Algebra II: Determinants and Linear Transformation

Gudfit

Contents

0 Vector Kinematics 4

0.1 Parametric Curves and Motion 4

0.2 Vector Calculus 6

0.3 Dynamics and Constraints 7

1 Geometric Foundations of Determinants 10

1.1 Geometric Proofs via Area 10

1.2 Axiomatic Area and Orientation 12

1.3 The Vector Formulation of Area 13

1.4 Polygonal Areas and Volume 17

1.5 The Geometry of Volume in $\mathbb{R}^3$ 21

1.6 Exercises 25

2 General Theory of Determinants 28

2.1 The Determinant Axioms 28

2.2 Uniqueness and the Leibniz Formula 30

2.3 Computational Techniques via Multilinearity 31

2.4 Fundamental Properties of the Determinant 32

2.5 Parity and the Symmetric Group 37

2.6 Abstract Volume Forms 39

2.7 Orientation and Euclidean Volume 41

2.8 Cofactor Expansion 42

2.9 Algebraic Properties 45

2.10 The Adjugate Matrix 48

2.11 Applications of the Determinant 50

2.12 Exercises 52

3 Linear Transformations 56

3.1 Definition and Elementary Properties 56

3.2 The Matrix Representation Theorem 58

3.3 Geometric Properties of Linear Maps 60

3.4 Injectivity and Surjectivity 62

3.5 Composition and Invertibility 65

3.6 The Algebra of Linear Mappings 67

3.7 Elementary Geometric Transformations 69

3.8 Coordinates and Change of Basis 71

3.9 Applications: Affine Geometry 76

3.10 Exercises 78

4 *Eigenvalues and Eigenvectors* 81
- 4.1 *Motivation: Matrix Powers and Fibonacci* 81
- 4.2 *Definitions and Existence* 82
- 4.3 *Diagonalisation* 87
- 4.4 *Abstract Spectral Theory* 90
- 4.5 *Eigenspaces and Multiplicity* 91
- 4.6 *Invariant Subspaces* 93
- 4.7 *Exercises* 94

5 *Decoupling and Geometric Evolution* 98
- 5.1 *Spectral Properties* 98
- 5.2 *Generalized Eigenvectors and the Jordan Structure* 102
- 5.3 *Diagonalisation and Similarity* 106
- 5.4 *Complex Eigenvalues and Canonical Forms* 107
- 5.5 *Polynomial Methods and Annihilating Polynomials* 109
- 5.6 *Exercises* 112

0

Vector Kinematics

In the preceding notes, we established the algebraic and geometric foundations of vectors in $\mathbb{R}^n$. We kick off these notes by demonstrating some of its uses by applying it to the study of motion. By treating position, velocity, and acceleration as vector quantities, we can describe complex dynamical systems without the cumbersome coordinate-by-coordinate analysis required in elementary calculus.

0.1 *Parametric Curves and Motion*

To describe motion, we require a notion of time. We assume the existence of a time parameter $t \in \mathbb{R}$, representing a continuum of moments ordered from past to future. The position of a particle in space is strictly a function of this parameter.

Definition 0.1. *Trajectory*.
The *trajectory* or *path* of a particle in $\mathbb{R}^n$ is a vector-valued function $\mathbf{r} : I \to \mathbb{R}^n$, where $I \subseteq \mathbb{R}$ is an interval of time.

$$\mathbf{r}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix}.$$

The vector $\mathbf{r}(t)$ is referred to as the *position vector* relative to a chosen origin O.

定義

Remark.
This definition generalises the parametric equation of a line discussed in previous chapters. A line is simply a trajectory where the dependence on t is linear.

Uniform Motion

The simplest form of motion occurs when a particle moves along a straight line at a constant speed.

Definition 0.2. ***Uniform Motion.***
A particle undergoes *uniform motion* if its position vector satisfies:

$$\mathbf{r}(t) = \mathbf{r}_0 + t\mathbf{v},$$

where $\mathbf{r}_0$ is the initial position (at $t = 0$) and $\mathbf{v}$ is a constant vector known as the *velocity*.

定義

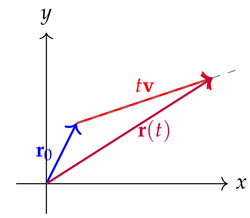

Figure 1: Uniform motion. The position $\mathbf{r}(t)$ is the vector sum of the initial position $\mathbf{r}_0$ and the displacement $t\mathbf{v}$.

Geometrically, this describes a line passing through $\mathbf{r}_0$ in the direction of $\mathbf{v}$. If we define the speed as the magnitude $\|\mathbf{v}\|$, then the distance travelled after time t is $\|t\mathbf{v}\| = |t|\|\mathbf{v}\|$, consistent with the scalar formula distance $=$ speed $\times$ time.

Relative Velocity

The vector nature of position allows us to define motion relative to different observers. Consider an observer at the origin O measuring the position of a particle P. The vector is $\boldsymbol{OP}$.
Now, suppose a second observer O' is observing the same particle P. This observer defines position relative to themselves, measuring $\boldsymbol{O'P}$.

Proposition 0.1. ***Composition of Position.***
Let O and O' be two reference points. For any point P:

$$\boldsymbol{OP} = \boldsymbol{OO'} + \boldsymbol{O'P}.$$

命題

Proof
This is an immediate consequence of the triangle law of vector addition.

■

See Chapter 3 of my *Geometry 1 notes.*

If the observer O' is moving with a constant velocity $\mathbf{u}$ relative to O, then $\boldsymbol{OO'} = \mathbf{r}_{O'} + t\mathbf{u}$. If the particle P is moving with velocity $\mathbf{v}$ relative to O, then $\boldsymbol{OP} = \mathbf{r}_P + t\mathbf{v}$. Substituting these into the composition law:

$$\mathbf{r}_P + t\mathbf{v} = (\mathbf{r}_{O'} + t\mathbf{u}) + \boldsymbol{O'P}.$$

Rearranging for the position relative to the moving observer:

$$\boldsymbol{O'P} = (\mathbf{r}_P - \mathbf{r}_{O'}) + t(\mathbf{v} - \mathbf{u}).$$

This implies that O' perceives the particle to be moving with velocity $\mathbf{v} - \mathbf{u}$.

Theorem 0.1. ***Relative Velocity.***
If a particle P moves with velocity $\mathbf{v}$ and an observer O' moves with velocity $\mathbf{u}$ (both relative to a fixed frame), the velocity of P relative to O' is:

$$\mathbf{v}_{\text{rel}} = \mathbf{v} - \mathbf{u}.$$

定理

Example 0.1. Galilean Relativity. Consider the motion of the Sun relative to the Earth. Let the Sun be fixed ($\mathbf{v}_S = \mathbf{0}$) and the Earth move with velocity $\mathbf{v}_E$.

1. Velocity of Sun relative to Earth: $\mathbf{v}_{S/E} = \mathbf{v}_S - \mathbf{v}_E = -\mathbf{v}_E$.
2. Velocity of Earth relative to Sun: $\mathbf{v}_{E/S} = \mathbf{v}_E - \mathbf{v}_S = \mathbf{v}_E$.

Historically, the debate over whether the Earth or Sun moved was effectively a debate over the choice of the origin $\mathbf{0}$. The relative velocity vector $\mathbf{v}_{rel}$ is invariant (up to sign) regardless of the choice of frame.

範例

Remark (Relativistic Disclaimer).

The addition of velocities $\mathbf{v} - \mathbf{u}$ holds only for speeds significantly lower than the speed of light c. As $v \to c$, the Galilean transformation must be replaced by the Lorentz transformation of Special Relativity.

For the purposes of this chapter, we assume $\|\mathbf{v}\| \ll c$.

0.2 *Vector Calculus*

To describe non-uniform motion, where velocity changes direction or magnitude, we must extend the calculus of differentiation to vector-valued functions.

Definition 0.3. ***Vector Derivative.***
Let $\mathbf{r}(t)$ be a vector function. The derivative $\mathbf{r}'(t)$ or $\frac{d\mathbf{r}}{dt}$ is defined by the limit:

$$\frac{d\mathbf{r}}{dt} = \lim_{\Delta t \to 0} \frac{\mathbf{r}(t + \Delta t) - \mathbf{r}(t)}{\Delta t}.$$

Provided the limit exists, $\mathbf{r}(t)$ is said to be differentiable.

定義

Since vector limits are computed component-wise, the derivative is simply the vector of scalar derivatives:

$$\mathbf{r}'(t) = \begin{bmatrix} x_1'(t) \\ \vdots \\ x_n'(t) \end{bmatrix}.$$

Instantaneous Velocity and Acceleration

We define the kinematic quantities in terms of these derivatives.

Definition 0.4. ***Velocity and Acceleration.***
Let $\mathbf{r}(t)$ be the position of a particle.

1. The *instantaneous velocity* $\mathbf{v}(t)$ is the first derivative of position:

$$\mathbf{v}(t) = \frac{d\mathbf{r}}{dt}.$$

   The direction of $\mathbf{v}(t)$ is tangent to the trajectory at $\mathbf{r}(t)$. The magnitude $\|\mathbf{v}(t)\|$ is the speed.

2. The *acceleration* $\mathbf{a}(t)$ is the first derivative of velocity (and the second derivative of position):

$$\mathbf{a}(t) = \frac{d\mathbf{v}}{dt} = \frac{d^2\mathbf{r}}{dt^2}.$$

定義

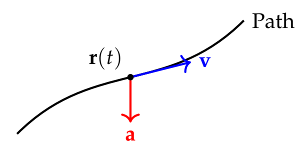

Figure 2: The velocity vector **v** is always tangent to the path. The acceleration **a** points "inside" the curve, reflecting the change in velocity.

Theorem 0.2. ***Differentiation Rules.***
For differentiable vector functions $\mathbf{u}(t), \mathbf{v}(t)$ and scalar function $f(t)$:

1. $\frac{d}{dt}(\mathbf{u} + \mathbf{v}) = \mathbf{u}' + \mathbf{v}'$.
2. $\frac{d}{dt}(f\mathbf{u}) = f'\mathbf{u} + f\mathbf{u}'$.
3. $\frac{d}{dt}(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u}' \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{v}'$ (Dot Product Rule).

定理

Proof

These follow directly from the definition of the derivative and the standard properties of limits and scalar derivatives applied component-wise.

■

0.3 *Dynamics and Constraints*

Newton's Second Law provides the link between the geometry of motion and the physical causes of motion (forces). It states that for a particle of constant mass m, the total force $\mathbf{F}$ acting on it determines the acceleration:

$$\mathbf{F} = m\mathbf{a} = m\frac{d^2\mathbf{r}}{dt^2}.$$

This is a second-order vector differential equation. To solve for the motion $\mathbf{r}(t)$, we require the force function and two initial conditions: position $\mathbf{r}_0$ and velocity $\mathbf{v}_0$.

Example 0.2. Projectile Motion. Consider a particle near the surface of the Earth subject only to constant gravity. We orient the y-axis vertically upwards. The acceleration is $\mathbf{a} = (0, -g)$, where g is the gravitational constant. Integrating $\mathbf{a}(t)$ with respect to t:

$$\mathbf{v}(t) = \int \mathbf{a}\, dt = \begin{bmatrix} 0 \\ -gt \end{bmatrix} + \mathbf{v}_0,$$

where $\mathbf{v}_0 = (u_x, u_y)$ is the initial velocity. Integrating again yields position:

$$\mathbf{r}(t) = \int \mathbf{v}(t)\, dt = \begin{bmatrix} 0 \\ -\frac{1}{2}gt^2 \end{bmatrix} + t\mathbf{v}_0 + \mathbf{r}_0.$$

In components (assuming $\mathbf{r}_0 = \mathbf{0}$):

$$x(t) = u_x t, \quad y(t) = u_y t - \frac{1}{2}gt^2.$$

Eliminating t (via $t = x/u_x$) reveals the trajectory is parabolic:

$$y = \frac{u_y}{u_x}x - \frac{g}{2u_x^2}x^2.$$

範例

Constrained Motion

Often, a particle is constrained to move on a specific surface or curve, such as a sphere or a circle. This constraint imposes a geometric condition on the velocity and acceleration vectors.

Theorem 0.3. ***Motion on a Sphere.***
A particle moves on the surface of a sphere centered at the origin (i.e., $\|\mathbf{r}(t)\| = \text{constant}$) if and only if its velocity is everywhere orthogonal to its position vector.

$$\mathbf{r}(t) \cdot \mathbf{v}(t) = 0.$$

定理

Proof
Let c be the radius of the sphere. The condition that the particle stays on the sphere is:

$$\|\mathbf{r}(t)\|^2 = \mathbf{r}(t) \cdot \mathbf{r}(t) = c^2.$$

Differentiating both sides with respect to t:

$$\frac{d}{dt}(\mathbf{r} \cdot \mathbf{r}) = \frac{d}{dt}(c^2).$$

Using the product rule for dot products:

$$\mathbf{r}' \cdot \mathbf{r} + \mathbf{r} \cdot \mathbf{r}' = 0 \implies 2\mathbf{v} \cdot \mathbf{r} = 0.$$

Thus, $\mathbf{v} \cdot \mathbf{r} = 0$. ■

The steps are reversible, so the converse holds (integration yields $\|\mathbf{r}\|^2 = \text{const}$).

Example 0.3. Circular Kinematics. Consider a particle moving in the plane $\mathbb{R}^2$ defined by:

$$x(t) = 3t^2, \quad y(t) = 2t^3.$$

We calculate the kinematic quantities:
- **Position:** $\mathbf{r}(t) = (3t^2, 2t^3)$.
- **Velocity:** $\mathbf{v}(t) = \mathbf{r}'(t) = (6t, 6t^2) = 6t(1, t)$.
- **Speed:** $\|\mathbf{v}(t)\| = \sqrt{(6t)^2 + (6t^2)^2} = 6|t|\sqrt{1+t^2}$.
- **Acceleration:** $\mathbf{a}(t) = \mathbf{v}'(t) = (6, 12t)$.

We can determine when the particle reaches a specific speed, say 12:

$$6|t|\sqrt{1+t^2} = 12 \implies |t|\sqrt{1+t^2} = 2.$$

Squaring both sides: $t^2(1+t^2) = 4 \implies (t^2)^2 + t^2 - 4 = 0$. Solving this quadratic in t^2:

$$t^2 = \frac{-1 \pm \sqrt{1+16}}{2} = \frac{-1+\sqrt{17}}{2}$$

(taking the positive root since $t^2 \geq 0$). Thus $t = \pm\sqrt{\frac{\sqrt{17}-1}{2}}$.

範例

We will learn more about this later in the mechanics notes.

1

Geometric Foundations of Determinants

In the previous volume of this text, we focused primarily on the algebraic utility of matrices: representing linear maps, solving systems of the form $A\mathbf{x} = \mathbf{b}$, and computing inverses. We now return to the geometric interpretation of these objects.

See *Algebra I: Matrices and Applications* for the algebraic treatment of linear systems.

Specifically, as hinted at in the foundational geometry notes, determinants are inextricably linked to the concepts of area and volume. It is crucial to observe that area and volume can be formalised independently of length (the metric axioms). One may deform a shape thereby changing the edge lengths and internal angles, yet preserving the area. In this sense, the concept of volume is more primitive than that of the inner product or distance.

Orientation

The algebraic formulation of area necessitates the concept of *orientation*. When calculating the area of a parallelogram spanned by two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^2$, the result of a determinant calculation is a *signed* value. The sign indicates whether the pair $(\mathbf{u}, \mathbf{v})$ is positively oriented (typically counter-clockwise) or negatively oriented (clockwise).

Thus, algebraic simplicity and generality imply geometric subtlety. We are forced to distinguish between "clockwise" and "counter-clockwise" in the plane if we wish to deal with areas algebraically. This leads to a somewhat surprising realisation: area and volume are, in a specific algebraic sense, more basic than length.

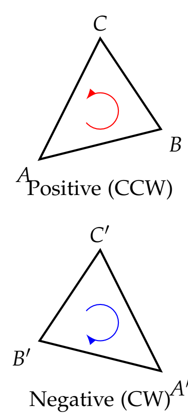

Figure 1.1: Orientation in the plane. Triangle ABC is oriented counter-clockwise, while $A'B'C'$ (where the traversal order is reversed) is clockwise.

1.1 *Geometric Proofs via Area*

The notions of area and volume are primitive and intuitive. As soon as we compare which container holds more or which field requires more seed, we are invoking these concepts. We now offer some classical examples to illustrate how area arguments can prove geometric theorems, relying on the "Sum of Parts Principle": the area of a union of disjoint sets is the sum of their individual areas.

Remark.

The "Sum of Parts Principle" often functions as a slogan rather than a axiom in elementary texts. While intuitively obvious, it requires careful definition in measure theory. We shall use it here in the classical geometric sense.

Example 1.1. Pythagoras via Rearrangement. Consider a square of side length $a+b$. We can partition this square in two distinct ways.

1. **Configuration A:** Place four identical right-angled triangles (with legs a, b and hypotenuse c) in the corners. The remaining uncovered area is a central square of side c.

2. **Configuration B:** Rearrange the four triangles into two rectangles of dimension $a \times b$. The remaining area consists of two smaller squares: one of side a and one of side b.

Since the total area of the large square is invariant, and the area occupied by the four triangles is constant, the remaining areas must be equal.

$$\text{Area(Central Square)} = \text{Area(Square } a) + \text{Area(Square } b) \implies c^2 = a^2 + b^2.$$

範例

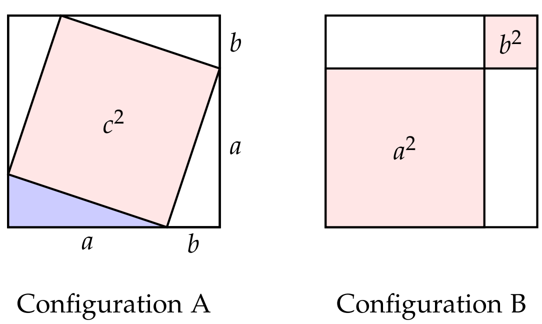

Figure 1.2: Visual proof of Pythagoras' Theorem. The red area in A (c^2) must equal the red area in B ($a^2 + b^2$) because the four triangles (white regions) are identical in both figures.

We can also express this algebraically without explicit rearrangement, simply by summing the parts of Configuration A.

Example 1.2. Pythagoras via Algebraic Area. Using the left side of *Figure 1.2*, the area of the large outer square can be computed directly as $(a+b)^2$. Alternatively, it is the sum of the inner square and the four triangles.

$$\begin{aligned}(a+b)^2 &= c^2 + 4\left(\frac{1}{2}ab\right)\\ a^2+2ab+b^2 &= c^2+2ab\\ a^2+b^2 &= c^2.\end{aligned}$$

This method relies on the algebraic expansion of area, linking the geometric decomposition to polynomial arithmetic.

範例

A final example demonstrates how area relates linear dimensions within a single figure.

Example 1.3. Altitudes of a Triangle. Consider a triangle ABC with side lengths a, b, c and corresponding altitudes h_a, h_b, h_c. We can compute the area of the triangle in two ways:

$$\text{Area} = \frac{1}{2}ah_a \quad \text{and} \quad \text{Area} = \frac{1}{2}bh_b.$$

Equating these expressions yields:

$$\frac{1}{2}ah_a = \frac{1}{2}bh_b \implies ah_a = bh_b \implies \frac{h_a}{h_b} = \frac{b}{a}.$$

This result states that the altitudes are inversely proportional to their corresponding bases. While this can be proven using similar triangles, the area argument is immediate and avoids the construction of auxiliary angles.

範例

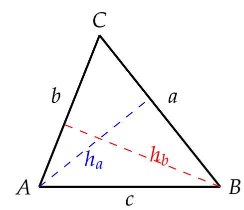

Figure 1.3: The altitudes of a triangle. The length of the altitude is inversely proportional to the side it intersects.

1.2 *Axiomatic Area and Orientation*

Note

Before formalising, we clarify some notation. From now on, "Area" will always mean signed area, denoted either by $\mathcal{A}(\cdot)$ for triangles or by the outer product $[\cdot, \cdot]$ for parallelograms. We will also freely identify a point with its position vector from a fixed origin O, writing A both for the point and for $\boldsymbol{OA}$.

The intuitive "Sum of Parts Principle" used in the previous section encounters a critical limitation when formalised algebraically. Consider a triangle ABC and a point P. If P lies inside the triangle, we intuitively accept that:

$$\mathcal{A}(ABC) = \mathcal{A}(PAB) + \mathcal{A}(PBC) + \mathcal{A}(PCA).$$

However, if P lies outside the triangle, this equality fails for positive area. To maintain such additive laws universally, irrespective of the point's location, we must assign a *sign* to the area, much as we assign a sign to coordinates on a line to satisfy $AC = AB + BC$ regardless of the order of points A, B, C.

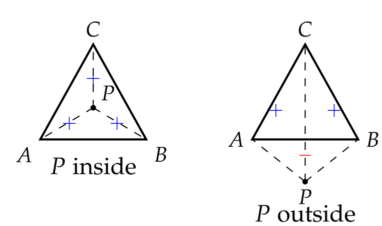

Figure 1.4: When P lies inside ABC, all sub-triangles share the same orientation. When P lies outside, triangle PAB has opposite orientation, requiring signed area for identity to hold

> **Definition 1.1.** ***Signed Area.***
> We postulate the existence of a signed area function $\mathcal{A}(A, B, C)$ for any three points A, B, C in the plane, satisfying the following axioms:
>
> 1. **Skew-Symmetry:** Permuting two vertices reverses the sign.
>
> $$\mathcal{A}(ABC) = -\mathcal{A}(BAC) = -\mathcal{A}(ACB) = -\mathcal{A}(CBA).$$
>
> Cyclic permutations preserve the sign:
>
> $$\mathcal{A}(ABC) = \mathcal{A}(BCA) = \mathcal{A}(CAB).$$
>
> Consequently, if two vertices coincide, the area is zero (e.g., $\mathcal{A}(AAB) = 0$).
>
> 2. **Additivity:** For any four points A, B, C, P:
>
> $$\mathcal{A}(ABC) = \mathcal{A}(PAB) + \mathcal{A}(PBC) + \mathcal{A}(PCA).$$
>
> 3. **Normalisation:** There exists at least one non-degenerate triangle (vertices are not collinear) with non-zero area.
>
> 定義

The orientation is customarily defined such that $\mathcal{A}(ABC)$ is positive if the traversal $A \to B \to C$ is counter-clockwise, and negative if clockwise. This convention aligns with the "right-hand rule" in mechanics.

1.3 *The Vector Formulation of Area*

While the classical geometric approach relying on cutting and pasting (dissection) provides intuitive proofs for theorems like Pythagoras', it is cumbersome for general calculation and lacks the algebraic structure necessary for higher dimensions. We therefore adopt a vector-theoretic approach. Instead of treating area as a property of a static shape, we view it as a function of the vectors that generate the shape.

The Outer Product

We consider the area of a parallelogram defined by two vectors $\mathbf{u}$ and $\mathbf{v}$ in a 2-dimensional vector space V (conceptually $\mathbb{R}^2$). We denote this oriented area by the bracket notation $[\mathbf{u}, \mathbf{v}]$. We postulate three fundamental properties that this function must satisfy.

> **Definition 1.2.** ***Area Axioms (Outer Product).***
> Let V be a 2-dimensional vector space. A function $[\cdot, \cdot] : V \times V \to \mathbb{R}$ is called an *outer product* (or an area function) if it satisfies the following axioms for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $k \in \mathbb{R}$: (We use "outer product"

here in the geometric sense of an oriented area function, not in the matrix sense $\mathbf{u}\mathbf{v}^\top$.)

1. **Skew-Symmetry:** $[\mathbf{u}, \mathbf{v}] = -[\mathbf{v}, \mathbf{u}]$.
2. **Linearity in the Second Argument:**

$$[\mathbf{u}, \mathbf{v} + k\mathbf{w}] = [\mathbf{u}, \mathbf{v}] + k[\mathbf{u}, \mathbf{w}].$$

3. **Non-Degeneracy:** $[\mathbf{u}, \mathbf{v}] = 0$ if and only if $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent.

定義

Remark.

These axioms encode our geometric intuition:

- *Skew-Symmetry* captures orientation. Swapping the order of vectors effectively "flips" the parallelogram over, reversing its signed area.
- *Linearity* captures the effect of scaling and shearing. Scaling one side of a parallelogram scales its area. Adding a multiple of $\mathbf{u}$ to $\mathbf{v}$ (a shear operation parallel to the base $\mathbf{u}$) preserves the height and thus the area.
- *Non-Degeneracy* ensures that "flat" parallelograms have zero area and "open" ones have non-zero area.

An immediate consequence of skew-symmetry is that the "square" of a vector vanishes.

Proposition 1.1. ***Vanishing Diagonal.***
For any vector $\mathbf{u} \in V$, $[\mathbf{u}, \mathbf{u}] = 0$.

命題

Proof

By skew-symmetry, $[\mathbf{u}, \mathbf{u}] = -[\mathbf{u}, \mathbf{u}]$. Adding $[\mathbf{u}, \mathbf{u}]$ to both sides gives $2[\mathbf{u}, \mathbf{u}] = 0$, implying $[\mathbf{u}, \mathbf{u}] = 0$.

■

This seemingly simple property is powerful. It allows us to perform "algebra" with these brackets where repeated factors annihilate the term.

Note

Linearity in the first argument follows from skew-symmetry and linearity in the second:

$$[\mathbf{u} + \mathbf{v}, \mathbf{w}] = -[\mathbf{w}, \mathbf{u} + \mathbf{v}] = -([\mathbf{w}, \mathbf{u}] + [\mathbf{w}, \mathbf{v}]) = [\mathbf{u}, \mathbf{w}] + [\mathbf{v}, \mathbf{w}].$$

Thus, the outer product is a *bilinear form*.

We see more of this later.

Derivation of the Determinant

One might ask if such a function even exists, or if there are too many. It turns out that once we fix a basis (a unit of measurement), the area function is unique up to scaling. This leads us directly to the determinant.

Theorem 1.1. ***Uniqueness of the Area Function.***
Let $\{\mathbf{e}_1, \mathbf{e}_2\}$ be a basis for V. For any two vectors $\mathbf{u} = x_1\mathbf{e}_1 + y_1\mathbf{e}_2$ and $\mathbf{v} = x_2\mathbf{e}_1 + y_2\mathbf{e}_2$, the outer product is given by:

$$[\mathbf{u}, \mathbf{v}] = (x_1y_2 - x_2y_1)[\mathbf{e}_1, \mathbf{e}_2].$$

定理

Proof

We expand the product using bilinearity:

$$\begin{aligned}[\mathbf{u}, \mathbf{v}] &= [x_1\mathbf{e}_1 + y_1\mathbf{e}_2, x_2\mathbf{e}_1 + y_2\mathbf{e}_2] \\ &= x_1[\mathbf{e}_1, x_2\mathbf{e}_1 + y_2\mathbf{e}_2] + y_1[\mathbf{e}_2, x_2\mathbf{e}_1 + y_2\mathbf{e}_2] \\ &= x_1x_2[\mathbf{e}_1, \mathbf{e}_1] + x_1y_2[\mathbf{e}_1, \mathbf{e}_2] + y_1x_2[\mathbf{e}_2, \mathbf{e}_1] + y_1y_2[\mathbf{e}_2, \mathbf{e}_2].\end{aligned}$$

Using the property that $[\mathbf{e}_i, \mathbf{e}_i] = 0$ and $[\mathbf{e}_2, \mathbf{e}_1] = -[\mathbf{e}_1, \mathbf{e}_2]$, this simplifies to:

$$[\mathbf{u}, \mathbf{v}] = x_1y_2[\mathbf{e}_1, \mathbf{e}_2] - y_1x_2[\mathbf{e}_1, \mathbf{e}_2] = (x_1y_2 - x_2y_1)[\mathbf{e}_1, \mathbf{e}_2].$$

■

We therefore define the determinant in dimension 2 by

$$\det \begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \end{bmatrix} := x_1y_2 - x_2y_1,$$

which is exactly the scalar giving the oriented area relative to the unit square.

This theorem establishes a profound connection: **Determinants are the unique scalars required to satisfy the geometric axioms of area.**

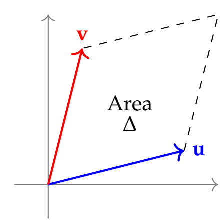

Figure 1.5: The area $\Delta = [\mathbf{u}, \mathbf{v}]$ is scaled relative to the unit square $[\mathbf{e}_1, \mathbf{e}_2]$.

Standardisation

In standard Euclidean space $\mathbb{R}^2$, we choose the standard basis $\mathbf{e}_1 = [1, 0]^T$ and $\mathbf{e}_2 = [0, 1]^T$. We normalise the area measure by declaring the area of the unit square to be 1:

$$[\mathbf{e}_1, \mathbf{e}_2] = 1.$$

With this normalisation, the area function becomes exactly the determinant:

$$\left[\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}, \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}\right] = \det \begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \end{bmatrix}.$$

Example 1.4. Shearing Preserves Area. Consider the rectangle defined by $\mathbf{u} = (2,0)^T$ and $\mathbf{v} = (0,3)^T$. Its area is:

$$[\mathbf{u}, \mathbf{v}] = \det \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} = 6.$$

Now, apply a shear operation: let $\mathbf{v}' = \mathbf{v} + 2\mathbf{u} = (0,3)^T + (4,0)^T = (4,3)^T$. The new area is:

$$[\mathbf{u}, \mathbf{v}'] = \det \begin{bmatrix} 2 & 4 \\ 0 & 3 \end{bmatrix} = 6.$$

This algebraic result matches the geometric fact that shearing a rectangle into a parallelogram does not change its area.

範例

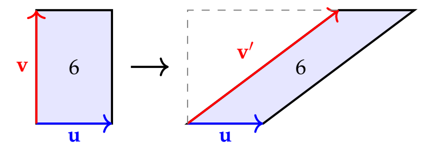

Figure 1.6: Shearing preserves area. Replacing $\mathbf{v}$ with $\mathbf{v}' = \mathbf{v} + 2\mathbf{u}$ transforms the rectangle into a parallelogram of equal area.

Barycentric Coordinates as Area Ratios

We can apply this vector formulation to revisit the geometry of the triangle. If we have a triangle ABC and a point P inside it, we can express P as a linear combination of the vertices using weights that sum to 1. These weights, known as barycentric coordinates, turn out to be ratios of oriented areas.

Proposition 1.2. ***Barycentric Area Formula.***
Let A, B, C be vertices of a non-degenerate triangle. Any point P in the plane of ABC can be written as:

$$P = \alpha A + \beta B + \gamma C, \quad \text{with } \alpha + \beta + \gamma = 1.$$

The coefficients are given by the ratio of signed areas:

$$\alpha = \frac{\mathcal{A}(P,B,C)}{\mathcal{A}(A,B,C)}, \quad \beta = \frac{\mathcal{A}(A,P,C)}{\mathcal{A}(A,B,C)}, \quad \gamma = \frac{\mathcal{A}(A,B,P)}{\mathcal{A}(A,B,C)}.$$

命題

Proof

Fix origin O and represent A, B, C, P by vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{p}$. From $P = \alpha A + \beta B + \gamma C$ and $\alpha + \beta + \gamma = 1$, we have:

$$\mathbf{p} = \alpha \mathbf{a} + \beta \mathbf{b} + \gamma \mathbf{c}, \quad \beta + \gamma = 1 - \alpha.$$

Therefore:

$$\mathbf{p} - \mathbf{b} = \alpha(\mathbf{a} - \mathbf{b}) + \gamma(\mathbf{c} - \mathbf{b}).$$

We compute $\mathcal{A}(PBC)$ using the outer product:

$$\begin{aligned}\mathcal{A}(PBC) &= \frac{1}{2}[\mathbf{p}-\mathbf{b},\mathbf{c}-\mathbf{b}] \\ &= \frac{1}{2}\left[\alpha(\mathbf{a}-\mathbf{b})+\gamma(\mathbf{c}-\mathbf{b}),\mathbf{c}-\mathbf{b}\right] \\ &= \frac{1}{2}\left(\alpha[\mathbf{a}-\mathbf{b},\mathbf{c}-\mathbf{b}]+\gamma[\mathbf{c}-\mathbf{b},\mathbf{c}-\mathbf{b}]\right).\end{aligned}$$

The second term vanishes since the outer product of a vector with itself is zero. The first term simplifies to $\alpha\mathcal{A}(ABC)$. Thus, $\alpha = \mathcal{A}(PBC)/\mathcal{A}(ABC)$, and similarly for β and γ.

■

Remark.

This interpretation explains why these coordinates are sometimes called *areal coordinates*. It provides a coordinate-free method for locating points relative to a simplex.

1.4 *Polygonal Areas and Volume*

Having established the algebraic properties of the outer product and its relation to the determinant in *theorem* 1.1, we now apply this formalism to general geometric figures. The signed area function allows us to define the area of arbitrary polygons and, by extension, volumes in higher dimensions, without resorting to geometric decomposition (cutting and pasting) for every case.

The Area of a Triangle

We previously defined the signed area of a triangle ABC implicitly through the additivity axiom. We now give a constructive definition using the outer product.

Definition 1.3. ***Vector Area of a Triangle.***
Let A, B, C be points in the plane. The signed area of the triangle $\triangle ABC$ is defined as:

$$\mathcal{A}(ABC) = \frac{1}{2}[\boldsymbol{AB},\boldsymbol{AC}].$$

定義

This definition satisfies the axioms of *Signed Area*. Skew-symmetry follows immediately from the properties of the outer product. We now derive a critical formula expressing this area in terms of position vectors relative to an arbitrary origin O.

Proposition 1.3. ***Origin Expansion Formula.***
Let O be any point in the plane, and let $\mathbf{a}, \mathbf{b}, \mathbf{c}$ be the position vectors of A, B, C relative to O. Then:

$$\mathcal{A}(ABC) = \mathcal{A}(OAB) + \mathcal{A}(OBC) + \mathcal{A}(OCA).$$

命題

Proof
By definition, $\boldsymbol{AB} = \mathbf{b} - \mathbf{a}$ and $\boldsymbol{AC} = \mathbf{c} - \mathbf{a}$. Substituting into the area definition and using the bilinearity of the outer product:

$$\begin{aligned} 2\mathcal{A}(ABC) &= [\mathbf{b} - \mathbf{a}, \mathbf{c} - \mathbf{a}] \\ &= [\mathbf{b}, \mathbf{c}] - [\mathbf{b}, \mathbf{a}] - [\mathbf{a}, \mathbf{c}] + [\mathbf{a}, \mathbf{a}]. \end{aligned}$$

Recall that $[\mathbf{a}, \mathbf{a}] = 0$ and $[\mathbf{u}, \mathbf{v}] = -[\mathbf{v}, \mathbf{u}]$. Thus:

$$2\mathcal{A}(ABC) = [\mathbf{a}, \mathbf{b}] + [\mathbf{b}, \mathbf{c}] + [\mathbf{c}, \mathbf{a}].$$

Dividing by 2, we recognise the terms as $\mathcal{A}(OAB)$, $\mathcal{A}(OBC)$, and $\mathcal{A}(OCA)$ respectively.

■

This result provides the "natural" expression for area in terms of position vectors. It implies that the area of a triangle is the sum of the signed areas swept out by the position vector as it traverses the boundary $A \to B \to C \to A$.

General Polygons

The Origin Expansion Formula suggests a method for defining the area of any polygon by summing the areas of triangles formed with the origin.

Definition 1.4. ***Polygon.***
A polygon σ is defined by a finite sequence of vertices $P_1, P_2, \ldots, P_n$ in the plane. The boundary of σ, denoted $\partial\sigma$, consists of the oriented line segments $P_1P_2, P_2P_3, \ldots, P_nP_1$.

定義

Definition 1.5. ***Area of a Polygon.***
The signed area of a polygon $\sigma = (P_1, \ldots, P_n)$ is defined by:

$$\text{Area}(\sigma) = \sum_{i=1}^{n} \mathcal{A}(O, P_i, P_{i+1}),$$

where $P_{n+1} = P_1$ and O is any point in the plane.

定義

For this definition to be mathematically sound, it must be independent of the choice of origin O.

Theorem 1.2. *Independence of Origin.*
The value of $\text{Area}(\sigma)$ is invariant under the change of origin. That is, for any two points O and Q:

$$\sum_{i=1}^{n} \mathcal{A}(O, P_i, P_{i+1}) = \sum_{i=1}^{n} \mathcal{A}(Q, P_i, P_{i+1}).$$

定理

Proof

Using the *proposition* 1.3 with Q as the "origin" for the triangle OP_iP_{i+1}, we can write:

$$\mathcal{A}(Q, P_i, P_{i+1}) = \mathcal{A}(O, P_i, P_{i+1}) + \mathcal{A}(O, P_{i+1}, Q) + \mathcal{A}(O, Q, P_i).$$

Summing over $i = 1$ to n:

$$\sum_{i=1}^{n} \mathcal{A}(Q, P_i, P_{i+1}) = \sum_{i=1}^{n} \mathcal{A}(O, P_i, P_{i+1}) + \sum_{i=1}^{n} \left(\mathcal{A}(O, P_{i+1}, Q) - \mathcal{A}(O, P_i, Q)\right),$$

where we have used the skew-symmetry $\mathcal{A}(O, Q, P_i) = -\mathcal{A}(O, P_i, Q)$. The second sum is a telescoping series. Since $P_{n+1} = P_1$, all terms cancel:

$$\sum_{i=1}^{n} \left(\mathcal{A}(O, P_{i+1}, Q) - \mathcal{A}(O, P_i, Q)\right) = 0.$$

Thus, the total area remains unchanged.

■

Example 1.5. Decomposition of Complex Shapes. Consider a non-convex polygon, such as a "star" shape or a polygon with self-intersections. The definition above remains valid. The sign of the resulting area reflects the direction in which the boundary winds around the interior (counter-clockwise versus clockwise). For simple polygons, the area satisfies the "Sum of Parts Principle" via internal cancellation.
If a polygon $ABCD$ is decomposed into two triangles ABC and ACD by the diagonal AC, the total area is:

$$\text{Area}(ABCD) = \mathcal{A}(OAB) + \mathcal{A}(OBC) + \mathcal{A}(OCD) + \mathcal{A}(ODA).$$

Using $\mathcal{A}(OAC) + \mathcal{A}(OCA) = 0$, we can insert these terms:

$$\text{Area}(ABCD) = (\mathcal{A}(OAB) + \mathcal{A}(OBC) + \mathcal{A}(OCA)) + (\mathcal{A}(OAC) + \mathcal{A}(OCD) + \mathcal{A}(ODA)).$$

$$\text{Area}(ABCD) = \mathcal{A}(ABC) + \mathcal{A}(ACD).$$

Thus, the area of the union is the sum of the components.

範例

Volume and The Triple Product

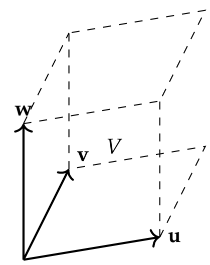

Figure 1.7: The volume of the parallelepiped is determined by the scalar triple product of its edges.

The geometric intuition of determinants extends naturally to 3-dimensional space. While the area of a parallelogram is given by a 2×2 determinant, the volume of a *parallelepiped* spanned by three vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$ is given by a 3×3 determinant.

Definition 1.6. ***Scalar Triple Product.***
For three vectors $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$, the signed volume of the parallelepiped they generate is:

$$\text{Vol}(\mathbf{u}, \mathbf{v}, \mathbf{w}) = \det \begin{bmatrix} | & | & | \\ \mathbf{u} & \mathbf{v} & \mathbf{w} \\ | & | & | \end{bmatrix}.$$

Algebraically, this is computed via the expansion:

$$\det \begin{bmatrix} u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \\ u_3 & v_3 & w_3 \end{bmatrix} = u_1 \det \begin{bmatrix} v_2 & w_2 \\ v_3 & w_3 \end{bmatrix} - u_2 \det \begin{bmatrix} v_1 & w_1 \\ v_3 & w_3 \end{bmatrix} + u_3 \det \begin{bmatrix} v_1 & w_1 \\ v_2 & w_2 \end{bmatrix}.$$

定義

Orientation in $\mathbb{R}^3$

Just as the sign of the 2D determinant indicates clockwise or counter-clockwise orientation, the sign of the 3D determinant indicates *handedness*.

- If $\det(\mathbf{u}, \mathbf{v}, \mathbf{w}) > 0$, the vectors form a **right-handed system**. This is consistent with the "right-hand rule": if the fingers of the right hand curl from $\mathbf{u}$ to $\mathbf{v}$, the thumb points in the direction of $\mathbf{w}$ (assuming $\mathbf{w}$ is on that side of the $\mathbf{uv}$-plane).
- If $\det(\mathbf{u}, \mathbf{v}, \mathbf{w}) < 0$, they form a **left-handed system**.
- If $\det(\mathbf{u}, \mathbf{v}, \mathbf{w}) = 0$, the vectors are *coplanar* (linearly dependent), and the volume is zero.

Example 1.6. Volume of a Box. Consider a rectangular box with side lengths l, w, h aligned with the axes. The defining vectors are $\mathbf{u} = (l, 0, 0)^T$, $\mathbf{v} = (0, w, 0)^T$, and $\mathbf{w} = (0, 0, h)^T$. The volume is:

$$\det \begin{bmatrix} l & 0 & 0 \\ 0 & w & 0 \\ 0 & 0 & h \end{bmatrix} = l(w \cdot h - 0) = lwh.$$

If we swap $\mathbf{u}$ and $\mathbf{v}$, the determinant becomes $-lwh$, reflecting the change in orientation from right-handed to left-handed.

範例

Remark.

This determinant formulation unifies the cross product and dot product from vector calculus. The scalar triple product can be written as $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})$. The cross product $\mathbf{v} \times \mathbf{w}$ produces an area vector normal to the base, and the dot product with $\mathbf{u}$ computes the "height" projected onto this normal, scaled by the base area.

Example 1.7. Cross Product via Determinants. The cross product $\mathbf{a} \times \mathbf{b}$ is often memorised using a heuristic determinant with basis vectors in the first row:

$$\mathbf{a} \times \mathbf{b} = \det \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{bmatrix} = (a_2b_3 - a_3b_2)\mathbf{e}_1 - (a_1b_3 - a_3b_1)\mathbf{e}_2 + (a_1b_2 - a_2b_1)\mathbf{e}_3.$$

While formally an abuse of notation (mixing vectors and scalars in a matrix), it correctly generates the algebraic expansion of the cross product.

範例

1.5 *The Geometry of Volume in* $\mathbb{R}^3$

We now extend the axiomatic framework established for the plane in *definition* 1.2 to three-dimensional space. While the signed area in $\mathbb{R}^2$ is a function of two vectors, the signed volume in $\mathbb{R}^3$ is a function of three.

Definition 1.7. ***Volume Axioms (Triple Product).***
Let V be a 3-dimensional vector space. A function $[\cdot,\cdot,\cdot] : V \times V \times V \to \mathbb{R}$ is called a *volume form* (or triple product) if it satisfies the following axioms for all $\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{x} \in V$ and $k \in \mathbb{R}$:

1. **Alternating Property:** Interchanging any two adjacent vectors reverses the sign.

$$[\mathbf{u}, \mathbf{v}, \mathbf{w}] = -[\mathbf{v}, \mathbf{u}, \mathbf{w}] = -[\mathbf{u}, \mathbf{w}, \mathbf{v}].$$

2. **Trilinearity:** The function is linear in each argument. For the first argument:

$$[\mathbf{u} + k\mathbf{x}, \mathbf{v}, \mathbf{w}] = [\mathbf{u}, \mathbf{v}, \mathbf{w}] + k[\mathbf{x}, \mathbf{v}, \mathbf{w}].$$

By the alternating property, linearity holds for the second and third

arguments as well.
3. **Non-Degeneracy:** $[\mathbf{u}, \mathbf{v}, \mathbf{w}] \neq 0$ if and only if $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is linearly independent.

定義

Note

The alternating property implies that if any two vectors are identical, the volume is zero. For instance, if $\mathbf{u} = \mathbf{v}$, then $[\mathbf{u}, \mathbf{u}, \mathbf{w}] = -[\mathbf{u}, \mathbf{u}, \mathbf{w}]$, which forces the value to be 0.

Uniqueness and the Determinant

Just as the area function in the plane is uniquely determined by the determinant relative to a basis (see *theorem* 1.1), the volume function in space is uniquely determined by the 3×3 determinant.

Theorem 1.3. ***Uniqueness of the Volume Form.***
Let $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ be a basis for $\mathbb{R}^3$, normalised such that $[\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3] = 1$. Then for any three vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$, the volume is given by the determinant of the matrix formed by their components:

$$[\mathbf{u}, \mathbf{v}, \mathbf{w}] = \det \begin{bmatrix} | & | & | \\ \mathbf{u} & \mathbf{v} & \mathbf{w} \\ | & | & | \end{bmatrix}.$$

定理

Proof

Let $\mathbf{u} = \sum_i u_i \mathbf{e}_i$, $\mathbf{v} = \sum_j v_j \mathbf{e}_j$, and $\mathbf{w} = \sum_k w_k \mathbf{e}_k$. Using trilinearity, we expand the product completely:

$$[\mathbf{u}, \mathbf{v}, \mathbf{w}] = \sum_{i=1}^{3} \sum_{j=1}^{3} \sum_{k=1}^{3} u_i v_j w_k [\mathbf{e}_i, \mathbf{e}_j, \mathbf{e}_k].$$

The sum contains $3^3 = 27$ terms. However, by the alternating property, any term where indices are repeated (e.g., $[\mathbf{e}_1, \mathbf{e}_1, \mathbf{e}_2]$) vanishes. The only non-zero terms correspond to permutations of distinct indices $\{1, 2, 3\}$. There are $3! = 6$ such permutations. We group them by sign:

- **Even permutations** (cyclic shifts of 123) preserve the sign:

$$[\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3] = [\mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_1] = [\mathbf{e}_3, \mathbf{e}_1, \mathbf{e}_2] = 1.$$

- **Odd permutations** (single swaps) reverse the sign:

$$[\mathbf{e}_2, \mathbf{e}_1, \mathbf{e}_3] = [\mathbf{e}_1, \mathbf{e}_3, \mathbf{e}_2] = [\mathbf{e}_3, \mathbf{e}_2, \mathbf{e}_1] = -1.$$

Substituting these values back into the sum yields:

$$[\mathbf{u}, \mathbf{v}, \mathbf{w}] = (u_1 v_2 w_3 + u_2 v_3 w_1 + u_3 v_1 w_2) - (u_2 v_1 w_3 + u_1 v_3 w_2 + u_3 v_2 w_1).$$

This is precisely the Leibniz expansion of the determinant $\det(\mathbf{u}, \mathbf{v}, \mathbf{w})$.

■

This theorem provides the justification for the definition of the Scalar Triple Product given in the previous section. The algebraic properties of the determinant are not arbitrary; they are the necessary consequences of the geometric axioms of volume.

Remark.

In more advanced contexts, this product is denoted by the wedge product $\mathbf{u} \wedge \mathbf{v} \wedge \mathbf{w}$. The space of such tri-vectors in $\mathbb{R}^3$ is 1-dimensional, which explains why the volume is unique up to a scalar factor.

Example 1.8. Linearly Dependent Vectors. Consider the vectors $\mathbf{u} = [1, 2, 3]^T$, $\mathbf{v} = [4, 5, 6]^T$, and $\mathbf{w} = [7, 8, 9]^T$. We can compute the volume $[\mathbf{u}, \mathbf{v}, \mathbf{w}]$ using column operations on the determinant, relying on the property that adding a multiple of one column to another preserves the volume (an immediate corollary of linearity and the vanishing of repeated factors).

$$\det \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix} \xrightarrow{C_2 \to C_2 - C_1} \det \begin{bmatrix} 1 & 3 & 7 \\ 2 & 3 & 8 \\ 3 & 3 & 9 \end{bmatrix} \xrightarrow{C_3 \to C_3 - C_1} \det \begin{bmatrix} 1 & 3 & 6 \\ 2 & 3 & 6 \\ 3 & 3 & 6 \end{bmatrix}.$$

Since the second and third columns are proportional (specifically $C_3 = 2C_2$), the determinant is 0. Thus, the vectors are coplanar.

範例

The Tetrahedron

The most elementary 3-dimensional solid is the tetrahedron, determined by four vertices O, A, B, C. Just as the area of a triangle is half the area of the parallelogram spanned by two vectors, the volume of a tetrahedron is a fraction of the parallelepiped spanned by three.

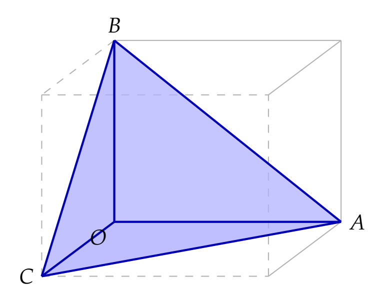

Figure 1.8: The tetrahedron $OABC$ occupies exactly $1/6$ of the volume of the parallelepiped spanned by $\boldsymbol{OA}, \boldsymbol{OB}, \boldsymbol{OC}$.

Definition 1.8. ***Volume of a Tetrahedron.***
The signed volume of the tetrahedron with vertices O, A, B, C is given by:

$$\mathrm{Vol}(OABC) = \frac{1}{6}[\boldsymbol{OA}, \boldsymbol{OB}, \boldsymbol{OC}].$$

定義

The factor of $1/6$ arises from elementary calculus (or the principle of Cavalieri): the volume of a pyramid is $\frac{1}{3} \times$ Base Area $\times$ Height. Since the base area of the triangle OAB is $\frac{1}{2}$ the parallelogram area, the total factor is $\frac{1}{3} \times \frac{1}{2} = \frac{1}{6}$.

We can use the algebraic properties of the volume form to solve geometric problems without explicit coordinates.

Example 1.9. The Midpoint Tetrahedron. Let $OABC$ be a tetrahedron. Let P, Q, R be the midpoints of the edges OA, OB, and OC respectively. It is immediate that the volume of $OPQR$ is $\frac{1}{8}$ the volume of $OABC$. Consider a more non-trivial construction: Let P, Q, R be the midpoints of the edges of the *face* triangle ABC. Specifically:

$$P = \frac{1}{2}(A + B), \quad Q = \frac{1}{2}(B + C), \quad R = \frac{1}{2}(C + A).$$

We wish to find the volume of the tetrahedron $OPQR$ relative to $OABC$. Using the definition:

$$\text{Vol}(OPQR) = \frac{1}{6}[\boldsymbol{OP}, \boldsymbol{OQ}, \boldsymbol{OR}] = \frac{1}{6}\left[\frac{\mathbf{a}+\mathbf{b}}{2}, \frac{\mathbf{b}+\mathbf{c}}{2}, \frac{\mathbf{c}+\mathbf{a}}{2}\right].$$

Factoring out the scalars ($\frac{1}{2}$ from each term):

$$\text{Vol}(OPQR) = \frac{1}{6} \cdot \frac{1}{8}[\mathbf{a}+\mathbf{b}, \mathbf{b}+\mathbf{c}, \mathbf{c}+\mathbf{a}].$$

We expand the bracket using linearity. Note that any term with repeated vectors will vanish.

$$\begin{aligned}[\mathbf{a}+\mathbf{b}, \mathbf{b}+\mathbf{c}, \mathbf{c}+\mathbf{a}] &= [\mathbf{a}, \mathbf{b}+\mathbf{c}, \mathbf{c}+\mathbf{a}] + [\mathbf{b}, \mathbf{b}+\mathbf{c}, \mathbf{c}+\mathbf{a}] \\ &= ([\mathbf{a}, \mathbf{b}, \mathbf{c}] + [\mathbf{a}, \mathbf{b}, \mathbf{a}] + [\mathbf{a}, \mathbf{c}, \mathbf{c}] + [\mathbf{a}, \mathbf{c}, \mathbf{a}]) \\ &\quad + ([\mathbf{b}, \mathbf{b}, \mathbf{c}] + [\mathbf{b}, \mathbf{b}, \mathbf{a}] + [\mathbf{b}, \mathbf{c}, \mathbf{c}] + [\mathbf{b}, \mathbf{c}, \mathbf{a}]).\end{aligned}$$

Terms like $[\mathbf{a}, \mathbf{b}, \mathbf{a}]$ vanish. The only survivors are:

$$[\mathbf{a}, \mathbf{b}, \mathbf{c}] \quad \text{and} \quad [\mathbf{b}, \mathbf{c}, \mathbf{a}].$$

By the cyclic property of the triple product, $[\mathbf{b}, \mathbf{c}, \mathbf{a}] = [\mathbf{a}, \mathbf{b}, \mathbf{c}]$. Thus the sum is $2[\mathbf{a}, \mathbf{b}, \mathbf{c}]$. Substituting back:

$$\text{Vol}(OPQR) = \frac{1}{48}(2[\mathbf{a}, \mathbf{b}, \mathbf{c}]) = \frac{1}{4}\left(\frac{1}{6}[\mathbf{a}, \mathbf{b}, \mathbf{c}]\right) = \frac{1}{4}\text{Vol}(OABC).$$

範例

Example 1.10. Tetrahedral Volume Ratio. Let O, A, B, C be the vertices of a tetrahedron. Let G be the centroid of the face triangle ABC, given by $\mathbf{g} = \frac{1}{3}(\mathbf{a} + \mathbf{b} + \mathbf{c})$, where position vectors are relative

to O. Prove that the volume of the tetrahedron $OABG$ is one-third the volume of $OABC$.
We compute the ratio:

$$\frac{\mathrm{Vol}(OABG)}{\mathrm{Vol}(OABC)} = \frac{[\mathbf{a},\mathbf{b},\mathbf{g}]}{[\mathbf{a},\mathbf{b},\mathbf{c}]} = \frac{[\mathbf{a},\mathbf{b},\frac{1}{3}(\mathbf{a}+\mathbf{b}+\mathbf{c})]}{[\mathbf{a},\mathbf{b},\mathbf{c}]}.$$

Using linearity in the third argument:

$$[\mathbf{a},\mathbf{b},\frac{1}{3}(\mathbf{a}+\mathbf{b}+\mathbf{c})] = \frac{1}{3}\left([\mathbf{a},\mathbf{b},\mathbf{a}] + [\mathbf{a},\mathbf{b},\mathbf{b}] + [\mathbf{a},\mathbf{b},\mathbf{c}]\right).$$

The first two terms vanish due to repeated vectors. We are left with $\frac{1}{3}[\mathbf{a},\mathbf{b},\mathbf{c}]$. Thus the ratio is $1/3$.

範例

1.6 Exercises

1. **Determinant Calculation.** Let A and B be the matrices defined by:

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

   Compute the values of the following determinants:

   (a) $\det(A)$ and $\det(B)$.

   (b) $\det(AB^2)$.

   (c) $\det(A+B)$.

   (d) $\det(A^{-1}(A+B))$.

2. **Cyclic Equations.** Solve the equation $\det(M) = 0$ for the variable x, where M is given by:

$$M = \begin{bmatrix} a-x & b-x & c \\ a-x & c & b-x \\ a & b-x & c-x \end{bmatrix}.$$

Hint: Use row and column operations to simplify the matrix before expanding. Look for common factors.

3. **Block Determinants.** Let A be an $n \times n$ matrix partitioned into blocks:

$$A = \begin{bmatrix} U & V \\ \mathbf{0} & X \end{bmatrix},$$

   where U is a $k \times k$ matrix, X is an $m \times m$ matrix (with $k + m = n$), and $\mathbf{0}$ is the zero matrix. Using the definition of the determinant (sum over permutations), prove that:

$$\det(A) = \det(U)\det(X).$$

> *Remark.*
> Consider which permutations $\sigma \in S_n$ yield non-zero terms. If σ maps an index from the first k rows to one of the last m columns, what is the value of the corresponding matrix entry?

4. **Roots of Unity.** Let ω be a complex root of the equation $x^3 - 1 = 0$ with $\omega \neq 1$. Consider the matrix:

$$C = \begin{bmatrix} 1 & \omega & \omega^2 \\ \omega & \omega^2 & 1 \\ \omega^2 & 1 & \omega \end{bmatrix}.$$

   (a) Evaluate $\det(C)$.

   (b) Show that the columns of C form a linearly dependent set over $\mathbb{C}$.

   (c) What is the geometric interpretation of the linear dependence of these complex vectors in $\mathbb{C}^3$?

5. **The Shoelace Formula.** Let $P_1 = (x_1, y_1), P_2 = (x_2, y_2), \ldots, P_n = (x_n, y_n)$ be the vertices of a polygon in $\mathbb{R}^2$ listed in counter-clockwise order. Using the vector definition of area for a polygon (Area $= \sum \mathcal{A}(O, P_i, P_{i+1})$), derive the coordinate formula:

$$\text{Area} = \frac{1}{2} \sum_{i=1}^{n} (x_i y_{i+1} - x_{i+1} y_i),$$

where $(x_{n+1}, y_{n+1}) = (x_1, y_1)$.

Expand the term $\mathcal{A}(O, P_i, P_{i+1})$ using the determinant of position vectors relative to the origin $O = (0, 0)$.

6. **Linearity of the Triple Product.** Let $\mathbf{u}, \mathbf{v}, \mathbf{w}$ be three vectors in $\mathbb{R}^3$.

   (a) Using the linearity and alternating properties of the volume form $[\cdot, \cdot, \cdot]$, prove the identity:

$$[\mathbf{u} + \mathbf{v}, \mathbf{v} + \mathbf{w}, \mathbf{w} + \mathbf{u}] = 2[\mathbf{u}, \mathbf{v}, \mathbf{w}].$$

   (b) Give a geometric interpretation of this result relating the volume of the parallelepiped spanned by the face diagonals to the volume of the original parallelepiped.

7. **Integer Coordinates.** Let $\triangle ABC$ be a triangle in the plane such that the coordinates of the vertices A, B, C are all integers.

   (a) Prove that the signed area $\mathcal{A}(ABC)$ is a rational number of the form $n/2$ for some integer n.

   (b) Deduce that an equilateral triangle cannot have all its vertices on integer lattice points.

8. **Cramer's Rule via Volume.** Consider the system of linear equations $A\mathbf{x} = \mathbf{b}$, where $A = [\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3]$ is a 3×3 matrix with column vectors $\mathbf{a}_i$, and $\mathbf{x} = [x_1, x_2, x_3]^T$.

(a) Express the vector $\mathbf{b}$ as $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3$.

(b) Compute the determinant $\det([\mathbf{b}, \mathbf{a}_2, \mathbf{a}_3])$ using the linearity of the determinant in the first column.

(c) Deduce Cramer's Rule: if $\det(A) \neq 0$, then

$$x_1 = \frac{\det([\mathbf{b}, \mathbf{a}_2, \mathbf{a}_3])}{\det(A)}.$$

9. **The Vector Triple Product Identity.** While the scalar triple product gives a volume, the *vector* triple product $\mathbf{a} \times (\mathbf{b} \times \mathbf{c})$ yields a vector. Using the property that $\mathbf{a} \times (\mathbf{b} \times \mathbf{c})$ must lie in the plane spanned by $\mathbf{b}$ and $\mathbf{c}$ (why?), and must be orthogonal to $\mathbf{a}$, prove the expansion:

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}.$$

You may assume the standard basis vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ and check the identity component-wise, or use the determinant definition of the cross product.

10. **Continuity of Orientation.** Let $A(t)$ be a continuous family of $n \times n$ invertible matrices for $t \in [0, 1]$. This means the entries $a_{ij}(t)$ are continuous functions of t. Prove that if $\det(A(0)) > 0$, then $\det(A(t)) > 0$ for all $t \in [0, 1]$.

> *Remark.*
>
> Use the fact that the determinant is a polynomial function of the entries, hence continuous. What property of the intermediate values of continuous functions prevents the determinant from jumping from positive to negative without crossing zero?

2

General Theory of Determinants

We have established in the previous chapter that area in $\mathbb{R}^2$ and volume in $\mathbb{R}^3$ are characterised by specific algebraic properties: skew-symmetry, multilinearity, and normalisation. We now extend these concepts to n-dimensional space. While our geometric intuition falters beyond three dimensions, the algebraic structure remains robust. We formally define the determinant not as a mere formula, but as the unique function satisfying these geometric axioms for $n \times n$ matrices.

2.1 The Determinant Axioms

Let $M_n(\mathbb{R})$ denote the set of $n \times n$ matrices with real entries. We view a matrix $A \in M_n(\mathbb{R})$ as an ordered list of its n column vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^n$.

$$A = [\mathbf{a}_1, \ldots, \mathbf{a}_n] = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}.$$

We seek a function $\det : M_n(\mathbb{R}) \to \mathbb{R}$, often denoted by the bracket notation $[\mathbf{a}_1, \ldots, \mathbf{a}_n]$, that generalises the volume forms of *definition 1.7*.

Definition 2.1. ***Determinant Functions.***
A function $D : (\mathbb{R}^n)^n \to \mathbb{R}$ is called a *determinant function* if it satisfies the following axioms:

1. **Skew-Symmetry (Alternating):** Interchanging any two adjacent columns reverses the sign.

$$D(\ldots, \mathbf{a}_i, \mathbf{a}_{i+1}, \ldots) = -D(\ldots, \mathbf{a}_{i+1}, \mathbf{a}_i, \ldots).$$

2. **Multilinearity:** The function is linear in each column argument. For the k-th column:

$$D(\ldots, \mathbf{u} + c\mathbf{v}, \ldots) = D(\ldots, \mathbf{u}, \ldots) + cD(\ldots, \mathbf{v}, \ldots).$$

3. **Normalisation:** The determinant of the identity matrix $I_n = [\mathbf{e}_1, \dots, \mathbf{e}_n]$ is 1.

$$D(\mathbf{e}_1, \dots, \mathbf{e}_n) = 1.$$

定義

Note

The term "determinant" typically refers to the unique function satisfying these axioms. We denote this value by $\det(A)$ or $|A|$.

Elementary Consequences

From these axioms alone, several key properties follow immediately, mirroring those derived for the triple product.

Proposition 2.1. ***Basic Properties.***
Let $[\cdot]$ be a function satisfying the skew-symmetry and multilinearity axioms.

1. **Vanishing Property:** If two columns are identical, the value is zero.
2. **Linear Dependence:** If the columns $\mathbf{a}_1, \dots, \mathbf{a}_n$ are linearly dependent, then $[\mathbf{a}_1, \dots, \mathbf{a}_n] = 0$.
3. **General Permutation:** For any permutation σ of indices $\{1, \dots, n\}$:

$$[\mathbf{a}_{\sigma(1)}, \dots, \mathbf{a}_{\sigma(n)}] = \operatorname{sgn}(\sigma)[\mathbf{a}_1, \dots, \mathbf{a}_n],$$

where $\operatorname{sgn}(\sigma)$ is $+1$ if σ is even and -1 if σ is odd.

命題

Proof

For (1), let $\mathbf{a}_i = \mathbf{a}_j$ with $i < j$. By repeatedly swapping adjacent columns, we can bring $\mathbf{a}_j$ next to $\mathbf{a}_i$. Let the resulting value be V. Swapping the two identical vectors gives $-V$, but since the vectors are identical, the matrix is unchanged, so $V = -V \implies V = 0$.
For (2), if the columns are dependent, one vector, say $\mathbf{a}_k$, can be written as a linear combination of the others: $\mathbf{a}_k = \sum_{j \neq k} c_j \mathbf{a}_j$. By multilinearity:

$$[\dots, \mathbf{a}_k, \dots] = \sum_{j \neq k} c_j [\dots, \mathbf{a}_j, \dots].$$

In each term of the sum, the column $\mathbf{a}_j$ appears twice (once at position j and once at position k). Thus every term vanishes.
For (3), any permutation can be decomposed into a sequence of transpositions (swaps). Each swap reverses the sign. Thus, if N swaps are required, the sign change is $(-1)^N$, which is precisely the sign of the permutation.

∎

2.2 *Uniqueness and the Leibniz Formula*

We now address the existence and uniqueness of such a function. By expressing the column vectors in terms of the standard basis, we can derive an explicit formula, proving uniqueness.
Let $A = [\mathbf{a}_1, \ldots, \mathbf{a}_n]$. We write each column $\mathbf{a}_j$ as:

$$\mathbf{a}_j = \sum_{i=1}^{n} a_{ij}\mathbf{e}_i.$$

Using multilinearity in the first column:

$$[\mathbf{a}_1, \ldots, \mathbf{a}_n] = \left[\sum_{i_1=1}^{n} a_{i_1 1}\mathbf{e}_{i_1}, \mathbf{a}_2, \ldots, \mathbf{a}_n\right] = \sum_{i_1=1}^{n} a_{i_1 1}[\mathbf{e}_{i_1}, \mathbf{a}_2, \ldots, \mathbf{a}_n].$$

Repeating this expansion for all n columns yields:

$$\det(A) = \sum_{i_1=1}^{n}\sum_{i_2=1}^{n} \cdots \sum_{i_n=1}^{n} a_{i_1 1}a_{i_2 2}\cdots a_{i_n n}[\mathbf{e}_{i_1}, \mathbf{e}_{i_2}, \ldots, \mathbf{e}_{i_n}].$$

The term $[\mathbf{e}_{i_1}, \ldots, \mathbf{e}_{i_n}]$ is non-zero only if the indices $i_1, \ldots, i_n$ are distinct (otherwise there is a repeated column). Thus, the sum restricts to permutations $\sigma \in S_n$ where $i_k = \sigma(k)$. Using the permutation property (*proposition* 2.1) and the normalisation $[\mathbf{e}_1, \ldots, \mathbf{e}_n] = 1$:

$$[\mathbf{e}_{\sigma(1)}, \ldots, \mathbf{e}_{\sigma(n)}] = \operatorname{sgn}(\sigma)[\mathbf{e}_1, \ldots, \mathbf{e}_n] = \operatorname{sgn}(\sigma).$$

Substituting this back, we obtain the **Leibniz Formula**.

> **Theorem 2.1.** ***Leibniz Formula for Determinants.***
> There exists a unique determinant function $\det : M_n(\mathbb{R}) \to \mathbb{R}$, given by:
>
> $$\det(A) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{j=1}^{n} a_{\sigma(j)j}.$$
>
> 定理

> *Remark.*
>
> While *theorem* 2.1 proves uniqueness, verifying that this formula satisfies the axioms (Existence) is a standard exercise in combinatorics.
>
> 1. **Normalisation:** For I_n, $a_{ij} = \delta_{ij}$. The only non-zero term in the sum corresponds to the identity permutation, yielding 1.
> 2. **Skew-Symmetry:** Swapping columns corresponds to composing σ with a transposition τ. Since $\operatorname{sgn}(\sigma \circ \tau) = -\operatorname{sgn}(\sigma)$, the entire sum changes sign.
> 3. **Multilinearity:** Each term in the sum contains exactly one factor

from the k-th column, ensuring linearity with respect to that column.

2.3 Computational Techniques via Multilinearity

Although the Leibniz formula defines the determinant, it involves $n!$ terms, making it computationally intractable for large n. The axioms themselves often provide a more efficient route for calculation, particularly using the property that adding a scalar multiple of one column to another preserves the determinant (an immediate corollary of multilinearity and the alternating property).

Example 2.1. Multilinearity in Action. Let us evaluate the following 4×4 determinant not by row reduction or cofactor expansion, but by direct application of the linearity axioms on the column vectors.

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

We identify the columns as vectors in $\mathbb{R}^4$. Observe the decomposition in terms of basis vectors $\mathbf{e}_i$:

$$\begin{aligned} \mathbf{c}_1 &= \mathbf{e}_1 + 3\mathbf{e}_2 \\ \mathbf{c}_2 &= 2\mathbf{e}_2 - \mathbf{e}_3 \\ \mathbf{c}_3 &= 2\mathbf{e}_1 + 4\mathbf{e}_4 \\ \mathbf{c}_4 &= 5\mathbf{e}_3 + 2\mathbf{e}_4 \end{aligned}$$

The determinant is $D = [\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3, \mathbf{c}_4]$. We expand using multilinearity.

$$D = [\mathbf{e}_1 + 3\mathbf{e}_2, \quad 2\mathbf{e}_2 - \mathbf{e}_3, \quad 2\mathbf{e}_1 + 4\mathbf{e}_4, \quad 5\mathbf{e}_3 + 2\mathbf{e}_4].$$

We can expand the first column:

$$D = [\mathbf{e}_1, \mathbf{c}_2, \mathbf{c}_3, \mathbf{c}_4] + 3[\mathbf{e}_2, \mathbf{c}_2, \mathbf{c}_3, \mathbf{c}_4].$$

However, let us look for terms that survive. A term survives only if it contains distinct basis vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4$ in some order. Consider the expansion of the determinant expression:

$$[\mathbf{e}_1 + 3\mathbf{e}_2, \quad 2\mathbf{e}_2 - \mathbf{e}_3, \quad 2\mathbf{e}_1 + 4\mathbf{e}_4, \quad 5\mathbf{e}_3 + 2\mathbf{e}_4].$$

Let us distribute terms using multilinearity. From $\mathbf{c}_3$, the $\mathbf{e}_1$ term will interact with $\mathbf{c}_1$. If we pick $\mathbf{e}_1$ from $\mathbf{c}_1$, we must pick $4\mathbf{e}_4$ from $\mathbf{c}_3$ (to avoid repeating $\mathbf{e}_1$). If we pick $3\mathbf{e}_2$ from $\mathbf{c}_1$, we can pick either $2\mathbf{e}_1$ or $4\mathbf{e}_4$ from $\mathbf{c}_3$.
Let us group by the choice from $\mathbf{c}_1$:

1. **Select $\mathbf{e}_1$ from $\mathbf{c}_1$:** We must not pick $\mathbf{e}_1$ from $\mathbf{c}_3$, so we must pick $4\mathbf{e}_4$. Current basis: $\{\mathbf{e}_1, \cdot, \mathbf{e}_4, \cdot\}$. From $\mathbf{c}_2 = 2\mathbf{e}_2 - \mathbf{e}_3$, we can pick either. From $\mathbf{c}_4 = 5\mathbf{e}_3 + 2\mathbf{e}_4$, we cannot pick $\mathbf{e}_4$ (already chosen). We must pick $5\mathbf{e}_3$. This forces us to pick $2\mathbf{e}_2$ from $\mathbf{c}_2$ (since $\mathbf{e}_3$ is taken by $\mathbf{c}_4$).

   Resulting term:

   $$[\mathbf{e}_1, 2\mathbf{e}_2, 4\mathbf{e}_4, 5\mathbf{e}_3] = (1)(2)(4)(5)[\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_4, \mathbf{e}_3].$$

   Since $[\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_4, \mathbf{e}_3] = -[\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4] = -1$, this term is $40(-1) = -40$.

2. **Select $3\mathbf{e}_2$ from $\mathbf{c}_1$:** Current basis: $\{\mathbf{e}_2, \cdot, \cdot, \cdot\}$. From $\mathbf{c}_2 = 2\mathbf{e}_2 - \mathbf{e}_3$, we cannot pick $2\mathbf{e}_2$. We must pick $-\mathbf{e}_3$. Current basis: $\{\mathbf{e}_2, \mathbf{e}_3, \cdot, \cdot\}$. From $\mathbf{c}_4 = 5\mathbf{e}_3 + 2\mathbf{e}_4$, we cannot pick $5\mathbf{e}_3$. We must pick $2\mathbf{e}_4$. Current basis: $\{\mathbf{e}_2, \mathbf{e}_3, \cdot, \mathbf{e}_4\}$. From $\mathbf{c}_3 = 2\mathbf{e}_1 + 4\mathbf{e}_4$, we cannot pick $4\mathbf{e}_4$. We must pick $2\mathbf{e}_1$.

   Resulting term:

   $$[3\mathbf{e}_2, -\mathbf{e}_3, 2\mathbf{e}_1, 2\mathbf{e}_4] = (3)(-1)(2)(2)[\mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_1, \mathbf{e}_4].$$

   To order $\mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_1, \mathbf{e}_4$: swap $\mathbf{e}_3 \leftrightarrow \mathbf{e}_1$ (one swap) $\to \mathbf{e}_2, \mathbf{e}_1, \mathbf{e}_3, \mathbf{e}_4$. Swap $\mathbf{e}_2 \leftrightarrow \mathbf{e}_1$ (second swap) $\to \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4$. Two swaps imply an even permutation (+1). Term value: $(-12)(1) = -12$.

Total Determinant $D = -40 - 12 = -52$.

範例

This method of "multilinear expansion" is often superior to standard algorithms for sparse matrices where columns are linear combinations of few basis vectors. It highlights that the determinant measures how the "volume" of the standard hypercube is distorted and mixed by the linear transformation.

2.4 *Fundamental Properties of the Determinant*

We now turn to the theoretical implications of the determinant axioms. Having established that the determinant is the unique alternating multilinear form normalised at the identity, we can use it to characterise linear dependence, solve systems of equations, and relate column properties to row properties.

Linear Independence

In *proposition* 2.1, we established that if the columns of a matrix A are linearly dependent, then $\det(A) = 0$. The converse is equally true, providing a complete algebraic criterion for linear independence.

Theorem 2.2. ***Determinant Criterion for Independence.***
Let $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^n$. The vectors are linearly dependent if and only if

$$[\mathbf{a}_1, \ldots, \mathbf{a}_n] = 0.$$

定理

Proof

The forward implication was proven in *proposition* 2.1. We prove the converse by contradiction. Suppose that $[\mathbf{a}_1, \ldots, \mathbf{a}_n] = 0$ but the vectors are linearly independent. Since they are n linearly independent vectors in an n-dimensional space, they form a basis for $\mathbb{R}^n$.
Consequently, the standard basis vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ can be expressed as linear combinations of the $\mathbf{a}_j$'s. For each $k \in \{1, \ldots, n\}$, there exist scalars b_{ki} such that:

$$\mathbf{e}_k = \sum_{i=1}^{n} b_{ki}\mathbf{a}_i.$$

Consider the determinant of the identity matrix $I = [\mathbf{e}_1, \ldots, \mathbf{e}_n]$. Using the expansion above:

$$1 = \det(I) = \left[\sum_{i_1} b_{1i_1}\mathbf{a}_{i_1}, \ldots, \sum_{i_n} b_{ni_n}\mathbf{a}_{i_n}\right].$$

By the multilinearity of the determinant, we may expand this into a sum of determinants of the form:

$$b_{1i_1} \cdots b_{ni_n}[\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_n}].$$

If any two indices in the sequence $i_1, \ldots, i_n$ are identical, the term vanishes. If all indices are distinct, the sequence is a permutation of $1, \ldots, n$. In this case,

$$[\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_n}] = \pm[\mathbf{a}_1, \ldots, \mathbf{a}_n].$$

However, by hypothesis, $[\mathbf{a}_1, \ldots, \mathbf{a}_n] = 0$. Thus, every term in the expansion is zero, implying $\det(I) = 0$, which contradicts the normalisation axiom ($\det(I) = 1$). We conclude that the vectors must be linearly dependent.

■

This theorem provides a computationally verifiable test for the invertibility of a matrix. A matrix A is invertible (non-singular) if and only if $\det(A) \neq 0$.

Cramer's Rule

The multilinearity of the determinant allows us to construct explicit formulas for the solutions of linear systems $A\mathbf{x} = \mathbf{b}$, provided $\det(A) \neq 0$.

Theorem 2.3. ***Cramer's Rule.***
Let $A = [\mathbf{a}_1, \ldots, \mathbf{a}_n]$ be an invertible $n \times n$ matrix, and let $\mathbf{b} \in \mathbb{R}^n$. The unique solution to the system

$$x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{b}$$

is given by

$$x_k = \frac{\det(A_k)}{\det(A)},$$

where A_k is the matrix obtained by replacing the k-th column of A with $\mathbf{b}$:

$$A_k = [\mathbf{a}_1, \ldots, \mathbf{a}_{k-1}, \mathbf{b}, \mathbf{a}_{k+1}, \ldots, \mathbf{a}_n].$$

定理

Proof

Since A is invertible, $\det(A) \neq 0$ and the columns $\{\mathbf{a}_i\}$ form a basis. The equation $A\mathbf{x} = \mathbf{b}$ is equivalent to:

$$\mathbf{b} = \sum_{i=1}^{n} x_i\mathbf{a}_i.$$

Consider the determinant of the matrix A_k. By linearity in the k-th column:

$$\begin{aligned}
\det(A_k) &= [\mathbf{a}_1, \ldots, \mathbf{b}, \ldots, \mathbf{a}_n] \\
&= \left[\mathbf{a}_1, \ldots, \sum_{i=1}^{n} x_i\mathbf{a}_i, \ldots, \mathbf{a}_n\right] \\
&= \sum_{i=1}^{n} x_i[\mathbf{a}_1, \ldots, \mathbf{a}_i, \ldots, \mathbf{a}_n].
\end{aligned}$$

Observe that for $i \neq k$, the column $\mathbf{a}_i$ appears twice (at position i and position k), causing the determinant to vanish. The only nonzero term corresponds to $i = k$:

$$\det(A_k) = x_k[\mathbf{a}_1, \ldots, \mathbf{a}_k, \ldots, \mathbf{a}_n] = x_k \det(A).$$

Solving for x_k yields the result.

■

Example 2.2. Solving a System via Cramer's Rule. Consider the system:

$$\begin{aligned} 3x + y - z &= 4 \\ x + y + z &= 0 \\ 2x + z &= 2 \end{aligned}$$

We can write this in vector form $x\mathbf{a}_1 + y\mathbf{a}_2 + z\mathbf{a}_3 = \mathbf{b}$, with coefficient matrix A. First, we compute the determinant of coefficients:

$$\Delta = \det(A) = \det \begin{bmatrix} 3 & 1 & -1 \\ 1 & 1 & 1 \\ 2 & 0 & 1 \end{bmatrix}.$$

Using column operations: $C_1 \to C_1 - C_2$ and $C_3 \to C_3 - C_2$:

$$\Delta = \det \begin{bmatrix} 2 & 1 & -2 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix} = 1 \cdot \det \begin{bmatrix} 2 & -2 \\ 2 & 1 \end{bmatrix} = 2(1) - (-2)(2) = 6.$$

Since $\Delta \neq 0$, a unique solution exists. To find y, we replace the second column with $\mathbf{b} = [4, 0, 2]^T$:

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

Expanding along the second row (which has a zero):

$$\Delta_y = -1 \det \begin{bmatrix} 4 & -1 \\ 2 & 1 \end{bmatrix} - 1 \det \begin{bmatrix} 3 & 4 \\ 2 & 2 \end{bmatrix} = -1(4 - (-2)) - 1(6 - 8) = -6 + 2 = -4.$$

Thus, $y = \frac{\Delta_y}{\Delta} = \frac{-4}{6} = -\frac{2}{3}$. (Computing x and z is similar).

範例

The Transpose Property

The axioms of the determinant are defined in terms of columns. However, there is a perfect duality between rows and columns.

Definition 2.2. ***Transpose.***
Let A be an $n \times n$ matrix with entries a_{ij}. The transpose A^T is the matrix with entries $b_{ij} = a_{ji}$.

定義

Theorem 2.4. ***Determinant of the Transpose.***

For any square matrix A,

$$\det(A) = \det(A^T).$$

定理

Proof

We utilise the Leibniz Formula. Let $A = (a_{ij})$ and $A^T = (b_{ij})$ where $b_{ij} = a_{ji}$.

$$\det(A^T) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} b_{\sigma(i)i} = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i\sigma(i)}.$$

We may rearrange the product term. Let $j = \sigma(i)$. Since σ is a bijection, as i ranges over $\{1, \ldots, n\}$, so does j. The relation $j = \sigma(i)$ implies $i = \sigma^{-1}(j)$. Thus:

$$\prod_{i=1}^{n} a_{i\sigma(i)} = \prod_{j=1}^{n} a_{\sigma^{-1}(j)j}.$$

The sum is over all permutations $\sigma \in S_n$. As σ ranges over the group S_n, so does its inverse $\tau = \sigma^{-1}$. Furthermore, $\operatorname{sgn}(\sigma) = \operatorname{sgn}(\sigma^{-1})$ (as we shall see in section 2.5, σ and σ^{-1} share the same decomposition into transpositions). Substituting τ for σ^{-1} in the sum:

$$\det(A^T) = \sum_{\tau \in S_n} \operatorname{sgn}(\tau) \prod_{j=1}^{n} a_{\tau(j)j}.$$

This is precisely the expression for $\det(A)$.

■

Proposition 2.2. ***Row Operations.***
Since $\det(A) = \det(A^T)$, every property established for columns applies equally to rows.

1. **Row Alternation:** Swapping two rows reverses the sign of the determinant.
2. **Row Multilinearity:** The determinant is linear in each row.
3. **Row Operations:** Adding a multiple of one row to another preserves the determinant.

命題

Proof

Let R be a row operation (e.g., swapping rows). Performing R on A is equivalent to performing the corresponding column operation on A^T. Since $\det(A) = \det(A^T)$, the effect on the determinant is identical.

■

Example 2.3. Determinant via Row Reduction. To evaluate

$$D = \det \begin{bmatrix} 0 & 3 & 7 \\ 1 & 2 & 3 \\ -1 & 8 & 6 \end{bmatrix},$$

we may transpose it to work with columns, or simply apply row operations directly. Swap R_1 and R_2 (sign change):

$$D = -\det \begin{bmatrix} 1 & 2 & 3 \\ 0 & 3 & 7 \\ -1 & 8 & 6 \end{bmatrix}.$$

Add R_1 to R_3:

$$D = -\det \begin{bmatrix} 1 & 2 & 3 \\ 0 & 3 & 7 \\ 0 & 10 & 9 \end{bmatrix}.$$

Now expand along the first column:

$$D = -(1) \det \begin{bmatrix} 3 & 7 \\ 10 & 9 \end{bmatrix} = -(27 - 70) = -(-43) = 43.$$

範例

2.5 *Parity and the Symmetric Group*

The foundation of the determinant, specifically the Leibniz Formula, relies entirely on the sign of a permutation. In *proposition* 2.1, we asserted that a permutation σ could be classified as even or odd based on the number of transpositions required to form it. We now provide the combinatorial justification for this claim, ensuring that the sign is well-defined and independent of the specific sequence of transpositions used.

Inversions and Parity

Let S_n denote the group of permutations of the set $\{1, \dots, n\}$. We represent a permutation $\sigma \in S_n$ by the sequence of values $\sigma(1), \sigma(2), \dots, \sigma(n)$.

Definition 2.3. ***Inversion.***
An *inversion* in a permutation σ is a pair of indices (i, j) such that $i < j$ but $\sigma(i) > \sigma(j)$. The *inversion count* $N(\sigma)$ is the total number of such pairs.

定義

Note

Intuitively, if we write the numbers in a row, an inversion occurs whenever a larger number precedes a smaller one. We can calculate $N(\sigma)$ by iterating through the sequence and counting how many "strangers" (larger numbers) sit to the left of each element.

Example 2.4. Calculating Parity. Consider the permutation $\sigma = 53241$ in S_5. We count the inversions for each position j (pairs (i, j) with $i < j$ and $\sigma(i) > \sigma(j)$):

- 5: No elements to the left. (0)
- 3: Preceded by 5. (1)
- 2: Preceded by 5, 3. (2)
- 4: Preceded by 5. (1)
- 1: Preceded by 5, 3, 2, 4. (4)

Total inversions $N(\sigma) = 0 + 1 + 2 + 1 + 4 = 8$. Since 8 is even, we call σ an *even permutation*.

範例

We define the *sign* of the permutation as $\mathrm{sgn}(\sigma) = (-1)^{N(\sigma)}$. To show that this matches the definition based on transpositions, we prove that swapping any two elements flips the parity.

Lemma 2.1. ***Adjacent Transposition.***
Let σ be a permutation and let τ be a transposition swapping two adjacent elements at positions k and $k+1$. Then:

$$N(\sigma \circ \tau) = N(\sigma) \pm 1.$$

Consequently, $\mathrm{sgn}(\sigma \circ \tau) = -\mathrm{sgn}(\sigma)$.

引理

Proof

Let the sequence be $\dots, a, b, \dots$, where $a = \sigma(k)$ and $b = \sigma(k+1)$. After the swap, the sequence becomes $\dots, b, a, \dots$. For any element x not at positions k or $k + 1$, its relative order with a and b remains unchanged. Thus, the only inversion pair affected is $(k, k+1)$ itself.

- If $a < b$, the pair was not an inversion, but becomes one ($b > a$). The count increases by 1.
- If $a > b$, the pair was an inversion, but ceases to be one. The count decreases by 1.

In either case, the parity of the total count reverses.

■

Lemma 2.2. ***General Transposition.***
Let τ be a transposition swapping any two elements. Then $\text{sgn}(\sigma \circ \tau) = -\text{sgn}(\sigma)$.

引理

Proof

Suppose we wish to swap elements at positions i and j with $i < j$. Let $k = j - i - 1$ be the number of elements between them.

$$\dots a \underbrace{x_1 \dots x_k}_{\text{intermediate}} b \dots$$

We can achieve this swap by a sequence of adjacent transpositions: 1. Move a to the right past each x_m until it is adjacent to b. This requires k swaps. 2. Swap a and b. (1 swap) 3. Move b (now at a's old position relative to x) to the left past each x_m. This requires k swaps. The total number of adjacent swaps is $2k + 1$. Since each adjacent swap reverses the sign (*lemma* 2.1), and $2k + 1$ is odd, the total sign change is $(-1)^{2k+1} = -1$.

■

Theorem 2.5. ***Well-definedness of Parity.***
If a permutation σ can be written as a product of m transpositions, then $\text{sgn}(\sigma) = (-1)^m$. Thus, m is always even or always odd for a fixed σ.

定理

Proof

The identity permutation has $N(\text{id}) = 0$, so $\text{sgn}(\text{id}) = +1$. If $\sigma = \tau_m \cdots \tau_1$, we apply the transpositions sequentially to the identity. Each step flips the sign. Thus $\text{sgn}(\sigma) = (-1)^m$.

■

This confirms that the determinant axiom of skew-symmetry is consistent: no matter how one permutes the columns to reach a canonical ordering, the resulting sign change is determined solely by the final permutation.

2.6 *Abstract Volume Forms*

We now lift the concept of the determinant from the specific case of matrices $M_n(\mathbb{R})$ to general n-dimensional vector spaces. This abstraction allows us to discuss volume without reference to a specific coordinate system.

Definition 2.4. ***Volume Form.***
Let V be an n-dimensional vector space over $\mathbb{R}$. A map $\omega : V^n \to \mathbb{R}$ is called a *volume form* (or an alternating n-linear form) if it satisfies:

1. **Multilinearity:** ω is linear in each of its n arguments.
2. **Alternating:** $\omega(\mathbf{v}_1, \ldots, \mathbf{v}_n) = 0$ if $\mathbf{v}_i = \mathbf{v}_j$ for any $i \neq j$.

定義

Note

Recall from *proposition* 2.1 that the alternating property implies skew-symmetry: swapping arguments reverses the sign.

A key result is that the vector space of such volume forms is one-dimensional. In geometric terms: once we decide the volume of a single non-degenerate parallelepiped (a basis), the volumes of all other parallelepipeds are fixed.

Theorem 2.6. ***Uniqueness of Volume Forms.***
Let V be an n-dimensional vector space and let $\mathcal{B} = \{\mathbf{f}_1, \ldots, \mathbf{f}_n\}$ be a basis for V. For any scalar $k \in \mathbb{R}$, there exists a *unique* volume form ω such that

$$\omega(\mathbf{f}_1, \ldots, \mathbf{f}_n) = k.$$

定理

Uniqueness.

Let $\mathbf{u}_1, \ldots, \mathbf{u}_n$ be arbitrary vectors in V. We express them in the basis $\mathcal{B}$:

$$\mathbf{u}_j = \sum_{i=1}^{n} a_{ij}\mathbf{f}_i.$$

Substituting these into ω and using multilinearity:

$$\omega(\mathbf{u}_1, \ldots, \mathbf{u}_n) = \omega\left(\sum_{i_1} a_{i_1 1}\mathbf{f}_{i_1}, \ldots, \sum_{i_n} a_{i_n n}\mathbf{f}_{i_n}\right) = \sum_{i_1, \ldots, i_n} a_{i_1 1} \cdots a_{i_n n}\omega(\mathbf{f}_{i_1}, \ldots, \mathbf{f}_{i_n}).$$

By the alternating property, terms vanish unless indices are distinct. Thus the sum is over permutations σ:

$$\omega(\mathbf{u}_1, \ldots, \mathbf{u}_n) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \left(\prod_{j=1}^{n} a_{\sigma(j)j}\right) \omega(\mathbf{f}_1, \ldots, \mathbf{f}_n).$$

The term in the parenthesis is exactly $\det(A)$, where $A = (a_{ij})$ is the matrix of coordinates of the $\mathbf{u}$'s relative to $\mathcal{B}$. Thus: $\omega(\mathbf{u}_1, \ldots, \mathbf{u}_n) = \det(A) \cdot k$. This formula is determined solely by k and the coordinates, proving uniqueness.

証明終

Existence.

Define $\omega(\mathbf{u}_1, \ldots, \mathbf{u}_n) = k \det(A)$. Since the determinant satisfies the multilinearity and alternating axioms, so does ω. Furthermore, if we input the basis vectors $\mathbf{f}_j$, the coordinate matrix A is the identity I, and $\det(I) = 1$, yielding $\omega(\mathbf{f}_1, \ldots, \mathbf{f}_n) = k$.

証明終

Remark.

If $k \neq 0$, the form ω is non-degenerate. This theorem implies that any two non-zero volume forms ω and ω' on V are proportional: $\omega' = c\omega$ for some $c \neq 0$. This scalar c represents a change in the "unit of measure."

2.7 *Orientation and Euclidean Volume*

In a general vector space, there is no natural "standard" volume; we must arbitrarily choose a basis and assign it a volume (usually 1). However, if the space is equipped with an inner product, the geometry becomes rigid enough to distinguish a canonical volume form.

Oriented Euclidean Spaces

Let V be a Euclidean space (a real vector space with an inner product $\langle \cdot, \cdot \rangle$). Recall that an *orthonormal basis* is a basis $\{\mathbf{e}_i\}$ where $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = \delta_{ij}$. We know that such bases exist. A natural question arises: do all orthonormal bases have the same "volume"?

Theorem 2.7. ***Determinant of Orthonormal Transition.***
Let $\mathcal{E} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be an orthonormal basis for V, and let ω be a volume form such that $\omega(\mathbf{e}_1, \ldots, \mathbf{e}_n) = 1$. If $\mathcal{F} = \{\mathbf{f}_1, \ldots, \mathbf{f}_n\}$ is any other orthonormal basis, then:

$$\omega(\mathbf{f}_1, \ldots, \mathbf{f}_n) = \pm 1.$$

定理

Proof

Let P be the transition matrix from $\mathcal{E}$ to $\mathcal{F}$, so $\mathbf{f}_j = \sum_i p_{ij} \mathbf{e}_i$. By the uniqueness proof above,

$$\omega(\mathbf{f}_1, \ldots, \mathbf{f}_n) = \det(P) \cdot \omega(\mathbf{e}_1, \ldots, \mathbf{e}_n) = \det(P).$$

Since both bases are orthonormal, the matrix P is orthogonal, meaning $P^T P = I$. Taking determinants:

$$\det(P^T P) = \det(P^T) \det(P) = (\det P)^2 = 1.$$

Thus $\det(P) = \pm 1$.

■

This dichotomy allows us to split the set of ordered orthonormal bases into two disjoint classes: those with determinant $+1$ and those with -1.

Definition 2.5. ***Orientation.***
An *orientation* on a Euclidean space V is a choice of one of the two equivalence classes of orthonormal bases. The bases in the chosen class are called *positively oriented* (or right-handed); the others are *negatively oriented* (or left-handed).

定義

An *oriented Euclidean space* is a Euclidean space equipped with a distinguished volume form Vol such that for any positively oriented orthonormal basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$:

$$\text{Vol}(\mathbf{e}_1, \ldots, \mathbf{e}_n) = 1.$$

Note

This reconciles the algebraic determinant with the geometric inner product.

- The inner product $\mathbf{u} \cdot \mathbf{v}$ measures lengths and angles.
- The volume form $[\mathbf{u}_1, \ldots, \mathbf{u}_n]$ measures signed content.

In an oriented Euclidean space, we can compute the volume of any parallelepiped using coordinates relative to any positively oriented orthonormal basis. If $\mathbf{u}_j = (u_{1j}, \ldots, u_{nj})^T$ in such a basis, the volume is simply the determinant of the matrix of components.

2.8 *Cofactor Expansion*

While the Leibniz Formula provides a closed-form expression for the determinant, it is inefficient for practical computation. We now introduce an inductive method known as *Laplace expansion* (or cofactor expansion), which reduces the determinant of an $n \times n$ matrix to a weighted sum of $(n-1) \times (n-1)$ determinants.

Definition 2.6. ***Minors and Cofactors.***
Let $A \in M_n(\mathbb{R})$.

1. The **minor** M_{ij} is the determinant of the $(n-1) \times (n-1)$ submatrix obtained by deleting the i-th row and j-th column of A.
2. The **cofactor** C_{ij} is the signed minor:

$$C_{ij} = (-1)^{i+j} M_{ij}.$$

定義

Theorem 2.8. ***Laplace Expansion.***
The determinant of A can be computed by expanding along any row i or any column j:

$$\det(A) = \sum_{k=1}^{n} a_{ik}C_{ik} \quad \text{(Expansion along row } i\text{)}$$

$$\det(A) = \sum_{k=1}^{n} a_{kj}C_{kj} \quad \text{(Expansion along column } j\text{)}$$

定理

Proof

We prove the column expansion case using the multilinearity axiom. Fix a column index j. We express the j-th column vector $\mathbf{a}_j$ in the standard basis:

$$\mathbf{a}_j = \sum_{k=1}^{n} a_{kj}\mathbf{e}_k.$$

By the linearity of the determinant in the j-th argument:

$$\det(A) = \det([\mathbf{a}_1, \ldots, \sum_{k=1}^{n} a_{kj}\mathbf{e}_k, \ldots, \mathbf{a}_n]) = \sum_{k=1}^{n} a_{kj} \det([\mathbf{a}_1, \ldots, \mathbf{e}_k, \ldots, \mathbf{a}_n]).$$

Let $D_k = \det([\mathbf{a}_1, \ldots, \mathbf{e}_k, \ldots, \mathbf{a}_n])$. This matrix has the standard basis vector $\mathbf{e}_k$ in the j-th column. To compute D_k, we move the j-th column to the n-th position (requiring $n - j$ swaps) and the k-th row to the n-th position (requiring $n - k$ swaps). Let the resulting matrix be A'.

$$A' = \begin{bmatrix} M_{kj}^{\text{sub}} & \mathbf{0} \\ \mathbf{v}^T & 1 \end{bmatrix},$$

where M_{kj}^{sub} is the submatrix used to define the minor M_{kj}. The total number of swaps is $(n - j) + (n - k) = 2n - (j + k)$. Since $2n$ is even, the sign change is $(-1)^{-(j+k)} = (-1)^{j+k}$. Thus, $D_k = (-1)^{j+k} \det(A')$. By the block structure (or inductive definition), $\det(A') = 1 \cdot \det(M_{kj}^{\text{sub}}) = M_{kj}$. Substituting back:

$$\det(A) = \sum_{k=1}^{n} a_{kj}(-1)^{j+k}M_{kj} = \sum_{k=1}^{n} a_{kj}C_{kj}.$$

Row expansion follows immediately from $\det(A) = \det(A^T)$.

■

Example 2.5. Recursive Calculation. Consider the matrix

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

Expanding along the first row:

$$\begin{aligned} \det(A) &= 1 \cdot C_{11} + 2 \cdot C_{12} + 3 \cdot C_{13} \\ &= 1 \begin{vmatrix} 5 & 6 \\ 8 & 9 \end{vmatrix} - 2 \begin{vmatrix} 4 & 6 \\ 7 & 9 \end{vmatrix} + 3 \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix} \\ &= 1(45 - 48) - 2(36 - 42) + 3(32 - 35) \\ &= -3 + 12 - 9 = 0. \end{aligned}$$

This confirms that the rows are linearly dependent (specifically, $R_2 = \frac{1}{2}(R_1 + R_3)$).

範例

Triangular Matrices

A significant consequence of the expansion theorem is the ease of computing determinants for triangular matrices.

Proposition 2.3. ***Determinant of Triangular Matrices.***
If A is an upper triangular, lower triangular, or diagonal matrix, its determinant is the product of its diagonal entries:

$$\det(A) = \prod_{i=1}^{n} a_{ii}.$$

命题

Proof

We proceed by induction on n. For $n = 1$, $\det(A) = a_{11}$. Assume the property holds for $(n-1) \times (n-1)$ matrices. Let A be lower triangular. Expanding along the first row, the only non-zero entry is a_{11} (since $a_{1j} = 0$ for $j > 1$).

$$\det(A) = a_{11}C_{11} = a_{11}(-1)^{1+1}M_{11} = a_{11}M_{11}.$$

The submatrix corresponding to M_{11} is also lower triangular with diagonal entries $a_{22}, \ldots, a_{nn}$. By the inductive hypothesis, $M_{11} = a_{22} \cdots a_{nn}$. Thus $\det(A) = a_{11}a_{22} \cdots a_{nn}$. The proof for upper triangular matrices is identical using column expansion.

■

2.9 *Algebraic Properties*

Note

In this section, we will utilise two key results from our earlier notes on matrix theory:

1. Any invertible matrix can be expressed as a product of elementary matrices.
2. A square matrix with a left or right inverse is automatically invertible.

The geometric definition of the determinant leads to powerful multiplicative properties, connecting the determinant to the group structure of invertible matrices $\mathrm{GL}_n(\mathbb{R})$.

The Product Theorem

Perhaps the most useful property of the determinant is that it preserves multiplication. To prove this, we utilise the decomposition of matrices into elementary matrices.

Lemma 2.3. ***Determinants of Elementary Matrices.***
Let E be an elementary matrix.

1. If E swaps two rows, $\det(E) = -1$.
2. If E scales a row by $k \neq 0$, $\det(E) = k$.
3. If E adds a multiple of one row to another, $\det(E) = 1$.

Moreover, for any matrix B, $\det(EB) = \det(E)\det(B)$.

引理

Proof

proposition 2.2 establishes the effect of row operations on the determinant:

1. Swapping rows negates the determinant. Thus $\det(EB) = -\det(B)$. Since E is obtained by swapping rows of I (where $\det(I) = 1$), $\det(E) = -1$. Hence $\det(EB) = \det(E)\det(B)$.
2. Scaling a row by k scales the determinant by k. $\det(EB) = k\det(B)$ and $\det(E) = k$.
3. Row addition leaves the determinant invariant. $\det(EB) = \det(B)$ and $\det(E) = 1$.

In all cases, the multiplicative property holds.

■

Theorem 2.9. ***The Multiplicative Property.***

For any $A, B \in M_n(\mathbb{R})$:

$$\det(AB) = \det(A)\det(B).$$

定理

We distinguish two cases based on the invertibility of A.

Case 1: A is not invertible.

By the *Determinant Criterion for Independence*, $\det(A) = 0$. If A is singular, then AB is also singular. (If AB were invertible, then $B(AB)^{-1}$ would be a right inverse for A. In our earlier notes, we proved that a square matrix with a right inverse is invertible, which would contradict A being singular). Thus $\det(AB) = 0$, and the equality $0 = 0 \cdot \det(B)$ holds.

証明終

Case 2: A is invertible.

As established in our previous notes on Gaussian elimination, any invertible matrix A can be written as a product of elementary matrices: $A = E_1 E_2 \cdots E_k$. Using *lemma* 2.3 inductively:

$$\begin{aligned}
\det(AB) &= \det(E_1 E_2 \cdots E_k B) \\
&= \det(E_1)\det(E_2 \cdots E_k B) \\
&\vdots \\
&= \det(E_1)\cdots\det(E_k)\det(B).
\end{aligned}$$

Setting $B = I$, we see that $\det(A) = \det(E_1)\cdots\det(E_k)$. Substituting this back yields $\det(AB) = \det(A)\det(B)$.

証明終

Corollary 2.1. *Determinant of the Inverse.* If A is invertible, then

$$\det(A^{-1}) = \frac{1}{\det(A)}.$$

推論

Proof

Since $AA^{-1} = I$, we have $\det(A)\det(A^{-1}) = \det(I) = 1$.

■

Corollary 2.2. *Invariance under Similarity.* If A and B are similar matrices, i.e., $B = P^{-1}AP$ for some invertible P, then $\det(B) = \det(A)$.

推論

Proof

$$\det(P^{-1}AP) = \det(P^{-1})\det(A)\det(P) = \frac{1}{\det(P)}\det(A)\det(P) = \det(A).$$

■

Remark.

This corollary is crucial for linear algebra: it implies that the determinant is an intrinsic property of the *linear operator*, independent of the basis chosen to represent it.

Block Matrices

We conclude this section with a useful result for matrices with block structures.

Proposition 2.4. ***Block Determinant.***
Let M be a block triangular matrix of the form

$$M = \begin{bmatrix} A & B \\ 0 & D \end{bmatrix},$$

where $A \in M_k(\mathbb{R})$ and $D \in M_{n-k}(\mathbb{R})$ are square matrices. Then

$$\det(M) = \det(A)\det(D).$$

命題

Proof

We can factorise M as:

$$\begin{bmatrix} A & B \\ 0 & D \end{bmatrix} = \begin{bmatrix} I_k & 0 \\ 0 & D \end{bmatrix}\begin{bmatrix} A & B \\ 0 & I_{n-k} \end{bmatrix}.$$

Let $M_1 = \begin{bmatrix} I_k & 0 \\ 0 & D \end{bmatrix}$. By applying cofactor expansion along the first k columns (which are standard basis vectors), $\det(M_1) = \det(D)$. Let $M_2 = \begin{bmatrix} A & B \\ 0 & I_{n-k} \end{bmatrix}$. By expanding along the last $n - k$ rows, $\det(M_2) = \det(A)$. By the Product Theorem, $\det(M) = \det(M_1)\det(M_2) = \det(D)\det(A)$.

■

Example 2.6. Block Example. Compute the determinant of:

$$M = \begin{bmatrix} 1 & 2 & 9 & 8 \\ 3 & 4 & 7 & 6 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 2 & 3 \end{bmatrix}.$$

Here $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ and $D = \begin{bmatrix} 5 & 0 \\ 2 & 3 \end{bmatrix}$. The zero block is in the bottom-left.

$$\det(M) = \det(A)\det(D) = (4-6)(15-0) = (-2)(15) = -30.$$

範例

2.10 *The Adjugate Matrix*

We have seen that the inverse of a matrix can be computed via Gaussian elimination or expressed element-wise using Cramer's rule. We now derive a closed-form algebraic expression for the inverse, known as the *adjugate* formula. This formula is of significant theoretical interest, though it is computationally expensive for large matrices.

Definition 2.7. ***Adjugate Matrix.***
Let $A \in M_n(\mathbb{R})$. The **adjugate** of A, denoted $\operatorname{adj}(A)$, is the transpose of the matrix of cofactors. That is,

$$(\operatorname{adj}(A))_{ij} = C_{ji} = (-1)^{i+j} M_{ji}.$$

定義

Remark.

Historically, this matrix was called the "adjoint". However, in modern linear algebra, "adjoint" typically refers to the Hermitian conjugate A^* (or $A^\dagger$) in the context of inner product spaces. To avoid ambiguity, we use "adjugate".

Example 2.7. 2×2 Adjugate. For $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, the cofactors are:

$$C_{11} = d, \quad C_{12} = -c, \quad C_{21} = -b, \quad C_{22} = a.$$

The matrix of cofactors is $\begin{bmatrix} d & -c \\ -b & a \end{bmatrix}$. Transposing gives:

$$\operatorname{adj}(A) = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

範例

The fundamental property of the adjugate is that it annihilates the matrix A up to a scalar factor.

Theorem 2.10. ***Adjugate Identity.***
For any $A \in M_n(\mathbb{R})$:

$$A\,\mathrm{adj}(A) = \mathrm{adj}(A)\,A = \det(A)I.$$

定理

Consider the product $P = A\,\mathrm{adj}(A)$. The (i, j)-th entry is the dot product of the i-th row of A and the j-th column of $\mathrm{adj}(A)$ (which is the j-th row of the cofactor matrix).

$$P_{ij} = \sum_{k=1}^{n} a_{ik}(\mathrm{adj}(A))_{kj} = \sum_{k=1}^{n} a_{ik}C_{jk}.$$

We analyse two cases:

Diagonal entries (i = j)

$$P_{ii} = \sum_{k=1}^{n} a_{ik}C_{ik}.$$

This is exactly the cofactor expansion of $\det(A)$ along the i-th row. Thus, $P_{ii} = \det(A)$.

証明終

Off-diagonal entries (i ≠ j)

$$P_{ij} = \sum_{k=1}^{n} a_{ik}C_{jk}.$$

This sum represents the determinant of a matrix A' obtained from A by replacing the j-th row with a copy of the i-th row. Since A' has two identical rows, $\det(A') = 0$. Thus, $P_{ij} = 0$.

証明終

Consequently, P is a diagonal matrix with $\det(A)$ on the diagonal, i.e., $P = \det(A)I$.

Corollary 2.3. Formula for the Inverse

推論

If A is invertible ($\det(A) \neq 0$), then:

$$A^{-1} = \frac{1}{\det(A)}\mathrm{adj}(A).$$

Proof

From *theorem 2.10*, $A\left(\frac{1}{\det(A)}\operatorname{adj}(A)\right) = I$. Since the inverse is unique, the result follows.

■

2.11 *Applications of the Determinant*

The determinant is not merely a criterion for invertibility; it encodes essential geometric and algebraic information about linear transformations and systems of vectors.

Volume and Linear Maps

As established in the geometric introduction, the determinant measures the scaling factor of volume.

Theorem 2.11. ***Volume of a Parallelepiped.***
Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear operator represented by the matrix A. If P is a parallelepiped in $\mathbb{R}^n$, then:

$$\operatorname{Vol}(T(P)) = |\det(A)| \cdot \operatorname{Vol}(P).$$

定理

Proof

It suffices to verify this for the unit hypercube generated by the standard basis vectors. Its image is the parallelepiped spanned by the columns of A. By definition, the volume of this parallelepiped is $|\det(A)|$. Since T is linear, the scaling factor applies to any parallelepiped (which is the image of the unit hypercube under some linear map).

■

Remark.

Using measure theory, this result extends to any measurable set S, giving $\operatorname{Vol}(T(S)) = |\det(A)| \cdot \operatorname{Vol}(S)$. We restrict our attention to parallelepipeds here.

Example 2.8. Area of a Polygon. Consider a parallelogram in $\mathbb{R}^2$ defined by a linear transformation of the unit square. If the vertices are $(0,0), (3,0), (0,1), (3,1)$ (a rectangle of area 3) and we apply $T(x,y) = (2x, x+y)$, the matrix is $A = \begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix}$ with $\det(A) = 2$. The new area is $2 \times 3 = 6$.

範例

Polynomial Interpolation

The determinant naturally appears in problems involving polynomial fitting. The condition for the existence of a unique polynomial passing through n points leads to the *Vandermonde determinant*.

Definition 2.8. ***Vandermonde Matrix.***
Given n scalars $x_1, \dots, x_n$, the Vandermonde matrix V is defined by $V_{ij} = x_i^{j-1}$:

$$V = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-1} \end{bmatrix}.$$

定義

Theorem 2.12. ***Vandermonde Determinant.***

$$\det(V) = \prod_{1 \le i < j \le n} (x_j - x_i).$$

定理

Proof

Let $\Delta_n(x_1, \dots, x_n) = \det(V)$. We proceed by induction on n. For $n = 1$, $\Delta_1 = \det[1] = 1$, which agrees with the empty product.
Assume $n \ge 2$ and that the formula holds for $n - 1$. Perform the column operations $C_k \to C_k - x_1 C_{k-1}$ for $k = n, n-1, \dots, 2$. These operations preserve the determinant, so the new matrix V' satisfies $\det(V') = \Delta_n$.
For $k \ge 2$, the (i, k)-entry of V' is

$$x_i^{k-1} - x_1 x_i^{k-2} = x_i^{k-2}(x_i - x_1).$$

When $i = 1$, this entry is $x_1^{k-2}(x_1 - x_1) = 0$, so the first row of V' is $(1, 0, \dots, 0)$. Expanding along the first row yields $\Delta_n = \det(W)$, where W is the $(n-1) \times (n-1)$ matrix obtained by deleting the first row and column of V'. Indexing rows by $i = 2, \dots, n$ and columns by $k = 2, \dots, n$, the entries are

$$W_{i,k} = x_i^{k-2}(x_i - x_1).$$

Each row i of W contains the common factor $(x_i - x_1)$. By multilinearity in the rows,

$$\det(W) = \left(\prod_{i=2}^{n} (x_i - x_1) \right) \det(U),$$

where U is the matrix obtained by dividing the i-th row of W by

$(x_i - x_1)$. Thus

$$U = \begin{bmatrix} 1 & x_2 & x_2^2 & \cdots & x_2^{n-2} \\ 1 & x_3 & x_3^2 & \cdots & x_3^{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{n-2} \end{bmatrix},$$

so $\det(U) = \Delta_{n-1}(x_2, \dots, x_n)$. Therefore

$$\Delta_n = \left(\prod_{i=2}^{n}(x_i - x_1)\right) \Delta_{n-1}(x_2, \dots, x_n).$$

Applying the induction hypothesis,

$$\Delta_{n-1}(x_2, \dots, x_n) = \prod_{2 \le i < j \le n} (x_j - x_i),$$

hence

$$\Delta_n = \left(\prod_{i=2}^{n}(x_i - x_1)\right) \left(\prod_{2 \le i < j \le n} (x_j - x_i)\right) = \prod_{1 \le i < j \le n} (x_j - x_i).$$

This completes the proof.

■

This result implies that V is invertible if and only if all x_i are distinct. This guarantees that there is a unique polynomial of degree $n-1$ passing through any n points with distinct x-coordinates.

Example 2.9. Equation of a Line. Three points (x, y), (x_1, y_1), and (x_2, y_2) are collinear if and only if the triangle they form has zero area. This condition is expressed by the vanishing of the determinant:

$$\det \begin{bmatrix} 1 & x & y \\ 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \end{bmatrix} = 0.$$

Performing row operations $R_2 \to R_2 - R_1$ and $R_3 \to R_3 - R_1$ leads to the two-point form of the line equation.

範例

2.12 *Exercises*

In the following exercises, assume all matrices are square and have real entries unless specified otherwise. You may use the properties established in the chapter (multilinearity, alternating property, product theorem, etc.).

1. **Elementary Row Operations.** Using elementary row operations to introduce zeros, compute the determinant of the following matrix:

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

2. **Solving for Singularity.** Find all values of x for which the following matrix is singular (i.e., has determinant zero):

Identify the structure of this matrix to solve it by inspection, rather than brute-force expansion.

$$M(x) = \begin{bmatrix} 1 & x & x^2 \\ 1 & 2 & 4 \\ 1 & 5 & 25 \end{bmatrix}.$$

3. **Determinant Arithmetic.** Let A and B be 4×4 matrices with $\det(A) = 3$ and $\det(B) = -2$. Compute the following values:
   (a) $\det(AB^T)$
   (b) $\det(2A^{-1})$
   (c) $\det(A^3 B^{-1} A^T)$
   (d) $\det(\text{adj}(A))$

4. **Permutation Parity.** Consider the permutation $\sigma \in S_8$ defined by:

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

   (a) Determine the number of inversions $N(\sigma)$.
   (b) Decompose σ into a product of disjoint cycles.
   (c) Determine the sign $\text{sgn}(\sigma)$.

5. **The Arrow Matrix.** Compute the determinant of the following $n \times n$ "arrow" matrix, which has entries a on the diagonal, ones in the last column and last row, and zeros elsewhere (except the (n, n) entry):

Use row operations to eliminate the ones in the last column.

$$D_n = \det \begin{bmatrix} a & 0 & \cdots & 0 & 1 \\ 0 & a & \cdots & 0 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & a & 1 \\ 1 & 1 & \cdots & 1 & 1 \end{bmatrix}.$$

6. **Skew-Symmetric Matrices.** A matrix A is called *skew-symmetric* if $A^T = -A$.
   (a) Prove that if A is a skew-symmetric matrix of *odd* dimension n, then $\det(A) = 0$.
   (b) Give an example of a 2×2 skew-symmetric matrix with non-zero determinant.

7. **The Matrix Determinant Lemma.** Let A be an invertible $n \times n$ matrix, and let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ be column vectors.
   (a) Using the block matrix identity $\begin{bmatrix} I & 0 \\ \mathbf{v}^T & 1 \end{bmatrix} \begin{bmatrix} I + \mathbf{u}\mathbf{v}^T & \mathbf{u} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} I & 0 \\ -\mathbf{v}^T & 1 \end{bmatrix} = \begin{bmatrix} I & \mathbf{u} \\ 0 & 1 + \mathbf{v}^T\mathbf{u} \end{bmatrix}$ (or a similar decomposition), prove that:

   $$\det(I + \mathbf{u}\mathbf{v}^T) = 1 + \mathbf{v}^T\mathbf{u}.$$

   (b) Deduce that for an invertible A, $\det(A + \mathbf{u}\mathbf{v}^T) = \det(A)(1 + \mathbf{v}^T A^{-1}\mathbf{u})$.
8. **Integrality of the Inverse.** Let A be a matrix with integer entries. Prove that A^{-1} exists and has integer entries if and only if $\det(A) = \pm 1$.

   Use the Adjugate Formula for the sufficient condition.
9. **Orthogonal Matrices.** A real matrix Q is *orthogonal* if $Q^T Q = I$.
   (a) Prove that $\det(Q) \in \{1, -1\}$.
   (b) Geometric Interpretation: If $\det(Q) = 1$, Q represents a rotation. If $\det(Q) = -1$, Q represents a reflection (or rotation-reflection). Verify this for 2×2 matrices.
10. **The Adjugate Determinant.** Let A be an $n \times n$ matrix.
   (a) Prove that $\det(\operatorname{adj}(A)) = (\det(A))^{n-1}$.
   (b) Prove that if $\det(A) \neq 0$, then $\operatorname{adj}(\operatorname{adj}(A)) = (\det(A))^{n-2} A$.

   Treat the singular and non-singular cases separately. For the singular case, argue that $\operatorname{adj}(A)$ cannot be invertible.
11. **Anti-Commuting Matrices.** Let A, B be $n \times n$ matrices such that $AB = -BA$.
   (a) Prove that if n is odd, then at least one of A or B must be singular.
   (b) Show that for $n = 2$, it is possible for both A and B to be non-singular.
12. **Schur Complement.** Let $M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ be a block matrix where A is an invertible $k \times k$ matrix.
   (a) Verify the decomposition:

   $$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}.$$

   (b) Deduce that $\det(M) = \det(A)\det(D - CA^{-1}B)$.
   (c) If A and C commute (i.e., $AC = CA$), prove that $\det(M) = \det(AD - CB)$.
13. **Tridiagonal Recurrence.** Let T_n be the determinant of the $n \times n$

tridiagonal matrix defined by:

$$T_n = \det \begin{bmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ 0 & -1 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 2 \end{bmatrix}.$$

(a) By expanding along the first row, derive the recurrence relation $T_n = 2T_{n-1} - T_{n-2}$.

(b) Calculate T_1 and T_2, and solve the recurrence to find a closed form formula for T_n.

14. Nilpotent Matrices. A matrix A is called *nilpotent* if there exists an integer $k \geq 1$ such that $A^k = 0$.

(a) Prove that if A is nilpotent, then $\det(A) = 0$.

(b) Prove that if A is nilpotent, then $\det(I + A) = 1$.

For part (b), consider the geometric series expansion or algebraic factors. Or, if you prefer a contrapositive approach, assume $\det(I + A) \neq 1$ and consider the eigenvalues (if known) or characteristic polynomial properties implied by nilpotency.

3
Linear Transformations

In the preceding chapters, we treated matrices primarily as static algebraic objects or arrays of numbers used to compute determinants and solve systems of linear equations. We now adopt a dynamic perspective, viewing matrices as operators that transform vectors. This shift in perspective from *state* to *process* is central to modern linear algebra. The determinant, previously defined axiomatically, will be understood as a measure of how these transformations distort volume.

3.1 *Definition and Elementary Properties*

We begin by formalising the notion of a structure-preserving map between vector spaces. While we principally work with $\mathbb{R}^n$, the definitions apply to general vector spaces.

A key insight from 19th century mathematics: study sets with structure and the maps preserving it. Linearity means the order of operations yields the same result.

> **Definition 3.1.** ***Linear Transformation.***
> Let V and W be vector spaces over $\mathbb{R}$. A mapping $T : V \to W$ is called a *linear transformation* (or linear map) if it satisfies the following two conditions for all $\mathbf{u}, \mathbf{v} \in V$ and scalars $c \in \mathbb{R}$:
> 1. **Additivity:** $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$.
> 2. **Homogeneity:** $T(c\mathbf{u}) = cT(\mathbf{u})$.
>
> These conditions can be combined into a single requirement:
>
> $$T(c\mathbf{u} + \mathbf{v}) = cT(\mathbf{u}) + T(\mathbf{v}).$$
>
> 定義

The words *map*, *mapping*, *function*, and *transformation* are synonymous; *transformation* is preferred in linear algebra.

The linearity conditions imply strong constraints on the behaviour of the map. The most immediate consequence is the preservation of the zero vector.

> **Proposition 3.1.** ***Preservation of the Origin.***
> If $T : V \to W$ is a linear transformation, then $T(\mathbf{0}_V) = \mathbf{0}_W$.
>
> 命題

Proof

Let $\mathbf{u} \in V$. Using additivity:

$$T(\mathbf{0}_V) = T(\mathbf{u} - \mathbf{u}) = T(\mathbf{u}) + T(-1 \cdot \mathbf{u}) = T(\mathbf{u}) - T(\mathbf{u}) = \mathbf{0}_W.$$

Alternatively, using homogeneity with $c = 0$: $T(\mathbf{0}) = T(0\mathbf{u}) = 0T(\mathbf{u}) = \mathbf{0}$.

■

A linear transformation preserves the structure of linear combinations indefinitely.

Proposition 3.2. ***Generalised Linearity.***
Let $T : V \to W$ be a linear transformation. For any scalars $c_1, \ldots, c_k$ and vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in V$:

$$T\left(\sum_{i=1}^{k} c_i \mathbf{v}_i\right) = \sum_{i=1}^{k} c_i T(\mathbf{v}_i).$$

命題

Proof

We proceed by induction on k. The base cases $k = 1, 2$ follow directly from the definition of linearity. Assume the property holds for k vectors. Consider a linear combination of $k+1$ vectors:

$$\mathbf{u} = \sum_{i=1}^{k+1} c_i \mathbf{v}_i = \left(\sum_{i=1}^{k} c_i \mathbf{v}_i\right) + c_{k+1}\mathbf{v}_{k+1}.$$

By the additivity axiom and the inductive hypothesis:

$$\begin{aligned} T(\mathbf{u}) &= T\left(\sum_{i=1}^{k} c_i \mathbf{v}_i\right) + T(c_{k+1}\mathbf{v}_{k+1}) \\ &= \sum_{i=1}^{k} c_i T(\mathbf{v}_i) + c_{k+1} T(\mathbf{v}_{k+1}) = \sum_{i=1}^{k+1} c_i T(\mathbf{v}_i). \end{aligned}$$

By the Principle of Mathematical Induction, the property holds for all $k \in \mathbb{N}$.

■

This proposition implies that a linear transformation maps subspaces to subspaces. Specifically, it maps the span of a set of vectors in the domain to the span of their images in the codomain.

The codomain W specifies the type of output (e.g., lists of m numbers), not which outputs actually occur — that is the *image*.

Proposition 3.3. ***Mapping of Spans.***
Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \subset V$. Then:

$$T(\text{span}(S)) = \text{span}\{T(\mathbf{v}_1), \ldots, T(\mathbf{v}_k)\}.$$

命題

Proof

Let $\mathbf{y} \in T(\text{span}(S))$. Then $\mathbf{y} = T(\mathbf{x})$ for some $\mathbf{x} \in \text{span}(S)$. Writing $\mathbf{x} = \sum c_i \mathbf{v}_i$, we have $\mathbf{y} = T(\sum c_i \mathbf{v}_i) = \sum c_i T(\mathbf{v}_i)$, which is in $\text{span}(T(S))$. Conversely, any element in $\text{span}(T(S))$ is of the form $\sum c_i T(\mathbf{v}_i)$. By linearity, this equals $T(\sum c_i \mathbf{v}_i)$, which is in $T(\text{span}(S))$.

■

Remark.

This result is geometrically significant. If $T : \mathbb{R}^n \to \mathbb{R}^m$, it ensures that lines through the origin map to lines (or points), and planes through the origin map to planes (or lines/points). The dimension of the image cannot exceed the dimension of the domain.

Example 3.1. Economic Linearity. Consider a simplified economic model of a checkout counter, defined by a map $T : \mathbb{R}^n \to \mathbb{R}$. The input $\mathbf{v} \in \mathbb{R}^n$ represents the quantities of n distinct products in a shopping cart. The output $T(\mathbf{v})$ is the total cost. If the pricing is strictly per-unit (no bulk discounts), the map is linear.

1. **Additivity:** $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$. The cost of two combined carts is the sum of their individual costs.

2. **Homogeneity:** $T(c\mathbf{v}) = cT(\mathbf{v})$. Purchasing c times the quantity results in c times the cost.

Real-world systems often violate linearity (e.g., "buy one get one free" implies $T(2\mathbf{v}) < 2T(\mathbf{v})$), but linearity remains the fundamental first-order approximation in modelling.

範例

3.2 *The Matrix Representation Theorem*

The abstract definition of a linear transformation is theoretically clean, but for computation in finite-dimensional spaces, we rely on matrix representations. We now establish the fundamental correspondence between linear transformations $T : \mathbb{R}^n \to \mathbb{R}^m$ and $m \times n$ matrices.

Theorem 3.1. ***Matrix Induced Maps.***
Let $A \in M_{m \times n}(\mathbb{R})$. The function $T : \mathbb{R}^n \to \mathbb{R}^m$ defined by matrix-vector multiplication,

$$T(\mathbf{x}) = A\mathbf{x},$$

is a linear transformation.

定理

Proof

This follows directly from the algebraic properties of matrix multiplication. For any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $c \in \mathbb{R}$:

$$A(\mathbf{x}+\mathbf{y}) = A\mathbf{x} + A\mathbf{y} \implies T(\mathbf{x}+\mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y}),$$
$$A(c\mathbf{x}) = c(A\mathbf{x}) \implies T(c\mathbf{x}) = cT(\mathbf{x}).$$

Thus T satisfies *definition* 3.1.

■

The converse is a profound result: *every* linear transformation between Euclidean spaces is a matrix transformation. Furthermore, the matrix is constructed explicitly by the action of T on the standard basis.

Theorem 3.2. ***The Standard Matrix Representation.***
Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. There exists a unique $m \times n$ matrix A, denoted by $[T]$, such that for all $\mathbf{x} \in \mathbb{R}^n$:

$$T(\mathbf{x}) = [T]\mathbf{x}.$$

We write vectors as columns so this reads naturally. With row vectors, we would need $T(\mathbf{x}) = \mathbf{x}[T]^\top$.

The columns of $[T]$ are the images of the standard basis vectors $\mathbf{e}_1, \dots, \mathbf{e}_n$ of $\mathbb{R}^n$:

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

定理

Proof

Let $\mathbf{x} \in \mathbb{R}^n$. We can expand $\mathbf{x}$ uniquely in the standard basis:

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n = \sum_{j=1}^{n} x_j\mathbf{e}_j.$$

Applying T and utilising *proposition* 3.2:

$$T(\mathbf{x}) = T\left(\sum_{j=1}^{n} x_j\mathbf{e}_j\right) = \sum_{j=1}^{n} x_j T(\mathbf{e}_j).$$

Recall the definition of matrix-vector multiplication. If A is the matrix with columns $\mathbf{c}_j = T(\mathbf{e}_j)$, then $A\mathbf{x}$ is precisely the linear combination of columns:

$$A\mathbf{x} = x_1\mathbf{c}_1 + \cdots + x_n\mathbf{c}_n = \sum_{j=1}^{n} x_j T(\mathbf{e}_j).$$

Comparing the two expressions, we see that $T(\mathbf{x}) = A\mathbf{x}$.

■

Remark.

This theorem reduces the study of linear transformations on finite-dimensional spaces to the study of matrices. The seemingly abstract "black box" of the function is fully characterised by its outputs on the n basis vectors.

Example 3.2. Constructing a Transformation. Suppose we wish to define a linear map $T : \mathbb{R}^2 \to \mathbb{R}^2$ that rotates the plane by 90° counter-clockwise. We determine the standard matrix by tracking the basis vectors:

1. $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ rotates to $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$.
2. $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ rotates to $\begin{bmatrix} -1 \\ 0 \end{bmatrix}$.

Thus, the matrix is:

$$[T] = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

To rotate an arbitrary vector $\mathbf{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$, we simply multiply:

$$T(\mathbf{v}) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} -3 \\ 2 \end{bmatrix}.$$

範例

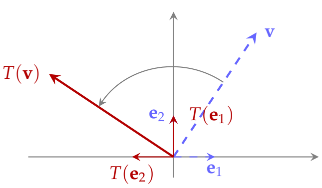

Figure 3.1: A 90° counter-clockwise rotation. Original vectors (dashed blue) map to their images (solid red).

3.3 *Geometric Properties of Linear Maps*

Linear transformations are rigid in their treatment of linear structures. We have already seen that they preserve subspaces. We now examine their effect on affine structures, such as lines not passing through the origin.

Proposition 3.4. ***Preservation of Line Segments.***
Let L be the line segment connecting vectors $\mathbf{p}$ and $\mathbf{q}$ in $\mathbb{R}^n$.

$$L = \{\mathbf{p} + t(\mathbf{q} - \mathbf{p}) \mid t \in [0, 1]\}.$$

If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, the image $T(L)$ is the line segment connecting $T(\mathbf{p})$ and $T(\mathbf{q})$.

命題

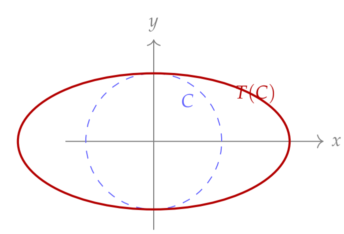

Figure 3.2: The matrix $\left[\begin{smallmatrix} 2 & 0 \\ 0 & 1 \end{smallmatrix}\right]$ stretches the unit circle into an ellipse.

Proof

Let $\mathbf{y} \in T(L)$. Then there exists $\mathbf{x} \in L$ such that $\mathbf{y} = T(\mathbf{x})$. We write $\mathbf{x} = \mathbf{p} + t(\mathbf{q} - \mathbf{p})$ for some $t \in [0, 1]$. By linearity:

$$\mathbf{y} = T(\mathbf{p} + t(\mathbf{q} - \mathbf{p})) = T(\mathbf{p}) + t(T(\mathbf{q}) - T(\mathbf{p})).$$

This is precisely the parameterisation of the segment from $T(\mathbf{p})$ to $T(\mathbf{q})$.

■

Note

If $T(\mathbf{p}) = T(\mathbf{q})$, the segment degenerates to a single point.

Example 3.3. Shear Transformations. Consider the shear transformation $S : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $S(\mathbf{e}_1) = \mathbf{e}_1$ and $S(\mathbf{e}_2) = \mathbf{e}_2 + k\mathbf{e}_1$. The matrix is:

$$[S] = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}.$$

A vertical line segment connecting $(1, 0)$ and $(1, 1)$ is mapped to the segment connecting $S(1, 0) = (1, 0)$ and $S(1, 1) = (1 + k, 1)$. The vertical line is "tilted" or sheared, but it remains a straight line segment. As noted in the previous chapter on determinants, such operations preserve area, consistent with $\det([S]) = 1$.

範例

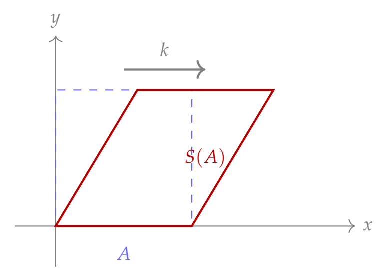

Figure 3.3: A horizontal shear with $k = 0.6$. The unit square (dashed) maps to a parallelogram of equal area.

Coordinate Vectors and Change of Basis

Note

While *theorem 3.2* defines the matrix with respect to the *standard* basis, linearity applies to *any* basis. This foreshadows the general theory of similarity transformations.

Let $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis for $\mathbb{R}^n$. Any vector $\mathbf{x}$ has a unique coordinate representation relative to $\mathcal{B}$, denoted $[\mathbf{x}]_{\mathcal{B}} = (c_1, \ldots, c_n)^T$, such that $\mathbf{x} = \sum c_i \mathbf{v}_i$. The linearity of T ensures that knowledge of $T(\mathbf{v}_i)$ is sufficient to determine $T(\mathbf{x})$ for all $\mathbf{x}$. Specifically:

$$T(\mathbf{x}) = \sum_{i=1}^{n} c_i T(\mathbf{v}_i).$$

This equation expresses the Fundamental Theorem of Linear Algebra in its coordinate-free guise: a linear map is completely determined by its action on a basis.

3.4 *Injectivity and Surjectivity*

Having established the correspondence between linear transformations and matrices via the *The Standard Matrix Representation*, we now turn to the structural properties of these maps. Specifically, we investigate how the algebraic properties of the matrix $[T]$ dictate the geometric behaviour of the transformation T, particularly regarding the uniqueness and existence of solutions to the equation $T(\mathbf{x}) = \mathbf{b}$.

Injectivity asks: "Is the solution unique?" Surjectivity asks: "Does a solution exist?" A bijective map answers both affirmatively.

The Kernel and Injectivity

A fundamental question for any function is whether it preserves distinctness of inputs. In the context of linear algebra, this property (injectivity), is inextricably linked to the mapping of the zero vector.

Definition 3.2. ***Kernel and Image.***
Let $T : V \to W$ be a linear transformation.

1. The **kernel** (or nullspace) of T, denoted $\ker(T)$, is the set of all vectors in V that map to the zero vector in W:

$$\ker(T) = \{\mathbf{v} \in V \mid T(\mathbf{v}) = \mathbf{0}_W\}.$$

2. The **image** (or range) of T, denoted $\operatorname{Im}(T)$, is the set of all vectors in W that are images of vectors in V:

$$\operatorname{Im}(T) = \{T(\mathbf{v}) \mid \mathbf{v} \in V\}.$$

定義

Theorem 3.3. ***Kernel Criterion for Injectivity.***
A linear transformation $T : V \to W$ is injective (one-to-one) if and only if its kernel is trivial, i.e.,

$$\ker(T) = \{\mathbf{0}_V\}.$$

定理

We prove both implications.

($\Leftarrow$)
Assume $\ker(T) = \{\mathbf{0}_V\}$. Suppose $T(\mathbf{u}) = T(\mathbf{v})$ for some $\mathbf{u}, \mathbf{v} \in V$. By the linearity of T:

$$T(\mathbf{u}) - T(\mathbf{v}) = \mathbf{0}_W \implies T(\mathbf{u} - \mathbf{v}) = \mathbf{0}_W.$$

This implies that $\mathbf{u} - \mathbf{v} \in \ker(T)$. By our assumption, the only element in the kernel is $\mathbf{0}_V$, so $\mathbf{u} - \mathbf{v} = \mathbf{0}_V$, which yields $\mathbf{u} = \mathbf{v}$. Thus, T is injective.

証明終

($\Rightarrow$)

Conversely, assume T is injective. We know from elementary properties that $T(\mathbf{0}_V) = \mathbf{0}_W$. If $\mathbf{x} \in \ker(T)$, then $T(\mathbf{x}) = \mathbf{0}_W = T(\mathbf{0}_V)$. Since T is injective, it must be that $\mathbf{x} = \mathbf{0}_V$. Thus, $\ker(T)$ contains only the zero vector.

証明終

Remark.

This theorem highlights a stark difference between linear and non-linear maps. For a general function like $f(x) = x^2$, knowing that $f(0) = 0$ is insufficient to determine injectivity. Linearity ensures that the local behaviour at the origin dictates the global behaviour of the map.

Matrix Characterisation

When $T : \mathbb{R}^n \to \mathbb{R}^m$ is represented by a matrix $A = [T]$, the abstract notions of injectivity and surjectivity translate directly into conditions on the columns of A.

Theorem 3.4. ***Matrix Rank and Transformation Properties.***
Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation with standard matrix $A = [T]$.

1. T is **injective** if and only if the columns of A are linearly independent.
2. T is **surjective** (onto $\mathbb{R}^m$) if and only if the columns of A span $\mathbb{R}^m$.

定理

Proof

For (1), recall from *theorem 3.3* that T is injective if and only if $T(\mathbf{x}) = \mathbf{0}$ implies $\mathbf{x} = \mathbf{0}$. In matrix terms, this equation is $A\mathbf{x} = \mathbf{0}$. Let $\mathbf{c}_1, \ldots, \mathbf{c}_n$ be the columns of A. The matrix-vector product is a linear combination of these columns:

$$x_1\mathbf{c}_1 + \cdots + x_n\mathbf{c}_n = \mathbf{0}.$$

The condition that the only solution is the trivial solution $\mathbf{x} = \mathbf{0}$ is precisely the definition of linear independence for the set $\{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$.
For (2), surjectivity means that for every $\mathbf{b} \in \mathbb{R}^m$, there exists $\mathbf{x} \in \mathbb{R}^n$ such that $T(\mathbf{x}) = \mathbf{b}$. Equivalently, the system $A\mathbf{x} = \mathbf{b}$ must be consistent for all $\mathbf{b}$. Since $A\mathbf{x}$ lies in the span of the columns of A, the condition that a solution exists for *every* $\mathbf{b}$ is equivalent to saying that the column space of A is the entire codomain $\mathbb{R}^m$.

■

Remark.

This connects directly to the Rank-Nullity Theorem. The dimension of the image, $\dim(\mathrm{Im}(T))$, is the rank of A. If the rank equals m, the map is surjective. If the rank equals n (implying the nullity is 0), the map is injective.

Example 3.4. Analyzing a Transformation in $\mathbb{R}^4$. Consider the linear transformation $L : \mathbb{R}^4 \to \mathbb{R}^4$ defined by:

$$L(t, x, y, z) = (t + x + y + z,\ z - x,\ 0,\ 3t - z).$$

To analyse this map, we first construct its standard matrix $[L]$ by evaluating L on the standard basis vectors of $\mathbb{R}^4$.

$$\begin{aligned} L(\mathbf{e}_1) &= L(1,0,0,0) = (1,0,0,3)^T \\ L(\mathbf{e}_2) &= L(0,1,0,0) = (1,-1,0,0)^T \\ L(\mathbf{e}_3) &= L(0,0,1,0) = (1,0,0,0)^T \\ L(\mathbf{e}_4) &= L(0,0,0,1) = (1,1,0,-1)^T \end{aligned}$$

Ideally, we arrange these as the columns of the matrix A:

$$A = [L] = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & -1 \end{bmatrix}.$$

We observe immediately that the third row consists entirely of zeros. This implies that for any vector $\mathbf{v}$, the third component of $L(\mathbf{v})$ is always 0. Consequently, L cannot map to any vector with a non-zero third component (e.g., $\mathbf{e}_3$). Thus, the columns do not span $\mathbb{R}^4$, and L is **not surjective**.
To check injectivity, we examine the linear independence of the columns.

範例

Note

The matrix is square; since it is not surjective (rank $<$ 4), it cannot be injective either (by the Rank-Nullity Theorem (once again see previous notes) or determinant properties). Explicitly, we can find a non-trivial element in the kernel by solving $A\mathbf{x} = \mathbf{0}$, revealing the dependency among columns.

Example 3.5. A Surjective Map. Consider the map $T : \mathbb{R}^3 \to \mathbb{R}^2$

given by $T(x, y, z) = (x + 2y, 3y + 4z)$. The standard matrix is:

$$[T] = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 3 & 4 \end{bmatrix}.$$

The columns are $\mathbf{c}_1 = (1, 0)^T$, $\mathbf{c}_2 = (2, 3)^T$, $\mathbf{c}_3 = (0, 4)^T$. Since $\mathbf{c}_1$ and $\mathbf{c}_3$ are clearly linearly independent (and form a basis for $\mathbb{R}^2$), the column space is $\mathbb{R}^2$. Thus, T is **surjective**. However, since there are 3 vectors in $\mathbb{R}^2$, they must be linearly dependent. Thus, T is **not injective**.

範例

3.5 *Composition and Invertibility*

The algebraic power of matrices stems from the fact that matrix multiplication corresponds to the composition of linear maps. If $S : U \to V$ and $T : V \to W$ are linear transformations represented by matrices B and A respectively, then the composite map $T \circ S : U \to W$ is linear and is represented by the product AB.

The order matters: $T \circ S$ means "apply S first, then T." In matrix form, this is $[T][S]$, reading right-to-left.

Proposition 3.5. ***Composition Property.***
Let S and T be linear transformations with standard matrices $[S]$ and $[T]$ respectively. Then:

$$[T \circ S] = [T][S].$$

命題

Proof

For any vector $\mathbf{x}$ in the domain of S:

$$(T \circ S)(\mathbf{x}) = T(S(\mathbf{x})) = T([S]\mathbf{x}) = [T]([S]\mathbf{x}) = ([T][S])\mathbf{x}.$$

By the uniqueness of the standard matrix (*theorem* 3.2), the matrix representing $T \circ S$ must be $[T][S]$.

■

This property allows us to seamlessly transfer the concept of matrix invertibility to linear transformations.

Theorem 3.5. ***Invertibility of Linear Transformations.***
Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation with standard matrix $A = [T]$. Then T is invertible if and only if A is invertible. Furthermore, the inverse transformation T^{-1} is induced by the inverse matrix:

$$[T^{-1}] = A^{-1}.$$

定理

(⇒)

Suppose T is invertible. Then there exists a map $T^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ such that $T \circ T^{-1} = \text{id}$ and $T^{-1} \circ T = \text{id}$. Let $B = [T^{-1}]$. Using the composition property:

$$[T \circ T^{-1}] = [T][T^{-1}] = AB.$$

Since the standard matrix of the identity map is the identity matrix I_n, we have $AB = I_n$. Similarly, $BA = I_n$. Thus A is invertible and $A^{-1} = B$.

証明終

(⇐)

Suppose A is invertible. Consider the linear transformation L induced by A^{-1}.

$$(T \circ L)(\mathbf{x}) = A(A^{-1}\mathbf{x}) = (AA^{-1})\mathbf{x} = I_n\mathbf{x} = \mathbf{x}.$$

$$(L \circ T)(\mathbf{x}) = A^{-1}(A\mathbf{x}) = (A^{-1}A)\mathbf{x} = I_n\mathbf{x} = \mathbf{x}.$$

Thus L acts as the inverse of T, so T is invertible.

証明終

Remark.

This theorem unifies the algebraic and geometric viewpoints. A linear transformation is an isomorphism (a bijective structure-preserving map) precisely when its determinant is non-zero.

Example 3.6. Inverting a Coordinate Map. Let $L : \mathbb{R}^3 \to \mathbb{R}^3$ be defined by $L(x, y, z) = (x, x + y, x + y + z)$. We wish to determine if L is invertible and find its inverse. The standard matrix is:

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

Since A is lower triangular with non-zero diagonal entries, $\det(A) = 1 \cdot 1 \cdot 1 = 1 \neq 0$. Thus L is invertible. We find A^{-1} using row reduction or inspection. Solving $A\mathbf{x} = \mathbf{y}$ for $\mathbf{x}$:

$$\begin{aligned} x &= y_1 \\ x + y &= y_2 \implies y = y_2 - x = y_2 - y_1 \\ x + y + z &= y_3 \implies z = y_3 - (x + y) = y_3 - y_2 \end{aligned}$$

Thus the inverse map is $L^{-1}(u, v, w) = (u, v - u, w - v)$. The matrix corresponds to:

$$[L^{-1}] = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix},$$

which is indeed A^{-1}.

範例

3.6 *The Algebra of Linear Mappings*

We have established that linear transformations are the structure-preserving maps between vector spaces. Just as we can add vectors and multiply them by scalars, we can perform algebraic operations on the transformations themselves. This algebraic structure on the set of mappings, denoted $\mathcal{L}(V, W)$, is what justifies the rules of matrix arithmetic derived in earlier chapters.

Composition and Matrix Multiplication

The most significant operation on functions is composition. When those functions are linear, their composition corresponds perfectly to the multiplication of their representative matrices. This result is not merely a happy coincidence; it is the *reason* matrix multiplication is defined the way it is.

Theorem 3.6. ***Composition corresponds to Matrix Multiplication.*** Let U, V, W be finite-dimensional vector spaces. Suppose $S : U \to V$ and $T : V \to W$ are linear transformations with standard matrices $[S]$ and $[T]$ respectively. Then the composite map $T \circ S : U \to W$ is linear, and its standard matrix is the product of the individual matrices:

$$[T \circ S] = [T][S].$$

定理

Note

This generalizes the earlier *Composition Property* to general vector spaces, confirming that the algebra of matrices mirrors the algebra of maps.

Proof

First, we verify linearity. Let $\mathbf{u}, \mathbf{v} \in U$ and $c \in \mathbb{R}$.

$$\begin{aligned}(T \circ S)(\mathbf{u} + c\mathbf{v}) &= T(S(\mathbf{u} + c\mathbf{v})) \\ &= T(S(\mathbf{u}) + cS(\mathbf{v})) && \text{(Linearity of } S\text{)} \\ &= T(S(\mathbf{u})) + cT(S(\mathbf{v})) && \text{(Linearity of } T\text{)} \\ &= (T \circ S)(\mathbf{u}) + c(T \circ S)(\mathbf{v}).\end{aligned}$$

Thus $T \circ S$ is linear. To determine its matrix $[T \circ S]$, we recall that the j-th column of a standard matrix is the image of the j-th basis vector $\mathbf{e}_j$.

$$\text{col}_j([T \circ S]) = (T \circ S)(\mathbf{e}_j) = T(S(\mathbf{e}_j)).$$

We know that $S(\mathbf{e}_j)$ is simply the j-th column of $[S]$, which we denote $\mathbf{s}_j$. Thus:

$$\text{col}_j([T \circ S]) = T(\mathbf{s}_j) = [T]\mathbf{s}_j.$$

By the definition of matrix-vector multiplication, $[T]\mathbf{s}_j$ is the j-th column of the matrix product $[T][S]$. Since the matrices agree on all columns, they are equal.

■

Remark.

This theorem allows us to visualize complex transformations as sequences of simpler ones. It also implies the associativity of matrix multiplication, inherited directly from the associativity of function composition: $(f \circ g) \circ h = f \circ (g \circ h)$.

Example 3.7. Stretching and Rotation. Consider the linear map S that stretches the plane by a factor of 2 in the x-direction, and the map R that rotates the plane by $\pi/4$ (45 degrees). The matrix for S is $\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$. The matrix for R is $\begin{bmatrix} \cos(\pi/4) & -\sin(\pi/4) \\ \sin(\pi/4) & \cos(\pi/4) \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$. The composite transformation $R \circ S$ (stretch then rotate) is given by:

$$[R \circ S] = [R][S] = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \sqrt{2} & -\frac{1}{\sqrt{2}} \\ \sqrt{2} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

Geometrically, this transforms the unit square into a tilted rectangle of length 2 and width 1.

範例

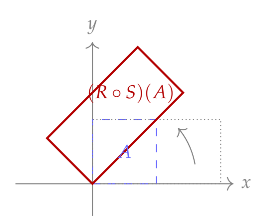

Figure 3.4: Stretching by 2 in the x-direction, then rotating by 45°. The dotted rectangle shows the intermediate step.

Addition and Scalar Multiplication

The set of linear transformations $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ itself forms a vector space.

Definition 3.3. ***Operations on Maps.***
Let $S, T : \mathbb{R}^n \to \mathbb{R}^m$ be linear transformations and $c \in \mathbb{R}$. We define:

1. **Sum:** $(S+T)(\mathbf{x}) = S(\mathbf{x}) + T(\mathbf{x})$.
2. **Scalar Multiple:** $(cS)(\mathbf{x}) = c \cdot S(\mathbf{x})$.

定義

Proposition 3.6. ***Linearity of Operations.***
If S and T are linear, then $S+T$ and cS are also linear. Furthermore, their matrices satisfy:

$$[S+T] = [S] + [T] \quad \text{and} \quad [cS] = c[S].$$

命題

Proof

For the sum, linearity follows from the commutativity of vector addition.

$$(S+T)(\mathbf{u}+\mathbf{v}) = S(\mathbf{u}+\mathbf{v}) + T(\mathbf{u}+\mathbf{v}) = (S(\mathbf{u}) + S(\mathbf{v})) + (T(\mathbf{u}) + T(\mathbf{v})).$$

Rearranging terms yields $(S+T)(\mathbf{u}) + (S+T)(\mathbf{v})$. Homogeneity is similar. For the matrices, consider the action on $\mathbf{e}_j$:

$$(S+T)(\mathbf{e}_j) = S(\mathbf{e}_j) + T(\mathbf{e}_j) = \mathrm{col}_j([S]) + \mathrm{col}_j([T]).$$

This is precisely the definition of matrix addition.

■

3.7 *Elementary Geometric Transformations*

Having established the algebraic machinery, we catalogue several fundamental linear transformations in $\mathbb{R}^2$ and $\mathbb{R}^3$. These serve as the building blocks for more complex operators.

Scaling and Reflection

The simplest transformations act diagonally on the standard basis.

Example 3.8. Scaling. A transformation T that scales the i-th coor-

Diagonal matrices are the easiest to analyse: each coordinate axis is an eigenvector, and the diagonal entries are eigenvalues.

dinate by a factor λ_i is represented by a diagonal matrix:

$$T(\mathbf{x}) = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)\mathbf{x} = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

- If $\lambda_i = c$ for all i, T is a **uniform scaling** (or homothety) by c. The matrix is cI_n.
- If $c > 1$, it is a dilation (expansion).
- If $0 < c < 1$, it is a contraction.

範例

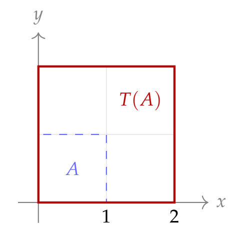

Figure 3.5: Uniform scaling by $c = 2$. The unit square maps to a square of side length 2 and area 4.

Example 3.9. Reflection. Reflections across coordinate axes (or hyperplanes) are achieved by setting specific $\lambda_i = -1$. In $\mathbb{R}^2$:

- Reflection across the y-axis ($x \to -x$): $\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$.
- Reflection across the origin ($\mathbf{x} \to -\mathbf{x}$): $\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$.

範例

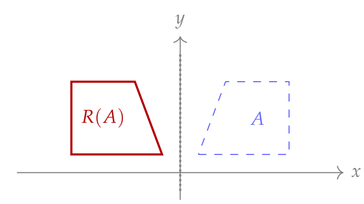

Figure 3.6: Reflection across the y-axis. The shape is mirrored; note the reversal of orientation.

Rotation

Rotations are fundamental orthogonal transformations that preserve lengths and the origin.

Theorem 3.7. ***Rotation Matrix in $\mathbb{R}^2$.***
Let $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ be the transformation that rotates the plane counter-clockwise by an angle θ. Its matrix is:

$$[R_\theta] = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

定理

Proof

We compute the image of the basis vectors $\mathbf{e}_1 = (1, 0)$ and $\mathbf{e}_2 = (0, 1)$. Using elementary trigonometry:

1. $\mathbf{e}_1$ lies on the positive x-axis. Rotating by θ moves it to $(\cos\theta, \sin\theta)$.
2. $\mathbf{e}_2$ lies on the positive y-axis ($\pi/2$ ahead of $\mathbf{e}_1$). Rotating by θ moves it to $(\cos(\theta + \pi/2), \sin(\theta + \pi/2)) = (-\sin\theta, \cos\theta)$.

Thus, $[R_\theta] = [\mathbf{e}'_1 \mid \mathbf{e}'_2] = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$.

■

Example 3.10. Trigonometric Addition Formulas. Since rotation satisfies $R_\alpha \circ R_\beta = R_{\alpha+\beta}$, we have $[R_\alpha][R_\beta] = [R_{\alpha+\beta}]$.

$$\begin{bmatrix} c_\alpha & -s_\alpha \\ s_\alpha & c_\alpha \end{bmatrix} \begin{bmatrix} c_\beta & -s_\beta \\ s_\beta & c_\beta \end{bmatrix} = \begin{bmatrix} c_{\alpha+\beta} & -s_{\alpha+\beta} \\ s_{\alpha+\beta} & c_{\alpha+\beta} \end{bmatrix}.$$

Computing the product yields the familiar identities, e.g., $\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta$.

範例

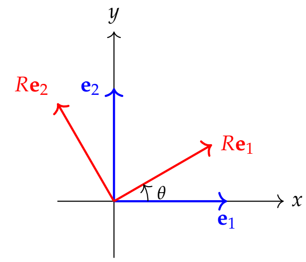

Figure 3.7: Rotation of the standard basis vectors.

3.8 *Coordinates and Change of Basis*

In *theorem 3.2*, we defined the "standard matrix" of a linear map. This assumed we were strictly using the standard basis $\mathcal{E} = \{\mathbf{e}_1, \dots, \mathbf{e}_n\}$. However, vector spaces often admit bases that are far more natural for a given problem (e.g., axes aligned with the symmetry of a crystal). We now generalise matrix representations to arbitrary bases.

The Coordinate Map

Let V be a vector space of dimension n, and let $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ be an ordered basis for V. By definition, any vector $\mathbf{v} \in V$ can be written uniquely as:

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

Definition 3.4. ***Coordinate Vector.***
The **coordinate vector** of $\mathbf{v}$ relative to $\mathcal{B}$, denoted $[\mathbf{v}]_\mathcal{B}$, is the column vector of coefficients in $\mathbb{R}^n$:

$$[\mathbf{v}]_\mathcal{B} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}.$$

The transformation $\Phi_\mathcal{B} : V \to \mathbb{R}^n$ defined by $\Phi_\mathcal{B}(\mathbf{v}) = [\mathbf{v}]_\mathcal{B}$ is a linear isomorphism called the **coordinate map**.

定義

Note

The order of basis vectors matters. Permuting $\mathcal{B}$ permutes the entries of $[\mathbf{v}]_\mathcal{B}$.

Example 3.11. Non-Standard Coordinates. Let $\mathcal{B} = \{(1,0),(1,1)\}$ in $\mathbb{R}^2$. Let $\mathbf{v} = (1,3)$. We seek c_1, c_2 such that:

$$\begin{bmatrix}1\\3\end{bmatrix} = c_1 \begin{bmatrix}1\\0\end{bmatrix} + c_2 \begin{bmatrix}1\\1\end{bmatrix}.$$

Solving this system (e.g., via row reduction), we find $c_2 = 3$ and $c_1 + c_2 = 1 \implies c_1 = -2$. Thus, $[\mathbf{v}]_\mathcal{B} = \begin{bmatrix}-2\\3\end{bmatrix}$. Note that $[\mathbf{v}]_\mathcal{E} = \begin{bmatrix}1\\3\end{bmatrix}$ relative to the standard basis.

範例

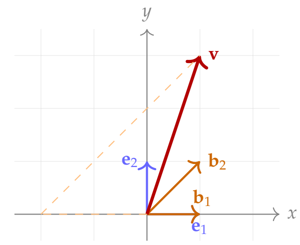

Figure 3.8: The vector $\mathbf{v} = (1,3)$ has coordinates $(1,3)$ in the standard basis but $(-2,3)$ in basis $\mathcal{B}$.

Change of Basis for Vectors

Given two bases $\mathcal{B}$ and $\mathcal{C}$ for $\mathbb{R}^n$, how do we translate coordinate vectors from one system to the other? Since coordinate maps are linear, the relationship must be a matrix multiplication.

Theorem 3.8. ***Change of Basis Matrix.***
Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ and $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$ be bases for $\mathbb{R}^n$. There exists a unique invertible matrix $P_{\mathcal{C}\leftarrow\mathcal{B}}$ such that for all $\mathbf{v} \in \mathbb{R}^n$:

$$[\mathbf{v}]_\mathcal{C} = P_{\mathcal{C}\leftarrow\mathcal{B}}[\mathbf{v}]_\mathcal{B}.$$

The columns of this matrix are the coordinate vectors of the $\mathcal{B}$-basis vectors relative to $\mathcal{C}$:

$$P_{\mathcal{C}\leftarrow\mathcal{B}} = [[\mathbf{b}_1]_\mathcal{C} \mid [\mathbf{b}_2]_\mathcal{C} \mid \cdots \mid [\mathbf{b}_n]_\mathcal{C}].$$

定理

Proof
Let $\mathbf{v} \in \mathbb{R}^n$. By definition, if $[\mathbf{v}]_\mathcal{B} = (x_1, \ldots, x_n)^T$, then $\mathbf{v} = \sum_{j=1}^n x_j \mathbf{b}_j$. Applying the coordinate map $\Phi_\mathcal{C}$ (which is linear) to both sides:

$$[\mathbf{v}]_\mathcal{C} = \left[\sum_{j=1}^n x_j \mathbf{b}_j\right]_\mathcal{C} = \sum_{j=1}^n x_j [\mathbf{b}_j]_\mathcal{C}.$$

This is exactly the matrix-vector product of the matrix with columns $[\mathbf{b}_j]_\mathcal{C}$ and the vector $\mathbf{x} = [\mathbf{v}]_\mathcal{B}$.

■

Remark.
When $\mathcal{C} = \mathcal{E}$ (the standard basis), the matrix $P_{\mathcal{E}\leftarrow\mathcal{B}}$ is simply the matrix formed by writing the vectors of $\mathcal{B}$ as columns. We often

denote this simply as $P_{\mathcal{B}}$. In this case:

$$\mathbf{v} = P_{\mathcal{B}}[\mathbf{v}]_{\mathcal{B}} \implies [\mathbf{v}]_{\mathcal{B}} = P_{\mathcal{B}}^{-1}\mathbf{v}.$$

Example 3.12. Calculating Coordinates via Matrices. Returning to example 3.11, with $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\} = \{(1,0),(1,1)\}$. The change of basis matrix to standard coordinates is $P_{\mathcal{B}} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. To find $[\mathbf{v}]_{\mathcal{B}}$ from $\mathbf{v} = (1,3)^T$, we compute $P_{\mathcal{B}}^{-1}$:

$$P_{\mathcal{B}}^{-1} = \frac{1}{1}\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}.$$

Then:

$$[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 1-3 \\ 3 \end{bmatrix} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}.$$

This matches our direct calculation.

範例

Matrix Representations of Linear Transformations

Just as vectors have coordinates relative to a basis, linear transformations have matrix representations relative to pairs of bases.

Definition 3.5. ***General Matrix Representation.***
Let $T : V \to W$ be a linear transformation. Let $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ be a basis for V and $\mathcal{C} = \{\mathbf{c}_1, \dots, \mathbf{c}_m\}$ be a basis for W. The **matrix of** T **relative to** $\mathcal{B}$ **and** $\mathcal{C}$, denoted $[T]_{\mathcal{C}\leftarrow\mathcal{B}}$ (or simply $[T]_{\mathcal{B},\mathcal{C}}$), is the $m \times n$ matrix defined by:

$$[T]_{\mathcal{C}\leftarrow\mathcal{B}} = [[T(\mathbf{b}_1)]_{\mathcal{C}} \mid [T(\mathbf{b}_2)]_{\mathcal{C}} \mid \cdots \mid [T(\mathbf{b}_n)]_{\mathcal{C}}].$$

This matrix satisfies the fundamental relation:

$$[T(\mathbf{v})]_{\mathcal{C}} = [T]_{\mathcal{C}\leftarrow\mathcal{B}}[\mathbf{v}]_{\mathcal{B}}.$$

定義

This definition is encapsulated by the following commutative diagram:

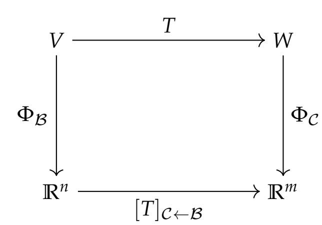

Figure 3.9: The matrix representation commutes with the coordinate maps.

Example 3.13. Derivative Operator. Let $V = P_2(\mathbb{R})$ (polynomials of degree ≤ 2) with basis $\mathcal{B} = \{1, x, x^2\}$. Let $D : V \to V$ be the derivative map $D(p) = p'$. To find $[D]_{\mathcal{B}}$, we differentiate the basis elements and find their coordinates in $\mathcal{B}$:

1. $D(1) = 0 \implies [0]_{\mathcal{B}} = (0,0,0)^T$.

2. $D(x) = 1 \implies [1]_{\mathcal{B}} = (1, 0, 0)^T$.
3. $D(x^2) = 2x \implies [2x]_{\mathcal{B}} = (0, 2, 0)^T$.

Thus:

$$[D]_{\mathcal{B}} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}.$$

Calculating the derivative of $p(x) = 3x^2 + 5x + 2$ via matrices:

$$[p]_{\mathcal{B}} = \begin{bmatrix} 2 \\ 5 \\ 3 \end{bmatrix} \implies [D]_{\mathcal{B}}[p]_{\mathcal{B}} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 5 \\ 3 \end{bmatrix} = \begin{bmatrix} 5 \\ 6 \\ 0 \end{bmatrix}.$$

This corresponds to $5(1) + 6(x) + 0(x^2) = 6x + 5$, which is indeed $\frac{d}{dx}(3x^2 + 5x + 2)$.

範例

Similarity

The most important case occurs when $T : V \to V$ is an operator on a single space, and we perform a change of basis from $\mathcal{B}$ to $\mathcal{C}$. How are the matrices $[T]_{\mathcal{B}}$ and $[T]_{\mathcal{C}}$ related?

Theorem 3.9. ***Similarity Transformation.***
Let $T : V \to V$ be a linear operator. Let $\mathcal{B}$ and $\mathcal{C}$ be bases for V, and let $P = P_{\mathcal{B} \leftarrow \mathcal{C}}$ be the change of basis matrix from $\mathcal{C}$ to $\mathcal{B}$. Then:

$$[T]_{\mathcal{C}} = P^{-1}[T]_{\mathcal{B}}P.$$

定理

Proof

Consider a vector $\mathbf{v}$. We can compute $[T(\mathbf{v})]_{\mathcal{C}}$ in two ways. 1. Directly: $[T(\mathbf{v})]_{\mathcal{C}} = [T]_{\mathcal{C}}[\mathbf{v}]_{\mathcal{C}}$. 2. Via $\mathcal{B}$: First convert $\mathbf{v}$ to $\mathcal{B}$-coordinates using P: $[\mathbf{v}]_{\mathcal{B}} = P[\mathbf{v}]_{\mathcal{C}}$. Apply $[T]_{\mathcal{B}}$ to get $[T(\mathbf{v})]_{\mathcal{B}}$. Then convert back to $\mathcal{C}$ using P^{-1}.

$$[T(\mathbf{v})]_{\mathcal{C}} = P^{-1}\left([T]_{\mathcal{B}}\left(P[\mathbf{v}]_{\mathcal{C}}\right)\right) = (P^{-1}[T]_{\mathcal{B}}P)[\mathbf{v}]_{\mathcal{C}}.$$

Since this holds for all $[\mathbf{v}]_{\mathcal{C}}$, the matrices are equal.

■

The characteristic polynomial is not merely an artifact of the matrix A but an intrinsic property of the linear transformation T itself.

Theorem 3.10. ***Similarity Invariance.***
Similar matrices have the same characteristic polynomial. Consequently,

they have the same eigenvalues (with the same algebraic multiplicities), the same determinant, and the same trace.

定理

Proof

Let $B = P^{-1}AP$. We compute the characteristic polynomial of B:

$$\begin{aligned}\det(B - \lambda I) &= \det(P^{-1}AP - \lambda P^{-1}IP) \\ &= \det(P^{-1}(A - \lambda I)P) \\ &= \det(P^{-1})\det(A - \lambda I)\det(P) \\ &= \frac{1}{\det(P)}\det(A - \lambda I)\det(P) \\ &= \det(A - \lambda I).\end{aligned}$$

Thus $p_B(\lambda) = p_A(\lambda)$.

■

Note

While similar matrices share the same characteristic polynomial, the converse is false. The matrices $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ both have characteristic polynomial $(\lambda - 1)^2$, but they are not similar.

Definition 3.6. ***Similarity.***
Two matrices $A, B \in M_n(\mathbb{R})$ are called **similar** if there exists an invertible matrix P such that $B = P^{-1}AP$.

定義

Similar matrices represent the *same* linear operator viewed from different coordinate systems. Consequently, they share coordinate-independent properties, such as the determinant, trace, and eigenvalues.

Similarity is an equivalence relation on matrices. The quest for canonical forms (e.g., Jordan normal form) is the search for the simplest representative in each equivalence class.

Example 3.14. Diagonalisation (Preview). Consider $T : \mathbb{R}^3 \to \mathbb{R}^3$ defined by $T(x, y, z) = (2x - 2y + 2z, x - z, 2x - 3y + 2z)$. In the standard basis, $[T]_{\mathcal{E}} = \begin{bmatrix} 2 & -2 & 2 \\ 1 & 0 & -1 \\ 2 & -3 & 2 \end{bmatrix}$. If we choose a specific basis $\mathcal{B} = \{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3\}$ (an eigenbasis), the matrix becomes:

$$[T]_{\mathcal{B}} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

This diagonal form reveals the geometric nature of T immediately: it stretches by 4 in the $\mathbf{f}_1$ direction, reflects/scales by -1 in $\mathbf{f}_2$, and leaves $\mathbf{f}_3$ invariant. Finding such a basis is the goal of spectral theory.

範例

3.9 Applications: Affine Geometry

The distinction between "linear" in the sense of calculus ($y = mx + c$) and "linear" in the sense of algebra ($T(c\mathbf{u} + \mathbf{v}) = cT(\mathbf{u}) + T(\mathbf{v})$) is a common source of confusion. A function of the form $f(x) = mx + c$ with $c \neq 0$ fails the linearity test because $f(0) = c \neq 0$. Such maps, which combine a linear transformation with a translation, are termed *affine*.

The equation $y = mx + c$ from introductory algebra describes a line with slope m and y-intercept c. This is the graph of an affine function, not a linear one (unless $c = 0$).

Affine Maps

Definition 3.7. ***Affine Map.***
A function $F : \mathbb{R}^n \to \mathbb{R}^m$ is called an *affine map* if it can be written in the form:

$$F(\mathbf{x}) = A\mathbf{x} + \mathbf{b},$$

where $A \in M_{m\times n}(\mathbb{R})$ is a matrix (representing a linear map) and $\mathbf{b} \in \mathbb{R}^m$ is a fixed translation vector. Equivalently, F is affine if the map $L(\mathbf{x}) = F(\mathbf{x}) - F(\mathbf{0})$ is linear.

定義

Geometrically, an affine transformation is a linear transformation followed by a shift. If F maps a subspace $U \subset \mathbb{R}^n$, the image $F(U)$ is not necessarily a subspace (as it may not contain the origin), but a "shifted" subspace.

Both affine and linear maps produce first-degree polynomial outputs. An affine map is linear precisely when $\mathbf{b} = \mathbf{0}$; conversely, every linear map is affine (with trivial translation).

Example 3.15. Affine vs Linear. The map $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x - y + 2 \\ 2x + y + 1 \end{bmatrix}$$

is affine, not linear. We can decompose it as:

$$T(\mathbf{x}) = \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

The linear part stretches and rotates the plane, while the vector $(2, 1)^T$ translates the result away from the origin.

範例

Parametrisation of Affine Subspaces

Linear algebra provides the natural language for describing flat geometric objects — lines, planes, and hyperplanes — in n-dimensional space. While a single equation (like $ax + by + cz = d$) describes a surface implicitly, a *parametrisation* describes it explicitly as the range of a function.

An affine subspace of dimension k is formed by translating a k-dimensional linear subspace.

Definition 3.8. ***Parametric Representation.***
An affine subspace $S \subset \mathbb{R}^n$ of dimension k can be parametrised by a function $\mathbf{r} : \mathbb{R}^k \to \mathbb{R}^n$:

$$\mathbf{r}(t_1, \ldots, t_k) = \mathbf{p}_0 + t_1\mathbf{v}_1 + t_2\mathbf{v}_2 + \cdots + t_k\mathbf{v}_k,$$

where $\mathbf{p}_0$ is a position vector (the "origin" of the subspace) and $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is a set of linearly independent direction vectors.

定義

Note

If the direction vectors are linearly dependent, the dimension of the image collapses to less than k, and the parametrisation is redundant.

Lines and Planes

1. **Lines ($k = 1$):** A line is determined by a point $\mathbf{p}_0$ and a direction vector $\mathbf{v}$.
$$\mathbf{r}(t) = \mathbf{p}_0 + t\mathbf{v}.$$

2. **Planes ($k = 2$):** A plane requires a point and two independent direction vectors.
$$\mathbf{r}(u, v) = \mathbf{p}_0 + u\mathbf{v}_1 + v\mathbf{v}_2.$$

Example 3.16. Degenerate Parametrisation. Consider the map $\mathbf{X} : \mathbb{R}^2 \to \mathbb{R}^3$ given by:

$$\mathbf{X}(u, v) = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + u \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + v \begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix}.$$

At first glance, this appears to describe a plane. However, the direction vectors are dependent: $\mathbf{v}_2 = 2\mathbf{v}_1$. We can rewrite the map as:

$$\mathbf{X}(u, v) = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + (u + 2v) \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

Letting $t = u + 2v$, we see this describes a line passing through $(1, 1, 1)$ in the direction $(1, 0, 1)$, not a plane. The rank of the matrix of direction vectors determines the true dimension of the object.

範例

This parametric perspective generalises immediately to higher dimensions. In relativity, for instance, the path of a particle is a "world-line" (dimension 1) in spacetime $\mathbb{R}^4$, while a "world-sheet" (dimension 2) might describe a string. The condition for these objects to be "flat" (affine) is precisely that their parametrisations are affine maps. Complex curves and surfaces arise when the maps become non-linear, requiring the tools of differential geometry (where the derivative $D\mathbf{X}$ provides a local linear approximation).

3.10 Exercises

1. **Linearity Check.** Determine whether the following maps are linear transformations. If linear, find the standard matrix. If not, provide a counter-example violating additivity or homogeneity.
   (a) $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T(x, y) = (2x + y, x - 3y)$.
   (b) $F : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $F(x, y) = (x + 1, y)$.
2. **Geometric Constructions.** Find the standard matrix $[T]$ for the following linear transformations on $\mathbb{R}^2$:
   (a) A clockwise rotation by $\pi/3$.
   (b) A reflection across the line $y = x$.
   (c) A projection onto the x-axis followed by a rotation by $\pi/2$ counter-clockwise.
   (d) A shear that maps $\mathbf{e}_1$ to $\mathbf{e}_1$ and $\mathbf{e}_2$ to $\mathbf{e}_2 + 3\mathbf{e}_1$.
3. **Kernel and Image.** Let $T : \mathbb{R}^4 \to \mathbb{R}^3$ be defined by the matrix:

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

   (a) Find a basis for the kernel of T. Is T injective?
   (b) Find a basis for the image of T. Is T surjective?
4. **Coordinates in Non-Standard Bases.** Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ be a basis for $\mathbb{R}^3$ given by:

$$\mathbf{b}_1 = \frac{1}{\sqrt{3}}(1, 1, 1), \quad \mathbf{b}_2 = \frac{1}{\sqrt{2}}(-1, 0, 1), \quad \mathbf{b}_3 = \frac{1}{\sqrt{6}}(-1, 2, -1).$$

   (a) Calculate the coordinate vector $[\mathbf{v}]_{\mathcal{B}}$ for $\mathbf{v} = (a, b, c)$.
5. **Matrix Representation.** Let $\mathcal{B}$ be the basis defined in Exercise 4. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear operator defined by:

$$T(x\mathbf{b}_1 + y\mathbf{b}_2 + z\mathbf{b}_3) = 6x\mathbf{b}_1 + 4y\mathbf{b}_2 + 12z\mathbf{b}_3.$$

(a) Write down the matrix $[T]_{\mathcal{B}}$ relative to the basis $\mathcal{B}$.

(b) Find the standard matrix $[T]_{\mathcal{E}}$ relative to the standard basis.

6. **Polynomial Operators.** Let $P_3(\mathbb{R})$ be the space of polynomials of degree at most 3. Consider the differentiation operator $D : P_3(\mathbb{R}) \to P_3(\mathbb{R})$ defined by $D(p) = p'$.

   (a) Find the matrix $[D]_{\mathcal{E}}$ relative to the standard basis $\{1, x, x^2, x^3\}$.

   (b) Show that D is nilpotent (i.e., $D^k = 0$ for some k). What is the smallest such k?

   (c) Determine the kernel and image of D^2.

7. **Change of Basis.** Let $\mathcal{B} = \{\mathbf{u}, \mathbf{v}\}$ and $\mathcal{C} = \{\mathbf{u} + \mathbf{v}, \mathbf{u} - \mathbf{v}\}$ be bases for a vector space V.

   (a) Find the change of basis matrix $P_{\mathcal{C} \leftarrow \mathcal{B}}$.

   (b) If a transformation T has matrix $[T]_{\mathcal{B}} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$, calculate $[T]_{\mathcal{C}}$ using the similarity formula.

8. **Affine Parametrisation.** Consider the plane W in $\mathbb{R}^3$ defined by the equation $x + 2y + 2z = 11$.

   (a) Express this plane as the image of an affine map $F : \mathbb{R}^2 \to \mathbb{R}^3$, explicitly identifying the translation vector and the linear part.

   (b) Find a pair of linearly independent tangent vectors $\mathbf{u}, \mathbf{v}$ (vectors parallel to the plane).

   (c) Construct a parametrisation $\gamma(t)$ for a circle of radius R lying in W centred at $(1, 2, 3)$.

9. **Rigid Motions.** A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is called *orthogonal* if it preserves lengths, i.e., $\|T(\mathbf{x})\| = \|\mathbf{x}\|$ for all $\mathbf{x}$.

   (a) Prove that an orthogonal transformation maps the unit sphere to the unit sphere.

   (b) Let $C_R(\mathbf{p})$ denote the circle in $\mathbb{R}^2$ of radius R centred at $\mathbf{p}$. Prove that if T is orthogonal, the image $T(C_R(\mathbf{p}))$ is a circle of radius R. Where is the new centre?

10. **Injectivity and Independence.** Let $T : V \to W$ be a linear transformation. Prove that T is injective if and only if it maps linearly independent sets to linearly independent sets. That is, for any linearly independent $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \subset V$, the set $\{T(\mathbf{v}_1), \ldots, T(\mathbf{v}_k)\}$ is linearly independent in W.

11. **Rank-Nullity and Composition.** Let $S : U \to V$ and $T : V \to W$ be linear transformations.

(a) Prove that $\ker(S) \subseteq \ker(T \circ S)$.

(b) Prove that $\mathrm{Im}(T \circ S) \subseteq \mathrm{Im}(T)$.

(c) If $U = V = W$ are finite-dimensional, prove that $\mathrm{rank}(T \circ S) \leq \min(\mathrm{rank}(T), \mathrm{rank}(S))$.

12. Projections and Idempotence. A linear operator $P : V \to V$ is called a *projection* (or idempotent) if $P^2 = P$ (i.e., $P(P(\mathbf{v})) = P(\mathbf{v})$).

(a) Prove that if P is a projection, then any vector $\mathbf{v} \in V$ can be uniquely written as $\mathbf{v} = \mathbf{k} + \mathbf{i}$, where $\mathbf{k} \in \ker(P)$ and $\mathbf{i} \in \mathrm{Im}(P)$.

Consider the vectors $P(\mathbf{v})$ and $\mathbf{v} - P(\mathbf{v})$.

(b) Deduce that $V = \ker(P) \oplus \mathrm{Im}(P)$.

13. Reflection across Arbitrary Lines. Let L be the line in $\mathbb{R}^2$ passing through the origin with angle ϕ to the x-axis.

(a) Write the matrix for the reflection R_L by viewing it as a composition: rotate the line to the x-axis, reflect across the x-axis, then rotate back.

(b) Show that the resulting matrix is $\begin{bmatrix} \cos(2\phi) & \sin(2\phi) \\ \sin(2\phi) & -\cos(2\phi) \end{bmatrix}$.

(c) Verify that $\det(R_L) = -1$ and $R_L^2 = I$.

4
Eigenvalues and Eigenvectors

When Werner Heisenberg formulated matrix mechanics in 1925, he was initially unaware of the algebraic structure he was employing; it was Max Born who identified the non-commutative multiplication tables as matrix algebra. The subsequent development of quantum mechanics relied heavily on the spectral theory of these operators. In this chapter, we explore the decomposition of linear operators into their fundamental components: eigenvalues and eigenvectors. These concepts allow us to decouple complex coupled systems into independent, manageable parts, providing deep insight into the long-term behaviour of dynamical systems, from population models to the geometry of linear maps.

4.1 Motivation: Matrix Powers and Fibonacci

In *theorem* 2.9, we established that matrix multiplication corresponds to the composition of linear maps. A frequent objective in applied mathematics is to evaluate the long-term behaviour of a system evolving under a constant linear rule, represented by the repeated application of a matrix A. That is, we wish to compute A^n for large n. Direct multiplication is computationally expensive and offers little geometric insight. However, if the matrix is *diagonal*, the calculation is trivial.

$$\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}^n = \begin{bmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{bmatrix}.$$

We seek a method to transform a general matrix into a diagonal one via a change of basis.

> **Example 4.1.** The Fibonacci Sequence. The Fibonacci numbers are defined recursively by $a_0 = 0, a_1 = 1$, and $a_{n+1} = a_n + a_{n-1}$. We propose to prove the closed-form expression (Binet's formula):
>
> $$a_n = \frac{1}{\sqrt{5}} \left[\left(\frac{1+\sqrt{5}}{2} \right)^n - \left(\frac{1-\sqrt{5}}{2} \right)^n \right].$$

This formula involves irrational numbers despite a_n always being an integer. We begin by casting the recursion as a matrix system. Let $\mathbf{x}_n = [a_n, a_{n+1}]^T$. Then:

$$\begin{bmatrix} a_n \\ a_{n+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} a_{n-1} \\ a_n \end{bmatrix}.$$

Let $A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$. By induction, $\mathbf{x}_n = A^n \mathbf{x}_0$, where $\mathbf{x}_0 = [0, 1]^T$. Consider the matrix

$$P = \begin{bmatrix} 1 & 1 \\ \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} \end{bmatrix}.$$

A direct calculation reveals that P diagonalises A. Specifically, if we let $\phi = \frac{1+\sqrt{5}}{2}$ and $\psi = \frac{1-\sqrt{5}}{2}$:

$$P^{-1}AP = \begin{bmatrix} \phi & 0 \\ 0 & \psi \end{bmatrix} = D.$$

From the similarity relation $A = PDP^{-1}$, we find $A^n = (PDP^{-1})^n = PD^nP^{-1}$.

$$\begin{aligned} \begin{bmatrix} a_n \\ a_{n+1} \end{bmatrix} &= \begin{bmatrix} 1 & 1 \\ \phi & \psi \end{bmatrix} \begin{bmatrix} \phi^n & 0 \\ 0 & \psi^n \end{bmatrix} \frac{1}{\psi - \phi} \begin{bmatrix} \psi & -1 \\ -\phi & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= \frac{1}{-\sqrt{5}} \begin{bmatrix} \phi^n & \psi^n \\ \phi^{n+1} & \psi^{n+1} \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} \\ &= \frac{1}{\sqrt{5}} \begin{bmatrix} \phi^n - \psi^n \\ \phi^{n+1} - \psi^{n+1} \end{bmatrix}. \end{aligned}$$

Reading off the first component yields the desired formula.

範例

Note

The scalar $\phi \approx 1.618$ is the Golden Ratio. Since $|\psi| \approx 0.618 < 1$, the term ψ^n vanishes for large n, implying that a_n grows exponentially at a rate determined solely by ϕ.

4.2 *Definitions and Existence*

The Fibonacci example works because there exist special vectors $\mathbf{v}$ such that the action of A is simply scalar multiplication: $A\mathbf{v} = \lambda \mathbf{v}$. In the basis of these vectors, the matrix becomes diagonal.

Definition 4.1. ***Eigenvalues and Eigenvectors.***
Let $A \in M_n(\mathbb{R})$. A scalar $\lambda \in \mathbb{C}$ is called an **eigenvalue** of A if there

exists a non-zero vector $\mathbf{v} \in \mathbb{C}^n$ such that:

$$A\mathbf{v} = \lambda\mathbf{v}.$$

The vector $\mathbf{v}$ is called an **eigenvector** corresponding to λ.

定義

Note

We exclude the zero vector from being an eigenvector, as $A\mathbf{0} = \lambda\mathbf{0}$ holds for any λ. However, the eigenvalue λ itself may be zero.

To find these scalars, we rewrite the defining equation as a homogeneous linear system:

$$(A - \lambda I)\mathbf{v} = \mathbf{0}.$$

For a non-zero solution $\mathbf{v}$ to exist, the matrix $A - \lambda I$ must be singular. This connects spectral theory immediately to the theory of determinants established in the previous chapter.

Theorem 4.1. ***The Characteristic Equation.***
A scalar λ is an eigenvalue of A if and only if λ satisfies the characteristic equation:

$$\det(A - \lambda I) = 0.$$

The polynomial $p_A(\lambda) = \det(A - \lambda I)$ is a polynomial of degree n, called the **characteristic polynomial** of A.

定理

Proof

This follows directly from the *Determinant Criterion for Independence* and *Matrix Rank and Transformation Properties*. The equation $(A - \lambda I)\mathbf{v} = \mathbf{0}$ has a non-trivial solution $\mathbf{v} \neq \mathbf{0}$ if and only if the columns of $A - \lambda I$ are linearly dependent, which is equivalent to the determinant vanishing.

■

Example 4.2. Calculating Eigensystems. Let $A = \begin{bmatrix} 3 & 0 \\ 8 & -1 \end{bmatrix}$.

1. **Find Eigenvalues:** Compute $\det(A - \lambda I)$.

$$\det \begin{bmatrix} 3-\lambda & 0 \\ 8 & -1-\lambda \end{bmatrix} = (3-\lambda)(-1-\lambda) = (\lambda - 3)(\lambda + 1).$$

The roots are $\lambda_1 = 3$ and $\lambda_2 = -1$.

2. **Find Eigenvectors for $\lambda_1 = 3$:** Solve $(A - 3I)\mathbf{v} = \mathbf{0}$.

$$\begin{bmatrix} 0 & 0 \\ 8 & -4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

This implies $8x - 4y = 0$ or $y = 2x$. The eigenspace is spanned by $\mathbf{v}_1 = [1, 2]^T$.

3. **Find Eigenvectors for $\lambda_2 = -1$:** Solve $(A + I)\mathbf{v} = \mathbf{0}$.

$$\begin{bmatrix} 4 & 0 \\ 8 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

This implies $4x = 0 \implies x = 0$. The variable y is free. The eigenspace is spanned by $\mathbf{v}_2 = [0, 1]^T$.

Geometrically, A stretches vectors along the line $y = 2x$ by a factor of 3 and reflects vectors along the y-axis ($x = 0$) while preserving their length.

範例

Example 4.3. Distinct Eigenvalues in $\mathbb{R}^3$. Consider the matrix

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

We compute the characteristic polynomial $p_A(\lambda) = \det(A - \lambda I)$:

$$\begin{aligned} \det \begin{bmatrix} 2-\lambda & 1 & 1 \\ 2 & 3-\lambda & 4 \\ -1 & -1 & -2-\lambda \end{bmatrix} &= (2-\lambda) \det \begin{bmatrix} 3-\lambda & 4 \\ -1 & -2-\lambda \end{bmatrix} - 1 \det \begin{bmatrix} 2 & 4 \\ -1 & -2-\lambda \end{bmatrix} + 1 \det \begin{bmatrix} 2 & 3-\lambda \\ -1 & -1 \end{bmatrix} \\ &= (2-\lambda)[(3-\lambda)(-2-\lambda) + 4] - [2(-2-\lambda) + 4] + [-2 + (3-\lambda)] \\ &= (2-\lambda)[\lambda^2 - \lambda - 2] - [-2\lambda] + [1-\lambda] \\ &= -\lambda^3 + 3\lambda^2 + \lambda - 3. \end{aligned}$$

By inspection, $\lambda = 1$ is a root. Factoring out $(\lambda - 1)$, we find:

$$p_A(\lambda) = -(\lambda - 1)(\lambda + 1)(\lambda - 3).$$

The eigenvalues are $1, -1, 3$. Since these are distinct, A is diagonalisable. We find the eigenvectors by solving $(A - \lambda I)\mathbf{x} = \mathbf{0}$ for each λ:

1. **For $\lambda = 1$:**

$$\begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 4 \\ -1 & -1 & -3 \end{bmatrix} \xrightarrow{\text{RREF}} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

This yields $x_3 = 0$ and $x_1 + x_2 = 0$. Eigenvector: $\mathbf{v}_1 = (1, -1, 0)^T$.

2. **For $\lambda = -1$:** The system yields $\mathbf{v}_2 = (0, 1, -1)^T$.

3. **For $\lambda = 3$:** The system yields $\mathbf{v}_3 = (2, 3, -1)^T$.

The eigenspaces E_1, E_{-1}, E_3 are all 1-dimensional.

範例

Discrete Dynamical Systems

Eigenvalues provide a natural language for analysing discrete dynamical systems of the form $\mathbf{x}_{k+1} = A\mathbf{x}_k$. The long-term behaviour of the state vector $\mathbf{x}_k$ is dominated by the eigenvalue with the largest magnitude (the dominant eigenvalue).

Attractors and Repellers

Revisiting the matrix $A = \begin{bmatrix} 3 & 0 \\ 8 & -1 \end{bmatrix}$ from the previous example, we can trace the trajectory of an arbitrary point $\mathbf{x}_0$. Since $\{\mathbf{v}_1, \mathbf{v}_2\}$ forms a basis for $\mathbb{R}^2$, we can write $\mathbf{x}_0 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$. Then:

$$\mathbf{x}_k = A^k\mathbf{x}_0 = c_1(3)^k\mathbf{v}_1 + c_2(-1)^k\mathbf{v}_2.$$

For large k, the term $3^k\mathbf{v}_1$ overwhelms $(-1)^k\mathbf{v}_2$ (provided $c_1 \neq 0$). Thus, almost all trajectories tend towards the direction of the eigenspace corresponding to $\lambda = 3$.

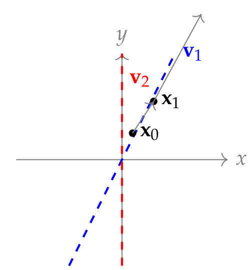

Figure 4.1: Trajectories are pulled towards the eigenspace of the dominant eigenvalue $\lambda = 3$.

Complex Eigenvalues and Rotation

If the characteristic polynomial has no real roots, the matrix does not scale any line in $\mathbb{R}^n$ simply by a real factor. Instead, the action typically involves rotation.

Example 4.4. Rotation and Cycles. Consider $A = \begin{bmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{bmatrix}$. The characteristic equation is:

$$(\lambda - 1/2)^2 + 3/4 = 0 \implies \lambda^2 - \lambda + 1 = 0.$$

The roots are $\lambda = \frac{1 \pm i\sqrt{3}}{2} = e^{\pm i\pi/3}$. Since the eigenvalues are complex numbers with modulus 1, the transformation preserves distance but rotates vectors by $60°$ ($\pi/3$ radians). A trajectory starting at $\mathbf{x}_0 = [1, 0]^T$ cycles through 6 distinct points before returning to $\mathbf{x}_0$:

$$\mathbf{x}_0 \to \mathbf{x}_1 \text{ (rotated } 60°) \to \cdots \to \mathbf{x}_6 = \mathbf{x}_0.$$

Complex eigenvalues in real matrices always occur in conjugate pairs and signify rotational components in the linear map.

範例

Stochastic Matrices and Steady States

An important class of dynamical systems arises in probability and statistical modelling, specifically Markov chains.

Definition 4.2. ***Stochastic Matrix.***
A square matrix P with non-negative entries is called a **stochastic matrix** if the sum of the entries in each column is 1. (We use this column-stochastic convention throughout.)

定義

This condition ensures that P maps probability vectors (vectors with non-negative entries summing to 1) to probability vectors. A fundamental property of these matrices is the guaranteed existence of a steady state.

Theorem 4.2. ***Steady State Existence.***
If P is a stochastic matrix, then $\lambda = 1$ is an eigenvalue of P. Consequently, there exists a non-zero vector $\mathbf{q}$ such that $P\mathbf{q} = \mathbf{q}$.

定理

Remark.

In Markov chain applications, under additional assumptions such as irreducibility and aperiodicity, this eigenvector can be chosen as a unique probability vector, called the **stationary distribution**.

Proof

We consider the transpose P^T. Since the columns of P sum to 1, the rows of P^T sum to 1. Let $\mathbf{u} = [1, 1, \dots, 1]^T$. Then:

$$P^T\mathbf{u} = \begin{bmatrix} \sum p_{1j} \\ \vdots \\ \sum p_{nj} \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = 1\mathbf{u}.$$

Thus, 1 is an eigenvalue of P^T. By *theorem* 2.4, $\det(P - 1I) = \det(P^T - 1I) = 0$, so 1 is also an eigenvalue of P.

■

Example 4.5. Modelling Viral Spread. Consider a population of laboratory mice partitioned into two states: Infected (I) and Non-infected (N). Suppose that each week:

- An infected mouse has an 80% chance of recovering (becoming N) and 20% chance of remaining infected.
- A non-infected mouse has a 10% chance of becoming infected and 90% chance of remaining non-infected.

The transition matrix P acts on the state vector $\mathbf{x}_k = [I_k, N_k]^T$:

$$P = \begin{bmatrix} 0.2 & 0.1 \\ 0.8 & 0.9 \end{bmatrix}.$$

(Note: Col 1 is transition from I, Col 2 from N). To find the long-term steady state $\mathbf{x}^*$, we solve $(P - I)\mathbf{x}^* = \mathbf{0}$:

$$\begin{bmatrix} -0.8 & 0.1 \\ 0.8 & -0.1 \end{bmatrix} \begin{bmatrix} I^* \\ N^* \end{bmatrix} = \mathbf{0}.$$

This yields $-0.8I^* + 0.1N^* = 0 \implies N^* = 8I^*$. The steady state is a scalar multiple of $[1, 8]^T$. Normalising for a population of 1000 mice:

$$I^* + 8I^* = 1000 \implies 9I^* = 1000 \implies I^* \approx 111, \quad N^* \approx 889.$$

Regardless of the initial infection rate, the system converges to this distribution.

範例

Example 4.6. Market Share Dynamics. Consider three competing tech companies A, B, C with a total market share. Customers switch annually according to the stochastic matrix P:

$$P = \begin{bmatrix} 0.7 & 0.1 & 0.1 \\ 0.2 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}.$$

We verify that columns sum to 1. To find the steady state, we solve $(P - I)\mathbf{q} = \mathbf{0}$.

$$P - I = \begin{bmatrix} -0.3 & 0.1 & 0.1 \\ 0.2 & -0.2 & 0.1 \\ 0.1 & 0.1 & -0.2 \end{bmatrix}.$$

Row reducing (scaling by 10 for convenience):

$$\begin{bmatrix} -3 & 1 & 1 \\ 2 & -2 & 1 \\ 1 & 1 & -2 \end{bmatrix} \xrightarrow{R_1 \leftrightarrow R_3} \begin{bmatrix} 1 & 1 & -2 \\ 2 & -2 & 1 \\ -3 & 1 & 1 \end{bmatrix} \xrightarrow[R_3+3R_1]{R_2-2R_1} \begin{bmatrix} 1 & 1 & -2 \\ 0 & -4 & 5 \\ 0 & 4 & -5 \end{bmatrix}.$$

This implies $-4q_2 + 5q_3 = 0 \implies q_2 = \frac{5}{4}q_3$. From R_1: $q_1 + \frac{5}{4}q_3 - 2q_3 = 0 \implies q_1 = \frac{3}{4}q_3$. Let $q_3 = 4$. Then $q_2 = 5$ and $q_1 = 3$. The steady state vector is $[3, 5, 4]^T$. Normalising: $3 + 5 + 4 = 12$, so the market shares stabilize at $A = 25\%$, $B \approx 41.7\%$, $C \approx 33.3\%$.

範例

4.3 *Diagonalisation*

We return to the algebraic structure suggested by the Fibonacci example. We say a matrix A is **diagonalisable** if it is similar to a diagonal matrix D. That is, there exists an invertible matrix P such that

$P^{-1}AP = D$.

Theorem 4.3. ***Diagonalisation Criterion.***
An $n \times n$ matrix A is diagonalisable if and only if A has n linearly independent eigenvectors. In this case, the columns of P are the eigenvectors, and the diagonal entries of D are the corresponding eigenvalues.

定理

Proof

Let $P = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ and $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Note that $AP = A[\mathbf{v}_1, \ldots, \mathbf{v}_n] = [A\mathbf{v}_1, \ldots, A\mathbf{v}_n]$. Also, $PD = [\mathbf{v}_1, \ldots, \mathbf{v}_n]D = [\lambda_1\mathbf{v}_1, \ldots, \lambda_n\mathbf{v}_n]$. Thus, $AP = PD$ if and only if $A\mathbf{v}_i = \lambda_i\mathbf{v}_i$ for all i. For P to be invertible, its columns (the eigenvectors) must be linearly independent. If these conditions hold, multiplication by P^{-1} yields $P^{-1}AP = D$.

■

Corollary 4.1. Distinct Eigenvalues

推論

If an $n \times n$ matrix has n distinct eigenvalues, it is diagonalisable. Let $A \in M_n(\mathbb{R})$ (or $M_n(\mathbb{C})$) have n distinct eigenvalues $\lambda_1, \ldots, \lambda_n$. For each λ_i, choose a corresponding eigenvector $\mathbf{v}_i \neq \mathbf{0}$ such that

$$A\mathbf{v}_i = \lambda_i\mathbf{v}_i, \qquad i = 1, \ldots, n.$$

By the *Diagonalisation Criterion*, it is enough to show that the eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent. We prove this by induction on the number of eigenvectors.

Base case ($m = 1$).

Any single non-zero vector is linearly independent, so the statement holds for one eigenvector.

証明終

Inductive step.

Assume that any collection of $m - 1$ eigenvectors corresponding to $m - 1$ *distinct* eigenvalues is linearly independent. We will show that any collection of m eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$ corresponding to distinct eigenvalues $\lambda_1, \ldots, \lambda_m$ is also linearly independent.
Suppose, for the sake of contradiction, that $\mathbf{v}_1, \ldots, \mathbf{v}_m$ are linearly *dependent*. Then there exist scalars $c_1, \ldots, c_m$, not all zero, such that

$$c_1\mathbf{v}_1 + \cdots + c_m\mathbf{v}_m = \mathbf{0}. \tag{4.1}$$

We may assume at least one of the c_i is non-zero; without loss of generality, we will not reorder the vectors. Apply the linear map A

to both sides of (4.1):

$$A(c_1\mathbf{v}_1 + \cdots + c_m\mathbf{v}_m) = c_1A\mathbf{v}_1 + \cdots + c_mA\mathbf{v}_m = c_1\lambda_1\mathbf{v}_1 + \cdots + c_m\lambda_m\mathbf{v}_m = \mathbf{0}.$$

Now consider the combination

$$A(c_1\mathbf{v}_1 + \cdots + c_m\mathbf{v}_m) - \lambda_m(c_1\mathbf{v}_1 + \cdots + c_m\mathbf{v}_m) = \mathbf{0}.$$

Substituting the two expressions above, we obtain

$$\begin{aligned}\mathbf{0} &= \big(c_1\lambda_1\mathbf{v}_1 + \cdots + c_m\lambda_m\mathbf{v}_m\big) - \lambda_m\big(c_1\mathbf{v}_1 + \cdots + c_m\mathbf{v}_m\big)\\ &= \sum_{i=1}^{m} c_i(\lambda_i - \lambda_m)\mathbf{v}_i.\end{aligned}$$

Observe that the coefficient of $\mathbf{v}_m$ is $c_m(\lambda_m - \lambda_m) = 0$, so the term involving $\mathbf{v}_m$ disappears. Thus we are left with

$$\sum_{i=1}^{m-1} c_i(\lambda_i - \lambda_m)\mathbf{v}_i = \mathbf{0}.$$

Since the eigenvalues are distinct, we have $\lambda_i - \lambda_m \neq 0$ for $i = 1, \ldots, m-1$. Hence each scalar $c_i(\lambda_i - \lambda_m)$ is zero if and only if $c_i = 0$. Therefore the above relation shows that

$$c_1(\lambda_1 - \lambda_m) = \cdots = c_{m-1}(\lambda_{m-1} - \lambda_m) = 0 \quad \Longrightarrow \quad c_1 = \cdots = c_{m-1} = 0.$$

Returning to (4.1), we now have

$$c_m\mathbf{v}_m = \mathbf{0}.$$

But $\mathbf{v}_m \neq \mathbf{0}$ by definition of an eigenvector, so this forces $c_m = 0$. Thus all coefficients $c_1, \ldots, c_m$ must be zero, which contradicts our assumption that they were not all zero. Therefore $\mathbf{v}_1, \ldots, \mathbf{v}_m$ are linearly independent.
By induction, any collection of eigenvectors corresponding to distinct eigenvalues is linearly independent, in particular the n eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of A. Hence A has n linearly independent eigenvectors. The Diagonalisation Criterion now implies that A is diagonalisable.

証明終

Note

Not all matrices are diagonalisable. For example, $J = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ has characteristic polynomial λ^2. The only eigenvalue is 0. If J were diagonalisable, it would be similar to the zero matrix, which implies $J = P0P^{-1} = 0$, a contradiction. Such matrices are called *defective* and require the Jordan Canonical Form for analysis.

4.4 Abstract Spectral Theory

While matrices provide a concrete computational framework, the concept of an eigenvalue is intrinsic to the linear operator itself, independent of any coordinate system. We therefore broaden our definition to general vector spaces.

Definition 4.3. ***Eigenvalues of a Linear Operator.***
Let V be a vector space and $T : V \to V$ be a linear transformation. A scalar λ is an **eigenvalue** of T if there exists a non-zero vector $\mathbf{v} \in V$ such that:

$$T(\mathbf{v}) = \lambda \mathbf{v}.$$

The vector $\mathbf{v}$ is an **eigenvector** (or eigenfunction, if V is a function space) corresponding to λ.

定義

This definition allows us to analyse operators on infinite-dimensional spaces, where the matrix determinant is not immediately available.

Example 4.7. The Derivative Operator. Let $V = C^\infty(\mathbb{R})$ be the space of smooth functions. Consider the derivative operator $D : V \to V$ defined by $D(f) = f'$. The eigenvalue equation is the differential equation:

$$\frac{df}{dt} = \lambda f.$$

Separation of variables yields the solution $f(t) = Ce^{\lambda t}$. Thus, for every $\lambda \in \mathbb{R}$, the function $f(t) = e^{\lambda t}$ is an eigenfunction of D. The spectrum of this operator is the entire real line.

範例

Example 4.8. The Transpose Operator. Let $V = M_n(\mathbb{R})$ be the vector space of $n \times n$ matrices. Define $T : V \to V$ by $T(A) = A^T$. We seek scalars λ and non-zero matrices A such that $A^T = \lambda A$. Applying the transpose again:

$$A = (A^T)^T = (\lambda A)^T = \lambda A^T = \lambda^2 A.$$

Since $A \neq 0$, we must have $\lambda^2 = 1$, so $\lambda = 1$ or $\lambda = -1$.

1. For $\lambda = 1$, $A^T = A$. The eigenvectors are the *symmetric* matrices.
2. For $\lambda = -1$, $A^T = -A$. The eigenvectors are the *skew-symmetric* matrices.

This spectral decomposition implies that any square matrix can be uniquely written as the sum of a symmetric and a skew-symmetric matrix: $A = \frac{1}{2}(A + A^T) + \frac{1}{2}(A - A^T)$.

範例

4.5 Eigenspaces and Multiplicity

Returning to finite-dimensional spaces, we refine our understanding of the solution set to $(A - \lambda I)\mathbf{v} = \mathbf{0}$.

Definition 4.4. ***Eigenspace.***
The **eigenspace** of A corresponding to an eigenvalue λ, denoted E_λ, is the kernel of the matrix $A - \lambda I$:

$$E_\lambda = \ker(A - \lambda I) = \{\mathbf{v} \in \mathbb{R}^n \mid A\mathbf{v} = \lambda\mathbf{v}\}.$$

定義

The eigenspace consists of the zero vector and all eigenvectors for λ. It is a subspace of $\mathbb{R}^n$.

Proposition 4.1. ***Equivalences for Eigenvalues.***
For $A \in M_n(\mathbb{R})$ and $\lambda \in \mathbb{C}$, the following are equivalent:

1. λ is an eigenvalue of A.
2. The kernel of $A - \lambda I$ is non-trivial.
3. $\det(A - \lambda I) = 0$.
4. The system $(A - \lambda I)\mathbf{x} = \mathbf{0}$ has infinitely many solutions.

命題

Algebraic vs. Geometric Multiplicity

A subtle but critical distinction arises when the characteristic polynomial has repeated roots.

Definition 4.5. ***Multiplicities.***
Let λ_0 be an eigenvalue of A.

1. The **algebraic multiplicity**, denoted $m_a(\lambda_0)$, is the multiplicity of λ_0 as a root of the characteristic polynomial $p_A(\lambda)$.
2. The **geometric multiplicity**, denoted $m_g(\lambda_0)$, is the dimension of the eigenspace E_{λ_0}.

定義

Note

It is a fundamental result that $1 \leq m_g(\lambda) \leq m_a(\lambda)$. If $m_g(\lambda) < m_a(\lambda)$ for any eigenvalue, the matrix is *defective* and not diagonalisable.

Example 4.9. Repeated Eigenvalues (Diagonalisable Case). Consider $A = \begin{bmatrix} 0 & 0 & -4 \\ 2 & 4 & 2 \\ 2 & 0 & 6 \end{bmatrix}$. The characteristic polynomial is:

$$\begin{aligned} \det(A - \lambda I) &= \det \begin{bmatrix} -\lambda & 0 & -4 \\ 2 & 4-\lambda & 2 \\ 2 & 0 & 6-\lambda \end{bmatrix} \\ &= (4-\lambda) \det \begin{bmatrix} -\lambda & -4 \\ 2 & 6-\lambda \end{bmatrix} \quad \text{(Expansion along Col 2)} \\ &= (4-\lambda)(-\lambda(6-\lambda)+8) = -(4-\lambda)(\lambda^2 - 6\lambda + 8) \\ &= -(\lambda-4)(\lambda-4)(\lambda-2). \end{aligned}$$

The eigenvalues are $\lambda = 4$ (algebraic multiplicity 2) and $\lambda = 2$ (algebraic multiplicity 1).

- For $\lambda = 2$: Solving $(A - 2I)\mathbf{v} = \mathbf{0}$ yields a 1-dimensional eigenspace spanned by $\mathbf{v}_1 = [-2, 1, 1]^T$.
- For $\lambda = 4$: We compute $\ker(A - 4I)$:

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

  The equation is $x + z = 0$. We have two free variables, y and z. The general solution is:

$$\mathbf{v} = \begin{bmatrix} -s \\ t \\ s \end{bmatrix} = s \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} + t \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

  The geometric multiplicity is 2, matching the algebraic multiplicity. Thus, A is diagonalisable.

範例

Example 4.10. Repeated Eigenvalues (Defective Case). Consider the shear matrix $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. The characteristic polynomial is $(1-\lambda)^2$, so $\lambda = 1$ has algebraic multiplicity 2. The eigenspace is the kernel of $A - I = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. The equation implies $y = 0$, so eigenvectors are of the form $[x, 0]^T$. The geometric multiplicity is 1. Since $1 < 2$, the matrix is defective and cannot be diagonalised.

範例

4.6 Invariant Subspaces

The geometric essence of the eigenvalue problem lies in the identification of lines passing through the origin that are mapped back onto themselves. We now formalise this concept and extend it to subspaces of higher dimension.

Definition 4.6. ***Invariant Subspace.***
Let $T : V \to V$ be a linear operator on a vector space V. A subspace $W \subseteq V$ is called invariant under T (or T-invariant) if T maps W into itself:

$$T(W) \subseteq W.$$

That is, for every $\mathbf{w} \in W$, the image $T(\mathbf{w})$ is also in W.

定義

The trivial subspaces $\{\mathbf{0}\}$ and V are always invariant under any operator. The study of eigenvalues is precisely the search for one-dimensional invariant subspaces.

Proposition 4.2. ***Eigenvectors and Invariant Lines.***
Let V be a vector space over $\mathbb{F}$. A one-dimensional subspace $W = \text{span}\{\mathbf{v}\}$ (where $\mathbf{v} \neq \mathbf{0}$) is invariant under T if and only if $\mathbf{v}$ is an eigenvector of T.

命題

($\Rightarrow$)

If W is invariant, then $T(\mathbf{v}) \in W$. Since W is spanned by $\mathbf{v}$, we must have $T(\mathbf{v}) = \lambda\mathbf{v}$ for some scalar $\lambda \in \mathbb{F}$. Thus $\mathbf{v}$ is an eigenvector.

証明終

($\Leftarrow$)

If $T(\mathbf{v}) = \lambda\mathbf{v}$, then for any $\mathbf{w} = c\mathbf{v} \in W$, $T(\mathbf{w}) = T(c\mathbf{v}) = c\lambda\mathbf{v} \in W$. Thus W is invariant.

証明終

Furthermore, the eigenspaces E_λ defined in *definition* 4.4 are themselves invariant subspaces. If $\mathbf{x} \in E_\lambda$, then $T(\mathbf{x}) = \lambda\mathbf{x}$. Since E_λ is a subspace, it is closed under scalar multiplication, so $\lambda\mathbf{x} \in E_\lambda$.

Example 4.11. The Integration Operator. We previously examined the differentiation operator on the space of smooth functions. Consider now the integration operator $T : C[0,1] \to C[0,1]$ defined by:

$$T(f)(x) = \int_0^x f(t)\,dt.$$

We investigate whether T admits any eigenvalues, or equivalently, any one-dimensional invariant subspaces. Suppose f is a non-zero eigenfunction with eigenvalue λ:

$$\int_0^x f(t)\,dt = \lambda f(x).$$

If $\lambda = 0$, differentiating both sides yields $f(x) = 0$, which contradicts $f \neq 0$. If $\lambda \neq 0$, the right-hand side is differentiable, so f must be differentiable. Differentiating with respect to x:

$$f(x) = \lambda f'(x) \implies f'(x) = \frac{1}{\lambda} f(x).$$

The general solution to this differential equation is $f(x) = ce^{x/\lambda}$. However, we must satisfy the boundary condition implied by the definition of T. Evaluating at $x = 0$:

$$T(f)(0) = \int_0^0 f(t)\,dt = 0 \implies \lambda f(0) = 0.$$

Since $\lambda \neq 0$, we have $f(0) = 0$. Substituting into the general solution:

$$ce^0 = c = 0.$$

Thus $f(x)$ is the zero function. We conclude that the integration operator has **no eigenvalues** and consequently no one-dimensional invariant subspaces. This highlights a fundamental difference between finite-dimensional and infinite-dimensional spaces; in the latter, operators need not have non-trivial invariant subspaces.

範例

Remark.

The decomposition of a vector space into a direct sum of invariant subspaces allows us to study the operator on each subspace independently. If $V = W_1 \oplus W_2$ where both W_1 and W_2 are T-invariant, the matrix of T relative to a basis adapted to this decomposition is block diagonal:

$$[T] = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}.$$

Diagonalisation is the maximal case of this reduction, where each W_i is one-dimensional.

4.7 Exercises

In the following exercises, matrices are assumed to be over $\mathbb{R}$ unless specified otherwise. When asked to find eigenvectors, finding a basis for each eigenspace is sufficient.

1. **Eigensystem Calculations.** For each of the following matrices, determine the characteristic polynomial, the eigenvalues, their algebraic and geometric multiplicities, and a basis for each eigenspace. Determine if the matrix is diagonalisable over $\mathbb{R}$. If it is, find the transition matrix P and the diagonal matrix D.

   (a) $A = \begin{bmatrix} 2 & 1 & 2 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$

   (b) $B = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 3 & 0 \\ -1 & 4 & -1 \end{bmatrix}$

   (c) $C = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$

2. **Inspection and Intuition.** Without performing the full determinant calculation, find the eigenvalues and at least one eigenvector for the following matrices.

   (a) $A = \begin{bmatrix} -1 & 6 \\ 0 & 5 \end{bmatrix}$

   (b) $B = \begin{bmatrix} 3 & 0 & 0 \\ -2 & 7 & 0 \\ 4 & 8 & 1 \end{bmatrix}$

   (c) $C = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$

   Hint: Consider the row sums and the effect of $C - I$.

3. **Matrix Powers and Inverses.** Let $A = \begin{bmatrix} 7 & 2 & 3 \\ 3 & 10 & 3 \\ 2 & 0 & 6 \end{bmatrix}$.

   (a) Find the eigenvalues and eigenspaces of A.
   (b) Deduce the eigenvalues and eigenspaces of A^2.
   (c) Deduce the eigenvalues and eigenspaces of A^{-1}.

4. **Functions of Matrices.** Let $M = \begin{bmatrix} -5 & 3 \\ 6 & -2 \end{bmatrix}$.

   (a) Find an invertible matrix P and a diagonal matrix D such that $M = PDP^{-1}$.
   (b) Find a "cube root" of M, i.e., a matrix R such that $R^3 = M$.
   (c) Compute the matrix exponential $e^M = \sum_{n=0}^{\infty} \frac{1}{n!} M^n$. Express the entries in terms of e^{λ_i}.

5. **Polynomial Spaces.** Let V be the vector space of real polynomials in x of degree at most d, where $d > 0$. Determine which of the following linear operators $T : V \to V$ are diagonalisable.

   Find the matrix representation of each operator in the basis $\{1, x, \ldots, x^d\}$.

(a) $T_1(f(x)) = xf'(x)$
(b) $T_2(f(x)) = f'(x)$
(c) $T_3(f(x)) = f(x+1)$
(d) $T_4(f(x)) = f(-x)$

6. **The Transpose Operator.** Let $V = M_n(\mathbb{R})$. Define the operator $S : V \to V$ by $S(A) = A^T$.

   (a) Show that the only possible eigenvalues are ± 1.
   (b) Describe the eigenspaces corresponding to these eigenvalues.
   (c) Prove that S is diagonalisable by showing that the sum of the dimensions of the eigenspaces equals $\dim(V)$.

7. **Cayley-Hamilton Preview.** Let $\chi_A(x) = \det(xI - A)$ be the characteristic polynomial of a matrix A.

   (a) Prove that if $\mathbf{v}$ is an eigenvector of A with eigenvalue λ, then $\chi_A(A)\mathbf{v} = \mathbf{0}$.
   (b) Deduce that if A is diagonalisable, then $\chi_A(A) = 0$ (the zero matrix).

8. **Simultaneous Diagonalisation.** Let $S, T : V \to V$ be diagonalisable linear operators on a finite-dimensional space V. Suppose that S and T commute, i.e., $ST = TS$.

   (a) Let λ be an eigenvalue of S with eigenspace E_λ. Prove that E_λ is invariant under T (i.e., if $\mathbf{v} \in E_\lambda$, then $T(\mathbf{v}) \in E_\lambda$).
   (b) Deduce that there exists a common basis of eigenvectors for both S and T.

9. **Involutions.** Let V be a finite-dimensional vector space and $S : V \to V$ be a linear map such that $S^2 = I$ (an involution).

   (a) Prove that the only possible eigenvalues of S are 1 and -1.
   (b) Define $U = \{\mathbf{u} \in V \mid S\mathbf{u} = \mathbf{u}\}$ and $W = \{\mathbf{w} \in V \mid S\mathbf{w} = -\mathbf{w}\}$. Prove that $V = U \oplus W$ by showing that any $\mathbf{v}$ can be written as $\mathbf{v} = \frac{1}{2}(\mathbf{v} + S\mathbf{v}) + \frac{1}{2}(\mathbf{v} - S\mathbf{v})$.
   (c) Conclude that S is always diagonalisable.

10. **Nilpotent Operators.** Let E be a square matrix such that $E^{k+1} = 0$ for some integer $k \geq 1$.

    (a) Show that the only eigenvalue of E is 0.
    (b) Prove that $I - \lambda E$ is invertible for any scalar λ.

Consider the geometric series expansion $(I - X)^{-1} = I + X + X^2 + \ldots$, which terminates if X is nilpotent.

11. **Derivative of Matrix Powers.** Let $A(t)$ be a differentiable matrix-valued function.

    (a) Verify the product rule for matrices: $\frac{d}{dt}(A(t)^2) = A'(t)A(t) + A(t)A'(t)$.
    (b) Let $A = \begin{bmatrix} t & t^2 \\ 1 & t^3 \end{bmatrix}$. Calculate A^2 explicitly, differentiate it, and

verify the formula above.

12. **The Spectral Mapping Theorem for Polynomials.** Let $T : V \to V$ be a linear operator and let $\mathbf{v}$ be an eigenvector of T corresponding to the eigenvalue λ.

    (a) Let $S : V \to V$ be another linear operator such that $\mathbf{v}$ is also an eigenvector of S with eigenvalue μ. Prove that $\mathbf{v}$ is an eigenvector of the operator $aT + bS$ with eigenvalue $a\lambda + b\mu$ for any scalars a, b.

    (b) Prove by induction that $\mathbf{v}$ is an eigenvector of T^n with eigenvalue λ^n for any $n \in \mathbb{N}$.

    (c) Let $P(x) = a_n x^n + \cdots + a_1 x + a_0$ be a polynomial. Prove that $\mathbf{v}$ is an eigenvector of the operator $P(T) = a_n T^n + \cdots + a_1 T + a_0 I$ with eigenvalue $P(\lambda)$.

This result allows us to analyse the convergence of matrix power series, such as the matrix exponential e^A, by examining the eigenvalues of the constituent terms.

13. **Square Roots of Operators.**

    (a) Let $V = \mathbb{R}^2$. Let $T : V \to V$ be the rotation of the plane anti-clockwise through an angle of $\pi/2$ radians. Show that T possesses no eigenvalues in $\mathbb{R}$. However, prove that every non-zero vector in V is an eigenvector for the operator T^2. What is the corresponding eigenvalue?

    (b) Let $T : V \to V$ be an arbitrary linear operator. Suppose that T^2 possesses a non-negative eigenvalue λ^2 (where $\lambda \geq 0$). Prove that at least one of λ or $-\lambda$ is an eigenvalue for T.

Consider the factorisation $T^2 - \lambda^2 I = (T - \lambda I)(T + \lambda I)$. If $(T^2 - \lambda^2 I)\mathbf{v} = \mathbf{0}$, apply the operators on the right-hand side to $\mathbf{v}$.

14. **Eigenvalues of Differential and Shift Operators.**

    (a) Let V be the vector space of all real functions differentiable on the interval $(0, 1)$. Let $T : V \to V$ be the operator defined by $T(f)(t) = tf'(t)$. Prove that every real number λ is an eigenvalue of T and find the corresponding eigenfunctions.

    (b) Let V_n be the space of real polynomials of degree strictly less than n. Define the shift operator $T : V_n \to V_n$ by $T(p)(t) = p(t+1)$. Prove that T has only the eigenvalue 1. What are the corresponding eigenfunctions?

Consider the effect of T on the leading coefficient of the polynomial.

15. **Linear Independence and Scalar Operators.**

    (a) Let $T : V \to V$ be a linear operator. Suppose $\mathbf{x}$ and $\mathbf{y}$ are eigenvectors of T corresponding to distinct eigenvalues λ and μ. Prove that a linear combination $a\mathbf{x} + b\mathbf{y}$ cannot be an eigenvector of T unless one of the scalars a or b is zero.

    (b) **Scalar Operators.** Let $T : V \to V$ be a linear operator with the property that *every* non-zero vector in V is an eigenvector of T. Prove that T must be a scalar multiple of the identity operator; that is, there exists a scalar c such that $T(\mathbf{v}) = c\mathbf{v}$ for all $\mathbf{v} \in V$.

Let $\mathbf{x}, \mathbf{y}$ be linearly independent vectors. We know $T\mathbf{x} = \lambda_x \mathbf{x}$ and $T\mathbf{y} = \lambda_y \mathbf{y}$. Consider the action of T on the sum $\mathbf{x} + \mathbf{y}$ to show $\lambda_x = \lambda_y$.

5
Decoupling and Geometric Evolution

The power of the eigenvector decomposition lies in its ability to decouple dynamical systems. If a matrix A is diagonalisable, we can analyse the evolution of the system $\mathbf{x}_{n+1} = A\mathbf{x}_n$ by working in the eigenbasis, where the coordinates evolve independently.

Revisiting the Fibonacci system from the previous chapter, we found eigenvalues $\phi = \frac{1+\sqrt{5}}{2}$ and $\psi = \frac{1-\sqrt{5}}{2}$ with eigenvectors $\mathbf{v}_1 = [1, \phi]^T$ and $\mathbf{v}_2 = [1, \psi]^T$. Any initial state $\mathbf{x}_0$ can be decomposed as:

$$\mathbf{x}_0 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2.$$

Applying A repeated n times yields:

$$\mathbf{x}_n = A^n(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = c_1A^n\mathbf{v}_1 + c_2A^n\mathbf{v}_2 = c_1\phi^n\mathbf{v}_1 + c_2\psi^n\mathbf{v}_2.$$

This formula reveals the geometry of the state space:

1. **Decoupling:** The coefficient c_1 grows by a factor of ϕ each step, while c_2 shrinks by a factor of ψ (since $|\psi| < 1$). The "interaction" between components is an artifact of the standard basis; in the eigenbasis, the system consists of two independent 1D systems.

2. **Dominance:** For large n, the term involving ψ^n vanishes. The state vector $\mathbf{x}_n$ aligns almost perfectly with $\mathbf{v}_1$.

$$\mathbf{x}_n \approx c_1\phi^n\mathbf{v}_1.$$

This decoupling principle applies generally. For any diagonalisable linear operator T, we can construct a basis such that the matrix of T is diagonal. In this basis, the operator acts as a simple scaling in each coordinate direction, stripping away the complexity of the original coupled system.

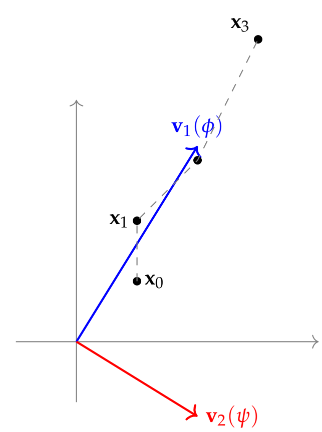

Figure 5.1: The Fibonacci sequence as stepping stones. The trajectory begins with "baby steps" but rapidly aligns with the dominant eigenvector $\mathbf{v}_1$, effectively ignoring the contracting direction $\mathbf{v}_2$.

5.1 Spectral Properties

Having established the definition of eigenvalues and their role in diagonalisation, we now deduce several theoretical consequences that relate the spectrum of a matrix to its macroscopic properties, such as the determinant and the trace.

Coefficients of the Characteristic Polynomial

The characteristic polynomial $p_A(\lambda) = \det(A - \lambda I)$ acts as a bridge between the matrix entries and its eigenvalues. By the Fundamental Theorem of Algebra, any polynomial of degree n with complex coefficients has exactly n roots in $\mathbb{C}$, counting multiplicity.

Proposition 5.1. ***Determinant and Eigenvalues.***
Let $A \in M_n(\mathbb{R})$ with eigenvalues $\lambda_1, \ldots, \lambda_n$ (repeated according to algebraic multiplicity). Then:

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

命題

Proof

Since $\lambda_1, \ldots, \lambda_n$ are the roots of $p_A(\lambda)$, we can factor the polynomial over $\mathbb{C}$:

$$\det(A - \lambda I) = c(\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_n).$$

The leading term of $\det(A - \lambda I)$ comes from the product of the diagonal entries $(a_{11} - \lambda) \cdots (a_{nn} - \lambda)$, which is $(-1)^n \lambda^n$. Thus $c = (-1)^n$. Setting $\lambda = 0$ in the factorisation:

$$\det(A) = p_A(0) = (-1)^n(-\lambda_1)(-\lambda_2) \cdots (-\lambda_n) = (-1)^n(-1)^n \prod_{i=1}^{n} \lambda_i = \prod_{i=1}^{n} \lambda_i.$$

■

Remark.

This proposition yields an immediate test for singularity: $\lambda = 0$ is an eigenvalue if and only if $\det(A) = 0$, consistent with our earlier determinant criterion for invertibility.

While not explicitly detailed in the definitions above, examining the coefficient of λ^{n-1} in the expansion of $\det(A - \lambda I)$ reveals another invariant. The sum of the roots equals the negative of the coefficient of λ^{n-1}, divided by the leading coefficient.

$$\sum_{i=1}^{n} \lambda_i = \mathrm{tr}(A) \equiv \sum_{i=1}^{n} a_{ii}.$$

Thus, the sum of the eigenvalues is the trace of the matrix.

Invariance Under Power Operations

The spectral mapping theorem states that operations applied to a matrix translate directly to operations on its eigenvalues.

Proposition 5.2. ***Eigenvalues of Powers.***
If λ is an eigenvalue of A with eigenvector $\mathbf{v}$, then λ^k is an eigenvalue of A^k with the same eigenvector $\mathbf{v}$, for any integer $k \geq 1$.

命題

Proof

We proceed by induction. The base case $k = 1$ is the definition $A\mathbf{v} = \lambda\mathbf{v}$. Assume $A^{k-1}\mathbf{v} = \lambda^{k-1}\mathbf{v}$. Then:

$$A^k\mathbf{v} = A(A^{k-1}\mathbf{v}) = A(\lambda^{k-1}\mathbf{v}) = \lambda^{k-1}(A\mathbf{v}) = \lambda^{k-1}(\lambda\mathbf{v}) = \lambda^k\mathbf{v}.$$

■

Triangular Matrices

For triangular matrices, the spectrum is immediately visible.

Proposition 5.3. ***Spectrum of Triangular Matrices.***
Let A be an upper or lower triangular matrix. The eigenvalues of A are exactly its diagonal entries.

命題

Proof

Consider the characteristic matrix $A - \lambda I$. If A is triangular, then $A - \lambda I$ is also triangular. The determinant of a triangular matrix is the product of its diagonal entries (see *Determinant of Triangular Matrices*):

$$\det(A - \lambda I) = \prod_{i=1}^{n}(a_{ii} - \lambda).$$

The roots of this polynomial are precisely $\lambda = a_{ii}$.

■

Example 5.1. Determinant Calculation via LU. This property simplifies determinant calculation significantly. If we perform an LU decomposition $A = LU$, the eigenvalues of L are all 1 (if unit triangular) and the eigenvalues of U are its diagonal entries. Since $\det(A) = \det(L)\det(U)$, the determinant is simply the product of the diagonal entries of U.

範例

The Trace and Polynomial Coefficients

While the determinant provides the constant term of the characteristic polynomial, the other coefficients encode equally fundamental invariants.

Proposition 5.4. ***Coefficients of the Characteristic Polynomial.***
Let $A \in M_n(\mathbb{R})$. The characteristic polynomial $p_A(\lambda) = \det(A - \lambda I)$ is a polynomial of degree n of the form:

$$p_A(\lambda) = (-1)^n \lambda^n + (-1)^{n-1} \mathrm{tr}(A) \lambda^{n-1} + \cdots + \det(A).$$

命題

Proof

The term $\det(A - \lambda I)$ is a sum of products involving entries from the matrix. The highest powers of λ arise solely from the product of the diagonal terms:

$$\prod_{i=1}^{n} (a_{ii} - \lambda) = (-1)^n \lambda^n + (-1)^{n-1} \left(\sum_{i=1}^{n} a_{ii} \right) \lambda^{n-1} + (\text{terms of degree } < n-1).$$

No other term in the determinant expansion (which involves permutations) contributes to λ^n or λ^{n-1}, as any permutation deviating from the diagonal must involve at least two off-diagonal elements, reducing the degree of λ by at least 2. The constant term is found by evaluating at $\lambda = 0$: $p_A(0) = \det(A)$.

■

Definition 5.1. ***Trace.***
The **trace** of a square matrix A, denoted $\mathrm{tr}(A)$, is the sum of its diagonal entries:

$$\mathrm{tr}(A) = \sum_{i=1}^{n} a_{ii}.$$

定義

Combining *proposition* 5.4 with the factorisation $p_A(\lambda) = (-1)^n \prod(\lambda - \lambda_i)$, we obtain a direct relationship between the eigenvalues and the trace.

Corollary 5.1. Trace as Sum of Eigenvalues

推論

The sum of the eigenvalues of A (counting algebraic multiplicity) equals the trace of A:

$$\sum_{i=1}^{n} \lambda_i = \mathrm{tr}(A).$$

Remark.

This property, along with $\prod \lambda_i = \det(A)$, provides a rapid sanity check for calculated eigenvalues. In example 4.3, the eigenvalues $1, -1, 3$ sum to 3. The trace of A is $2 + 3 + (-2) = 3$.

5.2 Generalized Eigenvectors and the Jordan Structure

We previously identified a class of *defective* matrices for which the geometric multiplicity of an eigenvalue is strictly less than its algebraic multiplicity. In such cases, there are insufficient eigenvectors to form a basis for $\mathbb{R}^n$, and the matrix cannot be diagonalised. To remedy this, we introduce *generalized eigenvectors*.

Generalized Eigenvectors

Definition 5.2. ***Generalized Eigenvector.***
Let $A \in M_n(\mathbb{R})$ and let λ be an eigenvalue of A. A non-zero vector $\mathbf{v}$ is a **generalized eigenvector of order** k if

$$(A - \lambda I)^k \mathbf{v} = \mathbf{0} \quad \text{and} \quad (A - \lambda I)^{k-1} \mathbf{v} \neq \mathbf{0}.$$

定義

Note that a generalized eigenvector of order 1 is a standard eigenvector.

Jordan Chains

Generalized eigenvectors naturally organize themselves into chains generated by the operator $A - \lambda I$.

Proposition 5.5. ***Jordan Chain Construction.***
Suppose A has an eigenvalue λ with eigenvector $\mathbf{v}_1$. If the equation

$$(A - \lambda I)\mathbf{v}_2 = \mathbf{v}_1$$

has a solution $\mathbf{v}_2$, then $\mathbf{v}_2$ is a generalized eigenvector of order 2. Furthermore, $\{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly independent.

命題

Proof

Multiplying the defining equation by $A - \lambda I$ gives:

$$(A - \lambda I)^2 \mathbf{v}_2 = (A - \lambda I)\mathbf{v}_1 = \mathbf{0}.$$

Since $(A - \lambda I)\mathbf{v}_2 = \mathbf{v}_1 \neq \mathbf{0}$, $\mathbf{v}_2$ is of order 2. To prove independence, assume $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = \mathbf{0}$. Applying $A - \lambda I$ to this equation:

$$c_1(A - \lambda I)\mathbf{v}_1 + c_2(A - \lambda I)\mathbf{v}_2 = \mathbf{0} \implies c_1\mathbf{0} + c_2\mathbf{v}_1 = \mathbf{0}.$$

Since $\mathbf{v}_1 \neq \mathbf{0}$, we must have $c_2 = 0$. Substituting back yields $c_1\mathbf{v}_1 = \mathbf{0} \implies c_1 = 0$.

■

This process can be iterated. A sequence $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ such that

$$(A - \lambda I)\mathbf{v}_j = \mathbf{v}_{j-1} \quad (\text{with } \mathbf{v}_0 = \mathbf{0})$$

is called a **Jordan chain** of length k. The vector $\mathbf{v}_1$ is the eigenvector, and $\mathbf{v}_k$ is the generalized eigenvector of order k.

Proposition 5.6. ***Independence of Chains.***
A Jordan chain $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ consists of linearly independent vectors.

命題

The proof generalizes the argument for $k = 2$ by repeatedly applying $A - \lambda I$ to peel off terms from a linear combination.

Example 5.2. Constructing a Jordan Basis. Consider the matrix from example 4.10:

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

We found $\lambda = 1$ with algebraic multiplicity 2 and geometric multiplicity 1. The only eigenvector is $\mathbf{u}_1 = [1, 0]^T$ (up to scaling). We seek a generalized eigenvector $\mathbf{u}_2$ such that $(A - I)\mathbf{u}_2 = \mathbf{u}_1$.

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

This yields the equation $v = 1$. The variable u is free; we choose $u = 0$ for simplicity. Thus $\mathbf{u}_2 = [0, 1]^T$. The set $\mathcal{J} = \{\mathbf{u}_1, \mathbf{u}_2\}$ forms a basis for $\mathbb{R}^2$. In this basis, the matrix of the transformation is:

$$P^{-1}AP = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

This is a **Jordan block**.

範例

The Jordan Normal Form

The generalized eigenvectors allow us to construct a basis for $\mathbb{R}^n$ even when the matrix is defective.

Theorem 5.1. ***Jordan Normal Form.***
Every matrix $A \in M_n(\mathbb{R})$ with real eigenvalues is similar to a block diagonal matrix J, called the **Jordan Canonical Form**:

$$J = \begin{bmatrix} J_1 & & 0 \\ & \ddots & \\ 0 & & J_p \end{bmatrix},$$

where each block J_i corresponds to a Jordan chain and has the form:

$$J_i = \begin{bmatrix} \lambda & 1 & & 0 \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda \end{bmatrix}.$$

The columns of the similarity matrix P constitute a **Jordan basis**, consisting of Jordan chains for each eigenvalue.

定理

Note

If A is diagonalisable, all Jordan blocks are 1×1, and J is simply a diagonal matrix. The presence of 1s on the super-diagonal indicates the "coupling" between generalized eigenvectors in a chain.

Example 5.3. A Larger Jordan Structure. Consider a 4×4 matrix A with $\lambda = 1$ appearing 4 times as a root of $p_A(\lambda)$. Suppose $(A - I)$ row reduces to show only two free variables. This implies there are two eigenvectors, $\mathbf{u}_1$ and $\mathbf{u}_2$. To form a basis, we need two more vectors. We look for chains. Suppose we solve $(A - I)\mathbf{u}_3 = \mathbf{u}_1$ to find $\mathbf{u}_3$, and $(A - I)\mathbf{u}_4 = \mathbf{u}_2$ to find $\mathbf{u}_4$. The Jordan basis is $\{\mathbf{u}_1, \mathbf{u}_3, \mathbf{u}_2, \mathbf{u}_4\}$. The resulting matrix J will have two 2×2 blocks:

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

The structure of the Jordan form (number and size of blocks) is completely determined by the dimensions of the kernels $\ker(A - \lambda I)^k$.

範例

Example 5.4. Explicit Jordan Construction (3x3). Let $A =$

$\begin{bmatrix} 2 & 1 & 0 \\ -1 & 4 & 0 \\ -1 & 1 & 3 \end{bmatrix}$. The characteristic polynomial is:

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

$= (3-\lambda)((2-\lambda)(4-\lambda)+1) = (3-\lambda)(\lambda^2 - 6\lambda + 9) = -(\lambda-3)^3$.

So $\lambda = 3$ with algebraic multiplicity 3. Solve $(A - 3I)\mathbf{v} = \mathbf{0}$:

$$A - 3I = \begin{bmatrix} -1 & 1 & 0 \\ -1 & 1 & 0 \\ -1 & 1 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The equation is $x - y = 0 \implies x = y$. z is free. Basis for eigenspace: $\mathbf{v}_1 = [1,1,0]^T$ and $\mathbf{v}_2 = [0,0,1]^T$. Geometric multiplicity is 2. Since $2 < 3$, the matrix is defective. We need one generalized eigenvector. We need a generalized eigenvector $\mathbf{u}_2$ such that $(A - 3I)\mathbf{u}_2 = \mathbf{u}_1$ where $\mathbf{u}_1$ is an eigenvector. We observe that the range of $A - 3I$ is spanned by $[-1,-1,-1]^T$. Our eigenvector $\mathbf{u}_1$ must lie in this range. We find a linear combination $\mathbf{u}_1 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = [c_1, c_1, c_2]^T$ that is a multiple of $[1,1,1]^T$. Setting $c_1 = c_2 = 1$ gives $\mathbf{u}_1 = [1,1,1]^T$, which is a valid eigenvector. Now we solve $(A - 3I)\mathbf{u}_2 = \mathbf{u}_1$:

$$\begin{bmatrix} -1 & 1 & 0 \\ -1 & 1 & 0 \\ -1 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

All rows imply $-x + y = 1$. Let $x = 0, y = 1, z = 0$. So $\mathbf{u}_2 = [0,1,0]^T$. We now have a Jordan chain $\{\mathbf{u}_1, \mathbf{u}_2\}$. To complete the basis, we need a third vector $\mathbf{w}$ that is an eigenvector linearly independent of $\mathbf{u}_1$. We can choose $\mathbf{w} = \mathbf{v}_2 = [0,0,1]^T$. The Jordan basis is $P = [\mathbf{u}_1, \mathbf{u}_2, \mathbf{w}] = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$. The Jordan form is $J = \begin{bmatrix} 3 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}$.

範例

5.3 *Diagonalisation and Similarity*

We have seen that if a matrix A has n linearly independent eigenvectors, it can be diagonalised. However, the true utility of this decomposition is not just computational but structural: it reveals that the matrix A is essentially a diagonal matrix "viewed from a different angle" (i.e., in a different basis).

Criteria for Real Diagonalisation

Proposition 5.7. ***Criterion for Real Diagonalisation.***
Let $A \in M_n(\mathbb{R})$. Suppose the characteristic polynomial $p_A(\lambda)$ factors completely into real linear factors:

$$p_A(\lambda) = (-1)^n(\lambda - \lambda_1)^{m_1} \cdots (\lambda - \lambda_k)^{m_k},$$

where $\lambda_1, \ldots, \lambda_k$ are distinct real eigenvalues with algebraic multiplicities m_j. Let $n_j = \dim(E_{\lambda_j})$ be the geometric multiplicity of λ_j. Then A is diagonalisable if and only if $m_j = n_j$ for all $j = 1, \ldots, k$.

命题

This condition ensures that the sum of the dimensions of the eigenspaces is exactly n, allowing us to construct a basis of eigenvectors for $\mathbb{R}^n$.

The Matrix of the Transformation

If the conditions of *proposition* 5.7 are met, we can explicitly construct the diagonalising matrix.

Theorem 5.2. ***Diagonalisation Formula.***
Suppose A has n linearly independent eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ corresponding to eigenvalues $\lambda_1, \ldots, \lambda_n$. Let $P = [\mathbf{v}_1 \mid \cdots \mid \mathbf{v}_n]$. Then P is invertible and

$$P^{-1}AP = D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n).$$

定理

Proof

Since the columns of P are linearly independent, P is invertible.

$$AP = A[\mathbf{v}_1 \mid \cdots \mid \mathbf{v}_n] = [\lambda_1\mathbf{v}_1 \mid \cdots \mid \lambda_n\mathbf{v}_n].$$

Also,

$$PD = [\mathbf{v}_1 \mid \cdots \mid \mathbf{v}_n] \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} = [\lambda_1\mathbf{v}_1 \mid \cdots \mid \lambda_n\mathbf{v}_n].$$

Thus $AP = PD$, which implies $P^{-1}AP = D$.

■

Example 5.5. Diagonalisation in Practice. Revisiting the matrix $A = \begin{bmatrix} 3 & 1 \\ 3 & 1 \end{bmatrix}$. The eigenvalues are $\lambda_1 = 0$ and $\lambda_2 = 4$. The eigenvectors are $\mathbf{u}_1 = [1, -3]^T$ and $\mathbf{u}_2 = [1, 1]^T$. Let $P = \begin{bmatrix} 1 & 1 \\ -3 & 1 \end{bmatrix}$. The determinant is $1 - (-3) = 4$, so

$$P^{-1} = \frac{1}{4} \begin{bmatrix} 1 & -1 \\ 3 & 1 \end{bmatrix}.$$

Computing the product:

$$P^{-1}AP = \frac{1}{4} \begin{bmatrix} 1 & -1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -3 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix}.$$

This confirms the diagonalisation.

範例

Remark.

Calculating the inverse P^{-1} is computationally expensive. However, in many applications (such as powers A^k), we only need the decomposition $A = PDP^{-1}$, and explicit inversion might be avoidable if we work with orthonormal bases (where $P^{-1} = P^T$).

5.4 *Complex Eigenvalues and Canonical Forms*

When a real matrix has complex eigenvalues, it cannot be diagonalised over $\mathbb{R}$. However, the complex eigenvectors still provide a structural decomposition involving rotations.

Complex Conjugate Pairs

Proposition 5.8. ***Conjugate Eigenpairs.***
Let $A \in M_n(\mathbb{R})$. If $\lambda = \alpha + i\beta$ (with $\beta \neq 0$) is an eigenvalue with eigenvector $\mathbf{v} = \mathbf{a} + i\mathbf{b}$, then $\bar{\lambda} = \alpha - i\beta$ is an eigenvalue with eigenvector $\bar{\mathbf{v}} = \mathbf{a} - i\mathbf{b}$.

命題

Proof

Since A is real, $\overline{A\mathbf{v}} = A\bar{\mathbf{v}}$. Also $\overline{\lambda\mathbf{v}} = \bar{\lambda}\bar{\mathbf{v}}$. Thus $A\bar{\mathbf{v}} = \bar{\lambda}\bar{\mathbf{v}}$. Since $\mathbf{v} \neq \mathbf{0}$ implies $\mathbf{a}, \mathbf{b}$ are not both zero, $\bar{\mathbf{v}} \neq \mathbf{0}$.

■

Furthermore, the real vectors $\mathbf{a}$ and $\mathbf{b}$ carry significant geometric information.

Proposition 5.9. ***Independence of Real and Imaginary Parts.***
If $\mathbf{v} = \mathbf{a} + i\mathbf{b}$ is an eigenvector for a complex eigenvalue λ ($\beta \neq 0$), then $\{\mathbf{a}, \mathbf{b}\}$ is a linearly independent set in $\mathbb{R}^n$.

命题

Proof

Suppose $c_1\mathbf{a} + c_2\mathbf{b} = \mathbf{0}$. We can express $\mathbf{a} = (\mathbf{v} + \bar{\mathbf{v}})/2$ and $\mathbf{b} = (\mathbf{v} - \bar{\mathbf{v}})/2i$. Substituting these into the linear combination yields a relation between $\mathbf{v}$ and $\bar{\mathbf{v}}$. Since $\mathbf{v}$ and $\bar{\mathbf{v}}$ correspond to distinct eigenvalues λ and $\bar{\lambda}$, they are linearly independent over $\mathbb{C}$. This forces the coefficients c_1, c_2 to vanish.

■

Real Canonical Form

Using the basis $\mathcal{B} = \{\mathbf{b}, \mathbf{a}\}$ (note the order for convention), the matrix A restricts to a block form that makes the rotation explicit.

Theorem 5.3. ***Real Block Diagonalisation.***
Let A have a complex eigenvalue $\lambda = \alpha + i\beta$ with eigenvector $\mathbf{v} = \mathbf{a} + i\mathbf{b}$. Let $P = [\mathbf{a} \mid \mathbf{b} \mid \cdots]$. Then the block of $P^{-1}AP$ corresponding to these vectors is:

$$C = \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix} = \sqrt{\alpha^2 + \beta^2} \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix}.$$

This represents a scaling by the magnitude $|\lambda| = \sqrt{\alpha^2 + \beta^2}$ and a rotation by the angle $\phi = \arg(\bar{\lambda})$ (which is $-\arg(\lambda)$).

定理

Proof

We have $A(\mathbf{a} + i\mathbf{b}) = (\alpha + i\beta)(\mathbf{a} + i\mathbf{b}) = (\alpha\mathbf{a} - \beta\mathbf{b}) + i(\beta\mathbf{a} + \alpha\mathbf{b})$. Equating real and imaginary parts:

$$A\mathbf{a} = \alpha\mathbf{a} - \beta\mathbf{b}, \quad A\mathbf{b} = \beta\mathbf{a} + \alpha\mathbf{b}.$$

Thus, relative to the basis $\{\mathbf{a}, \mathbf{b}\}$:

$$[A]_{\{\mathbf{a},\mathbf{b}\}} = \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}.$$

■

Example 5.6. Complex Decomposition. Let $A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$. Eigenvalues are $1 \pm i$. For $\lambda = 1+i$, $(A-(1+i)I)\mathbf{v} = \mathbf{0} \implies \begin{bmatrix} -i & -1 \\ 1 & -i \end{bmatrix}\mathbf{v} = \mathbf{0}$. This gives $v_1 = iv_2$. Let $v_2 = 1$, so $\mathbf{v} = \begin{bmatrix} i \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} + i\begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Here $\mathbf{a} = [0,1]^T$ and $\mathbf{b} = [1,0]^T$.

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

Check:

$$P^{-1}AP = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}.$$

This matches the canonical form with $\alpha = 1, \beta = 1$.

範例

This canonical form is robust: it allows us to decompose any real linear operator into a sum of scalings and rotations, avoiding complex arithmetic in the final result.

5.5 *Polynomial Methods and Annihilating Polynomials*

We have primarily determined eigenvalues by solving the characteristic equation $\det(A - \lambda I) = 0$. We now introduce an alternative algebraic approach that connects the linear dependence of matrix powers to polynomial factorisation. This framework provides a constructive method for finding eigenvectors and offers a precise criterion for diagonalisability based on the roots of specific polynomials.

Annihilating Polynomials of Vectors

Let $A \in M_n(\mathbb{C})$ and let $w \in \mathbb{C}^n$ be a non-zero vector. Consider the sequence of vectors $w, Aw, A^2w, \ldots$. Since the space is finite-dimensional, this sequence cannot be linearly independent indefinitely.

Definition 5.3. ***The Minimal Polynomial of a Vector.***
Let m be the smallest integer such that the set $\{w, Aw, \ldots, A^m w\}$ is linearly dependent. Then $A^m w$ can be uniquely expressed as a linear combination of the preceding vectors:

$$A^m w = -a_0 w - a_1 Aw - \cdots - a_{m-1}A^{m-1}w.$$

We define the polynomial $p_w(t)$ as:

$$p_w(t) = t^m + a_{m-1}t^{m-1} + \cdots + a_1 t + a_0.$$

This is the monic polynomial of lowest degree such that $p_w(A)w = \mathbf{0}$.

定義

Theorem 5.4. ***Constructive Existence of Eigenvectors.***
Let $p_w(t)$ be the polynomial defined above. If λ is a root of $p_w(t)$, then there exists a vector v constructed from w that is an eigenvector of A with eigenvalue λ.

定理

Proof

By the Fundamental Theorem of Algebra, we factor $p_w(t) = (t - \lambda)q(t)$, where $q(t)$ has degree $m - 1$. Define the vector $v = q(A)w$. First, we observe that $v \neq \mathbf{0}$. Since m is the minimal degree such that $A^k w$ are dependent, and $\deg(q) = m-1$, the vector $q(A)w$ is a non-trivial linear combination of linearly independent vectors $\{w, \ldots, A^{m-1}w\}$. Second, we apply $(A - \lambda I)$:

$$(A - \lambda I)v = (A - \lambda I)q(A)w = p_w(A)w = \mathbf{0}.$$

Thus $Av = \lambda v$, and v is an eigenvector.

■

Computational Procedure via Row Reduction

Finding the coefficients a_i is a problem of linear dependence, which is efficiently solved via row reduction. Form the matrix

$$M = [w \mid Aw \mid A^2 w \mid \cdots \mid A^n w].$$

Row reducing M reveals the first non-pivotal column. If column $m+1$ (corresponding to $A^m w$) is the first non-pivot, the coefficients in that column relative to the pivots give the linear relation:

$$A^m w = b_0 w + b_1 Aw + \cdots + b_{m-1}A^{m-1}w.$$

The polynomial is then $p_w(t) = t^m - \sum_{j=0}^{m-1} b_j t^j$.

Example 5.7. Eigenbasis Construction. Let $A = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}$.
We seek eigenvalues using $w = \mathbf{e}_1$. We compute the Krylov se-

quence:

$$\mathbf{e}_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}, \quad A\mathbf{e}_1 = \begin{bmatrix}1\\-1\\0\end{bmatrix}, \quad A^2\mathbf{e}_1 = \begin{bmatrix}2\\-3\\1\end{bmatrix}, \quad A^3\mathbf{e}_1 = \begin{bmatrix}5\\-9\\4\end{bmatrix}.$$

Forming the matrix $M = [\mathbf{e}_1, A\mathbf{e}_1, A^2\mathbf{e}_1, A^3\mathbf{e}_1]$ and row reducing yields:

$$\text{rref}(M) = \begin{bmatrix}1 & 0 & 0 & 0\\0 & 1 & 0 & -3\\0 & 0 & 1 & 4\end{bmatrix}.$$

The fourth column indicates $A^3\mathbf{e}_1 = 0\mathbf{e}_1 - 3A\mathbf{e}_1 + 4A^2\mathbf{e}_1$. Thus, $p(t) = t^3 - 4t^2 + 3t = t(t-1)(t-3)$. The roots are $0, 1, 3$. We construct eigenvectors for each:

1. $\lambda = 0$: $q(t) = \frac{p(t)}{t} = t^2 - 4t + 3$. $v_1 = (A^2 - 4A + 3I)\mathbf{e}_1 = [1, 1, 1]^T$.
2. $\lambda = 1$: $q(t) = \frac{p(t)}{t-1} = t^2 - 3t$. $v_2 = (A^2 - 3A)\mathbf{e}_1 = [-1, 0, 1]^T$.
3. $\lambda = 3$: $q(t) = \frac{p(t)}{t-3} = t^2 - t$. $v_3 = (A^2 - A)\mathbf{e}_1 = [1, -2, 1]^T$.

範例

Criterion for the Existence of an Eigenbasis

While the standard diagonalisation theorem relies on geometric multiplicities, the polynomial perspective offers a criterion based on the simplicity of roots. Let $p_i(t)$ denote the minimal annihilating polynomial for the standard basis vector $\mathbf{e}_i$.

Theorem 5.5. ***Polynomial Roots and Diagonalisability.***
Let $A \in M_n(\mathbb{C})$. There exists an eigenbasis for A if and only if for every $i \in \{1, \dots, n\}$, the roots of the polynomial $p_i(t)$ are all distinct (simple).

定理

Sufficiency:

Assume each $p_i(t)$ has distinct roots. Let $E_i = \text{span}\{\mathbf{e}_i, A\mathbf{e}_i, \dots\}$. The dimension of E_i is $\deg(p_i)$. The procedure above generates $\deg(p_i)$ linearly independent eigenvectors $v_{i,j}$ (one for each root of p_i) which span E_i. Since the union of subspaces E_i spans $\mathbb{C}^n$ (as it contains all $\mathbf{e}_i$), the union of all generated eigenvectors spans $\mathbb{C}^n$, ensuring an eigenbasis exists.

証明終

Necessity:

Let $p(t)$ be the polynomial with distinct roots $\prod(t - \lambda_k)$ covering all distinct eigenvalues of A. If A is diagonalisable, then $p(A) = 0$ (since $p(D) = 0$ requires only that p vanishes on the diagonal entries). Consequently, for any vector $\mathbf{x}$, $p(A)\mathbf{x} = \mathbf{0}$. The minimal polynomial $p_i(t)$ for $\mathbf{e}_i$ must therefore divide $p(t)$. Since $p(t)$ has distinct roots, any factor $p_i(t)$ must also have distinct roots.

証明終

Remark.

This relates to the concept of the *minimal polynomial* of a matrix, denoted $\mu_A(t)$. A matrix is diagonalisable if and only if $\mu_A(t)$ splits into distinct linear factors. The polynomials $p_i(t)$ are divisors of $\mu_A(t)$.

5.6 *Exercises*

In the following exercises, matrices are assumed to be over $\mathbb{R}$ unless specified otherwise. When asked to find eigenvectors, finding a basis for each eigenspace is sufficient.

1. **Defective Matrices and Jordan Chains.** Let $A = \begin{bmatrix} 1 & 4 \\ -1 & -3 \end{bmatrix}$.
   (a) Find the eigenvalue λ and show it has algebraic multiplicity 2 but geometric multiplicity 1.
   (b) Find an eigenvector $\mathbf{v}_1$.
   (c) Find a generalized eigenvector $\mathbf{v}_2$ such that $(A - \lambda I)\mathbf{v}_2 = \mathbf{v}_1$.
   (d) Form the matrix $P = [\mathbf{v}_1 \mid \mathbf{v}_2]$ and compute $J = P^{-1}AP$.
2. **Minimal Polynomials and Eigenbasis.** Let A be a block diagonal matrix given by $A = \operatorname{diag}(J_2(7), J_3(2))$.
   (a) Write down the matrix explicitly.
   (b) Find the characteristic polynomial.
   (c) Find the minimal polynomial (the lowest degree polynomial $m(t)$ such that $m(A) = 0$).
   (d) Verify that the minimal polynomial has repeated roots, consistent with the fact that A is not diagonalisable.
3. $\star$ **The Matrix Exponential for Jordan Blocks.** Let $J = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$.
   (a) Write $J = \lambda I + N$, where $N = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$.
   (b) Show that (λI) and N commute.

(c) Use the property $e^{X+Y} = e^X e^Y$ (valid for commuting matrices) to compute e^{tJ}.

(d) Generalize this to a 3×3 Jordan block.

4. $\star$ **Annihilating Polynomials.** Let $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -3 & 3 \end{bmatrix}$.

(a) Calculate the sequence of vectors $\mathbf{e}_1, A\mathbf{e}_1, A^2\mathbf{e}_1, A^3\mathbf{e}_1$.

(b) Find the linear dependence relation $A^3\mathbf{e}_1 = c_0\mathbf{e}_1 + c_1 A\mathbf{e}_1 + c_2 A^2\mathbf{e}_1$.

(c) Construct the polynomial $p(t) = t^3 - c_2 t^2 - c_1 t - c_0$.

(d) Find the roots of $p(t)$ and use the constructive theorem to find the eigenvectors of A.

5. **Spectral Analysis in $\mathbb{R}^3$.** For each of the following matrices, calculate the characteristic polynomial, the eigenvalues, and a basis for each eigenspace. Determine whether the matrix is diagonalisable over $\mathbb{R}$.

(a) $A = \begin{bmatrix} 2 & 1 & 1 \\ 2 & 3 & 2 \\ 3 & 3 & 4 \end{bmatrix}$

(b) $B = \begin{bmatrix} 2 & -1 & 1 \\ 0 & 3 & -1 \\ 2 & 1 & 3 \end{bmatrix}$

(c) $C = \begin{bmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix}$

6. **Physics Application: Pauli Matrices.** The Pauli spin matrices, fundamental to the quantum mechanical treatment of electron spin, are defined as:

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

(a) Verify that each matrix has eigenvalues 1 and -1.

(b) Prove that these matrices anti-commute (e.g., $\sigma_x\sigma_y = -\sigma_y\sigma_x$) and that the square of each is the identity matrix.

(c) **Generalisation:** Determine the form of the most general 2×2 matrix with complex entries having eigenvalues 1 and -1.

7. **The Trace and Characteristic Coefficients.** Let A be an $n \times n$ matrix with characteristic polynomial $p_A(\lambda) = \det(A - \lambda I) = c_n\lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_0$.

(a) Prove that $c_n = (-1)^n$.

(b) Prove that $c_0 = \det(A)$.

(c) Prove by induction that the coefficient of λ^{n-1} is given by $c_{n-1} = (-1)^{n-1}\text{tr}(A)$.
(d) Using these results, show that for 2×2 matrices, $p_A(\lambda) = \lambda^2 - \text{tr}(A)\lambda + \det(A)$.

8. **The Spectral Mapping Theorem.** Let $T : V \to V$ be a linear operator and let $\mathbf{v}$ be an eigenvector of T corresponding to the eigenvalue λ.

(a) Prove that $\mathbf{v}$ is an eigenvector of T^k with eigenvalue λ^k for any $k \in \mathbb{N}$.
(b) Let $P(x) = a_n x^n + \cdots + a_0$ be a polynomial. Prove that $\mathbf{v}$ is an eigenvector of the operator $P(T)$ with eigenvalue $P(\lambda)$.
(c) **Invertibility:** Prove that T is invertible if and only if $\lambda \neq 0$. If T is invertible, show that $\mathbf{v}$ is an eigenvector of T^{-1} with eigenvalue λ^{-1}.

9. **Commutativity of Spectra.** Let A and B be $n \times n$ matrices.

(a) Prove that A and its transpose A^T share the same characteristic polynomial and thus the same eigenvalues. Do they necessarily share the same eigenvectors?
(b) If A is non-singular, prove that AB and BA have the same eigenvalues.
(c) $\star$ Argue that AB and BA share the same characteristic polynomial even if A and B are singular.

Hint: Consider the similarity transformation using A.

Consider perturbation $A_\epsilon = A - \epsilon I$, or block matrix determinants.

10. **Similarity vs. Spectrum.** Two matrices are similar if they represent the same linear operator in different bases. Similar matrices must share the same eigenvalues (why?).

(a) Prove that the matrices

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

have identical eigenvalues but are *not* similar.
(b) Find a non-singular matrix C such that $C^{-1}MC$ is diagonal for:

$$M = \begin{bmatrix} 4 & 2 \\ 3 & 3 \end{bmatrix}.$$

(c) Explain why no such C exists for the matrix A in part (a).

11. **Complex Structures on Real Spaces.** Let A be an $n \times n$ matrix with real entries such that $A^2 = -I$.

(a) Prove that A is non-singular.
(b) Prove that A has no real eigenvalues.
(c) Prove that $\det(A) = 1$.
(d) Deduce that n must be an even number.

This algebraic structure defines a *complex structure* on a real vector space, allowing it to be treated as a complex vector space of dimension $n/2$.

12. **Operator Analysis: Differential Equations.** Let $V = C^\infty(\mathbb{R})$ be the space of smooth real functions.
    (a) Let $T(f) = f'$. Show that every real number λ is an eigenvalue of T. Determine the corresponding eigenspaces.
    (b) Let $S(f) = f''$. Show that every real number μ is an eigenvalue. Describe the eigenspace for $\mu > 0$ and $\mu < 0$.

13. **Operator Analysis: Integral Operators.** Let V be the space of continuous functions on $(-\infty, \infty)$ such that the specified integrals exist.
    (a) Let $T(f)(x) = \int_{-\infty}^{x} f(t)\,dt$. Prove that T has no eigenvalues.
    (b) $\star$ Let $S(f)(x) = \int_{-\infty}^{x} e^{x-t} f(t)\,dt$. Prove that every $\lambda > 1$ is an eigenvalue for S. Find the eigenfunctions.

Transform the integral equation $S(f) = \lambda f$ into a differential equation by differentiating with respect to x.