A TEXT IN MATHEMATICS
- LINEAR ALGEBRA -

K. ASAI

1. BASICS OF 3-DIMENSIONAL VECTORS

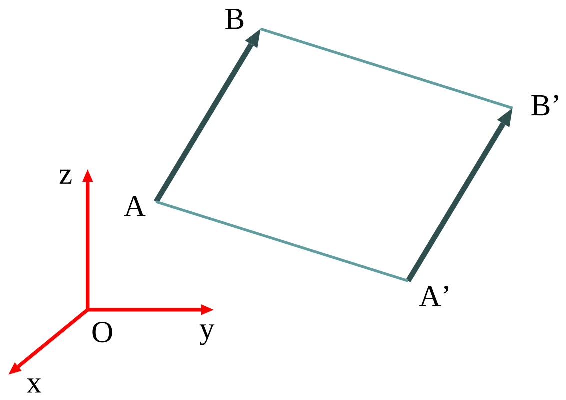
KEYWORDS: 3-DIMENSIONAL VECTORS, ADDITION, SCALAR MULTIPLICATION, MAGNITUDE, LENGTH, INNER PRODUCTS, VECTOR EQUATIONS OF LINES, DIRECTION VECTOR, EQUATIONS OF LINES, LINEAR COMBINATIONS, LINEAR INDEPENDENCE, LINEAR DEPENDENCE

1.1. **3-dimensional vectors.** A vector is a quantity that has both magnitude and direction. Two or three-dimensional vectors are elementary examples of vectors, which can be explained similarly, and so we study mainly 3-dimensional vectors in this chapter. Let A and B be two points in the 3-dimensional Euclidean space. An arrow from A to B is called a 3-dimensional vector or simply a vector, denoted by $\overrightarrow{AB}$. Here, A and B are called the initial and terminal points of this vector, respectively. The length of a line segment AB is called the magnitude or length of $\overrightarrow{AB}$. The direction from A to B is called the direction of $\overrightarrow{AB}$. As we have seen, the vector $\overrightarrow{AB}$ has magnitude and direction, but the information of the location of the vector is ignored. That is, any two vectors which are mapped to each other by translation are identified. If $\overrightarrow{AB}$ is translated to $\overrightarrow{A'B'}$ as in the figure, then we consider that $\overrightarrow{AB} = \overrightarrow{A'B'}$.

Letting $A = (x_0, y_0, z_0)$, $B = (x_1, y_1, z_1)$, by translation, we have $A' = (x_0 + a, y_0 + b, z_0 + c)$, $B' = (x_1 + a, y_1 + b, z_1 + c)$. If we calculate the differences between the coordinates of the initial and terminal points, and arrange them vertically, we have

$$\begin{pmatrix} x_1 - x_0 \\ y_1 - y_0 \\ z_1 - z_0 \end{pmatrix} = \begin{pmatrix} (x_1 + a) - (x_0 + a) \\ (y_1 + b) - (y_0 + b) \\ (z_1 + c) - (z_0 + c) \end{pmatrix}. \tag{1}$$

Conversely, if the sets of the differences between the coordinates of the initial and terminal points are the same, then we have



$$\begin{pmatrix} x_1 - x_0 \\ y_1 - y_0 \\ z_1 - z_0 \end{pmatrix} = \begin{pmatrix} (x_1 + a') - (x_0 + a) \\ (y_1 + b') - (y_0 + b) \\ (z_1 + c') - (z_0 + c) \end{pmatrix}. \tag{2}$$

Hence $a = a'$, $b = b'$, $c = c'$, and so two vectors are mapped to each other by translation, and therefore they are identified. In summary,

$$\begin{array}{ll} & \text{Two vectors are the same.} \\ \Longleftrightarrow & \text{Two vectors are mapped to each other by translation.} \\ \Longleftrightarrow & \begin{array}{l}\text{The sets of the differences between the coordinates} \\ \text{of the initial and terminal points are the same.}\end{array} \end{array} \tag{3}$$

Accordingly, we may use the differences between the coordinates of the initial and terminal points to represent a vector, say,

$$\overrightarrow{AB} = \begin{pmatrix} x_1 - x_0 \\ y_1 - y_0 \\ z_1 - z_0 \end{pmatrix}. \tag{4}$$

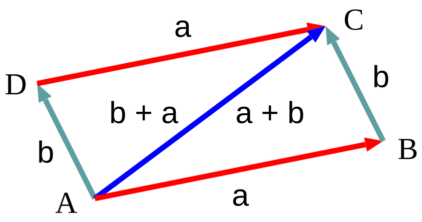
This is called the component form of a vector.

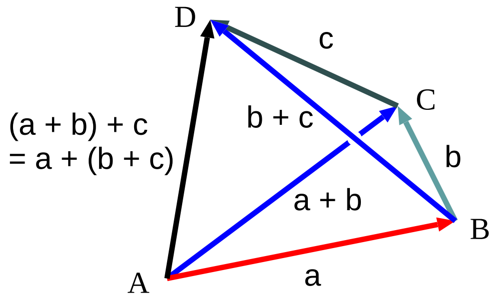
1.2. **Vector addition.** A 3-dimensional vector can take any point as the initial point. To be exact, for any 3-dimensional vector $\mathtt{a}$ and for any point A, there exists a unique point B such that $\mathtt{a} = \overrightarrow{AB}$. Then letting $\mathtt{a} = \overrightarrow{AB}$ and $\mathtt{b} = \overrightarrow{BC}$, we define the sum of $\mathtt{a}$ and $\mathtt{b}$ by

$$\mathtt{a} + \mathtt{b} = \overrightarrow{AB} + \overrightarrow{BC} = \overrightarrow{AC}\ . \tag{5}$$

That is to say, $\mathtt{a} + \mathtt{b}$ is obtained by connecting $\mathtt{a}$ and $\mathtt{b}$ by moving the initial point of $\mathtt{b}$ to the terminal point of $\mathtt{a}$. The operation $+$ is called (vector) addition. Using the following figure, we have

$$\begin{array}{cc} \mathtt{a} + \mathtt{b} = \mathtt{b} + \mathtt{a} & \text{(commutative law of addition)} \\ (\mathtt{a} + \mathtt{b}) + \mathtt{c} = \mathtt{a} + (\mathtt{b} + \mathtt{c}) & \text{(associative law of addition)} \end{array} \tag{6}$$





The associative law is proved also by calculation easily as follows:

$$(\overrightarrow{AB} + \overrightarrow{BC}) + \overrightarrow{CD} = \overrightarrow{AC} + \overrightarrow{CD} = \overrightarrow{AD} = \overrightarrow{AB} + \overrightarrow{BD} = \overrightarrow{AB} + (\overrightarrow{BC} + \overrightarrow{CD}). \quad (7)$$

By the associative law, any expression such as $\mathtt{a}_1 + \mathtt{a}_2 + \cdots + \mathtt{a}_n$ is independent of the way of adding parentheses, and therefore parentheses are often omitted.

Let $A = (x_0, y_0, z_0)$, $B = (x_1, y_1, z_1)$ and $C = (x_2, y_2, z_2)$. The identity $\overrightarrow{AB} + \overrightarrow{BC} = \overrightarrow{AC}$ is rewritten by the component form:

$$\begin{pmatrix} x_1 - x_0 \\ y_1 - y_0 \\ z_1 - z_0 \end{pmatrix} + \begin{pmatrix} x_2 - x_1 \\ y_2 - y_1 \\ z_2 - z_1 \end{pmatrix} = \begin{pmatrix} x_2 - x_0 \\ y_2 - y_0 \\ z_2 - z_0 \end{pmatrix}. \quad (8)$$

Hence we see that every component of $\overrightarrow{AC}$ is the sum of the corresponding components of $\overrightarrow{AB}$ and $\overrightarrow{BC}$. Therefore

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \\ a_3 + b_3 \end{pmatrix}. \quad (9)$$

There is a unique special 3-dimensional vector whose magnitude is 0 with no direction, denoted by $\mathtt{0}$ and called the zero vector. This is a 3-dimensional vector $\overrightarrow{AA} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ for an arbitrary point A.

For a 3-dimensional vector $\overrightarrow{AB}$, a 3-dimensional vector $\overrightarrow{BA}$ is called the inverse vector of $\overrightarrow{AB}$, denoted by

$$\overrightarrow{BA} = -\overrightarrow{AB}. \quad (10)$$

$-\overrightarrow{AB}$ has the same magnitude as $\overrightarrow{AB}$, and has the direction opposite to $\overrightarrow{AB}$. Obviously, we have the following.

$$\begin{aligned} \mathtt{a} + \mathtt{0} &= \mathtt{0} + \mathtt{a} = \mathtt{a} \\ \mathtt{a} + (-\mathtt{a}) &= (-\mathtt{a}) + \mathtt{a} = \mathtt{0} \end{aligned} \quad (11)$$

(exercise01) (1) Calculate $\overrightarrow{AB} + \overrightarrow{CD} + \overrightarrow{BC} + \overrightarrow{DE}$. (2) Prove that

$$\overrightarrow{A_1A_2} + \overrightarrow{A_2A_3} + \cdots + \overrightarrow{A_nA_{n+1}} = \overrightarrow{A_1A_{n+1}}. \quad (12)$$

(3) Calculate $\overrightarrow{A_1A_2} + \overrightarrow{A_2A_3} + \cdots + \overrightarrow{A_nA_1}$.

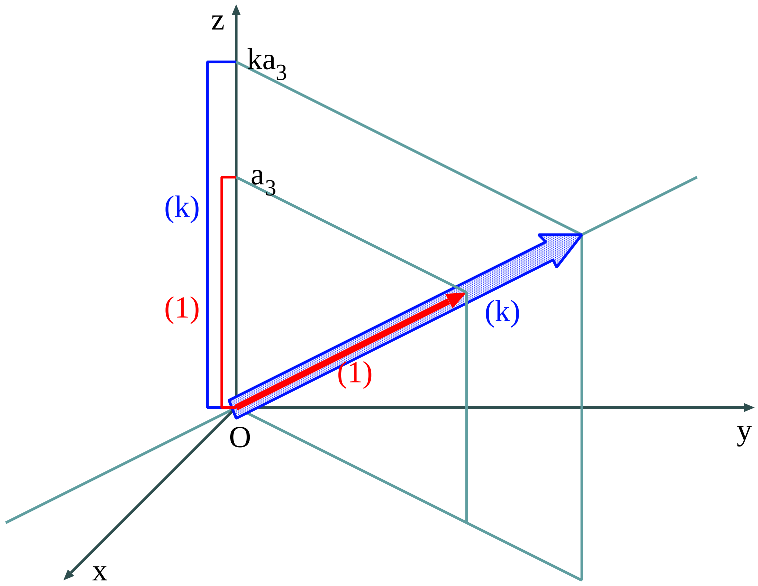
1.3. **Scalar multiplication.** Let $\mathtt{a}$ be a 3-dimensional vector and k be a positive real number. The scalar multiplication $k\mathtt{a}$ of $\mathtt{a}$ by k or the scalar multiplication $(-k)\mathtt{a}$ of $\mathtt{a}$ by $-k$ is defined by

$$\begin{aligned} k\mathtt{a} &= (\text{A vector with the same direction as } \mathtt{a}\text{, and of magnitude} \\ &\qquad k \text{ times the magnitude of } \mathtt{a}.) \\ (-k)\mathtt{a} &= (\text{A vector with the direction opposite to } \mathtt{a}\text{, and of magnitude} \\ &\qquad k \text{ times the magnitude of } \mathtt{a}.) \end{aligned} \tag{13}$$

In addition, define $0\mathtt{a} = \mathtt{0}$, $\pm k\mathtt{0} = \mathtt{0}$. For scalar multiplication, we have

$$k\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} ka_1 \\ ka_2 \\ ka_3 \end{pmatrix} \tag{14}$$

for every real number k. This is confirmed by the following figure, if $k > 0$ and for the z-axis, and the other axes similarly. If $k < 0$, it suffices to note that the direction changes to the opposite.



For addition and scalar multiplication, we have the following. Letting k, l be real numbers and $\mathtt{a}, \mathtt{b}$ be 3-dimensional vectors,

$$\begin{aligned} k(\mathtt{a}+\mathtt{b}) &= k\mathtt{a} + k\mathtt{b} \\ (k+l)\mathtt{a} &= k\mathtt{a} + l\mathtt{a} \\ (kl)\mathtt{a} &= k(l\mathtt{a}) \end{aligned} \tag{15}$$

(exercise02) (1) Show the above formulas by the use of the component form. (2) Show that $(-1)\mathtt{a} = -\mathtt{a}$. (3) Show the following.

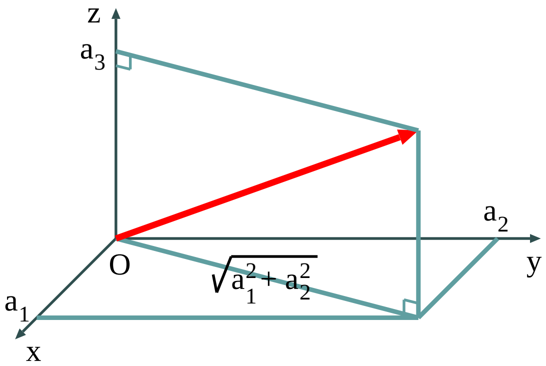
$$\begin{aligned} k(\mathtt{a}_1 + \mathtt{a}_2 + \cdots + \mathtt{a}_n) &= k\mathtt{a}_1 + k\mathtt{a}_2 + \cdots + k\mathtt{a}_n \\ (k_1 + k_2 + \cdots + k_n)\mathtt{a} &= k_1\mathtt{a} + k_2\mathtt{a} + \cdots + k_n\mathtt{a} \end{aligned} \tag{16}$$

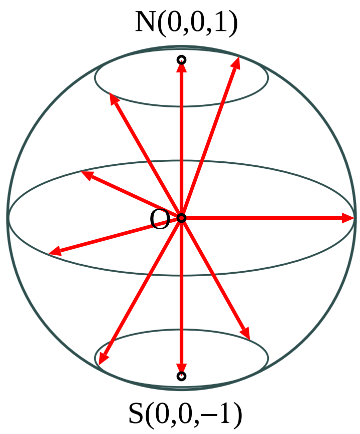
(note) We often write simply that $\mathtt{a} + (-\mathtt{b}) = \mathtt{a} - \mathtt{b}$.

1.4. **The magnitude of 3-dimensional vectors.** The magnitude of a vector $\mathtt{a}$ is denoted by $||\mathtt{a}||$. Letting $\mathtt{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$, by Pythagoras' theorem,

$$||\mathtt{a}|| = \sqrt{a_1^2 + a_2^2 + a_3^2}. \tag{17}$$

A vector of magnitude 1 is called a unit vector. For $\mathtt{a} \neq \mathtt{0}$, $\frac{1}{||\mathtt{a}||}\mathtt{a}$ is a unit vector with the same direction as $\mathtt{a}$.



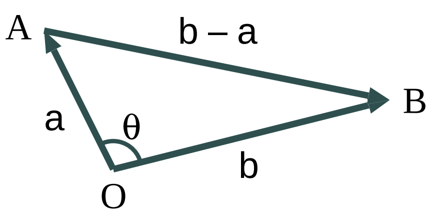


unit vectors

1.5. **Inner product.** For two 3-dimensional vectors $\mathtt{a}, \mathtt{b}$, let the angle of them be θ. We define the inner product $(\mathtt{a}, \mathtt{b})$ of $\mathtt{a}$ and $\mathtt{b}$ by

$$(\mathtt{a}, \mathtt{b}) = ||\mathtt{a}|| \cdot ||\mathtt{b}|| \cdot \cos\theta. \tag{18}$$

Let $\overrightarrow{OA} = \mathtt{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$, $\overrightarrow{OB} = \mathtt{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$, and apply the law of cosines to $\triangle OAB$.



$$\begin{aligned} ||\mathtt{b} - \mathtt{a}||^2 &= ||\mathtt{a}||^2 + ||\mathtt{b}||^2 - 2||\mathtt{a}|| \cdot ||\mathtt{b}|| \cdot \cos\theta. \\ \therefore \quad (\mathtt{a}, \mathtt{b}) &= ||\mathtt{a}|| \cdot ||\mathtt{b}|| \cdot \cos\theta \\ &= \frac{1}{2}\left(||\mathtt{a}||^2 + ||\mathtt{b}||^2 - ||\mathtt{b} - \mathtt{a}||^2\right) \\ &= \frac{1}{2}\left[(a_1^2 + a_2^2 + a_3^2) + (b_1^2 + b_2^2 + b_3^2) - (b_1 - a_1)^2 - (b_2 - a_2)^2 - (b_3 - a_3)^2\right] \\ &= a_1 b_1 + a_2 b_2 + a_3 b_3. \end{aligned} \tag{19}$$

If the angle of 3-dimensional vectors $\mathtt{a}$ and $\mathtt{b}$ is $\pi/2$, then they are called orthogonal or perpendicular, and written as $\mathtt{a} \perp \mathtt{b}$. If the angle of $\mathtt{a}$ and $\mathtt{b}$ is 0 or π, that is, these vectors have the same or opposite direction, then they are called parallel, and written as $\mathtt{a} \parallel \mathtt{b}$. For any vector $\mathtt{a}$, we promise that $\mathtt{a} \perp \mathtt{0}$ and $\mathtt{a} \parallel \mathtt{0}$. From the definition of inner product, we have the following.

$$\begin{array}{lcl} \mathtt{a} \perp \mathtt{b} & \Longleftrightarrow & (\mathtt{a}, \mathtt{b}) = 0 \\ \mathtt{a} \parallel \mathtt{b} & \Longleftrightarrow & (\mathtt{a}, \mathtt{b}) = \pm\, ||\mathtt{a}|| \cdot ||\mathtt{b}|| \end{array} \tag{20}$$

Let k be a real number. Inner product satisfies the following properties.

$$\begin{array}{c} (\mathtt{a}, \mathtt{b}) = (\mathtt{b}, \mathtt{a}) \\ (\mathtt{a}, \mathtt{b} + \mathtt{c}) = (\mathtt{a}, \mathtt{b}) + (\mathtt{a}, \mathtt{c}), \qquad (\mathtt{a} + \mathtt{b}, \mathtt{c}) = (\mathtt{a}, \mathtt{c}) + (\mathtt{b}, \mathtt{c}) \\ (\mathtt{a}, k\mathtt{b}) = (k\mathtt{a}, \mathtt{b}) = k(\mathtt{a}, \mathtt{b}) \\ (\mathtt{a}, \mathtt{a}) \geq 0 \quad \text{and} \quad (\mathtt{a}, \mathtt{a}) = 0 \Longleftrightarrow \mathtt{a} = \mathtt{0} \end{array} \tag{21}$$

(exercise03) (1) Prove (21). (2) Show the following.

$$\begin{array}{l} (\mathtt{a}, \mathtt{b}_1 + \mathtt{b}_2 + \cdots + \mathtt{b}_n) = (\mathtt{a}, \mathtt{b}_1) + (\mathtt{a}, \mathtt{b}_2) + \cdots + (\mathtt{a}, \mathtt{b}_n) \\ (\mathtt{a}_1 + \mathtt{a}_2 + \cdots + \mathtt{a}_n, \mathtt{b}) = (\mathtt{a}_1, \mathtt{b}) + (\mathtt{a}_2, \mathtt{b}) + \cdots + (\mathtt{a}_n, \mathtt{b}) \end{array} \tag{22}$$

(exercise04) Show the following properties concerning the magnitude and the inner product of vectors.

$$\begin{array}{ll} ||\mathtt{a}|| = \sqrt{(\mathtt{a}, \mathtt{a})} & \\ ||k\mathtt{a}|| = |k| \cdot ||\mathtt{a}|| & \\ |(\mathtt{a}, \mathtt{b})| \leq ||\mathtt{a}|| \cdot ||\mathtt{b}|| & \text{(Cauchy—Schwarz inequality)} \\ ||\mathtt{a} + \mathtt{b}|| \leq ||\mathtt{a}|| + ||\mathtt{b}|| & \text{(triangle inequality)} \end{array} \tag{23}$$

(exercise05) Let A, B, C, D be 4 points in the 3-dimensional space. Show the following.

$$AB^2 + CD^2 = AD^2 + BC^2 \iff \overrightarrow{AC} \perp \overrightarrow{BD} \tag{24}$$

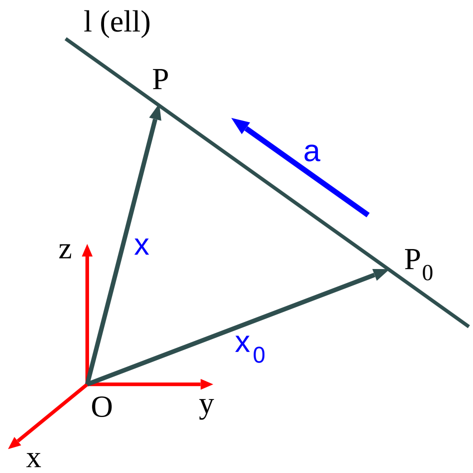
1.6. **Vector equations of lines in the 3-dimensional space.** For a point P in the 3-dimensional space, a vector $\overrightarrow{OP}$ is called the position vector of P. Let l be a line in the 3-dimensional space. An equation which express the position vector $\overrightarrow{OP} = \mathtt{x}$ for every point $P(x, y, z)$ on l is called the vector equation of the line l. To construct this equation, we need the direction vector $\mathtt{a}$ of l, which is a vector parallel to l, and the position vector of a point $P_0(x_0, y_0, z_0)$ on l. Then we have $\overrightarrow{P_0P} = t\,\mathtt{a}$ using real number t, and thus

$$\begin{array}{rll} \overrightarrow{OP} & = \overrightarrow{OP_0} + t\,\mathtt{a} & \text{or} \\ \mathtt{x} & = \mathtt{x}_0 + t\,\mathtt{a}. & \end{array} \tag{25}$$

This equation is called the vector equation of l, which is often given by the component form:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} + t \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}. \tag{26}$$

(exercise06) Let l be a line passing $(1, -2, 3)$, and parallel to $\begin{pmatrix} -1 \\ 3 \\ 2 \end{pmatrix}$. Find the vector equation of l.



1.7. **Equations of lines in the 3-dimensional space.** Let a_1, a_2, a_3 be nonzero real numbers. By the vector equation (26) of l, we have

$$\begin{cases} x = x_0 + ta_1 \\ y = y_0 + ta_2 \\ z = z_0 + ta_3 \end{cases} ; \qquad \therefore \quad \begin{cases} t = \frac{x-x_0}{a_1} \\ t = \frac{y-y_0}{a_2} \\ t = \frac{z-z_0}{a_3} \end{cases} \tag{27}$$

From this, it follows that

$$\frac{x - x_0}{a_1} = \frac{y - y_0}{a_2} = \frac{z - z_0}{a_3}. \tag{28}$$

Conversely, putting each side of (28) equal to t, we go back

$$(28) \Rightarrow (27) \Rightarrow (26)$$

and get the vector equation of l. Therefore we see that l is represented by (28), which is called the equation of l.

(note) If some of a_1, a_2, a_3 are equal to 0, since it is impossible to divide by zero, another equation is derived. For example, the cases $a_2 = 0$ and $a_1 = a_3 = 0$ implies the following equations, respectively.

$$\begin{cases} \dfrac{x - x_0}{a_1} = \dfrac{z - z_0}{a_3} \\ y = y_0 \end{cases} ; \qquad \begin{cases} x = x_0 \\ z = z_0 \end{cases} \tag{29}$$

(exercise07) Transform the following vector equation and (ordinary) equation of a line into an (ordinary) equation and a vector equation, respectively.

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix} + t \begin{pmatrix} -1 \\ 3 \\ 2 \end{pmatrix} \tag{30}$$

$$\frac{x-1}{-1} = \frac{y+2}{3} = \frac{z-3}{2} \tag{31}$$

1.8. **Linear combination.** We often express a 3-dimensional vector in terms of other 3-dimensional vectors. This expression may contain only two operations: addition and scalar multiplication as follows.

$$k_1\mathtt{a}_1 + k_2\mathtt{a}_2 + \cdots + k_n\mathtt{a}_n \tag{32}$$

This is called a linear combination of vectors $\mathtt{a}_1, \mathtt{a}_2, \ldots, \mathtt{a}_n$.

Three vectors $\mathtt{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\mathtt{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, $\mathtt{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ are called elementary vectors.

Every 3-dimensional vector is expressed as a linear combination of elementary vectors as follows.

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = x\mathtt{e}_1 + y\mathtt{e}_2 + z\mathtt{e}_3. \tag{33}$$

There exists other set of vectors which can express every vector. In general, any 3 vectors which can not be contained simultaneously in any plane, are called linearly independent. Given linearly independent 3 vectors, every 3-dimensional vector is expressed as a linear combination of them. If vectors are not linearly independent, then they are called linearly dependent. That is to say, 3 vectors are linearly dependent if they can be contained in some plane. In other words, they are linearly dependent if some vector of them is expressed as a linear combination of the rest vectors. Linearly dependent 3 vectors can not express all vectors.

(exercise08) Let $\mathtt{a} = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}$, $\mathtt{b} = \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}$, $\mathtt{c} = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$. Express $\mathtt{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ as a linear combination of $\mathtt{a}, \mathtt{b}, \mathtt{c}$.

(answer) $\mathtt{x} = \frac{4x+y-2z}{9}\mathtt{a} + \frac{-2x+4y+z}{9}\mathtt{b} + \frac{x-2y+4z}{9}\mathtt{c}$.

★

2. PLANES IN THE 3-DIMENSIONAL SPACE AND ITS EQUATIONS

⋆ 3 ⋆

KEYWORDS: VECTOR EQUATIONS OF PLANES, NORMAL VECTORS, EQUATIONS OF PLANES

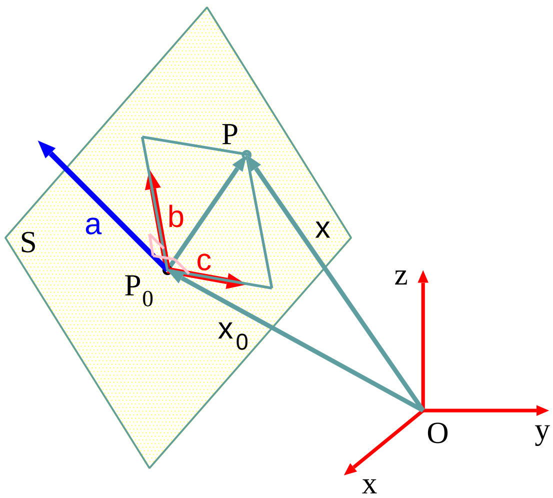


Figure 1

2.1. **Vector equations of planes in the 3-dimensional space.** Let S be a plane in the 3-dimensional space. Let $P_0(x_0, y_0, z_0)$ be a point on S, and its position vector be $\overrightarrow{OP_0} = \mathtt{x}_0$. Take two non-parallel vectors $\mathtt{b}$ and $\mathtt{c}$ contained in S. Let $P(x, y, z)$ be an arbitrary point on S and $\overrightarrow{OP} = \mathtt{x}$ be its position vector. Since $\overrightarrow{P_0P} = t\,\mathtt{b} + u\,\mathtt{c}$ $(t, u \in \mathbb{R})$, we have

$$\begin{aligned} \overrightarrow{OP} &= \overrightarrow{OP_0} + t\,\mathtt{b} + u\,\mathtt{c} \qquad \text{or} \\ \mathtt{x} &= \mathtt{x}_0 + t\,\mathtt{b} + u\,\mathtt{c}. \end{aligned} \tag{1}$$

Also, every point P expressed as above is always on S. Hence (1) is called the vector equation of S. (See Figure 1) In concrete problems, (1) is written using component form as follows.

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} + t \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} + u \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} \tag{2}$$

(exercise01) Let S be a plane which contains a point $(3, 2, -1)$ and suppose two vectors $\begin{pmatrix} 2 \\ -2 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$ are contained in S. Determine the vector equation of S.

2.2. **Equations of planes in the 3-dimensional space.** Let us find the ordinary equation of a plane S. A vector which is perpendicular to a plane S is called the normal vector to S. Let a be the normal vector to S. Similarly to Section 1, let $P_0(x_0, y_0, z_0)$ be a point on S, and $P(x, y, z)$ be an arbitrary point on S. Since a is the normal vector to S, we have

$$(\mathsf{a}, \overrightarrow{P_0P}) = 0. \tag{3}$$

Conversely, any point P satisfying (3) is always on S. Then letting $\mathsf{a} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$ and rewrite (3) as

$$(\mathsf{a}, \overrightarrow{P_0P}) = (\begin{pmatrix} a \\ b \\ c \end{pmatrix}, \begin{pmatrix} x - x_0 \\ y - y_0 \\ z - z_0 \end{pmatrix}) = 0. \tag{4}$$

$$\therefore \quad \boxed{a(x - x_0) + b(y - y_0) + c(z - z_0) = 0}. \tag{5}$$

This is the equation of a plane S.

Theorem 1. *The equation of a plane in the 3-dimensional space with the normal vector* $\begin{pmatrix} a \\ b \\ c \end{pmatrix}$ *passing through a point* (x_0, y_0, z_0) *is* (5).

Next, given the vector equation of a plane S, we modify this equation to be the ordinary equation of S. To do this, it suffices to determine the normal vector to S. For this purpose, find a vector a which is perpendicular to both of two vectors b and c, by the vector product $\mathsf{b} \times \mathsf{c}$, etc. Then a is perpendicular to arbitrary $\overrightarrow{P_0P} = t\,\mathsf{b} + u\,\mathsf{c}$. Indeed,

$$(\mathsf{a}, \overrightarrow{P_0P}) = (\mathsf{a}, t\,\mathsf{b} + u\,\mathsf{c}) = t(\mathsf{a}, \mathsf{b}) + u(\mathsf{a}, \mathsf{c}) = 0. \tag{6}$$

Hence a is the normal vector to S. The rest to do is to find a point P_0 on S to give the equation (5) of S.

The equation (5) of a plane is often written in the form:

$$\boxed{ax + by + cz = d.} \tag{7}$$

Indeed, (5) is easily modified to (7). Conversely, for given (7), taking some $(x, y, z) = (x_0, y_0, z_0)$ which satisfies (7), we have

$$ax_0 + by_0 + cz_0 = d \tag{8}$$

and (7)−(8) gives (5).
(exercise02) Let S be a plane expressed by the vector equation:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2 \\ -3 \\ 1 \end{pmatrix} + t \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} + u \begin{pmatrix} 3 \\ -1 \\ 0 \end{pmatrix}. \tag{9}$$

Determine the equation of S. (ans) $x + 3y - 2z = -9$.

2.3. **Converting the equation of a plane to the vector equation.** Given the equation of a plane, let us find the vector equation of it. Let S be a plane and $ax+by+cz=d$ be the equation of S. First we find some point $P_0(x_0, y_0, z_0)$ on S, and let its position vector be $\mathtt{x}_0$. The normal vector to S is $\mathtt{a} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$, and note that the vectors which are perpendicular to it should be contained in S by translation. Then taking non-parallel two vectors $\mathtt{b}$ and $\mathtt{c}$ which are perpendicular to the normal vector $\mathtt{a}$, and we have the vector equation of S:

$$\mathtt{x} = \mathtt{x}_0 + t\,\mathtt{b} + u\,\mathtt{c}. \tag{10}$$

The key is to find non-parallel two vectors perpendicular to the normal vector.

(exercise03) Determine the vector equation of a plane expressed by $x + 3y - 2z = -9$.

(note) The equation of a plane in the form $ax+by+cz=d$ is uniquely determined up to constant multiples. However, the vector equation of a plane is not uniquely determined because there are infinitely many choice of non-parallel two vectors contained in the plane.

⋆

3. LINEAR INDEPENDENCE OF 3-DIMENSIONAL VECTORS AND THE VOLUMES OF PARALLELEPIPEDS

⋆ 4 ⋆

KEYWORDS: LINEAR INDEPENDENCE, LINEAR DEPENDENCE, PARALLELEPIPED, VECTOR PRODUCTS, 3×3 MATRIX DETERMINANTS

3.1. **Linear independence.** Suppose three 3-dimensional vectors $\mathtt{a}, \mathtt{b}, \mathtt{c}$ have the same initial point. If $\mathtt{a}, \mathtt{b}, \mathtt{c}$ are not contained in any plane, then they are called linearly independent. If they are contained in some plane, then then they are called linearly dependent. A necessary and sufficient condition for three vectors to be linearly dependent is that some vector of them is expressed as a linear combination of the rest two vectors. Indeed, in that case, clearly they are contained in some plane, and conversely, if they are contained in some plane, it is clear that some vector of them is expressed as a linear combination of the rest.

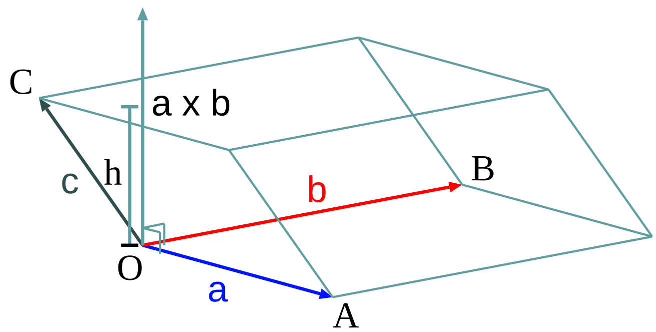


Figure 2

3.2. **Parallelepipeds.** Although we know about all the above mentioned, it is sometimes difficult to determine whether three vectors are linearly independent or not. In such a case, it is useful to consider the volume of a parallelepiped. Suppose there exist three vectors $\mathtt{a}, \mathtt{b}, \mathtt{c}$ as in the above figure. A hexahedron with these vectors as edges, such that every two faces facing each other are parallel, is called the parallelepiped spanned by $\mathtt{a}, \mathtt{b}$ and $\mathtt{c}$. If those three vectors are linearly independent, then the polyhedron never collapses and it has nonzero volume, and if those vectors are linearly dependent, then the polyhedron collapses and it has no volume. Therefore we can determine linear independence by the volume of parallelepipeds.

So let us study the volume V of the parallelepiped spanned by $\mathtt{a}, \mathtt{b}$ and $\mathtt{c}$. For this purpose, we introduce the vector products of two 3-dimensional vectors.

3.3. **Vector products.** Suppose that two 3-dimensional vectors $\mathtt{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$ and $\mathtt{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$ have the same initial point. Let S be the area of the parallelogram spanned by $\mathtt{a}$ and $\mathtt{b}$ (with $\mathtt{a}$ and $\mathtt{b}$ as edges). Let θ be the angle between $\mathtt{a}$ and $\mathtt{b}$. A vector of magnitude S, which is perpendicular to the parallelogram, directed by the right-hand rule,[1] is called the vector product (cross product) of $\mathtt{a}$ and $\mathtt{b}$ denoted by $\mathtt{a} \times \mathtt{b}$. Here we have

$$\begin{aligned} S^2 &= ||\mathtt{a}||^2 \cdot ||\mathtt{b}||^2 \cdot \sin^2\theta = ||\mathtt{a}||^2 \cdot ||\mathtt{b}||^2 \cdot (1-\cos^2\theta) = ||\mathtt{a}||^2 \cdot ||\mathtt{b}||^2 - (\mathtt{a},\mathtt{b})^2 \\ &= (a_1^2+a_2^2+a_3^2)(b_1^2+b_2^2+b_3^2) - (a_1b_1+a_2b_2+a_3b_3)^2 \\ &= \cdots = (a_2b_3-a_3b_2)^2 + (-a_1b_3+a_3b_1)^2 + (a_1b_2-a_2b_1)^2. \end{aligned} \tag{1}$$

In this way, we have calculated the area S. Then we can define the vector product $\mathtt{a} \times \mathtt{b}$ as follows.

$$\mathtt{a} \times \mathtt{b} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \times \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} a_2b_3 - a_3b_2 \\ -a_1b_3 + a_3b_1 \\ a_1b_2 - a_2b_1 \end{pmatrix} \tag{2}$$

Indeed, it is clear that $||\mathtt{a} \times \mathtt{b}|| = S$. Also we see that $(\mathtt{a} \times \mathtt{b}, \mathtt{a}) = (\mathtt{a} \times \mathtt{b}, \mathtt{b}) = 0$, and therefore $\mathtt{a} \times \mathtt{b}$ is perpendicular to the parallelogram spanned by $\mathtt{a}$ and $\mathtt{b}$. For confirmation,

$$(\mathtt{a} \times \mathtt{b}, \mathtt{a}) = (a_2b_3 - a_3b_2)a_1 + (-a_1b_3 + a_3b_1)a_2 + (a_1b_2 - a_2b_1)a_3 = 0. \tag{3}$$

Similarly, we have $(\mathtt{a} \times \mathtt{b}, \mathtt{b}) = 0$. In addition, we can confirm that the direction of $\mathtt{a} \times \mathtt{b}$ satisfies the right-hand rule. In the case of $\mathtt{e}_1 \times \mathtt{e}_2 = \mathtt{e}_3$, it is clear. In general, for given $\mathtt{a} \times \mathtt{b}$, we transform $\mathtt{e}_1$ and $\mathtt{e}_2$ into $\mathtt{a}$ and $\mathtt{b}$, respectively, continuously, keeping $\mathtt{a}$ and $\mathtt{b}$ are not parallel along the way. Then $\mathtt{e}_1 \times \mathtt{e}_2$ is transformed keeping perpendicular to the parallelogram spanned by $\mathtt{e}_1$ and $\mathtt{e}_2$ and never becomes $\mathtt{0}$ along the way. Consequently, $\mathtt{a} \times \mathtt{b}$ still satisfies the right-hand rule. Therefore the right-hand side of (2) is certainly the vector product of $\mathtt{a}$ and $\mathtt{b}$.

One way to remember the formula (2) is that put $\mathtt{a}$ and $\mathtt{b}$ side by side to make $\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \\ a_3 & b_3 \end{pmatrix}$. The i-th component of the vector product $\mathtt{a} \times \mathtt{b}$ is made by hiding the i-th row and calculate the rest determinant of 2×2 matrix. Here, note that, for the second component, we should inverse the signature.

(exercise01) Confirm that $(\mathtt{a} \times \mathtt{b}, \mathtt{b}) = 0$.

(exercise02) For several i, j, calculate $\mathtt{e}_i \times \mathtt{e}_j$. Also, calculate $\begin{pmatrix} 2 \\ -1 \\ 3 \end{pmatrix} \times \begin{pmatrix} 2 \\ 2 \\ 5 \end{pmatrix}$.

[1] A right screw is put on the initial point of $\mathtt{a}$ and $\mathtt{b}$, so that the screw is perpendicular to both $\mathtt{a}$ and $\mathtt{b}$, then the direction of $\mathtt{a} \times \mathtt{b}$ is the one in which the screw advances if it is rotated from $\mathtt{a}$ to $\mathtt{b}$ by the angle θ.

It is confirmed by (2) that the following laws of vector product hold. (1: is clear by the right-hand rule.)

$$\begin{array}{ll} 1: & \mathtt{a} \times \mathtt{b} = -(\mathtt{b} \times \mathtt{a}) \\ 2: & \mathtt{a} \times (\mathtt{b} + \mathtt{c}) = \mathtt{a} \times \mathtt{b} + \mathtt{a} \times \mathtt{c} \\ 3: & (\mathtt{a} + \mathtt{b}) \times \mathtt{c} = \mathtt{a} \times \mathtt{c} + \mathtt{b} \times \mathtt{c} \\ 4: & (k\mathtt{a}) \times \mathtt{b} = \mathtt{a} \times (k\mathtt{b}) = k(\mathtt{a} \times \mathtt{b}) \end{array} \tag{4}$$

(exercise03) Show these laws.
Proof of 2:.

$$\begin{aligned} \mathtt{a} \times (\mathtt{b} + \mathtt{c}) &= \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \times \begin{pmatrix} b_1 + c_1 \\ b_2 + c_2 \\ b_3 + c_3 \end{pmatrix} = \begin{pmatrix} a_2(b_3 + c_3) - a_3(b_2 + c_2) \\ -a_1(b_3 + c_3) + a_3(b_1 + c_1) \\ a_1(b_2 + c_2) - a_2(b_1 + c_1) \end{pmatrix} \\ &= \begin{pmatrix} a_2b_3 - a_3b_2 \\ -a_1b_3 + a_3b_1 \\ a_1b_2 - a_2b_1 \end{pmatrix} + \begin{pmatrix} a_2c_3 - a_3c_2 \\ -a_1c_3 + a_3c_1 \\ a_1c_2 - a_2c_1 \end{pmatrix} = \mathtt{a} \times \mathtt{b} + \mathtt{a} \times \mathtt{c}. \end{aligned} \tag{5}$$

3.4. **Volumes of parallelepipeds and 3×3 matrix determinants.** Using vector products, the volume V of a parallelepiped is calculated. For the parallelepiped in Figure 2, the bottom area S is calculated as $S = ||\mathtt{a} \times \mathtt{b}||$. Here, let h be the height of the parallelepiped, then $V = S \cdot h$. However, since the vector product $\mathtt{a} \times \mathtt{b}$ is perpendicular to the bottom, letting φ be the angle between $\mathtt{a} \times \mathtt{b}$ and $\mathtt{c}$, we have

$$V = S \cdot ||\mathtt{c}|| \cos\varphi = ||\mathtt{a} \times \mathtt{b}|| \cdot ||\mathtt{c}|| \cos\varphi = (\mathtt{a} \times \mathtt{b}, \mathtt{c}). \tag{6}$$

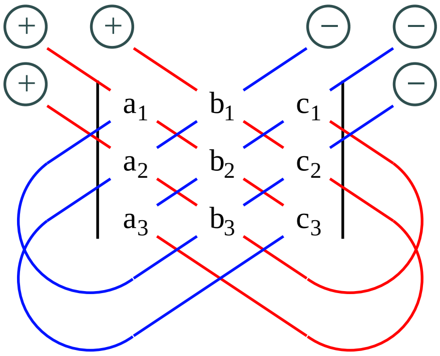
In this way, V is expressed by the vector product and inner product. The right-hand side of (6) has, exactly speaking, negative values ($(-1)\times$ volume) when $\pi/2 < \varphi \leq \pi$, thus it represents the signed volume of the parallelepiped. In addition, (6) is the definition of 3×3 (matrix) determinants, say,

$$\left| \begin{array}{ccc} \mathtt{a} & \mathtt{b} & \mathtt{c} \end{array} \right| = \left| \begin{array}{ccc} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{array} \right| = (\mathtt{a} \times \mathtt{b}, \mathtt{c}) = \begin{array}{l} a_1b_2c_3 + a_2b_3c_1 + a_3b_1c_2 \\ -a_1b_3c_2 - a_2b_1c_3 - a_3b_2c_1 \end{array}. \tag{7}$$

Accordingly, we have the following.

Theorem 1. *A necessary and sufficient condition for 3 vectors* $\mathtt{a}, \mathtt{b}, \mathtt{c}$ *to be linearly independent is that* $\left| \begin{array}{ccc} \mathtt{a} & \mathtt{b} & \mathtt{c} \end{array} \right| \neq 0$.

(note) The expansion formula which expands a 3×3 determinant by the rightmost-hand side of (7) is called Sarrus' rule. The following illustration helps you to remember it.



(exercise04) Let $\mathtt{a} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$, $\mathtt{b} = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}$ and $\mathtt{c} = \begin{pmatrix} 7 \\ 8 \\ t \end{pmatrix}$. Represent the necessary and sufficient condition for these vectors to be linearly independent by an expression with respect to t.

(ans) $\begin{vmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & t \end{vmatrix} = 1\cdot 5\cdot t + 2\cdot 6\cdot 7 + 3\cdot 4\cdot 8 - 1\cdot 6\cdot 8 - 2\cdot 4\cdot t - 3\cdot 5\cdot 7 = -3t + 27 \neq 0.$

$\therefore$ $\boxed{t \neq 9}$.

★

4. 3×3 MATRICES AND LINEAR TRANSFORMATIONS OF V^3

⋆ 5 ⋆

KEYWORDS: 3×3 MATRICES, V^3, LINEAR TRANSFORMATIONS, LINEARITY, COMPOSITION, LINEAR TRANSFORMATIONS DETERMINED BY MATRICES, MATRIX TRANSFORMATIONS, IMAGE OF A FIGURE BY A LINEAR TRANSFORMATION

4.1. 3×3 **matrices.** A square array of 9 real numbers with 3 rows and 3 columns is called a 3×3 real matrix. For simplicity, we use a term "3×3 matrix". Actually, a 3×3 matrix is the following square array of real numbers:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \tag{1}$$

Let us denote it by the upper case letter A. Define the multiplication or product of a 3×3 matrix A and a 3-dimensional vector $\mathtt{x}$ by

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} a_{11}x + a_{12}y + a_{13}z \\ a_{21}x + a_{22}y + a_{23}z \\ a_{31}x + a_{32}y + a_{33}z \end{pmatrix} \tag{2}$$

Furthermore, the (matrix) multiplication or (matrix) product of 3×3 matrices A and B is defined as follows. (Sorry, little bit complicated!)

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} =$$

$$\begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} & a_{11}b_{13} + a_{12}b_{23} + a_{13}b_{33} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} & a_{21}b_{13} + a_{22}b_{23} + a_{23}b_{33} \\ a_{31}b_{11} + a_{32}b_{21} + a_{33}b_{31} & a_{31}b_{12} + a_{32}b_{22} + a_{33}b_{32} & a_{31}b_{13} + a_{32}b_{23} + a_{33}b_{33} \end{pmatrix} \tag{3}$$

Let A, B and C be 3×3 matrices, $\mathtt{x}$, $\mathtt{y}$ be 3-dimensional vectors, and k be a real number. Then the following laws of multiplication of matrices or of a matrix and a vector hold.

$$\begin{array}{ll} (AB)C = A(BC) & \text{(associative law)} \\ (AB)\mathtt{x} = A(B\mathtt{x}) & \text{(associative law)} \\ A(\mathtt{x} + \mathtt{y}) = A\mathtt{x} + A\mathtt{y} & \text{(distributive law)} \\ A(k\mathtt{x}) = k(A\mathtt{x}) & \end{array} \tag{4}$$

(exercise01) Show the above laws.

4.2. **V^3.** Denote by V^3 the set of all 3-dimensional vectors. By mapping every point in the 3-dimensional space to its position vector, V^3 can be regarded as the set of all points in the 3-dimensional space, say, the 3-dimensional space itself. In this way, every 3-dimensional figure is represented as some set of 3-dimensional vectors. Vector equations of planes or lines are examples of them.

$$V^3 = \{\text{The set of all 3-dimensional vectors}\} = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mid x, y, z \in \mathbb{R} \right\} \tag{5}$$

4.3. **Linear transformations of V^3.** A mapping (function) T from V^3 to itself, which maps every element $\mathtt{x}$ of V^3 to some element $T(\mathtt{x})$ of V^3, is called a transformation of V^3. Here, for simplicity, $T(\mathtt{x})$ is sometimes written as $T\mathtt{x}$. In particular, if T has the following property "linearity", T is called a linear transformation of V^3.

$$\begin{aligned} T(\mathtt{x}+\mathtt{y}) &= T\mathtt{x} + T\mathtt{y} && (\mathtt{x}, \mathtt{y} \in V^3) \\ T(k\mathtt{x}) &= k(T\mathtt{x}) && (\mathtt{x} \in V^3,\ k \in \mathbb{R}) \end{aligned} \tag{6}$$

From the first formula, we have

$$\begin{aligned} T(\mathtt{x}+\mathtt{y}+\mathtt{z}) = {} & T((\mathtt{x}+\mathtt{y})+\mathtt{z}) \\ = {} & T(\mathtt{x}+\mathtt{y}) + T\mathtt{z} = T\mathtt{x} + T\mathtt{y} + T\mathtt{z}. \end{aligned} \tag{7}$$

Similarly, a linear transformation T satisfies that

$$T(\mathtt{x}_1 + \cdots + \mathtt{x}_n) = T\mathtt{x}_1 + \cdots + T\mathtt{x}_n. \tag{8}$$

Let T and S be two transformations of V^3, then the composition $S \circ T$ is defined as

$$(S \circ T)\mathtt{x} = S(T\mathtt{x}) \qquad (\mathtt{x} \in V^3). \tag{9}$$

$S \circ T$ is sometimes written simply as ST.

Theorem 1. *If T and S are linear transformations of V^3, then ST is also a linear transformation of V^3.*

Proof. For every $\mathtt{x}, \mathtt{y} \in V^3$ and $k \in \mathbb{R}$,

$$\begin{aligned} (ST)(\mathtt{x}+\mathtt{y}) &= S(T(\mathtt{x}+\mathtt{y})) = S(T\mathtt{x}+T\mathtt{y}) && \text{(By linearity of } T\text{)} \\ &= S(T\mathtt{x}) + S(T\mathtt{y}) && \text{(By linearity of } S\text{)} \\ &= (ST)\mathtt{x} + (ST)\mathtt{y}, \\ (ST)(k\mathtt{x}) &= S(T(k\mathtt{x})) = S(kT\mathtt{x}) && \text{(By linearity of } T\text{)} \\ &= kS(T\mathtt{x}) && \text{(By linearity of } S\text{)} \\ &= k(ST)\mathtt{x}. \quad \square \end{aligned} \tag{10}$$

4.4. 3×3 **matrices and linear transformations.** For a 3×3 matrix A, define a linear transformation T_A as follows.

$$T_A(\mathtt{x}) = A\mathtt{x} \quad (\mathtt{x} \in V^3) \tag{11}$$

T_A is called the linear transformation (matrix transformation) of V^3 determined by A. In other words, T_A is a linear transformation defined by the multiplication of A and vectors. If a linear transformation T is expressed as $T = T_A$ for some matrix A, then A is called the matrix of a linear transformation T.

It is clear from the third and fourth formulas of (4) that T_A satisfies the linearity property (6). In fact, the converse is valid.

Theorem 2. *Let T be a linear transformation of V^3. Then there exists a 3×3 matrix A such that $T = T_A$.*

Proof. Suppose T satisfies (6). It suffices to show that there exists a 3×3 matrix A such that for every $\mathtt{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$, $T\mathtt{x} = T_A(\mathtt{x})$. Letting the elementary vectors in V^3 be $\mathtt{e}_1$, $\mathtt{e}_2$ and $\mathtt{e}_3$, we have

$$\begin{aligned} T\mathtt{x} = T(x\mathtt{e}_1 + y\mathtt{e}_2 + z\mathtt{e}_3) = &\quad T(x\mathtt{e}_1) + T(y\mathtt{e}_2) + T(z\mathtt{e}_3) \qquad (12) \\ = &\quad xT\mathtt{e}_1 + yT\mathtt{e}_2 + zT\mathtt{e}_3. \qquad (13) \end{aligned}$$

Here, letting $T\mathtt{e}_j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ a_{3j} \end{pmatrix}$ $(j = 1, 2, 3)$, we have

$$\begin{aligned} (13) \quad &= x \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix} + y \begin{pmatrix} a_{12} \\ a_{22} \\ a_{32} \end{pmatrix} + z \begin{pmatrix} a_{13} \\ a_{23} \\ a_{33} \end{pmatrix} = \begin{pmatrix} a_{11}x + a_{12}y + a_{13}z \\ a_{21}x + a_{22}y + a_{23}z \\ a_{31}x + a_{32}y + a_{33}z \end{pmatrix} \\ &= A\mathtt{x} = T_A(\mathtt{x}). \quad \square \end{aligned} \tag{14}$$

(exercise02) Let T be a linear transformation defined by the following. Determine a matrix A satisfying that $T\mathtt{x} = A\mathtt{x}$ $(\mathtt{x} \in V^3)$.

$$(1)\ T\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \qquad (2)\ T\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2y + 3z \\ 3z + 2x \\ 2x + 3y \end{pmatrix}. \tag{15}$$

Theorem 3. *For linear transformations T_A, T_B of V^3, it holds that*

$$T_A T_B = T_{AB}. \tag{16}$$

Proof. For every $\mathtt{x} \in V^3$, we have

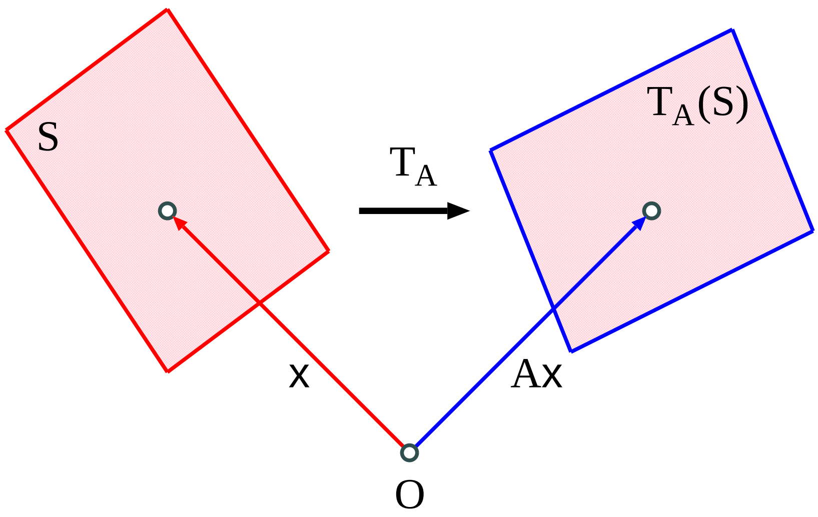
$$\begin{aligned} (T_A T_B)\mathtt{x} &= T_A(T_B(\mathtt{x})) = T_A(B\mathtt{x}) = A(B\mathtt{x}) \\ &= (AB)\mathtt{x} = T_{AB}(\mathtt{x}). \quad \square \end{aligned} \tag{17}$$

This theorem shows that the composition of linear transformations determined by matrices is determined by the product of matrices.

4.5. **Images of planes and lines by T_A.** Take a 3×3 matrix A. If V^3 is regarded as the set of all points in the 3-dimensional space, then a plane S in the 3-dimensional space is regarded as a subset of V^3. The set

$$T_A(S) = \{A\mathrm{x} \mid \mathrm{x} \text{ is the position vector of a point on } S\} \tag{18}$$

is called the image of S by (under) T_A, or the figure onto which S is mapped by T_A. When S is a line or something, definition of $T_A(S)$ is very similar. Using vector equations, we can determine $T_A(S)$ explicitly.



(exercise03) Let $A = \begin{pmatrix} 1 & 1 & 0 \\ 2 & -1 & 2 \\ 0 & 1 & 1 \end{pmatrix}$. Determine the equation of the image S' of a plane $S : x + 2y + 3z = 4$ by T_A.

(ans) The normal vector to S is $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$. Two vectors $\begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 3 \\ 0 \\ -1 \end{pmatrix}$ are perpendicular to the normal vector. S contains a point $(1, 0, 1)$. Hence the vector equation of S is the following.

$$\mathrm{x} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + t\begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix} + u\begin{pmatrix} 3 \\ 0 \\ -1 \end{pmatrix}. \qquad (t, u \in \mathbb{R}) \tag{19}$$

Therefore the vector equation of S' is the following.

$$\begin{aligned} \mathrm{x}' &= A\mathrm{x} = \begin{pmatrix} 1 & 1 & 0 \\ 2 & -1 & 2 \\ 0 & 1 & 1 \end{pmatrix} \left[\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + t\begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix} + u\begin{pmatrix} 3 \\ 0 \\ -1 \end{pmatrix} \right] \\ &= \begin{pmatrix} 1 & 1 & 0 \\ 2 & -1 & 2 \\ 0 & 1 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + t\begin{pmatrix} 1 & 1 & 0 \\ 2 & -1 & 2 \\ 0 & 1 & 1 \end{pmatrix}\begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix} + u\begin{pmatrix} 1 & 1 & 0 \\ 2 & -1 & 2 \\ 0 & 1 & 1 \end{pmatrix}\begin{pmatrix} 3 \\ 0 \\ -1 \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ 4 \\ 1 \end{pmatrix} + t\begin{pmatrix} 1 \\ 5 \\ -1 \end{pmatrix} + u\begin{pmatrix} 3 \\ 4 \\ -1 \end{pmatrix}. \end{aligned} \tag{20}$$

Next we transform it into the equation. The normal vector to S' is

$$\begin{pmatrix} 1 \\ 5 \\ -1 \end{pmatrix} \times \begin{pmatrix} 3 \\ 4 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ -2 \\ -11 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 \\ 2 \\ 11 \end{pmatrix}. \tag{21}$$

S' contains $(1, 4, 1)$. Hence the equation of S' is $\boxed{x - 1 + 2(y - 4) + 11(z - 1) = 0}$, i.e. $\boxed{x + 2y + 11z = 20}$.

(exercise04) Let A be as above. Determine the equation of the image l' of a line l : $\frac{x-1}{4} = \frac{y-2}{5} = \frac{z-3}{6}$ by T_A.

(ans) Letting $\frac{x-1}{4} = \frac{y-2}{5} = \frac{z-3}{6} = t$ (or since the direction vector of l is $\begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}$ and l contains $(1, 2, 3)$), the vector equation of l is as follows.

$$\mathtt{x} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + t \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}. \qquad (t \in \mathbb{R}) \tag{22}$$

Hence the vector equation of l' is

$$\begin{aligned} \mathtt{x}' = A\mathtt{x} \quad &= \begin{pmatrix} 1 & 1 & 0 \\ 2 & -1 & 2 \\ 0 & 1 & 1 \end{pmatrix} \left[\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + t \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} \right] \\ &= \begin{pmatrix} 1 & 1 & 0 \\ 2 & -1 & 2 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + t \begin{pmatrix} 1 & 1 & 0 \\ 2 & -1 & 2 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} \\ &= \begin{pmatrix} 3 \\ 6 \\ 5 \end{pmatrix} + t \begin{pmatrix} 9 \\ 15 \\ 11 \end{pmatrix}. \end{aligned} \tag{23}$$

Eliminating t (or since the direction vector of l' is $\begin{pmatrix} 9 \\ 15 \\ 11 \end{pmatrix}$ and l contains $(3, 6, 5)$), we have the equation of l': $\boxed{\frac{x-3}{9} = \frac{y-6}{15} = \frac{z-5}{11}}$.

★

5. MATRIX OPERATIONS AND ELEMENTARY OPERATIONS ON MATRICES

⋆ 13 ⋆

KEYWORDS: COMPLEX MATRICES, REAL MATRICES, ROW VECTORS, COLUMN VECTORS, TRANSPOSE, MATRIX ADDITION, SUM, SCALAR MULTIPLICATION, MATRIX MULTIPLICATION, PRODUCT, DIAGONAL MATRICES, SCALAR MATRICES, IDENTITY MATRIX, ZERO MATRIX, BLOCK (PARTITIONED) MATRICES, INVERSE MATRIX, NONSINGULAR (INVERTIBLE) MATRICES, TRACE, ELEMENTARY MATRICES, ELEMENTARY OPERATIONS, RANK, CANONICAL FORM, INNER PRODUCT, NORMAL MATRICES, HERMITIAN MATRICES, SYMMETRIC MATRICES, ALTERNATIVE MATRICES, UNITARY MATRICES, ORTHOGONAL MATRICES

5.1. $m \times n$ **matrices.** A horizontal array of numbers or symbols is called a row, and a vertical array of numbers or symbols is called a column. A rectangular array (1) of numbers consisting of m rows and n columns is called an $m \times n$ (read "m by n") matrix. The dimensions of an $m \times n$ matrix is defined to be $m \times n$, the number of its rows and the number of its columns. The rows of the matrix are called the first row, the second row, . . . , and the m-th row from top to bottom, and the columns of the matrix are called the first column, the second column, . . . , and the n-th column from left to right.

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots\cdots\cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots\cdots\cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots\cdots\cdots & a_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots\cdots\cdots & a_{mn} \end{pmatrix} \tag{1}$$

The individual items in a matrix is called its entries or elements. In particular, the entry on the intersection of the i-th row and the j-th column of a matrix is called its (i, j) entry. A matrix with complex number entries is called a complex matrix, whereas a matrix with real number entries is called a real matrix. By definition a real matrix is a complex matrix. An $n \times n$ matrix is also called a square matrix of order n, or more simply, a matrix of order n, whose dimension is defined to be n.

An $m \times 1$ matrix is called a column vector with m entries, and a $1 \times n$ matrix is called a row vector with n entries. A vector with complex/real number entries is called a complex/real vector. Accordingly, vectors are special cases of matrices. A 1×1 matrix is considered as an ordinary number (scalar).

We usually denote the matrix (1) by A. If the entries are b_{ij}, we denote it by B. For simplicity, we use the notation:

$$A = (a_{ij}) \qquad (1 \leq i \leq m,\ 1 \leq j \leq n). \tag{2}$$

This means that A is a matrix with (i, j) entry a_{ij}. The parenthesised item at end of (2) can be omitted if confusion does not occur. For convenience, we sometimes change the suffices as $A = (a_{jk})$, etc, which has the same meaning.

Two matrices A and B are identical, i.e. $A = B$, if they have the same dimension and the same (i, j) entries for every i, j.

For an $m \times n$ matrix $A = (a_{ij})$, the $n \times m$ matrix $B = (b_{ij})$, $b_{ij} = a_{ji}$ is called the transpose of A, denoted by $B = {}^tA$ or A^T. The j-th row of ${}^tA$ is the transpose of the j-th column of A, and the i-th column of ${}^tA$ is the transpose of the i-th row of A. For example,

$$\begin{pmatrix} 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}^T = \begin{pmatrix} 4 & 7 \\ 5 & 8 \\ 6 & 9 \end{pmatrix}. \tag{3}$$

It is clear that ${}^t({}^tA) = A$.

It is sometimes convenient to decompose (1) into row or column vectors. Letting

$$\mathsf{a}_j = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix}, \qquad \mathsf{a}'_i = \begin{pmatrix} a_{i1} & a_{i2} & \ldots & a_{in} \end{pmatrix}; \tag{4}$$

we have

$$A = \begin{pmatrix} \mathsf{a}_1 & \mathsf{a}_2 & \ldots & \mathsf{a}_n \end{pmatrix}, \qquad A = \begin{pmatrix} \mathsf{a}'_1 \\ \vdots \\ \mathsf{a}'_m \end{pmatrix}. \tag{5}$$

5.2. **Matrix addition and scalar multiplication.** Matrix addition and scalar multiplication of matrices are defined similarly to 3-dimensional vectors. Let $A = (a_{ij})$ and $B = (b_{ij})$ be $m \times n$ matrices, then the sum $A + B$ of A and B (addition of A and B) is defined to be the following $m \times n$ matrix.

$$\begin{aligned} &\begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & \ldots & b_{1n} \\ b_{21} & b_{22} & \ldots & b_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ b_{m1} & b_{m2} & \ldots & b_{mn} \end{pmatrix} \\ = &\begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \ldots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \ldots & a_{2n}+b_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \ldots & a_{mn}+b_{mn} \end{pmatrix} \end{aligned} \tag{6}$$

Let k be a scalar, then the scalar multiplication kA of A by k is defined to be the following $m \times n$ matrix.

$$k \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} = \begin{pmatrix} ka_{11} & ka_{12} & \ldots & ka_{1n} \\ ka_{21} & ka_{22} & \ldots & ka_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ ka_{m1} & ka_{m2} & \ldots & ka_{mn} \end{pmatrix} \tag{7}$$

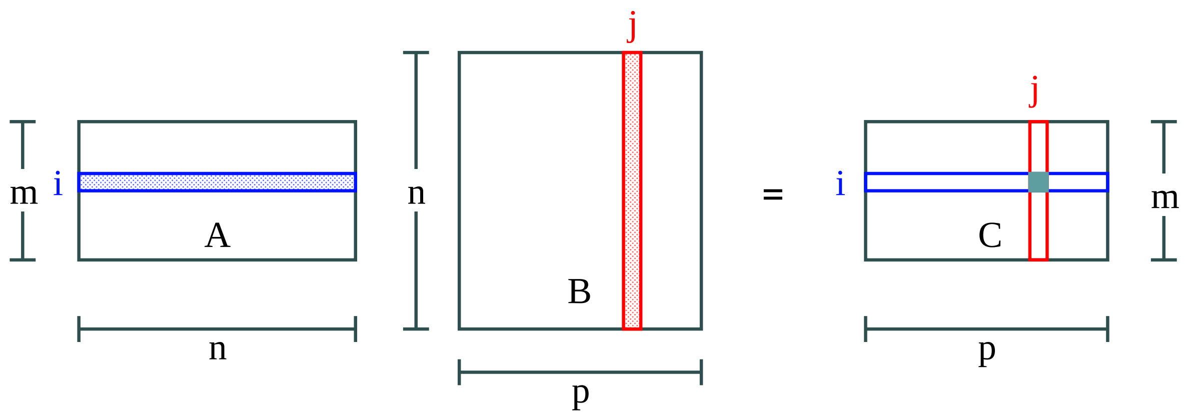
It is easily confirmed that the following laws of matrix addition and scalar multiplication hold.

$$\begin{array}{ll} A+B=B+A & \text{(commutative law)} \\ (A+B)+C=A+(B+C) & \text{(associative law)} \\ k(A+B)=kA+kB & \\ (k+l)A=kA+lA & \\ (kl)A=k(lA) & \end{array} \tag{8}$$

An $m \times n$ matrix with only 0 entries is called a zero matrix, denoted by O_{mn} or O. It is clear that $A+O=O+A=A$, $1A=A$, $0A=O$. A matrix B such that $A+B=B+A=O$ is denoted by $-A$. It holds that $-A=(-1)A$, $-(-A)=A$. A row/column vector with only 0 entries is called a zero vector, denoted by $\mathtt{0}$.
(exercise01) Show (8).
(note) We often write simply that $A+(-B)=A-B$.

5.3. **Matrix multiplication.** Let $A=(a_{ij})$ be an $m \times n$ matrix, $B=(b_{ij})$ be an $n \times p$ matrix. Then we define the product $AB=C=(c_{ij})$ of A and B (multiplication of A and B) as follows. C is an $m \times p$ matrix such that

$$c_{ij}=\sum_{k=1}^{n} a_{ik}b_{kj}. \tag{9}$$



(exercise02) Calculate the following.

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{pmatrix}\begin{pmatrix} 2 & 1 & 1 \\ -1 & 3 & -2 \\ 2 & -3 & 1 \end{pmatrix}+\begin{pmatrix} 1 & -3 & -1 \\ 2 & -1 & -4 \\ 1 & 1 & 2 \end{pmatrix}\begin{pmatrix} 2 & 1 & 1 \\ -1 & 3 & -2 \\ 2 & -3 & 1 \end{pmatrix} \\ +\begin{pmatrix} 1 & 1 & -2 \\ -4 & 1 & 3 \\ -4 & -2 & -1 \end{pmatrix}\begin{pmatrix} 2 & 1 & 1 \\ -1 & 3 & -2 \\ 2 & -3 & 1 \end{pmatrix} \tag{10}$$

(ans) Answer only: $\begin{pmatrix} 6 & 3 & 3 \\ -3 & 9 & -6 \\ 6 & -9 & 3 \end{pmatrix}$.

In general, $AB=BA$ does not hold for matrices A and B. If it holds, A and B are said to be commutative. The following laws of matrix addition and multiplication hold, provided that matrices have appropriate dimensions so that matrix operations can be performed. For example, in 1:, A, B and C are $m \times n$, $n \times p$ and $p \times q$, respectively, and in 2:, A is $m \times n$, and B, C are $n \times p$.

$$\begin{array}{lll} 1: & (AB)C = A(BC) & \text{(associative law)} \\ 2: & A(B+C) = AB + AC & \text{(distributive law)} \\ 3: & (A+B)C = AC + BC & \text{(distributive law)} \\ 4: & (kA)B = A(kB) = k(AB) & \end{array} \tag{11}$$

(exercise03) Prove (11).

Proof of 1:. It is clear that both sides are defined and they have the same dimension $m \times q$. Let $AB = (x_{ik})$, then $x_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}$. Hence letting $(AB)C = (y_{il})$, we have

$$\begin{aligned} y_{il} &= \textstyle\sum_{k=1}^{p} x_{ik} c_{kl} = \sum_{k=1}^{p} \left(\sum_{j=1}^{n} a_{ij} b_{jk} \right) c_{kl} \\ &= \textstyle\sum_{k=1}^{p} \sum_{j=1}^{n} a_{ij} b_{jk} c_{kl}. \end{aligned} \tag{12}$$

Next let $BC = (\tilde{x}_{jl})$, then $\tilde{x}_{jl} = \sum_{k=1}^{p} b_{jk} c_{kl}$. Hence letting $A(BC) = (\tilde{y}_{il})$, we have

$$\begin{aligned} \tilde{y}_{il} &= \textstyle\sum_{j=1}^{n} a_{ij} \tilde{x}_{jl} = \sum_{j=1}^{n} a_{ij} \left(\sum_{k=1}^{p} b_{jk} c_{kl} \right) \\ &= \textstyle\sum_{j=1}^{n} \sum_{k=1}^{p} a_{ij} b_{jk} c_{kl}. \end{aligned} \tag{13}$$

From (12) and (13), it follows that $y_{il} = \tilde{y}_{il}$. Therefore $(AB)C = A(BC)$. □

Proof of 2:. It is clear that both sides are defined and they have the same dimension $m \times p$. Let $B + C = (x_{jk})$, then $x_{jk} = b_{jk} + c_{jk}$. Hence letting $A(B + C) = (y_{ik})$, we have

$$y_{ik} = \sum_{j=1}^{n} a_{ij} x_{jk} = \sum_{j=1}^{n} a_{ij} (b_{jk} + c_{jk}) = \sum_{j=1}^{n} a_{ij} b_{jk} + \sum_{j=1}^{n} a_{ij} c_{jk}. \tag{14}$$

Next let $AB = (\tilde{x}_{ik})$, $AC = (\tilde{y}_{ik})$ and $AB + AC = (\tilde{z}_{ik})$, then

$$\begin{aligned} & \tilde{x}_{ik} = \textstyle\sum_{j=1}^{n} a_{ij} b_{jk}, \quad \tilde{y}_{ik} = \sum_{j=1}^{n} a_{ij} c_{jk}. \\ \therefore \quad & \tilde{z}_{ik} = \tilde{x}_{ik} + \tilde{y}_{ik} = \textstyle\sum_{j=1}^{n} a_{ij} b_{jk} + \sum_{j=1}^{n} a_{ij} c_{jk}. \end{aligned} \tag{15}$$

From (14) and (15), it follows that $y_{ik} = \tilde{z}_{ik}$. Therefore $A(B + C) = AB + AC$. □

According to (8) and (11), the associative laws of addition and multiplication hold. Thus the sum or product of several matrices does not depend on the way to insert parentheses.

$$\begin{array}{l} A_1 + A_2 + \cdots + A_s \\ A_1 A_2 \ldots A_s \end{array} \tag{16}$$

Hence in expressions as (16), parentheses are usually omitted. Moreover, by distributive law (11), we have

$$\begin{array}{l} A(B_1 + B_2 + \cdots + B_s) = AB_1 + AB_2 + \cdots + AB_s, \\ (A_1 + A_2 + \cdots + A_s)B = A_1 B + A_2 B + \cdots + A_s B. \end{array} \tag{17}$$

(exercise04) Prove the above by induction.

Let A be an $m \times n$ matrix, then obviously it holds that $AO_{np} = O_{mp}$, and $O_{lm}A = O_{ln}$.

The (i,i) entry of a matrix is called a diagonal entry. An $n\times n$ matrix, such that only diagonal entries may have nonzero values, is called a diagonal matrix of order n.

$$\begin{pmatrix} a_{11} & 0 & 0 & \dots & 0 \\ 0 & a_{22} & 0 & \dots & 0 \\ 0 & 0 & a_{33} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_{nn} \end{pmatrix} \tag{18}$$

This diagonal matrix is often simply expressed as follows.

$$\begin{pmatrix} a_{11} & & & O \\ & a_{22} & & \\ & & \ddots & \\ O & & & a_{nn} \end{pmatrix} \tag{19}$$

Here, O in the matrix does not mean a zero matrix, but it means the symbol O and the space around it has only zero entries. The above diagonal matrix is sometimes denoted more simply by $\mathrm{diag}(a_{11},\dots,a_{nn})$. It is clear that

$$\begin{aligned} &\mathrm{diag}(a_{11},\dots,a_{nn})+\mathrm{diag}(b_{11},\dots,b_{nn})=\mathrm{diag}(a_{11}+b_{11},\dots,a_{nn}+b_{nn}) \\ &\mathrm{diag}(a_{11},\dots,a_{nn})\,\mathrm{diag}(b_{11},\dots,b_{nn})=\mathrm{diag}(a_{11}b_{11},\dots,a_{nn}b_{nn}) \\ &(\mathrm{diag}(a_{11},\dots,a_{nn}))^s=\mathrm{diag}\left(a_{11}^s,\dots,a_{nn}^s\right) \quad (s=1,2,3,\dots) \quad (\Rightarrow 5.5). \end{aligned} \tag{20}$$

A diagonal matrix of order n where every diagonal entry is 1 is called the identity matrix of order n, denoted by E_n or E. Every column vector of E_n is called an n-dimensional elementary vector, denoted by $\mathsf{e}_1,\mathsf{e}_2,\dots,\mathsf{e}_n$ from left to right.

$$E_n=\begin{pmatrix} 1 & & & O \\ & 1 & & \\ & & \ddots & \\ O & & & 1 \end{pmatrix};\quad \mathsf{e}_1=\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix},\ \mathsf{e}_2=\begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix},\dots,\ \mathsf{e}_n=\begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \tag{21}$$

For every $m\times n$ matrix A, it holds that

$$AE_n=E_mA=A. \tag{22}$$

(exercise05) (1) Confirm (20). (2) Show (22). (3) Let D be a diagonal matrix of order n and A be a matrix of order n. Calculate AD and DA.

For some scalar k, a matrix kE_n is called a scalar matrix. By (11) and (22), we have

$$\begin{aligned} A(kE_n)=k(AE_n)=kA, \\ (kE_m)A=k(E_mA)=kA. \end{aligned} \tag{23}$$

That is, multiplying a matrix by a scalar matrix on the left or right causes only scalar multiplication. Conversely, only scalar matrices have such a property.

Theorem 1. *Let F be a matrix of order n, then*

$$F \text{ is a scalar matrix} \iff \text{For any matrix } X \text{ of order } n,\ FX=XF. \tag{24}$$

Proof. $\Rightarrow$ 5.11.

A square matrix $A = (a_{ij})$ is called upper triangular if it holds that $i > j \implies a_{ij} = 0$, and A is called lower triangular if it holds that $i < j \implies a_{ij} = 0$. A square matrix is called triangular if it is upper triangular or lower triangular. The sum or product of upper (respectively, lower) triangular matrices of the same order is also upper (respectively, lower) triangular.

5.4. **Block matrices.** If an $m \times n$ matrix A is parted by vertical or horizontal lines, then A is decomposed into several smaller matrices. For example,

$$A = \begin{array}{|c|c|c|l} \hline A_{11} & A_{12} & A_{13} & m_1 \\ \hline A_{21} & A_{22} & A_{23} & m_2 \\ \hline A_{31} & A_{32} & A_{33} & m_3 \\ \hline \multicolumn{1}{c}{n_1} & \multicolumn{1}{c}{n_2} & \multicolumn{1}{c}{n_3} & \end{array} = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix}. \tag{25}$$

This matrix expression is called a block matrix or a partitioned matrix. It is also simply denoted by $A = (A_{ij})$ $(1 \le i, j \le 3)$. In general, a block matrix has the form:

$$\begin{aligned} A &= \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1s} \\ A_{21} & A_{22} & \dots & A_{2s} \\ \multicolumn{4}{c}{\dotfill} \\ A_{r1} & A_{r2} & \dots & A_{rs} \end{pmatrix} \\ &= (A_{ij}) \quad (1 \le i \le r,\ 1 \le j \le s), \end{aligned} \tag{26}$$

which is called an $r \times s$ block matrix, where A_{ij} is called the (i, j) block or a submatrix of A.[1] Note that $r \times s$ is not the dimension of A but the number of blocks. A block matrix of dimension $n \times n$ is called a block matrix of order n.

Let $A = (A_{ij})$ $(1 \le i \le r,\ 1 \le j \le s)$ be an $r \times s$ block matrix, and A_{ij} be an $m_i \times n_j$ matrix for every i, j. Then the row partition of A is defined to be the sequence $(m_1, m_2, \dots, m_r)$, and the column partition of A is defined to be the sequence $(n_1, n_2, \dots, n_s)$. Any two matrices of the same dimension are called identically partitioned if they have the same row partition and the same column partition.

Let $B = (B_{ij})$ be a 3×3 block matrix, identically partitioned to (25). Then it holds that

$$A + B = \begin{pmatrix} A_{11} + B_{11} & A_{12} + B_{12} & A_{13} + B_{13} \\ A_{21} + B_{21} & A_{22} + B_{22} & A_{23} + B_{23} \\ A_{31} + B_{31} & A_{32} + B_{32} & A_{33} + B_{33} \end{pmatrix}. \tag{27}$$

In general, (including scalar multiplication) we have

Theorem 2. *Suppose two $r \times s$ block matrices $A = (A_{ij})$ and $B = (B_{ij})$ $(1 \le i \le r,\ 1 \le j \le s)$ are identically partitioned, and k is a scalar, then*

$$\begin{aligned} A + B &= (A_{ij} + B_{ij}) \qquad (1 \le i \le r,\ 1 \le j \le s), \\ kA &= (kA_{ij}) \qquad (1 \le i \le r,\ 1 \le j \le s). \end{aligned} \tag{28}$$

[1] A matrix and its arbitrary partitionings (block matrices) are completely identified with each other.

Next we consider multiplication of block matrices. Let A be a block matrix in (25), and let B be the following block matrix:

$$B = \begin{array}{|c|c|c} \hline B_{11} & B_{12} & n_1 \\ \hline B_{21} & B_{22} & n_2 \\ \hline B_{31} & B_{32} & n_3 \\ \hline p_1 & p_2 & \end{array} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \\ B_{31} & B_{32} \end{pmatrix}. \tag{29}$$

Here, note that the column partition of A is identical to the row partition of B. Then we can perform multiplication as follows.

$$AB = \begin{pmatrix} A_{11}B_{11} + A_{12}B_{21} + A_{13}B_{31} & A_{11}B_{12} + A_{12}B_{22} + A_{13}B_{32} \\ A_{21}B_{11} + A_{22}B_{21} + A_{23}B_{31} & A_{21}B_{12} + A_{22}B_{22} + A_{23}B_{32} \\ A_{31}B_{11} + A_{32}B_{21} + A_{33}B_{31} & A_{31}B_{12} + A_{32}B_{22} + A_{33}B_{32} \end{pmatrix} \tag{30}$$

Theorem 2'. *Let $A = (A_{ij})$ $(1 \le i \le r,\ 1 \le j \le s)$ be an $r \times s$ block matrix and $B = (B_{ij})$ $(1 \le i \le s,\ 1 \le j \le t)$ be an $s \times t$ block matrix, and suppose A_{ij} is $m_i \times n_j$ and B_{ij} is $n_i \times p_j$. Then*

$$AB = (C_{ij}) \quad (1 \le i \le r,\ 1 \le j \le t), \qquad C_{ij} = \sum_{k=1}^{s} A_{ik}B_{kj}. \tag{31}$$

Proof. Consider the case (30). Let $A = (a_{ij})$ and $B = (b_{ij})$. Focus on the (i, j) entry c_{ij} of AB. For simplicity, the i-th row of A is in the top blocks, and the j-th column of B is in the left-most blocks. By definition of matrix multiplication, (9) holds, which is rewritten as

$$c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj} = \sum_{k=1}^{n_1} a_{ik}b_{kj} + \sum_{k=n_1+1}^{n_1+n_2} a_{ik}b_{kj} + \sum_{k=n_1+n_2+1}^{n} a_{ik}b_{kj}. \tag{32}$$

However, these terms are the (i, j) entries of $A_{11}B_{11}$, $A_{12}B_{21}$ and $A_{13}B_{31}$, respectively. Consequently,

$$(\text{the } (i, j) \text{ entry of } AB) = (\text{the } (i, j) \text{ entry of } A_{11}B_{11} + A_{12}B_{21} + A_{13}B_{31}). \tag{33}$$

We can treat similarly the case where the (i, j) entry is located in the other block. □

(exercise06) Using block matrices, perform the matrix calculation below.

$$\begin{pmatrix} 2 & 0 & 0 & 3 \\ 0 & 2 & 0 & 2 \\ 0 & 0 & 2 & 1 \\ 1 & 0 & 1 & 3 \end{pmatrix} \begin{pmatrix} 3 & 2 & 1 & 4 \\ 2 & 1 & 3 & 4 \\ 1 & 3 & 2 & 3 \\ 1 & 0 & 0 & 3 \end{pmatrix} \tag{34}$$

(ans)

$$\begin{pmatrix} 2 & 0 & 0 & 3 \\ 0 & 2 & 0 & 2 \\ 0 & 0 & 2 & 1 \\ 1 & 0 & 1 & 3 \end{pmatrix}\begin{pmatrix} 3 & 2 & 1 & 4 \\ 2 & 1 & 3 & 4 \\ 1 & 3 & 2 & 3 \\ 1 & 0 & 0 & 3 \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}\begin{pmatrix} 3 & 2 & 1 & 4 \\ 2 & 1 & 3 & 4 \\ 1 & 3 & 2 & 3 \end{pmatrix} + \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 & 3 \end{pmatrix} \\ \begin{pmatrix} 1 & 0 & 1 \end{pmatrix}\begin{pmatrix} 3 & 2 & 1 & 4 \\ 2 & 1 & 3 & 4 \\ 1 & 3 & 2 & 3 \end{pmatrix} + \begin{pmatrix} 3 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 & 3 \end{pmatrix} \end{pmatrix}$$
$$= \begin{pmatrix} \begin{pmatrix} 6 & 4 & 2 & 8 \\ 4 & 2 & 6 & 8 \\ 2 & 6 & 4 & 6 \end{pmatrix} + \begin{pmatrix} 3 & 0 & 0 & 9 \\ 2 & 0 & 0 & 6 \\ 1 & 0 & 0 & 3 \end{pmatrix} \\ \begin{pmatrix} 4 & 5 & 3 & 7 \end{pmatrix} + \begin{pmatrix} 3 & 0 & 0 & 9 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 9 & 4 & 2 & 17 \\ 6 & 2 & 6 & 14 \\ 3 & 6 & 4 & 9 \\ 7 & 5 & 3 & 16 \end{pmatrix}. \tag{35}$$

A block matrix of order n is called a symmetrically partitioned matrix if the row partition is identical to the column partition. In other words, if a block matrix:

$$\begin{pmatrix} A_{11} & A_{12} & \dots & A_{1t} \\ A_{21} & A_{22} & \dots & A_{2t} \\ \dots & \dots & \dots & \dots \\ A_{t1} & A_{t2} & \dots & A_{tt} \end{pmatrix} \tag{36}$$

has square blocks $A_{11}, A_{22}, \dots, A_{tt}$, then it is called symmetrically partitioned. If several square matrices of order n are symmetrically and identically partitioned, then we can perform addition and multiplication keeping their symmetrically partitioned form.

A symmetrically partitioned matrix (36) satisfying that $i \neq j \Rightarrow A_{ij} = O$ is called a block diagonal matrix.[2] For two identically partitioned block diagonal matrices, we have

$$\begin{pmatrix} A_1 & & & O \\ & A_2 & & \\ & & \ddots & \\ O & & & A_t \end{pmatrix}\begin{pmatrix} B_1 & & & O \\ & B_2 & & \\ & & \ddots & \\ O & & & B_t \end{pmatrix} = \begin{pmatrix} A_1B_1 & & & O \\ & A_2B_2 & & \\ & & \ddots & \\ O & & & A_tB_t \end{pmatrix}. \tag{37}$$

Therefore for a natural number s, we have the following. (⇒ 5.5)

$$\begin{pmatrix} A_1 & & & O \\ & A_2 & & \\ & & \ddots & \\ O & & & A_t \end{pmatrix}^s = \begin{pmatrix} A_1^s & & & O \\ & A_2^s & & \\ & & \ddots & \\ O & & & A_t^s \end{pmatrix} \tag{38}$$

5.5. **Inverse matrices.** For a matrix A of order n, a matrix X of order n which satisfies the following is called the inverse (matrix) of A, denoted by A^{-1}.

$$AX = XA = E_n \tag{39}$$

By this definition, $X = A^{-1}$ and simultaneously, $A = X^{-1}$, that is, A and X are the inverse matrices of each other. Hence it holds that $A = (A^{-1})^{-1}$. However, not every matrix has an inverse. If A has an inverse, then A is called nonsingular, invertible or

[2]If a symmetrically partitioned matrix (36) satisfies that $i > j \Rightarrow A_{ij} = O$ (respectively, $i < j \Rightarrow A_{ij} = O$), then it is called a block upper triangular (respectively, lower triangular) matrix. If a matrix is block upper triangular or block lower triangular, then it is called block triangular.

nondegenerate, and the inverse of A is unique, because letting X and Y be inverses of A, then

$$Y = E_nY = (XA)Y = X(AY) = XE_n = X. \tag{40}$$

Suppose A and B are nonsingular matrices of order n, then $(AB)^{-1} = B^{-1}A^{-1}$. Indeed,

$$\begin{aligned}(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AE_nA^{-1} = AA^{-1} = E_n \\ (B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}E_nB = B^{-1}B = E_n.\end{aligned} \tag{41}$$

Similarly, suppose $A_1, A_2, \dots, A_s$ are nonsingular matrices of order n, then we see that

$$(A_1A_2 \dots A_s)^{-1} = A_s^{-1} \dots A_2^{-1}A_1^{-1}. \tag{42}$$

Therefore the product of several nonsingular matrices are also nonsingular.
(note) A square matrix that is not nonsingular is called singular or degenerate.

An explicit formula for the inverse of an arbitrary nonsingular matrix is given later. Here we consider several formulas for the inverses of simple block matrices. Let $A, C, A_1, A_2, \dots, A_t$ be nonsingular matrices of various orders, then

$$\begin{aligned}&\begin{pmatrix} A & B \\ O & C \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} & -A^{-1}BC^{-1} \\ O & C^{-1} \end{pmatrix}; \begin{pmatrix} A & O \\ O & C \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} & O \\ O & C^{-1} \end{pmatrix} \\ &\begin{pmatrix} A & O \\ B & C \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} & O \\ -C^{-1}BA^{-1} & C^{-1} \end{pmatrix} \\ &\begin{pmatrix} A_1 & & & O \\ & A_2 & & \\ & & \ddots & \\ O & & & A_t \end{pmatrix}^{-1} = \begin{pmatrix} A_1^{-1} & & & O \\ & A_2^{-1} & & \\ & & \ddots & \\ O & & & A_t^{-1} \end{pmatrix}.\end{aligned} \tag{43}$$

(exercise07) (1) Show those formulas by block matrix calculation. (2) Show that the inverse of a 2×2 nonsingular matrix is given by: $\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \dfrac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$ $(ad - bc \neq 0)$.

The product of s copies of a matrix A of order n is called the s-th power of A, denoted by A^s. The following exponential laws hold. Letting r, s be natural numbers,

$$A^rA^s = A^{r+s}, \qquad (A^r)^s = A^{rs}. \tag{44}$$

Define $A^0 = E_n$, and if A is nonsingular, define $A^{-s} = \left(A^{-1}\right)^s$, then (44) holds for every integers r, s. If $AB = BA$, it holds that $(AB)^s = A^sB^s$.[3]

The following theorem concerning nonsingular matrices is very useful.

Theorem 3. *Let A and B be matrices of order n. If $AB = E_n$, then $BA = E_n$, hence A and B are nonsingular and one is the inverse of the other.*

Proof. $\Rightarrow$ 5.11.

(note) We refrain to use this theorem especially for easy proof problems.

[3]If $AB = BA$, for nonnegative integer s, we have $(A + B)^s = \sum_{k=0}^{s} \binom{s}{k} A^{s-k}B^k$ (the binomial theorem), etc.

5.6. **Trace.** Let $A = (a_{ij})$ be a matrix of order n. The sum of all diagonal entries of A is called the trace of A, denoted by $\mathrm{tr}A$.

$$\mathrm{tr}A = a_{11} + a_{22} + \cdots + a_{nn}. \tag{45}$$

Let B be another matrix of order n, then the following holds.

$$\begin{aligned} &\mathrm{tr}(A+B) = \mathrm{tr}A + \mathrm{tr}B \\ &\mathrm{tr}(kA) = k\,\mathrm{tr}A \\ &\mathrm{tr}(AB) = \mathrm{tr}(BA) \\ &\mathrm{tr}A = \mathrm{tr}({}^tA) \end{aligned} \tag{46}$$

(exercise08) (1) Show the above formulas. (2) Let P be a nonsingular matrix of order n, then prove that $\mathrm{tr}(P^{-1}AP) = \mathrm{tr}A$.

Proof of the third formula of (46).

$$\mathrm{tr}(AB) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ji} = \sum_{j=1}^{n}\sum_{i=1}^{n} b_{ji}a_{ij} = \mathrm{tr}(BA) \quad \square \tag{47}$$

As we will learn later, the trace of a square matrix is the sum of all eigenvalues (including their multiplicities) of the matrix. (⇒ 12)

5.7. **Elementary operations on matrices.** The concept of elementary operations is very important in the theory of matrices. To begin with we introduce three kinds of elementary matrices.

$$
\begin{array}{lll}
P_n(i,j) & = \begin{array}{l}\text{A matrix given by interchanging the } i\text{-th row and} \\ \text{the } j\text{-th row of } E_n\text{, or equivalently, interchanging} \\ \text{the } i\text{-th column and the } j\text{-th column of } E_n.\end{array} & (i \neq j) \\
Q_n(i,c) & = \text{A matrix given by replacing the } (i,i) \text{ entry of } E_n \text{ by } c. & (c \neq 0) \\
R_n(i,j;c) & = \text{A matrix given by replacing the } (i,j) \text{ entry of } E_n \text{ by } c. & (i \neq j)
\end{array} \tag{48}
$$

$$
P_n(i,j) = \begin{array}{c} \\ \\ \\ i) \\ \\ \\ \\ j) \\ \\ \\ \\ \end{array}
\begin{array}{c}
\begin{array}{cccccccccc} & & & i) & & & j) & & & \end{array} \\
\begin{pmatrix}
1 & & & & & & & & \\
& \ddots & & \vdots & & & \vdots & & \\
& & 1 & & & & & & \\
& \cdots & & 0 & & \cdots & 1 & \cdots & \\
& & & & 1 & & & & \\
& & & \vdots & & \ddots & \vdots & & \\
& & & & & 1 & & & \\
& \cdots & & 1 & & \cdots & 0 & \cdots & \\
& & & & & & & 1 & \\
& & & \vdots & & & \vdots & \ddots & \\
& & & & & & & & 1
\end{pmatrix}
\end{array}
$$

$$
Q_n(i,c) = \begin{array}{c} \\ \\ \\ i) \\ \\ \\ \\ \end{array}
\begin{array}{c}
\begin{array}{ccccc} & & & i) & \end{array} \\
\begin{pmatrix}
1 & & & & & \\
& \ddots & & \vdots & & \\
& & 1 & & & \\
& \cdots & & c & & \cdots \\
& & & & 1 & \\
& & & \vdots & & \ddots \\
& & & & & & 1
\end{pmatrix}
\end{array} \tag{49}
$$

$$
R_n(i,j;c) = \begin{array}{c} \\ \\ i) \\ \\ \\ \end{array}
\begin{array}{c}
\begin{array}{cccc} & & & j) \end{array} \\
\begin{pmatrix}
1 & & & \\
& \ddots & \vdots & \\
& \cdots & c & \\
& & \ddots & \\
& & & 1
\end{pmatrix}
\end{array}
$$

As we have seen, all elementary matrices are obtained by minor changes of the identity matrix.

Here we give several examples of elementary matrices of order 3.

$$P_3(1,2) = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad Q_3(1,5) = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$
$$R_3(2,3;-4) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -4 \\ 0 & 0 & 1 \end{pmatrix} \tag{50}$$

Next we observe the result of the multiplication of an $m \times n$ matrix A on the left by an elementary matrix.

$$\begin{array}{lll} P_m(i,j)A & : \text{interchanges the } i\text{-th row and the } j\text{-th row of } A. & (i \neq j) \\ Q_m(i,c)A & : \text{multiplies the } i\text{-th row of } A \text{ by } c. & (c \neq 0) \\ R_m(i,j;c)A & : \begin{array}{l}\text{adds a scalar } c \text{ multiple of the } j\text{-th row of } A \\ \text{to the } i\text{-th row of } A.\end{array} & (i \neq j) \end{array} \tag{51}$$

Similarly, the multiplication of A on the right by an elementary matrix follows.

$$\begin{array}{lll} AP_n(i,j) & : \text{interchanges the } i\text{-th column and the } j\text{-th column of } A. & (i \neq j) \\ AQ_n(i,c) & : \text{multiplies the } i\text{-th column of } A \text{ by } c. & (c \neq 0) \\ AR_n(i,j;c) & : \begin{array}{l}\text{adds a scalar } c \text{ multiple of the } i\text{-th column of } A \\ \text{to the } j\text{-th column of } A.\end{array} & (i \neq j) \end{array} \tag{52}$$

Three kinds of operations in (51) are called elementary row operations, and three kinds of operations in (52) are called elementary column operations. Collectively, they are called elementary operations.

The following are examples of elementary operations on a matrix of order 3. Here, $\textcircled{i}$ denotes the i-th row, and $\boxed{j}$ denotes the j-th column.

$$P_3(2,3)A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 7 & 8 & 9 \\ 4 & 5 & 6 \end{pmatrix} \qquad \textcircled{2} \leftrightarrow \textcircled{3}$$
$$AP_3(1,3) = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 3 & 2 & 1 \\ 6 & 5 & 4 \\ 9 & 8 & 7 \end{pmatrix} \qquad \boxed{1} \leftrightarrow \boxed{3}$$
$$Q_3(2,c)A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 4c & 5c & 6c \\ 7 & 8 & 9 \end{pmatrix} \qquad c\,\textcircled{2}$$
$$AQ_3(3,c) = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & c \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3c \\ 4 & 5 & 6c \\ 7 & 8 & 9c \end{pmatrix} \qquad c\boxed{3}$$
$$R_3(1,3;c)A = \begin{pmatrix} 1 & 0 & c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = \begin{pmatrix} 1+7c & 2+8c & 3+9c \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \qquad \textcircled{1} + c\,\textcircled{3}$$
$$AR_3(3,2;c) = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & c & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2+3c & 3 \\ 4 & 5+6c & 6 \\ 7 & 8+9c & 9 \end{pmatrix} \qquad \boxed{2} + c\boxed{3} \tag{53}$$

(exercise09) Prove (51) and (52) using block matrices.

5.8. **Nonsingularity of elementary matrices.** Every elementary matrix is nonsingular, and its inverse is also an elementary matrix. The following show this fact.

$$\begin{aligned} &P_n(i,j)P_n(i,j) = E_n. \qquad \therefore\ (P_n(i,j))^{-1} = P_n(i,j). \\ &Q_n(i,c)Q_n(i,c^{-1}) = Q_n(i,c^{-1})Q_n(i,c) = E_n. \\ &\qquad\qquad \therefore\ (Q_n(i,c))^{-1} = Q_n(i,c^{-1}). \\ &R_n(i,j;c)R_n(i,j;-c) = R_n(i,j;-c)R_n(i,j;c) = E_n. \\ &\qquad\qquad \therefore\ (R_n(i,j;c))^{-1} = R_n(i,j;-c). \end{aligned} \tag{54}$$

These formulas can be shown by calculation, however, there is another way: the multiplication: $P_n(i,j)P_n(i,j)A$ first interchanges the i-th row and the j-th row of A, next interchanges the i-th row and the j-th row again, then returns it to the original A.

$$\therefore\ P_n(i,j)P_n(i,j)A = A. \qquad \therefore\ P_n(i,j)P_n(i,j) = E_n. \tag{55}$$

$Q_n(i,c^{-1})Q_n(i,c)A$ multiplies the i-th row by c, and multiplies it by c^{-1}, then returns it to the original. $R_n(i,j;-c)R_n(i,j;c)A$ adds the i-th row by the j-th row multiplied by c, and subtracts the i-th row by the j-th row multiplied by c, then returns it to the original. Consequently, (54) is proved.

Since the inverse of an elementary matrix is also an elementary matrix, the inverse operation of an elementary operation:

$$A \ \underset{P^{-1}\times}{\overset{P\times}{\rightleftarrows}}\ PA, \qquad A \ \underset{\times Q^{-1}}{\overset{\times Q}{\rightleftarrows}}\ AQ \tag{56}$$

is also an elementary operation. That is to say, elementary operations are invertible. In particular, the inverse of an elementary row operation is an elementary row operation, and the inverse of an elementary column operation is an elementary column operation.

Reversing the order of two operations does not change the result, then these operations are called commutative. In general, two elementary row operations, or two elementary column operations are not commutative, but an elementary row operation and an elementary column operation are commutative as below.

$$A \xrightarrow{P\times} PA \xrightarrow{\times Q} PAQ; \qquad A \xrightarrow{\times Q} AQ \xrightarrow{P\times} PAQ \tag{57}$$

5.9. **Ranks of matrices.** The rank of a matrix is defined by using elementary operations. Let A be an $m \times n$ matrix. Define the canonical form $F_{mn}(r)$ of a matrix as

$$F_{mn}(r) = \begin{pmatrix} E_r & O_{r,n-r} \\ O_{m-r,r} & O_{m-r,n-r} \end{pmatrix}. \tag{58}$$

This is an $m \times n$ matrix such that successive r diagonal entries from the upper left corner are 1, and the rest entries are 0. If A is transformed into $F_{mn}(r)$ by a sequence of elementary operations, say,

$$A \longrightarrow\longrightarrow \cdots \longrightarrow F_{mn}(r), \tag{59}$$

then $F_{mn}(r)$ is the canonical form of A, and r is called the rank of A, denoted by $r(A)$. Every matrix can be transformed into some canonical form. Note that $F_{mn}(r)$ has several forms depending on the values of m, n, r as follows.

$$\begin{aligned} &F_{mn}(m) = \begin{pmatrix} E_m & O_{m,n-m} \end{pmatrix} &\qquad& F_{mn}(n) = \begin{pmatrix} E_n \\ O_{m-n,n} \end{pmatrix} \\ &F_{nn}(n) = E_n && F_{mn}(0) = O_{mn} \end{aligned} \tag{60}$$

The following is a typical algorithm to find the rank of a matrix.

0: Given a matrix O_{mn}, then it is already $F_{mn}(0)$.
1: Given a matrix $A \neq O$, if its $(1,1)$ entry is not equal to 1, then make it equal to 1 by elementary operations.
2: Subtract a scalar multiple of the first row from every $i > 1$ th row to make every entry below the $(1,1)$ entry equal to 0. Then we have e_1 in the first column. Next subtract a scalar multiple of the first column from every $j > 1$ th column to make every entry on the right-hand side of the $(1,1)$ entry equal to 0. This technique is called the sweeping-out method, sweeping the first column and row.
3: If the $(2,2)$ entry is not equal to 1, then make it equal to 1 by elementary operations with respect to the second and subsequent rows and columns.
4: Subtract a scalar multiple of the second row from every $i > 2$ th row to make every entry below the $(2,2)$ entry equal to 0. Then we have e_2 in the second column. Next subtract a scalar multiple of the second column from every $j > 2$ th column to make every entry on the right-hand side of the $(2,2)$ entry equal to 0. The second column and row have been swept.
5: Iterating sweeping, we have a canonical form.

Even if we use this algorithm, there are many ways from a matrix to the canonical form. Also, it is better to avoid fractions or large magnitude numbers as much as possible. The following is an example of a sequence of elementary operations to determine the rank of a matrix.

$$\begin{aligned} A &= \begin{pmatrix} 2 & 3 & -1 & -4 \\ 5 & 2 & 1 & -3 \\ 4 & -5 & 5 & 6 \end{pmatrix} \longrightarrow \begin{pmatrix} 2 & 3 & 1 & -4 \\ 5 & 2 & -1 & -3 \\ 4 & -5 & -5 & 6 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 3 & 2 & -4 \\ -1 & 2 & 5 & -3 \\ -5 & -5 & 4 & 6 \end{pmatrix} \\ &\longrightarrow \begin{pmatrix} 1 & 3 & 2 & -4 \\ 0 & 5 & 7 & -7 \\ 0 & 10 & 14 & -14 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 5 & 7 & -7 \\ 0 & 10 & 14 & -14 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 7 & -7 \\ 0 & 2 & 14 & -14 \end{pmatrix} \\ &\longrightarrow \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 7 & -7 \\ 0 & 0 & 0 & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = F_{34}(2). \qquad \therefore \ r(A) = 2. \end{aligned} \tag{61}$$

The following theorem assures that the rank of a matrix is well-defined.

Theorem 4. *The canonical form of a matrix A is uniquely determined by A itself, independent of the way of transformation of A.*

Proof. Suppose an $m \times n$ matrix A is transformed into two canonical forms $F_{mn}(r)$ and $F_{mn}(s)$ $(r \leq s)$ by two sequences of elementary operations.

$$\begin{aligned} A \longrightarrow\longrightarrow \cdots \longrightarrow F_{mn}(r) \\ A \longrightarrow\longrightarrow \cdots \longrightarrow F_{mn}(s) \end{aligned} \tag{62}$$

Since elementary operations are invertible, starting with $F_{mn}(r)$ we have

$$F_{mn}(r) \longrightarrow \cdots \longrightarrow F_{mn}(s). \tag{63}$$

Elementary operations are given by multiplying by elementary matrices on the left or right. Thus letting P_i and Q_i be elementary matrices, we have

$$P_k \dots P_2 P_1 F_{mn}(r) Q_1 Q_2 \dots Q_l = F_{mn}(s). \tag{64}$$

Letting $P_k \dots P_2 P_1 = P$ and $Q_1 Q_2 \dots Q_l = Q$, we have

$$P F_{mn}(r) Q = F_{mn}(s). \tag{65}$$

Then letting P and Q be symmetrically partitioned,

$$\begin{aligned} \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix} \begin{pmatrix} E_r & O \\ O & O \end{pmatrix} \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} &= \begin{pmatrix} P_{11}Q_{11} & P_{11}Q_{12} \\ P_{21}Q_{11} & P_{21}Q_{12} \end{pmatrix} \\ &= \begin{pmatrix} E_s & O \\ O & O \end{pmatrix}. \end{aligned} \tag{66}$$

Since $r \leq s$, $P_{11}Q_{11} = E_r$. Hence by Theorem 3, P_{11} and Q_{11} are nonsingular. Thus from $P_{11}Q_{12} = O$ it follows that $Q_{12} = O$. Therefore $P_{21}Q_{12} = O$, which implies that $r = s$. □

5.10. **Rank and nonsingularity.** Nonsingularity of a square matrix is determined by its rank.

Theorem 5. *For a matrix A of order n,*

$$A \text{ is nonsingular} \iff r(A) = n. \tag{67}$$

Proof. $(\Rightarrow)$ By reduction to absurdity. Let A be a nonsingular matrix of order n. Suppose $r(A) = r < n$. By applying elementary operations on A, we have

$$A \longrightarrow\longrightarrow \cdots \longrightarrow F_{nn}(r). \tag{68}$$

Hence multiplying A by elementary matrices P_i and Q_i on both sides, we have

$$P_k \dots P_2 P_1 A Q_1 Q_2 \dots Q_l = PAQ = F_{nn}(r). \tag{69}$$

Since the product of several elementary matrices is nonsingular, P and Q are nonsingular. Also, as A is nonsingular, the product PAQ is nonsingular. Here letting $(PAQ)^{-1} = X$,

$$E_n = X(PAQ) = XF_{nn}(r) = \begin{pmatrix} * & \cdots & * & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ * & \cdots & * & 0 & \cdots & 0 \end{pmatrix}, \tag{70}$$

which is a contradiction. □

Proof. ($\Leftarrow$) Let A be a matrix of order n and of rank n. Similarly to the above proof, by nonsingular matrices P and Q, we have

$$PAQ = E_n. \qquad \therefore \quad A = P^{-1}Q^{-1}. \tag{71}$$

Hence A is the product of nonsingular matrices, and therefore A is nonsingular. □

Reviewing (71), P and Q are products of several elementary matrices, and their inverses P^{-1} and Q^{-1} are also products of several elementary matrices. Consequently, A is represented as a product of several elementary matrices. Therefore if $r(A) = n$, then A is a product of elementary matrices, and conversely, if A is a product of elementary matrices, then A is nonsingular and by Theorem 5, we have $r(A) = n$. Accordingly,

$$r(A) = n \iff A \text{ is a product of several elementary matrices} \tag{72}$$

In addition, slightly modifying (71), we have

$$QPA = E_n, \qquad AQP = E_n. \tag{73}$$

The first equation shows that A is transformed into E_n by only elementary row operations, and the second one shows that A is transformed into E_n by only elementary column operations. Summarizing those results, we have the following.

Theorem 5$^+$. *For a matrix A of order n, five conditions below are equivalent.*

(i) A is nonsingular. *(ii) $r(A) = n$.*
(iii) A is represented as a product of several elementary matrices.
(iv) A is transformed into E_n by only elementary row operations. (74)
(v) A is transformed into E_n by only elementary column operations.

We can determine the inverse matrix of a nonsingular matrix using the fact that every nonsingular matrix is transformed into the identity matrix by only elementary row operations. Let A be a nonsingular matrix of order n, and consider an $n \times 2n$ matrix $\begin{pmatrix} A & E_n \end{pmatrix}$. If it is transformed into a matrix $\begin{pmatrix} E_n & B \end{pmatrix}$ by only elementary row operations, then we have $B = A^{-1}$. The reason why is that, as elementary row operations are to multiply by elementary matrices,

$$\begin{aligned} &P_k \ldots P_2 P_1 \begin{pmatrix} A & E_n \end{pmatrix} = P \begin{pmatrix} A & E_n \end{pmatrix} = \begin{pmatrix} PA & P \end{pmatrix} = \begin{pmatrix} E_n & B \end{pmatrix}. \\ &\therefore \quad PA = E_n. \qquad \therefore \quad P = A^{-1} = B. \end{aligned} \tag{75}$$

(exercise10) Determine the inverse of the matrix $A = \begin{pmatrix} 3 & 1 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & 1 \end{pmatrix}$.

(ans)

$$\begin{pmatrix} A & E \end{pmatrix} = \begin{pmatrix} 3 & 1 & 2 & 1 & 0 & 0 \\ 2 & -1 & 2 & 0 & 1 & 0 \\ 2 & 2 & 1 & 0 & 0 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 2 & 0 & 1 & -1 & 0 \\ 2 & -1 & 2 & 0 & 1 & 0 \\ 2 & 2 & 1 & 0 & 0 & 1 \end{pmatrix} \longrightarrow$$

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

$$\begin{pmatrix} 1 & 2 & 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -2 & 1 & 2 \\ 0 & -2 & 1 & -2 & 2 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 & 5 & -3 & -4 \\ 0 & 1 & 0 & -2 & 1 & 2 \\ 0 & 0 & 1 & -6 & 4 & 5 \end{pmatrix}.$$

$$\therefore \quad A^{-1} = \begin{pmatrix} 5 & -3 & -4 \\ -2 & 1 & 2 \\ -6 & 4 & 5 \end{pmatrix}. \tag{76}$$

(note) If $\begin{pmatrix} A \\ E_n \end{pmatrix}$ is transformed into $\begin{pmatrix} E_n \\ B \end{pmatrix}$ by only elementary column operations, then $B = A^{-1}$. (Why?)

5.11. **Proofs of Theorems 1 and 3.** In this section, we give proofs of Theorem 1 in Section 5.3 and Theorem 3 in Section 5.5.

Proof of Theorem 1. (⇒) Clear by (23). □

Proof. (⇐) Suppose a matrix F of order n satisfies that $FX = XF$ for every matrix X of order n. Choose and fix a pair (i, j), and let X be a matrix of order n with a single nonzero entry, 1, at the (i, j)-th position. Letting $F = (f_{ij})$,

$$FX = \begin{pmatrix} 0 & \dots & 0 & f_{1i} & 0 & \dots & 0 \\ 0 & \dots & 0 & f_{2i} & 0 & \dots & 0 \\ \multicolumn{7}{c}{\dots\dots\dots\dots\dots} \\ 0 & \dots & 0 & f_{ni} & 0 & \dots & 0 \end{pmatrix} \;(\text{column } j) = i) \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ \multicolumn{5}{c}{\dots\dots\dots\dots\dots} \\ 0 & 0 & \dots & 0 & 0 \\ f_{j1} & f_{j2} & \dots & f_{j,n-1} & f_{jn} \\ 0 & 0 & \dots & 0 & 0 \\ \multicolumn{5}{c}{\dots\dots\dots\dots\dots} \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix} = XF. \tag{77}$$

Hence every entry except (i, j) of both sides is equal to 0. Therefore

$$\begin{aligned} f_{1i} = f_{2i} = \cdots = f_{i-1,i} = f_{i+1,i} = f_{i+2,i} = \cdots = f_{ni} = 0 \\ f_{j1} = f_{j2} = \cdots = f_{j,j-1} = f_{j,j+1} = f_{j,j+2} = \cdots = f_{jn} = 0. \end{aligned} \tag{78}$$

Comparing the (i, j) entries of both sides, we have

$$f_{ii} = f_{jj}. \tag{79}$$

Taking various (i, j), it is seen that F has nonzero entries only on the diagonal, and all entries on the diagonal have the same value. Therefore F is a scalar matrix. □

Proof of Theorem 3. By induction on n. Obviously, the proposition holds for $n = 1$. Suppose it holds for $n-1$. Let A and B be matrices of order n satisfying that $AB = E_n$. Then it is clear that A is transformed as follows by elementary operations.

$$A \longrightarrow\longrightarrow \cdots \longrightarrow \begin{pmatrix} 1 & {}^t\mathsf{0} \\ \mathsf{0} & A_1 \end{pmatrix} \tag{80}$$

Here, A_1 is a matrix of order $n-1$. Hence letting P and Q be the products of several elementary matrices, we have $PAQ = \begin{pmatrix} 1 & {}^t\mathsf{0} \\ \mathsf{0} & A_1 \end{pmatrix}$. Since $AB = E_n$,

$$(PAQ)(Q^{-1}BP^{-1}) = P(AQQ^{-1}B)P^{-1} = P(AB)P^{-1} = PE_nP^{-1} = E_n. \tag{81}$$

Letting $Q^{-1}BP^{-1} = R$, we have $(PAQ)R = E_n$. Using block matrices,

$$\begin{pmatrix} 1 & {}^t\mathsf{0} \\ \mathsf{0} & A_1 \end{pmatrix}\begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix} = \begin{pmatrix} 1 & {}^t\mathsf{0} \\ \mathsf{0} & E_{n-1} \end{pmatrix}. \tag{82}$$

Thus we have $A_1R_{22} = E_{n-1}$, and therefore, by the induction hypothesis, A_1 is non-singular and $R_{22} = A_1^{-1}$. Hence

$$\begin{pmatrix} 1 & {}^t\mathsf{0} \\ \mathsf{0} & R_{22} \end{pmatrix}\begin{pmatrix} 1 & {}^t\mathsf{0} \\ \mathsf{0} & A_1 \end{pmatrix} = \begin{pmatrix} 1 & {}^t\mathsf{0} \\ \mathsf{0} & E_{n-1} \end{pmatrix} = E_n. \tag{83}$$

Letting $\begin{pmatrix} 1 & {}^t\mathsf{0} \\ \mathsf{0} & R_{22} \end{pmatrix} = \tilde{R}$, we have $\tilde{R}(PAQ) = E_n$.

$$\therefore\ Q\tilde{R}(PAQ)Q^{-1} = E_n. \qquad \therefore\ (Q\tilde{R}P)A = E_n. \tag{84}$$

Consequently, there exists a matrix B' such that $B'A = E_n$. Since

$$B = E_nB = (B'A)B = B'(AB) = B'E_n = B', \tag{85}$$

we have $BA = E_n$. The induction is completed. □

5.12. **Inner products of vectors.** For a complex number $z = x + yi$, $\overline{z} = x - yi$ is called the complex conjugate of z. Denote by $|z|$ the absolute value (modulus, magnitude) of z, defined by $|z| = \sqrt{x^2 + y^2}$. Let w be another complex number, then we have the following.

$$\begin{gathered} \overline{z+w} = \overline{z} + \overline{w} \qquad \overline{zw} = \overline{z}\,\overline{w} \\ |z+w| \leq |z| + |w| \qquad |zw| = |z| \cdot |w| \\ z\,\overline{z} = |z|^2 \end{gathered} \tag{86}$$

In general, for a complex matrix $A = (a_{ij})$, the complex matrix with complex conjugate entries is denoted by $\overline{A}$, that is,

$$\overline{A} = (\overline{a_{ij}}). \tag{87}$$

Inner products of column vectors with n entries are defined in a similar manner to 3-dimensional vectors. For two complex column vectors with n entries $\mathtt{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$,

$\mathtt{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$, define the inner product of x and y by

$$(\mathtt{x}, \mathtt{y}) = {}^t\mathtt{x}\overline{\mathtt{y}} = x_1\overline{y_1} + \cdots + x_n\overline{y_n}. \tag{88}$$

In particular, for real vectors, we have

$$(\mathtt{x}, \mathtt{y}) = {}^t\mathtt{x}\mathtt{y} = x_1 y_1 + \cdots + x_n y_n. \tag{89}$$

To distinguish from inner products of real vectors, inner products of complex vectors are sometimes called Hermitian products.

Let c be a complex number, inner product satisfies the following.

$$\begin{array}{c} (\mathtt{x}, \mathtt{y}) = \overline{(\mathtt{y}, \mathtt{x})} \\ (\mathtt{x}, \mathtt{y}+\mathtt{z}) = (\mathtt{x}, \mathtt{y}) + (\mathtt{x}, \mathtt{z}) \qquad (\mathtt{x}+\mathtt{y}, \mathtt{z}) = (\mathtt{x}, \mathtt{z}) + (\mathtt{y}, \mathtt{z}) \\ (c\mathtt{x}, \mathtt{y}) = c(\mathtt{x}, \mathtt{y}) \qquad (\mathtt{x}, c\mathtt{y}) = \overline{c}(\mathtt{x}, \mathtt{y}) \\ (\mathtt{x}, \mathtt{x}) \geq 0 \quad \text{and} \quad (\mathtt{x}, \mathtt{x}) = 0 \iff \mathtt{x} = \mathtt{0} \end{array} \tag{90}$$

The property of the first three lines is called conjugate linearity. The property of the fourth line is called positivity. From positivity it follows that $\sqrt{(\mathtt{x}, \mathtt{x})}$ is always a nonnegative real number. Then write

$$\sqrt{(\mathtt{x}, \mathtt{x})} = ||\mathtt{x}|| \tag{91}$$

and is called the magnitude or norm of x. The following holds.

$$\begin{array}{ll} ||c\mathtt{x}|| = |c| \cdot ||\mathtt{x}|| & \\ |(\mathtt{x}, \mathtt{y})| \leq ||\mathtt{x}|| \cdot ||\mathtt{y}|| & \text{(Cauchy–Schwarz inequality)} \\ ||\mathtt{x}+\mathtt{y}|| \leq ||\mathtt{x}|| + ||\mathtt{y}|| & \text{(Triangle inequality)} \end{array} \tag{92}$$

If two vectors x and *tty* satisfy that $(\mathtt{x}, \mathtt{y}) = 0$, then x and y are perpendicular to each other. In addition, if there exists a scalar c such that $\mathtt{x} = c\mathtt{y}$ or $\mathtt{y} = c\mathtt{x}$, then x and y are parallel to each other.

A vector of magnitude 1 is called a unit vector. For a vector x, $\pm||\mathtt{x}||^{-1}\mathtt{x}$ is a unit vector which is parallel to x. In the case of complex vectors, $\omega||\mathtt{x}||^{-1}\mathtt{x}$ is a unit vector which is parallel to x for every complex number ω of absolute value 1.

(exercise11) (1) Show (90) and (92). (2) Find real unit vectors which are perpendicular to all of $\begin{pmatrix} 2 \\ 1 \\ 1 \\ -1 \end{pmatrix}$, $\begin{pmatrix} 1 \\ 2 \\ 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 2 \\ 1 \\ 3 \end{pmatrix}$.

(ans) (1) We show the second and third inequalities of (92).

Proof of Cauchy–Schwarz inequality. If $\mathtt{y} = \mathtt{0}$, then both sides are equal to 0. Hence suppose $\mathtt{y} \neq \mathtt{0}$.

$$\begin{aligned} 0 &\leq || \, ||\mathtt{y}||^2\mathtt{x} - (\mathtt{x}, \mathtt{y})\mathtt{y}||^2 \\ &= ||\mathtt{y}||^4||\mathtt{x}||^2 - ||\mathtt{y}||^2\overline{(\mathtt{x}, \mathtt{y})}(\mathtt{x}, \mathtt{y}) - (\mathtt{x}, \mathtt{y})||\mathtt{y}||^2(\mathtt{y}, \mathtt{x}) + (\mathtt{x}, \mathtt{y})\overline{(\mathtt{x}, \mathtt{y})}||\mathtt{y}||^2 \\ &= ||\mathtt{y}||^2(||\mathtt{x}||^2||\mathtt{y}||^2 - |(\mathtt{x}, \mathtt{y})|^2). \end{aligned} \tag{93}$$

$\therefore \quad |(\mathtt{x}, \mathtt{y})| \leq ||\mathtt{x}|| \cdot ||\mathtt{y}||. \quad \square$

Proof of Triangle inequality. It suffices to show that the square of the right-hand side subtracted by the square of the left-hand side is greater or equal to 0.

$$\begin{aligned}&(||\mathtt{x}|| + ||\mathtt{y}||)^2 - ||\mathtt{x}+\mathtt{y}||^2\\ &= ||\mathtt{x}||^2 + 2||\mathtt{x}||\cdot||\mathtt{y}|| + ||\mathtt{y}||^2 - ||\mathtt{x}||^2 - (\mathtt{x},\mathtt{y}) - (\mathtt{y},\mathtt{x}) - ||\mathtt{y}||^2\\ &= 2||\mathtt{x}||\cdot||\mathtt{y}|| - (\mathtt{x},\mathtt{y}) - \overline{(\mathtt{x},\mathtt{y})}\\ &\geq 2||\mathtt{x}||\cdot||\mathtt{y}|| - 2|(\mathtt{x},\mathtt{y})| \geq 0 \qquad \text{(By Cauchy–Schwarz inequality)} \quad \square\end{aligned} \tag{94}$$

(2) Shown the result only: $\pm\frac{1}{\sqrt{14}}\begin{pmatrix} 2\\ 0\\ -3\\ 1\end{pmatrix}$.

5.13. **Various matrices.** In general, for an $m\times n$ complex matrix A, an $n\times m$ complex matrix $\overline{{}^tA}$ is denoted by A^*, called the Hermitian adjoint (Hermitian transpose, Hermitian conjugate) of A., denoted by A^*. Transpose or Hermitian adjoint satisfies the following. Here, in the last line, A is assumed to be nonsingular.

$$\begin{array}{ccc} {}^t({}^tA) = A & {}^t(A+B) = {}^tA + {}^tB & {}^t(AB) = {}^tB\,{}^tA\\ (A^*)^* = A & (A+B)^* = A^* + B^* & (AB)^* = B^*A^*\\ {}^t(kA) = k\,{}^tA & & (kA)^* = \overline{k}A^*\\ ({}^tA)^{-1} = {}^t\left(A^{-1}\right) & & (A^*)^{-1} = \left(A^{-1}\right)^* \end{array} \tag{95}$$

(exercise12) (1) Prove the above equalities. (From the fourth line, it follows that A is nonsingular $\iff$ ${}^tA$ is nonsingular $\iff$ A^* is nonsingular.)
(2) Show that $r(A) = r({}^tA) = r(A^*)$.

Let A be a matrix of order n. Several names of matrices are given if some condition is satisfied.

condition	complex/real	name
$AA^* = A^*A$	complex	normal
$A^* = A$	complex	Hermitian
$A^* = -A$	complex	skew-Hermitian
${}^tA = A$	complex	complex symmetric
${}^tA = A$	real	real symmetric
${}^tA = -A$	complex	complex skew-symmetric
${}^tA = -A$	real	real skew-symmetric
$AA^* = A^*A = E$	complex	unitary
$A\,{}^tA = {}^tAA = E$	real	orthogonal
$i > j \Rightarrow a_{ij} = 0$	complex	upper triangular
$i < j \Rightarrow a_{ij} = 0$	complex	lower triangular

(exercise13) (1) Prove that the diagonal entries of a Hermite matrix is real. (2) Prove that the diagonal entries of a skew-symmetric matrix is 0. (3) Prove that Hermitian, skew-Hermitian, unitary and real skew-symmetric matrices are normal. (4) Show the following.

$$A \text{ is Hermitian} \iff iA \text{ is skew-Hermitian} \tag{96}$$

(5) Show that the product of upper (respectively, lower) triangular matrices of order n is also upper (respectively, lower) triangular.

$\bigstar$

6. SOLUTIONS TO SYSTEMS OF LINEAR EQUATIONS

$\star$ 3 $\star$

KEYWORDS: SYSTEMS OF LINEAR EQUATIONS, GENERAL SOLUTIONS, HOMOGENEOUS SYSTEMS OF LINEAR EQUATIONS, ELEMENTARY ROW OPERATIONS, COLUMN TRANSPOSITION

6.1. **Systems of linear equations.** A system of linear equations with variables $x_1, x_2, \ldots, x_n$ has the following form:

$$\left\{\begin{array}{ccccccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & a_{13}x_3 & + & \cdots & + & a_{1n}x_n & = & c_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & a_{23}x_3 & + & \cdots & + & a_{2n}x_n & = & c_2 \\ a_{31}x_1 & + & a_{32}x_2 & + & a_{33}x_3 & + & \cdots & + & a_{3n}x_n & = & c_3 \\ \multicolumn{11}{c}{\dotfill} \\ a_{m1}x_1 & + & a_{m2}x_2 & + & a_{m3}x_3 & + & \cdots & + & a_{mn}x_n & = & c_m \end{array}\right. \tag{1}$$

Let us solve this system of equations. Here, let us denote

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \multicolumn{4}{c}{\dotfill} \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}, \quad \mathsf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \mathsf{c} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{pmatrix}. \tag{2}$$

They are called the coefficient matrix, the variable vector and the constant vector, respectively. Then (1) is rewritten simply as

$$A\mathsf{x} = \mathsf{c}. \tag{3}$$

Hence, it suffices to solve this equation. To begin with, multiplying both sides of (3) by a nonsingular matrix P on the left, we have

$$PA\mathsf{x} = P\mathsf{c}. \tag{4}$$

However, multiplying again both sides by P^{-1} on the left, we retrieve (3). Therefore (3) and (4) are equivalent to each other. Namely,

$$A\mathsf{x} = \mathsf{c} \quad \underset{P^{-1}}{\overset{P}{\rightleftarrows}} \quad PA\mathsf{x} = P\mathsf{c}. \tag{5}$$

Thus as a result of transformation of (3), we hope to simplify (4) as much as possible. We use elementary row operations for this purpose. A nonsingular matrix P is expressed by the product of several elementary matrices as $P = P_1P_2\ldots P_s$. Hence, to multiply by P on the left is to repeat elementary row operations. Accordingly, consider the extended coefficient matrix $\widetilde{A} = \begin{pmatrix} A & \mathtt{c} \end{pmatrix}$ and perform elementary row operations on $\widetilde{A}$ to have

$$\begin{pmatrix} A & \mathtt{c} \end{pmatrix} \quad \underset{\longleftarrow \cdots \longleftarrow \atop P^{-1}}{\overset{P \atop \longrightarrow \cdots \longrightarrow}{}} \quad \begin{pmatrix} PA & P\mathtt{c} \end{pmatrix}. \tag{6}$$

Consequently we have $\begin{pmatrix} PA & P\mathtt{c} \end{pmatrix}$ (and we can go back). Therefore we can obtain the system (4) by performing elementary row operations on $\widetilde{A}$.

For convenience to solve (4), suppose $\begin{pmatrix} PA & P\mathtt{c} \end{pmatrix}$ is transformed into the following form:

$$\begin{pmatrix} PA & P\mathtt{c} \end{pmatrix} = \begin{pmatrix} E_r & B & \mathtt{d}_1 \\ & O & \mathtt{d}_2 \end{pmatrix}. \tag{7}$$

Then (4) has the form:

$$\begin{pmatrix} E_r & B \\ & O \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_r \\ x_{r+1} \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \mathtt{d}_1 \\ \mathtt{d}_2 \end{pmatrix}. \tag{8}$$

A simple block matrix calculation shows that

$$\begin{pmatrix} E_r \begin{pmatrix} x_1 \\ \vdots \\ x_r \end{pmatrix} + B \begin{pmatrix} x_{r+1} \\ \vdots \\ x_n \end{pmatrix} \\ \mathtt{0} \end{pmatrix} = \begin{pmatrix} \mathtt{d}_1 \\ \mathtt{d}_2 \end{pmatrix}. \tag{9}$$

The lower half of this equation, we have the condition for (3) to have a solution: $\mathtt{d}_2 = \mathtt{0}$. If $\mathtt{d}_2 \neq \mathtt{0}$, (3) has no solution.

Thus, suppose that $\mathtt{d}_2 = \mathtt{0}$ and consider the upper half of (9), then we have

$$\begin{pmatrix} x_1 \\ \vdots \\ x_r \end{pmatrix} = \mathtt{d}_1 - B \begin{pmatrix} x_{r+1} \\ \vdots \\ x_n \end{pmatrix}. \tag{10}$$

In order to solve this, taking arbitrary numbers α_i $(i = 1, \ldots, n-r)$, and let

$$\begin{pmatrix} x_{r+1} \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_{n-r} \end{pmatrix}. \tag{11}$$

Then from (10) it follows that

$$\begin{pmatrix} x_1 \\ \vdots \\ x_r \end{pmatrix} = \mathtt{d}_1 - B \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_{n-r} \end{pmatrix} = \mathtt{d}_1 - \begin{pmatrix} \mathtt{b}_1 & \mathtt{b}_2 & \ldots & \mathtt{b}_{n-r} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_{n-r} \end{pmatrix}. \tag{12}$$

Here, we set $B = \begin{pmatrix} \mathtt{b}_1 & \mathtt{b}_2 & \ldots & \mathtt{b}_{n-r} \end{pmatrix}$. Therefore

$$\begin{pmatrix} x_1 \\ \vdots \\ x_r \end{pmatrix} = \mathtt{d}_1 - \alpha_1 \mathtt{b}_1 - \alpha_2 \mathtt{b}_2 - \cdots - \alpha_{n-r}\mathtt{b}_{n-r}. \tag{13}$$

Consequently, (9) is solved. Since (9) is equivalent to (3), (3) is solved after all.

Noticing that (3) is actually solved under the condition that $\mathtt{d}_2 = \mathtt{0}$, it is a necessary and sufficient condition that (3) is solvable. In addition, by (7), we have

$$\mathtt{d}_2 = \mathtt{0} \iff r(A) = r\left(\widetilde{A}\right). \tag{14}$$

Hence $r(A) = r\left(\widetilde{A}\right)$ is a necessary and sufficient condition that (3) is solvable. Summarizing (11) and (13), we have the following formula.

Theorem 1. *Let A be an $m \times n$ matrix, and suppose $r(A) = r\left(\widetilde{A}\right) = r$. Then the general solution (an expression which describes all possible solutions) to* (3) *is given by*

$$\mathtt{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_r \\ x_{r+1} \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \mathtt{d}_1 \\ \mathtt{0} \end{pmatrix} + \alpha_1 \begin{pmatrix} -\mathtt{b}_1 \\ \mathtt{e}_1 \end{pmatrix} + \alpha_2 \begin{pmatrix} -\mathtt{b}_2 \\ \mathtt{e}_2 \end{pmatrix} + \cdots + \alpha_{n-r} \begin{pmatrix} -\mathtt{b}_{n-r} \\ \mathtt{e}_{n-r} \end{pmatrix}. \tag{15}$$

Here, $\alpha_1, \ldots, \alpha_{n-r}$ are arbitrary constants.

The arbitrary constants $\alpha_1, \ldots, \alpha_{n-r}$ in (15) are called parameters of this solution. The number of the parameters are $n - r =$ (the number of variables) $- r(A)$. If (15) contains at least one parameter, there are infinitely many solutions, and if it contains no parameters, there is a unique solution.

(note1) If the number of linear equations $= r(A)$, say, $m = r$, then the right-hand side of (7) has no zero rows, which is expressed in the form:

$$\begin{pmatrix} E_r & B & \mathtt{d}_1 \end{pmatrix}. \tag{16}$$

In this case, there is no $\mathtt{d}_2$, and therefore solutions exist, and is represented as (15).

(note2) A further special case is $n = r$ or $m = n = r$. Then the right-hand side of (7) has no B block, which is expressed in the form:

$$\begin{pmatrix} E_n & \mathtt{d}_1 \\ O & \mathtt{d}_2 \end{pmatrix} \qquad \text{or} \qquad \begin{pmatrix} E_n & \mathtt{d}_1 \end{pmatrix} \tag{17}$$

In the left case, there exists a solution subject to $\mathtt{d}_2 = \mathtt{0}$, while in the right case, there exists always a solution. Anyway, it is a unique solution expressed as $\mathtt{x} = \mathtt{d}_1$, because (17) corresponds to the equation $E_n\mathtt{x} = \mathtt{d}_1$, say, $\mathtt{x} = \mathtt{d}_1$. This case, where the solution is unique, corresponds to the case where (15) has no parameters.

6.2. **Homogeneous systems of linear equations.** A system of linear equations (1) with zero constant terms $c_i = 0$ $(i = 1, \ldots, m)$ is called a homogeneous system of linear equations (corresponding to the system (1)). A system which is not homogeneous is called inhomogeneous. A homogeneous system of linear equations is written with matrices as

$$A\mathtt{x} = \mathtt{0}. \tag{18}$$

When we solve this system, since $\widetilde{A}$ has the right-most column $\mathtt{0}$, after elementary row operations, the right-most column remains $\mathtt{0}$, and therefore $\mathtt{d}_1 = \mathtt{0}$ and $\mathtt{d}_2 = \mathtt{0}$. Hence (18) always has a solution, and its general solution is expressed as (15) without the first term. In particular, it has a solution $\mathtt{x} = \mathtt{0}$, called a trivial solution. In the solving process, it suffices to perform elementary row operations on a coefficient matrix A, instead of $\widetilde{A}$, to get the form $\begin{pmatrix} E_r & B \\ & O \end{pmatrix}$. Here, if A is transformed into $\begin{pmatrix} E_n \\ O \end{pmatrix}$ or E_n by elementary row operations, then by the above-mentioned (note2), there is a unique solution $\mathtt{x} = \mathtt{0}$. Otherwise, non-trivial solutions exist.

Theorem 2. *Let A be an $m \times n$ matrix, then*

$$r(A) < n \iff (18) \text{ has non-trivial solutions.} \tag{19}$$

If $m < n$, (18) *always has non-trivial solutions. In particular, if A is a matrix of order n, then*

$$A \text{ is singular} \iff (18) \text{ has non-trivial solutions.} \tag{20}$$

There is the following relationship between the solutions to (3) and (18).

Theorem 3. *The general solution to* (3) *is represented as the sum of a solution to* (3) *and the general solution to* (18)*.*

Proof. Indeed, the first term of (15) is a solution to (3), and the rest sum including parameters is the general solution to (18). Also, we can prove this theorem without (15), let $\mathtt{x}_0$ be a solution to (3), say, $A\mathtt{x}_0 = \mathtt{c}$, we have

$$A\mathtt{y} = \mathtt{0} \iff A(\mathtt{x}_0 + \mathtt{y}) = \mathtt{c}, \tag{21}$$

which proves the theorem. □

(exercise01) Solve the system of linear equations:

$$\begin{cases} 2x & +y & +z & & = & 5 \\ & -2y & -z & +w & = & -5 \\ x & & +2z & -w & = & 3 \end{cases} \tag{22}$$

(answer) Performing elementary row operations on $\widetilde{A}$, we have the following.

$$\begin{aligned} \widetilde{A} = \begin{pmatrix} 2 & 1 & 1 & 0 & 5 \\ 0 & -2 & -1 & 1 & -5 \\ 1 & 0 & 2 & -1 & 3 \end{pmatrix} &\longrightarrow \begin{pmatrix} 1 & 0 & 2 & -1 & 3 \\ 0 & -2 & -1 & 1 & -5 \\ 2 & 1 & 1 & 0 & 5 \end{pmatrix} \longrightarrow \\ \begin{pmatrix} 1 & 0 & 2 & -1 & 3 \\ 0 & -2 & -1 & 1 & -5 \\ 0 & 1 & -3 & 2 & -1 \end{pmatrix} &\longrightarrow \begin{pmatrix} 1 & 0 & 2 & -1 & 3 \\ 0 & 1 & -3 & 2 & -1 \\ 0 & -2 & -1 & 1 & -5 \end{pmatrix} \longrightarrow \\ \begin{pmatrix} 1 & 0 & 2 & -1 & 3 \\ 0 & 1 & -3 & 2 & -1 \\ 0 & 0 & -7 & 5 & -7 \end{pmatrix} &\longrightarrow \begin{pmatrix} 1 & 0 & 2 & -1 & 3 \\ 0 & 1 & -3 & 2 & -1 \\ 0 & 0 & 1 & -\frac{5}{7} & 1 \end{pmatrix} \longrightarrow \\ \begin{pmatrix} 1 & 0 & 0 & \frac{3}{7} & 1 \\ 0 & 1 & 0 & -\frac{1}{7} & 2 \\ 0 & 0 & 1 & -\frac{5}{7} & 1 \end{pmatrix}&. \end{aligned} \tag{23}$$

$$\therefore \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 0 \end{pmatrix} + \alpha \begin{pmatrix} -\frac{3}{7} \\ \frac{1}{7} \\ \frac{5}{7} \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 0 \end{pmatrix} + \tilde{\alpha} \begin{pmatrix} -3 \\ 1 \\ 5 \\ 7 \end{pmatrix}. \qquad (\alpha = 7\tilde{\alpha})$$

(exercise02) Solve the system of linear equations:

$$\begin{cases} 3x & +y & -4z & +5w & = & 2 \\ 8x & +y & -9z & +15w & = & 12 \end{cases} \tag{24}$$

(answer)

$$\begin{aligned} \widetilde{A} = \begin{pmatrix} 3 & 1 & -4 & 5 & 2 \\ 8 & 1 & -9 & 15 & 12 \end{pmatrix} &\longrightarrow \begin{pmatrix} 3 & 1 & -4 & 5 & 2 \\ -1 & -2 & 3 & 0 & 6 \end{pmatrix} \\ \longrightarrow \begin{pmatrix} -1 & -2 & 3 & 0 & 6 \\ 3 & 1 & -4 & 5 & 2 \end{pmatrix} &\longrightarrow \begin{pmatrix} 1 & 2 & -3 & 0 & -6 \\ 3 & 1 & -4 & 5 & 2 \end{pmatrix} \\ \longrightarrow \begin{pmatrix} 1 & 2 & -3 & 0 & -6 \\ 0 & -5 & 5 & 5 & 20 \end{pmatrix} &\longrightarrow \begin{pmatrix} 1 & 2 & -3 & 0 & -6 \\ 0 & 1 & -1 & -1 & -4 \end{pmatrix} \\ \longrightarrow \begin{pmatrix} 1 & 0 & -1 & 2 & 2 \\ 0 & 1 & -1 & -1 & -4 \end{pmatrix}&. \end{aligned} \tag{25}$$

$$\therefore \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 2 \\ -4 \\ 0 \\ 0 \end{pmatrix} + \alpha \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} -2 \\ 1 \\ 0 \\ 1 \end{pmatrix}.$$

(exercise03) Determine a necessary and sufficient condition for the following system of linear equations to have a solution, and solve it under the condition. Here, express the solution using a, b only.

$$\begin{cases} \quad\ \ 2y \ -z = a \\ 3x \ +y \ +z = b \\ 2x \qquad +z = c \end{cases} \tag{26}$$

(answer)

$$\begin{aligned} \widetilde{A} = \begin{pmatrix} 0 & 2 & -1 & a \\ 3 & 1 & 1 & b \\ 2 & 0 & 1 & c \end{pmatrix} &\longrightarrow \begin{pmatrix} 0 & 2 & -1 & a \\ 1 & 1 & 0 & b-c \\ 2 & 0 & 1 & c \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 1 & 0 & b-c \\ 0 & 2 & -1 & a \\ 2 & 0 & 1 & c \end{pmatrix} \\ \longrightarrow \begin{pmatrix} 1 & 1 & 0 & b-c \\ 0 & 2 & -1 & a \\ 0 & -2 & 1 & -2b+3c \end{pmatrix} &\longrightarrow \begin{pmatrix} 1 & 1 & 0 & b-c \\ 0 & 2 & -1 & a \\ 0 & 0 & 0 & a-2b+3c \end{pmatrix} \\ \longrightarrow \begin{pmatrix} 1 & 1 & 0 & b-c \\ 0 & 1 & -\frac{1}{2} & \frac{a}{2} \\ 0 & 0 & 0 & a-2b+3c \end{pmatrix} &\longrightarrow \begin{pmatrix} 1 & 0 & \frac{1}{2} & -\frac{a}{2}+b-c \\ 0 & 1 & -\frac{1}{2} & \frac{a}{2} \\ 0 & 0 & 0 & a-2b+3c \end{pmatrix}. \end{aligned} \tag{27}$$

Therefore the solvable condition is $a - 2b + 3c = 0$. Then

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -\frac{a}{2}+b-c \\ \frac{a}{2} \\ 0 \end{pmatrix} + \alpha \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ 1 \end{pmatrix} = \begin{pmatrix} -\frac{a}{6}+\frac{b}{3} \\ \frac{a}{2} \\ 0 \end{pmatrix} + \tilde{\alpha} \begin{pmatrix} -1 \\ 1 \\ 2 \end{pmatrix}. \tag{28}$$

6.3. **Exceptions and column permutations.** The above method is applicable for the most cases, however, for some systems of linear equations, $\widetilde{A}$ can not be transformed into the right-hand side of (7) by only elementary row operations. In such cases, using additional operation, permutation of columns of $\widetilde{A}$ except the rightmost column, makes it possible to transform $\widetilde{A}$ into the right-hand side of (7). Here, since the columns of A correspond to variables, if we permute columns, then we also permute variables similarly. This method can be used for the ordinary case which is solvable with only row operations, but to avoid confusion, it should not be blindly used. The following example helps you to understand this method.

(exercise04) Solve the system of linear equations:

$$\begin{cases} 2x \ -y \ +4z \ +3w = 5 \\ \ \ x \ -y \ +z \ +w = 5 \\ \qquad -y \ -2z \ +3w = 1 \end{cases} \tag{29}$$

(answer)

$$\widetilde{A} = \begin{pmatrix} 2 & -1 & 4 & 3 & 5 \\ 1 & -1 & 1 & 1 & 5 \\ 0 & -1 & -2 & 3 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & -1 & 1 & 1 & 5 \\ 2 & -1 & 4 & 3 & 5 \\ 0 & -1 & -2 & 3 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & -1 & 1 & 1 & 5 \\ 0 & 1 & 2 & 1 & -5 \\ 0 & -1 & -2 & 3 & 1 \end{pmatrix}$$

$$\longrightarrow \begin{matrix} x & y & z & w & \\ \end{matrix} \begin{pmatrix} 1 & 0 & 3 & 2 & 0 \\ 0 & 1 & 2 & 1 & -5 \\ 0 & 0 & 0 & 4 & -4 \end{pmatrix} \longrightarrow \begin{matrix} x & y & w & z & \\ \end{matrix} \begin{pmatrix} 1 & 0 & 2 & 3 & 0 \\ 0 & 1 & 1 & 2 & -5 \\ 0 & 0 & 4 & 0 & -4 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 2 & 3 & 0 \\ 0 & 1 & 1 & 2 & -5 \\ 0 & 0 & 1 & 0 & -1 \end{pmatrix}$$

$$\longrightarrow \begin{pmatrix} 1 & 0 & 0 & 3 & 2 \\ 0 & 1 & 0 & 2 & -4 \\ 0 & 0 & 1 & 0 & -1 \end{pmatrix}.$$

$$\therefore \quad \begin{pmatrix} x \\ y \\ w \\ z \end{pmatrix} = \begin{pmatrix} 2 \\ -4 \\ -1 \\ 0 \end{pmatrix} + \alpha \begin{pmatrix} -3 \\ -2 \\ 0 \\ 1 \end{pmatrix}. \qquad \therefore \quad \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 2 \\ -4 \\ 0 \\ -1 \end{pmatrix} + \alpha \begin{pmatrix} -3 \\ -2 \\ 1 \\ 0 \end{pmatrix}. \tag{30}$$

⋆

7. DETERMINANTS AND THEIR APPLICATIONS

⋆ 13 ⋆

KEYWORDS: DETERMINANTS, BIJECTIONS, PERMUTATIONS, PRODUCTS OF PERMUTATIONS, SIGN, TRANSPOSITION, EVEN OR ODD PERMUTATIONS, CYCLES, MULTILINEARITY, ANTISYMMETRY, ELEMENTARY OPERATIONS, SINGULARITY, COFACTOR EXPANSION, COFACTOR MATRICES, ADJUGATE MATRICES, INVERSE MATRICES, CRAMER'S RULE, SPECIAL DETERMINANTS

7.1. **Determinants.** For a square matrix A of order n, a polynomial in the entries of A, named the determinant of A, a determinant of order n or an $n \times n$ determinant, is often considered. Determinants are quite complicated, but using them, we can determine the singularity of matrices, or establish an explicit formula for inverse matrices. We denote by $|A|$ or $\det A$ the determinant of A. To begin with, we observe the definition of determinants. Letting $A = (a_{ij})$,

$$\begin{aligned} |A| &= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \dots a_{n\sigma(n)} \\ &= \sum_{j_1,\dots,j_n} \operatorname{sgn}\begin{pmatrix} 1 & 2 & \dots & n \\ j_1 & j_2 & \dots & j_n \end{pmatrix} a_{1j_1} a_{2j_2} \dots a_{nj_n}. \end{aligned} \tag{1}$$

Here, S_n under the summation symbol on the right-hand side of the first line denotes the set of all permutations of n letters, and the sum on the right-hand side of the second line runs over all permutations $j_1, j_2, \dots, j_n$ of $1, 2, \dots, n$. The right-hand sides of the first and the second lines represent a completely identical meaning, with just a different way of writing, and therefore either can be adopted as the definition of a determinant. Anyway, as the definition includes permutations, we should study them in advance.

7.2. **Bijections.** Let A, B be two sets. Let f be a function from A to B. We write this function as

$$f : A \longrightarrow B. \tag{2}$$

If f satisfies that $f(x) \neq f(x')$ for any distinct two elements x and x', then f is called injective or an injection. If there exists an element x of A such that $f(x) = y$ for every element y of B, then f is called surjective, onto, or a surjection. If f is both injective and surjective, then f is called bijective, a bijection, or a 1-1 (one-to-one) correspondence. If $f : A \longrightarrow B$ is bijective, then there exists a function $g : B \longrightarrow A$ such that

$$\begin{aligned} (g \circ f)(x) = g(f(x)) = x \qquad & (x \in A) \\ (f \circ g)(y) = f(g(y)) = y \qquad & (y \in B) \end{aligned} \tag{3}$$

This function g is called the inverse (function) of f, denoted by $g = f^{-1}$. Then it holds that f is also the inverse of g.

7.3. **Permutations.** For convenience, write $\{1, 2, \ldots, n\} = [n]$. A bijection from $[n]$ to $[n]$ is called a permutation of n letters. For example, letting $n = 3$, the function below is actually a bijection because different numbers are mapped to different ones.

$$\begin{aligned} 1 &\longrightarrow 3 \\ 2 &\longrightarrow 1 \\ 3 &\longrightarrow 2 \end{aligned} \tag{4}$$

Hence this is a permutation of three letters. This is nothing but permuting of order of letters 1,2,3. Therefore the number of permutations of three letters is $3! = 6$. The permutation (4) is denoted by

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}. \tag{5}$$

Denoting it by σ, we have $\sigma(1) = 3$, $\sigma(2) = 1$ and $\sigma(3) = 2$. In general, a permutation of n letters are expressed as follows. This expression has the same meaning if we change the order of columns, because only the pair of column itself is important.

$$\begin{pmatrix} 1 & 2 & 3 & \ldots & n \\ j_1 & j_2 & j_3 & \ldots & j_n \end{pmatrix} \tag{6}$$

For a permutation of (6), j_1 has n possibilities, j_2 has $(n-1)$ possibilities,..., j_n has 1 possibility, thus the total number of permutations of n letters is $n!$. The set of all those permutations are denoted by S_n. For example,

$$\begin{aligned} S_3 = \; & \left\{ \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, \right. \\ & \left. \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} \right\}. \end{aligned} \tag{7}$$

The first permutation does not change the order, and so it is called the identity (unit) permutation, denoted by e. In general, the identity permutation of n letters is as follows.

$$e = \begin{pmatrix} 1 & 2 & \ldots & n \\ 1 & 2 & \ldots & n \end{pmatrix}. \tag{8}$$

7.4. **Products of permutations.** A permutation is a bijection, that is a kind of function. Hence we can compose permutations. The product of permutations is defined to be the composition of them, say, letting σ and τ be permutations of n letters, then $\sigma\tau$ is defined by

$$(\sigma\tau)(x) = \sigma(\tau(x)) \qquad (x \in [n]). \tag{9}$$

This is a new permutation of n letters. In general, it does not hold that $\sigma\tau = \tau\sigma$. For example,

$$\begin{aligned} \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} &= \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} \\ \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} &= \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}. \end{aligned} \tag{10}$$

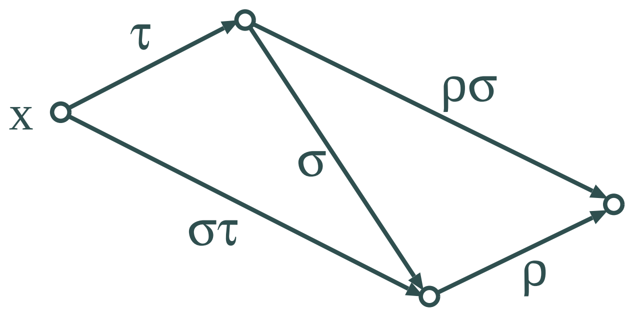
Obviously,

$$\sigma e = e\sigma = \sigma. \tag{11}$$

Theorem 1. *Let ρ, σ, τ be permutations, then the associative law holds:*

$$(\rho\sigma)\tau = \rho(\sigma\tau). \tag{12}$$

Proof. Take any $x \in [n]$. It follows from the following figure that $((\rho\sigma)\tau)(x) = (\rho(\sigma\tau))(x)$. $\therefore$ $(\rho\sigma)\tau = \rho(\sigma\tau)$. □



By Theorem 1, it is shown that the product of permutations:

$$\sigma_1\sigma_2 \dots \sigma_s \tag{13}$$

does not depend on the way to insert parentheses, and therefore parentheses are usually omitted.

Let σ be a permutation of n letters. Then there exists a permutation of n letters τ satisfying the following:

$$\sigma\tau = \tau\sigma = e. \tag{14}$$

This τ is called the inverse (permutation) of σ, denoted by σ^{-1}. Then it also holds that $\tau^{-1} = \sigma$. Hence σ and τ are the inverse of each other. It is easy to determine the inverse, it suffices to find simply the inverse function. For example,

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 2 & 3 & 1 \\ 1 & 2 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}. \tag{15}$$

Theorem 2. *It holds that*

$$(\sigma\tau)^{-1} = \tau^{-1}\sigma^{-1}. \tag{16}$$

Proof. Actually,

$$\begin{aligned}(\sigma\tau)(\tau^{-1}\sigma^{-1}) &= \sigma(\tau\tau^{-1})\sigma^{-1} = \sigma e\sigma^{-1} = \sigma\sigma^{-1} = e\\ (\tau^{-1}\sigma^{-1})(\sigma\tau) &= \tau^{-1}(\sigma^{-1}\sigma)\tau = \tau^{-1}e\tau = \tau^{-1}\tau = e. \quad \square\end{aligned} \tag{17}$$

Similarly, it is shown that

$$(\sigma_1\sigma_2\ldots\sigma_s)^{-1} = \sigma_s^{-1}\ldots\sigma_2^{-1}\sigma_1^{-1}. \tag{18}$$

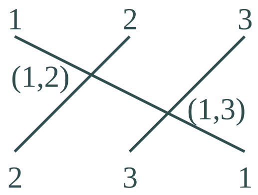
7.5. **The signs of permutations.** A permutation that interchanges only two letters and never changes the other letters, is called a transposition. There are three transpositions in S_3. A transposition that interchanges i and j is denoted by (i,j) or (j,i). The following is easily confirmed.

$$(i,j)^{-1} = (i,j) \tag{19}$$

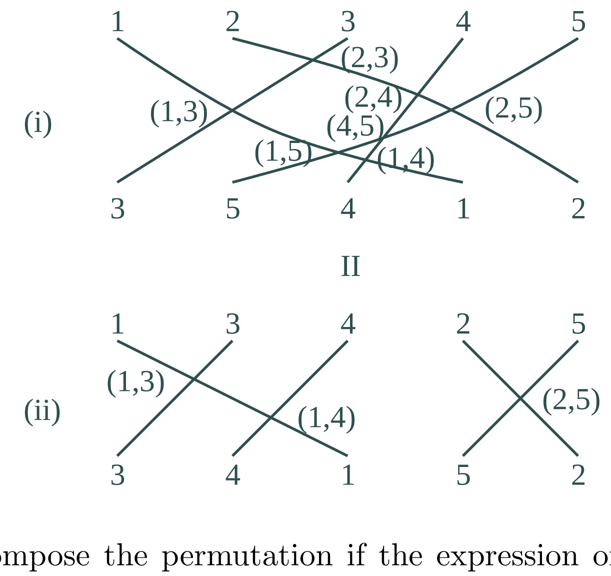
Every permutation of n letters can be expressed as the product of several transpositions. The method is as follows. We use the following examples for explanation.

$$\sigma = \begin{pmatrix} 1 & 2 & 3\\ 2 & 3 & 1\end{pmatrix}, \qquad \tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5\\ 3 & 5 & 4 & 1 & 2\end{pmatrix} \tag{20}$$

Draw the following figure I for σ, and write the transposition (i,j) at the intersection of the line connecting i's and the one connecting j's. If you read these transpositions from bottom to top, and write from left to right, then you have a decomposition into transpositions. Here, the lines may be curved, but monotonously go downwards, and it is forbidden that two lines touch or more than three lines intersect at one point.



I



II

In the figure II, it is easy to decompose the permutation if the expression of the permutation is transformed into the lower one. Consequently,

$$\sigma = (1,3)(1,2), \qquad \tau = \begin{cases} (1,4)(1,5)(4,5)(1,3)(2,5)(2,4)(2,3)\\ (1,4)(2,5)(1,3)\end{cases} \tag{21}$$

In the figure of a permutation, several transpositions located horizontally can be arranged freely as long as they are correctly arranged with the other transpositions. For example, the order of (1,3) and (2,5) in II can be inverted.

Let us consider the reason why the above operation gives a decomposition of a permutation into transpositions. In the figure of a permutation, the vertical axis represents time, where time passes downwards. The number 1 moves along the line, and reaches to the position of 1. The moves of the other numbers are very similar. At a certain time, the numbers from 1 to n are permuted into the arrangement of the numbers which appear as the intersections on the horizontal line corresponding to the time. If the intersections are 1,3,4,2,5, then the permutation

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 4 & 2 & 5 \end{pmatrix} \tag{22}$$

is performed up to the time of the horizontal line. And finally, the given permutation appears. In the permutation process in the figure, the change across the intersection is clearly a transposition. Accordingly, the permutation is expressed by piling up several transpositions.

In general, there are many ways to decompose a permutation into transpositions. Also, the number of transpositions in the product is not uniquely determined by the permutation. However, it is determined that whether the number of transpositions is even or odd (the parity of the number of transpositions) by the permutation. If a permutation is decomposed into even number of transpositions, then the permutation is called an even permutation, whereas if it is decomposed into odd number of transpositions, then it is called an odd permutation. Now define

$$\operatorname{sgn}(\sigma) = \begin{cases} 1 & (\sigma \text{ is an even permutation}) \\ -1 & (\sigma \text{ is an odd permutation}) \end{cases} \tag{23}$$

This is the sign of a permutation. The sign is determined by the number of the intersection in the figure of a permutation.

We explain the reason why the parity of the number of transpositions is determined by a permutation. Consider the product of all differences of $x_1, \dots, x_n$:

$$\Delta = \prod_{1 \le i < j \le n} (x_j - x_i). \tag{24}$$

It is called the difference product of n variables $x_1, \dots, x_n$. Let a permutation operate on Δ, e.g. if i is mapped to j, then replace x_i by x_j. It is confirmed that, by every transposition, Δ is always changed to $-De$. Hence if a permutation σ is decomposed as $\sigma = \sigma_1 \dots \sigma_s = \sigma'_1 \dots \sigma'_t$ for even s and odd t in two ways, then

$$\begin{aligned} \sigma\Delta = \sigma_1 \dots \sigma_s \Delta = (-1)^s \Delta = \Delta \\ \sigma\Delta = \sigma'_1 \dots \sigma'_t \Delta = (-1)^t \Delta = -\Delta, \end{aligned} \tag{25}$$

which is a contradiction.

Theorem 3. *Given two permutations σ and τ, we have the following.*

$$\begin{aligned} \operatorname{sgn}(\sigma\tau) &= \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau) \\ \operatorname{sgn}(\sigma^{-1}) &= \operatorname{sgn}(\sigma) \end{aligned} \tag{26}$$

Proof. Let $\sigma = \sigma_1\sigma_2\dots\sigma_s$ and $\tau = \tau_1\tau_2\dots\tau_t$ be decompositions into transpositions.

$$\begin{aligned} &\operatorname{sgn}(\sigma\tau) = \operatorname{sgn}(\sigma_1\dots\sigma_s\tau_1\dots\tau_t) = (-1)^{s+t} = (-1)^s(-1)^t = \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau) \\ &\operatorname{sgn}(\sigma^{-1}) = \operatorname{sgn}\left((\sigma_1\dots\sigma_s)^{-1}\right) = \operatorname{sgn}\left(\sigma_s^{-1}\dots\sigma_1^{-1}\right) = \operatorname{sgn}(\sigma_s\dots\sigma_1) = (-1)^s \\ &= \operatorname{sgn}(\sigma) \end{aligned} \tag{27}$$

Hence the theorem holds. For the second formula, it is clear if we notice that the figure of σ^{-1} is given by turning the figure of σ upside down. □

A permutation which substitutes letters as follows, while fixes the other letters, is called an s-cycle or a cycle of length s, denoted by $(i_1 i_2 \dots i_s)$.

$$i_1 \longrightarrow i_2 \longrightarrow \cdots \longrightarrow i_s \longrightarrow i_1 \tag{28}$$

Here, any cyclic reordering of the symbol $(i_1 i_2 \dots i_s)$ represents the same cycle. For example, $(235) = (352) = (523)$.

In general, an arbitrary permutation is expressed as a product of several cycles no two of which share common letters. For example, we can write as

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 5 & 1 & 6 & 4 & 2 & 3 & 8 & 7 \end{pmatrix} = (152)(36)(78)(4). \tag{29}$$

Such a decomposition of a permutation is called the cycle decomposition of a permutation, which is uniquely determined up to order of the factors. When a permutation is decomposed into cycles, the decreasing sequence of the lengths of those cycles is called the cycle type of the permutation. In the above case, the cycle type is $(3, 2, 2, 1)$.

If we decompose a permutation into disjoint cycles, and further decompose each cycle into transpositions, then we can decompose the permutation into transpositions. A cycle $(i_1 i_2 \dots i_s)$ is decomposed into $s - 1$ transpositions:

$$(i_1 i_2 \dots i_s) = (i_1, i_s) \dots (i_1, i_3)(i_1, i_2). \tag{30}$$

Hence if a permutation of n letters is decomposed into r cycles, then it is decomposed into $n - r$ transpositions, and therefore, we have

$$\operatorname{sgn}(\sigma) = (-1)^{n-r}. \tag{31}$$

(exercise01) (1) List up all permutations of 3 letters, and determine their signs.

(2) Decompose $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix}$ into transpositions and determine the sign.

(3) Confirm the Sarrus' rule for 3×3 determinants using the definition (1) of determinants.

7.6. **Basic properties of determinants.** With the above preparation, we can understand the meaning of the definition (1) of determinants. Intuitively speaking, a determinant of order n is a signed sum of $n!$ terms of the product of n entries selected exactly once from each row and column. However, (1) is not often used for actual calculation of determinants. The properties or formulas derived from (1) are often used for calculation of determinants.

Determinants have two properties "multilinearity" and "antisymmetry" as shown in the following.

Theorem 4. *(Multilinearity with respect to rows)*

$$\begin{vmatrix} \mathtt{a}'_1 \\ \vdots \\ \mathtt{a}'_i + \tilde{\mathtt{a}}'_i \\ \vdots \\ \mathtt{a}'_n \end{vmatrix} = \begin{vmatrix} \mathtt{a}'_1 \\ \vdots \\ \mathtt{a}'_i \\ \vdots \\ \mathtt{a}'_n \end{vmatrix} + \begin{vmatrix} \mathtt{a}'_1 \\ \vdots \\ \tilde{\mathtt{a}}'_i \\ \vdots \\ \mathtt{a}'_n \end{vmatrix}, \qquad \begin{vmatrix} \mathtt{a}'_1 \\ \vdots \\ c\mathtt{a}'_i \\ \vdots \\ \mathtt{a}'_n \end{vmatrix} = c \begin{vmatrix} \mathtt{a}'_1 \\ \vdots \\ \mathtt{a}'_i \\ \vdots \\ \mathtt{a}'_n \end{vmatrix}. \tag{32}$$

Theorem 4'. *(Multilinearity with respect to columns)*

$$\begin{aligned} \begin{vmatrix} \mathtt{a}_1 & \dots & \mathtt{a}_j + \tilde{\mathtt{a}}_j & \dots & \mathtt{a}_n \end{vmatrix} &= \begin{vmatrix} \mathtt{a}_1 & \dots & \mathtt{a}_j & \dots & \mathtt{a}_n \end{vmatrix} \\ &\quad + \begin{vmatrix} \mathtt{a}_1 & \dots & \tilde{\mathtt{a}}_j & \dots & \mathtt{a}_n \end{vmatrix}, \\ \begin{vmatrix} \mathtt{a}_1 & \dots & c\mathtt{a}_j & \dots & \mathtt{a}_n \end{vmatrix} &= c \begin{vmatrix} \mathtt{a}_1 & \dots & \mathtt{a}_j & \dots & \mathtt{a}_n \end{vmatrix}. \end{aligned} \tag{33}$$

Theorem 5. *(Antisymmetry with respect to rows) For $\tau \in S_n$,*

$$\begin{vmatrix} \mathtt{a}'_{\tau(1)} \\ \vdots \\ \mathtt{a}'_{\tau(n)} \end{vmatrix} = \operatorname{sgn}(\tau) \begin{vmatrix} \mathtt{a}'_1 \\ \vdots \\ \mathtt{a}'_n \end{vmatrix}. \tag{34}$$

Theorem 5'. *(Antisymmetry with respect to columns) For $\tau \in S_n$,*

$$\begin{vmatrix} \mathtt{a}_{\tau(1)} & \dots & \mathtt{a}_{\tau(n)} \end{vmatrix} = \operatorname{sgn}(\tau) \begin{vmatrix} \mathtt{a}_1 & \dots & \mathtt{a}_n \end{vmatrix}. \tag{35}$$

Here, using the first formula of (32) repeatedly, we have the following formula often used instead of the first one of (32).

$$\begin{vmatrix} \mathtt{a}'_1 \\ \vdots \\ \mathtt{a}'_i + \mathtt{b}'_i + \cdots + \mathtt{c}'_i \\ \vdots \\ \mathtt{a}'_n \end{vmatrix} = \begin{vmatrix} \mathtt{a}'_1 \\ \vdots \\ \mathtt{a}'_i \\ \vdots \\ \mathtt{a}'_n \end{vmatrix} + \begin{vmatrix} \mathtt{a}'_1 \\ \vdots \\ \mathtt{b}'_i \\ \vdots \\ \mathtt{a}'_n \end{vmatrix} + \cdots + \begin{vmatrix} \mathtt{a}'_1 \\ \vdots \\ \mathtt{c}'_i \\ \vdots \\ \mathtt{a}'_n \end{vmatrix} \tag{36}$$

Similarly, the following formula derived from the first formula of (33) is often used instead of the first one of (33).

$$\begin{aligned} \begin{vmatrix} \mathtt{a}_1 & \dots & \mathtt{a}_j + \mathtt{b}_j + \cdots + \mathtt{c}_j & \dots & \mathtt{a}_n \end{vmatrix} &= \begin{vmatrix} \mathtt{a}_1 & \dots & \mathtt{a}_j & \dots & \mathtt{a}_n \end{vmatrix} \\ &\quad + \begin{vmatrix} \mathtt{a}_1 & \dots & \mathtt{b}_j & \dots & \mathtt{a}_n \end{vmatrix} + \cdots + \begin{vmatrix} \mathtt{a}_1 & \dots & \mathtt{c}_j & \dots & \mathtt{a}_n \end{vmatrix} \end{aligned} \tag{37}$$

Proof of Theorem 4. From (1), it follows that

$$\begin{aligned}
\begin{vmatrix} \mathtt{a}'_1 \\ \vdots \\ \mathtt{a}'_i + \tilde{\mathtt{a}}'_i \\ \vdots \\ \mathtt{a}'_n \end{vmatrix} &= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} \dots \Big(a_{i\sigma(i)} + \tilde{a}_{i\sigma(i)} \Big) \dots a_{n\sigma(n)} \\
&= \sum_{\sigma \in S_n} \Big(\operatorname{sgn}(\sigma) a_{1\sigma(1)} \dots a_{i\sigma(i)} \dots a_{n\sigma(n)} + \operatorname{sgn}(\sigma) a_{1\sigma(1)} \dots \tilde{a}_{i\sigma(i)} \dots a_{n\sigma(n)} \Big) \\
&= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} \dots a_{i\sigma(i)} \dots a_{n\sigma(n)} + \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} \dots \tilde{a}_{i\sigma(i)} \dots a_{n\sigma(n)} \\
&= \begin{vmatrix} \mathtt{a}'_1 \\ \vdots \\ \mathtt{a}'_i \\ \vdots \\ \mathtt{a}'_n \end{vmatrix} + \begin{vmatrix} \mathtt{a}'_1 \\ \vdots \\ \tilde{\mathtt{a}}'_i \\ \vdots \\ \mathtt{a}'_n \end{vmatrix}. \qquad \text{Similarly, by (1),}
\end{aligned} \tag{38}$$

$$\begin{aligned}
\begin{vmatrix} \mathtt{a}'_1 \\ \vdots \\ c\mathtt{a}'_i \\ \vdots \\ \mathtt{a}'_n \end{vmatrix} &= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} \dots \Big(c a_{i\sigma(i)} \Big) \dots a_{n\sigma(n)} \\
&= c \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} \dots a_{i\sigma(i)} \dots a_{n\sigma(n)} = c \begin{vmatrix} \mathtt{a}'_1 \\ \vdots \\ \mathtt{a}'_n \end{vmatrix}. \quad \square
\end{aligned} \tag{39}$$

Proof of Theorem 5. From (1), it follows that

$$\begin{aligned}
\begin{vmatrix} \mathtt{a}'_{\tau(1)} \\ \vdots \\ \mathtt{a}'_{\tau(n)} \end{vmatrix} &= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{\tau(1)\sigma(1)} a_{\tau(2)\sigma(2)} \dots a_{\tau(n)\sigma(n)} \\
&\qquad \text{Applying the permutation } \tau^{-1} \text{ to the order of the product,} \\
&= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{\tau(\tau^{-1}(1)),\sigma(\tau^{-1}(1))}\, a_{\tau(\tau^{-1}(2)),\sigma(\tau^{-1}(2))} \dots a_{\tau(\tau^{-1}(n)),\sigma(\tau^{-1}(n))} \\
&= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{1,(\sigma\tau^{-1})(1)}\, a_{2,(\sigma\tau^{-1})(2)} \dots a_{n,(\sigma\tau^{-1})(n)} \\
&\qquad \text{Here, for fixed } \tau, \text{ if } \sigma \text{ runs over the set } S_n, \\
&\qquad \text{then } \rho = \sigma\tau^{-1} \text{ also runs over } S_n, \text{ thus} \\
&= \sum_{\rho \in S_n} \operatorname{sgn}(\rho\tau) a_{1\rho(1)}\, a_{2\rho(2)} \dots a_{n\rho(n)} \\
&= \sum_{\rho \in S_n} \operatorname{sgn}(\rho)\operatorname{sgn}(\tau) a_{1\rho(1)}\, a_{2\rho(2)} \dots a_{n\rho(n)} \\
&= \operatorname{sgn}(\tau) \sum_{\rho \in S_n} \operatorname{sgn}(\rho) a_{1\rho(1)}\, a_{2\rho(2)} \dots a_{n\rho(n)} = \operatorname{sgn}(\tau) \begin{vmatrix} \mathtt{a}'_1 \\ \vdots \\ \mathtt{a}'_n \end{vmatrix}. \quad \square
\end{aligned} \tag{40}$$

The properties with respect to columns are shown later. For antisymmetry, it is useful when τ is a transposition. Then the theorem is rewritten as

Theorem 5-. *A transposition of two rows of a determinant changes the sign of it, and a transposition of two columns of a determinant changes the sign of it.*

Next, we give a theorem for determinants of transposes.

Theorem 6. *Let A be a matrix of order n, then*

$$|A| = |{}^tA|. \tag{41}$$

Proof. By (1), we have

$$|{}^tA| = \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) a_{\sigma(1)1}a_{\sigma(2)2}\cdots a_{\sigma(n)n}$$

Applying the permutation σ^{-1} to the order of the product,

$$= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) a_{\sigma(\sigma^{-1}(1)),\sigma^{-1}(1)}a_{\sigma(\sigma^{-1}(2)),\sigma^{-1}(2)}\cdots a_{\sigma(\sigma^{-1}(n)),\sigma^{-1}(n)}$$

$$= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) a_{1,\sigma^{-1}(1)}a_{2,\sigma^{-1}(2)}\cdots a_{n,\sigma^{-1}(n)} \tag{42}$$

Here, σ runs over the set S_n, then $\rho = \sigma^{-1}$ also runs over S_n, thus

$$= \sum_{\rho\in S_n} \operatorname{sgn}(\rho^{-1}) a_{1\rho(1)}a_{2\rho(2)}\cdots a_{n\rho(n)}$$

$$= \sum_{\rho\in S_n} \operatorname{sgn}(\rho) a_{1\rho(1)}a_{2\rho(2)}\cdots a_{n\rho(n)} = |A|. \quad \square$$

From Theorem 6, we can derive the properties (33) and (35) concerning columns, because we have already shown the properties (32) and (34) concerning rows, and transposing both sides, we have the properties concerning columns. By Theorem 6, the determinant of a matrix is equal to the determinant of the transpose, and therefore we have the property concerning columns.

7.7. **Elementary operations and determinants.** By the property of determinants, we have the relationship between determinants and elementary operations.

R1: Interchanging two rows of a determinant changes the sign of the determinant. (multiplied by (-1))
R2: Multiplying one row of a determinant by c multiplies the determinant by c.
R3: Adding a scalar c multiple of some row to another row of a determinant preserves the value of the determinant.
C1: Interchanging two columns of a determinant changes the sign of the determinant. (multiplied by (-1))
C2: Multiplying one column of a determinant by c multiplies the determinant by c.
C3: Adding a scalar c multiple of some column to another column of a determinant preserves the value of the determinant.

Since R1,R2,C1,C2 are just the properties of determinants, we consider R3 and C3. First of all, we confirm that a determinant with two or more identical rows, or with two or more identical columns, is equal to 0. Let A be a square matrix with the identical i-th and j-th rows. If we interchange these rows, A is unchanged. However, by antisymmetry of a determinant,

$$|A| = -|A|, \qquad \therefore \quad |A| = 0. \tag{43}$$

Similarly, a determinant with several identical columns is equal to 0. Then let B be a matrix obtained by adding c multiple of the j-th row to the i-th row, then we have

$$|B| = \begin{vmatrix} \mathsf{a}'_1 \\ \vdots \\ \mathsf{a}'_i + c\mathsf{a}'_j \\ \vdots \\ \mathsf{a}'_j \\ \vdots \\ \mathsf{a}'_n \end{vmatrix} = \begin{vmatrix} \mathsf{a}'_1 \\ \vdots \\ \mathsf{a}'_i \\ \vdots \\ \mathsf{a}'_j \\ \vdots \\ \mathsf{a}'_n \end{vmatrix} + \begin{vmatrix} \mathsf{a}'_1 \\ \vdots \\ c\mathsf{a}'_j \\ \vdots \\ \mathsf{a}'_j \\ \vdots \\ \mathsf{a}'_n \end{vmatrix} = |A| + c \begin{vmatrix} \mathsf{a}'_1 \\ \vdots \\ \mathsf{a}'_j \\ \vdots \\ \mathsf{a}'_j \\ \vdots \\ \mathsf{a}'_n \end{vmatrix} = |A|. \tag{44}$$

A Similar formula with respect to columns holds.

Consider a determinant some of whose rows, i.e. the i-th row is a zero vector. According to (1), $a_{i\sigma(i)}$ is always equal to 0, hence all terms vanishes and the determinant vanishes. Transposing this result, a determinant some of whose columns is a zero vector also vanishes.

(exercise02) Calculate the determinant below using elementary operations.

$$\begin{vmatrix} 5 & 100 & 15 \\ 4 & 50 & 6 \\ 7 & 80 & 8 \end{vmatrix} \tag{45}$$

(ans)

$$\begin{aligned} \begin{vmatrix} 5 & 100 & 15 \\ 4 & 50 & 6 \\ 7 & 80 & 8 \end{vmatrix} &= 5 \cdot \begin{vmatrix} 1 & 20 & 3 \\ 4 & 50 & 6 \\ 7 & 80 & 8 \end{vmatrix} = 5 \cdot 10 \cdot \begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 8 \end{vmatrix} = 50 \cdot \begin{vmatrix} 1 & 2 & 1 \\ 4 & 5 & 1 \\ 7 & 8 & 0 \end{vmatrix} \\ &= 50 \cdot \begin{vmatrix} 1 & 1 & 1 \\ 4 & 1 & 1 \\ 7 & 1 & 0 \end{vmatrix} = 50 \cdot \begin{vmatrix} 1 & 1 & 1 \\ 3 & 0 & 0 \\ 7 & 1 & 0 \end{vmatrix} = 150. \end{aligned} \tag{46}$$

7.8. **Singularity and determinants.** Although it is not convenient to use (1) directly for determinant computation, it is sometimes useful for the determinant of a simple matrix A. Consider the determinant of E_n. For $A = E_n$, most terms of (1) vanish except the term such that all of

$$a_{1\sigma(1)}, a_{2\sigma(2)}, \ldots, a_{n\sigma(n)} \tag{47}$$

are diagonal entries, that is, $\sigma = e$. Accordingly, there is a unique nonzero term, and therefore

$$|E_n| = \mathrm{sgn}(e) 1 \ldots 1 = 1 \tag{48}$$

Similarly, for a diagonal matrix,

$$\begin{vmatrix} a_{11} & & & O \\ & a_{22} & & \\ & & \ddots & \\ O & & & a_{nn} \end{vmatrix} = \operatorname{sgn}(e)a_{11}a_{22}\dots a_{nn} = a_{11}a_{22}\dots a_{nn}. \tag{49}$$

Here, we consider a relationship between singularity of an $n \times n$ matrix A and its determinant. If we perform elementary operations on A, how does the determinant $|A|$ change? The answer is that only the sign changes or multiplied by c or unchanged, where $c \neq 0$. Therefore, the property that "$|A|$ is 0 or not" is unchanged before and after elementary operations. If A is nonsingular, then $r(A) = n$, thus it is transformed into E_n by elementary operations, and as $|E_n| = 1 \neq 0$, we have $|A| \neq 0$. To the contrary, if A is singular, then $r(A) = r < n$, thus it is transformed into $F_{nn}(r)$ by elementary operations, and as $|F_{nn}(r)| = 0$, we have $|A| = 0$.

Theorem 7. *Let A be a matrix of order n, then the following holds.*

$$\begin{array}{ccc} A \text{ is nonsingular} & \Longleftrightarrow & |A| \neq 0 \\ A \text{ is singular} & \Longleftrightarrow & |A| = 0 \end{array} \tag{50}$$

7.9. **Several important formulas.** Here we give several important formulas for determinants.

Theorem 8. *Let A and B be matrices of order n, then*

$$|AB| = |A||B|. \tag{51}$$

Theorem 9. *Let A be a matrix of order r and C be a matrix of order s, then*

$$\begin{vmatrix} A & B \\ O & C \end{vmatrix} = \begin{vmatrix} A & O \\ B' & C \end{vmatrix} = \begin{vmatrix} A & O \\ O & C \end{vmatrix} = |A||C|. \tag{52}$$

Proof of Theorem 8. First we prove when A is nonsingular. As A is nonsingular, by elementary row operations, we have

$$A \longrightarrow \cdots \longrightarrow E_n. \tag{53}$$

This is performed by multiplying several elementary matrices on the left, that is, multiplying A^{-1} on the left. By the very same operation as (53) on AB, we have

$$AB \longrightarrow \cdots \longrightarrow A^{-1}AB = B. \tag{54}$$

Incidentally, if the same elementary operation is performed on two matrices, then the values of their determinants are changed similarly, and consequently, multiplied by the same constant. Here, in (53), the determinant is multiplied by $\frac{1}{|A|}$, and thus from (54), it follows that

$$|AB| \cdot \frac{1}{|A|} = |B|. \qquad \therefore\ |AB| = |A||B|. \tag{55}$$

Next let A be singular. Then AB is also singular, because if not, there exists $X = (AB)^{-1}$ and

$$ABX = XAB = E_n, \tag{56}$$

and therefore, by Chapter 5, Theorem 3 (the theorem concerning nonsingular matrices),

$$A^{-1} = BX. \qquad \text{(contradiction)} \tag{57}$$

Accordingly, by Theorem 7, both sides of (51) are equal to 0. □

Proof of Theorem 9. Let us prove only the equality of the leftmost side and the rightmost side of (52), because the other equalities can be proved similarly. Let A and C be nonsingular. As A is nonsingular, performing elementary column operations on the columns of $\begin{vmatrix} A & B \\ O & C \end{vmatrix}$ which A shares, we have

$$\begin{vmatrix} A & B \\ O & C \end{vmatrix} \longrightarrow \overset{\times \frac{1}{|A|}}{\cdots} \longrightarrow \begin{vmatrix} E_r & B \\ O & C \end{vmatrix}. \tag{58}$$

By this operation, the value of the determinant is multiplied by $\frac{1}{|A|}$. Next, by elementary column operations, a block B on the right side of E_r is swept to be O. By this operations, the value of the determinant is unchanged. Lastly, as C is nonsingular, performing elementary column operations on the columns of $\begin{vmatrix} A & B \\ O & C \end{vmatrix}$ which C shares, and transform C into E_s. By this operation, the value of the determinant is multiplied by $\frac{1}{|C|}$. To summarize,

$$\begin{vmatrix} E_r & B \\ O & C \end{vmatrix} \longrightarrow \overset{\times 1}{\cdots} \longrightarrow \begin{vmatrix} E_r & O \\ O & C \end{vmatrix} \longrightarrow \overset{\times \frac{1}{|C|}}{\cdots} \longrightarrow \begin{vmatrix} E_r & O \\ O & E_s \end{vmatrix} = 1. \tag{59}$$

By (58),(59),

$$\begin{vmatrix} A & B \\ O & C \end{vmatrix} \cdot \frac{1}{|A|} \cdot \frac{1}{|C|} = 1. \qquad \therefore \begin{vmatrix} A & B \\ O & C \end{vmatrix} = |A||C|. \tag{60}$$

Next let A or C be singular. Then $\begin{pmatrix} A & B \\ O & C \end{pmatrix}$ is also singular. Because if not, we have

$$\begin{pmatrix} A & B \\ O & C \end{pmatrix}\begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix} = \begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix}\begin{pmatrix} A & B \\ O & C \end{pmatrix} = \begin{pmatrix} E_r & O \\ O & E_s \end{pmatrix}. \tag{61}$$

Hence

$$CX_{22} = E_s, \qquad X_{11}A = E_r, \tag{62}$$

and therefore, by Chapter 5, Theorem 3, A and C are nonsingular ((contradiction). Accordingly, by Theorem 7, the leftmost and the rightmost sides of (52) are equal to 0. □

(note) For upper or lower triangular matrices, we have

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ 0 & a_{22} & a_{23} & \dots & a_{2n} \\ 0 & 0 & a_{33} & \dots & a_{3n} \\ \multicolumn{5}{c}{\dotfill} \\ 0 & 0 & \dots & 0 & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{11} & 0 & 0 & \dots & 0 \\ a_{21} & a_{22} & 0 & \dots & 0 \\ a_{31} & a_{32} & a_{33} & \dots & 0 \\ \multicolumn{5}{c}{\dotfill} \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{vmatrix} = a_{11}a_{22}\dots a_{nn}. \quad (63)$$

Because, for upper triangular matrices, by the equality of the leftmost and the rightmost sides of Theorem 9, the given determinant is decomposed into smaller upper triangular determinants again and again, and consequently, equals to $a_{11}a_{22}\dots a_{nn}$. The lower triangular case is similar.

7.10. **Cofactor expansion of determinants.** In this section, we consider the method to expand a determinant using several determinants of smaller order. Let $A = (a_{ij})$ be a matrix of order n. Letting

$$A = \begin{pmatrix} \mathtt{a}_1 & \dots & \mathtt{a}_j & \dots & \mathtt{a}_n \end{pmatrix}, \quad (64)$$

and by elementary vectors $\mathtt{e}_i$, write

$$\mathtt{a}_j = \sum_{i=1}^{n} a_{ij}\mathtt{e}_i. \quad (65)$$

Then by multilinearity with respect to columns (37) and (33) (the second law), we have

$$\begin{aligned} |A| = & \begin{vmatrix} \mathtt{a}_1 & \dots & \overset{j)}{\sum_{i=1}^{n} a_{ij}\mathtt{e}_i} & \dots & \mathtt{a}_n \end{vmatrix} = \sum_{i=1}^{n} \begin{vmatrix} \mathtt{a}_1 & \dots & \overset{j)}{a_{ij}\mathtt{e}_i} & \dots & \mathtt{a}_n \end{vmatrix} \\ = & \sum_{i=1}^{n} a_{ij} \begin{vmatrix} \mathtt{a}_1 & \dots & \overset{j)}{\mathtt{e}_i} & \dots & \mathtt{a}_n \end{vmatrix}. \end{aligned} \quad (66)$$

Here, for simplicity, consider the i-th term of the sum. Interchange $\mathtt{e}_i$ on the j-th column by the left neighbor column again and again, then by $(j-1)$ times repetitions, $\mathtt{e}_i$ is moved to the first column. Namely,

$$a_{ij} \begin{vmatrix} \mathtt{a}_1 & \dots & \overset{j)}{\mathtt{e}_i} & \dots & \mathtt{a}_n \end{vmatrix} = (-1)^{j-1} a_{ij} \begin{vmatrix} \mathtt{e}_i & \mathtt{a}_1 & \dots & \hat{\mathtt{a}_j} & \dots & \mathtt{a}_n \end{vmatrix}. \quad (67)$$

Here, the symbol $\hat{\mathtt{a}_j}$ means that $\mathtt{a}_j$ is removed. (" ˆ " is sometimes called exclusion symbol.) Next, interchange the i-th row by the upper neighbor row again and again, then by $(i-1)$ times repetitions, the i-th row is moved to the first row, say,

$$(67) = (-1)^{i-1+j-1} a_{ij} \begin{vmatrix} 1 & a_{i1} & \dots & a_{i,j-1} & a_{i,j+1} & \dots & a_{in} \\ 0 & a_{11} & \dots & a_{1,j-1} & a_{1,j+1} & \dots & a_{1n} \\ 0 & a_{21} & \dots & a_{2,j-1} & a_{2,j+1} & \dots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & a_{i-1,1} & \dots & a_{i-1,j-1} & a_{i-1,j+1} & \dots & a_{i-1,n} \\ 0 & a_{i+1,1} & \dots & a_{i+1,j-1} & a_{i+1,j+1} & \dots & a_{i+1,n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & a_{n,1} & \dots & a_{n,j-1} & a_{n,j+1} & \dots & a_{n,n} \end{vmatrix} \tag{68}$$

Now by Theorem 9, we have

$$(68) = (-1)^{i+j} a_{ij} \begin{vmatrix} a_{11} & \dots & a_{1,j-1} & a_{1,j+1} & \dots & a_{1n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{i-1,1} & \dots & a_{i-1,j-1} & a_{i-1,j+1} & \dots & a_{i-1,n} \\ a_{i+1,1} & \dots & a_{i+1,j-1} & a_{i+1,j+1} & \dots & a_{i+1,n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n,1} & \dots & a_{n,j-1} & a_{n,j+1} & \dots & a_{n,n} \end{vmatrix}. \tag{69}$$

The determinant in (69) is the determinant of the matrix obtained by removing the i-th row and the j-th column from A. This determinant is called the (i,j) minor of A, denoted by Δ_{ij}. Furthermore, $(-1)^{i+j}\Delta_{ij}$ is called the (i,j) cofactor of A, denoted by $\tilde{a}_{ij}$. Using this notation, by (66)–(69), we have the first formula (70) of the following theorem.

Theorem 10. *For a matrix A of order n,*

$$|A| = a_{1j}\tilde{a}_{1j} + a_{2j}\tilde{a}_{2j} + \cdots + a_{nj}\tilde{a}_{nj} \tag{70}$$
$$|A| = a_{i1}\tilde{a}_{i1} + a_{i2}\tilde{a}_{i2} + \cdots + a_{in}\tilde{a}_{in}. \tag{71}$$

The second formula (71) is proved very similarly to the case of (70), or it is immediately obtained by transposing both sides of (70).

The formula (70) is called the cofactor expansion of $|A|$ along the j-th column, and the formula (71) is called the cofactor expansion of $|A|$ along the i-th row.

(exercise03) Compute the following determinant.

$$\begin{vmatrix} 2 & 3 & -1 & 1 \\ 3 & -1 & 1 & 2 \\ -1 & 1 & 2 & 3 \\ 1 & 2 & 3 & -1 \end{vmatrix} \tag{72}$$

(answer)

$$\begin{vmatrix} 2 & 3 & -1 & 1 \\ 3 & -1 & 1 & 2 \\ -1 & 1 & 2 & 3 \\ 1 & 2 & 3 & -1 \end{vmatrix} = \begin{vmatrix} 2 & 3 & -1 & 1 \\ 3 & -1 & 1 & 2 \\ -1 & 1 & 2 & 3 \\ 5 & 5 & 5 & 5 \end{vmatrix} = 5\begin{vmatrix} 2 & 3 & -1 & 1 \\ 3 & -1 & 1 & 2 \\ -1 & 1 & 2 & 3 \\ 1 & 1 & 1 & 1 \end{vmatrix}$$

$$= 5\begin{vmatrix} 3 & 4 & 0 & 2 \\ 2 & -2 & 0 & 1 \\ -3 & -1 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{vmatrix} \begin{matrix}\text{cofactor expansion} \\ = \\ \text{along the 3rd column}\end{matrix} \; 5\cdot(-1)^{4+3}1\begin{vmatrix} 3 & 4 & 2 \\ 2 & -2 & 1 \\ -3 & -1 & 1 \end{vmatrix} \tag{73}$$

$$= -5\begin{vmatrix} 3 & 4 & 2 \\ 5 & -1 & 0 \\ -3 & -1 & 1 \end{vmatrix} \begin{matrix}\text{cofactor expansion} \\ = \\ \text{along the 2nd row}\end{matrix}$$

$$-5\left((-1)^{2+1}5\begin{vmatrix} 4 & 2 \\ -1 & 1 \end{vmatrix} + (-1)^{2+2}(-1)\begin{vmatrix} 3 & 2 \\ -3 & 1 \end{vmatrix}\right) = -5(-30-9) = 195.$$

(note) Determinants of order greater than 3 does not satisfy Sarrus' rule.

7.11. **An explicit formula for inverse matrices.** So far, we have studied determinant computation, and as a byproduct, we can obtain an explicit formula for inverse matrices. The key formulas are (70), (71) and the following.

$$a_{1i}\tilde{a}_{1j} + a_{2i}\tilde{a}_{2j} + \cdots + a_{ni}\tilde{a}_{nj} = 0 \qquad (i \neq j) \tag{74}$$

$$a_{j1}\tilde{a}_{i1} + a_{j2}\tilde{a}_{i2} + \cdots + a_{jn}\tilde{a}_{in} = 0 \qquad (i \neq j) \tag{75}$$

To show the above equalities, consider a matrix B obtained by replacing the j-th column of A by the i-th column, and consider a matrix C obtained by replacing the i-th row of A by the j-th row. Since B and C have two identical columns and rows, respectively, their determinants are equal to 0. Then the cofactor expansion of $|B|$ along the j-th column gives (74) and the cofactor expansion of $|C|$ along the i-th row gives (75).

Here, for a matrix A of order n, consider the following matrix:

$$\operatorname{cof}(A) = \begin{pmatrix} \tilde{a}_{11} & \tilde{a}_{12} & \dots & \tilde{a}_{1n} \\ \tilde{a}_{21} & \tilde{a}_{22} & \dots & \tilde{a}_{2n} \\ \dots & \dots & \dots & \dots \\ \tilde{a}_{n1} & \tilde{a}_{n2} & \dots & \tilde{a}_{nn} \end{pmatrix}. \tag{76}$$

This matrix has the (i,j) cofactor of A as its (i,j) entry, and it is called the cofactor matrix of A. The transpose of the cofactor matrix is called the adjugate matrix of A, denoted by $\operatorname{adj}(A)$. In brief, $\operatorname{adj}(A) = {}^t(\tilde{a}_{ij})$. Then we have

$$A\operatorname{adj}(A) = \operatorname{adj}(A)\,A = |A|E_n. \tag{77}$$

To clarify the entries,

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \multicolumn{4}{c}{\dotfill} \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} \tilde{a}_{11} & \tilde{a}_{21} & \dots & \tilde{a}_{n1} \\ \tilde{a}_{12} & \tilde{a}_{22} & \dots & \tilde{a}_{n2} \\ \multicolumn{4}{c}{\dotfill} \\ \tilde{a}_{1n} & \tilde{a}_{2n} & \dots & \tilde{a}_{nn} \end{pmatrix} = \begin{pmatrix} |A| & & & O \\ & |A| & & \\ & & \ddots & \\ O & & & |A| \end{pmatrix}, \tag{78}$$

$$\begin{pmatrix} \tilde{a}_{11} & \tilde{a}_{21} & \dots & \tilde{a}_{n1} \\ \tilde{a}_{12} & \tilde{a}_{22} & \dots & \tilde{a}_{n2} \\ \multicolumn{4}{c}{\dotfill} \\ \tilde{a}_{1n} & \tilde{a}_{2n} & \dots & \tilde{a}_{nn} \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \multicolumn{4}{c}{\dotfill} \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} = \begin{pmatrix} |A| & & & O \\ & |A| & & \\ & & \ddots & \\ O & & & |A| \end{pmatrix}. \tag{79}$$

Indeed, in calculating the left-hand side of (78), the (i,i) entry is obtained by (71), and the (j,i) entry $(i \neq j)$ is obtained by (75). In calculating the left-hand side of (79), the (j,j) entry is obtained by (70), and the (j,i) entry $(i \neq j)$ is obtained by (74).

Here, if A is nonsingular, then $|A| \neq 0$ by Theorem 7, and dividing (77) by $|A|$, we have

$$A\left(\frac{1}{|A|}\,\mathrm{adj}\,(A)\right) = \left(\frac{1}{|A|}\,\mathrm{adj}\,(A)\right)A = E_n. \tag{80}$$

Theorem 11. *For a nonsingular matrix A,*

$$A^{-1} = \frac{1}{|A|}\,\mathrm{adj}\,(A)\,. \tag{81}$$

(exercise04) Compute the inverse of the following matrix A:

$$\begin{pmatrix} 4 & 5 & 6 \\ 7 & 8 & 8 \\ 1 & 2 & 3 \end{pmatrix} \tag{82}$$

(answer)

$$\begin{aligned} \begin{pmatrix} 4 & 5 & 6 \\ 7 & 8 & 8 \\ 1 & 2 & 3 \end{pmatrix}^{-1} &= \begin{vmatrix} 4 & 5 & 6 \\ 7 & 8 & 8 \\ 1 & 2 & 3 \end{vmatrix}^{-1} \begin{pmatrix} \begin{vmatrix} 8 & 8 \\ 2 & 3 \end{vmatrix} & -\begin{vmatrix} 5 & 6 \\ 2 & 3 \end{vmatrix} & \begin{vmatrix} 5 & 6 \\ 8 & 8 \end{vmatrix} \\ -\begin{vmatrix} 7 & 8 \\ 1 & 3 \end{vmatrix} & \begin{vmatrix} 4 & 6 \\ 1 & 3 \end{vmatrix} & -\begin{vmatrix} 4 & 6 \\ 7 & 8 \end{vmatrix} \\ \begin{vmatrix} 7 & 8 \\ 1 & 2 \end{vmatrix} & -\begin{vmatrix} 4 & 5 \\ 1 & 2 \end{vmatrix} & \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix} \end{pmatrix} \\ &= \frac{1}{3}\begin{pmatrix} 8 & -3 & -8 \\ -13 & 6 & 10 \\ 6 & -3 & -3 \end{pmatrix}. \end{aligned} \tag{83}$$

(note) The inverse matrix formula (81) is important in the theory, and for the order $n \leq 3$, it is also sufficiently practical. For larger n, however, the method by elementary operations is faster than this formula in many cases.

7.12. **Cramer's rule.** We can solve a system of linear equations for a nonsingular $n \times n$ matrix $A = (a_{ij})$:

$$A\mathtt{x} = \mathtt{c} \tag{84}$$

by the inverse matrix formula. First multiplying both sides of this equation by A^{-1} on the left,

$$\mathtt{x} = A^{-1}\mathtt{c} = |A|^{-1}\mathrm{adj}\,(A)\,\mathtt{c}. \tag{85}$$

Here, writing down the j-th component x_j of $\mathtt{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, we have

$$x_j = |A|^{-1}\left(\tilde{a}_{1j}c_1 + \tilde{a}_{2j}c_2 + \cdots + \tilde{a}_{nj}c_n\right). \tag{86}$$

Compare this with (70). Let $|A_j|$ denote the determinant given by replacing the j-th column of $|A|$ by $\mathtt{c}$. Then the numerator of (86) is the cofactor expansion of A_j along the j-th column, and therefore

Theorem 12. *(Cramer's rule)*

$$x_j = \frac{|A_j|}{|A|} = \frac{\overset{j)}{\begin{vmatrix} \mathtt{a}_1 & \ldots & \mathtt{c} & \ldots & \mathtt{a}_n \end{vmatrix}}}{\begin{vmatrix} \mathtt{a}_1 & \ldots & \mathtt{a}_n \end{vmatrix}}. \tag{87}$$

Note that this formula is valid only for the case that A is nonsingular.

(exercise05) Using Cramer's rule, solve the following system of linear equations.

$$\begin{cases} 3x_1 & -5x_2 & +x_3 & = & 8 \\ 5x_1 & +3x_2 & -5x_3 & = & 1 \\ 2x_1 & -2x_2 & +3x_3 & = & 7 \end{cases} \tag{88}$$

(answer)

$$|A| = \begin{vmatrix} 3 & -5 & 1 \\ 5 & 3 & -5 \\ 2 & -2 & 3 \end{vmatrix} = \begin{vmatrix} 3 & -2 & 1 \\ 5 & 8 & -5 \\ 2 & 0 & 3 \end{vmatrix} = 2\begin{vmatrix} 3 & -1 & 1 \\ 5 & 4 & -5 \\ 2 & 0 & 3 \end{vmatrix} = 2\begin{vmatrix} 4 & -1 & 1 \\ 0 & 4 & -5 \\ 5 & 0 & 3 \end{vmatrix}$$
$$= 2\begin{vmatrix} 4 & -1 & 0 \\ 0 & 4 & -1 \\ 5 & 0 & 3 \end{vmatrix} = 2(48+5) = 106. \tag{89}$$

$$|A_1| = \begin{vmatrix} 8 & -5 & 1 \\ 1 & 3 & -5 \\ 7 & -2 & 3 \end{vmatrix} = \begin{vmatrix} 1 & -3 & -2 \\ 1 & 3 & -5 \\ 7 & -2 & 3 \end{vmatrix} = \begin{vmatrix} 2 & 0 & -7 \\ 1 & 3 & -5 \\ 7 & -2 & 3 \end{vmatrix}$$
$$= 18 + 14 - 20 + 147 = 159.$$
$$|A_2| = \begin{vmatrix} 3 & 8 & 1 \\ 5 & 1 & -5 \\ 2 & 7 & 3 \end{vmatrix} = \begin{vmatrix} 3 & 8 & 4 \\ 5 & 1 & 0 \\ 2 & 7 & 5 \end{vmatrix} = \begin{vmatrix} 1 & 1 & -1 \\ 5 & 1 & 0 \\ 2 & 7 & 5 \end{vmatrix} = \begin{vmatrix} 0 & 0 & -1 \\ 5 & 1 & 0 \\ 7 & 12 & 5 \end{vmatrix}$$
$$= -60 + 7 = -53.$$
$$|A_3| = \begin{vmatrix} 3 & -5 & 8 \\ 5 & 3 & 1 \\ 2 & -2 & 7 \end{vmatrix} = \begin{vmatrix} 3 & -2 & 8 \\ 5 & 8 & 1 \\ 2 & 0 & 7 \end{vmatrix} = 2\begin{vmatrix} 1 & -1 & 1 \\ 5 & 4 & 1 \\ 2 & 0 & 7 \end{vmatrix} = 2\begin{vmatrix} 1 & -1 & 0 \\ 5 & 4 & -4 \\ 2 & 0 & 5 \end{vmatrix} \tag{89}$$
$$= 2(20 + 8 + 25) = 106.$$
$$\therefore \quad x_1 = \frac{|A_1|}{|A|} = \frac{159}{106} = \frac{3}{2},\ x_2 = \frac{|A_2|}{|A|} = \frac{-53}{106} = -\frac{1}{2},\ x_3 = \frac{|A_3|}{|A|} = \frac{106}{106} = 1,$$
$$\text{i.e.} \quad \mathtt{x} = \frac{1}{2}\begin{pmatrix} 3 \\ -1 \\ 2 \end{pmatrix}.$$

7.13. **Special determinants.** There are a lot of beautiful determinants. In this section, we show several examples of them. To begin with, we introduce a theorem to prove the formulas stated below.

Theorem 13. *(Factor theorem for several variables) For a polynomial $p(x_1, \ldots, x_n)$ in the variables $x_1, \ldots, x_n$, choose a variable, e.g. x_1, and if a polynomial $q(x_2, \ldots, x_n)$ in $(n-1)$ variables satisfies that $p(q, x_2, \ldots, x_n) = 0$, then $p(x_1, \ldots, x_n)$ is divisible by $x_1 - q(x_2, \ldots, x_n)$.*

7.13.1. **Vandermonde determinants.**

$$\begin{vmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \ldots & x_2^{n-1} \\ \multicolumn{5}{c}{\dotfill} \\ 1 & x_n & x_n^2 & \ldots & x_n^{n-1} \end{vmatrix} = \prod_{1 \le i < j \le n} (x_j - x_i) \tag{90}$$

(note) (1) The right-hand side is the difference product of n variables (24). (2) The left-hand side is sometimes written in the transposed form.

Proof. On the left-hand side, letting $x_j = x_i$ $(i \neq j)$, the determinant has the identical i-th and j-th row, and it vanishes. Therefore by the factor theorem for several variables, the left-hand side is divisible by $x_j - x_i$. For all i, j satisfying $1 \le i < j \le n$, there are $n(n-1)/2$ factors $x_j - x_i$, and every pair of them are relatively prime, thus the left-hand side is divisible by the product of them:

$$\prod_{1 \le i < j \le n} (x_j - x_i), \tag{91}$$

that is, the right-hand side. Here, the total degrees of both sides are equal to $n(n-1)/2$, and therefore the difference between both sides is a constant multiple. Since the coefficients of $x_2 x_3^2 \ldots x_n^{n-1}$ in both sides are equal to 1, both sides coincide. □

7.13.2. **Circulant determinants.**

$$\begin{vmatrix} x_0 & x_1 & x_2 & \ldots & x_{n-1} \\ x_{n-1} & x_0 & x_1 & \ldots & x_{n-2} \\ x_{n-2} & x_{n-1} & x_0 & \ldots & x_{n-3} \\ \multicolumn{5}{c}{\dotfill} \\ x_1 & x_2 & x_3 & \ldots & x_0 \end{vmatrix} = \prod_{\omega^n = 1} (x_0 + \omega x_1 + \omega^2 x_2 + \cdots + \omega^{n-1} x_{n-1}) \tag{92}$$

Here, the product on the right-hand side runs over n n-th roots of unity ω.

Proof. Take an arbitrary n-th root of unity ω. For $j = 2, \ldots, n$, add the first column by the j-th column multiplied by ω^{j-1} on the left-hand side. Then the $(i, 1)$ entry is expressed as

$$\omega^{i-1}(x_0 + \omega x_1 + \omega^2 x_2 + \cdots + \omega^{n-1} x_{n-1}). \tag{93}$$

Letting

$$x_0 = -\omega x_1 - \omega^2 x_2 - \cdots - \omega^{n-1} x_{n-1}, \tag{94}$$

then the first column vanishes and the left-hand side also vanishes. Hence by the factor theorem for several variables, the left-hand side is divisible by

$$g(\omega) = x_0 + \omega x_1 + \omega^2 x_2 + \cdots + \omega^{n-1} x_{n-1}. \tag{95}$$

For n values of ω, every pair of $g(\omega)$'s are relatively prime, and the left-hand side is divisible by the product of $g(\omega)$'s, i.e. the right-hand side. Here, the total degrees of both sides are equal to n, and therefore the difference between both sides is a constant multiple. Since the coefficients of x_0^n in both sides are equal to 1, both sides coincide. □

7.13.3. **Pfaffians.** Let A be a skew-symmetric matrix (satisfying ${}^tA = -A$) of order n. If n is odd, $|A| = 0$, whereas if n is even, $|A|$ is the square of some polynomial in the entries of A. That is to say, letting

$$A = \begin{pmatrix} 0 & a_{12} & a_{13} & \cdots & a_{1,n-1} & a_{1n} \\ -a_{12} & 0 & a_{23} & \cdots & a_{2,n-1} & a_{2n} \\ -a_{13} & -a_{23} & 0 & \cdots & a_{3,n-1} & a_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ -a_{1,n-1} & -a_{2,n-1} & -a_{3,n-1} & \cdots & 0 & a_{n-1,n} \\ -a_{1n} & -a_{2n} & -a_{3n} & \cdots & -a_{n-1,n} & 0 \end{pmatrix}, \tag{96}$$

then there exist two polynomials $p(a_{12}, a_{13}, \ldots, a_{n-1,n})$ in the entries of A, such that

$$|A| = [p(a_{12}, a_{13}, \ldots, a_{n-1,n})]^2. \tag{97}$$

One of the two polynomials $p(a_{12}, a_{13}, \ldots, a_{n-1,n})$ with $p(1, 1, \ldots, 1) = 1$ is called the Pfaffian of A, denoted by pf$A$. Examples for small n follows.

$$\mathrm{pf}\begin{pmatrix} 0 & a \\ -a & 0 \end{pmatrix} = a \qquad \mathrm{pf}\begin{pmatrix} 0 & a & b & c \\ -a & 0 & d & e \\ -b & -d & 0 & f \\ -c & -e & -f & 0 \end{pmatrix} = af - be + cd$$

$$\mathrm{pf}\begin{pmatrix} 0 & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ -a_{12} & 0 & a_{23} & a_{24} & a_{25} & a_{26} \\ -a_{13} & -a_{23} & 0 & a_{34} & a_{35} & a_{36} \\ -a_{14} & -a_{24} & -a_{34} & 0 & a_{45} & a_{46} \\ -a_{15} & -a_{25} & -a_{35} & -a_{45} & 0 & a_{56} \\ -a_{16} & -a_{26} & -a_{36} & -a_{46} & -a_{56} & 0 \end{pmatrix} = \begin{array}{l} a_{12}a_{34}a_{56} - a_{12}a_{35}a_{46} + a_{12}a_{36}a_{45} \\ -a_{13}a_{24}a_{56} + a_{13}a_{25}a_{46} - a_{13}a_{26}a_{45} \\ +a_{14}a_{23}a_{56} - a_{14}a_{25}a_{36} + a_{14}a_{26}a_{35} \\ -a_{15}a_{23}a_{46} + a_{15}a_{24}a_{36} - a_{15}a_{26}a_{34} \\ +a_{16}a_{23}a_{45} - a_{16}a_{24}a_{35} + a_{16}a_{25}a_{34} \end{array} \tag{98}$$

(exercise06) (1) If A is a skew-symmetric matrix of odd order, then show that $|A| = 0$.
(2) Prove the first and the second equalities of (98).

★

8. VECTOR SPACES, SUBSPACES, AND THEIR BASES

⋆ 12 ⋆

KEYWORDS: FIELDS, VECTOR SPACES, AXIOMS OF VECTOR SPACES, SUBSPACES, LINEAR INDEPENDENCE, LINEAR DEPENDENCE, FINITE-DIMENSIONAL VECTOR SPACES, BASES, DIMENSION, BASES OF K^n, SPANNED SUBSPACES, SUM OF SUBSPACES, INTERSECTION OF SUBSPACES, DIMENSION FORMULA, DIRECT SUM OF SUBSPACES

8.0. **Binary operations.** A set S is called closed under a binary operation if the operation is defined on S and for every two elements of S, the result of the operation is also in S. For example, the set of all integers $\mathbb{Z}$ is closed under addition, subtraction, and multiplication, whereas it is not closed under division. In mathematics, various operations and closed sets are often considered, and their properties are studied deeply.

8.1. **Vector spaces.** A field is, in short, a set closed under the four operations (division by zero is excluded), and satisfies commutativity, associativity, and distributivity of addition and multiplication.[1] The following are important examples of fields:

$$\begin{array}{l} \mathbb{Q} = \{\text{all rational numbers}\},\ \mathbb{R} = \{\text{all real numbers}\}, \\ \mathbb{C} = \{\text{all complex numbers}\}. \end{array} \tag{1}$$

They are called the field of rationals, the field of reals, and the field of complex numbers, respectively.

Let K be a field. If V is a set closed under addition and scalar multiplication by the element of K, and satisfied the following axioms $[L1]$–$[L8]$, then V is called a vector space over a scalar field (coefficient field) K.

[1] $x+y = y+x$, $(x+y)+z = x+(y+z)$, $xy = yx$, $(xy)z = x(yz)$, $x(y+z) = xy+xz$, $(x+y)z = xz+yz$.

$$
\begin{array}{llr}
[L1] & \mathtt{x}+\mathtt{y}=\mathtt{y}+\mathtt{x} & (\mathtt{x},\mathtt{y}\in V) \\
[L2] & (\mathtt{x}+\mathtt{y})+\mathtt{z}=\mathtt{x}+(\mathtt{y}+\mathtt{z}) & (\mathtt{x},\mathtt{y},\mathtt{z}\in V) \\
[L3] & \text{There exists a zero vector } \mathtt{0} \text{ in } V \text{ such that for every } \mathtt{x}\in V,\ \mathtt{x}+\mathtt{0}=\mathtt{0}+\mathtt{x}=\mathtt{x} & \\
[L4] & \text{For every } \mathtt{x}\in V, \text{ its inverse vector } -\mathtt{x} \text{ exists in } V \text{ and satisfies that } \mathtt{x}+(-\mathtt{x})=(-\mathtt{x})+\mathtt{x}=\mathtt{0} & \\
[L5] & k(\mathtt{x}+\mathtt{y})=k\mathtt{x}+k\mathtt{y} & (\mathtt{x},\mathtt{y}\in V;\ k\in K) \\
[L6] & (k+l)\mathtt{x}=k\mathtt{x}+l\mathtt{x} & (\mathtt{x}\in V;\ k,l\in K) \\
[L7] & (kl)\mathtt{x}=k(l\mathtt{x}) & (\mathtt{x}\in V;\ k,l\in K) \\
[L8] & 1\mathtt{x}=\mathtt{x} & (\mathtt{x}\in V)
\end{array}
\tag{2}
$$

However, in almost every case (especially in this text), if V is closed under addition and scalar multiplication, then the above axioms are satisfied automatically. Hence, it is actually valid that the definition of a vector space is to satisfy the condition that V is closed under addition and scalar multiplication, that is, satisfy the following:

For any $\mathtt{x},\mathtt{y}\in V$ and for any $k\in K$,

$$
\begin{aligned}
\mathtt{x}+\mathtt{y} &\in V \\
k\mathtt{x} &\in V.
\end{aligned}
\tag{3}
$$

If $K=\mathbb{C}$, V is called a complex vector space, and if $K=\mathbb{R}$, V is called a real vector space. We present the general theory of vector spaces over arbitrary K, because what K is does not influence the outline of the argument.

Elements of a vector space are called vectors. Vectors in a narrower sense handled so far are sometimes called numerical vectors to avoid the confusion. Elements of K are called scalars.

By [$L2$], the associative law of addition holds, and so the sum of several vectors does not depend on the way to insert parentheses.

$$
\mathtt{x}_1+\mathtt{x}_2+\cdots+\mathtt{x}_s \tag{4}
$$

Hence parentheses are usually omitted. Furthermore, by [$L1$], the commutative law of addition, the order of addition can be changed. The following are derived from the axioms of vector spaces.

(A) (Uniqueness of the zero vector) There exists a unique zero vector in V.

(B) (Uniqueness of the inverse vector) For every element $\mathtt{x}\in V$, there exists a unique inverse vector $-\mathtt{x}$.

(C) ((Two-sided) Cancellation law)
$\mathtt{x}+\mathtt{z}=\mathtt{y}+\mathtt{z} \implies \mathtt{x}=\mathtt{y};\quad \mathtt{z}+\mathtt{x}=\mathtt{z}+\mathtt{y} \implies \mathtt{x}=\mathtt{y}\quad (\mathtt{x},\mathtt{y},\mathtt{z}\in V)$

(D) $0\mathtt{x}=\mathtt{0}\quad(\mathtt{x}\in V);\qquad k\mathtt{0}=\mathtt{0}\quad(k\in K)$

(E) $k\mathtt{x}=\mathtt{0} \implies \mathtt{x}=\mathtt{0}$ or $k=0\qquad(\mathtt{x}\in V;\ k\in K)$

(F) $(-k)\mathtt{x}=k(-\mathtt{x})=-(k\mathtt{x})\qquad(\mathtt{x}\in V;\ k\in K)$

A typical example of a vector space over K is K^n defined as below, however, there are many other examples.[2]

$$K^n = \left\{ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mid x_1, \dots, x_n \in K \right\}, \tag{5}$$

This is a complex vector space $\mathbb{C}^n$ if $K = \mathbb{C}$, whereas it is a real vector space $\mathbb{R}^n$ if $K = \mathbb{R}$. In particular, $\mathbb{R}^3 = V^3$.

(exercise01) Are the following sets regarded as complex vector spaces or real vector spaces?

$$(1)\ V = \left\{ \begin{pmatrix} x \\ x+zi \\ z \end{pmatrix} \mid x, z \in \mathbb{R} \right\} \qquad (2)\ V = \left\{ \begin{pmatrix} x \\ x+zi \\ z \end{pmatrix} \mid x, z \in \mathbb{C} \right\} \tag{6}$$

(ans) (1) For any elements $\begin{pmatrix} x \\ x+zi \\ z \end{pmatrix}$ and $\begin{pmatrix} x' \\ x'+z'i \\ z' \end{pmatrix}$ of V, and any $k \in \mathbb{R}$,

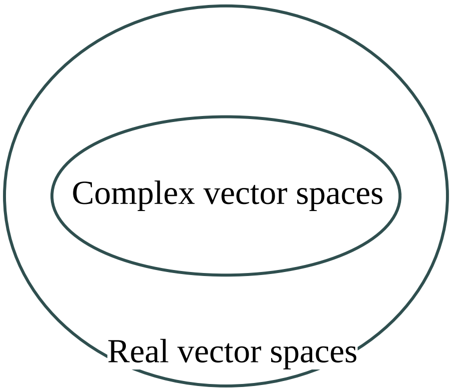
$$\begin{gathered} \begin{pmatrix} x \\ x+zi \\ z \end{pmatrix} + \begin{pmatrix} x' \\ x'+z'i \\ z' \end{pmatrix} = \begin{pmatrix} x+x' \\ x+x'+(z+z')i \\ z+z' \end{pmatrix} = \begin{pmatrix} \tilde{x} \\ \tilde{x}+\tilde{z}i \\ \tilde{z} \end{pmatrix} \in V \\ k\begin{pmatrix} x \\ x+zi \\ z \end{pmatrix} = \begin{pmatrix} kx \\ kx+kzi \\ kz \end{pmatrix} = \begin{pmatrix} x'' \\ x''+z''i \\ z'' \end{pmatrix} \in V. \end{gathered} \tag{7}$$

Hence V is a real vector space. However, for $i \in \mathbb{C}$,

$$i\begin{pmatrix} x \\ x+zi \\ z \end{pmatrix} = \begin{pmatrix} ix \\ -z+ix \\ iz \end{pmatrix} \notin V \quad \text{(Because the 1st or 3rd component is imaginary)}, \tag{8}$$

and therefore V is not a complex vector space.
(2) (Outline) Similarly to (1), since V is closed under addition and complex scalar multiplication, it is a complex vector space. Hence it is also regarded as a real vector space.

[2]Examples of vector spaces over K: (i) $M_{m,n}(K)$: the vector space consisting of all $m \times n$ matrices with entries in K, the zero vector is $O_{m,n}$, the inverse of $X \in M_{m,n}(K)$ is $-X$. In particular, write $M_{nn}(K) = M_n(K)$. (ii) $K[t]$: the vector space consisting of all polynomials in t with coefficients in K: the zero vector is 0, the inverse of $p(t) \in K[t]$ is $-p(t)$. (iii) K^X: the vector space consisting of all functions from a nonempty set X to K, the zero vector is $0(x) = 0$ $(x \in X)$, the inverse of $f \in K^X$ is $-f = (-1)f$. Here, we define that $(f+g)(x) = f(x) + g(x)$ and $(kf)(x) = kf(x)$ $(x \in X)$.



(note) If V is a complex vector space, then it holds that, for any $\mathtt{x}, \mathtt{y} \in V$ and for any complex number c, $\mathtt{x} + \mathtt{y} \in V$ and $c\mathtt{x} \in V$. Hence it is clear that for any real number k, $k\mathtt{x} \in V$ (because a real number is a complex number). Consequently, a complex vector space is regarded as also a real vector space. Conversely, if V is a real vector space and it holds that $\mathtt{x} \in V \Rightarrow i\mathtt{x} \in V$, then V is closed under complex scalar multiplication, and under the assumption that the axioms with respect to complex scalar multiplication hold, V is regarded also as a complex vector space.
(note) A singleton $\{\mathtt{0}\}$ is regarded as a special vector space, called the trivial vector space or zero vector space. All vector spaces not being trivial are non-trivial vector spaces.
(note) Hereafter, we use the symbols $\mathtt{e}_1, \mathtt{e}_2, \ldots, \mathtt{e}_n$ as elements of a vector space, or elementary vectors. The context determines which meaning is used.

8.2. **Subspaces.** Let V be a vector space over K. If a subset W of V is a vector space over K with respect to the same operations as V, then W is called a subspace of V. In other words, if W is closed under the operations (addition and scalar multiplication) of V, then W is called a subspace of V.

The zero vector space $\{\mathtt{0}\}$ and V itself are subspaces of V. The other subspaces are called proper subspaces.

(exercise02) Let $V = V^3$. Which of the following are subspaces of V.

$$(1)\ W_1 = \left\{ \begin{pmatrix} t \\ t+u \\ u \end{pmatrix} \mid t, u \in \mathbb{R} \right\} \qquad (2)\ W_2 = \left\{ t\begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix} + \begin{pmatrix} 3 \\ 3 \\ 1 \end{pmatrix} \mid t \in \mathbb{R} \right\}$$

$$(3)\ W_3 = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + t\begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix} + u\begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix} \mid t, u \in \mathbb{R} \right\}$$

(note) A subspace of V^3 is one of V^3 itself, a plane containing the origin, a line passing the origin, or the origin.

8.3. **Linear independence.** Let V be a vector space over K. For vectors $\mathtt{a}_1, \mathtt{a}_2, \ldots, \mathtt{a}_s$ of V, the expression:[3]

$$k_1\mathtt{a}_1 + k_2\mathtt{a}_2 + \cdots + k_s\mathtt{a}_s \qquad (k_1, \ldots, k_s \in K) \tag{9}$$

is called a linear combination of $\mathtt{a}_1, \mathtt{a}_2, \ldots, \mathtt{a}_s$ over K. Since V is closed under addition and scalar multiplication, all linear combinations are contained in V. A relationship among the vectors $\mathtt{a}_1, \mathtt{a}_2, \ldots, \mathtt{a}_s$:

$$k_1\mathtt{a}_1 + k_2\mathtt{a}_2 + \cdots + k_s\mathtt{a}_s = \mathtt{0} \tag{10}$$

is called a linear relation(ship) among the vectors over K.[4] This relation holds whenever all coefficients $k_1, \ldots, k_s$ are equal to zero. Such a relation is called the trivial linear relation. If there exists at least one nonzero coefficient, then it is called a non-trivial linear relation over K.

The vectors $\mathtt{a}_1, \mathtt{a}_2, \ldots, \mathtt{a}_s$ are defined to be linearly independent over K, if they have no non-trivial linear relations over K, whereas they are linearly dependent over K, if they have a non-trivial linear relation over K. The term "over K" is often omitted whenever there is no danger of confusion.

This definition is a generalization of linear independence or dependence of vectors in V^3.

The definition of linear independence of $\mathtt{a}_1, \mathtt{a}_2, \ldots, \mathtt{a}_s$ is rewritten as follows: For $k_1, \ldots, k_s \in K$,

$$\boxed{k_1\mathtt{a}_1 + k_2\mathtt{a}_2 + \cdots + k_s\mathtt{a}_s = \mathtt{0} \;\Rightarrow\; k_1 = k_2 = \cdots = k_s = 0}, \tag{11}$$

which is equivalent to the first definition.

(exercise03) (1) Are the four vectors $\begin{pmatrix}1\\2\\3\\4\end{pmatrix}, \begin{pmatrix}2\\3\\4\\1\end{pmatrix}, \begin{pmatrix}3\\4\\1\\2\end{pmatrix}, \begin{pmatrix}4\\1\\2\\3\end{pmatrix}$ linearly independent?
(2) Show that the following two conditions are equivalent. (i) $\mathtt{a}_1, \ldots, \mathtt{a}_s$ are linearly dependent. (ii) Some of $\mathtt{a}_1, \ldots, \mathtt{a}_s$ is expressed as a linear combination of the others.
(3) Show that if $\mathtt{a}_1, \ldots, \mathtt{a}_s$ are linearly independent, then any choice $\mathtt{a}_{i_1}, \ldots, \mathtt{a}_{i_t}$ from them are also linearly independent.
(ans) (1) Suppose that

$$k_1\begin{pmatrix}1\\2\\3\\4\end{pmatrix} + k_2\begin{pmatrix}2\\3\\4\\1\end{pmatrix} + k_3\begin{pmatrix}3\\4\\1\\2\end{pmatrix} + k_4\begin{pmatrix}4\\1\\2\\3\end{pmatrix} = \begin{pmatrix}0\\0\\0\\0\end{pmatrix}. \tag{12}$$

This is equivalent to

$$\begin{pmatrix}1&2&3&4\\2&3&4&1\\3&4&1&2\\4&1&2&3\end{pmatrix}\begin{pmatrix}k_1\\k_2\\k_3\\k_4\end{pmatrix} = \begin{pmatrix}0\\0\\0\\0\end{pmatrix}, \tag{13}$$

[3] If some term of (9) has the form $+(-k_i)\mathtt{a}_i$, then it is usually written as $k_1\mathtt{a}_1 + \cdots - k_i\mathtt{a}_i + \cdots + k_s\mathtt{a}_s$.
[4] Then we say that $\mathtt{a}_1, \mathtt{a}_2, \ldots, \mathtt{a}_s$ have a linear relation over K.

which is solved by using elementary operations: $k_1 = k_2 = k_3 = k_4 = 0$. (Or calculating the rank or determinant of the matrix, show that it is nonsingular, and therefore $k_1 = k_2 = k_3 = k_4 = 0$.) Accordingly, the given vectors are linearly independent.

Theorem 1. *Let* $\mathtt{a}_1, \mathtt{a}_2, \ldots, \mathtt{a}_s$ *be linearly independent vectors. If a vector* $\mathtt{a}$ *is not expressed as any linear combination of them, then* $\mathtt{a}_1, \mathtt{a}_2, \ldots, \mathtt{a}_s, \mathtt{a}$ *are linearly independent.*

Proof. Suppose $\mathtt{a}_1, \mathtt{a}_2, \ldots, \mathtt{a}_s$ are linearly independent and $\mathtt{a}$ is not expressed as any linear combination of them. Suppose $\mathtt{a}_1, \mathtt{a}_2, \ldots, \mathtt{a}_s, \mathtt{a}$ have a linear relation:

$$k_1\mathtt{a}_1 + k_2\mathtt{a}_2 + \cdots + k_s\mathtt{a}_s + k\mathtt{a} = \mathtt{0}. \tag{14}$$

If $k \neq 0$, then we have

$$\mathtt{a} = -\frac{k_1}{k}\mathtt{a}_1 - \frac{k_2}{k}\mathtt{a}_2 - \cdots - \frac{k_s}{k}\mathtt{a}_s, \tag{15}$$

which contradicts to the assumption. Hence $k = 0$. By (14),

$$k_1\mathtt{a}_1 + k_2\mathtt{a}_2 + \cdots + k_s\mathtt{a}_s = \mathtt{0}. \tag{16}$$

Since $\mathtt{a}_1, \mathtt{a}_2, \ldots, \mathtt{a}_s$ are linearly independent, we have $k_1 = k_2 = \cdots = k_s = 0$. Hence $\mathtt{a}_1, \mathtt{a}_2, \ldots, \mathtt{a}_s, \mathtt{a}$ are linearly independent. □

Theorem 2. *Suppose* $\mathtt{c}$ *is expressed as a linear combination of* $\mathtt{b}_1, \mathtt{b}_2, \ldots, \mathtt{b}_t$*, and every* $\mathtt{b}_j$ *is expressed as a linear combination of* $\mathtt{a}_1, \mathtt{a}_2, \ldots, \mathtt{a}_s$*. Then* $\mathtt{c}$ *is expressed as a linear combination of* $\mathtt{a}_1, \mathtt{a}_2, \ldots, \mathtt{a}_s$*.*

Proof. Suppose $\mathtt{c} = \sum_{j=1}^{t} k_j\mathtt{b}_j$ and $\mathtt{b}_j = \sum_{i=1}^{s} l_{ji}\mathtt{a}_i$, then

$$\mathtt{c} = \sum_{j=1}^{t} k_j \sum_{i=1}^{s} l_{ji}\mathtt{a}_i = \sum_{j=1}^{t}\sum_{i=1}^{s} k_j l_{ji}\mathtt{a}_i = \sum_{i=1}^{s}\sum_{j=1}^{t} k_j l_{ji}\mathtt{a}_i = \sum_{i=1}^{s}\left(\sum_{j=1}^{t} k_j l_{ji}\right)\mathtt{a}_i. \quad \square \tag{17}$$

8.4. **Finite-dimensional vector spaces.** If a vector space V has a property that finite number of vectors in V can be selected, so that every vector in V is expressed as a linear combination of the selected vectors, then V is called a finite-dimensional vector space. If a vector space is not finite-dimensional, then it is called infinite-dimensional. Hereafter, all vector spaces are assumed to be finite-dimensional.

8.5. **Bases of vector spaces.** A sequence of vectors $\mathtt{e}_1, \mathtt{e}_2, \dots, \mathtt{e}_n$ in V is called a basis of V if it satisfies the following properties:

I: $\mathtt{e}_1, \mathtt{e}_2, \dots, \mathtt{e}_n$ are linearly independent.
II: Any vector in V is expressed as a linear combination of $\mathtt{e}_1, \mathtt{e}_2, \dots, \mathtt{e}_n$.

These properties are joined as follows:

III: Any vector in V is expressed uniquely (up to order) as a linear combination of $\mathtt{e}_1, \mathtt{e}_2, \dots, \mathtt{e}_n$.

(exercise04) Show that I and II $\iff$ III.

(ans) *Proof.* ($\Rightarrow$) Suppose I and II hold. By II, any vector in V is expressed as a linear combination of $\mathtt{e}_i$'s. Then take any vector $\mathtt{x}$ in V, and suppose that it is expressed as follows:

$$\begin{aligned} \mathtt{x} &= k_1\mathtt{e}_1 + k_2\mathtt{e}_2 + \cdots + k_n\mathtt{e}_n \\ \mathtt{x} &= l_1\mathtt{e}_1 + l_2\mathtt{e}_2 + \cdots + l_n\mathtt{e}_n. \end{aligned} \tag{18}$$

Subtracting both sides, we have

$$\mathtt{0} = (k_1 - l_1)\mathtt{e}_1 + (k_2 - l_2)\mathtt{e}_2 + \cdots + (k_n - l_n)\mathtt{e}_n. \tag{19}$$

By I, $\mathtt{e}_1, \dots, \mathtt{e}_n$ are linearly independent, thus $k_1 - l_1 = k_2 - l_2 = \cdots = k_n - l_n = 0$, and therefore $k_1 = l_1$, $k_2 = l_2$, $\dots$, $k_n = l_n$. Hence III holds. □

Proof. ($\Leftarrow$) Suppose III holds. Then clearly II holds. Next letting

$$k_1\mathtt{e}_1 + k_2\mathtt{e}_2 + \cdots + k_n\mathtt{e}_n = \mathtt{0}, \tag{20}$$

then by III, $\mathtt{0}$ is expressed uniquely, and therefore all coefficients are equal to zero. This means that $\mathtt{e}_1, \dots, \mathtt{e}_n$ are linearly independent. □

(note) The elements of a basis are called basis vectors. The order of basis vectors is sensitive. A basis is usually represented as $\langle \mathtt{e}_1, \mathtt{e}_2, \dots, \mathtt{e}_n \rangle$ in $\langle \cdots \rangle$ brackets.

Theorem 3. *Let V be a vector space, and $\mathtt{a}_1, \mathtt{a}_2, \dots, \mathtt{a}_r$ be linearly independent vectors in V. Then we can attach several vectors in V to the vectors, so that $\langle \mathtt{a}_1, \mathtt{a}_2, \dots, \mathtt{a}_{r+s} \rangle$ is a basis of V.*

Proof. Since V is finite-dimensional, we can choose vectors $\mathtt{b}_1, \dots, \mathtt{b}_n$ so that every vector in V is expressible as a linear combination of them. If all of $\mathtt{b}_1, \dots, \mathtt{b}_n$ are expressible by $\mathtt{a}_1, \dots, \mathtt{a}_r$, then finish the operation. Otherwise, there exists some $\mathtt{b}_i$ which is not expressible by $\mathtt{a}_1, \dots, \mathtt{a}_r$, then from Theorem 1, it follows that we have linearly independent vectors $\mathtt{a}_1, \dots, \mathtt{a}_r, \mathtt{b}_i$. If all of $\mathtt{b}_1, \dots, \mathtt{b}_n$ are expressible by $\mathtt{a}_1, \dots, \mathtt{a}_r, \mathtt{b}_i$, then finish the operation. Otherwise, there exists some $\mathtt{b}_j$ which is not expressible by $\mathtt{a}_1, \dots, \mathtt{a}_r, \mathtt{b}_i$, then from Theorem 1, it follows that we have linearly independent vectors $\mathtt{a}_1, \dots, \mathtt{a}_r, \mathtt{b}_i, \mathtt{b}_j$. Repeating this operation, we have linearly independent vectors $\mathtt{a}_1, \dots, \mathtt{a}_r, \mathtt{a}_{r+1}, \dots, \mathtt{a}_{r+s}$ such that all of $\mathtt{b}_1, \dots, \mathtt{b}_n$ are expressible by them. Since every vector in V are expressible by $\mathtt{b}_1, \dots, \mathtt{b}_n$, by Theorem 2, every vector in V are expressible by $\mathtt{a}_1, \dots, \mathtt{a}_r, \mathtt{a}_{r+1}, \dots, \mathtt{a}_{r+s}$. Hence $\langle \mathtt{a}_1, \dots, \mathtt{a}_{r+s} \rangle$ is a basis of V. □

By Theorem 3, non-trivial vector spaces have their bases. The zero vector space has no bases.

8.6. **Dimensions of vector spaces and basis extensions.** We define the dimension of a vector space as follows:

Theorem 4. *Let V be a non-trivial vector space. There are many bases of V, however, the number of basis vectors is a constant independent of the bases, determined by V. The number of basis vectors is called the dimension of V, denoted by* $\dim V$.

Proof. $\Rightarrow$ 9.1.

(note) The dimension of zero vector space is defined to be 0. Non-trivial vector spaces have their bases, thus the dimensions of them are greater or equal to 1. Hence it holds that $\dim V = 0 \iff V = \{\mathtt{0}\}$.

Let V be a vector space, and $W \neq \{\mathtt{0}\}$ be a subspace of V. Take a basis $\langle \mathtt{a}_1, \ldots, \mathtt{a}_r \rangle$ of W. Then $\langle \mathtt{a}_1, \ldots, \mathtt{a}_r \rangle$ is linearly independent, thus by Theorem 3, a basis $\langle \mathtt{a}_1, \ldots, \mathtt{a}_r, \mathtt{a}_{r+1}, \ldots, \mathtt{a}_{r+s} \rangle$ of V can be made. Such a basis of V is called an extension of a basis of W.

Theorem 5. *Let V be a vector space and $W \neq \{\mathtt{0}\}$ be a subspace of V. We have a basis of V which is an extension of a basis of W.*

8.7. **Basis of K^n.** Let $\mathtt{e}_1, \ldots, \mathtt{e}_n$ be elementary vectors with n entries. Then $\langle \mathtt{e}_1, \ldots, \mathtt{e}_n \rangle$ is a basis of K^n. Thus it holds that

$$\dim K^n = n. \tag{21}$$

There are many bases of K^n, and the following is a test whether the vectors $\mathtt{a}_1, \ldots, \mathtt{a}_n$ in K^n form a basis of K^n or not.

Theorem 6.

$$\langle \mathtt{a}_1, \mathtt{a}_2, \ldots, \mathtt{a}_n \rangle \textit{ is a basis of } K^n \iff \begin{vmatrix} \mathtt{a}_1 & \mathtt{a}_2 & \ldots & \mathtt{a}_n \end{vmatrix} \neq 0 \tag{22}$$

Proof. $\Rightarrow$ 9.2.

(exercise05) $V = \left\{ \begin{pmatrix} x & y \\ z & w \end{pmatrix} \mid x, y, z, w \in K \right\}$ is a vector space over K. Confirm that $\langle \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \rangle$ is a basis of V. Hence $\dim V = 4$.

8.8. **Subspaces spanned by several vectors.** Let V be a vector space over K, and $\mathtt{a}_1, \dots, \mathtt{a}_s$ be vectors in V. Then the set:

$$W = \{k_1\mathtt{a}_1 + \cdots + k_s\mathtt{a}_s \mid k_1, \dots, k_s \in K\} \tag{23}$$

is a subspace of V, which is called a subspace spanned by $\mathtt{a}_1, \dots, \mathtt{a}_s$ (or W is spanned by $\mathtt{a}_1, \dots, \mathtt{a}_s$), and is denoted by $W = \mathrm{span}\{\mathtt{a}_1, \dots, \mathtt{a}_s\}$.[5] The set $\{\mathtt{a}_1, \dots, \mathtt{a}_s\}$ is called a generator of W. Using the terms, the condition II for bases is rewritten by II' or II"' as follows.

II': $\boxed{V \text{ is spanned by } \mathtt{e}_1, \dots, \mathtt{e}_n.}$

II": $\boxed{V = \mathrm{span}\{\mathtt{e}_1, \dots, \mathtt{e}_n\}.}$

(exercise06) Prove that the above mentioned W is a subspace of V.

(ans) For any vectors $\mathtt{x} = k_1\mathtt{a}_1 + \cdots + k_s\mathtt{a}_s$ and $\mathtt{x}' = k_1'\mathtt{a}_1 + \cdots + k_s'\mathtt{a}_s$, and for any $k \in K$, we have

$$\begin{aligned} \mathtt{x} + \mathtt{x}' = &\ (k_1\mathtt{a}_1 + \cdots + k_s\mathtt{a}_s) + (k_1'\mathtt{a}_1 + \cdots + k_s'\mathtt{a}_s) \\ = &\ (k_1 + k_1')\mathtt{a}_1 + \cdots + (k_s + k_s')\mathtt{a}_s = \tilde{k}_1\mathtt{a}_1 + \cdots + \tilde{k}_s\mathtt{a}_s \in W, \\ k\mathtt{x} = &\ k(k_1\mathtt{a}_1 + \cdots + k_s\mathtt{a}_s) = kk_1\mathtt{a}_1 + \cdots + kk_s\mathtt{a}_s \\ = &\ k_1''\mathtt{a}_1 + \cdots + k_s''\mathtt{a}_s \in W. \quad \square \end{aligned} \tag{24}$$

8.9. **Sum of subspaces.** Let V be a vector space over K, and let W_1, W_2 be two subspaces of V. The sum of W_1 and W_2 is defined as

$$W_1 + W_2 = \{\mathtt{w}_1 + \mathtt{w}_2 \mid \mathtt{w}_1 \in W_1,\ \mathtt{w}_2 \in W_2\}. \tag{25}$$

This is a subspace of V. The intersection $W_1 \cap W_2$ of W_1 and W_2 is also a subspace of V.
(exercise07) Show that $W_1 + W_2$ and $W_1 \cap W_2$ are subspaces of V.

Proof. We show that $W_1 + W_2$ is a subspace of V. Take any vectors $\mathtt{w}_1 + \mathtt{w}_2$ and $\mathtt{w}_1' + \mathtt{w}_2'$ ($\mathtt{w}_1, \mathtt{w}_1' \in W_1$, $\mathtt{w}_2, \mathtt{w}_2' \in W_2$) in $W_1 + W_2$, and take any $k \in K$. Since W_1 and W_2 are subspaces of V, we have $\mathtt{w}_1 + \mathtt{w}_1' \in W_1$ and $\mathtt{w}_2 + \mathtt{w}_2' \in W_2$. Hence

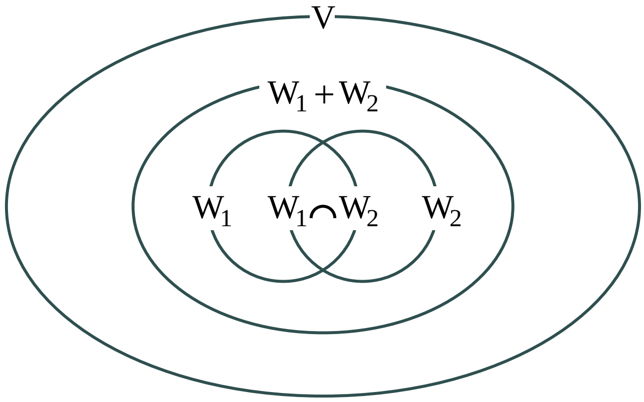
$$(\mathtt{w}_1 + \mathtt{w}_2) + (\mathtt{w}_1' + \mathtt{w}_2') = (\mathtt{w}_1 + \mathtt{w}_1') + (\mathtt{w}_2 + \mathtt{w}_2') \in W_1 + W_2. \tag{26}$$

Similarly, as W_1 and W_2 are subspaces, we have $\mathtt{w}_1 + \mathtt{w}_1' \in W_1$ and $\mathtt{w}_2 + \mathtt{w}_2' \in W_2$. Hence

$$k(\mathtt{w}_1 + \mathtt{w}_2) = k\mathtt{w}_1 + k\mathtt{w}_2 \in W_1 + W_2. \quad \square \tag{27}$$

Proof. Next we show that $W_1 \cap W_2$ is a subspace of V. Take any vectors $\mathtt{x}, \mathtt{y}$ in $W_1 \cap W_2$ and any $k \in K$. From $\mathtt{x}, \mathtt{y} \in W_1$, it follows that $\mathtt{x} + \mathtt{y} \in W_1$. Similarly, from $\mathtt{x}, \mathtt{y} \in W_2$ it follows that $\mathtt{x} + \mathtt{y} \in W_2$. Therefore $\mathtt{x} + \mathtt{y} \in W_1 \cap W_2$. In like manner, we have $k\mathtt{x} \in W_1$ and $k\mathtt{x} \in W_2$, and therefore $k\mathtt{x} \in W_1 \cap W_2$. $\square$

[5] $W = \mathrm{span}\{\mathtt{a}_1, \dots, \mathtt{a}_s\}$ is the intersection of all subspaces containing $\{\mathtt{a}_1, \dots, \mathtt{a}_s\}$. Hence W is also the minimal subspace containing $\{\mathtt{a}_1, \dots, \mathtt{a}_s\}$.



(note) The intersection of more than two subspaces is also a subspace.
(note) $W_1 \cup W_2$ is, in general, not a subspace of V.

The dimensions of subspaces satisfy the following, which is called the dimension formula for subspaces.

Theorem 7. *Let W_1 and W_2 be subspaces of V, then it holds that*

$$\dim W_1 + \dim W_2 = \dim(W_1 + W_2) + \dim(W_1 \cap W_2). \tag{28}$$

Proof. Let a basis of $W_1 \cap W_2$ be $\langle \mathtt{a}_1, \dots, \mathtt{a}_r \rangle$. By Theorem 5, extending this basis, we have a basis of W_1: $\langle \mathtt{a}_1, \dots, \mathtt{a}_r, \mathtt{b}_1, \dots, \mathtt{b}_s \rangle$ and a basis of W_2: $\langle \mathtt{a}_1, \dots, \mathtt{a}_r, \mathtt{c}_1, \dots, \mathtt{c}_t \rangle$. Here, we show that $\langle \mathtt{a}_1, \dots, \mathtt{a}_r, \mathtt{b}_1, \dots, \mathtt{b}_s, \mathtt{c}_1, \dots, \mathtt{c}_t \rangle$ is a basis of $W_1 + W_2$.

First we show it is linearly independent. Suppose there exists the following linear relation:

$$\begin{aligned} & k_1\mathtt{a}_1 + \cdots + k_r\mathtt{a}_r + l_1\mathtt{b}_1 + \cdots + l_s\mathtt{b}_s + m_1\mathtt{c}_1 + \cdots + m_t\mathtt{c}_t = \mathtt{0}. \\ \therefore \quad & k_1\mathtt{a}_1 + \cdots + k_r\mathtt{a}_r + l_1\mathtt{b}_1 + \cdots + l_s\mathtt{b}_s = -m_1\mathtt{c}_1 - \cdots - m_t\mathtt{c}_t. \end{aligned} \tag{29}$$

Here, it is clear that the left-hand side represents an element of W_1 and the right-hand side represents an element of W_2. Consequently, both sides represent an element of $W_1 \cap W_2$. However, since $\langle \mathtt{a}_1, \dots, \mathtt{a}_r, \mathtt{b}_1, \dots, \mathtt{b}_s \rangle$ is a basis of W_1, from the uniqueness of expression of the left-hand side by the basis, it follows that $l_1 = \cdots = l_s = 0$. Returning to (29), we have

$$k_1\mathtt{a}_1 + \cdots + k_r\mathtt{a}_r + m_1\mathtt{c}_1 + \cdots + m_t\mathtt{c}_t = \mathtt{0}. \tag{30}$$

Furthermore, as $\langle \mathtt{a}_1, \dots, \mathtt{a}_r, \mathtt{c}_1, \dots, \mathtt{c}_t \rangle$ is a basis of W_2, we have

$$k_1 = \cdots = k_r = m_1 = \cdots = m_t = 0. \tag{31}$$

Next, an arbitrary vector in $W_1 + W_2$ is expressed in the form: $\mathtt{w}_1 + \mathtt{w}_2$ ($\mathtt{w}_1 \in W_1$, $\mathtt{w}_2 \in W_2$). Now $\mathtt{w}_1$ and $\mathtt{w}_2$ are expressed, by using the basis, as follows:

$$\begin{aligned} \mathtt{w}_1 &= k_1\mathtt{a}_1 + \cdots + k_r\mathtt{a}_r + l_1\mathtt{b}_1 + \cdots + l_s\mathtt{b}_s, \\ \mathtt{w}_2 &= k_1'\mathtt{a}_1 + \cdots + k_r'\mathtt{a}_r + m_1\mathtt{c}_1 + \cdots + m_t\mathtt{c}_t. \end{aligned} \tag{32}$$

Thus

$$\begin{aligned} \mathtt{w}_1 + \mathtt{w}_2 \quad &= (k_1 + k_1')\mathtt{a}_1 + \cdots + (k_r + k_r')\mathtt{a}_r + l_1\mathtt{b}_1 + \cdots + l_s\mathtt{b}_s \\ &\quad + m_1\mathtt{c}_1 + \cdots + m_t\mathtt{c}_t, \end{aligned} \tag{33}$$

say, $\mathtt{w}_1 + \mathtt{w}_2$ is expressed as a linear combination of $\mathtt{a}_1, \dots, \mathtt{a}_r, \mathtt{b}_1, \dots, \mathtt{b}_s, \mathtt{c}_1, \dots, \mathtt{c}_t$.

Therefore we see that $\langle \mathtt{a}_1, \dots, \mathtt{a}_r, \mathtt{b}_1, \dots, \mathtt{b}_s, \mathtt{c}_1, \dots, \mathtt{c}_t \rangle$ is a basis of $W_1 + W_2$. Then noting the dimensions of subspaces,

$$\begin{aligned} &\dim W_1 + \dim W_2 = (r+s) + (r+t) = (r+s+t) + r \\ &= \dim(W_1 + W_2) + \dim(W_1 \cap W_2). \quad \square \end{aligned} \tag{34}$$

8.10. **Direct sum of subspaces.** If an arbitrary vector in $W = W_1 + W_2$ is uniquely expressed in the form:

$$\mathtt{w}_1 + \mathtt{w}_2 \qquad (\mathtt{w}_1 \in W_1,\ \mathtt{w}_2 \in W_2), \tag{35}$$

then W is called the direct sum of W_1 and W_2, denoted by $W = W_1 \oplus W_2$. This definition is equivalent to the following: for $\mathtt{w}_1 \in W_1$ and $\mathtt{w}_2 \in W_2$,

$$\mathtt{w}_1 + \mathtt{w}_2 = \mathtt{0} \ \Rightarrow\ \mathtt{w}_1 = \mathtt{w}_2 = \mathtt{0}. \tag{36}$$

Theorem 8. *Letting $W = W_1 + W_2$, we have*

$$W = W_1 \oplus W_2 \overset{(1)}{\iff} W_1 \cap W_2 = \{\mathtt{0}\} \overset{(2)}{\iff} \dim W = \dim W_1 + \dim W_2. \tag{37}$$

Proof. (2) is clear by Theorem 7. We show (1).
($\Rightarrow$) By reduction to absurdity. Let $W = W_1 \oplus W_2$. Suppose $W_1 \cap W_2 \neq \{\mathtt{0}\}$. Then there exists $\mathtt{a} \in W_1 \cap W_2$ such that $\mathtt{a} \neq \mathtt{0}$. Since $W_1 \cap W_2$ is a subspace, it is clear that $-\mathtt{a} \in W_1 \cap W_2$. Hence $\mathtt{0} \in W$ is expressed two ways as $\mathtt{0} = \mathtt{0} + \mathtt{0} = \mathtt{a} + (-\mathtt{a})$ by the sum of a vector in W_1 and a vector in W_2, which is a contradiction. $\square$
($\Leftarrow$) Let $W_1 \cap W_2 = \{\mathtt{0}\}$. For an arbitrary vector $\mathtt{w}$ in W, suppose

$$\mathtt{w} = \mathtt{w}_1 + \mathtt{w}_2 = \mathtt{w}_1' + \mathtt{w}_2' \qquad (\mathtt{w}_1, \mathtt{w}_1' \in W_1,\ \mathtt{w}_2, \mathtt{w}_2' \in W_2). \tag{38}$$

Then we have

$$W_1 \ni \mathtt{w}_1 - \mathtt{w}_1' = \mathtt{w}_2' - \mathtt{w}_2 \in W_2. \tag{39}$$

Hence both sides are a vector in $W_1 \cap W_2$, say, $\mathtt{0}$.

$$\therefore \quad \mathtt{w}_1 - \mathtt{w}_1' = \mathtt{w}_2' - \mathtt{w}_2 = \mathtt{0}. \qquad \therefore \quad \mathtt{w}_1 = \mathtt{w}_1', \quad \mathtt{w}_2 = \mathtt{w}_2'. \tag{40}$$

This shows that $W = W_1 \oplus W_2$. $\square$

8.11. **Sum or direct sum of several subspaces.** Let V be a vector space over K and $W_1, W_2, \dots, W_s$ be subspaces of V. We define the sum of the subspaces $W_1, \dots, W_s$ as

$$W_1 + \cdots + W_s = \{\mathtt{w}_1 + \cdots + \mathtt{w}_s \mid \ \mathtt{w}_i \in W_i \quad (i = 1, \dots, s)\}. \tag{41}$$

This is a subspace of V, because for any vectors $\mathtt{w}_1 + \cdots + \mathtt{w}_s$ and $\mathtt{w}_1' + \cdots + \mathtt{w}_s'$ in $W_1 + \cdots + W_s$, and any $k \in K$, we have

$$\begin{aligned} (\mathtt{w}_1 + \cdots + \mathtt{w}_s) + (\mathtt{w}_1' + \cdots + \mathtt{w}_s') &= (\mathtt{w}_1 + \mathtt{w}_1') + \cdots + (\mathtt{w}_s + \mathtt{w}_s') \in W_1 + \cdots + W_s, \\ k(\mathtt{w}_1 + \cdots + \mathtt{w}_s) &= k\mathtt{w}_1 + \cdots + k\mathtt{w}_s \in W_1 + \cdots + W_s. \end{aligned} \tag{42}$$

If an arbitrary vector in $W = W_1 + \cdots + W_s$ is uniquely expressed in the form:

$$\mathtt{w}_1 + \cdots + \mathtt{w}_s \qquad (\mathtt{w}_1 \in W_1, \dots, \mathtt{w}_s \in W_s), \tag{43}$$

then W is called the direct sum of $W_1, \dots, W_s$, denoted by $W = W_1 \oplus \cdots \oplus W_s$. This definition is equivalent to the following: for $\mathtt{w}_1 \in W_1, \dots, \mathtt{w}_s \in W_s$,

$$\mathtt{w}_1 + \cdots + \mathtt{w}_s = \mathtt{0} \;\Rightarrow\; \mathtt{w}_1 = \cdots = \mathtt{w_s} = \mathtt{0}. \tag{44}$$

The following holds.

Theorem 9. *A sequence made by joining bases of subspaces $W_1, \dots, W_s$ forms a basis of $W = W_1 \oplus \cdots \oplus W_s$.*

Proof. Let bases of $W_1, W_2, \dots, W_s$ be $\langle \mathtt{a}_1, \mathtt{a}_2, \dots, \mathtt{a}_p \rangle$, $\langle \mathtt{b}_1, \mathtt{b}_2, \dots, \mathtt{b}_q \rangle, \dots, \langle \mathtt{c}_1, \mathtt{c}_2, \dots, \mathtt{c}_r \rangle$, respectively. Joining these bases, we have $\mathbb{E} = \langle \mathtt{a}_1, \dots, \mathtt{a}_p, \mathtt{b}_1, \dots, \mathtt{b}_q, \dots, \mathtt{c}_1, \dots, \mathtt{c}_r \rangle$, which shall be proved to be a basis of $W = W_1 \oplus \cdots \oplus W_s$.

(To span W) For any vector $\mathtt{w}$ in W, we have $\mathtt{w} = \mathtt{w}_1 + \cdots + \mathtt{w}_s$ ($\mathtt{w}_i \in W_i$), and every $\mathtt{w}_i$ is expressed by $\mathbb{E}$, therefore $\mathtt{w}$ is also expressed by $\mathbb{E}$.

(Linear independence) Suppose $k_1\mathtt{a}_1 + \cdots + k_p\mathtt{a}_p + l_1\mathtt{b}_1 + \cdots + l_q\mathtt{b}_q + \cdots + m_1\mathtt{c}_1 + \cdots + m_r\mathtt{c}_r = \mathtt{0}$. Write $k_1\mathtt{a}_1 + \cdots + k_p\mathtt{a}_p = \mathtt{w}_1$, $l_1\mathtt{b}_1 + \cdots + l_q\mathtt{b}_q = \mathtt{w}_2$, $\dots$, then $\mathtt{w}_1 + \mathtt{w}_2 + \cdots + \mathtt{w}_s = \mathtt{0}$. However, since W is the direct sum of W_i's, by (44) $\mathtt{w}_1 = \mathtt{w}_2 = \cdots = \mathtt{w}_s = \mathtt{0}$. Therefore all coefficients $k_1, \dots, k_p, \dots, m_1, \dots, m_r$ vanish.
Consequently, $\mathbb{E}$ is a basis of W. □

Theorem 10. *Letting $W = W_1 + \cdots + W_s$, we have*

$$W = W_1 \oplus \cdots \oplus W_s \iff \dim W = \dim W_1 + \cdots + \dim W_s. \tag{45}$$

Proof. ($\Rightarrow$) It is clear by Theorem 9.
($\Leftarrow$) Suppose $\dim W = \dim W_1 + \cdots + \dim W_s$. As in the proof of Theorem 9, join bases of $W_1, W_2, \dots, W_s$ to have $\mathbb{E} = \langle \mathtt{a}_1, \dots, \mathtt{a}_p, \mathtt{b}_1, \dots, \mathtt{b}_q, \dots, \mathtt{c}_1, \dots, \mathtt{c}_r \rangle$, which we prove to be a basis of W.

(To span W) Similar to the proof of Theorem 9.

(Linear independence) If $\mathbb{E}$ is linearly dependent, then some vector in $\mathbb{E}$ is expressed by the other vectors in $\mathbb{E}$. Hence removing the vector from $\mathbb{E}$, we have $\mathbb{E}'$ which spans W. Repeating this process, we have a basis $\mathbb{E}^{(u)}$, however, from $\dim W = \dim W_1 + \cdots + \dim W_s$ it follows that $u = 0$. That is, $\mathbb{E}$ is a basis of W.

If an arbitrary vector $\mathtt{w}$ in W is expressed as $\mathtt{w} = \mathtt{w}_1 + \cdots + \mathtt{w}_s = \mathtt{w}'_1 + \cdots + \mathtt{w}'_s$, then expressing every $\mathtt{w}_i$ and $\mathtt{w}'_i$ by the above mentioned basis of W_i, $\mathtt{w}$ is expressed by $\mathbb{E}$ in two ways:

$$\begin{array}{l} \mathtt{w} = k_1\mathtt{a}_1 + \cdots + k_p\mathtt{a}_p + l_1\mathtt{b}_1 + \cdots + l_q\mathtt{b}_q + \cdots + m_1\mathtt{c}_1 + \cdots + m_r\mathtt{c}_r \\ \mathtt{w} = k'_1\mathtt{a}_1 + \cdots + k'_p\mathtt{a}_p + l'_1\mathtt{b}_1 + \cdots + l'_q\mathtt{b}_q + \cdots + m'_1\mathtt{c}_1 + \cdots + m'_r\mathtt{c}_r. \end{array} \tag{46}$$

Since corresponding coefficients are the same, we have $\mathtt{w}_i = \mathtt{w}'_i$ ($i = 1, \dots, s$). □

(exercise08) Let W and X be subspaces of $V = V^3$ defined as follows.

$$W = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mid 2x + 3y - 3z = 0 \right\}, \quad X = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mid \begin{array}{l} x - y + 6z = 0 \\ 3x + 2y + 3z = 0 \end{array} \right\} \tag{47}$$

(1) Determine a basis of $W + X$. (2) Determine a basis of $W \cap X$.
(3) Is $W + X$ the direct sum?
(ans) (1) First of all, to fine a basis of W, solve $2x + 3y - 3z = 0$.

$$\begin{pmatrix} 2 & 3 & -3 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & \frac{3}{2} & -\frac{3}{2} \end{pmatrix}. \qquad \therefore \ \mathtt{x} = \alpha \begin{pmatrix} -\frac{3}{2} \\ 1 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} \frac{3}{2} \\ 0 \\ 1 \end{pmatrix}$$
$$= \tilde{\alpha} \begin{pmatrix} -3 \\ 2 \\ 0 \end{pmatrix} + \tilde{\beta} \begin{pmatrix} 3 \\ 0 \\ 2 \end{pmatrix}. \quad \text{Hence a basis of } W \text{ is } \left\langle \begin{pmatrix} -3 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ 0 \\ 2 \end{pmatrix} \right\rangle. \tag{48}$$

Next we solve the system of linear equations of X.

$$\begin{pmatrix} 1 & -1 & 6 \\ 3 & 2 & 3 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & -1 & 6 \\ 0 & 5 & -15 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & -1 & 6 \\ 0 & 1 & -3 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & -3 \end{pmatrix}.$$
$$\therefore \ \mathtt{x} = \alpha \begin{pmatrix} -3 \\ 3 \\ 1 \end{pmatrix}. \quad \text{Hence a basis of } X \text{ is } \left\langle \begin{pmatrix} -3 \\ 3 \\ 1 \end{pmatrix} \right\rangle. \tag{49}$$

Finally, we find a basis of $W + X$. Consider the matrix consisting of bases of W and X, and transform it by elementary column operations into a "stair-like" form.

$$\begin{pmatrix} -3 & 3 & -3 \\ 2 & 0 & 3 \\ 0 & 2 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} -3 & 0 & 0 \\ 2 & 2 & 1 \\ 0 & 2 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} -3 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} \tag{50}$$

Therefore a basis of $W + X$ is $\left\langle \begin{pmatrix} -3 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right\rangle$.

(2) To determine $W \cap X$, it suffices to solve the following.

$$\begin{cases} 2x + 3y - 3z = 0 \\ x - y + 6z = 0 \\ 3x + 2y + 3z = 0 \end{cases} \tag{51}$$

$$\begin{pmatrix} 2 & 3 & -3 \\ 1 & -1 & 6 \\ 3 & 2 & 3 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & -1 & 6 \\ 2 & 3 & -3 \\ 3 & 2 & 3 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & -1 & 6 \\ 0 & 5 & -15 \\ 0 & 5 & -15 \end{pmatrix} \longrightarrow$$
$$\begin{pmatrix} 1 & -1 & 6 \\ 0 & 1 & -3 \\ 0 & 0 & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & -3 \\ 0 & 0 & 0 \end{pmatrix}. \qquad \therefore \ \mathtt{x} = \alpha \begin{pmatrix} -3 \\ 3 \\ 1 \end{pmatrix}. \tag{52}$$

Thus a basis of $W \cap X$ is $\left\langle \begin{pmatrix} -3 \\ 3 \\ 1 \end{pmatrix} \right\rangle$.

(3) $W + X$ is not the direct sum because $W \cap X \neq \{\mathtt{0}\}$.
(Alternative solution) $\dim(W+X) = 2$, $\dim W + \dim X = 2+1 = 3$. $\therefore \dim(W+X) \neq \dim W + \dim X$. Accordingly, $W + X$ is not the direct sum.

★

9. BASIC THEOREMS CONCERNING BASES

★ ★

KEYWORDS: BASES OF VECTOR SPACES, DIMENSION, BASIS OF K^n

9.1. **Dimensions of vector spaces.** To define the dimension of a vector space, the following theorem is needed.

Theorem 1. *Take a vector space V and fix it. Then the number of vectors contained in a basis of V is determined by V, independent of the choice of a basis. This number is called the dimension of V.*

Proof. By reduction to absurdity. Let $\mathbb{E} = \langle \mathtt{e}_1, \mathtt{e}_2, \ldots, \mathtt{e}_m \rangle$ and $\mathbb{F} = \langle \mathtt{f}_1, \mathtt{f}_2, \ldots, \mathtt{f}_n \rangle$ $(m < n)$ be two bases of V. Then there exists a vector in $\mathtt{e}_1, \ldots, \mathtt{e}_m$, such that it is not expressible by any linear combination of $\mathtt{f}_2, \ldots, \mathtt{f}_n$. Because if all of $\mathtt{e}_1, \ldots, \mathtt{e}_m$ are expressible by $\mathtt{f}_2, \ldots, \mathtt{f}_n$, then noting that $\mathtt{f}_1$ is expressible by $\mathtt{e}_1, \ldots, \mathtt{e}_m$, we see that $\mathtt{f}_1$ is expressible by $\mathtt{f}_2, \ldots, \mathtt{f}_n$, which contradicts linear independence of $\mathtt{f}_1, \ldots, \mathtt{f}_n$. Hence we choose a vector $\mathtt{e}_i$ which is not expressible by $\mathtt{f}_2, \ldots, \mathtt{f}_n$. Then we show that $\mathbb{F}' = \langle \mathtt{e}_i, \mathtt{f}_2, \mathtt{f}_3 \ldots, \mathtt{f}_n \rangle$ is a basis of V.

First of all, $\mathbb{F}'$ is made by adding a vector not expressible by linearly independent vectors, and therefore $\mathbb{F}'$ is linearly independent. (Chapter 8, Theorem 1)

Next we show that any vector of V is expressible by $\mathbb{F}'$. As $\mathbb{F}$ is a basis, $\mathtt{e}_i = c_1\mathtt{f}_1 + \cdots + c_n\mathtt{f}_n$. But if $c_1 = 0$, then e_i is expressed by $\mathtt{f}_2, \ldots, \mathtt{f}_n$, which contradicts the assumption, and thus, $c_1 \neq 0$. Hence $\mathtt{f}_1 = \frac{1}{c_1}(\mathtt{e}_i - c_2\mathtt{f}_2 - \cdots - c_n\mathtt{f}_n)$, say, $\mathtt{f}_1$ is expressed by $\mathbb{F}'$. Also, $\mathbb{F}'$ contains vectors $\mathtt{f}_2, \ldots, \mathtt{f}_n$, and any vector of V is expressed by $\mathtt{f}_1, \ldots, \mathtt{f}_n$. Consequently, any vector of V is expressed by $\mathbb{F}'$.

Therefore $\mathbb{F}'$ is a basis of V. As we make $'$ from $\mathbb{F}$, we can again replace $\mathtt{f}_2$ by some vector of $\mathtt{e}_1, \ldots, \mathtt{e}_m$, and have a new basis $\mathbb{F}''$ of V. Repeating this process, all vectors of $\mathbb{F}$ are replaced by some vectors of $\mathtt{e}_1, \ldots, \mathtt{e}_m$ again and again, and finally we have a basis $\mathbb{G} = \langle \mathtt{e}_{i_1}, \mathtt{e}_{i_2}, \ldots, \mathtt{e}_{i_n} \rangle$ of V. However, since $m < n$, $\mathbb{G}$ contains the same vectors, which contradicts that $\mathbb{G}$ is a basis. □

9.2. **Bases of K^n.** The following is a criterion to determine whether n vectors $\mathtt{a}_1, \mathtt{a}_2, \ldots, \mathtt{a}_n$ of K^n is a basis of K^n.

Theorem 2. *It holds that*

$$\langle \mathtt{a}_1, \mathtt{a}_2, \ldots, \mathtt{a}_n \rangle \text{ is a basis of } K^n \iff \begin{vmatrix} \mathtt{a}_1 & \mathtt{a}_2 & \ldots & \mathtt{a}_n \end{vmatrix} \neq 0. \tag{1}$$

Proof. Denote the left-hand side condition by (L), and the right-hand side condition by (R). Also consider the conditions below.

(R'): A matrix $A = \begin{pmatrix} \mathtt{a}_1 & \mathtt{a}_2 & \ldots & \mathtt{a}_n \end{pmatrix}$ is nonsingular.
(I): Vectors $\mathtt{a}_1, \mathtt{a}_2, \ldots, \mathtt{a}_n$ are linearly independent.
(II): Any vectors of K^n is expressed as a linear combination of $\mathtt{a}_1, \mathtt{a}_2, \ldots, \mathtt{a}_n$.

For these conditions, the following relations are proved, and consequently, we have the equivalence of (L) and (R). Here, it suffices to show the equivalences (i) and (ii).

$$\boxed{\boxed{\text{(L)}} \iff \left\{ \begin{array}{l} \text{(I)} \quad \overset{\text{(i)}}{\iff} \\ \\ \text{(II)} \quad \overset{\text{(ii)}}{\iff} \end{array} \right. \boxed{\text{(R')}} \iff \boxed{\text{(R)}}} \tag{2}$$

Proof of (i): ($\Longleftrightarrow$). Letting $A = \begin{pmatrix} \mathtt{a}_1 & \mathtt{a}_2 & \ldots & \mathtt{a}_n \end{pmatrix}$, we show the contraposition instead of (i), say,

$$A \text{ is singular } \iff \mathtt{a}_1, \ldots, \mathtt{a}_n \text{ are linearly dependent.} \tag{3}$$

By Chapter 6, Theorem 2, we have

$$A \text{ is singular } \iff A\mathtt{x} = \mathtt{0} \text{ has non-trivial solutions.} \tag{4}$$

Since the right-hand side conditions of (3) and (4) are equivalent, (3) holds. □

Proof of (i): ($\Rightarrow$). We also give an alternate proof of the above mentioned (i), not using the theory of systems of linear equations. First we prove (3):($\Rightarrow$) instead of (i):($\Rightarrow$). Suppose A is singular and $r(A) = r < n$. By elementary operations, A is transformed into $F_{nn}(r)$. Hence by nonsingular matrices P, Q, $PAQ = F_{nn}(r)$. Therefore

$$AQ = P^{-1}F_{nn}(r) = \begin{pmatrix} * & \cdots & * & 0 & \cdots & 0 \\ * & \cdots & * & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ * & \cdots & * & 0 & \cdots & 0 \end{pmatrix}. \tag{5}$$

Here, the number of zero column vectors of the right-hand side is $(n - r)$. Thus letting the $(r + 1)$-th column of Q be $\mathtt{q} = {}^t(q_1, \ldots, q_n)$, we have $A\mathtt{q} = \mathtt{0}$. Hence $q_1\mathtt{a}_1 + q_2\mathtt{a}_2 + \cdots + q_n\mathtt{a}_n = \mathtt{0}$. By nonsingularity of Q, $\mathtt{q} \neq \mathtt{0}$, say, there exists nonzero q_i. Consequently, $\mathtt{a}_1, \ldots, \mathtt{a}_n$ are linearly dependent. □

Proof of (i): ($\Leftarrow$). Let A be nonsingular. Suppose $k_1\mathtt{a}_1 + k_2\mathtt{a}_2 + \cdots + k_n\mathtt{a}_n = \mathtt{0}$. Letting $\mathtt{k} = {}^t(k_1, \ldots, k_n)$, we have $A\mathtt{k} = \mathtt{0}$, and A is nonsingular, we have $\mathtt{k} = \mathtt{0}$, say, $k_1 = k_2 = \cdots = k_n = 0$. Hence $\mathtt{a}_1, \ldots, \mathtt{a}_n$ are linearly independent. □

Proof of (ii): $(\Rightarrow)$. Any vectors of K^n are expressed as linear combinations of $\mathtt{a}_1, \mathtt{a}_2, \ldots, \mathtt{a}_n$. Hence elementary vectors with n entries: $\mathtt{e}_1, \mathtt{e}_2, \ldots, \mathtt{e}_n$ are expressed as linear combinations of $\mathtt{a}_1, \mathtt{a}_2, \ldots, \mathtt{a}_n$. This fact is written by matrices as

$$AQ = \begin{pmatrix} \mathtt{a}_1 & \ldots & \mathtt{a}_n \end{pmatrix} \begin{pmatrix} q_{11} & q_{12} & \cdots & q_{1n} \\ q_{21} & q_{22} & \cdots & q_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ q_{n1} & q_{n2} & \cdots & q_{nn} \end{pmatrix} = \begin{pmatrix} \mathtt{e}_1 & \ldots & \mathtt{e}_n \end{pmatrix} = E_n. \tag{6}$$

This shows that A is nonsingular. □

Proof of (ii): $(\Leftarrow)$. Let A be nonsingular. Take any vector $\mathtt{x}$ of K^n. Letting $k_1\mathtt{a}_1 + k_2\mathtt{a}_2 + \cdots + k_n\mathtt{a}_n = \mathtt{x}$, by matrices,

$$A\mathtt{k} = \mathtt{x}. \tag{7}$$

Since A is nonsingular, the coefficients are determined by $\mathtt{k} = A^{-1}\mathtt{x}$. Therefore $\mathtt{x}$ is expressed as a linear combination of $\mathtt{a}_1, \mathtt{a}_2, \ldots, \mathtt{a}_n$. □

(exercise01) Let $V = \mathbb{C}^4$. Determine the necessary and sufficient condition for

$\left\langle \begin{pmatrix} c \\ 3 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ c \\ 3 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ c \\ 3 \end{pmatrix}, \begin{pmatrix} 3 \\ 2 \\ 1 \\ c \end{pmatrix} \right\rangle$ to be a basis of V.

(ans)

$$\begin{aligned}
&\begin{vmatrix} c & 1 & 2 & 3 \\ 3 & c & 1 & 2 \\ 2 & 3 & c & 1 \\ 1 & 2 & 3 & c \end{vmatrix} = \begin{vmatrix} c & 1 & 2 & c+6 \\ 3 & c & 1 & c+6 \\ 2 & 3 & c & c+6 \\ 1 & 2 & 3 & c+6 \end{vmatrix} = (c+6)\begin{vmatrix} c & 1 & 2 & 1 \\ 3 & c & 1 & 1 \\ 2 & 3 & c & 1 \\ 1 & 2 & 3 & 1 \end{vmatrix} \\
&= (c+6)\begin{vmatrix} c & 1-c & 1 & 1 \\ 3 & c-3 & 1-c & 1 \\ 2 & 1 & c-3 & 1 \\ 1 & 1 & 1 & 1 \end{vmatrix} = (c+6)\begin{vmatrix} c-1 & -c & 0 & 1 \\ 2 & c-4 & -c & 1 \\ 1 & 0 & c-4 & 1 \\ 0 & 0 & 0 & 1 \end{vmatrix} \\
&= (c+6)\begin{vmatrix} c-1 & -c & 0 \\ 2 & c-4 & -c \\ 1 & 0 & c-4 \end{vmatrix} = (c+6)\left[(c-1)(c-4)^2 + c^2 + 2c(c-4)\right] \\
&= (c+6)(c^3 - 6c^2 + 16c - 16) = (c+6)(c-2)(c^2 - 4c + 8). \\
&\therefore \quad c \neq -6, 2, 2 \pm 2i.
\end{aligned} \tag{8}$$

★

10. FUNDAMENTALS OF LINEAR MAPPINGS

★ 9 ★

KEYWORDS: MAPPINGS, TRANSFORMATIONS, SURJECTIONS, INJECTIONS, BIJECTIONS, 1-1 CORRESPONDENCES, COMPOSITIONS, INVERSES, RESTRICTIONS, EXTENSIONS, LINEARITY, LINEAR MAPPINGS, SUM, SCALAR MULTIPLICATION, ISOMORPHISMS, ISOMORPHIC, IMAGES OF LINEAR MAPPINGS, KERNEL, RANK, LINEAR MAPPINGS BY MATRICES, BASES OF IMAGE AND KERNEL, MAXIMAL SYSTEM OF LINEARLY INDEPENDENT ELEMENTS, MINORS, DIMENSION FORMULA

10.1. **Mappings and transformations.** Let V and W be sets. An operation T that, for every element x of V, T associates x with one element $T(x)$ of W is called a mapping (function) from V to W. Denote this T symbolically as follows.

$$T : V \longrightarrow W \tag{1}$$

The sets V and W are called the domain and codomain of T, respectively. The subset:

$$\{T(x) \mid x \in V\} \tag{2}$$

of W is called the image of T, denoted by $\operatorname{Im} T$ or $T(V)$. We sometimes write $T(x)$ as Tx. For a subset X of V, the set:

$$T(X) = \{Tx \mid x \in X\} \tag{3}$$

is called the image of X by (under) T. For a subset Y, the set:

$$T^{-1}(Y) = \{x \in V \mid Tx \in Y\} \tag{4}$$

is called the inverse image of Y by (under) T. If $Y = \{y\}$, then we write $T^{-1}(\{y\}) = T^{-1}(y)$, say,

$$T^{-1}(y) = \{x \in V \mid Tx = y\}. \tag{5}$$

If T is a mapping from V to V, then T is called a transformation of V.

Let T be a mapping from V to W. If there exists $x \in V$ such that $Tx = y$ for every element y of W, then T is called a surjection. This definition is equivalent to $\operatorname{Im} T = W$. For $x, x' \in V$, if $x \neq x' \Rightarrow Tx \neq Tx'$ (or equivalently, $Tx = Tx' \Rightarrow x = x'$), then T is called an injection. If a surjection is also an injection, then it is called a bijection or 1-1 correspondence.

Let V, W, X be sets. For a mapping T from V to W and another mapping S from W to X, the composition $S \circ T$ of T and S is defined as follows.

$$(S \circ T)x = S(Tx) \qquad (x \in V) \tag{6}$$

This is regarded as a mapping from V to X, denoted also by ST, and also called a composite mapping. A composition of transformations is also called a composite transformation.

If a mapping $T : V \longrightarrow W$ is a bijection, then there exists a unique mapping $S : W \longrightarrow V$ such that

$$\begin{aligned} (ST)x = S(Tx) = x \qquad & (x \in V) \\ (TS)y = T(Sy) = y \qquad & (y \in W). \end{aligned} \tag{7}$$

This S is called the inverse (mapping) of T, denoted by $S = T^{-1}$. Then T is also the inverse of S, and therefore $(T^{-1})^{-1} = T$.

For a mapping $T : V \longrightarrow W$ and a subset X of V, the restriction $T|_X$ of T onto X is a mapping $T|_X : X \longrightarrow W$ such that $T|_X(x) = Tx$ $(x \in X)$. If S is a restriction of T onto some set, then T is called an extension of S.

10.2. **Linear mappings.** Let V and W be two vector spaces over K. If a mapping T from V to W satisfies the following property called linearity, then T is said to be a linear mapping from V to W.[1]

$$\begin{aligned} T(\mathtt{x} + \mathtt{y}) = T\mathtt{x} + T\mathtt{y} \qquad & (\mathtt{x}, \mathtt{y} \in V) \\ T(k\mathtt{x}) = k(T\mathtt{x}) \qquad & (\mathtt{x} \in V,\ k \in K) \end{aligned} \tag{8}$$

In particular, a linear mapping from V to V is called a linear transformation. By (8), a linear mapping T satisfies the following:

$$T(k_1\mathtt{x}_1 + \cdots + k_n\mathtt{x}_n) = k_1T\mathtt{x}_1 + \cdots + k_nT\mathtt{x}_n. \tag{9}$$

If a transformation T of V satisfies that

$$T\mathtt{x} = \mathtt{x} \qquad (\mathtt{x} \in V), \tag{10}$$

then T is called a identity transformation, denoted by I_V or I, which is a linear transformation of V.

(exercise01) (1) Show that $T(\mathtt{x}_1 + \cdots + \mathtt{x}_n) = T\mathtt{x}_1 + \cdots + T\mathtt{x}_n$ and (9). (2) Show that I is a linear transformation of V. (3) For a linear mapping $T : V \longrightarrow W$, show that $T\mathsf{0} = \mathsf{0}$.

10.3. **Composition, sum, and scalar multiplication.** Let V, W, X be three vector spaces over K, T be a linear mapping from V to W, and S be a linear mapping from W to X. Then we can get a composition $S \circ T = ST$ by the above mentioned (6).

Theorem 1. *The composition ST of a linear mapping T from V to W and a linear mapping S from W to X is a linear mapping from V to X.*

[1] We write $k(Tx) = kTx$ if confusion does not occur.

Proof. Since it is certain that ST is a mapping from V to X, we show the linearity of ST. For any $\mathtt{x}, \mathtt{y} \in V$,

$$\begin{aligned}(ST)(\mathtt{x}+\mathtt{y}) &= S(T(\mathtt{x}+\mathtt{y})) = S(T\mathtt{x}+T\mathtt{y}) = S(T\mathtt{x})+S(T\mathtt{y}) \\ &= (ST)\mathtt{x}+(ST)\mathtt{y}.\end{aligned} \tag{11}$$

For any $\mathtt{x} \in V$, $k \in K$,

$$(ST)(k\mathtt{x}) = S(T(k\mathtt{x})) = S(kT\mathtt{x}) = kS(T\mathtt{x}) = k(ST)\mathtt{x}. \tag{12}$$

□

Next, let S and T be linear mappings from V to W. The sum $S+T$ of S and T and the scalar multiplication aT $(a \in K)$ of T are mappings from V to W defined by

$$\begin{aligned}(S+T)\mathtt{x} &= S\mathtt{x}+T\mathtt{x} \qquad (\mathtt{x} \in V) \\ (aT)\mathtt{x} &= a(T\mathtt{x}) \qquad (\mathtt{x} \in V)\end{aligned} \tag{13}$$

Theorem 2. *For any linear mappings S, T from V to W, $S+T$ and aT $(a \in K)$ are linear mapping from V to W.*

Proof. (Case of $S+T$) For any $\mathtt{x}, \mathtt{y} \in V$, we have

$$\begin{aligned}(S+T)(\mathtt{x}+\mathtt{y}) &= S(\mathtt{x}+\mathtt{y})+T(\mathtt{x}+\mathtt{y}) = (S\mathtt{x}+S\mathtt{y})+(T\mathtt{x}+T\mathtt{y}) \\ &= (S\mathtt{x}+T\mathtt{x})+(S\mathtt{y}+T\mathtt{y}) \\ &= (S+T)\mathtt{x}+(S+T)\mathtt{y}.\end{aligned} \tag{14}$$

For any $\mathtt{x} \in V$ and $k \in K$, we have

$$\begin{aligned}(S+T)(k\mathtt{x}) &= S(k\mathtt{x})+T(k\mathtt{x}) = kS\mathtt{x}+kT\mathtt{x} \\ &= k(S\mathtt{x}+T\mathtt{x}) = k(S+T)\mathtt{x}.\end{aligned} \tag{15}$$

□

(exercise02) Show the case of aT.

Composition, sum, and scalar multiplication of linear mappings satisfy the following.

Theorem 3. *Let $T: V \longrightarrow W$, $\tilde{T}: V \longrightarrow W$, $\tilde{\tilde{T}}: V \longrightarrow W$, $S: W \longrightarrow X$, $\tilde{S}: W \longrightarrow X$ and $R: X \longrightarrow Y$ be linear mappings, and let $a \in K$, then*

$$\begin{gathered}(RS)T = R(ST) \\ T+\tilde{T} = \tilde{T}+T \qquad (T+\tilde{T})+\tilde{\tilde{T}} = T+(\tilde{T}+\tilde{\tilde{T}}) \\ S(T+\tilde{T}) = ST+S\tilde{T} \qquad (S+\tilde{S})T = ST+\tilde{S}T \\ a(T+\tilde{T}) = aT+a\tilde{T} \\ a(ST) = (aS)T = S(aT)\end{gathered} \tag{16}$$

Proof. The first equality is a basic property satisfied by general mappings. Here we show the fourth equality.

$$\begin{aligned}(S(T+\tilde{T}))\mathtt{x} &= S((T+\tilde{T})\mathtt{x}) = S(T\mathtt{x}+\tilde{T}\mathtt{x}) = S(T\mathtt{x})+S(\tilde{T}\mathtt{x}) \\ &= (ST)\mathtt{x}+(S\tilde{T})\mathtt{x} = (ST+S\tilde{T})\mathtt{x}.\end{aligned} \tag{17}$$

Hence $S(T+\tilde{T}) = ST+S\tilde{T}$. □

(exercise03) Show the rest equalities of (16).

By Theorem 3, the first formula (associative law for composition), any composition of several linear mappings allows to remove parentheses. This is valid for composition of general mappings. By the third formula (associative law for addition), any sum of several linear mappings also allows to remove parentheses, and by the second formula (commutative law for addition), the order of addition can be changed. Furthermore, by the fourth-fifth formulas (distributive law), we have the following.

$$\begin{aligned} S(T+\tilde{T}+\cdots+\tilde{\tilde{T}}) &= ST+S\tilde{T}+\cdots+S\tilde{\tilde{T}} \\ (S+\tilde{S}+\cdots+\tilde{\tilde{S}})T &= ST+\tilde{S}T+\cdots+\tilde{\tilde{S}}T \end{aligned} \tag{18}$$

According to Theorems 1-2, any mappings made by finite number of compositions, sums, and scalar multiplications of linear mappings are also linear mappings. This fact is valid, of course, for linear transformations. In particular, for a linear transformation T of V, letting $T^l = TT\dots T$ (l times), we see that the following is a linear transformation of V.

$$a_0T^s + a_1T^{s-1} + \cdots + a_{s-1}T + a_sI \qquad (a_0,\dots,a_s \in K) \tag{19}$$

10.4. **Isomorphisms.** If T is a linear mapping from V to W and also a bijection from V to W, then T is called an isomorphism from V to W (between V and W). For vector spaces V and W, if there exists an isomorphism from V to W, then V is called isomorphic to W (V and W are isomorphic), denoted by $V \simeq W$.

(note) The inverse of an isomorphism from V to W is an isomorphism from W to V.

Theorem 4. *For vector spaces V and W over K, it holds that*

$$V \simeq W \iff \dim V = \dim W \tag{20}$$

Theorem 4'. *Let $V \simeq W$ and $\phi : V \longrightarrow W$ be an isomorphism. For every basis $\{\mathtt{e}_1, \mathtt{e}_2, \dots, \mathtt{e}_n\}$, $\{\phi(\mathtt{e}_1), \phi(\mathtt{e}_2), \dots, \phi(\mathtt{e}_n)\}$ is a basis of W.*

Proof of Theorem 4'. (To span W.) As ϕ is an isomorphism, it is of course surjection. Hence for any $\mathtt{y} \in W$, there exists $\mathtt{x} \in V$ such that $\phi(\mathtt{x}) = \mathtt{y}$, and $\mathtt{x}$ is expressed as $\mathtt{x} = k_1\mathtt{e}_1 + \cdots + k_n\mathtt{e}_n$. Therefore

$$\begin{aligned} \mathtt{y} &= \phi(\mathtt{x}) = \phi(k_1\mathtt{e}_1 + \cdots + k_n\mathtt{e}_n) = \phi(k_1\mathtt{e}_1) + \cdots + \phi(k_n\mathtt{e}_n) \\ &= k_1\phi(\mathtt{e}_1) + \cdots + k_n\phi(\mathtt{e}_n). \end{aligned} \tag{21}$$

(Linear independence) Letting $k_1\phi(\mathtt{e}_1) + \cdots + k_n\phi(\mathtt{e}_n) = \mathtt{0}$, then $k_1\phi(\mathtt{e}_1) + \cdots + k_n\phi(\mathtt{e}_n) = \phi(k_1\mathtt{e}_1 + \cdots + k_n\mathtt{e}_n) = \mathtt{0}$. As ϕ is an isomorphism, it is injection, thus $k_1\mathtt{e}_1 + \cdots + k_n\mathtt{e}_n = \mathtt{0}$, and therefore $k_1 = \cdots = k_n = 0$. □

Proof of Theorem 4. (⇒) Letting $V \simeq W$, and $\phi : V \longrightarrow W$ be an isomorphism. By Theorem 4', we can make a basis $\langle \phi(\mathtt{e}_1), \phi(\mathtt{e}_2), \dots, \phi(\mathtt{e}_n) \rangle$ of W from a basis $\langle \mathtt{e}_1, \mathtt{e}_2, \dots, \mathtt{e}_n \rangle$ of V. Hence $\dim V = n = \dim W$.

($\Leftarrow$) Let $\dim V = \dim W$, then we can take a basis $\langle \mathtt{e}_1, \dots, \mathtt{e}_n \rangle$ of V and a basis $\langle \mathtt{f}_1, \dots, \mathtt{f}_n \rangle$ of W. Here, take a linear mapping $\phi : V \longrightarrow W$ defined by $\phi(\mathtt{e}_j) = \mathtt{f}_j$ ($j = 1, \dots, n$), then it is an isomorphism from V to W. Indeed, for an arbitrary $\mathtt{x} = k_1\mathtt{e}_1 + \cdots + k_n\mathtt{e}_n \in V$, ϕ is defined as follows.

$$\phi(\mathtt{x}) = k_1\mathtt{f}_1 + \cdots + k_n\mathtt{f}_n \tag{22}$$

Now for any $\mathtt{x}$, $\mathtt{y} = l_1\mathtt{e}_1 + \cdots + l_n\mathtt{e}_n \in V$, and $k \in K$, we have

$$\begin{aligned}
\phi(\mathtt{x}+\mathtt{y}) &= \phi((k_1\mathtt{e}_1 + \cdots + k_n\mathtt{e}_n) + (l_1\mathtt{e}_1 + \cdots + l_n\mathtt{e}_n)) \\
&= \phi((k_1+l_1)\mathtt{e}_1 + \cdots + (k_n+l_n)\mathtt{e}_n) \\
&= (k_1+l_1)\mathtt{f}_1 + \cdots + (k_n+l_n)\mathtt{f}_n \\
&= (k_1\mathtt{f}_1 + \cdots + k_n\mathtt{f}_n) + (l_1\mathtt{f}_1 + \cdots + l_n\mathtt{f}_n) \\
&= \phi(\mathtt{x}) + \phi(\mathtt{y}), \\
\phi(k\mathtt{x}) &= \phi(k(k_1\mathtt{e}_1 + \cdots + k_n\mathtt{e}_n)) = \phi(kk_1\mathtt{e}_1 + \cdots + kk_n\mathtt{e}_n) \\
&= kk_1\mathtt{f}_1 + \cdots + kk_n\mathtt{f}_n = k(k_1\mathtt{f}_1 + \cdots + k_n\mathtt{f}_n) \\
&= k\phi(\mathtt{x}).
\end{aligned} \tag{23}$$

Hence ϕ is a linear mapping form V to W. Next we show that ϕ is a bijection. First, by

$$\begin{aligned}
& \mathtt{x} = k_1\mathtt{e}_1 + \cdots + k_n\mathtt{e}_n \neq l_1\mathtt{e}_1 + \cdots + l_n\mathtt{e}_n = \mathtt{y} \\
\Longleftrightarrow \quad & \phi(\mathtt{x}) = k_1\mathtt{f}_1 + \cdots + k_n\mathtt{f}_n \neq l_1\mathtt{f}_1 + \cdots + l_n\mathtt{f}_n = \phi(\mathtt{y})
\end{aligned} \tag{24}$$

ϕ is an injection. Second, for any $\tilde{\mathtt{x}} = k_1\mathtt{f}_1 + \cdots + k_n\mathtt{f}_n \in W$, there exists $\mathtt{x} = k_1\mathtt{e}_1 + \cdots + k_n\mathtt{e}_n$ such that $\phi(\mathtt{x}) = \tilde{\mathtt{x}}$, thus ϕ is a surjection. Consequently, $\phi : V \longrightarrow W$ is an isomorphism, and therefore $V \simeq W$. □

10.5. **Images and kernels of linear mappings.** Let T be a linear mapping from V to W. The image $\operatorname{Im} T = T(V)$ of T is defined to be the above mentioned (2). More generally, for a subspace X of V, the image $T(X)$ of X by T is defined to be (3). Also, for a subspace Y of W, the inverse image $T^{-1}(Y)$ of Y by T is defined to be (4). Especially,

$$T^{-1}(\mathtt{0}) = \{\mathtt{x} \in V \mid T\mathtt{x} = \mathtt{0}\} \tag{25}$$

is called the kernel of T, denoted by $\operatorname{Ker} T$.

Theorem 5. *(i)* $\operatorname{Im} T$ *is a subspace of* W. *(ii)* $\operatorname{Ker} T$ *is a subspace of* V.
(iii) Let X *be a subspace of* V*, then* $T(X)$ *is a subspace of* W.
(iv) Let Y *be a subspace of* W*, then* $T^{-1}(Y)$ *is a subspace of* V*. ((i) and (ii) are special cases of (iii) and (iv), respectively.)*

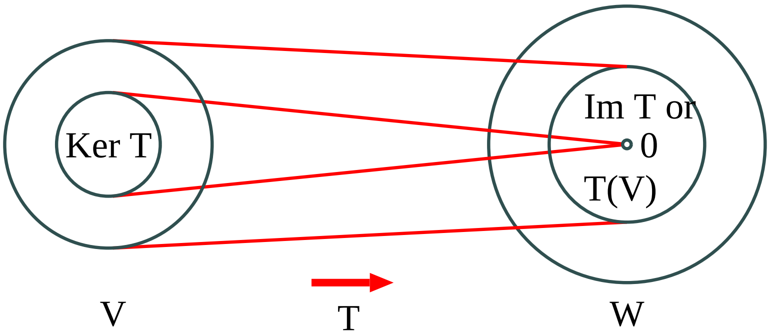
Proof. It suffices to prove that every case is closed under addition and scalar multiplication. (i) Take any elements $\mathtt{x}', \mathtt{y}'$ of $\operatorname{Im} T$, then there exists elements $\mathtt{x}, \mathtt{y} \in V$ which represent $\mathtt{x}' = T\mathtt{x}$, $\mathtt{y}' = T\mathtt{y}$. Therefore $\mathtt{x}' + \mathtt{y}' = T\mathtt{x} + T\mathtt{y} = T(\mathtt{x}+\mathtt{y}) \in \operatorname{Im} T$. Also, take any scalar k, then $k\mathtt{x}' = kT\mathtt{x} = T(k\mathtt{x}) \in \operatorname{Im} T$. Hence $\operatorname{Im} T$ is a subspace of W.

(ii) Take any elements $\mathtt{x}, \mathtt{y}$ of $\operatorname{Ker} T$. Then $T(\mathtt{x}+\mathtt{y}) = T\mathtt{x} + T\mathtt{y} = \mathtt{0} + \mathtt{0} = \mathtt{0}$, and thus $\mathtt{x} + \mathtt{y} \in \operatorname{Ker} T$. Further, for any scalar k, $T(k\mathtt{x}) = kT\mathtt{x} = k\mathtt{0} = \mathtt{0}$, and therefore $k\mathtt{x} \in \operatorname{Ker} T$. Hence $\operatorname{Ker} T$ is a subspace of V.

(iii) Similar to (i).

(iv) Take any elements $\mathtt{x}, \mathtt{y}$ of $T^{-1}(Y)$. Since $T\mathtt{x}, T\mathtt{y} \in Y$, $T(\mathtt{x}+\mathtt{y}) = T\mathtt{x} + T\mathtt{y} \in Y$. Hence $\mathtt{x}+\mathtt{y} \in T^{-1}(Y)$. Further, for any scalar k, $T(k\mathtt{x}) = kT\mathtt{x} \in Y$, and therefore $k\mathtt{x} \in T^{-1}(Y)$. Hence $T^{-1}(Y)$ is a subspace of V. □

(Definition) $\dim \operatorname{Im} T$ is called the rank of T, denoted by $r(T)$.



Theorem 6. *Let T be a linear mapping from V to W, then*

$$T \text{ is an isomorphism} \iff \operatorname{Ker} T = \{\mathtt{0}\} \text{ and } \operatorname{Im} T = W. \tag{26}$$

Proof. If T is a linear mapping, then

$$\begin{aligned} T \text{ is an isomorphism} &\iff T \text{ is a bijection} \\ &\iff T \text{ is a surjection and also an injection,} \end{aligned} \tag{27}$$

and it holds that

$$T \text{ is a surjection} \iff \operatorname{Im} T = W. \tag{28}$$

Hence, to prove Theorem 6, it suffices to show that

$$T \text{ is an injection} \iff \operatorname{Ker} T = \{\mathtt{0}\} \tag{29}$$

($\Rightarrow$) Let a linear mapping T be an injection. Since T is a linear mapping, $T\mathtt{0} = \mathtt{0}$. Since T is an injection, if $\mathtt{x} \neq \mathtt{0}$, then $T\mathtt{x} \neq \mathtt{0}$. Hence $\ker T = \{\mathtt{0}\}$.

($\Leftarrow$) Let a linear mapping T satisfy $\ker T = \{\mathtt{0}\}$. Suppose $T\mathtt{x} = T\mathtt{y}$. Then $T\mathtt{x} - T\mathtt{y} = T(\mathtt{x}-\mathtt{y}) = \mathtt{0}$. $\therefore$ $\mathtt{x}-\mathtt{y} \in \ker T = \{\mathtt{0}\}$. $\therefore$ $\mathtt{x} = \mathtt{y}$. Hence T is an injection. □

(exercise04) (1) Let $V = V^3 = \mathbb{R}^3$, $W = \{\text{all real } 2\times 2 \text{ matrices}\}$, then V and W are regarded as real vector spaces. Define a mapping $T : V \longrightarrow W$ by $T\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x & y \\ y & z \end{pmatrix}$, then show that T is a linear mapping.
(2) Let $Y = \{\text{all real symmetric } 2\times 2 \text{ matrices}\}$, then show that Y is a subspace of W. Also, show that a mapping $T : V \longrightarrow Y$ defined by (1) is an isomorphism.

10.6. **Linear mappings determined by matrices.** Let $V = K^n$ and $W = K^m$. Take an $m \times n$ matrix A with entries in K. Define a mapping T_A from V to W by

$$T_A(\mathtt{x}) = A\mathtt{x} \qquad (\mathtt{x} \in V). \tag{30}$$

T_A is a linear mapping from V to W, which is called the linear mapping (matrix mapping) determined by A. If a linear mapping T is expressed as $T = T_A$ for some matrix A, then A is called the matrix of a linear mapping T.

(exercise05) Show that T_A is a linear mapping.

Actually, it is known that the converse of (exercise05) holds, say,

Theorem 7. *Let $V = K^n$ and $W = K^m$. Any linear mapping from V to W is represented as T_A by some matrix A.*

Proof. Take an arbitrary linear mapping T from V to W. Denote by $\mathtt{e}_1, \mathtt{e}_2 \dots, \mathtt{e}_n$ the elementary vectors of V. Letting $T\mathtt{e}_j = a_j$, and determine a matrix A by

$$A = \begin{pmatrix} \mathtt{a}_1 & \mathtt{a}_2 & \dots & \mathtt{a}_n \end{pmatrix}. \tag{31}$$

Then for any $\mathtt{x} = x_1\mathtt{e}_1 + \cdots + x_n\mathtt{e}_n \in V$, by the linearity of T, we have

$$\begin{aligned} T\mathtt{x} &= T(x_1\mathtt{e}_1 + \cdots + x_n\mathtt{e}_n) = T(x_1\mathtt{e}_1) + \cdots + T(x_n\mathtt{e}_n) \\ &= x_1T\mathtt{e}_1 + \cdots + x_nT\mathtt{e}_n = x_1\mathtt{a}_1 + \cdots + x_n\mathtt{a}_n. \end{aligned} \tag{32}$$

On the other hand, we have

$$A\mathtt{x} = \begin{pmatrix} \mathtt{a}_1 & \mathtt{a}_2 & \dots & \mathtt{a}_n \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_1\mathtt{a}_1 + \cdots + x_n\mathtt{a}_n. \tag{33}$$

Hence $T\mathtt{x} = A\mathtt{x}$, and therefore $T = T_A$. □

Compositions, sums, and scalar multiplications of matrix mappings satisfy the following.

Theorem 8. *Let A, C be an $m \times n$ matrix, B be an $n \times p$ matrix, and $a \in K$, then*

$$\begin{gathered} T_AT_B = T_{AB} \\ T_A + T_C = T_{A+C} \\ aT_A = T_{aA}. \end{gathered} \tag{34}$$

Proof. Show the first equality. For any $\mathtt{x} \in K^p$,

$$(T_AT_B)\mathtt{x} = T_A(T_B(\mathtt{x})) = A(B\mathtt{x}) = (AB)\mathtt{x} = T_{AB}(\mathtt{x}). \tag{35}$$

Hence $T_AT_B = T_{AB}$. □

(exercise06) Show the second and the third equality of (34).

10.7. **Bases of the image and the kernel of T_A.** Let A be an $m \times n$ matrix. Consider the image and the kernel of the linear mapping determined by A:

$$T_A : K^n \longrightarrow K^m. \tag{36}$$

First of all, if A is represented as (31), then the image of T_A:

$$\operatorname{Im} T_A = \{A\mathtt{x} \mid \mathtt{x} \in K^n\} \tag{37}$$

is represented as follows.

$$\operatorname{Im} T_A = \{x_1\mathtt{a}_1 + \cdots + x_n\mathtt{a}_n \mid x_1, \ldots, x_n \in K\} = \operatorname{span}\{\mathtt{a}_1, \ldots, \mathtt{a}_n\} \tag{38}$$

Thus $\operatorname{Im} T_A$ is the subspace of K^m spanned by column vectors. To find a basis of it, it is a good way to perform elementary column operations on A to get a "stair-like" shape with nonzero entries. Then all columns each of which is not a zero vector form a basis of $\operatorname{Im} T_A$. The reason is as follows. When elementary column operation is performed on a matrix, every column vector after the operation is represented by the column vectors before the operation. Since the inverse of a column operation is also a column operation, every column vector before the operation is represented by the column vectors after the operation. Accordingly, the subspace spanned by the columns before the operation, and the subspace spanned by the columns after the operation, are the same at all. Hence if all column vectors except zero vectors are linearly independent after operations, then they form a basis of $\operatorname{Im} T_A$. Especially, if nonzero entries form a "stair-like" shape, they are clearly independent, and thus all column vectors except zero vectors are linearly independent. Hence these column vectors form a basis of $\operatorname{Im} T_A$.

In addition, if there is r "stair-like" column vectors, then the matrix is transformed into $F_{mn}(r)$. Hence the dimension of $\operatorname{Im} T_A$ is equal to the rank of A.

Theorem 9. *It holds that*

$$\dim \operatorname{Im} T_A = r(T_A) = r(A). \tag{39}$$

Next, noting the kernel of T_A:

$$\operatorname{Ker} T_A = \{\mathtt{x} \in K^n \mid A\mathtt{x} = \mathtt{0}\} \tag{40}$$

we see that it is nothing but the set of all solutions to $A\mathtt{x} = 0$. Hence if the general solution is determined to be

$$\mathtt{x} = \alpha_1\mathtt{g}_1 + \alpha_2\mathtt{g}_2 + \cdots + \alpha_s\mathtt{g}_s \tag{41}$$

using elementary operations, then $\langle \mathtt{g}_1, \mathtt{g}_2, \ldots, \mathtt{g}_s\rangle$ is a basis of $\ker T_A$. (Linear independence is clear from their form.) Here, by Chapter 6, Theorem 1, we have $s = n - r(A)$. Consequently, we have the following.

Theorem 10. *Let A be an $m \times n$ matrix, then*

$$\dim \operatorname{Ker} T_A = n - r(A). \tag{42}$$

This is nothing but the dimension formula (Theorem 12) for $T = T_A$.

(exercise07) Let $A = \begin{pmatrix} 3 & 1 & 4 & 2 \\ -1 & 0 & -1 & 0 \\ 1 & 1 & 2 & 2 \end{pmatrix}$. Find bases of the image and kernel of T_A.

(ans) (Basis of $\operatorname{Im} T_A$) Transforming A by elementary column operations,

$$A \longrightarrow \begin{pmatrix} 1 & 3 & 4 & 2 \\ 0 & -1 & -1 & 0 \\ 1 & 1 & 2 & 2 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 3 & 4 & 2 \\ 0 & -1 & -1 & 0 \\ 1 & 1 & 2 & 2 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & -1 & 0 \\ 1 & -2 & -2 & 0 \end{pmatrix}$$
$$\longrightarrow \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 1 & -2 & 0 & 0 \end{pmatrix}. \tag{43}$$

Hence a basis of $\operatorname{Im} T_A$ is $\langle \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ -2 \end{pmatrix} \rangle$.

(Basis of $\ker T_A$) Solve $A\mathtt{x} = \mathtt{0}$. Transforming A by elementary row operations,

$$A \longrightarrow \begin{pmatrix} -1 & 0 & -1 & 0 \\ 3 & 1 & 4 & 2 \\ 1 & 1 & 2 & 2 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 1 & 0 \\ 3 & 1 & 4 & 2 \\ 1 & 1 & 2 & 2 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 2 \\ 0 & 1 & 1 & 2 \end{pmatrix}$$
$$\longrightarrow \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \qquad \therefore \quad \mathtt{x} = \alpha \begin{pmatrix} -1 \\ -1 \\ 1 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} 0 \\ -2 \\ 0 \\ 1 \end{pmatrix}. \tag{44}$$

Hence a basis of $\ker T_A$ is $\langle \begin{pmatrix} -1 \\ -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -2 \\ 0 \\ 1 \end{pmatrix} \rangle$.

(note) A basis is not uniquely determined, and thus there are many other answers. For example, to find a basis of the image, it suffices to choose column vectors as much as possible, and for such a purpose, apply elementary row operations on A to make nonzero entries "stair-like" shape and select columns containing "stair corner" from the original A. This method is effective when the kernel is also needed, but column exchange for the kernel should be warned.

10.8. **Ranks of matrices, maximal system of linearly independent elements, minors.** For a subset S of a vector space V, consider all subspaces of V containing S, and take the intersection of them, then we have a new subspace W of V. This is called the subspace of V spanned by S, denoted by $W = \mathrm{span}(S)$. This is, in fact, the minimum subspace containing S. In other words, $\mathrm{span}(S)$ is a subspace consisting of all possible linear combinations of the elements of S.

Now, vectors $\mathtt{e}_1, \mathtt{e}_2, \ldots, \mathtt{e}_r$ chosen from S satisfy the following two conditions, then $\mathtt{e}_1 \mathtt{e}_2, \ldots, \mathtt{e}_r$ are called a maximal system of linearly independent elements of S.

I: $\mathtt{e}_1, \mathtt{e}_2, \ldots, \mathtt{e}_r$ are linearly independent.

II: Any vector in S is expressed as a linear combination of $\mathtt{e}_1, \mathtt{e}_2, \ldots, \mathtt{e}_r$.

This definition is very similar to the definition of bases, and consequently, every maximal system of linearly independent elements of S is shown to be a basis of $\mathrm{span}(S)$.

Let A be an $m \times n$ matrix, and S be the set of all column vectors of A. From the above consideration and Section 6, a maximal system of linearly independent elements of S is a basis of the subspace spanned by column vectors, say, $\mathrm{Im}\, T_A$. Consequently, the maximum number of linearly independent column vectors of A is equal to $\dim \mathrm{Im}\, T_A = r(A)$.[2] This argument can be applied to $B = {}^tA$. the maximum number of linearly independent column vectors of B (= the maximum number of linearly independent row vectors) is equal to $\dim \mathrm{Im}\, T_B = r(B)$. Here as $r(B) = r({}^tA) = r(A)$, the maximum number of liner independent column/row vectors of A is equal to $r(A)$.

Furthermore, we can associate $r(A)$ with determinants. Taking r rows and r columns from A, we have a determinant of order r, which is called an $r \times r$ minor of A. There are $\binom{m}{r} \times \binom{n}{r}$ $r \times r$ minors in A. Here, for fixed r, the property that all $r \times r$ minors are equal to 0, is preserved under elementary operations.[3] In other words, the property that there exists a nonzero $r \times r$ minor is also preserved under elementary operations. Consequently, the maximum order $s(A)$ of a nonzero minor is also preserved under such operations. Thus repeating elementary operations, we have $A \longrightarrow \cdots \longrightarrow F = F_{mn}(r)$, then it is clear that $r(F) = s(F)$, and therefore $r(A) = s(A)$.

Theorem 11. *For every $m \times n$ matrix A, the following quantity are the same. (i) $r(A)$. (ii) $\dim \mathrm{Im}\, T_A = r(A)$. (iii) The maximum number of linearly independent column vectors. (iv) The maximum number of linearly independent row vectors. (v) The maximum order of a nonzero minor.*

10.9. **The dimension formula.** For kernels and images of linear mappings, we have the following formula, which is known as the dimension formula for linear mappings.

Theorem 12. *Let $T : V \longrightarrow W$ be a linear mapping, then*

$$\dim V = \dim \mathrm{Im}\, T + \dim \mathrm{Ker}\, T. \tag{45}$$

$\dim \mathrm{Im}\, T$ *can be written as $r(T)$ for short.*

---

[2]If the maximum number of linearly independent column vectors are selected, then by Chapter 8, Theorem 1, the other column vectors are represented by those vectors. Hence they form a maximum system of linearly independent elements.

[3]To prove this, it suffices to show that if an elementary operation $A \longrightarrow B$ is performed, then every $r \times r$ minor Δ of B is expressed by a linear combination of $r \times r$ minors of A. If a used elementary operations is an interchange or scalar multiplication, it is easy to show. Consider an elementary operation that adds the j th row multiplied by c, to the i th row. Only the case that Δ contains the i th row should be confirmed. Then we have $\Delta = \Delta_1 + c\Delta_2$. Here, Δ_1 is an $r \times r$ minor made by choosing the same rows and columns as Δ, from A. Δ_2 is made by interchanging th i th row and the j th row, thus it is signed $r \times r$ minor of A, or the j th row occurs two times, which vanishes. The proof is completed. This argument is applicable to the column case. □

Proof. Let $\mathbb{E} = \langle \mathtt{e}_1, \ldots, \mathtt{e}_s \rangle$ be a basis of $\ker T$, and let $\tilde{\mathbb{E}} = \langle \mathtt{e}_1, \ldots, \mathtt{e}_n \rangle$ be a basis of V, which is an extension of $\mathbb{E}$. Further, let $T\mathtt{e}_i = \mathtt{f}_i$ $(i = s+1, \ldots, n)$. Then we show that $\mathbb{F} = \langle \mathtt{f}_{s+1}, \ldots, \mathtt{f}_n \rangle$ is a basis of $\operatorname{Im} T$. Take an arbitrary element $\mathtt{y}$ of $\operatorname{Im} T$. Then there exists an element $\mathtt{x}$ of V such that $T\mathtt{x} = \mathtt{y}$. Here $\mathtt{x}$ is represented by $\tilde{\mathbb{E}}$ as $\mathtt{x} = k_1\mathtt{e}_1 + \cdots + k_n\mathtt{e}_n$, and therefore, $\mathtt{y}$ is represented by $\mathbb{F}$ as follows.

$$\begin{aligned} \mathtt{y} = T\mathtt{x} &= T(k_1\mathtt{e}_1 + \cdots + k_n\mathtt{e}_n) = k_1T\mathtt{e}_1 + \cdots + k_sT\mathtt{e}_s + k_{s+1}T\mathtt{e}_{s+1} + \cdots + k_nT\mathtt{e}_n \\ &= k_{s+1}T\mathtt{e}_{s+1} + \cdots + k_nT\mathtt{e}_n = k_{s+1}\mathtt{f}_{s+1} + \cdots + k_n\mathtt{f}_n \end{aligned} \tag{46}$$

Next we show that the elements of $\mathbb{F}$ are linearly independent. Let $k_{s+1}\mathtt{f}_{s+1} + k_{s+2}\mathtt{f}_{s+2} + \cdots + k_n\mathtt{f}_n = \mathtt{0}$, then as $T\mathtt{e}_i = \mathtt{f}_i$, we have

$$\begin{aligned} \mathtt{0} &= k_{s+1}\mathtt{f}_{s+1} + \cdots + k_n\mathtt{f}_n = k_{s+1}T\mathtt{e}_{s+1} + \cdots + k_nT\mathtt{e}_n \\ &= T(k_{s+1}\mathtt{e}_{s+1} + \cdots + k_n\mathtt{e}_n) \end{aligned} \tag{47}$$

That is, $T(k_{s+1}\mathtt{e}_{s+1} + \cdots + k_n\mathtt{e}_n) = \mathtt{0}$. This means that $\mathtt{x} = k_{s+1}\mathtt{e}_{s+1} + \cdots + k_n\mathtt{e}_n \in \ker T$. Hence $\mathtt{x}$ is represented also by $\mathbb{E}$, a basis of $\ker T$. Consequently,

$$\mathtt{x} = k_{s+1}\mathtt{e}_{s+1} + \cdots + k_n\mathtt{e}_n = k_1\mathtt{e}_1 + \cdots + k_s\mathtt{e}_s, \tag{48}$$

and therefore $k_1\mathtt{e}_1 + \cdots + k_s\mathtt{e}_s - k_{s+1}\mathtt{e}_{s+1} - \cdots - k_n\mathtt{e}_n = \mathtt{0}$. Since $\tilde{\mathbb{E}} = \langle \mathtt{e}_1, \ldots, \mathtt{e}_n \rangle$ is a basis, $k_1 = k_2 = \cdots = k_n = 0$. Especially, as $k_{s+1} = \cdots = k_n = 0$, the elements of $\mathbb{F}$ are linearly independent.

Accordingly, $\mathbb{F}$ is a basis of $\operatorname{Im} T$. Hence $\dim \operatorname{Im} T = n - s$. Also, recall $\dim \ker T = s$, $\dim V = n$. From this it follows that $\dim \operatorname{Im} T + \dim \ker T = \dim V$. □

Finally, we give a corollary of this theorem. Let X be a subspace of V. Apply Theorem 12 to the restriction of T onto X, say, $T|_X$, then

$$\dim X = \dim \operatorname{Im} T|_X + \dim \operatorname{Ker} T|_X \tag{49}$$

Here, noting that $\operatorname{Im} T|_X = T(X)$, $\ker T|_X = X \cap \ker T$, we have the following.

Theorem 12'. *For a linear mapping $T : V \longrightarrow W$, and a subspace X of V, we have*

$$\dim X = \dim \operatorname{Im} T|_X + \dim \operatorname{Ker} T|_X \tag{50}$$

$$\dim X = \dim T(X) + \dim(X \cap \operatorname{Ker} T). \tag{51}$$

(exercise08) Let T be a linear transformation of V. Show that if $\ker T = \operatorname{Im} T$, then $\dim V$ is even.

★

11. BASE CHANGES AND LINEAR MAPPINGS

★ 6 ★

KEYWORDS: BASES, ISOMORPHISMS, BASE CHANGES, MATRICES WITH RESPECT TO BASES, IMAGES, KERNELS, INVARIANT SUBSPACES, CANONICAL FORMS OF MATRICES, OPERATIONS OF LINEAR TRANSFORMATIONS

11.1. **Bases and isomorphisms.** Throughout this chapter, let V and W be two linear spaces over K, and the dimensions of them be n and m, respectively. Let $T : V \longrightarrow W$ be a linear mapping from V to W. Let $\mathbb{E} = \langle \mathtt{e}_1, \mathtt{e}_2, \ldots, \mathtt{e}_n \rangle$ and $\tilde{\mathbb{E}} = \langle \tilde{\mathtt{e}}_1, \tilde{\mathtt{e}}_2, \ldots, \tilde{\mathtt{e}}_n \rangle$ be two bases of V, and let $\mathbb{F} = \langle \mathtt{f}_1, \mathtt{f}_2, \ldots, \mathtt{f}_n \rangle$ and $\tilde{\mathbb{F}} = \langle \tilde{\mathtt{f}}_1, \tilde{\mathtt{f}}_2, \ldots, \tilde{\mathtt{f}}_n \rangle$ be two bases of W.

Every element x of V is expressed uniquely by the linear combination of the basis $\mathbb{E}$ as follows.

$$\boxed{x = x_1\mathtt{e}_1 + x_2\mathtt{e}_2 + \cdots + x_n\mathtt{e}_n = \mathbb{E}\mathtt{x}} \qquad \left(\mathtt{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\right) \tag{1}$$

(Matrix calculation containing bases is performed regarding bases as row vectors.) By this equality, a mapping that corresponds x to $\mathtt{x}$ is determined:

$$\begin{array}{rcl} \phi_{\mathbb{E}} : V & \longrightarrow & K^n \\ x & \longmapsto & \mathtt{x} \end{array} \tag{2}$$

This mapping is an isomorphism, called the isomorphism determined by the basis $\mathbb{E}$. Similarly, by the following equality:

$$\boxed{x = \tilde{x}_1\tilde{\mathtt{e}}_1 + \tilde{x}_2\tilde{\mathtt{e}}_2 + \cdots + \tilde{x}_n\tilde{\mathtt{e}}_n = \tilde{\mathbb{E}}\tilde{\mathtt{x}}} \qquad \left(\tilde{\mathtt{x}} = \begin{pmatrix} \tilde{x}_1 \\ \vdots \\ \tilde{x}_n \end{pmatrix}\right) \tag{3}$$

the isomorphism determined by $\tilde{\mathbb{E}}$ is also defined.

Here notice that $\mathtt{x}$ and $\tilde{\mathtt{x}}$ are corresponding by an intermediation of x, that is,

$$\tilde{\mathtt{x}} \longmapsto x \longmapsto \mathtt{x} \tag{4}$$

This is represented by mappings as

$$\mathtt{x} = \phi_{\mathbb{E}}(\phi_{\tilde{\mathbb{E}}}^{-1}(\tilde{\mathtt{x}})) = (\phi_{\mathbb{E}} \circ \phi_{\tilde{\mathbb{E}}}^{-1})(\tilde{\mathtt{x}}). \tag{5}$$

This mapping is a composition of isomorphisms, and thus, it is an isomorphism. Hence it holds that $\mathtt{x} = P\tilde{\mathtt{x}}$, and is called the matrix of a base change from $\mathbb{E}$ to $\tilde{\mathbb{E}}$. In addition, P satisfy that $\tilde{\mathbb{E}} = \mathbb{E}P$. We show this below.

$$\begin{array}{ll} \text{Since } \mathtt{x} = P\,\tilde{\mathtt{x}}, \text{we have} & \mathbb{E}\mathtt{x} = \mathbb{E}(P\tilde{\mathtt{x}}) = (\mathbb{E}P)\tilde{\mathtt{x}}. \\ \text{Also, as } x = \mathbb{E}\mathtt{x} = \tilde{\mathbb{E}}\tilde{\mathtt{x}}, & (\mathbb{E}P)\tilde{\mathtt{x}} = \tilde{\mathbb{E}}\tilde{\mathtt{x}}. \\ \text{As this holds for all } \tilde{\mathtt{x}} \in K^n, & \mathbb{E}P = \tilde{\mathbb{E}}. \quad \square \end{array} \tag{6}$$

Similar consideration is possible for linear space W and its two bases $\mathbb{F}$ and $\tilde{\mathbb{F}}$. That is, for $y \in W$, by $y = \mathbb{F}\mathtt{y}$ we have a correspondence $y \longmapsto \mathtt{y}$, and by $y = \tilde{\mathbb{F}}\tilde{\mathtt{y}}$, we have $y \longmapsto \tilde{\mathtt{y}}$. Then the matrix Q of a base change from $\mathbb{F}$ to $\tilde{\mathbb{F}}$ is given by $\mathtt{y} = Q\tilde{\mathtt{y}}$ and satisfies $\tilde{\mathbb{F}} = \mathbb{F}Q$.

(exercise01) Let $V = \{\text{all polynomials in } t \text{ of degree less or equal to } 3\}$, and two bases $\mathbb{E} = \langle 1, t, t^2, t^3\rangle$ and $\tilde{\mathbb{E}} = \langle t^2+1, t^2-1, t^3+t, t^3-t\rangle$ are given.
(1) For $p = at^3 + bt^2 + ct + d$, determine $\mathtt{p} = \phi_{\mathbb{E}}(p)$ and $\tilde{\mathtt{p}} = \phi_{\tilde{\mathbb{E}}}(p)$.
(2) The matrix P of a base change from $\mathbb{E}$ to $\tilde{\mathbb{E}}$.

(ans) (1) By $p = \mathbb{E}\mathtt{p}$, we have

$$at^3 + bt^2 + ct + d = \begin{pmatrix} 1 & t & t^2 & t^3 \end{pmatrix}\begin{pmatrix} d \\ c \\ b \\ a \end{pmatrix}. \quad \therefore \quad \phi_{\mathbb{E}}(p) = \begin{pmatrix} d \\ c \\ b \\ a \end{pmatrix}. \tag{7}$$

Similarly, by $p = \tilde{\mathbb{E}}\tilde{\mathtt{p}}$, letting $\tilde{\mathtt{p}} = \begin{pmatrix} \tilde{p}_1 \\ \tilde{p}_2 \\ \tilde{p}_3 \\ \tilde{p}_4 \end{pmatrix}$, we have

$$\begin{aligned} &at^3 + bt^2 + ct + d = \begin{pmatrix} t^2+1 & t^2-1 & t^3+t & t^3-t \end{pmatrix}\begin{pmatrix} \tilde{p}_1 \\ \tilde{p}_2 \\ \tilde{p}_3 \\ \tilde{p}_4 \end{pmatrix} \\ &= (\tilde{p}_3 + \tilde{p}_4)t^3 + (\tilde{p}_1 + \tilde{p}_2)t^2 + (\tilde{p}_3 - \tilde{p}_4)t + (\tilde{p}_1 - \tilde{p}_2). \\ &\therefore \quad \tilde{p}_1 = \tfrac{b+d}{2}, \tilde{p}_2 = \tfrac{b-d}{2}, \tilde{p}_3 = \tfrac{a+c}{2}, \tilde{p}_4 = \tfrac{a-c}{2}. \qquad \therefore \quad \phi_{\tilde{\mathbb{E}}}(p) = \begin{pmatrix} \frac{b+d}{2} \\ \frac{b-d}{2} \\ \frac{a+c}{2} \\ \frac{a-c}{2} \end{pmatrix}. \end{aligned} \tag{8}$$

(2) From $\tilde{\mathbb{E}} = \mathbb{E}P$, it follows that

$$\begin{aligned} \begin{pmatrix} t^2+1 & t^2-1 & t^3+t & t^3-t \end{pmatrix} = \begin{pmatrix} 1 & t & t^2 & t^3 \end{pmatrix}\begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}. \\ \therefore \quad P = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}. \end{aligned} \tag{9}$$

11.2. **Bases and linear mappings.** Next, we consider a linear mapping T. If x and y satisfy $y = Tx$, then what relation $\mathtt{x}$ and $\mathtt{y}$ satisfy? Here, a correspondence from $\mathtt{x}$ to $\mathtt{y}$ is represented in the following diagram moving from K^n to K^m,

$$\begin{array}{ccc} V & \xrightarrow{T} & W \\ \uparrow \phi_{\mathbb{E}}^{-1} & & \downarrow \phi_{\mathbb{F}} \\ K^n & & K^m \end{array} \tag{10}$$

which gives a composition

$$\mathtt{x} \longmapsto x \longmapsto y \longmapsto \mathtt{y}. \tag{11}$$

Since this is a composition of linear mappings, it is a linear mapping, and therefore it is represented by a matrix A as

$$\mathtt{y} = A\mathtt{x}. \tag{12}$$

This matrix A is called the matrix of T with respect to bases $\mathbb{E}$ and $\mathbb{F}$, and satisfies that $T\mathbb{E} = \mathbb{F}A$. We show this below. Here, $T\mathbb{E}$ means that

$$T\mathbb{E} = T\begin{pmatrix} \mathtt{e}_1 & \mathtt{e}_2 & \dots & \mathtt{e}_n \end{pmatrix} = \begin{pmatrix} T\mathtt{e}_1 & T\mathtt{e}_2 & \dots & T\mathtt{e}_n \end{pmatrix}. \tag{13}$$

First of all,

$$y = Tx, \tag{14}$$

and the left-hand side and the right-hand side are calculated as

$$Tx = T(\mathbb{E}\mathtt{x}) = T(x_1\mathtt{e}_1 + \dots + x_n\mathtt{e}_n) \overset{\text{linearity of } T}{=} x_1(T\mathtt{e}_1) + \dots + x_n(T\mathtt{e}_n) = (T\mathbb{E})\mathtt{x}, \tag{15}$$

$$y = \mathbb{F}\mathtt{y} = \mathbb{F}(A\mathtt{x}) = (\mathbb{F}A)\mathtt{x}. \tag{16}$$

By (14),(15),(16)

$$(T\mathbb{E})\mathtt{x} = (\mathbb{F}A)\mathtt{x}. \tag{17}$$

This is valid for any element of $\mathtt{x}$ of K^n, thus

$$T\mathbb{E} = \mathbb{F}A. \quad \square \tag{18}$$

This argument is possible for the matrix of T with respect to $\tilde{\mathbb{E}}$ and $\tilde{\mathbb{F}}$, if $\tilde{\mathtt{y}} = B\tilde{\mathtt{x}}$, then $T\tilde{\mathbb{E}} = \tilde{\mathbb{F}}B$.

(note) If $V = W$ and $\mathbb{E} = \mathbb{F}$, then the matrix of T with respect to $\mathbb{E}$ and $\mathbb{F}$ is said to be simply the matrix of T with respect to $\mathbb{E}$.
(exercise02) Let V, $\mathbb{E}$, $\tilde{\mathbb{E}}$ be as in (exercise01), and let

$$W = \{ \text{ all polynomials in } t \text{ of degree less or equal to } 2\}. \tag{19}$$

Take two bases $\mathbb{F} = \langle 1, t, t^2 \rangle$, $\tilde{\mathbb{F}} = \langle t^2 + 1, t^2 - 1, t \rangle$. Let T be a linear mapping from V to W defined by $T(p(t)) = p'(2t)$. (1) Determine the matrix A of T with respect to $\mathbb{E}$ and $\mathbb{F}$.
(2) Determine the matrix B of T with respect to $\tilde{\mathbb{E}}$ and $\tilde{\mathbb{F}}$.
(3) Determine the matrix of a base change from $\mathbb{F}$ to $\tilde{\mathbb{F}}$, and confirm that $B = Q^{-1}AP$.

(ans) (1) By $T\mathbb{E} = \mathbb{F}A$, we have

$$T\mathbb{E} = T\begin{pmatrix} 1 & t & t^2 & t^3 \end{pmatrix} = \begin{pmatrix} T(1) & T(t) & T(t^2) & T(t^3) \end{pmatrix} = \begin{pmatrix} 0 & 1 & 4t & 12t^2 \end{pmatrix}$$
$$= \mathbb{F}A = \begin{pmatrix} 1 & t & t^2 \end{pmatrix}\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 12 \end{pmatrix}. \qquad \therefore\ A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 12 \end{pmatrix}. \tag{20}$$

(2) By $T\tilde{\mathbb{E}} = \tilde{\mathbb{F}}B$, we have

$$T\tilde{\mathbb{E}} = T\begin{pmatrix} t^2+1 & t^2-1 & t^3+t & t^3-t \end{pmatrix} =$$
$$\begin{pmatrix} T(t^2+1) & T(t^2-1) & T(t^3+t) & T(t^3-t) \end{pmatrix} = \begin{pmatrix} 4t & 4t & 12t^2+1 & 12t^2-1 \end{pmatrix}$$
$$= \tilde{\mathbb{F}}B = \begin{pmatrix} t^2+1 & t^2-1 & t \end{pmatrix}\begin{pmatrix} 0 & 0 & x & z \\ 0 & 0 & y & w \\ 4 & 4 & 0 & 0 \end{pmatrix}.$$
$$\therefore\ (x+y)t^2 + (x-y) = 12t^2+1, \quad (z+w)t^2 + (z-w) = 12t^2-1.$$
$$\therefore\ x = \tfrac{13}{2},\ y = \tfrac{11}{2},\ z = \tfrac{11}{2},\ w = \tfrac{13}{2}. \qquad \therefore\ B = \begin{pmatrix} 0 & 0 & \frac{13}{2} & \frac{11}{2} \\ 0 & 0 & \frac{11}{2} & \frac{13}{2} \\ 4 & 4 & 0 & 0 \end{pmatrix}. \tag{21}$$

(3) By $\tilde{\mathbb{F}} = \mathbb{F}Q$, we have

$$\begin{pmatrix} t^2+1 & t^2-1 & t \end{pmatrix} = \begin{pmatrix} 1 & t & t^2 \end{pmatrix}\begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$
$$\therefore\ Q = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}. \tag{22}$$

To show $B = Q^{-1}AP$, it suffices to show $QB = AP$.

$$\begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 0 & \frac{13}{2} & \frac{11}{2} \\ 0 & 0 & \frac{11}{2} & \frac{13}{2} \\ 4 & 4 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & -1 \\ 4 & 4 & 0 & 0 \\ 0 & 0 & 12 & 12 \end{pmatrix}$$
$$= \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 12 \end{pmatrix}\begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \quad \square \tag{23}$$

11.3. **Base changes.** Here, we consider the relationship of the above mentioned matrices A and B. Summarizing arguments so far, we have a commutative diagram (mapping and its composition is determined irrespective of the route):

$$\begin{array}{ccc} K^n & \xrightarrow{B} & K^m \\ \downarrow \phi_{\tilde{\mathbb{E}}}^{-1} & & \uparrow \phi_{\tilde{\mathbb{F}}} \\ V & \xrightarrow{T} & W \\ \downarrow \phi_{\mathbb{E}} & & \uparrow \phi_{\mathbb{F}}^{-1} \\ K^n & \xrightarrow{A} & K^m \end{array} \tag{24}$$

If we want to have a mapping from upper left K^n to upper right K^m, there are several ways containing passing B way, and passing A way, and resulting mappings are the same. Namely, two correspondence:

$$\begin{array}{ccccccc} \tilde{\mathtt{x}} & & & \stackrel{B}{\longmapsto} & & & \tilde{\mathtt{y}} \\ \tilde{\mathtt{x}} & \stackrel{P}{\longmapsto} & \mathtt{x} & \stackrel{A}{\longmapsto} & \mathtt{y} & \stackrel{Q^{-1}}{\longmapsto} & \tilde{\mathtt{y}} \end{array} \tag{25}$$

are the same, and therefore

$$\begin{cases} \tilde{\mathtt{y}} = B\tilde{\mathtt{x}}. \\ \tilde{\mathtt{y}} = Q^{-1}\mathtt{y} = Q^{-1}A\mathtt{x} = Q^{-1}AP\tilde{\mathtt{x}}. \end{cases} \tag{26}$$

are the same linear mappings, say, $B = Q^{-1}AP$.

If bases $\mathbb{E}, \mathbb{F}$ of V, W, respectively, are changed to $\tilde{\mathbb{E}}, \tilde{\mathbb{F}}$, then the matrix A representing T is changed to the matrix $Q^{-1}AP$. If $W = V$, then usually $\mathbb{E} = \mathbb{F}$, $\tilde{\mathbb{E}} = \tilde{\mathbb{F}}$ are supposed, and therefore A is changed to $P^{-1}AP$. This idea is applied to select a good basis and to represent T by an easier matrix.

11.4. **Images and kernels of linear mappings.** If a linear mapping T is not represented by multiplying vectors by a matrix, then the above method to reduce T to a matrix A. In particular, for the image $\operatorname{Im} T$ of T and the kernel $\operatorname{Ker} T$ of T, to determine bases of them, the following method is helpful.

(i) First, select bases $\mathbb{E}$ and $\mathbb{F}$ of V and W, respectively, and represent T by a matrix A.

(ii) Determine a basis $\mathbb{H} = \langle \mathtt{h}_1, \mathtt{h}_2, \ldots, \mathtt{h}_r \rangle$ of the image of T_A, and a basis $\mathbb{H} = \langle \mathtt{g}_1, \mathtt{g}_2, \ldots, \mathtt{g}_s \rangle$ of the kernel of T_A.

(iii) $\tilde{\mathbb{H}} = \langle \mathbb{F}\mathtt{h}_1, \mathbb{F}\mathtt{h}_2, \ldots, \mathbb{F}\mathtt{h}_r \rangle$ is a basis of the image of T, and $\tilde{\mathbb{G}} = \langle \mathbb{E}\mathtt{h}_1, \mathbb{E}\mathtt{h}_2, \ldots, \mathbb{E}\mathtt{h}_s \rangle$ is a basis of the kernel of T.

Proof. (1) By isomorphism $\phi_{\mathbb{F}}$, $\operatorname{Im} T$ is mapped to $\operatorname{Im} T_A$. It is clear by the diagram, but dare to show below.

$$\operatorname{Im} T_A = T_A(K^n) = (\phi_{\mathbb{F}} \circ T \circ \phi_{\mathbb{E}}^{-1})(K^n) = (\phi_{\mathbb{F}} \circ T)(V) = \phi_{\mathbb{F}}(T(V)) = \phi_{\mathbb{F}}(\operatorname{Im} T) \tag{27}$$

Hence $\phi_{\mathbb{F}}$ is an isomorphism from $\operatorname{Im} T$ to $\operatorname{Im} T_A$, and thus $\operatorname{Im} T \simeq \operatorname{Im} T_A$. A basis of $\operatorname{Im} T$ and a basis of $\operatorname{Im} T_A$ are mapped to each other by $\phi_{\mathbb{F}}$, we can make a basis $\langle \mathbb{F}\mathtt{h}_1, \mathbb{F}\mathtt{h} = 2, \ldots, \mathbb{F}\mathtt{h}_r \rangle$ of $\operatorname{Im} T$ by a basis $\langle \mathtt{h}_1, \mathtt{h}_2, \ldots, \mathtt{h}_r \rangle$ of $\operatorname{Im} T_A$.

(2) An isomorphism $\phi_{\mathbb{E}}$, $\operatorname{Ker} T$ is mapped to $\operatorname{Ker} T_A$. Because letting $x = \mathbb{E}\mathtt{x}$,

$$Tx = 0 \iff T\mathbb{E}\mathtt{x} = 0 \iff \mathbb{F}A\mathtt{x} = 0 \iff A\mathtt{x} = \mathtt{0} \tag{28}$$

Hence corresponding x and $\mathtt{x}$ are simultaneously belongs to the kernel or not. Accordingly, $\operatorname{Ker} T \simeq \operatorname{Ker} T_A$, and from a basis $\langle \mathtt{g}_1, \mathtt{g}_2, \ldots, \mathtt{g}_s \rangle$ of $\operatorname{Ker} T_A$, we have a basis $\langle \mathbb{E}\mathtt{g}_1, \mathbb{E}\mathtt{g}_2, \ldots, \mathbb{E}\mathtt{g}_s \rangle$ of $\operatorname{Ker} T$. □

(exercise03) Let T be defined in (exercise02). (1) Determine a basis of the image of T. (2) Determine a basis of the kernel of T.

(ans) (1) We determine a basis of the image of T_A, by elementary column operations on A transforming it into a stair-like form.

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 12 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \quad \therefore \ \mathbb{H} = \langle \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \rangle.$$

$$\therefore \ \tilde{\mathbb{H}} = \langle \mathbb{F} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \mathbb{F} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \mathbb{F} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \rangle = \langle 1, t, t^2 \rangle. \tag{29}$$

(2) To determine a basis of the kernel, we solve $A\mathtt{x} = \mathtt{0}$.

$$A = \begin{matrix} \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 12 \end{pmatrix} \\ \begin{matrix} x & y & z & w \end{matrix} \end{matrix} \longrightarrow \begin{matrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 12 & 0 \end{pmatrix} \\ \begin{matrix} y & z & w & x \end{matrix} \end{matrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

$$\therefore \ \begin{pmatrix} y \\ z \\ w \\ x \end{pmatrix} = \alpha \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}. \quad \therefore \ \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \alpha \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \quad \therefore \ \mathbb{G} = \langle \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \rangle.$$

$$\therefore \ \tilde{\mathbb{G}} = \langle \mathbb{E} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \rangle = \langle 1 \rangle. \tag{30}$$

(exercise04) Let V be a complex vector space made of all 2×2 complex matrices. $S = \begin{pmatrix} 3 & 2i \\ 2i & 1 \end{pmatrix}$, and define a transformation T of V by $TX = SX - XS$. (1) Show that T is a liner transformation of V.
(2) Taking a basis of V: $\mathbb{E} = \langle \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \rangle = \langle R_1, R_2, R_3, R_4 \rangle$, determine the matrix of T with respect to $\mathbb{E}$. (3) Finding bases of the image and the kernel of T_A, determine bases of the image and the kernel of T.

(ans) (1) For any $X, Y \in V$, and any $c \in \mathbb{C}$, we have

$$\begin{aligned} T(X+Y) = &\ S(X+Y)-(X+Y)S = SX+SY-XS-YS \\ = &\ (SX-XS)+(SY-YS) = TX+TY. \\ T(cX) = &\ ScX - cXS = cSX - cXS = c(SX-XS) = cTX. \end{aligned} \tag{31}$$

Hence T is a linear transformation of V. □

(2) For $X = \begin{pmatrix} x & y \\ z & w \end{pmatrix}$,

$$\begin{aligned} TX = &\ \begin{pmatrix} 3 & 2i \\ 2i & 1 \end{pmatrix}\begin{pmatrix} x & y \\ z & w \end{pmatrix} - \begin{pmatrix} x & y \\ z & w \end{pmatrix}\begin{pmatrix} 3 & 2i \\ 2i & 1 \end{pmatrix} \\ = &\ \begin{pmatrix} -2iy+2iz & -2ix+2y+2iw \\ 2ix-2z-2iw & 2iy-2iz \end{pmatrix}. \end{aligned} \tag{32}$$

Here, letting $X = R_i$,

$$\begin{aligned} \therefore\ T\mathbb{E} = &\ T\begin{pmatrix} R_1 & R_2 & R_3 & R_4 \end{pmatrix} = \begin{pmatrix} TR_1 & TR_2 & TR_3 & TR_4 \end{pmatrix} \\ = &\ \left(\begin{pmatrix} 0 & -2i \\ 2i & 0 \end{pmatrix} \begin{pmatrix} -2i & 2 \\ 0 & 2i \end{pmatrix} \begin{pmatrix} 2i & 0 \\ -2 & -2i \end{pmatrix} \begin{pmatrix} 0 & 2i \\ -2i & 0 \end{pmatrix} \right) \\ = &\ \mathbb{E}A = \begin{pmatrix} R_1 & R_2 & R_3 & R_4 \end{pmatrix} \begin{pmatrix} 0 & -2i & 2i & 0 \\ -2i & 2 & 0 & 2i \\ 2i & 0 & -2 & -2i \\ 0 & 2i & -2i & 0 \end{pmatrix}. \\ \therefore\ A = &\ \begin{pmatrix} 0 & -2i & 2i & 0 \\ -2i & 2 & 0 & 2i \\ 2i & 0 & -2 & -2i \\ 0 & 2i & -2i & 0 \end{pmatrix}. \end{aligned} \tag{33}$$

(3) To determine $\mathbb{H}$, A is transformed by elementary column operation,

$$\begin{aligned} &\begin{pmatrix} 0 & -2i & 2i & 0 \\ -2i & 2 & 0 & 2i \\ 2i & 0 & -2 & -2i \\ 0 & 2i & -2i & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 2i & -2i & 0 & 0 \\ 0 & 2 & -2i & 2i \\ -2 & 0 & 2i & -2i \\ -2i & 2i & 0 & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 2i & -2i & 0 & 0 \\ 0 & 2 & -2i & 0 \\ -2 & 0 & 2i & 0 \\ -2i & 2i & 0 & 0 \end{pmatrix} \\ &\longrightarrow \begin{pmatrix} 2i & 0 & 0 & 0 \\ 0 & 2 & -2i & 0 \\ -2 & -2 & 2i & 0 \\ -2i & 0 & 0 & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 2i & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ -2 & -2 & 0 & 0 \\ -2i & 0 & 0 & 0 \end{pmatrix}. \\ &\therefore\ \mathbb{H} = \left\langle \begin{pmatrix} 2i \\ 0 \\ -2 \\ -2i \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ -2 \\ 0 \end{pmatrix} \right\rangle. \\ &\therefore\ \tilde{\mathbb{H}} = \left\langle \mathbb{E}\begin{pmatrix} 2i \\ 0 \\ -2 \\ -2i \end{pmatrix}, \mathbb{E}\begin{pmatrix} 0 \\ 2 \\ -2 \\ 0 \end{pmatrix} \right\rangle = \left\langle \begin{pmatrix} 2i & 0 \\ -2 & -2i \end{pmatrix}, \begin{pmatrix} 0 & 2 \\ -2 & 0 \end{pmatrix} \right\rangle. \end{aligned} \tag{34}$$

To determine $\mathbb{G}$, we solve the equation $A\mathtt{x} = \mathtt{0}$.

$$\begin{pmatrix} 0 & -2i & 2i & 0 \\ -2i & 2 & 0 & 2i \\ 2i & 0 & -2 & -2i \\ 0 & 2i & -2i & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} -2i & 2 & 0 & 2i \\ 0 & -2i & 2i & 0 \\ 2i & 0 & -2 & -2i \\ 0 & 2i & -2i & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & i & 0 & -1 \\ 0 & -2i & 2i & 0 \\ 2i & 0 & -2 & -2i \\ 0 & 2i & -2i & 0 \end{pmatrix}$$

$$\longrightarrow \begin{pmatrix} 1 & i & 0 & -1 \\ 0 & -2i & 2i & 0 \\ 0 & 2 & -2 & 0 \\ 0 & 2i & -2i & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & i & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 2 & -2 & 0 \\ 0 & 2i & -2i & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & i & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

$$\therefore \quad \mathbb{G} = \langle \begin{pmatrix} -i \\ 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \rangle.$$

$$\therefore \quad \tilde{\mathbb{G}} = \langle \mathbb{E}\begin{pmatrix} -i \\ 1 \\ 1 \\ 0 \end{pmatrix}, \mathbb{E}\begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \rangle = \langle \begin{pmatrix} -i & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \rangle. \tag{35}$$

11.5. **Canonical forms of matrices.** For a linear mapping $T : V \longrightarrow W$, by well choosing bases of V and W, consider to simplify the matrix of T with respect to those bases. Chapter 10, Theorem 12, composition of bases in the proof is adopted. That is, take a basis $\langle \mathtt{e}_{r+1}, \ldots, \mathtt{e}_n \rangle$ of $\mathrm{Ker}\, T$, which is extend this to a basis $\mathbb{E} = \langle \mathtt{e}_1, \ldots, \mathtt{e}_r, \mathtt{e}_{r+1}, \ldots, \mathtt{e}_n \rangle$ of V. Further, $T\mathtt{e}_i = \mathtt{f}_i$ $(i = 1, \ldots, r)$, make a basis of $\mathrm{Im}\, T$. Finally, extending this basis to a basis of W. Then

$$\begin{aligned} T\mathbb{E} = & \begin{pmatrix} T\mathtt{e}_1 & \ldots & T\mathtt{e}_r & T\mathtt{e}_{r+1} & \ldots & T\mathtt{e}_n \end{pmatrix} \\ = & \begin{pmatrix} \mathtt{f}_1 & \ldots & \mathtt{f}_r & \mathtt{0} & \ldots & \mathtt{0} \end{pmatrix} \\ = & \begin{pmatrix} \mathtt{f}_1 & \ldots & \mathtt{f}_r & \mathtt{f}_{r+1} & \ldots & \mathtt{f}_m \end{pmatrix} \begin{pmatrix} E_r & O_{r,n-r} \\ O_{m-r,r} & O_{m-r,n-r} \end{pmatrix} = \mathbb{F} F_{mn}(r). \end{aligned} \tag{36}$$

Hence the matrix of T with respect to $\mathbb{E}$ and $\mathbb{F}$ is $F_{mn}(r)$. Especially, if $T = T_A$, a bases are regarded as nonsingular matrices,

$$AP = QF_{mn}(r). \qquad \therefore \quad Q^{-1}AP = F_{mn}(r). \tag{37}$$

This corresponds to the fact that a matrix A is transformed into a canonical form by elementary operations.

Next, consider that T is a linear transformation of V. In this case we choose the same basis of domain and codomain, thus the result is not so simple as above. Detailed contents shall be explained later, here we study basic principle. For $T : V \longrightarrow V$ if a subspace W of V satisfies that $T(W) \subset W$, then W is said to be T-invariant subspace. Take a basis of W is $\langle \mathtt{e}_1, \ldots, \mathtt{e}_s \rangle$, and extend it to a basis $\mathbb{E} = \langle \mathtt{e}_1, \ldots, \mathtt{e}_n \rangle$ of V. Then by

$$\begin{aligned} T\mathbb{E} = & \begin{pmatrix} T\mathtt{e}_1 & \ldots & T\mathtt{e}_s & T\mathtt{e}_{s+1} & \ldots & T\mathtt{e}_n \end{pmatrix} \\ = & \begin{pmatrix} \mathtt{e}_1 & \ldots & \mathtt{e}_s & \mathtt{e}_{s+1} & \ldots & \mathtt{e}_n \end{pmatrix} \begin{pmatrix} Q_{11} & Q_{12} \\ O_{n-s,s} & Q_{22} \end{pmatrix} = \mathbb{E}Q \end{aligned}, \tag{38}$$

we have the matrix Q of T with respect to $\mathbb{E}$.

More strongly, if V is the direct sum of two T-invariant subspaces W_1 and W_2, say,

$$V = W_1 \oplus W_2, \tag{39}$$

then arrange bases of those subspaces and make a basis $\mathbb{E} = \langle \mathtt{e}_1, \mathtt{e}_s, \mathtt{e}_{s+1}, \dots, \mathtt{e}_n \rangle$ of V, we have

$$\begin{aligned} T\mathbb{E} &= \begin{pmatrix} T\mathtt{e}_1 & \dots & T\mathtt{e}_s & T\mathtt{e}_{s+1} & \dots & T\mathtt{e}_n \end{pmatrix} \\ &= \begin{pmatrix} \mathtt{e}_1 & \dots & \mathtt{e}_s & \mathtt{e}_{s+1} & \dots & \mathtt{e}_n \end{pmatrix} \begin{pmatrix} Q_{11} & O_{s,n-s} \\ O_{n-s,s} & Q_{22} \end{pmatrix} = \mathbb{E}Q. \end{aligned} \tag{40}$$

This case Q is a block diagonal matrix. In general, if V is the direct sum of several subspaces

$$V = W_1 \oplus \cdots \oplus W_s, \tag{41}$$

then arranged bases of $W_1, \dots, W_s$ is a basis of V, and the matrix Q of T is the following.

$$\begin{pmatrix} Q_1 & & & O \\ & Q_2 & & \\ & & \ddots & \\ O & & & Q_s \end{pmatrix} \tag{42}$$

In particular, if $T = T_A$, then by the matrix P of arranged basis, we have

$$AP = PQ. \qquad \therefore \ P^{-1}AP = Q. \tag{43}$$

P is called a transformation matrix sending A to Q. Under a suitable condition, we can make Q a diagonal matrix, then this linear transformation T is said to be diagonalizable. (⇒ Chapter 12–13) However, a matrix is not always diagonalizable, if it is impossible, it can be Jordan's canonical form. (⇒ Chapter 14)

11.6. **Operations of linear mappings and matrices.** For linear transformations of T and S of V, let A and B be matrices with respect to $\mathbb{E}$, then

$$\begin{aligned} &(T+S)\mathbb{E} = T\mathbb{E} + S\mathbb{E} = \mathbb{E}A + \mathbb{E}B = \mathbb{E}(A+B) \\ &(TS)\mathbb{E} = T(S\mathbb{E}) = T(\mathbb{E}B) = (T\mathbb{E})B = (\mathbb{E}A)B = \mathbb{E}(AB) \\ &(kT)\mathbb{E} = k(T\mathbb{E}) = k(\mathbb{E}A) = \mathbb{E}(kA) \end{aligned} \tag{44}$$

Namely, if a basis in fixed, the operations such as the sum, composition, and scalar multiplication are simply corresponds to matrix operations. Especially, a polynomial of linear transformation T is (product is regarded as composition) $\Phi(T)$, using (44) repeatedly, we have

$$\Phi(T)\mathbb{E} = \mathbb{E}\Phi(A). \tag{45}$$

This result is used for the proof of Hamilton Cayley, etc.