Algebra IV: Linear

Gudfit

Contents

0

Review: Systems of Linear Equations

The primary objective of linear algebra is to solve the linear equation $Ax = b$. Many practical linear problems can be modelled in this manner.

0.1 *Matrices and Linear Systems*

I expect you to know the very basics of matrices; this includes sums, multiplication etc.

An array of objects is a collection where we track the row and column of each object. Formally, a finite sequence is a function from $\{1, \ldots, n\}$ to a set S. Similarly, an $m \times n$ array of objects in S corresponds to a function $a : \{1, \ldots, m\} \times \{1, \ldots, n\} \to S$, where $a(i, j)$ is the entry in the i-th row and j-th column.

Definition 0.1. ***Matrix.***
An $m \times n$ **matrix** is an array of objects with m rows and n columns. If A is an $m \times n$ matrix, then A_{ij} denotes the entry in the i-th row and j-th column. We write:

$$A = [A_{ij}] = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix}$$

The i-th row of A is denoted $\text{row}_i(A)$, and the j-th column is $\text{col}_j(A)$. The set of $m \times n$ matrices with entries in a set S is denoted $S^{m \times n}$. Specifically, $\mathbb{R}^{m \times n}$ and $\mathbb{C}^{m \times n}$ denote matrices with real and complex entries, respectively.

定義

Note

Distinguish between the matrix A (the array) and the entry A_{ij} (a scalar). The notation $A = [A_{ij}]$ indicates that A is the collection of all such components.

Example 0.1. Matrix Types.

- Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \in \mathbb{R}^{3\times 3}$.
- Let $B = \begin{bmatrix} i & 10 \\ 0 & 3+i \\ 11 & 12 \end{bmatrix} \in \mathbb{C}^{3\times 2}$.
- Matrices may contain objects other than numbers. For $f, g, h : \mathbb{R} \to \mathbb{R}$, $M = \begin{bmatrix} f & g \\ h & f \end{bmatrix}$ is a matrix of functions.

範例

Definition 0.2. ***Vector.***
A **vector** in $\mathbb{R}^n$ is an ordered list of n real numbers. We write vectors as columns:

$$v = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \in \mathbb{R}^n.$$

定義

Definition 0.3. ***System of Linear Equations.***
We use matrices to solve linear problems. Let $A \in \mathbb{R}^{m\times n}$. Let $b \in \mathbb{R}^m$ be a given vector, often called the **right-hand side vector**. We seek an unknown vector $x \in \mathbb{R}^n$ such that

$$Ax = b.$$

Explicitly, this represents a system of m linear equations in n unknowns:

$$\sum_{j=1}^{n} A_{ij}x_j = b_i \quad \text{for } i = 1, \dots, m.$$

定義

Definition 0.4. ***Matrix Equality.***
Two matrices A and B are equal, written $A = B$, if and only if they have the same dimensions and corresponding entries are identical: $A_{ij} = B_{ij}$ for all i, j.

定義

Notation 0.1. Block Matrices We often construct larger matrices by concatenating smaller ones. If A is $m \times n$ and B is $m \times k$, we write $[A \mid B]$ for the $m \times (n+k)$ matrix formed by placing B to the right of A.

記法

To develop intuition for linear systems, we first consider the case where $n = m = 2$, allowing for geometric visualization. Each equation in the system $Ax = b$ represents a line in the Euclidean plane $\mathbb{R}^2$.
Consider the system:

$$\begin{aligned} 2x_1 + 3x_2 &= 1, \\ 3x_1 + 2x_2 &= 2. \end{aligned}$$

Here, $A = \begin{bmatrix} 2 & 3 \\ 3 & 2 \end{bmatrix}$ and $b = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

· The first line passes through $(1/2, 0)$ and $(0, 1/3)$.
· The second line passes through $(2/3, 0)$ and $(0, 1)$.

Since the slopes are distinct ($-2/3$ vs $-3/2$), the lines are not parallel and must intersect at a unique point. This point represents the unique solution to the system.
However, not all systems possess a unique solution. Consider the system:

$$\begin{aligned} 2x_1 + 3x_2 &= 1, \\ 4x_1 + 6x_2 &= 2. \end{aligned}$$

The second equation is merely twice the first. Geometrically, these equations describe the same line. Consequently, every point on the line is a solution. The solution set S is infinite:

$$S = \left\{ (\alpha, \beta) \in \mathbb{R}^2 \;\middle|\; \beta = \frac{1}{3}(1 - 2\alpha),\ \alpha \in \mathbb{R} \right\}.$$

Finally, consider the inconsistent system:

$$\begin{aligned} 2x_1 + 3x_2 &= 1, \\ 4x_1 + 6x_2 &= 3. \end{aligned}$$

Here, the lines have identical slopes but distinct intercepts. They are parallel and never intersect. The solution set is empty.
This geometric intuition is valuable but limited to two or three dimensions. For systems with many variables, we require an algebraic method that can be automated.

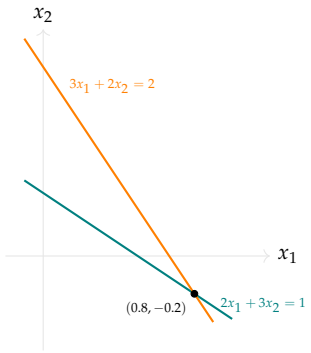


Figure 1: Geometric interpretation of a unique solution.

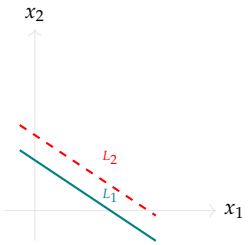


Figure 2: Parallel lines represent an inconsistent system.

0.2 *Gaussian Elimination*

We aim to formalise the method of elimination familiar from elementary algebra. The goal is to transform a system $Ax = b$ into an equivalent system $Cx = d$ where the matrix C has a structure that makes the system easy to solve. Specifically, we seek a C that is **upper triangular**.

Definition 0.5. ***Upper Triangular Matrix.***
A square matrix C is called **upper triangular** if all entries below the principal diagonal (the entries $C_{11}, C_{22}, \dots, C_{nn}$) are zero, i.e., $C_{ij} = 0$ for all $i > j$.

定義

Consider the reduction of our first example:

$$\text{(System 1)} \quad \begin{cases} 2x_1 + 3x_2 = 1 \\ 3x_1 + 2x_2 = 2 \end{cases}$$

Multiplying the first equation by 3 and the second by 2 yields $6x_1 + 9x_2 = 3$ and $6x_1 + 4x_2 = 4$. Subtracting the modified second from the first eliminates x_1. A more systematic approach (Gaussian elimination) typically proceeds to eliminate variables from subsequent equations. Subtracting $\frac{3}{2}$ times the first equation from the second yields:

$$\text{(System 2)} \quad \begin{cases} 2x_1 + 3x_2 = 1 \\ \quad -\frac{5}{2}x_2 = \frac{1}{2} \end{cases}$$

The coefficient matrix is now $C = \begin{bmatrix} 2 & 3 \\ 0 & -2.5 \end{bmatrix}$, which is upper triangular.

Back Substitution

Once the system is in the form $Cx = d$ with C upper triangular, we can solve it by **backward substitution**. Writing out the equations for an $n \times n$ system:

$$\begin{aligned} C_{11}x_1 + C_{12}x_2 + \cdots + C_{1n}x_n &= d_1 \\ C_{22}x_2 + \cdots + C_{2n}x_n &= d_2 \\ \vdots \quad &= \vdots \\ C_{nn}x_n &= d_n \end{aligned}$$

Provided $C_{nn} \neq 0$, we find $x_n = d_n / C_{nn}$. We then substitute this into the $(n-1)$-th equation to find x_{n-1}, and proceed recursively:

$$x_i = \frac{1}{C_{ii}} \left(d_i - \sum_{j=i+1}^{n} C_{ij} x_j \right).$$

Elementary Row Operations

To perform this reduction systematically, we define specific operations on the rows of the matrix A. These are the **Elementary Row**

Operations (EROs). We view an elementary row operation as a function $e : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$. There are three types of operations.

Definition 0.6. ***Type I: Scaling.***
Multiply the s-th row by a non-zero constant $\alpha \in \mathbb{R} \setminus \{0\}$. The function e_1 acts on A to produce a matrix A' with entries:

$$A'_{ij} = \begin{cases} \alpha A_{sj} & \text{if } i = s, \\ A_{ij} & \text{if } i \neq s. \end{cases}$$

定義

Definition 0.7. ***Type II: Replacement.***
Replace the s-th row by the sum of the s-th row and a multiple α of the t-th row (where $s \neq t$). The function e_2 produces A' with entries:

$$A'_{ij} = \begin{cases} A_{sj} + \alpha A_{tj} & \text{if } i = s, \\ A_{ij} & \text{if } i \neq s. \end{cases}$$

定義

Definition 0.8. ***Type III: Interchange.***
Interchange the s-th row and the t-th row. The function e_3 produces A' with entries:

$$A'_{ij} = \begin{cases} A_{tj} & \text{if } i = s, \\ A_{sj} & \text{if } i = t, \\ A_{ij} & \text{otherwise.} \end{cases}$$

定義

These operations are fundamental because they preserve the solution set of the linear system. Furthermore, they are reversible.

Proposition 0.1. ***Invertibility of EROs.***
Each elementary row operation e is a bijection. Its inverse e^{-1} is also an elementary row operation of the same type.

命題

Proof

We construct the inverses explicitly:

1. If e_1 multiplies row s by $\alpha \neq 0$, then the operation that multiplies row s by $1/\alpha$ recovers the original matrix. This is a Type I operation.
2. If e_2 adds α times row t to row s, then adding $-\alpha$ times row t to row s reverses the effect. This is a Type II operation.

3. If e_3 swaps rows s and t, applying the same swap again restores the original order. This is a Type III operation.

Since an inverse exists for every A, the map is a bijection.

■

Proposition 0.2. ***EROs Preserve Solution Sets.***
Let $[A \mid b]$ be the augmented matrix of a system $Ax = b$, and let $[A' \mid b']$ be obtained from $[A \mid b]$ by a single elementary row operation. Then $Ax = b$ and $A'x = b'$ have the same solution set.

命題

Proof

Each elementary row operation replaces one equation by an equivalent linear combination of the equations, or swaps equations. Any solution of $Ax = b$ satisfies all linear combinations of its equations, so it satisfies the transformed system. Since the operation is invertible, the reverse implication also holds. Thus the solution sets agree.

■

These operations allow us to transform matrices into simpler forms (Row Echelon Form) without altering the underlying relationships between variables.

Definition 0.9. ***Row Echelon Form.***
A matrix is in **row echelon form** if:
1. All non-zero rows are above any zero rows.
2. The leading (first non-zero) entry of each non-zero row is to the right of the leading entry of the row above it.
3. All entries below each leading entry are zero.

定義

0.3 *Equivalent Systems*

We now investigate the conditions under which two distinct systems of linear equations possess the same solution set. This concept is central to justifying the Gaussian elimination process.

Linear Combinations of Equations

Consider the system of m linear equations in n variables, denoted by (1):

$$\sum_{j=1}^{n} A_{ij}x_j = b_i \quad \text{for } i = 1, \ldots, m.$$

Definition 0.10. ***Linear Combination.***
Given equations $E_1, \ldots, E_m$, a **linear combination** is any equation of the form

$$\alpha_1 E_1 + \alpha_2 E_2 + \cdots + \alpha_m E_m,$$

where $\alpha_1, \ldots, \alpha_m$ are scalars.

定義

We can form a new equation by taking a linear combination of these m equations. Let $\alpha_1, \ldots, \alpha_m$ be scalars. Multiplying the i-th equation by α_i and summing over all i yields:

$$\sum_{i=1}^{m} \alpha_i \left(\sum_{j=1}^{n} A_{ij} x_j \right) = \sum_{i=1}^{m} \alpha_i b_i.$$

Rearranging the terms to group coefficients of each x_j, we obtain a new linear equation:

$$\sum_{j=1}^{n} \left(\sum_{i=1}^{m} \alpha_i A_{ij} \right) x_j = \sum_{i=1}^{m} \alpha_i b_i.$$

It is immediate that any solution x satisfying the original system (1) must also satisfy this new combined equation.
Now, consider a second system of k equations, denoted by (2):

$$\sum_{j=1}^{n} C_{pj} x_j = d_p \quad \text{for } p = 1, \ldots, k.$$

If every equation in system (2) can be obtained as a linear combination of the equations in system (1), then any solution to (1) is automatically a solution to (2). The converse holds if every equation in (1) is a linear combination of the equations in (2). This mutual dependency leads to the definition of equivalent systems.

Definition 0.11. ***Equivalent Systems.***
Two linear systems are said to be **equivalent** if every equation of each system can be expressed as a linear combination of the equations of the other system.

定義

Theorem 0.1. ***Equivalence Theorem.***
Equivalent linear systems have the same solution set.

定理

Proof

Let S_1 and S_2 be the solution sets of system (1) and system (2) respectively. If every equation in (2) is a linear combination of equa-

tions in (1), then any $x \in S_1$ satisfies all equations in (2). Thus $S_1 \subseteq S_2$. Conversely, if every equation in (1) is a linear combination of equations in (2), then any $x \in S_2$ satisfies all equations in (1). Thus $S_2 \subseteq S_1$. Therefore, $S_1 = S_2$.

■

Example 0.2. Non-Equivalent Systems. Consider system (1):

$$\begin{aligned} x_1 + 2x_2 + 5x_3 &= 0 \\ x_1 + 3x_2 + 8x_3 &= 0 \\ -x_1 + x_2 + 4x_3 &= 0 \end{aligned}$$

And system (2):

$$\begin{aligned} x_2 + 2x_3 &= 0 \\ x_1 - x_3 &= 0 \\ x_1 + x_3 &= 0 \end{aligned}$$

System (2) has the unique solution $x_1 = x_2 = x_3 = 0$. System (1), however, admits the non-trivial solution $x_1 = 1, x_2 = -3, x_3 = 1$. Since the solution sets differ ($S_2 \subsetneq S_1$), the systems are not equivalent. Specifically, the third equation of system (2) cannot be a linear combination of the equations of system (1).

範例

Row Equivalence

The algebraic process of taking linear combinations of equations corresponds directly to performing row operations on the augmented matrix of the system. We now formalise the relationship between matrices.

Definition 0.12. ***Row Equivalence.***
A matrix A is **row equivalent** to a matrix B, denoted $A \sim B$, if B can be obtained from A by a finite sequence of elementary row operations.

定義

Since each elementary row operation is invertible (*proposition* 0.1), the relation $\sim$ is symmetric. Since the identity operation is an ERO, it is reflexive. Since the composition of finite sequences is a finite sequence, it is transitive. Thus, row equivalence is an **equivalence relation** on the set of $m \times n$ matrices.
The connection between these two concepts is fundamental.

Theorem 0.2. ***Row Equivalence and Solutions.***
If two matrices A and C are row equivalent, then the homogeneous systems $Ax = 0$ and $Cx = 0$ have the same solution set.

定理

Proof

It suffices to prove this for a single elementary row operation, as the general result follows by induction. Let C be obtained from A by one operation e.

1. Each row of C is a linear combination of the rows of A (by definition of EROs). Thus, any solution to $Ax = 0$ is a solution to $Cx = 0$.
2. Since e is invertible, A can be obtained from C by the inverse operation e^{-1}. Thus, each row of A is a linear combination of the rows of C. Hence, any solution to $Cx = 0$ is a solution to $Ax = 0$.

The solution sets are therefore identical.

■

Remark.

For non-homogeneous systems $Ax = b$ and $Cx = d$, the systems are equivalent if the augmented matrices $[A|b]$ and $[C|d]$ are row equivalent.

Remark.

Equivalence in this sense is stronger than merely having the same solution set, but it is the notion naturally preserved by row operations.

Row Reduced Echelon Form

We now define the specific "simple structure" we aim to achieve through Gaussian elimination.

Definition 0.13. ***Row Reduced Echelon Form (RREF).***
A matrix R is called a **row reduced echelon matrix** (i.e., it is in **row reduced echelon form**, RREF) if it satisfies the following four conditions:

Leading Ones: The first non-zero entry of each non-zero row is 1. This entry is called the **leading entry** or pivot.

Zero Rows: All rows consisting entirely of zeros appear below all non-zero rows.

Pivot Columns: If a column contains a leading entry of some row, then all other entries in that column are 0.

Stepped Structure: Let the leading entry of the i-th non-zero row appear in column c_i. Then $c_1 < c_2 < c_3 < \cdots < c_r$. That is, the

leading entry of a lower row always appears to the right of the leading entry of a higher row.

定義

Example 0.3. Examples of RREF. The matrix

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

is in row reduced echelon form. The matrix

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

is also in RREF. The pivot columns are 1, 2, and 4.

範例

Example 0.4. Non-Examples.

- $\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$: Fails condition 4 (stepped structure). The pivot in row 2 is to the left of the pivot in row 1.
- $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \end{bmatrix}$: Fails condition 1. The leading entry of row 2 is 2, not 1.
- $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}$: Fails condition 3. Column 2 contains a leading entry but is not zero elsewhere.

範例

The utility of the RREF lies in the ease of reading off the solution set. For a system $Rx = 0$, the variables corresponding to pivot columns are called **basic variables**, while the others are **free variables**. The basic variables can be expressed explicitly in terms of the free variables.

Example 0.5. Solving a System. Solve $Ax = 0$ where

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

範例

Solution

Using row operations, we reduce A to the RREF:

$$R = \begin{bmatrix} 1 & 0 & 1/4 & 3/4 \\ 0 & 1 & -5/4 & -3/4 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The pivot columns are 1 and 2. Thus x_1, x_2 are basic, and x_3, x_4 are free. The equations are:

$$x_1 + \frac{1}{4}x_3 + \frac{3}{4}x_4 = 0 \implies x_1 = -\frac{1}{4}x_3 - \frac{3}{4}x_4$$
$$x_2 - \frac{5}{4}x_3 - \frac{3}{4}x_4 = 0 \implies x_2 = \frac{5}{4}x_3 + \frac{3}{4}x_4$$

Setting $x_3 = \alpha, x_4 = \beta$, the general solution is:

$$x = \alpha \begin{bmatrix} -1/4 \\ 5/4 \\ 1 \\ 0 \end{bmatrix} + \beta \begin{bmatrix} -3/4 \\ 3/4 \\ 0 \\ 1 \end{bmatrix}, \quad \alpha, \beta \in \mathbb{R}.$$

■

0.4 *Matrix Rank and System Consistency*

We have seen that reducing a matrix to its Row Reduced Echelon Form (RREF) provides a systematic way to solve linear systems. A central concept emerging from this process is the "rank" of a matrix, which essentially counts the number of independent constraints imposed by the linear system.
Before we rely on RREF for theoretical results, we must ensure that every matrix actually possesses one.

Theorem 0.3. ***Existence of RREF.***
Any $m \times n$ matrix A is row equivalent to a matrix in Row Reduced Echelon Form (RREF).

定理

Proof

We proceed by induction on the number of rows m. If A is the zero matrix, it is already in RREF. Suppose A is non-zero. Let k be the index of the first non-zero column of A. Select a row i such that $A_{ik} \neq 0$. By interchanging row 1 and row i, we ensure the entry in position $(1, k)$ is non-zero. Multiply row 1 by $1/A_{1k}$ to make the leading entry 1. Then, for each row $j > 1$, subtract A_{jk} times row 1 from row j. This clears all entries in column k below the first row. Now, consider the submatrix consisting of rows 2 through m. By

the induction hypothesis, this submatrix can be reduced to RREF. Finally, use row operations to clear any non-zero entries in the pivot columns of the submatrix that lie in row 1. The result satisfies all conditions of an RREF.

■

Remark.

While we stated existence, it is a non-trivial fact that the RREF of a matrix is **unique**. That is, row operations can take different paths, but they always lead to the same reduced matrix. We will assume this uniqueness for now.

Homogeneous Systems

Consider the homogeneous system $Ax = 0$. Let R be the RREF of A. Let r be the number of non-zero rows in R. This integer r is called the **row rank** of A. We will use $\mathbf{rank}(A)$ to mean this number. The variables corresponding to the leading ones (pivots) are determined by the system, while the remaining $n - r$ variables are "free".

Theorem 0.4. ***Non-Trivial Solutions.***
A homogeneous system of linear equations with fewer equations than unknowns always has a non-trivial solution.

定理

Proof

Let the system be $Ax = 0$, where A is $m \times n$ with $m < n$. Let R be the RREF of A. The number of non-zero rows r satisfies $r \leq m < n$. The number of free variables is $n - r > 0$. Since there is at least one free variable, we can assign it a non-zero value (e.g., 1), which determines the pivot variables uniquely. This constructs a solution where $x \neq 0$.

■

This result is fundamental. It tells us, for instance, that any set of $n + 1$ vectors in $\mathbb{R}^n$ must be linearly dependent, a concept we will formalise in the next chapter.

The Square Case

For a square matrix (n equations in n unknowns), the existence of non-trivial solutions is linked to the invertibility of the matrix.

Theorem 0.5. ***Square Homogeneous Systems.***
Let A be an $n \times n$ matrix. The system $Ax = 0$ has only the trivial solution $x = 0$ if and only if A is row equivalent to the identity matrix I_n.

定理

(⇐)

If $A \sim I_n$, then the system $Ax = 0$ is equivalent to $I_n x = 0$, which implies $x = 0$.

証明終

(⇒)

Suppose $Ax = 0$ has only the trivial solution. Let R be the RREF of A. The system $Rx = 0$ also has only the trivial solution. This implies there are no free variables. Thus, every column must be a pivot column. Since R is $n \times n$ and has n pivots, it must be I_n.

証明終

Non-Homogeneous Systems

We now turn to the general case $Ax = b$. To analyse this, we form the **augmented matrix** $[A \mid b]$ by appending b as an extra column. We perform row operations on this augmented matrix to obtain $[R \mid d]$, where R is the RREF of A.
The system $Ax = b$ is equivalent to $Rx = d$. Let r be the number of non-zero rows of R. The equations corresponding to the zero rows of R (rows $r + 1$ to m) take the form:

$$0 \cdot x_1 + \cdots + 0 \cdot x_n = d_i \quad \text{for } i = r + 1, \ldots, m.$$

These equations are satisfied if and only if $d_i = 0$.

Theorem 0.6. ***Consistency Condition.***
The system $Ax = b$ is consistent (has a solution) if and only if the vector d in the reduced augmented matrix satisfies $d_i = 0$ for all $i > r$, where r is the number of non-zero rows in the RREF of A.

定理

Proof

Let $[A \mid b]$ be the augmented matrix and let $[R \mid d]$ be its RREF, where R is the RREF of A. Row operations preserve solution sets, so $Ax = b$ is consistent if and only if $Rx = d$ is consistent.
By definition of r, the rows $r + 1, \ldots, m$ of R are zero rows. The corresponding equations in $Rx = d$ are

$$0 = d_i \quad \text{for } i = r + 1, \ldots, m.$$

If any $d_i \neq 0$ for $i > r$, the system is inconsistent. Conversely, if all $d_i = 0$ for $i > r$, then the remaining r equations (the non-zero rows) involve the pivot and free variables and always admit a solution.
Hence $Ax = b$ is consistent exactly when $d_i = 0$ for all $i > r$.

■

This condition effectively says that the rank of the augmented matrix

$[A \mid b]$ must equal the rank of A. If $\operatorname{rank}([A \mid b]) > \operatorname{rank}(A)$, the system is inconsistent.

Example 0.6. Consistent and Inconsistent Systems. Consider the system $Ax = b$ with

$$A = \begin{bmatrix} 1 & -2 & 1 \\ 2 & -4 & 2 \\ 1 & 1 & -3 \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} 2 \\ 4 \\ 2 \end{bmatrix}.$$

範例

Solution

Row reducing the augmented matrix:

$$\left[\begin{array}{ccc|c} 1 & -2 & 1 & 2 \\ 2 & -4 & 2 & 4 \\ 1 & 1 & -3 & 2 \end{array}\right] \xrightarrow{R_2-2R_1,\, R_3-R_1} \left[\begin{array}{ccc|c} 1 & -2 & 1 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 3 & -4 & 0 \end{array}\right].$$

The system is consistent since no row of the form $[0\ 0\ 0 \mid k]$ with $k \neq 0$ appears.

Now keep the same A but take $b = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$. Row reduction yields a row of the form $[0\ 0\ 0 \mid k]$ with $k \neq 0$, so the system is inconsistent. ■

Example 0.7. Consistency condition. Find the condition on b_1, b_2, b_3 for the following system to be consistent:

$$\begin{aligned} x + y + z &= b_1 \\ 2x + 2y + 2z &= b_2 \\ 3x + 3y + 3z &= b_3 \end{aligned}$$

Row reduction yields zero rows on the left. Consistency requires $b_2 - 2b_1 = 0$ and $b_3 - 3b_1 = 0$.

範例

0.5 *Elementary Matrices*

We now introduce the concept of elementary matrices, which provides a matrix-algebraic perspective on elementary row operations. This formalisation is not only theoretically elegant but also practical, as it allows us to represent row reduction as matrix multiplication—a form readily implemented in computational algorithms.

Definition 0.14. ***Elementary Matrix.***
An **elementary matrix** of order n is a matrix obtained by performing a single elementary row operation on the $n \times n$ identity matrix I_n.

定義

Note

The **order** (or size) of a matrix is its number of rows by number of columns. An $m \times n$ matrix has order $m \times n$, a **square matrix of order** n is an $n \times n$ matrix.

Definition 0.15. ***Identity Matrix.***
The **identity matrix** in $\mathbb{R}^{n\times n}$ is the $n \times n$ matrix I with entries

$$I_{ij} = \delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

We write I_n when the size needs emphasis. The symbol δ_{ij} is called the **Kronecker delta**.

定義

Since there are three types of elementary row operations, there are three types of elementary matrices.

Example 0.8. Examples of 2×2 Elementary Matrices. Starting from

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} :$$

1. **Type I (Scaling):** Multiply row 1 by $\alpha \neq 0$:

$$E_1 = \begin{bmatrix} \alpha & 0 \\ 0 & 1 \end{bmatrix}.$$

2. **Type II (Replacement):** Add α times row 2 to row 1:

$$E_2 = \begin{bmatrix} 1 & \alpha \\ 0 & 1 \end{bmatrix}.$$

3. **Type III (Interchange):** Swap row 1 and row 2:

$$E_3 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

範例

The fundamental property of elementary matrices is that they implement row operations via left multiplication.

Theorem 0.7. ***Matrix Multiplication Implements Row Operations.*** Let A be an $m \times n$ matrix, and let e be an elementary row operation. Let $E = e(I_m)$ be the corresponding elementary matrix. Then performing the operation e on A is equivalent to pre-multiplying A by E. That is:

$$e(A) = EA.$$

定理

Proof

We verify this for the Type II operation (Replacement). Let e be the operation "Replace row s by row $s \ + \ \alpha\times$ row t". The entries of the elementary matrix $E = e(I_m)$ are given by:

$$E_{ik} = \begin{cases} 1 & \text{if } i = k, \\ \alpha & \text{if } i = s, k = t, \\ 0 & \text{otherwise.} \end{cases}$$

Using the definition of matrix multiplication, the (i, j)-th entry of the product EA is:

$$(EA)_{ij} = \sum_{k=1}^{m} E_{ik} A_{kj}.$$

- If $i \neq s$, the only non-zero term in the sum is when $k = i$, where $E_{ii} = 1$. Thus $(EA)_{ij} = 1 \cdot A_{ij} = A_{ij}$. This matches the fact that rows other than s are unchanged.
- If $i = s$, the sum has two non-zero terms: $k = s$ (where $E_{ss} = 1$) and $k = t$ (where $E_{st} = \alpha$). Thus:

  $$(EA)_{sj} = 1 \cdot A_{sj} + \alpha \cdot A_{tj}.$$

  This is precisely the definition of the row operation e applied to A.

The proofs for Type I and Type III operations follow similarly.

■

This theorem allows us to view row reduction as a factorisation process. If a matrix B is obtained from A by a sequence of operations corresponding to elementary matrices $E_1, E_2, \ldots, E_k$, then:

$$B = E_k E_{k-1} \ldots E_1 A.$$

Invertibility

The relationship between elementary matrices and invertibility is central to linear algebra.

Proposition 0.3. ***Identity Acts as a Multiplicative Unit.***
If $X \in \mathbb{R}^{n\times p}$, then $XI_p = X$ and $I_n X = X$.

命題

Proof

For brevity, write $I = I_p$. The (i, j) entry of XI is

$$(XI)_{ij} = \sum_{k=1}^{p} X_{ik} I_{kj} = \sum_{k=1}^{p} X_{ik}\delta_{kj} = X_{ij},$$

since all other terms vanish. Thus $XI = X$. The proof of $I_n X = X$ is analogous.

■

Thus

Definition 0.16. ***Invertible Matrix.***
A square matrix A of order n is **invertible** if there exists a square matrix B of order n such that

$$AB = I_n \quad \text{and} \quad BA = I_n.$$

In this case, B is unique and is called the **inverse** of A, denoted A^{-1}.

定義

Proposition 0.4. ***Invertibility of Elementary Matrices.***
Every elementary matrix is invertible, and its inverse is an elementary matrix of the same type.

命題

Proof

Let E be an elementary matrix corresponding to the row operation e. From our discussion on row operations, we know e is a bijection with an inverse operation e^{-1} which is also an elementary row operation. Let $E' = e^{-1}(I)$. Then for any matrix X, $E'X$ performs e^{-1} on X. Consider the product $E'E$. This corresponds to performing e on I (getting E), and then performing e^{-1} on the result. Thus $E'E = e^{-1}(e(I)) = I$. Similarly, $EE' = I$. Thus E is invertible with inverse E'.

■

Characterisation of Invertible Matrices

We can now state the main theorem linking systems of equations, row reduction, and matrix inversion.

Theorem 0.8. ***The Invertible Matrix Theorem.***
Let A be a square matrix of order n. The following statements are equivalent:

1. A is invertible.
2. A is row equivalent to the identity matrix I_n.
3. A is a product of elementary matrices.
4. The homogeneous system $Ax = 0$ has only the trivial solution $x = 0$.
5. The system $Ax = b$ has a solution for every $b \in \mathbb{R}^n$.

定理

Proof

We prove the cyclic implications $(1) \implies (2) \implies (3) \implies (1)$, and link (4) and (5) separately.

(1) $\implies$ ***(2):*** Suppose A is invertible. Let R be the row reduced echelon form of A. Since row operations correspond to multiplication by invertible elementary matrices, $R = PA$ for some invertible matrix P. Since P and A are invertible, R is invertible. As established previously (*theorem* 0.5), the only invertible matrix in RREF of order n is I_n. Thus $A \sim I_n$.

(2) $\implies$ ***(3):*** If $A \sim I_n$, then $I_n = E_k \dots E_1 A$ for elementary matrices E_i. Multiplying by inverses:

$$A = E_1^{-1} \dots E_k^{-1} I_n = E_1^{-1} \dots E_k^{-1}.$$

Since the inverse of an elementary matrix is an elementary matrix, A is a product of elementary matrices.

(3) $\implies$ ***(1):*** If A is a product of elementary matrices, and each elementary matrix is invertible, then their product A is invertible.

(1) $\iff$ ***(4):*** If A is invertible, $Ax = 0 \implies A^{-1}Ax = A^{-1}0 \implies x = 0$. Conversely, if $Ax = 0$ has only the trivial solution, then $A \sim I_n$ (by *theorem* 0.5), which implies A is invertible by (2).

(1) $\iff$ ***(5):*** If A is invertible, for any b, let $x = A^{-1}b$. Then $Ax = A(A^{-1}b) = b$. Thus a solution exists. Conversely, if $Ax = b$ always has a solution, then for each column e_j of the identity matrix I_n, there exists a vector c_j such that $Ac_j = e_j$. Let C be the matrix with columns $c_1, \dots, c_n$. Then:

$$AC = A\begin{bmatrix} c_1 & \dots & c_n \end{bmatrix} = \begin{bmatrix} Ac_1 & \dots & Ac_n \end{bmatrix} = \begin{bmatrix} e_1 & \dots & e_n \end{bmatrix} = I_n.$$

Now consider the homogeneous system $Cx = 0$. Multiplying by A gives $ACx = A0 = 0$. Since $AC = I_n$, we have $I_n x = 0$, so $x = 0$. Thus, the homogeneous system $Cx = 0$ has only the trivial solution. By the equivalence $(1) \iff (4)$ applied to C, the matrix C is invertible. Since C is invertible and $AC = I_n$, we have $A = C^{-1}$. Therefore, A is invertible.

■

Proposition 0.5. ***One-Sided Inverses.***
Let A be a square matrix.
1. If $BA = I$, then $B = A^{-1}$.
2. If $AC = I$, then $C = A^{-1}$.

In other words, for square matrices, a left or right inverse suffices to prove invertibility.

命題

Remark.

The proof of this relies on the rank arguments from the homogeneous systems section. If $AC = I$, then for any x, $ACx = x$. If $Cx = 0$, then $x = 0$. Thus the homogeneous system $Cx = 0$ has only the trivial solution, implying C is invertible. Then $A = C^{-1}$ is uniquely determined.

Proof

We prove (2); the proof of (1) is analogous. Assume $AC = I_n$.
Then for any x, $ACx = x$. If $Cx = 0$, then $x = ACx = A0 = 0$, so the homogeneous system $Cx = 0$ has only the trivial solution. By *theorem* 0.5, this implies C is invertible.
Multiply $AC = I_n$ on the left by C^{-1} to get $C^{-1}AC = C^{-1}I_n$, hence $A = C^{-1}$. Therefore $C = A^{-1}$.

■

Computing the Inverse

The equivalence (1) $\iff$ (2) provides a practical algorithm for computing A^{-1}. Since $E_k \dots E_1 A = I_n$, we have $A^{-1} = E_k \dots E_1$. This means the *same sequence* of row operations that reduces A to I_n will transform I_n into A^{-1}.
Algorithm: Form the augmented matrix $[A \mid I_n]$. Apply row operations to reduce the left side to I_n. The right side will become A^{-1}.

$$[A \mid I_n] \xrightarrow{\text{RREF}} [I_n \mid A^{-1}].$$

If the left side cannot be reduced to I_n (i.e., a row of zeros appears), then A is not invertible.

Example 0.9. Inverse Calculation. Compute the inverse of

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

範例

Solution

Form $[A \mid I]$:

$$\left[\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{array}\right]$$

$R_3 \to R_3 - R_1$:

$$\left[\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & -1 & 1 & -1 & 0 & 1 \end{array}\right]$$

$R_3 \to R_3 + R_2$:

$$\left[\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 2 & -1 & 1 & 1 \end{array}\right]$$

Scale R_3 by $1/2$, then clear the columns above the pivots to get:

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1/2 & -1/2 & 1/2 \\ 0 & 1 & 0 & 1/2 & 1/2 & -1/2 \\ 0 & 0 & 1 & -1/2 & 1/2 & 1/2 \end{array}\right].$$

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

■

0.6 Exercises

1. **System Solving.** Use Gaussian elimination (row reduction) to solve the following systems.

   (a)
   $$\begin{cases} x_1 + 2x_2 + 3x_3 = 1 \\ 2x_1 + 2x_2 + 5x_3 = 2 \\ 3x_1 + 5x_2 + x_3 = 3 \end{cases}$$

   (b)
   $$\begin{cases} x_2 + x_3 + x_4 = 1 \\ x_1 + x_3 + x_4 = 2 \\ x_1 + x_2 + x_4 = 3 \\ x_1 + x_2 + x_3 = 4 \end{cases}$$

2. **Intersection of Planes.** Find the set of common points (the intersection) of the three planes in $\mathbb{R}^3$ given by:

   $$9x - 3y + z = 20, \quad x + y + z = 0, \quad -x + 2y + z = -10.$$

3. **Quadratic Fitting.** Determine whether the data points $(1,27),(2,16),(3,29),(4,q)$ can lie on a quadratic curve $y = ax^2 + bx + c$. Find the necessary condition on q and determine the quadratic function.

4. **Homogeneous vs Non-Homogeneous.** Consider the system:

$$\begin{cases} x + 3y + 2z = 4 \\ 2x + 5y - 3z = -1 \\ 4x + 11y + z = 7 \end{cases}$$

   Solve this system over $\mathbb{R}$. Let S be its solution set. Now replace the constants $(4,-1,7)$ with $(0,0,0)$ to form the associated homogeneous system. Let S_0 be its solution set. Describe both S and S_0 explicitly.

5. **Underdetermined System.** Solve the system and express the solution set as a linear combination of vectors:

$$\begin{cases} x_1 + x_2 + x_3 + x_4 = 0 \\ 2x_1 + x_2 - x_3 + 3x_4 = 0 \end{cases}$$

6. **General Solution.** Find the general solution of the system:

$$\begin{cases} x_1 + x_2 + x_3 + x_4 + x_5 = 1 \\ x_1 + 2x_2 + 3x_3 + 4x_4 + 5x_5 = 6 \\ x_1 - x_3 - 2x_4 - 3x_5 = -4 \end{cases}$$

7. **Triviality Check.** Without fully solving, determine whether each homogeneous system has a non-zero solution. Explain your reasoning based on rank or variable counts.

   (a)
$$\begin{cases} x + y + z = 0 \\ 2x + y + 5z = 0 \end{cases}$$

   (b)
$$\begin{cases} x + y + z = 0 \\ 2x + y + 5z = 0 \\ 3x + 2y + 6z = 0 \end{cases}$$

8. **Parameter Analysis.** For which values of λ does the following system have a solution?

$$\begin{cases} \lambda x_1 + x_2 + x_3 = 1 \\ x_1 + \lambda x_2 + x_3 = \lambda \\ x_1 + x_2 + \lambda x_3 = \lambda^2 \end{cases}$$

Identify the values of λ for which the solution is unique, and those for which there are infinitely many solutions.

9. **Consistency Condition.** For which values of t is the system consistent? When consistent, describe the solution set.

$$\begin{cases} x + y + z = 1 \\ 2x + 3y + z = 2 \\ 3x + 4y + 2z = t \end{cases}$$

10. **RREF Analysis.** Let A and b be defined as:

$$A = \begin{bmatrix} 1 & 2 & -1 \\ 2 & 4 & -2 \\ 1 & 1 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 2 \\ t \end{bmatrix}.$$

Reduce the augmented matrix $[A \mid b]$ to Row Reduced Echelon Form. Use the result to determine the values of t for which the system $Ax = b$ is consistent, and describe the number of solutions.

11. **Row Equivalence and Rank.** Prove that if A and B are row-equivalent matrices, then $\text{rank}(A) = \text{rank}(B)$. (Hint: compare the number of non-zero rows in their RREFs.)

1
Abstract Vector Spaces

Definition 1.1. ***Vector Space.***
Let F be a field (typically $\mathbb{R}$ or $\mathbb{C}$). A set V is called a **vector space** over F if it is equipped with two operations:
Vector Addition: A map $V \times V \to V$, denoted $(x, y) \mapsto x + y$.
Scalar Multiplication: A map $F \times V \to V$, denoted $(\lambda, x) \mapsto \lambda x$.
These operations must satisfy the following axioms for all $x, y, z \in V$ and $\alpha, \beta \in F$:

定義

Axiom 1. ***Commutativity.***
$x + y = y + x$.

公理

Axiom 2. ***Associativity of Addition.***
$(x + y) + z = x + (y + z)$.

公理

Axiom 3. ***Zero Vector.***
There exists an element $0 \in V$ such that $x + 0 = x$ for all $x \in V$.

公理

Axiom 4. ***Additive Inverses.***
For every $x \in V$, there exists an element $-x \in V$ such that $x + (-x) = 0$.

公理

Axiom 5. ***Unital Property.***
$1 \cdot x = x$, where 1 is the multiplicative identity of F.

公理

Axiom 6. ***Associativity of Scalar Multiplication.***
$(\alpha\beta)x = \alpha(\beta x)$.

公理

Axiom 7. ***Distributivity over Vectors.***
$\alpha(x+y) = \alpha x + \alpha y$.

公理

Axiom 8. ***Distributivity over Scalars.***
$(\alpha+\beta)x = \alpha x + \beta x$.

公理

The first four axioms assert that $(V,+)$ is an Abelian group. The remaining axioms govern the interaction between the field F and the group V.

Remark.

Strictly speaking, one should distinguish the symbols for operations in V from those in F. For instance, one might write $\oplus$ for vector addition and $\odot$ for scalar multiplication, reserving $+$ and $\cdot$ for the field operations. The distributive law (*axiom 8*) would then read $(\alpha+\beta)\odot x = (\alpha\odot x)\oplus(\beta\odot x)$. However, in practice, the context almost always clarifies which operation is intended.

To illustrate that vector spaces encompass more than just column vectors, we examine some non-standard structures.

Example 1.1. The Space of Positive Reals. Let $V = \mathbb{R}_+$ be the set of strictly positive real numbers. We define the vector operations as follows:

Vector Addition ($\oplus$): For $x, y \in \mathbb{R}_+$, define $x \oplus y = xy$ (standard real multiplication).

Scalar Multiplication ($\odot$): For $\lambda \in \mathbb{R}$ and $x \in \mathbb{R}_+$, define $\lambda \odot x = x^\lambda$.

範例

Solution

We verify the axioms. The "zero vector" is the element $e \in V$ such that $x \oplus e = x$, which corresponds to $x \cdot e = x$; thus, the zero vector is the real number 1. The additive inverse of x is x^{-1}. Distributivity holds:

$$\lambda \odot (x \oplus y) = (xy)^\lambda = x^\lambda y^\lambda = (\lambda \odot x) \oplus (\lambda \odot y).$$

Thus, $(\mathbb{R}_+, \oplus, \odot)$ is a vector space over $\mathbb{R}$. In this context, standard expressions like $2x$ would formally evaluate to x^2.

■

Example 1.2. The Complex Conjugate Space. Let V be a vector space over the complex numbers $\mathbb{C}$. We can construct a new vector space $\overline{V}$ which shares the same underlying set and additive group as V, but possesses a modified scalar multiplication.

範例

Solution

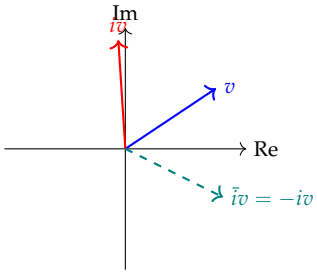
For any $\lambda \in \mathbb{C}$ and $x \in \overline{V}$, we define the operation $\odot$ by:

$$\lambda \odot x = \overline{\lambda} x,$$

where $\overline{\lambda}$ denotes the complex conjugate. Since the conjugation map $\lambda \mapsto \overline{\lambda}$ is a field automorphism (respecting addition and multiplication in $\mathbb{C}$), the vector space axioms are preserved. For example, associativity of scalar multiplication becomes:

$$(\alpha\beta) \odot x = \overline{(\alpha\beta)} x = (\overline{\alpha}\overline{\beta}) x = \overline{\alpha}(\overline{\beta} x) = \alpha \odot (\beta \odot x).$$

This structure is particularly important in the study of antilinear maps and dual spaces in quantum mechanics. ■



Scalar multiplication in V (red) vs $\overline{V}$ (teal) by $\lambda = i$.

Figure 1.1: Visualising the action of scalars in the conjugate space $\overline{V}$.

Several properties that appear intuitive from arithmetic can be rigorously derived from the axioms. These properties hold for any vector space, regardless of the nature of its elements.

Proposition 1.1. ***Basic Vector Arithmetic.***
Let V be a vector space over a field F. For all $x \in V$ and $\lambda \in F$:

1. The zero vector is unique.
2. $0 \cdot x = 0$. (The scalar 0 times any vector yields the zero vector).
3. $\lambda \cdot 0 = 0$. (Any scalar times the zero vector yields the zero vector).
4. If $\lambda x = 0$, then either $\lambda = 0$ or $x = 0$.
5. $(-1)x = -x$. (The scalar -1 produces the additive inverse).

命題

Proof

1. Suppose 0 and $0'$ both satisfy $x + 0 = x$ and $x + 0' = x$ for all $x \in V$. Then
$$0 = 0 + 0' = 0',$$
so the zero vector is unique.
2. Using *axiom 8*:
$$0 \cdot x = (0 + 0)x = 0 \cdot x + 0 \cdot x.$$
Adding the additive inverse $-(0 \cdot x)$ to both sides yields $0 = 0 \cdot x$.
3. Similarly, using *axiom 7*:
$$\lambda \cdot 0 = \lambda(0 + 0) = \lambda \cdot 0 + \lambda \cdot 0 \implies 0 = \lambda \cdot 0.$$

4. Suppose $\lambda x = 0$ and $\lambda \neq 0$. Since F is a field, λ has a multiplicative inverse λ^{-1}. Then:

$$x = 1 \cdot x = (\lambda^{-1}\lambda)x = \lambda^{-1}(\lambda x) = \lambda^{-1}0 = 0.$$

   Thus, if $\lambda \neq 0$, we must have $x = 0$.
5. We calculate the sum of x and $(-1)x$:

$$x + (-1)x = 1 \cdot x + (-1)x = (1 + (-1))x = 0 \cdot x = 0.$$

   If y is any element with $x + y = 0$, then

$$y = y + 0 = y + (x + (-1)x) = (y + x) + (-1)x = 0 + (-1)x = (-1)x,$$

   so the additive inverse of x is $(-1)x$.

■

Remark.

Property 5 allows us to define subtraction in vector spaces naturally as $x - y = x + (-1)y$. Furthermore, for any natural number n, the notation nx represents the n-fold sum $x + \cdots + x$, which is consistent with scalar multiplication by the integer n (interpreted as an element of the field F). If F has finite characteristic p, then $px = 0$ for all vectors x.

Remark.

The **characteristic** of a field F is the smallest positive integer p such that $p \cdot 1 = 0$ in F, i.e., 1 added to itself p times equals 0; if no such p exists, the characteristic is 0. For example, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ have characteristic 0, while $\mathbb{F}_p$ has characteristic p.

1.1 *Linear Combinations and Subspaces*

Definition 1.2. ***Linear Span.***
Let V be a vector space over a field F. Let $S \subseteq V$ be a subset of vectors (possibly infinite). A vector $v \in V$ is a **linear combination** of elements of S if it can be written as a finite sum:

$$v = \sum_{i=1}^{n} \lambda_i x_i,$$

where $\lambda_i \in F$ and $x_i \in S$. The set of all such linear combinations is called the **linear span** of S, denoted $\langle S \rangle$ or $\operatorname{span}(S)$.

定義

It is immediate from the axioms that $\langle S \rangle$ is closed under vector addition and scalar multiplication. If $v = \sum \lambda_i x_i$ and $w = \sum \mu_i x_i$ (padding

with zero coefficients if necessary to use the same vectors), then:

$$v + w = \sum(\lambda_i + \mu_i)x_i \quad \text{and} \quad \alpha v = \sum(\alpha\lambda_i)x_i.$$

This closure property is fundamental.

Definition 1.3. ***Subspace.***
Let V be a vector space. A subset $U \subseteq V$ is a **subspace** of V, denoted $U \leq V$, if U is non-empty and closed under the operations of V:
· If $u, v \in U$, then $u + v \in U$.
· If $u \in U$ and $\lambda \in F$, then $\lambda u \in U$.
A subspace U is called a **proper subspace** if $U \neq \{0\}$ and $U \neq V$.

定義

Definition 1.4. ***Row and Column Space.***
Let $A \in F^{m\times n}$. The **row space** of A is the subspace of F^n spanned by the rows of A. The **column space** of A is the subspace of F^m spanned by the columns of A.

定義

Remark.

To check if a subset is a subspace, it suffices to verify closure under addition and scalar multiplication. The existence of the zero vector follows from scalar multiplication by 0 ($0 \cdot u = 0 \in U$), and additive inverses follow from scalar multiplication by -1 ($(-1)u = -u \in U$).

The restriction of the operations of V to a subspace U makes U a vector space in its own right. The zero vector of V must lie in U (since $0 \cdot u = 0$). Furthermore, the intersection of any collection of subspaces is itself a subspace.

Proposition 1.2. ***Span as Smallest Subspace.***
For any subset $S \subseteq V$, the linear span $\langle S\rangle$ is the smallest subspace of V containing S. If S is already a subspace, then $\langle S\rangle = S$.

命題

Proof

First, $\langle S\rangle$ is a subspace: it is non-empty (if $S \neq \varnothing$, then $0 = 0 \cdot x \in \langle S\rangle$ for any $x \in S$; if $S = \varnothing$, then $\langle S\rangle = \{0\}$) and is closed under addition and scalar multiplication by construction of linear combinations. Also $S \subseteq \langle S\rangle$ since each $x \in S$ equals $1 \cdot x$. Let $U \leq V$ be any subspace with $S \subseteq U$. Since U is closed under linear combinations, every finite linear combination of elements of S lies in U, hence $\langle S\rangle \subseteq U$. Therefore $\langle S\rangle$ is the smallest subspace containing S. If S is already a subspace, then by the same argument $\langle S\rangle \subseteq S$, while always $S \subseteq \langle S\rangle$, so $\langle S\rangle = S$.

■

Example 1.3. Solution Space (Nullspace). As seen in chapter 0, the solution set of a homogeneous linear system $Ax = 0$ is a subset of $\mathbb{R}^n$ closed under addition and scalar multiplication. Thus, the solution set forms a subspace of $\mathbb{R}^n$, often called the **nullspace** of A.

範例

Example 1.4. Calculus Subspaces. Let $V = \mathbb{R}^{\mathbb{R}}$ be the space of all functions from $\mathbb{R}$ to $\mathbb{R}$.

- $C(\mathbb{R})$, the set of continuous functions, is a subspace of V.
- $C^1(\mathbb{R})$, the set of differentiable functions with continuous derivatives, is a subspace of $C(\mathbb{R})$.
- $C^\infty(\mathbb{R})$, the set of infinitely differentiable (smooth) functions, is a subspace of $C^1(\mathbb{R})$.

These nested subspaces play a crucial role in analysis.

範例

The universality of the vector space axioms allows us to treat diverse mathematical objects within the same framework.

Example 1.5. The Zero Space. Let $V = \{0\}$ and F be any field. Define $0 + 0 = 0$ and $\alpha \cdot 0 = 0$ for all $\alpha \in F$. This forms the trivial vector space.

範例

Example 1.6. Field Extensions. If K is a field extension of F (denoted K/F), then K is naturally a vector space over F.

- $\mathbb{C}$ is a vector space over $\mathbb{R}$. For example, $\{1, i\}$ spans $\mathbb{C}$ over $\mathbb{R}$.
- $\mathbb{R}$ is a vector space over $\mathbb{Q}$. This space is infinite-dimensional.

範例

Remark.

We say K is a **field extension** of F if F is a subfield of K (written $F \subseteq K$), meaning the operations on F agree with those in K. Equivalently, K is a field that contains a copy of F.

Example 1.7. Function Spaces. Let X be any non-empty set and F a field. The set of all functions $F^X = \{f : X \to F\}$ forms a vector space under pointwise operations:

$$(f + g)(x) = f(x) + g(x), \quad (\lambda f)(x) = \lambda f(x).$$

範例

Solution

If $X = \{1, \ldots, n\}$, this space is isomorphic to F^n. The function f corresponds to the tuple $(f(1), \ldots, f(n))$. The standard coordinate functions are the Kronecker delta functions δ_k, where

$$\delta_k(j) = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{if } j \neq k. \end{cases}$$

Thus any function can be written as

$$f = \sum_{k=1}^{n} f(k)\delta_k.$$

For infinite X, such a sum would be infinite, which is not defined in purely algebraic vector spaces. We typically study specific subspaces of F^X, such as:

- $\mathcal{C}(a, b)$: Continuous functions on an interval (a, b).
- $\mathcal{C}^1(a, b)$: Continuously differentiable functions on (a, b).
- Polynomials $F[x]$ or polynomials of degree at most n, denoted P_n.

■

Example 1.8. Integrable Functions. Let $\mathcal{R}[a, b]$ be the set of all Riemann-integrable functions on $[a, b]$. Standard results from calculus ensure that if f and g are integrable, then $f + g$ and αf are also integrable. Thus, $\mathcal{R}[a, b]$ is a vector space.

範例

Example 1.9. Matrix Spaces. The set of $m \times n$ matrices with entries in F, denoted $F^{m\times n}$, is a vector space under matrix addition and scalar multiplication. A particularly interesting subspace of the square matrices $M_n(\mathbb{Q})$ is the set of **semi-magic squares**: matrices where every row and column sums to the same constant $\sigma(A)$.

$$\mathrm{SMag}_n(\mathbb{Q}) = \left\{ A \in M_n(\mathbb{Q}) \;\middle|\; \forall i, j : \sum_k a_{ik} = \sum_k a_{kj} = \sigma(A) \right\}.$$

A **magic square** requires the main diagonal and anti-diagonal to also sum to $\sigma(A)$. These conditions define linear constraints, so the sets of semi-magic and magic squares form subspaces:

$$\mathrm{Mag}_n(\mathbb{Q}) \subseteq \mathrm{SMag}_n(\mathbb{Q}) \subseteq M_n(\mathbb{Q}).$$

範例

Geometric Interpretations

When $F = \mathbb{R}$, we refer to V as a **real vector space**. The most intuitive model is the set of geometric vectors (directed segments) in physical space. Addition follows the parallelogram law, and scalar multiplication scales length and reverses direction if negative.
When $F = \mathbb{C}$, we have a **complex vector space**. The one-dimensional space $\mathbb{C}^1$ can be visualised as the Argand plane $\mathbb{R}^2$. Scalar multiplication by a complex number $z = re^{i\theta}$ scales a vector by r and rotates it by θ.
However, the field need not be continuous.

Example 1.10. Finite Fields and Coding Theory. Let $F = \mathbb{F}_2 = \{0, 1\}$ be the binary field. The space $V = \mathbb{F}_2^n$ consists of binary strings of length n. Geometrically, these are the vertices of a unit hypercube in $\mathbb{R}^n$.
The subspace defined by the parity check equation:

$$\Pi_n = \{(\epsilon_1, \ldots, \epsilon_n) \in \mathbb{F}_2^n \mid \epsilon_1 + \cdots + \epsilon_n = 0\}$$

consists of all strings with an even number of 1s (since $1 + 1 = 0$ in $\mathbb{F}_2$). This is a simple error-detecting code. If a single bit is flipped during transmission, the parity sum becomes 1, alerting the receiver to an error. This geometric view of finite vector spaces underpins modern coding theory.

範例

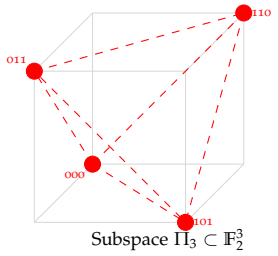


Figure 1.2: The even-parity subspace Π_3 forms a tetrahedron within the hypercube.

Non-Examples

It is instructive to examine sets that fail to be vector spaces.

Example 1.11. Affine Lines. Consider the set $S = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_2 = x_1 + 1\}$. This set is a line in the plane but does not pass through the origin (since $0 \neq 0 + 1$). Consequently, it does not contain the zero vector and cannot be a vector space. Furthermore, it is not closed under addition: if we take $x = (0, 1) \in S$ and $y = (1, 2) \in S$, their sum $x + y = (1, 3)$ satisfies $3 \neq 1 + 1$, so $x + y \notin S$.

範例

Example 1.12. Polynomials of Fixed Degree. Let V be the set of polynomials with degree *exactly* 2. This is not a vector space. Consider $p(t) = t^2 + t$ and $q(t) = -t^2 + t$. Both are in V, but their sum $(p + q)(t) = 2t$ has degree 1, so $p + q \notin V$. The set is not closed under addition. Moreover, the zero polynomial (degree $-\infty$ or undefined) is not in V.

範例

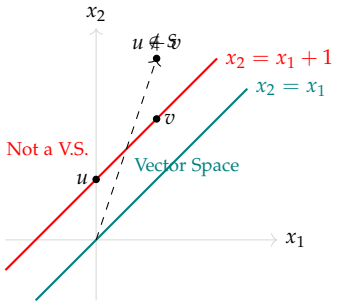


Figure 1.3: A line through the origin is a vector space (teal), while an affine line (red) is not.

Example 1.13. Hermitian Matrices. Consider the set of Hermitian matrices $H = \{A \in M_n(\mathbb{C}) \mid A^* = A\}$, where A^* denotes the conjugate transpose.
While H is closed under addition, it is **not** a subspace of the complex vector space $M_n(\mathbb{C})$. Taking $A = I$ (which is Hermitian) and scalar $\lambda = i$, we find $(\lambda A)^* = (iI)^* = -iI \neq \lambda A$. However, if we regard $M_n(\mathbb{C})$ as a vector space over $\mathbb{R}$, then H forms a subspace, as λA remains Hermitian for all real λ.

範例

Speaking of transpose:

Definition 1.5. ***Transpose.***
Let A be an $m \times n$ matrix. The **transpose** of A, denoted A^T, is the $n \times m$ matrix defined by:

$$(A^T)_{ji} = A_{ij} \quad \text{for all } 1 \leq i \leq m, 1 \leq j \leq n.$$

Intuitively, the transpose converts rows into columns and columns into rows.

定義

Proposition 1.3. ***Properties of Transpose.***
Let A, B be matrices of appropriate sizes and α be a scalar.
1. $(A^T)^T = A$.
2. $(A + B)^T = A^T + B^T$ and $(\alpha A)^T = \alpha A^T$.
3. $(AB)^T = B^T A^T$.
4. $\text{row}_j(A^T) = (\text{col}_j(A))^T$ and $\text{col}_i(A^T) = (\text{row}_i(A))^T$.

命題

Proof

1. The (i, j) entry of $(A^T)^T$ is the (j, i) entry of A^T, which is A_{ij}.
2. Follows from the linearity of the entry-wise operations.
3. Let $A \in F^{m \times k}$ and $B \in F^{k \times n}$. Then

$$((AB)^T)_{ij} = (AB)_{ji} = \sum_{r=1}^{k} A_{jr} B_{ri} = \sum_{r=1}^{k} (B^T)_{ir} (A^T)_{rj} = (B^T A^T)_{ij}.$$

4. The k-th entry of $\text{row}_j(A^T)$ is $(A^T)_{jk} = A_{kj}$, which is the k-th entry of $(\text{col}_j(A))^T$. Similarly, the k-th entry of $\text{col}_i(A^T)$ is $(A^T)_{ki} = A_{ik}$, which is the k-th entry of $(\text{row}_i(A))^T$.

■

1.2 *Linear Dependence and Dimension*

The structure of a vector space is determined by the relationships between its elements. The most fundamental such relationship is whether one vector can be built from others.

Definition 1.6. ***Linear Dependence.***
A finite set of vectors $\{v_1, \ldots, v_n\}$ in a vector space V is called **linearly dependent** if there exist scalars $\alpha_1, \ldots, \alpha_n \in F$, not all zero, such that:

$$\alpha_1 v_1 + \cdots + \alpha_n v_n = 0.$$

Such a relation is called a **non-trivial linear relation**. If the only solution to $\sum \alpha_i v_i = 0$ is the trivial solution $\alpha_1 = \cdots = \alpha_n = 0$, the set is called **linearly independent**.

定義

Remark.

If a set contains the zero vector, it is automatically linearly dependent. For instance, if $v_1 = 0$, we can choose $\alpha_1 = 1$ and all other $\alpha_i = 0$ to satisfy the equation.

Proposition 1.4. ***Basic Properties of Independence.***
Let S be a set of vectors in V.

1. The empty set $\varnothing$ is linearly independent.
2. If $0 \in S$, then S is linearly dependent.
3. If S is linearly independent, then every subset of S is linearly independent.
4. If S is linearly dependent, then every superset of S is linearly dependent.

命題

Proof

1. The condition for linear dependence requires the existence of non-zero scalars α_i such that $\sum \alpha_i v_i = 0$. Since there are no vectors in $\varnothing$, no such sum exists. Thus, the condition for dependence can never be satisfied.
2. Let $S = \{0, v_2, \ldots, v_n\}$. Consider the linear combination $1 \cdot 0 + 0 \cdot v_2 + \cdots + 0 \cdot v_n = 0$. Since the coefficient of the zero vector is non-zero ($1 \neq 0$), this is a non-trivial relation.
3. Let $A \subseteq S$. If A were dependent, there would be a non-trivial relation among elements of A. This same relation serves as a non-trivial relation for S (by setting coefficients of vectors in $S \setminus A$ to zero), contradicting the independence of S.
4. Let $S \subseteq B$. Since S is dependent, there exists a non-trivial relation $\sum \alpha_i s_i = 0$ with $s_i \in S$. This same sum is a linear com-

bination of vectors in B (with other coefficients zero), so B is dependent.

■

Lemma 1.1. ***Dependence of Two Vectors.***
Two vectors $u, v \in V$ are linearly dependent if and only if one is a scalar multiple of the other.

引理

($\Longrightarrow$)

Suppose u, v are dependent. Then $\alpha u + \beta v = 0$ with not both α, β zero. If $\alpha \neq 0$, then $u = (-\beta/\alpha)v$. If $\alpha = 0$, then $\beta \neq 0$, so $\beta v = 0 \implies v = 0$. In this case $v = 0 \cdot u$.

証明終

($\Longleftarrow$)

Suppose $u = \lambda v$. Then $1 \cdot u + (-\lambda)v = 0$. Since the coefficient of u is $1 \neq 0$, this is a non-trivial relation.

証明終

Example 1.14. Polynomial Independence via Differentiation. Consider the set $\{1, t, t^2, t^3\}$ in the space of real polynomials. Suppose

$$\alpha_0 + \alpha_1 t + \alpha_2 t^2 + \alpha_3 t^3 = 0$$

for all $t \in \mathbb{R}$. This is an identity of functions. Differentiating with respect to t repeatedly yields:

$$\begin{aligned} \alpha_1 + 2\alpha_2 t + 3\alpha_3 t^2 &= 0 \\ 2\alpha_2 + 6\alpha_3 t &= 0 \\ 6\alpha_3 &= 0 \end{aligned}$$

From the last equation, $\alpha_3 = 0$. Substituting back gives $\alpha_2 = 0$, then $\alpha_1 = 0$, and finally $\alpha_0 = 0$. Thus, the monomials are linearly independent. This method avoids the Fundamental Theorem of Algebra (roots).

範例

Proposition 1.5. ***Characterisation of Dependence.***
A set of non-zero vectors $\{v_1, \ldots, v_n\}$ is linearly dependent if and only if at least one vector v_k ($k \geq 2$) can be written as a linear combination of the **preceding** vectors $v_1, \ldots, v_{k-1}$.

命題

($\Longrightarrow$)

Suppose the set is dependent. Let k be the largest index such that

$\alpha_k \neq 0$ in a non-trivial relation $\sum_{i=1}^{n} \alpha_i v_i = 0$. Since $v_1 \neq 0$, we must have $k \geq 2$. Then:

$$v_k = -\alpha_k^{-1} \sum_{i=1}^{k-1} \alpha_i v_i.$$

証明終

($\Longleftarrow$)

If v_k is a linear combination of preceding vectors, then $v_k - \sum \beta_j v_j = 0$ is a non-trivial relation (coefficient of v_k is 1), so the set is dependent.

証明終

The size of linearly independent sets is bounded by the "capacity" of the space. The following theorem, fundamental to the theory, formalises this.

Lemma 1.2. ***Underdetermined Homogeneous Systems.***
Let F be a field and consider a homogeneous linear system with t equations in s unknowns over F. If $s > t$, then the system has a non-trivial solution (recall *theorem* 0.4).

引理

Proof

Row-reduce the coefficient matrix to echelon form. There are at most t pivots, so with $s > t$ there is at least one free variable. Assigning a non-zero value to a free variable produces a non-trivial solution.

■

Theorem 1.1. ***Steinitz Exchange Lemma.***
Let $\{e_1, \ldots, e_s\}$ be a linearly independent set in V, and let $\{f_1, \ldots, f_t\}$ be a set of vectors such that every e_i lies in $\operatorname{span}(\{f_1, \ldots, f_t\})$. Then $s \leq t$.

定理

Proof

Suppose for contradiction that $s > t$. Since each e_j is in the span of the f's, we can write:

$$e_j = \sum_{i=1}^{t} \alpha_{ij} f_i \quad \text{for } j = 1, \ldots, s.$$

Consider a linear combination of the e's equal to zero:

$$\sum_{j=1}^{s} x_j e_j = 0.$$

Substituting the expressions for e_j:

$$\sum_{j=1}^{s} x_j \left(\sum_{i=1}^{t} \alpha_{ij} f_i \right) = \sum_{i=1}^{t} \left(\sum_{j=1}^{s} \alpha_{ij} x_j \right) f_i = 0.$$

This equation is satisfied if the coefficients of each f_i vanish. Thus, we seek a solution to the homogeneous linear system:

$$\sum_{j=1}^{s} \alpha_{ij} x_j = 0 \quad \text{for } i = 1, \ldots, t.$$

This is a system of t equations in s unknowns. Since $s > t$, there are more unknowns than equations, so a non-trivial solution $(x_1, \ldots, x_s)$ exists by the previous lemma. This non-trivial solution implies $\sum x_j e_j = 0$, contradicting the linear independence of the e_j's. Thus, we must have $s \leq t$.

■

Corollary 1.1. *Invariance of Independent Set Size.* If two finite sets in V are equivalent (each spans the other) and linearly independent, they must have the same cardinality.

推論

Proof

Let $E = \{e_1, \ldots, e_s\}$ and $F = \{f_1, \ldots, f_t\}$ be linearly independent and assume each spans the other. Since $E \subseteq \text{span}(F)$ and E is independent, *Steinitz Exchange Lemma* gives $s \leq t$. Similarly, $F \subseteq \text{span}(E)$ and F independent implies $t \leq s$. Hence $s = t$.

■

We classify vector spaces by the size of their maximal independent sets.

Definition 1.7. ***Dimension.***
A vector space V is called **n-dimensional**, denoted $\dim V = n$, if there exists a set of n linearly independent vectors, and every set with more than n vectors is linearly dependent. If V contains arbitrarily large linearly independent sets, it is called **infinite-dimensional**. The zero space $\{0\}$ has dimension 0.

定義

Definition 1.8. ***Basis.***
Let V be an n-dimensional space. A set of vectors $\{e_1, \ldots, e_n\}$ is called a **basis** if it is linearly independent.

定義

Example 1.15. Infinite Dimensionality of $C[0,1]$. The space $C[0,1]$ of continuous real-valued functions on $[0,1]$ is infinite-dimensional. Consider the set of monomials $S = \{1, t, t^2, \dots\}$. As shown previously, any finite subset of S is linearly independent (via differentiation). If $C[0,1]$ had a finite dimension n, then any linearly independent set could have size at most n. However, S contains independent subsets of arbitrary size $k > n$, which is a contradiction. Thus, no finite basis exists.

範例

By definition, a basis is a maximal linearly independent set. It is also a minimal spanning set. The utility of a basis lies in the unique representation of vectors.

Theorem 1.2. ***Basis Representation and Extension.***
Let V be a vector space of dimension n.

1. If $\{e_1, \dots, e_n\}$ is a basis, then every vector $v \in V$ can be uniquely expressed as $v = \sum_{i=1}^{n} \lambda_i e_i$.
2. Any linearly independent set $\{f_1, \dots, f_s\}$ with $s < n$ can be extended to a basis of V.

定理

Proof

1. ***Existence:*** Since $\dim V = n$, the set $\{v, e_1, \dots, e_n\}$ contains $n+1$ vectors and must be linearly dependent. Thus $\alpha v + \sum \alpha_i e_i = 0$ for scalars not all zero. If $\alpha = 0$, the e_i would be dependent, which is false. Thus $\alpha \neq 0$, and we can solve for v:

$$v = \sum_{i=1}^{n} (-\alpha^{-1}\alpha_i) e_i.$$

   Uniqueness: Suppose $v = \sum \lambda_i e_i = \sum \mu_i e_i$. Subtracting gives $\sum(\lambda_i - \mu_i)e_i = 0$. By independence, $\lambda_i - \mu_i = 0$ for all i.
2. Consider the set $S = \{f_1, \dots, f_s\}$. If S spans V, then $s = n$ by *theorem* 1.1. If not, there exists e_{k_1} in a basis of V that is not in $\text{span}(S)$. Adjoin it to form $\{f_1, \dots, f_s, e_{k_1}\}$. Repeat this process until the set contains n vectors. The resulting set is independent and has size n, so it is a basis.

■

Corollary 1.2. *Subspace Dimensions.* If U is a subspace of a finite-dimensional space V, then $\dim U \leq \dim V$. If $\dim U = \dim V$, then $U = V$.

推論

Proof

Let $\dim V = n$ and let $\{u_1, \dots, u_k\}$ be a basis of U. Since $U \subseteq V$,

this set is linearly independent in V, so it can be extended to a basis of V of size n (by the extension part of *theorem* 1.2). Hence $k \leq n$, so $\dim U \leq \dim V$. If $\dim U = \dim V$, then a basis of U already has size n and thus is a basis of V, which implies $U = V$.

■

Corollary 1.3. *Basis Check by Size.* Let V be a vector space with $\dim V = n$. If S is a linearly independent subset of V with $|S| = n$, then S is a basis for V. (That is, spanning is automatic).

推論

Proof

If S were not a spanning set, we could extend it to a basis by adding at least one vector, resulting in a basis of size $> n$, which contradicts the unique dimension of V.

■

Example 1.16. Examples of Dimension.

- $\dim \mathbb{R}^n = n$. The standard basis is $\{e_1, \ldots, e_n\}$ where e_i has a 1 in the i-th position and 0 elsewhere.
- The space of $m \times n$ matrices has dimension mn.
- The space of polynomials P_n (degree $\leq n$) has dimension $n+1$. A basis is $\{1, x, \ldots, x^n\}$.

範例

Example 1.17. Subspaces of $\mathbb{R}^3$. We can classify all subspaces W of $\mathbb{R}^3$ by dimension:

- $\dim W = 0$: The zero subspace $\{0\}$ (the origin).
- $\dim W = 1$: A line passing through the origin.
- $\dim W = 2$: A plane passing through the origin.
- $\dim W = 3$: The entire space $\mathbb{R}^3$.

範例

Example 1.18. Homogeneous Polynomials. The space of homogeneous polynomials of degree k in m variables has dimension $\binom{k+m-1}{k}$.

範例

Proof

Write a homogeneous polynomial of degree k in m variables as a linear combination of monomials

$$x_1^{\alpha_1} \cdots x_m^{\alpha_m} \quad \text{with} \quad \alpha_1 + \cdots + \alpha_m = k,\ \alpha_i \in \mathbb{N}.$$

These monomials are linearly independent and span the space, so they form a basis. The number of m-tuples of nonnegative integers

with sum k equals $\binom{k+m-1}{k}$ (stars-and-bars), hence the dimension is $\binom{k+m-1}{k}$.

■

Remark.

For function spaces, the index set of a basis need not be finite. For a finite set X, the space F^X has dimension $|X|$ with basis $\{\delta_x \mid x \in X\}$, where $\delta_x(y) = 1$ if $y = x$ and 0 otherwise.

1.3 *Coordinates and Isomorphisms*

A basis provides a bridge between abstract vector spaces and the concrete coordinate spaces $\mathbb{R}^n$.

Definition 1.9. ***Coordinates.***
Let $(e_1, \ldots, e_n)$ be a basis of a vector space V over $\mathbb{R}$. For any vector $v \in V$, the unique scalars $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ such that

$$v = \lambda_1 e_1 + \cdots + \lambda_n e_n$$

are called the **coordinates** of the vector v relative to the given basis.

定義

This definition implies linearity: if $x = \sum \alpha_i e_i$ and $y = \sum \beta_i e_i$, then $x + y = \sum(\alpha_i + \beta_i)e_i$. Thus, adding vectors corresponds to adding their coordinates. Similarly, multiplying a vector by a scalar multiplies its coordinates by that scalar. The zero vector corresponds to the coordinates $(0, \ldots, 0)$.

Example 1.19. Polynomials. In the space P_n of polynomials of degree $\leq n$, the set $(1, t, \ldots, t^n)$ is a basis. The coordinates of $f(t) = a_0 + a_1 t + \cdots + a_n t^n$ are its coefficients $a_0, \ldots, a_n$.
However, using Taylor's formula, we can write:

$$f(t) = f(\alpha) + f'(\alpha)(t-\alpha) + \cdots + \frac{f^{(n-1)}(\alpha)}{(n-1)!}(t-\alpha)^{n-1}.$$

Relative to the basis $(1, t-\alpha, \ldots, (t-\alpha)^{n-1})$, the coordinates of the same polynomial are $f(\alpha), f'(\alpha), \ldots, \frac{f^{(n-1)}(\alpha)}{(n-1)!}$.

範例

A single vector space may have infinitely many bases. We investigate how coordinates change when we switch from one basis to another. Let $\mathcal{E} = (e_1, \ldots, e_n)$ and $\mathcal{E}' = (e'_1, \ldots, e'_n)$ be two bases of V. Each new

basis vector e'_j can be written uniquely in terms of the old basis $\mathcal{E}$:

$$e'_j = \sum_{i=1}^{n} a_{ij} e_i.$$

The coefficients $a_{ij} \in \mathbb{R}$ form the **transition matrix** A from the basis (e'_i) to the basis (e_i):

$$A = (a_{ij}) = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}.$$

It is crucial to note that the coordinates of the new basis vector e'_j relative to the old basis lie in the j-th column of A.
Let a vector v have coordinates $\lambda_1, \ldots, \lambda_n$ in the basis (e_i) and $\lambda'_1, \ldots, \lambda'_n$ in the basis (e'_i). We have:

$$v = \sum_{i=1}^{n} \lambda_i e_i = \sum_{j=1}^{n} \lambda'_j e'_j.$$

Substituting the expressions for e'_j:

$$v = \sum_{j=1}^{n} \lambda'_j \left(\sum_{i=1}^{n} a_{ij} e_i \right) = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} a_{ij} \lambda'_j \right) e_i.$$

Comparing the coefficients of e_i, we obtain:

$$\lambda_i = a_{i1}\lambda'_1 + a_{i2}\lambda'_2 + \cdots + a_{in}\lambda'_n.$$

In matrix notation, let X and X' be the columns of coordinates:

$$X = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix}, \quad X' = \begin{bmatrix} \lambda'_1 \\ \vdots \\ \lambda'_n \end{bmatrix}.$$

Then the relationship is $X = AX'$. Since both sets are bases, we can also express (e_i) in terms of (e'_i), which implies the matrix A is invertible. Thus, we have the inverse relationship:

Theorem 1.3. ***Coordinate Transformation.***
In the transition from basis $(e_1, \ldots, e_n)$ to basis $(e'_1, \ldots, e'_n)$ determined by matrix A, the new coordinates are expressed in terms of the original coordinates by:

$$X' = A^{-1}X.$$

定理

Proof

From the preceding computation, the coordinate columns satisfy $X = AX'$. Since the change-of-basis matrix A is invertible, multiply both sides by A^{-1} to obtain $X' = A^{-1}X$.

■

Remark.

The matrix A is invertible because it sends the coordinate basis in $\mathbb{R}^n$ to the coordinate columns of the new basis vectors, so its columns are linearly independent.

Isomorphisms

Using coordinates, we can identify any n-dimensional space with $\mathbb{R}^n$.

Definition 1.10. ***Isomorphism.***
Two vector spaces V and W over $\mathbb{R}$ are **isomorphic** if there exists a bijection $f : V \to W$ such that for all $\alpha, \beta \in \mathbb{R}$ and $u, v \in V$:

$$f(\alpha u + \beta v) = \alpha f(u) + \beta f(v).$$

Such a map f is called an isomorphism.

定義

Note

A map $f : V \to W$ is:

- **Injective** (or one-to-one) if $f(u) = f(v) \implies u = v$ for all $u, v \in V$.
- **Surjective** (or onto) if for every $w \in W$, there exists $v \in V$ such that $f(v) = w$.
- **Bijective** if it is both injective and surjective.

If f is an isomorphism, then $f^{-1} : W \to V$ is also an isomorphism. Dimension is invariant under isomorphism: if (e_i) is a basis for V, then $(f(e_i))$ is a basis for W.

Theorem 1.4. ***Classification of Finite Dimensional Spaces.***
All vector spaces of the same dimension n over $\mathbb{R}$ are isomorphic. Specifically, they are all isomorphic to the coordinate space $\mathbb{R}^n$.

定理

Proof

Let $(e_1, \ldots, e_n)$ be a basis for V. Define the map $f : V \to \mathbb{R}^n$ by mapping a vector $x = \sum \alpha_i e_i$ to its coordinate tuple $(\alpha_1, \ldots, \alpha_n)$. Since coordinates are unique, f is a bijection. Linearity follows from

the properties of coordinates:

$$f(\alpha x + \beta y) = \alpha(\alpha_1, \ldots, \alpha_n) + \beta(\beta_1, \ldots, \beta_n) = \alpha f(x) + \beta f(y).$$

Thus V is isomorphic to $\mathbb{R}^n$.

■

Note

While any two spaces of dimension n are isomorphic, the isomorphism depends on the choice of basis. An isomorphism defined without arbitrary choices is called **canonical** or **natural**.

1.4 *Operations on Subspaces*

We now consider how subspaces interact. Let $U, W \leq V$ be subspaces.

Intersection: The set $U \cap W$ is always a subspace. It is the largest subspace contained in both U and W.

Union: The set $U \cup W$ is generally *not* a subspace (e.g., the union of the x-axis and y-axis in $\mathbb{R}^2$ is not closed under addition).

Sum: The sum of subspaces is defined as:

$$U + W = \{u + w \mid u \in U, w \in W\}.$$

This is the smallest subspace containing both U and W. In fact, $U + W = \operatorname{span}(U \cup W)$.

The dimensions of these spaces are related by a fundamental formula analogous to the inclusion-exclusion principle for sets.

Theorem 1.5. ***Grassmann's Formula.***
Let U and W be finite-dimensional subspaces of V. Then:

$$\dim(U + W) = \dim U + \dim W - \dim(U \cap W).$$

定理

Proof

Let $\{v_1, \ldots, v_m\}$ be a basis for $U \cap W$, where $m = \dim(U \cap W)$. By *theorem* 1.2, we can extend this to a basis of U:

$$\mathcal{B}_U = \{v_1, \ldots, v_m, u_1, \ldots, u_{k-m}\}, \quad \text{where } k = \dim U.$$

Similarly, we extend it to a basis of W:

$$\mathcal{B}_W = \{v_1, \ldots, v_m, w_1, \ldots, w_{l-m}\}, \quad \text{where } l = \dim W.$$

We claim that the set $\mathcal{S} = \{v_1, \ldots, v_m, u_1, \ldots, u_{k-m}, w_1, \ldots, w_{l-m}\}$ is a basis for $U + W$. Clearly $\mathcal{S}$ spans $U + W$. To check independence, suppose:

$$\sum \gamma_i v_i + \sum \alpha_j u_j + \sum \beta_r w_r = 0.$$

Rewrite this as:

$$\sum \alpha_j u_j + \sum \gamma_i v_i = -\sum \beta_r w_r.$$

The left side is in U, and the right side is in W. Thus, the vector represented by this sum lies in $U \cap W$. Hence, $-\sum \beta_r w_r$ can be written as a linear combination of the basis vectors v_i of $U \cap W$:

$$-\sum \beta_r w_r = \sum \delta_i v_i \implies \sum \beta_r w_r + \sum \delta_i v_i = 0.$$

Since $\{v_i\} \cup \{w_r\}$ is a basis for W, all coefficients β_r (and δ_i) must be zero. The original equation reduces to $\sum \gamma_i v_i + \sum \alpha_j u_j = 0$. Since $\{v_i\} \cup \{u_j\}$ is a basis for U, all γ_i and α_j are zero. Thus $\mathcal{S}$ is independent. The dimension of $U + W$ is the size of $\mathcal{S}$:

$$\dim(U + W) = m + (k - m) + (l - m) = k + l - m.$$

■

Corollary 1.4. *Nontrivial Intersection.* If $\dim U + \dim W > \dim V$, then U and W must have a non-trivial intersection (i.e., $\dim(U \cap W) > 0$). For example, two planes in $\mathbb{R}^3$ through the origin must intersect in at least a line.

推論

Proof

By *theorem 1.5*,

$$\dim(U \cap W) = \dim U + \dim W - \dim(U + W).$$

Since $U + W \leq V$, we have $\dim(U + W) \leq \dim V$. Hence

$$\dim(U \cap W) \geq \dim U + \dim W - \dim V.$$

If $\dim U + \dim W > \dim V$, then the right-hand side is positive, so $\dim(U \cap W) > 0$.

■

Definition 1.11. ***Codimension.***
The **codimension** of a subspace $U \leq V$ is defined as $\operatorname{codim} U = \dim V - \dim U$. A subspace of codimension 1 is called a **hyperplane**.

定義

1.5 *Direct Sums and Quotient Spaces*

We have seen that the sum of subspaces $U + W$ is not always "efficient," in the sense that a vector in the sum may have multiple representations $u + w$. When representations are unique, the sum is called direct.

Definition 1.12. ***Direct Sum.***
Let $U_1, \ldots, U_m$ be subspaces of V. The sum $U = U_1 + \cdots + U_m$ is called a **direct sum**, denoted $U = U_1 \oplus \cdots \oplus U_m$, if every vector $u \in U$ can be uniquely written as:

$$u = u_1 + \cdots + u_m, \quad \text{where } u_i \in U_i.$$

定義

Proposition 1.6. ***Criteria for Direct Sums.***
The following conditions are equivalent for a sum $U = \sum_{i=1}^{m} U_i$:

1. The sum is direct.
2. The zero vector has a unique representation: if $\sum_{i=1}^{m} u_i = 0$ with $u_i \in U_i$, then $u_i = 0$ for all i.
3. For each k, the intersection of U_k with the sum of the other subspaces is zero:
$$U_k \cap \left(\sum_{j \neq k} U_j \right) = \{0\}.$$
4. If the spaces are finite-dimensional, the dimensions add up:
$$\dim \left(\sum_{i=1}^{m} U_i \right) = \sum_{i=1}^{m} \dim U_i.$$

命題

Proof

(1) $\Longrightarrow$ (2): If the sum is direct and $\sum_{i=1}^{m} u_i = 0$ with $u_i \in U_i$, then also $0 = \sum_{i=1}^{m} 0$. Uniqueness of representation in a direct sum forces $u_i = 0$ for all i.

(2) $\Longrightarrow$ (3): Suppose $x \in U_k \cap (\sum_{j \neq k} U_j)$. Then $x = u_k$ and $x = \sum_{j \neq k} u_j$ for some $u_j \in U_j$. Thus $0 = (-u_k) + \sum_{j \neq k} u_j$. By (2), each term is zero, so $x = 0$.

(3) $\Longrightarrow$ (1): Given $\sum_{i=1}^{m} u_i = \sum_{i=1}^{m} v_i$ with $u_i, v_i \in U_i$, subtract to get $\sum_{i=1}^{m} (u_i - v_i) = 0$. Fix k and rewrite

$$u_k - v_k = -\sum_{j \neq k} (u_j - v_j) \in U_k \cap \sum_{j \neq k} U_j.$$

By (3), $u_k - v_k = 0$. Since this holds for each k, all coordinates agree, so the representation is unique and the sum is direct.

(1) $\implies$ (4) *(finite-dimensional):* For $m = 2$, $U_1 \cap U_2 = \{0\}$, so *theorem 1.5* gives

$$\dim(U_1 + U_2) = \dim U_1 + \dim U_2.$$

Assume the formula holds for $m - 1$ subspaces and let $W = \sum_{i=1}^{m-1} U_i$. Directness implies $W \cap U_m = \{0\}$, so applying *theorem 1.5* to W and U_m gives

$$\dim\Big(\sum_{i=1}^{m} U_i\Big) = \dim W + \dim U_m = \sum_{i=1}^{m-1} \dim U_i + \dim U_m.$$

(4) $\implies$ (1) *(finite-dimensional):* Again let $W = \sum_{i=1}^{m-1} U_i$. By *theorem 1.5*,

$$\dim(W + U_m) = \dim W + \dim U_m - \dim(W \cap U_m).$$

Using (4) and the inductive hypothesis for W gives

$$\sum_{i=1}^{m} \dim U_i = \Big(\sum_{i=1}^{m-1} \dim U_i\Big) + \dim U_m - \dim(W \cap U_m),$$

so $\dim(W \cap U_m) = 0$. Thus $W \cap U_m = \{0\}$, and by the $m = 2$ case, the sum is direct. Inducting on m completes the proof. ■

Remark.

For two subspaces, the condition simplifies: $V = U \oplus W$ if and only if $U + W = V$ and $U \cap W = \{0\}$. In this case, W is called a **complement** of U in V. By the *theorem 1.2*, every subspace has a complement (extend a basis of U to a basis of V), but complements are not unique.

Quotient Space

The non-uniqueness of complements suggests we should look for an intrinsic object that captures the "difference" between V and a subspace U.

Definition 1.13. ***Quotient Space.***
Let U be a subspace of V. We define an equivalence relation on V:

$$v \sim w \iff v - w \in U.$$

The equivalence class of v is the set $v + U = \{v + u \mid u \in U\}$, called a **coset**. The set of all such cosets is denoted V/U. We define vector op-

erations on cosets:

$$(v+U)+(w+U)=(v+w)+U, \quad \lambda(v+U)=(\lambda v)+U.$$

These operations are well-defined (independent of the representative). The space V/U is called the **quotient space** of V by U.

定義

Theorem 1.6. ***Isomorphism of Complements.***
Let U be a subspace of V. If W is any complement of U (so $V = U \oplus W$), then W is isomorphic to V/U.

定理

Proof

Define the map $\pi : W \to V/U$ by $\pi(w) = w + U$. This map is linear.

Injectivity: If $\pi(w) = 0 + U$, then $w \in U$. Since $w \in W$ and $U \cap W = \{0\}$, we have $w = 0$.

Surjectivity: Let $v + U \in V/U$. Since $V = U \oplus W$, we can write $v = u + w$ with $u \in U, w \in W$. Then

$$v + U = (u + w) + U = w + (u + U) = w + U = \pi(w).$$

Thus π is an isomorphism.

■

Corollary 1.5. *Dimension of Quotients.* For finite-dimensional spaces:

$$\dim(V/U) = \dim V - \dim U = \operatorname{codim} U.$$

推論

Proof

Choose a complement W of U in V, so $V = U \oplus W$. By the previous theorem, V/U is isomorphic to W, hence $\dim(V/U) = \dim W$. By *theorem* 1.5,

$$\dim V = \dim U + \dim W,$$

so $\dim(V/U) = \dim V - \dim U$.

■

This formalises the idea that the quotient space "subtracts" the subspace U from V.

Example 1.20. Visualising Quotients. Let $V = \mathbb{R}^2$ and U be the x-axis (line $y = 0$). The cosets of U are lines parallel to the x-axis, of the form $\{(x, c) \mid x \in \mathbb{R}\}$.
Each coset is uniquely determined by its y-intercept c. Thus, the set

of cosets V/U can be identified with the y-axis (the complement $x = 0$), which is isomorphic to $\mathbb{R}$.

範例

1.6 Exercises

1. **Vector Space Verification.** Determine whether the following sets form vector spaces over $\mathbb{R}$:
   (a) The set of all polynomials $p(x) \in \mathbb{R}[x]$ with $p(0) = 1$.
   (b) The set of all functions $f : \mathbb{R} \to \mathbb{R}$ such that $f(x) \to 0$ as $x \to \infty$.
   (c) The set of all $n \times n$ matrices with trace zero (the trace of matrix $A = (a_{ij})$ is defined by $\operatorname{tr} A = a_{11} + a_{22} + \cdots + a_{nn}$).
   (d) The set of all $n \times n$ matrices with positive trace.
2. **Finite Fields and Subspace Counts.**
   (a) Let $\mathbb{F}_q$ be a finite field with q elements (where q is a prime power). Show that $|\mathbb{F}_q^n| = q^n$.
   (b) Let W be a k-dimensional vector space over $\mathbb{F}_q$. Count the number of ordered bases of W.
   (c) Conclude that the number of k-dimensional subspaces of an n-dimensional $\mathbb{F}_q$-vector space is the Gaussian binomial coefficient $\binom{n}{k}_q$.
3. **Dimension of Matrix Spaces.** Determine the dimension of the following subspaces of $M_n(\mathbb{R})$:
   (a) Symmetric matrices ($A = A^\top$).
   (b) Skew-symmetric matrices ($A = -A^\top$).
   (c) Traceless matrices ($\operatorname{tr}(A) = 0$).
4. **Polynomial Subspace.** Let P_n be the space of polynomials of degree at most n. Let $W = \{f \in P_n \mid f(1) = 0\}$. Find $\dim W$ and construct a basis for it.
5. **The Space $\mathbb{Q}[\theta]$.**
   (a) A polynomial $f \in \mathbb{Q}[t]$ is **irreducible** if it cannot be factored into two non-constant polynomials in $\mathbb{Q}[t]$. The **minimal polynomial** of θ is the monic polynomial of lowest degree in $\mathbb{Q}[t]$ having θ as a root. Explain why the minimal polynomial is unique (and irreducible).
   (b) Let d be the degree of the minimal polynomial of θ. Show

For (a): To prove uniqueness, suppose there are two; consider their difference and use a degree argument.

that $1, \theta, \ldots, \theta^{d-1}$ are linearly independent over $\mathbb{Q}$.

(c) Show that every element of $\mathbb{Q}[\theta]$ can be written as a $\mathbb{Q}$-linear combination of $1, \theta, \ldots, \theta^{d-1}$, and conclude $\dim_{\mathbb{Q}} \mathbb{Q}[\theta] = d$.

6. **Change of Basis.** Find the transition matrix from the standard basis $(1, t, \ldots, t^n)$ of P_n to the Taylor basis $(1, (t-\alpha), \ldots, (t-\alpha)^n)$.

7. **Coordinate Isomorphism.** Let V be the space of 2×2 symmetric matrices. Construct an explicit isomorphism from V to $\mathbb{R}^3$. What are the coordinates of the identity matrix under your map?

8. **Cancellation Failure.** Prove by counterexample that the cancellation law for direct sums fails: $U \oplus W_1 = U \oplus W_2$ does not imply $W_1 = W_2$. Consider lines in $\mathbb{R}^2$.

9. **Quotients of $\mathbb{R}[t]$.**

   (a) Describe a natural basis for $\mathbb{R}[t]$ as a vector space over $\mathbb{R}$, and recall the definition of a coset $p(t) + L$.

   (b) Give a criterion for when $\mathbb{R}[t]/L$ is finite-dimensional in terms of L.

   (c) Apply your criterion to determine whether the following quotient spaces are finite-dimensional, and find the dimension when it is finite:

      (i) $L = P_n$ (polynomials of degree $\leq n$),

      (ii) L is the subspace of polynomials divisible by t^n,

      (iii) L is the subspace of polynomials in t^2 (even polynomials).

10. **Codimension Formula.**

    (a) If $U \leq V$ has finite codimension, explain how $\operatorname{codim} U$ can be defined using $\dim(V/U)$.

    (b) Prove the formula

    $$\operatorname{codim}(U + W) + \operatorname{codim}(U \cap W) = \operatorname{codim} U + \operatorname{codim} W$$

    for finite-dimensional V.

    (c) Extend the argument to the case where V may be infinite-dimensional but U and W have finite codimension.

11. **Intersection of Subspaces.** Let $V_1, \ldots, V_k$ be subspaces of an n-dimensional space V. Prove that if $\sum_{i=1}^{k} \dim V_i > n(k-1)$, then the intersection $\bigcap_{i=1}^{k} V_i$ is non-trivial (contains a non-zero vector).

12. **Magic Squares.** Following the terminology of example 1.9, we

single out the obvious semi-magic matrices:

$$0, \quad E = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}, \quad D = \begin{bmatrix} & & & 1 \\ & & \iddots & \\ & 1 & & \\ 1 & & & \end{bmatrix}, \quad S = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}.$$

Determine the dimensions $\dim \mathrm{SMag}_n(\mathbb{Q})$ and $\dim \mathrm{Mag}_n(\mathbb{Q})$. Clearly, $\mathrm{SMag}_2(\mathbb{Q}) = \langle E, D \rangle_{\mathbb{Q}}$. In this case, $S = E + D$ is the only (up to a rational factor) magic matrix. For $n = 3$, consider the magic matrix:

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

Calculate the dimensions mentioned above for $n = 3$ and $n = 4$.

13. Direct Sum Decomposition. Prove the direct sum decomposition:

$$\mathrm{SMag}_n(\mathbb{Q}) = \mathrm{Mag}_n(\mathbb{Q}) \oplus \mathbb{Q}E \oplus \mathbb{Q}D.$$

14. Basis Extension. Let $S = \{1 + t, 1 + t^2\}$ be a subset of P_3 (polynomials of degree ≤ 3).

(a) Prove that S is linearly independent.

(b) Extend S to a basis for P_3.

15. Row and Column Spaces. Let A be the matrix

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

(a) Find a basis for the row space of A.

(b) Find a basis for the column space of A.

(c) Verify that $\dim(\mathrm{row}(A)) = \dim(\mathrm{col}(A))$.

16. Geometric Intersection. Let W_1 be the plane $x + y + z = 0$ and W_2 be the plane $x - y = 0$ in $\mathbb{R}^3$.

(a) Find a basis for the intersection $W_1 \cap W_2$.

(b) Determine $\dim W_1$ and $\dim W_2$.

(c) Verify Grassmann's formula: $\dim(W_1 + W_2) = \dim W_1 + \dim W_2 - \dim(W_1 \cap W_2)$.

For (c): What must $W_1 + W_2$ be if it contains two distinct planes?

2
Linear Maps

2.1 *Linear Maps and Operators*

We now turn our attention from the internal structure of vector spaces to the relationships between them. Linear maps are the fundamental functions that preserve the algebraic structure of vector spaces.

Definition 2.1. ***Linear Map.***
Let V and W be vector spaces over the same field F. A function $T : V \to W$ is called a **linear map** (or **linear transformation**) if it satisfies the following two conditions for all $u, v \in V$ and $\alpha \in F$:
Additivity: $T(u+v) = T(u) + T(v)$.
Homogeneity: $T(\alpha u) = \alpha T(u)$.
These conditions can be combined into a single requirement:

$$T(\alpha u + \beta v) = \alpha T(u) + \beta T(v) \quad \text{for all } u, v \in V \text{ and } \alpha, \beta \in F.$$

If $W = V$, the map T is often called a **linear operator**. The set of all linear maps from V to W is denoted by $\mathcal{L}(V, W)$ or $\mathrm{Hom}(V, W)$.

定義

Note

We use 0_V and 0_W to denote the zero vectors in V and W respectively. When there is no ambiguity, we simply write 0.

Proposition 2.1. ***Vector Space Structure of Maps.***
The set $\mathcal{L}(V, W)$ forms a vector space over F under pointwise operations. For $f, g \in \mathcal{L}(V, W)$ and $\lambda \in F$, we define:

$$(f+g)(x) = f(x) + g(x), \quad (\lambda f)(x) = \lambda f(x).$$

命題

Proof

The verification of the axioms is direct. For instance, additivity of the sum $f + g$ follows from the additivity of f and g and the commutativity of vector addition in W.

■

Proposition 2.2. ***Basic Properties.***
Let $T \in \mathcal{L}(V, W)$.
1. $T(0_V) = 0_W$.
2. $T(u - v) = T(u) - T(v)$.
3. Linearity extends to finite sums: $T\left(\sum_{i=1}^{k} a_i v_i\right) = \sum_{i=1}^{k} a_i T(v_i)$.

命題

Proof

1. $T(0) = T(0 + 0) = T(0) + T(0)$. Adding $-T(0)$ to both sides yields $T(0) = 0$.
2. $T(u - v) = T(u + (-1)v) = T(u) + (-1)T(v) = T(u) - T(v)$.
3. Follows by induction on k.

■

Example 2.1. Trivial Maps.
- The **zero map** $\mathcal{O} : V \to W$ defined by $\mathcal{O}(v) = 0_W$ for all v.
- The **identity map** $\mathcal{E} : V \to V$ defined by $\mathcal{E}(v) = v$.

範例

Example 2.2. Geometric Transformations in $\mathbb{R}^2$. Consider maps $T : \mathbb{R}^2 \to \mathbb{R}^2$.

Reflection: Let $T(x_1, x_2) = (x_1, -x_2)$. This reflects a vector across the x_1-axis.

Rotation: Let T_θ rotate a vector by an angle θ counter-clockwise. Using polar coordinates, if $x = (r\cos\alpha, r\sin\alpha)$, the rotated vector y is $(r\cos(\alpha + \theta), r\sin(\alpha + \theta))$. Expanding this:

$$y_1 = r\cos\alpha\cos\theta - r\sin\alpha\sin\theta = x_1\cos\theta - x_2\sin\theta$$
$$y_2 = r\sin\alpha\cos\theta + r\cos\alpha\sin\theta = x_2\cos\theta + x_1\sin\theta$$

This can be written as matrix multiplication $y = A_\theta x$, where

$$A_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

Since matrix multiplication is linear, rotation is a linear map.

範例

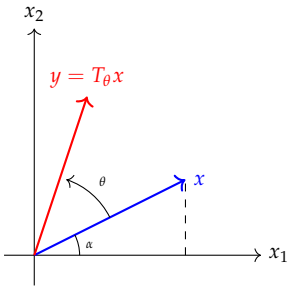


Figure 2.1: Rotation of a vector x by angle θ.

Example 2.3. Projections and Inclusions. Let $m < n$.
Projection: $P : \mathbb{R}^n \to \mathbb{R}^m$ defined by $P(x_1, \ldots, x_n) = (x_1, \ldots, x_m)$.
Natural Inclusion: $\iota : \mathbb{R}^m \to \mathbb{R}^n$ defined by $\iota(x_1, \ldots, x_m) = (x_1, \ldots, x_m, 0, \ldots, 0)$.
Ideally, projections "flatten" the space onto a subspace, while inclusions embed a smaller space into a larger one.

範例

Example 2.4. Differentiation and Integration. Let $C^1[0,1]$ be the space of continuously differentiable real functions. The map $D : C^1[0,1] \to C[0,1]$ defined by $D(f) = f'$ is linear:

$$D(\alpha f + \beta g) = \frac{d}{dt}(\alpha f + \beta g) = \alpha f' + \beta g' = \alpha D(f) + \beta D(g).$$

Similarly, the integral map $I : C[0,1] \to \mathbb{R}$ defined by $I(f) = \int_0^1 f(t)\,dt$ is a linear functional.

範例

Note

A **linear functional** on a vector space V over F is a linear map $f : V \to F$.

Example 2.5. The Transpose Map. Consider the space of matrices $F^{m\times n}$. The map $T : F^{m\times n} \to F^{n\times m}$ defined by $T(A) = A^T$ is linear.

$$T(\alpha A + \beta B) = (\alpha A + \beta B)^T = \alpha A^T + \beta B^T = \alpha T(A) + \beta T(B).$$

範例

Remark.

Not all geometric transformations are linear. The translation map $T(x) = x + b$ with $b \neq 0$ is not linear, as $T(0) = b \neq 0$. Such maps are called **affine**.

2.2 *Kernel and Image*

Associated with any linear map are two fundamental subspaces.

Definition 2.2. ***Kernel and Image.***
Let $f : V \to W$ be a linear map.
- The **kernel** (or nullspace) of f is $\operatorname{Ker} f = \{v \in V \mid f(v) = 0\}$.
- The **image** (or range) of f is $\operatorname{Im} f = \{w \in W \mid \exists v \in V, f(v) = w\}$.

定義

It is routine to verify that $\operatorname{Ker} f$ is a subspace of V and $\operatorname{Im} f$ is a subspace of W. For the image, if $w_1 = f(u_1)$ and $w_2 = f(u_2)$, then $\alpha w_1 + \beta w_2 = f(\alpha u_1 + \beta u_2) \in \operatorname{Im} f$.

Lemma 2.1. ***Injectivity and Kernel.***
A linear map f is injective if and only if $\operatorname{Ker} f = \{0\}$.

引理

$(\Longrightarrow)$

Since $f(0) = 0$, if f is injective, 0 is the unique element mapping to 0.

証明終

$(\Longleftarrow)$

Suppose $\operatorname{Ker} f = \{0\}$ and $f(x) = f(y)$. By linearity, $f(x-y) = 0$, so $x - y \in \operatorname{Ker} f$. Thus $x - y = 0$, implying $x = y$.

証明終

Example 2.6. Injectivity and Dimension. We investigate maps defined by variations of $f(x, y) = x \pm y$ to illustrate how dimension influences injectivity and surjectivity.

Higher to Lower Dimension ***($n > m$):*** Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be defined by $T(x) = (x_1 + x_2, x_1 - x_2)$. The kernel is determined by the system $x_1 + x_2 = 0$ and $x_1 - x_2 = 0$, which implies $x_1 = x_2 = 0$. However, x_3 is unconstrained. Thus $\operatorname{Ker} T = \operatorname{span}(e_3) \neq \{0\}$, so T is not injective. The system $T(x) = y$ is solvable for all $y \in \mathbb{R}^2$ (as the defining matrix is invertible), so T is surjective.

Lower to Higher Dimension ***($n < m$):*** Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be defined by $T(x) = (x_1 + x_2, x_1 - x_2, 0)$. The kernel requires $x_1 + x_2 = 0$ and $x_1 - x_2 = 0$, yielding $x = 0$. Thus T is injective (*lemma 2.1*). The range consists only of vectors with a zero third component (e.g., $(0, 0, 1)$ is not in the image), so T is not surjective.

Equal Dimension ***($n = m$):*** Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $T(x) = (x_1 + x_2, x_1 - x_2)$. The kernel is trivial ($x = 0$), so T is injective. The defining matrix is invertible, ensuring T is also surjective.

This highlights a general principle: a linear map cannot be injective if the domain is 'larger' than the codomain, nor surjective if the domain is 'smaller' (proved later in this section).

範例

Theorem 2.1. ***Mapping of Spanning Sets.***
Let $f : V \to W$ be linear and $U = \langle e_1, \ldots, e_s \rangle$ be a subspace of V. Then:

$$f(U) = \langle f(e_1), \ldots, f(e_s) \rangle.$$

Consequently, $\dim f(U) \leq \dim U$.

定理

Proof

Any $u \in U$ is of the form $\sum \alpha_i e_i$. Then $f(u) = \sum \alpha_i f(e_i)$, which lies in $\text{span}(\{f(e_i)\})$. Thus $f(U) \subseteq \langle f(e_i) \rangle$. Conversely, any linear combination of $f(e_i)$ is the image of the corresponding combination of e_i. Let B be a basis of U. Then $f(B)$ spans $f(U)$, so $\dim f(U) \leq |B| = \dim U$.

■

Definition 2.3. ***Rank.***
The **rank** of a linear map f, denoted $\text{rank}\, f$, is the dimension of its image, $\dim(\text{Im}\, f)$.

定義

2.3 *Matrix Representation*

We have seen that $m \times n$ matrices induce linear maps. Conversely, every linear map between finite-dimensional spaces can be represented by a matrix.

Theorem 2.2. ***Map Determined by Basis.***
Let V and W be vector spaces, with V finite-dimensional. Let $\{v_1, \ldots, v_n\}$ be a basis for V. For any vectors $w_1, \ldots, w_n \in W$, there exists a **unique** linear map $f : V \to W$ such that $f(v_i) = w_i$ for all i.

定理

Proof

Existence: For any $x \in V$, write $x = \sum \alpha_i v_i$ uniquely. Define $f(x) = \sum \alpha_i w_i$. Linearity is easily checked.

Uniqueness: If g is another such map, then for any $x = \sum \alpha_i v_i$, linearity forces $g(x) = \sum \alpha_i g(v_i) = \sum \alpha_i w_i = f(x)$.

■

Note

This theorem implies that two linear maps are equal if and only if they agree on a basis.

Remark.

The uniqueness of the coordinate representation $x = \sum \alpha_i v_i$ is essential for the existence proof. It ensures the map f is **well-defined**: if a vector x had multiple representations, the formula $f(x) = \sum \alpha_i w_i$ could yield different results for the same vector, violating the definition of a function.

Example 2.7. Constructing the Map. Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be defined by its action on the standard basis:

$$T(e_1) = (-1, 0), \quad T(e_2) = (1, 1), \quad T(e_3) = (0, 1).$$

For an arbitrary vector $x = (x_1, x_2, x_3) = x_1e_1 + x_2e_2 + x_3e_3$, we have:

$$\begin{aligned} T(x) &= x_1T(e_1) + x_2T(e_2) + x_3T(e_3) \\ &= x_1(-1, 0) + x_2(1, 1) + x_3(0, 1) \\ &= (-x_1 + x_2, x_2 + x_3). \end{aligned}$$

範例

Example 2.8. Matrix of Differentiation Operator. Consider the differentiation map $D : P_3 \to P_2$, where P_k is the space of polynomials of degree at most k. Let $\mathcal{B} = (1, t, t^2, t^3)$ be the basis for P_3 and $\mathcal{C} = (1, t, t^2)$ be the basis for P_2. We compute the image of each basis vector from $\mathcal{B}$:

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

The coordinate columns are $[0, 0, 0]^\top$, $[1, 0, 0]^\top$, $[0, 2, 0]^\top$, and $[0, 0, 3]^\top$. Thus, the matrix is:

$$M_D = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}.$$

範例

Let V and W be spaces with fixed bases $\mathcal{B}_V = (v_1, \ldots, v_n)$ and $\mathcal{B}_W = (w_1, \ldots, w_m)$. Let $f : V \to W$ be a linear map. We decompose the images of the basis vectors of V into the basis of W:

$$f(v_j) = \sum_{i=1}^{m} a_{ij} w_i \quad \text{for } j = 1, \ldots, n.$$

The $m \times n$ matrix $M_f = (a_{ij})$ is called the **matrix of** f relative to these bases. Specifically, the j-th column of M_f contains the coordinates of $f(v_j)$.

If $x = \sum x_j v_j$ is a vector in V with coordinate column X, and $y = f(x) = \sum y_i w_i$ has coordinate column Y, then the linearity of f implies:

$$Y = M_f X.$$

Theorem 2.3. ***Isomorphism of Maps and Matrices.***
Let V and W be finite-dimensional vector spaces over F with fixed bases. The map $\Phi : \mathcal{L}(V,W) \to F^{m\times n}$ given by $f \mapsto M_f$ is a vector space isomorphism. Consequently,

$$\dim \mathcal{L}(V,W) = (\dim V)(\dim W).$$

定理

Proof

The map is linear: $M_{\alpha f+\beta g} = \alpha M_f + \beta M_g$ because coordinates satisfy $(f+g)(v_j) = f(v_j) + g(v_j)$. Bijectivity follows from *theorem 2.2*: every matrix defines a unique set of images for the basis of V, which defines a unique map.

■

Theorem 2.4. ***Row Rank Equals Column Rank.***
For any matrix $A \in F^{m\times n}$, the dimension of the column space of A equals the dimension of the row space of A. This common value is the number of pivots in the RREF of A.

定理

Proof

Let R be the RREF of A. We know that $Ax = 0$ if and only if $Rx = 0$. This implies that the linear dependence relations among the columns of A are identical to those among the columns of R. The pivot columns of R are the standard basis vectors $e_1, \ldots, e_r$ (truncated to size m) and are clearly linearly independent. Thus, the corresponding columns of A are linearly independent. Every non-pivot column of R is a linear combination of the pivot columns to its left. Because the dependence relations are the same, the corresponding columns of A are linear combinations of the pivot columns of A. Therefore, the pivot columns of A form a basis for $\mathrm{col}(A)$. The size of this basis is the number of pivots, which is exactly the row rank of A (the number of non-zero rows in R).

■

Proposition 2.3. ***Consistency of Rank.***
Let $f \in \mathcal{L}(V,W)$ and let M_f be its matrix representation relative to any choice of bases. Then

$$\operatorname{rank} f = \operatorname{rank}(M_f).$$

命題

Proof

Fix bases $\mathcal{B}_V$ and $\mathcal{B}_W$. For any $x \in V$ with coordinate column X, the image $f(x)$ has coordinate column $Y = M_f X$. Thus the coordinate map $\psi : W \to F^m$ restricts to an isomorphism

$$\psi : \operatorname{Im} f \to \operatorname{Im}(M_f),$$

where $\operatorname{Im}(M_f)$ denotes the column space of M_f in F^m. Hence

$$\dim \operatorname{Im} f = \dim \operatorname{col}(M_f).$$

By *theorem* 2.4, $\dim \operatorname{col}(M_f)$ equals the row rank of M_f (the number of pivots). Therefore $\operatorname{rank} f = \operatorname{rank}(M_f)$.

■

Proposition 2.4. ***Composition and Matrix Multiplication.***
Let U, V, W be vector spaces with bases. Let $g : U \to V$ and $f : V \to W$ be linear maps with matrices M_g and M_f. Then the composition $f \circ g : U \to W$ corresponds to the matrix product:

$$M_{f \circ g} = M_f M_g.$$

命題

Proof

Let $x \in U$ with coordinate column X in the chosen basis of U. Then $g(x)$ has coordinates $Y = M_g X$ in the basis of V, and the image $f(g(x))$ has coordinates $Z = M_f Y = M_f(M_g X)$ in the basis of W. Thus $Z = (M_f M_g)X$ for all X, so the matrix of $f \circ g$ is $M_f M_g$.

■

This correspondence allows us to derive properties of maps from matrices and vice versa.

Theorem 2.5. ***Rank Inequalities.***
Let $f : V \to W$ and $g : U \to V$. Then:
1. $\dim \operatorname{Im}(f \circ g) \leq \dim \operatorname{Im} f$.
2. $\dim \operatorname{Im}(f \circ g) \leq \dim \operatorname{Im} g$.

定理

Proof

1. Since $\operatorname{Im}(f \circ g) \subseteq \operatorname{Im} f$, the dimension inequality is immediate.
2. Note that $\operatorname{Im}(f \circ g) = f(\operatorname{Im} g)$. By the "Mapping of Spanning Sets" theorem, applying f to the subspace $\operatorname{Im} g$ cannot increase its dimension. Thus $\dim f(\operatorname{Im} g) \leq \dim \operatorname{Im} g$.

■

2.4 *Dimension Theorem*

The dimensions of the kernel and image are fundamentally linked by the dimension of the domain.

Theorem 2.6. ***Rank-Nullity Theorem.***
Let V be a finite-dimensional vector space over the field F, and $f : V \to W$ be a linear map. Then $\operatorname{Ker} f$ and $\operatorname{Im} f$ are both finite-dimensional, and

$$\dim \operatorname{Ker} f + \dim \operatorname{Im} f = \dim V.$$

定理

Proof

Since $\operatorname{Ker} f$ is a subspace of V, $\dim \operatorname{Ker} f \leq \dim V$. Let $(e_1, \ldots, e_k)$ be a basis for $\operatorname{Ker} f$ (where $k = \dim \operatorname{Ker} f$). Extend this to a basis $(e_1, \ldots, e_k, e_{k+1}, \ldots, e_n)$ of V. Any vector in $\operatorname{Im} f$ is of the form

$$f\left(\sum_{i=1}^{n} \alpha_i e_i\right) = \sum_{i=1}^{n} \alpha_i f(e_i) = \sum_{i=k+1}^{n} \alpha_i f(e_i),$$

since $f(e_i) = 0$ for $i \leq k$. Thus, $S = \{f(e_{k+1}), \ldots, f(e_n)\}$ spans $\operatorname{Im} f$. To show independence, suppose $\sum_{j=k+1}^{n} \lambda_j f(e_j) = 0$. Then $f(\sum \lambda_j e_j) = 0$, implying $v = \sum \lambda_j e_j \in \operatorname{Ker} f$. Thus v can be written as a linear combination of the kernel basis: $\sum_{j=k+1}^{n} \lambda_j e_j = \sum_{i=1}^{k} \mu_i e_i$. Rearranging gives a linear dependence relation among the basis vectors of V:

$$\sum_{i=1}^{k} (-\mu_i) e_i + \sum_{j=k+1}^{n} \lambda_j e_j = 0.$$

Since the basis of V is independent, all coefficients must be zero, specifically $\lambda_j = 0$. Thus S is a basis for $\operatorname{Im} f$, and $\dim \operatorname{Im} f = n - k = \dim V - \dim \operatorname{Ker} f$.

■

Corollary 2.1. *Injectivity in Finite Dimensions.* If $\dim V < \infty$, the following are equivalent for a linear map $f : V \to W$:
1. f is injective.
2. $\dim V = \dim \operatorname{Im} f$.

推論

Proof

By the Rank-Nullity Theorem, $\dim V = \dim \operatorname{Im} f$ if and only if $\dim \operatorname{Ker} f = 0$, which is equivalent to $\operatorname{Ker} f = \{0\}$, i.e., injectivity.

■

Corollary 2.2. *Dimensional Constraints.* Let $f : V \to W$ be a linear map between finite-dimensional spaces with $\dim V = n$ and $\dim W =$

m.

1. If $n > m$, then f is not injective.
2. If $n < m$, then f is not surjective.

推論

Proof

1. By Rank-Nullity (*theorem* 2.6), $\dim \operatorname{Im} f = n - \dim \operatorname{Ker} f$. Since $\operatorname{Im} f \leq W$, we have $n - \dim \operatorname{Ker} f \leq m$. If f were injective, $\operatorname{Ker} f = \{0\}$, implying $n \leq m$. Thus $n > m$ forces a non-trivial kernel.
2. The rank is at most n. If $n < m$, then $\dim \operatorname{Im} f \leq n < m = \dim W$, so the image cannot be all of W.

■

Remark.

If $\dim V = \dim W$ (e.g., a linear operator $f : V \to V$), then injectivity implies $\dim \operatorname{Im} f = \dim V = \dim W$, so $\operatorname{Im} f = W$ (surjectivity). Thus, for operators on finite-dimensional spaces, injectivity $\iff$ surjectivity $\iff$ bijectivity.

2.5 *Isomorphisms*

We end this chapter with some proofs for the properties of isomorphism. We recall the definition of an isomorphism from the previous chapter. A linear map $T : V \to W$ is an isomorphism if it is bijective.

Proposition 2.5. ***Linearity of Inverse.***
If $f : V \to W$ is a bijective linear map, then its set-theoretic inverse $f^{-1} : W \to V$ is also linear.

命題

Proof

Let $x, y \in W$ and $\alpha \in F$. Since f is surjective, there exist unique $u, v \in V$ such that $f(u) = x$ and $f(v) = y$. Thus $f^{-1}(x) = u$ and $f^{-1}(y) = v$. Using the linearity of f:

$$f(u+v) = f(u) + f(v) = x + y \implies f^{-1}(x+y) = u + v = f^{-1}(x) + f^{-1}(y).$$

Similarly, $f(\alpha u) = \alpha f(u) = \alpha x \implies f^{-1}(\alpha x) = \alpha u = \alpha f^{-1}(x)$.

■

Proposition 2.6. ***Isomorphisms Preserve Bases.***
Let $f : V \to W$ be an isomorphism. If $\{e_1, \ldots, e_n\}$ is a basis for V, then $\{f(e_1), \ldots, f(e_n)\}$ is a basis for W.

命題

Proof

Let $B' = \{f(e_1), \ldots, f(e_n)\}$. Since f is surjective and $\{e_i\}$ spans V, the image of the span is the span of the image (*theorem* 2.1), so B' spans W. For independence, suppose $\sum \lambda_i f(e_i) = 0$. By linearity, $f(\sum \lambda_i e_i) = 0$. Since f is injective, $\operatorname{Ker} f = \{0\}$ (*lemma* 2.1), so $\sum \lambda_i e_i = 0$. The independence of $\{e_i\}$ implies all $\lambda_i = 0$.

■

Consequently, isomorphic finite-dimensional spaces have the same dimension.

2.6 *Exercises*

1. **Verification of Linearity.** Determine whether the following maps are linear:
   (a) $f : V \to V/L$, where $L \leq V$, defined by $f(v) = v + L$ (the canonical projection).
   (b) $f : P_n \to P_n$ defined by $f(u(t)) = tu'(t) - u(t)$ for $n \geq 1$. Find $\operatorname{Ker} f$ and $\operatorname{rank} f$ (for the trivial case $n = 0$, $f(c) = -c$ has trivial kernel and rank 1).
   (c) $f_C : M_n(F) \to M_n(F)$ defined by $f_C(X) = C^{-1}XC$, where C is invertible. Verify that $f_C(XY) = f_C(X)f_C(Y)$ (automorphism property).
2. **Geometric Transformations.** Find the matrix representation of the linear map $T : \mathbb{R}^2 \to \mathbb{R}^2$ that reflects vectors across the line $y = 2x$.
3. **Rank-Nullity Practice.** Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be defined by $T(x, y, z) = (x + y, y + z, z + x)$. Find a basis for $\operatorname{Ker} T$ and $\operatorname{Im} T$. Verify the Rank-Nullity Theorem.
4. **Injection and Surjection.** Let $T : V \to W$.
   (a) If T is injective and $\{v_1, \ldots, v_k\}$ is linearly independent in V, prove that $\{T(v_1), \ldots, T(v_k)\}$ is linearly independent in W.
   (b) If T is surjective and $\{v_1, \ldots, v_k\}$ spans V, prove that $\{T(v_1), \ldots, T(v_k)\}$ spans W.
5. **Map from Basis.** Let $V = \mathbb{R}^2$. Let e_1, e_2 be the standard basis. Define T by $T(e_1) = (2, 1)$ and $T(e_2) = (1, -1)$. Find $T(3, 4)$.
6. **Left/Right Multiplication Matrices (a.k.a. Kronecker product block form).** Identify the space $M_2(\mathbb{R})$ with $\mathbb{R}^4$ via the coordinate vector $\mathbf{x} = [x_1, x_2, x_3, x_4]^\top$ corresponding to

$$X = \begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix}.$$

Let

$$A = \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix}.$$

Define two linear maps on $M_2(\mathbb{R})$:

$$f_L(X) = AX, \quad f_R(X) = XA.$$

Verify that the matrix representations M_{f_L} and M_{f_R} relative to the standard basis of $\mathbb{R}^4$ are:

$$M_{f_L} = \begin{bmatrix} a_1 & 0 & a_2 & 0 \\ 0 & a_1 & 0 & a_2 \\ a_3 & 0 & a_4 & 0 \\ 0 & a_3 & 0 & a_4 \end{bmatrix}, \quad M_{f_R} = \begin{bmatrix} a_1 & a_3 & 0 & 0 \\ a_2 & a_4 & 0 & 0 \\ 0 & 0 & a_1 & a_3 \\ 0 & 0 & a_2 & a_4 \end{bmatrix}.$$

7. **Idempotent Maps.** A linear map $P : V \to V$ is called a projection if $P^2 = P$.
   (a) Prove that $V = \operatorname{Im} P \oplus \operatorname{Ker} P$.
   (b) If P is a projection, show that $\mathcal{E} - P$ is also a projection. What are its image and kernel?

8. **Rank Inequalities.** Let A, B be $n \times n$ matrices. Prove Sylvester's Rank Inequality:

$$\operatorname{rank}(A) + \operatorname{rank}(B) - n \leq \operatorname{rank}(AB) \leq \min(\operatorname{rank}(A), \operatorname{rank}(B)).$$

3
Linear Operator Algebra

We now shift our perspective from the general mapping between distinct vector spaces to the rich internal structure of maps from a space to itself. These endomorphisms form an algebraic structure that allows us to apply the tools of polynomial ring theory to linear algebra.

3.1 The Algebra of Operators

Throughout this chapter, let V be a vector space over a field F. We abbreviate the space of linear maps $\mathcal{L}(V, V)$ as $\mathcal{L}(V)$. The elements of this space are called **linear operators** or simply **operators**.

> **Notation 3.1.** Operator Notation We adopt the convention of using calligraphic Latin letters ($\mathcal{A}, \mathcal{B}, \mathcal{C}, \dots$) to denote linear operators. Their corresponding matrix representations with respect to a fixed basis (e_i) will be denoted by standard Roman capitals ($A, B, C, \dots$). If the basis is changed to (e'_i), the matrices will be denoted $A', B', \dots$. We denote the identity operator by $\mathcal{E}$ (where $\mathcal{E}x = x$ for all x) and the corresponding identity matrix by $E = (\delta_{ij})$. The action of an operator $\mathcal{A}$ on a vector x is written as $\mathcal{A}x$, omitting parentheses where no ambiguity arises.
>
> 記法

Algebraic Structure

We have previously established that $\mathcal{L}(V)$ is a vector space. However, operators can also be composed. This introduces a multiplicative structure compatible with vector addition.

> **Definition 3.1.** ***Algebra over a Field.***
> A ring K is called an **algebra** over a field F if K is equipped with a vector space structure over F such that scalar multiplication is compatible with ring multiplication:
>
> $$\lambda(xy) = (\lambda x)y = x(\lambda y)$$

for all $\lambda \in F$ and $x, y \in K$. If the multiplication is associative, it is an **associative algebra**.

定義

Proposition 3.1. ***The Operator Algebra.***
The set $\mathcal{L}(V)$ forms an associative algebra over F with identity $\mathcal{E}$. For all $\mathcal{A}, \mathcal{B}, \mathcal{C} \in \mathcal{L}(V)$ and $\alpha, \beta \in F$, the following hold:
Linearity: $\alpha(\mathcal{A} + \mathcal{B}) = \alpha\mathcal{A} + \alpha\mathcal{B}$ and $(\alpha + \beta)\mathcal{A} = \alpha\mathcal{A} + \beta\mathcal{A}$.
Associativity: $\mathcal{A}(\mathcal{B}\mathcal{C}) = (\mathcal{A}\mathcal{B})\mathcal{C}$.
Distributivity: $\mathcal{A}(\mathcal{B} + \mathcal{C}) = \mathcal{A}\mathcal{B} + \mathcal{A}\mathcal{C}$ and $(\mathcal{A} + \mathcal{B})\mathcal{C} = \mathcal{A}\mathcal{C} + \mathcal{B}\mathcal{C}$.
Scalar Compatibility: $\lambda(\mathcal{A}\mathcal{B}) = (\lambda\mathcal{A})\mathcal{B} = \mathcal{A}(\lambda\mathcal{B})$.
Furthermore, if $\dim V = n$, then $\dim \mathcal{L}(V) = n^2$.

命題

Proof

The vector space properties (1) follow from the definition of linear maps. The ring properties (2, 3) follow from the properties of function composition. For (4), we observe:

$$(\lambda(\mathcal{A}\mathcal{B}))x = \lambda(\mathcal{A}(\mathcal{B}x)) = \mathcal{A}(\lambda\mathcal{B}x) = \mathcal{A}((\lambda\mathcal{B})x),$$

which verifies the compatibility.

■

The correspondence between operators and matrices preserves this algebraic structure. If $\mathcal{A}$ and $\mathcal{B}$ are operators with matrices $A = (a_{ij})$ and $B = (b_{ij})$ relative to a basis (e_k), then the product operator $\mathcal{C} = \mathcal{A}\mathcal{B}$ corresponds to the matrix product $C = AB$.

Proof

Let $\mathcal{A}e_k = \sum_i a_{ik} e_i$ and $\mathcal{B}e_j = \sum_k b_{kj} e_k$. The action of the composite operator on a basis vector e_j is:

$$\begin{aligned}(\mathcal{A}\mathcal{B})e_j &= \mathcal{A}\left(\sum_k b_{kj} e_k\right) = \sum_k b_{kj} \mathcal{A} e_k \\ &= \sum_k b_{kj} \left(\sum_i a_{ik} e_i\right) = \sum_i \left(\sum_k a_{ik} b_{kj}\right) e_i.\end{aligned}$$

The coefficient of e_i is exactly the (i, j)-th entry of the matrix product AB.

■

3.2 *Fundamental Examples*

We examine several fundamental classes of operators.

Example 3.1. Zero and Scalar Operators.

1. The **zero operator** $\mathcal{O}$ maps every vector to the zero vector. Its rank is 0.
2. The **scalar operator** $\mathcal{A}_\lambda$ is defined by $\mathcal{A}_\lambda x = \lambda x$ for a fixed $\lambda \in F$. Its matrix in any basis is λE.

範例

Example 3.2. Rotation in the Plane. Recall the rotation map from Chapter 2. While we previously derived its matrix using coordinates, we can now view it algebraically by identifying $V = \mathbb{R}^2$ with the complex field $\mathbb{C}$. Identifying $\mathbb{R}^2$ with $\mathbb{C}$ via the basis $\{1, i\}$, this operation corresponds to multiplication by $e^{i\alpha}$. Explicitly, the map $z \mapsto e^{i\alpha} z$ yields:

$$\mathcal{A}(1) = \cos\alpha + i\sin\alpha, \quad \mathcal{A}(i) = i(\cos\alpha + i\sin\alpha) = -\sin\alpha + i\cos\alpha.$$

Thus, relative to the basis $(1, i)$ (identified with standard basis vectors e_1, e_2), the matrix is:

$$A = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}.$$

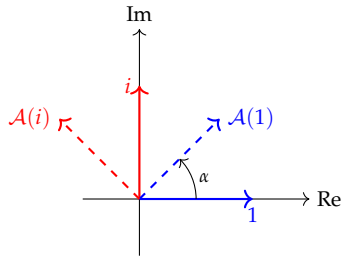


Figure 3.1: Action of the rotation operator $\mathcal{A}$ on the basis elements 1 and i.

範例

Example 3.3. Projections. Let $V = U \oplus W$. Every vector x has a unique decomposition $x = x_U + x_W$ with $x_U \in U, x_W \in W$. The **projection operator** $\mathcal{P}$ onto U along W is defined by $\mathcal{P}x = x_U$. Since $\mathcal{P}x_U = x_U$, applying the operator twice yields the same result: $\mathcal{P}^2 = \mathcal{P}$. Operators satisfying this idempotence property are characteristic of projections.

範例

Invertibility

An operator $\mathcal{B}$ is the **inverse** of $\mathcal{A}$ if $\mathcal{A}\mathcal{B} = \mathcal{B}\mathcal{A} = \mathcal{E}$. If such an operator exists, it is unique and denoted $\mathcal{A}^{-1}$. By previous results on linear maps, $\mathcal{A}$ is invertible if and only if $\operatorname{Ker}\mathcal{A} = \{0\}$. For operators on finite-dimensional spaces, this is equivalent to $\operatorname{rank}\mathcal{A} = \dim V$, or having nullity zero.

Example 3.4. The Differentiation Operator. Let P_n be the space of polynomials over F of degree at most $n - 1$. Let $\mathcal{D}$ be the differential operator defined by $\mathcal{D}(f) = f'$. The kernel of $\mathcal{D}$ is the subspace of constant polynomials, $\langle 1 \rangle$, which has dimension 1. The image of $\mathcal{D}$ is $P_{n-1} = \langle 1, t, \ldots, t^{n-2} \rangle$. Note that while the Rank-Nullity

Theorem holds:

$$\dim \operatorname{Ker} \mathcal{D} + \dim \operatorname{Im} \mathcal{D} = 1 + (n-1) = n = \dim P_n,$$

it is **not** true that $P_n = \operatorname{Ker} \mathcal{D} \oplus \operatorname{Im} \mathcal{D}$. Indeed, $\operatorname{Ker} \mathcal{D} \subset \operatorname{Im} \mathcal{D}$ (since constants are polynomials of degree 0), so their intersection is non-trivial. One must not conflate the arithmetic sum of dimensions with the direct sum of subspaces.

範例

Example 3.5. Infinite Dimensional Counterexamples. The equivalence between injectivity and surjectivity fails for infinite-dimensional spaces. Let P be the space of all real polynomials. Define the differentiation operator $\mathcal{D} : P \to P$ by $\mathcal{D}(f) = f'$. Since $\mathcal{D}(1) = 0$, the kernel is non-trivial ($\operatorname{Ker} \mathcal{D} \neq \{0\}$), so $\mathcal{D}$ is not injective. However, every polynomial has an antiderivative, so $\mathcal{D}$ is surjective. Conversely, define the integration operator $\mathcal{S} : P \to P$ by $(\mathcal{S}f)(t) = \int_0^t f(x)\,dx$. If $\mathcal{S}f = 0$, differentiating yields $f = 0$, so $\mathcal{S}$ is injective. However, the image of $\mathcal{S}$ consists only of polynomials with a zero constant term, so it is not surjective. Observe that $\mathcal{D}\mathcal{S} = \mathcal{E}$ (Fundamental Theorem of Calculus), so $\mathcal{S}$ is a right inverse of $\mathcal{D}$. However, $\mathcal{S}\mathcal{D} \neq \mathcal{E}$ since $\mathcal{S}\mathcal{D}(1) = \mathcal{S}(0) = 0 \neq 1$.

範例

3.3 *Polynomials of Operators*

The algebraic structure of $\mathcal{L}(V)$ allows us to substitute operators into polynomials. Let $f(t) = \sum_{i=0}^{m} a_i t^i \in F[t]$. We define the operator $f(\mathcal{A})$ by:

$$f(\mathcal{A}) = a_0 \mathcal{A}^m + a_1 \mathcal{A}^{m-1} + \cdots + a_{m-1}\mathcal{A} + a_m \mathcal{E}.$$

Here, $\mathcal{A}^k$ denotes the k-fold composition of $\mathcal{A}$, with $\mathcal{A}^0 = \mathcal{E}$.

Definition 3.2. ***Generated Subalgebra.***
The set of all polynomials in $\mathcal{A}$, denoted $F[\mathcal{A}]$, forms a subalgebra of $\mathcal{L}(V)$. It is the smallest subalgebra containing $\mathcal{A}$ and $\mathcal{E}$.

定義

Unlike the general algebra $\mathcal{L}(V)$, the subalgebra $F[\mathcal{A}]$ is **commutative**. For any $f, g \in F[t]$, we have $f(\mathcal{A})g(\mathcal{A}) = g(\mathcal{A})f(\mathcal{A})$, which follows from the fact that powers of $\mathcal{A}$ commute ($\mathcal{A}^k\mathcal{A}^l = \mathcal{A}^{k+l} = \mathcal{A}^l\mathcal{A}^k$).

The Minimal Polynomial

Since $\mathcal{L}(V)$ has finite dimension n^2, the powers $\mathcal{E}, \mathcal{A}, \mathcal{A}^2, \ldots, \mathcal{A}^{n^2}$ cannot be linearly independent. There must exist a non-trivial linear

combination equal to the zero operator $\mathcal{O}$. Thus, there exists a non-zero polynomial annihilating $\mathcal{A}$.

Definition 3.3. ***Minimal Polynomial.***
The **minimal polynomial** of $\mathcal{A}$, denoted $\mu_{\mathcal{A}}(t)$, is the unique monic polynomial of lowest degree such that $\mu_{\mathcal{A}}(\mathcal{A}) = \mathcal{O}$.

定義

Theorem 3.1. ***Properties of the Minimal Polynomial.***
Let $\mathcal{A} \in \mathcal{L}(V)$ and let $\mu_{\mathcal{A}}(t) = t^m + \mu_1 t^{m-1} + \cdots + \mu_m$ be its minimal polynomial.

1. The set $\{\mathcal{E}, \mathcal{A}, \ldots, \mathcal{A}^{m-1}\}$ is linearly independent.
2. $\dim F[\mathcal{A}] = m = \deg \mu_{\mathcal{A}}$.
3. If $f(t) \in F[t]$ annihilates $\mathcal{A}$ (i.e., $f(\mathcal{A}) = \mathcal{O}$), then $\mu_{\mathcal{A}}(t)$ divides $f(t)$.
4. $\mathcal{A}$ is invertible if and only if the constant term $\mu_m \neq 0$.

定理

Proof

1. Suppose $\sum_{i=0}^{m-1} \lambda_i \mathcal{A}^i = \mathcal{O}$. This corresponds to a polynomial $P(t)$ of degree less than m annihilating $\mathcal{A}$. By the minimality of m, $P(t)$ must be the zero polynomial.
2. The powers $\mathcal{A}^k$ for $k \geq m$ can be reduced to combinations of lower powers using the relation $\mathcal{A}^m = -\sum_{i=0}^{m-1} \mu_{m-i} \mathcal{A}^i$. Thus $\{\mathcal{E}, \ldots, \mathcal{A}^{m-1}\}$ spans $F[\mathcal{A}]$. With independence established in (1), it is a basis.
3. Perform Euclidean division: $f(t) = q(t)\mu_{\mathcal{A}}(t) + r(t)$, where $\deg r < \deg \mu_{\mathcal{A}}$ or $r = 0$. Substituting $\mathcal{A}$:

$$\mathcal{O} = f(\mathcal{A}) = q(\mathcal{A})\mu_{\mathcal{A}}(\mathcal{A}) + r(\mathcal{A}) = q(\mathcal{A})\mathcal{O} + r(\mathcal{A}) = r(\mathcal{A}).$$

   Since $r(t)$ has degree strictly less than m, $r(\mathcal{A}) = \mathcal{O}$ implies $r(t) = 0$. Thus $\mu_{\mathcal{A}} \mid f$.
4. ($\Longrightarrow$) Suppose $\mu_m \neq 0$. We have:

$$\mathcal{A}^m + \cdots + \mu_{m-1}\mathcal{A} + \mu_m \mathcal{E} = \mathcal{O}.$$

   Rearranging terms:

$$\mathcal{A}(\mathcal{A}^{m-1} + \cdots + \mu_{m-1}\mathcal{E}) = -\mu_m \mathcal{E}.$$

   Dividing by $-\mu_m$, we find an explicit inverse:

$$\mathcal{A}^{-1} = -\frac{1}{\mu_m}(\mathcal{A}^{m-1} + \cdots + \mu_{m-1}\mathcal{E}).$$

   ($\Longleftarrow$) Suppose $\mu_m = 0$. Then $\mu_{\mathcal{A}}(t) = tq(t)$ for some polynomial $q(t)$ of degree $m-1$. Thus $\mathcal{O} = \mathcal{A}q(\mathcal{A})$. Since $\deg q < m$,

$q(\mathcal{A}) \neq \mathcal{O}$. Therefore, $\mathcal{A}$ is a zero divisor in the algebra $\mathcal{L}(V)$, which implies it cannot be invertible. (Specifically, there exists a non-zero vector $y = q(\mathcal{A})x$ such that $\mathcal{A}y = 0$, so $\operatorname{Ker}\mathcal{A} \neq \{0\}$).

■

Remark.

The degree of the minimal polynomial satisfies $m \leq n^2$ simply because $\dim \mathcal{L}(V) = n^2$. However, a much stronger bound, $m \leq n$, holds. This is the content of the Cayley-Hamilton theorem, which we will explore in subsequent chapters.

3.4 *Operators and Change of Basis*

We have seen that a linear operator $\mathcal{A} : V \to V$ can be represented by a matrix A relative to a chosen basis. Since the choice of basis is arbitrary, it is crucial to understand how this matrix representation changes when the basis changes.

Note

Strictly speaking, when defining the matrix of an operator (or vector), the basis must be an **ordered basis**. Changing the order of vectors in the basis permutes the rows and columns of the corresponding matrix. We assume all bases are ordered sequences.

Definition 3.4. ***Similar Matrices.***
Let A and A' be $n \times n$ matrices over F. We say A' is **similar** to A, denoted $A' \sim A$, if there exists an invertible matrix B such that:

$$A' = B^{-1}AB.$$

定義

Similarity is an equivalence relation. It is reflexive ($A = E^{-1}AE$), symmetric ($A' = B^{-1}AB \implies A = (B^{-1})^{-1}A'B^{-1}$), and transitive ($A'' \sim A'$ and $A' \sim A \implies A'' \sim A$). This partitions the set of $n \times n$ matrices into equivalence classes.

Theorem 3.2. ***Change of Basis for Operators.***
Let $\mathcal{A}$ be a linear operator on V. Let A be the matrix of $\mathcal{A}$ relative to a basis $\mathcal{B} = (e_1, \ldots, e_n)$, and let A' be the matrix of $\mathcal{A}$ relative to a basis $\mathcal{B}' = (e'_1, \ldots, e'_n)$. Then $A' = B^{-1}AB$, where B is the transition matrix from $\mathcal{B}$ to $\mathcal{B}'$ (i.e., the columns of B are the coordinate vectors of e'_j in the basis $\mathcal{B}$). Consequently, two matrices represent the same linear operator in different bases if and only if they are similar.

定理

Proof

Let X and X' be the coordinate column vectors of $x \in V$ in bases $\mathcal{B}$ and $\mathcal{B}'$ respectively. The coordinate transformation is given by $X = BX'$. Let Y and Y' be the coordinates of $\mathcal{A}x$ in the respective bases. The operator action is represented by matrix multiplication: $Y = AX$ and $Y' = A'X'$. Since $Y = BY'$, we have:

$$BY' = Y = AX = A(BX') = (AB)X'.$$

Multiplying by B^{-1} gives $Y' = (B^{-1}AB)X'$. Since this holds for all X', we must have $A' = B^{-1}AB$.

■

Example 3.6. Numerical Coordinate Change. Consider $V = \mathbb{R}^3$. Let $x = (e, \pi, 0)$. Relative to the standard basis $\mathcal{B}_1 = (e_1, e_2, e_3)$, the coordinate vector is simply $[x]_{\mathcal{B}_1} = [e, \pi, 0]^\top$. Now consider the basis $\mathcal{B}_2 = (u_1, u_2, u_3)$ where $u_1 = (1, 1, 0)$, $u_2 = (1, -1, 0)$, and $u_3 = (0, 0, 1)$. We wish to find coordinates $\alpha_1, \alpha_2, \alpha_3$ such that $x = \alpha_1 u_1 + \alpha_2 u_2 + \alpha_3 u_3$. Writing this out:

$$(e, \pi, 0) = \alpha_1(1, 1, 0) + \alpha_2(1, -1, 0) + \alpha_3(0, 0, 1) = (\alpha_1 + \alpha_2, \alpha_1 - \alpha_2, \alpha_3).$$

Solving the system yields $\alpha_3 = 0$, $2\alpha_1 = e + \pi$, and $2\alpha_2 = e - \pi$. Thus, the coordinate vector relative to $\mathcal{B}_2$ is $[x]_{\mathcal{B}_2} = [\frac{e+\pi}{2}, \frac{e-\pi}{2}, 0]^\top$.

範例

Example 3.7. Powers of Matrices. Similarity is a powerful tool for computation. If $A' = B^{-1}AB$, then

$$(A')^k = (B^{-1}AB)^k = B^{-1}A(BB^{-1})A \dots AB = B^{-1}A^kB.$$

Ideally, we seek a basis where the matrix A' is diagonal, say $A' = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$. Then $(A')^k = \operatorname{diag}(\lambda_1^k, \dots, \lambda_n^k)$, making the computation of $A^k = B(A')^kB^{-1}$ trivial. This naturally extends to polynomials: if $f(t) \in F[t]$, then $f(A) = Bf(A')B^{-1}$.

範例

Invariants: Determinant and Trace

Since similar matrices represent the same underlying operator, properties shared by all matrices in a similarity class can be attributed to the operator itself.

Definition 3.5. ***Determinant and Trace of an Operator.***
Let $\mathcal{A}$ be a linear operator on a finite-dimensional space V. Let A be the matrix of $\mathcal{A}$ in any basis.

1. The **determinant** of $\mathcal{A}$ is $\det \mathcal{A} = \det A$. (We assume familiarity with the matrix determinant and its property $\det(XY) = \det X \det Y$; a coordinate-free treatment will be provided in a later chapter).
2. The **trace** of $\mathcal{A}$ is $\operatorname{tr} \mathcal{A} = \operatorname{tr} A = \sum_{i=1}^{n} a_{ii}$.

定義

Proposition 3.2. ***Well-Definedness.***
The determinant and trace are independent of the choice of basis.

命題

Proof

For the determinant, let $A' = B^{-1}AB$. Using the multiplicative property:

$$\det A' = \det(B^{-1}AB) = \det(B^{-1}) \det(A) \det(B) = \det(A) \det(B^{-1}B) = \det A.$$

For the trace, we use the cyclic property $\operatorname{tr}(XY) = \operatorname{tr}(YX)$.

$$\operatorname{tr}(B^{-1}AB) = \operatorname{tr}(B^{-1}(AB)) = \operatorname{tr}((AB)B^{-1}) = \operatorname{tr}(ABB^{-1}) = \operatorname{tr} A.$$

■

These invariants carry structural information. For instance, $\mathcal{A}$ is invertible if and only if $\det \mathcal{A} \neq 0$. The trace is a linear functional on $\mathcal{L}(V)$:

$$\operatorname{tr}(\alpha \mathcal{A} + \beta \mathcal{B}) = \alpha \operatorname{tr} \mathcal{A} + \beta \operatorname{tr} \mathcal{B}.$$

3.5 Nilpotent Operators and Commutators

A special class of operators plays a significant role in the structure theory of linear maps.

Definition 3.6. ***Nilpotent Operator.***
A linear operator $\mathcal{A}$ is called **nilpotent** if there exists a positive integer m such that $\mathcal{A}^m = \mathcal{O}$. The smallest such m is called the **nilpotency index**.

定義

For a nilpotent operator of index m, the minimal polynomial is $\mu_{\mathcal{A}}(t) = t^m$. Examples include the differentiation operator $\mathcal{D}$ on P_n (where $\mathcal{D}^n = \mathcal{O}$) and strictly upper triangular matrices.

Lie Algebras

The algebra $\mathcal{L}(V)$ is associative. However, we can define a new non-associative product that captures the "failure" of commutativity.

Definition 3.7. ***Commutator.***
The **commutator** of two operators $\mathcal{A}, \mathcal{B}$ is defined as:

$$[\mathcal{A}, \mathcal{B}] = \mathcal{A}\mathcal{B} - \mathcal{B}\mathcal{A}.$$

定義

Equipped with this operation, $\mathcal{L}(V)$ becomes a **Lie algebra**, denoted $\mathfrak{gl}(V)$ or $\mathfrak{gl}_n(F)$. The bracket satisfies:
Antisymmetry: $[\mathcal{A}, \mathcal{B}] = -[\mathcal{B}, \mathcal{A}]$ (which implies $[\mathcal{A}, \mathcal{A}] = 0$).
Jacobi Identity: $[[\mathcal{A}, \mathcal{B}], \mathcal{C}] + [[\mathcal{B}, \mathcal{C}], \mathcal{A}] + [[\mathcal{C}, \mathcal{A}], \mathcal{B}] = \mathcal{O}$.

Example 3.8. The Heisenberg Relation. In quantum mechanics, the position operator $\mathcal{X}$ (multiplication by x) and momentum operator $\mathcal{P}$ (differentiation) satisfy the canonical commutation relation $[\mathcal{P}, \mathcal{X}] = \mathcal{E}$. Let $V = F[t]$. Let $\mathcal{D}_t$ be differentiation and $\mathcal{F}_t$ be multiplication by t. Then:

$$[\mathcal{D}_t, \mathcal{F}_t]f(t) = \mathcal{D}_t(tf(t)) - t\mathcal{D}_t f(t) = (f(t) + tf'(t)) - tf'(t) = f(t).$$

Thus $[\mathcal{D}_t, \mathcal{F}_t] = \mathcal{E}$.

範例

Proposition 3.3. ***Trace Obstruction.***
If F has characteristic 0 (e.g., $\mathbb{R}$ or $\mathbb{C}$), there exist no operators $\mathcal{A}, \mathcal{B}$ on a **finite-dimensional** space V such that $[\mathcal{A}, \mathcal{B}] = \mathcal{E}$.

命題

Proof

Suppose such operators exist. Taking the trace of both sides:

$$\operatorname{tr}[\mathcal{A}, \mathcal{B}] = \operatorname{tr}(\mathcal{A}\mathcal{B} - \mathcal{B}\mathcal{A}) = \operatorname{tr}(\mathcal{A}\mathcal{B}) - \operatorname{tr}(\mathcal{B}\mathcal{A}) = 0.$$

However, $\operatorname{tr}(\mathcal{E}) = \dim V = n$. Since F has characteristic 0, $n \neq 0$, a contradiction.

■

Remark.

This result fails in characteristic p if p divides n. For example, if $n = p$, the matrices

$$J_p = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}, \quad N_p = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 2 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & p-1 & 0 \end{bmatrix}$$

satisfy $[J_p, N_p] = E_p$, as $\operatorname{tr}(E_p) = p \equiv 0 \pmod{p}$.

3.6 *Exercises*

1. **Kernel and Image Equality.** Let V be a finite-dimensional vector space and $\mathcal{T} \in \mathcal{L}(V)$ such that $\operatorname{Ker} \mathcal{T} = \operatorname{Im} \mathcal{T}$.
   (a) Prove that $\dim V$ must be even.
   (b) Construct an example of such an operator on $\mathbb{R}^2$.
   (c) Show that such an operator satisfies $\mathcal{T}^2 = \mathcal{O}$.
2. **One-Sided Inverses.** Let V be the space of all polynomials. Let $\mathcal{D}$ and $\mathcal{S}$ be the differentiation and integration operators defined in the text.
   (a) Verify explicitly that $\mathcal{D}$ is surjective.
   (b) Verify explicitly that $\mathcal{S}$ is not surjective.
3. **Trace Obstruction Matrices.** Verify that the matrices J_p and N_p introduced in the "Trace Obstruction" remark are indeed nilpotent of order p. Specifically, show that $J_p^p = N_p^p = 0$.
4. **Cyclic Property of Trace.** Prove that if A, B, C are matrices of size $n \times p$, $p \times q$, and $q \times n$ respectively, then:
$$\operatorname{tr}(ABC) = \operatorname{tr}(BCA) = \operatorname{tr}(CAB).$$
5. **Finite Field Automorphisms.** Interpret $GL_n(\mathbb{F}_p)$ as the group of automorphisms of an n-dimensional vector space V over the finite field $\mathbb{F}_p$.
   (a) Show that determining an automorphism is equivalent to choosing a basis for V.
   (b) Count the number of possible bases to find the order $|GL_n(\mathbb{F}_p)|$.
6. **The Special Linear Algebra.** Let $\mathfrak{sl}_n(F) = \{\mathcal{A} \in \mathcal{L}(V) \mid \operatorname{tr} \mathcal{A} = 0\}$.
   (a) Prove that $\mathfrak{sl}_n(F)$ is a subspace of $\mathcal{L}(V)$ of codimension 1.
   (b) Prove that it is a subalgebra of the Lie algebra $\mathfrak{gl}_n(F)$ (i.e., closed under the commutator bracket).
7. **Rank Intersection Formula.** Prove that for any linear operators $\mathcal{A}, \mathcal{B}$ on V:
$$\dim(\operatorname{Im} \mathcal{A} \cap \operatorname{Ker} \mathcal{B}) = \operatorname{rank} \mathcal{A} - \operatorname{rank} \mathcal{BA}.$$
   Consider the restriction of $\mathcal{B}$ to the subspace $\operatorname{Im} \mathcal{A}$.
8. **Frobenius Inequality.** Use the previous exercise to prove that for any operators $\mathcal{A}, \mathcal{B}, \mathcal{C}$ on V:
$$\operatorname{rank} \mathcal{BA} + \operatorname{rank} \mathcal{AC} \leq \operatorname{rank} \mathcal{A} + \operatorname{rank} \mathcal{BAC}.$$
9. **Iterated Kernels.** Prove that for any linear operator $\mathcal{A}$ and integer

$i \geq 1$:

$$\dim(\operatorname{Im} \mathcal{A}^{i-1} \cap \operatorname{Ker} \mathcal{A}) = \dim \operatorname{Ker} \mathcal{A}^i - \dim \operatorname{Ker} \mathcal{A}^{i-1}.$$

10. Field Extension and Similarity. Prove that if two real matrices $A, B \in M_n(\mathbb{R})$ are similar over $\mathbb{C}$ (i.e., $P^{-1}AP = B$ for some $P \in M_n(\mathbb{C})$), then they are similar over $\mathbb{R}$.

Write $P = X + iY$ and consider the polynomial $\det(X + tY)$.

11. Minimal Polynomial of a Vector. Let $\mu_{\mathcal{A}}(t)$ be the minimal polynomial of $\mathcal{A}$. For a vector $v \in V$, let $\mu_{\mathcal{A},v}(t)$ be the monic polynomial of lowest degree such that $\mu_{\mathcal{A},v}(\mathcal{A})v = 0$.

(a) Prove that $\mu_{\mathcal{A},v}(t)$ divides $\mu_{\mathcal{A}}(t)$ for any v.

(b) Prove that for any $u, v \in V$, $\mu_{\mathcal{A},u+v}(t)$ is the least common multiple of $\mu_{\mathcal{A},u}(t)$ and $\mu_{\mathcal{A},v}(t)$, provided the latter two are coprime.

(c) Conclude that there exists a vector $a \in V$ such that $\mu_{\mathcal{A},a}(t) = \mu_{\mathcal{A}}(t)$. (This vector is often called a cyclic vector if the degree is n).

12. Trace Zero and Main Diagonal. Let F be a field of characteristic zero. Prove that if $\operatorname{tr}(A) = 0$, then A is similar to a matrix with all zeros on the main diagonal. Proceed by induction:

(a) Show that if A is not a scalar multiple of the identity and $\operatorname{tr}(A) = 0$, there exists a vector x such that x and Ax are linearly independent.

(b) Use x as the first basis vector to show A is similar to a block matrix $\begin{bmatrix} 0 & * \\ * & A' \end{bmatrix}$.

(c) Apply the induction hypothesis to A'.

(d) Explain why the characteristic zero assumption is necessary.

13. Direct Sum Projections. Let $V = V_1 \oplus V_2$ and $W = W_1 \oplus W_2$ with $W_i \subseteq V_i$. Let $\mathcal{P}_2$ be the projection onto V_2 along V_1. Prove:

(a) If $V_1 = W_1 + (U \cap V_1)$ and $V_2 = W_2 + \mathcal{P}_2(U)$, then $V = W + U$.

(b) If $V = W + U$ and $\mathcal{P}_2(U) \cap W_2 = \{0\}$, then the decomposition in (a) holds.

14. The Center of the Algebra. The center of an algebra is the set of elements that commute with all other elements. Prove that the center of $\mathcal{L}(V)$ consists exactly of the scalar operators $\{\lambda\mathcal{E} \mid \lambda \in F\}$.

4
Dual Spaces

We have previously studied linear maps between arbitrary vector spaces. We now focus on the case where the codomain is the underlying field itself. This special class of maps reveals a deep symmetry inherent in vector spaces.

Definition 4.1. ***Linear Functional.***
Let V be a vector space over a field F. A map $f : V \to F$ is called a **linear functional** (or **linear form**) if it satisfies:

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y) \quad \text{for all } \alpha, \beta \in F \text{ and } x, y \in V.$$

定義

Let $(e_1, \ldots, e_n)$ be a basis for V. Any vector $x \in V$ can be uniquely expressed as $x = \sum_{i=1}^{n} \lambda_i e_i$. Applying a linear functional f to x yields:

$$f(x) = f\left(\sum_{i=1}^{n} \lambda_i e_i\right) = \sum_{i=1}^{n} \lambda_i f(e_i).$$

Let $\beta_i = f(e_i)$. These scalars are determined solely by the functional f and the choice of basis. Conversely, for any choice of scalars $\beta_1, \ldots, \beta_n \in F$, there exists a unique linear functional f such that $f(e_i) = \beta_i$. Thus, relative to a fixed basis, a linear functional is completely determined by the n-tuple $(\beta_1, \ldots, \beta_n)$.

Change of Basis

While the definition of a linear functional is independent of the basis, its coordinate representation depends on it. It is instructive to see how the coefficients β_i transform under a change of basis. Let $(e_1, \ldots, e_n)$ and $(e'_1, \ldots, e'_n)$ be two bases of V related by the transformation:

$$e'_j = \sum_{i=1}^{n} a_{ij} e_i, \quad j = 1, \ldots, n.$$

Let f be a linear functional. We define its coefficients with respect to the two bases as $\beta_i = f(e_i)$ and $\beta'_j = f(e'_j)$. Substituting the

expression for e'_j:

$$\begin{aligned}\beta'_j = f(e'_j) &= f\left(\sum_{i=1}^{n} a_{ij} e_i\right) \\ &= \sum_{i=1}^{n} a_{ij} f(e_i) \\ &= \sum_{i=1}^{n} a_{ij} \beta_i.\end{aligned}$$

In coordinate form, covector coefficients satisfy $\beta' = A^\top \beta$, whereas vector coordinates satisfy $\lambda' = A^{-1}\lambda$. Covectors thus transform with $A^\top$ (covariantly), contrasting with vectors, which transform with A^{-1} (contravariantly).

Remark.

This distinction leads to the terminology used in tensor calculus. Elements of V are often called **contravariant vectors** (indices upstairs) because their components transform inversely to the basis change. Elements of the dual space (linear functionals) are called **covariant vectors** or **covectors** (indices downstairs) because their components transform consistently with the basis change. In the language of tensors, a vector is a type $(0,1)$ tensor, and a linear functional is a type $(1,0)$ tensor.

Example 4.1. Standard Functionals.

- **Coordinate Space:** For $V = \mathbb{R}^n$, any linear functional f takes the form $f(x) = \sum_{i=1}^{n} a_i x_i$ for fixed scalars a_i. This can be written as the dot product $f(x) = a^\top x$ or matrix multiplication $[a]x$.
- **Trace:** On the space $M_n(\mathbb{R})$ of square matrices, the trace map $\operatorname{tr}(A) = \sum A_{ii}$ is a linear functional.
- **Integration:** On the space $C[0,1]$ of continuous functions, the definite integral $I(g) = \int_0^1 g(t)\,dt$ is a linear functional.
- **Evaluation:** On a function space such as P_n, the map $E_t(p) = p(t)$ for a fixed $t \in \mathbb{R}$ is a linear functional.

範例

4.1 *The Dual Space*

The set of all linear functionals on V can be equipped with vector space operations.

Definition 4.2. ***Dual Space.***
The set of all linear functionals on V, denoted by V^* (or sometimes V'),

forms a vector space over F with the operations:

$$(\alpha f + \beta g)(x) = \alpha f(x) + \beta g(x).$$

This space is called the **dual space** of V.

定義

Since a linear functional is determined by its values on a basis of V, there is a natural bijection between V^* and the coordinate space F^n. Specifically, fixing a basis $(e_1, \ldots, e_n)$ of V, the map $\theta : f \mapsto (f(e_1), \ldots, f(e_n))$ is an isomorphism. Consequently:

$$\dim V^* = \dim V = n.$$

We can construct a specific basis for V^* associated with a given basis of V.

Theorem 4.1. ***The Dual Basis.***
Let $(e_1, \ldots, e_n)$ be a basis of an n-dimensional vector space V. Define the linear functionals $e^1, \ldots, e^n \in V^*$ by their action on the basis vectors:

$$e^i(e_j) = \delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Then $(e^1, \ldots, e^n)$ forms a basis for V^*, called the **dual basis** to (e_i).

定理

Proof

Since $\dim V^* = n$, it suffices to show that $\{e^1, \ldots, e^n\}$ is linearly independent. Consider a linear combination equal to the zero functional:

$$\sum_{i=1}^{n} \lambda_i e^i = 0.$$

Applying this functional to the basis vector $e_k \in V$:

$$\left(\sum_{i=1}^{n} \lambda_i e^i \right)(e_k) = \sum_{i=1}^{n} \lambda_i e^i(e_k) = \sum_{i=1}^{n} \lambda_i \delta_{ik} = \lambda_k.$$

Since the functional is zero, it must evaluate to zero on all vectors. Thus $\lambda_k = 0$ for all $k = 1, \ldots, n$. Hence the set is linearly independent and forms a basis.

■

Example 4.2. Lagrange Interpolation. Consider $V = P_n(\mathbb{R})$, the space of polynomials of degree at most n (dimension $n + 1$). Let $t_0, \ldots, t_n$ be distinct real numbers. The evaluation functionals $L_i(p) = p(t_i)$ form a basis for V^*. The basis of V dual to $\{L_0, \ldots, L_n\}$ consists of the polynomials $l_0, \ldots, l_n$ such that $L_i(l_j) = \delta_{ij}$. Explicitly, $l_j(t_i) = \delta_{ij}$. These are precisely the Lagrange

interpolating polynomials:

$$l_j(t) = \prod_{k \neq j} \frac{t - t_k}{t_j - t_k}.$$

This establishes that any polynomial p can be reconstructed from its samples: $p = \sum p(t_i) l_i$.

範例

Example 4.3. Calculating the Dual Basis. Let $V = \mathbb{R}^3$ with basis $u_1 = (1, 0, -1)^\top$, $u_2 = (1, 1, 1)^\top$, and $u_3 = (2, 2, 0)^\top$. To find the dual basis f_1, f_2, f_3, we express each functional as $f_i(x) = c_i^\top x$ for some coefficient vector c_i. The condition $f_i(u_j) = \delta_{ij}$ becomes $c_i^\top u_j = \delta_{ij}$.
Arranging the basis vectors as columns of a matrix $U = [u_1 \mid u_2 \mid u_3]$ and the coefficient vectors as rows of a matrix C, we require $CU = I$. Thus $C = U^{-1}$.

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

Inverting U (e.g., by row reduction $[U \mid I] \to [I \mid U^{-1}]$) yields:

$$C = \begin{bmatrix} 1 & -1 & 0 \\ 1 & -1 & 1 \\ -0.5 & 1 & -0.5 \end{bmatrix}.$$

The rows of C give the functionals:

$$f_1(x) = x_1 - x_2, \quad f_2(x) = x_1 - x_2 + x_3, \quad f_3(x) = -\frac{1}{2}x_1 + x_2 - \frac{1}{2}x_3.$$

範例

Canonical Pairing

The relationship between V and V^* is symmetric. We introduce the notation (f, x) to denote the evaluation $f(x)$. This defines a map $V^* \times V \to F$:

$$(f, x) = f(x).$$

This map is **bilinear**, meaning it is linear in both arguments:

$$(\alpha f + \beta g, x) = \alpha(f, x) + \beta(g, x),$$
$$(f, \alpha x + \beta y) = \alpha(f, x) + \beta(f, y).$$

Such a pairing is called **canonical** because it depends only on the definition of the spaces, not on a choice of basis.

Using dual bases, we can express the pairing in terms of coordinates. Let $x = \sum \alpha_i e_i$ and $f = \sum \beta_i e^i$. Then:

$$(f, x) = \sum_{j=1}^{n} \beta_j \left(e^j, \sum_{i=1}^{n} \alpha_i e_i \right) = \sum_{j=1}^{n} \sum_{i=1}^{n} \beta_j \alpha_i \delta_{ji} = \sum_{k=1}^{n} \alpha_k \beta_k.$$

Furthermore, the coordinates themselves can be recovered via the pairing:

$$\alpha_k = (e^k, x) \quad \text{and} \quad \beta_k = (f, e_k).$$

Example 4.4. Polynomials and Derivatives. Let $V = P_n$ be the space of polynomials over $\mathbb{R}$ of degree less than n, with basis $(1, t, \ldots, t^{n-1})$. For any $\lambda \in \mathbb{R}$, the evaluation map $f_\lambda : \varphi \mapsto \varphi(\lambda)$ is a linear functional. However, a more convenient basis for V^* relates to derivatives. Define $e^k \in V^*$ by:

$$e^k(\varphi) = \frac{\varphi^{(k)}(0)}{k!}.$$

For the basis vector $e_j = t^j$, we have:

$$\frac{d^k}{dt^k}(t^j)\bigg|_{t=0} = \begin{cases} k! & \text{if } j = k, \\ 0 & \text{if } j \neq k. \end{cases}$$

Thus $e^k(t^j) = \delta_{kj}$, so $(e^0, \ldots, e^{n-1})$ is the dual basis to $(1, t, \ldots, t^{n-1})$. Generally, expanding φ in this basis corresponds to the Maclaurin series coefficients. If we instead use the basis $(1, t - \lambda, \ldots, (t - \lambda)^{n-1})$, the dual basis consists of functionals $\varphi \mapsto \varphi^{(k)}(\lambda)/k!$, corresponding to Taylor expansion at λ.

範例

The Double Dual and Reflexivity

Since V^* is a vector space, we can consider *its* dual space, $V^{**} = (V^*)^*$, called the **double dual** or **bidual**. Elements of V^{**} are linear functionals on V^*. While V^* is isomorphic to V (as they have the same dimension), constructing such an isomorphism requires choosing a basis. Remarkably, there exists a *natural* isomorphism between V and V^{**} that requires no basis choice.

Theorem 4.2. ***Reflexivity.***
Define the map $\varepsilon : V \to V^{**}$ by $\varepsilon(x) = \varepsilon_x$, where ε_x acts on $f \in V^*$ by:

$$\varepsilon_x(f) = f(x).$$

Then ε is a linear isomorphism.

定理

Proof

First, we verify linearity. For $x, y \in V$ and $\alpha, \beta \in F$, and for any $f \in V^*$:

$$\begin{aligned}\varepsilon_{\alpha x+\beta y}(f) &= f(\alpha x + \beta y) \\ &= \alpha f(x) + \beta f(y) \quad \text{(linearity of } f\text{)} \\ &= \alpha \varepsilon_x(f) + \beta \varepsilon_y(f) \\ &= (\alpha \varepsilon_x + \beta \varepsilon_y)(f).\end{aligned}$$

Thus $\varepsilon_{\alpha x+\beta y} = \alpha \varepsilon_x + \beta \varepsilon_y$.
To prove bijectivity, let (e_i) be a basis for V and (e^i) be the dual basis for V^*. We evaluate ε_{e_j} on the basis vectors of V^*:

$$\varepsilon_{e_j}(e^i) = e^i(e_j) = \delta_{ij}.$$

The functionals $(\varepsilon_{e_1}, \ldots, \varepsilon_{e_n})$ in V^{**} satisfy the condition of being the dual basis to $(e^1, \ldots, e^n)$. By *theorem* 4.1, they form a basis for V^{**}. Since ε maps a basis of V to a basis of V^{**}, it is an isomorphism.

■

Definition 4.3. ***Reflexive Space.***
A vector space V is called **reflexive** if the natural map $\varepsilon : V \to V^{**}$ is an isomorphism.

定義

Note

theorem 4.1 implies that all finite-dimensional vector spaces are reflexive. This property allows us to treat V and V^{**} as identical. The equation $x(f) = f(x)$ becomes an identity, reinforcing the symmetry of the pairing (f, x).

Annihilators and Linear Independence

The dual space provides powerful tools for characterizing linear independence and subspaces. We can associate a set of vectors in V with a determinant of values of functionals.

Definition 4.4. ***Annihilator.***
Let S be a subset of V. The **annihilator** of S, denoted S°, is the set of linear functionals in V^* that vanish on S:

$$S^\circ = \{f \in V^* \mid f(x) = 0 \text{ for all } x \in S\}.$$

It is immediate that S° is a subspace of V^*.

定義

Theorem 4.3. ***Dimension of the Annihilator.***
Let V be finite-dimensional and W be a subspace of V. Then:

$$\dim W + \dim W^\circ = \dim V.$$

定理

Proof

Let $(e_1, \ldots, e_k)$ be a basis for W. Extend this to a basis $(e_1, \ldots, e_n)$ for V. Let $(e^1, \ldots, e^n)$ be the corresponding dual basis for V^*. We claim that $(e^{k+1}, \ldots, e^n)$ is a basis for W°. First, for any $j > k$ and any $w = \sum_{i=1}^{k} \alpha_i e_i \in W$, we have $e^j(w) = \sum_{i=1}^{k} \alpha_i \delta_{ji} = 0$. Thus $e^j \in W^\circ$. Conversely, let $f = \sum_{i=1}^{n} \beta_i e^i \in W^\circ$. For any $j \leq k$, we have $0 = f(e_j) = \beta_j$. Thus $f = \sum_{i=k+1}^{n} \beta_i e^i$. This shows $W^\circ = \operatorname{span}(e^{k+1}, \ldots, e^n)$. Since these are basis elements, they are independent, so $\dim W^\circ = n - k = \dim V - \dim W$.

■

Definition 4.5. ***Hyperspace.***
A subspace of V with codimension 1 (dimension $n-1$) is called a **hyperspace** (or hyperplane).

定義

Corollary 4.1. *Kernels and Hyperspaces.* The kernel of any non-zero linear functional $f \in V^*$ is a hyperspace. Conversely, every hyperspace is the kernel of some non-zero linear functional.

推論

Proof

If $f \neq 0$, then $\operatorname{Im} f = F$ (dimension 1). By Rank-Nullity (*theorem 2.6*), $\dim \operatorname{Ker} f = \dim V - 1$. Conversely, if W is a hyperspace, $\dim W^\circ = \dim V - (n-1) = 1$. Let f be a non-zero element of W°. Then $W \subseteq \operatorname{Ker} f$. Since dimensions match, $W = \operatorname{Ker} f$.

■

Theorem 4.4. ***Dual Basis Independence Criterion.***
Let $(f_1, \ldots, f_n)$ be a basis of V^*. A set of vectors $\{a_1, \ldots, a_n\}$ in V is linearly independent if and only if the matrix $M = (f_i(a_j))$ is invertible.

定理

Proof

Define the linear map $\Phi : V \to F^n$ by evaluating the basis function-

als on a vector:

$$\Phi(v) = \begin{bmatrix} f_1(v) \\ \vdots \\ f_n(v) \end{bmatrix}.$$

We first show that Φ is an isomorphism.

Injectivity: Suppose $\Phi(v) = 0$. Then $f_i(v) = 0$ for all $i = 1, \ldots, n$. Since $\{f_i\}$ is a basis for V^*, any functional $g \in V^*$ can be expressed as $g = \sum c_i f_i$. Thus $g(v) = \sum c_i f_i(v) = 0$. Since $g(v) = 0$ for all $g \in V^*$, we must have $v = 0$. Thus $\operatorname{Ker}\Phi = \{0\}$, so Φ is injective.

Surjectivity: Since $\dim V = n$ and $\dim F^n = n$, an injective linear map between them is automatically an isomorphism.

Observe that the j-th column of the matrix M is exactly the coordinate vector $\Phi(a_j)$. Thus $M = [\Phi(a_1) \mid \cdots \mid \Phi(a_n)]$. The vectors $\{a_j\}$ are linearly independent in V if and only if their images $\{\Phi(a_j)\}$ are linearly independent in F^n (as Φ is an isomorphism). By the Invertible Matrix Theorem (*theorem* 0.5), the columns of a square matrix M are linearly independent if and only if M is invertible.

■

These results lead to a general rank criterion.

Theorem 4.5. ***Rank via Duality.***
Let $(f_1, \ldots, f_n)$ be a basis of V^*. For any set of vectors $\{a_1, \ldots, a_k\} \subseteq V$, the rank of the set of vectors is equal to the rank of the $n \times k$ matrix $M = (f_i(a_j))$.

定理

Proof

Let $\Phi : V \to F^n$ be the isomorphism defined in *theorem* 4.4. The rank of the set $\{a_1, \ldots, a_k\}$ is the dimension of their span $U = \langle a_1, \ldots, a_k \rangle$. Since Φ is an isomorphism, it preserves dimensions of subspaces:

$$\dim\langle a_1, \ldots, a_k \rangle = \dim\langle \Phi(a_1), \ldots, \Phi(a_k) \rangle.$$

The vectors $\Phi(a_j)$ are exactly the columns of the matrix M. The dimension of the span of these columns is the column rank of M, which is simply the rank of M.

■

Geometric Interpretation of Homogeneous Systems

We can reinterpret homogeneous linear systems using dual spaces. A system of m linear equations in n unknowns can be written abstractly as:

$$f_1(x) = 0, \quad \dots, \quad f_m(x) = 0,$$

where $x \in V$ and $f_i \in V^*$. The solution set is the subspace $U = \bigcap_{i=1}^{m} \operatorname{Ker} f_i$.

Theorem 4.6. ***Annihilators and Solution Spaces.***
Let $S = \{f_1, \dots, f_m\} \subseteq V^*$ be a set of functionals with rank r.

1. The subspace $U = \{x \in V \mid f_i(x) = 0 \quad \forall i\}$ has dimension $n - r$.
2. Every subspace $U \subseteq V$ of dimension k is the solution set of a system of $n - k$ independent linear equations.

定理

Proof

1. Assume without loss of generality that $f_1, \dots, f_r$ are linearly independent. They can be extended to a basis $(f_1, \dots, f_r, f_{r+1}, \dots, f_n)$ of V^*. Let $(e_1, \dots, e_n)$ be the dual basis in V. The condition $f_i(x) = 0$ for $i = 1, \dots, r$ implies that the first r coordinates of x in the basis (e_i) must be zero. Thus $x = \sum_{j=r+1}^{n} \lambda_j e_j$. The vectors $e_{r+1}, \dots, e_n$ are linearly independent and span U. Hence $\dim U = n - r$.
2. Let U be a subspace with basis $(e_1, \dots, e_k)$. Extend this to a basis $(e_1, \dots, e_n)$ of V. Let $(f_1, \dots, f_n)$ be the dual basis of V^*. A vector $x = \sum \lambda_i e_i$ lies in U if and only if $\lambda_{k+1} = \cdots = \lambda_n = 0$. Since $\lambda_j = f_j(x)$, this is equivalent to the system:

$$f_{k+1}(x) = 0, \quad \dots, \quad f_n(x) = 0.$$

These are $n - k$ linearly independent equations.

■

4.2 *The Transpose Map*

The dual space construction allows us to define the "dual" of a linear map. This is the abstract operator-theoretic origin of the matrix transpose.

Definition 4.6. ***Transpose Operator.***
Let $T : V \to W$ be a linear map. The **transpose** of T is the map $T^\top : W^* \to V^*$ defined by pre-composition:

$$(T^\top g)(v) = g(Tv) \quad \text{for all } g \in W^*, v \in V.$$

That is, $T^\top g = g \circ T$.

定義

It is straightforward to verify that $T^\top$ is a linear map. If $g, h \in W^*$, then $(T^\top(g+h))(v) = (g+h)(Tv) = g(Tv) + h(Tv) = (T^\top g)(v) + (T^\top h)(v)$.

Theorem 4.7. ***Matrix of the Transpose.***
Let V, W be finite-dimensional with bases $\mathcal{B}, \mathcal{C}$ respectively, and let $\mathcal{B}^*, \mathcal{C}^*$ be their dual bases. If A is the matrix of T relative to $\mathcal{B}, \mathcal{C}$, then the matrix of $T^\top$ relative to $\mathcal{C}^*, \mathcal{B}^*$ is the matrix transpose $A^\top$.

定理

Proof

Let $\mathcal{B} = (v_j)$ and $\mathcal{C} = (w_i)$. The matrix entries A_{ij} are defined by $Tv_j = \sum_i A_{ij} w_i$. Let $\mathcal{B}^* = (v^j)$ and $\mathcal{C}^* = (w^i)$. We compute the coordinates of $T^\top w^k$:

$$(T^\top w^k)(v_j) = w^k(Tv_j) = w^k\left(\sum_i A_{ij} w_i\right) = \sum_i A_{ij}\delta_{ki} = A_{kj}.$$

Thus $T^\top w^k = \sum_j A_{kj} v^j$. The coefficient of v^j is A_{kj}, which is the (j,k) entry of $A^\top$.

■

Theorem 4.8. ***Annihilator Relations and Rank.***
Let $T : V \to W$ be linear.
1. $\operatorname{Ker}(T^\top) = (\operatorname{Im} T)^\circ$.
2. $\operatorname{Im}(T^\top) = (\operatorname{Ker} T)^\circ$.
3. $\operatorname{rank}(T) = \operatorname{rank}(T^\top)$.

定理

Proof

1. $g \in \operatorname{Ker}(T^\top) \iff T^\top g = 0 \iff g(Tv) = 0 \,\forall v \in V \iff g(\operatorname{Im} T) = 0 \iff g \in (\operatorname{Im} T)^\circ$.
2. We use the double annihilator property. $\operatorname{Im}(T^\top)$ is a subspace of V^*.

   $$(\operatorname{Im} T^\top)^\circ = \{v \in V \mid (T^\top g)(v) = 0 \,\forall g \in W^*\} = \{v \in V \mid g(Tv) = 0 \,\forall g \in W^*\}.$$

   The only vector annihilated by all functionals is the zero vector, so $Tv = 0$, meaning $v \in \operatorname{Ker} T$. Thus $(\operatorname{Im} T^\top)^\circ = \operatorname{Ker} T$. Taking annihilators again yields $\operatorname{Im} T^\top = (\operatorname{Ker} T)^\circ$.
3. $\operatorname{rank}(T^\top) = \dim \operatorname{Im}(T^\top) = \dim(\operatorname{Ker} T)^\circ = \dim V - \dim \operatorname{Ker} T$. By Rank-Nullity (*theorem* 2.6), this equals $\operatorname{rank}(T)$.

■

Remark.

The equality $\text{rank}(T) = \text{rank}(T^\top)$ provides a conceptual proof that the row rank of a matrix equals its column rank (*theorem* 2.4). The column rank of A is $\text{rank}(T)$, and the row rank of A is the column rank of $A^\top$, which is $\text{rank}(T^\top)$.

4.3 Multilinear Maps

The concept of a linear functional, which maps a single vector to a scalar, can be generalised to functions accepting multiple vector arguments.

Definition 4.7. ***Multilinear Map.***
Let $V_1, \dots, V_p$ and U be vector spaces over a field F. A map

$$f : V_1 \times V_2 \times \cdots \times V_p \to U$$

is called p**-linear** (or **multilinear**) if it is linear in each argument independently. That is, for any fixed index i and fixed vectors $\mathbf{a}_j \in V_j$ (for $j \neq i$), the induced map

$$\mathbf{v} \mapsto f(\mathbf{a}_1, \dots, \mathbf{a}_{i-1}, \mathbf{v}, \mathbf{a}_{i+1}, \dots, \mathbf{a}_p)$$

is a linear map from V_i to U. Explicitly:

$$f(\dots, \alpha x + \beta y, \dots) = \alpha f(\dots, x, \dots) + \beta f(\dots, y, \dots).$$

The set of all such maps is denoted $\mathcal{L}(V_1, \dots, V_p; U)$.

定義

It is routine to verify that the sum of two p-linear maps and the scalar multiple of a p-linear map remain p-linear. Thus, $\mathcal{L}(V_1, \dots, V_p; U)$ forms a vector space.

Multilinear Forms

A particularly important case arises when the codomain is the underlying field itself, i.e., $V_1 = \cdots = V_p = U = F$. Such a map is called a **multilinear form** on $V_1 \times \cdots \times V_p$. The simplest example is the product map $f(v_1, \dots, v_p) = v_1 \cdots v_p$ on F^p. More generally, let $l^i \in V_i^*$ be linear functionals on V_i. We can construct a multilinear form by taking the product of their evaluations:

$$f(v_1, \dots, v_p) = l^1(v_1) l^2(v_2) \cdots l^p(v_p).$$

This specific form is called the **tensor product** of the functionals and is denoted by $l^1 \otimes l^2 \otimes \cdots \otimes l^p$.

Note

We define the tensor product here concretely as a multilinear map. This is a specific instance of the general tensor product of vector spaces.

When all domains are identical, say $V_i = V$, we denote the space of multilinear forms as $\mathcal{L}_p(V, F)$. In the language of tensors, elements of this space are tensors of type $(p, 0)$, also known as covariant tensors of order p.

Definition 4.8. ***Symmetry and Alternation.***
Let S_p denote the set of all permutations of $\{1, \ldots, p\}$, known as the **symmetric group**. For $\pi \in S_p$, let $\operatorname{sgn}(\pi)$ be the sign of the permutation ($+1$ if even, -1 if odd). A multilinear form $f \in \mathcal{L}_p(V, F)$ is called:
Symmetric if its value remains unchanged under any permutation of its arguments:

$$f(v_{\pi(1)}, \ldots, v_{\pi(p)}) = f(v_1, \ldots, v_p) \quad \text{for all } \pi \in S_p.$$

Skew-symmetric *(or* ***alternating****)* if swapping arguments introduces a sign determined by the parity of the permutation:

$$f(v_{\pi(1)}, \ldots, v_{\pi(p)}) = \operatorname{sgn}(\pi) f(v_1, \ldots, v_p).$$

定義

Example 4.5. Determinant as a Form. The determinant of a square matrix, when viewed as a function of its n column vectors, is the prototypical example of an alternating n-linear form on F^n.

範例

4.4 *Bilinear Forms*

We now restrict our attention to the case $p = 2$ with $V_1 = V_2 = V$.

Definition 4.9. ***Bilinear Form.***
A **bilinear form** on a vector space V is a map $f : V \times V \to F$ that is linear in both arguments. For all $u, v, w \in V$ and $\alpha, \beta \in F$:

$$f(\alpha u + \beta v, w) = \alpha f(u, w) + \beta f(v, w),$$
$$f(w, \alpha u + \beta v) = \alpha f(w, u) + \beta f(w, v).$$

定義

Note

In general, bilinear forms need not be commutative; that is, $f(u,v)$ is not necessarily equal to $f(v,u)$.

Matrix Representation

Let $(e_1,\dots,e_n)$ be a basis for V. Any two vectors $x,y \in V$ can be expanded in coordinates as $x = \sum x_i e_i$ and $y = \sum y_j e_j$. Using the bilinearity of f:

$$\begin{aligned} f(x,y) &= f\left(\sum_{i=1}^{n} x_i e_i, \sum_{j=1}^{n} y_j e_j\right) \\ &= \sum_{i=1}^{n} x_i f\left(e_i, \sum_{j=1}^{n} y_j e_j\right) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_j f(e_i,e_j). \end{aligned}$$

The n^2 scalars $f_{ij} = f(e_i,e_j)$ completely determine the form. We arrange these scalars into a matrix $F = (f_{ij})$. Let X and Y be the column vectors of coordinates for x and y respectively. The expression above can be written using matrix multiplication:

$$f(x,y) = \sum_{i,j} x_i f_{ij} y_j = X^{\top} F Y.$$

Proposition 4.1. ***Isomorphism with Matrices.***
Fixing a basis for V, there is a bijective linear correspondence between the space of bilinear forms $\mathcal{L}_2(V,F)$ and the space of matrices $M_n(F)$.

命题

Proof

The map $f \mapsto F = (f(e_i,e_j))$ is clearly linear. Conversely, given any matrix M, the function $g(x,y) = X^{\top} M Y$ defines a bilinear form. Since the scalars $f(e_i,e_j)$ are uniquely determined, this correspondence is an isomorphism.

■

Change of Basis and Congruence

The matrix representing a bilinear form depends on the choice of basis. We determine the transformation law for this matrix. Let $(e_1,\dots,e_n)$ be a basis for V and let $(e'_1,\dots,e'_n)$ be a new basis defined by the transition matrix $A = (a_{ij})$, such that:

$$e'_j = \sum_{i=1}^{n} a_{ij} e_i.$$

Let X and X' be the coordinate columns of a vector x in the old and new bases, respectively. The relationship between coordinates is given by $X = AX'$. Let F be the matrix of the form f in the basis (e_i), and let F' be the matrix in the basis (e'_i). We must have:

$$f(x,y) = X^\top FY = (X')^\top F'Y'.$$

Substituting the coordinate transformation $X = AX'$ and $Y = AY'$ into the left-hand side:

$$X^\top FY = (AX')^\top F(AY') = (X')^\top A^\top FAY'.$$

Comparing this with $(X')^\top F'Y'$, and noting that this holds for all X', Y', we deduce the relationship between F and F'.

Theorem 4.9. ***Transformation of Bilinear Forms.***
Let F be the matrix of a bilinear form relative to a basis $\mathcal{B}$, and let A be the transition matrix from $\mathcal{B}$ to a new basis $\mathcal{B}'$. The matrix of the form relative to $\mathcal{B}'$ is:

$$F' = A^\top FA.$$

定理

Proof

Let X', Y' be coordinate columns of x, y in the new basis $\mathcal{B}'$. Coordinates in the old basis satisfy $X = AX'$ and $Y = AY'$. Evaluating f in the old coordinates gives

$$f(x,y) = X^\top FY = (AX')^\top F(AY') = (X')^\top A^\top FAY'.$$

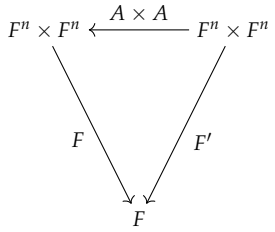
By definition, $f(x,y) = (X')^\top F'Y'$ in the new basis. Equality for all X', Y' forces $F' = A^\top FA$.

■

Definition 4.10. ***Congruence.***
Two square matrices A and B are called **congruent** if there exists an invertible matrix P such that $B = P^\top AP$.

定義

This transformation differs significantly from the similarity transformation $B = P^{-1}AP$ used for linear operators.



Change of basis for bilinear forms corresponds to matrix congruence $F' = A^\top FA$.

Figure 4.1: Commutative diagram illustrating the coordinate change.

Rank of a Bilinear Form

Since congruent matrices are related by multiplication with invertible matrices, their ranks are identical.

Corollary 4.2. *Invariance of Rank.* The rank of the matrix representing a bilinear form is independent of the basis. We define the **rank of the form** f to be the rank of any of its matrix representations.

推論

Proof

If F' and F are matrices of the same form in two bases, the transformation result above gives $F' = A^\top FA$ with A invertible. Multiplying by invertible matrices on the left or right does not change rank, so $\operatorname{rank} F' = \operatorname{rank} F$.

■

We can characterise the rank intrinsically without reference to matrices using the concept of the radical.

Definition 4.11. ***Left Radical.***
The **left radical** (or left kernel) of a bilinear form f is the set:

$$L_f = \{x \in V \mid f(x, y) = 0 \text{ for all } y \in V\}.$$

定義

It is straightforward to verify that L_f is a subspace of V.

Proposition 4.2. ***Rank-Nullity for Bilinear Forms.***
For a bilinear form f on an n-dimensional space V:

$$\operatorname{rank} f = n - \dim L_f.$$

命題

Proof

Fix a basis $(e_1, \ldots, e_n)$. A vector x belongs to L_f if and only if $f(x, e_j) = 0$ for all $j = 1, \ldots, n$. For each j, define the linear functional $f_j(x) = f(x, e_j)$. The condition $x \in L_f$ is equivalent to $x \in \bigcap_{j=1}^{n} \operatorname{Ker} f_j$. The coordinates of the functional f_j in the dual basis correspond to the j-th row of the matrix $F = (f(e_i, e_j))$. Specifically, $f_j(e_i) = f(e_i, e_j) = f_{ij}$. Let $S = \operatorname{span}(f_1, \ldots, f_n) \subseteq V^*$. The dimension of S is the row rank of F, which equals $\operatorname{rank} f$. By the theory of annihilators (specifically *theorem* 2.6), the dimension of the solution space to $f_j(x) = 0$ (which is L_f) is $\dim V - \dim S$. Thus, $\dim L_f = n - \operatorname{rank} f$.

■

4.5 *Annihilator Consequences*

The following corollaries are immediate from *theorem* 4.3.

Corollary 4.3. *Hyperplane Intersection.* If $\dim V = n$ and $\dim W = k$, then $W = \bigcap_{i=1}^{n-k} \operatorname{Ker} f_i$ for suitable non-zero functionals f_i. Each $\operatorname{Ker} f_i$ is a hyperplane.

推論

Proof

Let $(f_{k+1}, \dots, f_n)$ be the dual functionals corresponding to a basis extension as in *theorem* 4.3. Then $W = \{x \mid f_i(x) = 0 \text{ for } i = k+1, \dots, n\}$.

■

Corollary 4.4. *Equality via Annihilators.* For subspaces $W_1, W_2 \leq V$, $W_1 = W_2$ if and only if $W_1^\circ = W_2^\circ$.

推論

Proof

The forward implication is immediate. For the converse, $\dim W_1 = \dim V - \dim W_1^\circ = \dim V - \dim W_2^\circ = \dim W_2$. Also $W_1 \subseteq (W_1^\circ)^\circ = (W_2^\circ)^\circ$ (see *theorem* 4.10 below), so $W_1 \subseteq W_2$ and the dimensions force equality.

■

Example 4.6. Three Functionals in $\mathbb{R}^4$. Let $f_1(x) = x_1 + x_2 - x_3 + x_4$, $f_2(x) = x_1 - 2x_2$, $f_3(x) = 3x_2 + 2x_4$. The annihilated subspace $W = \{x \mid f_i(x) = 0,\ i = 1,2,3\}$ has basis $\{(-4,2,3,3)^\top\}$, so $\dim W = 1$ and $W^\circ = \operatorname{span}(f_1, f_2, f_3)$.

範例

Example 4.7. Annihilator in $\mathbb{R}^5$. Let $W = \operatorname{span}\{(1,1,-1,-1,1), (1,1,-1,-1,-1), (1,1,0,0,0), (0,0,0,0,2)\}$. Row reduction gives $\dim W = 3$. Functionals in W° have coordinates $(-\alpha, \alpha, -\beta, \beta, 0)$, so $W^\circ = \operatorname{span}\{x \mapsto x_2 - x_1,\ x \mapsto x_4 - x_3\}$.

範例

Theorem 4.10. ***Double Annihilator.***
For any subset $S \subseteq V$, $S^{\circ\circ} = \operatorname{span}(S)$. In particular, if $W \leq V$ then $W^{\circ\circ} = W$.

定理

Proof

The inclusion $\operatorname{span}(S) \subseteq S^{\circ\circ}$ holds because every $f \in S^\circ$ vanishes on S, hence on its span. For finite-dimensional V, *theorem* 4.3 applied to $W = \operatorname{span}(S)$ gives $\dim W^{\circ\circ} = \dim W$, so equality follows.

■

The isomorphism $V \to V^{**}$ given by $x \mapsto \varepsilon_x$ (see Reflexivity above) yields the following.

Corollary 4.5. *Representation of V^{**}.* Every $L \in V^{**}$ equals ε_x for a unique $x \in V$.

推論

Corollary 4.6. *Realising a Dual Basis.* Given any basis $(f_1, \dots, f_n)$ of V^*, there exists a unique basis $(u_1, \dots, u_n)$ of V such that $f_i(u_j) = \delta_{ij}$.

推論

Proof

Let $(\ell_1, \dots, \ell_n)$ be the dual basis of (f_i) in V^{**}. By reflexivity there are unique $u_j \in V$ with $\varepsilon_{u_j} = \ell_j$. Then $f_i(u_j) = \varepsilon_{u_j}(f_i) = \ell_j(f_i) = \delta_{ij}$, and the u_j are independent since the ℓ_j form a basis.

■

Definition 4.12. ***Right Radical.***
The **right radical** of f is

$$R_f = \{y \in V \mid f(x, y) = 0 \text{ for all } x \in V\}.$$

定義

Remark.

If f is symmetric (or alternating with a sign flip), then $L_f = R_f$. For a general bilinear form they may differ, so both radicals are relevant.

4.6 *Exercises*

1. **Trace Functional Representation.** Let $V = M_n(\mathbb{R})$. It is known that $\operatorname{tr} : V \to \mathbb{R}$ is a linear functional. Prove that any linear functional $f \in V^*$ can be uniquely represented as $f(X) = \operatorname{tr}(AX)$ for some fixed matrix $A \in M_n(\mathbb{R})$.
2. **Functionals on Polynomials.** Let P_n be the space of real polynomials of degree $< n$. Let $a(t)$ be a fixed polynomial. Determine which of the following maps $f : P_n \to \mathbb{R}$ are linear functionals:
   (a) $f(u) = \int_0^1 a(t)u(t)\,dt$.
   (b) $f(u) = \int_0^1 a(t)u(t^2)\,dt$.
   (c) $f(u) = \int_0^1 a(t)[u(t)]^2\,dt$.
   (d) $f(u) = \frac{d^3}{dt^3}[a(t)u(t)]\big|_{t=-1}$.
3. **Proportional Functionals.** Let $f, g \in V^*$. Prove that if $\operatorname{Ker} f = \operatorname{Ker} g$, then $g = \lambda f$ for some scalar λ.
4. **Coordinate Functional.** Prove that for any non-zero linear functional f on an n-dimensional space V, there exists a basis $(e_1, \dots, e_n)$

of V such that $f(\sum \alpha_i e_i) = \alpha_1$.

5. **Determining Functionals.** Let $x \in V$ be non-zero. Does the condition $f(x) = 1$ uniquely determine a functional $f \in V^*$?

6. **Map Representation.** Let $f_1, \dots, f_m \in V^*$. Define $T : V \to \mathbb{F}^m$ by $T(x) = (f_1(x), \dots, f_m(x))$. Show T is linear. Conversely, show any linear map $V \to \mathbb{F}^m$ is of this form.

7. **Polynomial Dual Basis.** Let $V = P_2(\mathbb{R})$. Define three functionals:

$$f_1(p) = \int_0^1 p(x)\,dx, \quad f_2(p) = \int_0^2 p(x)\,dx, \quad f_3(p) = \int_{-1}^1 p(x)\,dx.$$

   Show that $\{f_1, f_2, f_3\}$ is a basis for V^*. Find the basis of V to which it is dual.

8. **Evaluation Basis.** Let W be an n-dimensional subspace of the space of functions $S \to \mathbb{F}$. Show there exist points $x_1, \dots, x_n \in S$ and functions $f_1, \dots, f_n \in W$ such that $f_i(x_j) = \delta_{ij}$.

9. **Annihilator of a Sum Subspace.** Let $W \leq \mathbb{F}^n$ be the subspace of vectors with coordinate sum zero: $\sum x_i = 0$.

   (a) Describe W°. Show it consists of functionals $f(x) = c\sum x_i$.

   (b) Show that W^* can be identified with functionals on $\mathbb{F}^n$ of the form $f(x) = \sum c_i x_i$ where $\sum c_i = 0$.

10. **Concrete Annihilator.** Let $W = \operatorname{span}\{(1, 0, -1, 2), (2, 3, 1, 1)\} \subset \mathbb{R}^4$. Which functionals $f(x) = \sum c_i x_i$ belong to W°?

11. **Annihilator Algebra.** Let $W_1, W_2 \leq V$. Prove:

   (a) $(W_1 + W_2)^\circ = W_1^\circ \cap W_2^\circ$.

   (b) $(W_1 \cap W_2)^\circ = W_1^\circ + W_2^\circ$.

12. **Extension of Functionals.** Let $W \leq V$. If $f \in W^*$, prove there exists $g \in V^*$ such that $g|_W = f$.

13. **Zero Product implies Zero Factor.** Let f, g be linear functionals on a complex vector space V. If the product map $h(x) = f(x)g(x)$ is also linear, prove that either $f = 0$ or $g = 0$.

14. **Separating Vector.** Let $v_1, \dots, v_m$ be non-zero vectors in V. Prove there exists a functional $f \in V^*$ such that $f(v_i) \neq 0$ for all i.

15. **Trace Pairing Nondegeneracy.** Let $\langle X, Y \rangle = \operatorname{tr}(XY)$ on $M_n(F)$.

   (a) Show that $\langle \cdot, \cdot \rangle$ is a bilinear form.

   (b) Prove nondegeneracy: if $\langle X, Y \rangle = 0$ for all Y, then $X = 0$.

   (c) Conclude that the induced map $M_n(F) \to M_n(F)^*$, $X \mapsto (Y \mapsto \operatorname{tr}(XY))$, is an isomorphism (i.e., the trace pairing identifies $M_n(F)$ with its dual).

5
Eigenvalues and Diagonalisation

We now address the structure of linear operators on a finite-dimensional vector space V. Consistent with *Linear Operator Algebra*, calligraphic letters denote operators; fix $\mathcal{T} \in \mathcal{L}(V)$ and use A (or T when explicitly stated) for a matrix representing $\mathcal{T}$. Having established the correspondence between operators and matrices, a natural question arises: can we find a basis $\mathcal{B}$ of V such that the matrix representation of an operator is simple? The simplest non-scalar matrices are diagonal matrices.

> **Definition 5.1.** ***Diagonalisable Operator.***
> A linear operator $\mathcal{T} \in \mathcal{L}(V)$ is said to be **diagonalisable** if there exists a basis $\mathcal{B}$ of V such that the matrix of $\mathcal{T}$ relative to $\mathcal{B}$ is a diagonal matrix.
>
> 定義

Explicitly, if $\mathcal{B} = (u_1, \ldots, u_n)$ and the matrix is diagonal with entries $\lambda_1, \ldots, \lambda_n$, the action of the operator is described by:

$$\mathcal{T}u_j = \lambda_j u_j \quad \text{for } j = 1, \ldots, n.$$

This equation completely characterizes the operator.

5.1 *Eigenvalues and Eigenvectors*

The equation $Tx = \lambda x$ is central to the problem of diagonalisation.

> **Definition 5.2.** ***Eigenvalue and Eigenvector.***
> Let $\mathcal{T} \in \mathcal{L}(V)$ where V is a vector space over a field F. A scalar $\lambda \in F$ is called an **eigenvalue** (or characteristic value) of $\mathcal{T}$ if there exists a non-zero vector $x \in V$ such that:
>
> $$\mathcal{T}x = \lambda x.$$
>
> Any vector x satisfying this equation is called an **eigenvector** corresponding to λ.

定義

Note

We explicitly require $x \neq 0$. If $x = 0$, the equation $T(0) = \lambda \cdot 0$ holds for any λ, which is trivial.

Definition 5.3. ***Spectrum.***
The set of all eigenvalues of $\mathcal{T}$ is called the **spectrum** of $\mathcal{T}$ and is denoted by $\text{Spec}(\mathcal{T})$.

定義

This condition is equivalent to requiring that the kernel of the operator $\mathcal{T} - \lambda\mathcal{E}$ is non-trivial. That is,

$$(\mathcal{T} - \lambda\mathcal{E})x = 0 \quad \text{for some } x \neq 0.$$

For finite-dimensional spaces, a non-trivial kernel implies the operator is not invertible.

Example 5.1. Rotation in the Plane. Consider the rotation operator $\mathcal{T} : \mathbb{R}^2 \to \mathbb{R}^2$ which rotates vectors by 90° counter-clockwise:

$$\mathcal{T}(x_1, x_2) = (-x_2, x_1).$$

We seek a scalar $\lambda \in \mathbb{R}$ and a non-zero vector x such that $Tx = \lambda x$.

$$(-x_2, x_1) = (\lambda x_1, \lambda x_2) \implies \begin{cases} -x_2 = \lambda x_1 \\ x_1 = \lambda x_2 \end{cases}$$

Substituting the second equation into the first: $-x_2 = \lambda(\lambda x_2) = \lambda^2 x_2$, so $(1 + \lambda^2)x_2 = 0$. Since $\lambda \in \mathbb{R}$, $1 + \lambda^2 \neq 0$. Thus $x_2 = 0$, which implies $x_1 = 0$. Since the only solution is the zero vector, T has no eigenvalues over $\mathbb{R}$. The polynomial $t^2 + 1$ has no roots in $\mathbb{R}$. This highlights that the existence of eigenvalues depends on the algebraic closure of the underlying field F.

範例

Remark.

An **algebraic closure** $\overline{F}$ of a field F is a field extension in which every non-constant polynomial with coefficients in F has a root. For example, $\mathbb{C}$ is the algebraic closure of $\mathbb{R}$.

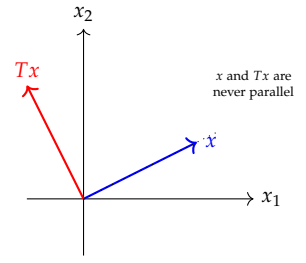


Figure 5.1: A 90° rotation has no invariant directions in $\mathbb{R}^2$.

The Characteristic Polynomial

To determine the eigenvalues, we use the determinant function (assumed known from prior matrix theory; see chapter 2 for our notation). The condition that $\mathcal{T} - \lambda I$ is not invertible is equivalent to

the vanishing of the characteristic determinant of any representing matrix A:

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

Let A be the matrix of $\mathcal{T}$ relative to some fixed basis. The eigenvalues are the roots of the equation $\det(\lambda I - A) = 0$.

Definition 5.4. ***Characteristic Polynomial.***
The **characteristic polynomial** of a matrix $A \in M_n(F)$ is defined as:

$$p_A(\lambda) = \det(\lambda I - A).$$

This is a monic polynomial of degree n.

定義

Definition 5.5. ***Algebraic Multiplicity.***
Let λ be a root of p_A. Its **algebraic multiplicity** is its multiplicity as a root of p_A.

定義

We must verify this definition is intrinsic to the operator T and not dependent on the choice of basis. Recall that if A and B represent the same operator in different bases, they are similar: $B = PAP^{-1}$ for some invertible P.

$$\begin{aligned} p_B(\lambda) &= \det(\lambda I - B) = \det(\lambda PIP^{-1} - PAP^{-1}) \\ &= \det(P(\lambda I - A)P^{-1}) \\ &= \det P \cdot \det(\lambda I - A) \cdot \det P^{-1} \\ &= \det(\lambda I - A) = p_A(\lambda). \end{aligned}$$

Thus, similar matrices have the same characteristic polynomial and the same eigenvalues. We may define $p_T(\lambda) = p_A(\lambda)$.

Example 5.2. Algebraic vs Geometric Multiplicity. Consider the matrix A representing an operator T on $\mathbb{R}^3$:

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

The characteristic polynomial is

$$p_A(\lambda) = (\lambda - 1)^2(\lambda - 2).$$

The eigenvalues are $\lambda_1 = 1$ (algebraic multiplicity 2) and $\lambda_2 = 2$ (algebraic multiplicity 1).
Eigenvectors:

1. For $\lambda = 1$: $(A - I)x = 0$ gives x_2 free, $x_1 = 0, x_3 = 0$, so the eigenspace is $\operatorname{span}\{(0, 1, 0)\}$ (geometric multiplicity 1).

2. For $\lambda = 2$: $(A - 2I)x = 0$ gives $x = (0, 0, 1)$ up to scaling (geometric multiplicity 1).

Although the eigenvalues account for dimension 3 algebraically, the eigenspaces provide only two independent eigenvectors. There is no basis of eigenvectors, so T is **not** diagonalisable.

範例

5.2 *Conditions for Diagonalisability*

The previous example demonstrates that the existence of eigenvalues is not sufficient for diagonalisability. We require enough eigenvectors to span the space.

Definition 5.6. ***Eigenspace.***
Let λ be an eigenvalue of T. The **eigenspace** corresponding to λ is the set of all eigenvectors corresponding to λ, together with the zero vector:

$$W_\lambda = \mathrm{Ker}(T - \lambda\mathcal{E}).$$

This is a subspace of V. Its dimension is called the **geometric multiplicity** of λ.

定義

Proposition 5.1. ***Polynomials of Operators.***
Let $f(t) \in F[t]$ be a polynomial. If x is an eigenvector of T with eigenvalue λ, then x is an eigenvector of the operator $f(T)$ with eigenvalue $f(\lambda)$.

命題

Proof

Let $f(t) = \sum_{k=0}^{m} a_k t^k$. Then $f(T) = \sum a_k T^k$. Since $Tx = \lambda x$, by induction $T^k x = \lambda^k x$. By linearity:

$$f(T)x = \left(\sum_{k=0}^{m} a_k T^k\right) x = \sum_{k=0}^{m} a_k (T^k x) = \sum_{k=0}^{m} a_k \lambda^k x = f(\lambda)x.$$

■

Theorem 5.1. ***Independence of Eigenvectors.***
Eigenvectors corresponding to distinct eigenvalues of $\mathcal{T}$ are linearly independent.

定理

Proof

Let $\lambda_1, \ldots, \lambda_k$ be distinct eigenvalues and $u_1, \ldots, u_k$ be correspond-

ing eigenvectors. Suppose there is a linear dependence relation:

$$\sum_{i=1}^{k} \alpha_i u_i = 0.$$

We prove all $\alpha_i = 0$. For a specific index j, construct the polynomial:

$$f_j(t) = \prod_{m \neq j} \frac{t - \lambda_m}{\lambda_j - \lambda_m}.$$

Since λ_i are distinct, the denominators are non-zero. Observe that $f_j(\lambda_m) = \delta_{jm}$ (Kronecker delta). Apply the operator $f_j(T)$ to the dependence relation:

$$f_j(T)\left(\sum_{i=1}^{k} \alpha_i u_i\right) = \sum_{i=1}^{k} \alpha_i f_j(T) u_i = \sum_{i=1}^{k} \alpha_i f_j(\lambda_i) u_i.$$

Since $f_j(\lambda_i) = 0$ for $i \neq j$ and 1 for $i = j$, the sum collapses to:

$$\alpha_j \cdot 1 \cdot u_j = 0.$$

Since $u_j \neq 0$, we must have $\alpha_j = 0$. This holds for all j, so the vectors are linearly independent.

■

This theorem implies that if $\dim V = n$ and $\mathcal{T}$ has n distinct eigenvalues, then $\mathcal{T}$ is necessarily diagonalisable (since the n corresponding eigenvectors form a basis).

Characterisation of Diagonalisability

In the general case where eigenvalues may repeat, diagonalisability is determined by the dimensions of the eigenspaces. Let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of $\mathcal{T}$ with algebraic multiplicities $d_1, \ldots, d_k$ (so the characteristic polynomial splits as $\prod(\lambda - \lambda_i)^{d_i}$). Let W_i be the eigenspace corresponding to λ_i.

Lemma 5.1. ***Sum of Eigenspaces.***
Let $W = W_1 + \cdots + W_k$. The sum is direct, i.e., $W = W_1 \oplus \cdots \oplus W_k$, and

$$\dim W = \sum_{i=1}^{k} \dim W_i.$$

引理

Proof

We must show that if $u_1 + \cdots + u_k = 0$ with $u_i \in W_i$, then each $u_i = 0$. The non-zero terms in the sum $u_1 + \cdots + u_k$ are eigenvectors corresponding to distinct eigenvalues. By *theorem* 5.1, they are

linearly independent. The only way their sum can be zero is if there are no non-zero terms. Thus $u_i = 0$ for all i.

■

Theorem 5.2. ***Diagonalisability Criterion.***
Let T be a linear operator on a finite-dimensional space V of dimension n. The following are equivalent:

1. T is diagonalisable.
2. The characteristic polynomial splits into linear factors over F, and for each eigenvalue λ_i, the geometric multiplicity equals the algebraic multiplicity:
$$\dim W_{\lambda_i} = d_i.$$
3. The sum of the eigenspaces is the whole space: $V = W_{\lambda_1} \oplus \cdots \oplus W_{\lambda_k}$.

定理

Proof

(1) $\implies$ (2): If $\mathcal{T}$ is diagonalisable, there is a basis $\mathcal{B}$ such that the matrix is diagonal with diagonal entries λ_i. The characteristic polynomial is $\prod(\lambda - \lambda_{ii})$, which splits. The number of times λ_i appears on the diagonal is d_i. The rank of $\mathcal{T} - \lambda_i\mathcal{E}$ is determined by the non-zero diagonal entries, implying the nullity (dimension of W_i) is exactly d_i.

(2) $\implies$ (3): We know $\sum d_i = n$ (degree of characteristic polynomial). If $\dim W_i = d_i$, then $\dim(W_1 \oplus \cdots \oplus W_k) = \sum d_i = n$. A subspace of dimension n in V is V itself.

(3) $\implies$ (1): If $V = \oplus W_i$, we can form a basis for V by taking the union of bases for each W_i. Since vectors in W_i are eigenvectors, this basis consists entirely of eigenvectors. Thus T is diagonalisable.

■

5.3 *The Minimal Polynomial*

We have seen that diagonalisability relies on the structural relationship between algebraic and geometric multiplicities. Another perspective involves the ideal of polynomials that annihilate $\mathcal{T}$. Recall from *definition 3.3* in *Linear Operator Algebra* that the **minimal polynomial** $\mu_{\mathcal{T}}(t)$ is the unique monic polynomial of lowest degree such that $\mu_{\mathcal{T}}(\mathcal{T}) = \mathcal{O}$. Existence and uniqueness were proved there; we use them without repetition.

Proposition 5.2. ***Properties of the Minimal Polynomial.***
Let $\mu_{\mathcal{T}}(t)$ be the minimal polynomial of $\mathcal{T}$.
1. If $f(t) \in F[t]$ satisfies $f(\mathcal{T}) = \mathcal{O}$, then $\mu_{\mathcal{T}}(t)$ divides $f(t)$.
2. The roots of $\mu_{\mathcal{T}}(t)$ are exactly the eigenvalues of $\mathcal{T}$.

命題

Proof

1. This was established in *Properties of the Minimal Polynomial*.
2. ***(λ eigenvalue $\implies \mu_{\mathcal{T}}(\lambda) = 0$):*** Let $\mathcal{T}x = \lambda x$ with $x \neq 0$. Then $\mu_{\mathcal{T}}(\mathcal{T})x = \mu_{\mathcal{T}}(\lambda)x$. Since $\mu_{\mathcal{T}}(\mathcal{T}) = \mathcal{O}$, we have $0 = \mu_{\mathcal{T}}(\lambda)x$. As $x \neq 0$, $\mu_{\mathcal{T}}(\lambda) = 0$.

   ($\mu_{\mathcal{T}}(\lambda) = 0 \implies \lambda$ eigenvalue): Write $\mu_{\mathcal{T}}(t) = (t - \lambda)q(t)$. Since $\deg q < \deg \mu_{\mathcal{T}}$, $q(\mathcal{T}) \neq \mathcal{O}$. Thus there exists $x \neq 0$ such that $y = q(\mathcal{T})x \neq 0$. Then $(\mathcal{T} - \lambda\mathcal{E})y = (\mathcal{T} - \lambda\mathcal{E})q(\mathcal{T})x = \mu_{\mathcal{T}}(\mathcal{T})x = 0$. Thus $\mathcal{T}y = \lambda y$, so λ is an eigenvalue.

■

Example 5.3. Minimal Polynomial Computations.
· Let

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

The characteristic polynomial is $p(t) = (t-1)^2$. Since $A - I \neq 0$, the minimal polynomial cannot be $t - 1$. Thus $\mu_A(t) = (t-1)^2$.

· Let

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

Direct computation gives $A^3 = 4A$ while $A^2 \neq 2A$ and $A^2 \neq -2A$ (indeed $A^2 + 2A$ has all entries 2). Hence $t^3 - 4t = t(t-2)(t+2)$ annihilates A, and no quadratic factor does. Thus $\mu_A(t) = t(t-2)(t+2)$. Since the minimal polynomial splits into distinct linear factors, A is diagonalisable.

範例

Before proving the main theorem, we recall a property of matrices related to the determinant. For any square matrix B, there exists a unique matrix called the **adjugate** of B, denoted $\operatorname{adj}(B)$, such that:

$$B \operatorname{adj}(B) = \operatorname{adj}(B) B = \det(B) I.$$

The entries of $\operatorname{adj}(B)$ are the cofactors of B; if the entries of B are polynomials in a variable t, then the entries of $\operatorname{adj}(B)$ are polynomials of degree one less than those of B.

Theorem 5.3. ***Cayley-Hamilton.***
Every linear operator satisfies its own characteristic polynomial. That is, if $p(t)$ is the characteristic polynomial of $\mathcal{T}$, then $p(\mathcal{T}) = \mathcal{O}$.

定理

Proof

Let A be the matrix of $\mathcal{T}$ in some basis. We consider the matrix characteristic polynomial $p(\lambda) = \det(\lambda I - A)$. Let $B(\lambda) = \lambda I - A$. The entries of $B(\lambda)$ are polynomials in λ of degree at most 1. Consequently, the entries of the adjugate matrix $\operatorname{adj}(B(\lambda))$ are polynomials in λ of degree at most $n-1$. We can thus write:

$$\operatorname{adj}(\lambda I - A) = C_{n-1}\lambda^{n-1} + C_{n-2}\lambda^{n-2} + \cdots + C_1\lambda + C_0,$$

where each C_k is a scalar matrix (independent of λ). We use the adjugate identity:

$$(\lambda I - A)\operatorname{adj}(\lambda I - A) = \det(\lambda I - A)I = p(\lambda)I.$$

Let $p(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_0$. Substituting the expressions:

$$(\lambda I - A)\sum_{k=0}^{n-1} C_k\lambda^k = (\lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_0)I.$$

Expanding the left side:

$$\sum_{k=0}^{n-1} C_k\lambda^{k+1} - \sum_{k=0}^{n-1} AC_k\lambda^k = \sum_{k=0}^{n-1} C_k\lambda^{k+1} - \sum_{j=0}^{n-1} AC_j\lambda^j.$$

Comparing the coefficients of like powers of λ:

$$\begin{aligned}
\text{Coeff of } \lambda^n: &\quad C_{n-1} = I \\
\text{Coeff of } \lambda^{n-1}: &\quad C_{n-2} - AC_{n-1} = a_{n-1}I \\
&\vdots \quad \vdots \\
\text{Coeff of } \lambda^k: &\quad C_{k-1} - AC_k = a_kI \\
&\vdots \quad \vdots \\
\text{Coeff of } \lambda^0: &\quad -AC_0 = a_0I
\end{aligned}$$

To form $p(A)$, we multiply the equation for λ^k by A^k on the left and sum them up:

$$\begin{aligned}
A^n(C_{n-1}) &= A^n \\
A^{n-1}(C_{n-2} - AC_{n-1}) &= a_{n-1}A^{n-1} \\
&\vdots \\
A^k(C_{k-1} - AC_k) &= a_kA^k \\
&\vdots \\
I(-AC_0) &= a_0I
\end{aligned}$$

Summing the left-hand sides creates a telescoping sum:

$$A^n C_{n-1} + \sum_{k=1}^{n-1} (A^k C_{k-1} - A^{k+1} C_k) - AC_0 = A^n C_{n-1} - A^n C_{n-1} + \cdots = 0.$$

Thus, the sum of the right-hand sides is also zero:

$$A^n + a_{n-1} A^{n-1} + \cdots + a_0 I = p(A) = 0.$$

■

Corollary 5.1. *Divisibility.* The minimal polynomial $\mu_T(t)$ divides the characteristic polynomial $p_T(t)$. Since they share the same roots, $\mu_T(t)$ contains every irreducible factor of $p_T(t)$ at least once.

推論

Proof

By Cayley-Hamilton, $p_T(T) = \mathcal{O}$. Property (1) of *proposition 5.2* then forces $\mu_T \mid p_T$.

■

Theorem 5.4. ***Diagonalisability and Minimal Polynomial.*** A linear operator T is diagonalisable if and only if its minimal polynomial $\mu_T(t)$ splits into distinct linear factors over F.

定理

Proof

($\Longrightarrow$) Suppose $\mathcal{T}$ is diagonalisable with distinct eigenvalues $\lambda_1, \ldots, \lambda_k$. Let $m(t) = \prod_{i=1}^{k} (t - \lambda_i)$. Since $\mathcal{T}$ is diagonalisable, there is a basis of eigenvectors v. For any eigenvector v_j with value λ_j, $m(\mathcal{T}) v_j = m(\lambda_j) v_j = 0$. Since $m(\mathcal{T})$ annihilates a basis, $m(\mathcal{T}) = \mathcal{O}$. Thus $\mu_{\mathcal{T}}(t)$ divides $m(t)$. Since roots of $\mu_{\mathcal{T}}$ are exactly the eigenvalues, $\mu_{\mathcal{T}}(t)$ must be exactly $m(t)$, which is a product of distinct linear factors.

($\Longleftarrow$) Write $\mu_{\mathcal{T}}(t) = \prod_{i=1}^{k} (t - \lambda_i)$ with pairwise distinct roots. Set $g_i(t) = \prod_{m \neq i} (t - \lambda_m)$. There exist polynomials $a_i(t)$ such that $\sum_{i=1}^{k} a_i(t) g_i(t) = 1$. Define $E_i = a_i(\mathcal{T}) g_i(\mathcal{T})$. Then $\sum E_i = \mathcal{E}$ and $E_i E_j = 0$ for $i \neq j$ by commutativity of polynomials in $\mathcal{T}$. Moreover, $(\mathcal{T} - \lambda_i \mathcal{E}) E_i = 0$, so $\operatorname{Im} E_i \subseteq \operatorname{Ker}(\mathcal{T} - \lambda_i \mathcal{E})$. If $x \in \operatorname{Ker}(\mathcal{T} - \lambda_i \mathcal{E})$, then $g_i(\mathcal{T}) x = g_i(\lambda_i) x = 0$ and hence $E_i x = x$. Thus $\operatorname{Im} E_i = \operatorname{Ker}(\mathcal{T} - \lambda_i \mathcal{E})$. If $x \in \operatorname{Ker}(\mathcal{T} - \lambda_i \mathcal{E})$, then $g_i(\mathcal{T}) x = g_i(\lambda_i) x = 0$ and hence $E_i x = x$. Thus $\operatorname{Im} E_i = \operatorname{Ker}(\mathcal{T} - \lambda_i \mathcal{E})$. If $\sum u_i = 0$ with $u_i \in \operatorname{Im} E_i$, applying E_j yields $u_j = 0$, so the images are independent. Therefore $V = \bigoplus_i \operatorname{Ker}(\mathcal{T} - \lambda_i \mathcal{E})$, which furnishes a basis of eigenvectors. Hence $\mathcal{T}$ is diagonalisable.

■

5.4 *Invariant Subspaces*

To delve deeper into the structure of operators, we investigate subspaces that are preserved by the operator.

Definition 5.7. ***Invariant Subspace.***
A subspace W of V is called **invariant** under an operator $\mathcal{T}$ if $\mathcal{T}(W) \subseteq W$. That is, for all $w \in W$, $\mathcal{T}w \in W$.

定義

Example 5.4. Basic Examples.
- The zero subspace $\{0\}$ and the entire space V are always invariant.
- The kernel $\operatorname{Ker}\mathcal{T}$ and the image $\operatorname{Im}\mathcal{T}$ are invariant.
- Any eigenspace $W_\lambda = \operatorname{Ker}(\mathcal{T} - \lambda\mathcal{E})$ is invariant.

範例

Example 5.5. Differentiation Chain. Let $\mathcal{D}$ be the differentiation operator on P_n (polynomials of degree at most n). Let $W_k = P_k$ be the subspace of polynomials of degree at most k. Since differentiating a polynomial lowers its degree, $\mathcal{D}(P_k) \subseteq P_{k-1} \subseteq P_k$. Thus, we have a complete chain (or flag) of invariant subspaces:

$$\{0\} \subset P_0 \subset P_1 \subset \cdots \subset P_n.$$

範例

The last example can be generalised. If S is an operator that commutes with $\mathcal{T}$ (i.e., $S\mathcal{T} = \mathcal{T}S$), then $\operatorname{Ker} S$ and $\operatorname{Im} S$ are invariant under $\mathcal{T}$. Since $\mathcal{T} - \lambda\mathcal{E}$ commutes with $\mathcal{T}$, eigenspaces are invariant. Invariant subspaces allow us to break down the operator into smaller components. If $V = W_1 \oplus W_2$ where W_1 and W_2 are invariant under $\mathcal{T}$, we can study the restrictions $\mathcal{T}|_{W_1}$ and $\mathcal{T}|_{W_2}$ independently.

Example 5.6. Non-Existence of Invariant Subspaces. Consider again the rotation by 90° in $\mathbb{R}^2$, defined by $\mathcal{T}(x_1, x_2) = (-x_2, x_1)$. If W is a non-trivial proper invariant subspace, it must be 1-dimensional (a line through the origin). Let $W = \operatorname{span}(v)$ for some $v \neq 0$. If W is invariant, then $\mathcal{T}v \in W$, so $\mathcal{T}v = \lambda v$ for some scalar λ. This implies v is an eigenvector. However, we established in *figure* 5.1 that $\mathcal{T}$ has no eigenvalues in $\mathbb{R}$. Thus, this operator has no proper non-trivial invariant subspaces.

範例

The Adjoint Operator

Recall from *Dual Spaces* that for any linear map $\mathcal{T} : V \to V$, the **transpose** (or **adjoint**) map $\mathcal{T}^\top : V^* \to V^*$ is defined by $(\mathcal{T}^\top f)(v) = f(\mathcal{T}v)$. It satisfies:

- $(\mathcal{T}+\mathcal{S})^\top = \mathcal{T}^\top + \mathcal{S}^\top$ and $(\alpha\mathcal{T})^\top = \alpha\mathcal{T}^\top$.
- $(\mathcal{T}\mathcal{S})^\top = \mathcal{S}^\top\mathcal{T}^\top$.
- $\mathcal{T}^{\top\top} = \mathcal{T}$ (under the canonical identification $V \cong V^{**}$).
- If A is the matrix of $\mathcal{T}$ in basis $\mathcal{B}$, then $A^\top$ is the matrix of $\mathcal{T}^\top$ in the dual basis $\mathcal{B}^*$.

Quotient Operators

Let W be an invariant subspace of $\mathcal{T}$. The operator $\mathcal{T}$ induces a natural linear operator on the quotient space V/W.

Definition 5.8. ***Quotient Operator.***
The **quotient operator** $\bar{\mathcal{T}} : V/W \to V/W$ is defined by:

$$\bar{\mathcal{T}}(v+W) = \mathcal{T}v + W.$$

This is well-defined because if $v + W = v' + W$, then $v - v' \in W$. Since W is invariant, $\mathcal{T}(v-v') \in W$, so $\mathcal{T}v - \mathcal{T}v' \in W$, implying $\mathcal{T}v + W = \mathcal{T}v' + W$.

定義

If we choose a basis $(e_1, \ldots, e_k)$ for W and extend it to a basis $(e_1, \ldots, e_n)$ for V, the matrix of $\mathcal{T}$ is block upper triangular:

$$A = \begin{bmatrix} A_W & B \\ 0 & A_{V/W} \end{bmatrix}, \quad \text{where } A_{V/W} \text{ represents } \bar{\mathcal{T}}.$$

Existence of Invariant Subspaces

The existence of eigenvalues (and thus 1-dimensional invariant subspaces) depends on the field.

Theorem 5.5. ***Invariant Subspaces over*** $\mathbb{R}$ ***and*** $\mathbb{C}$**.**
Let $\mathcal{T}$ be a linear operator on a finite-dimensional space V.

1. If $F = \mathbb{C}$, $\mathcal{T}$ has a 1-dimensional invariant subspace (an eigenspace).
2. If $F = \mathbb{R}$, $\mathcal{T}$ has an invariant subspace of dimension 1 or 2.

定理

Proof

1. Over $\mathbb{C}$, the characteristic polynomial splits into linear factors. Thus there is at least one root λ, providing an eigenvector and a 1-dimensional invariant subspace $\operatorname{span}(v)$.
2. Over $\mathbb{R}$, the minimal polynomial factors into linear terms $(t-\lambda)$

and irreducible quadratic terms $(t^2 - \alpha t - \beta)$ with $\alpha^2 + 4\beta < 0$. If there is a linear factor, we get an eigenvector (dim 1). If not, $\mu_{\mathcal{T}}(t)$ has a factor $q(t) = t^2 - \alpha t - \beta$. There exists u such that $q(\mathcal{T})u = 0$ but $u \neq 0$ (since $q(\mathcal{T})$ is not invertible on the kernel of the full minimal polynomial). Let $W = \text{span}(u, \mathcal{T}u)$. Since $\mathcal{T}^2 u = \alpha \mathcal{T}u + \beta u$, W is invariant. Its dimension is at most 2 (and at least 1 since $u \neq 0$).

■

Theorem 5.6. ***Invariant Hyperplanes.***
Every linear operator $\mathcal{T}$ on a finite-dimensional complex vector space has an invariant hyperplane (subspace of codimension 1).

定理

Proof

Consider the transpose operator $\mathcal{T}^\top : V^* \to V^*$. Since V^* is a complex vector space, $\mathcal{T}^\top$ has an eigenvector $f \in V^*$, so $\mathcal{T}^\top f = \lambda f$. Let $W = \text{Ker}\, f$. Since $f \neq 0$, W is a hyperplane. For any $x \in W$:

$$f(\mathcal{T}x) = (\mathcal{T}^\top f)(x) = (\lambda f)(x) = \lambda f(x) = 0.$$

Thus $\mathcal{T}x \in \text{Ker}\, f = W$, so W is invariant.

■

The T-Conductor

We introduce a set of polynomials associated with a vector and a subspace, generalising the minimal polynomial.

Definition 5.9. ***T-Conductor.***
Let W be an invariant subspace of $\mathcal{T}$ and let $y \in V$. The **$\mathcal{T}$-conductor** of y into W, denoted $S_{\mathcal{T}}(y, W)$, is the set of all polynomials $g(t) \in F[t]$ such that $g(\mathcal{T})y \in W$.

定義

The set $S_{\mathcal{T}}(y, W)$ is closed under addition and multiplication by any polynomial:

1. If g_1, g_2 are in the set, then $(g_1 + g_2)(\mathcal{T})y = g_1(\mathcal{T})y + g_2(\mathcal{T})y \in W$ (since W is a subspace).
2. If g is in the set and $f \in F[t]$, then $(fg)(\mathcal{T})y = f(\mathcal{T})(g(\mathcal{T})y)$. Since $g(\mathcal{T})y \in W$ and W is invariant under $f(\mathcal{T})$ (polynomials in $\mathcal{T}$ preserve invariant subspaces), the product is in W.

Among all non-zero polynomials in this set choose one of minimal degree and scale it to be monic; this generator is unique and is called the $\mathcal{T}$-conductor of y into W, denoted $g(t)$. Note that the minimal polynomial $\mu_{\mathcal{T}}(t)$ is always in this set (since $\mu_{\mathcal{T}}(\mathcal{T})y = 0 \in W$), so

$g(t)$ divides $\mu_{\mathcal{T}}(t)$.

Lemma 5.2. ***Linear Conductor Existence.***
Suppose the minimal polynomial $\mu_{\mathcal{T}}(t)$ splits into linear factors over F. Let W be a proper invariant subspace of $\mathcal{T}$. Then there exists a vector $x \in V \setminus W$ such that $(\mathcal{T} - \lambda I)x \in W$ for some eigenvalue λ.

引理

Proof

Pick any $y \in V \setminus W$. Let $g(t)$ be the $\mathcal{T}$-conductor of y into W. Since $y \notin W$, $g(t)$ is not a constant (otherwise $1 \cdot y \in W$). Since $g(t)$ divides $\mu_{\mathcal{T}}(t)$, and $\mu_{\mathcal{T}}(t)$ is a product of linear factors, $g(t)$ must have a linear factor. Write $g(t) = (t - \lambda)h(t)$. Since $\deg h < \deg g$, the vector $x = h(\mathcal{T})y$ is not in W (by minimality of the conductor). However,

$$(\mathcal{T} - \lambda\mathcal{E})x = (\mathcal{T} - \lambda\mathcal{E})h(\mathcal{T})y = g(\mathcal{T})y \in W.$$

Thus x satisfies the condition.

■

Triangulability

While not every operator is diagonalisable, a weaker form of simplification is almost always possible.

Definition 5.10. ***Triangulable Operator.***
An operator $\mathcal{T}$ is **triangulable** if there exists a basis $\mathcal{B}$ such that a representing matrix of $\mathcal{T}$ is upper triangular.

定義

Theorem 5.7. ***Triangulability Criterion.***
An operator $\mathcal{T}$ is triangulable if and only if its minimal polynomial (or equivalently, its characteristic polynomial) splits into linear factors over F.

定理

($\Longrightarrow$)

If $\mathcal{T}$ is triangulable, its matrix A is upper triangular. The characteristic polynomial is $\prod(t - a_{ii})$, which clearly splits. The minimal polynomial divides this, so it also splits.

証明終

($\Longleftarrow$)

We proceed by induction on $\dim V = n$. For $n = 1$, every matrix is triangular. Assume the result for dimension $n - 1$.
Let $W_0 = \{0\}$. Using the Linear Conductor Lemma, there exists

$x_1 \neq 0$ such that $(\mathcal{T} - \lambda_1 \mathcal{E})x_1 \in \{0\}$, i.e., $\mathcal{T}x_1 = \lambda_1 x_1$. Let $W_1 = \text{span}(x_1)$. This is an invariant subspace.
We repeat the process. Consider the quotient space V/W_1. The operator $\mathcal{T}$ induces an operator $\bar{\mathcal{T}}$ on the quotient. The minimal polynomial of $\bar{\mathcal{T}}$ divides $\mu_{\mathcal{T}}(t)$, so it splits. By the inductive hypothesis, there is a basis $(\bar{v}_2, \ldots, \bar{v}_n)$ for V/W_1 that triangulates $\bar{\mathcal{T}}$. Lifting these vectors back to V (and including x_1) gives a basis $(x_1, v_2, \ldots, v_n)$ in which a matrix of $\mathcal{T}$ is upper triangular.

証明終

Corollary 5.2. *Algebraically Closed Fields.* If F is **algebraically closed** (meaning every non-constant polynomial in $F[t]$ has a root in F, e.g., $\mathbb{C}$), every linear operator is triangulable.

推論

Proof

Over an algebraically closed field, every characteristic polynomial splits into linear factors. By the Triangulability Criterion, this implies the operator is triangulable.

■

5.5 *Direct Sum Decompositions and Projections*

We have characterised diagonalisability using the eigenspace decomposition $V = \bigoplus W_{\lambda_i}$. This is a specific instance of a more general structure: the decomposition of a vector space into independent subspaces, and its relationship with projection operators.

Definition 5.11. ***Independent Subspaces.***
A collection of subspaces $W_1, \ldots, W_k$ of V is called **independent** if the equation

$$w_1 + \cdots + w_k = 0 \quad (\text{with } w_i \in W_i)$$

implies that $w_i = 0$ for all i. Equivalently, every vector in the sum $W = W_1 + \cdots + W_k$ has a *unique* representation as a sum of vectors from the W_i.

定義

Recall that if $W_1, \ldots, W_k$ are independent, their sum is denoted by the direct sum symbol: $W = W_1 \oplus \cdots \oplus W_k$. If $W = V$, we say we have a direct sum decomposition of V.

Projections

A linear operator $E \in \mathcal{L}(V)$ is called a **projection** (or idempotent) if $E^2 = E$. Geometrically, E maps any vector onto the range of E by

"projecting" it, and applying it again changes nothing since the vector is already in the range.

Proposition 5.3. ***Properties of Projections.***
Let E be a projection.

1. $V = \operatorname{Im} E \oplus \operatorname{Ker} E$.
2. If $x \in \operatorname{Im} E$, then $Ex = x$.
3. The operator $\mathcal{E} - E$ is also a projection, with $\operatorname{Im}(\mathcal{E} - E) = \operatorname{Ker} E$ and $\operatorname{Ker}(\mathcal{E} - E) = \operatorname{Im} E$.
4. In a basis adapted to the decomposition $V = \operatorname{Im} E \oplus \operatorname{Ker} E$, the matrix of E is:
$$\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}, \quad \text{where } r = \operatorname{rank} E.$$
5. $\operatorname{tr} E = \operatorname{rank} E$.

命題

Proof

For any $v \in V$, we can write $v = Ev + (v - Ev)$. Clearly $Ev \in \operatorname{Im} E$. Also $E(v - Ev) = Ev - E^2v = Ev - Ev = 0$, so $v - Ev \in \operatorname{Ker} E$. To check independence, let $x \in \operatorname{Im} E \cap \operatorname{Ker} E$. Then $x = Ey$ for some y, and $Ex = 0$. Thus $0 = Ex = E(Ey) = E^2y = Ey = x$. The intersection is trivial, so the sum is direct. The matrix form and trace property follow immediately from choosing a basis $(e_1, \dots, e_r)$ for $\operatorname{Im} E$ and $(e_{r+1}, \dots, e_n)$ for $\operatorname{Ker} E$.

■

This concept generalises to multiple subspaces.

Theorem 5.8. ***Decomposition via Projections.***
Let V be a finite-dimensional vector space.

1. If $V = W_1 \oplus \cdots \oplus W_k$, there exist k linear operators $E_1, \dots, E_k$ such that:
   - $\mathcal{E} = E_1 + \cdots + E_k$.
   - $E_iE_j = \delta_{ij}E_i$ (so $E_i^2 = E_i$ and $E_iE_j = \mathcal{O}$ for $i \neq j$).
   - $\operatorname{Im} E_i = W_i$.
2. Conversely, if there exist operators $E_1, \dots, E_k$ satisfying the first two conditions, then V is the direct sum of their ranges $W_i = \operatorname{Im} E_i$.

定理

Proof

1. For any $v \in V$, write $v = w_1 + \cdots + w_k$ uniquely. Define $E_iv = w_i$. The properties follow directly from the uniqueness of the decomposition.
2. Given such operators, for any v, $v = \mathcal{E}v = \sum E_iv$, so $V = \sum W_i$. If $\sum w_i = 0$ with $w_i \in W_i$, apply E_j. Since $w_i \in \operatorname{Im} E_i$, $E_iw_i = w_i$ (as $E_i^2 = E_i$). Thus $E_j(\sum w_i) = \sum E_jw_i = \sum_{i \neq j} E_jE_iw_i' + E_jw_j =$

w_j. Since the sum is zero, $w_j = E_j(0) = 0$. Thus the subspaces are independent.

■

Relationship with Operators

Now we connect this to a linear operator T. We are interested in decompositions where each subspace W_i is invariant under T.

Proposition 5.4. ***Commutativity and Invariance.***
Let $V = \bigoplus W_i$ with associated projections E_i. The subspaces W_i are invariant under T if and only if T commutes with each projection E_i, i.e., $TE_i = E_iT$.

命題

Proof

If $\mathcal{T}E_i = E_i\mathcal{T}$, let $w \in W_i$. Then $w = E_iw$, so $\mathcal{T}w = \mathcal{T}E_iw = E_i(\mathcal{T}w)$. Thus $\mathcal{T}w \in \operatorname{Im} E_i = W_i$. Conversely, if W_i is invariant, for any v decomposed as $\sum w_j$, $\mathcal{T}v = \sum \mathcal{T}w_j$. Since $\mathcal{T}w_j \in W_j$, the i-th component of $\mathcal{T}v$ is $\mathcal{T}w_i$. Thus $E_i\mathcal{T}v = \mathcal{T}w_i$. Also $\mathcal{T}E_iv = \mathcal{T}w_i$. Hence $\mathcal{T}E_i = E_i\mathcal{T}$.

■

This leads to the spectral decomposition theorem for diagonalisable operators.

Theorem 5.9. ***Spectral Resolution.***
An operator T is diagonalisable with distinct eigenvalues $\lambda_1, \ldots, \lambda_k$ if and only if there exist non-zero projections $E_1, \ldots, E_k$ such that:
1. $\mathcal{E} = \sum E_i$.
2. $E_iE_j = 0$ for $i \neq j$.
3. $T = \sum_{i=1}^{k} \lambda_i E_i$.

In this case, $\operatorname{Im} E_i$ is exactly the eigenspace W_{λ_i}.

定理

($\Longrightarrow$)

Suppose T is diagonalisable with distinct eigenvalues. Let W_{λ_i} be the eigenspace and let E_i be the projection onto W_{λ_i} along the direct sum $\bigoplus_{m \neq i} W_{\lambda_m}$ (existence by *theorem 5.8*). Then $\mathcal{E} = \sum E_i$ and $E_iE_j = 0$ for $i \neq j$. For $x = \sum x_i$ with $x_i \in W_{\lambda_i}$, $Tx = \sum \lambda_i x_i = \sum \lambda_i E_i x$, so $T = \sum \lambda_i E_i$ and $\operatorname{Im} E_i = W_{\lambda_i}$.

証明終

($\Longleftarrow$)

Assume projections E_i satisfy the three conditions. For any x, write $x = \sum E_i x$ (since $\mathcal{E} = \sum E_i$). Then $Tx = \sum \lambda_i E_i x$. Thus $E_i x$ is an eigenvector with eigenvalue λ_i, and x is a sum of eigenvectors

from distinct eigenspaces. Independence of the ranges follows from $E_iE_j = 0$, so $V = \bigoplus \operatorname{Im} E_i$, giving a basis of eigenvectors and therefore diagonalisability.

証明終

This decomposition $T = \sum \lambda_i E_i$ allows us to define functions of operators easily: $f(T) = \sum f(\lambda_i)E_i$.

Primary Decomposition Theorem

Finally, we state the general decomposition theorem for operators whose minimal polynomial splits (or in general using irreducible factors). Recall the minimal polynomial $\mu_T(t) = \prod_{i=1}^{k} p_i(t)^{r_i}$, where $p_i(t)$ are distinct monic irreducible polynomials (in an algebraically closed field, $p_i(t) = t - \lambda_i$).

Theorem 5.10. ***Primary Decomposition.***
Let T be a linear operator on V with minimal polynomial $\mu_T(t) = p_1(t)^{r_1} \cdots p_k(t)^{r_k}$. Let $W_i = \operatorname{Ker}(p_i(T)^{r_i})$. Then:

1. $V = W_1 \oplus \cdots \oplus W_k$.
2. Each W_i is invariant under T.
3. Let $T_i = T|_{W_i}$. The minimal polynomial of T_i is $p_i(t)^{r_i}$.

定理

This theorem reduces the study of a general operator to the study of operators whose minimal polynomial is a power of an irreducible polynomial. In the case where the field is algebraically closed, each block corresponds to a single eigenvalue λ_i, and the operator on W_i is of the form $\lambda_i I + N_i$ where N_i is nilpotent. This leads directly to the Jordan Canonical Form.

5.6 *Exercises*

1. **Basic Computations.** For each of the following matrices, find the characteristic polynomial, the eigenvalues, and a basis for each eigenspace.

   (a)
   $$A = \begin{bmatrix} 2 & 0 & 0 \\ -16 & 8 & 7 \\ 0 & 0 & 1 \end{bmatrix} \text{ over } \mathbb{R}.$$

   (b)
   $$B = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \text{ over } \mathbb{C}.$$

(c)

$$C = \begin{bmatrix} 3 & 1 \\ -1 & 1 \end{bmatrix} \text{ over } \mathbb{R}.$$

2. **Trivial Operators.** Determine the characteristic and minimal polynomials for the identity operator $\mathcal{E}$ and the zero operator 0 on an n-dimensional space V.

3. **Triangular Matrices.** Prove that the eigenvalues of an upper triangular matrix are exactly its diagonal entries.

4. **Differentiation Operator.** Let $V = P_n(\mathbb{R})$ be the space of polynomials of degree at most n. Let $D : V \to V$ be the differentiation operator. Find the minimal polynomial of D.

5. **Projection Operator.** Let $P : \mathbb{R}^2 \to \mathbb{R}^2$ be the projection onto the x-axis along the y-axis ($P(x, y) = (x, 0)$). Find the minimal polynomial of P.

6. **Nilpotent Growth.** Let T be an operator on an n-dimensional space V. Prove that if $T^k = 0$ for some k, then $T^n = 0$. What is the characteristic polynomial of a nilpotent operator?

7. **Restriction Property.** Let W be a T-invariant subspace. Prove that the minimal polynomial of the restriction $T|_W$ divides the minimal polynomial of T.

8. **Idempotent Matrices.** Let A be an $n \times n$ matrix such that $A^2 = A$ (a projection).
   (a) Prove that A is similar to a diagonal matrix with entries o and 1.
   (b) Prove that $\operatorname{rank}(A) + \operatorname{rank}(I - A) = n$.

9. **Primary Decomposition Detail.** Let $p_T(x) = \prod (x - c_i)^{d_i}$ and $\mu_T(x) = \prod (x - c_i)^{r_i}$. Let $W_i = \operatorname{Ker}((T - c_i\mathcal{E})^{r_i})$.
   (a) Prove that $W_i = \{v \in V \mid (T - c_i\mathcal{E})^m v = 0 \text{ for some } m \geq 1\}$.
   (b) Prove that $\dim W_i = d_i$.

   For (b): The restriction $T|_{W_i} - c_i\mathcal{E}$ is nilpotent.

10. **Orthogonal System of Idempotents.** Let $\{A_1, \ldots, A_{m-1}\}$ be a set of matrices such that $A_i^2 = A_i$ and $A_i A_j = 0$ for $i \neq j$. Let $A = \sum A_i$. Prove that $A^2 = A$ and $AA_i = A_i A = A_i$. If we define $A_m = I - A$, prove that $\{A_1, \ldots, A_m\}$ is a complete orthogonal system of idempotents (i.e., sum to I).

11. **Multiplicative Maps on Matrices.** Let $D : M_n(\mathbb{R}) \to M_n(\mathbb{R})$ be a non-zero linear map satisfying $D(AB) = D(A)D(B)$. Prove there exists an invertible matrix C such that $D(X) = C^{-1}XC$.

12. **Stabilisation of Image.** Let $T : V \to V$ be linear. Suppose $\operatorname{Im} T^p = \operatorname{Im} T^{p+1}$ for some p. Prove that $V = \operatorname{Ker} T^p \oplus \operatorname{Im} T^p$.

13. **Cyclic Vector.** Let V be n-dimensional. Prove that if the operators $\mathcal{E}, T, T^2, \dots, T^{n-1}$ are linearly independent, then there exists a vector v such that $\{v, Tv, \dots, T^{n-1}v\}$ is a basis for V.

14. **Commuting Anti-Involutions.** Let A be a real $n \times n$ matrix with no real eigenvalues. Prove there exists a real matrix B such that $AB = BA$ and $B^2 = -I$. (This implies n is even and A defines a complex structure).

15. **Characteristic Polynomial of Products.** Prove that for any $A, B \in M_n(\mathbb{R})$, the matrices AB and BA have the same characteristic polynomial.

16. **Circulant Eigenvalues.** Find the eigenvalues of the circulant matrix
$$A = \begin{bmatrix} a_0 & a_1 & a_2 \\ a_2 & a_0 & a_1 \\ a_1 & a_2 & a_0 \end{bmatrix}$$
using the relation $A = a_0 I + a_1 P + a_2 P^2$, where $P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$.

17. **Semi-Magic Algebra.** Prove that the space of semi-magic squares $\mathrm{SMag}_n(\mathbb{Q})$ (matrices with constant row/column sums) is a subalgebra of $M_n(\mathbb{Q})$.

18. **Signed Similarity.** Let $A \in M_n(\mathbb{K})$ with $\operatorname{char}\mathbb{K} \neq 2$, and set
$$S(A) = \{DA \mid D = \operatorname{diag}(\varepsilon_1, \dots, \varepsilon_n),\ \varepsilon_i = \pm 1\}.$$

    (a) Prove: For any such D, $\det(DA - I) = \pm \det(A - D)$.

    (b) Consider the polynomial $p(t_1, \dots, t_n) = \det\big(A - \operatorname{diag}(t_1, \dots, t_n)\big)$. Show its highest-degree term is $(-1)^n t_1 \cdots t_n$, so p is not the zero polynomial.

    (c) Deduce that there exists a choice $\varepsilon_i \in \{\pm 1\}$ with $p(\varepsilon_1, \dots, \varepsilon_n) \neq 0$, hence some matrix in $S(A)$ has no eigenvalue 1.

> *Remark.*
>
> Evaluate p on all 2^n sign choices; use $\operatorname{char}\mathbb{K} \neq 2$.

6
Jordan Canonical Form

To understand the structure of a linear operator $\mathcal{T} : V \to V$, it is natural to seek a basis of V in which the matrix representation of $\mathcal{T}$ is as simple as possible. We have seen in chapter 5 that if the characteristic polynomial splits into distinct linear factors, the operator is diagonalisable. However, if eigenvalues repeat, diagonalisation is not guaranteed. We now assume the underlying field is algebraically closed (e.g., $F = \mathbb{C}$) and develop a canonical form that applies to *all* operators: the Jordan Canonical Form.

6.1 *Cayley-Hamilton Theorem*

We previously established the Triangulability Criterion in chapter 5. We recall the result here, noting that over $\mathbb{C}$, the condition of the characteristic polynomial splitting is always satisfied.

Proposition 6.1. ***Triangular Form.***
Recall from chapter 5 that an operator is triangulable if and only if its minimal polynomial splits into linear factors. Over an algebraically closed field like $\mathbb{C}$, this condition is always satisfied. Thus, for any linear operator $\mathcal{T} \in \mathcal{L}(V)$ on a finite-dimensional complex vector space, there exists a basis $\mathcal{B}$ such that the matrix of $\mathcal{T}$ relative to $\mathcal{B}$ is upper triangular.

命題

Proof (using Invariant Hyperplanes)

We proceed by induction on $n = \dim V$. For $n = 1$, the matrix is a scalar, which is trivially triangular. Assume the result holds for spaces of dimension $n - 1$. By the theorem on Invariant Hyperplanes (see *Eigenvalues and Diagonalisation*), the operator $\mathcal{T}$ admits an invariant hyperplane U (a subspace of dimension $n - 1$). By the inductive hypothesis, there exists a basis $(e_1, \ldots, e_{n-1})$ for U such that the restriction $\mathcal{T}|_U$ is upper triangular. Specifically:

$$\mathcal{T}e_i \in \operatorname{span}(e_1, \ldots, e_i) \quad \text{for } 1 \leq i \leq n-1.$$

We extend this to a basis for V by choosing any vector $e_n \notin U$. Since $\mathcal{T}e_n \in V$, we can write $\mathcal{T}e_n = \sum_{j=1}^{n} a_{jn}e_j$. The matrix of $\mathcal{T}$ in the basis $(e_1, \ldots, e_n)$ is therefore:

$$A = \begin{bmatrix} \lambda_1 & \cdots & * & a_{1n} \\ 0 & \lambda_2 & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

This is upper triangular; its diagonal entries coincide with the diagonal entries inherited from the induction step on U (hence are eigenvalues of $\mathcal{T}$).

■

This triangular structure allows for a succinct, coordinate-free proof of the Cayley-Hamilton theorem. Recall that we provided a matrix-based proof using the adjugate in chapter 5.

Theorem 6.1. ***Cayley-Hamilton (Alternative Proof).***
Let $\mathcal{T}$ be a linear operator on a finite-dimensional space V, and let $p_{\mathcal{T}}(t)$ be its characteristic polynomial. Then $\mathcal{T}$ annihilates its characteristic polynomial:

$$p_{\mathcal{T}}(\mathcal{T}) = \mathcal{O}.$$

定理

Proof

We provide an alternative, coordinate-free proof using the triangular form established above. Since $p_{\mathcal{T}}(t)$ is independent of the basis, we may assume $\mathcal{T}$ is represented by an upper triangular matrix A with diagonal entries $\lambda_1, \ldots, \lambda_n$. The characteristic polynomial is $p_{\mathcal{T}}(t) = \prod_{i=1}^{n}(t - \lambda_i)$. Consider the filtration of subspaces defined by the basis vectors:

$$V_k = \operatorname{span}(e_1, \ldots, e_k), \quad V_0 = \{0\}.$$

This forms a chain $V_0 \subset V_1 \subset \cdots \subset V_n = V$. Since A is upper triangular, $\mathcal{T}e_k = \lambda_k e_k + v$ where $v \in V_{k-1}$. Consequently, $(\mathcal{T} - \lambda_k\mathcal{E})e_k \in V_{k-1}$. For any $x \in V_k$, writing $x = \alpha e_k + w$ with $w \in V_{k-1}$, we observe:

$$(\mathcal{T} - \lambda_k\mathcal{E})x = \alpha(\mathcal{T} - \lambda_k\mathcal{E})e_k + (\mathcal{T} - \lambda_k\mathcal{E})w.$$

Since V_{k-1} is invariant under $\mathcal{T}$ (and thus under $\mathcal{T} - \lambda_k\mathcal{E}$), both terms lie in V_{k-1}. Hence:

$$(\mathcal{T} - \lambda_k\mathcal{E})V_k \subseteq V_{k-1}.$$

We evaluate $p_{\mathcal{T}}(\mathcal{T})$ by applying the factors successively to the

space V:

$$\begin{aligned} p_{\mathcal{T}}(\mathcal{T})V &= (\mathcal{T}-\lambda_1\mathcal{E})\cdots(\mathcal{T}-\lambda_n\mathcal{E})V_n \\ &\subseteq (\mathcal{T}-\lambda_1\mathcal{E})\cdots(\mathcal{T}-\lambda_{n-1}\mathcal{E})V_{n-1} \\ &\subseteq \dots \\ &\subseteq (\mathcal{T}-\lambda_1\mathcal{E})V_1 \\ &\subseteq V_0 = \{0\}. \end{aligned}$$

Thus $p_{\mathcal{T}}(\mathcal{T}) = \mathcal{O}$.

■

Remark (Minimal Polynomial Divisibility).

As established in chapter 5, the minimal polynomial $\mu_{\mathcal{T}}(t)$ divides the characteristic polynomial $p_{\mathcal{T}}(t)$. Furthermore, every root of $p_{\mathcal{T}}(t)$ (i.e., every eigenvalue) is a root of $\mu_{\mathcal{T}}(t)$.

Remark.

One might be tempted to prove *theorem* 6.1 by substituting matrix A for t in $\det(tI - A)$, yielding $\det(AI - A) = \det(0) = 0$. This reasoning is flawed; $p_A(t)$ is a scalar polynomial, while substitution of A yields a matrix equation. The equality must hold in the algebra of operators, not merely as a scalar value.

6.2 *Jordan Blocks and Nilpotent Operators*

To refine the triangular form, we analyze the structure of nilpotent operators.

Example 6.1. Nilpotent Structure. Let $\mathcal{N}$ be a nilpotent operator with index m (so $\mathcal{N}^m = \mathcal{O}$ but $\mathcal{N}^{m-1} \neq \mathcal{O}$). Pick v such that $\mathcal{N}^{m-1}v \neq 0$. The vectors

$$\mathcal{B} = (\mathcal{N}^{m-1}v, \mathcal{N}^{m-2}v, \dots, \mathcal{N}v, v)$$

are linearly independent. Applying $\mathcal{N}$ to this sequence shifts each vector to the left, annihilating the first. Letting $e_1 = \mathcal{N}^{m-1}v, \dots, e_m = v$, we have:

$$\mathcal{N}e_1 = 0, \quad \mathcal{N}e_k = e_{k-1} \text{ for } k > 1.$$

The matrix of $\mathcal{N}$ in this basis is a Jordan block with eigenvalue 0.

範例

Definition 6.1. ***Jordan Block.***
A **Jordan block** of size m corresponding to $\lambda \in \mathbb{C}$ is a matrix of the

form:

$$J_m(\lambda) = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & \lambda \end{bmatrix} \in \mathbb{C}^{m \times m}.$$

A matrix J is a **Jordan matrix** if it is a block diagonal matrix composed of Jordan blocks:

$$J = \operatorname{diag}(J_{m_1}(\lambda_1), \ldots, J_{m_k}(\lambda_k)).$$

定義

We say an operator $\mathcal{T}$ admits a **Jordan Canonical Form** if there exists a basis (a Jordan basis) in which its matrix is a Jordan matrix. Note that $J_m(\lambda) = \lambda I + N$, where N is the nilpotent matrix with 1s on the superdiagonal.

Example 6.2. Differentiation Space. Consider the space $D_n(\lambda)$ of functions of the form $e^{\lambda t} f(t)$, where $f(t)$ is a polynomial of degree less than n. The differentiation operator $\mathcal{D} = \frac{d}{dt}$ acts on this space. Using the product rule:

$$\frac{d}{dt}\left(\frac{t^k}{k!}e^{\lambda t}\right) = \frac{t^{k-1}}{(k-1)!}e^{\lambda t} + \lambda \frac{t^k}{k!}e^{\lambda t}.$$

Setting basis vectors $e_{k+1} = \frac{t^k}{k!}e^{\lambda t}$ for $k = 0, \ldots, n-1$, we see that $\mathcal{D}e_{k+1} = e_k + \lambda e_{k+1}$ (with $e_0 = 0$). Thus, the matrix of $\mathcal{D}$ in this basis is exactly $J_n(\lambda)$. This structure is fundamental to the theory of linear differential equations.

範例

Example 6.3. Functions of a Jordan Block. If $f(t)$ is a polynomial, the matrix $f(J_m(\lambda))$ has a convenient structure given by the Taylor expansion of f at λ:

$$f(J_m(\lambda)) = \begin{bmatrix} f(\lambda) & f'(\lambda) & \frac{f''(\lambda)}{2!} & \cdots & \frac{f^{(m-1)}(\lambda)}{(m-1)!} \\ 0 & f(\lambda) & f'(\lambda) & \cdots & \frac{f^{(m-2)}(\lambda)}{(m-2)!} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & f(\lambda) & f'(\lambda) \\ 0 & 0 & \cdots & 0 & f(\lambda) \end{bmatrix}.$$

This demonstrates that operating with Jordan blocks is computationally tractable.

範例

Theorem 6.2. ***Existence of Jordan Form.***
Any linear operator $\mathcal{T}$ on a finite-dimensional vector space over an algebraically closed field admits a Jordan Canonical Form. The form is unique up to the permutation of the Jordan blocks.

定理

Proof by induction on $n = \dim V$.

Base case $n = 1$.

Trivial: any 1×1 matrix is $J_1(\lambda)$.

証明終

Inductive step.

Assume the statement holds for all dimensions $< n$. Because the field is algebraically closed, pick an eigenvalue λ of $\mathcal{T}$ and set $\mathcal{U} = \mathcal{T} - \lambda\mathcal{E}$. Let $R = \operatorname{Im}\mathcal{U}$ and $K = \operatorname{Ker}\mathcal{U}$. Since λ is an eigenvalue, $K \neq \{0\}$ and $\dim R = n - \dim K < n$.

Jordan chains inside R. The subspace R is $\mathcal{T}$-invariant: for $y = \mathcal{U}x$ we have $\mathcal{T}y = \mathcal{U}\mathcal{T}x \in \operatorname{Im}\mathcal{U}$. By the inductive hypothesis, $\mathcal{T}|_R$ has a Jordan basis consisting of disjoint chains $\mathcal{C}_i = (v_1^{(i)}, \dots, v_{p_i}^{(i)})$ for eigenvalues μ_i (possibly equal or different from λ).

Extend the λ***-chains.*** If $\mu_i = \lambda$, take the head $v_1^{(i)} \in R$. Since $v_1^{(i)} \in \operatorname{Im}\mathcal{U}$, choose $w^{(i)}$ with $\mathcal{U}w^{(i)} = v_1^{(i)}$. Then

$$(\mathcal{T} - \lambda\mathcal{E})w^{(i)} = v_1^{(i)}, \qquad (\mathcal{T} - \lambda\mathcal{E})v_j^{(i)} = v_{j+1}^{(i)} \ (j < p_i), \qquad (\mathcal{T} - \lambda\mathcal{E})v_{p_i}^{(i)} = 0.$$

Thus $(w^{(i)}, v_1^{(i)}, \dots, v_{p_i}^{(i)})$ is a λ–Jordan chain of length $p_i + 1$. Independence is clear because applying $(\mathcal{T} - \lambda\mathcal{E})^{p_i}$ sends $w^{(i)}$ to $v_{p_i}^{(i)} \neq 0$ while killing the other chains.

If $\mu_i \neq \lambda$, we leave $\mathcal{C}_i$ unchanged. Chains for distinct eigenvalues are automatically independent (they live in distinct generalized eigenspaces of $\mathcal{T}|_R$).

Let $\mathcal{B}_1$ be the union of all extended λ-chains and the unchanged $\mu_i \neq \lambda$ chains. These still form a basis of R together with exactly one new vector for each λ-chain. Denote by c the number of λ-chains in R.

Add missing eigenvectors from K. The tails of the λ-chains in $\mathcal{B}_1$ lie in $K \cap R$ and are independent; extend them to a basis of K by adding vectors $z_1, \dots, z_q$ chosen outside $\operatorname{span}\mathcal{B}_1$. Each z_j is itself a λ-chain of length 1.

Step 4: Independence and spanning. *Independence:* Any non-trivial linear combination of vectors in $\mathcal{B}_1$ that lies in K must involve only the tail vectors (because applying a suitable power

of $\mathcal{U}$ kills earlier vectors but not the tail of its own chain). Since we extended those tails to a basis of K, adding z_j with $z_j \notin \operatorname{span} \mathcal{B}_1$ keeps the whole set independent.

Counting dimensions: $|\mathcal{B}_1| = \dim R + c$, and we add $q = \dim K - c$ vectors z_j. Total size $|\mathcal{B}_1| + q = \dim R + \dim K = n$, so independence implies the set is a basis of V.

In this basis, each extended λ-chain is a Jordan block for λ, each unchanged chain is a Jordan block for its μ_i, and each z_j is a 1×1 λ-block. Hence the matrix of $\mathcal{T}$ is block diagonal with Jordan blocks, as desired.

証明終

The structure of the Jordan form is intimately tied to the minimal polynomial. If $\mu_{\mathcal{T}}(t) = \prod_{i=1}^{p} (t - \lambda_i)^{m_i}$, then m_i corresponds to the size of the *largest* Jordan block associated with λ_i.

Corollary 6.1. *Diagonalisability Criterion.* A matrix A is diagonalisable if and only if its minimal polynomial $\mu_A(t)$ has no repeated roots (i.e., it is a product of distinct linear factors).

推論

Proof

If μ_A has no repeated roots, the largest Jordan block for any eigenvalue has size 1. Thus all blocks are 1×1, meaning A is diagonal. Conversely, if A is diagonal, it satisfies $\prod (A - \lambda_i I) = \mathcal{O}$ where the product is over distinct eigenvalues, so μ_A splits distinctly.

■

6.3 *Root Subspaces*

To prove the existence of the Jordan form, we decompose the space V into subspaces corresponding to each eigenvalue.

Definition 6.2. ***Root Subspace.***
Let $\lambda \in \operatorname{Spec}(\mathcal{T})$. The **root subspace** (or generalized eigenspace) $V(\lambda)$ is defined as:

$$V(\lambda) = \{v \in V \mid (\mathcal{T} - \lambda \mathcal{E})^k v = 0 \text{ for some } k \geq 1\}.$$

定義

Since V is finite-dimensional, this is equivalent to $V(\lambda) = \operatorname{Ker}((\mathcal{T} - \lambda\mathcal{E})^n)$. Clearly, the eigenspace $\operatorname{Ker}(\mathcal{T} - \lambda\mathcal{E})$ is contained in $V(\lambda)$.
We now provide the constructive proof for the Primary Decomposition Theorem stated in chapter 5.

Theorem 6.3. ***Primary Decomposition.***
Let $\mathcal{T}$ be a linear operator with characteristic polynomial $p_{\mathcal{T}}(t) = \prod_{i=1}^{p}(t - \lambda_i)^{n_i}$, where λ_i are distinct. Then V decomposes as the direct sum of invariant root subspaces:

$$V = V(\lambda_1) \oplus V(\lambda_2) \oplus \cdots \oplus V(\lambda_p).$$

Moreover, $\dim V(\lambda_i) = n_i$, and the restriction of $\mathcal{T} - \lambda_i\mathcal{E}$ to $V(\lambda_i)$ is nilpotent.

定理

Proof

For each i, define the polynomial $Q_i(t) = p_{\mathcal{T}}(t)/(t - \lambda_i)^{n_i} = \prod_{j\neq i}(t - \lambda_j)^{n_j}$. The polynomials $Q_1(t), \ldots, Q_p(t)$ share no common root, so their greatest common divisor is 1. By the Euclidean algorithm for polynomials (Bezout's identity), there exist polynomials $f_1(t), \ldots, f_p(t)$ such that:

$$\sum_{i=1}^{p} f_i(t)Q_i(t) = 1.$$

Substituting the operator $\mathcal{T}$:

$$\sum_{i=1}^{p} f_i(\mathcal{T})Q_i(\mathcal{T}) = \mathcal{E}.$$

Let $\mathcal{P}_i = f_i(\mathcal{T})Q_i(\mathcal{T})$. Note that $\mathcal{P}_i$ commutes with $\mathcal{T}$. We claim $W_i = \mathrm{Im}(\mathcal{P}_i)$ coincides with $V(\lambda_i)$. First, observe that $Q_i(t)$ contains the factor $(t - \lambda_j)^{n_j}$ for all $j \neq i$. By Cayley-Hamilton, $p_{\mathcal{T}}(\mathcal{T}) = \mathcal{O}$. Thus $(\mathcal{T} - \lambda_i\mathcal{E})^{n_i}Q_i(\mathcal{T}) = p_{\mathcal{T}}(\mathcal{T}) = \mathcal{O}$. Consequently, for any v, $(\mathcal{T} - \lambda_i\mathcal{E})^{n_i}\mathcal{P}_i v = f_i(\mathcal{T})(\mathcal{T} - \lambda_i\mathcal{E})^{n_i}Q_i(\mathcal{T})v = 0$. This implies $\mathrm{Im}(\mathcal{P}_i) \subseteq V(\lambda_i)$.
The identity $\sum \mathcal{P}_i = \mathcal{E}$ implies $V = \sum \mathrm{Im}(\mathcal{P}_i) = \sum V(\lambda_i)$. To show the sum is direct, suppose $v \in V(\lambda_i) \cap \sum_{j\neq i} V(\lambda_j)$. On $V(\lambda_j)$, the operator $(\mathcal{T} - \lambda_j\mathcal{E})$ is nilpotent. Because Q_i contains $(t - \lambda_j)^{n_j}$, we have $Q_i(\mathcal{T})|_{V(\lambda_j)} = 0$ for $j \neq i$. Conversely, on $V(\lambda_i)$ the nilpotent part of $(\mathcal{T} - \lambda_i\mathcal{E})$ commutes with $Q_i(\mathcal{T})$, and the scalar $Q_i(\lambda_i) \neq 0$ implies

$$Q_i(\mathcal{T})|_{V(\lambda_i)} = Q_i(\lambda_i)\mathcal{E} + (\text{nilpotent})$$

which is invertible on $V(\lambda_i)$ (a non-zero scalar plus nilpotent is invertible on a finite-dimensional space). Therefore $\mathcal{P}_i$ restricts to an automorphism of $V(\lambda_i)$ and vanishes on $V(\lambda_j)$ for $j \neq i$, so $\mathcal{P}_i$ is the projection onto $V(\lambda_i)$ along $\bigoplus_{j\neq i} V(\lambda_j)$. This proves the sum is direct and $W_i = V(\lambda_i)$.
Finally, on each $V(\lambda_i)$ the restriction of $\mathcal{T}$ equals $\lambda_i\mathcal{E}$ plus a nilpotent operator (because $(\mathcal{T} - \lambda_i\mathcal{E})^{n_i} = 0$ on that subspace). Thus

the classification of $\mathcal{T}$ reduces to nilpotent blocks, giving the Jordan form.

■

6.4 Cyclic Subspaces

To construct the Jordan Canonical Form, we break down nilpotent operators into simpler components. The fundamental building block is the cyclic subspace.

Definition 6.3. ***Cyclic Subspace.***
Let $\mathcal{N}$ be a nilpotent operator on V with nilpotency index m. For any vector $v \in V$, the **cyclic subspace** generated by v with respect to $\mathcal{N}$ is:

$$Z(v;\mathcal{N}) = \text{span}(v, \mathcal{N}v, \mathcal{N}^2v, \ldots, \mathcal{N}^{k-1}v),$$

where k is the smallest integer such that $\mathcal{N}^k v = 0$. Note that $k \leq m$.

定義

The vectors $(v, \mathcal{N}v, \ldots, \mathcal{N}^{k-1}v)$ form a basis for $Z(v;\mathcal{N})$. In the reversed order $(\mathcal{N}^{k-1}v, \ldots, v)$, the matrix of $\mathcal{N}$ restricted to this subspace is the Jordan block $J_k(0)$.

Theorem 6.4. ***Jordan Form for Nilpotent Operators.***
Let $\mathcal{N}$ be a nilpotent operator on a finite-dimensional vector space V. Then V admits a decomposition into a direct sum of cyclic subspaces:

$$V = Z(v_1;\mathcal{N}) \oplus Z(v_2;\mathcal{N}) \oplus \cdots \oplus Z(v_s;\mathcal{N}).$$

Consequently, there exists a basis in which the matrix of $\mathcal{N}$ is a direct sum of Jordan blocks with eigenvalue 0.

定理

Proof

We proceed by induction on $\dim V$. The base case is trivial. Assume the theorem holds for spaces of dimension less than n. Since $\mathcal{N}$ is nilpotent, its image $\text{Im}\,\mathcal{N}$ is a proper subspace of V (otherwise $\mathcal{N}$ would be surjective and thus invertible, contradicting nilpotency). Let $U = \text{Im}\,\mathcal{N}$. Since $\dim U < \dim V$, the inductive hypothesis implies U decomposes into cyclic subspaces:

$$U = Z(u_1;\mathcal{N}) \oplus \cdots \oplus Z(u_r;\mathcal{N}).$$

Let k_i be the dimension of $Z(u_i;\mathcal{N})$, so $\mathcal{N}^{k_i}u_i = 0$ and u_i generates the sequence $u_i, \mathcal{N}u_i, \ldots$. Since $u_i \in \text{Im}\,\mathcal{N}$, there exist vectors $v_i \in V$ such that $\mathcal{N}v_i = u_i$. Consider the new cyclic subspaces $Z(v_i;\mathcal{N})$.

The sequence generated by v_i is $v_i, u_i, \mathcal{N}u_i, \ldots$, which has length $k_i + 1$. Let $W = \bigoplus_{i=1}^{r} Z(v_i; \mathcal{N})$. We must determine if W covers all of V. Independence of these subspaces follows from a "highest nonzero iterate" argument: if $\sum_i x_i = 0$ with $x_i \in Z(v_i; \mathcal{N})$, apply $\mathcal{N}^{k-1}$ where k is maximal such that some x_i has a nonzero $\mathcal{N}^{k-1}$-iterate. Only one tail term survives (the tail of its chain), forcing all coefficients to be zero.
If $W \neq V$, we can find vectors in $\operatorname{Ker}\mathcal{N}$ that are not in W to complete the basis. Specifically, extend the independent set $\{\mathcal{N}^{k_i}u_i\}_{i=1}^{r}$ (which forms a basis for $\operatorname{Im}\mathcal{N} \cap \operatorname{Ker}\mathcal{N}$) to a full basis of $\operatorname{Ker}\mathcal{N}$ by adding vectors $z_1, \ldots, z_q$ chosen outside W. These z_j generate cyclic subspaces of dimension 1 (since $\mathcal{N}z_j = 0$).
Any vector in $W \cap \operatorname{span}(z_1, \ldots, z_q)$ would lie in both W and $\ker\mathcal{N}$; applying the same highest-iterate argument shows this forces the vector to be 0, so the sum remains direct. Finally,

$$V = W \oplus \operatorname{span}(z_1) \oplus \cdots \oplus \operatorname{span}(z_q),$$

so V is a direct sum of cyclic subspaces.

■

Combining this with the Primary Decomposition Theorem yields the full existence result. For any operator $\mathcal{T}$, $V = \bigoplus V(\lambda_i)$. On each $V(\lambda_i)$, the operator $\mathcal{T} - \lambda_i\mathcal{E}$ is nilpotent. Decomposing $V(\lambda_i)$ into cyclic subspaces for $\mathcal{T} - \lambda_i\mathcal{E}$ yields blocks of the form $J_m(\lambda_i)$.
While the existence of the Jordan form is guaranteed, proving uniqueness requires identifying invariants that do not depend on the choice of basis.

Theorem 6.5. ***Uniqueness of Jordan Form.***
The Jordan Canonical Form of an operator $\mathcal{T}$ is unique up to the order of the Jordan blocks.

定理

Proof

The number and sizes of the Jordan blocks corresponding to an eigenvalue λ are completely determined by the ranks of the powers of $(\mathcal{T} - \lambda\mathcal{E})$. Let $N(m, \lambda)$ be the number of Jordan blocks of size m for the eigenvalue λ. Let $r_k = \operatorname{rank}((\mathcal{T} - \lambda\mathcal{E})^k)$. The geometric multiplicity $\dim\operatorname{Ker}(\mathcal{T} - \lambda\mathcal{E}) = n - r_1$ is the total number of Jordan blocks for λ (since each block contributes exactly one eigenvector).
Generally, for a single block $J_m(\lambda)$, the rank of $(J_m(\lambda) - \lambda I)^k$ drops by 1 for each power until it becomes 0 for all $k \geq m$. Summing over all blocks, $r_{k-1} - r_k$ counts the number of blocks of size at least k;

taking a discrete difference isolates those of exact size m:

$$N(m,\lambda) = r_{m-1} - 2r_m + r_{m+1}.$$

Here $r_k = \operatorname{rank}((\mathcal{T} - \lambda\mathcal{E})^k)$, with $r_0 = n$ and $r_k = 0$ for sufficiently large k (specifically for $k \geq \dim V$). Since the ranks r_k are basis-independent invariants of $\mathcal{T}$, the numbers $N(m,\lambda)$ are unique.

■

The Jordan-Chevalley Decomposition

The Jordan Canonical Form allows us to decompose any operator into a diagonalisable part and a nilpotent part.

Theorem 6.6. ***Jordan-Chevalley Decomposition.***
Let $\mathcal{T}$ be a linear operator on a finite-dimensional vector space V over an algebraically closed field. There exist unique operators $\mathcal{S}$ (semisimple/diagonalisable) and $\mathcal{N}$ (nilpotent) such that:

$$\mathcal{T} = \mathcal{S} + \mathcal{N} \quad \text{and} \quad \mathcal{S}\mathcal{N} = \mathcal{N}\mathcal{S}.$$

Furthermore, $\mathcal{S}$ and $\mathcal{N}$ can be expressed as polynomials in $\mathcal{T}$.

定理

Proof

Consider the eigenspace decomposition $V = \bigoplus V(\lambda_i)$ given by the primary decomposition. Define $\mathcal{S}$ to be the operator that acts as $\lambda_i\mathcal{E}$ on each subspace $V(\lambda_i)$. Since V has a basis of eigenvectors for $\mathcal{S}$ (the union of bases for each $V(\lambda_i)$), $\mathcal{S}$ is diagonalisable. Define $\mathcal{N} = \mathcal{T} - \mathcal{S}$. On $V(\lambda_i)$, $\mathcal{N}$ acts as $\mathcal{T} - \lambda_i\mathcal{E}$, which is nilpotent by the definition of root subspaces. Since $\mathcal{N}$ is nilpotent on each invariant summand, it is nilpotent on V. Commutativity follows because $\mathcal{S}$ is a polynomial in $\mathcal{T}$ (using Lagrange interpolation to match eigenvalues on the blocks), and $\mathcal{T}$ commutes with itself. Uniqueness is proved by observing that if $\mathcal{T} = \mathcal{S}' + \mathcal{N}'$ is another such decomposition, $\mathcal{S} - \mathcal{S}' = \mathcal{N}' - \mathcal{N}$. The LHS is diagonalisable and the RHS is nilpotent; the only operator that is both is the zero operator.

■

Example 6.4. Computing the Decomposition via Polynomials.
Consider the operator T on $\mathbb{R}^3$ represented by the lower triangular matrix:

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

The characteristic polynomial is $p(t) = (t-2)^2(t+1)$. Since the

rank of $A - 2I$ is 2 (nullity 1), the geometric multiplicity of $\lambda = 2$ is 1, strictly less than the algebraic multiplicity 2. Thus, A is not diagonalisable. The minimal polynomial is $m(t) = (t-2)^2(t+1)$. To find the semisimple part S and nilpotent part N, we use the partial fraction decomposition (Bezout's identity) on the factors of $m(t)$. Let $p_1(t) = t+1$ and $p_2(t) = (t-2)^2$. We define $f_1(t) = p_2(t) = (t-2)^2$ and $f_2(t) = p_1(t) = t+1$. We seek polynomials g_1, g_2 such that:

$$f_1(t)g_1(t) + f_2(t)g_2(t) = 1.$$

Using the Euclidean algorithm, we find:

$$g_1(t) = \frac{1}{9}, \quad g_2(t) = \frac{-t+5}{9}.$$

Check: $\frac{1}{9}(t^2 - 4t + 4) + \frac{1}{9}(t+1)(-t+5) = \frac{1}{9}(t^2 - 4t + 4 - t^2 + 4t + 5) = 1$.
The projections onto the generalized eigenspaces are $E_1 = f_1(A)g_1(A)$ and $E_2 = f_2(A)g_2(A)$.

$$E_1 = \frac{1}{9}(A - 2I)^2, \quad E_2 = \frac{1}{9}(A + I)(5I - A).$$

Calculating these yields:

$$E_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The semisimple part S is the weighted sum of projections by eigenvalues:

$$S = (-1)E_1 + 2E_2 = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

(In this specific basis, S happens to be diagonal, though usually it is just diagonalisable). The nilpotent part is $N = A - S$:

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

One can verify $N^2 = 0$ and $SN = NS$.

範例

Remark.

The decomposition $A = S + N$ with $SN = NS$ is particularly useful

for computing matrix exponentials. Since S and N commute, $e^A = e^{S+N} = e^S e^N$. e^S is easily computed (diagonalise S), and e^N is a finite sum $\sum_{k=0}^{m-1} \frac{N^k}{k!}$ since N is nilpotent.

Example 6.5. Calculating Jordan Structure. Let A be a matrix with characteristic polynomial $p(\lambda) = (\lambda - 2)^4$. Suppose the ranks of $(A - 2I)^k$ are: $r_1 = 2$, $r_2 = 1$, $r_3 = 0$. Using the formula:

$$N(1,2) = r_0 - 2r_1 + r_2 = 4 - 2(2) + 1 = 1.$$
$$N(2,2) = r_1 - 2r_2 + r_3 = 2 - 2(1) + 0 = 0.$$
$$N(3,2) = r_2 - 2r_3 + r_4 = 1 - 0 + 0 = 1.$$

Thus, A has one block of size 1 and one block of size 3: $J \cong \text{diag}(J_3(2), J_1(2))$.

範例

When the field is not algebraically closed (e.g., $\mathbb{R}$), the Jordan form may not exist because eigenvalues might not lie in the field. In this case, we use the Rational Canonical Form, which relies on cyclic subspaces for the operator $\mathcal{T}$ itself rather than its nilpotent parts.

Definition 6.4. ***Cyclic Vector.***
A vector v is **cyclic** for an operator $\mathcal{T}$ if the vectors $v, \mathcal{T}v, \ldots, \mathcal{T}^{n-1}v$ form a basis for V. The space is then called a **cyclic space**.

定義

Example 6.6. Cyclic and Non-Cyclic Vectors. Consider the operator $\mathcal{T}$ on $\mathbb{R}^2$ defined by $\mathcal{T}(x_1, x_2) = (0, x_1)$.

- Let $v = e_1 = (1,0)$. Then $\mathcal{T}v = (0,1) = e_2$ and $\mathcal{T}^2 v = 0$. The set $\{e_1, e_2\}$ spans $\mathbb{R}^2$, so e_1 is a cyclic vector.
- Let $w = e_2 = (0,1)$. Then $\mathcal{T}w = (0,0)$. The cyclic subspace $Z(w;\mathcal{T})$ is $\text{span}(e_2)$, which is not the whole space. Thus e_2 is not cyclic.

範例

Proposition 6.2. ***Cyclic Subspace Properties.***
Let $Z(x;\mathcal{T})$ be the cyclic subspace generated by x.

1. $Z(x;\mathcal{T}) = \text{span}(x)$ if and only if x is an eigenvector of $\mathcal{T}$.
2. If V is a cyclic space with cyclic vector v, then the minimal polynomial $\mu_{\mathcal{T}}(t)$ equals the characteristic polynomial $p_{\mathcal{T}}(t)$, and both equal the $\mathcal{T}$-annihilator $p_v(t)$.

命題

Proof

1. ($\implies$) If $Z(x;\mathcal{T}) = \text{span}(x)$, then $\mathcal{T}x \in Z(x;\mathcal{T})$ implies $\mathcal{T}x =$

λx for some $\lambda \in F$. Since x generates the subspace (assumed non-zero), x is an eigenvector.

($\Longleftarrow$) If $\mathcal{T}x = \lambda x$, then for any polynomial $g(t)$, $g(\mathcal{T})x = g(\lambda)x \in \text{span}(x)$. Thus $Z(x;\mathcal{T}) \subseteq \text{span}(x)$, and equality holds.

2. Let $n = \dim V$. Since v is cyclic, the set $\mathcal{B} = \{v, \mathcal{T}v, \dots, \mathcal{T}^{n-1}v\}$ is linearly independent and forms a basis. The annihilator $p_v(t)$ is the unique monic polynomial of least degree such that $p_v(\mathcal{T})v = 0$. If $\deg p_v = k < n$, then $\mathcal{T}^k v$ would be a linear combination of $v, \dots, \mathcal{T}^{k-1}v$, contradicting the independence of $\mathcal{B}$. Thus $\deg p_v = n$.
Since $\mu_{\mathcal{T}}(\mathcal{T}) = \mathcal{O}$, we have $\mu_{\mathcal{T}}(\mathcal{T})v = 0$, so p_v divides $\mu_{\mathcal{T}}$. Conversely, for any $w \in V$, since v generates V, we can write $w = g(\mathcal{T})v$ for some polynomial g. Then:

$$p_v(\mathcal{T})w = p_v(\mathcal{T})g(\mathcal{T})v = g(\mathcal{T})p_v(\mathcal{T})v = g(\mathcal{T})0 = 0.$$

Thus $p_v(\mathcal{T})$ annihilates all vectors, so $\mu_{\mathcal{T}}$ divides p_v. Since they are both monic and divide each other, $\mu_{\mathcal{T}} = p_v$. Finally, $p_{\mathcal{T}}$ is a monic polynomial of degree n. By Cayley-Hamilton, $\mu_{\mathcal{T}}$ divides $p_{\mathcal{T}}$. Since $\deg \mu_{\mathcal{T}} = \deg p_v = n$, we must have $\mu_{\mathcal{T}} = p_{\mathcal{T}}$.

■

If V is cyclic with minimal polynomial $t^n - \sum_{i=0}^{n-1} a_i t^i$, the matrix of $\mathcal{T}$ in the basis $\mathcal{B} = (\mathcal{T}^{n-1}v, \dots, v)$ takes the form of a **companion matrix** (or cyclic block):

$$C(p) = \begin{bmatrix} a_{n-1} & 1 & 0 & \cdots & 0 \\ a_{n-2} & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ a_1 & 0 & 0 & \cdots & 1 \\ a_0 & 0 & 0 & \cdots & 0 \end{bmatrix}.$$

Every operator admits a decomposition $V = Z_1 \oplus \cdots \oplus Z_k$ where each Z_i is a cyclic subspace. The resulting block diagonal matrix of companion matrices is the Rational Canonical Form. Unlike Jordan form, this requires no field extension.

Finding the Similarity Matrix

Finding the matrix C such that $C^{-1}AC = J$ is equivalent to solving the linear system $AC = CJ$. In practice one constructs the basis of generalized eigenvectors explicitly. For a chain ending in eigenvector v (where $(A - \lambda I)v = 0$), one solves $(A - \lambda I)v_2 = v$, $(A - \lambda I)v_3 = v_2$, etc., moving upwards.

Example 6.7. Projection Matrix. Consider the matrix S of all ones ($S_{ij} = 1$). $S^2 = nS$. The minimal polynomial is $t(t-n)$. Since roots are distinct (0 and n), S is diagonalisable. Rank is 1, so the eigenvalue 0 has geometric multiplicity $n-1$. $J = \operatorname{diag}(n, 0, \dots, 0)$. The transition matrix C can be found by picking one eigenvector for $\lambda = n$ (e.g., $(1, \dots, 1)^T$) and $n-1$ independent vectors in the kernel of S (vectors summing to 0).

範例

6.5 *Admissibility and Decomposition*

To establish the cyclic decomposition for general operators (including the Rational Canonical Form), we formalise the conditions under which a subspace can be "split off" as a direct summand. This leads to the concept of T-admissibility, which is central to proving the general structure theorem without assuming field closure.

Definition 6.5. ***T-Conductor and Annihilator.***
Recall from chapter 5 that the **T-conductor** of a vector v into a subspace W is the set $S_T(v, W) = \{f \in F[t] \mid f(T)v \in W\}$. This set is non-empty (it contains the minimal polynomial of T) and is closed under addition and multiplication by any polynomial. Consequently, there exists a unique monic polynomial of lowest degree in $S_T(v, W)$ that divides every other polynomial in the set. We refer to this specific polynomial as the T-conductor. The **T-annihilator** of v, denoted $p_v(t)$, is the T-conductor of v into the zero subspace $\{0\}$. It is the unique monic polynomial of lowest degree such that $p_v(T)v = 0$.

定義

Definition 6.6. ***T-Admissibility.***
Let W be a T-invariant subspace of V. We say W is T-admissible if for every polynomial $f(t)$ and every vector $v \in V$, the condition $f(T)v \in W$ implies there exists a vector $w \in W$ such that $f(T)v = f(T)w$.

定義

Remark.

The motivation for this technical definition lies in the problem of finding invariant complements. If W is a T-invariant subspace, there does not necessarily exist a subspace W' such that $V = W \oplus W'$ and W' is also invariant. However, if W is part of such a decomposition, it must be T-admissible.

Lemma 6.1. ***Invariant Direct Sums imply Admissibility.***
Let $V = W \oplus W'$ where both W and W' are T-invariant. Then W is

T-admissible.

引理

Proof

Let $v \in V$ and $f(t)$ be a polynomial such that $f(T)v \in W$. Since $V = W \oplus W'$, we can write $v = w + w'$ uniquely, with $w \in W, w' \in W'$. By linearity, $f(T)v = f(T)w + f(T)w'$. Since W and W' are invariant, $f(T)w \in W$ and $f(T)w' \in W'$. We are given that $f(T)v \in W$. Rearranging the equation:

$$f(T)w' = f(T)v - f(T)w.$$

The RHS is in W (difference of two vectors in W). The LHS is in W'. Since $W \cap W' = \{0\}$, we must have $f(T)w' = 0$. Thus $f(T)v = f(T)w$. Since $w \in W$, the condition for T-admissibility is satisfied.

■

Intuitively, T-admissibility ensures that if a vector "looks like" it belongs to W relative to the action of polynomials in T, we can find a representative actually inside W that behaves identically under that polynomial action.

Theorem 6.7. ***Cyclic Decomposition Theorem.***
Let T be a linear operator on a finite-dimensional vector space V. There exist non-zero vectors $v_1, \ldots, v_r \in V$ with T-annihilators $p_1, \ldots, p_r$ such that:
1. $V = Z(v_1; T) \oplus Z(v_2; T) \oplus \cdots \oplus Z(v_r; T)$.
2. p_k divides p_{k-1} for $k = 2, \ldots, r$.

Furthermore, the integer r and the polynomials $p_1, \ldots, p_r$ are uniquely determined by T.

定理

The polynomials $p_i(t)$ are called the **invariant factors** of T. The first polynomial $p_1(t)$ is the minimal polynomial of T, and the product $\prod p_i(t)$ is the characteristic polynomial (up to a scalar factor).

Proof

The proof is constructive and relies on T-admissibility. We proceed by induction. Let $W_0 = \{0\}$, which is trivially T-admissible. Suppose we have constructed $W_{k-1} = Z(v_1; T) \oplus \cdots \oplus Z(v_{k-1}; T)$. If $W_{k-1} \neq V$, we find a vector v_k such that the T-conductor $S_T(v_k, W_{k-1})$ is maximal among all vectors in $V \setminus W_{k-1}$. Let $p_k = S_T(v_k, W_{k-1})$. We define $Z(v_k; T)$ and show that $W_k = W_{k-1} \oplus Z(v_k; T)$ is a direct sum and is T-admissible. The divisibility condition $p_k \mid p_{k-1}$ arises from the maximality of the conductor. Specifically, since v_{k-1} had a "larger" conductor at the

previous step, the structure of the cyclic subspaces enforces the divisibility chain. The process terminates when $W_r = V$. The uniqueness of the invariant factors $p_1, \dots, p_r$ follows from the fact that they are determined by the greatest common divisors of the minors of the matrix $tI - A$. Specifically, if $D_k(t)$ is the greatest common divisor of all $k \times k$ minors of $tI - A$, then $p_k(t)$ can be recovered from the quotients of these scalar invariants (specifically $p_{r-j+1}(t) = D_{n-j+1}(t)/D_{n-j}(t)$). Since determinants are basis-independent, these polynomials are unique to the operator.

■

The cyclic decomposition allows us to represent any operator by a matrix composed of companion matrices, regardless of whether the field is algebraically closed.

Theorem 6.8. ***Rational Canonical Form.***
Every square matrix A over a field F is similar to a block diagonal matrix

$$R = \begin{bmatrix} C(p_1) & 0 & \cdots & 0 \\ 0 & C(p_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & C(p_r) \end{bmatrix},$$

where $C(p_i)$ is the companion matrix of the polynomial $p_i(t)$, and $p_k(t)$ divides $p_{k-1}(t)$ for $k = 2, \dots, r$. This matrix R is called the **Rational Canonical Form** of A.

定理

Example 6.8. Rational Form Computation. Consider the matrix

$$B = \begin{bmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{bmatrix}.$$

The characteristic polynomial is $\chi_B(t) = (t-1)(t-2)^2$. The minimal polynomial is $\mu_B(t) = (t-1)(t-2) = t^2 - 3t + 2$. Since $\mu_B \neq \chi_B$, there must be more than one invariant factor. The invariant factors must satisfy $p_1 = \mu_B$ and $\prod p_i = \chi_B$, with $p_2 \mid p_1$. The only possibility is $p_1(t) = (t-1)(t-2)$ and $p_2(t) = (t-2)$. The companion matrix for $p_1(t) = t^2 - 3t + 2$ is:

$$C(p_1) = \begin{bmatrix} 0 & -2 \\ 1 & 3 \end{bmatrix}.$$

The companion matrix for $p_2(t) = t - 2$ is simply $[2]$. Thus, the Ra-

tional Canonical Form is:

$$R = \begin{bmatrix} 0 & -2 & 0 \\ 1 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

Note that B is diagonalisable (eigenvalues 1, 2, 2 with distinct eigenvectors), but the Rational Form groups the cyclic components explicitly.

範例

Remark.

The Rational Canonical Form is "rational" because it requires no field extension to compute. It relies only on operations within the field F, unlike the Jordan Canonical Form which may require splitting fields to find eigenvalues.

6.6 *Exercises*

1. **Determinant of the Shifted Matrix.** Let S be the $n \times n$ matrix of all ones from example 6.7. Compute the determinant of the matrix:

$$A = \begin{bmatrix} m & -1 & \cdots & -1 \\ -1 & m & \cdots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \cdots & m \end{bmatrix}$$

   Show that $\det A = \chi_S(m+1)$, where $\chi_S(t)$ is the characteristic polynomial of S.

2. **Classification of Nilpotent Matrices.** Up to similarity, the following matrices exhaust all nonzero 4×4 nilpotent matrices:

$$A_1 = J_2(0) \oplus J_1(0) \oplus J_1(0), \qquad A_2 = J_2(0) \oplus J_2(0),$$
$$A_3 = J_3(0) \oplus J_1(0), \qquad A_4 = J_4(0).$$

   Determine which A_i each of the following matrices is similar to:

$$M_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad M_2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$M_3 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix}, \quad M_4 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

3. **Reconstruction from Invariants.**
   (a) Given a characteristic polynomial $\chi_A(t) = (t-3)^4(t+2)$ and $\operatorname{rank}(A - 3I) = 2$, find the Jordan Canonical Form $J(A)$.
   (b) In the cases where $\operatorname{rank}(A - 3I) \in \{1, 3, 4\}$, can $J(A)$ be uniquely recovered? Explain why or why not.

4. **Comparison of Matrices.** Consider the matrices:
$$A = \begin{bmatrix} 6 & 2 & -2 \\ -2 & 2 & 2 \\ 2 & 2 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 6 & 2 & 2 \\ -2 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$
   (a) Show that A and B have the same characteristic polynomial.
   (b) Find the minimal polynomials $\mu_A(t)$ and $\mu_B(t)$.
   (c) Find the Jordan forms $J(A)$ and $J(B)$. Are A and B similar?

5. **Self-Duality.** Prove that every matrix $A \in M_n(\mathbb{C})$ is similar to its transpose A^T.

   It suffices to prove this for a single Jordan block.

6. **Roots of Unity.** Prove that for a matrix $A \in M_n(\mathbb{C})$, the relation $A^N = I$ holds if and only if A is diagonalisable and its eigenvalues are all N-th roots of unity.

7. **The Ring of Magic Squares.** Let $\operatorname{Mag}_n(\mathbb{Q})$ denote the set of $n \times n$ magic squares (matrices where all row sums, column sums, and main diagonal sums are equal). Verify directly that:
$$A = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 2 \\ 2 & 0 & 1 \end{bmatrix} \in \operatorname{Mag}_3(\mathbb{Q}),$$
   but $A^2 \notin \operatorname{Mag}_3(\mathbb{Q})$. Conclude that the set of magic squares is not closed under multiplication. (Contrast this with the set of semi-magic squares, which is closed under multiplication).

   Use the Cayley-Hamilton theorem to express A^m as a linear combination of I, A, A^2. Note that this property is specific to $n = 3$.

8. **Higher Order Magic Squares.** Verify that for any $m \geq 2$, the matrix A^m is not a magic square, where:
$$A = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \in \operatorname{Mag}_4(\mathbb{Q}).$$
   Using this, show that for all $n \geq 4$, there exists an $n \times n$ magic square matrix whose m-th power ($m \geq 2$) is not a magic square.

9. **Jordan-Chevalley Computation.** Let $A = \begin{bmatrix} 1 & 1 \\ -1 & 3 \end{bmatrix} \in M_2(\mathbb{R})$.
   (a) Determine the semisimple part S and the nilpotent part N of the decomposition $A = S + N$ such that $SN = NS$.

(b) Express S and N as polynomials in A.

10. **Jacobson's Lemma for 2×2.** Prove that for any three matrices $X, Y, Z \in M_2(\mathbb{R})$, the following identity holds:

$$[[X, Y]^2, Z] = 0,$$

where $[X, Y] = XY - YX$ is the commutator.

11. **Classification via Minimal Polynomials.** Let N_1 and N_2 be 3×3 nilpotent matrices over a field F. Prove that N_1 and N_2 are similar if and only if they have the same minimal polynomial. Give a counter-example to show this is false for 4×4 matrices.

12. **Constructing Jordan Forms.** If A is a complex 5×5 matrix with characteristic polynomial $f(x) = (x-2)^3(x+7)^2$ and minimal polynomial $p(x) = (x-2)^2(x+7)$, determine the Jordan Canonical Form of A.

13. **Enumerating Similarity Classes.** How many distinct similarity classes of 6×6 complex matrices exist that have the characteristic polynomial $\chi(x) = (x+2)^4(x-1)^2$?

14. **Differentiation Operator.** Let $V = P_{\leq 3}(\mathbb{C})$ be the space of polynomials of degree at most 3. Let $T : V \to V$ be the differentiation operator $T(f) = f'$. Find the Jordan form of T.

15. **Rational Canonical Form Calculation.** Find the minimal polynomials and the Rational Canonical Forms for the following real matrices:

$$A_1 = \begin{bmatrix} 0 & -1 & -1 \\ 1 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad A_3 = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}.$$

16. **Similarity Criteria.** Prove that if A and B are 3×3 matrices over a field F, a necessary and sufficient condition for similarity is that they share the same characteristic and minimal polynomials.

17. **Invariant Complements.** Let T be a linear operator on a finite-dimensional space V, with range R and null space N.

    (a) Prove that R has a complementary T-invariant subspace if and only if $R \cap N = \{0\}$.

    (b) If $R \cap N = \{0\}$, prove that N is the *unique* T-invariant subspace complementary to R.

18. **Non-Splitting Subspaces.** Let T be the linear operator on $\mathbb{R}^3$ represented by:

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

Let W be the null space of $T - 2\mathcal{E}$. Prove that W has no complementary T-invariant subspace.

Consider the action of T on the generalized eigenspace associated with eigenvalue 2.

19. **Cyclic Vectors in $\mathbb{F}^2$.** Let T be a linear operator on F^2. Prove that any non-zero vector which is not an eigenvector for T is a cyclic vector. Deduce that T has a cyclic vector unless T is a scalar multiple of the identity.

20. **Cyclicity Inheritance.** Prove that if T^2 has a cyclic vector, then T has a cyclic vector. Is the converse true?

21. **Nilpotent Cyclic Generators.** Let N be a nilpotent operator on an n-dimensional space V. Suppose $N^{n-1} \neq 0$. Let $a \in V$ such that $N^{n-1}a \neq 0$.

    (a) Prove that a is a cyclic vector for N.

    (b) Describe the matrix of N in the ordered basis $\{a, Na, \ldots, N^{n-1}a\}$.

22. **Companion Matrix Characteristic.** Give a direct proof (by expanding minors or induction) that the characteristic polynomial of the companion matrix $C(p)$ is exactly $p(t)$.

23. **Diagonalisability and Cyclicity.** Let T be a diagonalisable operator on an n-dimensional space V.

    (a) Prove that T has a cyclic vector if and only if T has n distinct eigenvalues.

    (b) If T has distinct eigenvalues, construct a cyclic vector explicitly as a sum of eigenvectors.

24. **The Double Commutant.** Let T be a linear operator on a finite-dimensional space V. Prove that if T has a cyclic vector, then any linear operator U which commutes with T (i.e., $UT = TU$) is a polynomial in T.

25. **Square Roots of Nilpotents.** Let N be an $n \times n$ matrix such that $N^n = 0$ but $N^{n-1} \neq 0$ (where $n \geq 2$). Prove that N has no square root; that is, there is no matrix A such that $A^2 = N$.

26. **Dimension of the Commutant.** Let $C(A) = \{X \in M_n(\mathbb{C}) \mid XA = AX\}$ be the commutant of A.

    (a) Use the result of **Double Commutant** to show that if A is regular (i.e., admits a cyclic vector), then $\dim C(A) = n$.

    (b) Prove that in general, $\dim C(A) \geq n$.

    (c) Show that $\dim C(A) = n$ if and only if the characteristic and minimal polynomials of A are identical.

27. **Common Eigenvectors.** Let $\{A_i\}_{i \in I}$ be a family of pairwise commuting linear operators on a non-zero finite-dimensional vector

Proceed by induction on $\dim V$. Consider the eigenspace of one non-scalar operator A_k.

space V over an algebraically closed field. Prove that there exists a non-zero vector $v \in V$ that is an eigenvector for every A_i.

28. The Matrix Exponential. For $A \in M_n(\mathbb{C})$, defined $e^A = \sum_{k=0}^{\infty} A^k/k!$.

(a) Use the Jordan form to calculate $e^{J_n(\lambda)}$.

(b) Prove that $\det(e^A) = e^{\mathrm{tr}(A)}$.

(c) Prove that the map $\exp : M_n(\mathbb{C}) \to GL_n(\mathbb{C})$ is surjective.

(d) Is the map $\exp : M_n(\mathbb{R}) \to GL_n(\mathbb{R})$ surjective? (Consider a matrix with negative determinant).

29. Shift Operators and Infinite Dimensions. Let $V = \mathbb{C}^{\mathbb{N}}$ be the space of complex sequences. Define the right shift $R(x_0, x_1, \dots) = (0, x_0, x_1, \dots)$ and the left shift $L(x_0, x_1, \dots) = (x_1, x_2, \dots)$.

(a) Find the point spectrum (eigenvalues) of R and L.

(b) Show that $LR = \mathcal{E}$ but $RL \neq \mathcal{E}$.

(c) Contrast the spectral behaviour of these operators with the finite-dimensional Jordan blocks.

30. Square Roots of the Identity. Let $A \in M_n(\mathbb{C})$.

(a) If $A^2 = I$, prove that A is diagonalisable. What are the possible eigenvalues?

(b) If A is a 2×2 matrix such that $A^2 = 0$, must A be the zero matrix?

7
Symmetric Bilinear and Quadratic Forms

In this chapter, we restrict our attention to specific classes of bilinear forms that arise naturally in geometry and physics: those possessing symmetry properties. This leads to the definition of quadratic forms, which generalize the notion of length and energy, and the problem of finding bases that simplify these forms to sums of squares.

7.1 Symmetry and Skew-Symmetry

Definition 7.1. ***Symmetric and Skew-Symmetric Forms.***
Let V be a vector space over a field F. A bilinear form $f : V \times V \to F$ is called:
Symmetric if $f(x, y) = f(y, x)$ for all $x, y \in V$.
Skew-symmetric if $f(x, y) = -f(y, x)$ for all $x, y \in V$.

定義

Let F be the matrix of f relative to a fixed basis. The symmetry condition translates directly to matrix operations. Since $f(x, y)$ is a scalar, it equals its own transpose. If f satisfies $f(x, y) = \epsilon f(y, x)$ with $\epsilon = \pm 1$, then:

$$X^\top F Y = f(x, y) = \epsilon f(y, x) = \epsilon (Y^\top F X) = \epsilon (Y^\top F X)^\top = \epsilon X^\top F^\top Y.$$

This holds for all coordinate vectors X, Y, implying $F^\top = \epsilon F$. Thus, symmetric forms correspond to symmetric matrices ($F^\top = F$) and skew-symmetric forms to skew-symmetric matrices ($F^\top = -F$). Crucially, this property is intrinsic to the form and independent of the basis. If $F' = A^\top F A$ is the matrix in a new basis (*theorem* 4.9):

$$(F')^\top = (A^\top F A)^\top = A^\top F^\top A = A^\top (\epsilon F) A = \epsilon F'.$$

Assumption. *Characteristic Not 2.* Throughout this chapter, we assume that the characteristic of the field F is not 2.

定

This assumption is necessary to distinguish between symmetry and skew-symmetry. If $\operatorname{char} F = 2$, then $1 = -1$, and the conditions $f(x,y) = f(y,x)$ and $f(x,y) = -f(y,x)$ are identical. In this case, the important distinction is between **symmetric** forms and **alternating** forms (where $f(x,x) = 0$ for all x). Every alternating form is symmetric, but the converse fails if diagonal elements are non-zero.

Theorem 7.1. ***Decomposition of Bilinear Forms.***
The space of bilinear forms $\mathcal{L}_2(V,F)$ decomposes as the direct sum of the subspace of symmetric forms $\mathcal{L}_2^+$ and the subspace of skew-symmetric forms $\mathcal{L}_2^-$:

$$\mathcal{L}_2(V,F) = \mathcal{L}_2^+(V,F) \oplus \mathcal{L}_2^-(V,F).$$

定理

Proof

First, we show the sum is direct. Let $f \in \mathcal{L}_2^+ \cap \mathcal{L}_2^-$. Then for any x, y, we have $f(x,y) = f(y,x)$ and $f(x,y) = -f(y,x)$. Summing these gives $2f(x,y) = 0$. Since $\operatorname{char} F \neq 2$, we can divide by 2 to conclude $f(x,y) = 0$. Thus the intersection is trivial.
To show they span $\mathcal{L}_2$, let f be an arbitrary bilinear form. We construct:

$$f_s(x,y) = \frac{1}{2}(f(x,y) + f(y,x)) \quad \text{and} \quad f_a(x,y) = \frac{1}{2}(f(x,y) - f(y,x)).$$

It is routine to verify that f_s is symmetric, f_a is skew-symmetric, and $f = f_s + f_a$.

■

7.2 *Quadratic Forms*

Symmetric bilinear forms are intimately related to functions of a single vector variable known as quadratic forms.

Definition 7.2. ***Quadratic Form.***
A function $q : V \to F$ is called a **quadratic form** if:
1. $q(-v) = q(v)$ for all $v \in V$.
2. The map $f : V \times V \to F$ defined by the **polarization identity**:

$$f(x,y) = \frac{1}{2}(q(x+y) - q(x) - q(y))$$

is a bilinear form.
The bilinear form f defined in (2) is called the **polar form** of q. Note that f is automatically symmetric.

定義

Conversely, given any symmetric bilinear form f, we can define a

function $q_f(x) = f(x,x)$. This function satisfies $q_f(-x) = f(-x,-x) = (-1)^2 f(x,x) = q_f(x)$. Furthermore, expanding $f(x+y, x+y)$ yields:

$$f(x+y,x+y) = f(x,x)+f(x,y)+f(y,x)+f(y,y) = q_f(x)+2f(x,y)+q_f(y).$$

Rearranging this recovers the polarization identity. Thus, the correspondence is bijective.

Theorem 7.2. ***Bijection between Quadratic and Symmetric Bilinear Forms.***

Every quadratic form q is uniquely determined by its polar bilinear form f, specifically via $q(x) = f(x,x)$.

定理

Proof

From the definition of polarization, set $y = -x$:

$$f(x,-x) = \frac{1}{2}(q(0) - q(x) - q(-x)).$$

Using bilinearity, $f(x,-x) = -f(x,x)$. Using the property $q(-x) = q(x)$:

$$-f(x,x) = \frac{1}{2}q(0) - q(x).$$

Since f is bilinear, $f(0,0) = 0$, implying $q(0) = 0$. Thus $q(x) = f(x,x)$.

■

Notation 7.1. Matrix of a Quadratic Form Let $\mathcal{B} = (e_1, \ldots, e_n)$ be a basis for V. The matrix of the quadratic form q is defined to be the matrix $F = (f_{ij})$ of its polar form f. Explicitly:

$$f_{ij} = \frac{1}{2}(q(e_i + e_j) - q(e_i) - q(e_j)).$$

In coordinates, if x has column vector X, then $q(x) = X^\top F X = \sum_{i,j} f_{ij} x_i x_j$.

記法

The **rank** of q is defined as the rank of its matrix F. This is well-defined because matrix rank is invariant under congruence $F' = A^\top F A$ (*corollary 4.2*).
The kernel of the polar form f is often called the **radical** or the **null space** of q, denoted L_q:

$$L_q = \{u \in V \mid f(u,v) = 0 \quad \forall v \in V\}.$$

Using the relation between rank and nullity for bilinear forms, we have $\operatorname{rank} q = \dim V - \dim L_q$.

7.3 *Canonical Forms*

We seek a basis in which the expression for $q(x)$ is as simple as possible. Ideally, we wish to eliminate the "mixed" terms $x_i x_j$ $(i \neq j)$ in the polynomial expansion.

Definition 7.3. ***Canonical Basis.***
A basis $(e_1, \ldots, e_n)$ is called a **canonical basis** for q if the matrix of q in this basis is diagonal. In such a basis,

$$q(x) = \sum_{i=1}^{n} \lambda_i x_i^2, \quad \text{where } \lambda_i = f(e_i, e_i).$$

定義

Theorem 7.3. ***Existence of Canonical Basis.***
Every symmetric bilinear form (and thus every quadratic form) on a finite-dimensional space V admits a canonical basis.

定理

We proceed by induction on $n = \dim V$.

Base Case:

For $n = 1$, any basis vector e_1 yields a 1×1 matrix, which is diagonal.

証明終

Inductive Step:

Assume the result holds for dimension $n - 1$. If f is the zero form, any basis is canonical (with all $\lambda_i = 0$). If $f \neq 0$, there exists a vector e_1 such that $q(e_1) = f(e_1, e_1) \neq 0$. (If $f(x, x) = 0$ for all x, then by polarization $f(x, y) = 0$ for all x, y, contradicting $f \neq 0$).
Consider the linear functional $\phi : V \to F$ defined by $\phi(y) = f(e_1, y)$. Since $f(e_1, e_1) \neq 0$, ϕ is not the zero functional. Let $L = \operatorname{Ker} \phi = \{y \in V \mid f(e_1, y) = 0\}$. By the Rank-Nullity Theorem, $\dim L = n - 1$. We restrict f to the subspace L. By the inductive hypothesis, L has a canonical basis $(e_2, \ldots, e_n)$ such that $f(e_i, e_j) = 0$ for $i \neq j$ where $i, j \geq 2$.
By the definition of L, $f(e_1, e_i) = 0$ for all $i \geq 2$. Since f is symmetric, $f(e_i, e_1) = 0$ as well. It remains to show that $(e_1, \ldots, e_n)$ is a basis for V. Suppose $\alpha_1 e_1 + \cdots + \alpha_n e_n = 0$. Applying the functional ϕ:

$$\phi\left(\sum \alpha_i e_i\right) = \alpha_1 f(e_1, e_1) + \sum_{i=2}^{n} \alpha_i f(e_1, e_i) = \alpha_1 f(e_1, e_1) = 0.$$

Since $f(e_1, e_1) \neq 0$, we must have $\alpha_1 = 0$. The remaining relation $\sum_{i=2}^{n} \alpha_i e_i = 0$ lies in L, and since $(e_2, \ldots, e_n)$ is a basis for L, all $\alpha_i = 0$. Thus the matrix of f in this basis is diagonal.

証明終

Corollary 7.1. *Diagonal Congruence.* For any symmetric matrix F, there exists an invertible matrix A such that $A^\top FA$ is diagonal.

推論

Proof

Let f be the symmetric bilinear form on F^n defined by $f(x,y) = X^\top FY$ relative to the standard basis. By *theorem* 7.3, there exists a canonical basis $\mathcal{C}$ for F^n such that the matrix of f in this basis, denote it D, is diagonal. Let A be the transition matrix from the standard basis to the canonical basis $\mathcal{C}$. According to the change of basis formula for bilinear forms (see *theorem* 4.9), the matrix representation transforms as $D = A^\top FA$. Since D is diagonal and A is invertible (being a transition matrix), the result holds.

■

Lagrange's Method of Completing the Square

While the theorem above guarantees existence, Lagrange provided an algorithmic method to compute the canonical basis by explicitly eliminating mixed terms. Given $q(x) = \sum_{i,j} f_{ij}x_ix_j$, we distinguish two cases for the recursive step:

Case 1: Pivot Exists ***($f_{11} \neq 0$).*** We group all terms involving x_1:

$$q(x) = f_{11}x_1^2 + 2x_1\sum_{j=2}^{n} f_{1j}x_j + \sum_{i,j=2}^{n} f_{ij}x_ix_j.$$

We complete the square for x_1:

$$q(x) = \frac{1}{f_{11}}\left(f_{11}x_1 + \sum_{j=2}^{n} f_{1j}x_j\right)^2 + q'(x_2,\dots,x_n).$$

The term q' contains no x_1. We define a coordinate change:

$$x_1' = f_{11}x_1 + \sum_{j=2}^{n} f_{1j}x_j, \quad x_k' = x_k \text{ for } k \geq 2.$$

This transformation is invertible. We then repeat the process for q' on the remaining variables.

Case 2: No Diagonal Pivot ***($f_{ii} = 0$ for all i).*** If q is not zero, there must be a mixed term $f_{ij}x_ix_j \neq 0$. Without loss of generality, assume $f_{12} \neq 0$. We apply the coordinate change:

$$x_1 = u_1 + u_2, \quad x_2 = u_1 - u_2, \quad x_k = u_k \text{ for } k \geq 2.$$

The term $2f_{12}x_1x_2$ becomes $2f_{12}(u_1^2 - u_2^2)$. The new expression for q now has non-zero coefficients for u_1^2 and u_2^2, allowing us to proceed as in Case 1.

Example 7.1. Canonical Reduction. Let $q(x) = x_1x_2$ in $\mathbb{R}^2$. The matrix is

$$\begin{bmatrix} 0 & 1/2 \\ 1/2 & 0 \end{bmatrix}.$$

Since diagonal entries are zero, we use Case 2. Let $x_1 = y_1 + y_2, x_2 = y_1 - y_2$.

$$q = (y_1 + y_2)(y_1 - y_2) = y_1^2 - y_2^2.$$

In the basis corresponding to y, the matrix is $\operatorname{diag}(1, -1)$.

範例

7.4 *Real Quadratic Forms and Inertia*

While *theorem 7.3* guarantees that any quadratic form over a field F (with $\operatorname{char} F \neq 2$) can be diagonalised, the specific diagonal entries λ_i depend on the algebraic structure of F. Over the field of real numbers $\mathbb{R}$, we can scale the basis vectors to normalize the non-zero coefficients to ± 1. Specifically, if $\lambda_i > 0$, the substitution $e_i' = \lambda_i^{-1/2} e_i$ yields a coefficient of 1. If $\lambda_i < 0$, the substitution $e_i' = (-\lambda_i)^{-1/2} e_i$ yields -1.

Theorem 7.4. ***Standard Form over*** $\mathbb{R}$**.**
Let q be a quadratic form on a finite-dimensional real vector space V. There exists a basis in which q takes the **standard form**:

$$q(x) = \sum_{i=1}^{s} x_i^2 - \sum_{j=s+1}^{r} x_j^2,$$

where $r = \operatorname{rank} q$ and $0 \leq s \leq r$.

定理

Proof

By *theorem 7.3*, there exists a basis $(u_1, \ldots, u_n)$ in which the matrix of q is diagonal. That is,

$$q(x) = \sum_{i=1}^{n} \lambda_i y_i^2,$$

where y_i are the coordinates relative to this basis. We reorder the basis vectors so that the positive coefficients appear first, followed by the negative coefficients, and finally the zeros. Let s be the num-

ber of positive coefficients and $r - s$ be the number of negative coefficients. (The remaining $n - r$ coefficients are zero). We define a new basis $(e_1, \dots, e_n)$ by scaling the vectors u_i:

- If $\lambda_i > 0$ (for $1 \le i \le s$), let $e_i = \frac{1}{\sqrt{\lambda_i}} u_i$. Then $f(e_i, e_i) = \frac{1}{\lambda_i} f(u_i, u_i) = 1$.
- If $\lambda_i < 0$ (for $s < i \le r$), let $e_i = \frac{1}{\sqrt{-\lambda_i}} u_i$. Then $f(e_i, e_i) = \frac{1}{-\lambda_i} f(u_i, u_i) = -1$.
- If $\lambda_i = 0$ (for $r < i \le n$), let $e_i = u_i$. Then $f(e_i, e_i) = 0$.

In this normalised basis, the diagonal entries are exactly s ones, followed by $r - s$ minus ones, followed by zeros. Thus q takes the stated form.

■

While the canonical basis is not unique, the integers s and r are invariants of the quadratic form. The invariance of the rank r is a consequence of the invariance of matrix rank. The invariance of s, the number of positive squares, is the content of Sylvester's Law of Inertia.

Theorem 7.5. ***Sylvester's Law of Inertia.***
Let q be a real quadratic form. The number of positive coefficients s appearing in any diagonal representation of q is an invariant of q, called the **positive index of inertia**.

定理

Proof

Suppose there exist two bases $\mathcal{B} = (e_1, \dots, e_n)$ and $\mathcal{C} = (u_1, \dots, u_n)$ giving distinct standard forms:

$$q(x) = x_1^2 + \cdots + x_s^2 - x_{s+1}^2 - \cdots - x_r^2 \quad \text{(relative to } \mathcal{B}\text{)},$$
$$q(x) = y_1^2 + \cdots + y_t^2 - y_{t+1}^2 - \cdots - y_r^2 \quad \text{(relative to } \mathcal{C}\text{)}.$$

Assume for contradiction that $s \neq t$; without loss of generality, let $s > t$. Consider the subspaces defined by the "positive" part of the first decomposition and the "non-positive" part of the second:

$$U = \operatorname{span}(e_1, \dots, e_s) \quad \text{and} \quad W = \operatorname{span}(u_{t+1}, \dots, u_n).$$

For any non-zero vector $u \in U$, we have $q(u) > 0$. For any vector $w \in W$, $q(w) \le 0$. Thus $U \cap W = \{0\}$. We compute the dimensions: $\dim U = s$ and $\dim W = n - t$. Using the dimension formula for sums of subspaces:

$$\dim(U + W) = \dim U + \dim W - \dim(U \cap W) = s + (n - t) - 0 = n + (s - t).$$

Since $s > t$, this implies $\dim(U+W) > n = \dim V$, which is impossible. Thus $s = t$.

■

Definition 7.4. ***Signature.***
The **signature** of a real quadratic form is the pair $(s, r-s)$, representing the number of positive and negative terms in its standard form. Sometimes the signature is defined as the integer $s-(r-s) = 2s-r$. The integer $r-s$ is called the **negative index of inertia**.

定義

7.5 *Definiteness and Sylvester's Criterion*

In applications such as optimisation and mechanics, the sign of the values taken by a quadratic form is of critical importance.

Theorem 7.6. ***Spectral Theorem.***
Let A be a real symmetric matrix. Then all eigenvalues of A are real, and there exists an orthogonal basis of $\mathbb{R}^n$ consisting of eigenvectors of A. In particular, A is diagonalisable by an orthogonal matrix.

定理

Recall that a basis $\{v_i\}$ is **orthogonal** if $v_i \cdot v_j = 0$ for $i \neq j$. A matrix P is **orthogonal** if $P^\top P = I$. The proof of this fundamental result is developed in the exercises.

Definition 7.5. ***Definiteness.***
A non-degenerate real quadratic form q on V is called:
Positive definite if $q(x) > 0$ for all $x \neq 0$.
Negative definite if $q(x) < 0$ for all $x \neq 0$.
Indefinite if q takes both positive and negative values.
If degeneracy is allowed, we say q is **positive semi-definite** if $q(x) \geq 0$ for all x.

定義

In terms of the standard form invariants, q is positive definite if and only if $s = r = n$ (signature $(n,0)$), and positive semi-definite if $s = r \leq n$ (signature $(r,0)$). A symmetric matrix A is called positive definite if its associated form $q(x) = x^\top A x$ is positive definite.

Theorem 7.7. ***Factorisation of Positive Definite Matrices.***
A real symmetric matrix F is positive definite if and only if there exists a non-singular matrix A such that $F = A^\top A$.

定理

Proof

If F is positive definite, its canonical form is the identity matrix I. Thus F is congruent to I, meaning there is an invertible P such that $P^\top F P = I$, or $F = (P^{-1})^\top (P^{-1})$. Setting $A = P^{-1}$ yields the result.

Conversely, if $F = A^\top A$ with A invertible, then for any $x \neq 0$:

$$x^\top F x = x^\top A^\top A x = (Ax)^\top (Ax) = \|Ax\|^2.$$

Since A is invertible, $Ax \neq 0$, so $\|Ax\|^2 > 0$.

■

Example 7.2. Stability of Critical Points. Let $\phi : \mathbb{R}^2 \to \mathbb{R}$ be a smooth function. The local behaviour of ϕ near a critical point x_0 (where $\nabla\phi(x_0) = 0$) is determined by the Hessian matrix H of second derivatives. The Taylor expansion gives:

$$\phi(x) \approx \phi(x_0) + \frac{1}{2}(x - x_0)^\top H (x - x_0).$$

The term $q(v) = v^\top H v$ is a quadratic form.
- If H is positive definite, x_0 is a local minimum.
- If H is negative definite, x_0 is a local maximum.
- If H is indefinite, x_0 is a saddle point.

範例

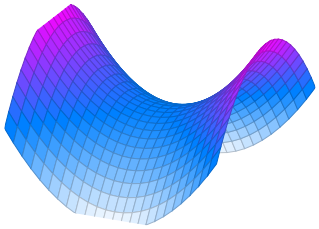


Figure 7.1: A saddle point corresponding to an indefinite quadratic form $q(x, y) = x^2 - y^2$.

Sylvester's Criterion

While eigenvalues provide a test for definiteness (all $\lambda_i > 0$ by the Spectral Theorem), computing them is non-trivial. Sylvester's criterion offers a determinant-based test.

Definition 7.6. ***Leading Principal Minors.***
Let $F = (f_{ij})$ be an $n \times n$ matrix. The **leading principal minors** are the determinants of the top-left $k \times k$ submatrices:

$$\Delta_k = \det \begin{bmatrix} f_{11} & \cdots & f_{1k} \\ \vdots & \ddots & \vdots \\ f_{k1} & \cdots & f_{kk} \end{bmatrix}, \quad k = 1, \ldots, n.$$

By convention, $\Delta_0 = 1$.

定義

Theorem 7.8. ***Jacobi's Method.***
Let q be a quadratic form with matrix F. If all leading principal minors Δ_k are non-zero, there exists a basis in which q has the diagonal form:

$$q(x) = \sum_{k=1}^{n} \frac{\Delta_{k-1}}{\Delta_k} y_k^2.$$

定理

Proof

We proceed by induction on n. Let $V_k = \operatorname{span}(e_1, \ldots, e_k)$. The restriction of q to V_{n-1} has matrix F_{n-1} (the top-left block) with minors $\Delta_1, \ldots, \Delta_{n-1}$. By hypothesis, these are non-zero. By the inductive hypothesis, there exists a basis $(u_1, \ldots, u_{n-1})$ for V_{n-1} diagonalising the restriction:

$$q|_{V_{n-1}}(u) = \sum_{k=1}^{n-1} \frac{\Delta_{k-1}}{\Delta_k} y_k^2.$$

This implies the polar form f satisfies $f(u_i, u_j) = 0$ for $i \neq j$ in this range. We seek a final basis vector u_n such that $f(u_i, u_n) = 0$ for all $i < n$. This requires u_n to be f-orthogonal to V_{n-1}. The conditions $f(e_i, u_n) = 0$ for $i = 1, \ldots, n-1$ form a system of $n-1$ linear equations. Since $\Delta_{n-1} \neq 0$, the non-degenerate restriction ensures a solution exists that is linearly independent of V_{n-1}. Let A be the change of basis matrix from the standard basis to $(u_1, \ldots, u_n)$. Since the new matrix F' is diagonal,

$$\det F' = \prod_{k=1}^{n} f(u_k, u_k).$$

Also $\det F' = \det(A^\top F A) = (\det A)^2 \Delta_n$. The product of the first $n-1$ diagonal entries is Δ_{n-1} (by applying the determinant relation to the restriction). Thus:

$$f(u_n, u_n) \cdot \Delta_{n-1} (\det A_{n-1})^2 \cdot (\text{scaling}) = \Delta_n (\det A)^2.$$

Proper normalisation of u_n ensures the term simplifies to $f(u_n, u_n) = \Delta_n / \Delta_{n-1}$.

■

Corollary 7.2. *Sylvester's Criterion.* A real quadratic form is positive definite if and only if all its leading principal minors are strictly positive:

$$\Delta_1 > 0, \quad \Delta_2 > 0, \quad \ldots, \quad \Delta_n > 0.$$

推論

Proof

If $\Delta_k > 0$ for all k, then the ratios Δ_k / Δ_{k-1} are all positive. By Jacobi's formula, the canonical coefficients are positive, so q is positive definite. Conversely, if q is positive definite, then restricted to any subspace V_k, it remains positive definite. The determinant of a positive definite matrix is positive (product of positive eigenvalues). Thus $\Delta_k = \det(F|_{V_k}) > 0$.

■

Structure of Skew-Symmetric Forms

We conclude this chapter by returning to skew-symmetric forms. Let f be a skew-symmetric bilinear form on V (so $f(x,y) = -f(y,x)$). As in the symmetric case, we can define the radical $V_0 = \operatorname{Ker} f$. We restrict our attention to the non-degenerate case where $V_0 = \{0\}$.

Theorem 7.9. ***Even Dimension of Non-Degenerate Forms.***
If V admits a non-degenerate skew-symmetric form f, then $\dim V$ must be even.

定理

Proof

Let F be the matrix of f. Then $F^\top = -F$.

$$\det F = \det(F^\top) = \det(-F) = (-1)^n \det F.$$

If n is odd, $\det F = -\det F$, implying $\det F = 0$ (since $\operatorname{char} F \neq 2$). This contradicts non-degeneracy. Thus n is even.

■

The canonical structure of such forms is built from 2-dimensional subspaces.

Definition 7.7. ***Hyperbolic Plane.***
A 2-dimensional subspace W equipped with a skew-symmetric form f is called a **hyperbolic plane** (or symplectic plane) if f is non-degenerate on W. It admits a basis (u,v) such that $f(u,v) = 1$. Since f is skew-symmetric, $f(u,u) = f(v,v) = 0$ and $f(v,u) = -1$. The matrix in this basis is $J_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ (or its transpose).

定義

Theorem 7.10. ***Symplectic Decomposition.***
Let V be a finite-dimensional space with a non-degenerate skew-symmetric form f. Then V decomposes into an orthogonal direct sum of hyperbolic planes:

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

where $\dim V = 2m$.

定理

Proof

We use induction on $\dim V$. Pick any non-zero vector e_1. Since f is non-degenerate, there exists e_2 such that $f(e_1,e_2) \neq 0$. Scale e_2 so that $f(e_1,e_2) = 1$. Let $W_1 = \operatorname{span}(e_1,e_2)$. The restriction of f to W_1 is non-degenerate (determinant 1). Let $W_1^\perp = \{x \in V \mid f(y,x) = 0 \quad \forall y \in W_1\}$. Since f is non-degenerate

on W_1, $V = W_1 \oplus W_1^\perp$. The restriction of f to $W_1^\perp$ remains non-degenerate. By the inductive hypothesis, $W_1^\perp$ decomposes into hyperbolic planes.

■

Corollary 7.3. *Canonical Form.* For any skew-symmetric matrix F, there exists an invertible matrix A such that

$$A^\top FA = \operatorname{diag}(J_2, \ldots, J_2, 0, \ldots, 0).$$

推論

Proof

Let f be the skew-symmetric bilinear form on F^n represented by the matrix F. By the Symplectic Decomposition Theorem, the space decomposes as an orthogonal direct sum of hyperbolic planes $W_1, \ldots, W_m$ and a radical subspace V_0:

$$F^n = W_1 \oplus \cdots \oplus W_m \oplus V_0.$$

For each hyperbolic plane W_k, there exists a basis (u_k, v_k) such that the restriction of f to W_k has matrix $J_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. For the radical V_0, the form is identically zero, so any basis yields a zero matrix. We construct a basis $\mathcal{B}$ for the entire space by concatenating these bases:

$$\mathcal{B} = (u_1, v_1, u_2, v_2, \ldots, u_m, v_m, z_1, \ldots, z_l),$$

where z_i form a basis for V_0. Since the decomposition is orthogonal, $f(x, y) = 0$ if x and y belong to distinct summands. Thus, the matrix of f relative to $\mathcal{B}$ is block diagonal, with m blocks of J_2 and zero blocks elsewhere. Let A be the transition matrix from the standard basis to $\mathcal{B}$. Then $A^\top FA$ is the matrix of f in the basis $\mathcal{B}$, which is the desired canonical form.

■

For a skew-symmetric matrix F of even dimension $2m$, the determinant is a perfect square of a polynomial in its entries. This polynomial is called the **Pfaffian**, denoted $\operatorname{Pf}(F)$. Specifically, $\det F = (\operatorname{Pf}(F))^2$. The Pfaffian satisfies the transformation property:

$$\operatorname{Pf}(A^\top FA) = \det(A) \operatorname{Pf}(F).$$

For the standard symplectic block matrix $J_{2m} = \operatorname{diag}(J_2, \ldots, J_2)$, we have $\operatorname{Pf}(J_{2m}) = 1$.

Example 7.3. Pfaffian in Dimension 4. For a 4×4 skew-symmetric

matrix $F = (f_{ij})$ with $f_{ji} = -f_{ij}$:

$$\mathrm{Pf}(F) = f_{12}f_{34} - f_{13}f_{24} + f_{14}f_{23}.$$

Squaring this expression yields $\det F$.

範例

7.6 *Exercises*

1. **Negative Definiteness via Minors.** Let F be the symmetric matrix of a real quadratic form q. Let $\Delta_1, \dots, \Delta_n = \det F$ be the leading principal minors of F. Prove that q and F are negative definite if and only if the signs of the minors alternate, starting with negative:
$$(-1)^k \Delta_k > 0 \quad \text{for all } k = 1, 2, \dots, n.$$
Consider the connection between F and $-F$ and apply Sylvester's criterion.

2. **Positive Entries vs. Positive Definite.** Give counter-examples to the following intuitions about positive definiteness:
   (a) A positive definite matrix $A = (a_{ij})$ such that some off-diagonal entry is negative ($a_{ij} < 0$).
   (b) A symmetric matrix $A = (a_{ij})$ with strictly positive entries ($a_{ij} > 0$ for all i, j) that is **not** positive definite.

3. **Parameterised Definiteness.** Find all values of $\lambda, \mu \in \mathbb{R}$ for which the following matrices are positive definite:
$$A = \begin{bmatrix} 1 & \lambda & \lambda \\ \lambda & 1 & \lambda \\ \lambda & \lambda & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 & \mu \\ 1 & \mu & 1 \\ \mu & 1 & 1 \end{bmatrix}.$$

4. **Composition of Forms.** Let $x = [x_1, x_2, x_3] \in \mathbb{C}^3$ and consider the cubic form $Q(x) = x_1^3 + x_2^3 + x_3^3 - 3x_1x_2x_3$. Let ε be a primitive cube root of unity.
   (a) Verify the factorisation:
   $$Q(x) = (x_1 + x_2 + x_3)(x_1 + \varepsilon x_2 + \varepsilon^2 x_3)(x_1 + \varepsilon^2 x_2 + \varepsilon x_3).$$
   (b) Prove the composition law: $Q(x)Q(y) = Q(z)$, where the components z_i are symmetric bilinear forms in x and y (i.e., $z_i = \sum_{j,k} a_{jk}^{(i)} x_j y_k$). Find the explicit expressions for z_1, z_2, z_3.

5. **Perturbation of Identity.** Let A be an arbitrary real symmetric matrix. Prove that there exists $\varepsilon_0 > 0$ such that for all $|\varepsilon| < \varepsilon_0$, the matrix $B = E + \varepsilon A$ is positive definite.
Consider the eigenvalues of B in relation to the eigenvalues of A.

6. **Rank and Signature Calculations.** For each of the following real quadratic forms, find the rank and signature using Lagrange's method (completing the square):
   (a) $q(x) = x_1^2 + 2x_1x_2 + 2x_2^2 + 4x_2x_3 + 5x_3^2$
   (b) $q(x) = x_1x_2 + x_2x_3 + x_3x_1$
   (c) $q(x) = \sum_{1 \le i < j \le 4} x_i x_j$
7. **Gram Matrices.** Let V be a real vector space equipped with a positive definite symmetric bilinear form $\langle \cdot, \cdot \rangle$ (an inner product). Let $v_1, \ldots, v_k$ be vectors in V. The Gram matrix is defined as $G_{ij} = \langle v_i, v_j \rangle$.
   (a) Prove that G is always positive semi-definite.
   (b) Prove that G is positive definite if and only if the vectors $v_1, \ldots, v_k$ are linearly independent.
8. **The Spectral Theorem.** In this exercise, we prove the Spectral Theorem for real symmetric matrices. Let A be a real symmetric matrix.
   (a) Prove that if $\lambda \in \mathbb{C}$ is an eigenvalue of A, then $\lambda \in \mathbb{R}$. (Hint: Consider $\bar{v}^\top A v$).
   (b) Prove that eigenvectors corresponding to distinct eigenvalues are orthogonal with respect to the standard dot product.
   (c) Let $f(x) = x^\top A x$ for $x \in \mathbb{R}^n$. Show that the maximum of $f(x)$ on the unit sphere S^{n-1} is attained at an eigenvector of A. (This establishes the existence of at least one real eigenvalue).
   (d) Let v be an eigenvector of A. Show that the subspace $W = v^\perp$ is invariant under A.
   (e) Use induction to conclude that there exists an orthogonal basis of $\mathbb{R}^n$ consisting of eigenvectors of A.
9. **Simultaneous Diagonalisation.** Let A and B be two real symmetric matrices.
   (a) Prove that if A is positive definite, there exists an invertible matrix P such that $P^\top A P = I$ and $P^\top B P$ is diagonal.
   (b) Give a counter-example to show that if neither matrix is positive definite, they may not be simultaneously diagonalisable by congruence, even if they are non-singular.

First reduce A to I, then apply the Spectral Theorem to the transformed B.

10. Pfaffian Identity. For the 4×4 skew-symmetric matrix:

$$F = \begin{bmatrix} 0 & a & b & c \\ -a & 0 & d & e \\ -b & -d & 0 & f \\ -c & -e & -f & 0 \end{bmatrix},$$

calculate the Pfaffian $\text{Pf}(F) = af - be + cd$ and verify explicitly that $(\text{Pf}(F))^2 = \det(F)$.