Beijing-Dublin International College (BDIC)

Linear Algebra

(*BDIC1014J & BDIC1022J*)

Xin LIU & Erjun Zhang

2017

Contents

Chapter 1 Revisit to high school mathematics

This chapter is a short review of some algebraic contents of high school mathematics, serving as a bridge between high school and university. Student are suggested to use this simple chapter to get themselves familiar with mathematics learning in the language of English.

In this chapter we will have the chance to pay a revisit to set theory （集合论）, notations of summation (the “big Σ” symbol) and product (the “big Π” symbol), and the technique of mathematical induction（数学归纳法）.

§ 1.1 Set theory

§ 1.1.1 Terminology

***Definition* 1.1.** A *set*（集合）is a collection of objects which are called members or elements of the set.

A set is usually represented as enclosed by a pair of braces { }. If an element x belongs to a set S, we write $x \in S$; otherwise, $x \notin S$.

A set is represented in two ways:

- Listing: e.g., $\{a, b, c, d\}$;
- Description: e.g., $\{n|n = 2k + 1, k$ — an positive integer$\}$ (set of odd natural integers).

***Definition* 1.2.** A set S is called an *empty set* or *null set* （空集）, denoted as $\varnothing$, if it has no elements.

***Definition* 1.3.** A set R is called a *subset*（子集）of another set S, denoted as $R \subseteq S$, if every element of R is also an element of S.

Empty set is a subset of any set S, $\varnothing \subseteq S$.

***Definition* 1.4.** A set R is called a *proper subset*（真子集）of another set S, denoted as $R \subset S$, if $R \subseteq S$ and there exists at least one element $x \in S$ but $x \notin R$.

✎ ***Example* 1.1.** Let $R = \{a, b, c\}$ and $S = \{a, b, c, d\}$. There are $R \subseteq R$ and $R \subset S$.

✎ ***Example* 1.2.** Let $A = \{n|n = 2k - 1, k$ an integer$\}$ and $B =$ set of integers. Then $A \subset B$.

§ 1.1.2 Sets of numbers

***Definition* 1.5.** $\mathbb{N}$ — The set of natural numbers（自然数）, $\{1, 2, 3, 4, \cdots\}$.

$\mathbb{N}$ is closed（封闭的）under the operations of addition and multiplication.

***Definition* 1.6.** $\mathbb{Z}$ — The set of integers（整数）, $\{\cdots, -3, -2, -1, 0, 1, 2, 3, \cdots\}$.

$\mathbb{Z}$ is closed under the operations of addition, subtraction and multiplication.

***Definition* 1.7.** $\mathbb{Q}$ — The set of rational numbers（有理数）, i.e., the numbers of the form $\frac{n}{m}$ where n and m are integers with $m \neq 0$.

Rational numbers include decimals which either terminate or repeat. Integers are a subset of the rational numbers, since they are of the form $\frac{n}{m}$ where $m = 1$.

$\mathbb{Q}$ is closed under the operations of addition, subtraction, multiplication and division, provided that division by zero is excluded.

***Definition* 1.8.** $\mathbb{R}$ — The set of real numbers（实数）, including all rational numbers and all irrational numbers.

***Definition* 1.9.** $\mathbb{C}$ — The set of complex numbers（复数）.

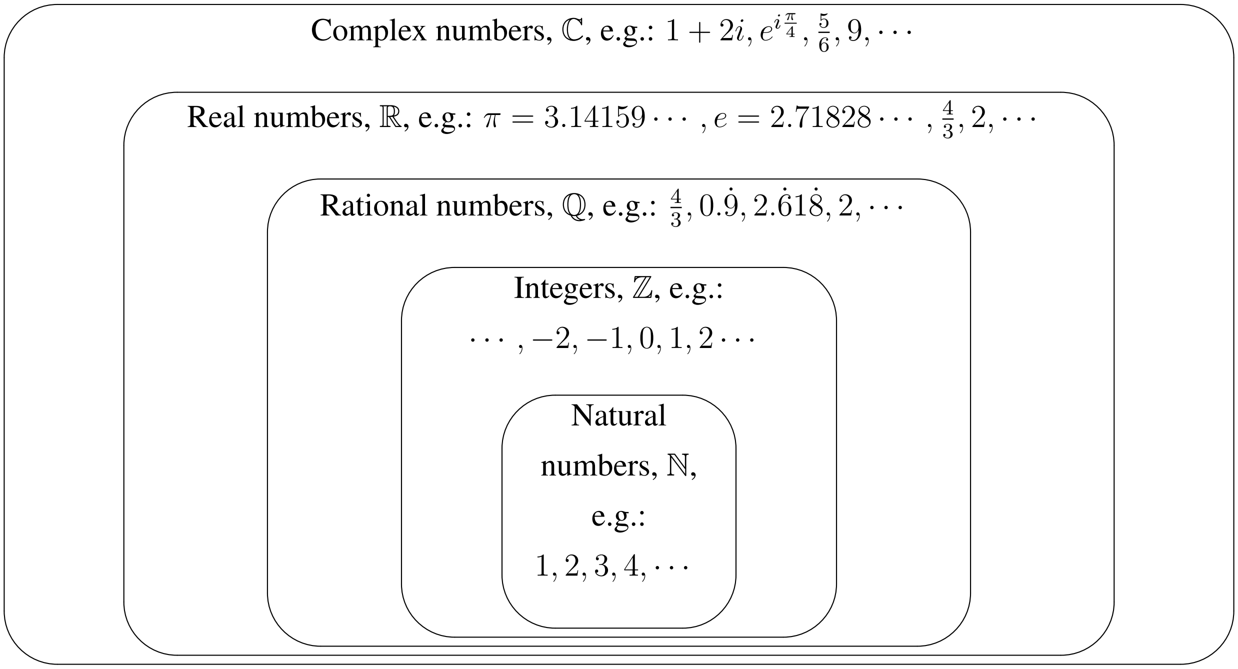
A complex number is a number that can be expressed in the form $a + bi$, where i is the imaginary unit satisfying $i^2 = -1$, and a and b are called the real and imaginary part, respectively, $a, b \in \mathbb{R}$.

Complex numbers extend the concept of the one-dimensional number line to the two-dimensional complex plane by using the horizontal axis for the real part and the vertical axis for the imaginary.

The inclusive relationships（包容关系）among the above sets of numbers is

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}, \tag{1.1.1}$$

as shown diagrammatically below.



§ 1.1.3 Union and intersection of sets

***Definition* 1.10.** The union（并集）of two sets A and B, denoted as $A \cup B$, is the set of all elements in A or in B:

$$A \cup B = \{x | x \in A \text{ or } x \in B\}. \tag{1.1.2}$$

Extending this concept, the union of a finite number of sets is defined by:

$$S = \bigcup_{\alpha=1}^{n} S_\alpha = \{x | x \in S_\alpha, \text{ for at least one value of } \alpha.\} \tag{1.1.3}$$

or

$$S = \bigcup_{\alpha \in A} S_\alpha, \qquad\qquad A \text{ the index set of } \alpha. \tag{1.1.4}$$

***Definition* 1.11.** The intersection（交集）of two sets A and B, denoted as $A \cap B$, is the set of all elements both in A and in B:

$$A \cap B = \{x | x \in A \text{ and } x \in B\}. \tag{1.1.5}$$

Extending this concept, the intersection of a finite number of sets is defined by:

$$S = \bigcap_{\alpha=1}^{n} S_\alpha = \{x | x \in S_\alpha, \alpha = 1, \cdots, n\} \tag{1.1.6}$$

or

$$S = \bigcap_{\alpha \in A} S_\alpha, \qquad\qquad A \text{ the index set of } \alpha. \tag{1.1.7}$$

***Definition* 1.12.** Let A and B be two sets. The complement（补集）of B with respect to A is defined as

$$A \setminus B = \{x | x \in A \text{ but } x \notin B\}. \tag{1.1.8}$$

And the complement of A with respect to B is defined as

$$B \setminus A = \{x | x \in B \text{ but } x \notin A\}. \tag{1.1.9}$$

✎ ***Example* 1.3.** Let $A = \{a, b, c, d\}$, $B = \{a, b, e, f\}$ and $C = \{a, b, c, d, g, h\}$. Then

$$A \cup B = \{a, b, c, d, e, f\}, \qquad A \cap B = \{a, b\}.$$

$$A \setminus B = \{c, d\}, \qquad B \setminus A = \{e, f\}, \qquad C \setminus A = \{g, h\}.$$

✎ ***Example* 1.4.** Let $n \in \mathbb{N}$ and X_n be an interval（区间）on the number line,

$$X_n = \left[-\frac{1}{n}, \frac{1}{n}\right] = \left\{x \in \mathbb{R} \middle| -\frac{1}{n} \leq x \leq \frac{1}{n}\right\}.$$

Then

$$\bigcup_{n=1}^{\infty} X_n = [-1, 1], \qquad \bigcap_{n=1}^{\infty} X_n = \{0\}.$$

Set-theoretical relationships are often shown via the so-called *Venn diagram*（韦恩图）:

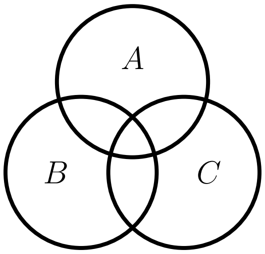


Figure 1.1: Venn diagram.

§ **1.1.4 Properties of set union and intersection**

Let A, B, C and U are sets, with $A, B, C \subseteq U$. We call U the universe or the universal set（全集）. Let $\bar{A}, \bar{B}$ and $\bar{C}$ denote the complements of A, B and C in U : $\bar{A} = U \backslash A, \bar{B} = U \backslash B, \bar{C} = U \backslash C$.

Then we have the following properties:

1. $\varnothing$ and universe:

$$\varnothing \cup A = A = A \cup \varnothing, \qquad \varnothing \cap A = \varnothing = A \cap \varnothing,$$
$$U \cup A = U = A \cup U, \qquad U \cap A = A = A \cap U. \tag{1.1.10}$$

2. Commutativity（可对易性）:

$$A \cup B = B \cup A, \qquad A \cap B = B \cap A. \tag{1.1.11}$$

3. Associativity（可结合性）:

$$(A \cup B) \cup C = A \cup (B \cup C), \qquad (A \cap B) \cap C = A \cap (B \cap C). \tag{1.1.12}$$

4. Distributivity（可分配性）:

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C), \qquad A \cup (B \cap C) = (A \cup B) \cap (A \cup C). \tag{1.1.13}$$

5. Transitivity（可传导性）:

$$A \subseteq B, B \subseteq C \Longrightarrow A \subseteq C. \tag{1.1.14}$$

6. Complements:

$$A \cap \bar{A} = \varnothing, \qquad A \cup \bar{A} = U; \qquad \overline{(\overline{A})} = A; \tag{1.1.15}$$

$$\overline{(A \cup B)} = \overline{A} \cap \overline{B}, \qquad \overline{(A \cap B)} = \overline{A} \cup \overline{B}. \tag{1.1.16}$$

Venn diagrams are a useful tool for understanding the above formulae; for example, see Fig.1.2 for the property of distributivity. Here the proofs for these formulae are ignored.

Figure 1.2: Venn diagrams showing union and intersection operations of sets: (*left*) $(B \cup C) \cap A = (B \cap A) \cup (C \cap A)$; (*right*) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

§ 1.1.5 Cartesian products of sets

***Definition* 1.13.** Let R and S be two sets. Their Cartesian product（笛卡尔积，直积）, denoted by $R \times S$, is defined as the set containing all ordered pairs (x, y) with $x \in R$ and $y \in S$:

$$R \times S = \{(x, y) | x \in R, y \in S\}. \tag{1.1.17}$$

Two ordered pairs (x, y) and (u, v) are equal if and only if $x = u$ in R and $y = v$ in S.

***Definition* 1.14.** The Cartesian product of a finite number of sets $S_1, S_2, \cdots, S_n$ are defined as the set of all ordered n-tuples（n元组）, where $x_i \in S_i, i = 1, \cdots, n$:

$$S_1 \times S_2 \times \cdots \times S_n = \{(x_1, x_2, \cdots, x_n) \, | x_i \in S_i, i = 1, \cdots, n\}. \tag{1.1.18}$$

If specially $S_i = S$ for all i (denoted as $\forall i$), the Cartesian product is usually expressed as S^n.

In the context off linear algebra and multivariable calculus, for example, computations are usually conducted within the space $\mathbb{R}^n$, the Euclidean n-space, which is the Cartesian product of n copies of the real number set $\mathbb{R}$.

✍ ***Exercises* 1.1.**

1. Let $U = \{0, 1, 2, \cdots, 9\}$ be the universal set, and $A = \{0, 2, 4, 6, 8\}$ be a subset of U. List out explicitly the elements of $\bar{A} = U \backslash A$.
2. Let P and Q be two sets, $P = \{\,(x, y) |\, y = x^2 + 1;\ x, y \in \mathbb{R}\}$, $Q = \{\,(x, y) |\, y = x + 1;\ x, y \in \mathbb{R}\}$. Try to find $P \cap Q$.

§ 1.2 Summation and product notations

§ 1.2.1 Summation notation — big Σ

The Greek letter sigma, Σ, is used to denote the sum (finite or infinite) of a set of objects of a set.

***Definition* 1.15.** Let $x_1, x_2, \cdots, x_n$ be n objects (numbers, vectors, matrices, etc.) of a set. We write their *finite series*（有限级数）as

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n. \tag{1.2.1}$$

The i in the symbol x_i is called a *subscript*（脚标）, which is a label/tag used to distinguish one element from another and order the elements in a particular way. In general, we refer to the subscripts as *dummy variables* （哑变量）.[*1]

When the upper summation bound of a finite series goes to infinity, which means we have an infinite sequence of numbers $\{a_i\}_{i=1}^{\infty}$ to add, we obtain an *infinite series*（无穷级数）:

$$\sum_{i=1}^{\infty} a_i = \lim_{n\to\infty} \sum_{i=1}^{n} a_i. \tag{1.2.2}$$

☞ **[Aside]:**

Such an infinite series raises questions of *convergence*（收敛）or *divergence*（发散）. The infinite series is said to be convergent if the sequence of partial sums（部分和）

$$\left\{\sum_{i=1}^{n} a_i\right\}_{n=1}^{\infty} = \{a_1, a_1 + a_2, \cdots, a_1 + a_2 + \cdots + a_n + \cdots\} \tag{1.2.3}$$

converges. Otherwise, the series is said to be divergent. ✌

[*1] "Dummy" means these variables are used only for counting the number of the variables but have no essential meaning. That is, the two ways below are the same to express the sum of l elements:

$$\sum_{i=1}^{l} ix^i = \sum_{n=1}^{l} nx^n.$$

There are basic properties of this big sigma notation that simplify our dealing with finite series.

***Theorem* 1.1.** Let $\{x_1, x_2, \cdots, x_n\}, \{y_1, y_2, \cdots, y_n\}$ and $\{z_1, z_2, \cdots, z_n\}$ be finite sets of mathematical objects, with the same defined addition. Let k be a constant. If the addition operation is commutative and associative, then

$$\sum_{i=1}^{n}(x_i + y_i + z_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i + \sum_{i=1}^{n} z_i; \tag{1.2.4}$$

$$\sum_{i=1}^{n} kx_i = k\sum_{i=1}^{n} x_i; \tag{1.2.5}$$

$$\sum_{i=1}^{n} k = nk. \tag{1.2.6}$$

Proof:

$$\begin{aligned}
\sum_{i=1}^{n}(x_i + y_i + z_i) &= (x_1 + y_1 + z_1) + (x_2 + y_2 + z_2) + \cdots + (x_n + y_n + z_n) \\
&= (x_1 + x_2 + \cdots + x_n) + (y_1 + y_2 + \cdots + y_n) + (z_1 + z_2 + \cdots + z_n) \\
&= \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i + \sum_{i=1}^{n} z_i; \\
\sum_{i=1}^{n} kx_i &= kx_1 + kx_2 + \cdots + kx_n = k\left(x_1 + x_2 + \cdots + x_n\right) = k\sum_{i=1}^{n} x_i; \\
\sum_{i=1}^{n} k &= \sum_{i=1}^{n} k \times 1 = k\sum_{i=1}^{n} 1 = k(\overbrace{1 + 1 + \cdots + 1}^{n \text{ copies}}) = kn.
\end{aligned}$$

□

§ 1.2.2 Commonly used series

Arithmetic series

This arithmetic series（等差数列）is generated by

$$a_{n+1} = a_n + d, \quad d \neq 0, \tag{1.2.7}$$

where d is called the common difference（公差）. Hence

$$\begin{aligned} a_n &= a_0 + nd, && (1.2.8) \\ \sum_{i=0}^{n} a_i &= (n+1)\,a_0 + \frac{n\,(n+1)}{2} d. && (1.2.9) \end{aligned}$$

✎ ***Example* 1.5.**

$$\begin{aligned} & 5 + 9 + 13 + 17 + \cdots + 61 + 65 \\ = & \sum_{i=0}^{15} (5 + 4i) = (15+1) \times 5 + \frac{15 \times (15+1)}{2} \times 4 \\ = & 560. \end{aligned}$$

This series can also be computed as

$$\begin{aligned} & 65 + 61 + \cdots + 17 + 13 + 9 + 5 \\ = & \sum_{i=0}^{15} [65 + (-4)\,i] = (15+1) \times 65 + \frac{15 \times (15+1)}{2} \times (-4) \\ = & 560. \end{aligned}$$

Geometric series

This geometric series（等比数列）is generated by

$$a_n = a_{n-1} r, \qquad (1.2.10)$$

where $a_0 \neq 0$, and r is called the common ratio（公比）, $|r| \neq 0, 1.$[*2] Hence

$$\begin{aligned} a_n &= a_0 r^n, && (1.2.11) \\ \sum_{i=0}^{n} a_i &= a_0 \frac{1 - r^{n+1}}{1 - r}. && (1.2.12) \end{aligned}$$

An obvious corollary is

$$\sum_{i=0}^{\infty} a_i = \begin{cases} \frac{a_0}{1-r}, \text{ converges,} & \text{when } |r| < 1; \\ \quad \text{diverges,} & \text{when } |r| > 1. \end{cases} \qquad (1.2.13)$$

[*2] If $r = 0$, the sequence is a trivial $\{a_0, 0, \cdots, 0\}$. If $r = 1$, the sequence is $\{a_0, a_0, \cdots, a_0\}$, and the series $\sum_{i=1}^{n} a_i = na_0$. If If $r = -1$, the sequence is $\{a_0, -a_0, a_0, -a_0, \cdots, \pm a_0\}$, and the series $\sum_{i=1}^{n} a_i = \begin{cases} a_0, & \text{when } n \text{ odd;} \\ 0, & \text{when } n \text{ even.} \end{cases}$

✎ ***Example* 1.6.**

$$\begin{aligned}
2+\frac{2}{3}+\frac{2}{9}+\cdots+\frac{2}{243} &= \sum_{i=0}^{5} 2\times\frac{1}{3^i} = 2\times\sum_{i=0}^{5}\frac{1}{3^i} = 2\times\frac{1-\frac{1}{3^{5+1}}}{1-\frac{1}{3}} = \frac{728}{243},\\
2-\frac{2}{3}+\frac{2}{9}-\cdots-\frac{2}{243} &= \sum_{i=0}^{5} 2\times\frac{1}{(-3)^i} = 2\times\sum_{i=0}^{5}\frac{1}{(-3)^i} = 2\times\frac{1-\frac{1}{(-3)^{5+1}}}{1-\frac{1}{(-3)}} = \frac{364}{243}.
\end{aligned}$$

☞ **[Aside]:**

In statistics, an important measure of the variation present in a finite set of discrete measurements makes use of a sum of the form

$$S=\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2, \tag{1.2.14}$$

where $\bar{x}$ is the so-called arithmetic mean（算术平均），

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i. \tag{1.2.15}$$

In the light of the properties of summation one can derive a form of S which is more convenient to use:

$$\begin{aligned}
\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2 &= \sum_{i=1}^{n}\left(x_i^2-2x_i\bar{x}+\bar{x}^2\right)=\sum_{i=1}^{n}x_i^2-2\sum_{i=1}^{n}x_i\bar{x}+\sum_{i=1}^{n}\bar{x}^2\\
&= \sum_{i=1}^{n}x_i^2-2\bar{x}\sum_{i=1}^{n}x_i+n\bar{x}^2=\sum_{i=1}^{n}x_i^2-2\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)\sum_{i=1}^{n}x_i+n\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)^2,
\end{aligned}$$

i.e.,

$$S=\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2=\sum_{i=1}^{n}x_i^2-\frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n}. \tag{1.2.16}$$

✌

§ 1.2.3 Product notation — big Π

The big pi notation, Π, is for finite and infinite products:

$$\prod_{i=1}^{n}a_i=a_1a_2\cdots a_n, \qquad \prod_{i=1}^{\infty}a_i=\lim_{n\to\infty}\prod_{i=1}^{n}a_i=a_1a_2\cdots a_n\cdots, \tag{1.2.17}$$

where the subscript i is a dummy variable.

✎ ***Example* 1.7.** The factorial of a positive integer n can be expressed as

$$n!=\prod_{i=1}^{n}i. \tag{1.2.18}$$

For example, $9!=\prod_{k=1}^{9}k$.

✎ ***Example* 1.8.** Let $a_1, a_2, \cdots, a_n$ be nonnegative real numbers. The famous *Arithmetic-Geometric Mean* (AGM) inequality reads

$$\sqrt[n]{a_1 a_2 \cdots a_n} \le \frac{a_1 + a_2 + \cdots + a_n}{n},$$

which can be expressed succinctly in terms of the Σ and Π notations:

$$\left(\prod_{i=1}^{n} a_i\right)^{\frac{1}{n}} \le \frac{1}{n}\sum_{i=1}^{n} a_i. \tag{1.2.19}$$

The “=” holds when $a_1 = a_2 = \cdots = a_n$.

§ **1.2.4 Two important techniques in summation computations**

Double series

***Definition* 1.16.** A double series or double sum（双重求和）is a series having terms depending on two subscripts (indices), written as, say,

$$\begin{aligned} S &= \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij} = \sum_{i=1}^{m}(a_{i1} + a_{i2} + \cdots + a_{in}) \\ &= (a_{11} + a_{12} + \cdots + a_{1n}) + (a_{21} + a_{22} + \cdots + a_{2n}) \\ &\quad + \cdots + (a_{m1} + a_{m2} + \cdots + a_{mn}), \end{aligned} \tag{1.2.20}$$

which sums up all the elements of $\{a_{ij}\}$. Other notations include $\sum_{i,j} a_{ij}$, $\sum_{1\le i,j\le n} a_{ij}$, etc..

An important property of double series is that the calculation can be conducted by swapping the order of summation:

$$\begin{aligned} S &= \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij} = \sum_{j=1}^{n}\sum_{i=1}^{m} a_{ij} = \sum_{j=1}^{n}(a_{1j} + a_{2j} + \cdots + a_{mj}) \\ &= (a_{11} + a_{21} + \cdots + a_{m1}) + (a_{12} + a_{22} + \cdots + a_{m2}) \\ &\quad + \cdots + (a_{1n} + a_{2n} + \cdots + a_{mn}). \end{aligned} \tag{1.2.21}$$

The calculations (1.2.20) and (1.2.21) give the same result since they are two ways to sum the same $(m \times n)$ elements below:

$$\begin{array}{cccccc} a_{11} & a_{12} & \cdots & a_{1n} & \rightarrow & \text{RS1} \\ a_{21} & a_{22} & \cdots & a_{2n} & \rightarrow & \text{RS2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & \rightarrow & \text{RS}m \\ \downarrow & \downarrow & \cdots & \downarrow & \vdots & \downarrow \\ \text{CS1} & \text{CS2} & \cdots & \text{CS}n & \rightarrow & S \end{array} \tag{1.2.22}$$

where the sums in the parentheses of (1.2.20) are row sums RSi and those in (1.2.21) are column sums CSj.

Moreover, some finite double series can be written as a product of series:

$$S = \sum_{i=1}^{m} \sum_{j=1}^{n} a_i b_j = \sum_{i=1}^{m} a_i \sum_{j=1}^{n} b_j. \tag{1.2.23}$$

Proof:

$$\begin{aligned} S &= \sum_{i=1}^{m} \sum_{j=1}^{n} a_i b_j \\ &= (a_1 b_1 + a_1 b_2 + \cdots + a_1 b_n) + (a_2 b_1 + a_2 b_2 + \cdots + a_2 b_n) \\ &\quad + \cdots + (a_m b_1 + a_m b_2 + \cdots + a_m b_n) \\ &= (a_1 + a_2 + \cdots + a_m)\, b_1 + (a_1 + a_2 + \cdots + a_m)\, b_2 + \cdots + (a_1 + a_2 + \cdots + a_m)\, b_n \\ &= \sum_{i=1}^{m} a_i\, (b_1 + b_2 + \cdots + b_n) = \sum_{i=1}^{m} a_i \sum_{j=1}^{n} b_j. \end{aligned}$$

□

Changing the appearance of a sum in terms of the Σ symbol

In practice we often need to change the appearance of a sum for the purpose of simplifying a sum or making the terms more clear to recognize. Let us show this in the light of the following example.

✎ ***Example* 1.9.** Let $\sum_{n=0}^{\infty} a_n x^n$ be a series. Suppose it satisfies

$$\sum_{n=0}^{\infty} a_n n x^{n-1} = \sum_{n=0}^{\infty} a_n x^n. \tag{1.2.24}$$

Show that the coefficients of the terms satisfy

$$a_n = \frac{1}{n} a_{n-1} \quad \text{for } n \geq 1.$$

Solution: To obtain the relationship between two coefficients we need to compare the corresponding terms, i.e., the terms with same exponential. From (1.2.24) this comparison cannot be done immediately, since the generic terms of the two sides, $a_n n x^{n-1}$ and $a_n x^n$, are not in the same exponential power. Hence we need to make a change to the LHS. Introducing another subscript m to replace n, with $m = n - 1$, i.e., $n = m + 1$,

$$\sum_{n=0}^{\infty} a_n n x^{n-1} = \sum_{m=-1}^{\infty} a_{m+1}\,(m+1)\, x^m = a_0 0 x^{-1} + \sum_{m=0}^{\infty} a_{m+1}\,(m+1)\, x^m.$$

Here the first term vanishes due to the 0, and the second term can be rewritten as

$$\sum_{n=0}^{\infty} a_{n+1}\left(n+1\right)x^n,$$

since m is a dummy subscript which has no essential meaning and can be safely replaced by n. Now we are at the point to compare the both sides of (1.2.24):

$$\sum_{n=0}^{\infty} a_{n+1}\left(n+1\right)x^n = \sum_{n=0}^{\infty} a_n x^n,$$

where the coefficients of the generic terms should be equal, $a_{n+1}\left(n+1\right) = a_n$, which is just

$$a_{n+1} = \frac{1}{(n+1)}a_n \text{ for } n \geq 0, \quad \text{i.e.,} \qquad a_n = \frac{1}{n}a_{n-1} \text{ for } n \geq 1.$$

□

✍ ***Exercises* 1.2.** Given that $\sum\limits_{k=1}^{17} a_k = 31$ and $\sum\limits_{k=1}^{17} b_k = 18$, evaluate $\sum\limits_{k=1}^{17}\left(3a_k - 2b_k\right)$.

§ 1.3 Mathematical induction

Mathematical induction（数学归纳法）is a very useful method of proving statements (often formulae) involving natural numbers. It is claimed that the Italian scientist Francesco Maurolico was the first European to use mathematical induction to provide rigorous proofs, whereas the English logician Augustus De Morgan (1806-1873) coined the phrase "mathematical induction" in the early nineteenth century.

The principle of mathematical induction as follows.

***Definition* 1.17.** Let $P\left(n\right)$ represent a statement relative to a positive integer n. If

1. $P\left(n_0\right)$ is true, where n_0 is the smallest integer for which the statement can be made, and
2. whenever $P\left(n\right)$ is true, it follows that $P\left(n+1\right)$ must also be true, then $P\left(n\right)$ is true for all $n \geq n_0$.

The assumption in the second step above — $P\left(n\right)$ is true — is called the *inductive hypothesis*.

The strategy of making an effective use of the induction is to try to express the statement $P\left(n+1\right)$ in terms of the statement $P\left(n\right)$ in a reasonable way.

✎ ***Example* 1.10.** Prove by mathematical induction on n that

$$\left(1-\frac{1}{2^2}\right)\times\left(1-\frac{1}{3^2}\right)\times\cdots\times\left(1-\frac{1}{n^2}\right) = \frac{n+1}{2n}, \qquad \text{where } n \geq 2, n \in \mathbb{Z}.$$

Proof:

Let $P\left(n\right)$ be the statement relative to n,

$$P\left(n\right):\prod_{i=2}^{n}\left(1-\frac{1}{i^2}\right)=\frac{n+1}{2n}.$$

1. Check if $P\left(2\right)$ is true: $P\left(2\right):\left(1-\frac{1}{2^2}\right)=\frac{2+1}{2\times2}$. Done.

2. The inductive hypothesis reads:

$$P\left(n\right):\prod_{i=2}^{n}\left(1-\frac{1}{i^2}\right)=\frac{n+1}{2n}.$$

Now check if $P\left(n+1\right)$ is true:

$$\begin{aligned}P\left(n+1\right)\quad &:\quad \prod_{i=2}^{n+1}\left(1-\frac{1}{i^2}\right)=\left[1-\frac{1}{\left(n+1\right)^2}\right]\prod_{i=2}^{n}\left(1-\frac{1}{i^2}\right)=\left[1-\frac{1}{\left(n+1\right)^2}\right]\frac{n+1}{2n}\\ &=\quad \frac{n+1}{2n}-\frac{1}{\left(n+1\right)^2}\frac{n+1}{2n}=\frac{n+1}{2n}-\frac{1}{n+1}\frac{1}{2n}=\frac{n\left(n+2\right)}{2n\left(n+1\right)}=\frac{\left(n+1\right)+1}{2\left(n+1\right)}.\end{aligned}$$

Done.

□

✎ ***Example* 1.11.** Prove by mathematical induction on n that

$$\frac{1}{1\times3}+\frac{1}{3\times5}+\cdots+\frac{1}{\left(2n-1\right)\times\left(2n+1\right)}=\frac{n}{2n+1}$$

where n is a positive integer.

Proof:

Let $P\left(n\right)$ be the statement relative to n,

$$P\left(n\right):\sum_{i=1}^{n}\frac{1}{\left(2i-1\right)\times\left(2i+1\right)}=\frac{n}{2n+1}.$$

1. Check if $P\left(1\right)$ is true:

$$P\left(1\right):\frac{1}{1\times3}=\frac{1}{2\times1+1}.$$

Done.

2. The inductive hypothesis reads:

$$P\left(n\right):\sum_{i=1}^{n}\frac{1}{\left(2i-1\right)\times\left(2i+1\right)}=\frac{n}{2n+1}.$$

Now check if $P\left(n+1\right)$ is true:

$$\begin{aligned}P\left(n+1\right)\quad &: \quad \sum_{i=1}^{n+1}\frac{1}{\left(2i-1\right)\times\left(2i+1\right)}\\ &= \quad \frac{1}{\left[2\left(n+1\right)-1\right]\times\left[2\left(n+1\right)+1\right]}+\sum_{i=1}^{n}\frac{1}{\left(2i-1\right)\times\left(2i+1\right)}\\ &= \quad \frac{1}{4\left(n+1\right)^{2}-1}+\frac{n}{2n+1}=\frac{2n^{2}+3n+1}{\left(2n+1\right)\left(2n+3\right)}=\frac{n+1}{2\left(n+1\right)+1}.\end{aligned}$$

Done.

☐

✎ ***Example* 1.12.** Prove by mathematical induction on n that the following statement is true:

A set of n elements has 2^n subsets.

Proof:

Let $P\left(n\right)$ be the statement relative to n:

$P\left(n\right)$: A set of n elements has 2^n subsets.

1. Check if $P\left(1\right)$ is true: A set of a unique element alway has two subsets, $\varnothing$ and itself. Hence the statement is true. Done.

2. The inductive hypothesis $P\left(n\right)$ reads: A set of n elements, denoted by $S=\{a_1,a_2,\cdots,a_n\}$, has 2^n subsets.

   Now check if $P\left(n+1\right)$ is true: Let $S=\{a_1,a_2,\cdots,a_n,a_{n+1}\}$ be a set of $n+1$ elements. The key observation is that every n-element subset $\hat{S}=\{a_1,a_2,\cdots,a_n\}$ is automatically a subset of S, hence this accounts for 2^n subsets of S by the inductive hypothesis. Furthermore, if R is any subset of $\hat{S}$, then $R\cup\{a_{n+1}\}$ is a subset of S. This gives us another 2^n subsets of S, for a total of $2^n+2^n=2^{n+1}$ subsets so far. In fact, we now have **all** the subsets of S because every subset of S either contains a_{n+1} or not. Therefore, $P\left(n+1\right)$ is true for all $n\in\mathbb{Z}^+$.

☐

✍ ***Exercises* 1.3.** Prove by mathematical induction that $n!>3^n$ for all integers $n\geq 7$.

Chapter 2 Introduction to vectors

We begin our investigation of linear algebra by studying vectors, the basic components of the theory and applications of linear algebra. The word vector has its origin in physics where it is used to denote a quantity having both magnitude and direction, such as force, velocity, etc..

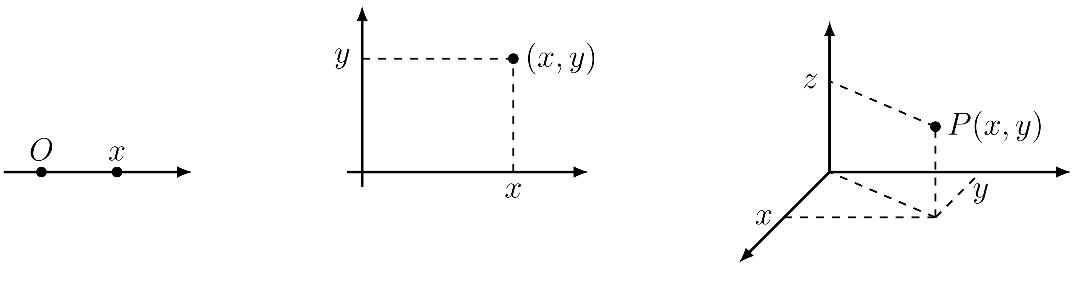
§ 2.1 Notations and basic concepts

§ 2.1.1 Points in space

A point on a line (1-dimensional space) can be represented by a single number, when the space is endowed with coordinates (i.e., a unit length is selected on the line).

A point in a plane (2-dimensional space) can be represented by a pair of numbers, when the space is endowed with coordinates.

A point in a 3-dimensional space can be represented by a triple of numbers, when the space is endowed with coordinates.



(a) (b) (c)

Figure 2.1: (a) Point on a line; (b) Point in a plane; (c) Point in a 3-dimensional space.

Generically, a point P in n-dimensional space can be represented by an n-tuple of numbers, $(x_1, x_2, \cdots, x_n)$, where the numbers $x_i, i = 1, \cdots, n$, are called the coordinates of P.

§ 2.1.2 Vectors

***Definition* 2.1.** A geometric vector（矢量）is a quantity that has both magnitude （大小）and direction（方向）.

Mathematically, a vector can be represented graphically in an n-dimensional space by a directed line segment or arrow that has its tail at one point and the head at a second point:

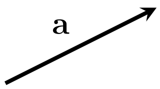
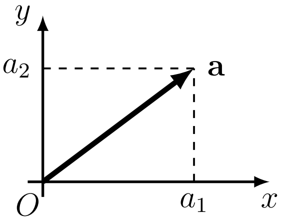


Figure 2.2: Vector.

Two vectors are equal if they have the same magnitude and direction, regardless of their position in space.[*1]

We use a lowercase bold letter (say, $\mathbf{a}$) or a lowercase letter with vector symbol (say, $\vec{a}$) to denote a vector, as shown in Figure 2.2. The magnitude of the vector is denoted by $|\mathbf{a}|$, while its direction is expressed by $\hat{\mathbf{a}}$ or $\frac{\mathbf{a}}{|\mathbf{a}|}$ (see §2.4). When the space is endowed with coordinates, the vector is expressed by an ordered finite list of real numbers, denoted by a column（列）$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$, where each number $a_i, i = 1, \cdots, n$, is called a component（分量）of the vector.



where $\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$.

Figure 2.3: Components of a vector.

It is useful to introduce the concept of *transpose*（转置）. For a column $\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$, its transpose,

---

[*1] Pay full attention to the difference between geometric vectors and physical vectors. Geometrically, two vectors having the same magnitude and direction are treated as same vectors, regardless of their initial positions in space. But physically, two vectors which are geometrically same but have different initial points are treated as different two vectors. Sample physical vectors include forces, velocities, momentum, etc..

denoted by $\mathbf{a}^T$, is a row（行）

$$\mathbf{a}^T = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}; \tag{2.1.1}$$

conversely, the column can also be regarded as the transpose of the row:

$$\mathbf{a} = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}^T = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}. \tag{2.1.2}$$

A vector can also be expressed by its endpoints. For example, in Fig.2.4 the vector $\mathbf{a}$ is expressed by its endpoints A and B as $\overrightarrow{AB}$, where A is called the initial point and B the terminal point. The components of $\mathbf{a}$ can be read from the coordinates of the endpoints, by $\mathbf{a} = \begin{pmatrix} 4-1 & 3-1 \end{pmatrix}^T = \begin{pmatrix} 3 & 2 \end{pmatrix}^T$.

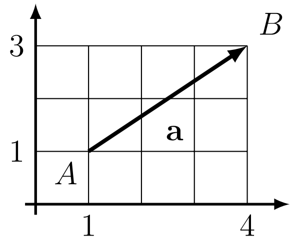


Figure 2.4: Vector expressed by its endpoints.

***Theorem* 2.1.** Two vectors $\overrightarrow{AB}$ and $\overrightarrow{CD}$ are equal, if and only if their components are equal.

Proof: Apparent. □

✎ ***Example* 2.1.** Let $A = (2, -1)$,$B = (4, 5)$,$C = (3, 0)$ and $D = (5, 6)$. Show that $\overrightarrow{AB} = \overrightarrow{CD}$.

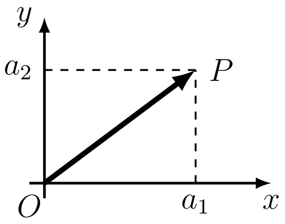
Solution:

Examining the x-components: $\left(\overrightarrow{AB}\right)_x = 4 - 2 = 2,\quad \left(\overrightarrow{CD}\right)_x = 5 - 3 = 2;$

Examining the y-components: $\left(\overrightarrow{AB}\right)_y = 5 - (-1) = 6,\quad \left(\overrightarrow{CD}\right)_y = 6 - 0 = 6.$

□

Specially, for a point P in space, its *position vector*（位置矢量）is given by $\overrightarrow{OP}$ where O denotes the origin of the space:

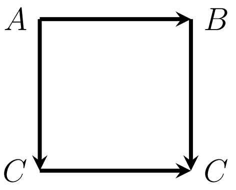


where the position vector of the point $P\,(a_1, a_2)$ is $\overrightarrow{OP} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$.

Figure 2.5: Position vector of a point.

✍ ***Exercises* 2.1.**

1. The edges of the square $\square ABCD$ are marked by vectors $\overrightarrow{AB}$, $\overrightarrow{AD}$, $\overrightarrow{BC}$ and $\overrightarrow{DC}$, as shown.



True or false:

(i) $\overrightarrow{AB} = \overrightarrow{BC}$; (ii) $\overrightarrow{AB} = \overrightarrow{CD}$; (iii) $\overrightarrow{AD} = \overrightarrow{BC}$.

2. If $|\mathbf{v}| = 2$, find $|\mathbf{u}|$ in each of the following cases.

(i) $\mathbf{u} = 3\mathbf{v}$; (ii) $\mathbf{u} = \frac{1}{2}\mathbf{v}$; (iii) $\mathbf{u} = -\mathbf{v}$; (iv) $\mathbf{v} = 3\mathbf{u}$.

§ 2.2 Addition and subtraction of vectors

§ 2.2.1 Vector addition

The addition of two vectors is conducted by means of the *parallelogram law* or the *triangle law*:

- Parallelogram law（平行四边形法则）: Let $\mathbf{a}$ and $\mathbf{b}$ be two vectors. Put their initial points at a same point O, such that $\mathbf{a}$ and $\mathbf{b}$ form the two adjacent sides of a parallelogram. Then the sum of $\mathbf{a}$ and $\mathbf{b}$, $\mathbf{a} + \mathbf{b}$, is represented both in magnitude and direction by the diagonal vector $\overrightarrow{OP}$ of the parallelogram. See Fig.2.6(a).

- Triangle law（三角形法则）: $\mathbf{a} + \mathbf{b}$ can also be obtained by placing the initial point of $\mathbf{b}$ at the terminal point $\mathbf{a}$ and connecting the very endpoints O and P. See Fig.2.6(b).

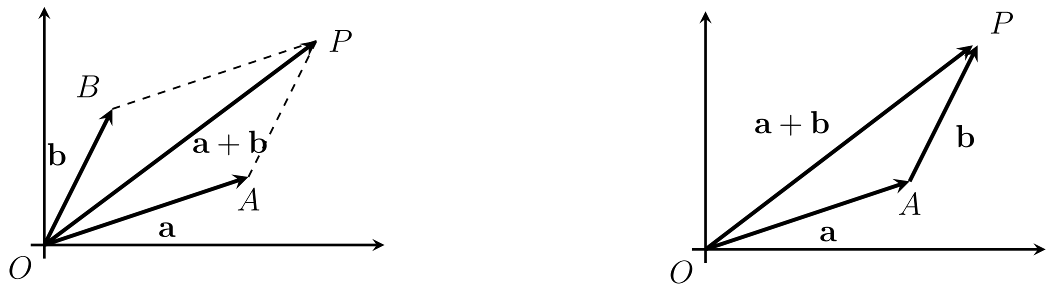


Figure 2.6: (*Left*) Vector addition using the parallelogram law: $\mathbf{a}+\mathbf{b}$ represents the sum of $\mathbf{a}$ and $\mathbf{b}$, which is obtained by placing the initial points of $\mathbf{a}$ and $\mathbf{b}$ at the same point O such that $\mathbf{a}+\mathbf{b}$ is given by the diagonal $\vec{OP}$; (*Right*) Vector addition using the triangle law: $\mathbf{a}+\mathbf{b}$ represents the sum of $\mathbf{a}$ and $\mathbf{b}$, which is obtained by placing the initial point of $\mathbf{b}$ at the terminal point $\mathbf{a}$, such that $\mathbf{a}+\mathbf{b}$ is given by the connection between the endpoints O and P.

In components, letting $\mathbf{a}$ and $\mathbf{b}$ be $\mathbf{a} = \begin{pmatrix} a_1 & a_2 \end{pmatrix}^T$ and $\mathbf{b} = \begin{pmatrix} b_1 & b_2 \end{pmatrix}^T$ respectively, their sum $\mathbf{a}+\mathbf{b}$ reads

$$\mathbf{a}+\mathbf{b} = \begin{pmatrix} a_1+b_1 & a_2+b_2 \end{pmatrix}^T. \tag{2.2.1}$$

✎ ***Example* 2.2.** Let $\mathbf{a} = \begin{pmatrix} 3 & 2 \end{pmatrix}^T$ and $\mathbf{b} = \begin{pmatrix} 5 & -1 \end{pmatrix}^T$. Their sum reads

$$\mathbf{a}+\mathbf{b} = \begin{pmatrix} 3+5 & 2+(-1) \end{pmatrix}^T = \begin{pmatrix} 8 & 1 \end{pmatrix}^T.$$

Vector addition has the following properties:

- Commutative law（交换律）:

$$\mathbf{a}+\mathbf{b} = \mathbf{b}+\mathbf{a}; \tag{2.2.2}$$

- Associative law（结合律）:

$$(\mathbf{a}+\mathbf{b})+\mathbf{c} = \mathbf{a}+(\mathbf{b}+\mathbf{c})\,. \tag{2.2.3}$$

☞ **[Aside]:**

The sum of three and more vectors can be illustrated by means of the triangle law. See Fig.2.7.

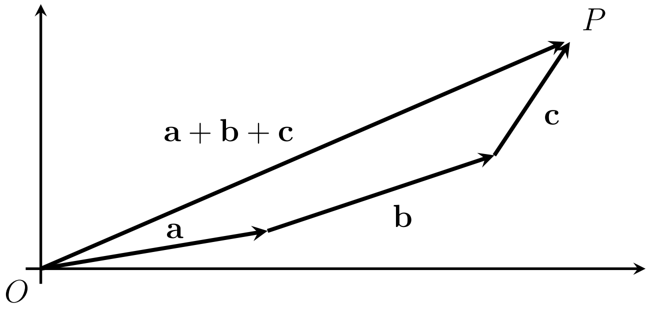


Figure 2.7: Addition of three vectors, $\mathbf{a}+\mathbf{b}+\mathbf{c}$, using the triangle law.

✎ ***Example* 2.3.** Let $\mathbf{a} = \begin{pmatrix} 3 & 2 \end{pmatrix}^T$, $\mathbf{b} = \begin{pmatrix} 5 & -1 \end{pmatrix}^T$ and $\mathbf{c} = \begin{pmatrix} -3 & 2 \end{pmatrix}^T$. Their sum reads

$$\mathbf{a}+\mathbf{b}+\mathbf{c} = \begin{pmatrix} 3+5+(-3) & 2+(-1)+2 \end{pmatrix}^T = \begin{pmatrix} 5 & 3 \end{pmatrix}^T.$$

✌

§ 2.2.2 Vector subtraction

The subtraction of two vectors is conducted as illustrated in Fig.2.8.

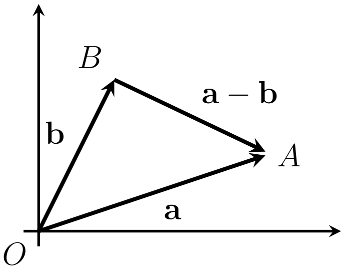


Figure 2.8: Vector subtraction: The difference between the vectors $\mathbf{a}$ and $\mathbf{b}$, $\mathbf{a}-\mathbf{b}$, is given by a diagonal of the parallelogram formed by $\mathbf{a}$ and $\mathbf{b}$.

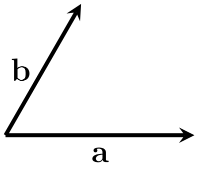
In components, letting $\mathbf{a}$ and $\mathbf{b}$ be $\mathbf{a} = \begin{pmatrix} a_1 & a_2 \end{pmatrix}^T$ and $\mathbf{b} = \begin{pmatrix} b_1 & b_2 \end{pmatrix}^T$ respectively, their difference $\mathbf{a}-\mathbf{b}$ reads

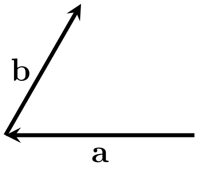
$$\mathbf{a}+\mathbf{b} = \begin{pmatrix} a_1-b_1 & a_2-b_2 \end{pmatrix}^T. \tag{2.2.4}$$

✎ ***Example* 2.4.** Let $\mathbf{a} = \begin{pmatrix} 3 & 2 \end{pmatrix}^T$ and $\mathbf{b} = \begin{pmatrix} 5 & -1 \end{pmatrix}^T$. Their difference reads

$$\mathbf{a}-\mathbf{b} = \begin{pmatrix} 3-5 & 2-(-1) \end{pmatrix}^T = \begin{pmatrix} -2 & 3 \end{pmatrix}^T.$$

✍ ***Exercises* 2.2.** Draw the vectors $\mathbf{a}+\mathbf{b}$ and $\mathbf{a}-\mathbf{b}$ on each diagram.





§ 2.3 Scalar multiplication and properties of vectors

§ 2.3.1 Scalar multiplication

Vectors have three types of multiplications:

- *scalar multiplication*（数乘，或标量乘积）;
- *dot product* （点乘）, also called *inner product*（内积）or *scalar product*（标量积）;
- *cross product*（叉乘）, also called *outer product*（外积）or vector product （矢量积）. [*2]

In this section we will introduce the first one; the last two will be introduced respectively in §§3.1 and 3.2.

***Definition* 2.2.** A scalar（标量）is a number. In this course a scalar is a real or complex number. Examples of scalars in science and technology include quantities like mass, energy, etc..

***Definition* 2.3.** Let k be a real number and $\mathbf{a}$ be an n-dimensional vector, $\mathbf{a} = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix}^T$. The scalar multiplication of $\mathbf{a}$ by k is defined by

$$k\mathbf{a} = k \begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix}^T = \begin{pmatrix} ka_1 & ka_2 & \cdots & ka_n \end{pmatrix}^T . \tag{2.3.1}$$

Notice: Scalar multiplication is to multiply a vector by a number, and the result is still a vector.

✎ ***Example* 2.5.** In physics, mass is a scalar and velocity is a vector. Suppose an object has a mass $m = 3$kg, and its velocity is $\mathbf{v} = \begin{pmatrix} 2 & 1 & 3 \end{pmatrix}^T$ m/s. Then the momentum（动量）of the object is given by

$$\mathbf{P} = m\mathbf{v} = 3 \times \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 6 \\ 3 \\ 9 \end{pmatrix} \quad (\text{kg} \cdot \text{m/s}) .$$

§ 2.3.2 Properties of vectors

- *Zero vector*:

  The zero vector $\mathbf{0}$ has zero length and points in every direction. For every vector $\mathbf{a}$,

  $$\mathbf{0} + \mathbf{a} = \mathbf{a}, \qquad \mathbf{a} - \mathbf{a} = \mathbf{a} + (-1)\mathbf{a} = \mathbf{0}. \tag{2.3.2}$$

  [Remark]:

  Be alert to the difference between 0 and $\mathbf{0}$. The former is a number, but the latter a vector.

[*2] The name "inner product" contrasts with the "outer product". The former takes a pair of vectors as the input and produces **a scalar as the output**; the latter takes a pair of vectors as the input but still produces **a vector as the output**. The names scalar product and vector product arise from this.

- *Negative vector*:

  The negative of $\mathbf{a}$ is defined as

  $$-\mathbf{a} = (-1)\mathbf{a}, \tag{2.3.3}$$

  which has the same length as $\mathbf{a}$, but pointing in the opposite direction.

  If P and Q are points then $\overrightarrow{PQ} = -\overrightarrow{QP}$.

- If $\mathbf{a}$ and $\mathbf{b}$ are vectors and λ and μ are scalars, then

$$\begin{aligned} &\lambda(\mu\mathbf{a}) = (\lambda\mu)\mathbf{a}, && \lambda(\mathbf{a}+\mathbf{b}) = \lambda\mathbf{a}+\lambda\mathbf{b}, && (\lambda+\mu)\mathbf{a} = \lambda\mathbf{a}+\mu\mathbf{a}; \\ &-(-\mathbf{a}) = \mathbf{a}, && 1\mathbf{a} = \mathbf{a}, && (-\lambda)\mathbf{a} = -(\lambda\mathbf{a}). \end{aligned} \tag{2.3.4}$$

✎ ***Example* 2.6.** Let two vectors be $\mathbf{a} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$ and $\mathbf{b} = \begin{pmatrix} 5 \\ -1 \end{pmatrix}$ and two numbers be $\lambda = -2$ and $\mu = 4$. Then,

$$\begin{aligned} &-2 \times \left[4\begin{pmatrix} 3 \\ 2 \end{pmatrix}\right] = (-2\times 4)\begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} -24 \\ -16 \end{pmatrix}, \\ &-2\left[\begin{pmatrix} 3 \\ 2 \end{pmatrix} + \begin{pmatrix} 5 \\ -1 \end{pmatrix}\right] = -2\begin{pmatrix} 3 \\ 2 \end{pmatrix} - 2\begin{pmatrix} 5 \\ -1 \end{pmatrix} = \begin{pmatrix} -16 \\ -2 \end{pmatrix}, \\ &(-2+4)\begin{pmatrix} 3 \\ 2 \end{pmatrix} = -2\begin{pmatrix} 3 \\ 2 \end{pmatrix} + 4\begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 6 \\ 4 \end{pmatrix}. \end{aligned}$$

✍ ***Exercises* 2.3.**

1. Simplify the following vector expressions.

   (i) $3\mathbf{a} + 2\mathbf{b} - 4\left(\mathbf{b} + \frac{1}{2}\mathbf{a}\right)$;   (ii) $-(\mathbf{w} - 6\mathbf{z}) - 2\mathbf{w} + \mathbf{v} - 2\mathbf{z}$.

2. If $|\mathbf{v}| = 2$, find $|\mathbf{u}|$ in each of the following cases.

   (i) $\mathbf{u} = 3\mathbf{v}$;   (ii) $\mathbf{u} = \frac{1}{2}\mathbf{v}$;   (iii) $\mathbf{u} = -3\mathbf{v}$;   (iii) $\mathbf{v} = -3\mathbf{u}$.

§ 2.4 Some important definitions and properties

***Definition* 2.4.** Let $\mathbf{a}$ be a vector. The *norm*（模长）of $\mathbf{a}$ is defined as its magnitude $|\mathbf{a}|$, which is a scalar.

***Definition* 2.5.** The *direction* of $\mathbf{a}$, denoted as $\hat{\mathrm{a}}$ (called "hat of $\mathbf{a}$" or simply "$\mathbf{a}$-hat"), is defined as the ratio $\frac{\mathbf{a}}{|\mathbf{a}|}$, where $|\mathbf{a}| \neq 0$.

Thus a vector $\mathbf{a}$ can be written as

$$\mathbf{a} = |\mathbf{a}|\, \hat{\mathbf{a}}. \tag{2.4.1}$$

This captures precisely the idea of $\mathbf{a}$ being characterized by its length and direction.

The norm of the zero vector **0** is 0, and its direction is indeterminate （不定的）.

Definition **2.6.** Two nonzero vectors $\mathbf{a}$ and $\mathbf{b}$ are *parallel vectors* （平行矢量）if and only if $\mathbf{a} = \lambda\mathbf{b}$ for some nonzero scalar λ.

If $\lambda > 0$, $\mathbf{a}$ and $\mathbf{b}$ are in the same direction; if $\lambda < 0$, $\mathbf{a}$ and $\mathbf{b}$ are in the opposite direction.

Definition **2.7.** A *unit vector*（单位矢量）is a vector of magnitude/norm 1.

A unit vector is usually used to represent a direction. The direction of a vector above, $\hat{\mathbf{a}}$ is a unit vector.

Definition **2.8.** Suppose a 3-dimensional space is endowed with the x-, y- and z-axes, which are mutually orthogonal（两两正交）. Usually we use $\mathbf{i}, \mathbf{j}$ and $\mathbf{k}$ to be the 3 unit vectors pointing in the positive x-, y- and z-axes.

Definition **2.9.** A vector $\mathbf{a} = \begin{pmatrix} a_1 & a_2 & a_3 \end{pmatrix}^T$, with the x-, y- and z-components being a_1, a_2 and a_3 respectively, can be expressed as

$$\mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}. \tag{2.4.2}$$

This is called the *Cartesian form* of the vector $\mathbf{a}$（笛卡尔形式或正交形式）. See Fig.2.9 below.

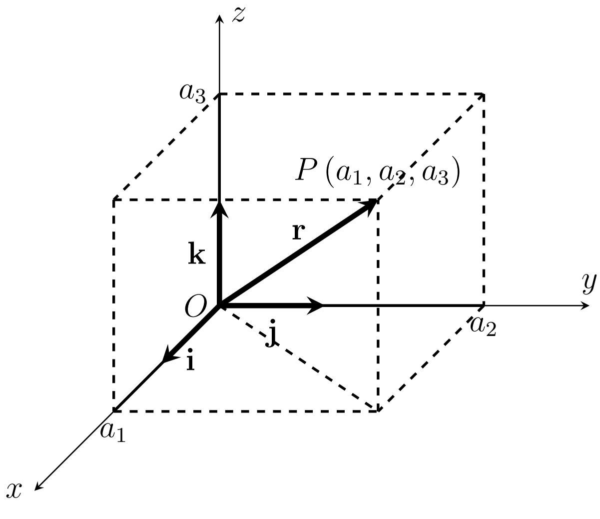


Figure 2.9: Projection onto coordinate axes.

Definition **2.10.** Let $\mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$. The norm/length of $\mathbf{a}$ is given by

$$|\mathbf{a}| = \sqrt{a_1^2 + a_2^2 + a_3^2}, \tag{2.4.3}$$

in the light of the Pythagoras theorem（毕达哥拉斯定理，中国称勾股定理）. In terms of $|\mathbf{a}|$, the unit vector $\hat{\mathbf{a}}$ is obtained by

$$\hat{\mathbf{a}} = \frac{\mathbf{a}}{|\mathbf{a}|} = \frac{1}{\sqrt{a_1^2 + a_2^2 + a_3^2}} \left(a_1 \mathbf{i} + a_2 \mathbf{j} + a_3 \mathbf{k} \right). \tag{2.4.4}$$

A position vector $\overrightarrow{OP} = \begin{pmatrix} a & b & c \end{pmatrix}^T$ can be written as

$$\overrightarrow{OP} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}. \tag{2.4.5}$$

And now we are able to conduct component-wise operations to vectors. Let $\mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$ and $\mathbf{b} = b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}$ be two vectors, and λ a scalar. To add, subtract or negate the vectors, we simply add, subtract or negate their respective components:

$$\mathbf{a} \pm \mathbf{b} \;=\; (a_1 \pm b_1)\,\mathbf{i} + (a_2 \pm b_2)\,\mathbf{j} + (a_3 \pm b_3)\,\mathbf{k}, \tag{2.4.6}$$

$$-\mathbf{a} \;=\; -a_1\mathbf{i} - a_2\mathbf{j} - a_3\mathbf{k}. \tag{2.4.7}$$

To multiply a vector by a scalar, we simply multiply each component by the scalar:

$$\lambda\mathbf{a} = \lambda a_1\mathbf{i} + \lambda a_2\mathbf{j} + \lambda a_3\mathbf{k}. \tag{2.4.8}$$

From (2.4.8) it is easy to prove that

$$|\lambda\mathbf{a}| = |\lambda|\,|\mathbf{a}|\,. \tag{2.4.9}$$

***Definition* 2.11.** Let $P\,(a_1, b_1, c_1)$ and $Q\,(a_2, b_2, c_2)$ be two points in space. Then

$$\overrightarrow{PQ} = (a_2 - a_1)\,\mathbf{i} + (b_2 - b_1)\,\mathbf{j} + (c_2 - c_1)\,\mathbf{k}. \tag{2.4.10}$$

The distance between the two points P and Q is defined as the length of the vector $\overrightarrow{PQ}$:

$$\left|\overrightarrow{PQ}\right| = \sqrt{(a_2 - a_1)^2 + (b_2 - b_1)^2 + (c_2 - c_1)^2}. \tag{2.4.11}$$

§ 2.4.1 Linear independence

***Definition* 2.12.** In the n-dimensional space $\mathbb{R}^n$, a set of k nonzero vectors, $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_k$ $(k \leq n)$, are called linearly independent（线性独立）if and only if the vector equation

$$\lambda_1\mathbf{a}_1 + \lambda_2\mathbf{a}_2 + \cdots + \lambda_k\mathbf{a}_k = 0, \qquad \lambda_1, \lambda_2, \cdots, \lambda_k \in \mathbb{R}, \tag{2.4.12}$$

holds when the scalar coefficients $\lambda_1 = \lambda_2 = \cdots = \lambda_k = 0$. Otherwise, the vectors $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_k$ are called linearly dependent（线性依赖）.

In $\mathbb{R}^2$, the linear independence of two nonzero vectors $\mathbf{a}$ and $\mathbf{b}$

$$\lambda\mathbf{a} + \rho\mathbf{b} = 0, \qquad \lambda, \rho \in \mathbb{R}, \tag{2.4.13}$$

implies $\mathbf{a}$ and $\mathbf{b}$ are not parallel.

✎ ***Example* 2.7.** Consider three vectors $\mathbf{a}_1 = \mathbf{i} + 2\mathbf{j} - \mathbf{k}, \mathbf{a}_2 = -\mathbf{i} + \mathbf{j}$ and $\mathbf{a}_3 = \mathbf{i} + 3\mathbf{j} - \mathbf{k}$ in $\mathbb{R}^3$. To determine the linear independence or dependence we have to investigate the solutions of $\lambda_1, \lambda_2, \lambda_3$ to the equation

$$\begin{aligned} \lambda_1\mathbf{a}_1 + \lambda_2\mathbf{a}_2 + \lambda_3\mathbf{a}_3 &= \lambda_1(\mathbf{i} + 2\mathbf{j} - \mathbf{k}) + \lambda_2(-\mathbf{i} + \mathbf{j}) + \lambda_3(\mathbf{i} + 3\mathbf{j} - \mathbf{k}) \\ &= (\lambda_1 - \lambda_2 + \lambda_3)\mathbf{i} + (2\lambda_1 + \lambda_2 + 3\lambda_3)\mathbf{j} + (-\lambda_1 - \lambda_3)\mathbf{k} = 0. \end{aligned}$$

The coefficients of $\mathbf{i}, \mathbf{j}$ and $\mathbf{k}$ should vanish separately, hence

$$\begin{cases} \lambda_1 - \lambda_2 + \lambda_3 = 0, \\ 2\lambda_1 + \lambda_2 + 3\lambda_3 = 0, \\ -\lambda_1 - \lambda_3 = 0. \end{cases}$$

The solutions are $\lambda_1 = \lambda_2 = \lambda_3 = 0$, meaning that $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ are linearly independent.

***Theorem* 2.2.** Let $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_k$ be a set of k nonzero vectors in the n-dimensional space $\mathbb{R}^n,\ k \leq n$. Then, $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_k$ are linearly independent if and only if any one of this set, $\mathbf{a}_p$ say, cannot be expressed by a linear combination of the other vectors, $\mathbf{a}_1, \cdots, \mathbf{a}_{p-1}, \mathbf{a}_{p+1}, \cdots, \mathbf{a}_k$.

This theorem is very important. In this regard, when we say a set of vectors are linearly independent, we are equivalently saying that any vector cannot be expressed by a linear combination of the other vectors.

Proof:

1. Sufficiency: Suppose $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_k$ are a set of linearly independent nonzero vectors, namely, $\lambda_1\mathbf{a}_1 + \lambda_2\mathbf{a}_2 + \cdots + \lambda_k\mathbf{a}_k = 0$ yields $\lambda_1 = \cdots = \lambda_k = 0$. To prove the sufficiency, we assume conversely that there exists a vector $\mathbf{a}_p$ that can be linearly expressed by the other vectors,

   $$\mathbf{a}_p = \rho_1\mathbf{a}_1 + \cdots + \rho_{p-1}\mathbf{a}_{p-1} + \rho_{p+1}\mathbf{a}_{p+1} + \cdots + \rho_k\mathbf{a}_k, \tag{2.4.14}$$

   where there exists at least one $\lambda_l \neq 0,\ l = 1, \cdots, p-1, p+1, \cdots, k$. Then rearranging the terms of (2.4.14) leads to

   $$\rho_1\mathbf{a}_1 + \cdots + \rho_{p-1}\mathbf{a}_{p-1} + \rho_{p+1}\mathbf{a}_{p+1} + \cdots + \rho_k\mathbf{a}_k - \mathbf{a}_p = 0,$$

   which contradicts the above given condition that all the coefficients should vanish, $\lambda_1 = \cdots = \lambda_k = 0$. Hence the above assumption does not hold.

2. Necessity: Suppose $\lambda_1\mathbf{a}_1 + \lambda_2\mathbf{a}_2 + \cdots + \lambda_k\mathbf{a}_k = 0$ implies $\lambda_1 = \cdots = \lambda_k = 0$. To prove that any one of $\mathbf{a}_1, \cdots, \mathbf{a}_k$ cannot be able to be expressed by other vectors of the set, we conversely assume

there exist at least two $\lambda_l, \lambda_p \neq 0,\ l = 1, \cdots, k$.[*3] Thus rearranging the terms we have

$$-\lambda_l \mathbf{a}_l = \cdots + \lambda_p \mathbf{a}_p + \cdots, \qquad \text{thus} \qquad \mathbf{a}_l = \cdots - \frac{\lambda_p}{\lambda_l}\mathbf{a}_p + \cdots .$$

There exists at least one coefficient $-\frac{\lambda_p}{\lambda_l} \neq 0$, which means $\mathbf{a}_l$ can be linearly expressed by the other vectors $\mathbf{a}_1, \cdots, \mathbf{a}_{l-1}, \mathbf{a}_{l+1}, \cdots, \mathbf{a}_k$. This contradicts the requirement that any $\mathbf{a}_l$ cannot be linearly expressed by other vectors. Hence the above assumption does not hold.

□

✍ ***Exercises* 2.4.**

1. Let P be the point $(3, 1, -2)$ and Q the point $(4, -2, 5)$ in space. The origin $(0, 0, 0)$ is denoted by O. Let $\mathbf{i}, \mathbf{j}$ and $\mathbf{k}$ be, as usual, the unit vectors in the positive x-, y- and z-directions, respectively.

   (i) Write down the position vectors $\overrightarrow{OP}$ and $\overrightarrow{OQ}$ in terms of $\mathbf{i}, \mathbf{j}$ and $\mathbf{k}$.

   (ii) Write down the displacement vector $\overrightarrow{PQ}$ in terms of $\mathbf{i}, \mathbf{j}$ and $\mathbf{k}$.

   (iii) Write down the coordinates of the point R such that $\overrightarrow{OR} = \overrightarrow{PQ}$.

2. Given that

   $$\mathbf{a} = 2\mathbf{i} - \mathbf{j} + 2\mathbf{k}, \mathbf{b} = \mathbf{i} + \mathbf{j} - \mathbf{k}, \mathbf{c} = 3\mathbf{i} - 4\mathbf{k},$$

   find

   (i) $\mathbf{a} + \mathbf{b}$; (ii) $\mathbf{a} + 3\mathbf{b} - 2\mathbf{c}$; (iii) $|\mathbf{a}|$; (iv) $\hat{\mathbf{a}}$; (v) $\hat{\mathbf{c}}$.

[*3] If there is only one coefficient $\lambda_l \neq 0$, i.e., $\lambda_1 \mathbf{a}_1 + \cdots + \lambda_k \mathbf{a}_k = \lambda_l \mathbf{a}_l = 0$, then the coefficient λ_l must be zero, since $\mathbf{a}_l$ is nonzero. This contradicts the precondition $\lambda_l \neq 0$. Hence it should be supposed that at least two coefficients $\lambda_l, \lambda_p \neq 0$.

Chapter 3 Dot product and cross product

§ 3.1 Dot product

Dot product is also called inner product or scalar product. “Inner” and “scalar” refer to the fact that the output of a pair of input vectors is a scalar, which is thought to be a *degenerated*（退化的）vector in a sense. This is in contrast with the outer product (also known as cross product) of §3.2, where the output of a pair of input vectors produces another vector.

§ 3.1.1 Algebraic definition of dot product

***Definition* 3.1.** Let $\mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$ and $\mathbf{b} = b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}$ be two vectors in 3 dimensions. The inner product of $\mathbf{a}$ and $\mathbf{b}$ is defined as

$$\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2 + a_3 b_3. \tag{3.1.1}$$

Notice:

- The output of a dot product is a scalar.
- In 2 dimensions, the inner product of $\mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j}$ and $\mathbf{b} = b_1\mathbf{i} + b_2\mathbf{j}$ is defined as

$$\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2. \tag{3.1.2}$$

Properties:

1. Zero vector:

$$\mathbf{0} \cdot \mathbf{a} = 0. \tag{3.1.3}$$

2. Unit vectors:

$$\mathbf{i} \cdot \mathbf{j} = \mathbf{j} \cdot \mathbf{k} = \mathbf{k} \cdot \mathbf{i} = 0, \qquad \mathbf{i} \cdot \mathbf{i} = \mathbf{j} \cdot \mathbf{j} = \mathbf{k} \cdot \mathbf{k} = 1. \tag{3.1.4}$$

3. Commutativity:

$$\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}. \tag{3.1.5}$$

Proof:

$$\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2 + a_3 b_3 = b_1 a_1 + b_2 a_2 + b_3 a_3 = \mathbf{b} \cdot \mathbf{a}. \tag{3.1.6}$$

□

4. Distributivity:

$$\mathbf{a}\cdot(\mathbf{b}+\mathbf{c})=\mathbf{a}\cdot\mathbf{b}+\mathbf{a}\cdot\mathbf{c}, \qquad (\mathbf{b}+\mathbf{c})\cdot\mathbf{a}=\mathbf{b}\cdot\mathbf{a}+\mathbf{c}\cdot\mathbf{a}. \tag{3.1.7}$$

Proof:

$$\begin{aligned}(\mathbf{b}+\mathbf{c})\cdot\mathbf{a} &= (b_1+c_1)\,a_1+(b_2+c_2)\,a_2+(b_3+c_3)\,a_3\\ &= b_1a_1+c_1a_1+b_2a_2+c_2a_2+b_3a_3+c_3a_3=\mathbf{b}\cdot\mathbf{a}+\mathbf{c}\cdot\mathbf{a}.\end{aligned}$$

□

✎ ***Example* 3.1.** Let $\mathbf{a}=2\mathbf{i}+3\mathbf{j}$ and $\mathbf{b}=-\mathbf{i}+4\mathbf{j}$. Thus

$$\mathbf{a}\cdot\mathbf{b}=2\times(-1)+3\times 4=10.$$

✎ ***Example* 3.2.** Let $\mathbf{a}=-\frac{1}{2}\mathbf{i}+\frac{\sqrt{3}}{2}\mathbf{j}$ and $\mathbf{b}=\mathbf{i}$. Thus

$$\mathbf{a}\cdot\mathbf{b}=\left(-\frac{1}{2}\right)\times 1+\frac{\sqrt{3}}{2}\times 0=-\frac{1}{2}.$$

✎ ***Example* 3.3.** The inner product of two nonzero vectors could be zero.

For example, let $\mathbf{a}=\mathbf{i}-2\mathbf{j}+2\mathbf{k}$ and $\mathbf{b}=2\mathbf{i}+5\mathbf{j}+4\mathbf{k}$. Thus

$$\mathbf{a}\cdot\mathbf{b}=1\times 2+(-2)\times 5+2\times 4=0.$$

***Definition* 3.2.** The norm of a vector $\mathbf{a}$ can also be defined as

$$|\mathbf{a}|=\sqrt{\mathbf{a}\cdot\mathbf{a}}=\sqrt{a_1a_1+a_2a_2+a_3a_3}. \tag{3.1.8}$$

This agrees to the definition of norm in Definition 2.10.

✎ ***Example* 3.4.** Suppose the vector

$$a=\begin{pmatrix} 1235, & 985, & 1050 \end{pmatrix}^T$$

holds four prices expressed in Euros, British pounds and Australian dollars, respectively. On a particular day, we know that the exchange rates of currency（汇率）are

$$\begin{aligned}\text{1 Euro} &= \$1.46420, \quad \text{1 British pound} = \$1.83637,\\ \text{1 Australian dollar} &= \$0.83580.\end{aligned}$$

How can we use vectors to find the total of the four prices in US dollars?

Arithmetically, we can calculate the total US dollars as follows:

$$\text{Total} = 1235 \times 1.46420 + 985 \times 1.83637 + 1050 \times 0.83580 = \$4494.70.$$

We can interprete this answer as the result of combining two vectors — one holding the original prices and the other carrying the currence conversion rates — in a way that multiplies the vectors' corresponding components and then adds the resulting products:

$$\begin{aligned} \text{Total} &= \overbrace{(1235\mathbf{i} + 985\mathbf{j} + 1050\mathbf{k})}^{\text{Price vector}} \cdot \overbrace{(1.46420\mathbf{i} + 1.83637\mathbf{j} + 0.83580\mathbf{k})}^{\text{Exchange rate vector}} \\ &= 1235 \times 1.46420 + 985 \times 1.83637 + 1050 \times 0.83580 = \underbrace{\$4494.70}_{\text{Total price (a scalar)}}. \end{aligned}$$

§ **3.1.2 Geometric definition of dot product**

***Definition* 3.3.** Let $\mathbf{a}$ and $\mathbf{b}$ be two vectors and θ the angle between them. The dot product of $\mathbf{a}$ and $\mathbf{b}$ is defined as

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}|\,|\mathbf{b}| \cos\theta. \tag{3.1.9}$$

This geometric definition is as illustrated in Fig.3.1.

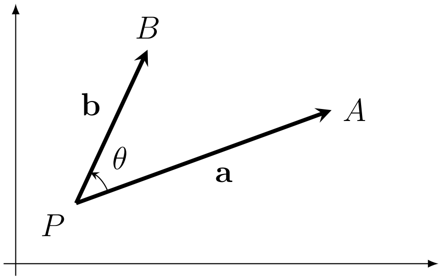


Figure 3.1: Geometric definition of inner product.

✎ ***Example* 3.5.** In the polar coordinate system, there are two vectors $\mathbf{a} = \left(2, \frac{\pi}{6}\right) = 2\left(\cos\frac{\pi}{6}\mathbf{i} + \sin\frac{\pi}{6}\mathbf{j}\right)$ and $\mathbf{b} = \left(3, \frac{\pi}{3}\right) = 3\left(\cos\frac{\pi}{3}\mathbf{i} + \sin\frac{\pi}{3}\mathbf{j}\right)$. The dot product between $\mathbf{a}$ and $\mathbf{b}$ is

$$\mathbf{a} \cdot \mathbf{b} = 2 \times 3 \cos\left(\frac{\pi}{3} - \frac{\pi}{6}\right) = 6 \cos\frac{\pi}{6} = 3\sqrt{3}.$$

☞ **[Aside]:**

The algebraic and geometric definitions of dot product are equivalent. This can be demonstrated as follows.

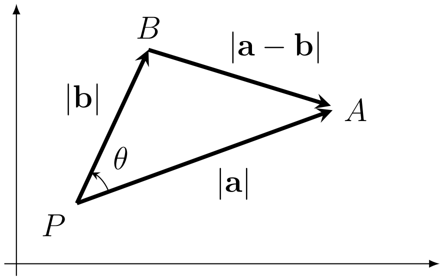


Figure 3.2: Proof of equivalence between algebraic and geometric definitions of dot product.

Let $\mathbf{a}$ and $\mathbf{b}$ be two vectors in 2 dimensions, $\mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j}$ and $\mathbf{b} = b_1\mathbf{i} + b_2\mathbf{j}$, and θ the angle between $\mathbf{a}$ and $\mathbf{b}$, as shown in Fig.3.2. Let $a = |\mathbf{a}|$, $b = |\mathbf{b}|$ and $c = |\mathbf{a}-\mathbf{b}|$. On the one hand, the algebraic definition reads

$$\mathbf{a}\cdot\mathbf{b} = a_1b_1 + a_2b_2.$$

On the other hand, the geometric definition reads $\mathbf{a}\cdot\mathbf{b} = ab\cos\theta$. Using the cosine rule in trigonometry we have

$$\begin{aligned} & 2ab\cos\theta = a^2 + b^2 - c^2 \\ = \; & \left(a_1^2 + a_2^2\right) + \left(b_1^2 + b_2^2\right) - \left[(b_1 - a_1)^2 + (b_2 - a_2)^2\right] \\ = \; & 2a_1b_1 + 2a_2b_2, \end{aligned}$$

$$\text{i.e.,} \qquad \mathbf{a}\cdot\mathbf{b}_{\,\text{geometric}} = ab\cos\theta = a_1b_1 + a_2b_2 = \mathbf{a}\cdot\mathbf{b}_{\,\text{algebraic}}\,.$$

✌

***Definition* 3.4.** The intersection angle θ between $\mathbf{a}$ and $\mathbf{b}$ can be computed by

$$\cos\theta = \frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{a}|\,|\mathbf{b}|}, \tag{3.1.10}$$

where $\mathbf{a}\cdot\mathbf{b}$ is computed by means of the algebraic definition of $\mathbf{a}\cdot\mathbf{b}$.

✎ ***Example* 3.6.** Let us re-consider Example3.2 by means of the geometric definition of dot product. The angle between $\mathbf{a} = -\frac{1}{2}\mathbf{i} + \frac{\sqrt{3}}{2}\mathbf{j}$ and $\mathbf{b} = \mathbf{i}$ is

$$\cos\theta = \frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{a}|\,|\mathbf{b}|} = \frac{\left(-\frac{1}{2}\mathbf{i} + \frac{\sqrt{3}}{2}\mathbf{j}\right)\cdot\mathbf{i}}{\sqrt{\left(-\frac{1}{2}\right)^2 + \left(\frac{\sqrt{3}}{2}\right)^2}\times 1} = -\frac{1}{2}, \qquad \text{hence} \qquad \theta = \frac{2\pi}{3}.$$

There are some special cases:

- The above definition of norm, $|\mathbf{a}|^2 = \mathbf{a}\cdot\mathbf{a}$, is a special case of intersection angle $\theta = 0$, since a vector $\mathbf{a}$ is parallel to itself.
- The inner product of two vectors is zero means their intersection angle is $\frac{\pi}{2}$, i.e., they are orthogonal to each other.

✎ ***Example* 3.7.** Let us re-consider the above Example3.3, where $\mathbf{a} = \mathbf{i} - 2\mathbf{j} + 2\mathbf{k}$ and $\mathbf{b} = 3\mathbf{i} + 5\mathbf{j} + 4\mathbf{k}$. The angle between them is given by

$$\cos\theta = \frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{a}|\,|\mathbf{b}|} = 0, \qquad \text{hence} \qquad \theta = \frac{\pi}{2}.$$

From the above property (3.1.4) one can see that the mutual intersection angles between the unit vectors $\mathbf{i}, \mathbf{j}$ and $\mathbf{k}$ are $\frac{\pi}{2}$, hence the x-, y- and z-axes form an orthonormal frame（正交标架）.

✎ ***Example* 3.8.** Inner product has wide applications in science and technology. For instance, in physics the quantity *work*（功）is defined as

$$w = \mathbf{f}\cdot\mathbf{s}, \tag{3.1.11}$$

where $\mathbf{f}$ is a force（力）applied on an object and $\mathbf{s}$ the displacement（位移）of the object. w, which is a scalar, is the work of $\mathbf{f}$ along $\mathbf{s}$, as show in Fig.3.3.

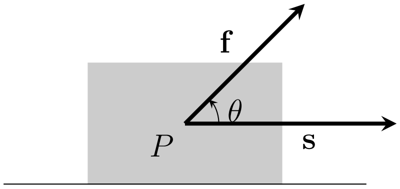


Figure 3.3: Application of inner product in physics — definition of *work*: A force $\mathbf{f}$ is exerted upon a body at the point P. The displacement of the body is $\mathbf{s}$. Then the work of $\mathbf{f}$ along $\mathbf{s}$ is defined as $w = \mathbf{f}\cdot\mathbf{s}$.

§ 3.1.3 Projection of vector

Let $\mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j} + a_1\mathbf{k}$ be a vector. a_1, a_2 and a_3 can be regarded as the inner products between $\mathbf{a}$ and the unit vectors of the x-, y- and z-axis, $\mathbf{i}, \mathbf{j}$ and $\mathbf{k}$, respectively:

$$a_1 = \mathbf{a}\cdot\mathbf{i} = |\mathbf{a}|\cos\alpha, \qquad a_2 = \mathbf{a}\cdot\mathbf{j} = |\mathbf{a}|\cos\beta, \qquad a_3 = \mathbf{a}\cdot\mathbf{k} = |\mathbf{a}|\cos\gamma, \tag{3.1.12}$$

where α, β and γ are the intersection angles between $\mathbf{a}$ and the x-, y- and z-axis, as shown in Fig.3.4. We call a_1 the scalar component（标量分量，或简称分量）of $\mathbf{a}$ onto the $\mathbf{i}$ direction, and $a_1\mathbf{i}$ the corresponding *vector projection*（矢量投影，或简称投影）. Similarly, a_2 the scalar component onto the $\mathbf{j}$ direction and $a_2\mathbf{j}$ the corresponding projection, and a_3 that onto the $\mathbf{k}$ direction, and $a_3\mathbf{k}$ the corresponding projection.

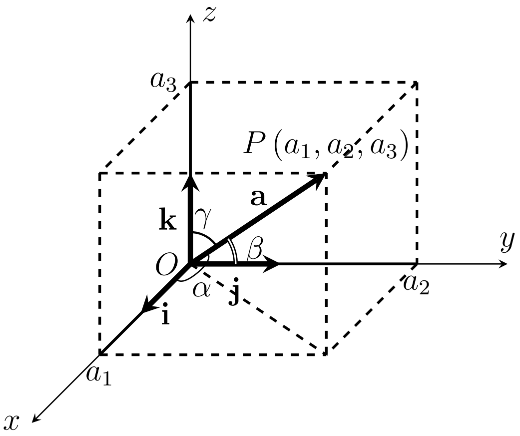


Figure 3.4: Projection of a vector onto coordinate axes: For a vector $\mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j} + a_1\mathbf{k}$, its intersection angle with the x-axis is denoted as α, that to the y-axis is β, and that to the z-axis is γ. The scalar components of $\mathbf{a}$ are $a_1 = |\mathbf{a}|\cos\alpha, a_2 = |\mathbf{a}|\cos\beta, a_3 = |\mathbf{a}|\cos\gamma$. The corresponding vector projections are $a_1\mathbf{i}, a_2\mathbf{j}$ and $a_3\mathbf{k}$, respectively.

Generally speaking, we have the following definitions for scalar component, vector projection and vector component of a vector orthogonal to another vector.

***Definition* 3.5.** The scalar component of a vector $\mathbf{a}$ onto another vector $\mathbf{b}$ is defined as

$$a_{\|} = \mathbf{a} \cdot \hat{\mathbf{b}} = \mathbf{a} \cdot \frac{\mathbf{b}}{|\mathbf{b}|}, \tag{3.1.13}$$

where $\hat{\mathbf{b}} = \frac{\mathbf{b}}{|\mathbf{b}|}$ is the unit vector of $\mathbf{b}$ (see Fig.3.5). The corresponding vector projection is defined as

$$a_{\|}\hat{\mathbf{b}} = \left(\mathbf{a} \cdot \frac{\mathbf{b}}{|\mathbf{b}|}\right) \frac{\mathbf{b}}{|\mathbf{b}|} = \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{b}|^2}\mathbf{b}. \tag{3.1.14}$$

***Definition* 3.6.** In 2 dimensions, the vector component of $\mathbf{a}$ orthogonal to $\mathbf{b}$ is defined as the difference between $\mathbf{a}$ and its vector projection onto $\mathbf{b}$ (see Fig.3.5):

$$a_{\perp}\hat{\mathbf{b}}_{\perp} = \mathbf{a} - a_{\|}\hat{\mathbf{b}} = \mathbf{a} - \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{b}|^2}\mathbf{b}, \tag{3.1.15}$$

where $\hat{\mathbf{b}}_{\perp}$ is the unit direction orthogonal to $\mathbf{b}$.

It is easy to check that $a_{\perp}\hat{\mathbf{b}}_{\perp}$ is perpendicular（垂直于）to $\mathbf{b}$:

$$\left(a_{\perp}\hat{\mathbf{b}}_{\perp}\right) \cdot \mathbf{b} = \mathbf{a} \cdot \mathbf{b} - \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{b}|^2}|\mathbf{b}|^2 = 0.$$

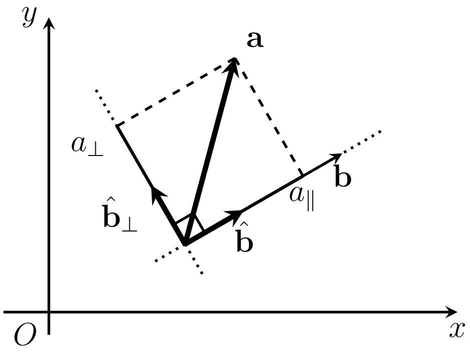


Figure 3.5: Projection of a vector onto another vector: Let $\mathbf{a}$ and $\mathbf{b}$ be two vectors in 2 dimensions. *The direction of* $\mathbf{b}$ is represented by the unit vector $\hat{\mathbf{b}}$, and *the direction orthogonal to* $\mathbf{b}$ is represented by the unit vector $\hat{\mathbf{b}}_\perp$. The scalar component of $\mathbf{a}$ onto $\mathbf{b}$ is denoted by $a_\|$, with the corresponding vector projection being $a_\|\hat{\mathbf{b}}$. The vector component of $\mathbf{a}$ orthogonal to $\mathbf{b}$ is denoted by $a_\perp\hat{\mathbf{b}}_\perp$.

✎ ***Example* 3.9.** In physics, the work $w = \mathbf{f} \cdot \mathbf{s} = |\mathbf{f}|\,|\mathbf{s}|\cos\theta$ can be thought of as $|\mathbf{s}|$ times the scalar component of $\mathbf{f}$ onto $\mathbf{s}$:

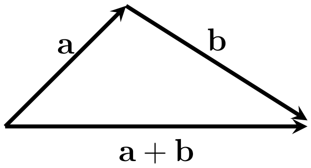
$$w = \mathbf{f} \cdot \mathbf{s} = f_\|\,|\mathbf{s}| = (|\mathbf{f}|\cos\theta)\,|\mathbf{s}|\,. \tag{3.1.16}$$

§ 3.1.4 Two important inequalities

In terms of the dot product we can prove the following two important inequalities （不等式）.

1. Edges of triangle:

$$|\mathbf{a} + \mathbf{b}| \le |\mathbf{a}| + |\mathbf{b}|\,. \tag{3.1.17}$$



Proof:

First, both $|\mathbf{a} + \mathbf{b}|$ and $|\mathbf{a}| + |\mathbf{b}|$ are positive numbers. Second, let the angle between $\mathbf{a}$ and $\mathbf{b}$ be θ. Then

$$\begin{aligned} |\mathbf{a} + \mathbf{b}|^2 &= (\mathbf{a} + \mathbf{b}) \cdot (\mathbf{a} + \mathbf{b}) = |\mathbf{a}|^2 + |\mathbf{b}|^2 + 2\mathbf{a} \cdot \mathbf{b}, \\ (|\mathbf{a}| + |\mathbf{b}|)^2 &= |\mathbf{a}|^2 + |\mathbf{b}|^2 + 2\,|\mathbf{a}|\,|\mathbf{b}|\,. \\ |\mathbf{a} + \mathbf{b}|^2 - (|\mathbf{a}| + |\mathbf{b}|)^2 &= 2\,(\mathbf{a} \cdot \mathbf{b} - |\mathbf{a}|\,|\mathbf{b}|) = 2\,|\mathbf{a}|\,|\mathbf{b}|\,(\cos\theta - 1) \le 0. \end{aligned}$$

Hence

$$|\mathbf{a} + \mathbf{b}| \le |\mathbf{a}| + |\mathbf{b}|\,.$$

□

2. Cauchy-Schwarz inequality（柯西-施瓦茨不等式）:

$$|\mathbf{a}\cdot\mathbf{b}| \leq |\mathbf{a}|\,|\mathbf{b}|\,. \tag{3.1.18}$$

Proof:

$$|\mathbf{a}\cdot\mathbf{b}| = |\mathbf{a}|\,|\mathbf{b}|\,|\cos\theta|\,. \qquad \text{Since } |\cos\theta| \leq 1, \quad \text{there is } |\mathbf{a}\cdot\mathbf{b}| \leq |\mathbf{a}|\,|\mathbf{b}|\,.$$

□

✍ ***Exercises* 3.1.**

1. Given that

$$\mathbf{u} = 2\mathbf{i} - \mathbf{j} + \mathbf{k}, \mathbf{v} = \mathbf{i} - 2\mathbf{j} + 2\mathbf{k}, \mathbf{w} = 3\mathbf{i} - \mathbf{k},$$

find

(i) $\mathbf{u}\cdot\mathbf{v}$ (ii) $\mathbf{u}\cdot\mathbf{w}$ (iii) $\mathbf{v}\cdot\mathbf{w}$ (iv) $\mathbf{u}\cdot\mathbf{u}$ (v) $\mathbf{v}\cdot\mathbf{v}$ (vi) $\mathbf{w}\cdot\mathbf{w}$

(vii) $|\mathbf{u}|$ (viii) $|\mathbf{v}|$ (ix) $|\mathbf{w}|$ (x) $\mathbf{u}\cdot(\mathbf{v}+\mathbf{w})$ (xi) $\mathbf{u}\cdot(\mathbf{v}-\mathbf{w})$

2. Let $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ be as in the previous exercise. Let α be the angle between $\mathbf{u}$ and $\mathbf{v}$, β be the angle between $\mathbf{u}$ and $\mathbf{w}$, and γ the angle between $\mathbf{v}$ and $\mathbf{w}$. Find

(i) $\cos\alpha$; (ii) $\cos\beta$; (iii) $\cos\gamma$.

3. Given that

$$\mathbf{a} = 2\mathbf{i} - \mathbf{j} + 2\mathbf{k}, \mathbf{b} = \mathbf{i} + \mathbf{j} - \mathbf{k}, \mathbf{c} = 3\mathbf{i} + 6\mathbf{j},$$

determine whether the following are true or false:

(i) The angle between $\mathbf{a}$ and $\mathbf{b}$ is acute（锐角）.

(ii) The angle between $\mathbf{b}$ and $\mathbf{c}$ is acute.

(iii) The vectors $\mathbf{a}$ and $\mathbf{c}$ are mutually perpendicular.

(iv) The angle between the vectors $\mathbf{a}+\mathbf{b}$ and $\mathbf{b}-\mathbf{c}$ is obtuse（钝角）.

§ 3.2 Cross product

Cross product, also called outer product and vector product, is another multiplication of vectors in 3 dimensions. Similar as dot product, it also has both algebraic and geometric definitions.

§ 3.2.1 Algebraic definition of cross product

***Definition* 3.7.** Let $\mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$ and $\mathbf{b} = b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}$ be two vectors in 3 dimensions. The cross product of them is defined as

$$\mathbf{a} \times \mathbf{b} = (a_2b_3 - a_3b_2)\,\mathbf{i} + (a_3b_1 - a_1b_3)\,\mathbf{j} + (a_1b_2 - a_2b_1)\,\mathbf{k}. \tag{3.2.1}$$

In terms of the language of determinants（行列式）(see Chapter 2 of General Reference Textbook [3]), (3.2.1) can be expressed as

$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} = \mathbf{i}\begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} - \mathbf{j}\begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} + \mathbf{k}\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}. \tag{3.2.2}$$

Cross product has the following properties: Let $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ be vectors in 3 dimensions, and λ a scalar.

1. Anti-commutativity（反对易性）:

$$\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}. \tag{3.2.3}$$

Notice: *Pay full attention to this unusual property!*

A natural corollary is

$$\mathbf{a} \times \mathbf{a} = 0. \tag{3.2.4}$$

2. Distributivity:

$$\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}. \tag{3.2.5}$$

3. Scalar multiplication:

$$\lambda\,(\mathbf{a} \times \mathbf{b}) = (\lambda\mathbf{a}) \times \mathbf{b} = \mathbf{a} \times (\lambda\mathbf{b})\,. \tag{3.2.6}$$

The unit vectors of the x-, y- and z-axes, $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$, form a right handed set（右手系）, as shown in Fig.3.6.

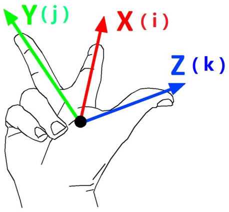


Figure 3.6: Right handed set formed by the unit vectors $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$.

Noticing that $|\mathbf{i}| = |\mathbf{j}| = |\mathbf{k}| = 1$ and they are mutually orthogonal, their cross products are given by

$$\mathbf{i} \times \mathbf{i} = \mathbf{j} \times \mathbf{j} = \mathbf{k} \times \mathbf{k} = \mathbf{0}; \tag{3.2.7}$$

$$\mathbf{i} \times \mathbf{j} = -\mathbf{j} \times \mathbf{i} = \mathbf{k}, \qquad \mathbf{j} \times \mathbf{k} = -\mathbf{k} \times \mathbf{j} = \mathbf{i}, \qquad \mathbf{k} \times \mathbf{i} = -\mathbf{i} \times \mathbf{k} = \mathbf{j}. \tag{3.2.8}$$

In terms of (3.2.7) and (3.2.8) one can easily prove (3.2.1):

Proof:

$$\begin{aligned}
\mathbf{a} \times \mathbf{b} &= (a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}) \times (b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}) \\
&= a_1b_1\mathbf{i} \times \mathbf{i} + a_1b_2\mathbf{i} \times \mathbf{j} + a_1b_3\mathbf{i} \times \mathbf{k} + a_2b_1\mathbf{j} \times \mathbf{i} + a_2b_2\mathbf{j} \times \mathbf{j} + a_2b_3\mathbf{j} \times \mathbf{k} \\
&\quad + a_3b_1\mathbf{k} \times \mathbf{i} + a_3b_2\mathbf{k} \times \mathbf{j} + a_3b_3\mathbf{k} \times \mathbf{k} \\
&= a_1b_2\mathbf{k} - a_1b_3\mathbf{j} - a_2b_1\mathbf{k} + a_2b_3\mathbf{i} + a_3b_1\mathbf{j} - a_3b_2\mathbf{i} \\
&= (a_2b_3 - a_3b_2)\,\mathbf{i} + (a_3b_1 - a_1b_3)\,\mathbf{j} + (a_1b_2 - a_2b_1)\,\mathbf{k}.
\end{aligned}$$

□

✎ ***Example* 3.10.**

$$\begin{aligned}
& (2\mathbf{i} + 3\mathbf{j} + 4\mathbf{k}) \times (5\mathbf{i} + 6\mathbf{j} + 7\mathbf{k}) \\
= \; & (3 \times 7 - 4 \times 6)\,\mathbf{i} + (4 \times 5 - 2 \times 7)\,\mathbf{j} + (2 \times 6 - 3 \times 5)\,\mathbf{k} = -3\mathbf{i} + 6\mathbf{j} - 3\mathbf{k}.
\end{aligned}$$

§ 3.2.2 Geometric definition of cross product

***Definition* 3.8.** Let $\mathbf{a}$ and $\mathbf{b}$ be two vectors in 3 dimensions, and the angle between them be θ. Their cross product reads

$$\mathbf{a} \times \mathbf{b} = |\mathbf{a}|\,|\mathbf{b}| \sin\theta\, \hat{\mathbf{c}}, \tag{3.2.9}$$

where the unit vector $\hat{\mathbf{c}}$, denoting the direction of $\mathbf{a} \times \mathbf{b}$, points upwards and is perpendicular to the plane which contains $\mathbf{a}$ and $\mathbf{b}$, as shown in Fig.3.7.

Notice: When deciding the direction of $\mathbf{a} \times \mathbf{b}$, we are using the *Right-hand rule*（右手定则）.

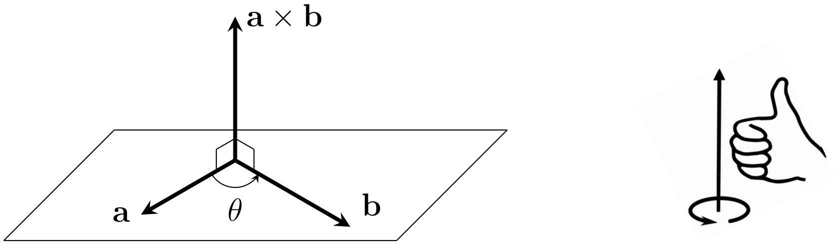


Figure 3.7: Geometric definition of cross product. *Left*: directions of $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{a} \times \mathbf{b}$. *Right*: the so-called right-hand rule used in deciding the direction of $\mathbf{a} \times \mathbf{b}$. The right-hand rule is also known as the right-hand grip rule and the corkscrew-rule. It says that when you wrap your right hand by rotating the fingers from the vector $\mathbf{a}$ to the vector $\mathbf{b}$, your thumb points in the direction of $\mathbf{a} \times \mathbf{b}$.

✎ ***Example* 3.11.** Suppose a 3-dimensional space is endowed with the cylindrical coordinates（柱坐标）(ρ, θ, z). Let $\mathbf{a}$ and $\mathbf{b}$ be two vectors in the $z = 0$ plane,

$$\mathbf{a} = \left(6, \frac{\pi}{3}, 0\right), \mathbf{b} = \left(4, \frac{\pi}{6}, 0\right).$$

Then their cross product is given by

$$\mathbf{a} \times \mathbf{b} = \left(0, 0, 4 \times 6 \sin\left(\frac{\pi}{6} - \frac{\pi}{3}\right)\right) = (0, 0, -12). \tag{3.2.10}$$

☞ **[Aside]:**

The geometric and algebraic definitions of cross product are equivalent.

Proof:

The complete proof for the generic case is tedious. The reader is encouraged to accomplish it by him/herself. In following we only examine a special case for the purpose of demonstrating the key step in the proof.

Let $\mathbf{a}$ and $\mathbf{b}$ be two vectors in the $z = 0$ plane in 3 dimensions,

$$\mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j}, \qquad \mathbf{b} = b_1\mathbf{i} + b_2\mathbf{j},$$

and the angle between them is denoted by θ, as shown in Fig.3.8.

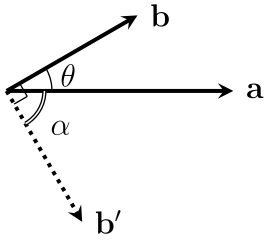


Figure 3.8: Proof of equivalence between geometric and algebraic definitions of cross product: $\mathbf{a}$ and $\mathbf{b}$ are two vectors and θ the angle in between them. $\mathbf{b}'$ is an introduced auxiliary vector. Let the component form of $\mathbf{b}$ be $\mathbf{b} = b_1\mathbf{i} + b_2\mathbf{j}$; then $\mathbf{b}'$ reads $\mathbf{b}' = b_2\mathbf{i} - b_1\mathbf{j}$. The $\mathbf{b}'$ has the same norm as $\mathbf{b}$, but is orthogonal to $\mathbf{b}$. The angle in between $\mathbf{a}$ and $\mathbf{b}'$ is α, complementary to θ.

- Algebraic definition: (3.2.1) gives

$$\mathbf{a} \times \mathbf{b} = (a_1 b_2 - a_2 b_1)\,\mathbf{k}, \qquad \text{noticing } a_3 = b_3 = 0. \tag{3.2.11}$$

- Geometric definition: The direction of $\mathbf{a} \times \mathbf{b}$ is undoubtedly the one towards the positive z-axis direction represented by $\mathbf{k}$. As far as the magnitude of $\mathbf{a} \times \mathbf{b}$ is concerned, we have

$$|\mathbf{a} \times \mathbf{b}| = |\mathbf{a}|\,|\mathbf{b}| \sin\theta. \tag{3.2.12}$$

In order to prove that (3.2.12) may reproduce the magnitude of (3.2.11), $a_1b_2 - a_2b_1$, we appeal to re-expressing the factor $\sin\theta$. Let us introduce an auxiliary vector $\mathbf{b}'$ as

$$\mathbf{b}' = b_2\mathbf{i} - b_1\mathbf{j}.$$

It is easy to check that $\mathbf{b}'$ has the same norm as $\mathbf{b}$ but is perpendicular to $\mathbf{b}$, as shown in Fig.3.8:

$$\left|\mathbf{b}'\right| = \sqrt{b_1^2 + b_2^2} = \left|\mathbf{b}\right|, \qquad \mathbf{b}' \cdot \mathbf{b} = (b_2\mathbf{i} - b_1\mathbf{j}) \cdot (b_1\mathbf{i} + b_2\mathbf{j}) = 0.$$

Apparently the angle between $\mathbf{b}'$ and $\mathbf{a}$ is α, the complementary angle of θ, which has $\sin\theta = \cos\alpha$. Thus, (3.2.12) turns to be

$$|\mathbf{a} \times \mathbf{b}| = |\mathbf{a}| \left|\mathbf{b}'\right| \cos\alpha = |\mathbf{a} \times \mathbf{b}| = \mathbf{a} \cdot \mathbf{b}'.$$

But on the other hand

$$\mathbf{a} \cdot \mathbf{b}' = (a_1\mathbf{i} + a_2\mathbf{j}) \cdot (b_2\mathbf{i} - b_1\mathbf{j}) = a_1b_2 - a_2b_1,$$

hence

$$|\mathbf{a} \times \mathbf{b}| = a_1b_2 - a_2b_1,$$

which completes the proof.

□

✌

✎ ***Example* 3.12.** Let us re-examine Example 3.11 above computed by using the geometric definition. The vector $\mathbf{a} = \left(6, \frac{\pi}{3}, 0\right)$ in cylindrical coordinates can be rewritten in the Cartesian coordinates（直角坐标）as

$$\mathbf{a} = 6\cos\frac{\pi}{3}\mathbf{i} + 6\sin\frac{\pi}{3}\mathbf{j} = 3\mathbf{i} + 3\sqrt{3}\mathbf{j}.$$

Similarly, $\mathbf{b} = \left(4, \frac{\pi}{3}, 0\right)$ has the Cartesian form

$$\mathbf{b} = 4\cos\frac{\pi}{6}\mathbf{i} + 4\sin\frac{\pi}{6}\mathbf{j} = 2\sqrt{3}\mathbf{i} + 2\mathbf{j}.$$

Therefore, in the light of the algebraic definition,

$$\mathbf{a} \times \mathbf{b} = \left(3 \times 2 - 3\sqrt{3} \times 2\sqrt{3}\right)\mathbf{k} = -12\mathbf{k},$$

in agreement to the result (3.2.10) computed in terms of the geometric definition.

§ 3.2.3 Application of cross product

Areas of parallelogram and triangle

Cross product is easy to use in computing the area of a parallelogram. Let $\square OACB$ be a parallelogram in 2 dimensions, of which two adjacent sides, $\overline{OA}$ and $\overline{OB}$, are denoted by two vectors $\mathbf{a}$ and $\mathbf{b}$, with the intersection angle being θ, as shown in the left of Fig.3.9. Geometrically, the area of the parallelogram can be given by the cross product of $\mathbf{a}$ and $\mathbf{b}$:

$$\begin{aligned} S_{\square OACB} &= \text{base} \cdot \text{height} = \overline{OA} \cdot \overline{BH} \\ &= \overline{OA}\,\overline{OB} \sin\theta = |\mathbf{a} \times \mathbf{b}| \,. \end{aligned} \tag{3.2.13}$$

Figure 3.9: Computations of areas of parallelogram and triangle in terms of cross product: *Left (parallelogram)*: The vectors $\mathbf{a}$ and $\mathbf{b}$ are adjacent sides of a parallelogram. The angle between $\mathbf{a}$ and $\mathbf{b}$ is θ. The area of the parallelogram reads $|\mathbf{a} \times \mathbf{b}|$; *Right (triangle)*: Similarly, the area of a triangle reads $\frac{1}{2} |\mathbf{a} \times \mathbf{b}|$.

Similarly, let ΔOAB be a triangle, and two of its sides are shown by two vectors $\mathbf{a}$ and $\mathbf{b}$ with their intersection angle θ, as shown in the right of Fig.3.9. The area of ΔOAB is given by the cross product of $\mathbf{a}$ and $\mathbf{b}$:

$$S_{\Delta OAB} = \frac{1}{2} |\mathbf{a}| \, |\mathbf{b}| \sin\theta = \frac{1}{2} |\mathbf{a} \times \mathbf{b}| \,. \tag{3.2.14}$$

Cross product has wide applications in physics too, as demonstrated by the following two examples.

Torque

Suppose a particle is located at position $\mathbf{r}$ relative to its axis of rotation. When a force $\mathbf{f}$ is applied to the particle, only the perpendicular component $\mathbf{f}_{\perp}$ produces a *torque*（力矩）. This torque $\tau = \mathbf{r} \times \mathbf{f}$ has magnitude $|\tau| = |\mathbf{r}| \, |\mathbf{f}| \sin\theta$ and is directed outward from the page, as shown in Fig.3.10.

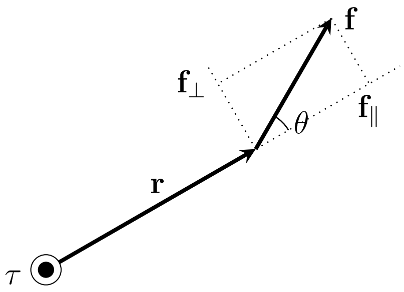


Figure 3.10: Definition of torque in physics.

Lorentz force

Suppose $\mathbf{B}$ is a magnetic field（磁场）pointing into the page, denoted by $\times$. A particle with charge $+q$ moves in this field with instantaneous velocity（瞬时速度）$\mathbf{v}$ perpendicular to $\mathbf{B}$. Then a force, the so-called Lorentz force （洛伦兹力）,

$$\mathbf{f} = q\mathbf{v} \times \mathbf{B} \tag{3.2.15}$$

will act on the particle, so that the particle will move along a circular trajectory（轨迹）, as shown in Fig.3.11.

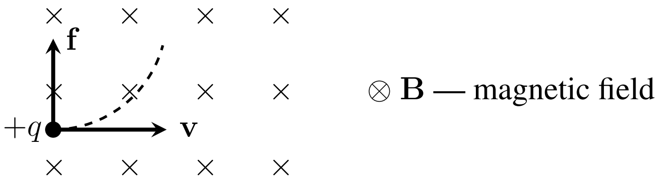


Figure 3.11: Lorentz force in physics.

✍ ***Exercises* 3.2.**

1. Write down

   (i) $\mathbf{i} \times \mathbf{j}$ (ii) $2\mathbf{i} \times 3\mathbf{j}$ (iii) $\mathbf{i} \times (-4\mathbf{j})$ (iv) $\mathbf{j} \times \mathbf{i}$ (v) $\mathbf{j} \times (-4\mathbf{i})$

   (vi) $\mathbf{j} \times \mathbf{k}$ (vii) $\mathbf{k} \times \mathbf{k}$ (viii) $\mathbf{k} \times (-\mathbf{k})$ (ix) $(-\mathbf{k}) \times \mathbf{i}$ (x) $(-\mathbf{k}) \times (-\mathbf{j})$

   (xi) $\mathbf{k} \times (\mathbf{i} + \mathbf{k})$ (xii) $(3\mathbf{j} - \mathbf{k}) \times 2\mathbf{j}$ (xiii) $(\mathbf{j} - \mathbf{k}) \times (\mathbf{k} + \mathbf{j})$

2. Evaluate

   (i) $\mathbf{i} \times (\mathbf{i} + \mathbf{j} + \mathbf{k})$ (ii) $(\mathbf{i} + \mathbf{j} + \mathbf{k}) \times (2\mathbf{i} + \mathbf{k})$

   (iii) $(2\mathbf{i} + 4\mathbf{j} - 3\mathbf{k}) \times (2\mathbf{i} - 3\mathbf{i} + 4\mathbf{k})$ (iv) $(\mathbf{i} - \mathbf{j} + 3\mathbf{k}) \times (3\mathbf{i} + \mathbf{i} - \mathbf{k})$

3. Given that

   $$\mathbf{a} = 2\mathbf{i} - \mathbf{j} + 2\mathbf{k}, \qquad \mathbf{b} = \mathbf{i} + \mathbf{j} - \mathbf{k},$$

   find

   (i) $|\mathbf{a}|$ (ii) $|\mathbf{b}|$ (iii) $\mathbf{a} \times \mathbf{b}$ (iv) $|\mathbf{a} \times \mathbf{b}|$

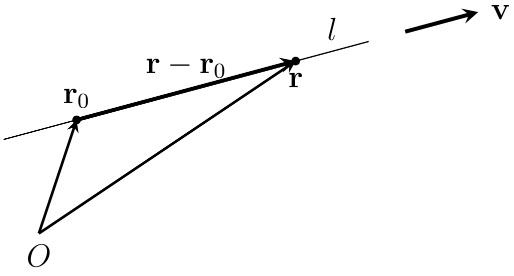
   (v) the sine of the angle between $\mathbf{a}$ and $\mathbf{b}$.

Chapter 4 Applications of vectors

In Chapters 2 and 3 we learnt definitions and properties of vectors, as well as scalar multiplication, dot product and cross product of vectors. In this chapter we will have a chance to see some of their applications in geometry, including equations of lines and planes and scalar triple product in 3 dimensions.

§ 4.1 Equations of straight lines in 3 dimensions

In §2.3.1 we learnt scalar multiplication of vectors, which will be used in this section to derive equations of straight lines. See Fig.4.1.



$$\mathbf{r}_0 = \mathbf{r}_0\,(x_0, y_0, z_0)$$
$$\mathbf{r} = \mathbf{r}\,(x, y, z)$$
$$\mathbf{v} = \mathbf{v}\,(a, b, c)$$

Figure 4.1: A straight line l in 3-dimensions: $\mathbf{r}_0\,(x_0, y_0, z_0)$ is a fixed point on l, and $\mathbf{r}\,(x, y, z)$ a running point on l. $\mathbf{v}\,(a, b, c)$ is a vector parallel to l.

Consider a straight line l in 3-dimensions. Let $\mathbf{r}_0\,(x_0, y_0, z_0)$ be a fixed point on l, and $\mathbf{v}\,(a, b, c)$ a vector parallel to l. Then for any point on l other than $\mathbf{r}_0$, denoted by $\mathbf{r}\,(x, y, z)$, its displacement from $\mathbf{r}_0$ should be proportional to $\mathbf{v}$:

$$\mathbf{r} - \mathbf{r}_0 = t\mathbf{v}, \qquad t \in \mathbb{R}. \tag{4.1.1}$$

Apparently, when the parameter t running out all values in $\mathbb{R}$, the moving point $\mathbf{r}$ travels along the whole line. Hence (4.1.1) gives an equation of the line l. Usually we rewrite (4.1.1) as

$$\mathbf{r} = \mathbf{r}_0 + t\mathbf{v} \qquad (t \in \mathbb{R})\,, \tag{4.1.2}$$

which is called the *(parametric) vector equation* of the straight line （直线的矢量方程）, with t being the parameter.

Given that $\mathbf{r}_0$, $\mathbf{r}$ and $\mathbf{v}$ have component forms,

$$\mathbf{r}_0 = x_0\mathbf{i} + y_0\mathbf{j} + z_0\mathbf{k}, \qquad \mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}, \qquad \mathbf{v} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}, \tag{4.1.3}$$

(4.1.2) reads

$$x\mathbf{i} + y\mathbf{j} + z\mathbf{k} = (x_0 + ta)\,\mathbf{i} + (y_0 + tb)\,\mathbf{j} + (z_0 + tc)\,\mathbf{k}. \tag{4.1.4}$$

This immediately leads to the following component form

$$\begin{cases} x = x_0 + ta, \\ y = y_0 + tb, \\ z = z_0 + tc, \end{cases} \qquad t \in \mathbb{R}. \tag{4.1.5}$$

(4.1.5) is called the *(parametric) scalar equation* or *component equation* of the straight line （直线的标量方程）.

Furthermore, from (4.1.5) we can eliminate the parameter t and obtain

$$\frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{c}. \tag{4.1.6}$$

(4.1.6) is called the *Cartesian equation* of the line （直线的笛卡尔方程）.

✎ ***Example* 4.1.** Find the parametric vector equation, scalar equation and the cartesian equation of the line passing through $(1, 3, -2)$ and parallel to $5\mathbf{i} - 7\mathbf{j} + 3\mathbf{k}$.

Solution: The position vector of the given point is $\mathbf{r}_0 = \mathbf{i}+3\mathbf{j}-2\mathbf{k}$, and the direction vector is $5\mathbf{i}-7\mathbf{j}+3\mathbf{k}$. The *parametric vector equation* of the line is thus

$$\mathbf{r} = (\mathbf{i} + 3\mathbf{j} - 2\mathbf{k}) + t\,(5\mathbf{i} - 7\mathbf{j} + 3\mathbf{k})\,, \qquad t \in \mathbb{R}.$$

In components the above expression is given by

$$x\mathbf{i} + y\mathbf{j} + z\mathbf{k} = (1 + 5t)\,\mathbf{i} + (3 - 7t)\,\mathbf{j} + (-2 + 3t)\,\mathbf{k},$$

which leads to the scalar equation of the line:

$$\begin{cases} x = 1 + 5t, \\ y = 3 - 7t, \\ z = -2 + 3t, \end{cases} \qquad t \in \mathbb{R}.$$

Eliminating the parameter t, we obtain the *Cartesian equation* of the line:

$$\frac{x - 1}{5} = \frac{y - 3}{-7} = \frac{z + 2}{3}.$$

□

✎ ***Example* 4.2.** (*"Two-point" equation of a straight line* （直线的两点式方程）) Find the parametric vector and the cartesian equations of the line passing through $A\,(7,4,2)$ and $B\,(8,6,5)$.

Solution: Let us choose the fixed point to be $A\,(7,4,2)$. For the direction vector it is computed as

$$\overrightarrow{AB} = (8-7)\,\mathbf{i} + (6-4)\,\mathbf{j} + (5-2)\,\mathbf{k} = \mathbf{i} + 2\mathbf{j} + 3\mathbf{k}.$$

Thus the parametric vector equation of the line is

$$\mathbf{r} = (7\mathbf{i} + 4\mathbf{j} + 2\mathbf{k}) + t\,(\mathbf{i} + 2\mathbf{j} + 3\mathbf{k})\,, \qquad t \in \mathbb{R}.$$

In components this equation is given by

$$x\mathbf{i} + y\mathbf{j} + z\mathbf{k} = (7+t)\,\mathbf{i} + (4+2t)\,\mathbf{j} + (2+3t)\,\mathbf{k}, \qquad t \in \mathbb{R},$$

so the parametric equation of the line is

$$\begin{cases} x = 7 + t, \\ y = 4 + 2t, \\ z = 2 + 3t, \end{cases} \qquad t \in \mathbb{R}.$$

And the cartesian equation reads

$$x - 7 = \frac{y-4}{2} = \frac{z-2}{3}.$$

□

There are some special cases for the equations of lines:

- If one of the components a, b and c of the direction vector $\mathbf{v}$ vanishes — say, $c = 0$ and $a, b \neq 0$ — eq.(4.1.6) becomes
$$\frac{x - x_0}{a} = \frac{y - y_0}{b}, \qquad z = z_0, \tag{4.1.7}$$
which gives a line lying in a plane parallel to the xy plane, $z = z_0$.

- If two of a, b and c vanish — say, $a, b = 0$ and $c \neq 0$ — eq.(4.1.5) becomes
$$x = x_0, \quad y = y_0, \quad z = z_0 + tc, \qquad t \in \mathbb{R}, \tag{4.1.8}$$
t being a parameter. This gives a line parallel to the z-axis and passing through the point (x_0, y_0, z_0).

✍ ***Exercises* 4.1.** Find parametric vector, parametric scalar and Cartesian equations of the line passing through the point $(2,3,5)$ in the direction of $\mathbf{i} + 3\mathbf{j} - \mathbf{k}$.

§ 4.2 Equations of planes in 3 dimensions

In this section we will derive equations of planes with the aid of dot and cross products.

Let Σ be a plane in 3 dimensions and $\mathbf{r}_0$ a point on it. Let $\mathbf{v}$, a non-zero vector, denote the normal vector of Σ （平面的法矢量）; that is, $\mathbf{v}$ be a vector perpendicular to Σ. From high school geometry we know that $\mathbf{v}$ is perpendicular to any line within the plane.

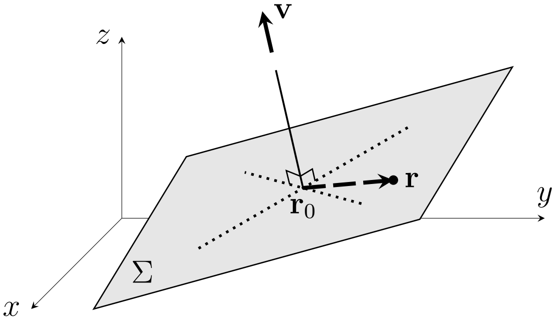


Figure 4.2: Using a point and a vector to determine a plane: Let Σ be a plane in 3 dimensions. $\mathbf{r}_0$ is fixed a point on Σ, and $\mathbf{v}$ a non-zero vector pointing in the normal direction of Σ (that is, $\mathbf{v}$ is perpendicular to any line residing within the plane). $\mathbf{r}$ is a moving point in Σ.

Consider a moving point $\mathbf{r}$ on Σ. $\mathbf{r} - \mathbf{r}_0$ is a vector lying within the plane, hence it should be perpendicular to the normal vector $\mathbf{v}$. This can be expressed by means of dot product as

$$(\mathbf{r} - \mathbf{r}_0) \cdot \mathbf{v} = 0. \tag{4.2.1}$$

When $\mathbf{r}$ moving over the whole Σ, eq.(4.2.1) gives the *vector equation* of the plane Σ （平面的矢量方程）.

Writing the vectors in components, $\mathbf{r}_0 = \mathbf{r}_0\left(x_0, y_0, z_0\right), \mathbf{r} = \mathbf{r}\left(x, y, z\right), \mathbf{v} = \mathbf{v}\left(a, b, c\right)$, we have

$$[(x\mathbf{i} + y\mathbf{j} + z\mathbf{k}) - (x_0\mathbf{i} + y_0\mathbf{j} + z_0\mathbf{k})] \cdot (a\mathbf{i} + b\mathbf{j} + c\mathbf{k}) = 0, \tag{4.2.2}$$

which simplifies to

$$a\left(x - x_0\right) + b\left(y - y_0\right) + c\left(z - z_0\right) = 0. \tag{4.2.3}$$

Eq.(4.2.3) is called the *Cartesian equation* of the plane （平面的笛卡尔方程）. Usually (4.2.3) is given by

$$ax + by + cz = d, \qquad \text{where} \qquad d = ax_0 + by_0 + cz_0. \tag{4.2.4}$$

There are some special cases for the equations of planes:

- If one of the components a, b and c of the normal vector $\mathbf{v}$ vanishes — say, $c = 0$ and $a, b \neq 0$ — eq.(4.2.4) becomes

  $$ax + by = d, \tag{4.2.5}$$

  which gives a plane parallel to the z-axis.

- If two of a, b and c vanish — say, $a, b = 0$ and $c \neq 0$ — eq.(4.2.4) becomes

$$z = \frac{d}{c}, \tag{4.2.6}$$

which gives a plane parallel to the xy-plane.

- If $d = 0$ and $a, b, c \neq 0$, eq.(4.2.4) becomes

$$ax + by + cz = 0, \tag{4.2.7}$$

which gives a plane passing through the origin $(0, 0, 0)$.

✎ ***Example* 4.3.** Find the vector equation and the Cartesian equation of the plane passing through $(1, 3, -2)$ and perpendicular to $5\mathbf{i} - 7\mathbf{j} + 3\mathbf{k}$.

Solution: The position vector of the given point is

$$\mathbf{r}_0 = \mathbf{i} + 3\mathbf{j} - 2\mathbf{k}.$$

The vector equation of the plane reads

$$[\mathbf{r} - (\mathbf{i} + 3\mathbf{j} - 2\mathbf{k})] \cdot (5\mathbf{i} - 7\mathbf{j} + 3\mathbf{k}) = 0$$

or

$$[(x-1)\,\mathbf{i} + (y-3)\,\mathbf{j} + (z+2)\,\mathbf{k}] \cdot (5\mathbf{i} - 7\mathbf{j} + 3\mathbf{k}) = 0.$$

Expanding the scalar product produces the Cartesian equation

$$5\,(x-1) - 7\,(y-3) + 3\,(z+2) = 0, \qquad \text{or} \qquad 5x - 7y + 3z + 22 = 0.$$

□

☞ **[Aside]:**

We can also derive the *"three-point" equation* of a plane （平面的三点式方程）(see Fig.4.3). Let $\mathbf{r}_0, \mathbf{r}_1$ and $\mathbf{r}_2$ be three points in the 3 dimensional space. To obtain the equation passing through the three points we need a point in the plane and a normal vector $\mathbf{v}$ of the plane. We can choose O to be the very point without loss of generality; for the normal vector $\mathbf{v}$, let us appeal to cross product. Namely, noticing $\mathbf{v}$ is perpendicular to any line within the plance, we have

$$\begin{aligned}
\mathbf{r}_1 - \mathbf{r}_0 &= (x_1 - x_0)\,\mathbf{i} + (y_1 - y_0)\,\mathbf{j} + (z_1 - z_0)\,\mathbf{k},\\
\mathbf{r}_2 - \mathbf{r}_0 &= (x_2 - x_0)\,\mathbf{i} + (y_2 - y_0)\,\mathbf{j} + (z_2 - z_0)\,\mathbf{k},
\end{aligned} \tag{4.2.8}$$

and therefore

$$\mathbf{v} = (\mathbf{r}_1 - \mathbf{r}_0) \times (\mathbf{r}_2 - \mathbf{r}_0) = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_1 - x_0 & y_1 - y_0 & z_1 - z_0 \\ x_2 - x_0 & y_2 - y_0 & z_2 - z_0 \end{vmatrix}. \tag{4.2.9}$$

Then the vector equation of the plane is achieved

$$(\mathbf{r} - \mathbf{r}_0) \cdot \mathbf{v} = ((x - x_0)\,\mathbf{i} + (y - y_0)\,\mathbf{j} + (z - z_0)\,\mathbf{k}) \cdot \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_1 - x_0 & y_1 - y_0 & z_1 - z_0 \\ x_2 - x_0 & y_2 - y_0 & z_2 - z_0 \end{vmatrix} = 0,$$

i.e.

$$\begin{vmatrix} x - x_0 & y - y_0 & z - z_0 \\ x_1 - x_0 & y_1 - y_0 & z_1 - z_0 \\ x_2 - x_0 & y_2 - y_0 & z_2 - z_0 \end{vmatrix} = 0. \tag{4.2.10}$$

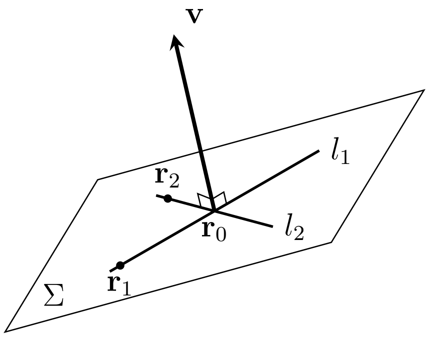


Figure 4.3: Using three points to determine a plane in 3 dimensions: Let $\mathbf{r}_0$, $\mathbf{r}_1$ and $\mathbf{r}_2$ be three points. l_1 is a line passing through $\mathbf{r}_0$ and $\mathbf{r}_1$, and l_2 another line passing through $\mathbf{r}_0$ and $\mathbf{r}_2$. In terms of the direction vectors of l_1 and l_2, we can determine the normal vector $\mathbf{v}$, and therefore the plane Σ that contains $\mathbf{r}_0$, $\mathbf{r}_1$ and $\mathbf{r}_2$.

✎ ***Example* 4.4.** Find the cartesian equation of the plane which passes through the point $A\,(1,1,1)\,, B\,(3,0,1)$ and $C\,(1,2,0)$.

Solution: Since the vectors $\overrightarrow{AB}$ and $\overrightarrow{AC}$ lie in the plane determined by A, B and C, a normal vector for the plane is the vector

$$\begin{aligned} \overrightarrow{AB} \times \overrightarrow{AC} &= [(3-1)\,\mathbf{i} + (0-1)\,\mathbf{j} + (1-1)\,\mathbf{k}] \times [(1-1)\,\mathbf{i} + (2-1)\,\mathbf{j} + (0-1)\,\mathbf{k}] \\ &= [2\mathbf{i} - \mathbf{j}] \times [\mathbf{j} - \mathbf{k}] = \mathbf{i} + 2\mathbf{j} + 2\mathbf{k}. \end{aligned}$$

Since the plane passes through the point $A\,(1,1,1)$ and is perpendicular to the vector $\mathbf{i} + 2\mathbf{j} + 2\mathbf{k}$, its Cartesian equation reads

$$1\,(x-1) + 2\,(y-1) + 2\,(z-1) = 0, \qquad \text{or} \qquad x + 2y + 2z - 5 = 0.$$

An alternative way is to directly use eq.(4.2.10) to obtain the result:

$$\begin{vmatrix} x-1 & y-1 & z-1 \\ 3-1 & 0-1 & 1-1 \\ 1-1 & 2-1 & 0-1 \end{vmatrix} = \begin{vmatrix} x-1 & y-1 & z-1 \\ 2 & -1 & 0 \\ 0 & 1 & -1 \end{vmatrix} = 0,$$

i.e.

$$1\,(x-1)+2\,(y-1)+2\,(z-1)=0, \qquad \text{or} \qquad x+2y+2z-5=0.$$

□

✌

✍ ***Exercises* 4.2.**

1. Find vector and Cartesian equations of the plane containing the point $(2,3,5)$ with normal vector $\mathbf{i}+3\mathbf{j}-\mathbf{k}$.

2. Let $P=(1,2,3)$, $Q=(-1,-2,-3)$ and $R=(4,-4,4)$.

   (i) Express $\overrightarrow{PQ}$ and $\overrightarrow{PR}$ in Cartesian form.

   (ii) Find the cross product $\overrightarrow{PQ}\times\overrightarrow{PR}$.

   (iii) Find the Cartesian equation of the plane containing P, Q, R.

§ 4.3 Scalar triple product (optional)

§ 4.3.1 Using scalar triple product to compute volume of parallelepiped

Let $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ be three vectors. Their *scalar triple product* （标量三重积; *triple product* for short）, also called *mixed product*（混合积）, is defined as

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

In eq.(3.2.13) of §3.2.3 we learnt that cross product is used to compute the area of a parallelogram; in the following we will see that scalar triple product is useful in computing the volume of a parallelepiped（平行六面体）in 3-dimensions.

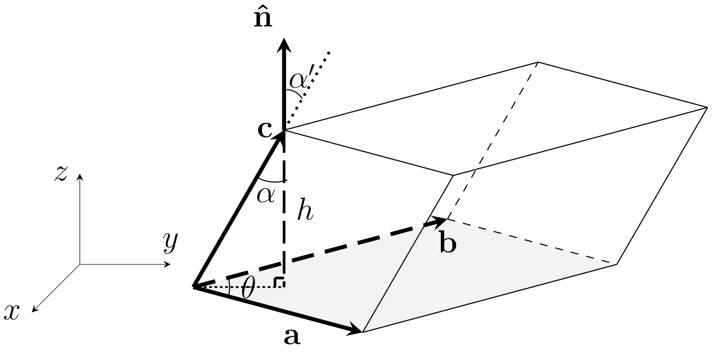


Figure 4.4: Computation of volume of 3-dimensional parallelepiped in terms of scalar triple product: Consider a parallelepiped with three adjacent edges represented by 3 vectors $\mathbf{a}, \mathbf{b}$ and $\mathbf{c}$. Its (signed) volume is given by the scalar triple product $\mathbf{c} \cdot (\mathbf{a} \times \mathbf{b})$.

In Fig.4.4, the parallelepiped has three adjacent edges represented by three vectors $\mathbf{a}, \mathbf{b}$ and $\mathbf{c}$. To compute its volume let us denote the parallelogram generated by $\mathbf{a}$ and $\mathbf{b}$ as a base（底）of the body (the grey one in Fig.4.4). The line segment h as the height （高）corresponding to the base. According to eq.(3.2.13) the area of the base can be computed by means of the cross product $\mathbf{a} \times \mathbf{b}$, that is, the magnitude $|\mathbf{a} \times \mathbf{b}| = |\mathbf{a}|\,|\mathbf{b}|\,|\sin\theta|$ gives the area and the direction of $\mathbf{a}\times\mathbf{b}$ gives the normal direction（法向）of the area (denoted by the unit vector $\hat{\mathbf{n}}$). In summary,

$$\mathbf{a} \times \mathbf{b} = |\mathbf{a} \times \mathbf{b}|\,\hat{\mathbf{n}}. \tag{4.3.1}$$

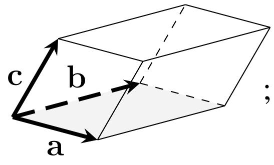
Then the scalar triple product is given by

$$\mathbf{c} \cdot (\mathbf{a} \times \mathbf{b}) = |\mathbf{a} \times \mathbf{b}|\,(\mathbf{c} \cdot \hat{\mathbf{n}}) = |\mathbf{a} \times \mathbf{b}|\,|\mathbf{c}| \cos\alpha'. \tag{4.3.2}$$

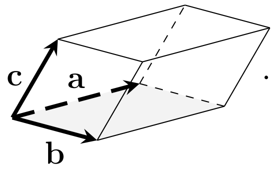
It is seen that α and α' are a pair of vertical angles（对顶角）, $\alpha = \alpha'$, and the magnitude of $|\mathbf{c}| \cos\alpha$ equals the height h. Hence the triple product gives the signed volume of the parallelepiped,

$$\mathbf{c} \cdot (\mathbf{a} \times \mathbf{b}) = \pm\,|\mathbf{a} \times \mathbf{b}|\,h, \tag{4.3.3}$$

where:

“$+$” holds when $\cos\alpha > 0$, i.e., when $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ form a right-handed set, 

;

“$-$” holds when $\cos\alpha < 0$, i.e., when $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ form a left-handed set, 

.

For simplicity, we can use the absolute value, $|\mathbf{c} \cdot (\mathbf{a} \times \mathbf{b})|$, to compute the volume of a parallelepiped.

§ 4.3.2 Invariance of scalar triple product under cyclic permutation of operands

As we know, cross product has the determinant definition, eq.(3.2.2). Thus the scalar triple product has a precise determinant expression

$$\begin{aligned} \mathbf{c}\cdot(\mathbf{a}\times\mathbf{b}) &= (c_1\mathbf{i}+c_2\mathbf{j}+c_3\mathbf{k})\cdot\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} = \begin{vmatrix} c_1 & c_2 & c_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} \\ &= a_1b_2c_3 + a_2b_3c_1 + a_3b_1c_2 - a_1b_3c_2 - a_3b_2c_1 - a_2b_1c_3. \end{aligned} \tag{4.3.4}$$

The reader is encouraged to prove (4.3.4) by expanding the components in the cross and dot products.

In the light of the properties of determinants (see Chapter 2 of General Reference Textbook [3]), this determinant is invariant under the following cyclic permutations of rows,

$$\begin{vmatrix} c_1 & c_2 & c_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = \begin{vmatrix} b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \\ a_1 & a_2 & a_3 \end{vmatrix}, \tag{4.3.5}$$

hence we have an important property: scalar triple product is invariant under circular shift of its three operands,

$$\mathbf{c}\cdot(\mathbf{a}\times\mathbf{b}) = \mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) = \mathbf{b}\cdot(\mathbf{c}\times\mathbf{a}). \tag{4.3.6}$$

Furthermore, taking into account the anti-symmetry of cross product, we have

$$\begin{aligned} &\mathbf{c}\cdot(\mathbf{a}\times\mathbf{b}) = \mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) = \mathbf{b}\cdot(\mathbf{c}\times\mathbf{a}) \\ = \ &-\mathbf{c}\cdot(\mathbf{b}\times\mathbf{a}) = -\mathbf{b}\cdot(\mathbf{a}\times\mathbf{c}) = -\mathbf{a}\cdot(\mathbf{c}\times\mathbf{b}). \end{aligned} \tag{4.3.7}$$

In terms of (4.3.7) there is an immediate corollary:

$$\mathbf{a}\cdot(\mathbf{a}\times\mathbf{b}) = \mathbf{b}\cdot(\mathbf{a}\times\mathbf{a}) = 0. \tag{4.3.8}$$

✎ ***Example* 4.5.** Consider a parallelepiped generated by three vectors: $\mathbf{a} = 4\mathbf{i}$, $\mathbf{b} = 2\mathbf{i} + 2\mathbf{j}$ and $\mathbf{c} = \mathbf{i} + \sqrt{3}\mathbf{k}$. Compute its volume.

Solution: In the light of (4.3.4) and (4.3.5) we have

$$\begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = \begin{vmatrix} 4 & 0 & 0 \\ 2 & 2 & 0 \\ 1 & 0 & \sqrt{3} \end{vmatrix} = 8\sqrt{3}.$$

This result can be easily checked by drawing a picture for the parallelepiped and recognizing a pair of base and height to compute its volume. □

✍ ***Exercises* 4.3.** Given the vectors $\mathbf{a} = 2\mathbf{i} - \mathbf{j}$, $\mathbf{b} = \mathbf{i} + \mathbf{j} + \mathbf{k}$ and $\mathbf{b} = -2\mathbf{i} + \mathbf{k}$, evaluate

(i) $(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}$;

(ii) $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$;

(iii) $\mathbf{a} \cdot (\mathbf{a} \times \mathbf{b})$.

Chapter 5 Solving linear equations

From this chapter we will enter into the major part of this course of *Linear Algebra*. Let us start from solving linear systems, i.e., systems of linear equations（线性方程组）.

§ 5.1 System of linear equations — Gaussian elimination and matrices

§ 5.1.1 Definitions

***Definition* 5.1.** *A linear equation in n unknowns*（未知数）is

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b \tag{5.1.1}$$

where $a_1, \cdots, a_n$ are real coefficients, and $x_1, \cdots, x_n$ are unknowns.

***Definition* 5.2.** When $b = 0$, specially, the equation is called *a homogeneous equation*（齐次方程）; otherwise, when $b \neq 0$, it is an *inhomogeneous equation*（非齐次方程）.

***Definition* 5.3** (Flatness & linearity).

- In 2D, it is a line, in Fig.5.1(a)

$$a_1x + a_2y = b.$$

- In 3D, it is a plane, in Fig.5.1(b)

$$a_1x + a_2y + a_3z = b.$$

- In n dimensions, it is a super plane（超平面）in the n-dimensional space

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b.$$

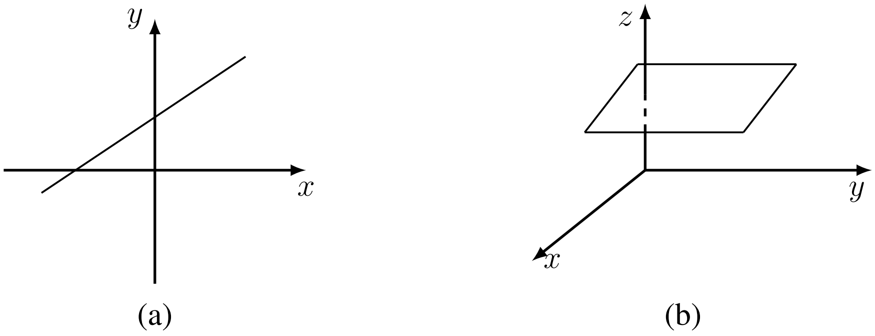


Figure 5.1: (a) A line in two dimensions. (b) A plane in three dimensions.

The methods and techniques we study here have roots in linear equations, as well as wide applications in other areas. In all these areas, flatness and linearity are the key properties, hence this course we are learning is named as *Linear Algebra*.

***Definition* 5.4.** A system of m linear equations in n unknowns (or, called an $m \times n$ linear system) is given by

$$\begin{aligned} a_{11}x_1 \;+\; a_{12}x_2 + \cdots + \; a_{1n}x_n &= b_1, \\ a_{21}x_1 \;+\; a_{22}x_2 + \cdots + \; a_{2n}x_n &= b_2, \\ \vdots \qquad\quad \vdots \qquad\qquad\quad \vdots \quad &= \vdots, \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m, \end{aligned} \tag{5.1.2}$$

where a_{ij} and b_i are real numbers, and x_j's are unknowns, $i = 1, \cdots, m$ and $j = 1, \cdots, n.$

***Definition* 5.5.** Specially, when $b_1 = b_2 = \cdots = b_m = 0$, the system is called a homogenous system.

One should notice that it is unnecessary that $m = n$. In fact,

(a) $m = n$: E.g., in a 2×2 case,

$$\begin{cases} \;\; x_1 + 2x_2 = 5, \\ 2x_1 + 3x_2 = 8. \end{cases}$$

(b) $m < n$: E.g., in a 2×3 case,

$$\begin{cases} \;\; x_1 - x_2 + x_3 = 2, \\ 2x_1 + x_2 - x_3 = 4. \end{cases}$$

(c) $m > n$: E.g., in a 3×2 case,

$$\begin{cases} x_1 + x_2 = 2, \\ x_1 - x_2 = 1, \\ x_1 = 4. \end{cases}$$

***Definition* 5.6** (Solution of linear system)**.** The solution to an $m \times n$ system is an n-tuple,

$$(x_1, x_2, \cdots, x_n)^T = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \tag{5.1.3}$$

For example, the solutions to the above Cases (a)–(c) are, respectively:

(a) $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$; (b) $\begin{pmatrix} 2 \\ \alpha \\ \alpha \end{pmatrix}$, $\alpha \in \mathbb{R}$ is a parameter; (c) no solution. (*Why? See below.*)

***Definition* 5.7** (Consistency 一致性/ inconsistency 不一致性)**.** A linear systems is called inconsistent, if it has no solutions; otherwise, it is called consistent if it has at least one solution.

Thus, the above Cases (a) and (b) are consistent, but Case (c) is inconsistent.

***Definition* 5.8** (Solution set 解集)**.** The set of all solutions of a linear system is called the solution set.

Obviously, the solution set of an inconsistent system is $\emptyset$.

§ 5.1.2 Geometric understanding of linear equations and solution set

Here some typical cases are illustrated for instance for your better understanding.

- 2×2 systems:

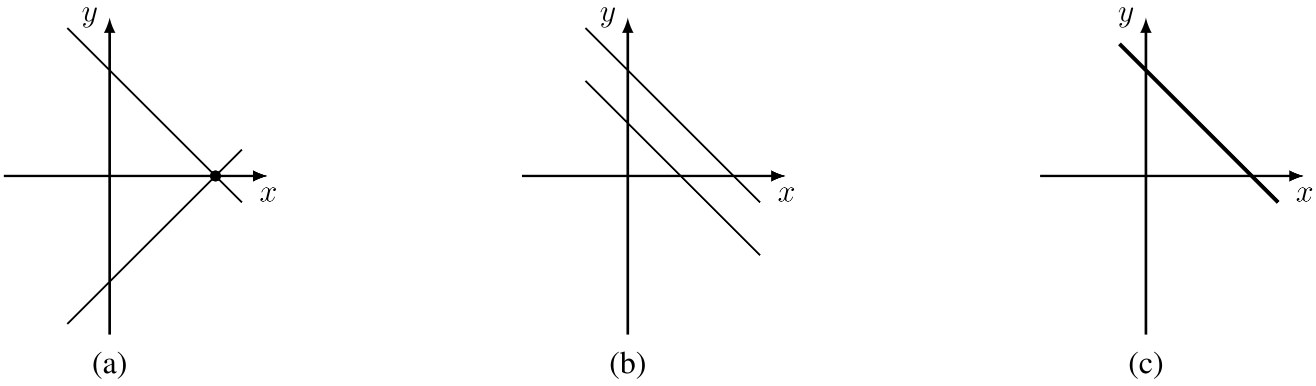


Figure 5.2: In two dimensions, a linear equation in two unknowns corresponds to a straight line. (a) Two lines intersecting at a point means the solution of the linear system is unique. (b) Two lines being parallel means the linear system is inconsistent, i.e., the solution set is empty. (c) Two lines coinciding means the linear system has infinitely many solutions.

Typical examples for the above three cases are the following:

(a) $\begin{cases} x_1 + x_2 & = 2 \\ x_1 - x_2 & = 2 \end{cases}$ (b) $\begin{cases} x_1 + x_2 = 2 \\ x_1 + x_2 = 1 \end{cases}$ (c) $\begin{cases} x_1 + x_2 = 2 \\ -x_1 - x_2 = -2 \end{cases}$

- 2×3 systems:

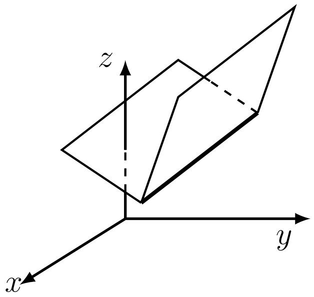


Figure 5.3: In three dimensions, a linear equation in three unknowns corresponds to a plane. Two planes intersecting at a line means the system has infinitely many solutions.

The reader is invited to find the case of two parallel planes and the corresponding solution set.

- 3×3 systems:

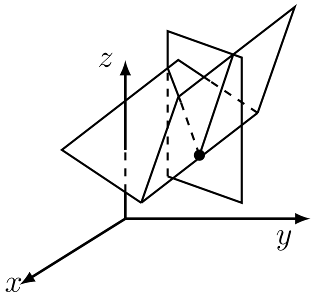


Figure 5.4: In three dimensions, a linear equation in three unknowns corresponds to a plane. Three planes intersecting at a point means the system has a unique solution.

The reader is invited to consider the case of two parallel planes intersecting with a third plane, as well as the corresponding solution set. Other cases include the situation of three parallel planes, and so on.

§ 5.1.3 Equivalent systems

***Definition* 5.9.** Two systems of equations involving the same number of unknowns are said to be equivalent if they have the same solution set.

Therefore, in order to solve a complicated system of equations, one can make *appropriate changes* to produce an *equivalent but simpler* version of the original one. Let us start from an example.

✎ ***Example* 5.1.**

$$\begin{cases} 3x_1 + 2x_2 - x_3 = -2 \\ -3x_1 - x_2 + x_3 = 5 \\ 3x_1 + 2x_2 + x_3 = 2 \end{cases} \qquad \Longrightarrow \qquad \begin{cases} 3x_1 + 2x_2 - x_3 = -2 \\ x_2 = 3 \\ 2x_3 = 4 \end{cases}$$

These two systems of linear equations, although having different looking, share the same solution set and therefore are equivalent. Obviously the right system is much easier to deal with, hence when solving the left one we can wisely change it into the right equivalent one and find the solution.

The next question is *how to make appropriate changes*? The rules are listed below in Table 5.1.

Tree rules	Notations
Swapping two rows/equations	$R_i \leftrightarrow R_j$
Multiplying a row/equation with a nonzero constant	$R_j \Rightarrow \alpha R_j$
Adding a multiple of a row/equation to another row/equation	$R_i \Rightarrow R_i + \alpha R_j$

Table 5.1: The three rules that apply to transformations turning a system to another equivalent one. *Note: The symbol "⇒" is often replaced by ":=" in literature, to denote that it is a re-evaluation symbol, instead of an equating operator.*

It is seen that these three rules exactly correspond to the Gaussian eliminations（高斯消元法）we learnt in high school mathematics. In the light of them we can change/transform a linear system to an equivalent but much simpler version to find the solution.

§ 5.1.4 $n \times n$ systems

Next, a natural question may arise:

> *What does the "simplest" system looks like? What is the criterion for "simple"?*

The generic and exact answer to this question is *the echelon form*（阶梯形式）, i.e., a *staircase-shaped* or *stepwise* form.

However, this case is too generic to begin with, since $m \neq n$ is permitted (wherein the shape of the coefficient matrix（系数矩阵）is unnecessarily a strict triangle). Hence let us leave the generic case to the next section, but start from a special but simpler case, $m = n$, in this section as a preparation. Namely, we will discuss an $n \times n$ system, where the coefficient matrix is a strict triangle.

Special answers in $n \times n$ system

For an $n \times n$ system, if it has exactly one solution, which is a fact we know beforehand, we can try to obtain the desired "simple form" by appealing to the "triangular system". Here *triangular* means, say,

$$\begin{aligned} 3x_1 + 2x_2 + x_3 &= 1 \\ x_2 - x_3 &= 2 \\ 2x_3 &= 4 \end{aligned} \tag{5.1.4}$$

Obviously this system is easy to solve. The formal definition for triangular shape is as follows.

***Definition* 5.10** (Strictly triangular form)**.** An $n \times n$ system is a strictly triangular form if the entries of the lower triangle are 0's and the diagonals are non-zeros.

[Remark]:

There is no restriction for the upper triangle. That is, the system looks like

$$\begin{array}{cccccc} * & \bullet & \bullet & = & \bullet \\ 0 & * & \bullet & = & \bullet \\ 0 & 0 & * & = & \bullet \end{array}$$

where the symbol "•" could be either 0 or nonzero.

Now we are at the stage to answer the question: *Why a triangular system is simple?* The reason is that it can be solved by means of *back substitution*（反向迭代法）, as shown by example below.

✎ ***Example* 5.2.**

$$\begin{aligned} 3x_1 + 2x_2 + x_3 &= 1 \qquad (R_1) \\ x_2 - x_3 &= 2 \qquad (R_2) \\ 2x_3 &= 4 \qquad (R_3) \end{aligned} \tag{5.1.5}$$

where R_1 denotes Row 1, and so forth.

Solution:

From R_3, we obtain: $x_3 = 2;$

Then using back substitution we obtain from R_2 : $x_2 - 2 = 2 \quad \Longrightarrow \quad x_2 = 4;$ ☐

Finally, from R_1 we obtain: $3x_1 + 2 \times 4 + 2 = 1 \quad \Longrightarrow \quad x_1 = -3.$

✎ ***Example* 5.3.**

$$\begin{aligned} 2x_1 - x_2 + 3x_3 - 2x_4 &= 1 \qquad (R_1) \\ x_2 - 2x_3 + 3x_4 &= 2 \qquad (R_2) \\ 4x_3 + 3x_4 &= 3 \qquad (R_3) \\ 4x_4 &= 4 \qquad (R_4) \end{aligned} \tag{5.1.6}$$

Solution:

$$R_4 : \quad x_4 = 1 \;\Rightarrow\; R_3 : \quad x_3 = 0 \;\Rightarrow\; R_2 : \quad x_2 = -1 \;\Rightarrow\; R_1 : \quad x_1 = 1 \tag{5.1.7}$$

☐

Having got an idea about the advantage of a triangular form, we always hope to turn a given form into a triangular one. Then, next, *how to do this job?* Our strategy is to take the *Gaussian elimination*, as mentioned.

✎ ***Example* 5.4.**

$$\begin{aligned} x_1 + 2x_2 + \ \ x_3 &= \ \ 3 \qquad (R_1)\\ 3x_1 - \ \ x_2 - 3x_3 &= -1 \qquad (R_2)\\ 2x_1 + 3x_2 + \ \ x_3 &= \ \ 4 \qquad (R_3) \end{aligned}$$

Solution:

$$\begin{array}{llll} (R_2) - 3(R_1): & -7x_2 - 6x_3 = -10 & & (R_4)\\ (R_3) - 2(R_1): & -\ x_2 - \ x_3 = -\ 2 & & (R_5)\\ (R_5) + \frac{1}{7}(R_4): & -\frac{1}{7}x_3 = -\ \frac{4}{7} & \text{i.e.,} \quad x_3 = 4 & (R_6) \end{array}$$

Hence (R_1), (R_4) and (R_6) form a triangle:

$$\begin{aligned} x_1 + 2x_2 + \ \ x_3 &= \ \ 3\\ -\ 7x_2 - 6x_3 &= -\ 10\\ x_3 &= \ \ 4 \end{aligned} \tag{5.1.8}$$

Solving this triangular system by back substitution we achieve

$$x_3 = 4, \qquad x_2 = -2, \qquad x_1 = 3.$$

☐

§ 5.1.5 Matrix: an introduction

Introduction to matrix

In solving a system of linear equations we see that the unknowns x_1, x_2, x_3 are dummy variables and therefore are ignorable; only the coefficients matter in the computation. Therefore the coefficients are extracted from the system to form a rectangular array like:

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

Such an array is called the coefficient matrix of a linear system.

***Definition* 5.11** (Matrix). A matrix of size $m \times n$ is a rectangular array of entries given by

$$\underbrace{\left.\begin{pmatrix} * & * & \cdots & * \\ * & * & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \cdots & * \end{pmatrix}\right\}}_{n \text{ columns}} m \text{ rows} \tag{5.1.9}$$

Specially, when $m = n$, this $n \times n$ array is called a square matrix.

Augmented matrix

***Definition* 5.12** (Augmented matrix 增广矩阵). Let A be an $m \times n$ matrix, and B be an $m \times r$ matrix,

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}, \qquad B = \begin{pmatrix} b_{11} & \cdots & b_{1r} \\ \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{mr} \end{pmatrix}.$$

Attaching B to A forms a new $m \times (n + r)$ matrix $(A|B)$, called the augmented matrix

$$(A|B) = \left(\begin{array}{ccc|ccc} a_{11} & \cdots & a_{1n} & b_{11} & \cdots & b_{1r} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} & b_{m1} & \cdots & b_{mr} \end{array}\right)$$

Augmented matrix is useful in solving linear systems

The most commonly used augmented matrix is a special case

$$\left(\begin{array}{ccc|c} a_{11} & \cdots & a_{1n} & b_1 \\ \vdots & \ddots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} & b_m \end{array}\right)$$

where $\begin{pmatrix} b_1 & \cdots & b_m \end{pmatrix}^T$ is the right-hand side of the studied linear system. This special augmented matrix is used to solve the linear system as follows.

Solving a linear system in terms of an augmented matrix

***Theorem* 5.1.** The Elementary Row Operations (ERO's)（基本行变换）for solving a system of linear equations are given by the three rules of Table 5.1:

1. $R_i \longleftrightarrow R_j$: Swapping/interchanging rows.
2. $R_i \Longrightarrow \alpha R_i$: Multiplying a row by a nonzero real number α.
3. $R_i \Longrightarrow R_i + \alpha R_j$: Adding the multiple of a row to another.

✎ ***Example* 5.5.** $\begin{cases} x_1 + 2x_2 + x_3 = 3 \\ 3x_1 - x_2 - 3x_3 = -1 \\ 2x_1 + 3x_2 + x_3 = 4 \end{cases}$

Solution:

The coefficients of the linear system are extracted to form an augmented matrix:

$$
\begin{aligned}
&\left(\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 3 & -1 & -3 & -1 \\ 2 & 3 & 1 & 4 \end{array}\right) \\
\xrightarrow{R_2 \Longrightarrow R_2 - 3R_1} &\left(\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 3 & -1 & -3 & -10 \\ 2 & 3 & 1 & 4 \end{array}\right) \xrightarrow{R_3 \Longrightarrow R_3 - 2R_1} \left(\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 0 & -7 & -6 & -10 \\ 0 & -1 & -1 & -2 \end{array}\right) \\
\xrightarrow{R_3 \Longrightarrow 7R_3 - R_2} &\left(\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 0 & -7 & -6 & -10 \\ 0 & 0 & -1 & -4 \end{array}\right) \xrightarrow[R_2 \Longrightarrow -R_2]{R_3 \Longrightarrow -R_3} \left(\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 0 & 7 & 6 & 10 \\ 0 & 0 & 1 & 4 \end{array}\right)
\end{aligned}
\tag{5.1.10}
$$

Then the solutions of x_1, x_2, x_3 are immediately obtained by means of back substitution. □

✎ ***Example* 5.6.** Solve the following linear system:

$$
\begin{array}{rrrrcr}
 & x_2 & -x_3 & +x_4 & = & 0 \\
x_1 & +x_2 & +x_3 & +x_4 & = & 6 \\
2x_1 & +4x_2 & +x_3 & -2x_4 & = & -1 \\
3x_1 & +x_2 & -2x_3 & +2x_4 & = & 3
\end{array}
$$

Solution:

The augmented matrix of coefficients reads

$$\left(\begin{array}{cccc|c} 0 & -1 & -1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 6 \\ 2 & 4 & 1 & -2 & -1 \\ 3 & 1 & -2 & 2 & 3 \end{array}\right)$$

$$\xrightarrow{R_1\longleftrightarrow R_2} \left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 6 \\ 0 & -1 & -1 & 1 & 0 \\ 2 & 4 & 1 & -2 & -1 \\ 3 & 1 & -2 & 2 & 3 \end{array}\right) \xrightarrow[R_4\Longrightarrow R_4-3R_1]{R_3\longleftrightarrow R_3-2R_1} \left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 6 \\ 0 & -1 & -1 & 1 & 0 \\ 0 & 2 & -1 & -4 & -13 \\ 0 & -2 & -5 & -1 & -15 \end{array}\right)$$

$$\xrightarrow[R_4\Longrightarrow R_4-2R_2]{R_3\longleftrightarrow R_3+2R_2} \left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 6 \\ 0 & -1 & -1 & 1 & 0 \\ 0 & 0 & -3 & -2 & -13 \\ 0 & 0 & -3 & -3 & -15 \end{array}\right) \xrightarrow{R_4\longleftrightarrow R_4-R_3} \left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 6 \\ 0 & -1 & -1 & 1 & 0 \\ 0 & 0 & -3 & -2 & -13 \\ 0 & 0 & 0 & -1 & -2 \end{array}\right)$$

Then, in terms of back substitution we obtain the solution, $(x_1, x_2, x_3, x_4)^T = (2, -1, 3, 2)^T$. ☐

The above procedure can be illustrated as follows:

$$\left(\begin{array}{cccc|c} \circ & \circ & \circ & \circ & \circ \\ \circ & \circ & \circ & \circ & \circ \\ \circ & \circ & \circ & \circ & \circ \\ \circ & \circ & \circ & \circ & \circ \end{array}\right) \longrightarrow \left(\begin{array}{cccc|c} \circ & \circ & \circ & \circ & \circ \\ 0 & \circ & \circ & \circ & \circ \\ 0 & \circ & \circ & \circ & \circ \\ 0 & \circ & \circ & \circ & \circ \end{array}\right)$$

$$\longrightarrow \left(\begin{array}{cccc|c} \circ & \circ & \circ & \circ & \circ \\ 0 & \circ & \circ & \circ & \circ \\ 0 & 0 & \circ & \circ & \circ \\ 0 & 0 & \circ & \circ & \circ \end{array}\right) \longrightarrow \left(\begin{array}{cccc|c} \circ & \circ & \circ & \circ & \circ \\ 0 & \circ & \circ & \circ & \circ \\ 0 & 0 & \circ & \circ & \circ \\ 0 & 0 & 0 & \circ & \circ \end{array}\right)$$

Then, back substitution leads to the final solution.

§ 5.2 Row echelon form

§ 5.2.1 Introduction

In § 5.1 we learnt the strictly triangular form of an $n \times n$ system, $\begin{pmatrix} * & \bullet & \cdots & \bullet \\ & * & \cdots & \bullet \\ & & \ddots & \vdots \\ & & & * \end{pmatrix}$. However, in practice some *easier situations may occur*. For instance, firstly,

$$\begin{pmatrix} * & \bullet & \bullet & \bullet & \cdots & \bullet \\ & 0 & \bullet & \bullet & \cdots & \bullet \\ & & \bigstar & \bullet & \cdots & \bullet \\ & & & \ddots & & \vdots \\ & & & & & \bullet \end{pmatrix} \xrightarrow[\text{wider staircase appears}]{\text{strictly triangular form is broken}} \begin{pmatrix} * & \bullet & \bullet & \bullet & \cdots & \bullet \\ & & * & \bullet & \cdots & \bullet \\ & & & * & \cdots & \bullet \\ & & & & \ddots & \vdots \\ & & & & & * \end{pmatrix}$$

A diagonal entry might be zero, as shown; let us denote the row contains this zero as R_0. Then the $\bigstar$ entry in the row immediately below R_0 can be eliminated by making use of the row R_0. The blue polyline indicates the staircase that contains a wider platform.

Secondly, an all-zero row may occur, as shown below. It can be moved to the very bottom, thanks to the Gaussian Elimination Rule 1 for swapping.

$$\begin{pmatrix} \bullet & \bullet & \cdots & \bullet \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \bullet & \bullet & \cdots & \bullet \end{pmatrix}$$

Therefore, we might come across a matrix in the form of (5.2.1):

$$\begin{pmatrix} * & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ 0 & 0 & * & \bullet & \bullet & \bullet & \bullet & \bullet \\ 0 & 0 & 0 & * & \bullet & \bullet & \bullet & \bullet \\ 0 & 0 & 0 & 0 & 0 & 0 & * & \bullet \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{5.2.1}$$

Obviously, such an echelon-shaped matrix, not strictly triangular, is even *easier* to solve than a triangular one. A matrix in the form of (5.2.1) is called an *echelon form*（阶梯形式）.

[Remark]:

1. The name *row echelon form* is in contrast with another name, *column echelon form*. The former is more popular in practice since it is commonly used in matrix transformations. Hence when talking about a row echelon form, people are used to calling it as an *echelon form* for short.

2. In the matrix (5.2.1), a $*$ entry is called a *pivot entry*（枢元）or leading entry. It is the first nonzero entry in its respective row.

   A pivot entry is nonzero for sure, but must it be a 1? The answer is NO in some textbooks, such as *David C. Lay, 2012*, but YES in the others, such as *Steven Leon, 2010*. In the present course we follow Leon's notation; hence an echelon form is given by

$$\begin{pmatrix} 1 & \bullet & \bullet & \bullet & \bullet \\ 0 & 0 & 0 & 1 & \bullet \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

3. Being an echelon form, unnecessarily strictly triangular, a matrix need not be a square matrix, but could be $m \times n$ with $m \neq n$. For instance, a 3×5 matrix,

$$\begin{pmatrix} 1 & \bullet & \bullet & \bullet & \bullet \\ 0 & 0 & 0 & 1 & \bullet \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Now we are at the stage to present the formal definition of echelon form.

***Definition* 5.13.** An $m \times n$ *echelon form*（阶梯形式）satisfies the following requirements:

I. All nonzero rows are above all the all-zero rows.
II. In any nonzero row, the leading entry is 1.
III. In any nonzero row, the leading 1 is further to the right than the leading 1 in any above row.

You may feel it is not easy to memorize the above requirements I–III. No worries, the following shape helps:

$$\begin{pmatrix} 1 & \bullet & \bullet & \bullet & \bullet & \bullet \\ 0 & 0 & 0 & 1 & \bullet & \bullet \\ 0 & 0 & 0 & 0 & 1 & \bullet \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

✍ ***Exercises* 5.1.** These forms are echelon forms:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & -2 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 2 & 4 \\ 0 & 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},$$

but these are not:

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 1 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 5 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

Why?

§ 5.2.2 Reduced row echelon form

Already achieved the row echelon form, we are now ready to find the solution of a linear system by means of back substitution.

✎ ***Example* 5.7.** Solve the solution of the following echelon form by using back substitution:

$$\begin{pmatrix} 1 & 3 & 2 & 5 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 3 & 2 & \boxed{0} \\ 0 & 0 & 1 & \boxed{0} \\ 0 & 0 & 0 & 1 \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 3 & \boxed{0} & \boxed{0} \\ 0 & 0 & 1 & \boxed{0} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

The last matrix is called a *reduced echelon form*（约化阶梯形式）. Obviously this is the final result of the back substitution operation, where the solution can be immediately read off, as shown by the following examples.

✎ ***Example* 5.8.** Augmented matrix of a reduced echelon form

- $$\left(\begin{array}{ccc|c} 1 & & & 3 \\ & 1 & & 1 \\ & & 1 & -1 \end{array}\right) \Longleftrightarrow \begin{cases} x_1 & & & = 3 \\ & x_2 & & = 1 \\ & & x_3 & = -1 \end{cases}$$

- $$\left(\begin{array}{cccc|c} 1 & 3 & & & 3 \\ & & 1 & & 1 \\ & & & 1 & -1 \end{array}\right) \Leftrightarrow \begin{cases} x_1 & +3x_2 & & & = 3 \\ & & x_3 & & = 1 \\ & & & x_4 & = -1 \end{cases}$$

i.e.,

$$\begin{cases} x_1 &= \quad 3-3x_2 \\ x_3 &= \quad 1 \\ x_4 &=- \quad 1 \end{cases} \quad \xLeftrightarrow{\text{Let}\quad x_2=t} \quad \begin{cases} x_1 &= \ 3-3t, \\ x_2 &= \ t, \\ x_3 &= \ 1, \\ x_4 &= -\ 1, \end{cases} \qquad t\in\mathbb{R}.$$

Now we are at the stage of introducing the formal definition of reduced echelon forms.

***Definition* 5.14.** A *reduced row echelon form*, or *reduced echelon form*（约化阶梯形式）for short, satisfies the following requirements:

1. It must be a row echelon form.
2. In each row that contains a leading entry 1, that 1 must be the only non-zero entry in the column where it resides.

The following illustration is useful for memorizing this definition:

$$\begin{pmatrix} 1 & \bullet & \bullet & 0 & \bullet & 0 & \bullet & 0 \\ 0 & 0 & 0 & 1 & \bullet & 0 & \bullet & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & \bullet & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Now, in summary, we use the following procedure to solve a linear system.

***Definition* 5.15** (Gauss-Jordan reduction 高斯–约当约化过程)**.**

$$\text{Matrix} \ \overset{\text{ERO's}}{\Longrightarrow} \ \text{Row echelon form} \ \overset{\text{ERO's}}{\Longrightarrow} \ \text{Reduced echelon form}$$

✎ ***Example* 5.9.** Use the Gauss-Jordan reduction method to solve the following linear system:

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

Solution:

$$\left(\begin{array}{cccc|c} -1 & 1 & -1 & 3 & 0 \\ 3 & 1 & -1 & -1 & 0 \\ 2 & -1 & -2 & -1 & 0 \end{array}\right) \longrightarrow \left(\begin{array}{cccc|c} -1 & 1 & -1 & 3 & 0 \\ 0 & 4 & -4 & 8 & 0 \\ 0 & 1 & -4 & 5 & 0 \end{array}\right)$$

$$\longrightarrow \left(\begin{array}{cccc|c} -1 & 1 & -1 & 3 & 0 \\ 0 & 1 & -1 & 2 & 0 \\ 0 & 1 & -4 & 5 & 0 \end{array}\right) \longrightarrow \left(\begin{array}{cccc|c} 1 & -1 & 1 & -3 & 0 \\ 0 & 1 & -1 & 2 & 0 \\ 0 & 0 & -3 & 3 & 0 \end{array}\right)$$

$$\longrightarrow \left(\begin{array}{rrrr|r} 1 & -1 & 1 & -3 & 0 \\ 0 & 1 & -1 & 2 & 0 \\ 0 & 0 & 1 & -1 & 0 \end{array}\right) \longrightarrow \left(\begin{array}{rrrr|r} 1 & -1 & 0 & -2 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 \end{array}\right)$$

$$\longrightarrow \left(\begin{array}{rrrr|r} 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 \end{array}\right)$$

This means the following system is equivalent to the original one:

$$\begin{cases} x_1 \qquad\qquad -x_4 = 0 \\ \qquad x_2 \qquad +x_4 = 0 \\ \qquad\qquad x_3 \; -x_4 = 0 \end{cases}$$

Introducing $x_4 = t$, we achieve a nontrivial solution:

$$x_1 = t, \qquad x_2 = -t, \qquad x_3 = t, \qquad x_4 = t, \qquad \text{where} \qquad t \in \mathbb{R}.$$

$\square$

The above example is a homogenous case. For an inhomogeneous system the Gauss-Jordan reduction is similar.

§ 5.2.3 Applications

✎ ***Example* 5.10.** *Traffic flow* (Page 17, S. Leon, 2010).

In a downtown section of a certain city, two sets of one-way streets intersect as shown in Figure 5.5. The average hourly volume of traffic entering and leaving this section during rush hour is given in the diagram. Determine the amount of traffic between each of the four intersection.

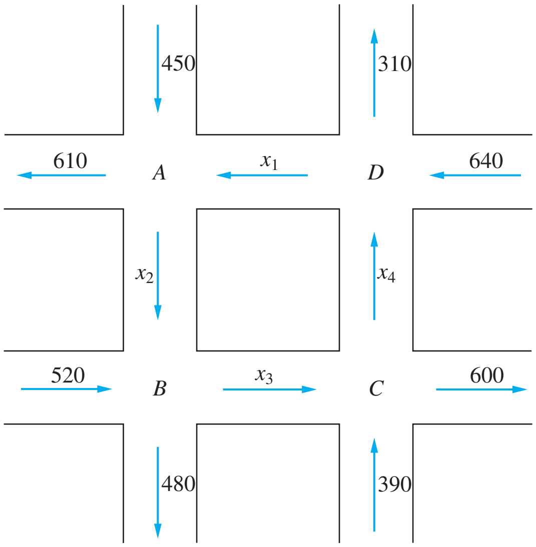


Figure 5.5: Traffic flow of a certain city

Solution: At each intersection point, the number of automobiles entering must be the same as the number leaving. For example, at the intersection D, the number of automobiles entering is $x_4 + 640$ and the number leaving is $x_1 + 310$. Thus,

$$x_4 + 640 = x_1 + 310 \quad (\text{intersection } D)$$

Similarly, we can get the other three equations from the other three intersections, i.e., A, B and C. The following system of equations is obtained:

$$\begin{cases} x_1 + 450 = x_2 + 610 \\ x_2 + 520 = x_3 + 480 \\ x_3 + 390 = x_4 + 600 \\ x_4 + 640 = x_1 + 310 \end{cases} \Longrightarrow \left\{ \begin{array}{rrrrcr} x_1- & x_2 & & & = & 160 \\ & x_2 - & x_3 & & = & -40 \\ & & x_3 & -x_4 & = & 210 \\ -x_1 & & & +x_4 & = & -330 \end{array} \right.$$

This can be solved by means of the augmented matrix:

$$\left(\begin{array}{rrrr|r} 1 & -1 & 0 & 0 & 160 \\ 0 & 1 & -1 & 0 & -40 \\ 0 & 0 & 1 & -1 & 210 \\ -1 & 0 & 0 & 1 & -330 \end{array}\right) \xrightarrow[+R_3+R_4]{R_4 \Rightarrow R_1+R_2} \left(\begin{array}{rrrr|r} 1 & -1 & 0 & 0 & 160 \\ 0 & 1 & -1 & 0 & -40 \\ 0 & 0 & 1 & -1 & 210 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

$$\xrightarrow{R_2 \Rightarrow R_2+R_3} \left(\begin{array}{rrrr|r} 1 & -1 & 0 & 0 & 160 \\ 0 & 1 & 0 & -1 & 170 \\ 0 & 0 & 1 & -1 & 210 \end{array}\right) \xrightarrow{R_1 \Rightarrow R_1+R_2} \left(\begin{array}{rrrr|r} 1 & 0 & 0 & -1 & 330 \\ 0 & 1 & 0 & -1 & 170 \\ 0 & 0 & 1 & -1 & 210 \end{array}\right),$$

i.e.,

$$\begin{cases} x_1 \qquad\quad - x_4 = & 330 \\ \quad x_2 \qquad - x_4 = & 170. \\ \qquad\quad x_3 - x_4 = & 210 \end{cases}$$

Introducing $x_4 = t$, we have

$$\begin{cases} x_1 = 330 + t, \\ x_2 = 170 + t, \\ x_3 = 210 + t, \\ x_4 = t, \end{cases} \qquad \text{with } t \in \mathbb{R}, \text{ a time parameter.}$$

The system is consistent; and since there is a free variable, there are many possible solutions — in agreement to our common sense. For instance, if $x_4 = t = 200$, we have

$$x_1 = 530, \qquad x_2 = 370, \qquad x_3 = 410.$$

□

Chapter 6　Matrix arithmatic

In the preceding chapter we saw the origin of matrices from solving systems of linear equations. In this chapter let us emphasize on the properties of the matrices themselves. We will study their addition, subtraction, scalar multiplication and matrix multiplication, as well as other operations with respective features.

§ 6.1　Definition and properties of matrices

§ 6.1.1　Matrix notation

***Definition* 6.1.** An $m \times n$ matrix is given by an array

$$A = (a_{ij}) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \tag{6.1.1}$$

where a_{ij} denotes the entries of the matrix, where $a_{ij} \in \mathbb{R}$ or $\mathbb{C}$, with $i = 1, \cdots, m$ and $j = 1, \cdots, n$. An equivalent and convenient notation is

$$A \in \mathbb{R}^{m\times n} \qquad \text{or} \qquad A \in \mathbb{C}^{m\times n}. \tag{6.1.2}$$

***Definition* 6.2** (Equity)**.** Let A and B be two $m \times n$ matrices. A is equal to B iff every entry of them are equal, i.e.,

$$A = B \iff a_{ij} = b_{ij}, \qquad \forall\ i = 1, \cdots, m;\ j = 1, \cdots, n. \tag{6.1.3}$$

***Definition* 6.3.** A *column vector*（列矢量）is an $m \times 1$ matrix.

An example is $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$.

***Definition* 6.4.** A *row vector*（行矢量）is a $1 \times m$ matrix.

An example is $\begin{pmatrix} 1, 2, 3 \end{pmatrix}$.

§ 6.1.2 Addition, subtraction and scalar multiplication of matrices

***Definition* 6.5** (Addition)**.** Let $A = (a_{ij})$ and $B = (b_{ij})$ be two matrices of the same size $m \times n$. Their sum is defined as

$$(A+B)_{ij} = a_{ij} + b_{ij}, \qquad \forall\ i = 1, \cdots, m;\ j = 1, \cdots, n. \tag{6.1.4}$$

***Definition* 6.6** (Subtraction)**.** Similarly,

$$(A-B)_{ij} = a_{ij} - b_{ij}, \qquad \forall\ i = 1, \cdots, m;\ j = 1, \cdots, n. \tag{6.1.5}$$

✎ ***Example* 6.1.**

$$\begin{pmatrix} 3 & 2 & 1 \\ 4 & 5 & 6 \end{pmatrix} + \begin{pmatrix} 2 & 2 & 2 \\ 1 & 2 & 3 \end{pmatrix} = \begin{pmatrix} 3+2 & 2+2 & 1+2 \\ 4+1 & 5+2 & 6+3 \end{pmatrix} = \begin{pmatrix} 5 & 4 & 3 \\ 5 & 7 & 9 \end{pmatrix}.$$

But $\begin{pmatrix} 3 & 2 & 1 \\ 4 & 5 & 6 \end{pmatrix}$ and $\begin{pmatrix} 2 & 1 \\ 2 & 2 \\ 2 & 3 \end{pmatrix}$ are not able to perform addition or subtraction, because they are not of the same size.

***Definition* 6.7** (Zero matrix)**.**

$$\mathbf{0} = \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{pmatrix}_{m \times n} \tag{6.1.6}$$

Obviously,

$$\mathbf{0} \pm A = \pm A, \qquad A \pm \mathbf{0} = \pm A, \qquad \text{where } \mathbf{0}, A \in \mathbb{R}^{m \times n}.$$

***Definition* 6.8** (Scalar multiplication)**.** Let A be a matrix and let α be a scalar, $\alpha \in \mathbb{R}, A \in \mathbb{R}^{m \times n}$. Then their scalar multiplication is defined as

$$\alpha A = \begin{pmatrix} \alpha a_{11} & \cdots & \alpha a_{1n} \\ \vdots & & \vdots \\ \alpha a_{m1} & \cdots & \alpha a_{mn} \end{pmatrix} = (\alpha a_{ij}). \tag{6.1.7}$$

✎ ***Example* 6.2.**

$$3 \begin{pmatrix} 4 & 8 & 2 \\ 6 & 8 & 10 \end{pmatrix} = \begin{pmatrix} 12 & 24 & 6 \\ 18 & 24 & 30 \end{pmatrix}.$$

§ 6.1.3 Matrix multiplication

In this subsection we will learn matrix multiplication. Please pay full attention to its difference from the scalar multiplication.

***Definition* 6.9** (Matrix multiplication)**.** Let $A = (a_{ij})$ be an $m \times n$ matrix and $B = (b_{jk})$ an $n \times r$ matrix. Their product $AB = C = (c_{ik})$ is defined as

$$c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}, \qquad i = 1, \cdots, m;\ j = 1, \cdots, n. \tag{6.1.8}$$

i.e.,

$$\begin{pmatrix} & \cdots & \\ \vdots & c_{ik} & \vdots \\ & \cdots & \end{pmatrix} = \begin{pmatrix} & & \cdots\cdots\cdots & \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ & & \cdots\cdots\cdots & \end{pmatrix} \begin{pmatrix} \vdots & b_{1k} & \vdots \\ \vdots & b_{2k} & \vdots \\ \vdots & \vdots & \vdots \\ \vdots & b_{nk} & \vdots \end{pmatrix} \tag{6.1.9}$$

with

$$c_{ik} = a_{i1} b_{1k} + a_{i2} b_{2k} + \cdots + a_{in} b_{nk} = \sum_{j=1}^{n} a_{ij} b_{jk}. \tag{6.1.10}$$

You may find that (6.1.10) looks like an inner product; we will discuss this in detail in the coming subsection §6.4.

✎ ***Example* 6.3.** Consider two matrices

$$A = \begin{pmatrix} 3 & -2 \\ 2 & 4 \\ 1 & -3 \end{pmatrix}, \qquad B = \begin{pmatrix} -2 & 1 & 3 \\ 4 & 1 & 6 \end{pmatrix}.$$

Their multiplication AB is permitted to perform, since the number of the columns of A equals to that of the rows of B. The result AB should be a 3×3 matrix:

$$\begin{aligned} AB &= \left(\begin{array}{c} 3 \quad -2 \\ \boxed{2 \quad\ \ 4} \\ 1 \quad -3 \end{array}\right) \left(\begin{matrix} -2 & 1 \\ 4 & 1 \end{matrix}\ \boxed{\begin{matrix} 3 \\ 6 \end{matrix}}\right) \\ &= \begin{pmatrix} 3 \times (-2) + (-2) \times 4 & 3 \times 1 + (-2) \times 1 & 3 \times 3 + (-2) \times 6 \\ 2 \times (-2) + \quad 4 \ \times 4 & 2 \times 1 + \quad 4 \ \times 1 & \boxed{2 \times 3 + \quad 4 \ \times 6} \\ 1 \times (-2) + (-3) \times 4 & 1 \times 1 + (-3) \times 1 & 1 \times 3 + (-3) \times 6 \end{pmatrix} \\ &= \begin{pmatrix} -14 & 1 & -3 \\ 12 & 6 & \boxed{30} \\ -14 & -2 & -15 \end{pmatrix}, \end{aligned}$$

where the blue boxes highlight the procedure of producing a particular entry–$(2, 3)$ of the matrix AB.

✎ ***Example* 6.4.** Similarly, the multiplication

$$BA = \begin{pmatrix} -2 & 1 & 3 \\ 4 & 1 & 6 \end{pmatrix} \begin{pmatrix} 3 & -2 \\ 2 & 4 \\ 1 & -3 \end{pmatrix}$$

is also permitted to perform. The result is a 2×2 matrix,

$$BA = \begin{pmatrix} -1 & -1 \\ 20 & -22 \end{pmatrix}.$$

A fact that should be stressed is that, as we see,

$$AB \neq BA,$$

where AB is of size 3×3 while BA is of 2×2. This non-commutativity between two matrices is *very common* in matrix arithmetic.

Corollary 1. *If both AB and BA are doable, there must be a fact that A has size $m \times n$ and B of size $n \times m$. It is forbidden that A has size $m \times n$ and B of size $n \times r$ but $m \neq r$.*

For instance, $\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{pmatrix}$ is permitted, but $\begin{pmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ is forbidden.

Computational rule

Scalar multiplication and matrix multiplication have higher priority than addition and subtraction operations, when parentheses are absent. For instance,

$$A + BC = \begin{pmatrix} 3 & 4 \\ 1 & 2 \end{pmatrix} + \underbrace{\begin{pmatrix} 1 & 3 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} -2 & 1 \\ 3 & 2 \end{pmatrix}}_{\text{multiplication first}} = \overbrace{\begin{pmatrix} 3 & 4 \\ 1 & 2 \end{pmatrix} + \begin{pmatrix} 7 & 7 \\ -1 & 4 \end{pmatrix}}^{\text{then addition}} = \begin{pmatrix} 10 & 11 \\ 0 & 6 \end{pmatrix}.$$

Application of matrix multiplication

A company manufactures three products. Its production expenses are divided into three categories. In each category, an estimate is given for the cost of producing a single item of each product. An estimate is also made of the amount of each product to be produced per quarter. These estimates are given in Tables 6.1 and 6.2. At its stockholders' meeting, the company would like to present a single table showing the total costs for each quarter in each of the three categories: raw materials, labor, and overhead.

Matrix M - Production cost as per each item A, B and C:

	A	B	C
Raw material	0.10	0.30	0.15
Labor	0.30	0.40	0.25
Overhead and Misscellaneous	0.10	0.20	0.15

Table 6.1: Production cost as per item A, B and C

Matrix P-Amount of each product produced per quarter (every season):

	Season			
	Summer	Fall	Winter	Spring
Product A	4000	4500	4500	4000
Product B	2000	2600	2400	2200
Product C	5800	6200	6000	6000

Table 6.2: Amount of each product produced per quarter

Therefore, in order to obtain a table summing up all the cost of the products A, B and C (as follows), we need to conduct a matrix multiplication as follows:

$$MP = \begin{pmatrix} 0.10 & 0.30 & 0.15 \\ 0.30 & 0.40 & 0.25 \\ 0.10 & 0.20 & 0.15 \end{pmatrix} \begin{pmatrix} 4000 & 4500 & 4500 & 4000 \\ 2000 & 2600 & 2400 & 2200 \\ 5800 & 6200 & 6000 & 6000 \end{pmatrix}$$

The result forms a 3×4 matrix. See Table 6.3.

	Summer	Fall	Winter	Spring	Yearly sum
Raw material	1870	2160	2070	1960	8060
Labor	3450	3940	3810	3580	14780
Overhead and Misscellaneous	1670	1900	1830	1740	7140
Total cost:	6990	8000	7710	7280	29980

Table 6.3: Cost Accounting

§ 6.2 Matrix representation of a system of linear equations

In this subsection the matrix representation of a linear system will be presented. Let us begin with the multiplication of an $m \times n$ matrix and an $n \times 1$ matrix (i.e., a column vector). Their product is an $m \times 1$ matrix (also a column vector):

$$\begin{pmatrix} \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \end{pmatrix} \begin{pmatrix} \bullet \\ \bullet \\ \bullet \\ \bullet \end{pmatrix} = \begin{pmatrix} \bullet \\ \bullet \end{pmatrix}$$

This is exactly what happens in the case of a linear system:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 = b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 = b_2 \end{cases} \tag{6.2.11}$$

$$\Longleftrightarrow \quad \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}. \tag{6.2.12}$$

Hence (6.2.12) is a matrix representation for (6.2.11).

In the language of matrices, a linear system

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = b_2 \\ \qquad\qquad \vdots & \quad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = b_m \end{cases} \tag{6.2.13}$$

has a matrix representation

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}, \tag{6.2.14}$$

and then be precisely expressed by

$$A\mathbf{x} = \mathbf{b}, \tag{6.2.15}$$

where

$$A \in \mathbb{R}^{m\times n}, \qquad \mathbf{x} \in \mathbb{R}^n, \qquad \mathbf{b} \in \mathbb{R}^m. \tag{6.2.16}$$

A further view of a linear system in terms of inner products will be given in Section §6.4.

§ 6.3 Transpose of matrix

***Definition* 6.10.** Let $A = (a_{ij})$ be an $m \times n$ matrix, and $B = (b_{ji})$ be an $n \times m$ matrix. B is the transpose of A iff

$$b_{ji} = a_{ij}, \qquad \forall\ i = 1, \cdots, m;\ j = 1, \cdots, n. \tag{6.3.17}$$

B is denoted as A^T.

✎ ***Example* 6.5.** Let $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$. Then $A^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$.

Specially, the transpose of an $m \times 1$ matrix (column vector) is a $1 \times m$ matrix (row vector); and vice versa.

✎ ***Example* 6.6.** $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}^T = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix}$, and $\begin{pmatrix} 1 & 2 & 3 \end{pmatrix}^T = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$.

This agrees to our past knowledge about the transpose of a vector.

***Theorem* 6.1.** Algebraic properties of matrix transposes:

1. $A^T = A$
2. $(\alpha A)^T = \alpha A^T$
3. $(A + B)^T = A^T + B^T$
4. $(AB)^T = B^T A^T$

Proof:

The proofs for Properties 1–3 are straightforward, by simply considering the locations of the entries. Proof for Property 4: Let $A = (a_{ij}),\ B = (b_{jk}),\ i = 1, \cdots, m,\ j = 1, \cdots, n,\ k = 1, \cdots, r$. Then,

$$\begin{aligned} AB &= (c_{ik}), & &\text{with } c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}, \\ (AB)^T &= (d_{ki}), & &\text{with } d_{ki} = c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}. \end{aligned} \tag{6.3.18}$$

On the other hand,

$$B^T A^T = \sum_{j=1}^{n} b_{kj} a_{ji} = d_{ki} = c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}. \tag{6.3.19}$$

Comparing (6.3.18) and (6.3.19) yields

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

□

✎ ***Example* 6.7.** Use the following matrices to verify $(AB)^T = B^T A^T$.

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 3 & 3 & 5 \\ 2 & 4 & 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 0 & 2 \\ 2 & 1 & 1 \\ 5 & 4 & 1 \end{pmatrix}.$$

Solution:

$$\begin{aligned} AB &= \begin{pmatrix} 10 & 6 & 5 \\ 34 & 23 & 14 \\ 15 & 8 & 9 \end{pmatrix}, \qquad \text{hence} \qquad (AB)^T = \begin{pmatrix} 10 & 34 & 15 \\ 6 & 23 & 8 \\ 5 & 14 & 9 \end{pmatrix}; \\ B^T A^T &= \begin{pmatrix} 1 & 2 & 5 \\ 0 & 1 & 4 \\ 2 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 & 2 \\ 2 & 3 & 4 \\ 1 & 5 & 1 \end{pmatrix} = \begin{pmatrix} 10 & 34 & 15 \\ 6 & 27 & 8 \\ 5 & 14 & 9 \end{pmatrix}. \end{aligned}$$

□

Symmetric and anti-symmetric matrices

Symmetric and anti-symmetric matrices have broad applications in science and technology, such as in physics (e.g., condensed matter physics, cosmology) and electricity & electronics.

***Definition* 6.11** (*Symmetric matrix*（对称矩阵）)**.** A matrix A is said to be symmetric iff

$$A^T = A. \tag{6.3.20}$$

Obviously, a symmetric matrix must be a square matrix, i.e., of the size $n \times n$.

✎ ***Example* 6.8.** Below are two examples of symmetric matrices.

$$\begin{pmatrix} 2 & 3 & 4 \\ 3 & 1 & 5 \\ 4 & 5 & 3 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 6 \\ 6 & -4 \end{pmatrix}.$$

***Theorem* 6.2.** Let A be a symmetric matrix, $A \in \mathbb{R}^{n\times n}$, i.e., A has n^2 entries. Then A has $\frac{n(n+1)}{2}$ degrees of freedom, i.e., A has $\frac{n(n+1)}{2}$ free entries.

***Definition* 6.12** (*Anti-symmetric matrix*（反对称矩阵）)**.** A matrix is said to be anti-symmetric iff

$$A^T = -A. \tag{6.3.21}$$

An anti-symmetric matrix is also a square matrix.

✎ ***Example* 6.9.** Below is an example of anti-symmetric matrix.

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

***Theorem* 6.3.** Let A be an anti-symmetric matrix, $A \in \mathbb{R}^{n\times n}$, i.e., A has n^2 entries. Then:

- The diagonal entries of A are all zeroes,

$$\begin{pmatrix} 0 & \bullet & \bullet \\ \bullet & \ddots & \bullet \\ \bullet & \bullet & 0 \end{pmatrix}.$$

- A has $\frac{n(n-1)}{2}$ degrees of freedom, i.e., A has $\frac{n(n-1)}{2}$ free entries.

✎ ***Example* 6.10.** (Networks and graph theory. See Page 54 of *S. Leon, 2010*.)

Graph theory is an important area of applied mathematics. It is used to model problems in virtually all the applied sciences. Graph theory is particularly useful in applications involving communication networks.

A *graph* is defined to be a set of points called *vertices*, together with a set of unordered pairs of vertices, which are referred to as edges. Figure 6.1 gives a geometrical representation of a graph. We can think of the vertices V_1, V_2, V_3, V_4, and V_5 as corresponding to the nodes in a communication network.

The line segments joining the vertices correspond to the edges:

$$\{V_1, V_2\}\ , \{V_2, V_5\}\ , \{V_3, V_4\}\ , \{V_3, V_5\}\ , \{V_4, V_5\}$$

Each edge represents a direct communication link between two nodes of the network.

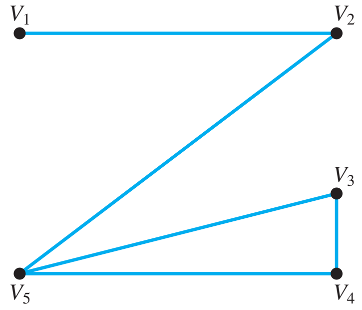


Figure 6.1: Networks and graphs

An actual communication network could involve a large number of vertices and edges. Indeed, if there are millions of vertices, a graphical picture of the network would be quite confusing. An alternative is to use a matrix representation for the network. If the graph contains a total of n vertices, we can define an $n \times n$ matirx A by

$$a_{ij} = \begin{cases} 1 & \text{if } \{V_i, V_j\} \text{ is an edge of the graph} \\ 0 & \text{if three is no edge joining of the vertices } V_i \text{ and } V_j \end{cases}$$

The matrix A is called the *adjacency matrix* of the graph. The adjacency matrix for the graph in Figure 6.1 is given by

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

Note that the matrix A is symmetric. Indeed, any adjacency matrix must be symmetric, for if $\{V_i, V_j\}$ is an edge of the graph, then $a_{ij} = a_{ji} = 1$ and $a_{ij} = a_{ji} = 0$ if there is no edge joining V_i and V_j. In either case, $a_{ij} = a_{ji}$.

We can think of a walk on a graph as a sequence of edges linking one vertex to another. For example, in Figure 6.1 the edges $\{V_1, V_2\}$, $\{V_2, V_5\}$ represent a walk from vertex V_1 to vertex V_5. The length of the walk is said to be 2, since it consists of two edges. A simple way to describe the walk is to indicate the movement between vertices by arrows. Thus, $V_1 \longrightarrow V_2 \longrightarrow V_5$ denotes a walk of length 2 from V_1 to V_5. Similarly, $V_4 \longrightarrow V_5 \longrightarrow V_2 \longrightarrow V_1$ represents a walk of length 3 from V_4 to V_1. It is possible to traverse the same edges more than once in a walk. For example, $V_5 \longrightarrow V_3 \longrightarrow V_5 \longrightarrow V_3$ is a walk of length 3 from V_5 to V_3. In general, by taking powers of the adjacency matrix, we can determine the number of walks of any specified length between two vertices.

§ 6.4 New view of matrix in term of inner product of vectors

After achieving basic knowledge of matrices and vectors, we are at the point to have a new view of a matrix. Let A be an $m \times n$ matrix,

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

Regarding every column of A as a column matrix, we have

$$A = \begin{pmatrix} \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} & \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} & \cdots & \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} \end{pmatrix}. \tag{6.4.22}$$

Introducing a set of column vectors $\mathbf{v}_j$, with $j = 1, \cdots, n$,

$$\mathbf{v}_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \ \mathbf{v}_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix}, \cdots, \mathbf{v}_n = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} \tag{6.4.23}$$

the A of (6.4.22) is rewritten as

$$A = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{pmatrix}. \tag{6.4.24}$$

This is a column vector representation for the matrix A.

Alternatively, one can also regard every row of A as a row matrix, and rewrite A as

$$A = \begin{pmatrix} (a_{11} & a_{12} & \cdots & a_{1n}) \\ (a_{21} & a_{22} & \cdots & a_{2n}) \\ \vdots & & & \vdots \\ (a_{m1} & a_{m2} & \cdots & a_{mn}) \end{pmatrix}. \tag{6.4.25}$$

Introducing a set of row vectors $\mathbf{w}_i^T$, $\ i = 1, \cdots, m$,

$$\begin{aligned} \mathbf{w}_1^T &= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \end{pmatrix}, \\ \mathbf{w}_2^T &= \begin{pmatrix} a_{21} & a_{22} & \cdots & a_{2n} \end{pmatrix}, \\ &\ \vdots \\ \mathbf{w}_m^T &= \begin{pmatrix} a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \end{aligned} \tag{6.4.26}$$

where the symbol "T" stands for the transpose of a column matrix, we achieve

$$A = \begin{pmatrix} \mathbf{w}_1^T \\ \mathbf{w}_2^T \\ \vdots \\ \mathbf{w}_m^T \end{pmatrix}. \tag{6.4.27}$$

This is a row vector representation for the matrix A.

✎ ***Example* 6.11.** Consider a matrix $A = \begin{pmatrix} 3 & 2 & 5 \\ -1 & 8 & 4 \end{pmatrix}$. Its column and row vector representations are given by, respectively,

$$A = \begin{cases} \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{pmatrix}, & \text{with } \mathbf{v}_1 = \begin{pmatrix} 3 \\ -1 \end{pmatrix}, \ \mathbf{v}_2 = \begin{pmatrix} 2 \\ 8 \end{pmatrix}, \ \mathbf{v}_3 = \begin{pmatrix} 5 \\ 4 \end{pmatrix}, \\ \begin{pmatrix} \mathbf{w}_1^T \\ \mathbf{w}_2^T \end{pmatrix}, & \text{with } \mathbf{w}_1^T = \begin{pmatrix} 3 & 2 & 5 \end{pmatrix}, \ \mathbf{w}_2^T = \begin{pmatrix} -1 & 8 & 4 \end{pmatrix}. \end{cases} \tag{6.4.28}$$

New look at linear systems in terms of inner product of vectors

In the light of the row vector representation (6.4.27), a linear system $A\mathbf{x} = \mathbf{b}$ can be re-expressed as

$$\begin{pmatrix} \mathbf{w}_1^T \\ \mathbf{w}_2^T \\ \vdots \\ \mathbf{w}_m^T \end{pmatrix} \mathbf{x} = \mathbf{b}, \qquad \text{i.e.,} \qquad \begin{pmatrix} \mathbf{w}_1^T\mathbf{x} \\ \vdots \\ \mathbf{w}_m^T\mathbf{x} \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}, \tag{6.4.29}$$

where $\mathbf{w}_i^T$ is a row matrix and $\mathbf{x}$ a column matrix, and

$$\mathbf{w}_i^T\mathbf{x} = \begin{pmatrix} a_{i1} & a_{i_2} & \cdots & a_{in} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = b_i. \tag{6.4.30}$$

This is reminiscent of an inner product between $\mathbf{w}_i^T$ and $\mathbf{x}$, hence $\mathbf{w}_i^T\mathbf{x}$ is a matrix realization of the inner product $\mathbf{w}_i^T \cdot \mathbf{x}$.

Furthermore, for two matrices

$$A = (a_{ij}), \quad B = (b_{jk}), \qquad i = 1, \cdots, m, \ j = 1, \cdots, n, \ k = 1, \cdots, r,$$

let their product be $C = AB = (c_{ik})$. Then the entries of C can be re-expressed by inner products between row vectors and column vectors:

$$AB = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & \vdots & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1r} \\ b_{21} & b_{22} & \cdots & b_{2r} \\ & \vdots & & \\ b_{n1} & b_{n2} & \cdots & b_{nr} \end{pmatrix} \tag{6.4.31}$$

$$= \begin{pmatrix} \mathbf{w}_1^T \\ \mathbf{w}_2^T \\ \vdots \\ \mathbf{w}_m^T \end{pmatrix} \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_r \end{pmatrix} = \begin{pmatrix} \mathbf{w}_1^T\mathbf{v}_1 & \mathbf{w}_1^T\mathbf{v}_2 & \cdots & \mathbf{w}_1^T\mathbf{v}_r \\ \mathbf{w}_2^T\mathbf{v}_1 & \mathbf{w}_2^T\mathbf{v}_2 & \cdots & \mathbf{w}_2^T\mathbf{v}_r \\ & \vdots & & \\ \mathbf{w}_m^T\mathbf{v}_1 & \mathbf{w}_m^T\mathbf{v}_2 & \cdots & \mathbf{w}_m^T\mathbf{v}_r \end{pmatrix}, \tag{6.4.32}$$

where $\begin{pmatrix} \mathbf{w}_1^T \\ \vdots \\ \mathbf{w}_m^T \end{pmatrix}$ is the row representation of A, and $\begin{pmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_r \end{pmatrix}$ the column representation of B. Namely,

$$AB = C = (c_{ik}) = \left(\mathbf{w}_i^T\mathbf{v}_k\right), \qquad i = 1, \cdots, m;\ k = 1, \cdots, r. \tag{6.4.33}$$

§ 6.5 Theorem on consistency of linear system

Recollect the linear combination of n vectors:

$$\mathbf{a} = c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_n\mathbf{a}_n \qquad \text{with } \ c_j \in \mathbb{R}. \tag{6.5.34}$$

Suppose $\mathbf{a}_1 \cdots \mathbf{a}_n$ have column matrix realizations:

$$\mathbf{a}_1 = \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix}, \quad \cdots, \quad \mathbf{a}_n = \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix}.$$

$\mathbf{a} = c_1\mathbf{a}_1 + \cdots + c_n\mathbf{a}_n$ is called a linear combination of column matrices. Then we have a theorem on consistency of a system of linear equations.

***Theorem* 6.4.** [Consistency of linear system]
A linear system $A\mathbf{x} = \mathbf{b}$ is consistent iff $\mathbf{b}$ is a linear combination of the column vectors of A. Namely, with

$$A = \begin{pmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{pmatrix} = \left[\begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix} \cdots \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix} \right], \qquad \mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix},$$

the linear system $A\mathbf{x} = \mathbf{b}$ is consistent iff $\mathbf{b}$ can be expressed as a linear combination

$$\mathbf{b} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n. \tag{6.5.35}$$

An equivalent statement of this theorem is that $\mathbf{b}$ is a vector within the space spanned by the vectors $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$.

Proof: Let $\mathbf{x} = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix}^T$. The linear system $A\mathbf{x} = \mathbf{b}$ can be rewritten as

$$A\mathbf{x} = \begin{pmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n = \mathbf{b}. \tag{6.5.36}$$

This means the set of numbers $\{x_1, x_2, \cdots, x_n\}$ can be found iff $\mathbf{b}$ is a linear combination of the vectors $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$. ☐

✎ ***Example* 6.12.** Consider a linear system

$$\begin{cases} x_1 + 2x_2 = 1, \\ 2x_1 + 4x_2 = 1, \end{cases} \qquad \text{which is equivalent to} \qquad \begin{cases} x_1 + 2x_2 = 1, \\ x_1 + 2x_2 = \frac{1}{2}. \end{cases} \tag{6.5.37}$$

Obviously this system has no solution. Now let us use Theorem 6.4 to confirm this fact.

The system reads

$$\begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix}.$$

The coefficient matrix

$$A = (\mathbf{v}_1, \mathbf{v}_2), \quad \text{with} \quad \mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \mathbf{v}_2 = 2\begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

If the system is consistent, the vector $\mathbf{b}$ should be able to be expressed as a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$.

$$\mathbf{b} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2, \qquad \text{i.e.,} \qquad \begin{pmatrix} 1 \\ 1/2 \end{pmatrix} = c_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + 2c_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

i.e.,

$$\begin{cases} 1 = & c_1 + 2c_2, \\ \dfrac{1}{2} = & c_1 + 2c_2. \end{cases} \tag{6.5.38}$$

However, this induced linear system (6.5.38) is inconsistent, hence the original system (6.5.37) has no solution.

Chapter 7 Matrix Algebra

In this chapter we will learn some important matrices, including the identity matrix, inverse of matrix and three types of elementary matrices. Some important operations of matrices will be introduced as well.

§ 7.1 Matrix Algebra

§ 7.1.1 Properties

Let A, B, C be matrices, and α, β scalars. A, B, C are of appropriate sizes for the operations below.

- Commutativity:
$$A + B = B + A.$$
- Associativity (for addition):
$$(A + B) + C = A + (B + C).$$
- Associativity (for multiplication):
$$(AB)C = A(BC).$$
- Distributivity (for matrices):
$$\begin{aligned} A(B+C) &= AB + AC, \\ (A+B)C &= AC + BC. \end{aligned}$$
- Scalar multiplication:
$$\begin{aligned} (\alpha\beta)A &= \alpha(\beta A), \\ \alpha(AB) &= (\alpha A)B = A(\alpha B). \end{aligned}$$
- Distributivity (for scalars with matrices):
$$\begin{aligned} (\alpha+\beta)A &= \alpha A + \beta A, \\ \alpha(A+B) &= \alpha A + \alpha B. \end{aligned}$$

Proof: Ignored. □

[Remark]: A notation:

$$A^k = \underbrace{AA\cdots A}_{k \text{ copies}}.$$

✎ ***Example* 7.1.** Let us pay attention to an interesting matrix

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}. \tag{7.1.1}$$

It can be checked that

$$\begin{aligned} A^2 &= \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}, \\ A^3 &= \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} = \begin{pmatrix} 4 & 4 \\ 4 & 4 \end{pmatrix}, \\ &\vdots \\ A^n &= \begin{pmatrix} 2^{n-1} & 2^{n-1} \\ 2^{n-1} & 2^{n-1} \end{pmatrix} = 2^{n-1}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}. \end{aligned} \tag{7.1.2}$$

§ 7.1.2 Identity matrix

***Definition* 7.1** (Identity matrix)**.** Let $I = (a_{ij})$ be an $n \times n$ matrix, $i, j = 1, \cdots, n$. I is an identity matrix iff

$$a_{ij} = \begin{cases} 1, & \text{if} \quad i = j, \\ 0, & \text{if} \quad i \neq j. \end{cases} \tag{7.1.3}$$

The identity matrix I has an explicit form,

$$I = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix}. \tag{7.1.4}$$

This matrix is named after the following property

$$IA = AI = A, \qquad \forall A \in \mathbb{R}^{n\times n}. \tag{7.1.5}$$

[Remark]:

1. An identity matrix acts as a "1" in matrices, hence people intend to denote it as $\mathbf{1}$.

2. Sometimes I is written as I_n or $I_{n\times n}$, for the purpose of highlighting its size.

Knonecker's symbol δ_{ij}:

***Definition* 7.2** (Knonecker δ function)**.**

$$\delta_{ij} = \begin{cases} 1, & \text{when } i = j. \\ 0, & \text{when } i \neq j. \end{cases} \quad (i, j \in \mathbb{N}) \tag{7.1.6}$$

The Knonecker δ-function has important applications in mathematics. The entries of the identity matrix I can be exactly expressed by

$$I = (\delta_{ij}), \qquad i, j = 1, \cdots, n. \tag{7.1.7}$$

Unit vector $\mathbf{e}_j$:

***Definition* 7.3.** A vector

$$\mathbf{e}_j = \begin{pmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{pmatrix}^T, \tag{7.1.8}$$

i.e., only the jth entry is 1 while all the others are 0.

This unit vector $\mathbf{e}_j$ has wide applications. For instance, the identity matrix I can be delivered by $\mathbf{e}_j$'s:

$$I = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix} = \left[\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \\ 0 \end{pmatrix} \cdots \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \right].$$

i.e.,

$$I = \begin{pmatrix} \mathbf{e}_1, & \mathbf{e}_2, & \cdots, & \mathbf{e}_n \end{pmatrix}. \tag{7.1.9}$$

§ 7.1.3 Matrix inversion

A nonzero real number a has an inverse

$$b = a^{-1} \quad \text{if} \quad ab = 1, \qquad \text{where} \quad a \neq 0.$$

Similarly, a matrix may also possess its inverse.

***Definition* 7.4** (Inverse of matrix)**.** Let A be an $n \times n$ matrix. A is said to be invertible iff there exists another matrix B such that

$$AB = BA = I. \tag{7.1.10}$$

This B, also of size $n \times n$, is called the (multiplicative) inverse of A, denoted as A^{-1}.

***Theorem* 7.1.** The inverse of a matrix is unique if it exists.

Proof:

Suppose conversely there exist two $n \times n$ matrices B and C satisfying $AB = BA = I$ and $AC = CA = I$. Then we have

$$B = BI = B(AC) = IC = C, \qquad \text{namely } B \text{ and } C \text{ are identical.}$$

□

✎ ***Example* 7.2.** Let $A = \begin{pmatrix} 2 & 4 \\ 3 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} -\frac{1}{10} & \frac{2}{5} \\ \frac{3}{10} & -\frac{1}{5} \end{pmatrix}$. It is easy to check that

$$AB = BA = I, \qquad \text{i.e.,} \qquad B = A^{-1}.$$

✎ ***Example* 7.3.** Let $A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & -2 & 5 \\ 0 & 1 & -4 \\ 0 & 0 & 1 \end{pmatrix}$ be two matrices in echelon form. It is seen that

$$AB = BA = I, \qquad \text{i.e.,} \qquad B = A^{-1}.$$

✎ ***Example* 7.4.** A counter example（反例）:

The matrix $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ is not invertible.

Proof: (By contradiction（反证法）)

Suppose conversely that A has an inverse $B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$. Then there should be

$$BA = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} b_{11} & 0 \\ b_{21} & 0 \end{pmatrix} \neq \begin{pmatrix} 1 & \\ & 1 \end{pmatrix} = I,$$

Contradiction. Hence the matrix A is not invertible.

□

***Definition* 7.5** (Singularity)**.** An $n \times n$ matrix is said to be singular iff it is not invertible.

Therefore, an invertible matrix is also called a nonsingular matrix.

[Remark]:

How to tell a matrix is singular or nonsingular? We will learn shortly

$$\det A \begin{cases} \neq 0, & A \text{ invertible}, \quad \text{i.e.,} \quad \text{nonsingular} \;, \\ = 0, & A \text{ singular}, \quad \text{i.e.,} \quad \text{having no inverse} \;, \end{cases} \tag{7.1.11}$$

where $\det A$ is the determinant（行列式）of A. See the next chapter on *Determinants*.

***Theorem* 7.2.** If A and B are both nonsingular $n \times n$ matrices, then their product AB is also nonsingular (i.e., invertible),

$$(AB)^{-1} = B^{-1}A^{-1}. \tag{7.1.12}$$

Proof: It can be verified that

$$(AB)^{-1}(AB) = B^{-1}A^{-1}AB = I, \qquad (AB)(AB)^{-1} = ABB^{-1}A^{-1} = I.$$

□

Generally, for k copies of $n \times n$ matrices,

$$(A_1A_2\cdots A_k)^{-1} = A_k^{-1}A_{k-1}^{-1}\cdots A_2^{-1}A_1^{-1}, \qquad k \in \mathbb{N}. \tag{7.1.13}$$

In particular,

$$(A^k)^{-1} = (A^{-1})^k = A^{-k}. \tag{7.1.14}$$

§ 7.2 Elementary matrices

In this section, we will see the functioning（功能）of a matrix when it is left- or right-multiplied to another matrix. We will learn a particular kind of matrices, the *elementary matrices*（初等矩阵）, which are able to realize the elementary row operations (ERO's) in the way of matrix left-multiplications. In terms of these elementary matrices, the inverse of a given matrix can be obtained; moreover, they are also key in the so-called *Lower-Upper (LU) decompositions*（上下分解）, which are a particular type of triangular factorizations（三角分解）of matrices.

§ 7.2.1 Three types of elementary matrices

Description of mission

For a given non-singular $n \times n$ matrix, how to find its inverse? Previously, when solving a linear system, we used to adopt the Gauss-Jordan reduction.

$$\boxed{\begin{array}{c}\text{Original linear system } \xrightarrow[\text{row operations}]{\text{elementary}} \text{Row echelon form} \\ \xrightarrow[\text{row operations}]{\text{elementary}} \text{Reduced row echelon form}\end{array}}$$

However, these "elementary row operations" are expressed merely in words. Can we perform them in terms of matrices? The answer is yes; what we need are the so-called elementary matrices of three types, corresponding to the three types of elementary row operations.

Intuitive view of matrix functioning

Let us begin with an example to get an intuitive view of the functioning of an elementary matrix, where you can see that the locations of the entries really matter.

✎ ***Example* 7.5.**

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \xrightarrow{R_1 \longleftrightarrow R_2} \begin{pmatrix} a_{21} & a_{22} & a_{23} \\ a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

This shows that a Type-I elementary row operation, swapping the first two rows of $A = (a_{ij})$, can be delivered by left-multiplying to A a matrix $\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. This method can be generalized to generic.

Elementary matrices

Type-I: Swapping two rows: $R_i \longleftrightarrow R_j$:

$$\begin{array}{cc}
 & \begin{array}{ccccccccc} & & & i^{\text{th}} & & j^{\text{th}} & & & \end{array} \\
\begin{array}{c} \\ \\ \\ i^{\text{th}} \\ \\ j^{\text{th}} \\ \\ \\ \\ \end{array} &
\begin{pmatrix}
1 & & & & & & & & \\
 & \ddots & & & & & & & \\
 & & 1 & & & & & & \\
 & & & 0 & \cdots & 1 & & & \\
 & & & \vdots & & \vdots & & & \\
 & & & 1 & \cdots & 0 & & & \\
 & & & & & & 1 & & \\
 & & & & & & & \ddots & \\
 & & & & & & & & 1
\end{pmatrix}
\end{array}
\begin{pmatrix}
\begin{matrix} a_{11} & a_{12} & \cdots & a_{1n} \end{matrix} \\
\begin{matrix} \vdots & \vdots & & \vdots \end{matrix} \\
\boxed{\begin{matrix} a_{i1} & a_{i2} & \cdots & a_{in} \end{matrix}} \\
\begin{matrix} \vdots & \vdots & & \vdots \end{matrix} \\
\boxed{\begin{matrix} a_{j1} & a_{j2} & \cdots & a_{jn} \end{matrix}} \\
\begin{matrix} \vdots & \vdots & & \vdots \end{matrix} \\
\begin{matrix} a_{m1} & a_{m2} & \cdots & a_{mn} \end{matrix}
\end{pmatrix}$$

$$\xlongequal{R_i\longleftrightarrow R_j}
\begin{pmatrix}
\begin{matrix} a_{11} & a_{12} & \cdots & a_{1n} \end{matrix} \\
\begin{matrix} \vdots & \vdots & & \vdots \end{matrix} \\
\boxed{\begin{matrix} a_{j1} & a_{j2} & \cdots & a_{jn} \end{matrix}} \\
\begin{matrix} \vdots & \vdots & & \vdots \end{matrix} \\
\boxed{\begin{matrix} a_{i1} & a_{i2} & \cdots & a_{in} \end{matrix}} \\
\begin{matrix} \vdots & \vdots & & \vdots \end{matrix} \\
\begin{matrix} a_{m1} & a_{m2} & \cdots & a_{mn} \end{matrix}
\end{pmatrix}. \tag{7.2.1}$$

Type-II: Rescaling a row R_i by a scalar α : $R_i \Longrightarrow \alpha R_i$

$$\begin{array}{cc}
 & \begin{array}{ccccccc} & & & i^{th} & & & \end{array} \\
\begin{array}{c} \\ \\ \\ i^{th} \\ \\ \\ \\ \end{array} &
\begin{pmatrix}
1 & & & & & & \\
 & \ddots & & & & & \\
 & & 0 & & & & \\
 & & & \alpha & & & \\
 & & & & 1 & & \\
 & & & & & \ddots & \\
 & & & & & & 1
\end{pmatrix}
\end{array}
\begin{pmatrix}
\begin{matrix} a_{11} & a_{12} & \cdots & a_{1n} \end{matrix} \\
\begin{matrix} \vdots & \vdots & & \vdots \end{matrix} \\
\boxed{\begin{matrix} a_{i1} & a_{i2} & \cdots & a_{in} \end{matrix}} \\
\begin{matrix} \vdots & \vdots & & \vdots \end{matrix} \\
\begin{matrix} a_{m1} & a_{m2} & \cdots & a_{mn} \end{matrix}
\end{pmatrix}$$

$$\xlongequal{R_i\Longrightarrow\alpha R_i}\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ \alpha a_{i1} & \alpha a_{i2} & \cdots & \alpha a_{in} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \tag{7.2.2}$$

✎ ***Example* 7.6.** Multiply the third row with a number 3,

$$\begin{pmatrix} 1 & & \\ & 1 & \\ & & 3 \end{pmatrix}\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 3a_{31} & 3a_{32} & 3a_{33} \end{pmatrix}.$$

Type-III: Multiplying one row with α and adding it to another row: $R_i \Longrightarrow R_i + \alpha R_j$.

$$\begin{array}{cc} & \begin{matrix} & & i^{th} & & j^{th} & & \end{matrix} \\ \begin{matrix} \\ \\ j^{th} \\ \\ \\ \\ \\ \end{matrix} & \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & \cdots & \alpha & & \\ & & & \ddots & \vdots & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \end{array} \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & & \vdots \\ a_{j1} & a_{j2} & \cdots & a_{jn} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

$$\xlongequal{R_i\Longrightarrow R_i+\alpha R_j}\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{i1}+\alpha a_{j1} & a_{i2}+\alpha a_{j2} & \cdots & a_{in}+\alpha a_{jn} \\ \vdots & \vdots & & \vdots \\ a_{j1} & a_{j2} & \cdots & a_{jn} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}. \tag{7.2.3}$$

✎ ***Example* 7.7.** Multiply a number 3 to the third row and then add it to the first row,

$$\begin{pmatrix} 1 & & 3 \\ & 1 & \\ & & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{11}+3a_{31} & a_{12}+3a_{32} & a_{13}+3a_{33} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

The elementary matrices are summarized as follows.

***Definition* 7.6.** Three types of *elementary matrices*, to realize elementary row operations:

$$E_{\text{I}} = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & \cdots & 1 & & \\ & & \vdots & & \vdots & & \\ & & 1 & \cdots & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}, \tag{7.2.4}$$

$$E_{\text{II}} = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & \alpha & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix}, \tag{7.2.5}$$

$$E_{\text{III}} = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & \cdots & \alpha & & \\ & & & \ddots & \vdots & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}. \tag{7.2.6}$$

☞ **[Aside]:**

How about right-multiplying an elementary matrix to another matrix? It realizes *elementary column operations*. See Page 59 of *S. Leon, 2010*.

Type-I: Swapping two columns C_i and C_j:

$$\begin{array}{c} C_i \xleftrightarrow{\text{exchange}} C_j \\ \begin{pmatrix} a_{11} & \cdots & a_{1i} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2i} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & & \vdots & & \vdots & & \vdots \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{m1} & \cdots & a_{mi} & \cdots & a_{mj} & \cdots & a_{mn} \end{pmatrix} \end{array} \begin{array}{cc} \begin{matrix} & & & i^{th} & & j^{th} & & & \end{matrix} & \\ \begin{pmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 1 & & & & & & \\ & & & 0 & \cdots & 1 & & & \\ & & & \vdots & & \vdots & & & \\ & & & 1 & \cdots & 0 & & & \\ & & & & & & 1 & & \\ & & & & & & & \ddots & \\ & & & & & & & & 1 \end{pmatrix} & \begin{matrix} \\ \\ \\ i^{th} \\ \\ j^{th} \\ \\ \\ \\ \end{matrix} \end{array}$$

$$= \begin{pmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1i} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2j} & \cdots & a_{2i} & \cdots & a_{2n} \\ \vdots & & \vdots & & & & \\ a_{m1} & \cdots & a_{mj} & \cdots & a_{mi} & \cdots & a_{mn} \end{pmatrix}.$$

i.e.,

$$AE^{\text{I}} = \tilde{A}, \tag{7.2.7}$$

where E^{I} is a Type-I elementary matrix, and $\tilde{A}$ a matrix column-equivalent to A.

Type-II: Rescaling a column C_i with a scalar α:

$$\begin{array}{c} C_i \longrightarrow \alpha C_i \\ \begin{pmatrix} a_{11} & \cdots & a_{1i} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2i} & \cdots & a_{2n} \\ \vdots & & \vdots & & \vdots \\ a_{m1} & \cdots & a_{mi} & \cdots & a_{mn} \end{pmatrix} \end{array} \begin{array}{cc} \begin{matrix} & & i^{th} & & \end{matrix} & \\ \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & \alpha & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix} & \begin{matrix} \\ \\ i^{th} \\ \\ \\ \end{matrix} \end{array}$$

$$= \begin{pmatrix} a_{11} & \cdots & \alpha a_{1i} & \cdots & a_{1n} \\ a_{21} & \cdots & \alpha a_{2i} & \cdots & a_{2n} \\ \vdots & & \vdots & & \vdots \\ a_{m1} & \cdots & \alpha a_{mi} & \cdots & a_{mn} \end{pmatrix}.$$

i.e.,

$$AE^{\text{ II}} = \tilde{A}, \tag{7.2.8}$$

where $E^{\text{ II}}$ is a Type-II elementary matrix, and $\tilde{A}$ a matrix column-equivalent to A.

Type-III: Multiplying a column C_j with a scalar α and then adding it to another column C_i:

$$C_i \Longrightarrow C_i + \alpha C_j$$

$$\begin{pmatrix} a_{11} & \cdots & a_{1i} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2i} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & & \vdots & & \vdots & & \vdots \\ \vdots & & \vdots & & \vdots & & \vdots \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{m1} & \cdots & a_{mi} & \cdots & a_{mj} & \cdots & a_{mn} \end{pmatrix} \begin{array}{cccccccc} & & i^{th} & & j^{th} & & & \\ 1 & & & & & & & \\ & \ddots & & & & & & \\ & & 1 & & & & & i^{th} \\ & & \vdots & \ddots & & & & \\ & & \alpha & \cdots & 1 & & & j^{th} \\ & & & & & \ddots & & \\ & & & & & & 1 & \end{array}$$

$$= \begin{pmatrix} a_{11} & \cdots & a_{1i} + \alpha a_{1j} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2i} + \alpha a_{2j} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & & \vdots & & & & \vdots \\ \vdots & & \vdots & & \vdots & & \vdots \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{m1} & \cdots & a_{mi} + \alpha a_{mj} & \cdots & a_{mj} & \cdots & a_{mn} \end{pmatrix}.$$

i.e.,

$$AE^{\text{ III}} = \tilde{A}, \tag{7.2.9}$$

where $E^{\text{ III}}$ is a Type-III elementary matrix, and $\tilde{A}$ a matrix column-equivalent to A.

✌

§ 7.2.2 Inverse of elementary matrices

The inverse of an elementary matrix is also an elementary matrix, as shown below.

- **Type-I**:

$$\begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & \cdots & 1 & & \\ & & \vdots & & \vdots & & \\ & & 1 & \cdots & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & \cdots & 1 & & \\ & & \vdots & & \vdots & & \\ & & 1 & \cdots & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}. \qquad (7.2.10)$$

[Intuitive understanding]: *Performing a swapping operation* $R_i \longleftrightarrow R_j$ *twice* recovers the original status. Hence the inverse of a Type-I elementary matrix is itself.

- **Type-II**:

$$\begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & \alpha & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & \frac{1}{\alpha} & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix}. \qquad (7.2.11)$$

[Intuitive understanding]: The function of a Type-II elementary matrix is to *make a rescale by* α, hence its inverse must be a rescale by $\frac{1}{\alpha}$ (for $\alpha \neq 0$).

- **Type-III**:

$$\begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & \cdots & \alpha & & \\ & & & \ddots & \vdots & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & \cdots & -\alpha & & \\ & & & \ddots & \vdots & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}. \qquad (7.2.12)$$

[Intuitive understanding]: The function of a Type-III elementary matrix is *to add a row* αR_j *to another row* R_i. Hence its inverse operation must be *subtracting an* αR_j *from* R_i, or equivalently, *adding* $-\alpha R_j$ *to* R_i.

The above results are based on the following theorem.

***Theorem* 7.3.**

1. If E is an elementary matrix, then E is nonsingular.
2. E^{-1} exists and is also an elementary matrix, of the same type as E.

Proof:

- If E is an elementary matrix of Type-I constructed from the identity I by interchanging the i^{th} and j^{th} rows, then E can be transformed back to I by interchanging the same rows again. Therefore, $EE = I$, with E being its own inverse.

- If E is the elementary matrix of Type-II constructed by multiplying the i^{th} row of I by a nonzero scalar α, then E can be transformed into the identity matrix by multiplying either its i^{th} row or i^{th} column by $1/\alpha$, if $\alpha \neq 0$. Thus,

$$E^{-1} = \begin{array}{cc} \begin{array}{ccccccc} & & & i^{\text{th}}\text{ column} & & & \end{array} & \\ \left(\begin{array}{ccccccc} 1 & & & & & & \\ & \ddots & & & & \mathbf{0} & \\ & & 1 & & & & \\ & & & 1/\alpha & & & \\ & & & & 1 & & \\ & \mathbf{0} & & & & \ddots & \\ & & & & & & 1 \end{array}\right) & \begin{array}{c} \\ \\ \\ i^{\text{th}}\text{ row} \\ \\ \\ \\ \end{array} \end{array}$$

- Finally, if E is the elementary matrix of Type-III constructed from I by adding to the j^{th} row the

m multiple of the i^{th} row,

$$E = \begin{array}{cl} \begin{array}{ccccccc} & & i^{\text{th}} & & j^{\text{th}} & & \end{array} & \\ \begin{pmatrix} 1 & & & & & & \\ \vdots & \ddots & & & \mathbf{0} & & \\ 0 & \cdots & 1 & & & & \\ \vdots & & & \ddots & & & \\ 0 & \cdots & m & \cdots & 1 & & \\ \vdots & & & & & \ddots & \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix} & \begin{array}{l} \\ \\ i^{\text{th}} \text{ row} \\ \\ j^{\text{th}} \text{ row} \\ \\ \\ \end{array} \end{array}$$

then E can be transformed back to I, either by subtracting from the j^{th} row the m multiple of the i^{th} row, or by subtracting from the i^{th} column the m multiple of the j^{th} column, i.e.,

$$E^{-1} = \begin{pmatrix} 1 & & & & & & \\ \vdots & \ddots & & & \mathbf{0} & & \\ 0 & \cdots & 1 & & & & \\ \vdots & & & \ddots & & & \\ 0 & \cdots & -m & \cdots & 1 & & \\ \vdots & & & & & \ddots & \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}$$

□

§ 7.2.3 Applications of elementary matrices

In Section §5.2 we learnt a fact that a sequence of elementary row operations are able to turn a square matrix into an echelon from or a reduced echelon form, as well as deriving the inverse of a matrix. Now, knowing that an elementary row operation can be realized by an elementary matrix, one can use a finite sequence of elementary matrices $E_1, E_2, \cdots, E_k$ to achieve the same objective.

***Definition* 7.7.** [Row equivalent] A matrix B is called row equivalent to another matrix A, denoted as $A \overset{\text{row}}{=\!=} B$, iff there exists a finite sequence of elementary matrices $E_1, E_2, \cdots, E_k$ such that

$$B = E_k E_{k-1} \cdots E_2 E_1 A. \tag{7.2.13}$$

Properties:

- If $B \overset{\text{row}}{=\!=} A$, then $A \overset{\text{row}}{=\!=} B$.

- If $A \overset{\text{row}}{=\!=} B$ and $B \overset{\text{row}}{=\!=} C$, then $A \overset{\text{row}}{=\!=} C$.

Application 1: Inverse of matrix

In Definition (7.2.13), when A is invertible and $B = I$ specially, we have

$$I = E_k E_{k-1} \ \cdots \ E_2 E_1 \ A, \tag{7.2.14}$$

which means the inverse can be given by

$$A^{-1} = E_k E_{k-1} \ \cdots \ E_2 E_1. \tag{7.2.15}$$

The next question is *how to construct this sequence* $E_k E_{k-1} \ \cdots \ E_2 E_1$. The answer is to *translate elementary row operations into elementary matrices*. Let us consider the following steps:

Step 1: Simultaneously left-multiply the matrices A and I with E_1.

Step 2: Simultaneously left-multiply $E_1 A$ and $E_1 I$ with E_2.

$\vdots$

Step k: Simultaneously left-multiply $E_{k-1} \cdots E_1 A$ and $E_{k-1} \cdots E_1 I$ with E_k.

This procedure is summarized in the table below:

	A	I
E_1:	$E_1 \ A$	$E_1 I = E_1$
E_2:	$E_2 \ E_1 \ A$	$E_2 \ E_1$
$\vdots$	$\vdots$	$\vdots$
E_k:	$I = E_k \ E_{k-1} \cdots E_1 \ A$	$E_k \ E_{k-1} \cdots E_1$

It is seen that the final result $E_k \ E_{k-1} \cdots E_1 \ A$ exactly gives the identity I, while the $E_k \ E_{k-1} \cdots E_1$ exactly the A^{-1}, according to (7.2.14) and (7.2.15). However, the above procedure is indeed mimicking the procedure of obtaining the inverse of a matrix A through an augmented matrix $(A|I)$ and a sequence of elementary row operations (ERO's) in Section §7.1.3:

$$(A|I) \quad \xRightarrow{\text{ERO }_k, \ \cdots, \ \text{ERO }_1} \quad \left(I|A^{-1}\right).$$

Therefore, in conclusion, we have achieved another method to obtain the inverse of a matrix A:

A. Recording all the elementary row operations, $\text{ERO}_1, \cdots, \text{ERO}_k$, in transforming A into the identity I;

B. Translating every elementary row operation, ERO_i, into an elementary matrix E_i.

C. Writing the E_i's in a sequence to get the inverse $A^{-1} = E_k E_{k-1} \cdots E_2 E_1$.

$$
\begin{aligned}
(A|I) &\xlongequal{\qquad} \left(\begin{array}{cccc|cccc}
a_{11} & a_{12} & \cdots & a_{1n} & 1 & & & \\
a_{21} & a_{22} & \cdots & a_{2n} & & 1 & & \\
\vdots & \vdots & & \vdots & & & \ddots & \\
a_{n1} & a_{n2} & \cdots & a_{nn} & & & & 1
\end{array}\right) \\
&\xlongequal[\text{Operation } l]{\text{Operation } k, \cdots,} \left(\begin{array}{cccc|cccc}
1 & & & & & & & \\
& 1 & & & & & & \\
& & \ddots & & E_k & E_{k-1} & \cdots & E_2 \quad E_1 \\
& & & 1 & & & &
\end{array}\right) \\
&\xlongequal{\qquad} (I \mid A^{-1})
\end{aligned}
$$

✎ ***Example* 7.8.** Compute the inverse of the matrix $A = \begin{pmatrix} 1 & 4 & 3 \\ -1 & -2 & 0 \\ 2 & 2 & 3 \end{pmatrix}$.

Solution: First, let us use the method of augmented matrix and elementary row operations to find A^{-1}:

$$
\left(\begin{array}{ccc|ccc}
1 & 4 & 3 & 1 & & \\
-1 & -2 & 0 & & 1 & \\
2 & 2 & 3 & & & 1
\end{array}\right)
$$

$$
\xrightarrow[R_3 \Longrightarrow R_3 - 2R_1 \ \ \textbf{(b)}]{R_2 \Longrightarrow R_1 + R_2 \ \ \textbf{(a)}}
\left(\begin{array}{ccc|ccc}
1 & 4 & 3 & 1 & & \\
0 & 2 & 3 & 1 & 1 & \\
0 & -6 & -3 & -2 & 0 & 1
\end{array}\right)
\xrightarrow{R_3 \Longrightarrow R_3 + 3R_2 \ \ \textbf{(c)}}
\left(\begin{array}{ccc|ccc}
1 & 4 & 3 & 1 & & \\
0 & 2 & 3 & 1 & 1 & \\
0 & 0 & 6 & 1 & 3 & 1
\end{array}\right)
$$

$$
\xrightarrow[R_3 \Longrightarrow \frac{1}{6} R_3 \ \ \textbf{(e)}]{R_2 \Longrightarrow \frac{1}{2} R_2 \ \ \textbf{(d)}}
\left(\begin{array}{ccc|ccc}
1 & 4 & 3 & 1 & & \\
0 & 1 & \frac{3}{2} & \frac{1}{2} & \frac{1}{2} & \\
0 & 0 & 1 & \frac{1}{6} & \frac{1}{2} & \frac{1}{6}
\end{array}\right)
\xrightarrow[R_1 \Longrightarrow R_1 - 3R_3 \ \ \textbf{(g)}]{R_2 \Longrightarrow R_2 - \frac{3}{2} R_3 \ \ \textbf{(f)}}
\left(\begin{array}{ccc|ccc}
1 & 4 & 0 & \frac{1}{2} & -\frac{3}{2} & -\frac{1}{2} \\
0 & 1 & 0 & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\
0 & 0 & 1 & \frac{1}{6} & \frac{1}{2} & \frac{1}{6}
\end{array}\right)
$$

$$
\xrightarrow{R_1 \Longrightarrow R_1 - 4R_2 \ \ \textbf{(h)}}
\left(\begin{array}{ccc|ccc}
1 & 0 & 0 & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\
0 & 1 & 0 & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\
0 & 0 & 1 & \frac{1}{6} & \frac{1}{2} & \frac{1}{6}
\end{array}\right)
= (I \mid A^{-1}),
$$

hence the inverse reads

$$A^{-1} = \begin{pmatrix} -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{6} \end{pmatrix}. \qquad (7.2.16)$$

In the above procedure, the ERO's **(a)–(h)** can be translated into eight elementary matrices:

$$\begin{aligned} E_{(a)} &= \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad E_{(b)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix}, \quad E_{(c)} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 3 & 1 \end{pmatrix}, \\ E_{(d)} &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad E_{(e)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{6} \end{pmatrix}, \quad E_{(f)} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & -\frac{3}{2} \\ 0 & 0 & 1 \end{pmatrix}, \\ E_{(g)} &= \begin{pmatrix} 1 & 0 & -3 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad E_{(h)} = \begin{pmatrix} 1 & -4 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \end{aligned} \qquad (7.2.17)$$

Then, putting $E_{(a)}$–$E_{(h)}$ together we obtain the inverse A^{-1}:

$$A^{-1} = E_{(h)}\, E_{(g)}\, E_{(f)}\, E_{(e)}\, E_{(d)}\, E_{(c)}\, E_{(b)}\, E_{(a)} = \begin{pmatrix} -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{6} \end{pmatrix}, \qquad (7.2.18)$$

which reproduces the above result (7.2.16).

It is seen that, of course, the method of (7.2.18) is more tedious than that of (7.2.16), hence usually we adopt the latter method in practice when searching for an inverse matrix.

□

The above application is based on the follow theorem.

***Theorem* 7.4.** [Equivalent conditions for non-singularity] Let A be an $n \times n$ matrix, the following statements are equivalent:

- A is nonsingular;
- A is row equivalent to I, i.e., A can be transformed into I through a series of elementary row operations by left-multiplying a finite sequence of elementary matrices;
- The homogeneous system of equations $A\mathbf{x} = \mathbf{0}$ has one and only one solution, i.e., the trivial solution $\mathbf{x} = \mathbf{0}$.

Proof: Ignored. ☐

Application 2: Solving linear system

Consider a linear system, $A\mathbf{x} = \mathbf{b}$, where A is an $n \times n$ matrix and $\mathbf{x}$ and $\mathbf{b}$ are column matrices. If the system has a unique solution, then it is given by

$$\mathbf{x} = A^{-1}\mathbf{b}. \tag{7.2.19}$$

In terms of the above result (7.2.15), we have

$$\mathbf{x} = E_k \, E_{k-1} \, \cdots \, E_2 \, E_1 \, \mathbf{b}. \tag{7.2.20}$$

This result can be regarded as mimicking our former procedure of Gauss-Jordan reduction of Section §5.2.2 to solve a linear system.

§ 7.3 Triangularization of matrix: LU decomposition

***Definition* 7.8** (Upper triangular). An $n \times n$ (square) matrix $A = (a_{ij})$ is said to be upper triangular iff

$$a_{ij} = 0, \qquad \forall\, i > j, \quad i, j = 1, \cdots, n, \qquad \text{which is of the form} \begin{pmatrix} * & * & * & * \\ & * & * & * \\ \mathbf{0} & & * & * \\ & & & * \end{pmatrix}. \tag{7.3.21}$$

***Definition* 7.9** (Lower triangular). An $n \times n$ (square) matrix $A = (a_{ij})$ is said to be lower triangular iff

$$a_{ij} = 0, \qquad \forall\, i < j, \quad i, j = 1, \cdots, n, \qquad \text{which is of the form} \begin{pmatrix} * & & & \\ * & * & & \mathbf{0} \\ * & * & * & \\ * & * & * & * \end{pmatrix}. \tag{7.3.22}$$

Lower-upper (LU) decomposition

Let us first re-examine the Type-III elementary matrices. There are a few facts:

1. An upper triangular Type-III matrix,

$$\begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & \cdots & \alpha & & \\ & & & \ddots & \vdots & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}, \tag{7.3.23}$$

is able to add a lower row to an upper row via left-multiplication, while a lower triangular one,

$$\begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & \vdots & \ddots & & & \\ & & \alpha & \cdots & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}, \tag{7.3.24}$$

is able to **add an upper row to a lower row** via left-multiplication.

2. The inverse of an upper triangular matrix is still an upper one, while that of a lower triangular matrix is still a lower one.

$$\begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & \alpha & & \\ & & & \ddots & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & -\alpha & & \\ & & & \ddots & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \tag{7.3.25}$$

$$\begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & \ddots & & & \\ & & \alpha & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & \ddots & & & \\ & & -\alpha & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \tag{7.3.26}$$

3. The product of upper triangular matrices is still an upper one, while that of lower triangular matrices is still a lower one:

$$\begin{pmatrix} 1 & * & * \\ & \ddots & * \\ & & 1 \end{pmatrix} \begin{pmatrix} 1 & * & * \\ & \ddots & * \\ & & 1 \end{pmatrix} = \begin{pmatrix} 1 & * & * \\ & \ddots & * \\ & & 1 \end{pmatrix}, \tag{7.3.27}$$

$$\begin{pmatrix} 1 & & \\ * & \ddots & \\ * & * & 1 \end{pmatrix} \begin{pmatrix} 1 & & \\ * & \ddots & \\ * & * & 1 \end{pmatrix} = \begin{pmatrix} 1 & & \\ * & \ddots & \\ * & * & 1 \end{pmatrix}. \tag{7.3.28}$$

These facts will be the basis for the LU decomposition below.

Second, let us re-examine the procedure of obtaining a strictly triangular form (which is an echelon form). For example, consider a mid-step

$$\begin{pmatrix} * & * & * & * \\ 0 & \bigstar & * & * \\ 0 & \bullet & * & * \\ 0 & \bullet & * & * \end{pmatrix} \Longrightarrow \begin{pmatrix} * & * & * & * \\ 0 & \bigstar & * & * \\ 0 & 0 & * & * \\ 0 & 0 & * & * \end{pmatrix}. \tag{7.3.29}$$

What we did is actually to use the $\bigstar$ entry to eliminate the two $\bullet$ entries by making use of a Type-III elementary row operation.

Third, on the other hand, a Type-III operation can be realized by an elementary matrix as pointed out by the above Fact 1, hence the step (7.3.29) can be done alternatively via left-multiplying an elementary matrix (7.3.24) to the LHS of (7.3.29).

Therefore, one can see that the Gauss-Jordan reduction procedure can be represented by a finite sequence of Type-III elementary matrices left-multiplied to the square matrix being studied:

$$E_k E_{k-1} \cdots E_2 E_1 \begin{pmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{pmatrix} = \begin{pmatrix} * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \\ 0 & 0 & 0 & * \end{pmatrix}. \tag{7.3.30}$$

It is seen that the RHS of (7.3.30) is an upper triangular matrix, and every E_i on the LHS is a lower triangular matrix.

Finally, since every elementary matrix E_i is invertible, (7.3.30) becomes

$$\begin{pmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{pmatrix} = E_1^{-1} E_2^{-1} \cdots E_{k-1}^{-1} E_k^{-1} \begin{pmatrix} * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \\ 0 & 0 & 0 & * \end{pmatrix}. \tag{7.3.31}$$

As asserted by the above Facts 2 and 3, the inverse E_i^{-1} is also a lower triangular Type-III elementary matrix, and the product of E_i^{-1}'s is lower triangular matrix. Hence (7.3.32) indeed looks like

$$\begin{pmatrix} * & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{pmatrix} = \underset{\text{Lower } \triangle}{\begin{pmatrix} * & & & \\ * & * & & \\ * & * & * & \\ * & * & * & * \end{pmatrix}} \underset{\text{Upper } \triangle}{\begin{pmatrix} * & * & * & * \\ & * & * & * \\ & & * & * \\ & & & * \end{pmatrix}}. \tag{7.3.32}$$

Now let us have the formal definition and theorem.

***Definition* 7.10.** A strictly lower triangular matrix with all 1's on the diagonal is called a unit lower triangular matrix, denoted as L.

***Theorem* 7.5.** If an $n \times n$ matrix A can be reduced to a strictly upper triangular form using *only* Type-III row operations, then it is possible to represent this reduction in terms of the so-called *lower-upper (LU) decomposition*:

$$A = LU, \tag{7.3.33}$$

where L and U are respectively a strictly lower- and upper-triangular matrix, as shown in (7.3.32).

Proof: Suppose a matrix A is able to be turned into a strictly upper triangular matrix through a finite sequence of k elementary row operations (ERO's),

$$A \xrightarrow[k \text{ elementary row operations}]{\text{a sequence of}} U = \begin{pmatrix} * & * & * & * \\ & * & * & * \\ \mathbf{0} & & \ddots & * \\ & & & * \end{pmatrix}.$$

Each ERO can be represented by an elementary matrix $E_i,\ i = 1, \cdots, k$, hence

$$E_k\, E_{k-1} \,\cdots\, E_1 A = U,$$

i.e.,

$$A = (E_k\, E_{k-1} \,\cdots\, E_1)^{-1} U = E_1^{-1}\, E_2^{-1} \,\cdots\, E_{k-1}^{-1}\, E_k^{-1}\, U.$$

Noticing the E_i's are lower triangular matrices, their inverses as well as the product of the inverses are also lower triangular matrices. Thus, A has an LU decomposition,

$$A = LU,$$

where L is a lower triangular matrix,

$$L = E_1^{-1} E_2^{-1} \cdots E_{k-1}^{-1} E_k^{-1} = \begin{pmatrix} * & & & \\ * & * & \mathbf{0} & \\ * & * & \ddots & \\ * & * & * & * \end{pmatrix}.$$

Chapter 8 Determinants

Determinants are not strange to us. In Chapter 3 we introduced the usage of determinants without formal definition, to give the algebraic expression of a cross product of vectors. In this chapter we will learn the formal definition of determinants as well as their properties.

§ 8.1 Introduction and definitions

Determinants arise from the attempts to derive the inverse of a nonsingular $n \times n$ (square) matrix.

A toy model: 2×2 *matrix case*

Starting from a 2×2 matrix

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

let us try to use Gaussian elimination to deduce its inverse:

$$\begin{aligned} & \left(\begin{array}{cc|cc} a & b & 1 & \\ c & d & & 1 \end{array}\right) \\ \xrightarrow{R_1 \Longrightarrow \frac{1}{a} R_1} & \left(\begin{array}{cc|cc} 1 & \frac{b}{a} & \frac{1}{a} & \\ c & d & & 1 \end{array}\right) \xrightarrow{R_2 \Longrightarrow R_2 - cR_1} \left(\begin{array}{cc|cc} 1 & \frac{b}{a} & \frac{1}{a} & 0 \\ 0 & \frac{ad-bc}{a} & -\frac{c}{a} & 1 \end{array}\right) \\ \xrightarrow{R_2 \Longrightarrow \frac{a}{\Delta} R_2} & \left(\begin{array}{cc|cc} 1 & \frac{b}{a} & \frac{1}{a} & 0 \\ 0 & 1 & -\frac{c}{\Delta} & \frac{a}{\Delta} \end{array}\right) \xrightarrow{R_1 \Longrightarrow R_1 - \frac{a}{b} R_2} \left(\begin{array}{cc|cc} 1 & 0 & \frac{d}{\Delta} & -\frac{b}{\Delta} \\ 0 & 1 & -\frac{c}{\Delta} & \frac{a}{\Delta} \end{array}\right), \end{aligned} \tag{8.1.1}$$

where a symbol Δ is introduced,

$$\Delta = ad - bc. \tag{8.1.2}$$

It is seen from (8.1.1) that, when $a \neq 0$ and $\Delta \neq 0$, the inverse of the matrix A exists and reads

$$A^{-1} = \frac{1}{\Delta} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \tag{8.1.3}$$

Apparently a, b, c, d cannot be zero at the same time; otherwise this zero matrix is nonsense. Hence one can always use the technique of "switching unknowns" and "swapping rows" to obtain a nonzero entry a, that is, the condition $a \neq 0$ can be guaranteed for sure.

Now the key is the requirement $\Delta \neq 0$. Obviously (8.1.3) asserts that the matrix A is singular if $\Delta = 0$, hence Δ becomes a key number. We define it to be the determinant of A:

$$\det A \equiv \mid A \mid = \Delta = ad - bc. \tag{8.1.4}$$

We see a fact

$$A \text{ nonsingular} \quad \overset{\text{iff}}{\Longleftrightarrow} \quad \det A \neq 0. \tag{8.1.5}$$

✎ ***Example* 8.1.** The determinant of a matrix $A = \begin{pmatrix} 3 & 4 \\ 2 & 1 \end{pmatrix}$ is computed as:

$$\det A = \begin{vmatrix} 3 & 4 \\ 2 & 1 \end{vmatrix} = 3 - 8 = -5.$$

<u>3×3 *matrix case*</u>

Now consider a 3×3 matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}. \tag{8.1.6}$$

For this 3×3 case, similarly, we can perform a Gaussian elimination. After the first step we have

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \longrightarrow \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ 0 & \frac{a_{11}a_{22}-a_{21}a_{12}}{a_{11}} & \frac{a_{11}a_{23}-a_{21}a_{13}}{a_{11}} \\ 0 & \frac{a_{11}a_{32}-a_{31}a_{12}}{a_{11}} & \frac{a_{11}a_{33}-a_{31}a_{13}}{a_{11}} \end{pmatrix}. \tag{8.1.7}$$

where all the entries of the first column vanish except the top entry, a_{11}. It is possible to guarantee $a_{11} \neq 0$ based on the same reason as in the 2×2 case. Then, subsequently, we need to keep going and turn the lower-right corner, i.e., the dashed-boxed 2×2 block, into a row echelon form $\begin{pmatrix} 1 & \\ & 1 \end{pmatrix}$

This means we need to repeat the above steps (8.1.1) to (8.1.4) for a new 2×2 matrix

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}, \tag{8.1.8}$$

with

$$\begin{aligned} A &= \frac{a_{11}a_{22} - a_{21}a_{12}}{a_{11}}, & B = \frac{a_{11}a_{23} - a_{21}a_{13}}{a_{11}}, \\ C &= \frac{a_{11}a_{32} - a_{31}a_{12}}{a_{11}}, & D = \frac{a_{11}a_{33} - a_{31}a_{13}}{a_{11}}. \end{aligned} \tag{8.1.9}$$

Obviously, the invertibility of the 3×3 matrix (8.1.6) is guaranteed by a condition $a_{11} \neq 0$ together with the invertibility of the 2×2 matrix (8.1.8), while the latter is decided by the 2×2 determinant

$$\det \begin{pmatrix} A & B \\ C & D \end{pmatrix} = AD - BC \neq 0. \tag{8.1.10}$$

The requirement (8.1.10) means

$$a_{11}a_{22}a_{33} - a_{11}a_{32}a_{23} - a_{12}a_{21}a_{33} + a_{12}a_{31}a_{23} + a_{13}a_{21}a_{32} - a_{13}a_{31}a_{22} \neq 0. \tag{8.1.11}$$

Hence, introducing the determinant of the 3×3 matrix A as

$$\begin{aligned} \det A = & \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \\ = \; & a_{11}a_{22}a_{33} - a_{11}a_{32}a_{23} - a_{12}a_{21}a_{33} + a_{12}a_{31}a_{23} + a_{13}a_{21}a_{32} - a_{13}a_{31}a_{22} \neq 0 \end{aligned} \tag{8.1.12}$$

We see the fact again,

$$A \text{ nonsingular} \quad \overset{\text{iff}}{\Longleftrightarrow} \quad \det A \neq 0. \tag{8.1.13}$$

Generic $n \times n$ case

Consider a generic $n \times n$ matrix

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}. \tag{8.1.14}$$

Let us re-examine its 2×2 and 3×3 cases with the aid of the above expressions (8.1.4) and (8.1.12).

- Case 2×2 :

$$\begin{aligned} \det A &= a_{11}a_{22} - a_{12}a_{21} \\ &= a_{11}M_{11} - a_{12}M_{12} \\ &= a_{11}A_{11} + a_{12}A_{12}, \end{aligned} \tag{8.1.15}$$

  where we introduce

$$A_{11} = M_{11} = +a_{22}, \qquad A_{12} = -M_{12} = -a_{21}. \tag{8.1.16}$$

- Case 3×3 :

$$\det A = a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{31}a_{22}). \tag{8.1.17}$$

Introducing

$$M_{11} = \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix}, \quad M_{12} = \begin{pmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{pmatrix}, \quad M_{13} = \begin{pmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix} \tag{8.1.18}$$

We have

$$\det A = a_{11} \ \det \ M_{11} - a_{12} \ \det \ M_{12} + a_{13} \ \det \ M_{13}. \tag{8.1.19}$$

Further, defining

$$A_{11} = (-1)^{1+1} \ \det M_{11}, \quad A_{12} = (-1)^{1+2} \ \det M_{12}, \quad A_{13} = (-1)^{1+3} \ \det M_{13}, \tag{8.1.20}$$

we have

$$\det A = a_{11}A_{11} + a_{12}A_{12} + a_{13}A_{13}. \tag{8.1.21}$$

The above cases demonstrate a fact that the 2×2 case is based on the 1×1 case, while the 3×3 case is based on the determinant of the 2×2 case, i.e., $\det M_{11}$, $\det M_{12}$ and $\det M_{13}$. Therefore, an inductive conclusion can be drawn for the generic case $n \times n$:

***Definition* 8.1** (Minor 余子式, cofactor 代数余子式).
Let $A = (a_{ij})$ be an $n \times n$ matrix, and M_{ij} the $(n-1) \times (n-1)$ matrix obtained from A by deleting the row and column containing the entry a_{ij}. That is

$$\begin{aligned} &\text{Minor of } a_{ij}: && \det M_{ij}. \\ &\text{Cofactor of } a_{ij}: && A_{ij} = (-1)^{i+j} \det M_{ij}. \end{aligned} \tag{8.1.22}$$

[Remark]: The sign $(-1)^{i+j}$ can be illustrated by the following matrix, in which the $+$ and $-$ are in an alternating order:

$$\begin{pmatrix} + & - & + & - & \cdots \\ - & + & - & + & \cdots \\ + & - & + & - & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{8.1.23}$$

Then we achieve the formal definition of the determinant of a matrix A.

***Theorem* 8.1.** [Cofactor expansion of determinant （行列式的）代数余子式展开]
Let A be an $n \times n$ matrix, $n \geq 2$. Its determinant $\det A$ can be expressed as a cofactor expansion in terms of any row (say, Row i) or any column (say, Column j) of A:

$$\begin{aligned} \det A &= a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in} && \Longleftarrow \text{Row } i, \qquad \forall i, j = 1, \cdots, n \\ &= a_{1j}A_{1j} + a_{2j}A_{2j} + \cdots + a_{nj}A_{nj} && \Longleftarrow \text{Column } j. \end{aligned} \tag{8.1.24}$$

✎ ***Example* 8.2.** Computation of a determinant:

$$\det\begin{pmatrix} 0 & 2 & 3 & 0 \\ 0 & 4 & 5 & 0 \\ 0 & 1 & 0 & 3 \\ 2 & 0 & 1 & 3 \end{pmatrix} = -2\begin{vmatrix} 2 & 3 & 0 \\ 4 & 5 & 0 \\ 1 & 0 & 3 \end{vmatrix}$$

$$= \quad -2 \times 3\begin{vmatrix} 2 & 3 \\ 4 & 5 \end{vmatrix} = -6 \times (10 - 12) = 12.$$

Alternatively,

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

Now, equipped with the tool of determinants, let us go back to our original mission of finding the inverse of a matrix.

***Theorem* 8.2.** An $n \times n$ matrix A is singular iff

$$\det A = 0. \tag{8.1.25}$$

Proof: Ignored. Readers are referred to other textbooks for details.

§ 8.2 Properties of determinants

In this section, some theorems and properties of determinants will be listed out. Most proofs will be ignored; interested readers are referred to textbooks for details. Our emphasis is focused on how to memorize and use them in practice.

§ 8.2.1 Determinant of matrix transpose

***Theorem* 8.3.** If A is an $n \times n$ matrix, then its transpose has the same determinant,

$$\det A^T = \det A. \tag{8.2.1}$$

Proof: Ignored.

Hint: A simple way is to appeal to a formula

$$\det A = \frac{1}{n!}\epsilon_{i_1\cdots i_n}\epsilon_{j_1\cdots j_n}A_{i_1j_1}\cdots A_{i_nj_n}, \tag{8.2.2}$$

where

$$\epsilon_{i_1\cdots i_n} = \begin{cases} 1, & \text{when } i_1\cdots i_n \text{ are an } \textbf{even} \text{ permutation（置换）of } 1,\cdots,n; \\ -1, & \text{when } i_1\cdots i_n \text{ are an } \textbf{odd} \text{ permutation of } 1,\cdots,n. \end{cases} \tag{8.2.3}$$

☐

***Theorem* 8.4.** Let A be a triangular matrix,

$$A = \begin{pmatrix} a_{11} & & * \\ & a_{22} & \\ \mathbf{0} & & \ddots & \\ & & & a_{nn} \end{pmatrix} \quad \text{or} \quad A = \begin{pmatrix} a_{11} & & \mathbf{0} \\ & a_{22} & \\ & & \ddots & \\ * & & & a_{nn} \end{pmatrix}.$$

Its determinant equals to the product of all its diagonal entries,

$$\det A = a_{11}a_{22}\cdots a_{nn}. \tag{8.2.4}$$

Proof: This fact is easily seen in terms of the cofactor expansion, where only the diagonal entries, $a_{11}, a_{22}, \cdots, a_{nn}$, will have contributions to each step of the determinant computation.

☐

***Corollary* 8.4.1.** A diagonal matrix $A = \text{diag}\{a_{11}, a_{22}, \cdots, a_{nn}\}$ also has

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

Proof: Ignored. A diagonal matrix is a special case of upper or lower triangular matrix.

***Theorem* 8.5.** Let A be an $n \times n$ matrix.

1. If A has a row or column which consists of all zeros, then $\det A = 0$.
2. If A has two identical rows or columns, then $\det A = 0$.

Proof:

1. This fact is easily seen in terms of cofactor expansion.
2. This can be proved by means of mathematical induction:
   - When $n = 2$, consider a 2×2 matrix with, say, identical rows, $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{11} & a_{12} \end{pmatrix}$. Then its determinant reads

     $$\det A = a_{11}a_{12} - a_{11}a_{12} = 0.$$

   - Suppose the statement is true for $n = k$, i.e., $\det A = 0$ if A is a $k \times k$ matrix with identical rows or columns.
   - Examine the case $n = k + 1$. Consider a $(k+1) \times (k+1)$ matrix

     $$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} & a_{1,k+1} \\ a_{21} & a_{22} & \cdots & a_{2k} & a_{2,k+1} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ a_{k+1,1} & a_{k+1,2} & \cdots & a_{k+1,k} & a_{k+1,k+1} \end{pmatrix},$$

     whose top two rows are identical, i.e.,

     $$a_{11} = a_{21}, \qquad a_{12} = a_{22}, \qquad \cdots, \qquad a_{1,k+1} = a_{2,k+1}.$$

Then the cofactor expansion for A along the bottom row will prove the fact:

$$\begin{aligned}\det A &= (\text{ sign })a_{k+1,1}\det\begin{pmatrix} a_{12} & \cdots & a_{1,k+1} \\ a_{22} & \cdots & a_{2,k+1} \\ \vdots & \cdots & \vdots \\ a_{k2} & \cdots & a_{k,k+1}\end{pmatrix} + \cdots \\ &\quad + (\text{ sign })a_{k+1,k+1}\det\begin{pmatrix} a_{11} & \cdots & a_{1k} \\ a_{21} & \cdots & a_{2k} \\ \vdots & \cdots & \vdots \\ a_{k1} & \cdots & a_{kk}\end{pmatrix} \\ &= 0,\end{aligned}$$

where the (sign) means the corresponding sign appearing there, and each $k \times k$ determinant vanishes thanks to the inductive hypothesis.

□

[Remark]: Another proof is to be given in the following property regarding the Type-I operation.

***Theorem* 8.6.** Let A be an $n \times n$ matrix, and A_{ij} denote the cofactor of $a_{ij},\ \ i, j = 1, \cdots, n$. Then

$$a_{i1}A_{i'1} + a_{i2}A_{i'2} + \cdots + a_{in}A_{i'n} = \begin{cases} \det A, & \text{if } i = i', \\ 0, & \text{if } i \neq i'. \end{cases} \tag{8.2.5}$$

[Remark]:

- This theorem means the inner product between a row vector $\vec{R_i} = \begin{pmatrix} a_{i1} & \cdots & a_{in} \end{pmatrix}$ and its own cofactor row $\vec{A_i} = \begin{pmatrix} A_{i,1} & \cdots & A_{i,n} \end{pmatrix}$ will give the determinant; otherwise, the inner product between $\vec{R_i}$ and $\vec{A_{i'}}$ — the cofactor row of another row $\vec{R_{i'}},\ i \neq i'$ — will vanish. Namely, the row vector set, $\{\vec{R_i}\}$, and the cofactor row vector set, $\{\vec{A_{i'}}\}$, form an orthogonal relation.
- A similar conclusion can be drawn for cofactor expansions along columns, i.e.,

$$a_{1j}A_{1j'} + a_{2j}A_{2j'} + \cdots + a_{nj}A_{nj'} = \begin{cases} \det A, & \text{if } j = j', \\ 0, & \text{if } j \neq j'. \end{cases} \tag{8.2.6}$$

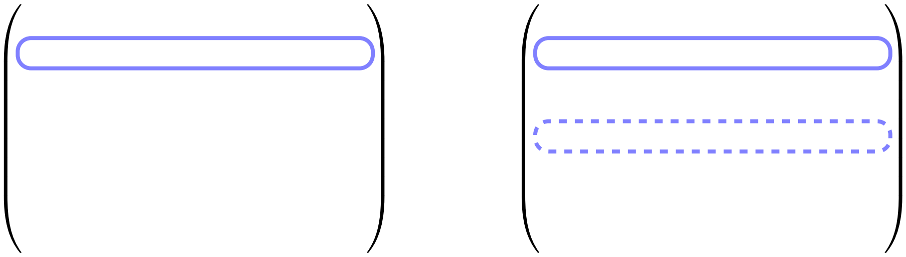
Proof:

1. Case $i = i'$:
   This is nothing but the definition of cofactor expansion of a determinant, (8.1):
   $$a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in} = \det A.$$
   For the further proof below, let us notice such a fact:
   If $\det A \neq 0$, there must be no identical rows in A, due to Theorem 8.5. Therefore, for an entry a_{ij}, its cofactor A_{ij} should contain no rows (or columns) coming from the same row (or column) containing a_{ij}, i.e.,
   $$A_{ij} = (-1)^{i+j} \begin{vmatrix} a_{11} & \cdots & a_{1,j-1} & a_{1,j+1} & \cdots & a_{1n} \\ \vdots & & & & & \vdots \\ a_{i-1,1} & \cdots & a_{i-1,j-1} & a_{i-1,j+1} & \cdots & a_{i-1,n} \\ a_{i+1,1} & \cdots & a_{i+1,j-1} & a_{i+1,j+1} & \cdots & a_{i+1,n} \\ \vdots & & & & & \vdots \\ a_{n1} & \cdots & a_{n,j-1} & a_{n,j+1} & \cdots & a_{nn} \end{vmatrix}.$$

2. Case $i \neq j$:
   In this case,
   $$a_{i1}A_{i'1} + a_{i2}A_{i'2} + \cdots + a_{in}A_{i'n}, \qquad i \neq i', \tag{8.2.7}$$
   since $i \neq i'$, the cofactor $A_{i'j}$ must contains a row coming from the same row containing a_{ij}, i.e.,
   $$A_{i'j} = (-1)^{i'+j} \begin{vmatrix} a_{11} & \cdots & a_{1,j-1} & a_{1,j+1} & \cdots & a_{1n} \\ \vdots & & \vdots & & & \vdots \\ a_{i1} & \cdots & a_{i,j-1} & a_{i,j+1} & \cdots & a_{in} \\ \vdots & & \vdots & & & \vdots \\ a_{n1} & \cdots & a_{n,j-1} & a_{n,j+1} & \cdots & a_{nn} \end{vmatrix}.$$
   Therefore, (8.2.7) indeed gives the determinant of a matrix containing two identical rows, which is definitely a zero:
   $$a_{i1}A_{i'1} + a_{i2}A_{i'2} + \cdots + a_{in}A_{i'n} = \det \begin{pmatrix} & \cdots\cdots\cdots & & \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ & \cdots\cdots\cdots & & \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ & \cdots\cdots\cdots & & \end{pmatrix} = 0. \tag{8.2.8}$$

□

[Remark]: The above proof can be illustrated as follows.



[Left] Case $i = i'$:

For a row represented by the solid box, its cofactors do not contain that row itself. Hence

$$a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in} = \det A.$$

[Right] Case $i \neq i'$:

This is exactly equivalent to the case that a matrix contains two identical rows. Namely, when doing the cofactor expansion along a row represented by the solid box, the cofactors cannot avoid taking values from the dash-boxed row. Since the two boxed rows are identical, the determinant must vanish,

$$a_{i1}A_{i'1} + a_{i2}A_{i'2} + \cdots + a_{in}A_{i'n} = 0, \qquad i \neq i'.$$

§ **8.2.2 Three types of row operations**

Type I*: Row exchange $R_i \longleftrightarrow R_j$*

***Theorem* 8.7.** Let A be an $n \times n$ matrix, and E_{ij}^{I} an elementary matrix to realize a Type I operation of row exchange, $R_i \longleftrightarrow R_j$. Then,

$$\det\left(E_{ij}^{\mathrm{I}}\, A\right) = \det E_{ij}^{\mathrm{I}}\, \det A = -\det A, \qquad 1 \le i, j \le n. \tag{8.2.9}$$

[Remark]: The meaning of this expression is twofold.

I. A determinant changes its sign under a row exchange, i.e.,

$$\det \tilde{A} = -\det A, \tag{8.2.10}$$

where $\tilde{A}$ is obtained by exchanging two rows of A.

II. The determinant of the matrix product equals to the product of the separate determinants of the matrices,

$$\det E_{ij}^{\mathrm{I}} = -1, \qquad \text{thus} \qquad \det\left(E_{ij}^{\mathrm{I}}\, A\right) = \det E_{ij}^{\mathrm{I}}\, \det A. \tag{8.2.11}$$

Proof: Let us prove (8.2.9) by verifying the two parts (8.2.10) and (8.2.11) in turn.

Step I: For an $n \times n$ matrix A, to prove $\det \tilde{A} = -\det A$, we appeal to mathematical induction.

(1) When $n = 2$,

$$\det A = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}, \qquad \det \tilde{A} = \begin{vmatrix} a_{21} & a_{22} \\ a_{11} & a_{12} \end{vmatrix} = a_{21}a_{12} - a_{11}a_{22},$$

hence

$$\det \tilde{A} = -\det A.$$

(2) Inductive hypothesis: Assume the statement is true for $n = k$, where $k \ge 2$.

(3) Examination of the case $n = k + 1$: Suppose $\tilde{A}$ is obtained from A through $R_i \longleftrightarrow R_j$. Then we can always row-expand $\det A$ and $\det \tilde{A}$ along a row R_l other than R_i and R_j, i.e.,

$l \neq i, j$:

$$\begin{aligned}
\det A &=\!=\!=\!=\!= \begin{vmatrix} \vdots & & \\ a_{i1} & \cdots & a_{i,k+1} \\ \vdots & & \\ a_{j1} & \cdots & a_{j,k+1} \\ \vdots & & \\ a_{l1} & \cdots & a_{l,k+1} \\ \vdots & & \end{vmatrix} \\
&=\!=\!=\!=\!= (-1)^{l+1} \begin{vmatrix} \vdots & & \\ a_{i2} & \cdots & a_{i,k+1} \\ \vdots & & \\ a_{j2} & \cdots & a_{j,k+1} \\ \vdots & & \end{vmatrix} + \cdots + (-1)^{l+k+1} \begin{vmatrix} \vdots & & \\ a_{i1} & \cdots & a_{i,k} \\ \vdots & & \\ a_{j1} & \cdots & a_{j,k} \\ \vdots & & \end{vmatrix} \\
&\overset{\text{Inductive}}{\underset{\text{hypothesis}}{=\!=\!=\!=\!=}} -(-1)^{l+1} \begin{vmatrix} \vdots & & \\ a_{j2} & \cdots & a_{j,k+1} \\ \vdots & & \\ a_{i2} & \cdots & a_{i,k+1} \\ \vdots & & \end{vmatrix} - \cdots - (-1)^{l+k+1} \begin{vmatrix} \vdots & & \\ a_{j1} & \cdots & a_{j,k} \\ \vdots & & \\ a_{i1} & \cdots & a_{i,k} \\ \vdots & & \end{vmatrix} \\
&=\!=\!=\!=\!= - \begin{vmatrix} \vdots & & \\ a_{j1} & \cdots & a_{j,k+1} \\ \vdots & & \\ a_{i1} & \cdots & a_{i,k+1} \\ \vdots & & \\ a_{l1} & \cdots & a_{l,k+1} \\ \vdots & & \end{vmatrix} = -\det \tilde{A}.
\end{aligned}$$

Hence the statement holds true for all integers $n \geq 2$.

Step II: The determinant of E^{I}_{ij} is computed as follows:

$$\det E^{\mathrm{I}}_{ij} = \begin{vmatrix} 1 &&&&&&&&&& \\ & \ddots &&&&&&&&& \\ && 1 &&&&&&&& \\ &&& 0 && \cdots && 1 &&& \\ &&&& 1 &&&&&& \\ &&& \vdots && \ddots && \vdots &&& \\ &&&&&& 1 &&&& \\ &&& 1 && \cdots && 0 &&& \\ &&&&&&&& 1 && \\ &&&&&&&&& \ddots & \\ &&&&&&&&&& 1 \end{vmatrix} = \begin{vmatrix} 0 && \cdots && 1 \\ & 1 &&& \\ \vdots && \ddots && \vdots \\ &&& 1 & \\ 1 && \cdots && 0 \end{vmatrix}$$

$$= \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = -1.$$

Therefore, the statement (8.2.9) holds true. ☐

***Type II**: Row rescaling:* $R_i \Longrightarrow \alpha R_i, \quad \alpha \neq 0$

***Theorem* 8.8.** Let A be an $n \times n$ matrix, and $E^{\mathrm{II}}_{i,\alpha}$ an elementary matrix to realize a Type II operation of row rescaling, $R_i \Longrightarrow \alpha R_i, \alpha \neq 0$. Then,

$$\det\left(E^{\mathrm{II}}_{i,\alpha} A\right) = \det E^{\mathrm{II}}_{i,\alpha} \det A = \alpha \det A, \qquad \alpha \neq 0. \tag{8.2.12}$$

[Remark]: The meaning of this expression is twofold.

I. Multiplying one row of a matrix with a scalar, equals to multiplying its determinant with the scalar, i.e.,

$$\det \tilde{A} = \alpha \det A, \qquad \alpha \neq 0, \tag{8.2.13}$$

where $\tilde{A}$ is obtained by rescaling the ith row of A with α.

II. The determinant of the matrix product equals to the product of the separate determinants of the matrices,

$$\det E^{\mathrm{II}}_{i,\alpha} = \alpha, \qquad \text{thus} \qquad \det\left(E^{\mathrm{II}}_{i,\alpha} A\right) = \det E^{\mathrm{II}}_{i,\alpha} \det A. \tag{8.2.14}$$

Proof: Let us prove (8.2.12) by verifying the two parts (8.2.13) and (8.2.14) in turn.

Step I: Let us examine the row-expansion of $\det \tilde{A}$ along the particular rescaled row:

$$\begin{aligned}\det \tilde{A} &= \begin{vmatrix} \vdots & & \\ \alpha a_{i1} & \cdots & \alpha a_{i1} \\ \vdots & & \end{vmatrix} = (-1)^{i+1}\alpha\,(\text{cofactor of } \ a_{i1}) + \cdots + (-1)^{i+n}\alpha\,(\text{cofactor of } \ a_{in}) \\ &= \alpha\left[(-1)^{i+1}\,(\text{cofactor of } \ a_{i1}) + \cdots + (-1)^{i+n}\,(\text{cofactor of } \ a_{in})\right] = \alpha \det A.\end{aligned}$$

Step II: Obviously, due to the computation of the determinant of a diagonal matrix, we have

$$\det E^{\text{II}}_{i,\alpha} = \begin{vmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & \alpha & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{vmatrix} = \alpha. \tag{8.2.15}$$

Therefore, the statement (8.2.12) holds true. □

***Corollary* 8.8.1.** Let A be an $n \times n$ matrix, and α a nonzero scalar. The determinant of the scalar multiplication αA reads

$$\det(\alpha A) = \alpha^n \det A. \tag{8.2.16}$$

Proof: Obvious. Notice, $\det(\alpha A) \neq \alpha \det A, \quad \text{when} \quad n \neq 1.$

Type III: *Adding a rescaled row to another row:* $R_i \Longrightarrow R_i + \alpha R_j$

***Theorem* 8.9.** Let A be an $n \times n$ matrix, and $E^{\text{III}}_{i,\alpha j}$ an elementary matrix to realize a Type III operation, $R_i \Longrightarrow R_i + \alpha R_j, \alpha \neq 0$. Then,

$$\det\left(E^{\text{III}}_{i,\alpha j} A\right) = \det E^{\text{III}}_{i,\alpha j} \det A = \det A. \tag{8.2.17}$$

[Remark]: The meaning of this expression is twofold.

I. Multiplying one row with a scalar and then adding it to another row, does not change the determinant of the matrix,

$$\det \tilde{A} = \det A, \tag{8.2.18}$$

where $\tilde{A}$ is obtained by performing a Type-III operation to A.

II. The determinant of the matrix product equals to the product of the separate determinants of the matrices,

$$\det E^{\text{III}}_{i,\alpha j} = 1, \qquad \text{thus} \qquad \det\left(E^{\text{III}}_{i,\alpha j} A\right) = \det E^{\text{III}}_{i,\alpha j} \det A. \tag{8.2.19}$$

Proof: Let us prove (8.2.17) by verifying the two parts (8.2.18) and (8.2.19) in turn.

Step 1: Let us prove $\det \tilde{A} = \det A$ by row-expand $\tilde{A}$ along the particular row:

$$\begin{aligned}
\det \tilde{A} \quad &= \quad \begin{vmatrix} \vdots & & \\ a_{i1} + \alpha a_{j1} & \cdots & a_{in} + \alpha a_{jn} \\ \vdots & & \\ a_{j1} & \cdots & a_{jn} \\ \vdots & & \end{vmatrix} \\
&\overset{\text{row expansion}}{=\!=\!=\!=} \quad \sum_{l=1}^{n} (-1)^{i+l} \left(a_{il} + \alpha a_{jl}\right) (\text{cofactor of site} \quad (i,l)) \\
&= \quad \sum_{l=1}^{n} (-1)^{i+l} a_{il} \,(\text{cofactor of site} \quad (i,l)) + \sum_{l=1}^{n} (-1)^{i+l} \alpha a_{jl} \,(\text{cofactor of site} \quad (i,l)) \\
&= \quad \sum_{l=1}^{n} (-1)^{i+l} a_{il} \,(\text{cofactor of site} \quad (i,l)) + \sum_{l=1}^{n} (-1)^{i+l} \alpha \begin{vmatrix} \vdots & & \\ a_{j1} & \cdots & a_{jn} \\ \vdots & & \\ a_{j1} & \cdots & a_{jn} \\ \vdots & & \end{vmatrix} \\
&\overset{\text{identical rows}}{=\!=\!=\!=} \quad \det A + 0 = \det A.
\end{aligned}$$

Step 2: Obviously, thanks to the computation rule of a triangular matrix,

$$\det E^{\mathrm{III}}_{i,\alpha j} = \begin{vmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & \vdots & \ddots & & \\ & & \alpha & \cdots & 1 & \\ & & & & \ddots & \\ & & & & & 1 \end{vmatrix} \quad \text{or} \quad \begin{vmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & \cdots & \alpha & \\ & & & \ddots & \vdots & \\ & & & & 1 & \\ & & & & & \ddots & \\ & & & & & & 1 \end{vmatrix} = 1. \qquad (8.2.20)$$

Therefore, the statement (8.2.17) holds true.

□

Summary of three types of elementary matrices

Now let us summarize the properties of the determinant of an elementary matrix.

***Theorem* 8.10.** Let E be an $n \times n$ elementary matrix, and A an $n \times n$ matrix. Then,

1.
$$\det E = \begin{cases} -1, & \text{if } E \text{ is Type-I,} \\ \alpha \neq 0, & \text{if } E \text{ is Type-II,} \\ 1, & \text{if } E \text{ is Type-III.} \end{cases} \qquad (8.2.21)$$

2.
$$\det(EA) = \det E \det A. \qquad (8.2.22)$$

Proof: Ignored. See above (8.2.9)–(8.2.17). □

§ 8.2.3 Determinant of product of matrices

Based on the above theorems regarding the determinants of elementary matrices, we have the following important conclusions.

***Theorem* 8.11.** An $n \times n$ matrix A is singular iff its determinant vanishes, i.e.,

$$A \text{ singular} \quad \Longleftrightarrow \quad \det A = 0. \tag{8.2.23}$$

Proof:

Let U be the reduced echelon form, which is an $n \times n$ matrix, induced by A through a sequence of k elementary row operations

$$A \xrightarrow{ERO-1,\ ERO-2,\ ,\cdots,\ ERO-k} U.$$

Since every $ERO - i$ can be realized by an elementary matrix E_i, we indeed have

$$E_k E_{k-1} \cdots E_1 A = U, \qquad \text{namely,} \qquad A = E_1^{-1} E_2^{-1} \cdots E_k^{-1} U,$$

where $E_1^{-1}, \cdots, E_k^{-1}$ are also elementary matrices. Introducing a new symbol, $\tilde{E}_i = E_i^{-1},\ i = 1, \cdots, k,$

$$A = \tilde{E}_1 \tilde{E}_2 \cdots \tilde{E}_k U.$$

Taking the determinants of both sides of the above equation, we have

$$\det A = \det\left(\tilde{E}_1 \tilde{E}_2 \cdots \tilde{E}_k U\right) = \det \tilde{E}_1 \det \tilde{E}_2 \cdots \det \tilde{E}_k \det U.$$

Since all the evaluations $\det \tilde{E}_i$'s are nonzero as shown in (8.2.21), we have

$$\det A = 0 \Longleftrightarrow \det U = 0. \tag{8.2.24}$$

Thus,

- if A is singular, its induced reduced echelon form U must contains at least one all-zero row, like

$$U = \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & 1 \\ 0 & 0 & \cdots & 0 \end{pmatrix}.$$

  Such a U must have $\det U = 0$. Hence $\det A = 0$.

- If A is non-singular, its induced reduced echelon form U must be an identity matrix

  $$U = \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix},$$

  which has $\det U = 1 \neq 0$. Hence $\det A \neq 0$.

Therefore, A singular $\iff \det A = 0$, the statement holds true. □

***Theorem* 8.12.** If A and B are $n \times n$ matrices, then

$$\det(AB) = \det A \det B. \tag{8.2.25}$$

Proof: Let us examine the different cases of the matrix A.

- Case 1 — A is singular:

  In this case we have $\det A = 0$, due to (8.2.23). Hence the RHS of (8.2.25) vanishes.

  For the LHS, let A_0 be the reduced echelon form row-equivalent to A, i.e.,

  $$A = E_k E_{k-1} \cdots E_1 A_0,$$

  where $E_1, \cdots, E_k$ are the needed elementary matrices. Then A_0 must contain all-zero rows in the bottom, say,

  $$A_0 = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \\ 0 & & \cdots & & 0 \end{pmatrix}.$$

  Therefore,

  $$\det(AB) = \det\left(E_k \cdots E_1 A_0 B\right) = \det\left(E_k \cdots E_1\right) \det\left(A_0 B\right).$$

  Given that A_0 contains all-zero rows, $A_0 B$ must contain all-zero rows, hence $A_0 B$ is singular, i.e.,

  $$\det\left(A_0 B\right) = 0.$$

  Therefore, the LHS vanishes too, and (8.2.25) holds true.

- Case 2 — A is non-singular:

  In this case the reduced echelon form A_0 is an identity matrix, thus $A = E_k \cdots E_1 I$. Therefore,

  $$\det(AB) = \det\left(E_k \cdots E_1 I B\right) = \det\left(E_k \cdots E_1\right) \det\left(B\right) = \det A \det B.$$

  (8.2.25) holds true.

□

More efficient way to compute a determinant based on row and column operations

Thanks to the property (8.2.1), $\det A^T = \det A$, we can similarly have elementary column operations.

In (7.2.7)–(7.2.9) of the last chapter, it is shown that a column operation can be done by right-multiplying an elementary matrix E to a matrix A. Moreover, it is easy to see that the transpose of an elementary matrix is also an elementary matrix, hence we have the following useful conclusion.

***Theorem* 8.13.** (*Determinants under Type-I elementary column operation*)
Let A and $\tilde{A}$ be two $n \times n$ matrices, where $\tilde{A}$ is obtained from A via an elementary column operation. Let E_{ij}^{I} be a Type-I elementary matrix of size $n \times n$, which realizes the function of swapping two columns C_i and C_j when right-multiplied to A:

$$AE_{ij}^{\mathrm{I}} \overset{c_i \leftrightarrow c_j}{=\!=\!=} \tilde{A}.$$

Then

$$\det \tilde{A} = -\det A, \qquad \text{i.e.,} \qquad \det A \xrightarrow{c_i \leftrightarrow c_j} -\det A. \tag{8.2.26}$$

Proof: Let $E_{ij}^{\mathrm{I,T}}$ be the transpose of E_{ij}^{I}, which has the function of swapping two rows R_i and R_j when left-multiplied to a matrix. Then, in the light of (8.2.9), we have

$$\det \tilde{A} = \det\left(AE_{ij}^{\mathrm{I}}\right) \overset{\text{Transpose}}{=\!=\!=\!=} \det\left(E_{ij}^{\mathrm{I,T}} A^T\right) = -\det A^T = -\det A.$$

□

***Theorem* 8.14.** (*Determinants under Type-II elementary column operation*)
Let A and $\tilde{A}$ be two $n \times n$ matrices, where $\tilde{A}$ is obtained from A via an elementary column operation. Let $E_{i,\alpha}^{\text{II}}$ be a Type-II elementary matrix of size $n \times n$, which realizes the function of rescaling a column C_i by α when right-multiplied to A:

$$AE_{i,\alpha}^{\text{II}} \overset{c_i \Rightarrow \alpha c_i}{=\!=\!=\!=} \tilde{A}.$$

Then

$$\det \tilde{A} = \alpha \det A, \qquad \text{i.e.,} \qquad \det A \xrightarrow{c_i \Rightarrow \alpha c_i} \alpha \det A. \tag{8.2.27}$$

Proof: Let $E_{i,\alpha}^{\text{II,T}}$ be the transpose of $E_{i,\alpha}^{\text{II}}$, which has the function of rescaling a row R_i by α when left-multiplied to a matrix. Then, in the light of (8.2.12), we have

$$\det \tilde{A} = \det\left(AE_{i,\alpha}^{\text{II}}\right) \overset{\text{Transpose}}{=\!=\!=\!=} \det\left(E_{i,\alpha}^{\text{II,T}}\, A^T\right) = \alpha \det A^T = \alpha \det A.$$

□

***Theorem* 8.15.** (*Determinants under Type-III elementary column operation*)
Let A and $\tilde{A}$ be two $n \times n$ matrices, where $\tilde{A}$ is obtained from A via an elementary column operation. Let E_{ij}^{III} be a Type-III elementary matrix of size $n \times n$, which realizes the function of adding to a column C_i another column C_j rescaled by α, when right-multiplied to A:

$$AE_{i,\alpha j}^{\text{III}} \overset{c_i \Rightarrow c_i + \alpha c_j}{=\!=\!=\!=} \tilde{A}.$$

Then

$$\det \tilde{A} = \det A, \qquad \text{i.e.,} \qquad \det A \xrightarrow{c_i \Rightarrow c_i + \alpha c_j} \det A. \tag{8.2.28}$$

Proof: Let $E_{i,\alpha j}^{\text{III,T}}$ be the transpose of $E_{i,\alpha j}^{\text{III}}$, which has the function of adding to a row R_i another row R_j rescaled by α, when left-multiplied to a matrix. Then, in the light of (8.2.17), we have

$$\det \tilde{A} = \det\left(AE_{i,\alpha j}^{\text{III}}\right) \overset{\text{Transpose}}{=\!=\!=\!=} \det\left(E_{i,\alpha j}^{\text{III,T}}\, A^T\right) = \det A^T = \det A.$$

□

Equipped with row and column operations, we are able to simplify a determinant as much as possible, for example, to the extent of triangular matrix, and then easily compute it.

✎ ***Example* 8.3.**

$$\begin{vmatrix} 2 & 1 & 3 \\ 4 & 2 & 1 \\ 6 & -3 & 4 \end{vmatrix} = \begin{vmatrix} 2 & 1 & 3 \\ 0 & 0 & -5 \\ 0 & -6 & -5 \end{vmatrix} = - \begin{vmatrix} 2 & 1 & 3 \\ 0 & -6 & -5 \\ 0 & 0 & -5 \end{vmatrix} = \begin{vmatrix} 2 & 1 & -3 \\ 0 & -6 & 5 \\ 0 & 0 & 5 \end{vmatrix} = -60.$$

The following Table 8.1 demonstrates a comparison between the two methods of computing a determinant — *cofactor expansion* versus *row and column operations* (see Page 96 of *S. Leon, 2010*). It is seen that, as the size of a determinant increases, the advantage of the approach of row and column operations becomes more and more explicit.

	Cofactors		Row/column operations	
n	Additions	Multiplications	Additions	Multiplications and Divisions
2	1	2	1	3
3	5	9	5	10
4	23	40	14	23
5	119	205	30	44
10	3,628,799	6,235,300	285	339

Table 8.1: Comparison between the two methods of computing a determinant: *cofactor expansion* vs. *row and column operations*.

§ 8.3 Applications of determinants

§ 8.3.1 Cross product

In (3.2.2) of Chapter 3, the algebraic definition of cross product（叉乘）is presented in terms of determinants, before we learn the definition of determinants in this chapter. Here let us recall the algebraic definition of determinants as well as its row-expansion:

$$\mathbf{a}\times\mathbf{b}=\begin{vmatrix}\mathbf{i}&\mathbf{j}&\mathbf{k}\\a_1&a_2&a_3\\b_1&b_2&b_3\end{vmatrix}=\mathbf{i}\begin{vmatrix}a_2&a_3\\b_2&b_3\end{vmatrix}-\mathbf{j}\begin{vmatrix}a_1&a_3\\b_1&b_3\end{vmatrix}+\mathbf{k}\begin{vmatrix}a_1&a_2\\b_1&b_2\end{vmatrix}. \tag{8.3.29}$$

§ 8.3.2 Adjoint of matrix

The concept of adjoint of a matrix builds upon the orthogonal relations (8.2.5) and (8.2.6) before:

$$\text{Expansion along row:}\qquad \frac{1}{\det A}\left(a_{i1}A_{i'1}+\cdots+a_{in}A_{i'n}\right)=\begin{cases}1, & \text{if } i=i';\\ 0, & \text{otherwise}\end{cases}. \tag{8.3.30}$$

$$\text{Expansion along column:}\qquad \frac{1}{\det A}\left(a_{1j}A_{1j'}+\cdots+a_{nj}A_{nj'}\right)=\begin{cases}1, & \text{if } j=j';\\ 0, & \text{otherwise}\end{cases}. \tag{8.3.31}$$

These imply the existence of *orthonormal relations*（正交归一关系）in between $(a_{i1},\cdots,a_{in})^T$ and $(A_{i'1},\cdots,A_{i'n})^T$ and in between $(a_{1j},\cdots,a_{nj})^T$ and $(A_{1j'},\cdots,A_{nj'})^T$, based on which we introduce the following concept of *adjoint matrix*.

***Definition* 8.2** (Adjoint matrix 伴随矩阵)**.** Let A be an $n\times n$ matrix,

$$A=\begin{pmatrix}a_{11}&a_{12}&\cdots&a_{1n}\\a_{21}&a_{22}&\cdots&a_{2n}\\\vdots&\vdots&\cdots&\vdots\\a_{n1}&a_{n2}&\cdots&a_{nn}\end{pmatrix}.$$

Its *adjoint matrix*, or *adjoint* for short, denoted as $\operatorname{adj} A$, is defined in terms of the cofactors of A:

$$\operatorname{adj}A=\begin{pmatrix}A_{11}&A_{12}&\cdots&A_{1n}\\A_{21}&A_{22}&\cdots&A_{2n}\\\vdots&\vdots&\cdots&\vdots\\A_{n1}&A_{n2}&\cdots&A_{nn}\end{pmatrix}^T=\begin{pmatrix}A_{11}&A_{21}&\cdots&A_{n1}\\A_{12}&A_{22}&\cdots&A_{n2}\\\vdots&\vdots&\cdots&\vdots\\A_{1n}&A_{2n}&\cdots&A_{nn}\end{pmatrix}, \tag{8.3.32}$$

where A_{ij} is the cofactor of the matrix entry a_{ij}.

Then, in the light of the orthonormal relations (8.3.30) and (8.3.31) above we immediately draw the following important conclusion.

***Theorem* 8.16.**

$$A \operatorname{adj} A = \operatorname{adj} A\, A = \det A\, I, \qquad \text{i.e.,} \qquad \frac{1}{\det A} A \ \operatorname{adj} A = \frac{1}{\det A} \operatorname{adj} A\, A = I. \tag{8.3.33}$$

This provides an alternative expression for the inverse of a matrix A:

$$A^{-1} = \frac{1}{\det A} \operatorname{adj} A, \qquad \text{when} \quad \det A \neq 0. \tag{8.3.34}$$

Proof: Ignored. ☐

✎ ***Example* 8.4.** Use the adjoint matrix approach to find the inverse of the matrix

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

Solution: First, it is easy to compute the determinant of A,

$$\det A = 5.$$

Second, the adjoint matrix is computed as

$$\operatorname{adj} A = \begin{pmatrix} \begin{vmatrix} 2 & 2 \\ 2 & 3 \end{vmatrix} & -\begin{vmatrix} 3 & 2 \\ 1 & 3 \end{vmatrix} & \begin{vmatrix} 3 & 2 \\ 1 & 2 \end{vmatrix} \\ -\begin{vmatrix} 1 & 2 \\ 2 & 3 \end{vmatrix} & \begin{vmatrix} 2 & 2 \\ 1 & 3 \end{vmatrix} & -\begin{vmatrix} 2 & 1 \\ 3 & 2 \end{vmatrix} \\ \begin{vmatrix} 1 & 2 \\ 2 & 2 \end{vmatrix} & -\begin{vmatrix} 2 & 2 \\ 3 & 2 \end{vmatrix} & \begin{vmatrix} 2 & 1 \\ 3 & 2 \end{vmatrix} \end{pmatrix}^T = \begin{pmatrix} 2 & 1 & -2 \\ -7 & 4 & 2 \\ 4 & -3 & 1 \end{pmatrix}.$$

Therefore,

$$A^{-1} = \frac{1}{5} \begin{pmatrix} 2 & 1 & -2 \\ -7 & 4 & 2 \\ 4 & -3 & 1 \end{pmatrix}.$$

The reader is invited to have a double check of this result by using the previous method of row reduction.

☐

§ **8.3.3 Cramer's rule**

In light of determinant, we can also derive a rule to obtain the solution of a linear system of equations. This rule is formal, but sometimes also practical.

***Theorem* 8.17.** (*Cramer's rule* 克莱姆法则)
Let $A\mathbf{x} = \mathbf{b}$ be a linear system of equations:

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix},$$

where A is an $n \times n$ nonsingular matrix. Then the solution of $\mathbf{x} = (x_1, x_2, \cdots, x_n)^T$ is given by

$$x_i = \frac{\det A_i}{\det A},\ i = 1, \cdots, n, \tag{8.3.35}$$

where A_i is the matrix obtained by replacing the i^{th} column of A with the column $\mathbf{b} = (b_1, b_2, \cdots, b_n)^T$.

Proof: Formally, the solution to the equation $A\mathbf{x} = \mathbf{b}$ can be expressed as

$$\mathbf{x} = A^{-1}\mathbf{b}, \qquad \text{when } A \text{ invertible.}$$

In the light of our expression for a matrix inverse of (8.3.34), $A^{-1} = \frac{1}{\det A}\text{adj } A$, we have

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \frac{1}{\det A}\,\text{adj}\,A\,\mathbf{b} = \frac{1}{\det A}\begin{pmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & a_{22} & \cdots & A_{n2} \\ \vdots & \vdots & \cdots & \vdots \\ A_{1n} & a_{2n} & \cdots & A_{nn} \end{pmatrix}\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix},$$

i.e.,

$$x_i = \frac{1}{\det A}(A_{1i}b_1 + A_{2i}b_2 + \cdots + A_{ni}b_n). \tag{8.3.36}$$

Comparing this with the definition of a determinant

$$\det A = a_{1i}A_{1i} + a_{2i}A_{2i} + \cdots + a_{ni}A_{ni},$$

it is seen that (8.3.36) can be obtained by playing the replacement

$$a_{\bullet i} \Longrightarrow b_{\bullet i},$$

namely,

$$x_i = \frac{1}{\det A} \det A_i,$$

where

$$A_i = \begin{pmatrix} a_{11} & \cdots & b_{1i} & \cdots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \cdots & b_{ni} & \cdots & a_{nn} \end{pmatrix}.$$

□

✎ ***Example* 8.5.** Solve the following linear system $A\mathbf{x} = \mathbf{b}$:

$$\begin{pmatrix} 1 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 5 \\ 6 \\ 9 \end{pmatrix}.$$

Solution:

$$\det A = \begin{vmatrix} 1 & 2 & 1 \\ 2 & 2 & 1 \\ 1 & 2 & 3 \end{vmatrix} = -4; \qquad \det A_1 = \begin{vmatrix} 5 & 2 & 1 \\ 6 & 2 & 1 \\ 9 & 2 & 3 \end{vmatrix} = -4;$$

$$\det A_2 = \begin{vmatrix} 1 & 5 & 1 \\ 2 & 6 & 1 \\ 1 & 9 & 3 \end{vmatrix} = -4; \qquad \det A_3 = \begin{vmatrix} 1 & 2 & 5 \\ 2 & 2 & 6 \\ 1 & 2 & 9 \end{vmatrix} = -8.$$

Therefore,

$$x_1 = \frac{\det A_1}{\det A} = \frac{-4}{-4} = 1;$$

$$x_2 = \frac{\det A_2}{\det A} = \frac{-4}{-4} = 1;$$

$$x_3 = \frac{\det A_3}{\det A} = \frac{-8}{-4} = 2.$$

□

Chapter 9 Eigenvalues and eigenvectors

§ 9.1 Introduction: significance of eigenvalue problems

Eigenvalue（本征值）problems — eigenvalues and corresponding eigenvectors（本征矢量）— are of extreme importance and have wide applications in natural sciences, engineering technology and social sciences, to mention just a few. The prefix *eigen-* is a German word meaning *own* and *proper*.

- **Application 1** — Principal axis of rigid body（刚体的主轴）:

  In material mechanics（材料力学）, we have the stress-strain analysis（应力-应变分析）:

  - Hooke's law（胡克定律）— simple colinear （共线）case:

    $$f = -kx. \tag{9.1.1}$$

  - Generic case: Young's modules（杨氏模量）, etc.

    $$\mathbf{T} = \sigma \mathbf{f}, \tag{9.1.2}$$

    where σ is the so-called Cauchy's stress tensor（柯西应力张量）, a matrix.

    Only in some particular directions the force and stress are colinear.

- **Application 2** — Aircraft design:

  See Page 289–291 of *S. Leon, 2010.*

  As shown in Figure 9.1, a space shuttle has three types of rotations:

$$\text{Roll } (R) \quad \text{— Rotation about the } x\text{-axis,} \tag{9.1.3}$$

$$\text{Pitch } (P) \quad \text{— Rotation about the } y\text{-axis,} \tag{9.1.4}$$

$$\text{Yaw } (Y) \quad \text{— Rotation about the } z\text{-axis.} \tag{9.1.5}$$

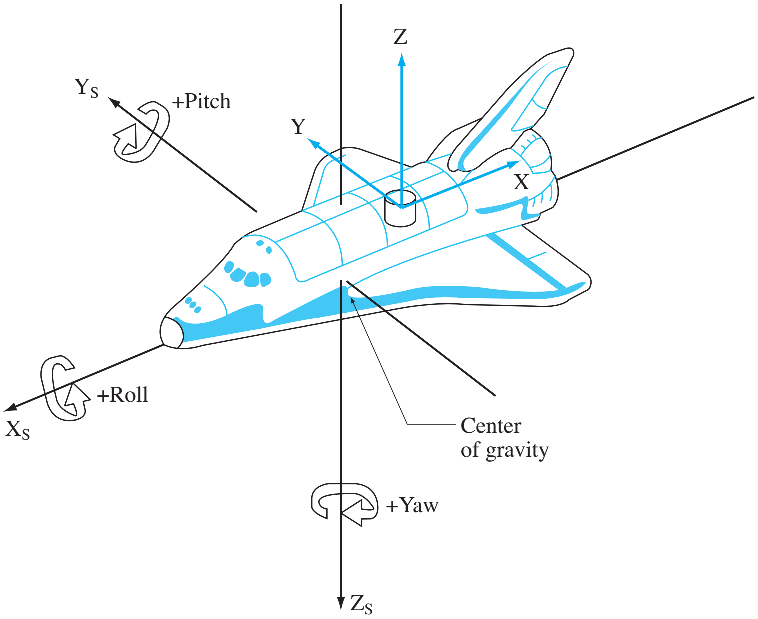


Figure 9.1: Three types of rotations of a space shuttle: roll, denoted as X_s, the rotations about the x-axis; pitch, denoted as Y_s, those about the y-axis; yaw, denoted as Z_s, those about the z-axis.

Yaw (Y), *pitch* (P) and *roll* (R) are used to describe the rotations of a space shuttle from its initial position to a new orientation. A combination of a yaw followed by a pitch and then a roll could be represented by a product $Q = YPR$. Figure 9.1 shows the axes for the yaw, pitch and roll, Z_S, Y_S and X_S, respectively, with the origin of system at the center of mass.

The Y, P and R transformations may reorient the shuttle from its initial position; however, rather than performing three separate rotations, it is more efficient to use only one rotation. This needs to determine a new single axis of rotation N and an angle of rotation β about N. The new axis should be determined by computing the eigenvectors of the transformation matrix Q.

- Other applications:
  - in digital image processing. See Pages 347–348 in the textbook of *S. Leon, 2010*.
  - in wires of music instruments, like guitar strings, where different modes of wire vibration（弦震动模式）form the musical scale （音阶）. Another example is the opening holes of a flute.
  - in social sciences, e.g., marital status computation.

Basically, for all practical models, if the problem contains a matrix (non-singular), its eigenvalue problem is usually of great importance. Moreover, besides applications in practice, eigenvalues and eigenvectors are important in theoretical studies of mathematics, such as diagonalization of matrices, etc, to be shown in this chapter.

§ 9.2 Definitions of eigenvalues and eigenvectors

§ 9.2.1 Definitions

***Definition* 9.1** (Eigenvalue and eigenvector)**.** Let A be an $n \times n$ matrix. A scalar λ is said to be an eigenvalue or characteristic value（特征值）of A, if there exists a nonzero column vector $\mathbf{x}$ such that

$$A\mathbf{x} = \lambda\mathbf{x}. \tag{9.2.1}$$

This vector $\mathbf{x}$ is called an eigenvector, or characteristic vector（特征矢量）, corresponding to the eigenvalue λ.

✎ ***Example* 9.1.** Let $A = \begin{pmatrix} 4 & -2 \\ 1 & 1 \end{pmatrix}$ be a matrix. It can be checked that not all vectors $\begin{pmatrix} x \\ y \end{pmatrix}$ satisfy

$$A\begin{pmatrix} x \\ y \end{pmatrix} = \lambda\begin{pmatrix} x \\ y \end{pmatrix}, \qquad \text{where} \qquad \lambda \in \mathbb{R}.$$

✎ ***Example* 9.2.** It is easy to check that $\begin{pmatrix} 2 & 3 \end{pmatrix}^T$ is of no hope to realize the definition (9.2.1),

$$\begin{pmatrix} 4 & -2 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} -4 \\ 5 \end{pmatrix} \neq \lambda\begin{pmatrix} 2 \\ 3 \end{pmatrix}.$$

But there do exist two families of vectors

$$\mathbf{x} = \alpha\begin{pmatrix} 2 \\ 1 \end{pmatrix} \quad \text{and} \quad \beta\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \text{where} \quad \alpha, \beta \neq 0, \ \ \alpha, \beta \in \mathbb{R},$$

satisfying

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

Let us have a double check:

$$\begin{aligned} A\mathbf{x} &= \begin{pmatrix} 4 & -2 \\ 1 & 1 \end{pmatrix}\left[\alpha\begin{pmatrix} 2 \\ 1 \end{pmatrix}\right] = \alpha\begin{pmatrix} 4 & -2 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 2 \\ 1 \end{pmatrix} \\ &= \alpha\begin{pmatrix} 6 \\ 3 \end{pmatrix} = 3\left[\alpha\begin{pmatrix} 2 \\ 1 \end{pmatrix}\right] = 3\mathbf{x}; \end{aligned}$$

$$\begin{aligned} A\mathbf{y} &= \begin{pmatrix} 4 & -2 \\ 1 & 1 \end{pmatrix}\left[\beta\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right] = \beta\begin{pmatrix} 4 & -2 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= \beta\begin{pmatrix} 2 \\ 2 \end{pmatrix} = 2\left[\beta\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right] = 2\mathbf{y}. \end{aligned}$$

§ 9.2.2 How to find eigenvalues and corresponding eigenvectors

We have the following strategy to do the job:

Step 1 Find all the eigenvalues λ for a given matrix A from the definition (9.2.1), $A\mathbf{x} = \lambda\mathbf{x}$.

Step 2 For each λ, solve out the corresponding eigenvector(s).

In the following these two steps will be narrated in detail.

Step 1: Solving eigenvalues

Rewrite the requirement $A\mathbf{x} = \lambda\mathbf{x}$ as

$$(A - \lambda I)\,\mathbf{x} = \mathbf{0}. \tag{9.2.2}$$

This is a homogeneous linear system. It is known that

- If the matrix $A - \lambda I$ is nonsingular, the vector $\mathbf{x}$ has only a trivial solution

$$\mathbf{x} = \mathbf{0} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{9.2.3}$$

- If $\mathbf{x}$ has non-zero solutions, the matrix $A - \lambda I$ must be singular.

  It is known well that

$$A \text{ singular} \quad \overset{\text{iff}}{\Longleftrightarrow} \quad \det A = 0. \tag{9.2.4}$$

  Hence to guarantee $\mathbf{x}$ has nonzero solution, there must be

$$\det(A - \lambda I) = 0. \tag{9.2.5}$$

  This is the key equation to solve the eigenvalues λ.

✎ ***Example* 9.3.** Let us continue to consider our previous example

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

We have

$$\det(A - \lambda I) = \begin{vmatrix} 4-\lambda & -2 \\ 1 & 1-\lambda \end{vmatrix} = (\lambda - 2)(\lambda - 3) = 0.$$

The solutions to this algebraic equation are $\lambda = 2, 3$, which agree to our previous guess of eigenvalues.

Step 2: Solving eigenvector(s)

After the eigenvalues λ's are achieved, the eigenvector(s) corresponding to each λ can be solved through

$$(A - \lambda I)\,\mathbf{x} = \mathbf{0}. \tag{9.2.6}$$

The procedure will be shown via the examples below.

✎ ***Example* 9.4.** The example above, $A = \begin{pmatrix} 4 & -2 \\ 1 & 1 \end{pmatrix}$, which has two eigenvalues $\lambda_1 = 2, \quad \lambda_2 = 3$.

Solution:

- For $\lambda_1 = 2$, the equation $(A - \lambda I)\,\mathbf{x} = \mathbf{0}$ reads

  $$(A - 2I)\,\mathbf{x} = \begin{pmatrix} 4-2 & -2 \\ 1 & 1-2 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{9.2.7}$$

  i.e.,

  $$\begin{aligned} & \begin{pmatrix} 2 & -2 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ \xLeftrightarrow{\text{row equivalent}} \quad & \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ \Longleftrightarrow \quad & \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \end{aligned}$$

  The solution is $x_1 = x_2$, i.e.,

  $$\mathbf{v}_1 = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} t \\ t \end{pmatrix}, \qquad \text{i.e.,} \qquad \mathbf{v}_1 = t \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad t \in \mathbb{R},$$

  where t is an introduced parameter to represent the freedom in the solution evaluation.

- For $\lambda_2 = 3$, the equation $(A - \lambda I)\,\mathbf{x} = \mathbf{0}$ turns to be

  $$(A - 3I)\mathbf{x} = \begin{pmatrix} 4-3 & -2 \\ 1 & 1-3 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

  i.e.,

  $$\begin{aligned} & \begin{pmatrix} 1 & -2 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ \xLeftrightarrow{\text{row equivalent}} \quad & \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ \Longleftrightarrow \quad & \begin{pmatrix} 1 & -2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \end{aligned}$$

The solution is $x_1 = 2x_2$, i.e.,

$$\mathbf{v}_2 = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2t \\ t \end{pmatrix}, \qquad \text{i.e.,} \qquad \mathbf{v}_2 = t \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad t \in \mathbb{R}.$$

- In summary,
  - the eigenvector corresponding to $\lambda = 2$ is $\mathbf{v}_1 = t \begin{pmatrix} 1 \\ 1 \end{pmatrix}$;
  - the eigenvector corresponding to $\lambda = 3$ is $\mathbf{v}_2 = t \begin{pmatrix} 2 \\ 1 \end{pmatrix}$.

This conclusion agrees to our previous results.

□

✎ ***Example* 9.5.** Find the eigenvalues and the corresponding eigenvectors of the matrix

$$A = \begin{pmatrix} 3 & 2 \\ 3 & -2 \end{pmatrix}. \tag{9.2.8}$$

Solution: The characteristic equation reads

$$\begin{aligned} \det(A - \lambda I) &= \begin{vmatrix} 3-\lambda & 2 \\ 3 & -2-\lambda \end{vmatrix} = -(3-\lambda)(2+\lambda) - 6 \\ &= \lambda^2 - \lambda - 12 = (\lambda - 4)(\lambda + 3). \end{aligned}$$

Hence the eigenvalues are $\lambda = 4, -3$.

- For $\lambda_1 = 4$: We have

$$(A - \lambda I)\mathbf{x} = \begin{pmatrix} -1 & 2 \\ 3 & -6 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

i.e.

$$\begin{pmatrix} -1 & 2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

The solution is $x_1 = 2x_2$, i.e.,

$$\mathbf{v}_1 = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = t \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad t \in \mathbb{R}.$$

[Remark]:

From the viewpoint of the next chapter, we can say that the solution space of $\mathbf{v}_1$ is $\mathrm{span}\{\begin{pmatrix} 2 \\ 1 \end{pmatrix}\}$.

- For $\lambda_2 = -3$: We have

  $$(A+3I)\mathbf{x} = \begin{pmatrix} 6 & 2 \\ 3 & 1 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

  i.e.,

  $$\begin{pmatrix} 3 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

  The solution is $x_1 = -\frac{1}{3}x_2$, i.e.,

  $$\mathbf{v}_2 = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = t' \begin{pmatrix} -\frac{1}{3} \\ 1 \end{pmatrix}, \quad \text{i.e.,} \quad \mathbf{v}_2 = t \begin{pmatrix} -1 \\ 3 \end{pmatrix}, \quad t', t \in \mathbb{R}.$$

  [Remark]:

  As per the next chapter, the solution space of $\mathbf{v}_2$ is $\mathrm{span}\{\begin{pmatrix} -1 \\ 3 \end{pmatrix}\}$.

□

✎ ***Example* 9.6.** Find the eigenvalues and corresponding eigenvectors of

$$A = \begin{pmatrix} 2 & -3 & 1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{pmatrix}. \tag{9.2.9}$$

Solution: The characteristic equation reads

$$\begin{aligned} \det(A-\lambda I) &= \begin{vmatrix} 2-\lambda & -3 & 1 \\ 1 & -2-\lambda & 1 \\ 1 & -3 & 2-\lambda \end{vmatrix} = -\lambda^3 + 2\lambda^2 - \lambda \\ &= -\lambda(\lambda-1)^2 = 0. \end{aligned}$$

This means the eigenvalues are $\lambda_1 = 0$ and $\lambda_2 = \lambda_3 = 1$, where λ_2 and λ_3 are repeated roots.

- For $\lambda = 0$: We have

  $$(A - 0I)\mathbf{x} = \mathbf{0},$$

  i.e.,

  $$\begin{aligned} &\begin{pmatrix} 2 & -3 & 1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \\ \Longleftrightarrow \quad &\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \end{aligned}$$

Thus the solution is given by

$$\mathbf{v}_1 = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = t \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad t \in \mathbb{R}. \qquad \text{Hence, the solution space of } \mathbf{v}_1 = \text{span}\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \}.$$

- For the repeated roots $\lambda_2 = \lambda_3 = 1$: We have $(A - I)\mathbf{x} = \mathbf{0}$, i.e.

$$\begin{pmatrix} 1 & -3 & 1 \\ 1 & -3 & 1 \\ 1 & -3 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad \Longleftrightarrow \quad \begin{pmatrix} 1 & -3 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Notice: The two all-zero rows imply two free parameters need to be introduced.

Hence, the solution reads $x_1 = 3x_2 - x_3$.

Introducing two parameters s and t as $x_2 = s, \ \ x_3 = t$, the solution is expressed as

$$\mathbf{v}_{2,3} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 3s - t \\ s \\ t \end{pmatrix} = s \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \quad s,\, t \in \mathbb{R}.$$

This means the solution space of $\mathbf{v}_{2,3}$ is given by

$$\text{span}\{ \begin{pmatrix} 3 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \}.$$

□

Complex eigenvalues

Let us start from an example.

✎ ***Example* 9.7.** Find the eigenvalues and corresponding eigenvectors of

$$A = \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}. \tag{9.2.10}$$

Solution: The characteristic equation reads

$$\det(A - \lambda I) = \begin{vmatrix} 1-\lambda & 2 \\ -2 & 1-\lambda \end{vmatrix} = \lambda^2 - 2\lambda + 5,$$

which has a pair of complex conjugate roots, $\lambda_{\pm} = 1 \pm 2i$.

- For $\lambda_+ = 1 + 2i$: We have $(A - \lambda_+ I)\mathbf{x} = \mathbf{0}$, i.e.,

$$\begin{pmatrix} -2i & 2 \\ -2 & -2i \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \Longleftrightarrow \quad \begin{pmatrix} 1 & i \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

  Hence the solution is given by

$$\mathbf{v}_+ = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = t \begin{pmatrix} -i \\ 1 \end{pmatrix}, \qquad t \in \mathbb{R}.$$

- For $\lambda_- = 1 - 2i$: We have $(A - \lambda_- I)\mathbf{x} = \mathbf{0}$, i.e.

$$\begin{pmatrix} 2i & 2 \\ -2 & 2i \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \Longleftrightarrow \quad \begin{pmatrix} -1 & i \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

  Hence the solution is given by

$$\mathbf{v}_- = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = t \begin{pmatrix} i \\ 1 \end{pmatrix}, \qquad t \in \mathbb{R}.$$

- In summary,
  - the eigenvector $\mathbf{v}_+ = t \begin{pmatrix} -i \\ 1 \end{pmatrix}$ can be equivalently expressed by $t \begin{pmatrix} 1 \\ i \end{pmatrix}$.
  - the eigenvector $\mathbf{v}_- = t \begin{pmatrix} 1 \\ i \end{pmatrix}$ can be equivalently expressed by $t \begin{pmatrix} 1 \\ -i \end{pmatrix}$.

□

From this example we see the emergence of complex conjugate of entries:

$$\lambda_1 = \lambda_2^*, \qquad \mathbf{v}_+ = \mathbf{v}_-^*. \tag{9.2.11}$$

Is this occasional? No, it is a common fact.

***Definition* 9.2** (Complex conjugate)**.**
Let $\lambda = a + bi$, where $a, b \in \mathbb{R}$. Its complex conjugate (c.c.) is defined as

$$\lambda^* = \overline{\lambda} = a - bi. \tag{9.2.12}$$

Let $A = (a_{ij})$ be an $m \times n$ matrix, $a_{ij} \in \mathbb{C}$. Its complex conjugate matrix is defined as

$$A^* = (a_{ij}^*), \qquad \text{or} \qquad \overline{A} = (\overline{a}_{ij}). \tag{9.2.13}$$

Specially,

$$A^* = A, \qquad \text{when} \qquad A \in \mathbb{R}^{m \times n}. \tag{9.2.14}$$

Since the complex conjugate operation is related only to the number of an entry, but irrelevant to its location/position in a matrix, we have

$$(AB)^* = B^* A^*. \tag{9.2.15}$$

***Theorem* 9.1.** Let A be a real matrix, and (λ, z) a pair of eigenvalue and eigenvector for A. Then the complex conjugate, $\left(\overline{\lambda}, \overline{z}\right)$, is also a pair of eigenvalue and eigenvector for A.

Proof:

$$A\overline{z} = \overline{A}\overline{z} = \overline{Az} = \overline{\lambda z} = \overline{\lambda}\overline{z} = \lambda^* \overline{z}.$$

□

[Remark]: Let us recall a fact: for a real coefficient polynomial, all its complex roots occur in conjugate pairs.

§ 9.2.3 Properties of eigenvalues and eigenvectors

Product of eigenvalues

As mentioned, to find the eigenvalues of a matrix A, we need to appeal to the characteristic polynomial

$$\mathrm{p}(\lambda) = \det(A - \lambda I) = \begin{vmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{vmatrix}.$$

Suppose the polynomial $p(\lambda)$ has n roots $\lambda_1, \lambda_2, \cdots, \lambda_n$, they should have the following factorized form according to the theory of polynomials:

$$\begin{aligned} p(\lambda) &= (\lambda_1 - \lambda)(\lambda_2 - \lambda) \cdots (\lambda_n - \lambda) \\ &= (-1)^n (\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_n). \end{aligned} \tag{9.2.16}$$

Taking $\lambda = 0$, we immediately draw a conclusion that the product of all eigenvalues gives the determinant of the matrix, i.e.,

$$p(0) = \lambda_1 \lambda_2 \cdots \lambda_n = \det A. \tag{9.2.17}$$

☞ **[Aside]:**

This conclusion (9.2.17) can also be proved in terms of the coming similarity transformation of (9.2.34). Indeed, let D be the diagonalization of A, with X being the needed similarity transformation matrix. We have

$$\begin{aligned}\det D &= \det\ \left(XAX^{-1}\right) = \det X\ \det A\ \det X^{-1} \\ &= \det X\ \det X^{-1}\ \det A = \det(XX^{-1})\det A = \det A.\end{aligned} \tag{9.2.18}$$

Given that

$$D = \operatorname{diag}\left(\lambda_1, \lambda_2, \cdots, \lambda_n\right), \tag{9.2.19}$$

we have

$$\det A = \lambda_1\lambda_2\cdots\lambda_n. \tag{9.2.20}$$

✌

Sum of eigenvalues · Trace

Trace（迹）is a very important concept in the study of matrices.

***Definition* 9.3** (Trace of matrix)**.** Let A be an $n \times n$ matrix,

$$A = (a_{ij}) = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}.$$

The trace of A is defined as the sum of the diagonal entries of A，

$$\operatorname{tr} A \equiv \text{Tr}\ A = \sum_{i=1}^{n} a_{ii}. \tag{9.2.21}$$

✎ ***Example* 9.8.**

$$\operatorname{tr}\begin{pmatrix} 5 & -18 \\ 1 & 1 \end{pmatrix} = 5 + 1 = 6.$$

***Theorem* 9.2.** Let $A = (a_{ij})$ be an $n \times n$ matrix. Its trace equals to the sum of its eigenvalues

$$\sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} a_{ii}. \tag{9.2.22}$$

Proof:

- **Method 1**:
  - Let us examine the following polynomial obtained from the characteristic determinant:

$$\begin{aligned} p(\lambda) &= \det(A - \lambda I) = \begin{vmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \\ & & & \end{vmatrix} \\ &= (a_{11} - \lambda)(a_{22} - \lambda) \cdots (a_{nn} - \lambda) + \textsf{others}, \end{aligned} \tag{9.2.23}$$

    $\textsf{others}$ referring to those terms in which the order of λ is no higher than $(n-2)$. Therefore, as far as λ^{n-1} is concerned, we only need to find it in the first term

$$(a_{11} - \lambda)(a_{22} - \lambda) \cdots (a_{nn} - \lambda), \tag{9.2.24}$$

    in the expansion of which, the λ^{n-1} term reads

$$(-1)^{n-1} \lambda^{n-1} \sum_{i=1}^{n} a_{ii}. \tag{9.2.25}$$

    Hence the coefficient of $(-\lambda)^{n-1}$ is given by

$$\sum_{i=1}^{n} a_{ii}. \tag{9.2.26}$$

  - On the other hand, as mentioned,

$$p(\lambda) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \cdots (\lambda_n - \lambda). \tag{9.2.27}$$

    In its expansion the coefficient of the $(-\lambda)^{n-1}$ term reads

$$\sum_{i=1}^{n} \lambda_i. \tag{9.2.28}$$

  - In comparison, the two ways (9.2.26) and (9.2.28) should give the same result, hence

$$\sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} a_{ii}. \tag{9.2.29}$$

- **Method 2:** (*With the aid of the coming similarity transformation of* (9.2.34) *and diagonalization of matrices in the next section*)

  Let A be an $n \times n$ matrix, and D its diagonalization,

$$X^{-1}AX = D, \qquad D = \operatorname{diag}\{\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n\}, \tag{9.2.30}$$

  L acting as the similarity transformation matrix. Then (9.2.30) leads to

$$\operatorname{tr} D = \operatorname{tr}\left(X^{-1}AX\right) = \operatorname{tr}\left(XX^{-1}A\right) = \operatorname{tr} A, \tag{9.2.31}$$

  namely,

$$\sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} a_{ii}, \tag{9.2.32}$$

  thanks to a property of the trace operation,

$$\operatorname{tr}(AB) = \operatorname{tr}(BA), \qquad \forall A, B \in \mathbb{R}^{n\times n}. \tag{9.2.33}$$

□

[Remark]: *Generalization of* (9.2.33)

$$\begin{aligned}
\operatorname{tr}\left(A_1A_2\cdots A_k\right) &= \operatorname{tr}\left(A_kA_1A_2\cdots A_{k-1}\right) \\
&= \operatorname{tr}\left(A_{k-1}A_kA_1A_2\cdots A_{k-2}\right) \\
&= \cdots \\
&= \operatorname{tr}\left(A_2A_3\cdots A_kA_1\right).
\end{aligned}$$

Similarity transformation

***Definition* 9.4.** A matrix B is said to be *similar* to another matrix A if there exists a nonsingular matrix X such that

$$B = X^{-1}AX. \tag{9.2.34}$$

$X^{-1}AX$ is called a similarity transformation（相似变换）, and X a similarity transformation matrix.

Based on similarity transformations we can close this section by presenting an important result about eigenvalues of similar matrices.

***Theorem* 9.3.** Let A and B be two $n \times n$ matrices. If they are similar to each other, they have the same characteristic polynomial and therefore the same eigenvalues,

$$p_B(\lambda) = p_A(\lambda). \tag{9.2.35}$$

Proof:

$$\begin{aligned} p_B(\lambda) &= \det(B - \lambda I) = \det\left(X^{-1}AX - \lambda I\right) = \det\left(X^{-1}AX - \lambda X^{-1}IX\right) \\ &= \det\left[X^{-1}(A - \lambda I)X\right] = \det(A - \lambda I) = p_A(\lambda). \end{aligned}$$

□

✎ ***Example* 9.9.** Consider a matrix $A = \begin{pmatrix} 2 & 1 \\ 0 & 3 \end{pmatrix}$. Check the conclusion (9.2.35) under the similarity transformation

$$X = \begin{pmatrix} 5 & 3 \\ 3 & 2 \end{pmatrix}, \qquad X^{-1} = \begin{pmatrix} 2 & -3 \\ -3 & 5 \end{pmatrix}.$$

Solution: The eigenvalues of A are computed as

$$\lambda_1^A = 2, \qquad \lambda_2^A = 2.$$

Let B be the matrix similar to A under the similarity matrix X

$$B = X^{-1}AX = \begin{pmatrix} -1 & -2 \\ 6 & 6 \end{pmatrix},$$

which has the eigenvalues

$$\lambda_1^B = 2, \quad \lambda_2^B = 2.$$

Obviously (9.2.35) holds true,

$$\{\lambda_1^A, \lambda_2^A\} = \{\lambda_1^B, \lambda_2^B\}.$$

□

§ 9.3 Diagonalization of matrices

§ 9.3.1 Procedure of diagonalization

Matrix diagonalization（矩阵对角化）is crucial and has wide applications in practice. Let us start from an example.

✎ ***Example* 9.10.** Consider a matrix

$$A = \begin{pmatrix} 3 & -1 & -2 \\ 2 & 0 & -2 \\ 2 & -1 & -1 \end{pmatrix}. \tag{9.3.1}$$

It is not easy to compute its exponential A^k, when k is a big integer, say, 1000.

However, if A is diagonalizable then this problem is easy to solve. Indeed, A can be diagonalized as $D = \operatorname{diag}\{0, 1, 1\}$, then D^k is immediately achieved as:

$$D^k = \begin{pmatrix} 0^k & & \\ & 1^k & \\ & & 1^k \end{pmatrix}. \tag{9.3.2}$$

From this example we see the significance of studying diagonalization of a matrix.

***Definition* 9.5.** An $n \times n$ matrix A is said to be diagonalizable if there exists a nonsingular matrix X and a diagonal matrix D, such that A and D are linked by X through a similarity transformation

$$X^{-1}AX = D. \tag{9.3.3}$$

This is called the diagonalization of A, or equivalently, A is diagonalized by X.

***Theorem* 9.4.** [*Procedure of diagonalizing a matrix*]
An $n \times n$ matrix A is diagonalizable, iff A has n linearly independent eigenvectors which form a nonsingular matrix

$$X = \left[\begin{pmatrix} \\ \mathbf{v}_1 \\ \\ \end{pmatrix} \begin{pmatrix} \\ \mathbf{v}_2 \\ \\ \end{pmatrix} \cdots \begin{pmatrix} \\ \mathbf{v}_n \\ \\ \end{pmatrix} \right], \qquad \det X \neq 0. \tag{9.3.4}$$

Proof: Notice, the proof below indeed provides a method to diagonalize a matrix.

Suppose A has n eigenvalues λ_i, $i = 1, 2, \cdots, n$; here λ_i's are permitted to be equal. Let $\mathbf{v}_j$ be the eigenvector corresponding to λ_i, i.e.,

$$A\mathbf{v}_i = \lambda_i \mathbf{v}_i. \tag{9.3.5}$$

First, considering a matrix formed by the $\mathbf{v}_i$'s,

$$X = \left[\begin{pmatrix} \\ \mathbf{v}_1 \\ \\ \end{pmatrix} \begin{pmatrix} \\ \mathbf{v}_2 \\ \\ \end{pmatrix} \cdots \begin{pmatrix} \\ \mathbf{v}_n \\ \\ \end{pmatrix} \right], \qquad \mathbf{v}_i\text{'s being column vectors,} \tag{9.3.6}$$

we have the action of A upon X,

$$\begin{aligned} AX &= A \left[\begin{pmatrix} \\ \mathbf{v}_1 \\ \\ \end{pmatrix} \begin{pmatrix} \\ \mathbf{v}_2 \\ \\ \end{pmatrix} \cdots \begin{pmatrix} \\ \mathbf{v}_n \\ \\ \end{pmatrix} \right] = \left[A \begin{pmatrix} \\ \mathbf{v}_1 \\ \\ \end{pmatrix} A \begin{pmatrix} \\ \mathbf{v}_2 \\ \\ \end{pmatrix} \cdots A \begin{pmatrix} \\ \mathbf{v}_n \\ \\ \end{pmatrix} \right] \\ &= \left[\lambda_1 \begin{pmatrix} \\ \mathbf{v}_1 \\ \\ \end{pmatrix} \lambda_2 \begin{pmatrix} \\ \mathbf{v}_2 \\ \\ \end{pmatrix} \cdots \lambda_n \begin{pmatrix} \\ \mathbf{v}_n \\ \\ \end{pmatrix} \right]. \end{aligned} \tag{9.3.7}$$

Second, suppose

$$X^{-1} = \begin{pmatrix} (\mathbf{w}_1^T) \\ (\mathbf{w}_2^T) \\ \vdots \\ (\mathbf{w}_n^T) \end{pmatrix}, \qquad \mathbf{w}_i^T\text{'s being row vectors,} \tag{9.3.8}$$

satisfying

$$X^{-1}X = I, \qquad \text{i.e.,} \qquad \mathbf{w}_i^T \mathbf{v}_j = \delta_{ij}. \tag{9.3.9}$$

Then $X^{-1}AX$ immediately yields the diagonalization of the matrix A:

$$\begin{aligned} X^{-1}AX &= \begin{pmatrix} (\mathbf{w}_1^T) \\ (\mathbf{w}_2^T) \\ \vdots \\ (\mathbf{w}_n^T) \end{pmatrix} A \left[\begin{pmatrix} \\ \mathbf{v}_1 \\ \\ \end{pmatrix} \begin{pmatrix} \\ \mathbf{v}_2 \\ \\ \end{pmatrix} \cdots \begin{pmatrix} \\ \mathbf{v}_n \\ \\ \end{pmatrix} \right] \\ &= \begin{pmatrix} (\mathbf{w}_1^T) \\ (\mathbf{w}_2^T) \\ \vdots \\ (\mathbf{w}_n^T) \end{pmatrix} \left[\lambda_1 \begin{pmatrix} \\ \mathbf{v}_1 \\ \\ \end{pmatrix} \lambda_2 \begin{pmatrix} \\ \mathbf{v}_2 \\ \\ \end{pmatrix} \cdots \lambda_n \begin{pmatrix} \\ \mathbf{v}_n \\ \\ \end{pmatrix} \right] \\ &= \begin{pmatrix} \boxed{\lambda_1 \mathbf{w}_1^T \mathbf{v}_1} & \lambda_2 \mathbf{w}_1^T \mathbf{v}_2 & \cdots & \lambda_n \mathbf{w}_1^T \mathbf{v}_n \\ \lambda_1 \mathbf{w}_2^T \mathbf{v}_1 & \boxed{\lambda_2 \mathbf{w}_2^T \mathbf{v}_2} & \cdots & \lambda_n \mathbf{w}_2^T \mathbf{v}_n \\ \vdots & & & \\ \lambda_1 \mathbf{w}_n^T \mathbf{v}_1 & \lambda_2 \mathbf{w}_n^T \mathbf{v}_2 & \cdots & \boxed{\lambda_n \mathbf{w}_n^T \mathbf{v}_n} \end{pmatrix}. \end{aligned} \tag{9.3.10}$$

Since

$$\mathbf{w}_i^T \mathbf{v}_j = \delta_{ij} = \begin{cases} 1, & \text{when } i = j, \\ 0, & \text{when } i \neq j, \end{cases} \tag{9.3.11}$$

we have

$$X^{-1}AX = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} = D. \tag{9.3.12}$$

This completes the diagonalization

$$X^{-1}AX = D, \qquad \text{i.e.,} \qquad A = XDX^{-1}. \tag{9.3.13}$$

□

[Remarks]:

- The above procedure makes sense when X is nonsingular, i.e., when X^{-1} is achievable.

  Obviously, if A has l eigenvectors only, $l < n$, it is impossible to have A is diagonalizable. The condition to keep A diagonalizable is that

  the n eigenvectors which form the matrix X must be linearly independent.

- The diagonalizing matrix X is not unique, due to a fact that a different ordering of the eigenvectors leads to a different matrix X', and therefore a different matrix D',

$$D' = \begin{pmatrix} \lambda_1' & & & \\ & \lambda_2' & & \\ & & \ddots & \\ & & & \lambda_n' \end{pmatrix}, \tag{9.3.14}$$

where $\{\lambda_1', \lambda_2', \cdots, \lambda_n'\}$ is a re-ordering of $\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$, illustrated as

$$X' = \left[\begin{pmatrix} \\ \mathbf{v}_1' \\ \\ \end{pmatrix} \begin{pmatrix} \\ \mathbf{v}_2' \\ \\ \end{pmatrix} \cdots \begin{pmatrix} \\ \mathbf{v}_n' \\ \\ \end{pmatrix} \right] \Longrightarrow D' = \begin{pmatrix} \lambda_1' & & & \\ & \lambda_2' & & \\ & & \ddots & \\ & & & \lambda_n' \end{pmatrix}. \tag{9.3.15}$$

§ 9.3.2 Applications

As mentioned, equipped with the diagonalization $A = XDX^{-1}$, it is easier to evaluate a power A^k:

$$\begin{aligned} A^k &= \left(XDX^{-1}\right)^k = \overbrace{\left(XDX^{-1}\right)\left(XDX^{-1}\right)\cdots\left(XDX^{-1}\right)}^{k \text{ copies}} \\ &= XD\left(X^{-1}X\right)D\left(X^{-1}X\right)D\ \cdots\ D\left(X^{-1}X\right)DX^{-1} \\ &= XD^kX^{-1} = X \begin{pmatrix} \lambda_1^k & & & \\ & \lambda_2^k & & \\ & & \ddots & \\ & & & \lambda_n^k \end{pmatrix} X^{-1}. \end{aligned} \tag{9.3.16}$$

In the following, we will use a few examples to show the procedure of diagonalizing a given matrix, as well as the convenience of computing a power of a matrix once its diagonalization is achievable.

✎ ***Example* 9.11.** Find the diagonalization of a 2×2 matrix

$$A = \begin{pmatrix} 2 & -3 \\ 2 & -5 \end{pmatrix}. \tag{9.3.17}$$

Solution:

1. Use the characteristic equation, $\det(A - \lambda I) = 0$, to solve out the eigenvalues:

$$\lambda_1 = 1, \qquad \lambda_2 = -4. \tag{9.3.18}$$

2. For $\lambda_1 = 1$: Solve out the corresponding eigenvector

$$\mathbf{v}_1 = \begin{pmatrix} 3 \\ 1 \end{pmatrix}. \tag{9.3.19}$$

For $\lambda_2 = -4$: Solve out its eigenvector

$$\mathbf{v}_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}. \tag{9.3.20}$$

3. Use $\mathbf{v}_1$ and $\mathbf{v}_2$ to form a diagonalizing matrix

$$X = \left[\begin{pmatrix} \mathbf{v}_1 \end{pmatrix} \begin{pmatrix} \mathbf{v}_2 \end{pmatrix} \right] = \left[\begin{pmatrix} 3 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right]. \tag{9.3.21}$$

4. Check the nonsingularity of X:

$$\det X = 5 \neq 0.$$

5. Use the row reduction method to find the inverse X^{-1}:

$$X^{-1} = \frac{1}{5} \begin{pmatrix} 2 & -1 \\ -1 & 3 \end{pmatrix}. \tag{9.3.22}$$

6. Complete the diagonalization $X^{-1}AX = D$:

$$\frac{1}{5} \begin{pmatrix} 2 & -1 \\ -1 & 3 \end{pmatrix} \begin{pmatrix} 2 & -3 \\ 2 & -5 \end{pmatrix} \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -4 \end{pmatrix}, \tag{9.3.23}$$

i.e.,

$$\begin{pmatrix} 2 & -3 \\ 2 & -5 \end{pmatrix} = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -4 \end{pmatrix} \left[\frac{1}{5} \begin{pmatrix} 2 & -1 \\ -1 & 3 \end{pmatrix} \right]. \tag{9.3.24}$$

□

✎ ***Example* 9.12.** Consider a 3×3 matrix

$$A = \begin{pmatrix} 3 & -1 & -2 \\ 2 & 0 & -2 \\ 2 & -1 & -1 \end{pmatrix}. \tag{9.3.25}$$

Use its diagonalization to evaluate the power A^{10}.

Solution:

1. Use the characteristic equation $\det(A - \lambda I) = 0$ to find the eigenvalues:

$$\lambda_1 = 0, \qquad \lambda_2 = \lambda_3 = 1 \quad \text{(duplicate roots).} \tag{9.3.26}$$

2. Find the eigenvectors:

$$\lambda_1 \;\to\; \mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}; \qquad \lambda_2 = \lambda_3 \;\to\; \mathbf{v}_2 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \;\; \mathbf{v}_3 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}. \tag{9.3.27}$$

3. Use $\mathbf{v}_1,\ \mathbf{v}_2,\ \mathbf{v}_3$ to form a diagonalizing matrix

$$X = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix}. \tag{9.3.28}$$

4. Non-singularity check:

$$\det X = -1 \neq 0. \tag{9.3.29}$$

5. Find the inverse of X:

$$X^{-1} = \begin{pmatrix} -2 & 1 & 2 \\ 1 & 0 & -1 \\ 2 & -1 & -1 \end{pmatrix}. \tag{9.3.30}$$

6. Finally, complete the diagonalization of $A : X^{-1}AX = D$, i.e.,

$$\begin{pmatrix} -2 & 1 & 2 \\ 1 & 0 & -1 \\ 2 & -1 & -1 \end{pmatrix}\begin{pmatrix} 3 & -1 & -2 \\ 2 & 0 & -2 \\ 2 & -1 & -1 \end{pmatrix}\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & & \\ & 1 & \\ & & 1 \end{pmatrix}; \tag{9.3.31}$$

or equivalently, $A = XDX^{-1}$:

$$\begin{pmatrix} 3 & -1 & -2 \\ 2 & 0 & -2 \\ 2 & -1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & & \\ & 1 & \\ & & 1 \end{pmatrix}\begin{pmatrix} -2 & 1 & 2 \\ 1 & 0 & -1 \\ 2 & -1 & -1 \end{pmatrix}. \tag{9.3.32}$$

Furthermore, on the top of the above result of diagonalization, one can compute

$$A^{10} = \left(XDX^{-1}\right)^{10} = XD^{10}X^{-1} = X\begin{pmatrix} 0^{10} & & \\ & 1^{10} & \\ & & 1^{10} \end{pmatrix} = XDX^{-1} = A. \tag{9.3.33}$$

□

Summary: Program to achieve diagonalization of matrix

— This demands almost all our knowledge already learnt about matrices.

STEPS:

1. For a given matrix A, use the characteristic equation $\det(A - \lambda I) = 0$ to solve its eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$. *Notice:* Some λ_i's are permitted to duplicate.
2. For each eigenvalue λ_i, find its corresponding eigenvector $\mathbf{v}_i$.
3. Use the eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ to form a diagonalizing matrix $X = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$.
4. For safety, check $\det X \neq 0$ to ensure the non-singularity of X. If failed, the original A is not diagonalizable.
5. Solve out the inverse X^{-1}, by means of elementary row operations OR adjoint matrices.
6. Finally, complete the diagonalization of A :

$$X^{-1}AX = D, \quad \text{i.e.,} \quad A = XDX^{-1}, \qquad \text{where} \qquad D = \mathrm{diag}\left\{\lambda_1, \lambda_2, \cdots, \lambda_n\right\}.$$

Extra steps

Furthermore,

- if asked to find A^k, one has

$$A^k = (XDX^{-1})^k = XD^kX^{-1} = X \begin{pmatrix} \lambda_1^k & & & \\ & \lambda_2^k & & \\ & & \ddots & \\ & & & \lambda_n^k \end{pmatrix} X^{-1}.$$

- if asked to find the quadratic form (*see the next section*)

$$ax^2 + 2bxy + cy^2 = \lambda_1 x'^2 + \lambda_2 y'^2,$$

one has

$$\begin{array}{ccccc} \mathbf{x}^T A\mathbf{x} & = & \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} & = & \mathbf{x}^T Q^T D Q\mathbf{x} \\ & & \circlearrowright & & \| \\ \mathbf{x}'^T D\mathbf{x}' & = & \begin{pmatrix} x' & y' \end{pmatrix} \begin{pmatrix} \lambda_1 & \\ & \lambda_2 \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} & = & (Q\mathbf{x})^T D(Q\mathbf{x}) \end{array}$$

§ 9.4 Quadratic forms

This section is an application of matrix diagonalization. Emphasis will be placed on diagonalization of a quadratic form as well as its geometric meaning.

§ 9.4.1 Quadratic forms

***Definition* 9.6.** A quadratic form（二次型）in two variables is given by

$$ax^2 + 2bxy + cy^2, \qquad a, b, c \in \mathbb{R}, \tag{9.4.1}$$

which can be written in a matrix form as

$$ax^2 + 2bxy + cy^2 = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \mathbf{x}^T A \mathbf{x}, \tag{9.4.2}$$

where

$$\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}, \qquad A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}.$$

[Remarks]:

- The textbook of *S. Leon 2010* starts from a quadratic equation

  $$ax^2 + 2bxy + cy^2 + dx + ey + f = 0. \tag{9.4.3}$$

  Here the extra 0th and 1st order terms $dx + ey + f$ actually indicate translation of diagrams, which belongs to knowledge of high school maths. Hence it is nonsense to repeat this redundant and even misleading information here; to highlight the quadratic form, we safely start from (9.4.1).

- The $2bxy$ in (9.4.1) is called a crossing term; in (9.4.2) it corresponds to the off-diagonal entries in the matrix A. Obviously, a quadratic form without a crossing term is much easier to deal with than a one with a crossing term,

With crossing term		Without crossing term
$ax^2 + 2bxy + cy^2$	Versus	$a'x'^2 + b'y'^2$
complicated		easy

  In following you will see that a form without crossings actually corresponds to a regular-looking quadratic curve whose symmetric axes fall upon the coordinate axes, while a form with crossings corresponds to a curve undergoing a rotation of the regular one to an angle $\theta \neq 0$, as shown in Figure 9.2. In this regard we try to develop a method to turn a form with crossings into a form without them, to simply the studied problem and highlight the geometric essence therein.

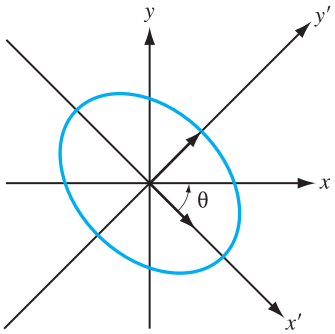


Figure 9.2: A curve undergoing a rotation with angle $\theta \neq 0$: x' and y' are the symmetric axes of the ellipse, and x and y are the coordinate axes. There exists an angle θ in between the xy- and $x'y'$-coordinate frames.

Diagonalization of a quadratic form

The crossing term arises from the nonzero off-diagonal entries b in the matrix A in (9.4.2); thus diagonalization of A implies to eliminate the off-diagonal entries.

Let the diagonalization of A be

$$QAQ^{-1} = D, \quad \text{i.e.,} \quad A = Q^{-1}DQ, \qquad \text{where} \quad D = \begin{pmatrix} \lambda_1 & \\ & \lambda_2 \end{pmatrix}. \tag{9.4.4}$$

Here Q is the diagonalizing matrix of A.

[Remark]: *Notice, Q is different from the X in* (9.3.3)*, who are transpose of each other,*

$$Q = X^{-1}. \tag{9.4.5}$$

The purpose of using the inverse of X is for the convenience of (9.4.7) *below.*

Then (9.4.2) becomes

$$\mathbf{x}^T A\mathbf{x} = \mathbf{x}^T Q^{-1} DQ\mathbf{x}. \tag{9.4.6}$$

We have a *hypothesis*:

$$\text{if} \quad Q^{-1} = Q^T, \qquad \text{then} \qquad \mathbf{x}^T Q^{-1} = \mathbf{x}^T Q^T = (Q\mathbf{x})^T, \tag{9.4.7}$$

which immediately yields an important result

$$\mathbf{x}^T A\mathbf{x} = (Q\mathbf{x})^T D\,(Q\mathbf{x}). \tag{9.4.8}$$

Defining $\mathbf{x}' = Q\mathbf{x}$, (9.4.8) precisely means

$$\mathbf{x}^T A\mathbf{x} = \mathbf{x}'^T D\mathbf{x}'. \tag{9.4.9}$$

Then the problem is how to meet the requirement of the above hypothesis (9.4.7). In this regard we need to appeal to the concept of *orthogonal matrix*（正交矩阵）.

***Definition* 9.7** (Orthogonal matrix).

An $n \times n$ matrix Q is said to be orthogonal, iff the column vectors of Q form an orthonormal set in $\mathbb{R}^n$.

Orthogonal matrices have an important feature.

***Theorem* 9.5.** Let Q be an orthogonal matrix. Then

$$Q^T Q = I, \qquad \text{i.e.,} \qquad Q^T = Q^{-1}. \tag{9.4.10}$$

Proof: Let $Q = \left[\left(\mathbf{v}_i\right)\right]$, where $\mathbf{v}_i$'s are column vectors orthonormal to each other, i.e.,

$$\mathbf{v}_i^T \mathbf{v}_j = \delta_{ij}.$$

Then,

$$Q^T Q = (a_{ij}), \qquad \text{where each} \qquad a_{ij} = \mathbf{v}_i^T \mathbf{v}_j = \delta_{ij},$$

hence

$$Q^T Q = I, \qquad \text{i.e.,} \qquad Q^T = Q^{-1}.$$

□

Then the theorem below gives the condition under which the hypothesis holds true.

***Theorem* 9.6.** If A is a real symmetric matrix（实对称矩阵）

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}, \qquad a, b, c \in \mathbb{R},$$

then A can be diagonalized by an orthogonal matrix Q:

$$A = Q^T D Q, \qquad \text{where } D \text{ is diagonal, and } Q \text{ is orthogonal,} \quad Q^T = Q^{-1}. \tag{9.4.11}$$

Proof: Ignored. Interested reader is referred to the proof on Pages 332–333 of the textbook of *S. Leon, 2010*.

□

Thus, the hypothesis (9.4.9) holds true

$$\mathbf{x}^T A \mathbf{x} = \mathbf{x}'^T D \mathbf{x}', \qquad \text{where } \mathbf{x}' = Q\mathbf{x}, \quad \text{with } D \text{ diagonal and } Q \text{ orthogonal, } Q^T = Q^{-1}. \tag{9.4.12}$$

§ 9.4.2 Geometric meaning of orthogonal transformation in 2D: rotation in plane

In this subsection the reader is referred to Page 353–357 of the textbook of S. Leon, 2010.

***Theorem* 9.7.** If Q is a 2×2 orthogonal matrix, it can be expressed as a *rotational matrix*（旋转矩阵）,

$$Q = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \tag{9.4.13}$$

Proof: Let $Q = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = [\mathbf{v}_1, \ \ \mathbf{v}_2]$, where $\mathbf{v}_j$ denotes a column vector, $\mathbf{v}_j = \begin{pmatrix} a_{1j} \\ a_{2j} \end{pmatrix}$, $j = 1, 2.$ Definition 9.7 on Page 158 requires

$$a_{11}a_{11} + a_{21}a_{21} = 1, \qquad a_{12}a_{22} + a_{21}a_{22} = 1, \qquad a_{11}a_{12} + a_{21}a_{22} = 0.$$

The solution to the above three equations is

$$Q = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \qquad \text{with } \theta \in \mathbb{R} \text{ as a free parameter.}$$

□

[Remarks]:

1. Q arises from the diagonalizing matrix X of A, with X constructed in terms of eigenvectors of A. Previously, when solving an eigenvector of A we intend to take benefit of integer entries, such as

   $$X = \left[\begin{pmatrix} 1 \\ 2 \end{pmatrix} \begin{pmatrix} 3 \\ 4 \end{pmatrix} \right],$$

   where $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ comes from $t\begin{pmatrix} 1 \\ 2 \end{pmatrix}$, with $t = 1$; $\begin{pmatrix} 3 \\ 4 \end{pmatrix}$ comes from $t\begin{pmatrix} 3 \\ 4 \end{pmatrix}$, with $t = 1$. We were able to take $t = 1$ because integers are easy to read and there was no particular requirement for t, hence t is free to choose.
   However, in the present case there exists a requirement $\cos^2\theta + \sin^2\theta = 1$, which removes the above freedom. t needs to act as a normalization factor by taking a particular evaluation.
2. It can be checked that Q is an orthogonal matrix satisfying $Q^TQ = I$:

   $$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I.$$

3. Geometric meaning — rotational matrix: The transformation

$$\mathbf{x}' = Q\mathbf{x}$$

describes a rotation from a vector $\mathbf{x}$ to another $\mathbf{x}'$. It is noticed that for this 2×2 matrix Q its determinant is $\det Q = 1$, which guarantees

$$\mid \mathbf{x}' \mid = \mid \mathbf{x} \mid, \tag{9.4.14}$$

namely, the length of the vector $\mathbf{x}$ is preserved.

4. An immediate corollary:
   Geometrically, the inverse of a θ-rotation must be a $(-\theta)$-rotation. Algebraic computation confirms this fact

$$Q(-\theta) = \begin{pmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos\theta \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} = Q(\theta)^T,$$

i.e.,

$$[Q(\theta)]^T = Q(-\theta). \tag{9.4.15}$$

✎ ***Example* 9.13.** Consider the curve described by a quadratic form

$$3x^2 + 2xy + 3y^2 - 8 = 0. \tag{9.4.16}$$

Try to determine the shape of this quadratic curve.

Solution: (9.4.16) can be rewritten as

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 8.$$

It is easy to find the eigenvalues of the matrix A: $\lambda_1 = 2, \ \ \lambda_2 = 4$. Hence A should have the following diagonalization

$$A = Q^T D Q,$$

i.e.,

$$\begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} 2 & \\ & 4 \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

In light of the eigenvectors it is easy to evaluate $\theta = \frac{\pi}{4}$, and hence $\cos\theta = \sin\theta = \frac{1}{\sqrt{2}}$.

Defining

$$\mathbf{x}' = \begin{pmatrix} x' \\ y' \end{pmatrix} = Q\mathbf{x} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$
$$= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} x - y \\ x + y \end{pmatrix},$$

we have

$$3x^2 + 2xy + 3y^2 = \begin{pmatrix} x' & y' \end{pmatrix} \begin{pmatrix} 2 & \\ & 4 \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} = 8,$$

i.e.,

$$2x'^2 + 4y'^2 = 8 \quad \Longleftrightarrow \quad \frac{x'^2}{2^2} + \frac{y'^2}{\sqrt{2}^2} = 1.$$

It is an ellipse, centered at the origin, with the long and short half axes as 2 and $\sqrt{2}$, respectively. The shape of this curve is indeed the one in Figure 9.2.

□

§ **9.4.3 Positive and negative definiteness**

Furthermore we have the following definitions, which will find important applications in optimization of multi-variable functions in Calculus.

***Definition* 9.8.** Let $f(\mathbf{x})$ be a quadratic form,

$$f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}. \tag{9.4.17}$$

- Definiteness:

  $f(\mathbf{x})$ is said to be definite if it takes on only one sign as $\mathbf{x}$ varies over all nonzero vectors in $\mathbb{R}^n$; otherwise, $f(\mathbf{x})$ is said to be indefinite.

- Positive definiteness（正定性）:

$$\mathbf{x}^T A\mathbf{x} > 0, \qquad \forall \mathbf{x} \in \mathbb{R}^n, \quad \mathbf{x} \neq 0. \tag{9.4.18}$$

- Positive semi-definiteness（半正定性）:

$$\mathbf{x}^T A\mathbf{x} \geq 0, \qquad \forall \mathbf{x} \in \mathbb{R}^n, \quad \mathbf{x} \neq 0. \tag{9.4.19}$$

- Negative definiteness（负定性）:

$$\mathbf{x}^T A\mathbf{x} < 0, \qquad \forall \mathbf{x} \in \mathbb{R}^n, \quad \mathbf{x} \neq 0. \tag{9.4.20}$$

- Negative semi-definiteness（半负定性）:

$$\mathbf{x}^T A\mathbf{x} \leq 0, \qquad \forall \mathbf{x} \in \mathbb{R}^n, \quad \mathbf{x} \neq 0. \tag{9.4.21}$$

The following theorem shows the relationship between positive definiteness and eigenvalues.

***Theorem* 9.8.** A real symmetric matrix A is positive definite, iff all eigenvalues of A are positive, i.e.,

$$A = Q^T D Q = Q^T \begin{pmatrix} \lambda_1 & \\ & \lambda_2 \end{pmatrix} Q, \qquad \text{where } \lambda_1, \lambda_2 > 0. \tag{9.4.22}$$

Proof: Ignored. □

Chapter 10 Vector space (*Brief review*)

In this chapter we will briefly review some profound but important concepts in linear algebra. Interested reader is referred to general reference textbooks for details, and our subsequent courses on theory of algebra.

§ 10.1 Rank of matrix

§ 10.1.1 Non-zero rows in reduced echelon form

Let us recall the procedure of obtaining the row echelon form of a matrix A:

$$\begin{pmatrix} * & * & \cdots & * \\ * & * & \cdots & * \\ \vdots & \vdots & & \vdots \\ * & * & \cdots & * \end{pmatrix} \xrightarrow[\text{elimination}]{\text{Gaussian}} \left(\begin{array}{cccccc} 1 & * & * & & \text{\Large $*$} & \\ & & 1 & * & & \\ & & & 1 & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 \\ & & \multicolumn{2}{c}{\cdots\cdots} & & \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right) \begin{array}{l} \left.\vphantom{\begin{array}{c}1\\1\\1\end{array}}\right\} \text{non-zero-row vects} \\ \left.\vphantom{\begin{array}{c}0\\ \cdots\\0\end{array}}\right\} \text{all-zero-row vects} \end{array}$$

We see important facts from the procedure:

- The number of the non-zero row vectors has been reduced to the minimum extent. All the non-zero rows are linearly independent of each other.
- An all-zero row is achieved actually by expressing it as a linear combination of the non-zero rows.

§ 10.1.2 Revisit to linear independence of vectors

Let us recollect the basic ideas of linear independence of vectors (see Chapter 2).

- In one dimension:

  Any two given non-zero vectors, say $\mathbf{v}$ and $\mathbf{a}$, are linear dependent. Let $\mathbf{a}$ be the basis, then $\mathbf{v}$ can be expressed as:

$$\mathbf{v} = \alpha\mathbf{a}, \qquad \alpha \in \mathbb{R}. \tag{10.1.1}$$

We can say this basis $\{\mathbf{a}\}$ spans a 1-dimensional space（张开一维空间）, denoted as $\mathrm{span}\,\{\mathbf{a}\}$. Then (10.1.1) can be equivalently expressed by

$$\mathbf{v} \subset \mathrm{span}\,\{\mathbf{a}\}\,. \tag{10.1.2}$$

- In two dimensions:

  Let $\mathbf{a}_1$ and $\mathbf{a}_2$ be two independent vectors, then $\{\mathbf{a}_1, \mathbf{a}_2\}$ forms a basis. Any vector $\mathbf{v}$ can be expanded onto this basis:

$$\mathbf{v} = \alpha\mathbf{a_1} + \beta\mathbf{a_2}, \qquad \alpha, \beta \in \mathbb{R}. \tag{10.1.3}$$

  The basis $\{\mathbf{a}_1, \mathbf{a}_2\}$ spans a 2-dimensional space, denoted as $\mathrm{span}\,\{\mathbf{a}_1, \mathbf{a}_2\}$. (10.1.3) can be equivalently expressed by

$$\mathbf{v} \subset \mathrm{span}\,\{\mathbf{a}_1, \mathbf{a}_2\}\,. \tag{10.1.4}$$

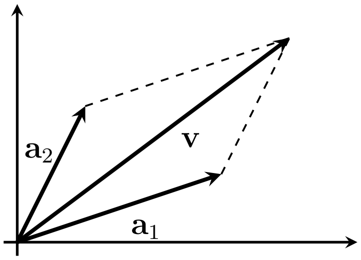


Figure 10.1: $\mathbf{a}_1$, $\mathbf{a}_2$ and $\mathbf{v}$ are three vectors in two dimensions. $\mathbf{a}_1$ and $\mathbf{a}_2$ are linearly independent, forming a basis $\{\mathbf{a}_1, \mathbf{a}_2\}$. $\mathbf{v}$ is linearly dependent on $\mathbf{a}_1$ and $\mathbf{a}_2$, namely, it can be expanded onto the basis $\{\mathbf{a}_1, \mathbf{a}_2\}$ as: $\mathbf{v} = \alpha\mathbf{a_1} + \beta\mathbf{a_2}$, where α and β are two coefficients, $\alpha, \beta \in \mathbb{R}$.

- In three dimensions:

  Let $\mathbf{a}_1,\ \mathbf{a}_2$ and $\mathbf{a}_3$ be three independent vectors, then $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ forms a basis. Any vector $\mathbf{v}$ can be expanded onto this basis:

$$\mathbf{v} = \alpha\mathbf{a_1} + \beta\mathbf{a_2} + \gamma\mathbf{a_3}, \qquad \alpha, \beta, \gamma \in \mathbb{R}. \tag{10.1.5}$$

  The basis $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ spans a 3-dimensional space, denoted as $\mathrm{span}\,\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$. (10.1.5) can be equivalently expressed by

$$\mathbf{v} \subset \mathrm{span}\,\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}\,. \tag{10.1.6}$$

Linear independent row vectors span a row vector space（行矢量空间）or row space.

✎ ***Example* 10.1.** Consider a matrix composed of four row vectors $\mathbf{v}_1^T, \cdots, \mathbf{v}_4^T$:

$$\begin{pmatrix} * & \cdots & * \\ * & \cdots & * \\ * & \cdots & * \\ * & \cdots & * \end{pmatrix} \begin{matrix} \leftarrow \mathbf{v}_1^T \\ \leftarrow \mathbf{v}_2^T \\ \leftarrow \mathbf{v}_3^T \\ \leftarrow \mathbf{v}_4^T \end{matrix}$$

Suppose we know the minimum number of linear independent row vectors is 3, and $\mathbf{v}_1^T$ and $\mathbf{v}_2^T$ are independent. Then we have two choices for when deciding the basis:

- Basis $\left\{\mathbf{v}_1^T, \mathbf{v}_2^T, \mathbf{v}_3^T\right\}$: Then
$$\mathbf{v}_4^T \subset \operatorname{span}\left\{\mathbf{v}_1^T, \mathbf{v}_2^T, \mathbf{v}_3^T\right\}.$$
- Basis $\left\{\mathbf{v}_1^T, \mathbf{v}_2^T, \mathbf{v}_4^T\right\}$: Then
$$\mathbf{v}_3^T \subset \operatorname{span}\left\{\mathbf{v}_1^T, \mathbf{v}_2^T, \mathbf{v}_4^T\right\}.$$

Always keep in mind, although the choices are different, the number of minimum linearly independent rows keeps intact. One can regard this minimum number, in a sense, as the number of the **useful** vectors among all the vectors.

§ 10.1.3 Definition of rank of matrix

This number of *useful* vectors is exactly the *rank* of a matrix（矩阵的秩）.

***Definition* 10.1** (Rank of matrix)**.**
The rank of a matrix A, denoted as $\operatorname{rank} A$, is the dimension of the row space of A, i.e.,

$$\operatorname{rank} A = \text{number of independent row vectors.} \tag{10.1.7}$$

Then, how to achieve the rank of a matrix? We have two ways, seemingly identical.

- Using the Gaussian elimination to reduce the matrix to generate zero rows as many as possible:

$$\left(\begin{array}{ccc} * & \cdots & * \\ \vdots & & \vdots \\ * & \cdots & * \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{array}\right) \begin{array}{l} \left.\begin{array}{l} \\ \\ \\ \end{array}\right\} \operatorname{rank} A \\ \\ \\ \\ \end{array} \tag{10.1.8}$$

- Obtaining the row echelon form of the matrix

$$\left(\begin{array}{cccccc} 1 & * & * & & \text{\huge $*$} & \\ & & 1 & * & & \\ & & & 1 & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \multicolumn{6}{c}{\cdots\cdots\cdots} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right) \begin{array}{l} \left.\begin{array}{l} \\ \\ \\ \end{array}\right\} \operatorname{rank} A \\ \\ \\ \\ \end{array} \tag{10.1.9}$$

✎ ***Example* 10.2.** Determine the rank of

$$A = \begin{pmatrix} 1 & 5 & 2 & 4 \\ 3 & 3 & 7 & -1 \\ 4 & 8 & 9 & 3 \\ 2 & -2 & 5 & -5 \end{pmatrix} \begin{matrix} \leftarrow R_1 \\ \leftarrow R_2 \\ \leftarrow R_3 \\ \leftarrow R_4 \end{matrix}$$

Solution: It is seen that R_1 and R_2 are linear independent, but

$$R_3 = R_1 + R_2, \qquad R_4 = R_2 - R_1.$$

Hence

$$\begin{pmatrix} 1 & 5 & 2 & 4 \\ 3 & 3 & 7 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}1\\3\end{matrix}}\right\} \text{non-zero-row vectors} \\ \vphantom{0} \\ \vphantom{0} \end{matrix}$$

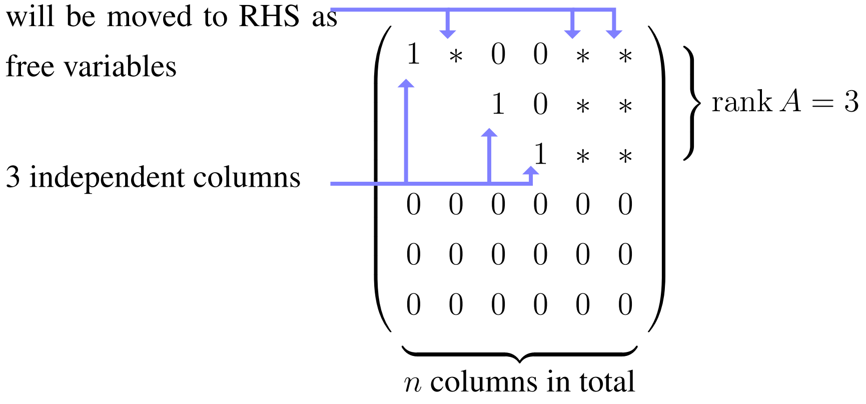
The rank reads

$$\operatorname{rank} A = 2.$$

☐

§ 10.2 Column space

The following fact is easier to see in a reduced row echelon form.



As well known, the * columns (i.e., the columns on plateaux) are to be moved to the RHS of equations to serve as free variables — each free variable (* column) corresponds to a free parameter to introduce — and leave only the *leading* 1 (pivot) columns at the LHS. By this doing, the entries left at the LHS form a strict triangle:

$$\operatorname{rank} A \left\{ \left(\begin{array}{c|c} \begin{matrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{matrix} & \begin{matrix} * & \cdots & * \\ * & \cdots & * \\ \vdots & & \vdots \\ * & \cdots & * \end{matrix} \end{array} \right) \right. \atop \qquad \underbrace{\hphantom{1 \quad 1 \quad \ddots \quad 1}}_{\text{independent}} \; \underbrace{\hphantom{* \cdots *}}_{\text{free variables}}$$

Hence, the number of the *independent columns* are equal to that of the *independent rows*,

$$\operatorname{rank} A = \text{Number of independent rows} = \text{Number of independent columns}. \tag{10.2.1}$$

Furthermore, denote

$$n = \text{total number of columns} \;, \qquad \dim N(A) = \text{Number of free variable columns}, \tag{10.2.2}$$

where $N(A)$ stands for the space of the free variables, called the *nullity space*（零化度空间）. Then we achieve an important relation among the dimensions:

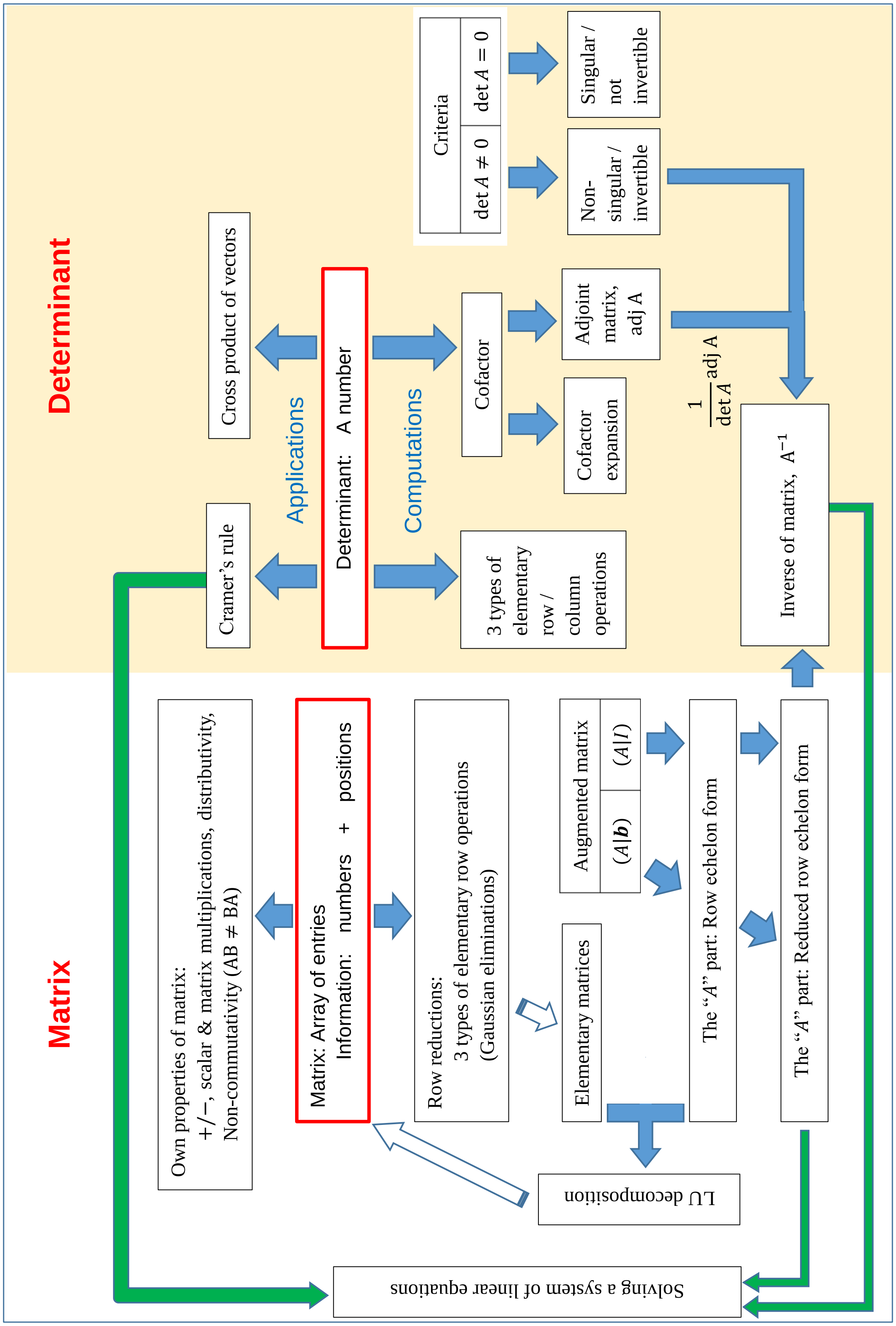
$$n = \operatorname{rank} A + \dim N(A). \tag{10.2.3}$$

This is called the rank-nullity theorem（秩-零化度定理）.

Obviously, for a square matrix $A \in \mathbb{R}^{n\times n}$, its rank computed via rows (row rank) equals to that computed via columns (column rank),

$$\operatorname{rank} A_{\text{row}} = \operatorname{rank} A_{\text{column}} \;. \tag{10.2.4}$$

Revision of Linear Algebra



General reference textbooks

[1] S. Lang: *Introduction to linear algebra*, 3th ed., Springer Science+Business Media Inc., 1987.

[2] D.C. Lay: *Linear algebra and its applications*, 4th ed., Pearson Education Inc., 2012.

[3] S. Leon: *Linear algebra with applications*, 8th ed., Pearson, Prentice-Hall, 2010.

[4] G. Strang: *Linear algebra and its applications*, 4th ed., Thomson Learning Inc. 2006.

[5] 北京大学数学系几何与代数教研室前代数小组编，王萼芳 石生明修订：《高等代数》（第3版），高等教育出版社，2003。

[6] 李尚志：《线性代数（数学专业用）》，高等教育出版社，2006。

[7] 李尚志：《线性代数（非数学专业用）》，高等教育出版社，2011。

Index