perform addition, multiplication, scalar multiplication, transposition, and taking determinants.

⚠**Caution** *$A(\lambda)$ is invertible if and only if $|A(\lambda)|$ is a <u>non-zero constant</u>, it is worth noting that this is non-zero constant, not just any non-zero polynomial.*

> **Definition 8.2 (Equivalent Polynomial Matrices)**
>
> Let $A(\lambda) \in F[\lambda]^{m \times n}$ and $B(\lambda) \in F[\lambda]^{m \times n}$ be two polynomial matrices. If they can be transformed into each other through a series of elementary row and column operations, then they are called **equivalent polynomial matrices**, denoted as $A(\lambda) \cong B(\lambda)$. ♣

📖**Note** *Obviously, $A(\lambda) \cong B(\lambda)$ if and only if their ranks are equal:* $\operatorname{rank} A(\lambda) = \operatorname{rank} B(\lambda)$.

> **Definition 8.3 (Smith Normal Form (Canonical Form of Polynomial Matrices))**
>
> Let $A(\lambda) \in F[\lambda]^{m \times n}$ be a polynomial matrix with rank $r \geqslant 1$. If $A(\lambda)$ is equivalent to the following polynomial matrix:
> $$\begin{pmatrix} d_1(\lambda) & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & d_2(\lambda) & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_r(\lambda) & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix},$$
> where $d_1(\lambda), d_2(\lambda), \ldots, d_r(\lambda)$ are monic polynomials over field F such that $d_1(\lambda)|d_2(\lambda)|\cdots|d_r(\lambda)$, then this polynomial matrix is called the **Smith normal form** of $A(\lambda)$. ♣

It is easy to verify that the Smith normal form of polynomial matrices exists and is unique up to the order of the invariant divisors.

8.2 Invariant Divisors and Rational Canonical Form

¶ **Invariant Divisors**

> **Definition 8.4 (k Order Determinant Divisor)**
>
> Let $A(\lambda) \in F[\lambda]^{m \times n}$ be a polynomial matrix. The monic greatest common divisor of all k-order minors of $A(\lambda)$ is called the k **order determinant divisor** of $A(\lambda)$, denoted as $D_k(\lambda)$.
>
> ♣

> **Definition 8.5 (Invariant Divisors)**
>
> Let $A(\lambda) \in F[\lambda]^{m \times n}$ be a polynomial matrix, and let $D_1(\lambda), D_2(\lambda), \ldots, D_r(\lambda)$ be its $1, 2, \ldots, r$ order determinant divisors, where $r = \operatorname{rank} A(\lambda)$. The polynomials
> $$d_1(\lambda) = D_1(\lambda), \quad d_2(\lambda) = \frac{D_2(\lambda)}{D_1(\lambda)}, \quad \ldots, \quad d_r(\lambda) = \frac{D_r(\lambda)}{D_{r-1}(\lambda)}$$
> are called the **invariant divisors** of $A(\lambda)$.
>
> ♣

☞ **Note** *In fact, for the canonical form of equivalent polynomial matrices,*
$$\operatorname{diag}\{d_1(\lambda), d_2(\lambda), \ldots, d_r(\lambda), 0, \ldots, 0\},$$
the invariant divisors are just the non-zero diagonal elements of the canonical form.

¶ **Rational Canonical Form**

> **Definition 8.6 (Companion Matrix)**
>
> The Frobenius companion matrix of the monic polynomial
> $$p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$
> is the square matrix defined as follows:
> $$C(p) = \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -a_{n-1} \end{pmatrix}.$$
>
> ♣

Some authors use the transpose of this matrix, $C(p)^T$, which is more convenient for some purposes such as linear recurrence relations.

$C(p)$ is defined from the coefficients of $p(x)$, while the characteristic polynomial as well as the minimal polynomial of $C(p)$ are equal to $p(x)$. In this sense, the matrix $C(p)$ and the polynomial $p(x)$ are "companions".

✐ **Property**

1. *The characteristic polynomial and minimal polynomial of $C(p)$ is $p(x)$ (monic).*
2.
3.

> **Definition 8.7 (Frobenius Normal Form (Rational Canonical Form))**
>
> Let V be an n-dimensional linear space over field F, and let $\mathcal{A} \in \operatorname{Hom}(V)$. If there exists a basis of V such that

the matrix representation of $\mathcal{A}$ under this basis is:

$$F = \begin{pmatrix} C(p_1) & 0 & \cdots & 0 \\ 0 & C(p_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & C(p_s) \end{pmatrix},$$

where $p_1, p_2, \ldots, p_s$ are monic polynomials over field F such that $p_1|p_2|\cdots|p_s$, then F is called the **Frobenius normal form** (or rational canonical form) of $\mathcal{A}$.

♣

¶ Geometric Interpretation

> **Definition 8.8 (Cyclic Subspace)**
>
> Let V be a linear space over field F, and let $\mathcal{A} \in \mathrm{Hom}(V)$. For a vector $0 \neq \xi \in V$, the subspace
>
> $$C(\mathcal{A}, \xi) = \langle \xi, \mathcal{A}(\xi), \mathcal{A}^2(\xi), \ldots \rangle = \{f(\mathcal{A})(\xi)|f(x) \in F[x]\}$$
>
> is called the **cyclic subspace** generated by ξ under the linear transformation $\mathcal{A}$.
> If $C(\mathcal{A}, \xi) = V$, then C is called a **cyclic space**.

♣

Obviously, $C(\mathcal{A}, \xi)$ is the smallest $\mathcal{A}$-subspace containing ξ.

✐ **Property** *If* $\dim C(\mathcal{A}, \xi) = r$, *then* $\{\xi, \mathcal{A}(\xi), \ldots, \mathcal{A}^{r-1}(\xi)\}$ *is a basis of* $C(\mathcal{A}, \xi)$.

✎ *Proof*

Let $k_0 = \max\{k \in \mathbb{N}^*|\xi, \mathcal{A}(\xi), \ldots, \mathcal{A}^{k-1}(\xi) \text{ is linearly independent}\}$, then according to properties of linear independence (3.1-6) and mathematical induction, we can prove that for any $k \geqslant k_0$, $\mathcal{A}^k$ is a linear combination of $\xi, \mathcal{A}(\xi), \ldots, \mathcal{A}^{k_0-1}(\xi)$. Therefore, $\{\xi, \mathcal{A}(\xi), \ldots, \mathcal{A}^{k_0-1}(\xi)\}$ is a basis of $C(\mathcal{A}, \xi)$ and $\dim C(\mathcal{A}, \xi) = k_0$.

∎

Next, we will show that the sufficient and necessary conditions for cyclic subspace and cyclic space.

> **Proposition 8.1**
>
> Let V_1 be $\mathcal{A}$-subspace of linear space V over field F, then V_1 is a cyclic subspace if and only if the matrix representation of $\mathcal{A}|_{V_1}$ under some basis of V_1 is a companion matrix.

♠

Generally, let the set of invariant divisors of $\mathcal{A}$ be denoted as

$$\{1, \cdots, 1, d_1(x), d_2(x), \ldots, d_k(x)\},$$

where $d_i(x)$ are monic polynomials in $F[x]$ such that

$$d_1(x)|d_2(x)|\cdots|d_k(x).$$

Then according to theory of rational canonical form, there exists a basis of V such that the matrix representation of $\mathcal{A}$ under this basis is:

$$C = \mathrm{diag}\{C(d_1), C(d_2), \ldots, C(d_k)\}.$$

Combining with the above proposition, we can conclude that there exists a decomposition of V into a direct sum of cyclic subspaces:

$$V = \bigoplus_{i=1}^{k} C(\mathcal{A}, \xi_i),$$

such that the matrix representation of $\mathcal{A}|_{C(\mathcal{A},\xi_i)}$ under the basis $\{\xi_i, \mathcal{A}(\xi_i), \ldots, \mathcal{A}^{r_i-1}(\xi_i)\}$ is the companion matrix $C(d_i)$, where $r_i = \dim C(\mathcal{A}, \xi_i)$.

This is the geometric interpretation of rational canonical form.

Proposition 8.2

Let V be an n-dimensional linear space over field F, and let $\mathcal{A} \in \mathrm{Hom}(V)$. The following statements are equivalent:

1. There exists a vector $\xi \in V$ such that $\{\xi, \mathcal{A}(\xi), \ldots, \mathcal{A}^{n-1}(\xi)\}$ is a basis of V, i.e., V is a cyclic space generated by ξ;
2. For any eigenvalue λ of $\mathcal{A}$, the geometric multiplicity is 1;
3. The characteristic polynomial of $\mathcal{A}$ is equal to the minimal polynomial of $\mathcal{A}$.

♠

🖊 *Proof*

1⇒2 There exists a vector $\xi \in V$ such that $\{\xi, \mathcal{A}(\xi), \ldots, \mathcal{A}^{n-1}(\xi)\}$ is a basis of V. The matrix representation of $\mathcal{A}$ under this basis is a companion matrix,

$$C(p) = \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -a_{n-1} \end{pmatrix}.$$

For any $c \in F$, we have

$$cE - C(p) = \begin{pmatrix} c & 0 & \cdots & 0 & a_0 \\ -1 & c & \cdots & 0 & a_1 \\ 0 & -1 & \cdots & 0 & a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & c & a_{n-1} \end{pmatrix}.$$

The minor of order $n-1$ in the bottom left corner is non-zero, thus the rank of $cE - C(p)$ is at least $n-1$. Therefore, the solution space of the equation

$$(cE - C(p))X = 0$$

is $n - \mathrm{rank}(cE - C(p)) \leqslant 1$. For any eigenvalue λ of $\mathcal{A}$,

$$(\lambda E - C(p))X = 0$$

has non-zero solutions definitely, thus its solution space is $\geqslant 1$.

Hence, the dimension of solution space of the equation

$$(\lambda E - C(p))X = 0$$

is exactly 1,i.e., the geometric multiplicity of eigenvalue λ is 1.

2⇒1 Let the distinct eigenvalues of $\mathcal{A}$ be $\lambda_1, \lambda_2, \ldots, \lambda_k$, and let the algebraic multiplicities of these eigenvalues be $r_1, r_2, \ldots, r_k$, respectively. Obviously, $r_1 + r_2 + \cdots + r_k = n$ and the characteristic polynomial of $\mathcal{A}$ is

$$f(x) = (x - \lambda_1)^{r_1}(x - \lambda_2)^{r_2} \cdots (x - \lambda_k)^{r_k}.$$

According to the root space decomposition theorem (7.5),

$$V = V_{\lambda_1}^{(r_1)} \oplus V_{\lambda_2}^{(r_2)} \oplus \cdots \oplus V_{\lambda_k}^{(r_k)},$$

where $V_{\lambda_i}^{(r_i)}$ is the root subspace of order r_i corresponding to eigenvalue λ_i, which is an $\mathcal{A}$-subspace of dimension r_i apparently.

Since the geometric multiplicity of eigenvalue λ_i is 1, the Jordan canonical form of $\mathcal{A}|_{V_{\lambda_i}^{(r_i)}}$ is

$$J_{r_i}(\lambda_i) = \begin{pmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 \\ 0 & 0 & \lambda_i & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_i \end{pmatrix}_{r_i \times r_i}.$$

Since the equivalence between Jordan block and cyclic space, there exists $\alpha_i \in V_{\lambda_i}^{(r_i)}$ such that

$$J_{r_i}(\lambda_i) = C(\mathcal{A}, \alpha_i) = \langle \alpha_i, (\mathcal{A} - \lambda_i \mathcal{E})(\alpha_i), \ldots, (\mathcal{A} - \lambda_i \mathcal{E})^{r_i - 1}(\alpha_i) \rangle,$$

that is, $(\mathcal{A} - \lambda_i \mathcal{E})^{r_i - 1}(\alpha_i) \neq 0$.

Let $\alpha = \alpha_1 + \alpha_2 + \cdots + \alpha_k$, then we will prove that $\{\alpha, \mathcal{A}(\alpha), \ldots, \mathcal{A}^{n-1}(\alpha)\}$ is a basis of V.

If not, they are linearly dependent. Then there exist $p(x) \in F[x]$, $\deg p(x) \leqslant n - 1$, such that

$$p(\mathcal{A})(\alpha) = 0.$$

Since $V_{\lambda_i}^{(r_i)}$ is an $\mathcal{A}$-subspace, we have $p(\mathcal{A})(\alpha) \in V_{\lambda_i}^{(r_i)}$.

Since

$$0 = p(\mathcal{A})(\alpha) = p(\mathcal{A})(\alpha_1) + p(\mathcal{A})(\alpha_2) + \cdots + p(\mathcal{A})(\alpha_k),$$

we have $p(\mathcal{A})(\alpha_i) = 0$ for all i.

Since $(\mathcal{A} - \lambda_i \mathcal{E})^{r_i - 1}(\alpha_i) \neq 0$, $(x - \lambda_i)^{r_i}$ is the minimal degree monic polynomial of $\mathcal{A}$ annihilating α_i, thus

$$(x - \lambda_i)^{r_i} | p(x).$$

Since when $i \neq j$, $(x - \lambda_i)$ and $(x - \lambda_j)$ are coprime, we have

$$(x - \lambda_1)^{r_1}(x - \lambda_2)^{r_2} \cdots (x - \lambda_k)^{r_k} | p(x),$$

i.e., $f(x)|p(x)$, which contradicts $\deg p(x) < n = \deg f(x)$ and $p(x) \neq 0$. ■

8.3 Elementary Divisors and Jordan Canonical Form

¶ **Elementary Divisors**

> **Definition 8.9 (Elementary Divisors)**
>
> Let $A(\lambda) \in F[\lambda]^{m \times n}$ be a polynomial matrix, and let $d_1(\lambda), d_2(\lambda), \ldots, d_r(\lambda)$ be its invariant divisors, where $r = \operatorname{rank} A(\lambda)$. If the factorization of each invariant divisor into irreducible polynomials over field F is:
>
> $$d_i(\lambda) = p_{i1}(\lambda)^{k_{i1}} p_{i2}(\lambda)^{k_{i2}} \cdots p_{is_i}(\lambda)^{k_{is_i}},$$
>
> then the polynomials
>
> $$p_{ij}(\lambda)^{k_{ij}}, \quad i = 1, 2, \ldots, r; \quad j = 1, 2, \ldots, s_i$$
>
> are called the **elementary divisors** of $A(\lambda)$. ♣

¶ `Jordan Canonical Form`

Lemma 8.1

Let V be a linear space over complex number field F, and let $\mathcal{A} \in \text{Hom}(V)$ and $\mathcal{A}$ is nilpotent. Then there exists a basis of V such that the matrix representation of $\mathcal{A}$ under this basis is

$$\text{diag}\{J_{r_1}(0), J_{r_2}(0), \ldots, J_{r_s}(0)\},$$

where $J_{r_i}(0)$ is the Jordan block of size r_i corresponding to eigenvalue 0.

♡

Combine with Theorem 7.5, we have the following theorem:

Theorem 8.1 (Existence of Jordan Canonical Form)

Let V be an n-dimensional linear space over complex number field F, and let $\mathcal{A} \in \text{Hom}(V)$. Then there exists a basis of V: $\{e_1, e_2, \ldots, e_n\}$, such that the matrix representation of $\mathcal{A}$ under this basis is Jordan canonical form:

$$J = \text{diag}\{J_{r_1}(\lambda_1), J_{r_2}(\lambda_2), \ldots, J_{r_s}(\lambda_s)\},$$

♡

¶ `Geometric Interpretation`

From perspective of space decomposition, a linear space V can be decomposed into a direct sum of root subspaces, and according to the proof of Lemma 8.1, each root subspace can be further decomposed into a direct sum of cyclic subspaces generated by generalized eigenvectors. Therefore, V can be decomposed into a direct sum of cyclic subspaces generated by generalized eigenvectors,

$$V = C(\mathcal{A} - \lambda_1 \mathcal{E}, \xi_1) \oplus C(\mathcal{A} - \lambda_2 \mathcal{E}, \xi_2) \oplus \cdots \oplus C(\mathcal{A} - \lambda_s \mathcal{E}, \xi_s),$$

where cyclic subspace $C(\mathcal{A} - \lambda_i \mathcal{E}, \xi_i)$ one-to-one corresponds to a Jordan block $J_{r_i}(\lambda_i)$, which is geometric interpretation of Jordan canonical form.

8.4 Finding Jordan Canonical Form and Transition Matrices

¶ `Finding Jordan Canonical Form`
¶ `Finding Transition Matrices`

8.5 Complete Set of Invariant Quantities for Matrix Similarity

Chapter 9　Quadratic Forms

9.1　Quadratic Forms and Their Standard Forms

> **Definition 9.1 (Quadratic Form)**
>
> Let P be a number field, a quadratic homogeneous polynomial in n variables over P^a:
>
> $$f(x_1, x_2, \cdots, x_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$$
>
> $$= a_{11}x_1^2 + 2a_{12}x_1x_2 + \cdots + 2a_{1n}x_1x_n + a_{22}x_2^2 + \cdots + 2a_{2n}x_2x_n + \cdots + a_{nn}x_n^2,$$
>
> is called a **quadratic form** in n variables over field P.
>
> It can be expressed in matrix form as:
>
> $$f(x_1, x_2, \cdots, x_n) = X^{\mathrm{T}}AX,$$
>
> where
>
> $$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad A = (a_{ij})_{n \times n}, \quad a_{ij} = a_{ji} \quad (1 \leqslant i, j \leqslant n).$$
>
> ---
>
> ^{a}That is, the coefficients of the polynomial belong to the field P.
>
> ♣

It is easy to verify that the matrix A of a quadratic form is symmetric.

🍷 **Note**　*In fact, for any square matrix B[1], we have:*

$$X^T B X = X^T \left(\frac{B + B^T}{2} \right) X + X^T \left(\frac{B - B^T}{2} \right) X$$

$$= X^T \left(\frac{B + B^T}{2} \right) X + 0$$

$$= X^T \left(\frac{B + B^T}{2} \right) X.$$

It shows that any quadratic form can be represented by a symmetric matrix.

---

[1]For a skew-symmetric matrix S ($S^T = -S$),

$$X^T S X = -(X^T S X)^T = -X^T S^T X = -X^T S X \implies X^T S X = 0.$$

9.2 Canonical Forms

9.3 Definite Quadratic Forms

Definition 9.2 (Positive Definite Quadratic Form)

A real quadratic form $f(x_1, x_2, \cdots, x_n) = X^{\mathrm{T}} A X$ is called **positive definite** if:

$$f(x_1, x_2, \cdots, x_n) > 0, \quad \forall X \neq 0.$$

And A is called a **positive definite matrix**.

♣

Theorem 9.1 (Sufficient and Necessary Condition for Positive Definiteness)

A real quadratic form $f(x_1, x_2, \cdots, x_n) = X^{\mathrm{T}} A X$ is positive definite if and only if:

1. The positive inertia index of f is n;
2. A is congruent to the identity matrix E;
3. All eigenvalues of A are positive;
4. All leading principal minors[a] of A are positive.

---

[a]The leading principal minors of a matrix are the determinants of the top-left $k \times k$ submatrices for $k = 1, 2, \ldots, n$.

♡

Chapter 10   Inner Product Spaces

10.1 Bilinear Forms

Definition 10.1 (Bilinear Form)

Let V be a linear space over field F. A function $f : V \times V \to F$ is called a **bilinear form** on V if:
1. For any fixed $\beta \in V$, the function $f(\cdot, \beta) : V \to F$ defined by $f(\alpha, \beta)$ is a linear function on V;
2. For any fixed $\alpha \in V$, the function $f(\alpha, \cdot) : V \to F$ defined by $f(\alpha, \beta)$ is a linear function on V. ♣

10.2 Real Inner Product Spaces

Definition 10.2 (Real Inner Product Space)

A **real inner product space** is a real linear space V equipped with a function $(\cdot, \cdot) : V \times V \to \mathbb{R}$ satisfying the following properties:

Positivity $(\alpha, \alpha) \geq 0, \quad \forall \alpha \in V$, and $(\alpha, \alpha) = 0$ if and only if $\alpha = 0$;

Symmetry $(\alpha, \beta) = (\beta, \alpha), \quad \forall \alpha, \beta \in V$;

Linearity in the First Argument $(k_1\alpha_1 + k_2\alpha_2, \beta) = k_1(\alpha_1, \beta) + k_2(\alpha_2, \beta), \quad \forall k_1, k_2 \in \mathbb{R}, \alpha_1, \alpha_2, \beta \in V$.

The function $(\cdot, \cdot)$ is called the (real) **inner product**[a] on V.

Real inner product spaces with finite dimensions are called **Euclidean spaces**.

---
[a]The inner product can be also defined as a positive-definite bilinear form. ♣

Definition 10.3 (Normed Linear Space)

A real **normed linear space** is a real linear space V equipped with a function $\|\cdot\| : V \to \mathbb{R}$ satisfying the following properties[a]:

Positivity $\|\alpha\| \geq 0, \quad \forall \alpha \in V$, and $\|\alpha\| = 0$ if and only if $\alpha = 0$;

Homogeneity $\|k\alpha\| = |k|\|\alpha\|, \quad \forall k \in \mathbb{R}, \alpha \in V$;

Triangle Inequality $\|\alpha + \beta\| \leq \|\alpha\| + \|\beta\|, \quad \forall \alpha, \beta \in V$.

The function $\|\cdot\|$ is called the (vector) **norm**[b] on V.

---
[a]Similarly, the definition of norm can be given in complex linear spaces.

[b]If replace positivity with semi-positivity in the above definition, i.e., $\|\alpha\| \geq 0, \quad \forall \alpha \in V$, then we get the definition of **semi-norm**. ♣

In Euclidean spaces, the norm can be induced by the inner product:

$$\|\alpha\| = (\alpha, \alpha)^{\frac{1}{2}}, \quad \forall \alpha \in V,$$

which is called the **Euclidean norm**.

⚡**Remark** The definition of Euclidean space can be also derived from normed linear space or metric space.

> **Theorem 10.1 (Cauchy-Буняко́вский-Schwarz Inequality)**
>
> Let V be an inner product space[a]. For any vectors $\alpha, \beta \in V$, the following inequality holds:
>
> $$|(\alpha, \beta)| \leq \|\alpha\|\|\beta\|,$$
>
> with equality if and only if α and β are linearly dependent.
>
> ---
>
> [a]It <u>does not</u> require in real inner product spaces.

✑ Note

1. *In linear space $\mathbb{R}^n$, for vectors $\alpha, \beta \in \mathbb{R}^n$, define the inner product as:*

$$(\alpha, \beta) = \alpha^{\mathrm{T}}\beta = \sum_{i=1}^{n} x_i y_i,$$

*then $\mathbb{R}^n$ forms a Euclidean space, called the **standard Euclidean space**. We still denote it as $\mathbb{R}^n$ without confusion. In this case, Cauchy-Буняко́вский-Schwarz inequality becomes:*

$$\left|\sum_{i=1}^{n} x_i y_i\right| \leqslant \left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^{n} y_i^2\right)^{\frac{1}{2}}.$$

2. *In $C[a, b]$, which is the linear space of continuous real-valued functions on interval $[a, b]$, for functions $f(x), g(x) \in C[a, b]$, define the inner product as:*

$$(f, g) = \int_a^b f(x)g(x)\,\mathrm{d}x,$$

then $C[a, b]$ forms a Euclidean space.

In this case, Cauchy-Буняко́вский-Schwarz inequality becomes:

$$\left|\int_a^b f(x)g(x)\,\mathrm{d}x\right| \leqslant \left(\int_a^b f^2(x)\,\mathrm{d}x\right)^{\frac{1}{2}} \left(\int_a^b g^2(x)\,\mathrm{d}x\right)^{\frac{1}{2}}.$$

Some concepts in real inner product spaces can be derived:

Distance The distance between two vectors $\alpha, \beta \in V$ is defined as:

$$d(\alpha, \beta) = \|\alpha - \beta\|$$

Angle The angle θ between two non-zero vectors $\alpha, \beta \in V$ is defined using the inner product:

$$\cos\theta = \frac{(\alpha, \beta)}{\|\alpha\|\|\beta\|}, 0 \leq \theta \leq \pi.$$

Orthogonality Two vectors $\alpha, \beta \in V$ are said to be orthogonal if:

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

denoted as $\alpha \perp \beta$.

10.3 Metric Matrices and Orthonormal Bases

¶ Metric Matrices

> **Definition 10.4 (Gram Matrix)**
>
> Let V be a n-dimensional Euclidean space, and let $\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\}$ be a basis of V. The matrix
>
> $$G = ((\varepsilon_i, \varepsilon_j))_{n \times n} = \begin{pmatrix} (\varepsilon_1, \varepsilon_1) & (\varepsilon_1, \varepsilon_2) & \cdots & (\varepsilon_1, \varepsilon_n) \\ (\varepsilon_2, \varepsilon_1) & (\varepsilon_2, \varepsilon_2) & \cdots & (\varepsilon_2, \varepsilon_n) \\ \vdots & \vdots & \ddots & \vdots \\ (\varepsilon_n, \varepsilon_1) & (\varepsilon_n, \varepsilon_2) & \cdots & (\varepsilon_n, \varepsilon_n) \end{pmatrix}$$
>
> is called the **Gram matrix** (or metric matrix) of the inner product on V under the basis $\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\}$. ♣

For all vectors

$$\alpha = x_1 \varepsilon_1 + x_2 \varepsilon_2 + \cdots + x_n \varepsilon_n,$$

$$\beta = y_1 \varepsilon_1 + y_2 \varepsilon_2 + \cdots + y_n \varepsilon_n \in V,$$

since $(\alpha, \beta) = \sum_{i=1}^{n} \sum_{j=1}^{n} (\varepsilon_i, \varepsilon_j) x_i y_j$, it can be expressed in matrix form as:

$$(\alpha, \beta) = X^{\mathrm{T}} G Y, \quad X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}.$$

✎ **Property**

1. *Gram matrix G is symmetric;*
2. *G is positive semi-definite;*
3. *In V, Gram matrices under different bases are congruent;*
4. *$|G| \geqslant 0$, equality holds if and only if the basis is linearly dependent.*

¶ `Orthonormal Bases`

> **Definition 10.5 (Orthonormal Vector Set and Orthonormal Basis)**
>
> Let V be a n-dimensional real inner product space. A set of non-zero vectors $\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\}$ in V is called an **orthonormal vector set** if they are pairwise orthogonal.
>
> A orthonormal vector set with n vectors is called a **orthonormal basis** of V. Orthonormal basis made up of unit vectors is called a **orthonormal basis** of V. ♣

✎ **Property** *In n-dimensional Euclidean space,*

1. *Orthonormal vector set is linearly independent;*
2. *A set of basis is orthonormal basis if and only if its Gram matrix is the identity matrix;*
3. *Standard orthonormal basis always exists.*
4. *Any orthonormal vector set can be extended to a orthonormal basis.*

   In n-dimensional Euclidean space, let $\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\}$ be a orthonormal basis, then

- for all $\alpha \in V : \alpha = \sum_{i=1}^{n} (\alpha, \varepsilon_i) \varepsilon_i$, which is called the **Fourier expansion** of vector α under the basis;
- if the coordinates of α, β under this basis are

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix},$$

then
$$(\alpha, \beta) = X^{\mathrm{T}}Y = \sum_{i=1}^{n} x_i y_i.$$

> **Definition 10.6 (Orthonormal Matrix)**
>
> A square matrix $T \in \mathbb{R}^{n \times n}$ is called an **orthonormal matrix** if:
> $$T^{\mathrm{T}}T = TT^{\mathrm{T}} = E,$$
> where E is the identity matrix of order n. ♣

Let V be a n-dimensional Euclidean space, $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ and $\eta_1, \eta_2, \cdots, \eta_n$ be orthonormal bases of V, and let T be the transition matrix between them, i.e.,
$$(\eta_1, \eta_2, \cdots, \eta_n) = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)T.$$
Then
1. T is an orthonormal matrix;
2. T is upper triangular matrix;

> **Theorem 10.2**
>
> Let $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ be a orthonormal basis of n-dimensional Euclidean space V, and
> $$(\eta_1, \eta_2, \cdots, \eta_n) = (\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n)T.$$
> Then $\eta_1, \eta_2, \cdots, \eta_n$ is a orthonormal basis of V if and only if T is an orthonormal matrix. ♡

¶ `Gram-Schmidt Process`

> **Theorem 10.3**
>
> For any basis $\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\}$ of n-dimensional Euclidean space V, there exists a orthonormal basis $\{\eta_1, \eta_2, \ldots, \eta_n\}$ such that:
> $$\langle \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k \rangle = \langle \eta_1, \eta_2, \ldots, \eta_k \rangle, \quad k = 1, 2, \ldots, n.$$ ♡

✎ *Proof*

■

Example 10.1 Let $n \geqslant 2$, A is an n−order real symmetric matrix. $\alpha = (a_1, a_2, \ldots, a_n)^{\mathrm{T}}$, $\beta = (b_1, b_2, \ldots, b_n)^{\mathrm{T}}$, which are two eigenvectors of A corresponding to different eigenvalues λ_1, λ_2 respectively. Let
$$B = \begin{pmatrix} a_1 + b_1 & a_1 + b_2 & \cdots & a_1 + b_n \\ a_2 + b_1 & a_2 + b_2 & \cdots & a_2 + b_n \\ \vdots & \vdots & \ddots & \vdots \\ a_n + b_1 & a_n + b_2 & \cdots & a_n + b_n \end{pmatrix},$$
find all eigenvalues of matrix B.

✎ *Solution* Let $\xi_1 = \frac{\alpha}{\|\alpha\|}$, $\xi_2 = \frac{\beta}{\|\beta\|}$, and $S_a = \sum_{i=1}^{n} a_i$, $S_b = \sum_{i=1}^{n} b_i$.

It is easy to see that ξ_1, ξ_2 are orthonormal unit vectors.

Extend $\{\xi_1, \xi_2\}$ to an orthonormal basis $\{\xi_1, \xi_2, \ldots, \xi_n\}$ of $\mathbb{R}^n$.

Let $P = (\xi_1, \xi_2, \ldots, \xi_n)$, then P is an orthonormal matrix.

B can be expressed as:

$$B = \alpha e^{\mathrm{T}} + e\beta^{\mathrm{T}},$$

where $e = (1, 1, \ldots, 1)^{\mathrm{T}}$.

Therefore,

$$P^{-1}BP = P^{\mathrm{T}}BP = \begin{pmatrix} \xi_1^{\mathrm{T}} \\ \xi_2^{\mathrm{T}} \\ \vdots \\ \xi_n^{\mathrm{T}} \end{pmatrix} \begin{pmatrix} B\xi_1 & B\xi_2 & \cdots & B\xi_n \end{pmatrix}.$$

Since

$$B\xi_1 = (\alpha e^{\mathrm{T}} + e\beta^{\mathrm{T}})\frac{a_1}{\|\alpha\|} = \frac{\alpha(e^{\mathrm{T}}\alpha)}{\|\alpha\|} + \frac{e(\beta^{\mathrm{T}}\alpha)}{\|\alpha\|} = \frac{S_a}{\|\alpha\|}\alpha + 0 = S_a\|\alpha\|\xi_1,$$

and similarly, $B\xi_2 = S_b\|\beta\|\xi_2$. For $k \geqslant 3$, we have

$$B\xi_k = (\alpha e^{\mathrm{T}} + e\beta^{\mathrm{T}})\xi_k = \alpha(e^{\mathrm{T}}\xi_k) + e(\beta^{\mathrm{T}}\xi_k) = 0 + 0 = 0.$$

Then

$$P^{\mathrm{T}}BP = \begin{pmatrix} S_a & 0 & 0 & \cdots & 0 \\ 0 & S_b & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

Therefore, the eigenvalues of matrix B are:

$$\lambda_1 = S_a, \quad \lambda_2 = S_b, \quad \lambda_3 = 0, \quad \ldots, \quad \lambda_n = 0.$$

$\square$

10.4 Isomorphism of Real Inner Product Spaces

10.5 Orthogonal Completion and Orthogonal Projection

¶ **Orthogonal Completion**

> **Definition 10.7 (Orthogonal Complement)**
>
> Let V be a real inner product space, and let W be a subspace of V. The set
>
> $$W^{\perp} = \{\alpha \in V \mid (\alpha, \beta) = 0, \quad \forall \beta \in W\}$$
>
> is called the **orthogonal complement**[a] of W in V, and $W^{\perp}$ is also a subspace of V, called the **orthogonal subspace** of W in V.
>
> ---
>
> [a]Another equivalent definition is:
> $$W^{\perp} \perp W \text{ and } V = W + W^{\perp}.$$
>
> ♣

🖉 **Property**

1. *There exists a unique orthogonal complement $W^{\perp}$ of W in Euclidean space V;*
2. $W \oplus W^{\perp} = V$;

3. $\left(W^{\perp}\right)^{\perp} = W$

4. $(V_1 + V_2)^{\perp} = V_1^{\perp} \cap V_2^{\perp}, \quad (V_1 \cap V_2)^{\perp} = V_1^{\perp} + V_2^{\perp}$, where V_1, V_2 are subspaces of V.

¶ **Least Squares Method**

> *Definition 10.8 (Orthogonal Projection)*
>
> Let V be a real inner product space, and let W be a subspace of V. For any vector $\alpha \in V$, if there exists a vector $\beta \in W$ such that:
> $$\alpha - \beta \in W^{\perp},$$
> then β is called the **orthogonal projection** of α onto W, denoted as $\beta = \operatorname{proj}_W \alpha$.

> *Definition 10.9 (Best Approximation Element)*
>
> Let V be a real inner product space, and let W be a subspace of V. For any vector $\alpha \in V$, if there exists a vector $\beta \in W$ such that:
> $$\|\alpha - \beta\| = \min_{\gamma \in W} \|\alpha - \gamma\|,$$
> then β is called the **best approximation element** of α in W.

🖋 **Property** *In Euclidean space V, let W be an n-dimensional subspace of V. Take an orthogonal basis $\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\}$ of W.*

1. *For any vector $\alpha \in V$, the orthogonal projection of α onto W exists and is unique, and it is also the best approximation element of α in W;*

2. *The orthogonal projection of α onto W can be expressed as:*
$$\operatorname{proj}_W \alpha = \sum_{i=1}^{n} \frac{(\alpha, \varepsilon_i)}{(\varepsilon_i, \varepsilon_i)} \varepsilon_i =: \sum_{i=1}^{n} c_i \varepsilon_i.$$

3. *The remainder of the orthogonal projection satisfies:*
$$\|\alpha - \operatorname{proj}_W \alpha\|^2 = \|\alpha\|^2 - \|\operatorname{proj}_W \alpha\|^2 = \|\alpha\|^2 - \sum_{i=1}^{n} c_i^2 \|\varepsilon_i\|^2.$$

10.6 Orthogonal Transformations and Symmetric Transformations

¶ `Orthogonal Transformations`

¶ `Symmetric Transformations`

> *Definition 10.10 (Symmetric Transformation)*
>
> Let V be a real inner product space, and let $\mathcal{A} \in \operatorname{Hom}(V)$. If
> $$(\mathcal{A}\alpha, \beta) = (\alpha, \mathcal{A}\beta), \quad \forall \alpha, \beta \in V,$$
> then $\mathcal{A}$ is called a **symmetric transformation** on V.

Obviously, symmetric transformations are linear transformations.

> *Lemma 10.1*
>
> 1. If A is real symmetric matrix, then A has and only has n real eigenvalues (counting multiplicities).
>
> 2. Let V be a n-dimensional Euclidean space, then any linear transformation $\mathcal{A} : V \to V$ is symmetric if

and only if there exists a orthonormal basis of V such that the matrix representation of $\mathcal{A}$ under this basis is a symmetric matrix.

3. If $\mathcal{A}$ is a symmetric transformation on a real inner product space V, then if W is an invariant subspace of V under $\mathcal{A}$, then its orthogonal complement $W^{\perp}$ is also an invariant subspace of V under $\mathcal{A}$.

4. If A is a symmetric transformation on a Euclidean space V, then eigenvectors corresponding to distinct eigenvalues are definitely orthogonal.

♡

Theorem 10.4 (Orthogonal Diagonalization of Symmetric Transformations)

Let V be a n-dimensional Euclidean space, and let $\mathcal{A} : V \to V$ be a symmetric transformation on V. Then there exists a orthonormal basis of V such that the matrix representation of $\mathcal{A}$ under this basis is a diagonal matrix. From the perspective of matrices, let $A \in \mathbb{R}^{n \times n}$ be a real symmetric matrix. Then there exists an orthogonal matrix T such that:

$$T^{-1}AT = T^{\mathrm{T}}AT = \Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix},$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A.

♡

🏺 **Note** *According to the theorem above, for all real quadratic forms $f(x_1, x_2, \ldots, x_n) = X^{\mathrm{T}}AX$, there exists an orthogonal transformation $X = TY$ such that:*

$$f(x_1, x_2, \ldots, x_n) = Y^{\mathrm{T}}\Lambda Y = \sum_{i=1}^{n} \lambda_i y_i^2.$$

10.7 Unitary Spaces and Normal Transformations

¶ `Unitary Spaces`

Definition 10.11 (Complex Inner Product Space)

A **complex inner product space** is a complex linear space V equipped with a function $(\cdot, \cdot) : V \times V \to \mathbb{C}$ satisfying the following properties:

Positivity $(\alpha, \alpha) \geq 0, \quad \forall \alpha \in V$, and $(\alpha, \alpha) = 0$ if and only if $\alpha = 0$;

Conjugate Symmetry (Hermitian) $(\alpha, \beta) = \overline{(\beta, \alpha)}, \quad \forall \alpha, \beta \in V$;

Linearity in the First Argument $(k_1\alpha_1 + k_2\alpha_2, \beta) = k_1(\alpha_1, \beta) + k_2(\alpha_2, \beta), \quad \forall k_1, k_2 \in \mathbb{C}, \alpha_1, \alpha_2, \beta \in V$.

The function $(\cdot, \cdot)$ is called the (complex) **inner product** on V.

Complex inner product spaces with finite dimensions are called **unitary spaces**.

♣

The norm in unitary spaces can be also induced by the inner product:

$$\|\alpha\| = (\alpha, \alpha)^{\frac{1}{2}}, \quad \forall \alpha \in V.$$

¶ `Normal Transformation and Diagonalization`

Define the conjugate transpose (Hermitian adjoint) of a complex matrix $A \in \mathbb{C}^{m \times n}$ as:

$$A^{\dagger} = \overline{A}^{\mathrm{T}}.$$

Similarly, for a linear transformation $\mathcal{A} \in \mathrm{Hom}(V)$ on a complex inner product space V, the **adjoint transformation** $\mathcal{A}^\dagger$ of $\mathcal{A}$ is defined as:

$$(\mathcal{A}\alpha, \beta) = (\alpha, \mathcal{A}^\dagger\beta), \quad \forall \alpha, \beta \in V.$$

Definition 10.12 (Normal Transformation and Normal Matrix)

Let V be a complex inner product space, and let $\mathcal{A} \in \mathrm{Hom}(V)$. If

$$\mathcal{A}\mathcal{A}^\dagger = \mathcal{A}^\dagger\mathcal{A},$$

then $\mathcal{A}$ is called a **normal transformation** on V.

A square matrix $A \in \mathbb{C}^{n \times n}$ is called a **normal matrix** if:

$$AA^\dagger = A^\dagger A.$$

♣

Theorem 10.5

Let $\mathcal{A}$ be a normal transformation on a unitary space V. Then there exists an orthonormal basis of V such that the matrix representation of $\mathcal{A}$ under this basis is a diagonal matrix.

From the perspective of matrices, let $A \in \mathbb{C}^{n \times n}$ be a normal matrix. Then there exists a unitary matrix U such that:

$$U^{-1}AU = U^\dagger AU = \Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix},$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A.

♡

¶ **Special Normal Transformation**

Definition 10.13 (Unitary Transformation and Unitary Matrix)

Let V be a complex inner product space, and let $\mathcal{A} \in \mathrm{Hom}(V)$. If

$$\mathcal{A}^\dagger = \mathcal{A}^{-1}, \quad \text{or} \quad (\mathcal{A}\alpha, \mathcal{A}\beta) = (\alpha, \beta), \quad \forall \alpha, \beta \in V,$$

then $\mathcal{A}$ is called a **unitary transformation** on V.

A square matrix $A \in \mathbb{C}^{n \times n}$ is called a **unitary matrix** if:

$$A^\dagger = A^{-1}, \quad \text{or} \quad AA^\dagger = A^\dagger A = E.$$

♣

Obviously, unitary transformations are generalizations of orthogonal transformations in real inner product spaces.

Definition 10.14 (Hermitian Transformation and Hermitian Matrix)

Let V be a complex inner product space, and let $\mathcal{A} \in \mathrm{Hom}(V)$. If

$$(\mathcal{A}\alpha, \beta) = (\alpha, \mathcal{A}\beta), \quad \forall \alpha, \beta \in V,$$

then $\mathcal{A}$ is called a **Hermitian transformation** on V.

A square matrix $A \in \mathbb{C}^{n \times n}$ is called a **Hermitian matrix** if:

$$A^\dagger = A.$$

♣

Obviously, Hermitian transformations are generalizations of symmetric transformations in real inner product spaces.

10.8 Symplectic Spaces

Bibliography

[1] 丘维声. 高等代数 *(2nd edition)* , 清华大学出版社, 2019.

[2] 谢启鸿, 姚慕生, 吴泉水. 高等代数学 *(4th edition)* , 复旦大学出版社, 2022.

[3] 王萼芳, 石生明. 高等代数 *(5th edition)* , 高等教育出版社, 2019.

[4] 樊启斌. 高等代数典型问题与方法 *(1st edition)* , 高等教育出版社, 2021.

[5] Wikipedia. https://en.wikipedia.org/wiki/.