Linear Algebra

Min Yan

September 9, 2019

Contents

Chapter 1

System of Linear Equations

We learned how to solve equations in high school. The usual idea is to simplify equations sufficiently until they become easy to solve. After systematically analysing how to simplify linear equations (row operations), we determine the simplest equations we can get at the end (row echelon form). Then we may answer what kind of solution the equations have from the shape of the simplest equations.

Most calculations in linear algebra are merely solving linear equations in various guise. It is critical to fully understand this most basic calculation process.

1.1 Gauss Elimination

Quantities are often related by equations. The following are *linear equations*

$$\begin{gathered} x + 3y = 5, \\ 2x + 4y = 6, \\ u + 3v = 5, \\ x_1 - 2x_2 + 5x_3 + 10x_4 - 4x_5 = 12, \\ 2x + 3y - z = 1. \end{gathered}$$

Note that the third is essentially the same as the first, with the only difference in notations for *variables*. The following are also linear equations because they can be rewritten as the linear equations above

$$\begin{gathered} x + 3(y-1) = 2, \\ 2x = 6 - 4y, \\ 3(v-2) = -u - 1, \\ x_1 + 5x_3 + 10x_4 = 12 + 2x_2 + 4x_5, \\ 2(x-2) + 3(y+1) = z. \end{gathered}$$

The following are *non-linear* equations

$$x^2 + 3y^2 = 5,$$
$$2x^3 + 4y^4 = 6,$$
$$3u^2 + v^2 + 2uv = 5,$$
$$\sqrt{x} + \sqrt{2 + y^2} + \sqrt[3]{z} = 3,$$
$$\sin x + y \cos y = 0.$$

Specifically, they are respectively quadratic, quartic, quadratic, algebraic, and transcendental equations.

Exercise 1.1. Which ones are linear equations?

1. $3x - \sqrt{2}y - 1 = 0.$
2. $1 = 2x - y.$
3. $1 = 2xy.$
4. $(x + y - 1)(x + 1) = 0.$
5. $x = y.$
6. $3(x - 1) = 4(y + 1).$
7. $e^{x+y} = 1.$
8. $e^x y = 1.$
9. $ex + 1 = \pi y.$

Sometimes, several quantities are related by several equations, which we call a *system of equations*. If all the equations are linear, then we have a *system of linear equations*. The usual way of solving a system of equations is to first simplify the system by eliminating variables. The process is called *Gaussian elimination*. Then we solve the simplified equations one by one, by substituting the solution to simpler equations to more complex equations. The process is called *back substitution*.

Example 1.1.1. The following is a system of 2 linear equations in 2 variables.

$$x + 3y = 5,$$
$$2x + 4y = 6.$$

We may eliminate x by $\text{Eq}_2 - 2\text{Eq}_1$ (the second equation subtracting twice of the first equation). The result is

$$-2y = (2 - 2 \cdot 1)x + (4 - 2 \cdot 3)y = 6 - 2 \cdot 5 = -4.$$

Then we get $y = \frac{-4}{-2} = 2$. Substituting $y = 2$ into the first equation, we get $x + 3 \cdot 2 = 5$, from which we get $x = -1$. Therefore the solution is $x = -1$ and $y = 2$.

Alternatively, we may first use $4\text{Eq}_1 - 3\text{Eq}_2$ to eliminate y

$$-2x = (4 \cdot 1 - 3 \cdot 2)x + (4 \cdot 3 - 3 \cdot 4)y = 4 \cdot 5 - 3 \cdot 6 = 2.$$

Then we get $x = -1$. Substituting into Eq_1, we get $-1 + 3y = 5$, which implies $y = 2$.

Exercise 1.2. Suppose x and y satisfy Eq_1 and Eq_2. Why do they also satisfy $\mathrm{Eq}_2 - 2\mathrm{Eq}_1$? This justifies the elimination in the example.

Exercise 1.3. Solve systems of equations. What observations can you make by comparing with Example 1.1.1?

1. $\begin{aligned} x_1 + 3x_2 &= 5, \\ 2x_1 + 4x_2 &= 6. \end{aligned}$

2. $\begin{aligned} 2x + 4y &= 6, \\ x + 3y &= 5. \end{aligned}$

3. $\begin{aligned} 10x + 30y &= 50, \\ 2x + \ \ 4y &= 6. \end{aligned}$

4. $\begin{aligned} x + 3y &= 5, \\ 2x + 4y &= 6, \\ x + 3y &= 5. \end{aligned}$

5. $\begin{aligned} x + 3y &= 5, \\ -2x - 4y &= -6. \end{aligned}$

6. $\begin{aligned} x + 3y &= 5, \\ 2x + 4y &= 6, \\ 0 &= 0. \end{aligned}$

Example 1.1.2. The following is a system of 3 linear equations in 3 variables.

$$\begin{aligned} x + 4y + 7z &= 10, \\ 2x + 5y + 8z &= 11, \\ 3x + 6y + 9z &= 12. \end{aligned}$$

We apply $\mathrm{Eq}_2 - 2\mathrm{Eq}_1$ and $\mathrm{Eq}_3 - 3\mathrm{Eq}_1$ to eliminate x in Eq_2 and Eq_3

$$\begin{aligned} x + 4y + \ \ 7z &= 10, \\ - 3y - \ \ 6z &= -9, \\ - 6y - 12z &= -18. \end{aligned}$$

Then we further apply $\mathrm{Eq}_3 - 2\mathrm{Eq}_2$ to eliminate y in Eq_3

$$\begin{aligned} x + 4y + 7z &= 10, \\ - 3y - 6z &= -9, \\ 0 &= 0. \end{aligned}$$

It happens that z is also eliminated in Eq_3, and the equation becomes an identity. Then we solve the remaining simplest Eq_2 to get $y = 3 - 2z$. Substituting into Eq_1, we get $x + 4(-2z + 3) + 7z = 10$. Then $x = -2 + z$, and we get the general solution

$$x = -2 + z, \quad y = 3 - 2z, \quad z \text{ arbitrary}.$$

Example 1.1.3. The following is a system of 3 linear equations in 4 variables.

$$\begin{aligned} x + 4y + 7z + 10w &= 0, \\ 2x + 5y + 8z + 11w &= 0, \\ 3x + 6y + 9z + 12w &= 0. \end{aligned}$$

The system is homogeneous because the *right side* consists of only 0. Homogeneous equations have the property that $x = y = z = w = 0$ is always a solution.

We may use the same elimination in Example 1.1.2. However, we may also choose to use different elimination. By $\text{Eq}_1 - \text{Eq}_2$ and $\text{Eq}_2 - \text{Eq}_3$, we get

$$\begin{aligned} -x - \quad y - \quad z - \quad w &= 0, \\ -x - \quad y - \quad z - \quad w &= 0, \\ -3x + 6y + 9z + 12z &= 0. \end{aligned}$$

Then by $\text{Eq}_1 - \text{Eq}_2$ and $\text{Eq}_3 + 3\text{Eq}_2$, we get

$$\begin{aligned} 0 &= 0, \\ -x - \quad y - \quad z - \quad w &= 0, \\ 3y + 6z + 9w &= 0. \end{aligned}$$

While this is good enough for solving equations, we may make cosmetic improvements by using $\text{Eq}_1 \leftrightarrow \text{Eq}_2$ (exchange the first and second equations) and $\text{Eq}_2 \leftrightarrow \text{Eq}_3$ to rearrange the equations from the most complicated to the simplest

$$\begin{aligned} -x - \quad y - \quad z - \quad w &= 0, \\ 3y + 6z + 9w &= 0 \\ 0 &= 0. \end{aligned}$$

We may also simplify the coefficients by $-\text{Eq}_1$ (multiplying -1 to the first equation) and $\frac{1}{3}\text{Eq}_2$

$$\begin{aligned} x + y + \quad z + \quad w &= 0, \\ y + 2z + 3w &= 0, \\ 0 &= 0. \end{aligned}$$

Then we get solution

$$x = z + 2w, \quad y = -2z - 3w, \quad z, w \text{ arbitrary}.$$

Exercise 1.4. Use the elimination in Example 1.1.3 to solve the system in Example 1.1.2. You should get the same solution.

Exercise 1.5. In Examples 1.1.2 and 1.1.3, we use three kinds of modifications on systems of equations: $\text{Eq}_2 - 2\text{Eq}_1$, $\text{Eq}_1 \leftrightarrow \text{Eq}_2$, $\frac{1}{3}\text{Eq}_2$. Explain that these modifications preserve solutions of the system.

Exercise 1.6. Solve systems of equations.

1. $$\begin{aligned} 3x + 6y + 9z &= 12, \\ x + 4y + 7z &= 10, \\ 2x + 5y + 8z &= 11. \end{aligned}$$

2. $$\begin{aligned} x + 4y + 7z &= 10, \\ 2x + 5y + 8z &= 11. \end{aligned}$$

3. $$\begin{aligned} x_1 + 4x_2 + 7x_3 &= 10, \\ 2x_1 + 5x_2 + 8x_3 &= 11, \\ x_1 + 4x_2 + 7x_3 &= 10, \\ 3x_1 + 6x_2 + 9x_3 &= 12. \end{aligned}$$

4. $x_1 + 4x_2 + 7x_3 = 10.$

5. $3x_1 + 6x_2 + 9x_3 = 12.$

6. $$\begin{aligned} x_1 + 4x_2 + 7x_3 &= 10, \\ 3x_1 + 6x_2 + 9x_3 &= 12. \end{aligned}$$

1.2 Augmented Matrix

A *linear equation* in variables $x_1, x_2, \ldots, x_n$ is of the form

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

where the *coefficients* $a_1, a_2, \ldots, a_n$ and the *right side* b are numbers. A *system of linear equations* is a collection of linear equations involving the same variables. The following is a system of m linear equations in n variables.

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

Exercise 1.7. Rewrite systems of linear equations in the standard form.

1. $2x + 1 = 3y$, $4 = 3(x - 1) - y$.
2. $1 + u = 2(v - 1) + w$, $2w = 1 + 3v$.
3. $x = y = z$.
4. $x_1 + 2x_2 = x_2 + 2x_3 = x_3 + 2x_1$.
5. $x_1 = x_2 = \cdots = x_n = 1$.
6. $x_1 + x_2 = x_2 + x_3 = \cdots = x_{n-1} + x_n$.

Exercise 1.8. Suppose u_1, u_2, u_3 and v_1, v_2, v_3 are solutions of $a_1x_1 + a_2x_2 + a_3x_3 = b_1$ and $a_1x_1 + a_2x_2 + a_3x_3 = b_2$. Explain that $u_1 + v_1, u_2 + v_2, u_3 + v_3$ is a solution of $a_1x_1 + a_2x_2 + a_3x_3 = b_1 + b_2$.

Exercise 1.9. Suppose $u_1, u_2, \ldots, u_n$ and $v_1, v_2, \ldots, v_n$ are solutions of $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b_1$ and $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b_2$. Write down a linear equation satisfied by $u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n$.

Exercise 1.10. Explain that the sum of two solutions of a *homogeneous* linear equation $a_1x_1 + a_2x_2 + a_3x_3 = 0$ is still a solution of the equation. What about the sum of two solutions of $a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0$?

Exercise 1.11. Suppose u_1, u_2, u_3 is a solution of $a_1x_1 + a_2x_2 + a_3x_3 = b$. For a number c, write down a linear equation satisfied by cu_1, cu_2, cu_3. Extend your observation to a linear equation of n variables.

Exercise 1.12. For a number c, explain that multiplying c to a solution of a homogeneous linear equation is still a solution of the equation.

We know that changing notations for variables does not really change equations. In other words, the essential information about a system of linear equations is the *coefficients* and the numbers on the right side. They form the *coefficient matrix* and the *right side vector*

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \quad \vec{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

The m equations correspond to the m *rows* of A, the n variables correspond to the n *columns* of A, and A is an $m \times n$ matrix. Then we may write the corresponding system of linear equation as $A\vec{x} = \vec{b}$, with

$$A\vec{x} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix}.$$

For example, we have

$$\begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 + 4x_2 + 7x_3 \\ 2x_1 + 5x_2 + 8x_3 \\ 3x_1 + 6x_2 + 9x_3 \end{pmatrix}.$$

Then the following equality

$$A\vec{x} = \begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 10 \\ 11 \\ 12 \end{pmatrix} = \vec{b}$$

means the system of linear equations in Example 1.1.2.

The whole system of linear equations corresponds to the *augmented matrix*

$$(A\ \vec{b}) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{pmatrix}.$$

The first n columns of the augmented matrix correspond to variables, and the m rows of the augmented matrix correspond to equations. The following are the augmented matrices for systems of linear equations in Examples 1.1.1, 1.1.2, 1.1.3

$$\begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}, \quad \begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{pmatrix}, \quad \begin{pmatrix} 1 & 4 & 7 & 10 & 0 \\ 2 & 5 & 8 & 11 & 0 \\ 3 & 6 & 9 & 12 & 0 \end{pmatrix}.$$

Example 1.2.1 (Vandermonde matrix). Given some data, we would like to fit data into a polynomial. For example, we wish to find a quadratic polynomial $f(t) = x_0 + x_1 t + x_2 t^2$, such that $f(t_0) = b_0$, $f(t_1) = b_1$, $f(t_2) = b_2$. The problem becomes finding suitable coefficients x_0, x_1, x_2 in the polynomial, such that

$$\begin{aligned} x_0 + t_0 x_1 + t_0^2 x_2 &= b_0, \\ x_0 + t_1 x_1 + t_1^2 x_2 &= b_1, \\ x_0 + t_2 x_1 + t_2^2 x_2 &= b_2. \end{aligned}$$

The system of linear equations has

$$A = V(t_0, t_1, t_2) = \begin{pmatrix} 1 & t_0 & t_0^2 \\ 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \end{pmatrix}, \quad \vec{b} = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix}.$$

The matrix V is the *Vandermonde matrix*. The general Vandermonde matrix is

$$V(t_0, t_1, t_2, \dots, t_n) = \begin{pmatrix} 1 & t_0 & t_0^2 & \cdots & t_0^n \\ 1 & t_1 & t_1^2 & \cdots & t_1^n \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & t_n & t_n^2 & \cdots & t_n^n \end{pmatrix}.$$

Exercise 1.13. Write down augmented matrices of systems of linear equations in Exercises 1.3, 1.6, 1.7.

Exercise 1.14. Write down $A\vec{x}$.

1. $\begin{pmatrix} 1 & 3 & 5 & 7 \\ 2 & 4 & 6 & 8 \end{pmatrix}$.
2. $\begin{pmatrix} 1 & 3 & 0 & 7 \\ 2 & 4 & 0 & 8 \end{pmatrix}$.
3. $\begin{pmatrix} 1 & 0 & 5 & 7 \\ 2 & 0 & 6 & 8 \end{pmatrix}$.
4. $\begin{pmatrix} 1 & 2 & 3 & 4 \end{pmatrix}$.
5. $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$.
6. $\begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{pmatrix}$.
7. $\begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$.
8. $\begin{pmatrix} 1 & 4 & 0 \\ 2 & 5 & 0 \\ 3 & 6 & 0 \end{pmatrix}$.
9. $\begin{pmatrix} 0 & 1 & 4 \\ 0 & 2 & 5 \\ 0 & 3 & 6 \end{pmatrix}$.
10. $\begin{pmatrix} 1 & 0 & 3 \\ 0 & 2 & 4 \end{pmatrix}$.
11. $\begin{pmatrix} 0 & 2 & 4 \\ 1 & 0 & 3 \end{pmatrix}$.
12. $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$.
13. $\begin{pmatrix} 1 & 0 & 0 & 4 \\ 0 & 2 & 0 & 5 \\ 0 & 0 & 3 & 6 \end{pmatrix}$.

Exercise 1.15. Write down systems of linear equations with matrices in Exercise 1.14 as augmented matrices.

1.3 Row Operation

Since a system of linear equations is equivalent to its augmented matrix, and equations in the system correspond to rows of the matrix, the manipulations (Gaussian eliminations) of equations in Section 1.1 correspond to manipulations of rows in the matrix.

Example 1.3.1. The augmented matrix for the system of linear equations in Example 1.1.2 is

$$(A\ \vec{b}) = \begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{pmatrix}.$$

The operations $\text{Eq}_2 - 2\text{Eq}_1$ and $\text{Eq}_3 - 3\text{Eq}_1$ correspond to $\text{Row}_2 - 2\text{Row}_1$ (second row subtracting twice of first row) and $\text{Row}_3 - 3\text{Row}_1$

$$\begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{pmatrix} \xrightarrow[\text{Row}_3-3\text{Row}_1]{\text{Row}_2-2\text{Row}_1} \begin{pmatrix} 1 & 4 & 7 & 10 \\ 0 & -3 & -6 & -9 \\ 0 & -6 & -12 & -18 \end{pmatrix}.$$

The further operation $\text{Eq}_3 - 2\text{Eq}_2$ corresponds to

$$\begin{pmatrix} 1 & 4 & 7 & 10 \\ 0 & -3 & -6 & -9 \\ 0 & -6 & -12 & -18 \end{pmatrix} \xrightarrow{\text{Row}_3-2\text{Row}_2} \begin{pmatrix} 1 & 4 & 7 & 10 \\ 0 & -3 & -6 & -9 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Example 1.3.2. The Gaussian eliminations in Example 1.1.3 correspond to

$$\begin{aligned} \begin{pmatrix} 1 & 4 & 7 & 10 & 0 \\ 2 & 5 & 8 & 11 & 0 \\ 3 & 6 & 9 & 12 & 0 \end{pmatrix} &\xrightarrow[\text{Row}_2-\text{Row}_3]{\text{Row}_1-\text{Row}_2} \begin{pmatrix} -1 & -1 & -1 & -1 & 0 \\ -1 & -1 & -1 & -1 & 0 \\ 3 & 6 & 9 & 12 & 0 \end{pmatrix} \\ &\xrightarrow[\text{Row}_3+3\text{Row}_2]{\text{Row}_1-\text{Row}_2} \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & -1 & -1 & 0 \\ 0 & 3 & 6 & 9 & 0 \end{pmatrix} \\ &\xrightarrow[\text{Row}_2\leftrightarrow\text{Row}_3]{\text{Row}_1\leftrightarrow\text{Row}_2} \begin{pmatrix} -1 & -1 & -1 & -1 & 0 \\ 0 & 3 & 6 & 9 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \xrightarrow[\frac{1}{3}\text{Row}_2]{-\text{Row}_1} \begin{pmatrix} 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 2 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \end{aligned}$$

Note that the same operations give another way of simplifying the augmented matrix in Example 1.3.1

$$\begin{aligned} \begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{pmatrix} &\to \begin{pmatrix} -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 \\ 3 & 6 & 9 & 12 \end{pmatrix} \to \begin{pmatrix} 0 & 0 & 0 & 0 \\ -1 & -1 & -1 & -1 \\ 0 & 3 & 6 & 9 \end{pmatrix} \\ &\to \begin{pmatrix} -1 & -1 & -1 & -1 \\ 0 & 3 & 6 & 9 \\ 0 & 0 & 0 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \end{aligned}$$

Of course, we may also add three 0's to the operations in Example 1.3.1 to get alternative operations for the augmented matrix in this example.

The operations on matrices in Examples 1.3.1 and 1.3.2 are *row operations*. We used three row operations in the examples

1. $\text{Row}_i \leftrightarrow \text{Row}_j$: Exchange i-th row and j-th row.

2. $c\text{Row}_i$: Multiply $c \neq 0$ to i-th row.

3. $\text{Row}_i + c\text{Row}_j$: Add c multiple of j-th row to i-th row.

The row operations are allowed because they do not change solutions of corresponding systems of linear equations.

We remark that the third operation is the most useful for simplifying matrices. The first and second operations are used for further cosmetic improvements.

Exercise 1.16. You have solved systems in Exercises 1.3 and 1.6 by Gaussian eliminations. Can you write down the corresponding row operations?

Exercise 1.17. Explain that all row operations can be reversed by row operations.

Exercise 1.18. Explain that row operations on augmented matrix do not change solutions.

1.4 Row Echelon Form

A system is simpler if more coefficients are 0. Therefore the goal of row operations is to produce as many 0 as possible. Moreover, we may measure the complication of a single linear equation by looking at the location of the *leading coefficient* (the first nonzero coefficient in a row). The equation is longer (or more complicated) if the leading coefficient is further to the left.

The key tool for producing 0 is the third operation, that uses some multiple of a nonzero coefficient to "kill" other coefficients in the same column. For example, in the first step in Example 1.3.1, we apply the third operation to the coefficient 1 to kill the coefficients 2 and 3 in the first column. We indicate the idea by presenting the row operations in Example 1.3.1 as the following, where we use $\bullet$ to indicate nonzero leading coefficients, and use $*$ to indicate any (zero or nonzero) numbers

$$\begin{pmatrix} \bullet & * & * & * \\ \bullet & * & * & * \\ \bullet & * & * & * \end{pmatrix} \to \begin{pmatrix} \bullet & * & * & * \\ 0 & \bullet & * & * \\ 0 & \bullet & * & * \end{pmatrix} \to \begin{pmatrix} \bullet & * & * & * \\ 0 & \bullet & * & * \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

We notice that, as long as we have more than one $\bullet$ in the same column, then we may apply the third operation to reduce the number of $\bullet$ in the column. We may repeat such reductions until no two $\bullet$ are in the same column. This means that the shape of the last matrix cannot be further improved by row operations.

Definition 1.4.1. The *row echelon form* is the simplest *shape* of matrix one can get by row operations, and the lengths of rows are arranged from longest to shortest.

Examples 1.3.1 and 1.3.2 show that different row operations on the same matrix may give different row echelon form matrices, but the shapes of these matrices are the same. The rigorous explanation of this fact is given in Section 1.5.

In a row echelon form, the entries occupied by $\bullet$ are *pivots*. The rows and columns containing pivots are *pivot rows* and *pivot columns*. In the row echelon form above, the pivots are the $(1,1)$ and $(2,2)$ entries, the first and second rows are pivot, and the first and second columns are pivot.

The following are all 3×4 row echelon forms (there are 15)

$$\begin{pmatrix} \bullet & * & * & * \\ 0 & \bullet & * & * \\ 0 & 0 & \bullet & * \end{pmatrix} \begin{pmatrix} \bullet & * & * & * \\ 0 & \bullet & * & * \\ 0 & 0 & 0 & \bullet \end{pmatrix} \begin{pmatrix} \bullet & * & * & * \\ 0 & \bullet & * & * \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \bullet & * & * & * \\ 0 & 0 & \bullet & * \\ 0 & 0 & 0 & \bullet \end{pmatrix} \begin{pmatrix} \bullet & * & * & * \\ 0 & 0 & \bullet & * \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} \bullet & * & * & * \\ 0 & 0 & 0 & \bullet \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \bullet & * & * & * \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & \bullet & * & * \\ 0 & 0 & \bullet & * \\ 0 & 0 & 0 & \bullet \end{pmatrix} \begin{pmatrix} 0 & \bullet & * & * \\ 0 & 0 & \bullet & * \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & \bullet & * & * \\ 0 & 0 & 0 & \bullet \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & \bullet & * & * \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & \bullet & * \\ 0 & 0 & 0 & \bullet \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & \bullet & * \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & \bullet \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Exercise 1.19. Which ones are row echelon forms? Which ones are not?

$$\begin{pmatrix} 0 & \bullet & * & * \\ 0 & \bullet & * & * \\ 0 & 0 & \bullet & * \end{pmatrix} \begin{pmatrix} 0 & \bullet & * & * \\ 0 & 0 & 0 & \bullet \\ 0 & 0 & \bullet & * \end{pmatrix} \begin{pmatrix} \bullet & * & * & * \\ 0 & \bullet & * & * \end{pmatrix} \begin{pmatrix} 0 & \bullet & * & * \\ \bullet & * & * & * \end{pmatrix} \begin{pmatrix} \bullet & * & * \\ \bullet & * & * \\ 0 & 0 & \bullet \end{pmatrix} \begin{pmatrix} \bullet & * & * \\ 0 & \bullet & * \\ 0 & 0 & \bullet \end{pmatrix}$$

Exercise 1.20. Find row echelon forms of systems of linear equations in Exercises 1.3, 1.6.

Exercise 1.21. Write down all 2×3 row echelon forms, and all 3×2 row echelon forms. How many $n \times 2$ row echelon forms are there?

Exercise 1.22. Write down all 3×3 row echelon forms. How many $n \times 3$ row echelon forms are there?

Now we discuss how the row echelon form of the augmented matrix tells us the existence and uniqueness of solutions. We consider a system of 3 equations in 5 variables. The augmented matrix has size 3×6. Suppose the row echelon form is

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

Then the third equation is $0 = \bullet$. Since the pivot $\bullet \neq 0$, this is a contradiction. Therefore the system has no solution. In general, if the row echelon form of the augmented matrix has a row $(0\ 0\ \cdots\ 0\ \bullet)$, i.e., the last column is pivot, then the system has no solution.

So we turn to an example that the last column is not pivot

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

The last equation is $\bullet x_5 = *$. Since the pivot $\bullet \neq 0$, we may divide the number and get the unique value of x_5

$$x_5 = \frac{*}{\bullet} = d_5.$$

The second equation is $\bullet x_3 + *x_4 + *x_5 = *$. Substituting $x_5 = d_5$ and dividing the nonzero coefficient $\bullet$ of x_3, we get

$$x_3 = d_3 + c_{34}x_4, \quad x_4 \text{ arbitrary}.$$

The first equation is $\bullet x_1 + *x_2 + *x_3 + *x_4 + *x_5 = *$. Substituting the formulae for x_3, x_5 and dividing the nonzero coefficient $\bullet$ of x_1, we get

$$x_1 = d_1 + c_{12}x_2 + c_{14}x_4, \quad x_2 \text{ arbitrary}.$$

We conclude the system has solution, and the solution is of the form

$$x_1 = d_1 + c_{12}x_2 + c_{14}x_4, \quad x_3 = d_3 + c_{34}x_4, \quad x_5 = d_5, \quad x_2, x_4 \text{ arbitrary}.$$

Naturally, we call x_2, x_4 *free variables* and call x_1, x_3, x_5 *non-free variables*, and have the obvious correspondence between variables and columns of coefficient matrix

$$\text{free (variable)} \leftrightarrow \text{non-pivot (column)}, \quad \text{non-free (variable)} \leftrightarrow \text{pivot (column)}.$$

In general, the solution of $A\vec{x} = \vec{b}$ has the following possibilities, determined by the row echelon form of the augmented matrix $(A\ \vec{b})$.

1. If $\vec{b}$ is a pivot column, then $A\vec{x} = \vec{b}$ has no solution.

2. If $\vec{b}$ is not a pivot column, then $A\vec{x} = \vec{b}$ has solution. Furthermore,

   - If A has non-pivot columns, then the solution is not unique.
   - If all columns of A are pivot, then the solution is unique.

Theorem 1.4.2. *A system of linear equations $A\vec{x} = \vec{b}$ has solution if and only if the last column of the row echelon form of the augmented matrix $(A\ \vec{b})$ is not pivot.*

For the uniqueness of solution, we know the property is the same as all columns of A being pivot. In particular, the uniqueness is independent of the right side $\vec{b}$. This means that the solution of $A\vec{x} = \vec{b}$ is unique if and only if the solution of the homogeneous equation $A\vec{x} = \vec{0}$ is unique. Since the homogeneous equation always has the trivial solution $\vec{0}$, we get the following.

Theorem 1.4.3. *For a matrix A, the following are equivalent.*

1. *Solution of $A\vec{x} = \vec{b}$ is unique.*
2. *$A\vec{x} = \vec{0}$ has only the trivial solution.*
3. *All columns of A are pivot.*

Example 1.4.1. Consider the system of linear equations

$$\begin{aligned} x + 4y + 7z &= 10, \\ 2x + 5y + 8z &= 11, \\ 3x + 6y + az &= b. \end{aligned}$$

The same row operations in Example 1.3.1 gives

$$\begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & a & b \end{pmatrix} \to \begin{pmatrix} 1 & 4 & 7 & 10 \\ 0 & -3 & -6 & -9 \\ 0 & -6 & a-21 & b-30 \end{pmatrix} \to \begin{pmatrix} 1 & 4 & 7 & 10 \\ 0 & -3 & -6 & -9 \\ 0 & 0 & a-9 & b-12 \end{pmatrix}.$$

The row echelon form depends on the values of a and b. If $a \neq 9$, then the row echelon form is

$$\begin{pmatrix} \bullet & * & * & * \\ 0 & \bullet & * & * \\ 0 & 0 & \bullet & * \end{pmatrix},$$

and the system has unique solution. If $a = 9$, then the result is

$$\begin{pmatrix} \bullet & * & * & * \\ 0 & \bullet & * & * \\ 0 & 0 & 0 & b-12 \end{pmatrix},$$

and we have two possible row echelon forms

$$\begin{pmatrix} \bullet & * & * & * \\ 0 & \bullet & * & * \\ 0 & 0 & 0 & \bullet \end{pmatrix} \text{ if } b \neq 12; \quad \begin{pmatrix} \bullet & * & * & * \\ 0 & \bullet & * & * \\ 0 & 0 & 0 & 0 \end{pmatrix} \text{ if } b = 12.$$

We conclude that the system has no solution when $a = 9, b \neq 12$, and the system has non-unique solution when $a = 9, b = 12$ (we may choose x, y to be non-free variables, expressed in terms of free variable z).

Example 1.4.2. Consider the system of linear equations

$$\begin{aligned} x + 4y + 7z &= b_1, \\ 2x + 5y + 8z &= b_2, \\ 3x + 6y + 9z &= b_3. \end{aligned}$$

By the row operations in Example 1.3.1, we get

$$\begin{pmatrix} 1 & 4 & 7 & b_1 \\ 2 & 5 & 8 & b_2 \\ 3 & 6 & 9 & b_3 \end{pmatrix} \to \begin{pmatrix} 1 & 4 & 7 & 10 \\ 0 & -3 & -6 & b_2 - 2b_1 \\ 0 & -6 & -12 & b_3 - 3b_1 \end{pmatrix} \to \begin{pmatrix} 1 & 4 & 7 & 10 \\ 0 & -3 & -6 & -9 \\ 0 & 0 & 0 & b_1 - 2b_2 + b_3 \end{pmatrix}.$$

The row echelon form is

$$\begin{pmatrix} \bullet & * & * & * \\ 0 & \bullet & * & * \\ 0 & 0 & 0 & \bullet \end{pmatrix} \text{ if } b_1 - 2b_2 + b_3 \neq 0; \quad \begin{pmatrix} \bullet & * & * & * \\ 0 & \bullet & * & * \\ 0 & 0 & 0 & 0 \end{pmatrix} \text{ if } b_1 - 2b_2 + b_3 = 0.$$

We conclude that the system has solution if and only if $b_1 - 2b_2 + b_3 = 0$, and the solution is not unique because z can be the free variable.

Exercise 1.23. From the row echelon form of the augmented matrix, determine the existence and uniqueness of solutions of the corresponding system of linear equations.

$$\begin{pmatrix} \bullet & * & * & * \\ 0 & 0 & \bullet & * \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & \bullet & * & * \\ 0 & 0 & \bullet & * \end{pmatrix} \begin{pmatrix} 0 & \bullet & * & * \\ 0 & 0 & 0 & \bullet \end{pmatrix} \begin{pmatrix} \bullet & * & * \\ 0 & \bullet & * \end{pmatrix} \begin{pmatrix} 0 & 0 & \bullet \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \bullet & * & * \\ 0 & \bullet & * \\ 0 & 0 & \bullet \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

Exercise 1.24. Determine the existence and uniqueness of solutions of the systems.

1. $$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 1, \\ 2x_1 + 3x_2 + x_3 &= 1, \\ 3x_1 + x_2 + 2x_3 &= 1. \end{aligned}$$

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

3. $$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 1, \\ 3x_1 + x_2 + 2x_3 &= 1. \end{aligned}$$

4. $$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 1, \\ 2x_1 + 3x_2 + x_3 &= 1, \\ 3x_1 + x_2 + 2x_3 &= 1, \\ x_1 + x_2 + x_3 &= 0. \end{aligned}$$

5. $$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 1, \\ 2x_1 + 3x_2 + x_3 &= 2, \\ 3x_1 + x_2 + 2x_3 &= 3. \end{aligned}$$

6. $$\begin{aligned} x_1 + 2x_2 + 3x_3 + 4x_3 &= 1, \\ 4x_1 + x_2 + 2x_3 + 3x_3 &= 1, \\ 3x_1 + 4x_2 + x_3 + 2x_3 &= 1. \end{aligned}$$

Exercise 1.25. Determine the existence and uniqueness of solutions of the systems.

1. $$\begin{aligned} x_1 + 4x_2 + 7x_3 &= 10, \\ 2x_1 + 5x_2 + 8x_3 &= b, \\ 3x_1 + 6x_2 + ax_3 &= 12. \end{aligned}$$

2. $$\begin{aligned} x_1 + 7x_3 &= 10, \\ 2x_1 + 8x_3 &= b, \\ 3x_1 + ax_3 &= 12. \end{aligned}$$

3. $$\begin{aligned} x_1 + 4x_2 + 7x_3 &= 10, \\ 3x_1 + 6x_2 + ax_3 &= 12. \end{aligned}$$

4. $$\begin{aligned} x_1 + 4x_2 + 7x_3 &= 10, \\ 2x_1 + 5x_2 + 8x_3 &= b, \\ 3x_1 + 6x_2 + ax_3 &= 12, \\ x_1 + x_2 + x_3 &= 0. \end{aligned}$$

5. $\begin{aligned} x_1 + 4x_2 + 7x_3 &= 10, \\ 2x_1 + 5x_2 + 8x_3 &= 11, \\ 3x_1 + ax_2 + bx_3 &= 12. \end{aligned}$

6. $\begin{aligned} x_1 + 4x_2 + 7x_3 + 10x_4 &= 0, \\ 2x_1 + 5x_2 + 8x_3 + 11x_4 &= 0, \\ 3x_1 + 6x_2 + ax_3 + 12x_4 &= 0. \end{aligned}$

Exercise 1.26. Determine the existence and uniqueness of solutions of the systems.

1. $\begin{aligned} x + 2y + 3z &= b_1, \\ 4x + 5y + 6z &= b_2, \\ 7x + 8y + 9z &= b_3. \end{aligned}$

2. $\begin{aligned} x + 2y &= b_1, \\ 4x + 5y &= b_2, \\ 7x + 8y &= b_3. \end{aligned}$

3. $\begin{aligned} x + 2y + 3z &= b_1, \\ 4x + 5y + 6z &= b_2. \end{aligned}$

4. $\begin{aligned} x + 2y + 3z &= b_1, \\ 4x + 5y + 6z &= b_2, \\ 7x + 8y + 9z &= b_3, \\ 10x + 11y + 12z &= b_4. \end{aligned}$

Example 1.4.3. For

$$A = \begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & a \end{pmatrix},$$

we ask when $A\vec{x} = \vec{b}$ has solution for all the right side $\vec{b}$. By the row operation in Example 1.3.1, we have

$$(A\ \vec{b}) = \begin{pmatrix} 1 & 4 & 7 & 10 & b_1 \\ 2 & 5 & 8 & 11 & b_2 \\ 3 & 6 & 9 & 12 & b_3 \end{pmatrix} \to \begin{pmatrix} 1 & 4 & 7 & 10 & b_1' \\ 0 & -3 & -6 & -9 & b_2' \\ 0 & 0 & 0 & a-12 & b_3' \end{pmatrix}.$$

If $a \neq 12$, then all rows of A are pivot, and solution exists for all $\vec{b}$. If $a = 12$, then there is no solution when $b_3' \neq 0$. It is in fact possible to find suitable b_1, b_2, b_3 such that $b_3' \neq 0$. The reason is that row operations can be reversed by similar row operations. We may postulate $\vec{b}' = (b_1', b_2', b_3') = (0, 0, 1)$ and apply reverse row operations on $\vec{b}'$ to get this suitable $\vec{b}$. Therefore for $a = 12$, $A\vec{x} = \vec{b}$ may not have solution for some right side.

The example shows the following result.

Theorem 1.4.4. *For a matrix A, the following are equivalent.*

1. *$A\vec{x} = \vec{b}$ has solution for all $\vec{b}$.*
2. *All rows of A are pivot.*

Exercise 1.27. For matrix in Exercise 1.14, determine whether $A\vec{x} = \vec{b}$ has solution for all $\vec{b}$.

Combining Theorems 1.4.3 and 1.4.4, we get the following.

Theorem 1.4.5. *For a matrix A, any two of the following imply the third.*

1. *A is a square matrix.*
2. *$A\vec{x} = \vec{b}$ has solution for all $\vec{b}$.*
3. *Solution of $A\vec{x} = \vec{b}$ is unique.*

1.5 Reduced Row Echelon Form

Although the shape of row echelon form cannot be further improved, individual terms can still be improved. In the following 3×4 row echelon form, we may use the $(2,2)$-pivot to kill the $(1,2)$-entry $*$ above the pivot

$$\begin{pmatrix} \bullet & * & * & * \\ 0 & \bullet & * & * \\ 0 & 0 & 0 & 0 \end{pmatrix} \xrightarrow{\text{Row}_1 + c\text{Row}_2} \begin{pmatrix} \bullet & 0 & * & * \\ 0 & \bullet & * & * \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Then we may further divide rows by pivot coefficients to simplify $\bullet$ to 1

$$\begin{pmatrix} \bullet & 0 & * & * \\ 0 & \bullet & * & * \\ 0 & 0 & 0 & 0 \end{pmatrix} \xrightarrow[c_2\text{Row}_2]{c_1\text{Row}_1} \begin{pmatrix} 1 & 0 & * & * \\ 0 & 1 & * & * \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Definition 1.5.1. The *reduced row echelon form* is the simplest matrix one can get by row operations, and the lengths of rows are arranged from longest to shortest.

Reduced row echelon forms are characterised by the property that the pivots are occupied by 1, and the entries above pivots are occupied by 0. The following are all 3×4 reduced row echelon forms

$$\begin{pmatrix} 1 & 0 & 0 & * \\ 0 & 1 & 0 & * \\ 0 & 0 & 1 & * \end{pmatrix} \begin{pmatrix} 1 & 0 & * & 0 \\ 0 & 1 & * & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & * & * \\ 0 & 1 & * & * \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & * & * & * \\ 0 & 0 & 1 & * \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & * & 0 & * \\ 0 & 0 & 1 & * \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 1 & * & * & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & * & * & * \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & * \\ 0 & 0 & 1 & * \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & * & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 1 & * & * \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 & * \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Example 1.5.1. The row echelon form in Example 1.3.1 can be further simplified

$$\begin{pmatrix}1&4&7&10\\2&5&8&11\\3&6&9&12\end{pmatrix}\to\begin{pmatrix}1&4&7&10\\0&-3&-6&-9\\0&0&0&0\end{pmatrix}\xrightarrow{-\frac{1}{3}\text{Row}_2}\begin{pmatrix}1&4&7&10\\0&1&2&3\\0&0&0&0\end{pmatrix}$$
$$\xrightarrow{\text{Row}_1-4\text{Row}_2}\begin{pmatrix}1&0&-1&-2\\0&1&2&3\\0&0&0&0\end{pmatrix}.$$

The reduced echelon form at the end corresponds to linear equations

$$x-z=-2,\quad y+2z=3.$$

By moving terms around (and without calculation), we get the general solution in Example 1.1.2

$$x=-2+z,\quad y=-3+2z,\quad z\text{ arbitrary}.$$

More generally, suppose a system has reduced row echelon form

$$\begin{pmatrix}1&0&c_1&d_1\\0&1&c_2&d_2\\0&0&0&0\end{pmatrix}.$$

Then we get equivalent and simplest equations $x+c_1z=d_1$ and $y+c_2z=d_2$. Moving the z terms to the right, we get the general solution

$$x=d_1-c_1z,\quad y=d_2-c_2z,\quad z\text{ arbitrary}.$$

The example suggests that reduced row echelon form is equivalent to general solution. For example, the reduced row echelon form

$$\begin{pmatrix}1&c_{12}&0&c_{14}&0&d_1\\0&0&1&c_{23}&0&d_2\\0&0&0&0&1&d_3\end{pmatrix}$$

means general solution

$$x_1=d_1-c_{12}x_2-c_{14}x_4,\quad x_3=d_2-c_{34}x_4,\quad x_5=d_3,\quad x_2,x_4\text{ arbitrary}.$$

Since solution of a system of linear equations is not changed by row operations, the coefficients c_i,d_j in the solution above are independent of row operations. Therefore the reduced row echelon form is independent of row operations.

Theorem 1.5.2. *Every matrix has unique reduced row echelon form.*

Exercise 1.28. Find reduced row echelon forms of systems of linear equations in Exercises 1.3, 1.6.

Exercise 1.29. Write down all 2×3 reduced row echelon forms, all 3×2 reduced row echelon forms, and all $n\times2$ reduced row echelon forms are there?

Exercise 1.30. Write down all $n\times3$ reduced row echelon forms.

Exercise 1.31. Given reduced the row echelon form of the augmented matrix, find the general solution.

1. $\begin{pmatrix}1 & c_1 & 0 & d_1\\ 0 & 0 & 1 & d_2\end{pmatrix}$.
2. $\begin{pmatrix}1 & c_1 & 0 & d_1 & 0\\ 0 & 0 & 1 & d_2 & 0\end{pmatrix}$.
3. $\begin{pmatrix}1 & c_1 & c_2 & d_1\\ 0 & 0 & 0 & 0\end{pmatrix}$.
4. $\begin{pmatrix}1 & c_1 & 0 & c_2 & d_1\\ 0 & 0 & 1 & c_3 & d_2\end{pmatrix}$.
5. $\begin{pmatrix}0 & 1 & 0 & c_1 & d_1\\ 0 & 0 & 1 & c_2 & d_2\end{pmatrix}$.
6. $\begin{pmatrix}1 & 0 & 0 & d_1\\ 0 & 1 & 0 & d_2\\ 0 & 0 & 1 & d_3\end{pmatrix}$.
7. $\begin{pmatrix}1 & 0 & c_1 & d_1\\ 0 & 1 & c_2 & d_2\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{pmatrix}$.
8. $\begin{pmatrix}1 & 0 & c_1 & 0 & c_2 & d_1\\ 0 & 1 & c_3 & 0 & c_4 & d_2\\ 0 & 0 & 0 & 1 & c_5 & d_3\\ 0 & 0 & 0 & 0 & 0 & 0\end{pmatrix}$.

Exercise 1.32. Given the general solution, find the reduced row echelon form of the augmented matrix.

1. $x_1 = -x_3$, $x_2 = 1 + x_3$; x_3 arbitrary.
2. $x_1 = -x_3$, $x_2 = 1 + x_3$; x_3, x_4 arbitrary.
3. $x_2 = -x_4$, $x_3 = 1 + x_4$; x_1, x_4 arbitrary.
4. $x_2 = -x_4$, $x_3 = x_4 - x_5$; x_1, x_4, x_5 arbitrary.
5. $x_1 = 1 - x_2 + 2x_5$, $x_3 = 1 + 2x_5$, $x_4 = -3 + x_5$; x_2, x_5 arbitrary.
6. $x_1 = 1 + 2x_2 + 3x_4$, $x_3 = 4 + 5x_4 + 6x_5$; x_2, x_4, x_5 arbitrary.

1.6 Rank

Definition 1.6.1. The *rank* of a matrix A, denoted $\text{rank}A$, is the number of pivots in the row echelon form.

By Theorem 1.5.2, the rank is independent of the choice of row operations. Therefore the concept is well defined.

Since each row and column has at most one pivot, we have

$$\begin{aligned}\text{rank}A &= \text{number of pivots}\\ &= \text{number of pivot rows}\\ &= \text{number of pivot columns}.\end{aligned}$$

For an $m \times n$ matrix A, this implies

$$\text{rank}A \leq \min\{m, n\}.$$

If rank$A$ equals the maximal value $\min\{m, n\}$, then we say A has *full rank*.

Since free variables correspond to non-pivot columns of A, the number of free variables in the general solution of $A\vec{x} = \vec{b}$ is $n - \text{rank}A$. Then we get more precise version of Theorem 1.4.2.

Proposition 1.6.2. *Suppose A is an $m \times n$ matrix. Then $A\vec{x} = \vec{b}$ has solution if and only if* $\text{rank}(A\ \vec{b}) = m$. *Moreover, the general solution has $n - \text{rank}A$ free variables. In particular, the solution is unique if and only if* $\text{rank}A = n$.

Rank is the "essential size" of a system of equations. For example, the following system appears to have 3 equations and 4 variables

$$\begin{aligned} x_1 + 2x_2 + 3x_3 + 4x_4 &= 0, \\ x_1 + 2x_2 + 3x_3 + 4x_4 &= 0, \\ x_1 + 2x_2 + 3x_3 + 4x_4 &= 0. \end{aligned}$$

However, since all equations are the same, the system is essentially only 1 equation $x_1 + 2x_2 + 3x_3 + 4x_4 = 0$. Indeed the rank of the coefficient matrix is 1.

For another example, we may start a system of two equations

$$\begin{aligned} x + 2y + 3z &= 4, \\ 5x + 6y + 7z &= 8. \end{aligned}$$

Then we add two more equations 3Eq_1 and $\text{Eq}_2 - 2\text{Eq}_1$ to the system and get four equations

$$\begin{aligned} x + 2y + 3z &= 4, \\ 5x + 6y + 7z &= 8, \\ 3x + 6y + 9z &= 12, \\ 3x + 2y + z &= 0. \end{aligned}$$

Although new system appears to be larger, we all know that the new system is essentially the same as the old system. The larger size of four equations is only an illusion, and the essential size is two.

The row operations reduces any system to its "core". The size of this core is the essential size of the original system. This is the intuition of the concept of rank.

Example 1.6.1. Without any calculation, we may conclude that the solution of the homogeneous system of 3 equations in 4 variables in Example 1.1.3 cannot be unique. Specifically, we have $\text{rank}A \leq \min\{3, 4\} = 3 < 4$. By Theorem 1.6.2, the solution has $4 - \text{rank}A \geq 1$ free variables.

In general, if A is an $m \times n$ matrix and $m < n$, then the solution of $A\vec{x} = \vec{b}$ is not unique. This is consistent with the intuition that, in order to uniquely determine n variables, we must have at least n equations.

Exercise 1.33. What is a matrix of rank 0?

Exercise 1.34. Write down all 3×4 row echelon forms of rank 2.

Exercise 1.35. Write down all full rank 2×3, 3×2, and 3×3 row echelon forms.

Exercise 1.36. Find the rank.

1. $\begin{pmatrix} 1 \end{pmatrix}$.
2. $\begin{pmatrix} 0 \end{pmatrix}$.
3. $\begin{pmatrix} 1 & 2 & 3 \end{pmatrix}$.
4. $\begin{pmatrix} 0 & 0 & 0 \end{pmatrix}$.
5. $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$.
6. $\begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 4 & 5 & 6 \end{pmatrix}$.
7. $\begin{pmatrix} 1 & 2 \\ 4 & 5 \end{pmatrix}$.
8. $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$.
9. $\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$.
10. $\begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$.
11. $\begin{pmatrix} 1 & 0 & 4 \\ 2 & 0 & 5 \\ 3 & 0 & 6 \end{pmatrix}$.
12. $\begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{pmatrix}$.
13. $\begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{pmatrix}$.
14. $\begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \end{pmatrix}$.
15. $\begin{pmatrix} 1 & 4 & 10 & 7 \\ 2 & 5 & 11 & 8 \\ 3 & 6 & 12 & 9 \end{pmatrix}$.
16. $\begin{pmatrix} 1 & 7 & 10 & 4 \\ 2 & 8 & 11 & 5 \\ 3 & 9 & 12 & 6 \end{pmatrix}$.
17. $\begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}$.
18. $\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}$.

Chapter 2

Euclidean Space

Linear equations can be visualised as lines, planes, etc. The underlying geometric concept is Euclidean space. We establish the basic geometric languages, and then give geometrical interpretations of the results in Chapter 1. The most important concept is subspace of Euclidean space. We define basis and dimension of subspace, and calculate for the four basic subspaces associated to a matrix.

2.1 Geometry of Linear Equation

Using cartesian coordinate system, we may express any point on a plane as a pair of real numbers. Then an equation of two variables gives a curve on the plane. In particular, a linear equation gives a straight line.

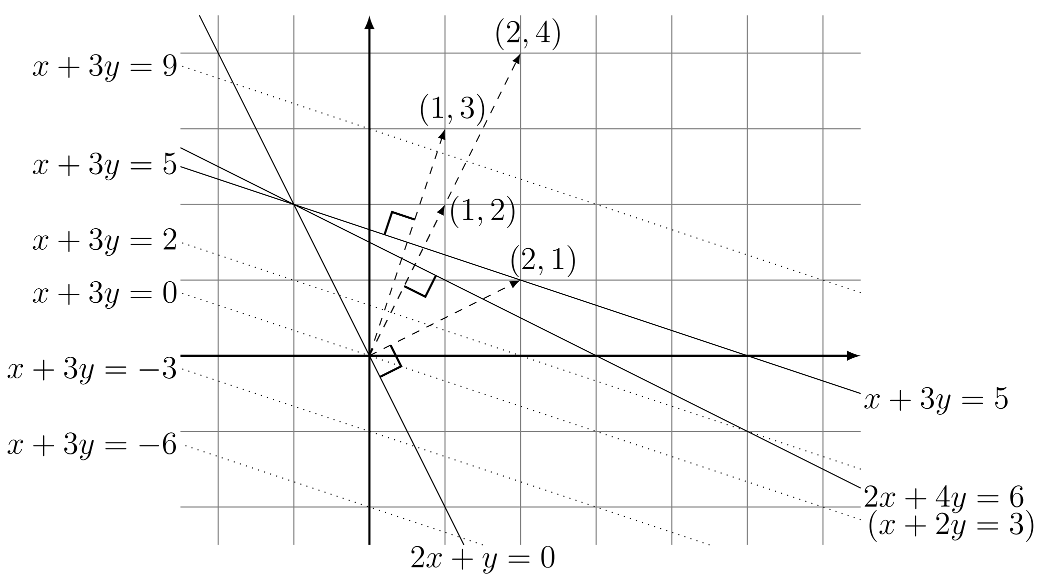

Figure 2.1.1: Linear equations are lines in $\mathbb{R}^2$.

Figure 2.1.1 shows straight lines given by linear equations $x+3y=5$, $2x+4y=6$, $x+2y=3$ (same line as $2x+4y=6$), $2x+y=0$. The first two equations appear in Example 1.1.1. We also indicate by dashed arrows the pairs of coefficients in the equations. We find that the line $ax+by=c$ (a or b is not 0) is orthogonal (or perpendicular) to the direction (a,b). Moreover, fixing a,b and changing c means parallelly moving the line. See the dotted lines, which are $x+3y=c$ for various c.

A system of several equations is then a collection of lines, and the solution means the intersection of all the lines. For example, we can clearly see from the picture that the system in Example 1.1.1 has unique solution $x=-1, y=2$. Moreover, the systems

$$\begin{aligned} x+3y&=2,\\ 2x+4y&=6, \end{aligned} \qquad \begin{aligned} x+3y&=5,\\ 2x+4y&=6,\\ 2x+\ \ y&=0, \end{aligned}$$

have respective unique solutions $(x,y)=(5,-1)$ and $(x,y)=(-1,2)$. We also see that solutions of the system $2x+4y=6$, $x+2y=3$ is the whole line (solution is not unique). Finally, we see that the systems

$$\begin{aligned} x+3y&=5,\\ x+3y&=2, \end{aligned} \qquad \begin{aligned} x+3y&=2,\\ 2x+4y&=6,\\ 2x+\ \ y&=0, \end{aligned}$$

have no solution because the intersections of lines are empty.

Exercise 2.1. Draw the line parallel to $x+3y=5$ and passing through the origin $(0,0)$. Find the intersection with $2x+4y=6$. Interpret your result in terms of a system of linear equations.

Exercise 2.2. Draw the line parallel to $x+3y=5$ and passing through the origin $(0,0)$. Draw the line parallel to $2x+4y=6$ and passing through $(1,0)$. Find the intersection of two lines and interpret your result in terms of a system of linear equations.

Exercise 2.3. Draw the lines orthogonal to $(2,1)$ and passing through $(a,1)$ for $a=-1,0,1,2,3$. For which a do this line and the lines $x+3y=5$, $2x+4y=6$ have common intersection. Interpret your result in terms of a system of linear equations, and redo the problem in terms of row operations and row echelon form.

Exercise 2.4. What are the equations of lines orthogonal to the line $x+3y=5$? In general, what are the equations of lines orthogonal $ax+by=c$?

Exercise 2.5. When is the whole plane solution of $ax+by=c$?

Exercise 2.6. Suppose at least one of a_1,b_1 is not 0, and at least one of a_2,b_2 is not 0. Explain that $a_1x+b_1y=c_1$ and $a_2x+b_2y=c_2$ give the same line (i.e., the solution of system is a line) if and only if there is λ, such that $a_2=\lambda a_1$, $b_2=\lambda b_1$, $c_2=\lambda c_1$.

To visualise solutions of linear equations of three variables, we need to use $\mathbb{R}^3$, the space we live in. An equation $ax + by + cz = d$ (one of a, b, c not 0) is a plane orthogonal to the direction (a, b, c). Moreover, fixing a, b, c and changing d means parallelly moving the plane.

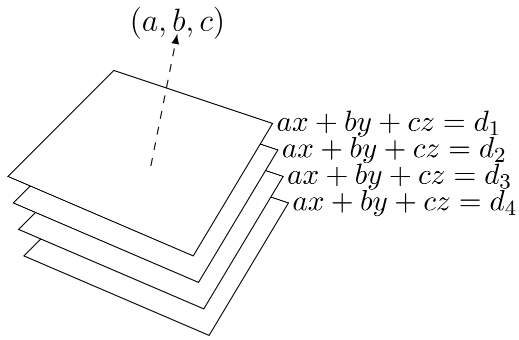

Figure 2.1.2: The planes $ax + by + cz = d$ are parallel for various d.

A system of linear equations of three variables is a collection of planes, and the solution means the intersection of all the planes. The intersection can be empty (no solution), single point (unique solution), a line (one free parameter), or a plane (two free parameters). The solution is a plane if all equations give the same plane (see Exercise 2.9).

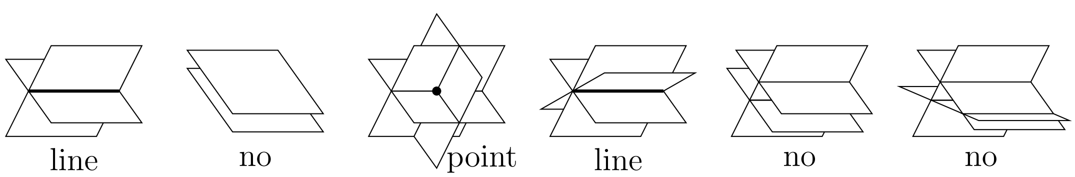

Figure 2.1.3: Solution of system of three variables is intersection of planes.

Exercise 2.7. Find the plane parallel to $x+4y+7z = 10$ and passing through $(1, 1, 1)$. Does the plane and the planes $2x + 5y + 8z = 11$, $3x + 6y + 9z = 12$ have common intersection? Interpret your result in terms of a system of linear equations.

Exercise 2.8. Find the plane orthogonal to $(1, 1, 1)$ and passing through $(1, 1, 1)$. Find the intersection of the plane and three planes $x + 4y + 7z = 10$, $2x + 5y + 8z = 11$, $3x + 6y + 9z = 12$. Interpret your result in terms of a system of linear equations.

Exercise 2.9. Suppose at least one of a_1, b_1, c_1 is not 0, and at least one of a_2, b_2, c_2 is not 0. Explain that $a_1x + b_1y + c_1z = d_1$ and $a_2x + b_2y + c_2z = d_2$ give the same plane if and only if there is λ, such that $a_2 = \lambda a_1$, $b_2 = \lambda b_1$, $c_2 = \lambda c_1$, $d_2 = \lambda d_1$.

2.2 Euclidean Vector

Solution of a system of linear equations is empty, single point, line, plane, or higher versions of infinite flat "things" (for linear equations of more than three variables). The general language for describing this is the following.

Definition 2.2.1. The *Euclidean space* $\mathbb{R}^n$ of dimension n is the collection of all n-tuples of real numbers, called *vectors*

$$\mathbb{R}^n = \{(x_1, x_2, \dots, x_n) \colon x_i \in \mathbb{R}\} = \{ \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} : x_i \in \mathbb{R}\}.$$

An n-tuple is a *Euclidean vector*, and the i-th term x_i is the i-th *coordinate.*

Usually the coordinates are arranged horizontally in a vector. For calculation purposes, however, it is often more convenient to arrange the coordinates vertically.

The space $\mathbb{R}^1$ is a straight line. The space $\mathbb{R}^3$ is the world we are living in. The space $\mathbb{R}^0$ is a single point.

Two Euclidean vectors of the same dimension can be added (called *addition*)

$$(x_1, x_2, \dots, x_n) + (y_1, y_2, \dots, y_n) = (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n).$$

We may also multiply a number to a Euclidean vector (called *scalar multiplication*)

$$a(x_1, x_2, \dots, x_n) = (ax_1, ax_2, \dots, ax_n).$$

The two operations satisfy the usual properties

$$(\vec{x} + \vec{y}) + \vec{z} = \vec{x} + (\vec{y} + \vec{z}), \quad \vec{x} + \vec{y} = \vec{y} + \vec{x}, \quad a(\vec{x} + \vec{y}) = a\vec{x} + a\vec{y}, \quad \dots.$$

By repeatedly using two operations, we get *linear combination*

$$a_1\vec{x}_1 + a_2\vec{x}_2 + \cdots + a_k\vec{x}_k.$$

The origin of a Euclidean space is the *zero vector*

$$\vec{0} = (0, 0, \dots, 0).$$

The vector is characterised by the property

$$\vec{x} + \vec{0} = \vec{x} = \vec{0} + \vec{x}.$$

Exercise 2.10. Verify identities, at least in $\mathbb{R}^2$.

1. $(\vec{x}+\vec{y})+\vec{z}=\vec{x}+(\vec{y}+\vec{z})$.
2. $\vec{x}+\vec{y}=\vec{y}+\vec{x}$.
3. $a(\vec{x}+\vec{y})=a\vec{x}+a\vec{y}$.
4. $(a+b)\vec{x}=a\vec{x}+b\vec{x}$.
5. $(ab)\vec{x}=a(b\vec{x})$.

Exercise 2.11. Explain that a linear combination of linear combinations is still a linear combination. For example

$$a_1(b_1\vec{x}+c_1\vec{y}+d_1\vec{z})+a_2(b_2\vec{x}+c_2\vec{y}+d_2\vec{z})$$

is still a linear combination of $\vec{x},\vec{y},\vec{z}$.

Figure 2.2.1 shows that sums $(1,2)+(3,1)=(4,3)$ and $(3,1)+(2,-2)=(5,-1)$ are geometrically given by parallelograms. Moreover, multiplying scalars $2(3,1)=(6,2)$ and $0.5(3,1)=(1.5,0.5)$ means stretching and shrinking. The negative $-(1,2)=(-1,-2)$ and $-(1.5,0.5)=(-1.5,-0.5)$ means opposite direction.

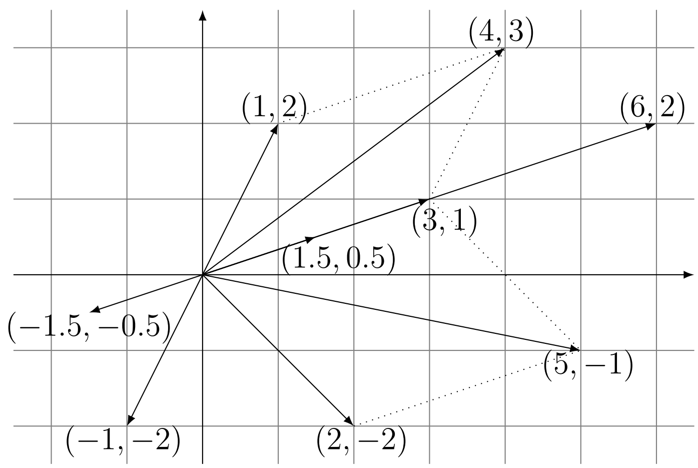

Figure 2.2.1: Euclidean space $\mathbb{R}^2$.

Example 2.2.1. A straight line is given by a point $\vec{x}_0$ on the line and the direction $\vec{v}$ of the line

$$\vec{x}=\vec{x}_0+t\vec{v},\quad t\in\mathbb{R}.$$

For example, the diagonal line of $\mathbb{R}^2$ passes $\vec{x}_0=(0,0)$ and has diagonal direction $\vec{v}=(1,1)$. Therefore the parameterised equation for the diagonal line is

$$\vec{x}=(0,0)+t(1,1)=(t,t),\quad t\in\mathbb{R}.$$

For another example, to get the line passing $(1,2,3)$ and $(4,5,6)$, we may take $\vec{x}_0=(1,2,3)$ and the direction $\vec{v}=(4,5,6)-(1,2,3)=(3,3,3)$. Here we note that the direction of a line is given by the difference between any two points on the line. Then the line is

$$\vec{x}=(1,2,3)+t(3,3,3)=(1+3t,2+3t,3+3t),\quad t\in\mathbb{R}.$$

Example 2.2.2. The general solution of the system of linear equations in Example 1.1.2 is $x = -2 + z$, $y = -3 + 2z$, with z arbitrary. The solution can be rewritten in vector form

$$\vec{x} = (x, y, z) = (-2 + z, 3 - 2z, z) = (-2, 3, 0) + z(1, -2, 1), \quad z \in \mathbb{R}.$$

By Example 2.2.1, this is a line.

The system in Example 1.1.3 has solution $x = z + 2w$, $y = -2z - 3w$, with z, w arbitrary. This can be interpreted as all linear combinations of two vectors

$$\vec{x} = (x, y, z, w) = (z + 2w, -2z - 3w, z, w) = z(1, -2, 1, 0) + w(2, -3, 0, 1).$$

The solution moves in two directions, and gives a plane.

Exercise 2.12. Find the parameterised equation for line.

1. Passing $(0, 0, 0)$ and in direction $(1, 2, 3)$.
2. Passing $(1, 2, 3, 4)$ and $(4, 3, 2, 1)$.
3. Passing $(1, 2)$ and orthogonal to $x + y = 0$.
4. Passing $(1, 2, 3, 4)$ and parallel to $\vec{x} = (t, t, t, t)$, $t \in \mathbb{R}$.

Example 2.2.3. For fixed $\vec{a}$, the movement $\vec{x} \to \vec{a} + \vec{x}$ means shifting by $\vec{a}$. Figure 2.2.2 shows the shifting of $2x + y = 0$ by $\vec{a} = (2, 1)$ is $2x + y = 5$. The picture also shows the shifting of a triangle and a disk by $\vec{a} = (2, 1)$.

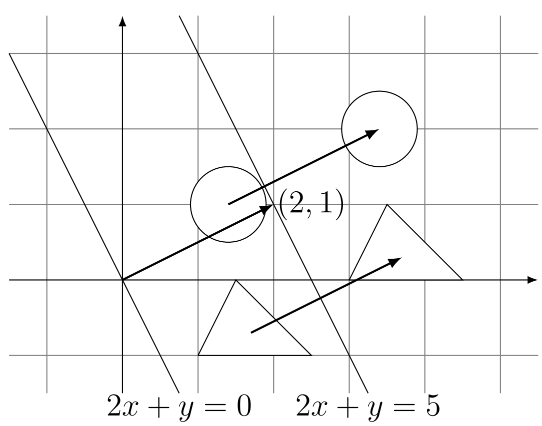

Figure 2.2.2: Shifting by $\vec{a} = (2, 3)$ in $\mathbb{R}^2$.

Figure 2.2.2 shows the effect of scaling. The scaling of $2x + y = 0$ by any $c \neq 0$ is still the line itself. The scaling of $2x + y = 5$ by any $c = -1$ is $2x + y = -5$, and the scaling by $c = 2$ is $2x + y = 10$.

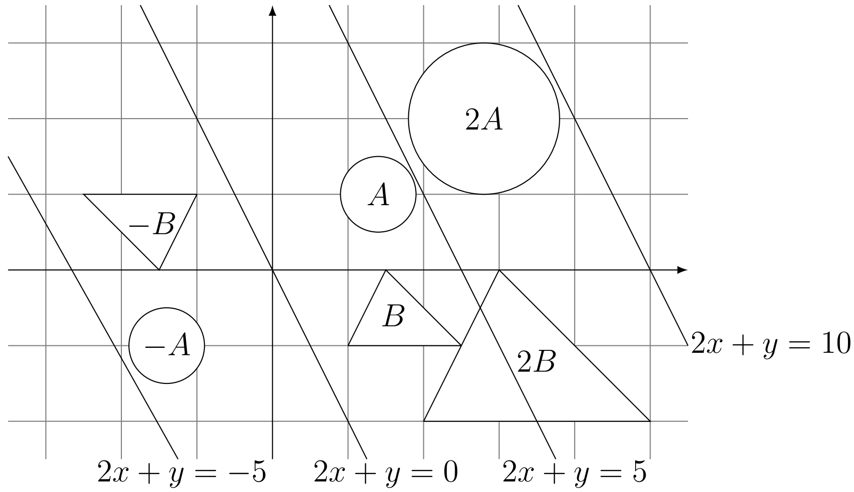

Figure 2.2.3: Scaling by $c = -1$ and $c = 2$ in $\mathbb{R}^2$.

Exercise 2.13. Draw the shifting of the following subsets by $(2, 3)$, and draw the scalings by $c = 0.5, 2, -2$.

1. Single point $(1, 1)$.
2. Line $x + y = 0$.
3. Line $x + y = 1$.
4. Disk of radius 1, center $(0, 0)$.
5. Disk of radius 1, center $(0, 1)$.
6. Triangle with vertices $(1, 1), (-1, 0), (0, -1)$.
7. Triangle with vertices $(1, 1), (0, 1), (1, 0)$.

Example 2.2.4. Any vector in $\mathbb{R}^2$ is a linear combination

$$\begin{pmatrix} x \\ y \end{pmatrix} = x \begin{pmatrix} 1 \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

In general, we have *standard basis vectors* in $\mathbb{R}^n$

$$\vec{e}_1 = (1, 0, \dots, 0), \quad \vec{e}_2 = (0, 1, \dots, 0), \quad \dots, \quad \vec{e}_n = (0, 0, \dots, 1),$$

and

$$(x_1, x_2, \dots, x_n) = x_1\vec{e}_1 + x_2\vec{e}_2 + \cdots + x_n\vec{e}_n.$$

Example 2.2.5. The system of linear equations in Example 1.1.2 can be interpreted as the equality of two Euclidean vectors

$$\begin{pmatrix} x_1 + 4x_2 + 7x_3 \\ 2x_1 + 5x_2 + 8x_3 \\ 3x_1 + 6x_2 + 9x_3 \end{pmatrix} = \begin{pmatrix} 10 \\ 11 \\ 12 \end{pmatrix}.$$

However, the left is a linear combination of three columns of the coefficient matrix

$$\begin{pmatrix} x_1 + 4x_2 + 7x_3 \\ 2x_1 + 5x_2 + 8x_3 \\ 3x_1 + 6x_2 + 9x_3 \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + x_2 \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} + x_3 \begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix}.$$

Therefore the system means expressing a vector $(10, 11, 12)$ as a linear combination of $(1, 2, 3)$, $(4, 5, 6)$, $(7, 8, 9)$.

The example shows that an $m \times n$ matrix corresponds to a collection n *column vectors* in $\mathbb{R}^m$

$$A = (\vec{v}_1 \ \vec{v}_2 \ \cdots \ \vec{v}_n), \quad \vec{v}_i \in \mathbb{R}^m.$$

Then the left side of a system of linear equations with coefficient matrix A is a linear combination of column vectors of A

$$A\vec{x} = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n,$$

and the system becomes expressing the right side as a linear combination

$$x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n = \vec{b}.$$

Exercise 2.14. Interpret systems of linear equations in Exercises 1.3, 1.6, 1.7 as expressing vectors as linear combinations of some other vectors. Determine whether the solution is a point, line, plane, etc.

Exercise 2.15. Can you express the given vector $\vec{b}$ as a linear combination of vectors $\vec{a}_i$? Is the expression unique?

1. $\vec{b} = (1, 2, 3)$, $\vec{a}_1 = (1, 0, 0)$, $\vec{a}_2 = (0, 1, 0)$, $\vec{a}_3 = (0, 0, 1)$.
2. $\vec{b} = (1, 2, 3)$, $\vec{a}_1 = (0, 0, 1)$, $\vec{a}_2 = (0, 1, 0)$, $\vec{a}_3 = (1, 0, 0)$.
3. $\vec{b} = (3, 2, 1)$, $\vec{a}_1 = (1, 0, 0)$, $\vec{a}_2 = (0, 1, 0)$, $\vec{a}_3 = (0, 0, 1)$.
4. $\vec{b} = (1, 2, 3)$, $\vec{a}_1 = (1, 0, 0)$, $\vec{a}_2 = (0, 1, 0)$.
5. $\vec{b} = (1, 2, 3)$, $\vec{a}_1 = (1, 0, 0)$, $\vec{a}_2 = (0, 1, 0)$, $\vec{a}_3 = (0, 0, 1)$, $\vec{a}_4 = (0, 1, 1)$.

Exercise 2.16. Can you express the given vector $\vec{b}$ as a linear combination of vectors $\vec{a}_i$? Is the expression unique?

1. $\vec{b} = (1, 2, 3)$, $\vec{a}_1 = (1, 2, 3)$, $\vec{a}_2 = (4, 5, 6)$, $\vec{a}_3 = (7, 8, 9)$.
2. $\vec{b} = (3, 2, 1)$, $\vec{a}_1 = (1, 2, 3)$, $\vec{a}_2 = (4, 5, 6)$, $\vec{a}_3 = (7, 8, 9)$.
3. $\vec{b} = (1, 2, 3)$, $\vec{a}_1 = (7, 8, 9)$, $\vec{a}_2 = (4, 5, 6)$, $\vec{a}_3 = (1, 2, 3)$.
4. $\vec{b} = (1, 2, 3)$, $\vec{a}_1 = (1, 2, 3)$, $\vec{a}_2 = (4, 5, 6)$.
5. $\vec{b} = (3, 2, 1)$, $\vec{a}_1 = (1, 2, 3)$, $\vec{a}_2 = (4, 5, 6)$.
6. $\vec{b} = (3, 2, 1)$, $\vec{a}_1 = (1, 2, 3)$, $\vec{a}_2 = (4, 5, 6)$, $\vec{a}_3 = (7, 8, 0)$.

Exercise 2.17. Find the exact condition that $\vec{b} = (b_1, b_2, b_3)$ is a linear combination of $\vec{a}_1 = (1, 2, 3)$, $\vec{a}_2 = (4, 5, 6)$, $\vec{a}_3 = (7, 8, 9)$.

2.3 Dot Product

The *dot product* of two Euclidean spaces is

$$\vec{x} \cdot \vec{y} = (x_1, x_2, \dots, x_n) \cdot (y_1, y_2, \dots, y_n) = x_1y_1 + x_2y_2 + \cdots + x_ny_n.$$

The result of the operation is a number instead of vector. This is different from addition and scalar multiplication.

The dot product gives geometry to the Euclidean space. By geometry, we mean sizes such as length, angle, area, etc. For example, the *length* (or *norm*) of a vector is

$$\|\vec{x}\| = \sqrt{\vec{x} \cdot \vec{x}} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

The angle θ between two nonzero vectors $\vec{x}$ and $\vec{y}$ is given by

$$\cos\theta = \frac{\vec{x} \cdot \vec{y}}{\|\vec{x}\|\|\vec{y}\|}.$$

This is justified by the standard fact that the plane vectors $(1, 0)$ and $(\cos\theta, \sin\theta)$ have angle θ

$$\frac{(1, 0) \cdot (\cos\theta, \sin\theta)}{\|(1, 0)\|\|(\cos\theta, \sin\theta)\|} = \frac{1 \cdot \cos\theta + 0 \cdot \sin\theta)}{\sqrt{1^2 + 0^2}\sqrt{\cos^2\theta + \sin^2\theta}} = \cos\theta.$$

The angle further gives the area of the parallelogram spanned by two vectors (base $\|\vec{x}\|$ and height $\|\vec{y}\|\sin\theta$)

$$\begin{aligned}\text{Area} &= \|\vec{x}\|\|\vec{y}\|\sin\theta = \|\vec{x}\|\|\vec{y}\|\sqrt{1 - \cos^2\theta} = \|\vec{x}\|\|\vec{y}\|\sqrt{1 - \frac{(\vec{x} \cdot \vec{y})^2}{\|\vec{x}\|^2\|\vec{y}\|^2}} \\ &= \sqrt{\|\vec{x}\|^2\|\vec{y}\|^2 - (\vec{x} \cdot \vec{y})^2} = \sqrt{(\vec{x} \cdot \vec{x})(\vec{y} \cdot \vec{y}) - (\vec{x} \cdot \vec{y})^2}.\end{aligned}$$

Example 2.3.1. Let $\vec{x} = (1, 2)$, $\vec{y} = (3, 4)$. Then

$$\vec{x} \cdot \vec{y} = 1 \cdot 3 + 2 \cdot 4 = 11, \quad \|\vec{x}\| = \sqrt{1^2 + 2^2} = \sqrt{5}, \quad \|\vec{y}\| = \sqrt{3^2 + 4^2} = 5.$$

Moreover, the angle θ between the two vectors is

$$\cos\theta = \frac{11}{\sqrt{5} \cdot 5}, \quad \theta \approx 10.305^\circ.$$

The area of the triangle spanned by the two vectors is

$$\frac{1}{2}\sqrt{\sqrt{5}^2 \cdot 5^2 - 11^2} = 1.$$

Example 2.3.2. Let $\vec{x} = (1, 2, 3)$, $\vec{y} = (4, 5, 6)$, $\vec{z} = (1, -2, 1)$. Then

$$\vec{x} \cdot \vec{y} = 32, \quad \vec{x} \cdot \vec{z} = 0, \quad \vec{y} \cdot \vec{z} = 0.$$

The angle between $\vec{x}$ and $\vec{z}$ is

$$\cos\theta = \frac{0}{\|\vec{x}\|\|\vec{z}\|} = 0, \quad \theta = 90^\circ.$$

For the area of the triangle with $\vec{x}, \vec{y}, \vec{z}$ as vertices, we take $\vec{v} = \vec{y} - \vec{x} = (3, 3, 3)$, $\vec{w} = \vec{z} - \vec{x} = (0, -4, -2)$. The area of the triangle is half of the area of the parallelogram spanned by $\vec{v}$ and $\vec{w}$

$$\frac{1}{2}\sqrt{(\vec{v} \cdot \vec{v})(\vec{w} \cdot \vec{w}) - (\vec{v} \cdot \vec{w})^2} = \frac{1}{2}\sqrt{27 \cdot 20 - (-18)^2} = 3\sqrt{6}.$$

Exercise 2.18. Calculate the area of the triangle in Example 2.3.3 in another way, for example, by using $\vec{x} - \vec{y}$ and $\vec{z} - \vec{y}$.

Exercise 2.19. Show that $\vec{x} \cdot \vec{x} \geq 0$, and the equality holds if and only if $\vec{x} = \vec{0}$.

The dot product is clearly *symmetric*

$$\vec{x} \cdot \vec{y} = \vec{y} \cdot \vec{x}.$$

It is also *linear* in the first vector

$$(\vec{x} + \vec{y}) \cdot \vec{z} = \vec{x} \cdot \vec{z} + \vec{y} \cdot \vec{z}, \quad (c\vec{x}) \cdot \vec{y} = c(\vec{x} \cdot \vec{y}).$$

This linear property is the same as

$$(a\vec{x} + b\vec{y}) \cdot \vec{z} = a\vec{x} \cdot \vec{z} + b\vec{y} \cdot \vec{z}.$$

The following verifies this combined property for $n = 3$

$$\begin{aligned}
&(a(x_1, x_2, x_3) + b(y_1, y_2, y_3)) \cdot (z_1, z_2, z_3) \\
&= (ax_1 + by_1, ax_2 + by_2, ax_3 + by_3) \cdot (z_1, z_2, z_3) \\
&= ax_1z_1 + by_1z_1 + ax_2z_2 + by_2z_2 + ax_3z_3 + by_3z_3 \\
&= ax_1z_1 + ax_2z_2 + ax_3z_3 + by_1z_1 + by_2z_2 + by_3z_3 \\
&= a(x_1, x_2, x_3) \cdot (z_1, z_2, z_3) + b(y_1, y_2, y_3) \cdot (z_1, z_2, z_3).
\end{aligned}$$

By symmetry, the linear property for the first vector implies the linear property for the second vector

$$\vec{x} \cdot (a\vec{y} + b\vec{z}) = a\vec{x} \cdot \vec{y} + b\vec{x} \cdot \vec{z}, \quad \vec{x} \cdot (c\vec{y}) = c(\vec{x} \cdot \vec{y}).$$

Again this can be combined to give

$$\vec{x} \cdot (a\vec{y} + b\vec{z}) = a\vec{x} \cdot \vec{y} + b\vec{x} \cdot \vec{z}.$$

Due to the linear property in both vectors, we say that the dot product is *bilinear*.

Example 2.3.3. The length of vector is defined by the dot product. Conversely, the dot product can be written in terms of length. We have

$$\begin{aligned}
\|\vec{x} + \vec{y}\|^2 &= (\vec{x} + \vec{y}) \cdot (\vec{x} + \vec{y}) \\
&= \vec{x} \cdot \vec{x} + \vec{x} \cdot \vec{y} + \vec{y} \cdot \vec{x} + \vec{y} \cdot \vec{y} = \|\vec{x}\|^2 + \|\vec{y}\|^2 + 2\vec{x} \cdot \vec{y}.
\end{aligned}$$

Therefore

$$\vec{x} \cdot \vec{y} = \frac{1}{2}(\|\vec{x} + \vec{y}\|^2 - \|\vec{x}\|^2 - \|\vec{y}\|^2).$$

This is called the *polarization identity*.

Exercise 2.20. Prove another polarization identity $\vec{x} \cdot \vec{y} = \frac{1}{4}(\|\vec{x} + \vec{y}\|^2 - \|\vec{x} - \vec{y}\|^2)$.

Next we explain the relation between the (bi-)linear property of dot product and linear equation.

A linear equation such as $x+2y+3z = 4$ can be expressed as $(1, 2, 3)\cdot(x, y, z) = 4$. In general, a linear equation is

$$\vec{a} \cdot \vec{x} = a_1x_1 + a_2x_2 + \cdots + a_nx_n = b.$$

Since dot product is linear in the second vector, the *linear function* $l(\vec{x}) = \vec{a} \cdot \vec{x}$ has the following property

$$l(\vec{x} + \vec{y}) = l(\vec{x}) + l(\vec{y}), \quad l(c\vec{x}) = cl(\vec{x}), \tag{2.3.1}$$

or

$$l(a\vec{x} + b\vec{y}) = a\, l(\vec{x}) + b\, l(\vec{y}),$$

For example, for

$$\vec{a}_1 = (1, 4, 7), \quad \vec{a}_2 = (2, 5, 8), \quad \vec{a}_3 = (3, 6, 9),$$

the three equations in Example 1.1.2 can be regarded as $\vec{a}_1 \cdot \vec{x} = 10$, $\vec{a}_2 \cdot \vec{x} = 11$, $\vec{a}_3 \cdot \vec{x} = 12$. The system means finding vectors such that the dot products with some given vector have the designated values. The augmented matrix is

$$\begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{pmatrix} = \begin{pmatrix} \vec{a}_1 & 10 \\ \vec{a}_2 & 11 \\ \vec{a}_3 & 12 \end{pmatrix},$$

and $\vec{a}_1, \vec{a}_2, \vec{a}_3$ are the row vectors of the coefficient matrix.

In a system of linear equations $A\vec{x} = \vec{b}$, each row of $A\vec{x}$ satisfies (2.3.1). This means

$$A(\vec{x} + \vec{y}) = A\vec{x} + A\vec{y}, \quad A(c\vec{x}) = cA\vec{x}, \tag{2.3.2}$$

or

$$A(a\vec{x} + b\vec{y}) = aA\vec{x} + bA\vec{y}.$$

Exercise 2.21. Calculate $A\vec{v}, A\vec{w}, A(\vec{v} + \vec{w}), A(2\vec{v} - \vec{w}), 2A\vec{v} - A\vec{w}$.

1. $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$, $\vec{v} = (1, 0)$, $\vec{w} = (0, 1)$.

2. $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$, $\vec{v} = (1, -1)$, $\vec{w} = (2, -1)$.

3. $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$, $\vec{v} = (1, 1, 1)$, $\vec{w} = (1, -2, 1)$.

4. $A = \begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{pmatrix}$, $\vec{v} = (1, 1, 1)$, $\vec{w} = (1, -2, 1)$.

Exercise 2.22. Explain that, if $\vec{v}$ and $\vec{w}$ are solutions of homogeneous system of linear equations $A\vec{x} = \vec{0}$, then $a\vec{v} + b\vec{w}$ is also a solution.

Two direction are *orthogonal* if the angle between them is 90°. By the formula of angle and $\cos 90^\circ = 0$, we get the following definition.

Definition 2.3.1. Two vectors $\vec{x}$ and $\vec{y}$ are *orthogonal*, and denoted $\vec{x} \perp \vec{y}$, if $\vec{x} \cdot \vec{y} = 0$. A collection of vectors $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\}$ is an *orthogonal set* if $\vec{v}_i \perp \vec{v}_j$ for any $i \neq j$.

A *homogeneous linear equation* $l(\vec{x}) = \vec{a} \cdot \vec{x} = 0$ means all vectors orthogonal to $\vec{a}$. For a general linear equation $l(\vec{x}) = \vec{a} \cdot \vec{x} = b$, if both $\vec{x}$ and $\vec{y}$ are solutions, then

$$l(\vec{x} - \vec{y}) = \vec{a} \cdot (\vec{x} - \vec{y}) = \vec{a} \cdot \vec{x} - \vec{a} \cdot \vec{y} = b - b = 0.$$

This means $\vec{x} - \vec{y}$ is orthogonal to the "coefficient vector" $\vec{a}$, and explains that all solutions form a hyperplane orthogonal to $\vec{a}$.

More generally, A *homogeneous system of linear equations* $A\vec{x} = \vec{0}$ means all vectors orthogonal to all the rows of A. This fact will be expressed as $\text{Nul}A = (\text{Row}A)^{\perp}$.

The most famous theorem about orthogonal vectors is the following theorem. The theorem follows from the calculation in Example 2.3.3.

Theorem 2.3.2 (Pythagorean Theorem)**.** *If $\vec{x} \perp \vec{y}$, then $\|\vec{x} + \vec{y}\|^2 = \|\vec{x}\|^2 + \|\vec{y}\|^2$.*

Exercise 2.23. Show that $\vec{x} \perp \vec{x}$ if and only if $\vec{x} = \vec{0}$.

Exercise 2.24. Show that $\vec{x} \perp \vec{y}$ and $\vec{x} \perp \vec{z}$ implies $\vec{x} \perp (a\vec{y} + b\vec{z})$.

Exercise 2.25. Show that $\|\vec{x} + \vec{y}\|^2 = \|\vec{x}\|^2 + \|\vec{y}\|^2$ implies $\vec{x} \perp \vec{y}$.

Exercise 2.26. If $\vec{x}, \vec{y}, \vec{z}$ form an orthogonal set, show that $\|\vec{x}+\vec{y}+\vec{z}\|^2 = \|\vec{x}\|^2+\|\vec{y}\|^2+\|\vec{z}\|^2$.

Finally, since $|\cos\theta| \le 1$, the definition of angle must be justified by the following result.

Theorem 2.3.3 (Cauchy-Schwartz Inequality)**.** *For any $\vec{x}$ and $\vec{y}$, we have*

$$|\vec{x} \cdot \vec{y}| \le \|\vec{x}\|\|\vec{y}\|.$$

For $n = 2$, the following verifies the Cauchy-Schwartz Inequality.

$$\begin{aligned}
\|\vec{x}\|^2\|\vec{y}\|^2 - (\vec{x} \cdot \vec{y})^2 &= (x_1^2 + x_2^2)(y_1^2 + y_2^2) - (x_1y_1 + x_2y_2)^2 \\
&= (x_1^2y_1^2 + x_1^2y_2^2 + x_2^2y_1^2 + x_2^2y_2^2) - (x_1^2y_1^2 + x_2^2y_2^2 + 2x_1y_1x_2y_2) \\
&= x_1^2y_2^2 + x_2^2y_1^2 - 2x_1y_1x_2y_2 \\
&= (x_1y_2 - x_2y_1)^2 \ge 0.
\end{aligned}$$

Exercise 2.27. Verify Cauchy-Schwartz inequality for $\mathbb{R}^3$.

Exercise 2.28. What is the condition for the Cauchy-Schwartz inequality to become equality?

2.4 Subspace

Geometrically, the solution of a system of linear equations is empty, one point, or an infinite flat thing. We note that the solution can be shifted (see Example 2.2.3) to $H \subset \mathbb{R}^n$ that contains the origin. Then

$$\{\text{solutions of } A\vec{x} = \vec{b}\} = \vec{x}_0 + H = \{\vec{x}_0 + \vec{v} \colon \vec{v} \in H\}.$$

Here H is empty, origin, or an infinite flat thing containing the origin. The property is precisely characterised by the the following.

Definition 2.4.1. A subset $H \subset \mathbb{R}^n$ is a *subspace* if

$$\vec{u}, \vec{v} \in H,\ c \in \mathbb{R} \implies \vec{u} + \vec{v} \in H,\ c\vec{v} \in H.$$

We remark that, by taking $c = 0$, a subspace contains the origin $\vec{0}$.

Exercise 2.29. Determine whether the following is a subspace of the Euclidean space.

1. $H = \{(x, y) \colon 2x + 3y = 0\}$.
2. $H = \{(x, y) \colon 2x + 3y = 1\}$.
3. $H = \{(2x, 3y) \colon 2x + 3y = 0\}$.
4. $H = \{(2x, 3y) \colon 2x + 3y = 1\}$.
5. $H = \{(x + y, x - y) \colon 2x + 3y = 0\}$.
6. $H = \{(x + y, x - y) \colon 2x + 3y = 1\}$.
7. $H = \{(x, y) \colon x^2 + y^2 \le 1\}$.
8. $H = \{(x, y) \colon x^2 + y^2 = 1\}$.
9. $H = \{(2x, 3y) \colon x^2 + y^2 = 0\}$.
10. $H = \{(x + 2, y) \colon 2x + 3y = 4\}$.

From geometric picture, we know the solution of $A\vec{x} = \vec{b}$ is $\vec{x}_0 + H$ in general. Now we see this fact again from the viewpoint of algebraic calculation.

Suppose a system $A\vec{x} = \vec{b}$ (of 3 equations in 5 variables) has reduced row echelon form

$$(A\ \vec{b}) \to \begin{pmatrix} 1 & c_{12} & 0 & c_{14} & 0 & d_1 \\ 0 & 0 & 1 & c_{24} & 0 & d_2 \\ 0 & 0 & 0 & 0 & 1 & d_3 \end{pmatrix}.$$

Then we get the general solution

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} d_1 - c_{12}x_2 - c_{14}x_4 \\ x_2 \\ d_2 - c_{24}x_4 \\ x_4 \\ d_3 \end{pmatrix} = \begin{pmatrix} d_1 \\ 0 \\ d_2 \\ 0 \\ d_3 \end{pmatrix} + x_2 \begin{pmatrix} -c_{12} \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} -c_{14} \\ 0 \\ -c_{24} \\ 1 \\ 0 \end{pmatrix}.$$

The collection of all solutions is $\vec{x}_0 + H$, which is the shifting of H by $\vec{x}_0$

$$\vec{x}_0 = \begin{pmatrix} d_1 \\ 0 \\ d_2 \\ 0 \\ d_3 \end{pmatrix}, \quad H = \{y_1 \begin{pmatrix} -c_{12} \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + y_2 \begin{pmatrix} -c_{14} \\ 0 \\ -c_{24} \\ 1 \\ 0 \end{pmatrix} : y_1, y_2 \in \mathbb{R}\}.$$

The subspace H is a plane in $\mathbb{R}^5$ passing through the origin.

In general, the solution of $A\vec{x} = \vec{b}$ is

$$\vec{x} = \vec{x}_0 + y_1\vec{v}_1 + y_2\vec{v}_2 + \cdots + y_k\vec{v}_k, \tag{2.4.1}$$

where $y_1, y_2, \dots, y_k$ are the free variables, corresponding to the non-pivot columns of A. By taking all $y_i = 0$, we find that $\vec{x}_0$ is *one special solution*. The collection of all solutions is $\vec{x}_0 + H$, with

$$H = \{y_1\vec{v}_1 + y_2\vec{v}_2 + \cdots + y_k\vec{v}_k \colon y_i \in \mathbb{R}\}. \tag{2.4.2}$$

Since a linear combination of linear combinations is still a linear combination (see Exercise 2.11), this H is a subspace.

Definition 2.4.2. The *span* of a collection of vectors $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\}$ in $\mathbb{R}^n$ is the subspace

$$\begin{aligned} \text{Span}\alpha &= \text{Span}\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\} \\ &= \{x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_k\vec{v}_k \colon x_i \in \mathbb{R}\} = \mathbb{R}\vec{v}_1 + \mathbb{R}\vec{v}_2 + \cdots + \mathbb{R}\vec{v}_k. \end{aligned}$$

If $\vec{v}_i$ are the columns of an $n \times k$-matrix $A = (\vec{v}_1 \ \vec{v}_2 \ \cdots \ \vec{v}_k)$, then $x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_k\vec{v}_k = A\vec{x}$, and the span is the *column space* of the matrix

$$\begin{aligned} \text{Col}A &= \text{Span}\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\} \\ &= \{A\vec{x} \colon \vec{x} \in \mathbb{R}^n\} \\ &= \{\vec{b} \in \mathbb{R}^n \colon A\vec{x} = \vec{b} \text{ has solution}\}. \end{aligned}$$

Proposition 2.4.3. *A system $A\vec{x} = \vec{b}$ has solution if and only if $\vec{b} \in \text{Col}A$.*

Example 2.4.1. The span of $(1,2,3), (4,5,6), (7,8,9), (10,11,a)$ is the column space of the matrix

$$A = \begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & a \end{pmatrix}.$$

The span consists of all $\vec{b} \in \mathbb{R}^3$ such that the system $A\vec{x} = \vec{b}$ has solution. Using the row operations in (1.3.1), we get

$$(A, \vec{b}) = \begin{pmatrix} 1 & 4 & 7 & 10 & b_1 \\ 2 & 5 & 8 & 11 & b_2 \\ 3 & 6 & 9 & a & b_3 \end{pmatrix} \to \begin{pmatrix} 1 & 4 & 7 & 10 & b_1 \\ 0 & -3 & -6 & -9 & b_2 - 2b_1 \\ 0 & 0 & 0 & a - 12 & b_3 - 2b_2 + b_1 \end{pmatrix}.$$

By Theorem 1.4.2, if $a \neq 12$, then $A\vec{x} = \vec{b}$ always has solution (for all right side $\vec{b}$. This means $\text{Col}A = \mathbb{R}^3$, i.e., the four vectors span the whole Euclidean space. If $a = 12$, however, then $A\vec{x} = \vec{b}$ has solution if and only if $b_3 - 2b_2 + b_1 = 0$. Therefore

$$\text{Col} \begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{pmatrix} = \{(b_1, b_2, b_3) \colon b_3 - 2b_2 + b_1 = 0\}.$$

For example, $(1, 2, 1)$ and $(13, 14, 15)$ are in the span, and $(1, 0, 0)$ is not in the span.

Exercise 2.30. Determine whether $\text{Col}A$ is the whole Euclidean space, and whether the vector $\vec{v}$ is in $\text{Col}A$.

1. $A = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$, $\vec{v} = (a, 1)$.
2. $A = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$, $\vec{v} = (1, a)$.
3. $A = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}$, $\vec{v} = (a, 1)$.
4. $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}$, $\vec{v} = (1, a, b)$.
5. $A = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$, $\vec{v} = (1, a, b)$.
6. $A = \begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{pmatrix}$, $\vec{v} = (1, a, b)$.

Another viewpoint of H is that, since H is a shifting of the solution of $A\vec{x} = \vec{b}$, H is the solution of $A\vec{x} = \vec{b}'$ for a different right side $\vec{b}'$. On the other hand, since H contains $\vec{0}$, the origin $\vec{0}$ is a solution of $A\vec{x} = \vec{b}'$. Substituting $\vec{x} = \vec{0}$ into $A\vec{x} = \vec{b}'$, we get $\vec{b}' = \vec{0}$. We conclude that H is actually the solution of a *homogeneous system* $A\vec{x} = \vec{0}$.

Definition 2.4.4. The *null space* of an $m \times n$ matrix A is

$$\text{Nul}A = \{\text{solutions of } A\vec{x} = \vec{0}\} = \{\vec{v} \in \mathbb{R}^n \colon A\vec{v} = \vec{0}\} \subset \mathbb{R}^n.$$

If the general solution of $A\vec{x} = \vec{b}$ is given by (2.4.1), then the null space is given by the span (2.4.2).

The following repeats Theorem 1.4.3, and can be compared with Proposition 2.4.3.

Proposition 2.4.5. *The solution of* $A\vec{x}=\vec{b}$ *is unique if and only if* $\text{Nul}A=\{\vec{0}\}$.

Example 2.4.2. For $x+3y=5$, we have $A=(1\ 3)$, $\vec{x}_0=(5,0)$ and

$$\text{Nul}A=\{(x,y)\colon x+3y=0\}=\{y(-3,1)\colon y\in\mathbb{R}\}=\mathbb{R}(-3,1)=\text{Span}\{(-3,1)\}.$$

For $x+4y+7z=10$, we have $A=(1\ 4\ 7)$, $\vec{x}_0=(10,0,0)$ and

$$\begin{aligned}\text{Nul}A&=\{(x,y)\colon x+4y+7z=0\}=\{y(-4,1,0)+z(-7,0,1)\colon y,z\in\mathbb{R}\}\\&=\mathbb{R}(-4,1,0)+\mathbb{R}(-7,0,1)=\text{Span}\{(-4,1,0),\ (-7,0,1)\}.\end{aligned}$$

For the system in Example 1.1.2, we have

$$A=\begin{pmatrix}1&4&7\\2&5&8\\3&6&9\end{pmatrix},\quad \vec{x}_0=(-2,3,0),\quad \text{Nul}A=\{y(1,-2,1)\colon y\in\mathbb{R}\}=\mathbb{R}(1,-2,1).$$

Exercise 2.31. For the systems in Exercise 1.24 that have solutions, express the general solution as $\vec{x}_0+H$, and express H as a span.

Exercise 2.32. For the reduced row echelon forms in Exercise 1.31 of augmented matrix, express the general solution as $\vec{x}_0+H$, and express H as a span.

Exercise 2.33. For the general solution in Exercise 1.32, express the general solution as $\vec{x}_0+H$, and express H as a span.

Exercise 2.34. Determine whether the vector $\vec{v}$ is in $\text{Nul}A$.

1. $A=\begin{pmatrix}1&3\\2&4\end{pmatrix}$, $\vec{v}=(3,-1)$.
2. $A=\begin{pmatrix}1&3\\2&4\end{pmatrix}$, $\vec{v}=(0,0)$.
3. $A=\begin{pmatrix}1&3\end{pmatrix}$, $\vec{v}=(3,-1)$.
4. $A=\begin{pmatrix}0&1&3\\0&2&4\end{pmatrix}$, $\vec{v}=(3,-1,0)$.
5. $A=\begin{pmatrix}0&1&3\\0&2&4\end{pmatrix}$, $\vec{v}=(3,0,0)$.
6. $A=\begin{pmatrix}1&3&5\\2&4&6\end{pmatrix}$, $\vec{v}=(1,-2,1)$.

Exercise 2.35. Determine the condition for vector $\vec{v}$ to be in $\text{Nul}A$.

1. $A=\begin{pmatrix}1&3&5\\2&4&6\end{pmatrix}$, $\vec{v}=(1,a,1)$.
2. $A=\begin{pmatrix}1&2&3\\4&5&6\end{pmatrix}$, $\vec{v}=(1,a,1)$.
3. $A=\begin{pmatrix}1&2&3\\4&5&6\end{pmatrix}$, $\vec{v}=(1,2a,-a)$.
4. $A=\begin{pmatrix}1&2&3\\4&5&6\end{pmatrix}$, $\vec{v}=(-a,2a,-a)$.

Exercise 2.36. Rephrase the conclusion of Exercise 2.16 as whether a vector is in some span, in some column space, and whether some null space is trivial.

2.5 Dimension of Subspace

We have the intuitive sense that a point, line, plane, have dimensions 0, 1, 2. We regard a line having 1 direction, given by 1 vector, and a plane having 2 directions, given by 2 vectors. In general, we express a subspace as $H = \text{Span}\alpha$, for a collection of vectors $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\}$. The collection provides all the directions of H. However, we cannot simply define $\dim H = k$ (the number of vectors in α) because some directions are superfluous. For example, our usual ground plane can be covered by the horizontal, vertical, and diagonal directions. However, the dimension of the ground plane is not 3, because horizontal and vertical directions already cover the whole plane, and the diagonal is in fact not needed. What we need is the "essential size" of α, similar to the rank of a matrix or a system of linear equations in Section 1.5.

Definition 2.5.1. A collection of vectors $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\}$ in a subspace H of $\mathbb{R}^n$ is a *basis* of H if it is a minimal spanning set. The *dimension* $\dim H$ is the number of vectors in a basis.

The collection $\alpha = \{\text{horizontal}, \text{vertical}, \text{diagonal}\}$ spans the ground plane. However, α is not minimal, because $\beta = \{\text{horizontal}, \text{vertical}\}$ is a strictly smaller spanning set. On the other hand, β is minimal because any smaller collection has at most one vector, and cannot span the whole plane. Therefore the dimension of the ground plane is 2.

Example 2.5.1. In Example 2.2.4, we see that $H = \mathbb{R}^3$ is spanned by

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

If we delete $\vec{e}_3$, then we get

$$\text{Span}\{\vec{e}_1, \vec{e}_2\} = \{x_1(1,0,0) + x_2(0,1,0) = (x_1, x_2, 0) \colon x_1, x_2 \in \mathbb{R}\} \neq \mathbb{R}^3.$$

Similarly, deleting $\vec{e}_1$ or $\vec{e}_2$ also does not span $\mathbb{R}^3$. Therefore $\vec{e}_1, \vec{e}_2, \vec{e}_3$ is a minimal spanning set of $\mathbb{R}^3$, i.e., a basis.

In general, the vectors $\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n$ in Example 2.2.4 form a basis of $\mathbb{R}^n$, called the *standard basis*.

Example 2.5.2. Let H be the span of $\vec{v}_1 = (1,2,3), \vec{v}_2 = (4,5,6), \vec{v}_3 = (7,8,9), \vec{v}_4 = (10,11,12)$. To find a basis, we consider the row operation in Example 1.3.1

$$(\vec{v}_1\ \vec{v}_2\ \vec{v}_3\ \vec{v}_4) = \begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{pmatrix} \to \begin{pmatrix} 1 & 4 & 7 & 10 \\ 0 & -3 & -6 & -9 \\ 0 & -6 & -12 & -18 \end{pmatrix} \to \begin{pmatrix} 1 & 4 & 7 & 10 \\ 0 & -3 & -6 & -9 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

We regard $(\vec{v}_1\ \vec{v}_2\ \vec{v}_3)$ as the augmented matrix of the system $y_1\vec{v}_1 + y_2\vec{v}_2 = \vec{v}_3$. By restricting the row operations to the first three columns, we get the row echelon form of the augmented matrix. Then we see that the system has solution, which means that $\vec{v}_3$ is a linear combination of $\vec{v}_1, \vec{v}_2$.

Similarly, we may regard the first, second and fourth columns as the augmented matrix of a system and conclude that $\vec{v}_4$ is also a linear combination of $\vec{v}_1, \vec{v}_2$.

Now we have

$$\begin{aligned} x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{v}_3 + x_4\vec{v}_4 &= x_1\vec{v}_1 + x_2\vec{v}_2 + x_3(y_1\vec{v}_1 + y_2\vec{v}_2) + x_4(z_1\vec{v}_1 + z_2\vec{v}_2) \\ &= (x_1 + x_3y_1 + x_4z_1)\vec{v}_1 + (x_2 + x_3y_2 + x_4z_2)\vec{v}_2. \end{aligned}$$

The left side is all the vectors in H, and the equality means that H is already spanned by $\beta = \{\vec{v}_1, \vec{v}_2\}$, or $H = \mathbb{R}\vec{v}_1 + \mathbb{R}\vec{v}_2$. In particular, α is not a minimal spanning set.

It remains to argue that β is minimal. If we delete $\vec{v}_2$ from β, then we get $\vec{v}_1$ only. By considering the row operation for the system $x_1\vec{v}_1 = \vec{v}_2$ (the first and second columns above), we find that $\vec{v}_2 \in H$ is not in the span $\mathbb{R}\vec{v}_1$. Therefore $\mathbb{R}\vec{v}_1 \neq H$. Similarly, we have $\mathbb{R}\vec{v}_2 \neq H$. Therefore any collection smaller than β cannot span H. We conclude that the spanning set β of H is minimal, and $\dim H = 2$.

In general, the span of $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\}$ is the columns space Col$A$ of the matrix $A = (\vec{v}_1\ \vec{v}_2\ \cdots\ \vec{v}_k)$. The example shows that we may find a basis (minimal spanning set) of the column space by the following steps.

1. Row operation on A to identify pivots.

2. Pivot columns of A form a basis of Col$A$.

The number of pivots is the rank, and we get

$$\dim \mathrm{Col}A = \mathrm{rank}A.$$

We also call this number the *rank* of set α

$$\mathrm{rank}\alpha = \dim \mathrm{Span}\alpha.$$

Exercise 2.37. Use row operations to find a minimal spanning set.

1. $(1, 2), (2, 1)$.

2. $(1, 2), (-1, -2)$.

3. $(1, 1, 0)$, $(1, 0, 1)$, $(0, 1, 1)$.

4. $(1, 1, -1)$, $(1, -1, 1)$, $(-1, 1, 1)$.

5. $(1, 1, -2)$, $(1, -2, 1)$, $(-2, 1, 1)$.

6. $(1, 1, 1, -2)$, $(1, 1, -2, 1)$, $(1, -2, 1, 1)$, $(-2, 1, 1, 1)$.

7. $(1,1,1,-3)$, $(1,1,-3,1)$, $(1,-3,1,1)$, $(-3,1,1,1)$.

8. $(1,1,1)$, $(2,2,2)$, $(3,3,3)$, $(4,4,4)$.

9. $(1,-1,0,0)$, $(1,0,-1,0)$, $(0,1,-1,0)$, $(1,0,0,-1)$, $(0,1,0,-1)$, $(0,0,1,-1)$.

Exercise 2.38. Find a basis of the column space in Exercise 1.36.

We need to justify the definition of dimension by showing that two minimal spanning sets of the same subspace H have the same number of vectors. Example 2.5.2 and the subsequent discussion show that, if $H = \text{Span}\alpha$, then the number of vectors in a minimal spanning subset within α is $\text{rank}\alpha$. Therefore the justification is a consequence of the following statement: If $\text{Span}\alpha \subset \text{Span}\beta$, then $\text{rank}\alpha \le \text{rank}\beta$. In terms of the corresponding matrices, this means that, if $\text{Col}A \subset \text{Col}B$, then $\text{rank}A \le \text{rank}B$.

Let us illustrate the statement for the case A and B are 4×3 and $\text{rank}B = 2$. The row operations on B give two pivots, such as (the other row echelon forms are also possible)

$$B \to \begin{pmatrix} \bullet & * & * \\ 0 & 0 & \bullet \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Applying the same row operations to A, we get

$$(B\ A) \to \begin{pmatrix} \bullet & * & * & * & * & * \\ 0 & 0 & \bullet & * & * & * \\ 0 & 0 & 0 & c_{31} & c_{32} & c_{33} \\ 0 & 0 & 0 & c_{41} & c_{42} & c_{43} \end{pmatrix}.$$

If some $c_{ij} \neq 0$, say $c_{i2} \neq 0$, then for the second column $\vec{v}$ of A, we have row operations

$$(B\ \vec{v}) \to \begin{pmatrix} \bullet & * & * & * \\ 0 & 0 & \bullet & * \\ 0 & 0 & 0 & c_{32} \\ 0 & 0 & 0 & c_{42} \end{pmatrix} \to \begin{pmatrix} \bullet & * & * & * \\ 0 & 0 & \bullet & * \\ 0 & 0 & 0 & \bullet \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

By Proposition 2.4.3, this implies that $\vec{v} \notin \text{Col}B$, contradicting to $\text{Col}A \subset \text{Col}B$. Therefore all $c_{ij} = 0$, and we get row operations

$$(B\ \vec{v}) \to \begin{pmatrix} \bullet & * & * & * \\ 0 & \bullet & * & * \\ 0 & 0 & \bullet & * \\ 0 & 0 & 0 & \bullet \end{pmatrix}.$$

This implies that $\vec{v} \notin \text{Col}B$, contradicting to $\text{Col}A = \text{Col}B$. Therefore $A' = O$, and we have row operation

$$A \to \begin{pmatrix} * & * & * \\ * & * & * \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

This implies $\text{rank}A \leq 2 = \text{rank}B$.

The argument for the statement in general is the same. Therefore the concept of dimension is well defined. In fact, the statement says that, if $\alpha \subset H$, then $\text{rank}A \leq \dim H$. Denoting $H' = \text{Span}\alpha$, we have the following.

Theorem 2.5.2. *If $H' \subset H$, then $\dim H' \leq \dim H$.*

Suppose α spans H. Then a minimal spanning set $\alpha' \subset \alpha$ is a basis of H. Therefore

$$\dim H = \#\alpha' \leq \#\alpha.$$

Moreover, we know $\#\alpha' = \#\alpha$ (i.e., α is a basis) if and only if $\alpha' = \alpha$. Therefore we have the following result.

Theorem 2.5.3. *If a set of k vectors span a subspace H, then $k \geq \dim H$. Moreover, if $k = \dim H$, then α is a basis of H.*

Exercise 2.39. Combine Theorems 2.5.2 and 2.5.3 to show that, if $H' \subset H$ and $\dim H' = \dim H$, then $H' = H$.

2.6 Linear Independence

The calculation in Section 2.5 shows that columns of A form a basis of $\text{Col}A$ if and only if all columns are pivot. By Theorem 1.4.2, this means the uniqueness of the solution of $A\vec{x} = \vec{b}$. By $A\vec{x} = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_k\vec{v}_k$, the uniqueness means the following property.

Definition 2.6.1. Vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k$ are *linearly independent* if

$$x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_k\vec{v}_k = y_1\vec{v}_1 + y_2\vec{v}_2 + \cdots + y_k\vec{v}_k \implies x_1 = y_1, x_2 = y_2, \ldots, x_k = y_k.$$

If the vectors are not linearly independent, then we say they are *linearly dependent*.

We introduce the concept in order to have the following result.

Theorem 2.6.2. *A set of vectors in a subspace $H \subset \mathbb{R}^n$ is a basis of H if and only if they span H and are linearly independent.*

Example 2.6.1. Suppose a homogeneous system $A\vec{x} = \vec{0}$ (of 3 equations in 5 variables) has reduced row echelon form

$$(A\,\vec{0}) \to \begin{pmatrix} 1 & c_{12} & 0 & c_{14} & 0 & 0 \\ 0 & 0 & 1 & c_{24} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Then we get general solution

$$\vec{x} = (x_1, x_2, x_3, x_4, x_5) = x_2(-c_{12}, 1, 0, 0, 0) + x_4(-c_{14}, 0, -c_{24}, 1, 0), \quad x_2, x_4 \in \mathbb{R}.$$

This implies

$$\text{Nul}A = \text{Span}\{\vec{v}_1, \vec{v}_2\}, \quad \vec{v}_1 = (-c_{12}, 1, 0, 0, 0),\ \vec{v}_2 = x_4(-c_{14}, 0, -c_{P24}, 1, 0).$$

On the other hand, for $\vec{x} = x_2\vec{v}_1 + x_4\vec{v}_2$ and $\vec{y} = y_2\vec{v}_1 + y_4\vec{v}_2 \in \text{Nul}A$, we have

$$x_2\vec{v}_1 + x_4\vec{v}_2 = y_2\vec{v}_1 + y_4\vec{v}_2 \implies \vec{x} = \vec{y} \implies x_2 = y_2,\ x_4 = y_4.$$

The second equality is due to the fact that equal vectors have equal coordinates. Therefore $\vec{v}_1, \vec{v}_2$ are linearly independent. By Theorem 2.6.2, the two vectors form a basis of Nul$A$.

The example shows that, if $A\vec{x} = \vec{b}$ has solution

$$\vec{x} = \vec{x}_0 + y_1\vec{v}_1 + y_2\vec{v}_2 + \cdots + y_k\vec{v}_k,$$

where $y_1, y_2, \dots, y_k$ are free variables, then $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k$ form a basis of Nul$A$. Since free variables correspond to non-pivot columns of A, if A has n columns, then we have

$$\begin{aligned} \dim \text{Nul}A &= \text{number of non-pivot columns of } A \\ &= n - \text{number of pivot columns of } A = n - \text{rank}A. \end{aligned}$$

Exercise 2.40. Find a basis and then the dimension of the null space of the matrix in Exercise 1.36.

Example 2.6.2. In Example 1.4.1, we have the row operations

$$\begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & a & b \end{pmatrix} \to \begin{pmatrix} 1 & 4 & 7 & 10 \\ 0 & -3 & -6 & -9 \\ 0 & 0 & a-9 & b-12 \end{pmatrix}.$$

The first three columns $(1,2,3), (4,5,6), (7,8,a)$ are linearly independent if and only if the three columns are pivot. The condition is exactly $a \neq 9$. Similarly, the first, second, and fourth columns $(1,2,3), (4,5,6), (10,11,b)$ are linearly independent if and only if $b \neq 12$.

Exercise 2.41. Determine the linear independence of vectors in Exercise 2.37.

Example 2.6.3. We examine the row operations in Example 1.3.1

$$A = (\vec{v}_1\ \vec{v}_2\ \vec{v}_3\ \vec{v}_4) = \begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{pmatrix} \to \begin{pmatrix} 1 & 4 & 7 & 10 \\ 0 & -3 & -6 & -9 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

By restricting to $(\vec{v}_1\ \vec{v}_2)$, we find that all columns are pivot, so that $\beta = \{\vec{v}_1, \vec{v}_2\}$ is linearly independent. By restricting to $(\vec{v}_1\ \vec{v}_2\ \vec{v}_3)$, we find that not all columns are pivot, so that adding $\vec{v}_3$ to β makes β linearly dependent. By similar argument, adding $\vec{v}_4$ to β also makes β linearly dependent. Therefore α is maximal independent subset of the four column vectors.

By Example 2.5.2, we know β is a basis of the column space Col$A$. The example suggests the following characterisation of basis of subspace in terms of linear independence. The characterisation complements the definition of basis.

Proposition 2.6.3. *A set of vectors in a subspace $H \subset \mathbb{R}^n$ is a basis of H if and only if it is a maximal independent set.*

The maximal independence has two possible meanings. Both are valid for the proposition.

1. If $H = \text{Span}\alpha$, then $\beta \subset \alpha$ is a basis of H if and only if β is linearly independent, and adding any more vector in α to β makes β linearly dependent.

2. A set $\alpha \subset H$ is a basis of H if and only if α is linearly independent, and adding any vector of H into α makes α linearly dependent.

Example 2.6.3 belongs to the first meaning, and the similar argument justifies the first meaning. For the second meaning, we need criteria for linear independence and linear dependence.

Using $A\vec{x} = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_k\vec{v}_k$ and Theorems 1.4.3 and 2.4.5, we get the following criterion for linear independence. The criterion is the special case $y_1 = y_2 = \cdots = y_k = 0$ of the definition.

Proposition 2.6.4. *Vectors $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k$ are linearly independent if and only if*

$$x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_k\vec{v}_k = \vec{0} \implies x_1 = x_2 = \cdots = x_k = 0.$$

Example 2.6.4. The linear independence of a single vector $\vec{v}$ means

$$x\vec{v} = \vec{0} \implies x = 0.$$

This happens precisely when $\vec{v} \neq \vec{0}$.

The linear *dependence* of two vectors $\vec{u}$ and $\vec{v}$ means there are x, y, not all 0, such that

$$x\vec{u} + y\vec{v} = \vec{0}.$$

If $x \neq 0$, then we have $\vec{u} = c\vec{v}$ for $c = -\frac{y}{x}$. If $y \neq 0$, then we have $\vec{v} = c\vec{u}$ for $c = -\frac{x}{y}$. Therefore linear dependence means one vector is a scalar multiple of another, i.e., the two vectors are parallel. We conclude that two vectors are linearly independent if and only if they are not parallel.

For linear dependence, we have the following criterion.

Proposition 2.6.5. *The following are equivalent for vectors $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k$.*

1. *The vectors are linearly dependent.*
2. *It is possible to have $x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_k\vec{v}_k = \vec{0}$, with some $x_i \neq 0$.*
3. *One vector is a linear combination of the others.*

By Proposition 2.6.4, the first and second statements are equivalent. Next we explain the second and third statements are the same.

In the second statement, if x_i is the last nonzero coefficient, then

$$x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_i\vec{v}_i = \vec{0}, \quad x_i \neq 0.$$

This implies $\vec{v}_i$ is a linear combination of $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_{i-1}$

$$\vec{v}_i = -\tfrac{x_1}{x_i}\vec{v}_1 - \tfrac{x_2}{x_i}\vec{v}_2 - \cdots - \tfrac{x_{i-1}}{x_i}\vec{v}_{i-1}.$$

Conversely, suppose the third statement holds, then for some i, we have

$$\vec{v}_i = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_{i-1}\vec{v}_{i-1}$$

This is the same as

$$x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_{i-1}\vec{v}_{i-1} - \vec{v}_i = \vec{0}.$$

Since the coefficient of $\vec{v}_i$ is not 0, the second statement holds.

Example 2.6.5. The vectors $\vec{v}_1 = (1, 2, 3), \vec{v}_2 = (4, 5, 6), \vec{v}_3 = (7, 8, 9)$ are linearly dependent because $\vec{v}_1 - 2\vec{v}_2 + \vec{v}_3 = \vec{0}$. This the same as $\vec{v}_3 = -\vec{v}_1 + 2\vec{v}_2$.

Back to the second meaning of Proposition 2.6.3, by Theorem 2.6.2, we need to explain that, a linearly independent set $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\} \subset H$ spans H if and only if $\alpha \cup \{\vec{v}\}$ is linearly dependent for any $\vec{v} \in H$.

If α spans H, then $\vec{v} \in H$ is a linear combination of α. By Proposition 2.6.5, we find that $\alpha \cup \{\vec{v}\}$ is linearly dependent. Conversely, if $\alpha \cup \{\vec{v}\}$ is linearly dependent, then we have

$$x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_k\vec{v}_k + x\vec{v} = \vec{0}, \quad x_i \neq 0,$$

where at least one of $x_1, x_2, \ldots, x_k, x$ is nonzero. If $x = 0$, then the linear independence of α and Proposition 2.6.4 are contradictory. Therefore $x \neq 0$, so that $\vec{v}$ can be expressed as a linear combination of α. This shows that α spans H.

By the second meaning Proposition 2.6.3, we may start with a linearly independent set $\alpha \subset H$. If α does not span H, then we may add vector $\vec{v} \in H - \text{Span}\alpha$, such that the bigger set $\alpha \cup \{\vec{v}\}$ in H is still linearly independent. We keep doing this until getting a maximal linear independent set in H. This shows that any linearly independent set in H can be extended to a basis of H. Therefore we have the following result. You may compare with Theorem 2.5.3.

Theorem 2.6.6. *If a set of k vectors in a subspace H are linearly independent, then $k \leq \dim H$. Moreover, if $k = \dim H$, then α is a basis of H.*

Example 2.6.6. The vectors $(1,2)$, $(3,4)$ are not scalar multiples of each other. By Example 2.6.4, we have two linearly independent vectors in $\mathbb{R}^2$. By Theorem 2.6.6, the two vectors form a basis of $\mathbb{R}^2$. This implies that the system

$$\begin{aligned} x + 3y &= b_1, \\ 2x + 4y &= b_2, \end{aligned}$$

has unique solution for all b_1, b_2.

2.7 Row Space and Column Space

Given an $m \times n$ matrix A, the columns of A span a subspace $\text{Col}A \subset \mathbb{R}^m$, and the rows of A also span a subspace $\text{Row}A \subset \mathbb{R}^n$, called *row space*. For example, if

$$A = \begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{pmatrix},$$

then

$$\begin{aligned} \text{Col}A &= \text{Span}\{(1,2,3),\ (4,5,6),\ (7,8,9),\ (10,11,12)\} \subset \mathbb{R}^3, \\ \text{Row}A &= \text{Span}\{(1,4,7,10),\ (2,5,8,11),\ (3,6,9,12)\} \subset \mathbb{R}^4. \end{aligned}$$

The two are clearly related by the transpose

$$\text{Row}A = \text{Col}A^T, \quad \text{Col}A = \text{Row}A^T.$$

In fact, for any matrix A, we may introduce four subspaces $\text{Col}A, \text{Row}A, \text{Nul}A, \text{Nul}A^T$.

Row operations on a matrix is the same as *column operations* on its transpose. The following shows that column operations do not change the spanned subspace.

Proposition 2.7.1. *The subspace* $\text{Span}\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k\}$ *is not changed under the following operations.*

1. *Exchange:* $\{\dots, \vec{v}_i, \dots, \vec{v}_j, \dots\} \to \{\dots, \vec{v}_j, \dots, \vec{v}_i, \dots\}$.
2. *Scale:* $\{\dots, \vec{v}_i, \dots\} \to \{\dots, c\vec{v}_i, \dots\}$, $c \neq 0$.
3. *Add a scale multiple:* $\{\dots, \vec{v}_i, \dots, \vec{v}_j, \dots\} \to \{\dots, \vec{v}_i + c\vec{v}_j, \dots, \vec{v}_j, \dots\}$.

For the proof, we need to argue that a linear combination of one side is also a linear combination of the other side. The following shows that, for three vectors, a linear combination of right side is also a linear combination of left side:

$$\begin{aligned}
x_1\vec{v}_1 + x_2\vec{v}_3 + x_3\vec{v}_2 &= x_1\vec{v}_1 + x_3\vec{v}_2 + x_2\vec{v}_3, \\
x_1\vec{v}_1 + x_2(c\vec{v}_2) + x_3\vec{v}_3 &= x_1\vec{v}_1 + cx_2\vec{v}_2 + x_3\vec{v}_3, \\
x_1\vec{v}_1 + x_2(\vec{v}_2 + c\vec{v}_3) + x_3\vec{v}_3 &= x_1\vec{v}_1 + x_2\vec{v}_2 + (cx_2 + x_3)\vec{v}_3.
\end{aligned}$$

The other way around is similar.

Example 2.7.1. To get a basis of the column space of the matrix in Example 2.5.2, we apply column operations

$$\begin{aligned}
A = \begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{pmatrix}
&\xrightarrow[\text{Col}_2 - \text{Col}_1]{\substack{\text{Col}_4 - \text{Col}_3 \\ \text{Col}_3 - \text{Col}_2}}
\begin{pmatrix} 1 & 3 & 3 & 3 \\ 2 & 3 & 3 & 3 \\ 3 & 3 & 3 & 3 \end{pmatrix}
\xrightarrow[\text{Col}_3 - \text{Col}_2]{\text{Col}_4 - \text{Col}_3}
\begin{pmatrix} 1 & 3 & 0 & 0 \\ 2 & 3 & 0 & 0 \\ 3 & 3 & 0 & 0 \end{pmatrix} \\
&\xrightarrow[\text{Col}_1 \leftrightarrow \text{Col}_2]{\frac{1}{3}\text{Col}_2}
\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 1 & 3 & 0 & 0 \end{pmatrix}
\xrightarrow{\text{Col}_2 - \text{Col}_1}
\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 \end{pmatrix}.
\end{aligned}$$

The right side is a *column echelon form*, the simplest shape you can get by column operations. By Proposition 2.7.1, we have

$$\text{Col}A = \text{Span}\{(1, 1, 1), (0, 1, 2)\}.$$

Moreover, the following shows that the two vectors are linearly independent

$$\begin{aligned}
& x_1(1, 1, 1) + x_2(0, 1, 2) = (0, 0, 0) \\
\implies & x_1 = 0 && \text{(look at first coordinate)} \\
\implies & x_2(0, 1, 2) = (0, 0, 0) && \text{(substitute } x_1 = 0\text{)} \\
\implies & x_2 = 0. && \text{(look at second coordinate)}
\end{aligned}$$

Therefore the two non-zero columns $(1,1,1),(0,1,2)$ of the column echelon form form a basis of Col$A$.

The column operations on A can be regarded as row operations on A^T

$$A^T = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix} \xrightarrow{\text{row op}} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

By the method in Section 2.5, we get a basis $(1,4,7,10),(2,5,8,11)$ for the subspace $\text{Col}A^T = \text{Row}A \subset \mathbb{R}^4$.

The example suggests the following alternative way of finding a basis of Col$A$:

1. Column operation on A to get column echelon form C.
2. Non-zero columns of C form a basis of Col$A$.

Since basis of column space can be calculated by both row and column operations, we get the following result.

Theorem 2.7.2. $\text{rank}A^T = \text{rank}A$.

Example 2.7.2. Theorem 2.7.2 shows that the rank is not changed by row and column operations. We have the column operations on the Vandermonde matrix in Example 1.2.1

$$V = \begin{pmatrix} 1 & t_0 & t_0^2 \\ 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \end{pmatrix} \xrightarrow[\text{Col}_2 - t_0\text{Col}_1]{\text{Col}_3 - t_0\text{Col}_2} \begin{pmatrix} 1 & 0 & 0 \\ 1 & t_1 - t_0 & t_1(t_1 - t_0) \\ 1 & t_2 - t_0 & t_2(t_2 - t_0) \end{pmatrix}$$

$$\xrightarrow{\text{Col}_3 - t_1\text{Col}_2} \begin{pmatrix} 1 & 0 & 0 \\ 1 & t_1 - t_0 & 0 \\ 1 & t_2 - t_0 & (t_2 - t_1)(t_2 - t_0) \end{pmatrix}.$$

If t_0, t_1, t_2 are distinct, then $\text{rank}V = 3$. If two of t_0, t_1, t_2 are the same but the third is distinct, then $\text{rank}V = 2$. If all t_0, t_1, t_2 are the same, then $\text{rank}V = 1$.

In general, the rank of $V(t_0, t_1, t_2, \dots, t_n)$ is the number of values of $t_0, t_1, t_2, \dots, t_n$, not counting duplicates.

Exercise 2.42. Find bases of $\text{Col}A, \text{Row}A, \text{Nul}A, \text{Nul}A^T$.

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

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

6. $A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{pmatrix}$.

7. $A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}$.

8. $A = \begin{pmatrix} -1 & 0 & 1 \\ 1 & -1 & 0 \\ 0 & -1 & 1 \end{pmatrix}$.

9. $A = \begin{pmatrix} -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}$.

10. $A = \begin{pmatrix} -1 & 0 \\ 1 & -1 \\ 0 & -1 \end{pmatrix}$.

Chapter 3

Linear Transformation

Chapter 1 provided equation viewpoint. Chapter 2 provided vector viewpoint. This chapter provides the third viewpoint of linear transformation.

Section 1: Linear relation between variables. Linear transformation. Matrix.

Section 2: Operation on linear transformation, becoming operation on matrix.

Section 3: Subspaces associated to linear transformation and matrix. Rank. Calculation.

Section 4: Inverse transformation and inverse matrix. Calculation.

Section 5: Elementary matrix. *LU* factorisation

Section 6: Block matrix

3.1 Matrix of Linear Transformation

We view $A\vec{x}$ as a formula that transforms $\vec{x}$ to $A\vec{x}$.

Example 3.1.1. The flipping of $\mathbb{R}^2$ with respect to the x axis is $T(x,y) = (x,-y)$. The transformation may be denoted

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ -y \end{pmatrix} = \begin{pmatrix} 1\cdot x + 0\cdot y \\ 0\cdot x + (-1)\cdot y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.$$

Therefore we have $T(\vec{x}) = A\vec{x}$ for

$$A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \vec{x} \in \mathbb{R}^2.$$

Similarly, the flipping of the second coordinate in $\mathbb{R}^3$ is given by

$$T\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 \\ -x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

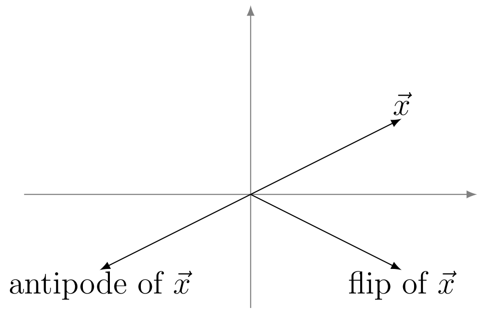

Figure 3.1.1: Flipping and antipode in $\mathbb{R}^2$.

Example 3.1.2. The *identity transformation* $T(\vec{x}) = \vec{x}$ fixes the vector. The following gives the corresponding matrix for the identity on $\mathbb{R}^3$

$$I\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} = \begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} = \begin{pmatrix}1\cdot x_1+0\cdot x_2+0\cdot x_3\\0\cdot x_1+1\cdot x_2+0\cdot x_3\\0\cdot x_1+0\cdot x_2+1\cdot x_3\end{pmatrix} = \begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}.$$

In general, the identity transformation is given by the *identity matrix*

$$I = \begin{pmatrix}1&0&\cdots&0\\0&1&\cdots&0\\\vdots&\vdots&&\vdots\\0&0&\cdots&1\end{pmatrix}.$$

The columns of I is the standard basis of $\mathbb{R}^n$ in Examples 2.2.4 and 2.5.1.

The *antipode transformation* is $T(\vec{x}) = -\vec{x}$. On $\mathbb{R}^3$, this is

$$T\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} = \begin{pmatrix}-x_1\\-x_2\\-x_3\end{pmatrix} = \begin{pmatrix}(-1)\cdot x_1+0\cdot x_2+0\cdot x_3\\0\cdot x_1+(-1)\cdot x_2+0\cdot x_3\\0\cdot x_1+0\cdot x_2+(-1)\cdot x_3\end{pmatrix} = \begin{pmatrix}-1&0&0\\0&-1&0\\0&0&-1\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}.$$

In general, the identity transformation is given by the negative of the identity matrix $-I$.

Example 3.1.3. We may embed straight line $\mathbb{R}^1$ into plane $\mathbb{R}^2$ as horizontal axis or vertical axis. We get transformations $E_h, E_v \colon \mathbb{R}^1 \to \mathbb{R}^2$ between different Euclidean spaces

$$E_h(x) = \begin{pmatrix}x\\0\end{pmatrix} = \begin{pmatrix}1\cdot x\\0\cdot x\end{pmatrix} = \begin{pmatrix}1\\0\end{pmatrix}(x), \quad E_v(x) = \begin{pmatrix}0\\x\end{pmatrix} = \begin{pmatrix}0\cdot x\\1\cdot x\end{pmatrix} = \begin{pmatrix}0\\1\end{pmatrix}(x)$$

The transformations are given by 2×1 matrices. The embedding into the diagonal is also given by a 2×1 matrix

$$E_h(x) = \begin{pmatrix}x\\x\end{pmatrix} = \begin{pmatrix}1\cdot x\\1\cdot x\end{pmatrix} = \begin{pmatrix}1\\1\end{pmatrix}(x)$$

We may also project plane to the two axis, and get transformations $P_h, P_v \colon \mathbb{R}^2 \to \mathbb{R}^1$

$$P_h \begin{pmatrix} x \\ y \end{pmatrix} = (x) = (1\ 0) \begin{pmatrix} x \\ y \end{pmatrix}, \quad P_v \begin{pmatrix} x \\ y \end{pmatrix} = (y) = (0\ 1) \begin{pmatrix} x \\ y \end{pmatrix}.$$

The transformations are given by 1×2 matrices.

Note that in expressing $T(\vec{x}$ as $A\vec{x}$, we require the entries of A to be constants (i.e., not involving $\vec{x}$, and $\vec{x}$ to be the original vector we start with. The following expressions are not regarded as $A\vec{x}$

$$\begin{pmatrix} x+2y \\ 3x+4y \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 3 & 2 \end{pmatrix} \begin{pmatrix} x \\ 2y \end{pmatrix}, \quad \begin{pmatrix} x+y \\ xy \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & x \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

The first expression must be revised to

$$\begin{pmatrix} x+2y \\ 3x+4y \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} x \\ 2 \end{pmatrix},$$

and the second expression can never be of the form $A\vec{x}$.

What kind of transformation is given by the formula $T(\vec{x}) = A\vec{x}$? Suppose $A = (\vec{v}_1\ \vec{v}_2\ \vec{v}_3)$, $\vec{x} = (x_1, x_2, x_3)$. Then $T(\vec{x}) = A\vec{x} = x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{v}_3$ satisfies

$$\begin{aligned} T(\vec{x}+\vec{y}) &= T(x_1+y_1, x_2+y_2, x_3+y_3) \\ &= (x_1+y_1)\vec{v}_1 + (x_2+y_2)\vec{v}_2 + (x_3+y_3)\vec{v}_3 \\ &= x_1\vec{v}_1 + y_1\vec{v}_1 + x_2\vec{v}_2 + y_2\vec{v}_2 + +x_3\vec{v}_3 + y_3\vec{v}_3 \\ &= (x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{v}_3) + (y_1\vec{v}_1 + y_2\vec{v}_2 + y_3\vec{v}_3) = T(\vec{x}) + T(\vec{y}); \\ T(c\vec{x}) &= T(cx_1, cx_2, cx_3) \\ &= cx_1\vec{v}_1 + cx_2\vec{v}_2 + cx_3\vec{v}_3 \\ &= c(x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{v}_3) = cT(\vec{x}). \end{aligned}$$

We see that a transformation given by the formula $T(\vec{x}) = A\vec{x}$ preserves addition and scalar multiplication. Geometrically, such transformation preserves parallelogram and scaling.

Definition 3.1.1. A tranformations $L \colon \mathbb{R}^n \to \mathbb{R}^m$ is *linear* if

$$L(\vec{x}+\vec{y}) = L(\vec{x}) + L(\vec{y}), \quad L(c\vec{x}) = cL(\vec{x}).$$

We change T to L to indicate "linear" property. Combining the two properties, a linear transformation preserves linear combinations

$$L(x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_k\vec{v}_k) = x_1L(\vec{v}_1) + x_2L(\vec{v}_2) + \cdots + x_kL(\vec{v}_k).$$

From Examples 2.2.4 and 2.5.1, recall that any vector $\vec{x} \in \mathbb{R}^n$ is unique linear combination of the standard basis vectors

$$\vec{x} = (x_1, x_2, \dots, x_n) = x_1\vec{e}_1 + x_2\vec{e}_2 + \cdots + x_n\vec{e}_n.$$

Applying a linear transformation $L\colon \mathbb{R}^n \to \mathbb{R}^m$ to the linear combinations, we get

$$\begin{aligned} L(\vec{x}) &= x_1L(\vec{e}_1) + x_2L(\vec{e}_2) + \cdots + x_nL(\vec{e}_n) \\ &= x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n && (\vec{v}_i = L(\vec{e}_i)) \\ &= A\vec{x}. && (A = (\vec{v}_1\ \vec{v}_2\ \cdots\ \vec{v}_n)) \end{aligned}$$

We conclude that a transformation is linear if and only if it is given by the formula $A\vec{x}$:

$$L(\vec{x}) = A\vec{x} \longleftrightarrow A = (L(\vec{e}_1)\ L(\vec{e}_2)\ \cdots\ L(\vec{e}_n)).$$

We call A the *matrix of linear transformation*, and denote

$$A = [L].$$

Example 3.1.4. The flipping of $\mathbb{R}^2$ with respect to the x-axis in Example 3.1.1 clearly preserves addition and scalar multiplication, and is therefore a linear transformation. By $L(1,0) = (1,0)$ and $L(0,1) = (0,-1)$, the matrix of flipping is

$$(L(\vec{e}_1)\ L(\vec{e}_2)) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Example 3.1.5. The antipodal transformation $L(\vec{x}) = -\vec{x}$ preserves addition and scalar multiplication, and is therefore a linear transformation. By $L(\vec{e}_i) = -\vec{e}_i$, its matrix is given by

$$(-\vec{e}_1\ \ -\vec{e}_2\ \cdots\ -\vec{e}_n) = \begin{pmatrix} -1 & 0 & \cdots & 0 \\ 0 & -1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & -1 \end{pmatrix}.$$

Example 3.1.6. The rotation R_θ of $\mathbb{R}^2$ by angle θ preserves addition and scalar multiplication, and is therefore a linear transformation. The rotation of $\vec{e}_1 = (1,0)$ is the unit vector $(\cos\theta, \sin\theta)$ at angle θ. The rotation of $\vec{e}_2 = (0,2)$ is the unit vector $(-\sin\theta, \cos\theta)$ at angle $\theta + \frac{\pi}{2}$. Therefore the matrix of rotation is

$$R_\theta = (R_\theta(\vec{e}_1)\ R_\theta(\vec{e}_2)) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

In other words, the rotation of (x,y) by angle θ is given by the formula

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x\cos\theta - y\sin\theta \\ x\sin\theta + y\cos\theta \end{pmatrix},$$

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

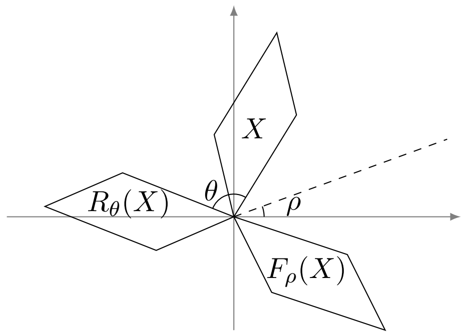

Figure 3.1.2: Rotation and flipping of $\mathbb{R}^2$.

Example 3.1.7. A reflection of $\mathbb{R}^2$ with re4spect to a linear through the origin preserves addition and scalar multiplication, and is therefore a linear transformation. If the linear has angle ρ, then the reflection F_ρ takes $\vec{e}_1$ to the unit vector at angle 2ρ, and also takes $\vec{e}_2$ to the unit vector at angle $2\rho - \frac{\pi}{2}$. Therefore the matrix of reflection is

$$F_\rho = \begin{pmatrix} \cos 2\rho & \cos\left(2\rho - \frac{\pi}{2}\right) \\ \sin 2\rho & \sin\left(2\rho - \frac{\pi}{2}\right) \end{pmatrix} = \begin{pmatrix} \cos 2\rho & \sin 2\rho \\ \sin 2\rho & -\cos 2\rho \end{pmatrix}.$$

Example 3.1.8. The linear transformation taking $\vec{e}_1 = (1, 0)$ to $\vec{v}_1 = (1, 2)$ and taking $\vec{e}_2 = (0, 1)$ to $\vec{v}_2 = (3, 4)$ is given by matrix

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

The reverse transformation L is also linear, and satisfies $L(\vec{v}_1) = \vec{e}_1, L(\vec{v}_2) = \vec{e}_2$. To find the $L(\vec{e}_1)$, the first column of the matrix of L, we try to express the first standard basis vector as $\vec{e}_1 = x_1\vec{v}_1 + x_2\vec{v}_2$. Then we can get

$$L(\vec{e}_1) = x_1 L(\vec{v}_1) + x_2 L(\vec{v}_2) = x_1\vec{e}_1 + x_2\vec{e}_2 = (x_1, x_2).$$

We find that the first column of the matrix of L is in fact the solution of the system $x_1\vec{v}_1 + x_2\vec{v}_2 = \vec{e}_1$. This can be done by row operations on the augmented matrix $(\vec{v}_1\ \vec{v}_2\ \vec{e}_1)$. Similarly, the second column of the matrix of L is the solution obtained by row operations on another augmented matrix $(\vec{v}_1\ \vec{v}_2\ \vec{e}_2)$. We may combine the two row operations

$$(\vec{v}_1\ \vec{v}_2\ \vec{e}_1\ \vec{e}_2) = (A\ I) = \begin{pmatrix} 1 & 3 & 1 & 0 \\ 2 & 4 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & -2 & \frac{3}{2} \\ 0 & 1 & 1 & -\frac{1}{2} \end{pmatrix} = (\vec{e}_1\ \vec{e}_2\ \vec{w}_1\ \vec{w}_2) = (I\ B).$$

Restricting the row operation to the first three columns, we find the solution of $x_1\vec{v}_1 + x_2\vec{v}_2 = \vec{e}_1$ is the third column $\vec{w}_1 = (-2, 1)$ on the right side. Similarly, the solution of $x_1\vec{v}_1 + x_2\vec{v}_2 = \vec{e}_1$ is the fourth column $\vec{w}_2 = (\frac{3}{2}, -\frac{1}{2})$ on the right side. The matrix of L is then

$$A^{-1} = (\vec{w}_1\ \vec{w}_2) = \begin{pmatrix} -2 & \frac{3}{2} \\ 1 & -\frac{1}{2} \end{pmatrix}.$$

The notation A^{-1} indicates the inverse, and is called the *inverse matrix* of A.

Exercise 3.1. Find matrix of flipping of $\mathbb{R}^3$ with respect to the (x, y)-plane.

Exercise 3.2. Find matrix of the linear transformation of $\mathbb{R}^3$ that multiplies every vector by 5. What about $\mathbb{R}^n$?

Exercise 3.3. Find matrix of linear transformation.

1. $L(1, 0) = (1, 2)$, $L(0, 1) = (3, 4)$.
2. $L(1, 2) = (1, 0)$, $L(3, 4) = (0, 1)$.
3. $L(1, 2) = (3, 4)$, $L(3, 4) = (1, 2)$.

Exercise 3.4. Find matrix of linear transformation.

1. $L(1, 0) = (1, 2, 3)$, $L(0, 1) = (4, 5, 6)$.
2. $L(0, 1) = (1, 2, 3)$, $L(1, 0) = (4, 5, 6)$.
3. $L(1, 0, 0) = (1, 2)$, $L(0, 1, 0) = (3, 4)$, $L(0, 0, 1) = (5, 6)$.
4. $L(1, 0, 0) = (1, 2)$, $L(0, 0, 1) = (3, 4)$, $L(0, 1, 0) = (5, 6)$.
5. $L(1, 0, 0) = (1, 2)$, $L(1, 1, 0) = (3, 4)$, $L(1, 1, 1) = (5, 6)$.

3.2 Matrix Operation

Transformations can be combined to produce new transformations. First, two linear transformations $L, K: \mathbb{R}^n \to \mathbb{R}^m$ can be added

$$(L + K)(\vec{x}) = L(\vec{x}) + K(\vec{x}).$$

The following verifies that $L + K$ preserves addition

$$\begin{aligned} (L + K)(\vec{x} + \vec{y}) &= L(\vec{x} + \vec{y}) + K(\vec{x} + \vec{y}) && \text{(definition of } L + K\text{)} \\ &= L(\vec{x}) + L(\vec{y}) + K(\vec{x}) + K(\vec{y}) && (L, K \text{preserve addition}) \\ &= L(\vec{x}) + K(\vec{x}) + L(\vec{y}) + K(\vec{y}) \\ &= (L + K)(\vec{x}) + (L + K)(\vec{y}). && \text{(definition of } L + K\text{)} \end{aligned}$$

Similarly, we may verify $(L+K)(c\vec{x}) = c(L+K)(\vec{x})$. Therefore $L+K$ is also a linear transformation.

Corresponding to the addition of linear transformations, we have the addition of matrices. Specifically, if

$$\begin{aligned} A = [L] &= (\vec{v}_1\ \vec{v}_2\ \cdots\ \vec{v}_n), & \vec{v}_i &= L(\vec{e}_i), \\ B = [K] &= (\vec{w}_1\ \vec{w}_2\ \cdots\ \vec{w}_n), & \vec{w}_i &= K(\vec{e}_i), \end{aligned}$$

then by $(L+K)(\vec{e}_i) = L(\vec{e}_i) + K(\vec{e}_i) = \vec{v}_i + \vec{w}_i$, we define

$$\begin{aligned} A + B = [L+K] &= ((L+K)(\vec{e}_1)\ (L+K)(\vec{e}_2)\ \cdots\ (L+K)(\vec{e}_n)) \\ &= (\vec{v}_1 + \vec{w}_1\ \vec{v}_2 + \vec{w}_2\ \cdots\ \vec{v}_n + \vec{w}_n). \end{aligned}$$

For linear transformations from $\mathbb{R}^3$ to $\mathbb{R}^2$, the addition corresponds to the following addition of 2×3 matrices

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{pmatrix} = \begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} & a_{13}+b_{13} \\ a_{21}+b_{21} & a_{22}+b_{22} & a_{23}+b_{23} \end{pmatrix}.$$

Similarly, we may define the scalar multiplication of a linear transformation

$$(cL)(\vec{x}) = c(L(\vec{x})).$$

We may verify cL is still a linear transformation. Then for $A = [L] = (\vec{v}_1\ \vec{v}_2\ \cdots\ \vec{v}_n)$, the matrix for cL is

$$cA = [cL] = (c\vec{v}_1\ c\vec{v}_2\ \cdots\ c\vec{v}_n).$$

For 2×3 matrix, this means

$$c\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} = \begin{pmatrix} ca_{11} & ca_{12} & ca_{13} \\ ca_{21} & ca_{22} & ca_{23} \end{pmatrix}.$$

Exercise 3.5. Add matrices that can be added together.

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

Exercise 3.6. Verify that $L+K = K+L$. Then explain $A+B = B+A$.

Exercise 3.7. Verify that $c(L+K) = cL+cK$. What does this tell you about addition and scalar multiplication of matrices?

Exercise 3.8. Verify that $(a+b)A = aA+bA$. What does this tell you about addition and scalar multiplication of linear transformations?

Given maps $L\colon \mathbb{R}^n \to \mathbb{R}^m$ and $K\colon \mathbb{R}^k \to \mathbb{R}^n$, we have the *composition*

$$(L \circ K)(\vec{x}) = L(K(\vec{x}))\colon \mathbb{R}^k \to \mathbb{R}^n \to \mathbb{R}^m.$$

The following shows that, if L, K are linear, then the composition also preserves addition

$$\begin{aligned}
(L \circ K)(\vec{x} + \vec{y}) &= L(K(\vec{x} + \vec{y})) && \text{(definition of composition)} \\
&= L(K(\vec{x}) + K(\vec{y})) && \text{(K preserves addition)} \\
&= (L \circ K)(\vec{x}) + (L \circ K)(\vec{y}). && \text{(L preserves addition)}
\end{aligned}$$

We can similarly verify that $L \circ K$ preserves scalar multiplication. Therefore $L \circ K$ is a lienar transformation.

The *multiplication of matrices* corresponds to the composition of linear transformations. Specifically, if $A = [L]$ ($m \times n$ matrix) and $B = [K] = (\vec{w}_1\ \vec{w}_2\ \cdots\ \vec{w}_k)$ ($n \times k$ matrix), then the multiplication

$$\begin{aligned}
AB = [L \circ K] &= (L(K(\vec{e}_1))\ L(K(\vec{e}_2))\ \cdots\ L(K(\vec{e}_k))) \\
&= (L(\vec{w}_1)\ L(\vec{w}_2)\ \cdots\ L(\vec{w}_k)) \\
&= (A\vec{w}_1\ A\vec{w}_2\ \cdots\ A\vec{w}_k).
\end{aligned}$$

The columns of AB is obtained by applying A to the columns of B like the left of system of equations. For example, we have

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} & a_{11}b_{13} + a_{12}b_{23} \\ a_{21}b_{21} + a_{22}b_{21} & a_{21}b_{22} + a_{22}b_{22} & a_{21}b_{23} + a_{22}b_{23} \end{pmatrix}.$$

The formula shows that matrix product is the dot products of rows of the first matrix and columns of the second matrix.

$$\begin{pmatrix} \vec{a}_1 \\ \vec{a}_2 \\ \vdots \\ \vec{a}_m \end{pmatrix} \begin{pmatrix} \vec{b}_1\ \vec{b}_2\ \cdots\ \vec{b}_n \end{pmatrix} = \begin{pmatrix} \vec{a}_1 \cdot \vec{b}_1 & \vec{a}_1 \cdot \vec{b}_2 & \cdots & \vec{a}_1 \cdot \vec{b}_n \\ \vec{a}_2 \cdot \vec{b}_1 & \vec{a}_2 \cdot \vec{b}_2 & \cdots & \vec{a}_2 \cdot \vec{b}_n \\ \vdots & \vdots & & \vdots \\ \vec{a}_m \cdot \vec{b}_1 & \vec{a}_m \cdot \vec{b}_2 & \cdots & \vec{a}_m \cdot \vec{b}_n \end{pmatrix}.$$

Example 3.2.1. Any map composed with the identity is the map itself. In terms of matrix, we have $IA = A = AI$. Specifically, we have

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Example 3.2.2. The composition of two rotations is still a rotation: $R_{\theta_1} \circ R_{\theta_2} = R_{\theta_1+\theta_2}$. Correspondingly, we have matrix product

$$\begin{pmatrix} \cos\theta_1 & -\sin\theta_1 \\ \sin\theta_1 & \cos\theta_1 \end{pmatrix} \begin{pmatrix} \cos\theta_2 & -\sin\theta_2 \\ \sin\theta_2 & \cos\theta_2 \end{pmatrix} = \begin{pmatrix} \cos(\theta_1+\theta_2) & -\sin(\theta_1+\theta_2) \\ \sin(\theta_1+\theta_2) & \cos(\theta_1+\theta_2) \end{pmatrix}.$$

Comparing the two sides, we get

$$\cos(\theta_1+\theta_2) = \cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2,$$
$$\sin(\theta_1+\theta_2) = \sin\theta_1\cos\theta_2 + \cos\theta_1\sin\theta_2.$$

Example 3.2.3. The composition has associativity property

$$\begin{aligned}((L\circ K)\circ M)(\vec{x}) &= (L\circ K)(M(\vec{x})) = L(K(M(\vec{x})))\\ &= L((K\circ M)(\vec{x})) = (L\circ(K\circ M))(\vec{x}).\end{aligned}$$

Correspondingly, the product of matrices satisfy $(AB)C = A(BC)$.

On the other hand, the composition of maps is generally not commutative: $L\circ K \neq K\circ L$. For example, we have $F_0\circ R_\theta = F_{-\frac{\theta}{2}}$ and $R_\theta\circ F_0 = F_{\frac{\theta}{2}}$. Correspondingly, we have

$$\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix} \neq \begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}.$$

Exercise 3.9. Multiply matrices in Exercise 3.5 that can be multiplied together.

Exercise 3.10. Flipping twice gives identity. What does this tell you about trigonometric functions?

Exercise 3.11. Verify that composition of linear transformations has the following properties

$$(L+K)\circ M = L\circ M + K\circ M,\quad M\circ(L+K) = M\circ L + M\circ K,\quad L\circ(cK) = c(L\circ K) = (cL)\circ K.$$

What do these tell you about the product of matrices?

3.3 Range and Kernel

An $m\times n$ matrix A gives subspaces

$$\text{Col}A = \{A\vec{x}\colon \vec{x}\in\mathbb{R}^n\} = \{\vec{b}\in\mathbb{R}^m\colon A\vec{x}=\vec{b}\text{ has solution}\}\subset\mathbb{R}^m,$$
$$\text{Nul}A = \{\vec{x}\in\mathbb{R}^n\colon A\vec{x}=\vec{0}\}\subset\mathbb{R}^n.$$

For the corresponding linear transformation $L\colon \mathbb{R}^n\to\mathbb{R}^m$, we have $L(\vec{x}) = A\vec{x}$ and the corresponding suspaces, called *range* and *kernel*

$$\text{Ran}L = \{L(\vec{x})\colon \vec{x}\in\mathbb{R}^n\} = L(\mathbb{R}^n)\subset\mathbb{R}^m,$$
$$\text{Ker}L = \{\vec{x}\in\mathbb{R}^n\colon L(\vec{x})=\vec{0}\} = L^{-1}(\vec{0})\subset\mathbb{R}^n.$$

From matrix, we expect the range and kernel to be subspaces. The following is direct argument that they are closed under addition

$$\begin{aligned}
\vec{u}, \vec{v} \in \text{Ran}L &\implies \vec{u} = L(\vec{x}),\ \vec{v} = L(\vec{y}) \text{ for some } \vec{x}, \vec{y} \in \mathbb{R}^n \\
&\implies \vec{u} + \vec{v} = L(\vec{x}) + L(\vec{y}) = L(\vec{x} + \vec{y}) \in \text{Ran}L; \\
\vec{x}, \vec{y} \in \text{Ker}L &\implies L(\vec{x}) = \vec{0},\ L(\vec{y}) = \vec{0} \\
&\implies L(\vec{x} + \vec{y}) = L(\vec{x}) + L(\vec{y}) = \vec{0} + \vec{0} = \vec{0} \\
&\implies \vec{x} + \vec{y} \in \text{Ker}L.
\end{aligned}$$

A linear transformation is a map with special property. A map $f\colon X \to Y$ is *onto* (or *surjective*) if each element in Y comes from some element in X. In other words, for any $y \in Y$, there is $x \in X$, such that $y = f(x)$.

A map $f\colon X \to Y$ is *one-to-one* (or *injective*) if each element in Y comes from at most one element in X. In other words, if $f(x_1) = f(x_2)$, then $x_1 = x_2$. Equivalently, this means $x_1 \neq x_2$ implies $f(x_1) \neq f(x_2)$.

Example 3.3.1. Consider the map

$$\text{Instructor}\colon \text{Courses} \to \text{Professors}.$$

The map is onto if every professor teaches some course. The map is one-to-one if each professor teaches at most one course.

Example 3.3.2. The identity and antipode are onto and one-to-one. The rotation and flipping are onto and one-to-one.

The embeddings of $\mathbb{R}^1$ into $\mathbb{R}^2$ in Example 3.1.3 are not onto, but is one-to-one. The projections of $\mathbb{R}^2$ to $\mathbb{R}^1$ are onto, but not one-to-one. If we view the projection in Example 3.1.3 as still inside $\mathbb{R}^2$, then the formulae for $P_h, P_v\colon \mathbb{R}^2 \to \mathbb{R}^2$ are given by

$$P_h \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \quad P_v \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ y \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

After the change of viewpoint, the projections are neither onto nor one-to-one.

We may also consider $f(x) = y$ as an equation. Onto means that, for any right side y, the equation has solution. Therefore a linear transformation $L\colon \mathbb{R}^n \to \mathbb{R}^m$ is onto means $\text{Ran}L = \mathbb{R}^m$. The following is a "dictionary" between different viewpoints.

Proposition 3.3.1. *Suppose $L\colon \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, and*

$$A = [L] = (\vec{v}_1\ \vec{v}_2\ \cdots\ \vec{v}_n).$$

Then the following are equivalent.

1. *L is onto, or* $\text{Ran}L = \mathbb{R}^m$.
2. $A\vec{x} = \vec{b}$ *has solution for all* $\vec{b}$.
3. $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ *span* $\mathbb{R}^m$.
4. *All rows of A are pivot.*

The last statement is the calculation of concepts.

One-to-one means that the solution of equation is unique. Then we have the following dictionary.

Proposition 3.3.2. *Suppose* $L\colon \mathbb{R}^n \to \mathbb{R}^m$ *is a linear transformation, and*

$$A = [L] = (\vec{v}_1\ \vec{v}_2\ \cdots\ \vec{v}_n).$$

Then the following are equivalent.

1. *L is one-to-one, or* $\text{Ker}L = \{\vec{0}\}$.
2. *Solution of* $A\vec{x} = \vec{b}$ *is unique.*
3. $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ *are linearly independent.*
4. *All columns of A are pivot.*

3.4 Inverse

Some linear transformations can be reversed. For example, rotation of $\mathbb{R}^2$ by angle θ can be reversed by the rotation by angle $-\theta$.

In general, the *inverse* of a map $f\colon X \to Y$ is a map $g\colon Y \to X$, such that

$$g(f(x)) = x, \quad f(g(y)) = y.$$

The property also means the composition $g \circ f$ is the identity map on X, and $f \circ g$ is the identity map on Y. We denote the inverse map by $g = f^{-1}$.

If a map has inverse, then we say the map is *invertible*.

Theorem 3.4.1. *A map f is invertible if and only if it is onto and one-to-one.*

Onto and one-to-one means surjective and injective. Therefore we also call an invertible map *bijective*.

Example 3.4.1. The map in Example 3.3.1

$$\text{Instructor}\colon \text{Courses} \to \text{Professors}$$

is invertible if and only if every professor teaches exactly one course. In such case, $\text{Instructor}^{-1}(\text{me}) = \text{linear algebra}$.

By definition, a linear transformation $L\colon \mathbb{R}^n \to \mathbb{R}^m$ is invertible if there is a map $K\colon \mathbb{R}^m \to \mathbb{R}^n$, such that $K \circ L$ is the identity on $\mathbb{R}^n$ and $L \circ K$ is the identity on $\mathbb{R}^m$. The inverse map K is necessarily a linear transformation. For $\vec{x}, \vec{y} \in \mathbb{R}^m$, by L linear and $L \circ K$ being identity, we have

$$L(K(\vec{x}) + K(\vec{y})) = L(K(\vec{x})) + L(K(\vec{y})) = \vec{x} + \vec{y} = L(K(\vec{x} + \vec{y})).$$

Then by $K \circ L$ being identity, we have

$$K(\vec{x}) + K(\vec{y}) = K(L(K(\vec{x}) + K(\vec{y}))) = K(L(K(\vec{x} + \vec{y}))) = K(\vec{x} + \vec{y}).$$

In the second equality, we substitute the earlier equality. Similar argument shows $K(c\vec{x}) = cK(\vec{x})$.

By Theorem 3.4.1, invertibility is the same as Propositions 3.3.1 and 3.3.2 combined.

Proposition 3.4.2. *Suppose $L\colon \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, and*

$$A = [L] = (\vec{v}_1 \; \vec{v}_2 \; \cdots \; \vec{v}_n).$$

Then the following are equivalent.

1. *L is invertible.*
2. *$A\vec{x} = \vec{b}$ has unique solution for all $\vec{b}$.*
3. *$\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ is a basis of $\mathbb{R}^m$.*
4. *All rows and columns of A are pivot.*

By Theorem **??**, the invertibility implies $m = n$, or A is a square matrix. By Theorems 2.6.4 and 2.6.5, *under the assumption* $m = n$, one half (onto, always existence, span, all rows pivot) of the invertibility is equivalent to the other half (one-to-one, uniqueness, linear independence, all columns pivot) of the invertibility, and is also equivalent to the whole invertibility.

Corresponding to the inverse of linear transformation, a matrix A is *invertible* if there is a matrix B, such that $AB = I$ and $BA = I$, where I is the identity matrix in Example 3.1.2. Since A must be a square matrix, the two identities have the same size, and we get $AB = I = BA$. We denote the inverse matrix by $B = A^{-1}$. Moreover, by the fourth statement in Proposition 3.4.2, the reduced row echelon form of an invertible matrix is I.

Proposition 3.4.3. *An invertible matrix must be square. Moreover, if A is a square matrix, then the following are equivalent.*

1. *A is invertible.*
2. *$AB = I$ for some B.*
3. *$BA = I$ for some B.*

Moreover, the matrix B in the second or third must be the inverse.

If $AB = I$, then the following shows that $A(B\vec{b}) = I\vec{b} = \vec{b}$ shows that $A\vec{x} = \vec{b}$ has solution $\vec{x} = B\vec{b}$ for all $\vec{b}$. If $BA = I$, then

$$A\vec{x} = \vec{b},\ A\vec{y} = \vec{b} \implies A\vec{x} = A\vec{y} \implies \vec{x} = BA\vec{x} = BA\vec{y} = \vec{y}$$

shows that solution of $A\vec{x} = \vec{b}$ is unique. In case A is a square matrix, both are equivalent to the invertibility of A. Moreover, if $AB = I = B'A$, then $B = BI = BAB' = IB' = B'$. Therefore $AB = I$ and $BA = I$ implies each other, and both imply that $B = A^{-1}$.

Example 3.4.2. In Example 3.1.6, the rotation of $\mathbb{R}^2$ by angle θ is given by the matrix R_θ. Since the inverse of rotation by θ is clearly the rotation by $-\theta$, we get

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \quad R_\theta^{-1} = R_{-\theta} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

The equality $R_\theta R_{-\theta} = I = R_{-\theta}R_\theta$ means $\cos^2\theta + \sin^2\theta = 1$.

Example 3.4.3. In Example 3.1.7, the inverse of flipping of $\mathbb{R}^2$ is clearly the flipping itself. We get

$$F_\rho^{-1} = F_\rho = \begin{pmatrix} \cos 2\rho & \sin 2\rho \\ \sin 2\rho & -\cos 2\rho \end{pmatrix}.$$

The equation $F_\rho^2 = I$ means $\cos^2 2\rho + \sin^2 2\rho = 1$.

Example 3.4.4. We try to find the inverse of the matrix in Example 3.1.8

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

The corresponding linear transformation L satisfies

$$L(\vec{e}_1) = \vec{v}_1 = (1, 2), \quad L(\vec{e}_2) = \vec{v}_2 = (3, 4).$$

The inverse linear transformation satisfies

$$L^{-1}(\vec{v}_1) = \vec{e}_1, \quad L^{-1}(\vec{v}_2) = \vec{e}_2,$$

and the corresponding matrix $[L^{-1}] = A^{-1}$.

The inverse L^{-1} here is the linear transformation L in Example 3.1.8. In the earlier example, the matrix $B = A^{-1}$ was obtained by row operations

$$(A\ I) = \begin{pmatrix} 1 & 3 & 1 & 0 \\ 2 & 4 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & -2 & \frac{3}{2} \\ 0 & 1 & 1 & -\frac{1}{2} \end{pmatrix} = (I\ B).$$

We get

$$\begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}^{-1} = \begin{pmatrix} -2 & \frac{3}{2} \\ 1 & -\frac{1}{2} \end{pmatrix}.$$

In Example 3.4.4, we did not explain why A is invertible. Instead, we gave the method for computing the inverse of a square matrix A: Form the matrix $(A\ I)$ and apply row operation, until the A part becomes I

$$(A\ I) \to (I\ B).$$

Then $B = A^{-1}$. Note that row operations can change A to I, so that the method works, if and only if A is invertible (i.e., I is the reduced row echelon of A).

Example 3.4.5. The row operations

$$\begin{pmatrix} 1 & a & 0 & 1 & 0 & 0 \\ 0 & 1 & a & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & a & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & -a \\ 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & 1 & -a & a^2 \\ 0 & 1 & 0 & 0 & 1 & -a \\ 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

imply

$$\begin{pmatrix} 1 & a & 0 \\ 0 & 1 & a \\ 0 & 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & -a & a^2 \\ 0 & 1 & -a \\ 0 & 0 & 1 \end{pmatrix}.$$

Example 3.4.6. The orthogonal projection $L\colon \mathbb{R}^3 \to \mathbb{R}^3$ onto the plane (subspace) H given by $x + y + z = 0$ is a linear transformation. The matrix of L is $[L] = (L(\vec{e}_1)\ L(\vec{e}_2)\ L(\vec{e}_3))$.

It is not immediately clear what the orthogonal projections of the standard basis vectors are. What is clear is that

$$L(\vec{v}) = \vec{v} \text{ for } \vec{v} \in H, \quad L(\vec{w}) = \vec{0} \text{ for } \vec{w} \perp H.$$

For example, we have

$$\vec{v}_1 = (1, -1, 0) \in H, \quad \vec{v}_2 = (1, 0, -1) \in H, \quad \vec{v}_3 = (1, 1, 1) \perp H.$$

Therefore

$$L(\vec{v}_1) = \vec{v}_1, \quad L(\vec{v}_2) = \vec{v}_2, \quad L(\vec{v}_3) = \vec{0}.$$

If $\vec{e}_1 = x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{v}_3$, then we may get $L(\vec{e}_1) = x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{0} = x_1\vec{v}_1 + x_2\vec{v}_2$. The problem becomes solving $A\vec{x} = \vec{e}_1$ for

$$A = (\vec{v}_1\ \vec{v}_2\ \vec{v}_3) = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix}.$$

We need to do the same for $\vec{e}_2$ and $\vec{e}_3$. This means we may solve three systems of linear equations (same A but different right sides) together by the row operation

$$(A\ I) = (\vec{v}_1\ \vec{v}_2\ \vec{v}_3\ \vec{e}_1\ \vec{e}_2\ \vec{e}_3) = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 & \frac{1}{3} & -\frac{2}{3} & \frac{1}{3} \\ 0 & 1 & 0 & \frac{1}{3} & \frac{1}{3} & -\frac{2}{3} \\ 0 & 0 & 1 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}.$$

This gives

$$A^{-1} = \frac{1}{3}\begin{pmatrix} 1 & -2 & 1 \\ 1 & 1 & -2 \\ 1 & 1 & 1 \end{pmatrix}.$$

By looking at the first four columns, we get

$$\vec{e}_1 = \frac{1}{3}(\vec{v}_1 + \vec{v}_2 + \vec{v}_3), \quad L(\vec{e}_1) = \frac{1}{3}(\vec{v}_1 + \vec{v}_2) = \frac{1}{3}(2, -1, -1).$$

Similarly, we may get

$$L(\vec{e}_2) = \frac{1}{3}(-2\vec{v}_1 + \vec{v}_2) = \frac{1}{3}(-1, 2, -1), \quad L(\vec{e}_3) = \frac{1}{3}(\vec{v}_1 - 2\vec{v}_2) = \frac{1}{3}(-1, -1, 2),$$

and

$$[L] = \frac{1}{3}\begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}.$$

Exercise 3.12. Find orthogonal projection of $\mathbb{R}^2$ onto $ax + by = 0$.

3.5 Block Matrix

Consider the linear transformation $L\colon \mathbb{R}^4 \to \mathbb{R}^4$ that is rotation by θ in the first two coordinates, and flipping with respect to the diagonal in the last two coordinates. Then we have

$$L\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} x_1\cos\theta - x_2\sin\theta \\ x_1\sin\theta + x_2\cos\theta \\ x_4 \\ x_3 \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta & 0 & 0 \\ \sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}.$$

The matrix of L is clearly decomposed as

$$[L] = \begin{pmatrix} R & O \\ O & F \end{pmatrix}, \quad R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \quad F = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad O = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

In general, consider a linear transformation $L\colon \mathbb{R}^4 \to \mathbb{R}^4$, given by matrix A. The restriction of L to the first 2-coordinates is also a linear transformation

$$L_1 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = L \begin{pmatrix} x_1 \\ x_2 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \\ a_{31}x_1 + a_{32}x_2 \\ a_{41}x_1 + a_{42}x_2 \end{pmatrix} : \mathbb{R}^2 \to \mathbb{R}^4.$$

Moreover, further taking only the first two coordinates of L_1 also gives a linear transformation

$$L_{11} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : \mathbb{R}^2 \to \mathbb{R}^2.$$

We may also take the last two coordinates of L_1 to get another linear transformation

$$L_{21} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a_{31}x_1 + a_{32}x_2 \\ a_{41}x_1 + a_{42}x_2 \end{pmatrix} = \begin{pmatrix} a_{31} & a_{32} \\ a_{41} & a_{42} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} : \mathbb{R}^2 \to \mathbb{R}^2.$$

If we start by restricting L to the last two coordinates to get a linear transformation L_2, and then taking the first two and last two coordinates of L_2, we get

$$\begin{aligned} L_{12} \begin{pmatrix} x_3 \\ x_4 \end{pmatrix} &= \begin{pmatrix} a_{13} & a_{14} \\ a_{23} & a_{24} \end{pmatrix} \begin{pmatrix} x_3 \\ x_4 \end{pmatrix}, \\ L_{22} \begin{pmatrix} x_3 \\ x_4 \end{pmatrix} &= \begin{pmatrix} a_{33} & a_{34} \\ a_{43} & a_{44} \end{pmatrix} \begin{pmatrix} x_3 \\ x_4 \end{pmatrix}. \end{aligned}$$

Denote the first two coordinates by $\vec{x}_1 = (x_1, x_2)$ and last two coordinates by $\vec{x}_2 = (x_3, x_4)$, we get

$$\begin{aligned} L \begin{pmatrix} \vec{x}_1 \\ \vec{x}_2 \end{pmatrix} &= L \begin{pmatrix} \vec{x}_1 \\ \vec{0} \end{pmatrix} + L \begin{pmatrix} \vec{0} \\ \vec{x}_2 \end{pmatrix} = L_1(\vec{x}_1) + L_2(\vec{x}_2) = \begin{pmatrix} L_{11}(\vec{x}_1) \\ L_{21}(\vec{x}_1) \end{pmatrix} + \begin{pmatrix} L_{12}(\vec{x}_2) \\ L_{22}(\vec{x}_2) \end{pmatrix} \\ &= \begin{pmatrix} L_{11}(\vec{x}_1) + L_{12}(\vec{x}_2) \\ L_{21}(\vec{x}_1) + L_{22}(\vec{x}_2) \end{pmatrix} = \begin{pmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{pmatrix} \begin{pmatrix} \vec{x}_1 \\ \vec{x}_2 \end{pmatrix}. \end{aligned}$$

Correspondingly, the matrix of L is

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad A_{ij} = [L_{ij}].$$

This is a *block matrix*.

Sometimes, it may be convenient to view a Euclidean vector as combining several Euclidean vectors together

$$\vec{x} = (\vec{x}_1, \vec{x}_2, \dots, \vec{x}_k) \in \mathbb{R}^n, \quad \vec{x}_i \in \mathbb{R}^{n_i}, \quad n_1 + n_2 + \cdots + n_k = n.$$

For a linear transformation $L\colon \mathbb{R}^n \to \mathbb{R}^m$, it may be convenient to have a combination in the source $\mathbb{R}^n$ as above, and another combination in the target $\mathbb{R}^m$

$$\vec{y} = (\vec{y}_1, \vec{y}_2, \dots, \vec{y}_l) \in \mathbb{R}^m, \quad \vec{y}_j \in \mathbb{R}^{m_j}, \quad m_1 + m_2 + \cdots + m_l = m.$$

Then we have block form of linear transformation

$$L = \begin{pmatrix} L_{11} & L_{12} & \cdots & L_{1k} \\ L_{21} & L_{22} & \cdots & L_{2k} \\ \vdots & \vdots & & \vdots \\ L_{l1} & L_{l2} & \cdots & L_{lk} \end{pmatrix}, \quad L_{ji}\colon \mathbb{R}^{n_i} \to \mathbb{R}^{m_j}.$$

Correspondingly, we have the block matrix

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1k} \\ A_{21} & A_{22} & \cdots & A_{2k} \\ \vdots & \vdots & & \vdots \\ A_{l1} & A_{l2} & \cdots & A_{lk} \end{pmatrix}, \quad A_{ji} = [L_{ji}]\ (m_j \times n_i \text{ matrix}).$$

The operations of block matrices are the same as the operations of usual matrices (with numbers as entries).

$$a\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} + b\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = \begin{pmatrix} aA_{11} + bB_{11} & aA_{12} + bB_{12} \\ aA_{21} + bB_{21} & aA_{22} + bB_{22} \end{pmatrix},$$

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = \begin{pmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{pmatrix}.$$

The only thing we need to be careful about is that blocks should have matching size, and matrix multiplications is generally not commutative.

Example 3.5.1. We have

$$I_{n_1+n_2} = \begin{pmatrix} I_{n_1} & O \\ O & I_{n_2} \end{pmatrix}$$

Example 3.5.2. We have

$$\begin{pmatrix} I & A \\ O & I \end{pmatrix}\begin{pmatrix} I & B \\ O & I \end{pmatrix} = \begin{pmatrix} I \cdot I + A \cdot O & I \cdot B + A \cdot I \\ O \cdot I + B \cdot O & O \cdot A + B \cdot I \end{pmatrix} = \begin{pmatrix} I & A + B \\ O & I \end{pmatrix}.$$

In particular, we have

$$\begin{pmatrix} I & A \\ O & I \end{pmatrix}^{-1} = \begin{pmatrix} I & -A \\ O & I \end{pmatrix}.$$

Example 3.5.3. We have

$$\begin{pmatrix} A & O & O \\ O & B & O \\ O & O & C \end{pmatrix} \begin{pmatrix} X & O & O \\ O & Y & O \\ O & O & Z \end{pmatrix} = \begin{pmatrix} AX & O & O \\ O & BY & O \\ O & O & CZ \end{pmatrix}.$$

In particular, $\begin{pmatrix} A & O & O \\ O & B & O \\ O & O & C \end{pmatrix}$ is invertible if and only if A, B, C are invertible, and

$$\begin{pmatrix} A & O & O \\ O & B & O \\ O & O & C \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} & O & O \\ O & B^{-1} & O \\ O & O & C^{-1} \end{pmatrix}.$$

Exercise 3.13. Find the inverse of $\begin{pmatrix} I & O & O \\ A & I & O \\ B & C & I \end{pmatrix}$.

Exercise 3.14. Find the condition for $\begin{pmatrix} A & B \\ O & C \end{pmatrix}$ to be invertible.

Chapter 4

Orthogonality

Content

4.1 Orthogonal Basis

Suppose $\alpha = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ is an orthogonal set, and

$$\vec{x} = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_k\vec{v}_k.$$

Then

$$\begin{aligned}\vec{x} \cdot \vec{v}_1 &= x_1\vec{v}_1 \cdot \vec{v}_1 + x_2\vec{v}_2 \cdot \vec{v}_1 + \cdots + x_k\vec{v}_k \cdot \vec{v}_1 \\ &= x_1\vec{v}_1 \cdot \vec{v}_1 + x_2 0 + \cdots + x_k 0 = x_1\vec{v}_1 \cdot \vec{v}_1.\end{aligned}$$

This gives the following result.

Proposition 4.1.1. *If $\alpha = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ is an orthogonal set, and $\vec{x} \in Span\alpha$, then*

$$\vec{x} = \frac{\vec{x} \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1 + \frac{\vec{x} \cdot \vec{v}_2}{\vec{v}_2 \cdot \vec{v}_2}\vec{v}_2 + \cdots + \frac{\vec{x} \cdot \vec{v}_k}{\vec{v}_k \cdot \vec{v}_k}\vec{v}_k.$$

The proposition tells us that, in a linear combination of orthogonal vectors, the coefficients can be easily calculated by dot product.

4.2 Orthogonal Complement

4.3 Orthogonal Basis

Content

4.4 Orthogonal Basis

Content

Chapter 5

Determinant

Content

5.1 Signed Volume

A *parallelogram* in $\mathbb{R}^2$ is spanned by two vectors

$$P(\vec{v},\vec{w}) = \{x\vec{v} + y\vec{w} \colon 0 \le x, y \le 1\}.$$

Let $\vec{v} = (a, b)$ and $\vec{w} = (c, d)$. The parallelogram $P(\vec{v},\vec{w})$ has four vertices $\vec{0}$, $\vec{v}$, $\vec{w}$ and $\vec{u} = \vec{v} + \vec{w}$.

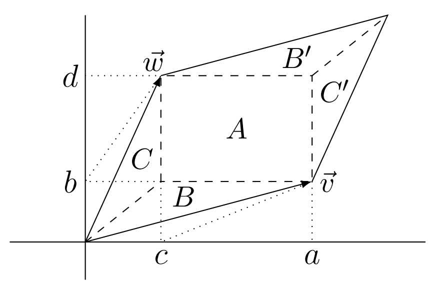

Figure 5.1.1: Area of parallelogram.

To find the area of parallelogram, we divide it into one rectangle A and four triangles B, B', C, C'. The triangle B and B' are identical and therefore have the same area. Moreover, the area of triangle B is half of the dotted rectangle below A, because they have the same base and same height. Therefore the areas of B and B' together is the area of the dotted rectangle below A. By the same reason, the areas of C and C' together is the area of the dotted rectangle on the left of A. The area of parallelogram is then the sum of the areas of the rectangle A, the dotted rectangle below A, and the dotted rectangle on the left of A. This sum is clearly

$$ad - bc = \det \begin{pmatrix} a & c \\ b & d \end{pmatrix}.$$

We note that the formula for the area can also be obtained by using dot product (see the verification of Cauchy-Schwartz inequality after Theorem 2.3.3)

$$\begin{aligned}\text{Area}(P(\vec{v},\vec{w})) &= \sqrt{(\vec{v}\cdot\vec{v})(\vec{w}\cdot\vec{w})-(\vec{v}\cdot\vec{w})^2}\\ &= \sqrt{(a^2+b^2)(c^2+d^2)-(ac+bd)^2}\\ &= \sqrt{a^2d^2+b^2c^2-2abcd} = |ad-bc|.\end{aligned}$$

The area should always be non-negative. However, $ad-bc$ may become negative. In fact, if we exchange $\vec{v}$ and $\vec{w}$ (so that a and c are exchanged, and b and d are exchanged), then $P(\vec{w},\vec{v})$ and $P(\vec{v},\vec{w})$ are the same parallelograms, and should ahve the same area. However, the exchange of vectors gives $cb-da=-(ad-bc)$.

The formula

$$ad-bc = \det\begin{pmatrix} a & c\\ b & d\end{pmatrix} = \det(\vec{v}\ \vec{w})$$

is called *determinant*. It is the area of the parallelogram $P(\vec{v},\vec{w})$ together with a sign. Moreover, the sign is determined by the relative positions of the two vectors.

1. If $\vec{v}$ moves to $\vec{w}$ in *counterclockwise direction*, then $\det(\vec{v}\ \vec{w}) = \text{Area}(P(\vec{v},\vec{w}))$.

2. If $\vec{v}$ moves to $\vec{w}$ in *clockwise direction*, then $\det(\vec{v}\ \vec{w}) = -\text{Area}(P(\vec{v},\vec{w}))$.

The two directions of $\mathbb{R}^2$ are the two *orientations*. We say counterclockwise is *positively oriented* and clockwise is *negatively oriented*.

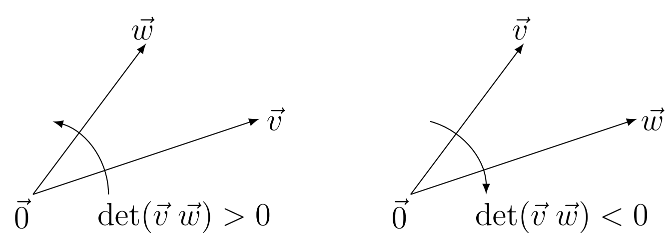

Figure 5.1.2: Sign of determinant of 2×2 matrix.

An $n\times k$ matrix $A = (\vec{v}_1\ \vec{v}_2\ \cdots\ \vec{v}_k)$ is the same as an ordered vector set $\alpha = \{\vec{v}_1,\vec{v}_2,\dots,\vec{v}_k\}$ in $\mathbb{R}^n$. The *parallelotope* spanned by k vectors is

$$P(A) = P(\alpha) = \{x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_k\vec{v}_k \colon 0\le x_i\le 1\}.$$

For the special case $k=n$, the parallelotope is generally an n-dimensional body inside $\mathbb{R}^n$, and has n-dimensional volume. This volume is supposed to be the absolute value of the determinant.

The sign of determinant is supposed to be given by the orientation of the (ordered) vectors.

1. In $\mathbb{R}^1$, rightward is positive orientation, leftward is negative orientation.

2. In $\mathbb{R}^2$, counterclockwise is positive orientation, clockwise is negative orientation.

3. In $\mathbb{R}^3$, right hand rule is positive orientation, left hand rule is negative orientation.

In $\mathbb{R}^n$, the standard basis $\epsilon = \{\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n\}$ represents positive orientation, and switching any two vectors changes orientation. For example, $\{\vec{e}_2, \vec{e}_1, \vec{e}_3, \dots, \vec{e}_n\}$ represents negative orientation.

Definition 5.1.1. The determinant of an $n \times n$ matrix A is the real number $\det A$ specified by the following.

1. The absolute value $|\det A|$ is the n-dimensional volume of the parallelotope spanned by column vectors.

2. The sign of $\det A$ is positive if column vectors represent positive orientation, and is negative if they represent negative orientation.

Example 5.1.1. The columns of the identity matrix I form the standard basis ϵ. The parallelotope $P(I)$ spanned by ϵ is the cube of unit side length, and therefore has volume $|\det I| = 1$. Since ϵ represents positive orientation, we also have $\det I > 0$. Therefore $\det I = 1$.

Example 5.1.2. Since exchanging two columns do not change the parallelotope and therefore the volume, but does change the orientation, this introduces negative sign.

$$\det(\cdots \vec{w} \cdots \vec{v} \cdots) = -\det(\cdots \vec{v} \cdots \vec{w} \cdots).$$

For example,

$$\det \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = \det(\vec{e}_2\ \vec{e}_3\ \vec{e}_1) = -\det(\vec{e}_2\ \vec{e}_1\ \vec{e}_3) = \det(\vec{e}_1\ \vec{e}_2\ \vec{e}_3) = \det I = 1.$$

The parallelotope $P(A)$ of an $n \times n$ matrix A is generally an n-dimensional body. If the column vectors are linearly dependent, however, then one direction is a linear combination of the other directions (Proposition 2.6.5), and therefore the parallelotope collapses to strictly lower dimension. The strictly lower dimension is the same as $P(A)$ having zero n-dimensional volume. By Theorem 2.6.4 and Proposition 3.4.2, we conclude the following.

Theorem 5.1.2. *A square matrix is invertible if and only if its determinant is nonzero.*

Finally, if A is not a square matrix, then $P(A)$ is generally a k-dimensional piece in $\mathbb{R}^n$. The piece still has k-dimensional volume, but has no orientation. The volume is $\sqrt{\det A^T A}$.

5.2 Geometric Property of Determinant

Based on the interpretation as signed volume, we know how column operations changes the determinant.

Proposition 5.2.1. *Let $A = (\vec{v}_1 \ \vec{v}_2 \ \cdots \ \vec{v}_n)$ be a square matrix (i.e., n vectors in $\mathbb{R}^n$). The determinant has the following properties.*

1. $\det(\vec{v}_2 \ \vec{v}_1 \ \vec{v}_3 \ \cdots) = -\det A$.
2. $\det(c\vec{v}_1 \ \vec{v}_2 \ \vec{v}_3 \ \cdots) = c \det A$.
3. $\det(\vec{v}_1 + c\vec{v}_2 \ \vec{v}_2 \ \vec{v}_3 \ \cdots) = \det A$.

The proposition only describes column operation on the first two columns. The properties also apply to column operation on any two columns.

Since each step of column operation involves only two columns, we provide geometrical explanation for two vectors in $\mathbb{R}^2$. The spirit of explanation applies to the general case.

The first property is already given by Figure 5.1.2. The parallelogram is not changed, and the orientation is changed.

In the second property, the parallelogram is stretched in $\vec{v}_1$ direction by c, so that the volume is multiplied by $|c|$. The orientation depends on comparing the directions of $\vec{v}_1$ and $c\vec{v}_1$. The directions are the same (and same orientation) if $c > 0$, and are different (and different orientation) if $c < 0$. The combination of volume and direction gives the second property.

In the third property, the parallelogram keep the "base" $\vec{v}_2$ and only shifts the side parallel to the base. The shifting does not change the distance to the base, and therefore preserves the volume. Moreover, it is clear the orientation is still preserved. Therefore the determinant remains the same.

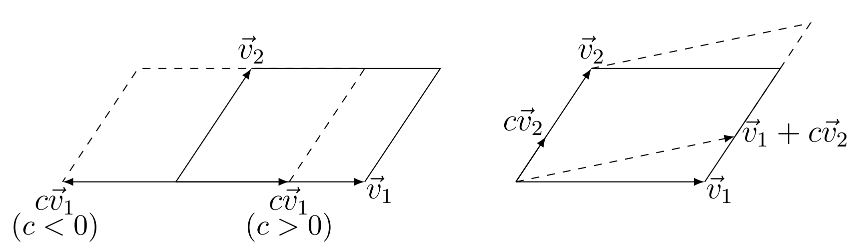

Figure 5.2.1: Column operations on parallelogram.

Example 5.2.1. By column operations, we have

$$\det\begin{pmatrix}1&4&7\\2&5&8\\3&6&a\end{pmatrix}\xlongequal[\text{Col}_3-7\text{Col}_1]{\text{Col}_2-4\text{Col}_1}\det\begin{pmatrix}1&0&0\\2&-3&-6\\3&-6&a-21\end{pmatrix}$$
$$\xlongequal[3\text{Col}_2]{\text{Col}_3-2\text{Col}_2}-3\det\begin{pmatrix}1&0&0\\2&1&0\\3&2&a-9\end{pmatrix}\xlongequal{(a-9)\text{Col}_3}-3(a-9)\det\begin{pmatrix}1&0&0\\2&1&0\\3&2&1\end{pmatrix}$$
$$\xlongequal[\text{Col}_1-3\text{Col}_3]{\substack{\text{Col}_2-2\text{Col}_3\\ \text{Col}_1-2\text{Col}_2}}-3(a-9)\det\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}=-3(a-9).$$

By Theorem 5.1.2, the matrix is invertible if and only if $a\neq 9$.

Example 5.2.2. In general, the last column operations in Example 5.2.1 tells us that the determinant of a *lower triangular matrix* is the product of diagonal entries

$$\det\begin{pmatrix}a_1&0&\cdots&0\\ *&a_2&\cdots&0\\ \vdots&\vdots&&\vdots\\ *&*&\cdots&a_n\end{pmatrix}=a_1a_2\cdots a_n\det\begin{pmatrix}1&0&\cdots&0\\ *&1&\cdots&0\\ \vdots&\vdots&&\vdots\\ *&*&\cdots&1\end{pmatrix}$$
$$=a_1a_2\cdots a_n\det\begin{pmatrix}1&0&\cdots&0\\ 0&1&\cdots&0\\ \vdots&\vdots&&\vdots\\ 0&0&\cdots&1\end{pmatrix}=a_1a_2\cdots a_n.$$

The argument above assumes all $a_i\neq 0$. If some $a_i=0$, then the matrix is not invertible. By Theorem 5.1.2, we get $\det A=0=a_1a_2\cdots a_n$.

The same argument also applies to *upper triangular matrix*

$$\det\begin{pmatrix}a_1&*&\cdots&*\\ 0&a_2&\cdots&*\\ \vdots&\vdots&&\vdots\\ 0&0&\cdots&a_n\end{pmatrix}=a_1a_2\cdots a_n=\det\begin{pmatrix}a_1&0&\cdots&0\\ *&a_2&\cdots&0\\ \vdots&\vdots&&\vdots\\ *&*&\cdots&a_n\end{pmatrix}.$$

Now we turn to the column operation.

Proposition 5.2.2. *For row operations, the determinant has the same properties as the column operations.*

Figure 5.2.2 shows the effect of three row operations on a 2×2 matrix. The general situation is similar.

The first operation $\text{Row}_1 \leftrightarrow \text{Row}_2$ is $(x_1, x_2) \to (x_2, x_1)$, which is flipping with respect to the diagonal. The flipping does not change the volume of parallelogram, but reverses the orientation. Therefore, we get

$$\det \begin{pmatrix} x_2 & y_2 \\ x_1 & y_1 \end{pmatrix} = -\det \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \end{pmatrix}.$$

The second operation $c\text{Row}_1$ is $(x_1, x_2) \to (cx_1, x_2)$. This changes volume by multiplying $|c|$, and preserves or reverses the orientation accordingly to $c > 0$ or $c < 0$. Therefore we get

$$\det \begin{pmatrix} cx_1 & cy_1 \\ x_2 & y_2 \end{pmatrix} = c \det \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \end{pmatrix}.$$

The third operation $\text{Row}_1 + c\text{Row}_2$ is $(x_1, x_2) \to (x_1 + cx_2, x_2)$. To compare the volumes of the two parallelograms, we dive each into two equal triangles (by dotted lines). Each triangle has volume $\frac{1}{2}ah$. Therefore the two parallelograms have the same volume. Moreover, it is clear the orientation is preserved. Therefore the determinant is preserved

$$\det \begin{pmatrix} x_1 + cx_2 & y_1 + cy_2 \\ x_2 & y_2 \end{pmatrix} = \det \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \end{pmatrix}.$$

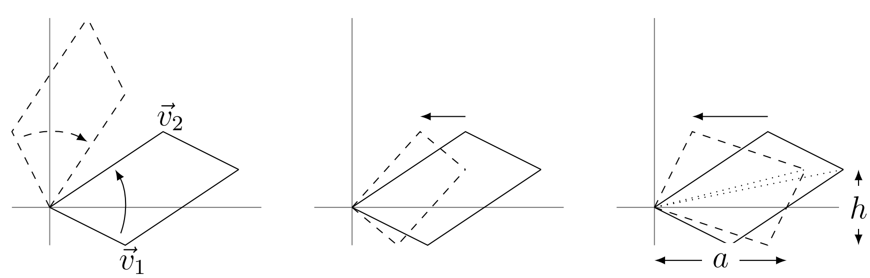

Figure 5.2.2: Row operations on parallelogram.

The row and column operations can change any matrix to upper to lower triangular. Since determinant behaves the same for row and column operations, and Example 5.2.2 shows that $\det A^T = \det A$ for upper or lower triangular A, we get the following.

Proposition 5.2.3. $\det A^T = \det A$.

Example 5.2.3. We calculate the determinant in Example 5.2.1 by mixing row and column operations, and the determinant of upper or lower triangular matrix

$$\det\begin{pmatrix}1&4&7\\2&5&8\\3&6&a\end{pmatrix} \xlongequal[\text{Row}_2-\text{Row}_1]{\text{Row}_3-\text{Row}_2} \det\begin{pmatrix}1&4&7\\1&1&1\\1&1&a-8\end{pmatrix} \xlongequal[\text{Col}_2-\text{Col}_1]{\text{Col}_3-\text{Col}_2} \det\begin{pmatrix}1&3&3\\1&0&0\\1&0&a-9\end{pmatrix}$$

$$\xlongequal[\text{Row}_2\leftrightarrow\text{Row}_3]{\substack{\text{Col}_1\leftrightarrow\text{Col}_2\\ \text{Col}_2\leftrightarrow\text{Col}_3}} -\det\begin{pmatrix}3&3&1\\0&a-9&1\\0&0&1\end{pmatrix} = -3\cdot(a-9)\cdot 1 = -3(a-9).$$

Note that the negative sign after the third equality is due to odd number of exchanges.

Proposition 5.2.4. *The determinant has the following properties.*

1. $\det\begin{pmatrix}A&O\\O&B\end{pmatrix} = \det A\det B.$

2. $\det AB = \det A\det B.$

The parallelotope $P\begin{pmatrix}A&O\\O&B\end{pmatrix}$ can be regarded as a "super-rectangle" with $P(A)$ and one side and $P(B)$ as another side. See Figure 5.2.3. The orientation of $\mathbb{R}^{n_1}$ (the first n_1 coordinates) followed by the orientation of $\mathbb{R}^{n_2}$ (the last n_2 coordinates) is the orientation of $\mathbb{R}^{n_1+n_2}$. Therefore the fourth statement follows.

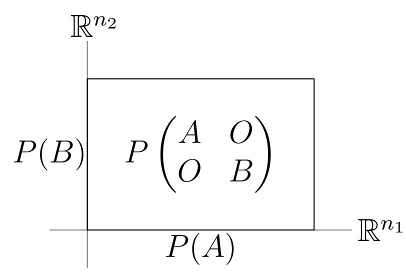

Figure 5.2.3: Geometric property of determinant.

The last statement follows from the fact that a linear transformation $L(\vec{x}) = A\vec{x}\colon \mathbb{R}^n \to \mathbb{R}^n$ multiplies the volume by $|\det A|$[1]: $\text{vol}(L(X)) = |\det A|\text{vol}(X)$. Since L takes column vectors of B to column vectors of AB, it takes $P(B)$ to $P(AB)$. Therefore $L(P(B)) = P(AB)$, and

$$|\det AB| = \text{vol}(P(AB)) = \text{vol}(L(P(B))) = |\det A|\text{vol}(P(B)) = |\det A||\det B|.$$

This gives the fifth statement at least subject to further verification of sign.

[1]This requires further investigation on the true meaning of volume, a topic in measure theory.

Example 5.2.4. Suppose A and B are square matrices. If A is invertible, then we have

$$\begin{pmatrix} A & X \\ O & B \end{pmatrix} \begin{pmatrix} I & O \\ -A^{-1}X & I \end{pmatrix} = \begin{pmatrix} A & O \\ O & B \end{pmatrix}.$$

We have

$$\det \begin{pmatrix} I & O \\ -A^{-1}X & I \end{pmatrix} = 1, \quad \det \begin{pmatrix} A & O \\ O & B \end{pmatrix} = \det A \det B,$$

because the first is lower triangular, and the second follows from the first statement of Proposition 5.2.4. Then by the second statement of Proposition 5.2.4, we get

$$\det \begin{pmatrix} A & X \\ O & B \end{pmatrix} = \det \begin{pmatrix} A & X \\ O & B \end{pmatrix} \det \begin{pmatrix} I & O \\ -A^{-1}X & I \end{pmatrix} = \det \begin{pmatrix} A & O \\ O & B \end{pmatrix} = \det A \det B.$$

In fact, it is possible to use column operation of third kind to change $\det \begin{pmatrix} A & X \\ O & B \end{pmatrix}$ to $\begin{pmatrix} A & O \\ O & B \end{pmatrix}$. By Proposition 5.2.1, the operation does not change determinant.

If A is not invertible, then $\det \begin{pmatrix} A & X \\ O & B \end{pmatrix}$ is also not invertible (see Exercise 3.14). Then by Theorem 5.1.2, we have

$$\det \begin{pmatrix} A & X \\ O & B \end{pmatrix} = 0 = \det A,$$

and the equality above still holds.

In general, if $A_1, A_2, \dots, A_k$ are square matrices, then

$$\det \begin{pmatrix} A_1 & * & \cdots & * \\ O & A_2 & \cdots & * \\ \vdots & \vdots & & \vdots \\ O & O & \cdots & A_n \end{pmatrix} = \det A_1 \det A_2 \cdots \det A_n = \det \begin{pmatrix} A_1 & O & \cdots & O \\ * & A_2 & \cdots & O \\ \vdots & \vdots & & \vdots \\ * & * & \cdots & A_n \end{pmatrix}.$$

5.3 Algebra of Determinant

We have discussed determinant from the geometric view. Now we derive the algebraic formula.

The determinant of 1×1 matrix is $\det(x) = x$.

We already know the determinant of 2×2 matrix

$$\det \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \end{pmatrix} = x_1y_2 - x_2y_1.$$

For 3×3 matrix, we assume $x_{11} \neq 0$ and carry out column operations

$$\begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix} \to \begin{pmatrix} x_{11} & 0 & 0 \\ x_{21} & x_{22} - \frac{x_{12}x_{21}}{x_{11}} & x_{23} - \frac{x_{13}x_{21}}{x_{11}} \\ x_{31} & x_{32} - \frac{x_{12}x_{31}}{x_{11}} & x_{33} - \frac{x_{13}x_{31}}{x_{11}} \end{pmatrix}.$$

Then we get

$$\det\begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix} = \det(x_{11})\det\begin{pmatrix} x_{22}-\frac{x_{12}x_{21}}{x_{11}} & x_{23}-\frac{x_{13}x_{21}}{x_{11}} \\ x_{32}-\frac{x_{12}x_{31}}{x_{11}} & x_{33}-\frac{x_{13}x_{31}}{x_{11}} \end{pmatrix}$$
$$= x_{11}\left[\left(x_{22}-\frac{x_{12}x_{21}}{x_{11}}\right)\left(x_{33}-\frac{x_{13}x_{31}}{x_{11}}\right)-\left(x_{23}-\frac{x_{12}x_{31}}{x_{11}}\right)\left(x_{32}-\frac{x_{13}x_{21}}{x_{11}}\right)\right]$$
$$= x_{11}\left[x_{22}x_{33}-x_{23}x_{32}+\frac{1}{x_{11}}(-x_{22}x_{13}x_{31}-x_{12}x_{21}x_{33}+x_{23}x_{13}x_{21}+x_{12}x_{31}x_{32})\right]$$
$$= x_{11}(x_{22}x_{33}-x_{23}x_{32})+x_{21}(-x_{12}x_{33}+x_{23}x_{13})+x_{31}(-x_{22}x_{13}+x_{12}x_{32})$$
$$= x_{11}\det\begin{pmatrix} x_{22} & x_{23} \\ x_{32} & x_{33} \end{pmatrix} - x_{21}\det\begin{pmatrix} x_{12} & x_{13} \\ x_{32} & x_{33} \end{pmatrix} + x_{31}\det\begin{pmatrix} x_{12} & x_{13} \\ x_{22} & x_{23} \end{pmatrix}.$$

Note that the determinant of a 3×3 matrix is of the form $a_1x_{11} + a_2x_{21} + a_3x_{31}$. Therefore the determinant is a linear function of the first column vector. By exchanging columns, we see that the determinant is also a linear function of the second column, and a linear function of the third column.

$$\det\begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix}$$
$$= -x_{12}\det\begin{pmatrix} x_{21} & x_{23} \\ x_{31} & x_{33} \end{pmatrix} + x_{22}\det\begin{pmatrix} x_{11} & x_{13} \\ x_{31} & x_{33} \end{pmatrix} + x_{32}\det\begin{pmatrix} x_{11} & x_{13} \\ x_{21} & x_{23} \end{pmatrix}$$
$$= x_{13}\det\begin{pmatrix} x_{21} & x_{22} \\ x_{31} & x_{32} \end{pmatrix} - x_{23}\det\begin{pmatrix} x_{11} & x_{12} \\ x_{31} & x_{32} \end{pmatrix} + x_{33}\det\begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix}.$$

By Proposition 5.2.3, we further know that the determinant is a linear function of each row of the matrix.

By the formulae of determinant of up to 3×3 matrices, and the first statement of Proposition 5.2.1, we get the following algebraic definition of determinant.

Definition 5.3.1. The determinant of $n \times n$ matrix $A = (\vec{v}_1\ \vec{v}_2\ \cdots\ \vec{v}_n)$ is a function $D(A)$ satisfying the following properties

1. Multilinear: $D(A)$ is linear in each column vector of A
$$D(\cdots\ a\vec{v}_{1i} + b\vec{v}_{2i}\ \cdots) = aD(\cdots\ \vec{v}_{1i}\ \cdots) + bD(\cdots\ \vec{v}_{2i}\ \cdots).$$
2. Alternating: Switching two columns changes sign
$$D(\cdots\ \vec{v}_j\ \cdots\ \vec{v}_i\ \cdots) = -D(\cdots\ \vec{v}_i\ \cdots\ \vec{v}_j\ \cdots).$$
3. Normal: $D(I) = 1$.

If two columns are equal, then the alternating property implies $D(\cdots\ \vec{v}\ \cdots\ \vec{v}\ \cdots)$ is the negative of itself. Therefore

$$D(\cdots\ \vec{v}\ \cdots\ \vec{v}\ \cdots) = 0. \tag{5.3.1}$$

To justify the definition, we use the multilinear and alternating properties to derive the formula for D. For 2×2 matrix, we let $\vec{e}_1 = (1,0)$, $\vec{e}_2 = (0,1)$ be the standard basis. Then

$$\begin{aligned}
D\begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} &= D(x_{11}\vec{e}_1 + x_{21}\vec{e}_2, x_{12}\vec{e}_1 + x_{22}\vec{e}_2) \\
&= x_{11}D(\vec{e}_1, x_{12}\vec{e}_1 + x_{22}\vec{e}_2) + x_{21}D(\vec{e}_2, x_{12}\vec{e}_1 + x_{22}\vec{e}_2) \\
&= x_{11}x_{12}D(\vec{e}_1,\vec{e}_1) + x_{11}x_{22}D(\vec{e}_1,\vec{e}_2) + x_{21}x_{12}D(\vec{e}_2,\vec{e}_1) + x_{21}x_{22}D(\vec{e}_2,\vec{e}_2) \\
&= x_{11}x_{12}0 + x_{11}x_{22}D(\vec{e}_1,\vec{e}_2) - x_{21}x_{12}D(\vec{e}_1,\vec{e}_2) + x_{21}x_{22}0 \\
&= (x_{11}x_{22} - x_{21}x_{12})D(\vec{e}_1,\vec{e}_2).
\end{aligned}$$

The second equality is by linear in the first column. The third equality is by linear in the second column. The fourth equality is by alternating property and (5.3.1). If the normal property is also satisfied, then $D(\vec{e}_1,\vec{e}_2) = D(I) = 1$, and we get the usual determinant.

For 3×3 matrix, we have

$$\begin{aligned}
&D\begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix} \\
&= D(x_{11}\vec{e}_1 + x_{21}\vec{e}_2 + x_{31}\vec{e}_3, x_{12}\vec{e}_1 + x_{22}\vec{e}_2 + x_{32}\vec{e}_3, x_{13}\vec{e}_1 + x_{23}\vec{e}_2 + x_{33}\vec{e}_3) \\
&= x_{11}x_{22}x_{33}D(\vec{e}_1,\vec{e}_2,\vec{e}_3) + x_{11}x_{32}x_{23}D(\vec{e}_1,\vec{e}_3,\vec{e}_2) \\
&+ x_{21}x_{12}x_{33}D(\vec{e}_2,\vec{e}_1,\vec{e}_3) + x_{21}x_{32}x_{13}D(\vec{e}_2,\vec{e}_3,\vec{e}_1) \\
&+ x_{31}x_{12}x_{23}D(\vec{e}_3,\vec{e}_1,\vec{e}_2) + x_{31}x_{22}x_{13}D(\vec{e}_3,\vec{e}_2,\vec{e}_1) \\
&= (x_{11}x_{22}x_{33} - x_{11}x_{32}x_{23} - x_{21}x_{12}x_{33} \\
&+ x_{21}x_{32}x_{13} + x_{31}x_{12}x_{23} - x_{31}x_{22}x_{13})D(\vec{e}_1,\vec{e}_2,\vec{e}_3).
\end{aligned}$$

In the second equality, we use (5.3.1) to get $D(\vec{e}_1,\vec{e}_1,\vec{e}_2) = D(\vec{e}_2,\vec{e}_2,\vec{e}_3) = 0$, etc. In the third equality, we use alternating property to get

$$D(\vec{e}_3,\vec{e}_1,\vec{e}_2) = -D(\vec{e}_1,\vec{e}_3,\vec{e}_2) = D(\vec{e}_1,\vec{e}_2,\vec{e}_3),$$

and so on. If the normal property is also satisfied, then $D(\vec{e}_1,\vec{e}_2,\vec{e}_3) = D(I) = 1$, and we get the usual determinant.

The calculation shows that, in general, the multilinear and alternating properties imply

$$D\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{pmatrix} = D(I)\sum \text{sign}(i_1, i_2, \dots, i_n)x_{i_1 1}x_{i_2 2}\dots x_{i_n n}.$$

Here $(i_1, i_2, \dots, i_n)$ are *rearrangements* of $(1, 2, \dots, n)$. Moreover, the number of exchanges needed to change $(i_1, i_2, \dots, i_n)$ to $(1, 2, \dots, n)$ gives

$$\operatorname{sign}(i_1, i_2, \dots, i_n) = \begin{cases} 1, & \text{even number of exchanges,} \\ -1, & \text{odd number of exchanges.} \end{cases}$$

For example, the exchanges

$$(24153) \to (21453) \to (12453) \to (12435) \to (12345)$$

imply $\operatorname{sign}(24153) = 1$, and the exchanges

$$(43251) \to (13254) \to (13245) \to (12345)$$

imply $\operatorname{sign}(43251) = -1$.

Example 5.3.1. The determinant of 4×4 matrix has 24 terms

$$\begin{aligned}
&\det \begin{pmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ x_{31} & x_{32} & x_{33} & x_{34} \\ x_{41} & x_{42} & x_{43} & x_{44} \end{pmatrix} \\
&= x_{11}x_{22}x_{33}x_{44} + x_{21}x_{32}x_{43}x_{14} + x_{31}x_{42}x_{13}x_{24} + x_{41}x_{12}x_{23}x_{34} \\
&+ x_{11}x_{32}x_{43}x_{24} + x_{31}x_{42}x_{23}x_{14} + x_{41}x_{22}x_{13}x_{34} + x_{21}x_{12}x_{33}x_{44} \\
&+ x_{11}x_{42}x_{23}x_{34} + x_{41}x_{22}x_{33}x_{14} + x_{21}x_{32}x_{13}x_{44} + x_{31}x_{12}x_{43}x_{24} \\
&- x_{11}x_{22}x_{43}x_{34} - x_{21}x_{42}x_{33}x_{14} - x_{41}x_{32}x_{13}x_{24} - x_{31}x_{12}x_{23}x_{44} \\
&- x_{11}x_{42}x_{33}x_{24} - x_{41}x_{32}x_{23}x_{14} - x_{31}x_{22}x_{13}x_{44} - x_{21}x_{12}x_{43}x_{34} \\
&- x_{11}x_{32}x_{23}x_{44} - x_{31}x_{22}x_{43}x_{14} - x_{21}x_{42}x_{13}x_{34} - x_{41}x_{12}x_{33}x_{24}.
\end{aligned}$$

By $\det A^T = \det A$. All the discussions also applies to rows. The determinant is multilinear and alternating in row vectors.

The determinant is linear in the first column (the second equality uses alternating property for rows)

$$\begin{aligned}
&\det \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix} \\
&= \det \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ 0 & x_{22} & x_{23} \\ 0 & x_{32} & x_{33} \end{pmatrix} + \det \begin{pmatrix} 0 & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ 0 & x_{32} & x_{33} \end{pmatrix} + \det \begin{pmatrix} 0 & x_{12} & x_{13} \\ 0 & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix} \\
&= \det \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ 0 & x_{22} & x_{23} \\ 0 & x_{32} & x_{33} \end{pmatrix} - \det \begin{pmatrix} x_{21} & x_{22} & x_{23} \\ 0 & x_{12} & x_{13} \\ 0 & x_{32} & x_{33} \end{pmatrix} + \det \begin{pmatrix} x_{31} & x_{32} & x_{33} \\ 0 & x_{12} & x_{13} \\ 0 & x_{22} & x_{23} \end{pmatrix} \\
&= x_{11} \det \begin{pmatrix} x_{22} & x_{23} \\ x_{32} & x_{33} \end{pmatrix} - x_{21} \det \begin{pmatrix} x_{12} & x_{13} \\ x_{32} & x_{33} \end{pmatrix} + x_{31} \det \begin{pmatrix} x_{12} & x_{13} \\ x_{22} & x_{23} \end{pmatrix}.
\end{aligned}$$

We already derived this formula before, by geometric properties. We also derived similar formulae for the second and third columns.

In general, let A_{ij} be the matrix obtained by deleting the i-th row and j-th column of an $n] \times n$ matrix A. Then we have *cofactor expansion* along i-th column

$$\det A = (-1)^{1+i} x_{1i} \det A_{1i} + (-1)^{2+i} x_{2i} \det A_{2i} + \cdots + (-1)^{n+i} x_{ni} \det A_{ni}.$$

We also have cofactor expansion along i-th row

$$\det A = (-1)^{i+1} x_{i1} \det A_{i1} + (-1)^{i+2} x_{i2} \det A_{i2} + \cdots + (-1)^{i+n} x_{in} \det A_{in}.$$

Example 5.3.2. Cofactor expansion is the most convenient along rows or columns with only one nonzero entry.

$$\begin{aligned}
\det \begin{pmatrix} t-1 & 2 & 4 \\ 2 & t-4 & 2 \\ 4 & 2 & t-1 \end{pmatrix} &\xlongequal{\text{Col}_1 - \text{Col}_3} \det \begin{pmatrix} t-5 & 2 & 4 \\ 0 & t-4 & 2 \\ -t+5 & 2 & t-1 \end{pmatrix} \\
&\xlongequal{\text{Row}_3 + \text{Row}_1} \det \begin{pmatrix} t-5 & 2 & 4 \\ 0 & t-4 & 2 \\ 0 & 4 & t+3 \end{pmatrix} \\
&\xlongequal{\text{cofactor Col}_1} (t-5) \det \begin{pmatrix} t-4 & 2 \\ 4 & t+3 \end{pmatrix} \\
&= (t-5)(t^2 - t - 20) = (t-5)^2(t+4).
\end{aligned}$$

Example 5.3.3. We calculate the determinant of the 4×4 Vandermonde matrix in Example 1.2.1

$$\begin{aligned}
\det \begin{pmatrix} 1 & t_0 & t_0^2 & t_0^3 \\ 1 & t_1 & t_1^2 & t_1^3 \\ 1 & t_2 & t_2^2 & t_2^3 \\ 1 & t_3 & t_3^2 & t_3^3 \end{pmatrix} &\xlongequal[\text{Col}_2 - t_0\text{Col}_1]{\substack{\text{Col}_4 - t_0\text{Col}_3 \\ \text{Col}_3 - t_0\text{Col}_2}} \det \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & t_1 - t_0 & t_1(t_1 - t_0) & t_1^2(t_1 - t_0) \\ 1 & t_2 - t_0 & t_2(t_2 - t_0) & t_2^2(t_2 - t_0) \\ 1 & t_3 - t_0 & t_3(t_3 - t_0) & t_3^2(t_3 - t_0) \end{pmatrix} \\
&= \det \begin{pmatrix} t_1 - t_0 & t_1(t_1 - t_0) & t_1^2(t_1 - t_0) \\ t_2 - t_0 & t_2(t_2 - t_0) & t_2^2(t_2 - t_0) \\ t_3 - t_0 & t_3(t_3 - t_0) & t_3^2(t_3 - t_0) \end{pmatrix} \\
&\xlongequal[(t_3 - t_0)\text{Row}_3]{\substack{(t_1 - t_0)\text{Row}_1 \\ (t_2 - t_0)\text{Row}_2}} (t_1 - t_0)(t_2 - t_0)(t_3 - t_0) \det \begin{pmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ 1 & t_3 & t_3^2 \end{pmatrix}.
\end{aligned}$$

The second equality use the cofactor expansion along the first row. We find that the calculation is reduced to the determinant of a 3×3 Vandermonde matrix. In general, by induction, we have

$$\det \begin{pmatrix} 1 & t_0 & t_0^2 & \cdots & t_0^n \\ 1 & t_1 & t_1^2 & \cdots & t_1^n \\ 1 & t_2 & t_2^2 & \cdots & t_2^n \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & t_n & t_n^2 & \cdots & t_n^n \end{pmatrix} = \prod_{i<j} (t_j - t_i).$$

Exercise 5.1. Write down cofactor expansions of a 3×3 matrix along rows.

Chapter 6

Eigenvctor

Chapter 7

Vector Space

Content

Section 1: General vector space. Subspace.
Section 2: Linear transformation. Range, null.
Section 3: Basis. Coordinate. Dimension. Rank.
Section 4: Matrix of linear transformation with respect to basis. Change of basis.