The Little Book of Linear Algebra

Version 0.1.1

Duc-Tam Nguyen

September 4, 2025

Contents

Chapter 1. Vectors

1.1 Scalars and Vectors

A scalar is a single numerical quantity, most often taken from the real numbers, denoted by $\mathbb{R}$. Scalars are
the fundamental building blocks of arithmetic: they can be added, subtracted, multiplied, and, except in the case of
zero, divided. In linear algebra, scalars play the role of coefficients, scaling factors, and entries of larger
structures such as vectors and matrices. They provide the weights by which more complex objects are measured and
combined. A vector is an ordered collection of scalars, arranged either in a row or a column. When the scalars are real
numbers, the vector is said to belong to *real n-dimensional space*, written

$$\mathbb{R}^n = \{(x_1, x_2, \ldots, x_n) \mid x_i \in \mathbb{R}\}.$$

An element of $\mathbb{R}^n$ is called a vector of dimension n or an n-vector. The number n is called the
dimension of the vector space. Thus $\mathbb{R}^2$ is the space of all ordered pairs of real numbers, $\mathbb{R}^3$ the
space of all ordered triples, and so on.

Example 1.1.1.

- A 2-dimensional vector: $(3, -1) \in \mathbb{R}^2$.

- A 3-dimensional vector: $(2, 0, 5) \in \mathbb{R}^3$.

- A 1-dimensional vector: $(7) \in \mathbb{R}^1$, which corresponds to the scalar 7 itself.

Vectors are often written vertically in column form, which emphasizes their role in matrix multiplication:

$$\mathbf{v} = \begin{bmatrix} 2 \\ 0 \\ 5 \end{bmatrix} \in \mathbb{R}^3.$$

The vertical layout makes the structure clearer when we consider linear combinations or multiply matrices by vectors.

Geometric Interpretation

In $\mathbb{R}^2$, a vector (x_1, x_2) can be visualized as an arrow starting at the origin $(0, 0)$ and ending at the
point (x_1, x_2). Its length corresponds to the distance from the origin, and its orientation gives a direction in the
plane. In $\mathbb{R}^3$, the same picture extends into three dimensions: a vector is an arrow from

the origin
to (x_1, x_2, x_3). Beyond three dimensions, direct visualization is no longer possible, but the algebraic rules of
vectors remain identical. Even though we cannot draw a vector in $\mathbb{R}^{10}$, it behaves under addition, scaling,
and transformation exactly as a 2- or 3-dimensional vector does. This abstract point of view is what allows linear
algebra to apply to data science, physics, and machine learning, where data often lives in very high-dimensional spaces.

Thus a vector may be regarded in three complementary ways:

1. As a point in space, described by its coordinates.

2. As a displacement or arrow, described by a direction and a length.

3. As an abstract element of a vector space, whose properties follow algebraic rules independent of geometry.

Notation

- Vectors are written in boldface lowercase letters: $\mathbf{v}, \mathbf{w}, \mathbf{x}$.

- The i-th entry of a vector $\mathbf{v}$ is written v_i, where indices begin at 1.

- The set of all n-dimensional vectors over $\mathbb{R}$ is denoted $\mathbb{R}^n$.

- Column vectors will be the default form unless otherwise stated.

Why begin here?

Scalars and vectors form the atoms of linear algebra. Every structure we will build-vector spaces, linear
transformations, matrices, eigenvalues-relies on the basic notions of number and ordered collection of numbers. Once
vectors are understood, we can define operations such as addition and scalar multiplication, then generalize to
subspaces, bases, and coordinate systems. Eventually, this framework grows into the full theory of linear algebra, with
powerful applications to geometry, computation, and data.

Exercises 1.1

1. Write three different vectors in $\mathbb{R}^2$ and sketch them as arrows from the origin. Identify their coordinates
   explicitly.

2. Give an example of a vector in $\mathbb{R}^4$. Can you visualize it directly? Explain why high-dimensional
   visualization is challenging.

3. Let $\mathbf{v} = (4, -3, 2)$. Write $\mathbf{v}$ in column form and state v_1, v_2, v_3.

4. In what sense is the set $\mathbb{R}^1$ both a line and a vector space? Illustrate with examples.

5. Consider the vector $\mathbf{u} = (1, 1, \ldots, 1) \in \mathbb{R}^n$. What is special about this vector when n is
large? What might it represent in applications?

1.2 Vector Addition and Scalar Multiplication

Vectors in linear algebra are not static objects; their power comes from the operations we can perform on them. Two
fundamental operations define the structure of vector spaces: addition and scalar multiplication. These operations
satisfy simple but far-reaching rules that underpin the entire subject.

Vector Addition

Given two vectors of the same dimension, their sum is obtained by adding corresponding entries. Formally, if

$$\mathbf{u} = (u_1, u_2, \ldots, u_n), \quad \mathbf{v} = (v_1, v_2, \ldots, v_n),$$

then their sum is

$$\mathbf{u} + \mathbf{v} = (u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n).$$

Example 1.2.1.
Let $\mathbf{u} = (2, -1, 3)$ and $\mathbf{v} = (4, 0, -5)$. Then

$$\mathbf{u} + \mathbf{v} = (2 + 4, -1 + 0, 3 + (-5)) = (6, -1, -2).$$

Geometrically, vector addition corresponds to the *parallelogram rule*. If we draw both vectors as arrows from the
origin, then placing the tail of one vector at the head of the other produces the sum. The diagonal of the parallelogram
they form represents the resulting vector.

Scalar Multiplication

Multiplying a vector by a scalar stretches or shrinks the vector while preserving its direction, unless the scalar is
negative, in which case the vector is also reversed. If $c \in \mathbb{R}$ and

$$\mathbf{v} = (v_1, v_2, \ldots, v_n),$$

then

$$cv = (cv_1, cv_2, \ldots, cv_n).$$

Example 1.2.2.
Let $\mathbf{v} = (3, -2)$ and $c = -2$. Then

$$c\mathbf{v} = -2(3, -2) = (-6, 4).$$

This corresponds to flipping the vector through the origin and doubling its length.

Linear Combinations

The interaction of addition and scalar multiplication allows us to form *linear combinations*.
A linear combination of
vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ is any vector of the form

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k, \quad c_i \in \mathbb{R}.$$

Linear combinations are the mechanism by which we generate new vectors from existing
ones. The span of a set of
vectors-the collection of all their linear combinations-will later lead us to the idea of a
subspace.

Example 1.2.3.
Let $\mathbf{v}_1 = (1, 0)$ and $\mathbf{v}_2 = (0, 1)$. Then any vector $(a, b) \in \mathbb{R}^2$ can be expressed as

$$a\mathbf{v}_1 + b\mathbf{v}_2.$$

Thus $(1, 0)$ and $(0, 1)$ form the basic building blocks of the plane.

Notation

- Addition: $\mathbf{u} + \mathbf{v}$ means component-wise addition.

- Scalar multiplication: $c\mathbf{v}$ scales each entry of $\mathbf{v}$ by c.

- Linear combination: a sum of the form $c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k$.

Why this matters

Vector addition and scalar multiplication are the defining operations of linear algebra.
They give structure to vector
spaces, allow us to describe geometric phenomena like translation and scaling, and provide
the foundation for solving
systems of equations. Everything that follows-basis, dimension, transformations-builds on
these simple but profound
rules.

Exercises 1.2

1. Compute $\mathbf{u} + \mathbf{v}$ where $\mathbf{u} = (1, 2, 3)$ and $\mathbf{v} = (4, -1, 0)$.

2. Find $3\mathbf{v}$ where $\mathbf{v} = (-2,5)$. Sketch both vectors to illustrate the scaling.

3. Show that $(5, 7)$ can be written as a linear combination of $(1, 0)$ and $(0, 1)$.

4. Write $(4, 4)$ as a linear combination of $(1, 1)$ and $(1, -1)$.

5. Prove that if $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$,
   then $(c + d)(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v} + d\mathbf{u} + d\mathbf{v}$ for
   scalars $c, d \in \mathbb{R}$.

1.3 Dot Product, Norms, and Angles

The dot product is the fundamental operation that links algebra and geometry in vector spaces. It allows us to measure
lengths, compute angles, and determine orthogonality. From this single definition flow the notions of *norm* and
angle, which give geometry to abstract vector spaces.

The Dot Product

For two vectors in $\mathbb{R}^n$, the dot product (also called the inner product) is defined by

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

Equivalently, in matrix notation:

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v}.$$

Example 1.3.1.
Let $\mathbf{u} = (2, -1, 3)$ and $\mathbf{v} = (4, 0, -2)$. Then

$$\mathbf{u} \cdot \mathbf{v} = 2 \cdot 4 + (-1) \cdot 0 + 3 \cdot (-2) = 8 - 6 = 2.$$

The dot product outputs a single scalar, not another vector.

Norms (Length of a Vector)

The *Euclidean norm* of a vector is the square root of its dot product with itself:

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}.$$

This generalizes the Pythagorean theorem to arbitrary dimensions.

Example 1.3.2.
For $\mathbf{v} = (3, 4)$,

$$\|\mathbf{v}\| = \sqrt{3^2 + 4^2} = \sqrt{25} = 5.$$

This is exactly the length of the vector as an arrow in the plane.

Angles Between Vectors

The dot product also encodes the angle between two vectors. For nonzero vectors $\mathbf{u}, \mathbf{v}$,

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \, \|\mathbf{v}\| \cos\theta,$$

where θ is the angle between them. Thus,

$$\cos\theta = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}.$$

Example 1.3.3.
Let $\mathbf{u} = (1, 0)$ and $\mathbf{v} = (0, 1)$. Then

$$\mathbf{u} \cdot \mathbf{v} = 0, \quad \|\mathbf{u}\| = 1, \quad \|\mathbf{v}\| = 1.$$

Hence

$$\cos\theta = \frac{0}{1 \cdot 1} = 0 \quad \Rightarrow \quad \theta = \frac{\pi}{2}.$$

The vectors are perpendicular.

Orthogonality

Two vectors are said to be orthogonal if their dot product is zero:

$$\mathbf{u} \cdot \mathbf{v} = 0.$$

Orthogonality generalizes the idea of perpendicularity from geometry to higher dimensions.

Notation

- Dot product: $\mathbf{u} \cdot \mathbf{v}$.
- Norm (length): $|\mathbf{v}|$.
- Orthogonality: $\mathbf{u} \perp \mathbf{v}$ if $\mathbf{u} \cdot \mathbf{v} = 0$.

Why this matters

The dot product turns vector spaces into geometric objects: vectors gain lengths, angles, and notions of
perpendicularity. This foundation will later support the study of orthogonal projections, Gram–Schmidt
orthogonalization, eigenvectors, and least squares problems.

Exercises 1.3

1. Compute $\mathbf{u} \cdot \mathbf{v}$ for $\mathbf{u} = (1, 2, 3)$, $\mathbf{v} = (4, 5, 6)$.

2. Find the norm of $\mathbf{v} = (2, -2, 1)$.

3. Determine whether $\mathbf{u} = (1, 1, 0)$ and $\mathbf{v} = (1, -1, 2)$ are orthogonal.

4. Let $\mathbf{u} = (3, 4)$, $\mathbf{v} = (4, 3)$. Compute the angle between them.

5. Prove that $|\mathbf{u} + \mathbf{v}|^2 = |\mathbf{u}|^2 + |\mathbf{v}|^2 + 2\mathbf{u} \cdot \mathbf{v}$. This
identity is the algebraic version of the Law of Cosines.

1.4 Orthogonality

Orthogonality captures the notion of perpendicularity in vector spaces. It is one of the most important geometric ideas
in linear algebra, allowing us to decompose vectors, define projections, and construct special bases with elegant
properties.

Definition

Two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are said to be orthogonal if their dot product is zero:

$$\mathbf{u} \cdot \mathbf{v} = 0.$$

This condition ensures that the angle between them is $\pi/2$ radians (90 degrees).

Example 1.4.1.
In $\mathbb{R}^2$, the vectors $(1, 2)$ and $(2, -1)$ are orthogonal since

$$(1, 2) \cdot (2, -1) = 1 \cdot 2 + 2 \cdot (-1) = 0.$$

Orthogonal Sets

A collection of vectors is called orthogonal if every distinct pair of vectors in the set is orthogonal. If, in
addition, each vector has norm 1, the set is called orthonormal.

Example 1.4.2.
In $\mathbb{R}^3$, the standard basis vectors

$$\mathbf{e}_1 = (1, 0, 0), \quad \mathbf{e}_2 = (0, 1, 0), \quad \mathbf{e}_3 = (0, 0, 1)$$

form an orthonormal set: each has length 1, and their dot products vanish when the indices differ.

Projections

Orthogonality makes possible the decomposition of a vector into two components: one parallel to another vector, and one
orthogonal to it. Given a nonzero vector $\mathbf{u}$ and any vector $\mathbf{v}$, the projection of $\mathbf{v}$ onto $\mathbf{u}$ is

$$\text{proj}_{\mathbf{u}}(\mathbf{v}) = \frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{u} \cdot \mathbf{u}} \mathbf{u}.$$

The difference

$$\mathbf{v} - \text{proj}_{\mathbf{u}}(\mathbf{v})$$

is orthogonal to $\mathbf{u}$. Thus every vector can be decomposed uniquely into a parallel and perpendicular part with
respect to another vector.

Example 1.4.3.
Let $\mathbf{u} = (1, 0)$, $\mathbf{v} = (2, 3)$. Then

$$\text{proj}_{\mathbf{u}}(\mathbf{v}) = \frac{(1, 0) \cdot (2, 3)}{(1, 0) \cdot (1, 0)}(1, 0) = \frac{2}{1}(1, 0) = (2, 0).$$

Thus

$$\mathbf{v} = (2, 3) = (2, 0) + (0, 3),$$

where $(2, 0)$ is parallel to $(1, 0)$ and $(0, 3)$ is orthogonal to it.

Orthogonal Decomposition

In general, if $\mathbf{u} \neq \mathbf{0}$ and $\mathbf{v} \in \mathbb{R}^n$, then

$$\mathbf{v} = \text{proj\_}\mathbf{u}(\mathbf{v}) + \left(\mathbf{v} - \text{proj\_}\mathbf{u}(\mathbf{v})\right),$$

where the first term is parallel to $\mathbf{u}$ and the second term is orthogonal. This decomposition underlies methods
such as least squares approximation and the Gram–Schmidt process.

Notation

- **u** ⊥ **v**: vectors **u** and **v** are orthogonal.

- An orthogonal set: vectors pairwise orthogonal.

- An orthonormal set: pairwise orthogonal, each of norm 1.

Why this matters

Orthogonality gives structure to vector spaces. It provides a way to separate independent directions cleanly, simplify
computations, and minimize errors in approximations. Many powerful algorithms in numerical linear algebra and data
science (QR decomposition, least squares regression, PCA) rely on orthogonality.

Exercises 1.4

1. Verify that the vectors $(1, 2, 2)$ and $(2, 0, -1)$ are orthogonal.

2. Find the projection of $(3, 4)$ onto $(1, 1)$.

3. Show that any two distinct standard basis vectors in $\mathbb{R}^n$ are orthogonal.

4. Decompose $(5, 2)$ into components parallel and orthogonal to $(2, 1)$.

5. Let **u**, **v** be orthogonal nonzero vectors.
   (a) Show that $(\mathbf{u} + \mathbf{v}) \cdot (\mathbf{u} - \mathbf{v}) = \|\mathbf{u}\|^2 - \|\mathbf{v}\|^2$.
   (b) For what condition on **u** and **v** does $(\mathbf{u} + \mathbf{v}) \cdot (\mathbf{u} - \mathbf{v}) = 0$?

Chapter 2. Matrices

2.1 Definition and Notation

Matrices are the central objects of linear algebra, providing a compact way to represent and manipulate linear
transformations, systems of equations, and structured data. A matrix is a rectangular array of numbers arranged in rows
and columns.

Formal Definition

An $m \times n$ matrix is an array with m rows and n columns, written

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

Each entry a_{ij} is a scalar, located in the i-th row and j-th column. The size (or dimension) of the matrix is
denoted by $m \times n$.

- If $m = n$, the matrix is square.

- If $m = 1$, the matrix is a row vector.

- If $n = 1$, the matrix is a column vector.

Thus, vectors are simply special cases of matrices.

Examples

Example 2.1.1. A 2×3 matrix:

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

Here, $a_{12} = -2$, $a_{23} = 5$, and the matrix has 2 rows, 3 columns.

Example 2.1.2. A 3×3 square matrix:

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

This will later serve as the representation of a linear transformation on $\mathbb{R}^3$.

Indexing and Notation

- Matrices are denoted by uppercase bold letters: A, B, C.

- Entries are written as a_{ij}, with the row index first, column index second.

- The set of all real $m \times n$ matrices is denoted $\mathbb{R}^{m \times n}$.

Thus, a matrix is a function $A : 1, \ldots, m \times 1, \ldots, n \to \mathbb{R}$, assigning a scalar to each row-column
position.

Why this matters

Matrices generalize vectors and give us a language for describing linear operations systematically. They encode systems
of equations, rotations, projections, and transformations of data. With matrices, algebra and geometry come together: a
single compact object can represent both numerical data and functional rules.

Exercises 2.1

1. Write a $3 \times 2 matrix of your choice and identify its entries a\_{ij}$.

2. Is every vector a matrix? Is every matrix a vector? Explain.

3. Which of the following are square
matrices: $A \in \mathbb{R}^{4 \times 4}$, $B \in \mathbb{R}^{3 \times 5}$, $C \in \mathbb{R}^{1 \times 1}$?

4. Let

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

What kind of matrix is this?

1. Consider the matrix

$$E = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

Express $e_{11}, e_{12}, e_{21}, e_{22}$ explicitly.

2.2 Matrix Addition and Multiplication

Once matrices are defined, the next step is to understand how they combine. Just as vectors gain meaning through
addition and scalar multiplication, matrices become powerful through two operations: addition and multiplication.

Matrix Addition

Two matrices of the same size are added by adding corresponding entries. If

$$A = [a_{ij}] \in \mathbb{R}^{m \times n}, \quad B = [b_{ij}] \in \mathbb{R}^{m \times n},$$

then

$$A + B = [a_{ij} + b_{ij}] \in \mathbb{R}^{m \times n}.$$

Example 2.2.1.
Let

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

Then

$$A + B = \begin{bmatrix} 1 + (-1) & 2 + 0 \\ 3 + 5 & 4 + 2 \end{bmatrix} = \begin{bmatrix} 0 & 2 \\ 8 & 6 \end{bmatrix}.$$

Matrix addition is commutative $(A+B = B+A)$ and associative $((A+B)+C = A+(B+C))$. The zero matrix, with all entries 0, acts as the additive identity.

Scalar Multiplication

For a scalar $c \in \mathbb{R}$ and a matrix $A = [[a_{ij}]$, we define

$$cA = [c \cdot a_{ij}].$$

This stretches or shrinks all entries of the matrix uniformly.

Example 2.2.2.
If

$$A = \begin{bmatrix} 2 & -1 \\ 0 & 3 \end{bmatrix}, \quad c = -2,$$

then

$$cA = \begin{bmatrix} -4 & 2 \\ 0 & -6 \end{bmatrix}.$$

Matrix Multiplication

The defining operation of matrices is multiplication. If

$$A \in \mathbb{R}^{m \times n}, \quad B \in \mathbb{R}^{n \times p},$$

then their product is the $m \times p$ matrix

$$AB = C = [c_{ij}], \quad c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

Thus, the entry in the i-th row and j-th column of AB is the dot product of the i-th row of A with the j-th
column of B.

Example 2.2.3.
Let

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

Then

$$AB = \begin{bmatrix} 1 \cdot 4 + 2 \cdot 2 & 1 \cdot (-1) + 2 \cdot 5 \\ 0 \cdot 4 + 3 \cdot 2 & 0 \cdot (-1) + 3 \cdot 5 \end{bmatrix} = \begin{bmatrix} 8 & 9 \\ 6 & 15 \end{bmatrix}.$$

Notice that matrix multiplication is not commutative in general: $AB \neq BA$. Sometimes BA may not even be defined if
dimensions do not align.

Geometric Meaning

Matrix multiplication corresponds to the composition of linear transformations. If A transforms vectors
in $\mathbb{R}^n$ and B transforms vectors in $\mathbb{R}^p$, then AB represents applying B first, then A. This
makes matrices the algebraic language of transformations.

Notation

- Matrix sum: $A + B$.

- Scalar multiple: cA.

- Product: AB, defined only when the number of columns of A equals the number of rows of B.

Why this matters

Matrix multiplication is the core mechanism of linear algebra: it encodes how transformations combine, how systems of
equations are solved, and how data flows in modern algorithms. Addition and scalar multiplication make matrices into a
vector space, while multiplication gives them an algebraic structure rich enough to model geometry, computation, and
networks.

Exercises 2.2

1. Compute $A + B$ for

$$A = \begin{bmatrix} 2 & 3 \\ -1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 4 & -2 \\ 5 & 7 \end{bmatrix}.$$

1. Find $3A$ where

$$A = \begin{bmatrix} 1 & -4 \\ 2 & 6 \end{bmatrix}.$$

1. Multiply

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

1. Verify with an explicit example that $AB \neq BA$.

2. Prove that matrix multiplication is distributive: $A(B + C) = AB + AC$.

2.3 Transpose and Inverse

Two special operations on matrices-the transpose and the inverse-give rise to deep algebraic and geometric properties.
The transpose rearranges a matrix by flipping it across its main diagonal, while the inverse, when it exists, acts as
the undo operation for matrix multiplication.

The Transpose

The transpose of an $m \times n$ matrix $A = [a_{ij}]$ is the $n \times m$ matrix $A^T = [a_{ji}]$, obtained by swapping
rows and columns.

Formally,

$$(A^T)\_ij = a\_ji.$$

Example 2.3.1.
If

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

then

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

Properties of the Transpose.

1. $(A^T)^T = A$.

2. $(A + B)^T = A^T + B^T$.

3. $(cA)^T = cA^T$, for scalar c.

4. $(AB)^T = B^T A^T$.

The last rule is crucial: the order reverses.

The Inverse

A square matrix $A \in \mathbb{R}^{n \times n}$ is said to be invertible (or nonsingular) if there exists another matrix A^{-1} such that

$$AA^{-1} = A^{-1}A = I_n,$$

where I_n is the $n \times n$ identity matrix. In this case, A^{-1} is called the inverse of A.

Not every matrix is invertible. A necessary condition is that $\det(A) \neq 0$, a fact that will be developed in Chapter 6.

Example 2.3.2.
Let

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

Its determinant is $\det(A) = (1)(4) - (2)(3) = -2 \neq 0$. The inverse is

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ 1.5 & -0.5 \end{bmatrix}.$$

Verification:

$$AA^{-1} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} -2 & 1 \\ 1.5 & -0.5 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Geometric Meaning

- The transpose corresponds to reflecting a linear transformation across the diagonal. For vectors, it switches between row and column forms.

- The inverse, when it exists, corresponds to reversing a linear transformation. For example, if A scales and rotates vectors, A^{-1} rescales and rotates them back.

Notation

- Transpose: A^T.
- Inverse: A^{-1}, defined only for invertible square matrices.
- Identity: I_n, acts as the multiplicative identity.

Why this matters

The transpose allows us to define symmetric and orthogonal matrices, central to geometry and numerical methods. The inverse underlies the solution of linear systems, encoding the idea of undoing a transformation. Together, these operations set the stage for determinants, eigenvalues, and orthogonalization.

Exercises 2.3

1. Compute the transpose of

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

1. Verify that $(AB)^T = B^T A^T$ for

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix}.$$

1. Determine whether

$$C = \begin{bmatrix} 2 & 1 \\ 4 & 2 \end{bmatrix}$$

is invertible. If so, find C^{-1}.

1. Find the inverse of

$$D = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix},$$

and explain its geometric action on vectors in the plane.

1. Prove that if A is invertible, then so is A^T, and $(A^T)^{-1} = (A^{-1})^T$.

2.4 Special Matrices

Certain matrices occur so frequently in theory and applications that they are given special names. Recognizing their
properties allows us to simplify computations and understand the structure of linear transformations more clearly.

The Identity Matrix

The identity matrix I_n is the $n \times n$ matrix with ones on the diagonal and zeros elsewhere:

$$I_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

It acts as the multiplicative identity:

$$AI_n = I_nA = A, \quad \text{for all } A \in \mathbb{R}^{n \times n}.$$

Geometrically, I_n represents the transformation that leaves every vector unchanged.

Diagonal Matrices

A diagonal matrix has all off-diagonal entries zero:

$$D = \begin{bmatrix} d_{11} & 0 & \cdots & 0 \\ 0 & d_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_{nn} \end{bmatrix}.$$

Multiplication by a diagonal matrix scales each coordinate independently:

$$D\mathbf{x} = (d_{11}x_1, d_{22}x_2, \ldots, d_{nn}x_n).$$

20

Example 2.4.1.
Let

$$D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 1 \\ 4 \\ -2 \end{bmatrix}.$$

Then

$$D\mathbf{x} = \begin{bmatrix} 2 \\ 12 \\ 2 \end{bmatrix}.$$

Permutation Matrices

A permutation matrix is obtained by permuting the rows of the identity matrix. Multiplying a vector by a permutation
matrix reorders its coordinates.

Example 2.4.2.
Let

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

Then

$$P \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} b \\ a \\ c \end{bmatrix}.$$

Thus, P swaps the first two coordinates.

Permutation matrices are always invertible; their inverses are simply their transposes.

Symmetric and Skew-Symmetric Matrices

A matrix is symmetric if

$$A^T = A,$$

and skew-symmetric if
Symmetric matrices appear in quadratic forms and optimization, while skew-symmetric matrices describe rotations and
cross products in geometry.

21

Orthogonal Matrices

A square matrix Q is orthogonal if

$$Q^T Q = Q Q^T = I.$$

Equivalently, the rows (and columns) of Q form an orthonormal set. Orthogonal matrices preserve lengths and angles;
they represent rotations and reflections.

Example 2.4.3.
The rotation matrix in the plane:

$$R(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

is orthogonal, since

$$R(\theta)^T R(\theta) = I_2.$$

Why this matters

Special matrices serve as the building blocks of linear algebra. Identity matrices define the neutral element, diagonal
matrices simplify computations, permutation matrices reorder data, symmetric and orthogonal matrices describe
fundamental geometric structures. Much of modern applied mathematics reduces complex problems to operations involving
these simple forms.

Exercises 2.4

1. Show that the product of two diagonal matrices is diagonal, and compute an example.

2. Find the permutation matrix that cycles (a, b, c) into (b, c, a).

3. Prove that every permutation matrix is invertible and its inverse is its transpose.

4. Verify that

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

is orthogonal. What geometric transformation does it represent?
5. Determine whether

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

are symmetric, skew-symmetric, or neither.

Chapter 3. Systems of Linear Equations

3.1 Linear Systems and Solutions

One of the central motivations for linear algebra is solving systems of linear equations. These systems arise naturally
in science, engineering, and data analysis whenever multiple constraints interact. Matrices provide a compact language
for expressing and solving them.

Linear Systems

A linear system consists of equations where each unknown appears only to the first power and with no products between
variables. A general system of m equations in n unknowns can be written as:

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1,$$
$$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2,$$
$$\vdots$$
$$a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m.$$

Here the coefficients a_{ij} and constants b_i are scalars, and the unknowns are $x_1, x_2, \ldots, x_n$.

Matrix Form

The system can be expressed compactly as:

$$A\mathbf{x} = \mathbf{b},$$

where

- $A \in \mathbb{R}^{m \times n}$ is the coefficient matrix $[a_{ij}]$,
- $\mathbf{x} \in \mathbb{R}^n$ is the column vector of unknowns,
- $\mathbf{b} \in \mathbb{R}^m$ is the column vector of constants.

This formulation turns the problem of solving equations into analyzing the action of a matrix.

Example 3.1.1.
The system

$$\begin{cases} x + 2y = 5, \\ 3x - y = 4 \end{cases}$$

can be written as

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

Types of Solutions

A linear system may have:

1. No solution (inconsistent): The equations conflict. Example:

$$\begin{cases} x + y = 1 \\ x + y = 2 \end{cases}$$

This system has no solution.

1. Exactly one solution (unique): The system's equations intersect at a single point. Example: The following coefficient matrix:

$$\begin{bmatrix} 1 & 2 \\ 3 & -1 \end{bmatrix}$$

has a unique solution.

1. Infinitely many solutions: The equations describe overlapping constraints (e.g., multiple equations representing the same line or plane).

The nature of the solution depends on the rank of A and its relation to the augmented matrix $(A|\mathbf{b})$, which we will study later.

Geometric Interpretation

- In $\mathbb{R}^2$, each linear equation represents a line. Solving a system means finding intersection points of
  lines.

- In $\mathbb{R}^3$, each equation represents a plane. A system may have no solution (parallel planes), one solution (a
  unique intersection point), or infinitely many (a line of intersection).

- In higher dimensions, the picture generalizes: solutions form intersections of hyperplanes.

Why this matters

Linear systems are the practical foundation of linear algebra. They appear in balancing chemical reactions, circuit
analysis, least-squares regression, optimization, and computer graphics. Understanding how to represent and classify
their solutions is the first step toward systematic solution methods like Gaussian elimination.

Exercises 3.1

1. Write the following system in matrix form:

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

1. Determine whether the system

$$\begin{cases} x + y = 1, \\ 2x + 2y = 2 \end{cases}$$

has no solution, one solution, or infinitely many solutions.

1. Geometrically interpret the system

$$\begin{cases} x + y = 3, \\ x - y = 1 \end{cases}$$

in the plane.

1. Solve the system

$$\begin{cases} 2x + y = 1, \\ x - y = 4 \end{cases}$$

and check your solution.

1. In $\mathbb{R}^3$, describe the solution set of

$$\begin{cases} x + y + z = 0, \\ 2x + 2y + 2z = 0 \end{cases}$$

What geometric object does it represent?

3.2 Gaussian Elimination

To solve linear systems efficiently, we use Gaussian elimination: a systematic method of transforming a system into a
simpler equivalent one whose solutions are easier to see. The method relies on elementary row operations that preserve
the solution set.

Elementary Row Operations

On an augmented matrix $(A|\mathbf{b})$, we are allowed three operations:

1. Row swapping: interchange two rows.

2. Row scaling: multiply a row by a nonzero scalar.

3. Row replacement: replace one row by itself plus a multiple of another row.

These operations correspond to re-expressing equations in different but equivalent forms.

Row Echelon Form

A matrix is in row echelon form (REF) if:

1. All nonzero rows are above any zero rows.

2. Each leading entry (the first nonzero number from the left in a row) is to the right of the leading entry in the row
   above.

3. All entries below a leading entry are zero.

Further, if each leading entry is 1 and is the only nonzero entry in its column, the matrix is in reduced row echelon
form (RREF).

Algorithm of Gaussian Elimination

1. Write the augmented matrix for the system.

2. Use row operations to create zeros below each pivot (the leading entry in a row).

3. Continue column by column until the matrix is in echelon form.

4. Solve by back substitution: starting from the last pivot equation and working upward.

If we continue to RREF, the solution can be read off directly.

Example

Example 3.2.1. Solve

$$\begin{cases} x + 2y - z = 3, \\ 2x + y + z = 7, \\ 3x - y + 2z = 4. \end{cases}$$

Step 1. Augmented matrix

$$\begin{bmatrix} 1 & 2 & -1 & 3 \\ 2 & 1 & 1 & 7 \\ 3 & -1 & 2 & 4 \end{bmatrix}.$$

Step 2. Eliminate below the first pivot

Subtract 2 times row 1 from row 2, and 3 times row 1 from row 3:

$$\begin{bmatrix} 1 & 2 & -1 & 3 \\ 0 & -3 & 3 & 1 \\ 0 & -7 & 5 & -5 \end{bmatrix}.$$

Step 3. Pivot in column 2

Divide row 2 by -3:

$$\left[\begin{array}{ccc|c} 1 & 2 & -1 & 3 \\ 0 & 1 & -1 & -\frac{1}{3} \\ 0 & -7 & 5 & -5 \end{array}\right].$$

Add 7 times row 2 to row 3:

$$\left[\begin{array}{ccc|c} 1 & 2 & -1 & 3 \\ 0 & 1 & -1 & -\frac{1}{3} \\ 0 & 0 & -2 & -\frac{22}{3} \end{array}\right].$$

Step 4. Pivot in column 3

Divide row 3 by -2:

$$\left[\begin{array}{ccc|c} 1 & 2 & -1 & 3 \\ 0 & 1 & -1 & -\frac{1}{3} \\ 0 & 0 & 1 & \frac{11}{3} \end{array}\right].$$

Step 5. Back substitution

From the last row:

$$z = \tfrac{11}{3}.$$

Second row:

$$y - z = -\tfrac{1}{3} \implies y = -\tfrac{1}{3} + \tfrac{11}{3} = \tfrac{10}{3}.$$

First row:

$$x + 2y - z = 3 \implies x + 2 \cdot \tfrac{10}{3} - \tfrac{11}{3} = 3.$$

So

$$x + \tfrac{20}{3} - \tfrac{11}{3} = 3 \implies x + 3 = 3 \implies x = 0.$$

Solution:

$$(x, y, z) = \left(0, \tfrac{10}{3}, \tfrac{11}{3}\right).$$

Why this matters

Gaussian elimination is the foundation of computational linear algebra. It reduces complex systems to a form where
solutions are visible, and it forms the basis for algorithms used in numerical analysis, scientific computing, and
machine learning.

Exercises 3.2

1. Solve by Gaussian elimination:

$$\begin{cases} x + y = 2, \\ 2x - y = 0. \end{cases}$$

1. Reduce the following augmented matrix to REF:

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 6 \\ 2 & -1 & 3 & 14 \\ 1 & 4 & -2 & -2 \end{array}\right].$$

1. Show that Gaussian elimination always produces either:

- a unique solution,

- infinitely many solutions, or

- a contradiction (no solution).

1. Use Gaussian elimination to find all solutions of

$$\begin{cases} x + y + z = 0, \\ 2x + y + z = 1. \end{cases}$$

1. Explain why pivoting (choosing the largest available pivot element) is useful in numerical computation.

3.3 Rank and Consistency

Gaussian elimination not only provides solutions but also reveals the structure of a linear system. Two key ideas are
the rank of a matrix and the consistency of a system. Rank measures the amount of independent information in the
equations, while consistency determines whether the system has at least one solution.

Rank of a Matrix

The rank of a matrix is the number of leading pivots in its row echelon form. Equivalently, it is the maximum number of
linearly independent rows or columns.

Formally,

$$\text{rank}(A) = \dim(\text{row space of } A) = \dim(\text{column space of } A).$$

The rank tells us the effective dimension of the space spanned by the rows (or columns).

Example 3.3.1.
For

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix},$$

row reduction gives

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

Thus, $\text{rank}(A) = 1$, since all rows are multiples of the first.

Consistency of Linear Systems

Consider the system $A\mathbf{x} = \mathbf{b}$.
The system is consistent (has at least one solution) if and only if

$$\text{rank}(A) = \text{rank}(A|\mathbf{b}),$$

where $(A|\mathbf{b})$ is the augmented matrix.
If the ranks differ, the system is inconsistent.

- If $\text{rank}(A) = \text{rank}(A|\mathbf{b}) = n$ (number of unknowns), the system has a unique solution.

- If $\text{rank}(A) = \text{rank}(A|\mathbf{b}) < n$, the system has infinitely many solutions.

Example

Example 3.3.2.
Consider

$$\begin{cases} x + y + z = 1, \\ 2x + 2y + 2z = 2, \\ x + y + z = 3. \end{cases}$$

The augmented matrix is

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 \\ 1 & 1 & 1 & 3 \end{array} \right].$$

Row reduction gives

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

Here, $\text{rank}(A) = 1$, but $\text{rank}(A|\mathbf{b}) = 2$. Since the ranks differ, the system is inconsistent: no
solution exists.

Example with Infinite Solutions

Example 3.3.3.
For
$$\begin{cases} x + y = 2, \\ 2x + 2y = 4, \end{cases}$$
the augmented matrix reduces to

$$\left[\begin{array}{cc|c} 1 & 1 & 2 \\ 0 & 0 & 0 \end{array}\right].$$

Here, $\text{rank}(A) = \text{rank}(A|\mathbf{b}) = 1 < 2$. Thus, infinitely many solutions exist, forming a line.

Why this matters

Rank is a measure of independence: it tells us how many truly distinct equations or directions are present. Consistency
explains when equations align versus when they contradict. These concepts connect linear systems to vector spaces and
prepare for the ideas of dimension, basis, and the Rank–Nullity Theorem.

Exercises 3.3

1. Compute the rank of

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

1. Determine whether the system
$$\begin{cases} x + y + z = 1, \\ 2x + 3y + z = 2, \\ 3x + 5y + 2z = 3 \end{cases}$$
is consistent.

31

1. Show that the rank of the identity matrix I_n is n.

2. Give an example of a system in $\mathbb{R}^3$ with infinitely many solutions, and explain why it satisfies the rank
condition.

3. Prove that for any matrix $A \in \mathbb{R}^{m \times n}$,
$\text{rank}(A) \leq \min(m, n)$.

3.4 Homogeneous Systems

A homogeneous system is a linear system in which all constant terms are zero:

$$A\mathbf{x} = \mathbf{0},$$

where $A \in \mathbb{R}^{m \times n}$, and $\mathbf{0}$ is the zero vector in $\mathbb{R}^m$.

The Trivial Solution

Every homogeneous system has at least one solution:

$$\mathbf{x} = \mathbf{0}.$$

This is called the trivial solution. The interesting question is whether *nontrivial solutions* (nonzero vectors) exist.

Existence of Nontrivial Solutions

Nontrivial solutions exist precisely when the number of unknowns exceeds the rank of the coefficient matrix:

$$\text{rank}(A) < n.$$

In this case, there are infinitely many solutions, forming a subspace of $\mathbb{R}^n$. The dimension of this solution
space is

$$\dim(\text{null}(A)) = n - \text{rank}(A),$$

where null(A) is the set of all solutions to $A\mathbf{x} = 0$. This set is called the null space or kernel of A.

Example

Example 3.4.1.
Consider
$$\begin{cases} x + y + z = 0, \\ 2x + y - z = 0. \end{cases}$$

The augmented matrix is

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

Row reduction:

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & -1 & -3 & 0 \end{array}\right] \quad \rightarrow \quad \left[\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 0 & 1 & 3 & 0 \end{array}\right].$$

So the system is equivalent to:
$$\begin{cases} x + y + z = 0, \\ y + 3z = 0. \end{cases}$$

From the second equation, $y = -3z$. Substituting into the first:
$x - 3z + z = 0 \implies x = 2z.$

Thus solutions are:

$$(x, y, z) = z(2, -3, 1), \quad z \in \mathbb{R}.$$

The null space is the line spanned by the vector $(2, -3, 1)$.

Geometric Interpretation

The solution set of a homogeneous system is always a subspace of $\mathbb{R}^n$.

- If $\text{rank}(A) = n$, the only solution is the zero vector.

- If $\text{rank}(A) = n - 1$, the solution set is a line through the origin.

- If $\text{rank}(A) = n - 2$, the solution set is a plane through the origin.

More generally, the null space has dimension $n - \text{rank}(A)$, known as the nullity.

Why this matters

Homogeneous systems are central to understanding vector spaces, subspaces, and dimension.
They lead directly to the
concepts of kernel, null space, and linear dependence. In applications, homogeneous
systems appear in equilibrium
problems, eigenvalue equations, and computer graphics transformations.

Exercises 3.4

1. Solve the homogeneous system

$$\begin{cases} x + 2y - z = 0, \\ 2x + 4y - 2z = 0. \end{cases}$$

What is the dimension of its solution space?

1. Find all solutions of

$$\begin{cases} x - y + z = 0, \\ 2x + y - z = 0. \end{cases}$$

1. Show that the solution set of any homogeneous system is a subspace of $\mathbb{R}^n$.

2. Suppose A is a $3 \times 3 matrix with \text{rank}(A) = 2. What is the dimension of the null space of A

3. For

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

compute a basis for the null space of A.

Chapter 4. Vector Spaces

4.1 Definition of a Vector Space

Up to now we have studied vectors and matrices concretely in $\mathbb{R}^n$. The next step is to move beyond coordinates
and define vector spaces in full generality. A vector space is an abstract setting where the familiar rules of addition
and scalar multiplication hold, regardless of whether the elements are geometric vectors, polynomials, functions, or
other objects.

Formal Definition

A vector space over the real numbers $\mathbb{R}$ is a set V equipped with two operations:

1. Vector addition: For any $\mathbf{u}, \mathbf{v} \in V$, there is a vector $\mathbf{u} + \mathbf{v} \in V$.

2. Scalar multiplication: For any scalar $c \in \mathbb{R}$ and any $\mathbf{v} \in V$, there is a vector $c\mathbf{v} \in V$.

These operations must satisfy the following axioms (for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and all scalars $a, b \in \mathbb{R}$):

1. Commutativity of addition: $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.

2. Associativity of addition: $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$.

3. Additive identity: There exists a zero vector $\mathbf{0} \in V$ such that $\mathbf{v} + \mathbf{0} = \mathbf{v}$.

4. Additive inverses: For each $\mathbf{v} \in V$, there exists $(-\mathbf{v} \in V$ such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$.

5. Compatibility of scalar multiplication: $a(b\mathbf{v}) = (ab)\mathbf{v}$.

6. Identity element of scalars: $1 \cdot \mathbf{v} = \mathbf{v}$.

7. Distributivity over vector addition: $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$.

8. Distributivity over scalar addition: $(a + b)\mathbf{v} = a\mathbf{v} + b\mathbf{v}$.

If a set V with operations satisfies all eight axioms, we call it a vector space.

Examples

Example 4.1.1. Standard Euclidean space
$\mathbb{R}^n$ with ordinary addition and scalar multiplication is a vector space. This is the model case from which the
axioms are abstracted.

Example 4.1.2. Polynomials
The set of all polynomials with real coefficients, denoted $\mathbb{R}[x]$, forms a vector space.

Addition and scalar
multiplication are defined term by term.

Example 4.1.3. Functions
The set of all real-valued functions on an interval, e.g. $f : [0, 1] \to \mathbb{R}$, forms a vector space, since
functions can be added and scaled pointwise.

Non-Examples

Not every set with operations qualifies. For instance, the set of positive real numbers under usual addition is not a
vector space, because additive inverses (negative numbers) are missing. The axioms must all hold.

Geometric Interpretation

In familiar cases like $\mathbb{R}^2$ or $\mathbb{R}^3$, vector spaces provide the stage for geometry: vectors can be
added, scaled, and combined to form lines, planes, and higher-dimensional structures. In abstract settings like function
spaces, the same algebraic rules let us apply geometric intuition to infinite-dimensional problems.

Why this matters

The concept of vector space unifies seemingly different mathematical objects under a single framework. Whether dealing
with forces in physics, signals in engineering, or data in machine learning, the common language of vector spaces allows
us to use the same techniques everywhere.

Exercises 4.1

1. Verify that $\mathbb{R}^2$ with standard addition and scalar multiplication satisfies all eight vector space axioms.

2. Show that the set of integers $\mathbb{Z}$ with ordinary operations is not a vector space over $\mathbb{R}$. Which
   axiom fails?

3. Consider the set of all polynomials of degree at most 3. Show it forms a vector space over $\mathbb{R}$. What is its
   dimension?

4. Give an example of a vector space where the vectors are not geometric objects.

5. Prove that in any vector space, the zero vector is unique.

4.2 Subspaces

A subspace is a smaller vector space living inside a larger one. Just as lines and planes naturally sit inside
three-dimensional space, subspaces generalize these ideas to higher dimensions and more abstract settings.

Definition

Let V be a vector space. A subset $W \subseteq V$ is called a subspace of V if:

1. $\mathbf{0} \in W$ (contains the zero vector),

2. For all $\mathbf{u}, \mathbf{v} \in W$, the sum $\mathbf{u} + \mathbf{v} \in W$ (closed under addition),

3. For all scalars $c \in \mathbb{R}$ and vectors $\mathbf{v} \in W$, the product $c\mathbf{v} \in W$ (closed under scalar multiplication).

If these hold, then W is itself a vector space with the inherited operations.

Examples

Example 4.2.1. Line through the origin in $\mathbb{R}^2$
The set

$$W = \{(t, 2t) \mid t \in \mathbb{R}\}$$

is a subspace of $\mathbb{R}^2$. It contains the zero vector, is closed under addition, and is closed under scalar
multiplication.

Example 4.2.2. The x–y plane in $\mathbb{R}^3$
The set

$$W = \{(x, y, 0) \mid x, y \in \mathbb{R}\}$$

is a subspace of $\mathbb{R}^3$. It is the collection of all vectors lying in the plane through the origin parallel to
the x–y plane.

Example 4.2.3. Null space of a matrix
For a matrix $A \in \mathbb{R}^{m \times n}$, the null space

$$\{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{0}\}$$

is a subspace of $\mathbb{R}^n$. This subspace represents all solutions to the homogeneous system.

Non-Examples

Not every subset is a subspace.

- The set $(x, y) \in \mathbb{R}^2 \mid x \geq 0$ is not a subspace: it is not closed under scalar multiplication (a
  negative scalar breaks the condition).

- Any line in $\mathbb{R}^2$ that does not pass through the origin is not a subspace, because it does not
  contain $\mathbf{0}$.

Geometric Interpretation

Subspaces are the linear structures inside vector spaces.

- In $\mathbb{R}^2$, the subspaces are: the zero vector, any line through the origin, or the entire plane.

- In $\mathbb{R}^3$, the subspaces are: the zero vector, any line through the origin, any plane through the origin, or
  the entire space.

- In higher dimensions, the same principle applies: subspaces are the flat linear pieces through the origin.

Why this matters

Subspaces capture the essential structure of linear problems. Column spaces, row spaces, and null spaces are all
subspaces. Much of linear algebra consists of understanding how these subspaces intersect, span, and complement each
other.

Exercises 4.2

1. Prove that the set $W = (x, 0) \mid x \in \mathbb{R} \subseteq \mathbb{R}^2$ is a subspace.

2. Show that the line $(1 + t, 2t) \mid t \in \mathbb{R}$ is not a subspace of $\mathbb{R}^2$. Which condition fails?

3. Determine whether the set of all vectors $(x, y, z) \in \mathbb{R}^3$ satisfying $x + y + z = 0$ is a subspace.

4. For the matrix

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

Describe the null space of A as a subspace of $\mathbb{R}^3$.

1. List all possible subspaces of $\mathbb{R}^2$.

4.3 Span, Basis, Dimension

The ideas of span, basis, and dimension provide the language for describing the size and structure of subspaces.
Together, they tell us how a vector space is generated, how many building blocks it requires, and how those blocks can be chosen.

Span

Given a set of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k \subseteq V$, the span is the collection of all linear combinations:

$$\text{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} = \{c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k \mid c_i \in \mathbb{R}\}.$$

The span is always a subspace of V, namely the smallest subspace containing those vectors.

Example 4.3.1.
In $\mathbb{R}^2$, $\text{span}(1, 0) = \{(x, 0) \mid x \in \mathbb{R}\}$, the x-axis.
Similarly, $\text{span}\{(1, 0), (0, 1)\} = \mathbb{R}^2$.

Basis

A basis of a vector space V is a set of vectors that:

1. Span V.

2. Are linearly independent (no vector in the set is a linear combination of the others).

If either condition fails, the set is not a basis.

Example 4.3.2.
In $\mathbb{R}^3$, the standard unit vectors

$$\mathbf{e}_1 = (1, 0, 0), \quad \mathbf{e}_2 = (0, 1, 0), \quad \mathbf{e}_3 = (0, 0, 1)$$

form a basis. Every vector (x, y, z) can be uniquely written as

$$x\mathbf{e}_1 + y\mathbf{e}_2 + z\mathbf{e}_3.$$

Dimension

The dimension of a vector space V, written $\dim(V)$, is the number of vectors in any basis of V. This number is
well-defined: all bases of a vector space have the same cardinality.

Examples 4.3.3.

- $\dim(\mathbb{R}^2) = 2$, with basis $(1, 0), (0, 1)$.

- $\dim(\mathbb{R}^3) = 3$, with basis $(1,0,0), (0,1,0), (0,0,1)$.
- The set of polynomials of degree at most 3 has dimension 4, with basis $(1, x, x^2, x^3)$.

Geometric Interpretation

- The span is like the reach of a set of vectors.
- A basis is the minimal set of directions needed to reach everything in the space.
- The dimension is the count of those independent directions.

Lines, planes, and higher-dimensional flats can all be described in terms of span, basis, and dimension.

Why this matters

These concepts classify vector spaces and subspaces in terms of size and structure. Many theorems in linear algebra-such
as the Rank–Nullity Theorem-are consequences of understanding span, basis, and dimension. In practical terms, bases are
how we encode data in coordinates, and dimension tells us how much freedom a system truly has.

Exercises 4.3

1. Show that $(1,0,0)$, $(0,1,0)$, $(1,1,0)$ span the xy-plane in $\mathbb{R}^3$. Are they a basis?
2. Find a basis for the line $\{(2t, -3t, t) : t \in \mathbb{R}\}$ in $\mathbb{R}^3$.
3. Determine the dimension of the subspace of $\mathbb{R}^3$ defined by $x + y + z = 0$.
4. Prove that any two different bases of $\mathbb{R}^n$ must contain exactly n vectors.
5. Give a basis for the set of polynomials of degree ≤ 2. What is its dimension?

4.4 Coordinates

Once a basis for a vector space is chosen, every vector can be expressed uniquely as a linear combination of the basis
vectors. The coefficients in this combination are called the coordinates of the vector relative to that basis.
Coordinates allow us to move between the abstract world of vector spaces and the concrete world of numbers.

Coordinates Relative to a Basis

Let V be a vector space, and let

$$\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$$

be an ordered basis for V. Every vector $\mathbf{u} \in V$ can be written uniquely as

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

The scalars $(c_1, c_2, \ldots, c_n)$ are the coordinates of $\mathbf{u}$ relative to $\mathcal{B}$, written

$$[\mathbf{u}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}.$$

Example in $\mathbb{R}^2$

Example 4.4.1.
Let the basis be

$$\mathcal{B} = \{(1, 1), (1, -1)\}.$$

To find the coordinates of $\mathbf{u} = (3, 1)$ relative to $\mathcal{B}$, solve

$$(3, 1) = c_1(1, 1) + c_2(1, -1).$$

This gives the system

$$\begin{cases} c_1 + c_2 = 3, \\ c_1 - c_2 = 1. \end{cases}$$

Adding: $2c\_1 = 4 \implies c\_1 = 2. Then c\_2 = 1$.

So,

$$[\mathbf{u}]_{\mathcal{B}} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

Standard Coordinates

In $\mathbb{R}^n$, the standard basis is

$$\mathbf{e}_1 = (1, 0, \ldots, 0), \quad \mathbf{e}_2 = (0, 1, 0, \ldots, 0), \ldots, \mathbf{e}_n = (0, \ldots, 0, 1).$$

Relative to this basis, the coordinates of a vector are simply its entries. Thus, column vectors are coordinate
representations by default.

Change of Basis

If $\mathcal{B} = \mathbf{v}_1, \ldots, \mathbf{v}_n$ is a basis of $\mathbb{R}^n$, the change of basis matrix is

$$P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix},$$

with basis vectors as columns. For any vector $\mathbf{u}$,

$$\mathbf{u} = P[\mathbf{u}]_\mathcal{B}, \qquad [\mathbf{u}]_\mathcal{B} = P^{-1}\mathbf{u}.$$

Thus, switching between bases reduces to matrix multiplication.

Geometric Interpretation

Coordinates are the address of a vector relative to a chosen set of directions. Different bases are like different
coordinate systems: Cartesian, rotated, skewed, or scaled. The same vector may look very different numerically depending
on the basis, but its geometric identity is unchanged.

Why this matters

Coordinates turn abstract vectors into concrete numerical data. Changing basis is the algebraic language for rotations
of axes, diagonalization of matrices, and principal component analysis in data science. Mastery of coordinates is
essential for moving fluidly between geometry, algebra, and computation.

Exercises 4.4

1. Express $(4, 2)$ in terms of the basis $(1, 1), (1, -1)$.

2. Find the coordinates of $(1, 2, 3)$ relative to the standard basis of $\mathbb{R}^3$.

3. If $\mathcal{B} = \{(2, 0), (0, 3)\}$, compute $[(4, 6)]_\mathcal{B}$.

4. Construct the change of basis matrix from the standard basis of $\mathbb{R}^2$ to $\mathcal{B} = \{(1, 1), (1, -1)\}$.

5. Prove that coordinate representation with respect to a basis is unique.

Chapter 5. Linear Transformations

5.1 Functions that Preserve Linearity

A central theme of linear algebra is understanding linear transformations: functions between vector spaces that preserve
their algebraic structure. These transformations generalize the idea of matrix multiplication and capture the essence of
linear behavior.

Definition

Let V and W be vector spaces over $\mathbb{R}$. A function

$$T : V \to W$$

is called a linear transformation (or linear map) if for all vectors $\mathbf{u}, \mathbf{v} \in V$ and all scalars $c \in \mathbb{R}$:

1. Additivity:

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

1. Homogeneity:

$$T(c\mathbf{u}) = cT(\mathbf{u}).$$

If both conditions hold, then T automatically respects linear combinations:

$$T(c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k) = c_1 T(\mathbf{v}_1) + \cdots + c_k T(\mathbf{v}_k).$$

Examples

Example 5.1.1. Scaling in $\mathbb{R}^2$.
Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$T(x, y) = (2x, 2y).$$

This doubles the length of every vector, preserving direction. It is linear.

Example 5.1.2. Rotation.

Let $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ be

$$R_\theta(x, y) = (x \cos \theta - y \sin \theta, \ x \sin \theta + y \cos \theta).$$

This rotates vectors by angle θ. It satisfies additivity and homogeneity, hence is linear.

Example 5.1.3. Differentiation.

Let $D : \mathbb{R}[x] \to \mathbb{R}[x]$ be differentiation: $D(p(x)) = p'(x)$.

Since derivatives respect addition and scalar multiples, differentiation is a linear transformation.

Non-Example

The map $S : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$S(x, y) = (x^2, y^2)$$

is not linear, because $S(\mathbf{u} + \mathbf{v}) \neq S(\mathbf{u}) + S(\mathbf{v})$ in general.

Geometric Interpretation

Linear transformations are exactly those that preserve the origin, lines through the origin, and proportions along those
lines. They include familiar operations: scaling, rotations, reflections, shears, and projections. Nonlinear
transformations bend or curve space, breaking these properties.

Why this matters

Linear transformations unify geometry, algebra, and computation. They explain how matrices act on vectors, how data can
be rotated or projected, and how systems evolve under linear rules. Much of linear algebra is devoted to understanding
these transformations, their representations, and their invariants.

Exercises 5.1

1. Verify that $T(x, y) = (3x - y, 2y)$ is a linear transformation on $\mathbb{R}^2$.

2. Show that $T(x, y) = (x + 1, y)$ is not linear. Which axiom fails?

3. Prove that if T and S are linear transformations, then so is $T + S$.

4. Give an example of a linear transformation from $\mathbb{R}^3$ to $\mathbb{R}^2$.

5. Let $T : \mathbb{R}[x] \to \mathbb{R}[x]$ be integration:

$$T(p(x)) = \int_0^x p(t), dt.$$

Prove that T is a linear transformation.

5.2 Matrix Representation of Linear Maps

Every linear transformation between finite-dimensional vector spaces can be represented
by a matrix. This correspondence
is one of the central insights of linear algebra: it lets us use the tools of matrix arithmetic
to study abstract
transformations.

From Linear Map to Matrix

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Choose the standard
basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ of $\mathbb{R}^n$, where $\mathbf{e}_i$ has a 1 in the i-th position
and 0 elsewhere.

The action of T on each basis vector determines the entire transformation:

$$T(\mathbf{e}\_j) = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}.$$

Placing these outputs as columns gives the matrix of T:

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

Then for any vector $\mathbf{x} \in \mathbb{R}^n$:

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

Examples

Example 5.2.1. Scaling in $\mathbb{R}^2$.
Let $T(x, y) = (2x, 3y)$. Then

$$T(\mathbf{e}_1) = (2, 0), \quad T(\mathbf{e}_2) = (0, 3).$$

So the matrix is

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

Example 5.2.2. Rotation in the plane.
The rotation transformation $R_\theta(x, y) = (x \cos \theta - y \sin \theta,\ x \sin \theta + y \cos \theta)$ has matrix

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

Example 5.2.3. Projection onto the x-axis.
The map $P(x, y) = (x, 0)$ corresponds to

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

Change of Basis

Matrix representations depend on the chosen basis. If $\mathcal{B}$ and $\mathcal{C}$ are bases of $\mathbb{R}^n$
and $\mathbb{R}^m$, then the matrix of $T : \mathbb{R}^n \to \mathbb{R}^m$ with respect to these bases is obtained by
expressing $T(\mathbf{v}_j)$ in terms of $\mathcal{C}$ for each $\mathbf{v}_j \in \mathcal{B}$. Changing bases
corresponds to conjugating the matrix by the appropriate change-of-basis matrices.

Geometric Interpretation

Matrices are not just convenient notation-they *are* linear maps once a basis is fixed. Every rotation, reflection,
projection, shear, or scaling corresponds to multiplying by a specific matrix. Thus, studying linear transformations
reduces to studying their matrices.

Why this matters

Matrix representations make linear transformations computable. They connect abstract definitions to explicit
calculations, enabling algorithms for solving systems, finding eigenvalues, and performing decompositions. Applications
from graphics to machine learning depend on this translation.

Exercises 5.2

1. Find the matrix representation of $T : \mathbb{R}^2 \to \mathbb{R}^2$, $T(x, y) = (x + y, x - y)$.

2. Determine the matrix of the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^2$, $T(x, y, z) = (x + z, y - 2z)$.

3. What matrix represents reflection across the line $y = x$ in $\mathbb{R}^2$?

4. Show that the matrix of the identity transformation on $\mathbb{R}^n$ is I_n.

5. For the differentiation map $D : \mathbb{R}_2[x] \to \mathbb{R}_1[x]$, where $\mathbb{R}_k[x]$ is the space of polynomials of degree at most k, find the matrix of D relative to the bases $\{1, x, x^2\}$ and $\{1, x\}$.

5.3 Kernel and Image

To understand a linear transformation deeply, we must examine what it kills and what it produces. These ideas are
captured by the kernel and the image, two fundamental subspaces associated with any linear map.

The Kernel

The kernel (or null space) of a linear transformation $T : V \to W$ is the set of all vectors in V that map to the zero
vector in W:

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

The kernel is always a subspace of V. It measures the degeneracy of the transformation- directions that collapse to
nothing.

Example 5.3.1.
Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be defined by

$$T(x, y, z) = (x + y, y + z).$$

In matrix form,

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

To find the kernel, solve

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

This gives the equations $x + y = 0$, $y + z = 0$. Hence $x = -y, z = -y$. The kernel is

$$\ker(T) = \{(-t, t, -t) \mid t \in \mathbb{R}\},$$

a line in $\mathbb{R}^3$.

The Image

The image (or range) of a linear transformation $T : V \to W$ is the set of all outputs:

$$\operatorname{im}(T) = \{T(\mathbf{v}) \mid \mathbf{v} \in V\} \subseteq W.$$

Equivalently, it is the span of the columns of the representing matrix. The image is always a subspace of W.

Example 5.3.2.
For the same transformation as above,

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

the columns are $(1,0)$, $(1,1)$, and $(0,1)$. Since $(1,1) = (1,0) + (0,1)$, the image is

$$\mathrm{im}(T) = \mathrm{span}\{(1,0),(0,1)\} = \mathbb{R}^2.$$

Dimension Formula (Rank–Nullity Theorem)

For a linear transformation $T : V \to W$ with V finite-dimensional,

$$\dim(\ker(T)) + \dim(\mathrm{im}(T)) = \dim(V).$$

This fundamental result connects the lost directions (kernel) with the achieved directions (image).

Geometric Interpretation

- The kernel describes how the transformation flattens space (e.g., projecting a 3D object onto a plane).

- The image describes the target subspace reached by the transformation.

- The rank–nullity theorem quantifies the tradeoff: the more dimensions collapse, the fewer remain in the image.

Why this matters

Kernel and image capture the essence of a linear map. They classify transformations, explain when systems have unique or
infinite solutions, and form the backbone of important results like the Rank–Nullity Theorem, diagonalization, and
spectral theory.

Exercises 5.3

1. Find the kernel and image of $T : \mathbb{R}^2 \to \mathbb{R}^2$, $T(x,y) = (x - y, x + y)$.

2. Let

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

Find bases for $\ker(A)$ and $\operatorname{im}(A)$.

1. For the projection map $P(x, y, z) = (x, y, 0)$, describe the kernel and image.

2. Prove that $\ker(T)$ and $\operatorname{im}(T)$ are always subspaces.

3. Verify the Rank–Nullity Theorem for the transformation in Example 5.3.1.

5.4 Change of Basis

Linear transformations can look very different depending on the coordinate system we use.
The process of rewriting
vectors and transformations relative to a new basis is called a change of basis. This concept
lies at the heart of
diagonalization, orthogonalization, and many computational techniques.

Coordinate Change

Suppose V is an n-dimensional vector space, and let $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a
basis. Every vector $\mathbf{x} \in V$ has a coordinate vector $[\mathbf{x}]_\mathcal{B} \in \mathbb{R}^n$.

If P is the change-of-basis matrix from $\mathcal{B}$ to the standard basis, then

$$\mathbf{x} = P[\mathbf{x}]_\mathcal{B}.$$

Equivalently,

$$[\mathbf{x}]_\mathcal{B} = P^{-1}\mathbf{x}.$$

Here, P has the basis vectors of $\mathcal{B}$ as its columns:

$$P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}.$$

Transformation of Matrices

Let $T : V \to V$ be a linear transformation. Suppose its matrix in the standard basis is A.
In the
basis $\mathcal{B}$, the representing matrix becomes

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

Thus, changing basis corresponds to a similarity transformation of the matrix.

Example

Example 5.4.1.
Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be given by

$$T(x, y) = (3x + y, x + y).$$

In the standard basis, its matrix is

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

Now consider the basis $\mathcal{B} = \{(1, 1), (1, -1)\}$. The change-of-basis matrix is

$$P = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Then

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

Computing gives

$$[T]_{\mathcal{B}} = \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix}.$$

In this new basis, the transformation is diagonal: one direction is scaled by 4, the other collapsed to 0.

Geometric Interpretation

Change of basis is like rotating or skewing your coordinate grid. The underlying transformation does not change, but its
description in numbers becomes simpler or more complicated depending on the basis. Finding a basis that simplifies a
transformation (often a diagonal basis) is a key theme in linear algebra.

Why this matters

Change of basis connects the abstract notion of similarity to practical computation. It is the tool that allows us to
diagonalize matrices, compute eigenvalues, and simplify complex transformations. In applications, it corresponds to
choosing a more natural coordinate system-whether in geometry, physics, or machine learning.

Exercises 5.4

1. Let

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

Compute its representation in the basis $\{(1,0),(1,1)\}$.

1. Find the change-of-basis matrix from the standard basis of $\mathbb{R}^2$ to $\{(2,1),(1,1)\}$.

2. Prove that similar matrices (related by $P^{-1}AP$) represent the same linear transformation under different bases.

3. Diagonalize the matrix

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

in the basis $\{(1,1),(1,-1)\}$.

1. In $\mathbb{R}^3$, let $\mathcal{B} = \{(1,0,0),(1,1,0),(1,1,1)\}$. Construct the change-of-basis matrix P and compute P^{-1}.

Chapter 6. Determinants

6.1 Motivation and Geometric Meaning

Determinants are numerical values associated with square matrices. At first they may appear as a complicated formula,
but their importance comes from what they measure: determinants encode scaling, orientation, and invertibility of linear
transformations. They bridge algebra and geometry.

Determinants of 2×2 Matrices

For a 2×2 matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

the determinant is defined as

$$\det(A) = ad - bc.$$

Geometric meaning: If A represents a linear transformation of the plane, then $|\det(A)|$ is the area scaling factor.
For example, if $\det(A) = 2$, areas of shapes are doubled. If $\det(A) = 0$, the transformation collapses the plane to
a line: all area is lost.

Determinants of 3×3 Matrices

For

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix},$$

the determinant can be computed as

$$\det(A) = a(ei - fh) - b(di - fg) + c(dh - eg).$$

Geometric meaning: In $\mathbb{R}^3$, $|\det(A)|$ is the volume scaling factor. If $\det(A) < 0$, orientation is
reversed (a handedness flip), such as turning a right-handed coordinate system into a left-handed one.

General Case

For $A \in \mathbb{R}^{n \times n}$, the determinant is a scalar that measures how the linear transformation given
by A scales n-dimensional volume.

- If $\det(A) = 0$: the transformation squashes space into a lower dimension, so A is not invertible.

- If $\det(A) > 0$: volume is scaled by $\det(A)$, orientation preserved.

- If $\det(A) < 0$: volume is scaled by $|\det(A)|$, orientation reversed.

Visual Examples

1. Shear in $\mathbb{R}^2$:
   $$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$
   Then $\det(A) = 1$. The transformation slants the unit square into a parallelogram but preserves area.

2. Projection in $\mathbb{R}^2$:
   $$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$
   Then $\det(A) = 0$. The unit square collapses into a line segment: area vanishes.

3. Rotation in $\mathbb{R}^2$:
   $$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$
   Then $\det(R_\theta) = 1$. Rotations preserve area and orientation.

Why this matters

The determinant is not just a formula-it is a measure of transformation. It tells us whether a matrix is invertible, how
it distorts space, and whether it flips orientation. This geometric insight makes the determinant indispensable in
analysis, geometry, and applied mathematics.

Exercises 6.1

1. Compute the determinant of

$$\begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}$$

What area scaling factor does it represent?

1. Find the determinant of the shear matrix

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

What happens to the area of the unit square?

1. For the 3×3 matrix $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$ Compute the determinant. How does it scale volume in $\mathbb{R}$

2. Show that any rotation matrix in $\mathbb{R}^2$ has determinant 1.

3. Give an example of a 2×2 matrix with determinant -1. What geometric action does it represent?

6.2 Properties of Determinants

Beyond their geometric meaning, determinants satisfy a collection of algebraic rules that make them powerful tools in
linear algebra. These properties allow us to compute efficiently, test invertibility, and understand how determinants
behave under matrix operations.

Basic Properties

Let $A, B \in \mathbb{R}^{n \times n}$, and let $c \in \mathbb{R}$. Then:

1. Identity:

$$\det(I_n) = 1.$$

1. Triangular matrices:
   If A is upper or lower triangular, then

$$\det(A) = a_{11}a_{22}\cdots a_{nn}.$$

1. Row/column swap:
   Interchanging two rows (or columns) multiplies the determinant by -1.

2. Row/column scaling:
   Multiplying a row (or column) by a scalar c multiplies the determinant by c.

3. Row/column addition:
   Adding a multiple of one row to another does not change the determinant.

4. Transpose:

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

1. Multiplicativity:

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

1. Invertibility:
   A is invertible if and only if $\det(A) \neq 0$.

Example Computations

Example 6.2.1.
For

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

A is lower triangular, so

$$\det(A) = 2 \cdot 3 \cdot 5 = 30.$$

Example 6.2.2.
Let

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

Then

$$\det(B) = 1 \cdot 4 - 2 \cdot 3 = -2, \quad \det(C) = -1.$$

Since CB is obtained by swapping rows of B,

$$\det(CB) = -\det(B) = 2.$$

This matches the multiplicativity rule: $\det(CB) = \det(C)\det(B) = (-1)(-2) = 2$.

Geometric Insights

- Row swaps: flipping orientation of space.
- Scaling a row: stretching space in one direction.
- Row replacement: sliding hyperplanes without altering volume.
- Multiplicativity: performing two transformations multiplies their scaling factors.

These properties make determinants both computationally manageable and geometrically interpretable.

Why this matters

Determinant properties connect computation with geometry and theory. They explain why Gaussian elimination works, why
invertibility is equivalent to nonzero determinant, and why determinants naturally arise in areas like volume
computation, eigenvalue theory, and differential equations.

Exercises 6.2

1. Compute the determinant of

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

1. Show that if two rows of a square matrix are identical, then its determinant is zero.

2. Verify $\det(A^T) = \det(A)$ for

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

1. If A is invertible, prove that

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

1. Suppose A is a $3\times 3 matrix with \det(A) = 5. What is \det(2A)$?

6.3 Cofactor Expansion

While determinants of small matrices can be computed directly from formulas, larger matrices require a systematic
method. The cofactor expansion (also known as Laplace expansion) provides a recursive way to compute determinants by
breaking them into smaller ones.

Minors and Cofactors

For an $n \times n$ matrix $A = [a_{ij}]$:

- The minor M_{ij} is the determinant of the $(n-1)\times(n-1)$ matrix obtained by deleting the i-th row and j
  -th column of A.

- The cofactor C_{ij} is defined by

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

The sign factor $(-1)^{i+j}$ alternates in a checkerboard pattern:

$$\begin{bmatrix} + & - & + & - & \cdots \\ - & + & - & + & \cdots \\ + & - & + & - & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Cofactor Expansion Formula

The determinant of A can be computed by expanding along any row or any column:

$$\det(A) = \sum_{j=1}^{n} a_{ij} C_{ij} \quad \text{(expansion along row } i\text{)},$$

$$\det(A) = \sum_{i=1}^{n} a_{ij} C_{ij} \quad \text{(expansion along column } j\text{)}.$$

Example

Example 6.3.1.
Compute

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

Expand along the first row:

$$\det(A) = 1 \cdot C_{11} + 2 \cdot C_{12} + 3 \cdot C_{13}.$$

- For C_{11}:

$$M_{11} = \det \begin{bmatrix} 4 & 5 \\ 0 & 6 \end{bmatrix} = 24$$

so $C_{11} = (+1)(24) = 24$.

- For C_{12}:

$$M_{12} = \det \begin{bmatrix} 0 & 5 \\ 1 & 6 \end{bmatrix} = 0 - 5 = -5$$

so $C_{12} = (-1)(-5) = 5$.

- For C_{13}:

$$M_{13} = \det \begin{bmatrix} 0 & 4 \\ 1 & 0 \end{bmatrix} = 0 - 4 = -4$$

so $C_{13} = (+1)(-4) = -4$.

Thus,

$$\det(A) = 1(24) + 2(5) + 3(-4) = 24 + 10 - 12 = 22.$$

Properties of Cofactor Expansion

1. Expansion along any row or column yields the same result.

2. The cofactor expansion provides a recursive definition of determinant: a determinant of size n is expressed in
   terms of determinants of size $n - 1$.

3. Cofactors are fundamental in constructing the adjugate matrix, which gives a formula for inverses:

$$A^{-1} = \frac{1}{\det(A)} \operatorname{adj}(A), \quad \text{where } \operatorname{adj}(A) = [C_{ji}].$$

Geometric Interpretation

Cofactor expansion breaks down the determinant into contributions from sub-volumes defined by fixing one row or column
at a time. Each cofactor measures how that row/column influences the overall volume scaling.

Why this matters

Cofactor expansion generalizes the small-matrix formulas and provides a conceptual definition of determinants. While not
the most efficient way to compute determinants for large matrices, it is essential for theory, proofs, and connections
to adjugates, Cramer's rule, and classical geometry.

Exercises 6.3

1. Compute the determinant of

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

by cofactor expansion along the first column.

1. Verify that expanding along the second row of Example 6.3.1 gives the same determinant.

2. Prove that expansion along any row gives the same value.

3. Show that if a row of a matrix is zero, then its determinant is zero.

4. Use cofactor expansion to prove that $\det(A) = \det(A^T)$.

6.4 Applications (Volume, Invertibility Test)

Determinants are not merely algebraic curiosities; they have concrete geometric and computational uses. Two of the most
important applications are measuring volumes and testing invertibility of matrices.

Determinants as Volume Scalers

Given vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in \mathbb{R}^n$, arrange them as columns of a matrix:

$$A = \begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \\ | & | & & | \end{bmatrix}.$$

Then $|\det(A)|$ equals the volume of the parallelepiped spanned by these vectors.

- In $\mathbb{R}^2$, $|\det(A)|$ gives the area of the parallelogram spanned by $\mathbf{v}_1, \mathbf{v}_2$.

- In $\mathbb{R}^3$, $|\det(A)|$ gives the volume of the parallelepiped spanned
  by $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$.

- In higher dimensions, it generalizes to n-dimensional volume (hypervolume).

Example 6.4.1.
Let

$$\mathbf{v}_1 = (1, 0, 0), \quad \mathbf{v}_2 = (1, 1, 0), \quad \mathbf{v}_3 = (1, 1, 1).$$

Then

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \quad \det(A) = 1.$$

So the parallelepiped has volume 1, even though the vectors are not orthogonal.

Invertibility Test

A square matrix A is invertible if and only if $\det(A) \neq 0$.

- If $\det(A) = 0$: the transformation collapses space into a lower dimension (area/volume is zero). No inverse exists.

- If $\det(A) \neq 0$: the transformation scales volume by $|\det(A)|$, and is reversible.

Example 6.4.2.
The matrix

$$B = \begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix}$$

has determinant $\det(B) = 2 \cdot 2 - 4 \cdot 1 = 0$.
Thus, B is not invertible. Geometrically, the two column vectors are collinear, spanning only a line
in $\mathbb{R}^2$.

Cramer's Rule

Determinants also provide an explicit formula for solving systems of linear equations when the matrix is invertible.
For $A\mathbf{x} = \mathbf{b}$ with $A \in \mathbb{R}^{n \times n}$:

$$x_i = \frac{\det(A_i)}{\det(A)},$$

where A_i is obtained by replacing the i-th column of A with $\mathbf{b}$.
While inefficient computationally, Cramer's rule highlights the determinant's role in solutions and uniqueness.

Orientation

The sign of $\det(A)$ indicates whether a transformation preserves or reverses orientation. For example, a reflection in
the plane has determinant -1, flipping handedness.

Why this matters

Determinants condense key information: they measure scaling, test invertibility, and track orientation. These insights
are indispensable in geometry (areas and volumes), analysis (Jacobian determinants in calculus), and computation (
solving systems and checking singularity).

Exercises 6.4

1. Compute the area of the parallelogram spanned by $(2, 1)$ and $(1, 3)$.

2. Find the volume of the parallelepiped spanned by $(1, 0, 0), (1, 1, 0), (1, 1, 1)$.

3. Determine whether the matrix
$$\begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}$$
is invertible. Justify using determinants.

1. Use Cramer's rule to solve
$$\begin{cases} x + y = 3, \\ 2x - y = 0. \end{cases}$$

1. Explain geometrically why a determinant of zero implies no inverse exists.

Chapter 7. Inner Product Spaces

7.1 Inner Products and Norms

To extend the geometric ideas of length, distance, and angle beyond $\mathbb{R}^2$ and $\mathbb{R}^3$, we introduce
inner products. Inner products provide a way of measuring similarity between vectors, while norms derived from them
measure length. These concepts are the foundation of geometry inside vector spaces.

Inner Product

An inner product on a real vector space V is a function

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$$

that assigns to each pair of vectors $(\mathbf{u}, \mathbf{v})$ a real number, subject to the following properties:

1. Symmetry:
   $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$.

2. Linearity in the first argument:
   $\langle a\mathbf{u} + b\mathbf{w}, \mathbf{v} \rangle = a\langle \mathbf{u}, \mathbf{v} \rangle + b\langle \mathbf{w}, \mathbf{v} \rangle$.

3. Positive-definiteness:
   $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$, and equality holds if and only if $\mathbf{v} = \mathbf{0}$.

The standard inner product on $\mathbb{R}^n$ is the dot product:

$$\langle \mathbf{u}, \mathbf{v} \rangle = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

Norms

The norm of a vector is its length, defined in terms of the inner product:

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}.$$

For the dot product in $\mathbb{R}^n$:

$$\|(x_1, x_2, \ldots, x_n)\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

Angles Between Vectors

The inner product allows us to define the angle θ between two nonzero vectors $\mathbf{u}, \mathbf{v}$ by

$$\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\| \, \|\mathbf{v}\|}.$$

Thus, two vectors are orthogonal if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$.

Examples

Example 7.1.1.
In $\mathbb{R}^2$, with $\mathbf{u} = (1, 2)$, $\mathbf{v} = (3, 4)$:

$$\langle \mathbf{u}, \mathbf{v} \rangle = 1 \cdot 3 + 2 \cdot 4 = 11.$$

$$\|\mathbf{u}\| = \sqrt{1^2 + 2^2} = \sqrt{5}, \quad \|\mathbf{v}\| = \sqrt{3^2 + 4^2} = 5.$$

So,

$$\cos\theta = \frac{11}{\sqrt{5} \cdot 5}.$$

Example 7.1.2.
In the function space $C[0, 1]$, the inner product

$$\langle f, g \rangle = \int_0^1 f(x)g(x)\, dx$$

defines a length

$$\|f\| = \sqrt{\int_0^1 f(x)^2 dx}.$$

This generalizes geometry to infinite-dimensional spaces.

Geometric Interpretation

- Inner product: measures similarity between vectors.
- Norm: length of a vector.
- Angle: measure of alignment between two directions.

These concepts unify algebraic operations with geometric intuition.

Why this matters

Inner products and norms allow us to extend geometry into abstract vector spaces. They form the basis of orthogonality,
projections, Fourier series, least squares approximation, and many applications in physics and machine learning.

Exercises 7.1

1. Compute $\langle (2,-1,3), (1,4,0) \rangle$. Then find the angle between them.

2. Show that $\|(x,y)\| = \sqrt{x^2 + y^2}$ satisfies the properties of a norm.

3. In $\mathbb{R}^3$, verify that $(1,1,0)$ and $(1,-1,0)$ are orthogonal.

4. In $C[0,1]$, compute $\langle f, g \rangle$ for $f(x) = x$, $g(x) = 1$.

5. Prove the Cauchy–Schwarz inequality:

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \le \|\mathbf{u}\| \, \|\mathbf{v}\|.$$

7.2 Orthogonal Projections

One of the most useful applications of inner products is the notion of orthogonal projection. Projection allows us to
approximate a vector by another lying in a subspace, minimizing error in the sense of distance. This idea underpins
geometry, statistics, and numerical analysis.

Projection onto a Line

Let $\mathbf{u} \in \mathbb{R}^n$ be a nonzero vector. The line spanned by $\mathbf{u}$ is

$$L = \{c\mathbf{u} \mid c \in \mathbb{R}\}.$$

Given a vector $\mathbf{v}$, the projection of $\mathbf{v}$ onto $\mathbf{u}$ is the vector in L closest to $\mathbf{v}$. Geometrically, it is the shadow of $\mathbf{v}$ on the line.

The formula is

$$\text{proj}_{\mathbf{u}}(\mathbf{v}) = \frac{\langle \mathbf{v}, \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle} \mathbf{u}.$$

The error vector $\mathbf{v} - \text{proj}_{\mathbf{u}}(\mathbf{v})$ is orthogonal to $\mathbf{u}$.

Example 7.2.1

Let $\mathbf{u} = (1, 2)$, $\mathbf{v} = (3, 1)$.

$$\langle \mathbf{v}, \mathbf{u} \rangle = 3 \cdot 1 + 1 \cdot 2 = 5, \quad \langle \mathbf{u}, \mathbf{u} \rangle = 1^2 + 2^2 = 5.$$

So

$$\text{proj}_{\mathbf{u}}(\mathbf{v}) = \frac{5}{5}(1, 2) = (1, 2).$$

The error vector is $(3, 1) - (1, 2) = (2, -1)$, which is orthogonal to $(1, 2)$.

Projection onto a Subspace

Suppose $W \subseteq \mathbb{R}^n$ is a subspace with orthonormal basis $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$. The projection of a vector $\mathbf{v}$ onto W is

$$\text{proj}_W(\mathbf{v}) = \langle \mathbf{v}, \mathbf{w}_1 \rangle \mathbf{w}_1 + \cdots + \langle \mathbf{v}, \mathbf{w}_k \rangle \mathbf{w}_k.$$

This is the unique vector in W closest to $\mathbf{v}$. The difference $\mathbf{v} - \text{proj}_W(\mathbf{v})$ is orthogonal to all of W.

Least Squares Approximation

Orthogonal projection explains the method of least squares. To solve an overdetermined system $A\mathbf{x} \approx \mathbf{b}$, we seek the $\mathbf{x}$ that makes $A\mathbf{x}$ the projection of $\mathbf{b}$ onto the column space of A. This gives the normal equations

$$A^T A \mathbf{x} = A^T \mathbf{b}.$$

Thus, least squares is just projection in disguise.

Geometric Interpretation

- Projection finds the closest point in a subspace to a given vector.

- It minimizes distance (error) in the sense of Euclidean norm.

- Orthogonality ensures the error vector points directly away from the subspace.

Why this matters

Orthogonal projection is central in both pure and applied mathematics. It underlies the geometry of subspaces, the
theory of Fourier series, regression in statistics, and approximation methods in numerical linear algebra. Whenever we
fit data with a simpler model, projection is at work.

Exercises 7.2

1. Compute the projection of $(2, 3)$ onto the vector $(1, 1)$.

2. Show that $\mathbf{v} - \text{proj}_{\mathbf{u}}(\mathbf{v})$ is orthogonal to $\mathbf{u}$.

3. Let $W = \text{span}\{(1, 0, 0), (0, 1, 0)\} \subseteq \mathbb{R}^3$. Find the projection of $(1, 2, 3)$ onto W.

4. Explain why least squares fitting corresponds to projection onto the column space of A.

5. Prove that projection onto a subspace W is unique: there is exactly one closest vector in W to a
given $\mathbf{v}$.

7.3 Gram–Schmidt Process

The Gram–Schmidt process is a systematic way to turn any linearly independent set of vectors into an orthonormal basis.
This is especially useful because orthonormal bases simplify computations: inner products become simple coordinate
comparisons, and projections take clean forms.

The Idea

Given a linearly independent set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ in an inner product
space, we want to construct an orthonormal set $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ that spans the same
subspace.

We proceed step by step:

1. Start with $\mathbf{v}_1$, normalize it to get $\mathbf{u}_1$.

2. Subtract from $\mathbf{v}_2$ its projection onto $\mathbf{u}_1$, leaving a vector orthogonal to $\mathbf{u}_1$.
   Normalize to get $\mathbf{u}_2$.

3. For each $\mathbf{v}_k$, subtract projections onto all previously
   constructed $\mathbf{u}_1, \ldots, \mathbf{u}_{k-1}$, then normalize.

The Algorithm

For $k = 1, 2, \ldots, n$:

$$\mathbf{w}_k = \mathbf{v}_k - \sum_{j=1}^{k-1} \langle \mathbf{v}_k, \mathbf{u}_j \rangle \mathbf{u}_j,$$

$$\mathbf{u}_k = \frac{\mathbf{w}_k}{\|\mathbf{w}_k\|}.$$

The result $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is an orthonormal basis of the span of the original vectors.

Example 7.3.1

Take $\mathbf{v}_1 = (1, 1, 0)$, $\mathbf{v}_2 = (1, 0, 1)$, $\mathbf{v}_3 = (0, 1, 1)$ in $\mathbb{R}^3$.

1. Normalize $\mathbf{v}_1$:

$$\mathbf{u}_1 = \frac{1}{\sqrt{2}}(1, 1, 0).$$

1. Subtract projection of $\mathbf{v}_2$ on $\mathbf{u}_1$:

$$\mathbf{w}_2 = \mathbf{v}_2 - \langle \mathbf{v}_2, \mathbf{u}_1 \rangle \mathbf{u}_1.$$

$$\langle \mathbf{v}_2, \mathbf{u}_1 \rangle = \frac{1}{\sqrt{2}}(1 \cdot 1 + 0 \cdot 1 + 1 \cdot 0) = \tfrac{1}{\sqrt{2}}.$$

So

$$\mathbf{w}_2 = (1, 0, 1) - \tfrac{1}{\sqrt{2}} \cdot \tfrac{1}{\sqrt{2}}(1, 1, 0) = (1, 0, 1) - \tfrac{1}{2}(1, 1, 0) = \left(\tfrac{1}{2}, -\tfrac{1}{2}, 1 \right).$$

Normalize:

$$\mathbf{u}_2 = \frac{1}{\sqrt{\tfrac{1}{4} + \tfrac{1}{4} + 1}} \left(\tfrac{1}{2}, -\tfrac{1}{2}, 1 \right) = \frac{1}{\sqrt{\tfrac{3}{2}}} \left(\tfrac{1}{2}, -\tfrac{1}{2}, 1 \right).$$

1. Subtract projections from $\mathbf{v}_3$:

$$\mathbf{w}_3 = \mathbf{v}_3 - \langle \mathbf{v}_3, \mathbf{u}_1 \rangle \mathbf{u}_1 - \langle \mathbf{v}_3, \mathbf{u}_2 \rangle \mathbf{u}_2.$$

After computing, normalize to obtain $\mathbf{u}_3$.

The result is an orthonormal basis of the span of $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$.

Geometric Interpretation

Gram–Schmidt is like straightening out a set of vectors: you start with the original directions and adjust each new
vector to be perpendicular to all previous ones. Then you scale to unit length. The process ensures orthogonality while
preserving the span.

Why this matters

Orthonormal bases simplify inner products, projections, and computations in general. They make coordinate systems easier
to work with and are crucial in numerical methods, QR decomposition, Fourier analysis, and statistics (orthogonal
polynomials, principal component analysis).

Exercises 7.3

1. Apply Gram–Schmidt to $(1, 0), (1, 1)$ in $\mathbb{R}^2$.

2. Orthogonalize $(1, 1, 1), (1, 0, 1)$ in $\mathbb{R}^3$.

3. Prove that each step of Gram–Schmidt yields a vector orthogonal to all previous ones.

4. Show that Gram–Schmidt preserves the span of the original vectors.

5. Explain how Gram–Schmidt leads to the QR decomposition of a matrix.

7.4 Orthonormal Bases

An orthonormal basis is a basis of a vector space in which all vectors are both orthogonal to each other and have unit
length. Such bases are the most convenient possible coordinate systems: computations involving inner products,
projections, and norms become exceptionally simple.

Definition

A set of vectors $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ in an inner product space V is called an
orthonormal basis if

1. $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = 0$ whenever $i \neq j$ (orthogonality),

2. $\|\mathbf{u}_i\| = 1$ for all i (normalization),

3. The set spans V.

Examples

Example 7.4.1. In $\mathbb{R}^2$, the standard basis

$$\mathbf{e}_1 = (1, 0), \quad \mathbf{e}_2 = (0, 1)$$

is orthonormal under the dot product.

Example 7.4.2. In $\mathbb{R}^3$, the standard basis

$$\mathbf{e}_1 = (1, 0, 0), \quad \mathbf{e}_2 = (0, 1, 0), \quad \mathbf{e}_3 = (0, 0, 1)$$

is orthonormal.

Example 7.4.3. Fourier basis on functions:

$$\{1, \cos x, \sin x, \cos 2x, \sin 2x, \dots\}$$

is an orthogonal set in the space of square-integrable functions on $[-\pi, \pi]$ with inner product

$$\langle f, g \rangle = \int_{-\pi}^{\pi} f(x)g(x)\, dx.$$

After normalization, it becomes an orthonormal basis.

Properties

1. Coordinate simplicity: If $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ is an orthonormal basis of V, then any vector $\mathbf{v} \in V$ has coordinates

$$[\mathbf{v}] = \begin{bmatrix} \langle \mathbf{v}, \mathbf{u}_1 \rangle \\ \vdots \\ \langle \mathbf{v}, \mathbf{u}_n \rangle \end{bmatrix}.$$

That is, coordinates are just inner products.

2. Parseval's identity:
For any $\mathbf{v} \in V$,

$$\|\mathbf{v}\|^2 = \sum_{i=1}^{n} |\langle \mathbf{v}, \mathbf{u}_i \rangle|^2.$$

3. Projections:
The orthogonal projection onto the span of
$\mathbf{u}_1, \dots, \mathbf{u}_k$ is

$$\text{proj}(\mathbf{v}) = \sum_{i=1}^{k} \langle \mathbf{v}, \mathbf{u}_i \rangle \mathbf{u}_i.$$

Constructing Orthonormal Bases

- Start with any linearly independent set, then apply the Gram–Schmidt process to obtain an orthonormal set spanning the same subspace.

- In practice, orthonormal bases are often chosen for numerical stability and simplicity of computation.

Geometric Interpretation

An orthonormal basis is like a perfectly aligned and equally scaled coordinate system. Distances and angles are computed directly using coordinates without correction factors. They are the ideal rulers of linear algebra.

Why this matters

Orthonormal bases simplify every aspect of linear algebra: solving systems, computing projections, expanding functions, diagonalizing symmetric matrices, and working with Fourier series. In data science, principal component analysis produces orthonormal directions capturing maximum variance.

Exercises 7.4

1. Verify that $(1/sqrt2)(1, 1)$ and $(1/sqrt2)(1, -1)$ form an orthonormal basis of $\mathbb{R}^2$.

2. Express $(3, 4)$ in terms of the orthonormal basis $\{(1/sqrt2)(1, 1), (1/sqrt2)(1, -1)\}$.

3. Prove Parseval's identity for $mathbbR^n$ with the dot product.

4. Find an orthonormal basis for the plane $x + y + z = 0$ in $mathbbR^3$.

5. Explain why orthonormal bases are numerically more stable than arbitrary bases in computations.

Chapter 8. Eigenvalues and eigenvectors

8.1 Definitions and Intuition

The concepts of eigenvalues and eigenvectors reveal the most fundamental behavior of linear transformations. They
identify the special directions in which a transformation acts by simple stretching or compressing, without rotation or
distortion.

Definition

Let $T : V \to V$ be a linear transformation on a vector space V. A nonzero vector $\mathbf{v} \in V$ is called an
eigenvector of T if

$$T(\mathbf{v}) = \lambda \mathbf{v}$$

for some scalar $\lambda \in \mathbb{R}$ (or $\mathbb{C}$). The scalar λ is the eigenvalue corresponding
to $\mathbf{v}$.

Equivalently, if A is the matrix of T, then eigenvalues and eigenvectors satisfy

$$A\mathbf{v} = \lambda \mathbf{v}.$$

Basic Examples

Example 8.1.1.
Let

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

Then

$$A(1,0)^T = 2(1,0)^T, \quad A(0,1)^T = 3(0,1)^T.$$

So $(1,0)$ is an eigenvector with eigenvalue $2, and (0,1) is an eigenvector with eigenvalue $3.

Example 8.1.2.
Rotation matrix in $\mathbb{R}^2$:

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

If $\theta \neq 0, \pi$, R_θ has no real eigenvalues: every vector is rotated, not scaled. Over $\mathbb{C}$, however, it has eigenvalues $e^{i\theta}, e^{-i\theta}$.

Algebraic Formulation

Eigenvalues arise from solving the characteristic equation:

$$\det(A - \lambda I) = 0.$$

This polynomial in λ is the characteristic polynomial. Its roots are the eigenvalues.

Geometric Intuition

- Eigenvectors are directions that remain unchanged in orientation under a transformation; only their length is scaled.
- Eigenvalues tell us the scaling factor along those directions.
- If a matrix has many independent eigenvectors, it can often be simplified (diagonalized) by changing basis.

Applications in Geometry and Science

- Stretching along principal axes of an ellipse (quadratic forms).
- Stable directions of dynamical systems.
- Principal components in statistics and machine learning.
- Quantum mechanics, where observables correspond to operators with eigenvalues.

Why this matters

Eigenvalues and eigenvectors are a bridge between algebra and geometry. They provide a lens for understanding linear
transformations in their simplest form. Nearly every application of linear algebra-differential equations, statistics,
physics, computer science-relies on eigen-analysis.

Exercises 8.1

1. Find the eigenvalues and eigenvectors of
$\begin{bmatrix} 4 & 0 \\ 0 & -1 \end{bmatrix}$.

2. Show that every scalar multiple of an eigenvector is again an eigenvector for the same eigenvalue.

3. Verify that the rotation matrix R_θ has no real eigenvalues unless $\theta = 0$ or π.

4. Compute the characteristic polynomial of
$\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$.

5. Explain geometrically what eigenvectors and eigenvalues represent for the shear matrix
$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

8.2 Diagonalization

A central goal in linear algebra is to simplify the action of a matrix by choosing a good basis. Diagonalization is the
process of rewriting a matrix so that it acts by simple scaling along independent directions. This makes computations
such as powers, exponentials, and solving differential equations far easier.

Definition

A square matrix $A \in \mathbb{R}^{n \times n}$ is diagonalizable if there exists an invertible matrix P such that

$$P^{-1}AP = D,$$

where D is a diagonal matrix.

The diagonal entries of D are eigenvalues of A, and the columns of P are the corresponding eigenvectors.

When is a Matrix Diagonalizable?

- A matrix is diagonalizable if it has n linearly independent eigenvectors.

- Equivalently, the sum of the dimensions of its eigenspaces equals n.

- Symmetric matrices (over $\mathbb{R}$) are always diagonalizable, with an orthonormal basis of eigenvectors.

Example 8.2.1

Let

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

1. Characteristic polynomial:

$$\det(A - \lambda I) = (4 - \lambda)(2 - \lambda).$$

So eigenvalues are $\lambda_1 = 4$, $\lambda_2 = 2$.

1. Eigenvectors:

- For $\lambda = 4$, solve $(A - 4I)\mathbf{v} = 0$:
$$\begin{bmatrix} 0 & 1 \\ 0 & -2 \end{bmatrix} \mathbf{v} = 0, \text{ giving } \mathbf{v}_1 = (1, 0).$$

- For $\lambda = 2$: $(A - 2I)\mathbf{v} = 0$, giving $\mathbf{v}_2 = (1, -2)$.

1. Construct $P = \begin{bmatrix} 1 & 1 \\ 0 & -2 \end{bmatrix}$. Then

$$P^{-1}AP = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}.$$

Thus, A is diagonalizable.

Why Diagonalize?

- Computing powers:
  If $A = PDP^{-1}$, then

$$A^k = PD^kP^{-1}.$$

  Since D is diagonal, D^k is easy to compute.

- Matrix exponentials:
  $e^A = Pe^DP^{-1}$, useful in solving differential equations.

- Understanding geometry:
  Diagonalization reveals the directions along which a transformation stretches or compresses space independently.

Non-Diagonalizable Example

Not all matrices can be diagonalized.

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

has only one eigenvalue $\lambda = 1$, with eigenspace dimension 1. Since $n = 2$ but we only have 1 independent
eigenvector, A is not diagonalizable.

Geometric Interpretation

Diagonalization means we have found a basis of eigenvectors. In this basis, the matrix acts by simple scaling along each
coordinate axis. It transforms complicated motion into independent 1D motions.

Why this matters

Diagonalization is a cornerstone of linear algebra. It simplifies computation, reveals structure, and is the starting
point for the spectral theorem, Jordan form, and many applications in physics, engineering, and data science.

Exercises 8.2

1. Diagonalize

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

2. Determine whether

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

is diagonalizable. Why or why not?

3. Find A^5 for

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

using diagonalization.

4. Show that any $n \times n$ matrix with n distinct eigenvalues is diagonalizable.

5. Explain why real symmetric matrices are always diagonalizable.

8.3 Characteristic Polynomials

The key to finding eigenvalues is the characteristic polynomial of a matrix. This polynomial encodes the values
of λ for which the matrix $A - \lambda I$ fails to be invertible.

Definition

For an $n \times n$ matrix A, the characteristic polynomial is

$$p_A(\lambda) = \det(A - \lambda I).$$

The roots of $p_A(\lambda)$ are the eigenvalues of A.

Examples

Example 8.3.1.
Let

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

Then

$$p_A(\lambda) = \det \begin{bmatrix} 2 - \lambda & 1 \\ 1 & 2 - \lambda \end{bmatrix} = (2 - \lambda)^2 - 1 = \lambda^2 - 4\lambda + 3.$$

Thus eigenvalues are $\lambda = 1, 3$.

Example 8.3.2.
For

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

(rotation by 90°),

$$p_A(\lambda) = \det \begin{bmatrix} -\lambda & -1 \\ 1 & -\lambda \end{bmatrix} = \lambda^2 + 1.$$

Eigenvalues are $\lambda = \pm i$. No real eigenvalues exist, consistent with pure rotation.

Example 8.3.3.
For a triangular matrix

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

the determinant is simply the product of diagonal entries minus λ:

$$p_A(\lambda) = (2 - \lambda)(3 - \lambda)(4 - \lambda).$$

So eigenvalues are $2, 3, 4$.

Properties

1. The characteristic polynomial of an $n \times n$ matrix has degree n.

2. The sum of the eigenvalues (counted with multiplicity) equals the trace of A:

$$\text{tr}(A) = \lambda_1 + \cdots + \lambda_n.$$

3. The product of the eigenvalues equals the determinant of A:

$$\det(A) = \lambda_1 \cdots \lambda_n.$$

4. Similar matrices have the same characteristic polynomial, hence the same eigenvalues.

Geometric Interpretation

The characteristic polynomial captures when $A - \lambda I$ collapses space: its determinant is zero precisely when the
transformation $A - \lambda I$ is singular. Thus, eigenvalues mark the critical scalings where the matrix loses
invertibility.

Why this matters

Characteristic polynomials provide the computational tool to extract eigenvalues. They connect matrix invariants (trace
and determinant) with geometry, and form the foundation for diagonalization, spectral theorems, and stability analysis
in dynamical systems.

Exercises 8.3

1. Compute the characteristic polynomial of

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

2. Verify that the sum of the eigenvalues of
$\begin{bmatrix} 5 & 0 \\ 0 & -2 \end{bmatrix}$
equals its trace, and their product equals its determinant.

3. Show that for any triangular matrix, the eigenvalues are just the diagonal entries.

4. Prove that if A and B are similar matrices, then $p_A(\lambda) = p_B(\lambda)$.

5. Compute the characteristic polynomial of
$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$

8.4 Applications (Differential Equations, Markov Chains)

Eigenvalues and eigenvectors are not only central to the theory of linear algebra-they are indispensable tools across
mathematics and applied science. Two classic applications are solving systems of differential equations and analyzing
Markov chains.

Linear Differential Equations

Consider the system

$$\frac{d\mathbf{x}}{dt} = A\mathbf{x},$$

where A is an $n \times n$ matrix and $\mathbf{x}(t)$ is a vector-valued function.

If $\mathbf{v}$ is an eigenvector of A with eigenvalue λ, then the function

$$\mathbf{x}(t) = e^{\lambda t}\mathbf{v}$$

is a solution.

- Eigenvalues determine the growth or decay rate:
    - If $\lambda < 0$, solutions decay (stable).
    - If $\lambda > 0$, solutions grow (unstable).
    - If λ is complex, oscillations occur.

By combining eigenvector solutions, we can solve general initial conditions.

Example 8.4.1.
Let

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

Then eigenvalues are $2, -1 with eigenvectors (1,0), (0,1)$. Solutions are

$$\mathbf{x}(t) = c_1 e^{2t}(1, 0) + c_2 e^{-t}(0, 1).$$

Thus one component grows exponentially, the other decays.

Markov Chains

A Markov chain is described by a stochastic matrix P, where each column sums to 1 and entries are nonnegative.
If $\mathbf{x}_k$ represents the probability distribution after k steps, then

$$\mathbf{x}_{k+1} = P\mathbf{x}_k.$$

Iterating gives

$$\mathbf{x}_k = P^k\mathbf{x}_0.$$

Understanding long-term behavior reduces to analyzing powers of P.

- The eigenvalue $\lambda = 1$ always exists. Its eigenvector gives the steady-state distribution.

- All other eigenvalues satisfy $|\lambda| \leq 1$. Their influence decays as $k \to \infty$.

Example 8.4.2.
Consider

$$P = \begin{bmatrix} 0.9 & 0.5 \\ 0.1 & 0.5 \end{bmatrix}.$$

Eigenvalues are $\lambda_1 = 1$, $\lambda_2 = 0.4$. The eigenvector for $\lambda = 1$ is proportional to $(5, 1)$.
Normalizing gives the steady state

$$\pi = \left(\tfrac{5}{6}, \tfrac{1}{6} \right).$$

Thus, regardless of the starting distribution, the chain converges to π.

Geometric Interpretation

- In differential equations, eigenvalues determine the time evolution: exponential growth, decay, or oscillation.

- In Markov chains, eigenvalues determine the long-term equilibrium of stochastic processes.

Why this matters

Eigenvalue methods turn complex iterative or dynamical systems into tractable problems. In physics, engineering, and
finance, they describe stability and resonance. In computer science and statistics, they power algorithms from Google's
PageRank to modern machine learning.

Exercises 8.4

1. Solve $\frac{d}{dt}\mathbf{x} = \begin{bmatrix} 3 & 0 \\ 0 & -2 \end{bmatrix}\mathbf{x}$.

2. Show that if A has a complex eigenvalue $\alpha \pm i\beta$, then solutions of $\frac{d}{dt}\mathbf{x} = A\mathbf{x}$ involve oscillations of frequency β.

3. Find the steady-state distribution of

$$P = \begin{bmatrix} 0.7 & 0.2 \\ 0.3 & 0.8 \end{bmatrix}.$$

4. Prove that for any stochastic matrix P, 1 is always an eigenvalue.

5. Explain why all eigenvalues of a stochastic matrix satisfy $|\lambda| \leq 1$.

Chapter 9. Quadratic Forms and Spectral Theorems

9.1 Quadratic Forms

A quadratic form is a polynomial of degree two in several variables, expressed neatly using matrices. Quadratic forms
appear throughout mathematics: in optimization, geometry of conic sections, statistics (variance), and physics (energy
functions).

Definition

Let A be an $n \times n$ symmetric matrix and $\mathbf{x} \in \mathbb{R}^n$. The quadratic form associated with A is

$$Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}.$$

Expanded,

$$Q(\mathbf{x}) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j.$$

Because A is symmetric $(a_{ij} = a_{ji})$, the cross-terms can be grouped naturally.

Examples

Example 9.1.1.
For

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix},$$

$$Q(x, y) = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 2x^2 + 2xy + 3y^2.$$

Example 9.1.2.
The quadratic form

$$Q(x, y) = x^2 + y^2$$

corresponds to the matrix $A = I_2$. It measures squared Euclidean distance from the origin.

Example 9.1.3.
The conic section equation

$$4x^2 + 2xy + 5y^2 = 1$$

is described by the quadratic form $\mathbf{x}^T A \mathbf{x} = 1$ with

$$A = \begin{bmatrix} 4 & 1 \\ 1 & 5 \end{bmatrix}.$$

Diagonalization of Quadratic Forms

By choosing a new basis consisting of eigenvectors of A, we can rewrite the quadratic form without cross terms.
If $A = PDP^{-1}$ with D diagonal, then

$$Q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x} = (P^{-1}\mathbf{x})^T D (P^{-1}\mathbf{x}).$$

Thus quadratic forms can always be expressed as a sum of weighted squares:

$$Q(\mathbf{y}) = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2,$$

where λ_i are the eigenvalues of A.

Geometric Interpretation

Quadratic forms describe geometric shapes:

- In 2D: ellipses, parabolas, hyperbolas.

- In 3D: ellipsoids, paraboloids, hyperboloids.

- In higher dimensions: generalizations of ellipsoids.

Diagonalization aligns the coordinate axes with the principal axes of the shape.

Why this matters

Quadratic forms unify geometry and algebra. They are central in optimization (minimizing energy functions), statistics (
covariance matrices and variance), mechanics (kinetic energy), and numerical analysis. Understanding quadratic forms
leads directly to the spectral theorem.

Exercises 9.1

1. Write the quadratic form $Q(x, y) = 3x^2 + 4xy + y^2$ as $\mathbf{x}^T A \mathbf{x}$ for some symmetric matrix A.

2. For $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$, compute $Q(x, y)$ explicitly.

3. Diagonalize the quadratic form $Q(x, y) = 2x^2 + 2xy + 3y^2$.

4. Identify the conic section given by $Q(x, y) = x^2 - y^2$.

5. Show that if A is symmetric, quadratic forms defined by A and A^T are identical.

9.2 Positive Definite Matrices

Quadratic forms are especially important when their associated matrices are positive definite, since these guarantee
positivity of energy, distance, or variance. Positive definiteness is a cornerstone in optimization, numerical analysis,
and statistics.

Definition

A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is called:

- Positive definite if

$$\mathbf{x}^T A \mathbf{x} > 0 \quad \text{for all nonzero } \mathbf{x} \in \mathbb{R}^n.$$

- Positive semidefinite if

$$\mathbf{x}^T A \mathbf{x} \geq 0 \quad \text{for all } \mathbf{x}.$$

Similarly, negative definite (always < 0) and indefinite (can be both < 0 and > 0) matrices are defined.

Examples

Example 9.2.1.

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

is positive definite, since

$$Q(x, y) = 2x^2 + 3y^2 > 0$$

for all $(x, y) \neq (0, 0)$.

Example 9.2.2.

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

has quadratic form

$$Q(x, y) = x^2 + 4xy + y^2.$$

This matrix is not positive definite, since $Q(1, -1) = -2 < 0$.

Characterizations

For a symmetric matrix A:

1. Eigenvalue test: A is positive definite if and only if all eigenvalues of A are positive.

2. Principal minors test (Sylvester's criterion): A is positive definite if and only if all leading principal minors (
   determinants of top-left $k \times k$ submatrices) are positive.

3. Cholesky factorization: A is positive definite if and only if it can be written as

$$A = R^T R,$$

where R is an upper triangular matrix with positive diagonal entries.

Geometric Interpretation

- Positive definite matrices correspond to quadratic forms that define ellipsoids centered at the origin.

- Positive semidefinite matrices define flattened ellipsoids (possibly degenerate).

- Indefinite matrices define hyperbolas or saddle-shaped surfaces.

Applications

- Optimization: Hessians of convex functions are positive semidefinite; strict convexity corresponds to positive
  definite Hessians.

- Statistics: Covariance matrices are positive semidefinite.

- Numerical methods: Cholesky decomposition is widely used to solve systems with positive definite matrices efficiently.

Why this matters

Positive definiteness provides stability and guarantees in mathematics and computation. It ensures energy functions are
bounded below, optimization problems have unique solutions, and statistical models are meaningful.

Exercises 9.2

1. Use Sylvester's criterion to check whether

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

   is positive definite.

2. Determine whether

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

   is positive definite, semidefinite, or indefinite.

3. Find the eigenvalues of

$$A = \begin{bmatrix} 4 & 2 \\ 2 & 3 \end{bmatrix},$$

   and use them to classify definiteness.

4. Prove that all diagonal matrices with positive entries are positive definite.

5. Show that if A is positive definite, then so is $P^T A P$ for any invertible matrix P.

9.3 Spectral Theorem

The spectral theorem is one of the most powerful results in linear algebra. It states that symmetric matrices can always
be diagonalized by an orthogonal basis of eigenvectors. This links algebra (eigenvalues), geometry (orthogonal
directions), and applications (stability, optimization, statistics).

Statement of the Spectral Theorem

If $A \in \mathbb{R}^{n \times n}$ is symmetric ($A^T = A$), then:

1. All eigenvalues of A are real.

2. There exists an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A.

3. Thus, A can be written as

$$A = Q \Lambda Q^T,$$

   where Q is an orthogonal matrix ($Q^T Q = I$) and Λ is diagonal with eigenvalues of
   A on the diagonal.

Consequences

- Symmetric matrices are always diagonalizable, and the diagonalization is numerically stable.

- Quadratic forms $\mathbf{x}^T A \mathbf{x}$ can be expressed in terms of eigenvalues and eigenvectors, showing
  ellipsoids aligned with eigen-directions.

- Positive definiteness can be checked by confirming that all eigenvalues are positive.

Example 9.3.1

Let

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

1. Characteristic polynomial:

$$p(\lambda) = (2 - \lambda)^2 - 1 = \lambda^2 - 4\lambda + 3.$$

Eigenvalues: $\lambda_1 = 1, \ \lambda_2 = 3$.

1. Eigenvectors:

- For $\lambda = 1$: solve $(A - I)\mathbf{v} = 0$, giving $(1, -1)$.

- For $\lambda = 3$: solve $(A - 3I)\mathbf{v} = 0$, giving $(1, 1)$.

1. Normalize eigenvectors:

$$\mathbf{u}_1 = \tfrac{1}{\sqrt{2}}(1, -1), \quad \mathbf{u}_2 = \tfrac{1}{\sqrt{2}}(1, 1).$$

1. Then

$$Q = \begin{bmatrix} \tfrac{1}{\sqrt{2}} & \tfrac{1}{\sqrt{2}} 6pt] - \tfrac{1}{\sqrt{2}} & \tfrac{1}{\sqrt{2}} \end{bmatrix}, \quad \Lambda = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}.$$

So

$$A = Q\Lambda Q^T.$$

Geometric Interpretation

The spectral theorem says every symmetric matrix acts like independent scaling along orthogonal directions. In geometry,
this corresponds to stretching space along perpendicular axes.

- Ellipses, ellipsoids, and quadratic surfaces can be fully understood via eigenvalues and eigenvectors.

- Orthogonality ensures directions remain perpendicular after transformation.

Applications

- Optimization: The spectral theorem underlies classification of critical points via eigenvalues of the Hessian.

- PCA (Principal Component Analysis): Data covariance matrices are symmetric, and PCA finds orthogonal directions of maximum variance.

- Differential equations & physics: Symmetric operators correspond to measurable quantities with real eigenvalues (stability, energy).

Why this matters

The spectral theorem guarantees that symmetric matrices are as simple as possible: they can always be analyzed in terms
of real, orthogonal eigenvectors. This provides both deep theoretical insight and powerful computational tools.

Exercises 9.3

1. Diagonalize

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

    using the spectral theorem.

2. Prove that all eigenvalues of a real symmetric matrix are real.

3. Show that eigenvectors corresponding to distinct eigenvalues of a symmetric matrix are orthogonal.

4. Explain geometrically how the spectral theorem describes ellipsoids defined by quadratic forms.

5. Apply the spectral theorem to the covariance matrix

$$\Sigma = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix},$$

    and interpret the eigenvectors as principal directions of variance.

9.4 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a widely used technique in data science, machine learning, and statistics. At its
core, PCA is an application of the spectral theorem to covariance matrices: it finds orthogonal directions (principal
components) that capture the maximum variance in data.

The Idea

Given a dataset of vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m \in \mathbb{R}^n$:

1. Center the data by subtracting the mean vector $\bar{\mathbf{x}}$.

2. Form the covariance matrix

$$\Sigma = \frac{1}{m} \sum_{i=1}^{m} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T.$$

3. Apply the spectral theorem: $\Sigma = Q \Lambda Q^T$.

   - Columns of Q are orthonormal eigenvectors (principal directions).

   - Eigenvalues in Λ measure variance explained by each direction.

The first principal component is the eigenvector corresponding to the largest eigenvalue; it is the direction of maximum
variance.

Example 9.4.1

Suppose we have two-dimensional data points roughly aligned along the line $y = x$. The covariance matrix is
approximately

$$\Sigma = \begin{bmatrix} 2 & 1.9 \\ 1.9 & 2 \end{bmatrix}.$$

Eigenvalues are about $3.9 and $0.1. The eigenvector for $\lambda = 3.9$ is approximately $(1, 1)/\sqrt{2}$.

- First principal component: the line $y = x$.

- Most variance lies along this direction.

- Second component is nearly orthogonal $(y = -x)$, but variance there is tiny.

Thus PCA reduces the data to essentially one dimension.

Applications of PCA

1. Dimensionality reduction: Represent data with fewer features while retaining most variance.

2. Noise reduction: Small eigenvalues correspond to noise; discarding them filters data.

3. Visualization: Projecting high-dimensional data onto top 2 or 3 principal components reveals structure.

4. Compression: PCA is used in image and signal compression.

Connection to the Spectral Theorem

The covariance matrix Σ is always symmetric and positive semidefinite. Hence by the spectral theorem, it has an
orthonormal basis of eigenvectors and nonnegative real eigenvalues. PCA is nothing more than re-expressing data in this
eigenbasis.

Why this matters

PCA demonstrates how abstract linear algebra directly powers modern applications. Eigenvalues and eigenvectors give a
practical method for simplifying data, revealing patterns, and reducing complexity. It is one of the most important
algorithms derived from the spectral theorem.

Exercises 9.4

1. Show that the covariance matrix is symmetric and positive semidefinite.

2. Compute the covariance matrix of the dataset $(1, 2), (2, 3), (3, 4)$, and find its eigenvalues and eigenvectors.

3. Explain why the first principal component captures the maximum variance.

4. In image compression, explain how PCA can reduce storage by keeping only the top k principal components.

5. Prove that the sum of the eigenvalues of the covariance matrix equals the total variance of the dataset.

Chapter 10. Linear Algebra in Practice

10.1 Computer Graphics (Rotations, Projections)

Linear algebra is the language of modern computer graphics. Every image rendered on a
screen, every 3D model rotated or
projected, is ultimately the result of applying matrices to vectors. Rotations, reflections,
scalings, and projections
are all linear transformations, making matrices the natural tool for manipulating geometry.

Rotations in 2D

A counterclockwise rotation by an angle θ in the plane is represented by

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

For any vector $\mathbf{v} \in \mathbb{R}^2$, the rotated vector is

$$\mathbf{v}' = R_\theta \mathbf{v}.$$

This preserves lengths and angles, since R_θ is orthogonal with determinant 1.

Rotations in 3D

In three dimensions, rotations are represented by $3 \times 3 orthogonal matrices with determinant $1.
For example, a
rotation about the z-axis is

$$R_z(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Similar formulas exist for rotations about the x- and y-axes.

More general 3D rotations can be described by axis–angle representation or quaternions,
but the underlying idea is still
linear transformations represented by matrices.

Projections

To display 3D objects on a 2D screen, we use projections:

1. Orthogonal projection: drops the z-coordinate, mapping $(x, y, z) \mapsto (x, y)$.

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

2. Perspective projection: mimics the effect of a camera. A point (x, y, z) projects to

$$\left(\frac{x}{z}, \frac{y}{z} \right),$$

capturing how distant objects appear smaller.

These operations are linear (orthogonal projection) or nearly linear (perspective projection becomes linear in
homogeneous coordinates).

Homogeneous Coordinates

To unify translations and projections with linear transformations, computer graphics uses homogeneous coordinates. A 3D
point (x, y, z) is represented as a 4D vector $(x, y, z, 1)$. Transformations are then 4×4 matrices, which can
represent rotations, scalings, and translations in a single framework.

Example: Translation by (a, b, c):

$$T = \begin{bmatrix} 1 & 0 & 0 & a \\ 0 & 1 & 0 & b \\ 0 & 0 & 1 & c \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Geometric Interpretation

- Rotations preserve shape and size, only changing orientation.

- Projections reduce dimension: from 3D world space to 2D screen space.

- Homogeneous coordinates allow us to combine multiple transformations (rotation + translation + projection) into a
  single matrix multiplication.

Why this matters

Linear algebra enables all real-time graphics: video games, simulations, CAD software, and movie effects. By chaining
simple matrix operations, complex transformations are applied efficiently to millions of points per second.

Exercises 10.1

1. Write the rotation matrix for a 90° counterclockwise rotation in $\mathbb{R}^2$. Apply it to $(1,0)$.

2. Rotate the point $(1,1,0)$ about the z-axis by 180°.

3. Show that the determinant of any 2D or 3D rotation matrix is 1.

4. Derive the orthogonal projection matrix from $\mathbb{R}^3$ to the xy-plane.

5. Explain how homogeneous coordinates allow translations to be represented as matrix multiplications.

10.2 Data Science (Dimensionality Reduction, Least Squares)

Linear algebra provides the foundation for many data science techniques. Two of the most important are dimensionality
reduction, where high-dimensional datasets are compressed while preserving essential information, and the least squares
method, which underlies regression and model fitting.

Dimensionality Reduction

High-dimensional data often contains redundancy: many features are correlated, meaning the data essentially lies near a
lower-dimensional subspace. Dimensionality reduction identifies these subspaces.

- PCA (Principal Component Analysis):
  As introduced earlier, PCA diagonalizes the covariance matrix of the data.

  – Eigenvectors (principal components) define orthogonal directions of maximum variance.

  – Eigenvalues measure how much variance lies along each direction.

  – Keeping only the top k components reduces data from n-dimensional space to k-dimensional space while
  retaining most variability.

Example 10.2.1. A dataset of 1000 images, each with 1024 pixels, may have most variance captured by just 50 eigenvectors
of the covariance matrix. Projecting onto these components compresses the data while preserving essential features.

Least Squares

Often, we have more equations than unknowns-an overdetermined system:

$$A\mathbf{x} \approx \mathbf{b}, \quad A \in \mathbb{R}^{m \times n}, \ m > n.$$

An exact solution may not exist. Instead, we seek $\mathbf{x}$ that minimizes the error

$$\|A\mathbf{x} - \mathbf{b}\|^2.$$

This leads to the normal equations:

$$A^T A\mathbf{x} = A^T\mathbf{b}.$$

The solution is the orthogonal projection of $\mathbf{b}$ onto the column space of A.

Example 10.2.2

Fit a line $y = mx + c$ to data points (x_i, y_i).

Matrix form:

$$A = \begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} m \\ c \end{bmatrix}.$$

Solve $A^T A\mathbf{x} = A^T\mathbf{b}$. This yields the best-fit line in the least squares sense.

Geometric Interpretation

- Dimensionality reduction: Find the best subspace capturing most variance.

- Least squares: Project the target vector onto the subspace spanned by predictors.

Both are projection problems, solved using inner products and orthogonality.

Why this matters

Dimensionality reduction makes large datasets tractable, filters noise, and reveals structure. Least squares fitting
powers regression, statistics, and machine learning. Both rely directly on eigenvalues, eigenvectors, and
projections-core tools of linear algebra.

Exercises 10.2

1. Explain why PCA reduces noise in datasets by discarding small eigenvalue components.

2. Compute the least squares solution to fitting a line through $(0, 0), (1, 1), (2, 2)$.

3. Show that the least squares solution is unique if and only if $A^T A$ is invertible.

4. Prove that the least squares solution minimizes the squared error by projection arguments.

5. Apply PCA to the data points $(1, 0), (2, 1), (3, 2)$ and find the first principal component.

10.3 Networks and Markov Chains

Graphs and networks provide a natural setting where linear algebra comes to life. From modeling flows and connectivity
to predicting long-term behavior, matrices translate network structure into algebraic form. Markov chains, already
introduced in Section 8.4, are a central example of networks evolving over time.

Adjacency Matrices

A network (graph) with n nodes can be represented by an adjacency matrix $A \in \mathbb{R}^{n \times n}$:

$$A_{ij} = \begin{cases} 1 & \text{if there is an edge from node } i \text{ to node } j \\ 0 & \text{otherwise.} \end{cases}$$

For weighted graphs, entries may be positive weights instead of $0/1$.

- The number of walks of length k from node i to node j is given by the entry $(A^k)_{ij}$.

- Powers of adjacency matrices thus encode connectivity over time.

Laplacian Matrices

Another important matrix is the graph Laplacian:

$$L = D - A,$$

where D is the diagonal degree matrix ($D_{ii} = \text{degree}(i)$).

- L is symmetric and positive semidefinite.

- The smallest eigenvalue is always $0, with eigenvector (1,1,\dots,1)$.

- The multiplicity of eigenvalue 0 equals the number of connected components in the graph.

This connection between eigenvalues and connectivity forms the basis of spectral graph theory.

Markov Chains on Graphs

A Markov chain can be viewed as a random walk on a graph. If P is the transition matrix where P_{ij} is the
probability of moving from node i to node j, then

$$\mathbf{x}_{k+1} = P\mathbf{x}_k$$

describes the distribution of positions after k steps.

- The steady-state distribution is given by the eigenvector of P with eigenvalue 1.

- The speed of convergence depends on the gap between the largest eigenvalue (which is always 1) and the second
  largest eigenvalue.

Example 10.3.1

Consider a simple 3-node cycle graph:

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

This Markov chain cycles deterministically among the nodes. Eigenvalues are the cube roots of
unity: $1, e^{2\pi i/3}, e^{4\pi i/3}$. The eigenvalue 1 corresponds to the steady state, which is the uniform
distribution $(1/3, 1/3, 1/3)$.

Applications

- Search engines: Google's PageRank algorithm models the web as a Markov chain, where steady-state probabilities rank
  pages.

- Network analysis: Eigenvalues of adjacency or Laplacian matrices reveal communities, bottlenecks, and robustness.

- Epidemiology and information flow: Random walks model how diseases or ideas spread through networks.

Why this matters

Linear algebra transforms network problems into matrix problems. Eigenvalues and eigenvectors reveal connectivity, flow,
stability, and long-term dynamics. Networks are everywhere-social media, biology, finance,

and the internet-so these
tools are indispensable.

1. Write the adjacency matrix of a square graph with 4 nodes. Compute A^2 and interpret the entries.

2. Show that the Laplacian of a connected graph has exactly one zero eigenvalue.

3. Find the steady-state distribution of the Markov chain with

$$P = \begin{bmatrix} 0.5 & 0.5 \\ 0.4 & 0.6 \end{bmatrix}.$$

4. Explain how eigenvalues of the Laplacian can detect disconnected components of a graph.

5. Describe how PageRank modifies the transition matrix of the web graph to ensure a unique steady-state distribution.

10.4 Machine Learning Connections

Modern machine learning is built on linear algebra. From the representation of data as matrices to the optimization of
large-scale models, nearly every step relies on concepts such as vector spaces, projections, eigenvalues, and matrix
decompositions.

Data as Matrices

A dataset with m examples and n features is represented as a matrix $X \in \mathbb{R}^{m \times n}$:

$$X = \begin{bmatrix} \mathbf{x}_1^T & - \\ \mathbf{x}_2^T & - \\ \vdots & \\ \mathbf{x}_m^T & - \end{bmatrix},$$

where each row $\mathbf{x}_i \in \mathbb{R}^n$ is a feature vector. Linear algebra provides tools to analyze, compress,
and transform this data.

Linear Models

At the heart of machine learning are linear predictors:

$\hat{y} = X\mathbf{w},$

where $\mathbf{w}$ is the weight vector. Training often involves solving a least squares problem or a regularized
variant such as ridge regression:

$\min_{\mathbf{w}} \|X\mathbf{w} - \mathbf{y}\|^2 + \lambda \|\mathbf{w}\|^2.$

This is solved efficiently using matrix factorizations.

Singular Value Decomposition (SVD)

The SVD of a matrix X is

$X = U\Sigma V^T,$

where U, V are orthogonal and Σ is diagonal with nonnegative entries (singular values).

- Singular values measure the importance of directions in feature space.
- SVD is used for dimensionality reduction (low-rank approximations), topic modeling, and recommender systems.

Eigenvalues in Machine Learning

- PCA (Principal Component Analysis): diagonalization of the covariance matrix identifies directions of maximal
  variance.
- Spectral clustering: uses eigenvectors of the Laplacian to group data points into clusters.
- Stability analysis: eigenvalues of Hessian matrices determine whether optimization converges to a minimum.

Neural Networks

Even deep learning, though nonlinear, uses linear algebra at its core:

- Each layer is a matrix multiplication followed by a nonlinear activation.
- Training requires computing gradients, which are expressed in terms of matrix calculus.
- Backpropagation is essentially repeated applications of the chain rule with linear algebra.

Why this matters

Machine learning models often involve datasets with millions of features and parameters. Linear algebra provides the
algorithms and abstractions that make training and inference possible. Without it, large-scale computation in AI would
be intractable.

Exercises 10.4

1. Show that ridge regression leads to the normal equations

$$(X^T X + \lambda I)\mathbf{w} = X^T \mathbf{y}.$$

1. Explain how SVD can be used to compress an image represented as a matrix of pixel intensities.

2. For a covariance matrix Σ, show why its eigenvalues represent variances along principal components.

3. Give an example of how eigenvectors of the Laplacian matrix can be used for clustering a small graph.

4. In a neural network with one hidden layer, write the forward pass in matrix form.