Linear Algebra

Min Yan

December 4, 2025

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

1.1.1 Linear Equation

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

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

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

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

Geometrically, we may plot all the solutions of an equation and get a graph in the Euclidean space. The graphs of linear equations are generally infinite and flat subsets of the Euclidean space.

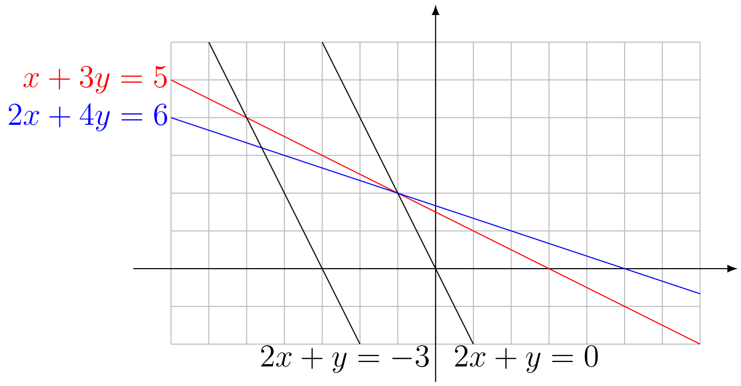

Figure 1.1.1: Graphs of linear equations.

The following are *non-linear* equations

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

Specifically, they are respectively quadratic, quartic, quadratic, algebraic, and transcendental equations. The graph of a nonlinear equation is generally curved.

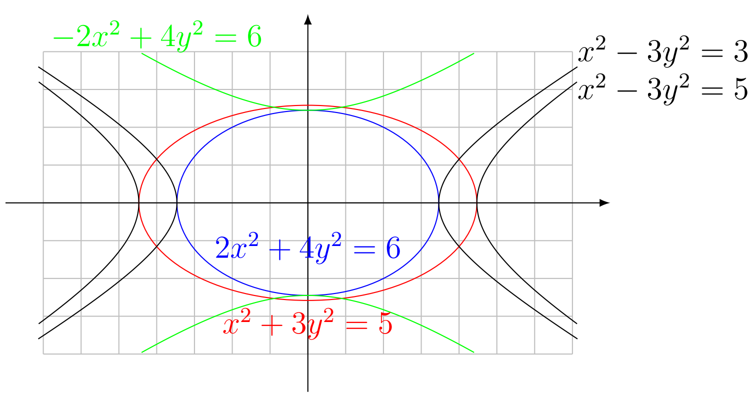

Figure 1.1.2: Graph of nonlinear equation.

Exercise 1.1. Which equations are linear?

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. $ex + 1 = \pi y$.

1.1.2 System of Linear Equations in Two Variables

If quantities are related by several equations, then we get a *system of equations.* If all the equations are linear, then we get 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 two linear equations in two variables.

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

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

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

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 $4\mathrm{Eq}_1 - 3\mathrm{Eq}_2$ to eliminate y

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

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

The solution is reflected in Figure 1.1.1. Solutions of $x + 3y = 5$ and $2x + 4y = 6$ form two lines, and the solution of the system is the unique intersection point $(x, y) = (-1, 2)$ of the two lines. By looking at the intersection, we also know the following system has unique solution

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

and the following system has no solution

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

In Figure 1.1.3, we see that a two variable linear equation $ax+by=c$ is a straight line (if one of a, b is nonzero) perpendicular to the "coefficient direction" (a, b). If we take various c on the right side, then we get parallel straight lines.

We also see that, if the coefficient directions of two linear equations are not parallel, then the system of two linear equations in two variables has unique solution. If the coefficient directions of two linear equations are parallel, then the linear equations represent parallel lines. In this case, the system may have no solution, or may have the whole line as the solution.

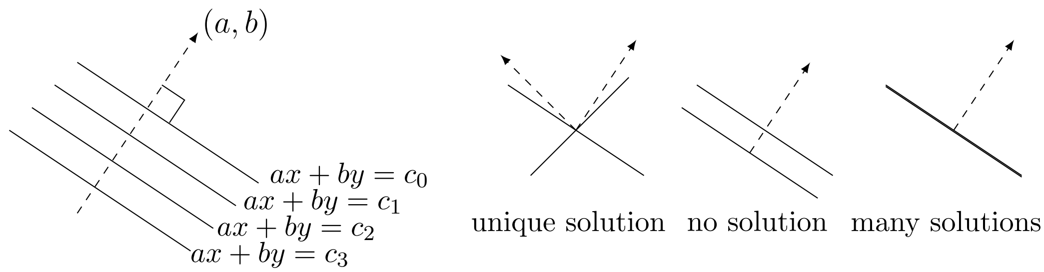

Figure 1.1.3: $ax+by=c$, and solutions of systems of two linear equations.

Exercise 1.2. Solve the systems, by using the graphs as well as by the Gaussian elimination.

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

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

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

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

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

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

Exercise 1.3. Solve systems of linear equations 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. $\begin{aligned} x+3y&=5,\\ 2x+4y&=6,\\ 0&=0. \end{aligned}$

Exercise 1.4. Find values of the right side b_i, such that the system has solution. Is the solution unique?

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

2. $\begin{aligned} x+3y&=b_1,\\ 2x+4y&=b_2,\\ x+\ \ y&=b_3. \end{aligned}$

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

4. $\begin{aligned} x+3y&=b_1,\\ 2x+4y&=b_2,\\ 2x-\ \ y&=b_3. \end{aligned}$

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

Exercise 1.5. By drawing various combinations of straight lines, what kind of solutions can a system of three linear equations in two variables have?

Exercise 1.6. What does the equation $0x + 0y = c$ represent?

1.1.3 System of Linear Equations in Three Variables

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

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

We slightly change the system to

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

Then we similarly carry out $\mathrm{Eq}_2 - 2\mathrm{Eq}_1$, $\mathrm{Eq}_3 - 3\mathrm{Eq}_1$, $\mathrm{Eq}_3 - 2\mathrm{Eq}_2$ to get

$$\begin{aligned} x + 4y + \ \ 7z &= 10, \\ - 3y - \ \ 6z &= -9, \\ - 6y - 12z &= -19. \end{aligned} \quad , \quad \begin{aligned} x + 4y + 7z &= 10, \\ - 3y - 6z &= -9, \\ 0 &= -1. \end{aligned}$$

Since we arrive at a contradiction $0 = -1$, the system has no solution.

We change the system again to

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

We carry out the similar eliminations to get

$$\begin{aligned} x + 4y + 7z &= 10, \\ -\,3y - 6z &= -9, \\ -\,6y - 13z &= -18. \end{aligned}, \qquad \begin{aligned} x + 4y + 7z &= 10, \\ -\,3y - 6z &= -9, \\ -z &= 0. \end{aligned}$$

The last equation gives $z = 0$. Substituting $z = 0$ into the second equation, we get $y = 3$. Substituting $y = 3$ and $z = 0$ into the first equation, we get $x = -2$. We conclude the system has unique solution $(x, y, z) = (-2, 3, 0)$.

In Figure 1.1.4, we see that a three variable linear equation $ax + by + cz = d$ is a plane in $\mathbb{R}^3$ (if one of a, b, c is nonzero) perpendicular to the "coefficient direction" (a, b, c). If we take various d on the right side, then we get parallel planes.

We also see various possibilities for the solutions of a system of linear equations in three variables. The three systems in Example 1.1.2 correspond to the fourth, sixth, and third pictures on the right of Figure 1.1.4.

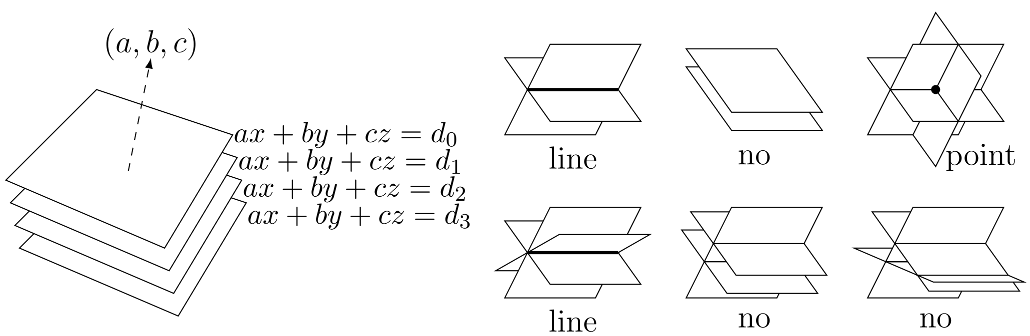

Figure 1.1.4: $ax + by + cz = d$, and solutions of systems of three linear equations.

Exercise 1.7. Solve systems of linear equations.

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

2. $x_1 + 2x_2 + 3x_3 = 4.$

3. $9x_1 + 10x_2 + 11x_3 = 12.$

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

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

6. $$\begin{aligned} 5x + 6y + 7z &= 8, \\ x + 2y + 3z &= 4, \\ 13x + 14y + 15z &= 16, \\ 9x + 10y + 11z &= 12. \end{aligned}$$

Exercise 1.8. Can you find a, such that the system has solution? Is the solution unique?

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

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

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

1.1.4 Homogeneous System of Linear Equations

Example 1.1.3. The following is a system of three linear equations in four variables.

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

The system is called *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 (see Exercise 1.9). To illustrate various possibilities, however, we choose to use different elimination. By $\text{Eq}_1 - \text{Eq}_2$ and $\text{Eq}_2 - \text{Eq}_3$, we get

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

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

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

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

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

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

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

Then we get the solution

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

Exercise 1.9. Use the elimination in Example 1.1.3 to solve the systems in Example 1.1.2. You should get the same solution.

Exercise 1.10. Solve the homogeneous systems of linear equations.

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

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

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

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

1.2 Row Operation

1.2.1 Augmented Matrix

The key information about a system of linear equations is the coefficients on the left and the constants on the right side. We assemble these numbers into a *matrix*. The system in Example 1.1.1 corresponds to a 2×3 matrix

$$\begin{aligned} x + 3y &= 5 \\ 2x + 4y &= 6 \end{aligned} \iff \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}.$$

The first system in Example 1.1.2 corresponds to a 3×4 matrix

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

The last column of the matrices above have different meaning from the other columns. To indicate the difference, for the systems in Examples 1.1.1 and 1.1.2, we denote the *coefficient matrix* A and the *right side vector* $\vec{b}$

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

The columns of A correspond to the variables of the system. The matrices corresponding to the systems in Examples 1.1.1 and 1.1.2 are the *augmented matrices*

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

Then we denote the system as $A\vec{x} = \vec{b}$

$$A\vec{x} = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1x + 3y \\ 2x + 4y \end{pmatrix} = \vec{b} = \begin{pmatrix} 5 \\ 6 \end{pmatrix};$$

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

The second equality is the definition of $A\vec{x}$. The following is the general case of two equations in three variables.

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}, \quad \vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix},$$

$$A\vec{x} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \end{pmatrix}.$$

Later on, the combination $A\vec{x}$ will be interpreted as matrix product. For the moment, it is an integrated notation.

Exercise 1.11. Write down augmented matrices of systems of linear equations in Exercises 1.2, 1.3, 1.7, 1.10.

Exercise 1.12. 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 1.13. Write down systems of linear equations with matrices in Exercise 1.12 as augmented matrices.

1.2.2 Row Operation

The Gaussian elimination on a system of linear equations is equivalent to row operations on the augmented matrix.

Example 1.2.1. The operations $\text{Eq}_2 - 2\text{Eq}_1$ and $\text{Eq}_3 - 3\text{Eq}_1$ in Example 1.1.2 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

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

The end result is the augmented matrix of the system

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

Example 1.2.2. 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 1.2.1

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

Of course, we may also add three 0's to the operations in Example 1.2.1 to get alternative operations for the augmented matrix in this example.

The operations on matrices in Examples 1.2.1 and 1.2.2 are *row operations*. Three row operations were used 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 do not change solutions of corresponding systems of linear equations.

Exercise 1.14. You have solved systems in Exercises 1.2, 1.7, 1.10 by Gaussian eliminations. Can you write down the corresponding row operations?

Exercise 1.15. Let

$$A_1 = \begin{pmatrix} 3 & 5 \\ -2 & 8 \\ 1 & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 3 & 5 & 0 \\ -2 & 8 & 4 \\ 1 & 0 & -2 \end{pmatrix}.$$

Find the results B_i of the row operations. The find the row operations that change B_i back to A_i.

1. $\text{Row}_2 \leftrightarrow \text{Row}_3$.
2. $\text{Row}_1 \leftrightarrow \text{Row}_3$.
3. -3Row_1.
4. $\frac{1}{2}\text{Row}_2$.
5. $\text{Row}_2 + 2\text{Row}_3$.
6. $\text{Row}_3 - 2\text{Row}_1$.

1.2.3 Row Echelon Form

A system of linear equations is simpler if more coefficients are 0. Therefore the goal of row operations is to produce as many 0 as possible. We use the third operation to achieve this, by using some multiple of a nonzero coefficient to "kill" other nonzero coefficients in the same column. The first and second operations are used for further cosmetic improvements.

Example 1.2.3. Let us examine the details of the simplification in Example 1.2.1

$$\begin{pmatrix} \bullet & * & * & * \\ \bullet & * & * & * \\ \bullet & * & * & * \end{pmatrix} = \begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{pmatrix} \to \begin{pmatrix} \bullet & * & * & * \\ 0 & \bullet & * & * \\ 0 & \bullet & * & * \end{pmatrix} = \begin{pmatrix} 1 & 4 & 7 & 10 \\ 0 & -3 & -6 & -9 \\ 0 & -6 & -12 & -18 \end{pmatrix}.$$

In each row, we highlight the first nonzero coefficient $\bullet \neq 0$. The later entries $*$ indicate any real number. We can use 1 to cancel 2 and 3, by multiplying $-\frac{2}{1}$ and $-\frac{3}{1}$ to the first row and then adding to the second and third rows. The key condition is that the coefficient $1 \neq 0$.

In general, we may always use a nonzero entry in a column to eliminate the other entries in the same column. For example, in Example 1.2.1, we continue using one $\bullet$ in the second column to eliminate the other $\bullet$ in the second column

$$\begin{pmatrix} \bullet & * & * & * \\ 0 & \bullet & * & * \\ 0 & \bullet & * & * \end{pmatrix} = \begin{pmatrix} 1 & 4 & 7 & 10 \\ 0 & -3 & -6 & -9 \\ 0 & -6 & -12 & -18 \end{pmatrix} \to \begin{pmatrix} \bullet & * & * & * \\ 0 & \bullet & * & * \\ 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 4 & 7 & 10 \\ 0 & -3 & -6 & -9 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

We note that the end result has the shape of upside down staircase. As we go down the rows, the positions of the first nonzero coefficients are more and more to the right. In terms of the corresponding linear equations, this means the equations are listed from the longest (most complicated) to the shortest (least complicated).

Note that we have not yet fully used -3 to eliminate all the other terms. The following row operations further simplify the entries in the matrix

$$\begin{aligned} & \begin{pmatrix} \bullet & * & * & * \\ 0 & \bullet & * & * \\ 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 4 & 7 & 10 \\ 0 & -3 & -6 & -9 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\ \xrightarrow{\frac{1}{3}\mathrm{Row}_2} & \begin{pmatrix} 1 & * & * & * \\ 0 & 1 & * & * \\ 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 4 & 7 & 10 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\ \xrightarrow{\mathrm{Row}_1 - 4\mathrm{Row}_2} & \begin{pmatrix} 1 & 0 & * & * \\ 0 & 1 & * & * \\ 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \end{aligned}$$

We remark that the operations simplify $\bullet$ to 1, and eliminate all entries above $\bullet$. Moreover, the shape of the matrix remains the same.

Definition 1.2.1. The *row echelon form* (REF) is the simplest *shape* of matrix one can get by row operations. The *reduced row echelon form* (RREF) is the simplest matrix one can get by row operations. In both forms, the lengths of rows are arranged from longest to shortest.

In a row echelon form, the entries occupied by $\bullet$ are called *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 rows, and the first and second columns are pivot columns.

The simplest shape is always the upside down staircase. In Example 1.2.2, we use different row operations to get

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

Both matrices on the right are row echelon forms. This suggests the shape is independent of the choice of row operations. Moreover, if we further simplify the entries

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix} \xrightarrow{\text{Row}_1 - \text{Row}_2} \begin{pmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

then we get the reduced row echelon form. This suggests the reduced echelon form is unique.

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

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

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

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

The following are the corresponding reduced row echelon forms

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

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

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

There is at most one pivot in any row: A row is pivot if it contains a nonzero entry. The first nonzero entry is the pivot. A row is not pivot if all entries in the row are 0.

There is at most one pivot in any column: If there are more than "pivots", then we may use one to eliminate the others.

Therefore we get

$$\text{number of pivots} = \text{number of pivot rows} = \text{number of pivot columns}.$$

Example 1.2.4. In Example 1.1.2, we tried to solve slightly different systems of linear equations. Here we apply different row operations to the augmented matrices

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

$$\xrightarrow[\frac{1}{3}\text{Row}_2]{-\text{Row}_1} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & \frac{8}{3} \\ 0 & 0 & 0 & 1 \end{pmatrix} \xrightarrow[\text{Row}_1 - \text{Row}_3]{\text{Row}_1 - \text{Row}_2} \begin{pmatrix} 1 & 0 & -1 & -\frac{5}{3} \\ 0 & 1 & 2 & \frac{8}{3} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\xrightarrow[\text{Row}_2 - \frac{8}{3}\text{Row}_2]{\text{Row}_1 + \frac{5}{3}\text{Row}_2} \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

By the third matrix, we already get the row echelon form. The last matrix is the reduced row echelon form.

Exercise 1.16. 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 1.17. Find the row echelon forms and reduced row echelon forms of systems of linear equations in Exercises 1.2, 1.7, 1.10.

Exercise 1.18. Find the row echelon forms and reduced row echelon forms of the matrices in Exercise 1.12.

Exercise 1.19. Write down all 2×3 row echelon forms and reduced row echelon forms, and all 3×2 row echelon forms and reduced row echelon forms. How many $n \times 2$ row echelon forms are there?

Exercise 1.20. Write down all 3×3 row echelon forms and reduced row echelon forms. How many $n \times 3$ row echelon forms are there?

1.3 Existence and Uniqueness

1.3.1 Reduced Row Echelon Form is Solution

We simplify the augmented matrix in order to find the solution. In Section 1.2.3, we know the simplest augmented matrix is the reduced row echelon form. We may directly read the solution from the reduced row echelon form.

Example 1.3.1. The reduced row echelon form of the system in Example 1.1.2 and the corresponding system are the following

$$\begin{pmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad \begin{aligned} x_1 \quad - \quad x_3 &= -2, \\ x_2 + 2x_3 &= 3, \\ 0 &= 0. \end{aligned}$$

The last equation $0 = 0$ is trivial, and we get the general solution by moving x_3 in the first and second equations to the right

$$x_1 = -2 + x_3, \quad x_2 = 3 - 2x_3, \quad x_3 \text{ arbitrary.}$$

More generally, suppose a system (of three equations in three variables) has the reduced row echelon form

$$\begin{pmatrix} 1 & 0 & c_1 & d_1 \\ 0 & 1 & c_2 & d_2 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad \begin{aligned} x_1 \quad + c_1x_3 &= d_1, \\ x_2 + c_2x_3 &= d_2, \\ 0 &= 0. \end{aligned}$$

Then we get the general solution

$$x_1 = d_1 - c_1x_3, \quad x_2 = d_2 - c_2x_3, \quad x_3 \text{ arbitrary.}$$

We see that the reduced row echelon form is equivalent to the general solution. For a more complicated 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 the general solution

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

We note that the *free* variables x_2, x_4 correspond to that non-pivot columns. The *non-free* variables, which correspond to the pivot columns, are written in terms of the free variables.

Example 1.3.2. In Example 1.2.4, we get the reduced row echelon form of the second system in Example 1.1.2

$$\begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad \begin{aligned} x_1 \quad - \quad x_3 &= 0, \\ x_2 + 2x_3 &= 0, \\ 0 &= 1. \end{aligned}$$

Since $0 = 1$ is a contradiction, we know the system has no solution. In fact, the contradiction already appears in the first row echelon form

$$\begin{pmatrix} -1 & -1 & -1 & -1 \\ 0 & 3 & 6 & 8 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad \begin{aligned} x_1 - x_2 - x_3 &= -1, \\ 3x_2 + 6x_3 &= 8, \\ 0 &= 1. \end{aligned}$$

In general, if a system has row echelon form in the last column, i.e., the row echelon form has a row $(0\,0\,0\,\bullet)$

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

then we get an equation $0 = \bullet$. Since this is a contradiction, the system has no solution.

Example 1.3.3. The third system in Example 1.1.2 has the following reduced row echelon form

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

The system corresponding to the reduced row echelon form directly gives the *unique* solution

$$x_1 = -2, \quad x_2 = 3, \quad x_3 = 0.$$

The solution is actually the last column of the reduced row echelon form.

In general, if all columns are pivot, then the last column of the reduced row echelon form is the unique solution

$$\begin{pmatrix} 1 & 0 & 0 & d_1 \\ 0 & 1 & 0 & d_2 \\ 0 & 0 & 1 & d_3 \end{pmatrix}, \qquad \begin{aligned} x_1 &= d_1, \\ x_2 &= d_2,. \\ x_3 &= d_3. \end{aligned}$$

Exercise 1.21. Given the reduced 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 1.22. 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.

1.3.2 Criteria for Existence and Uniqueness

We have some observations from Examples 1.3.1, 1.3.2, 1.3.3.

We have two questions for any system of linear equations $A\vec{x} = \vec{b}$:

1. *Existence*: Does the system have solution?
2. *Uniqueness*: Is the solution unique?

We note that the second question makes sense only if the answer to the first question is affirmative.

Example 1.3.1 shows that, if the last column $\vec{b}$ of the augmented matrix $(A\ \vec{b})$ is not pivot, then the system has solution. Example 1.3.2 shows that, if $\vec{b}$ is a pivot column, then the system has no solution. Therefore we get the following criterion for the existence of solution.

Theorem 1.3.1. *A system of linear equations $A\vec{x} = \vec{b}$ has solution if and only if $\vec{b}$ is not a pivot column of the augmented matrix $(A\ \vec{b})$.*

Examples 1.3.1 and 1.3.3 suggest that the uniqueness of the solution is related to the non-pivot columns of A. Specifically, the *columns* of A correspond to the *variables* of the system. Then from the way the solution is read off from the reduced row echelon form, we get the correspondence

variable	column
free	non-pivot
non-free	pivot

The uniqueness means no freedom in the solution, which means all columns are pivot. This is the equivalence between the first and third statements below.

Theorem 1.3.2. *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 $\vec{x} = \vec{0}$.*
3. *All columns of A are pivot.*

The correspondence shows that the right side $\vec{b}$ only affects the existence, and the uniqueness criterion depends only on A. Therefore the uniqueness of the solution of $A\vec{x} = \vec{b}$ is the same as the uniqueness of the solution of the homogeneous system $A\vec{x} = \vec{0}$. Since $A\vec{x} = \vec{0}$ always has the *trivial solution* $\vec{x} = \vec{0}$, the uniqueness means the trivial solution is the only solution.

Exercise 1.23. From the row echelon forms of the augmented matrices, determine the existence and uniqueness of solutions of the corresponding systems 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. In case the solution is not unique, identify the non-free variable.

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. For the augmented matrices, determine the existence and uniqueness of the solutions of the corresponding systems of linear equations.

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

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

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

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

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

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

Example 1.3.4. Consider the system of linear equations

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

By the same row operations in Example 1.2.1, we get

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

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

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

Therefore the condition for the system to have solution is $a \neq 9$, or $a = 9$ and $b = 12$. Moreover, the condition for the solution to be unique is $a \neq 9$. If $a = 9$ and $b = 12$, then we may choose x_3 as the free variable and express x_1, x_2 in terms of x_3.

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

1. $\begin{aligned} x_1 + ax_2 &= 2, \\ 3x_1 - 6x_2 &= b. \end{aligned}$

2. $\begin{aligned} x_1 + 3x_2 &= 2, \\ 3x_1 - 6x_2 &= a, \\ 2x_1 + x_2 &= b. \end{aligned}$

3. $\begin{aligned} x_1 + 3x_2 &= 2, \\ 3x_1 - 6x_2 &= a, \\ x_1 + x_2 &= b. \end{aligned}$

4. $\begin{aligned} x_1 + 3x_2 &= 2, \\ 3x_1 + ax_2 &= b. \end{aligned}$

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

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

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

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

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

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

1.3.3 Criteria for Existence for All Right Side

Example 1.3.5. Consider the system of linear equations

$$\begin{aligned} x_1 + 4x_2 + 7x_3 + 10x_4 &= b_1, \\ 2x_1 + 5x_2 + 8x_3 + 11x_4 &= b_2, \\ 3x_1 + 6x_2 + 9x_3 + ax_4 &= b_3. \end{aligned}$$

By the same row operations in Example 1.2.1, we get

$$\begin{aligned} \begin{pmatrix} 1 & 4 & 7 & 10 & b_1 \\ 2 & 5 & 8 & 11 & b_2 \\ 3 & 6 & 9 & a & b_3 \end{pmatrix} &\to \begin{pmatrix} 1 & 4 & 7 & 10 & b_1 \\ 0 & -3 & -6 & -9 & b_2 - 2b_1 \\ 0 & -6 & -12 & a-30 & b_3 - 3b_1 \end{pmatrix} \\ &\to \begin{pmatrix} 1 & 4 & 7 & 10 & b_1 \\ 0 & -3 & -6 & -9 & b_2 - 2b_1 \\ 0 & 0 & 0 & a-12 & b_1 - 2b_2 + b_3 \end{pmatrix}. \end{aligned}$$

If $a \neq 12$, then the row echelon form is

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

and the solution exists for all right side b_1, b_2, b_3. If $a = 12$, then we get

$$\begin{pmatrix} \bullet & * & * & * & * \\ 0 & \bullet & * & * & * \\ 0 & 0 & 0 & 0 & b_1 - 2b_2 + b_3 \end{pmatrix}.$$

The system does not have solution if $b_1 - 2b_2 + b_3 \neq 0$. Therefore the condition for the system to have solution for all b_1, b_2, b_3 is $a \neq 12$.

Exercise 1.27. Determine the condition for b_i, such that the systems have solutions.

1. $\begin{aligned} x_1 + 3x_2 &= b_1, \\ 3x_1 - 6x_2 &= b_2. \end{aligned}$

2. $\begin{aligned} x_1 + 2x_2 &= b_1, \\ 3x_1 + 6x_2 &= b_2. \end{aligned}$

3. $\begin{aligned} x_1 + 3x_2 &= b_1, \\ 3x_1 - 6x_2 &= b_2, \\ x_1 + 2x_2 &= b_3. \end{aligned}$

4. $\begin{aligned} x_1 + 2x_2 + 3x_3 &= b_1, \\ 4x_1 + 5x_2 + 6x_3 &= b_2, \\ 7x_1 + 8x_2 + 9x_3 &= b_3. \end{aligned}$

5. $\begin{aligned} x_1 + 2x_2 &= b_1, \\ 4x_1 + 5x_2 &= b_2, \\ 7x_1 + 8x_2 &= b_3. \end{aligned}$

6. $\begin{aligned} x_1 + 2x_2 + 3x_3 &= b_1, \\ 4x_1 + 5x_2 + 6x_3 &= b_2. \end{aligned}$

7. $\begin{aligned} x_1 + 2x_2 + 3x_3 &= b_1, \\ 4x_1 + 5x_2 + 6x_3 &= b_2, \\ 7x_1 + 8x_2 + 9x_3 &= b_3, \\ 10x_1 + 11x_2 + 12x_3 &= b_4. \end{aligned}$

Example 1.3.5 and the subsequent exercises suggest the following criterion for the *always existence* of solutions.

Theorem 1.3.3. *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.*

The second statement means the row echelon form of A has no row like $(0\ 0\ \cdots\ 0)$.

Exercise 1.28. For matrices in Exercises 1.12 and 1.25, determine whether $A\vec{x} = \vec{b}$ has solution for all $\vec{b}$.

Exercise 1.29. Determine the condition on a, such that the systems have solutions for all the right sides.

1. $\begin{aligned} x_1 + 3x_2 &= b_1, \\ 3x_1 + ax_2 &= b_2. \end{aligned}$

2. $\begin{aligned} x_1 + ax_2 &= b_1, \\ 3x_1 - 6x_2 &= b_2. \end{aligned}$

3. $\begin{aligned} x_1 + ax_2 &= b_1, \\ 3x_1 + ax_2 &= b_2. \end{aligned}$

4. $\begin{aligned} x_1 + 3x_2 &= b_1, \\ 3x_1 + ax_2 &= b_2, \\ x_1 + 2x_2 &= b_3. \end{aligned}$

5. $\begin{aligned} x_1 + 2x_2 + 3x_3 &= b_1, \\ 4x_1 + 5x_2 + ax_3 &= b_2, \\ 7x_1 + 8x_2 + 9x_3 &= b_3. \end{aligned}$

6. $$\begin{aligned} x_1 + 2x_2 + ax_3 &= b_1, \\ 4x_1 + 5x_2 + ax_3 &= b_2. \end{aligned}$$

7. $$\begin{aligned} x_1 + 2x_2 + ax_3 &= b_1, \\ 4x_1 + ax_2 + ax_3 &= b_2. \end{aligned}$$

8. $$\begin{aligned} x_1 + \ \ 2x_2 + \ \ 3x_3 &= b_1, \\ 4x_1 + \ \ 5x_2 + \ \ ax_3 &= b_2, \\ 7x_1 + \ \ 8x_2 + \ \ 9x_3 &= b_3, \\ 10x_1 + 11x_2 + 12x_3 &= b_4. \end{aligned}$$

1.4 Rank

1.4.1 Essential Size

Rank is the "essential size" of a system. For example, the following system appears to have three equations and four 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 one equation $x_1+2x_2+3x_3+4x_4 = 0$. Therefore the rank of the system (as well as the augmented matrix, and the coefficient matrix) is one.

For another example, we may start with 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 the new system appears to be larger, it is essentially the same as the old system. The larger size of four equations is only an illusion, and the rank of the system of four equations is two.

The row operations reduce any system to the row echelon form, which is the "core" of the system. The size of this core is the essential size. This is the intuition of the concept of rank.

Definition 1.4.1. The *rank* of a matrix A, denoted $\text{rank}A$, is the number of pivots in the row echelon form.

For an $m \times n$ matrix A, we always have

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

If the size of A is $m \times n$, then we have

$$\begin{aligned}\text{rank}A &= \text{number of pivot rows} \le m,\\ \text{rank}A &= \text{number of pivot columns} \le n.\end{aligned}$$

Therefore we get

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

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

Example 1.4.1. In Example 1.3.4, we have the following row operations

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

We further discussed the row echelon form. Then we get

$$\text{rank}\begin{pmatrix}1 & 4 & 7 & 10\\ 2 & 5 & 8 & 11\\ 3 & 6 & a & b\end{pmatrix} = \begin{cases}3, & \text{if } a \ne 9 \text{ or } b \ne 12\\ 2, & \text{if } a = 9 \text{ and } b = 12\end{cases}.$$

If we restrict the row operations to the first three columns, then we get

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

Then it is easy to get

$$\text{rank}\begin{pmatrix}1 & 4 & 7\\ 2 & 5 & 8\\ 3 & 6 & a\end{pmatrix} = \begin{cases}3, & \text{if } a \ne 9\\ 2, & \text{if } a = 9\end{cases}.$$

Exercise 1.30. What is a matrix of rank 0?

Exercise 1.31. Write down all 3×4 row echelon forms of rank 2.

Exercise 1.32. Write down all full rank 2×3, 3×2, and 3×3 row echelon forms.

Exercise 1.33. Find the ranks.

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

Exercise 1.34. Find the ranks of the matrices in Exercises 1.12 and 1.25.

Exercise 1.35. Find the ranks of systems of linear equations in Exercise 1.26.

1.4.2 Existence and Uniqueness in Terms of Rank

We interpret the criteria for the existence and uniqueness in terms of the rank.

Theorem 1.4.2. *Suppose $A\vec{x} = \vec{b}$ is a system of m linear equations in n variables, i.e., A is an $m \times n$ matrix.*

1. *$A\vec{x} = \vec{b}$ has solution if and only if* $\text{rank}(A\ \vec{b}) = \text{rank}A$.
2. *$A\vec{x} = \vec{b}$ has solution for all $\vec{b}$ if and only if* $\text{rank}A = m$.
3. *The solution of $A\vec{x} = \vec{b}$ is unique if and only if* $\text{rank}A = n$.

Theorem 1.3.1 says that $A\vec{x} = \vec{b}$ has solution if and only if $\vec{b}$ is not a pivot column of the augmented matrix $(A\ \vec{b})$. In other words, all pivots of $(A\ \vec{b})$ lie in A. This is the first statement above.

Theorem 1.3.3 says that $A\vec{x} = \vec{b}$ has solution for all $\vec{b}$ if and only if all rows of A are pivot. This is the second statement above.

Theorem 1.3.2 says that the solution of $A\vec{x} = \vec{b}$ is unique if and only if all columns of A are pivot. This is the third statement above.

In fact, according to the correspondence before 1.3.2, the general solution has $n - \text{rank}A$ free variables. Of course the uniqueness means $n - \text{rank}A = 0$.

Since $\text{rank}A \le \min\{m, n\}$, Theorem 1.4.2 implies the following.

Theorem 1.4.3. *Suppose $A\vec{x} = \vec{b}$ is a system of m linear equations in n variables.*

1. *If $A\vec{x} = \vec{b}$ has solution for all $\vec{b}$, then $m \le n$.*

2. *If the solution of* $A\vec{x} = \vec{b}$ *is unique, then* $m \geq n$.

The first statement is the same as that $m < n$ implies 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.

The first statement is the same as that $m > n$ implies $A\vec{x} = \vec{b}$ has no solution for some choice of $\vec{b}$. Intuitively, if the input $\vec{x}$ has n degrees of freedom, then the output $A\vec{x}$ cannot have m degrees of freedom. Therefore the output cannot be all $\vec{b}$.

Example 1.4.2. The system of linear equations in Example 1.3.4 is

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

In Example 1.4.1, we calculated the ranks of the augmented matrix $(A\ \vec{b})$ and the coefficient matrix A. We find $\text{rank}(A\ \vec{b}) = \text{rank}A$ in the following two cases

$$\begin{aligned} a \neq 9\colon \text{rank}(A\ \vec{b}) &= \text{rank}A = 3, \\ a = 9 \text{ and } b = 12\colon \text{rank}(A\ \vec{b}) &= \text{rank}A = 2. \end{aligned}$$

This is the same as the condition for the existence of solution in Example 1.3.4.

Example 1.4.3. Without any calculation, we know the solution of the following system cannot be unique

$$\begin{aligned} x_1 + 4x_2 + 7x_3 + 10x_4 &= 13, \\ 2x_1 + 5x_2 + 8x_3 + 11x_4 &= 14, \\ 3x_1 + 6x_2 + ax_3 + \quad bx_4 &= 15. \end{aligned}$$

The reason is that

$$\text{number of equations} = 3 < 4 = \text{number of variables}.$$

Example 1.4.4. Without any calculation, we know the system

$$\begin{aligned} x_1 + 5x_2 + \quad 9x_3 &= b_1, \\ 2x_1 + 6x_2 + 10x_3 &= b_2, \\ 3x_1 + 7x_2 + \quad ax_3 &= b_3, \\ 4x_1 + 8x_2 + \quad bx_3 &= b_4. \end{aligned}$$

has no solution for some choice of b_i. The reason is that

$$\text{number of equations} = 4 > 3 = \text{number of variables}.$$

The best case for a matrix A is that $A\vec{x}=\vec{b}$ has unique solution for all $\vec{b}$. In this case, by Theorem 1.4.3, we get $m=n$, i.e., A is a *square matrix*.

Conversely, suppose $m=n$. Then $\text{rank}A=m$ if and only if $\text{rank}A=n$. This implies the second and third properties in Theorem 1.4.2 are equivalent. Therefore we conclude the following.

Theorem 1.4.4. *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.*

Exercise 1.36. For the given matrix A, without any row operation, can you say something about the existence or uniqueness of the solutions of $A\vec{x}=\vec{b}$?

1. $\begin{pmatrix} 1 & 2 & 3 & 4 \\ \sqrt{1} & \sqrt{2} & \sqrt{3} & \sqrt{4} \\ 1^2 & 2^2 & 3^2 & 4^2 \end{pmatrix}$.
2. $\begin{pmatrix} 1 & \sqrt{1} & \sin 1 \\ 2 & \sqrt{2} & \sin 2 \\ 3 & \sqrt{3} & \sin 3 \\ 4 & \sqrt{4} & \sin 4 \end{pmatrix}$.
3. $\begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ \sqrt{1} & \sqrt{2} & \sqrt{3} \end{pmatrix}$.
4. $\begin{pmatrix} 1 & 0 & \sqrt{1} \\ 2 & 0 & \sqrt{2} \\ 3 & 0 & \sqrt{3} \end{pmatrix}$.
5. $\begin{pmatrix} 1.2 & \sqrt{3} & 0 & \pi \\ 0 & -12 & \sqrt{5} & e^2 \\ -9 & \sqrt{\pi} & e^{-2} & 0 \end{pmatrix}$.
6. $\begin{pmatrix} 1.2 & 0 & -9 \\ \sqrt{3} & -12 & \sqrt{\pi} \\ 0 & \sqrt{5} & e^{-2} \\ \pi & e^2 & 0 \end{pmatrix}$.

Exercise 1.37. Can you find a, such that $A\vec{x}=\vec{b}$ has unique solution for all $\vec{b}$?

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

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 Euclidean Vector

2.1.1 Euclidean Space

Definition 2.1.1. The *Euclidean space* $\mathbb{R}^n$ of dimension n is the collection of all n-tuples of real numbers, called *vectors*

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

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 Euclidean space $\mathbb{R}^1$ is a straight line. The space $\mathbb{R}^2$ is a plane. 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*)

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

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

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

The two operations satisfy the usual properties

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

The following verifies the first equality (associativity of the addition) in $\mathbb{R}^2$

$$\left(\begin{pmatrix}x_1\\x_2\end{pmatrix}+\begin{pmatrix}y_1\\y_2\end{pmatrix}\right)+\begin{pmatrix}z_1\\z_2\end{pmatrix}=\begin{pmatrix}x_1+y_1\\x_2+y_2\end{pmatrix}+\begin{pmatrix}z_1\\z_2\end{pmatrix}=\begin{pmatrix}(x_1+y_1)+z_1\\(x_2+y_2)+z_2\end{pmatrix}=\begin{pmatrix}x_1+y_1+z_1\\x_2+y_2+z_2\end{pmatrix},$$
$$\begin{pmatrix}x_1\\x_2\end{pmatrix}+\left(\begin{pmatrix}y_1\\y_2\end{pmatrix}+\begin{pmatrix}z_1\\z_2\end{pmatrix}\right)=\begin{pmatrix}x_1\\x_2\end{pmatrix}+\begin{pmatrix}y_1+z_1\\y_2+z_2\end{pmatrix}=\begin{pmatrix}x_1+(y_1+z_1)\\x_2+(y_2+z_2)\end{pmatrix}=\begin{pmatrix}x_1+y_1+z_1\\x_2+y_2+z_2\end{pmatrix}.$$

The right sides are equal.

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

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

The vector is characterised by the property

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

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

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

Exercise 2.1. Verify identities, at least in $\mathbb{R}^2$.

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

Exercise 2.2. Explain that a linear combination of linear combinations is still a linear combination. For example

$$c_1(a_1\vec{x}_1+a_2\vec{x}_2+a_3\vec{x}_3)+c_2(b_1\vec{x}_1+b_2\vec{x}_2+b_3\vec{x}_3)$$

is still a linear combination of $\vec{x}_1,\vec{x}_2,\vec{x}_2$.

Figure 2.1.1 shows that the additions

$$(1,2)+(3,1)=(4,3),\quad (3,1)+(2,-2)=(5,-1)$$

are geometrically given by parallelograms. Moreover, the scalar multiplications

$$2(3,1)=(6,2),\quad 0.5(3,1)=(1.5,0.5)$$

means stretching and shrinking. The negatives (i.e., multiplication by the scalar -1)

$$-(1,2)=(-1,-2),\quad -(1.5,0.5)=(-1.5,-0.5)$$

means the opposite direction.

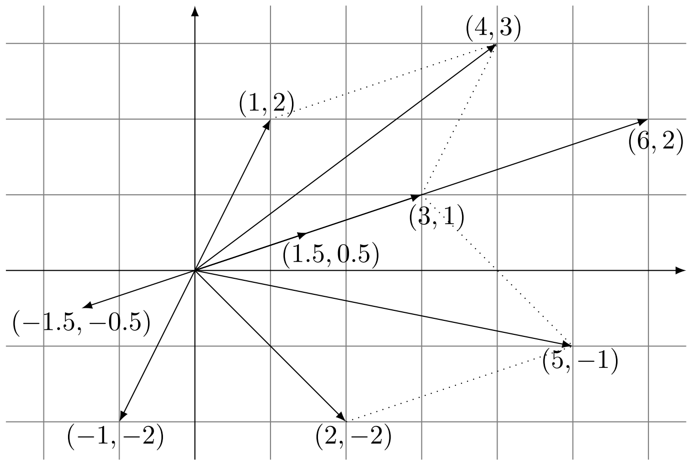

Figure 2.1.1: Euclidean space $\mathbb{R}^2$.

Example 2.1.1. 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 2.1.2. 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 the origin $\vec{x}_0 = (0, 0)$ and has the 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 $\vec{v} = (4, 5, 6) - (1, 2, 3) = (3, 3, 3)$, which 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 2.1.3. 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 2.1.2, this is a line.

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

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

If $\vec{u}$ and $\vec{v}$ are not parallel, then Figure 2.1.2 shows all the linear combinations $x\vec{u} + y\vec{v}$ form a plane. All the solutions of the system in Example 1.1.3 form a plane inside $\mathbb{R}^4$.

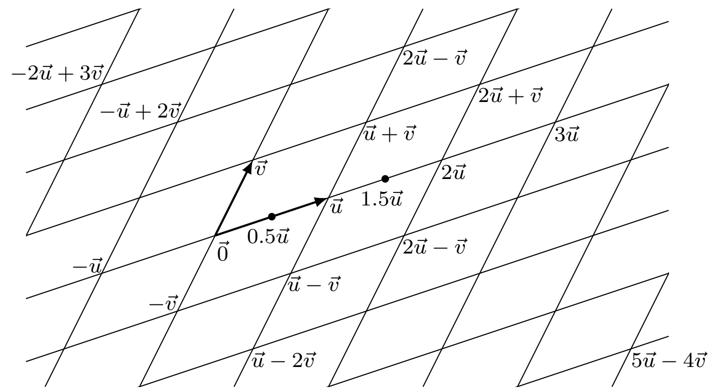

Figure 2.1.2: All the linear combinations $x\vec{u} + y\vec{v}$.

Exercise 2.3. Find the parameterised equations for lines.

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 $\vec{x} = (t, t, t, t)$, $t \in \mathbb{R}$.

2.1.2 System of Linear Equations is Linear Combination

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

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

The second equality means that $A\vec{x}$ is the integrated notation for the left side of the system of linear equations.

In the third equality, two vectors are equal if and only if all the coordinates are equal. Therefore the third equality is exactly the system $A\vec{x} = \vec{b}$.

The first equality is the calculation of linear combinations of the *column vectors* of A. The coefficients x_i are the variables, or coordinates of the vector variable $\vec{x}$.

Generally, a system of linear equations is interpreted as

$$x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{v}_3 = (\vec{v}_1\ \vec{v}_2\ \vec{v}_3)\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \vec{b}.$$

Solving the system means expressing the right side $\vec{b}$ as a linear combination of the column vectors of A.

Of course, the interpretation can be further generally to the case 3 is replaced by n.

$$x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n = (\vec{v}_1\ \vec{v}_2\ \cdots\ \vec{v}_n)\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \vec{b}.$$

The $m \times n$ coefficient matrix is considered as n column vectors in $\mathbb{R}^m$

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

The left side $A\vec{x}$ has the following property.

Proposition 2.1.2. *$A(a\vec{x} + b\vec{y}) = aA\vec{x} + bA\vec{y}$.*

We verify the property for the case $A = (\vec{v}_1\ \vec{v}_2\ \vec{v}_3)$. By

$$a\vec{x} + b\vec{y} = a\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + b\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} ax_1 + by_1 \\ ax_2 + by_2 \\ ax_3 + by_3 \end{pmatrix},$$

we get

$$\begin{aligned} A(a\vec{x} + b\vec{y}) &= (\vec{v}_1\ \vec{v}_2\ \vec{v}_3)\begin{pmatrix} ax_1 + by_1 \\ ax_2 + by_2 \\ ax_3 + by_3 \end{pmatrix} \\ &= (ax_1 + by_1)\vec{v}_1 + (ax_2 + by_2)\vec{v}_2 + (ax_3 + by_3)\vec{v}_3 \\ &= ax_1\vec{v}_1 + by_1\vec{v}_1 + ax_2\vec{v}_2 + by_2\vec{v}_2 + ax_3\vec{v}_3 + by_3\vec{v}_3 \\ &= a(x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{v}_3) + b(y_1\vec{v}_1 + y_2\vec{v}_2 + y_3\vec{v}_3) \\ &= a(\vec{v}_1\ \vec{v}_2\ \vec{v}_3)\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + b(\vec{v}_1\ \vec{v}_2\ \vec{v}_3)\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = aA\vec{x} + bA\vec{y}. \end{aligned}$$

We remark that $A\vec{x}$, $A\vec{y}$, $A(a\vec{x}+b\vec{y})$ are all linear combinations of $\vec{v}_i$. Therefore the proposition shows the linear combination of linear combinations is still a linear combination.

Exercise 2.4. For the matrices A in Exercise 1.12, express $A\vec{x}$ as linear combinations of column vectors.

Exercise 2.5. Interpret systems of linear equations in Exercises 1.2, 1.3, 1.7 as expressing vectors as linear combinations of some other vectors.

Exercise 2.6. Can you express the given vector $\vec{b}$ as a linear combination of vectors $\vec{v}_i$? Is the expression unique?

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

Exercise 2.7. Can you express the given vector $\vec{b}$ as a linear combination of vectors $\vec{v}_i$? Is the expression unique?

1. $\vec{b}=(10,11,12)$, $\vec{v}_1=(1,2,3)$, $\vec{v}_2=(4,5,6)$, $\vec{v}_3=(7,8,9)$.
2. $\vec{b}=(12,11,10)$, $\vec{v}_1=(1,2,3)$, $\vec{v}_2=(4,5,6)$, $\vec{v}_3=(7,8,9)$.
3. $\vec{b}=(10,12,11)$, $\vec{v}_1=(1,2,3)$, $\vec{v}_2=(4,5,6)$, $\vec{v}_3=(7,8,9)$.
4. $\vec{b}=(10,11,12)$, $\vec{v}_1=(7,8,9)$, $\vec{v}_2=(4,5,6)$, $\vec{v}_3=(1,2,3)$.
5. $\vec{b}=(10,11,12)$, $\vec{v}_1=(1,2,3)$, $\vec{v}_2=(4,5,6)$.
6. $\vec{b}=(10,12,11)$, $\vec{v}_1=(1,2,3)$, $\vec{v}_2=(4,5,6)$.
7. $\vec{b}=(10,11,12)$, $\vec{v}_1=(1,2,3)$, $\vec{v}_2=(4,5,6)$, $\vec{v}_3=(7,8,0)$.

Exercise 2.8. Find the exact condition that $\vec{b}=(b_1,b_2,b_3)$ is a linear combination of $\vec{v}_1=(1,2,3)$, $\vec{v}_2=(4,5,6)$, $\vec{v}_3=(7,8,9)$.

2.2 Span and Linear Independence

2.2.1 Existence is Span

Consider an $m\times n$ matrix

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

The system $A\vec{x}=\vec{b}$ has solution means that $\vec{b}$ can be written as $A\vec{x}$. Therefore we collect all the vectors in $\mathbb{R}^m$ of the form $A\vec{x}$, and get the *column space* of A

$$\text{Col}A=\{A\vec{x}\colon \vec{x}\in\mathbb{R}^n\}\subset\mathbb{R}^m.$$

By default, the column space measures the existence of solution.

Proposition 2.2.1. *A system of linear equations $A\vec{x}=\vec{b}$ has solution if and only if the right side $\vec{b}\in\text{Col}A$.*

For example, $A\vec{x}=\vec{b}$ has solution for all the right side $\vec{b}$ means $\text{Col}A=\mathbb{R}^m$.

By $A\vec{x}=x_1\vec{v}_1+x_2\vec{v}_2+\cdots+x_n\vec{v}_n$, the column space is all the linear combinations of the column vectors.

Definition 2.2.2. The *span* of a vector set $\alpha=\{\vec{v}_1,\vec{v}_2,\dots,\vec{v}_n\}$ is all the linear combinations of α

$$\text{Span}\alpha=\mathbb{R}\vec{v}_1+\mathbb{R}\vec{v}_2+\cdots+\mathbb{R}\vec{v}_n=\{x_1\vec{v}_1+x_2\vec{v}_2+\cdots+x_n\vec{v}_n\colon x_i\in\mathbb{R}\}.$$

The span $\mathbb{R}\vec{v}$ of a single nonzero vector $\vec{v}$ is the straight line passing through the origin and in the direction $\vec{v}$. Figure 2.1.2 shows that the span $\mathbb{R}\vec{u}+\mathbb{R}\vec{v}$ of two non-parallel vectors $\vec{u}$ and $\vec{v}$ is an infinite plane passing through the origin.

Exercise 2.9. For $\alpha=\{\vec{v}_1,\vec{v}_2,\dots,\vec{v}_n\}$, explain $\vec{v}_i\in\text{Span}\alpha$.

Exercise 2.10. When does the span consist of single zero vector $\vec{0}$?

Exercise 2.11. What is the span of parallel and nonzero vectors $\vec{v}$ and $a\vec{v}$?

Exercise 2.12. Show that $\mathbb{R}\vec{u}+\mathbb{R}\vec{v}+\mathbb{R}(a\vec{u}+b\vec{v})=\mathbb{R}\vec{u}+\mathbb{R}\vec{v}$. More generally, we have

$$\vec{w}\in\text{Span}\alpha \implies \text{Span}(\alpha\cup\{\vec{w}\})=\text{Span}\alpha.$$

The converse is also true.

Example 2.2.1. The span of the standard basis vectors $\vec{e}_1,\vec{e}_2,\dots,\vec{e}_n$ in Example 2.1.1 is the whole $\mathbb{R}^n$.

Example 2.2.2. 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 Example 1.2.1, 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 1.3.1, if $a \neq 12$, then $A\vec{x} = \vec{b}$ has solution for all right side $\vec{b}$. This means $\mathrm{Col}A = \mathbb{R}^3$, 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

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

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

Exercise 2.13. Determine whether $\mathrm{Col}A$ is the whole Euclidean space, and whether the vector $\vec{v}$ is in $\mathrm{Col}A$.

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

Exercise 2.14. Determine whether column vectors span the whole space. Do the row vectors span the whole space?

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

2.2.2 Uniqueness is Linear Independence

The existence of solution of systems of linear equations correspond to the span. The uniqueness of the solution means and two solutions $\vec{x}$ and $\vec{y}$ are equal

$$A\vec{x} = \vec{b} \text{ and } A\vec{y} = \vec{b} \implies \vec{x} = \vec{y}.$$

The left side is the same as $A\vec{x} = A\vec{y}$. Writing $A\vec{x}$ and $A\vec{y}$ as linear combinations of the column vectors, with $\vec{x}$ and $\vec{y}$ as coefficients, the uniqueness corresponds to the following.

Definition 2.2.3. Vectors $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ are *linearly independent* if

$$\begin{aligned} & x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n = y_1\vec{v}_1 + y_2\vec{v}_2 + \cdots + y_n\vec{v}_n \\ \implies & x_1 = y_1,\ x_2 = y_2,\ \dots,\ x_n = y_n. \end{aligned}$$

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

Since the linear independence is equivalent to the uniqueness of solution, by Theorem 1.3.2, the calculational criterion for the linear independence of $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ is the same as that all columns of the matrix $(\vec{v}_1\ \vec{v}_2\ \cdots\ \vec{v}_n)$ are pivot.

The following criterion for linear independence is the special case $y_1 = y_2 = \cdots = y_n = 0$ of the definition, and is also the second statement in Theorem 1.3.2.

Proposition 2.2.4. *Vectors $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ are linearly independent if and only if*

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

If the criterion is satisfied, then the following proves the linear independence

$$\begin{aligned} & x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n = y_1\vec{v}_1 + y_2\vec{v}_2 + \cdots + y_n\vec{v}_n \\ \implies & (x_1 - y_1)\vec{v}_1 + (x_2 - y_2)\vec{v}_2 + \cdots + (x_2 - y_2)\vec{v}_n = \vec{0}, \\ \implies & x_1 - y_1 = x_1 - y_1 = \cdots = x_n - y_n = 0 \\ \implies & x_1 = y_1, x_2 = y_2, \dots, x_n = y_n. \end{aligned}$$

The criterion is used in the second $\implies$.

The following is the criteria for linear dependence.

Proposition 2.2.5. *The following are equivalent for vectors $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$.*

1. *The vectors are linearly dependent.*

2. *It is possible to have $x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n = \vec{0}$, with some $x_i \neq 0$.*

3. *One vector is a linear combination of the others.*

The equivalence between the first two statements is Proposition 2.2.4. If the third statement holds, then for some i, we have

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

This is the same as

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

where $x_i = -1 \neq 0$. Therefore the second statement holds. Conversely, if the second statement holds, then we let x_i be the last nonzero coefficient and get

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

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

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

and the third statement holds.

The proof shows that, in the third statement, we can always find a vector that can be expressed as a linear combination of *previous* vectors

$$\vec{v}_i \in \mathbb{R}\vec{v}_1 + \mathbb{R}\vec{v}_2 + \cdots + \mathbb{R}\vec{v}_{i-1} \text{ for some } i.$$

Example 2.2.3. By Proposition 2.2.4, the linear independence of a single vector $\vec{v}$ means

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

This means exactly $\vec{v} \neq \vec{0}$.

By Proposition 2.2.5, the linear dependence of two vectors means one vector is a linear combination of the other. Since the linear combination of one vector is simply a scalar multiplication, this means exactly the two vectors are parallel. Therefore two vectors are linearly independent if and only if they are not parallel.

Exercise 2.15. Suppose $\alpha \subset \beta$. Explain the following. (You may assume $\alpha = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$, and $\beta = \{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4, \vec{v}_5\}$)

1. If α spans $\mathbb{R}^m$, then β spans $\mathbb{R}^m$.
2. If β is linearly independent, then α is linearly independent.

Then explain that, if α contains $\vec{0}$, then α is linearly dependent.

Exercise 2.16. Find three vectors, such that any two of the three are linearly independent, and all three are linearly dependent.

Example 2.2.4. The vectors $\vec{v}_1 = (1,2,3), \vec{v}_2 = (4,5,6), \vec{v}_3 = (7,8,9)$ are linearly dependent because $\vec{v}_1 - 2\vec{v}_2 + \vec{v}_3 = \vec{0}$ (second statement of Proposition 2.2.5). This the same as $\vec{v}_3 = -\vec{v}_1 + 2\vec{v}_2$ (third statement of Proposition 2.2.5).

Example 2.2.5. To determine the linear independence of three vectors $\vec{v}_1 = (1,2,3,4)$, $\vec{v}_2 = (5,6,7,8)$, $\vec{v}_3 = (9,10,11,a)$ in $\mathbb{R}^4$, we use the vectors as the columns of a matrix, and do row operations

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

The linear independence means the uniqueness of solution. The criterion for uniqueness is that all columns of A are pivot. Therefore $\vec{v}_1, \vec{v}_2, \vec{v}_3$ are linearly independent if and only if $a \neq 12$.

We remark that, if $a = 12$, then we regard $(\vec{v}_1\ \vec{v}_2\ \vec{v}_3)$ as the augmented matrix with $B = (\vec{v}_1\ \vec{v}_2)$ as the coefficient matrix, and with $\vec{v}_3$ as the right side. Then the row operation

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

can be interpreted as that the linear system

$$B\vec{x} = x_1\vec{v}_1 + x_2\vec{v}_2 = \vec{v}_3$$

has solution. In other words, $\vec{v}_3$ is a linear combination of $\vec{v}_1$ and $\vec{v}_2$, and the third statement in Proposition 2.2.5 holds.

Example 2.2.6. The four vectors in $\mathbb{R}^3$ in Example 2.2.2 must be linearly dependent, because by Example 1.4.3, the solution of the corresponding $A\vec{x} = \vec{b}$ is not unique. This is also confirmed by the row operations in Examples 1.2.1 and 1.3.4

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

where the third column is not pivot. In fact, by the same row operations, we get

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

By regarding $(\vec{v}_1\ \vec{v}_2\ \vec{v}_3)$ as the augmented matrix of a system of three equations in two variables, this shows $\vec{v}_3$ is a linear combination of $\vec{v}_1$ and $\vec{v}_2$. By Proposition 2.2.5, this implies the four vectors are linearly dependent. In fact, this also implies $\vec{v}_1, \vec{v}_2, \vec{v}_3$ are linearly dependent.

By the same row operations, we get

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

This shows that $\vec{v}_1, \vec{v}_2, \vec{v}_4$ are linearly independent if and only if $a \neq 12$. Since the four vectors are always linearly dependent, we know $\{\vec{v}_1, \vec{v}_2, \vec{v}_4\}$ is maximal linearly independent subset of $\{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4\}$.

Exercise 2.17. Determine the linear independence of the columns and rows of the matrices in Exercise 2.14.

Exercise 2.18. Find the conditions on a, b, such that the vectors are linearly independent.

1. $(1,2), (3,a)$.
2. $(1,2), (a,b)$.
3. $(1,1,0)$, $(1,0,1)$, $(0,a,b)$.
4. $(1,2,3)$, $(3,2,1)$, $(a,1,b)$, $(1,a,2)$.
5. $(1,1,-2,2)$, $(1,-2,1,2)$, $(-2,a,b,1)$.
6. $(1,1,-2,2)$, $(1,-2,1,2)$, $(-2,a,b,1)$, $(1,1,1,-3)$.

2.3 Subspace of $\mathbb{R}^m$

2.3.1 Subspace, Column Space, Null Space

The span and the solution set of a system of linear equations are either single point or an infinite and flat subset of the Euclidean space. The span always contains the zero vector $\vec{0}$, and the solution set may not contain the origin $\vec{0}$ (if the system is not homogeneous). The span is a subspace H of the Euclidean space, and the solution set is the shifting $\vec{x}_0 + H$ of a subspace.

Definition 2.3.1. A subset $H \subset \mathbb{R}^m$ is a *subspace* if

$$\vec{x}, \vec{y} \in H,\ a, b \in \mathbb{R} \implies a\vec{x} + b\vec{y} \in H.$$

In Proposition 2.1.2, we have the equality

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

Then for any two vectors $A\vec{x}, A\vec{y} \in \text{Col}A$, their linear combination $aA\vec{x} + bA\vec{y} = A(a\vec{x} + b\vec{y})$ is still in $\text{Col}A$. Therefore $\text{Col}A$ is a subspace of $\mathbb{R}^m$. Correspondingly, the span of vectors is a subspace.

On the other hand, we have

$$A\vec{x} = A\vec{y} = \vec{0} \implies A(a\vec{x} + b\vec{y}) = aA\vec{x} + bA\vec{y} = a\vec{0} + b\vec{0} = \vec{0}$$

This shows that the *null space*

$$\text{Nul}A = \{\vec{x} \in \mathbb{R}^n : A\vec{x} = \vec{0}\} \subset \mathbb{R}^n$$

is also a subspace.

Any $m \times n$ matrix A is associated with four subspaces

$$\text{Col}A \subset \mathbb{R}^m, \quad \text{Nul}A \subset \mathbb{R}^n, \quad \text{Col}A^T \subset \mathbb{R}^n, \quad \text{Nul}A^T \subset \mathbb{R}^m.$$

Since $\text{Col}A^T$ is the span of the rows of A, we also call it the *row space* and denote $\text{Row}A$.

Example 2.3.1. A subspace contains the origin $0\vec{x} + 0\vec{y} = \vec{0}$. For example, since the straight line $\{(x, y) : x + 2y = 3\}$ misses the origin, it is not a subspace of $\mathbb{R}^2$.

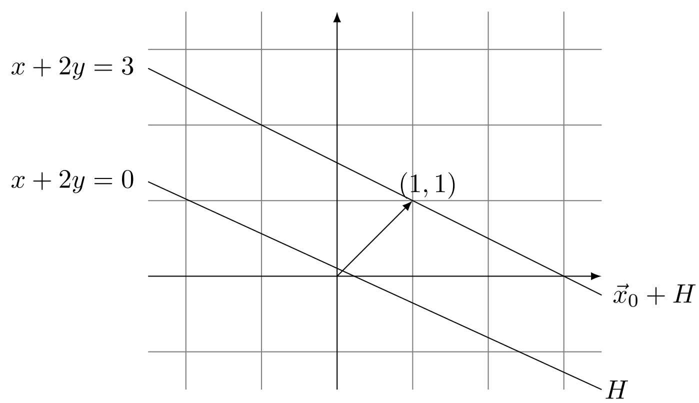

Figure 2.3.1: Non-homogeneous line is the shift of homogeneous line.

On the other hand, the straight line $H = \{(x, y) : x + 2y = 0\}$ passes through the origin and is a subspace of $\mathbb{R}^2$

$$\begin{aligned}(x_1, y_1), (x_2, y_2) \in H &\implies x_1 + 2y_1 = 0 \text{ and } x_2 + 2y_2 = 0 \\ &\implies (ax_1 + bx_2) + 2(ay_1 + by_2) = ax_1 + 2ay_1 + bx_2 + 2by_2 \\ &\qquad = a(x_1 + 2y_1) + b(x_2 + 2y_2) = 0 \\ &\implies a(x_1, y_1) + b(x_2, y_2) = (ax_1 + bx_2, ay_1 + by_2) \in H.\end{aligned}$$

The conceptual way of arguing that H is a subspace is that H is the null space of the 1×2 matrix $(1\ 2)$.

Example 2.3.2. The unit disk $H = \{(x, y) : x^2 + y^2 \leq 1\}$ contains the origin. However, we have $(1, 0), (0, 1) \in H$ but $(1, 0) + (0, 1) = (1, 1) \notin H$. In fact, we also have $(1, 0) \in H$ but $2(1, 0) = (2, 0) \notin H$. Either is the reason for H not to be a subspace of $\mathbb{R}^2$.

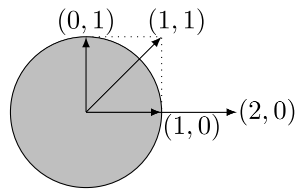

Figure 2.3.2: Unit disk is not subspace.

Exercise 2.19. Determine whether the subsets are subspaces of Euclidean spaces.

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

Exercise 2.20. Suppose H_1, H_2 are subspaces of $\mathbb{R}^m$, show that the intersection $H_1 \cap H_2$ is also a subspace of $\mathbb{R}^m$. However, $H_1 \cup H_2$ is generally not a subspace.

2.3.2 General Solution is Shift of Null Space

Suppose the system $A\vec{x} = \vec{b}$ has a solution $\vec{x}_0$. Then for any solution $\vec{x}$, we have

$$A(\vec{x} - \vec{x}_0) = A\vec{x} - A\vec{x}_0 = \vec{b} - \vec{b} = \vec{0}.$$

This means $\vec{x} - \vec{x}_0 \in \text{Nul}A$, or $\vec{x} \in \vec{x}_0 + \text{Nul}A$. Therefore the general solution of $A\vec{x} = \vec{b}$ is the shifting of the null space $\text{Nul}A$ by one solution $\vec{x}_0$.

In Figure 2.3.1, we see the general solution of $x + 2y = 3$ is the shifting of the solutions of the homogeneous space (i.e., null space) $x + 2y = 0$ by a solution $\vec{x}_0 = (1, 1)$.

Example 2.3.3. Consider the system $A\vec{x} = \vec{b}$ in Example 1.1.2, given by

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

The general solution is

$$x_1 = -2 + x_3, \quad x_2 = 3 - 2x_3, \quad x_3 \in \mathbb{R}.$$

We express the solution in vector form (this is done in Example 2.1.3)

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -2 + x_3 \\ 3 - 2x_3 \\ x_3 \end{pmatrix} = \begin{pmatrix} -2 \\ 3 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}.$$

The general solution consists of two parts. The first part $\vec{x}_0 = (-2,3,0)$ is one solution satisfying $A\vec{x}_0 = \vec{b}$ and $x_3 = 0$. We have

$$\text{Nul}A = \{x_3 \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} : x_3 \in \mathbb{R}\} = \mathbb{R} \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix},$$

and the general solution of $A\vec{x} = \vec{b}$ is $(2,3,0) + \mathbb{R}(1,-2,1)$.

Example 2.3.4. Suppose the augmented matrix of a system $A\vec{x} = \vec{b}$ (of 3 equations in 5 variables) has reduced row echelon form

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

In Example 1.3.1, 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 $\vec{x}_0$

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

Note that vectors in H correspond to solutions of the homogeneous system, with $d_1 = d_2 = d_3 = 0$. Therefore $H = \text{Nul}A$ is the null space.

Proposition 2.3.2. *Let $\vec{x}_0$ be one solution of $A\vec{x} = \vec{b}$. Then all the solutions of the system is $\vec{x}_0 + \text{Nul}A$.*

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

$$\vec{x} = \vec{x}_0 + y_1\vec{v}_1 + y_2\vec{v}_2 + \cdots + y_k\vec{v}_k, \tag{2.3.1}$$

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 solution, and

$$\begin{aligned} \text{Nul}A &= \{y_1\vec{v}_1 + y_2\vec{v}_2 + \cdots + y_k\vec{v}_k \colon y_i \in \mathbb{R}\} \\ &= \mathbb{R}\vec{v}_1 + \mathbb{R}\vec{v}_2 + \cdots + \mathbb{R}\vec{v}_k. \end{aligned} \tag{2.3.2}$$

The null space Nul$A$ represents the changes between various solutions of $A\vec{x} = \vec{b}$. Since uniqueness means no change, we get the following, which is actually the second part of Theorem 1.3.2.

Proposition 2.3.3. *The solution of $A\vec{x} = \vec{b}$ is unique if and only if* $\mathrm{Nul}A = \{\vec{0}\}$.

Exercise 2.21. Determine whether the vector $\vec{v}$ is in Nul$A$.

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

Exercise 2.22. Determine the condition for vector $\vec{v}$ to be in Nul$A$.

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

Exercise 2.23. For the reduced row echelon forms in Exercise 1.21 of augmented matrix, express the general solution as $\vec{x}_0 + H$, and express H as a span.

Exercise 2.24. For the general solution in Exercise 1.22, express the general solution as $\vec{x}_0 + H$, and express H as a span.

Exercise 2.25. 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.

2.4 Basis

2.4.1 Minimal Span and Maximal Independence

We have the intuition that a point, line, plane, have respective 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, a subspace H of $\mathbb{R}^m$ is spanned by a collection of vectors $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$, which provides all the directions of H. However, we cannot simply define $\dim H = n$, the number of vectors in α, because some directions are superfluous. For example, we should remove the zero vector from α because it does not contribute to directions. For another example, our usual ground plane can be covered by the front, right, and diagonal directions. However, the dimension of the ground plane is not 3, because front and right 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 1.4. This means the second property in the theorem below.

Theorem 2.4.1. *For a set of vectors α in a subspace H of $\mathbb{R}^n$, the following are equivalent:*

1. *α spans H and is linearly independent.*
2. *α is a minimal spanning set of H.*
3. *α is a maximal linearly independent set of H.*

We define the collection in the theorem a *basis* of H, and the *dimension* $\dim H$ of H is the number of vectors in a basis.

By minimal spanning set, we mean α spans H, and any subset strict smaller than α no longer spans H. In other words, after removing any vector $\vec{v} \in \alpha$, the set $\alpha' = \alpha - \{\vec{v}\}$ does not span H. Exercise 2.12 says that $\vec{v} \in \text{Span}\alpha'$ if and only if α' and $\alpha = \alpha' \cup \{\vec{v}\}$ have the same span, i.e., α' spans $H = \text{Span}\alpha$. Therefore α' does not span H if and only if $\vec{v}$ is not a linear combination of α', the other vectors in α. By Proposition 2.2.5, this means α is linearly independent.

Consider the set α as all the people in a company, and the linear combination as the collaboration work done by these people. Then α spanning H means the company can achieve all the targets in H. Minimal spanning set means removing any one person from the company will not achieve all the targets. Therefore the company is the most efficient in terms of achieve all the targets in H.

By maximal independent set, we mean α is a linearly independent set inside H, and any set strictly larger than α is linearly dependent. In other words, for any $\vec{v} \in H$, the set $\alpha' = \alpha \cup \{\vec{v}\}$ is linearly dependent. Since α is linearly independent, any vector in α is not a linear combination of the other vectors in α. Then by considering $\vec{v}$ as the last vector in $\alpha' = \alpha \cup \{\vec{v}\}$, and using Proposition 2.2.5 and the remark before Example 4.2.4, we know the linear dependence of α' means $\vec{v}$ is a linear combination of α. Since $\vec{v}$ can be any vector in H, this means α spans H.

Again consider vectors in α as people. Linear independence means no vector is a linear combination of the others. In other words, there is some work done by this person that cannot be done by the others in the company, i.e., the person is necessary for company to achieve H. The maximal independence means the contribution by any new hiring can already be done by existing people. Therefore there is no need to hire more people, and company is already sufficient to achieve all the targets, i.e., spanning H.

A basis is the analogue of a company with the best combination of people: can do all the things, and every people is necessary.

2.4.2 Basis of Euclidean Space

Example 2.4.1. We have

$$(x_1, x_2, x_3) = x_1(1,0,0) + x_2(0,1,0) + x_3(0,0,1) = x_1\vec{e}_1 + x_2\vec{e}_2 + x_3\vec{e}_3.$$

The equality shows that the standard basis vectors $\vec{e}_1, \vec{e}_2, \vec{e}_3$ in Example 2.1.1 span $\mathbb{R}^3$. It also shows that the vectors are linearly independent

$$x_1\vec{e}_1 + x_2\vec{e}_2 + x_3\vec{e}_3 = \vec{0} \implies (x_1, x_2, x_3) = \vec{0} \implies x_1 = x_2 = x_3 = 0.$$

Therefore $\vec{e}_1, \vec{e}_2, \vec{e}_3$ form a basis of $\mathbb{R}^3$, and $\dim \mathbb{R}^3 = 3$.

In general, the standard basis of $\mathbb{R}^n$ is $\epsilon = \{\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n\}$.

Example 2.4.2. To find when the columns of a matrix form a basis of $\mathbb{R}^3$, we use row operations (see Example 1.3.4)

$$A = \begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & a \end{pmatrix} \to \begin{pmatrix} 1 & 4 & 7 \\ 0 & -3 & -6 \\ 0 & 0 & a-9 \end{pmatrix}.$$

The columns span $\mathbb{R}^3$ is and only if $A\vec{x} = \vec{b}$ has solution for all $\vec{b}$. By Theorem 1.3.3, this means all rows of A are pivot. The condition is $a \neq 9$.

The columns are linearly independent if and only if the solution of $A\vec{x} = \vec{b}$ is unique. By Theorem 1.3.2, this means all columns of A are pivot. The condition is also $a \neq 9$.

In fact, by Theorem 1.4.4, for the square matrix A, columns spanning $\mathbb{R}^3$ is equivalent to columns being linearly independent.

Therefore columns of A form a basis of $\mathbb{R}^3$ if and only if $a \neq 9$.

In general, the columns of a matrix A form a basis of the Euclidean space if and only if the matrix is square, and all columns (or all rows) are pivot.

Exercise 2.26. Explain that the columns of an $m \times n$ matrix A is a basis of $\mathbb{R}^m$ if and only if the rows of A is a basis of $\mathbb{R}^n$.

Exercise 2.27. Determine whether the vectors form a basis of the Euclidean space.

1. $(0,1), (2,3)$.

2. $(0,2), (1,3)$.

3. $(0,1,2),(2,1,0),(1,1,1)$.
4. $(0,1,1),(1,1,0),(1,1,1)$.
5. $(0,1,2),(2,1,0),(0,1,1),(1,1,0)$.
6. $(0,1,0,2),(2,0,1,0),(1,1,1,1)$.
7. $(0,1,0,2),(2,0,1,0),(1,1,1,1),(1,0,0,1)$.
8. $(0,1,0,2),(2,0,1,0),(1,1,1,1),(1,0,0,2)$.
9. $(0,1,0,2),(2,0,1,0),(1,1,1,1),(0,1,2,0)$.

Exercise 2.28. Determine the condition on a, such that the vectors form a basis of the Euclidean space.

1. $(0,1),(2,a)$.
2. $(a,2),(1,3)$.
3. $(0,1,2),(2,1,0),(a,a,a)$.
4. $(0,1,1),(1,1,0),(a,a,a)$.
5. $(0,1,2),(2,1,0),(0,1,a),(a,1,0)$.
6. $(0,a,0,2),(2,0,a,0),(1,1,1,1)$.
7. $(0,1,0,2),(2,0,1,0),(1,1,1,1),(a,0,0,a^2)$.
8. $(0,1,0,2),(2,0,1,0),(1,1,1,1),(1,0,0,a)$.
9. $(0,1,0,2),(2,0,1,0),(1,1,1,1),(0,1,a,0)$.

2.4.3 Basis of Column Space: First Method

Example 2.4.3. Consider the columns space

$$A = \begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{pmatrix} = (\vec{v}_1\ \vec{v}_2\ \vec{v}_3\ \vec{v}_4), \quad H = \mathrm{Col}A = \mathbb{R}\vec{v}_1 + \mathbb{R}\vec{v}_2 + \mathbb{R}\vec{v}_3 + \mathbb{R}\vec{v}_4.$$

To find a basis, we consider the row operations in Example 1.2.1

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

We regard $(\vec{v}_1\ \vec{v}_2\ \vec{v}_3)$ as the augmented matrix of the system $y_1\vec{v}_1 + y_2\vec{v}_2 = \vec{v}_3$, and restrict the row operations to the first three columns. 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 $(\vec{v}_1\ \vec{v}_2\ \vec{v}_4)$ as the augmented matrix of the system $z_1\vec{v}_1 + z_2\vec{v}_2 = \vec{v}_4$, and similar reason shows that $\vec{v}_4$ is also a linear combination of $\vec{v}_1, \vec{v}_2$. Then all the vectors in H are linear combinations of $\vec{v}_1$ and $\vec{v}_2$ (see Exercise 2.12)

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

Therefore $H = \mathbb{R}\vec{v}_1 + \mathbb{R}\vec{v}_2$ is spanned by $\vec{v}_1$ and $\vec{v}_2$.

If we restrict the row operations to the matrix $(\vec{v}_1\ \vec{v}_2)$, then we see both columns are pivot. Therefore $\vec{v}_1, \vec{v}_2$ are linearly independent.

We conclude $\{\vec{v}_1, \vec{v}_2\}$ is a basis of H, and $\dim H = 2$.

In general, a basis of the column space

$$\mathrm{Col}A = \mathbb{R}\vec{v}_1 + \mathbb{R}\vec{v}_2 + \cdots + \mathbb{R}\vec{v}_n, \quad A = (\vec{v}_1\ \vec{v}_2\ \cdots\ \vec{v}_n),$$

is given by the pivot columns of A. Therefore we have the following.

Proposition 2.4.2. $\dim \mathrm{Col}A = \mathrm{rank}A$.

Due to the proposition, we also define the *rank of vector set* $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ by

$$\mathrm{rank}\alpha = \dim \mathrm{Span}\alpha = \dim(\mathbb{R}\vec{v}_1 + \mathbb{R}\vec{v}_2 + \cdots + \mathbb{R}\vec{v}_n).$$

An $m \times n$ matrix A correspond to n vectors in $\mathbb{R}^m$. Then the inequality $\mathrm{rank}A \le \min\{m, n\}$ becomes $\mathrm{rank}\alpha \le \min\{m, n\}$.

Exercise 2.29. What is the subspace of dimension 0?

Exercise 2.30. Explain that α is a basis of $\mathrm{Span}\alpha$ if and only if α is linearly independent.

Exercise 2.31. Find bases and dimensions of the spans.

1. $\mathbb{R}(1, 2) + \mathbb{R}(2, 1)$.
2. $\mathbb{R}(1, 2) + \mathbb{R}(-1, -2)$.
3. $\mathbb{R}(1, 1, 0) + \mathbb{R}(1, 0, 1) + \mathbb{R}(0, 1, 1)$.
4. $\mathbb{R}(1, 1, -1) + \mathbb{R}(1, -1, 1) + \mathbb{R}(-1, 1, 1)$.
5. $\mathbb{R}(1, 1, -2) + \mathbb{R}(1, -2, 1) + \mathbb{R}(-2, 1, 1)$.
6. $\mathbb{R}(1, 1, 1, -2) + \mathbb{R}(1, 1, -2, 1) + \mathbb{R}(1, -2, 1, 1) + \mathbb{R}(-2, 1, 1, 1)$.
7. $\mathbb{R}(1, 1, 1, -3) + \mathbb{R}(1, 1, -3, 1) + \mathbb{R}(1, -3, 1, 1) + \mathbb{R}(-3, 1, 1, 1)$.

8. $\mathbb{R}(1,1,1)+\mathbb{R}(2,2,2)+\mathbb{R}(3,3,3)+\mathbb{R}(4,4,4)$.

9. $\mathbb{R}(1,-1,0,0)+\mathbb{R}(1,0,-1,0)+\mathbb{R}(0,1,-1,0)+\mathbb{R}(1,0,0,-1)+\mathbb{R}(0,1,0,-1)+\mathbb{R}(0,0,1,-1)$.

Exercise 2.32. Find bases of column spaces.

1. $\begin{pmatrix} 0 & 2 & 6 \\ 2 & 2 & 16 \\ 1 & 0 & 5 \end{pmatrix}$

2. $\begin{pmatrix} 2 & 1 & -7 \\ 2 & 3 & -5 \\ 1 & 2 & 4 \end{pmatrix}$

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

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

5. $\begin{pmatrix} 1 & 6 & 3 & -4 \\ -3 & 2 & -2 & -8 \\ 4 & -1 & 3 & 9 \end{pmatrix}$

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

7. $\begin{pmatrix} 1 & 4 & 0 & 2 & -1 \\ 5 & 20 & 2 & 8 & 8 \\ 3 & 12 & 1 & 5 & 5 \\ 2 & 8 & 1 & 3 & 2 \end{pmatrix}$

8. $\begin{pmatrix} 1 & 2 & 1 & 3 \\ 2 & 4 & 2 & 6 \\ 3 & 3 & 0 & 3 \\ 11 & 15 & 4 & 19 \\ -3 & 2 & 5 & -2 \end{pmatrix}$

2.4.4 Property of Dimension

We need to justify the definition of dimension by showing that two bases α and β of the same subspace H have the same number of vectors. In fact, we will prove that, if α and β are bases of H' and H, and $H' \subset H$, then $\#\alpha \le \#\beta$. This means the following and justifies the definition of dimension.

Theorem 2.4.3. *If $H' \subset H$, then $\dim H' \le \dim H$. Moreover, if $\dim H' = \dim H$, then $H' = H$.*

Let $H' = \text{Col}A$ and $H = \text{Col}B$. By the discussion in Section 2.4.3, we know $\dim H' = \text{rank}A$ and $\dim H = \text{rank}B$. Therefore the theorem becomes

$$\text{Col}A \subset \text{Col}B \implies \text{rank}A \le \text{rank}B.$$

We consider the special case A and B are 4×5 matrices, and $\text{rank}B = 3$. Then the row echelon form of B has three pivots. For example, we assume row operations

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

Applying the same row operations to $A = (\vec{v}_1\ \vec{v}_2\ \vec{v}_3\ \vec{v}_4)$, we get

$$(B\ A) = (B\ \vec{v}_1\ \vec{v}_2\ \vec{v}_3\ \vec{v}_4) \to \begin{pmatrix} \bullet & * & * & * & * & * & * & * \\ 0 & \bullet & * & * & * & * & * & * \\ 0 & 0 & 0 & \bullet & * & * & * & * \\ 0 & 0 & 0 & 0 & a_1 & a_2 & a_3 & a_4 \\ 0 & 0 & 0 & 0 & b_1 & b_2 & b_3 & b_4 \end{pmatrix}.$$

The row operations restrict to

$$(B\ \vec{v}_1) \to \begin{pmatrix} \bullet & * & * & * & * \\ 0 & \bullet & * & * & * \\ 0 & 0 & 0 & \bullet & * \\ 0 & 0 & 0 & 0 & a_1 \\ 0 & 0 & 0 & 0 & b_1 \end{pmatrix}.$$

Since $\vec{v}_1 \in \text{Col}A \subset \text{Col}B$, we know $B\vec{x} = \vec{v}_1$ has solution. This implies $a_1 = b_1 = 0$.

By similar reason, we get all $a_i = b_i = 0$. Then the restriction of the row operations to A become

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

This implies A has at most three pivots. Therefore $\text{rank}A \le 3 = \text{rank}B$. This proves the first part of Theorem 2.4.3.

The second part is a consequence of the following.

Theorem 2.4.4. *Suppose α is a set of vectors in a subspace H.*

1. *If α spans H, then $\#\alpha \ge \dim H$.*
2. *If α is linearly independent, then $\#\alpha \le \dim H$.*

Moreover, in both cases, $\#\alpha = \dim H$ implies α is a basis of H.

For the case $H = \mathbb{R}^m$, and α is the column vectors of $m \times n$-matrix A. Then $\dim H = m$, and $n = \#\alpha$, and the theorem is Theorems 1.4.3 and 1.4.4. In particular, if α is n vectors in $\mathbb{R}^n$, then α is a basis of $\mathbb{R}^n$ if and only if α spans $\mathbb{R}^n$, and if and only if α is linearly independent.

Theorem 2.4.4 is a consequence of Theorem 2.4.1. A spanning set α of H can be reduced to a minimal spanning set α' of H. Then α' is a basis of H, and

$$\alpha \supset \alpha' \implies \#\alpha \ge \#\alpha' = \dim H.$$

On the other hand, a linearly independent α in H can be increased to a maximal linearly independent set α' in H. Then α' is a basis of H, and

$$\alpha \subset \alpha' \implies \#\alpha \le \#\alpha' = \dim H.$$

If $\#\alpha = \dim H = \#\alpha'$, then in both cases, the inclusion $\alpha \supset \alpha'$ or $\alpha \subset \alpha'$ implies $\alpha = \alpha'$ is a basis of H.

Now we can explain the second part of Theorem 2.4.3. Suppose $H' \subset H$ and $\dim H' = \dim H$. Let α be a basis of H'. Then α is a linearly independent set in H satisfying $\#\alpha = \dim H' = \dim H$. By the last sentence in Theorem 2.4.4, we know α is a basis of H. Then α spans both H' and H. This implies $H' = H$.

2.4.5 Basis of Null Space

Example 2.4.4. Example 2.4.1 calculates a basis of the column space Col$A$. We calculated the null space in and after Example 2.3.3. Specifically, 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 the general solution

$$\vec{x} = x_2 \begin{pmatrix} -c_{12} \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} -c_{14} \\ 0 \\ -c_{24} \\ 1 \\ 0 \end{pmatrix} = x_2\vec{v}_1 + x_4\vec{v}_2, \quad \text{Nul}A = \mathbb{R}\vec{v}_1 + \mathbb{R}\vec{v}_2.$$

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 and form a basis of Nul$A$.

As a concrete example, the reduced row echelon form

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

shows that Nul$A$ has a basis given by $(1, -2, 1, 0)$ and $(2, -3, 0, 1)$.

Example 2.4.4 clearly shows that the dimension of the null space of $m \times n$ matrix A is the number of free variables. Since free variables correspond to non-pivot columns (see the correspondence after Theorem 1.3.1), and the total number of variables is the number n of columns of matrix A, we get the following.

Proposition 2.4.5. $\dim \mathrm{Nul}A = n - \mathrm{rank}A$.

By Proposition 2.4.2, we have $\dim \mathrm{Col}A = \mathrm{rank}A$. Therefore

$$\dim \mathrm{Col}A + \dim \mathrm{Nul}A = n.$$

Exercise 2.33. Find bases of the null spaces of the matrices in Exercise 2.32.

2.4.6 Basis of Column Space: Second Method

While the columns of A span a subspace $\mathrm{Col}A \subset \mathbb{R}^m$, the rows of A also span a subspace $\mathrm{Row}A \subset \mathbb{R}^n$, called *row space*. For example, if

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

then

$$\mathrm{Col}A = \mathbb{R}\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + \mathbb{R}\begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} + \mathbb{R}\begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix} + \mathbb{R}\begin{pmatrix} 10 \\ 11 \\ 12 \end{pmatrix} \subset \mathbb{R}^3,$$
$$\mathrm{Row}A = \mathbb{R}\begin{pmatrix} 1 \\ 4 \\ 7 \\ 10 \end{pmatrix} + \mathbb{R}\begin{pmatrix} 2 \\ 5 \\ 8 \\ 11 \end{pmatrix} + \mathbb{R}\begin{pmatrix} 3 \\ 6 \\ 9 \\ 12 \end{pmatrix} \subset \mathbb{R}^4.$$

Denote the *transpose* A^T of matrix by exchanging rows and columns

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

Then we have

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

In fact, we may introduce four subspaces

$$\mathrm{Col}A \subset \mathbb{R}^m, \quad \mathrm{Row}A \subset \mathbb{R}^n, \quad \mathrm{Nul}A \subset \mathbb{R}^n, \quad \mathrm{Nul}A^T \subset \mathbb{R}^n.$$

To find a basis of the row space $\mathrm{Row}A = \mathrm{Col}A^T$, we may certainly carry out row operations on A^T, i.e., *column operations* on A, and then find pivot columns of A. However, the row operations in A, already used in Example 2.4.3 to get a basis of the column space $\mathrm{Col}A$, can also be used. The key is that column operations do not change the column space.

Proposition 2.4.6. *The subspace*

$$\mathrm{Span}\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\} = \mathbb{R}\vec{v}_1 + \mathbb{R}\vec{v}_2 + \cdots + \mathbb{R}\vec{v}_n$$

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 the right side is also a linear combination of the 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 2.4.5. To get a basis of the row space of the matrix in Example 2.4.3, we interpret the row operations in the earlier example as column operations

$$A^T = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix} \xrightarrow{\text{col op}} \begin{pmatrix} 1 & 0 & 0 \\ 4 & -3 & 0 \\ 7 & -6 & 0 \\ 10 & -9 & 0 \end{pmatrix} = (\vec{w}_1 \ \vec{w}_2 \ \vec{0}).$$

The result is the *column echelon form*. By Proposition 2.4.6, we get

$$\mathrm{Row}A = \mathrm{Col}A^T = \mathbb{R}\vec{w}_1 + \mathbb{R}\vec{w}_2 + \mathbb{R}\vec{0} = \mathbb{R}\vec{w}_1 + \mathbb{R}\vec{w}_2.$$

Moreover, the following shows that $\vec{w}_1, \vec{w}_2$ are linearly independent

$$\begin{aligned} & x_1\vec{w}_1 + x_2\vec{w}_2 = \vec{0} \\ \implies & 1x_1 = 0 \implies x_1 = 0 && \text{(look at the first coordinate)} \\ \implies & x_2\vec{w}_2 = \vec{0} && \text{(substitute } x_1 = 0) \\ \implies & -3x_2 = 0 \implies x_2 = 0. && \text{(look at the second coordinate)} \end{aligned}$$

The key point here is that the multiplying the pivot terms $1, 3$ to the coefficients x_1, x_2 should give 0. Therefore the nonzero columns $\vec{w}_1, \vec{w}_2$ in the column echelon form of A^T form a basis of the row space.

We can apply the similar idea to the column space. We carry out the column operations

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

The nonzero columns $(1, 1, 1), (0, 1, 2)$ on the right form a basis of the column space $\text{Col}A$.

The example shows the basis of column space can be calculated by either row or column operations. Since the dimension of the column space is the rank of the matrix (and calculated by row operations), we get the following result.

Theorem 2.4.7. $\text{rank}A^T = \text{rank}A$.

Exercise 2.34. Suppose the rank of an $m \times n$ matrix A is r. What are the dimensions of $\text{Col}A, \text{Row}A, \text{Nul}A, \text{Nul}A^T$?

Exercise 2.35. Find bases of $\text{Col}A, \text{Row}A, \text{Nul}A, \text{Nul}A^T$.

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

Exercise 2.36. Use the row operations that you did for Exercise 2.32 to find bases of row spaces.

2.5 Sum and Direct Sum

2.5.1 Generalisation of Span and Linear Independence

The *sum* of two subspaces H_1 and H_2 is

$$H_1 + H_2 = \{\vec{x}_1 + \vec{x}_2 \colon \vec{x}_1 \in H_1,\ \vec{x}_2 \in H_2\}.$$

Consider $\vec{x} = \vec{x}_1 + \vec{x}_2 \in H_1 + H_2$ and $\vec{y} = \vec{y}_1 + \vec{y}_2 \in H_1 + H_2$, with $\vec{x}_1, \vec{y}_1 \in H_1$ and $\vec{x}_2, \vec{y}_2 \in H_2$. The following shows $a\vec{x} + b\vec{y} \in H_1 + H_2$

$$a\vec{x} + b\vec{y} = (a\vec{x}_1 + a\vec{x}_2) + (b\vec{y}_1 + b\vec{y}_2) = (a\vec{x}_1 + b\vec{y}_1) + (a\vec{x}_2 + b\vec{y}_2) \in H_1 + H_2.$$

Therefore the sum is a subspace.

Definition 2.5.1. The *sum* of subspaces $H_1, H_2, \ldots, H_n \subset \mathbb{R}^m$ is

$$H_1 + H_2 + \cdots + H_n = \{\vec{x}_1 + \vec{x}_2 + \cdots + \vec{x}_n \colon \vec{x}_i \in H_i\}.$$

The sum is *direct* if for $\vec{x}_i, \vec{y}_i \in H_i$, we have

$$\vec{x}_1 + \vec{x}_2 + \cdots + \vec{x}_n = \vec{y}_1 + \vec{y}_2 + \cdots + \vec{y}_n \implies \vec{x}_i = \vec{y}_i.$$

In case of direct sum, we denote $H_1 \oplus H_2 \oplus \cdots \oplus H_n$.

Suppose $H_i = \mathbb{R}\vec{v}_i$ are spanned by single vectors. Then

$$\vec{x}_i = x_i\vec{v}_i, \quad \vec{x}_1 + \vec{x}_2 + \cdots + \vec{x}_n = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n.$$

Therefore the sum $\mathbb{R}\vec{v}_1 \oplus \mathbb{R}\vec{v}_2 \oplus \cdots \oplus \mathbb{R}\vec{v}_n$ of subspaces is the span of the vectors. If we further know $\vec{v}_i \neq \vec{0}$ (i.e., all H_i are 1-dimensional), then the direct sum $\mathbb{R}\vec{v}_1 + \mathbb{R}\vec{v}_2 + \cdots + \mathbb{R}\vec{v}_n$ means

$$\begin{aligned} & x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n = y_1\vec{v}_1 + y_2\vec{v}_2 + \cdots + y_n\vec{v}_n \\ \implies & x_1\vec{v}_1 = y_1\vec{v}_1,\ x_2\vec{v}_2 = y_2\vec{v}_2,\ \ldots,\ x_n\vec{v}_n = y_n\vec{v}_n \\ \iff & x_1 = y_1,\ x_2 = y_2,\ \ldots,\ x_n = y_n. \end{aligned}$$

This is exactly the linear independence of $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$.

Therefore the sum and direct sum generalise span and linear independence.

Example 2.5.1. For $k = 1$, we have the sum H_1 of a single vector space H_1. The single sum is always direct.

For $H_1 + H_2$ to be direct, we require

$$\vec{x}_1 + \vec{x}_2 = \vec{y}_1 + \vec{y}_2 \implies \vec{x}_1 = \vec{y}_1,\ \vec{x}_2 = \vec{y}_2.$$

Let $\vec{z}_1 = \vec{x}_1 - \vec{y}_1 \in H_1$ and $\vec{z}_2 = \vec{x}_2 - \vec{y}_2 \in H_2$. Then the condition becomes

$$\vec{z}_1 + \vec{z}_2 = \vec{0} \implies \vec{z}_1 = \vec{z}_2 = \vec{0}.$$

The equality on the left means $\vec{z}_1 = -\vec{z}_2$, which is a vector in the intersection subspace (see Exercise 2.20) $H_1 \cap H_2$. Therefore the condition above means exactly $H_1 \cap H_2 = \{\vec{0}\}$. This is the criterion for the sum $H_1 + H_2$ of two subspaces to be direct.

The example is the generalisation of Example 4.2.4.

Exercise 2.37. Show that the sum of subspaces has the following properties

$$H_1 + H_2 = H_2 + H_1, \quad (H_1 + H_2) + H_3 = H_1 + H_2 + H_3 = H_1 + (H_2 + H_3).$$

Exercise 2.38. Show that $H_1 \subset H_1 + H_2$ and $H_2 \subset H_1 + H_2$. Then show that $H_1 + H_2 = H_1$ if and only if $H_2 \subset H_1$.

Exercise 2.39. Show that $\mathrm{Span}(\alpha \cup \beta) = \mathrm{Span}\alpha + \mathrm{Span}\beta$. You may assume α has two vectors and β has three vectors. Then use Theorem 2.4.4 to prove $\dim(H + H') \le \dim H + \dim H'$.

Exercise 2.40. Let $(A\ B)$ be the matrix formed by A followed by B. Show that $\mathrm{Col}(A\ B) = \mathrm{Col}A + \mathrm{Col}B$. You may assume A has two columns and B has three columns. Then prove $\mathrm{rank}(A\ B) \le \mathrm{rank}A + \mathrm{rank}B$.

Exercise 2.41. Let $\begin{pmatrix} A \\ B \end{pmatrix}$ be the matrix formed by A on top of B. Show that $\mathrm{Nul}\begin{pmatrix} A \\ B \end{pmatrix} = \mathrm{Nul}A \cap \mathrm{Nul}B$. You may assume A is a 3×3 matrix and B is a 2×3 matrix.

The following are the generalisations of Proposition 2.2.4 and 2.2.5.

Proposition 2.5.2. *The sum $H_1 + H_2 + \cdots + H_n$ is direct if and only if for $\vec{x}_i \in H_i$, we have*

$$\vec{x}_1 + \vec{x}_2 + \cdots + \vec{x}_n = \vec{0} \implies \vec{x}_i = \vec{0}.$$

Proposition 2.5.3. *The following are equivalent for subspaces $H_1, H_2, \dots, H_n$.*

1. *The sum $H_1 + H_2 + \cdots + H_n$ is not direct.*

2. *It is possible to have $x_i \in H_i$ and some $x_i \neq 0$, such that $\vec{x}_1 + \vec{x}_2 + \cdots + \vec{x}_n = \vec{0}$.*

3. *A nonzero vector in some H_i is also in the sum of the other subspaces.*

2.5.2 Layers of Direct Sum

We have

$$(H_1 + H_2) + (H_3 + H_4 + H_5) = H_1 + H_2 + H_3 + H_4 + H_5.$$

The left side is two layers of sums, and the right has only one layer of sum. We will argue that the direct sum properties of the two sides are equivalent:

1. $H_1 \oplus H_2$, $H_3 \oplus H_4 \oplus H_5$, $(H_1 + H_2) \oplus (H_3 + H_4 + H_5)$ are direct.

2. $H_1 \oplus H_2 \oplus H_3 \oplus H_4 \oplus H_5$ is direct.

The following shows the first direct sum property implies the second

$$\begin{aligned}
&\vec{x}_1+\vec{x}_2+\vec{x}_3+\vec{x}_4+\vec{x}_5=\vec{y}_1+\vec{y}_2+\vec{y}_3+\vec{y}_4+\vec{y}_5\\
\Longrightarrow\ &(\vec{x}_1+\vec{x}_2)+(\vec{x}_3+\vec{x}_4+\vec{x}_5)=(\vec{y}_1+\vec{y}_2)+(\vec{y}_3+\vec{y}_4+\vec{y}_5)\\
\Longrightarrow\ &\vec{x}_1+\vec{x}_2=\vec{y}_1+\vec{y}_2,\ \vec{x}_3+\vec{x}_4+\vec{x}_5=\vec{y}_3+\vec{y}_4+\vec{y}_5\\
\Longrightarrow\ &\vec{x}_1=\vec{y}_1,\ \vec{x}_2=\vec{y}_2,\ \vec{x}_3=\vec{y}_3,\ \vec{x}_4=\vec{y}_4,\ \vec{x}_5=\vec{y}_5.
\end{aligned}$$

The second $\Longrightarrow$ is due to $(H_1+H_2)\oplus(H_3+H_4+H_5)$, and the third $\Longrightarrow$ is due to $H_1\oplus H_2$ and $H_3\oplus H_4\oplus H_5$.

Suppose the second direct sum property holds. The following shows the direct sum $(H_1+H_2)\oplus(H_3+H_4+H_5)$

$$\begin{aligned}
&(\vec{x}_1+\vec{x}_2)+(\vec{x}_3+\vec{x}_4+\vec{x}_5)=(\vec{y}_1+\vec{y}_2)+(\vec{y}_3+\vec{y}_4+\vec{y}_5)\\
\Longrightarrow\ &\vec{x}_1+\vec{x}_2+\vec{x}_3+\vec{x}_4+\vec{x}_5=\vec{y}_1+\vec{y}_2+\vec{y}_3+\vec{y}_4+\vec{y}_5\\
\Longrightarrow\ &\vec{x}_1=\vec{y}_1,\ \vec{x}_2=\vec{y}_2,\ \vec{x}_3=\vec{y}_3,\ \vec{x}_4=\vec{y}_4,\ \vec{x}_5=\vec{y}_5\\
\Longrightarrow\ &\vec{x}_1+\vec{x}_2+\vec{x}_3=\vec{y}_1+\vec{y}_2+\vec{y}_3,\ \vec{x}_4+\vec{x}_5=\vec{y}_4+\vec{y}_5.
\end{aligned}$$

The direct sum $H_1\oplus H_2\oplus H_3\oplus H_4\oplus H_5$ is used in the second $\Longrightarrow$. Moreover, the following shows the direct sum $H_1\oplus H_2$

$$\begin{aligned}
\vec{x}_1+\vec{x}_2=\vec{y}_1+\vec{y}_2 \implies\ &\vec{x}_1+\vec{x}_2+\vec{0}+\vec{0}+\vec{0}=\vec{y}_1+\vec{y}_2+\vec{0}+\vec{0}+\vec{0}\\
\implies\ &\vec{x}_1=\vec{y}_1,\ \vec{x}_2=\vec{y}_2,\ \vec{0}=\vec{0},\ \vec{0}=\vec{0},\ \vec{0}=\vec{0}\\
\implies\ &\vec{x}_1=\vec{y}_1,\ \vec{x}_2=\vec{y}_2.
\end{aligned}$$

The argument for the direct sum $H_3\oplus H_4\oplus H_5$ is similar.

We could have more layers of sums. The similar argument shows that the directness of the sums at all the layers is equivalent to the directness of the whole sum.

Example 2.5.2. By Example 2.4.2, we know $\vec{v}_1=(1,2,3),\vec{v}_2=(4,5,6),\vec{v}_3=(7,8,10)$ form a basis of $\mathbb{R}^3$. Therefore $\mathbb{R}^3=\mathbb{R}\vec{v}_1\oplus\mathbb{R}\vec{v}_2\oplus\mathbb{R}\vec{v}_3$ is a direct sum. This implies the direct sum

$$\mathbb{R}^3=(\mathbb{R}\vec{v}_1+\mathbb{R}\vec{v}_2)\oplus\mathbb{R}\vec{v}_3=\mathrm{Col}\begin{pmatrix}1&4\\2&5\\3&6\end{pmatrix}\oplus\mathrm{Col}\begin{pmatrix}7\\8\\10\end{pmatrix},$$

and the direct sum

$$\mathbb{R}^3=\mathbb{R}\vec{v}_1\oplus(\mathbb{R}\vec{v}_2+\mathbb{R}\vec{v}_3)=\mathrm{Row}(1\ 2\ 3)\oplus\mathrm{Row}\begin{pmatrix}4&5&6\\7&8&10\end{pmatrix}.$$

2.5.3 Dimension of Sum

We specialise the discussion in Section 2.5.2 to the sum $\mathbb{R}\vec{v}_1+\mathbb{R}\vec{v}_2+\mathbb{R}\vec{v}_3+\mathbb{R}\vec{v}_4+\mathbb{R}\vec{v}_5$. In other words, we take H_i in Section 2.5.2 to be $H_i=\mathbb{R}\vec{v}_i$, with $\vec{v}_i\neq\vec{0}$. Section 2.5.2 shows the following are equivalent (the following H_i are different from H_i in Section 2.5.2).

1. $H_1=\mathbb{R}\vec{v}_1+\mathbb{R}\vec{v}_2$, $H_2=\mathbb{R}\vec{v}_3+\mathbb{R}\vec{v}_4+\mathbb{R}\vec{v}_5$, and H_1+H_2 are direct.

2. $\mathbb{R}\vec{v}_1+\mathbb{R}\vec{v}_2+\mathbb{R}\vec{v}_3+\mathbb{R}\vec{v}_4+\mathbb{R}\vec{v}_5$ is direct.

Since sum and direct sum extends span and independence, we know the directness of $\mathbb{R}\vec{v}_1+\mathbb{R}\vec{v}_2$, $\mathbb{R}\vec{v}_3+\mathbb{R}\vec{v}_4+\mathbb{R}\vec{v}_5$ and $\mathbb{R}\vec{v}_1+\mathbb{R}\vec{v}_2+\mathbb{R}\vec{v}_3+\mathbb{R}\vec{v}_4+\mathbb{R}\vec{v}_5$ mean $\alpha=\{\vec{v}_1,\vec{v}_2\}$, $\beta=\{\vec{v}_3,\vec{v}_4,\vec{v}_5\}$, and $\alpha\cup\beta=\{\vec{v}_1,\vec{v}_2,\vec{v}_3,\vec{v}_4,\vec{v}_5\}$ are bases of the subspaces H_1, H_2, H_1+H_2. Therefore the equivalence of the two statements can be rephrased as the following: If α and β are bases of H_1 and H_2, then H_1+H_2 is direct if and only if $\alpha\cup\beta$ is a basis of H_1+H_2. This leads to the following general result.

Proposition 2.5.4. *Suppose α_i is a basis of H_i. Then the sum $H_1+H_2+\cdots+H_n$ is direct if and only if $\alpha_1\cup\alpha_2\cup\cdots\cup\alpha_n$ is a basis of the sum.*

For bases α_i of H_i, we have

$$\begin{aligned}\dim H_1+\dim H_2+\cdots+\dim H_n &= \#\alpha_1+\#\alpha_2+\cdots+\#\alpha_n\\ &= \#(\alpha_1\cup\alpha_2\cup\cdots\cup\alpha_n)\\ &\geq \dim(H_1+H_2+\cdots+H_n).\end{aligned}$$

Here the second $\geq$ is due to $\alpha_1\cup\alpha_2\cup\cdots\cup\alpha_n$ spanning the sum and the first part of Theorem 2.4.4. The inequality is the first part of the following theorem.

Theorem 2.5.5. *We have* $\dim(H_1+H_2+\cdots+H_n)\leq\dim H_1+\dim H_2+\cdots+\dim H_n$. *Moreover, the sum is direct if and only if the equality holds.*

For the second part, we note that $\geq$ becomes the equality if and only if

$$\#(\alpha_1\cup\alpha_2\cup\cdots\cup\alpha_n)=\dim(H_1+H_2+\cdots+H_n).$$

Since the set $\alpha_1\cup\alpha_2\cup\cdots\cup\alpha_n$ spans the sum $H_1+H_2+\cdots+H_n$, by Theorem 2.4.4, the equality implies the set is a basis of the sum. By Proposition 2.5.4, this is equivalent to that the sum $H_1+H_2+\cdots+H_n$ is direct.

In case $n=2$, the difference between the two sides of the inequality can be made precise.

Theorem 2.5.6. $\dim(H_1+H_2)=\dim H_1+\dim H_2-\dim(H_1\cap H_2)$.

To explain the equality, let α be a basis of the intersection subspace $H_1 \cap H_2$. Then α is linearly independent, and can be extended to (maximal linear independent sets) a basis $\alpha \cup \beta_1$ of H_1 and a basis $\alpha \cup \beta_2$ of H_2.

Let

$$K_0 = H_1 \cap H_2 = \text{Span}\alpha, \quad K_1 = \text{Span}\beta_1, \quad K_2 = \text{Span}\beta_2.$$

By Proposition 2.5.4, we know $H_1 = K_0 \oplus K_1$ and $H_2 = K_0 \oplus K_2$ are direct sums. By Exercise 2.5.1, this means $K_0 \cap K_1 = K_0 \cap K_2 = \{\vec{0}\}$.

We have $H_1 + H_2 = K_0 + K_1 + K_2$. We claim that the right side is a direct sum $K_0 \oplus K_1 \oplus K_2$. By Proposition 2.5.2, this means

$$\vec{x}_0 + \vec{x}_1 + \vec{x}_2 = \vec{0}, \quad \vec{x}_0 \in K_0,\ \vec{x}_1 \in K_1,\ \vec{x}_2 \in K_2,$$

imply $\vec{x}_0 = \vec{x}_1 = \vec{x}_2 = \vec{0}$.

We have $\vec{x}_0 + \vec{x}_1 = -\vec{x}_2$. Since the left side is inside H_1, and the right side inside K_2, both sides are inside (the first equality is due to $K_2 \subset H_2$)

$$H_1 \cap K_2 = H_1 \cap H_2 \cap K_2 = K_0 \cap K_2 = \{\vec{0}\}.$$

Therefore we get $\vec{x}_2 = \vec{0}$. By the similar reason, we get $\vec{x}_1 = \vec{0}$. Then $\vec{x}_0 = -\vec{x}_1 - \vec{x}_2 = \vec{0}$.

Now we get the equality in the theorem by counting

$$\begin{aligned}\dim(H_1 + H_2) &= \dim K_0 \oplus K_1 \oplus K_2 \\ &= \#(\alpha \cup \beta_1 \cup \beta_2) \\ &= \#(\alpha \cup \beta_1) + \#(\alpha \cup \beta_2) - \#\alpha \\ &= \dim H_1 + \dim H_2 - \dim(H_1 \cap H_2).\end{aligned}$$

The direct sum $K_0 \oplus K_1 \oplus K_2$ is used in the second equality.

Chapter 3

Linear Transformation

Chapter 1 provides equation viewpoint. Chapter 2 provides vector viewpoint. This chapter provides the third viewpoint of linear transformation. The three viewpoints are equivalent.

3.1 Matrix of Linear Transformation

3.1.1 Transformation Given by Matrix

We view the left side $A\vec{x}$ of a system of linear equations as a formula that transforms $\vec{x} \in \mathbb{R}^n$ to $A\vec{x} \in \mathbb{R}^m$.

Example 3.1.1. The flip 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 flip of the second coordinate in $\mathbb{R}^3$ is given by

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

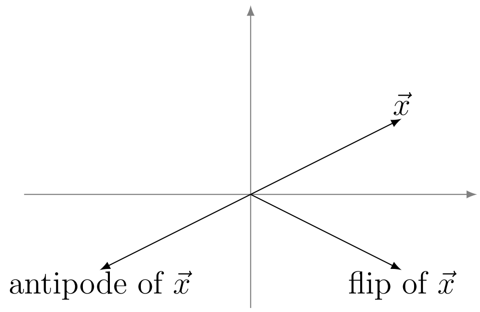

Figure 3.1.1: Flipping and antipode in $\mathbb{R}^2$.

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 2.1.1 and 2.4.1.

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 antipode transformation is given by the negative $-I$ of the identity matrix.

Exercise 3.1. Find the matrices for the flip of $\mathbb{R}^2$ with respect to the y-axis, and for the flip with respect to the diagonal line $x = y$.

Example 3.1.3. We may embed straight line $\mathbb{R}^1$ into the plane $\mathbb{R}^2$ as the horizontal axis or the 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

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

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

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

The transformations are given by 1×2 matrices.

Example 3.1.4. We apply the row operation $\mathrm{Row}_2 + c\mathrm{Row}_4$ to the (vertical) vector $\vec{x} \in \mathbb{R}^4$ and get

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \to \begin{pmatrix} x_1 \\ x_2 + cx_4 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & c \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}.$$

Therefore the row operation is a linear transformation with the matrix

$$E = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & c \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

All the row operations are of the for $\vec{x} \to E\vec{x}$. We call such matrices E *elementary matrices.*

Exercise 3.2. Describe the elementary matrices for applying the row operations on a vector in $\mathbb{R}^4$.

1. $\mathrm{Row}_4 + c\mathrm{Row}_2$.
2. $\mathrm{Row}_1 - c\mathrm{Row}_4$.
3. $\mathrm{Row}_1 \leftrightarrow \mathrm{Row}_2$.
4. $\mathrm{Row}_2 \leftrightarrow \mathrm{Row}_4$.
5. $c\mathrm{Row}_2$.
6. $-c\mathrm{Row}_4$.

Describe the general elementary matrices for row operations on vectors in $\mathbb{R}^n$.

Note that in expressing $T(\vec{x})$ as $A\vec{x}$, we require the entries of A to be constants, i.e., not involving $\vec{x}$. 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

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

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

3.1.2 Linear Transformation is Equivalent to Matrix

Not all transformations are of the form $A\vec{x}$. By Proposition 2.1.2, the transformation $A\vec{x}$ has the following property.

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

$$L(\vec{x}+\vec{y}) = L(\vec{x}) + L(\vec{y}), \quad L(a\vec{x}) = aL(\vec{x}).$$

Geometrically, the two properties mean that L preserves parallelograms and scalings.

Combining the two properties, a linear transformation preserves linear combinations

$$L(a_1\vec{x}_1 + a_2\vec{x}_2 + \cdots + a_n\vec{x}_n) = a_1L(\vec{x}_1) + a_2L(\vec{x}_2) + \cdots + a_nL(\vec{x}_n).$$

In Examples 2.1.1 and 2.4.1, we know any vector $\vec{x} \in \mathbb{R}^n$ is the 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 combination, 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 the one-to-one correspondence between linear transformations $L\colon \mathbb{R}^m \to \mathbb{R}^n$ and $m \times n$ matrices A:

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

We call $A = [L]$ the *matrix of linear transformation L*.

Example 3.1.5. A linear transformation $l\colon \mathbb{R}^n \to \mathbb{R}$ is called a *linear functional*. It is given by an $1 \times n$ matrix. Specifically, for $n = 3$, a linear functional is

$$l(x_1, x_2, x_3) = (a_1\ a_2\ a_3)\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = a_1x_1 + a_2x_2 + a_3x_3, \quad a_i = l(\vec{e}_i).$$

Exercise 3.3. Explain that a linear transformation $L\colon \mathbb{R} \to \mathbb{R}^n$ is the scaling of a vector.

Example 3.1.6. The flip 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 $L(\vec{e}_1) = \vec{e}_1$ and $L(\vec{e}_2) = -\vec{e}_2$, the matrix of the flip is

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

Example 3.1.7. 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

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

Example 3.1.8. 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 $\vec{e}_2 = (0, 1)$ 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 $(x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta)$.

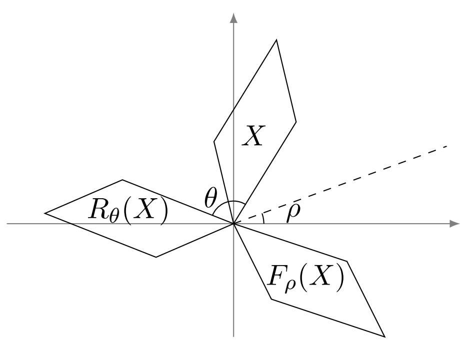

Figure 3.1.2: Rotation and flipping of $\mathbb{R}^2$.

Exercise 3.4. Use the correspondence between the linear transformation and the matrix to find the matrices for the flips in Exercise 3.1. What is the matrix of the flip with respect to the line of angle ρ (see Figure 3.1.2)?

Exercise 3.5. Find the matrix of flipping of $\mathbb{R}^3$ with respect to the (x, y)-plane.

Exercise 3.6. Find the matrix of the linear transformation of $\mathbb{R}^3$ that multiplies every vector by 5. What about $\mathbb{R}^n$?

Exercise 3.7. The *shear* transformation is a linear transformation that changes the unit square (normal line in Figure 3.1.3) to the gray parallelogram. Find the matrix of shear transformation.

There are other shear transformations, such that the one that changes the unit square to the dashed parallelogram. Find the matrix of this shear transformation.

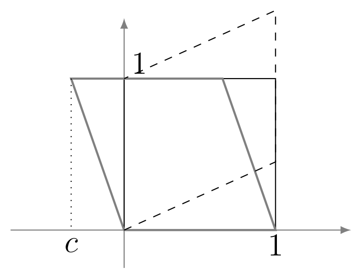

Figure 3.1.3: Shear transformation.

3.1.3 Calculation of Matrix of Linear Transformation

Example 3.1.9. 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 the 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 first column $L(\vec{e}_1)$ of the matrix $[L]$, 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

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

This means that $L(\vec{e}_1)$ is in fact the solution (x_1, x_2) of the system $x_1\vec{v}_1 + x_2\vec{v}_2 = \vec{e}_1$. We can find the solution $L(\vec{e}_1)$ by row operations on the augmented matrix $(\vec{v}_1\ \vec{v}_2\ \vec{e}_1)$. Similarly, the second column $L(\vec{e}_2)$ 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 operations to the first three columns, we find that the solution $L(\vec{e}_1)$ 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 $L(\vec{e}_2)$ of $x_1\vec{v}_1 + x_2\vec{v}_2 = \vec{e}_2$ is the fourth column $\vec{w}_2 = (\frac{3}{2}, -\frac{1}{2})$ on the right side. Then

$$[L] = (\vec{w}_1\ \vec{w}_2) = B = \begin{pmatrix} -2 & \frac{3}{2} \\ 1 & -\frac{1}{2} \end{pmatrix}.$$

This is the *inverse matrix* A^{-1} of A.

Example 3.1.10. We wish to find the matrix $[L]$ of the linear operator L on $\mathbb{R}^2$ that satisfies

$$L\begin{pmatrix}1\\2\end{pmatrix}=\begin{pmatrix}5\\6\end{pmatrix},\quad L\begin{pmatrix}3\\4\end{pmatrix}=\begin{pmatrix}7\\8\end{pmatrix}.$$

In Example 3.1.9, we already know

$$\vec{e}_1=-2\begin{pmatrix}1\\2\end{pmatrix}+1\begin{pmatrix}3\\4\end{pmatrix},\quad \vec{e}_2=\frac{3}{2}\begin{pmatrix}1\\2\end{pmatrix}-\frac{1}{2}\begin{pmatrix}3\\4\end{pmatrix}.$$

Following the same idea as the previous examples, we get

$$\begin{aligned}L(\vec{e}_1)&=-2L\begin{pmatrix}1\\2\end{pmatrix}+1L\begin{pmatrix}3\\4\end{pmatrix}=-2\begin{pmatrix}5\\6\end{pmatrix}+\begin{pmatrix}7\\8\end{pmatrix}=\begin{pmatrix}-3\\-4\end{pmatrix},\\ L(\vec{e}_2)&=\frac{3}{2}L\begin{pmatrix}1\\2\end{pmatrix}-\frac{1}{2}L\begin{pmatrix}3\\4\end{pmatrix}=\frac{3}{2}\begin{pmatrix}5\\6\end{pmatrix}-\frac{1}{2}\begin{pmatrix}7\\8\end{pmatrix}=\begin{pmatrix}4\\5\end{pmatrix}.\end{aligned}$$

Then

$$[L]=(L(\vec{e}_1)\ L(\vec{e}_2))=\begin{pmatrix}-3&4\\-4&5\end{pmatrix}.$$

A more direct method of finding $[L]$ makes use of the fact that linear transformations are preserved by column operations. Specifically, in the first matrix of the following column operation, applying L to the upper part gives the lower part.

$$\begin{pmatrix}\vec{v}_1&\vec{v}_2\\L(\vec{v}_1)&L(\vec{v}_2)\end{pmatrix}\xrightarrow{\mathrm{Col}_1+c\mathrm{Col}_2}\begin{pmatrix}\vec{v}_1+c\vec{v}_2&\vec{v}_2\\L(\vec{v}_1)+cL(\vec{v}_2)&L(\vec{v}_2)\end{pmatrix}=\begin{pmatrix}\vec{v}_1+c\vec{v}_2&\vec{v}_2\\L(\vec{v}_1+c\vec{v}_2)&L(\vec{v}_2)\end{pmatrix}.$$

After the column operation, in the matrix on the right, we also find that applying L to the upper part gives the lower part. The same observation applies to the other column operations

$$\begin{aligned}&\begin{pmatrix}\vec{v}_1&\vec{v}_2\\L(\vec{v}_1)&L(\vec{v}_2)\end{pmatrix}\xrightarrow{\mathrm{Col}_1\leftrightarrow\mathrm{Col}_2}\begin{pmatrix}\vec{v}_2&\vec{v}_1\\L(\vec{v}_2)&L(\vec{v}_1)\end{pmatrix},\\ &\begin{pmatrix}\vec{v}_1&\vec{v}_2\\L(\vec{v}_1)&L(\vec{v}_2)\end{pmatrix}\xrightarrow{c\mathrm{Col}_1}\begin{pmatrix}c\vec{v}_1&\vec{v}_2\\cL(\vec{v}_1)&L(\vec{v}_2)\end{pmatrix}=\begin{pmatrix}c\vec{v}_1&\vec{v}_2\\L(c\vec{v}_1)&L(\vec{v}_2)\end{pmatrix}.\end{aligned}$$

Therefore, if we have column operations

$$\begin{pmatrix}\vec{v}_1&\vec{v}_2\\\vec{w}_1&\vec{w}_2\end{pmatrix}\xrightarrow{\text{col op}}\begin{pmatrix}\vec{e}_1&\vec{e}_2\\\vec{u}_1&\vec{u}_2\end{pmatrix}=\begin{pmatrix}I\\A\end{pmatrix},$$

then $L(\vec{e}_i)=L(\vec{u}_i)$, and therefore $A=(\vec{u}_1\ \vec{u}_2)$ is the matrix of L.

Here is the specific column operations

$$\begin{pmatrix}1&3\\2&4\\5&7\\6&8\end{pmatrix}\xrightarrow{\mathrm{Col}_2-3\mathrm{Col}_1}\begin{pmatrix}1&0\\2&-2\\5&-8\\6&-10\end{pmatrix}\xrightarrow[-\frac{1}{2}\mathrm{Col}_2]{\mathrm{Col}_1+\mathrm{Col}_2}\begin{pmatrix}1&0\\0&1\\-3&4\\-4&5\end{pmatrix}.$$

We see that the lower half of the matrix is $[L]$.

Exercise 3.8. Find the matrices of linear transformations.

1. $L(1,0)=(1,2)$, $L(0,1)=(3,4)$.
2. $L(1,2)=(0,1)$, $L(3,4)=(1,0)$.
3. $L(1,2)=(3,4)$, $L(3,4)=(1,2)$.
4. $L(1,0)=(1,2,3)$, $L(0,1)=(4,5,6)$.
5. $L(0,1)=(1,2,3)$, $L(1,0)=(4,5,6)$.
6. $L(1,2)=(1,-2,1)$, $L(3,4)=(-2,1,3)$.
7. $L(1,0,0)=(1,2)$, $L(0,1,0)=(3,4)$, $L(0,0,1)=(5,6)$.
8. $L(1,0,0)=(1,2)$, $L(0,0,1)=(3,4)$, $L(0,1,0)=(5,6)$.
9. $L(1,0,0)=(1,2)$, $L(1,1,0)=(3,4)$, $L(1,1,1)=(5,6)$.
10. $L(1,2,0)=(1,2)$, $L(0,1,2)=(3,4)$, $L(0,0,1)=(5,6)$.
11. $L(1,-1,5)=(1,0,1)$, $L(-2,5,-4)=(0,1,-2)$, $L(-1,7,8)=(3,-2,0)$.
12. $L(1,-1,5)=(1,3)$, $L(-2,5,-4)=(-1,0)$, $L(-1,7,8)=(2,1)$.

Example 3.1.11. Suppose a linear transformation $L\colon \mathbb{R}^3 \to \mathbb{R}^3$ satisfies

$$L\begin{pmatrix}1\\2\\3\end{pmatrix}=\begin{pmatrix}0\\1\\2\end{pmatrix},\quad L\begin{pmatrix}4\\5\\6\end{pmatrix}=\begin{pmatrix}0\\4\\5\end{pmatrix}.$$

Then by

$$\begin{pmatrix}7\\8\\9\end{pmatrix}=-\begin{pmatrix}1\\2\\3\end{pmatrix}+2\begin{pmatrix}4\\5\\6\end{pmatrix},$$

we have

$$L\begin{pmatrix}7\\8\\9\end{pmatrix}=-L\begin{pmatrix}1\\2\\3\end{pmatrix}+2L\begin{pmatrix}4\\5\\6\end{pmatrix}=-\begin{pmatrix}0\\1\\2\end{pmatrix}+2\begin{pmatrix}0\\4\\5\end{pmatrix}=\begin{pmatrix}0\\7\\8\end{pmatrix}.$$

In particular, there is no linear transformation satisfying

$$L\begin{pmatrix}1\\2\\3\end{pmatrix}=\begin{pmatrix}0\\1\\2\end{pmatrix},\quad L\begin{pmatrix}4\\5\\6\end{pmatrix}=\begin{pmatrix}0\\4\\5\end{pmatrix},\quad L\begin{pmatrix}7\\8\\9\end{pmatrix}=\begin{pmatrix}0\\7\\7\end{pmatrix}.$$

Therefore the type of question in Example 3.1.10 may not always have answer. The key is that, if there is a linear relation

$$x_1\vec{v}_1+x_2\vec{v}_2+x_3\vec{v}_3=\vec{0},$$

then L should also satisfy the same linear relation

$$x_1 L(\vec{v}_1) + x_2 L(\vec{v}_2) + x_3 L(\vec{v}_3) = \vec{0}.$$

Of course, if the three vectors are linearly independent, then we have

$$x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{v}_3 = \vec{0} \implies x_1 = x_2 = x_3 = 0 \implies x_1 L(\vec{v}_1) + x_2 L(\vec{v}_2) + x_3 L(\vec{v}_3) = \vec{0}.$$

Then there is no constraint on the values $L(\vec{v}_1), L(\vec{v}_2), L(\vec{v}_3)$. When the three vectors are linearly dependent, then the three values cannot be arbitrarily chosen.

Exercise 3.9. Can you find linear transformations satisfying the given equalities?

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

Example 3.1.12. The orthogonal projection of $\mathbb{R}^3$ to the plane $H\colon x+y+z=0$ inside $\mathbb{R}^3$ is a linear transformation $P\colon \mathbb{R}^3 \to \mathbb{R}^3$. The columns of the matrix $[P]$ are the projections of the standard basis vectors $\vec{e}_i$ to the plane. These projections are not easy to see directly. On the other hand, we can easily find the projections of some other vectors.

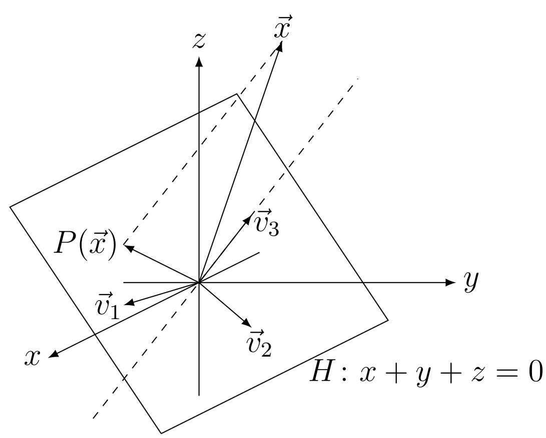

Figure 3.1.4: Projection to the plane $x+y+z=0$.

First, the vectors $\vec{v}_1 = (1,-1,0)$ and $\vec{v}_2 = (1,0,-1)$ are in H because they satisfy $x+y+z=0$. Since the projection clearly fixes the vectors in H, we get $P(\vec{v}_1) = \vec{v}_1$ and $P(\vec{v}_2) = \vec{v}_2$.

Second, the vector $\vec{v}_3 = (1,1,1)$ is orthogonal to H. Since the projection kills all vectors orthogonal to H, we get $P(\vec{v}_3) = \vec{0}$.

By $P(\vec{v}_1) = \vec{v}_1$, $P(\vec{v}_2) = \vec{v}_2$, $P(\vec{v}_3) = \vec{0}$, and the method in Example 3.1.3, we carry out column operations

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

Then we conclude

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

Exercise 3.10. Find the matrix of orthogonal projection of $\mathbb{R}^2$ to the line $x + y = 0$. How about the orthogonal projection to the diagonal $x = y$?

Exercise 3.11. Find the matrix of the orthogonal projection of $\mathbb{R}^3$ to the plane $H\colon x + y + 2z = 0$.

Exercise 3.12. Find the matrix of flip of $\mathbb{R}^3$ with respect to the plane $H\colon x + y + z = 0$.

Exercise 3.13. Find the linear transformation that takes the vectors $(0,1,2)$, $(1,2,0)$, $(2,0,1)$ to the next one, in circular way.

3.2 Matrix Operation

3.2.1 Sum of Matrix is Sum of Linear Transformation

The *addition* (or *sum*) of two linear transformations $L, K\colon \mathbb{R}^n \to \mathbb{R}^m$ is

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

The following shows that $L+K$ is a linear transformation

$$\begin{aligned}(L+K)(a\vec{x}+b\vec{y}) &= L(a\vec{x}+b\vec{y})+K(a\vec{x}+b\vec{y}) && \text{(definition of } L+K\text{)}\\ &= (aL(\vec{x})+bL(\vec{y}))+(aK(\vec{x})+bK(\vec{y})) && (L, K \text{ are linear})\\ &= a(L(\vec{x})+K(\vec{x}))+b(L(\vec{y})+K(\vec{y}))\\ &= a(L+K)(\vec{x})+b(L+K)(\vec{y}). && \text{(definition of } L+K\text{)}\end{aligned}$$

Similarly, the *scalar multiplication* cL is

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

We can easily verify that cL is also a linear transformation.

Definition 3.2.1. If A, B are $m \times n$ matrices for the linear transformations $L, K\colon \mathbb{R}^n \to \mathbb{R}^m$, then the *addition* $A+B$ is the matrix of the linear transformation $L+K$, and the *scalar multiplication* cA is the matrix of the linear transformation cL.

The definition means

$$[L+K] = [L]+[K], \quad [cL] = c[L].$$

Consider

$$\begin{aligned} L\begin{pmatrix}x_1\\x_2\end{pmatrix} &= \begin{pmatrix}1&3\\2&4\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix} = \begin{pmatrix}1x_1+3x_2\\2x_1+4x_2\end{pmatrix},\\ K\begin{pmatrix}x_1\\x_2\end{pmatrix} &= \begin{pmatrix}5&7\\6&8\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix} = \begin{pmatrix}5x_1+7x_2\\6x_1+8x_2\end{pmatrix}.\end{aligned}$$

We have

$$\begin{aligned} (L+K)\begin{pmatrix}x_1\\x_2\end{pmatrix} &= \begin{pmatrix}1x_1+3x_2\\2x_1+4x_2\end{pmatrix}+\begin{pmatrix}5x_1+7x_2\\6x_1+8x_2\end{pmatrix} = \begin{pmatrix}(1+5)x_1+(3+7)x_2\\(2+6)x_1+(4+8)x_2\end{pmatrix},\\ 10L\begin{pmatrix}x_1\\x_2\end{pmatrix} &= 10\begin{pmatrix}1x_1+3x_2\\2x_1+4x_2\end{pmatrix} = \begin{pmatrix}10\cdot 1x_1+10\cdot 3x_2\\10\cdot 2x_1+10\cdot 4x_2\end{pmatrix}.\end{aligned}$$

Therefore the matrix addition and scalar multiplication

$$\begin{aligned} \begin{pmatrix}1&3\\2&4\end{pmatrix}+\begin{pmatrix}5&7\\6&8\end{pmatrix} &= \begin{pmatrix}1+5&3+7\\2+6&4+8\end{pmatrix} = \begin{pmatrix}6&10\\8&12\end{pmatrix},\\ 10\begin{pmatrix}1&3\\2&4\end{pmatrix} &= \begin{pmatrix}10\cdot 1+10\cdot 3\\10\cdot 2+10\cdot 4\end{pmatrix} = \begin{pmatrix}10&30\\20&40\end{pmatrix}.\end{aligned}$$

We see the general pattern

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

In terms of columns of matrices, for

$$A = (\vec{v}_1 \ \vec{v}_2 \ \cdots \ \vec{v}_n), \quad B = (\vec{w}_1 \ \vec{w}_2 \ \cdots \ \vec{w}_n),$$

we have

$$A + B = (\vec{v}_1 + \vec{w}_1 \ \vec{v}_2 + \vec{w}_2 \ \cdots \ \vec{v}_n + \vec{w}_n), \quad cA = (c\vec{v}_1 \ c\vec{v}_2 \ \cdots \ c\vec{v}_n).$$

The same is true for the rows of matrices.

Note that we can add two matrices only if they have the same size. Matrices of different sizes cannot be added together.

Exercise 3.14. 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. $\begin{pmatrix} 1 & -1 \\ -1 & 0 \\ 0 & 1 \end{pmatrix}$.

Exercise 3.15. For a linear transformation L, verify that cL is still a linear transformation.

Exercise 3.16. Verify the transpose of matrix satisfies $(aA+bB)^T = aA^T + bB^T$. You may try this for 2×3 matrices. Then the general pattern is clear.

The interpretation in terms of linear transformation requires the dot product, which will be discussed in Section 5.1.2.

Exercise 3.17. Prove that $\mathrm{Col}(A+B) \subset \mathrm{Col}A + \mathrm{Col}B$. Then prove $\mathrm{rank}(A+B) \le \mathrm{rank}A + \mathrm{rank}B$. You may assume A and B have three columns.

The addition of linear transformations have the commutative property

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

Correspondingly, the addition of matrices have the commutative property

$$A + B = B + A.$$

It turns out addition and scalar multiplication of linear transformations have properties similar to numbers

$$(L+K)+M = L+(K+M), \quad a(L+K) = aL + aK, \quad (a+b)L = aL + bL.$$

Then we have the same properties for matrices

$$(A+B)+C = A+(B+C), \quad a(A+B) = aA + aB, \quad (a+b)A = aA + bA.$$

Exercise 3.18. Verify $(L+K)+M = L+(K+M)$, $a(L+K) = aL+aK$, $(a+b)L = aL+bL$.

3.2.2 Multiplication of Matrix is Composition of Linear Transformation

The *composition* of two linear transformations $L\colon \mathbb{R}^n \to \mathbb{R}^m$ and $K\colon \mathbb{R}^k \to \mathbb{R}^n$ is

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

The following shows that the composition is a linear transformation

$$\begin{aligned}
(L \circ K)(a\vec{x} + b\vec{y}) &= L(K(a\vec{x} + b\vec{y})) && \text{(definition of composition)}\\
&= L(aK(\vec{x}) + bK(\vec{y})) && (K \text{ is linear})\\
&= aL(K(\vec{x})) + b(L(K(\vec{y})) && (L \text{ is linear})\\
&= a(L \circ K)(\vec{x}) + b(L \circ K)(\vec{y}). && \text{(definition of composition)}
\end{aligned}$$

Definition 3.2.2. If A and B are $m \times n$ and $k \times n$ matrices for the linear transformations $L\colon \mathbb{R}^n \to \mathbb{R}^m$ and $K\colon \mathbb{R}^k \to \mathbb{R}^n$, then the *multiplication* AB is the $m \times k$ matrix of the linear transformation $L \circ K\colon \mathbb{R}^k \to \mathbb{R}^m$.

The definition means

$$[L \circ K] = [L][K].$$

Consider

$$\begin{aligned}
\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = L\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1y_1 + 3y_2 \\ 2y_1 + 4y_2 \end{pmatrix},\\
\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = K\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 5 & 7 \\ 6 & 8 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 5x_1 + 7x_2 \\ 6x_1 + 8x_2 \end{pmatrix},
\end{aligned}$$

The composition is the substituion

$$\begin{aligned}
(L \circ K)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= \begin{pmatrix} 1(5x_1 + 7x_2) + 3(6x_1 + 8x_2) \\ 2(5x_1 + 7x_2) + 4(6x_1 + 8x_2) \end{pmatrix}\\
&= \begin{pmatrix} (1 \cdot 5 + 3 \cdot 6)x_1 + (1 \cdot 7 + 3 \cdot 8)x_2 \\ (2 \cdot 5 + 4 \cdot 6)x_1 + (2 \cdot 7 + 4 \cdot 8)x_2 \end{pmatrix}.
\end{aligned}$$

Therefore the matrix multiplication

$$\begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}\begin{pmatrix} 5 & 7 \\ 6 & 8 \end{pmatrix} = \begin{pmatrix} 1 \cdot 5 + 3 \cdot 6 & 1 \cdot 7 + 3 \cdot 8 \\ 2 \cdot 5 + 4 \cdot 6 & 2 \cdot 7 + 4 \cdot 8 \end{pmatrix} = \begin{pmatrix} 23 & 31 \\ 34 & 46 \end{pmatrix}.$$

We see the general pattern

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

Basically the entries of AB are the (dot) products of the rows of A and columns of B. We note that the sizes of the matrices must be matched in order for the multiplication to make sense. The reason is that, in the composition $L \circ K$, the domain of L and the range of K must be the same.

In terms of columns of the matrix B in AB, we note that

$$\begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \end{pmatrix} = \begin{pmatrix} 1 \cdot x_1 + 3 \cdot x_2 & 1 \cdot y_1 + 3 \cdot y_2 \\ 2 \cdot x_1 + 4 \cdot x_2 & 2 \cdot y_1 + 4 \cdot y_2 \end{pmatrix},$$

where

$$\begin{pmatrix} 1 \cdot x_1 + 3 \cdot x_2 \\ 2 \cdot x_1 + 4 \cdot x_1 \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \begin{pmatrix} 1 \cdot y_1 + 3 \cdot y_2 \\ 2 \cdot y_1 + 4 \cdot y_1 \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

are the left sides $A\vec{x}$, $A\vec{y}$ of the system of linear equations. In general, we have

$$A(\vec{v}_1 \ \vec{v}_2 \ \cdots \ \vec{v}_k) = (A\vec{v}_1 \ A\vec{v}_2 \ \cdots \ A\vec{v}_k).$$

In particular, $A\vec{x}$ is the multiplication of the $m \times n$ matrix A and the $n \times 1$ matrix $\vec{x}$.

Example 3.2.1. Any map composed with the identity is the map itself. In terms of matrix, this means $IA = A = AI$. 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 the matrix multiplication

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

Comparing the two sides, we get

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

Exercise 3.19. Multiply matrices in Exercise 3.14 that can be multiplied together.

Exercise 3.20. Flip twice gives the identity. Use the matrix of flip in Exercise 3.4 to get trigonometric identities.

Example 3.2.3. In Example 3.1.4, applying a row operation to a vertical vector $\vec{x}$ is the same as $E\vec{x}$ for an elementary matrix E. If we apply this to all the column vectors in a matrix X, then we find a row operation on a matrix is the same as multiplying an elementary matrix on the left.

$$EX = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & c \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \\ x_4 & y_4 & z_4 \end{pmatrix} = \begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 + cx_4 & y_2 + cy_4 & z_2 + cz_4 \\ x_3 & y_3 & z_3 \\ x_4 & y_4 & z_4 \end{pmatrix}.$$

In particular, $E = EI$ s obtained by applying the row operation to the identity matrix I.

Exercise 3.21. Let X be a 4×3 matrix. For each row operation in Exercise 3.2, find the corresponding E, such that the row operation on X is EX.

Exercise 3.22. We apply several row operations to a 4×3 matrix X consecutively

1. $\text{Row}_4 + 2\text{Row}_2$, $\text{Row}_1 \leftrightarrow \text{Row}_2$.
2. $\text{Row}_1 \leftrightarrow \text{Row}_2$, $\text{Row}_4 + 2\text{Row}_2$.
3. 2Row_2, $\text{Row}_3 - \text{Row}_2$.
4. $\text{Row}_3 - \text{Row}_2$, 2Row_2.
5. $\text{Row}_4 + 2\text{Row}_2$, $\text{Row}_3 - \text{Row}_4$, 2Row_3.
6. 2Row_3, $\text{Row}_4 + 2\text{Row}_2$, $\text{Row}_3 - \text{Row}_4$.

Find matrix M, such that the result is MX.

Example 3.2.4. The composition has the 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 matrix multiplication satisfies $(AB)C = A(BC)$.

On the other hand, we generally have $L \circ K \neq K \circ L$, i.e., the composition is generally not commutative. 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 3.23. Verify that composition of linear transformations has the following properties

$$\begin{aligned} (L + K) \circ M &= L \circ M + K \circ M, \\ M \circ (L + K) &= M \circ L + M \circ K, \\ L \circ (aK) &= a(L \circ K) = (aL) \circ K. \end{aligned}$$

What do these tell you about the matrix multiplication?

Exercise 3.24. Verify the transpose of matrix satisfies $(AB)^T = B^T A^T$. You may try this for the case A and B are 2×2 matrices. Then the general pattern is clear.

The interpretation in terms of linear transformation requires the dot product, which will be discussed in Section 5.1.2.

Exercise 3.25. The transpose X^T exchanges the rows and columns of X, and we have $(AB)^T = B^T A^T$. Use the facts to interpret column operations as multiplying elementary matrices to the right. Then find the matrices for the following column operations

1. $\text{Col}_4 + c\text{Col}_2$.
2. $\text{Col}_1 - c\text{Col}_4$.
3. $\text{Col}_1 \leftrightarrow \text{Col}_2$.
4. $\text{Col}_2 \leftrightarrow \text{Col}_4$.
5. $c\text{Col}_2$.
6. $-c\text{Col}_4$.

Exercise 3.26. Prove that $\text{Col}AB \subset \text{Col}A$. Then prove $\text{rank}AB \le \text{rank}A$. Then use transpose to prove $\text{rank}AB \le \text{rank}B$.

3.3 Onto and One-to-One

3.3.1 Image and Preimage of Map

A linear transformation $L\colon \mathbb{R}^n \to \mathbb{R}^m$ corresponds to an $m \times n$ matrix A. The subspaces associated to A

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

correspond to the *range* and the *kernel* subspaces of L

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

The following are the direct verifications that these are subspaces

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

A linear transformation is a map with special property. In general, a *map* $f\colon X \to Y$ has the *domain* X and *range*[1]

$$f(X) = \{f(x)\colon \text{all } x \in X\} \subset Y.$$

[1]Sometimes Y is called the range. We do not use this definition in this course.

More generally, for any subset $A \subset X$, we may define the *image*

$$f(A) = \{f(x) \colon \text{all } x \in A\} \subset Y.$$

For any subset $B \subset Y$, we may define the *preimage*

$$f^{-1}(B) = \{x \in X \colon f(x) \in B\} \subset X.$$

Example 3.3.1. Consider the map

$$f = \text{Instructor} \colon X = \text{Courses} \to Y = \text{Professors}.$$

The range $f(X)$ is all the professors who teach some courses, the preimage $f^{-1}(\text{me})$ is all the courses I am teaching, including linear algebra $\in Y$.

Let $A \subset X$ be the subset of all mathematics courses. Then the image $f(A)$ is all the professors who teach mathematics courses.

Let $B \subset Y$ be all the mathematics professors. Then $f^{-1}(B)$ are all the course taught by mathematics professors.

The inclusion $f(A) \subset B$ means mathematics courses are taught only by mathematics professors, i.e., non-mathematics professors do not teach mathematics courses. The inclusion $f(A) \supset B$ means mathematics professors only teach mathematics courses, i.e., do not teach non-mathematics courses.

What is the meaning of $f(A) = B$?

Exercise 3.27. Find domain, range, $f(A)$, $f^{-1}(B)$.

1. Citizen: People $\to$ Countries. $A =$ Asians. $B =$ Asia.
2. Age: People $\to$ Natural numbers. $A =$ University students. $B = \{19, 20, 21, 22\}$.
3. Maker: Products $\to$ Manufacturers. $A =$ Japanese cars. $B =$ names starting with B.
4. Capital City: Countries $\to$ Cities. $A =$ names starting with A. $B =$ names starting with B.

Example 3.3.2. Let $L \colon \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation given by the matrix $\begin{pmatrix} 1.5 & 1 \\ -1 & 1 \end{pmatrix}$.

Let A be the standard unit square with $\vec{e}_1 = (1, 0)$ and $\vec{e}_2 = (0, 1)$ as the two sides. Then $L(A)$ is the parallelogram with $L(\vec{e}_1) = (1.5, -1)$ and $\vec{e}_2 = (1, 1)$ as the two sides.

Let B be the same standard unit square. Then we find $L^{-1}(\vec{e}_1) = \{(0.4, 0.4)\}$ and $L^{-1}(\vec{e}_2) = \{(-0.4, 0.6)\}$ by solving the systems $L(\vec{x}) = \vec{e}_1$ and $L(\vec{x}) = \vec{e}_2$. Then $L^{-1}(B)$ is the parallelogram with the vectors $(0.4, 0.4)$ and $(-0.4, 0.6)$ as the two sides.

Figure 3.3.1 gives $L(A)$ and $L^{-1}(B)$ in case A and B are the unit circles centered at the origin.

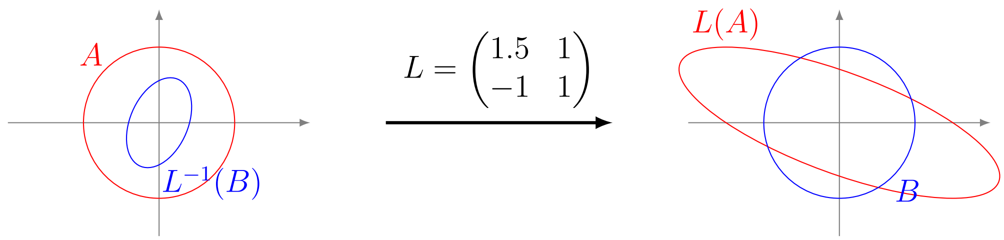

Figure 3.3.1: Image and preimage under a linear transformation.

Exercise 3.28. For the map $f(x) = x^2 \colon \mathbb{R} \to \mathbb{R}$ and the following subsets A, find $f(A)$ and $f^{-1}(A)$.

1. $\mathbb{R}$.
2. $\{0\}$.
3. $\{1\}$.
4. $\{-1\}$.
5. $[0, 1]$.
6. $[-1, 1]$.
7. $(1, +\infty)$.
8. $[0, +\infty)$.

Exercise 3.29. For the linear transformations in Example 3.1.3, find the domains and ranges. Then find $E_h^{-1}([0,1])$, $E_d^{-1}([0,1])$, $P_h(\mathbb{R} \times 0)$, $P_h(\text{diagonal})$, $P_h(0 \times \mathbb{R})$, $P_h^{-1}([0,1])$.

3.3.2 Onto and One-to-One

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 $y = f(x)$. This can be regarded as that the equation $f(x) = y$, with x as the variable, has solution for any given right side y.

The onto property for a linear transformation $L(\vec{x}) = A\vec{x} \colon \mathbb{R}^n \to \mathbb{R}^m$ means that the system of linear equations $A\vec{x} = \vec{b}$ has solution for all $\vec{b}$. By Theorems 1.3.3 and 1.4.2, we get a "dictionary" between different viewpoints.

Proposition 3.3.1. *Suppose a linear transformation $L \colon \mathbb{R}^n \to \mathbb{R}^m$ has corresponding matrix $A = (\vec{v}_1\ \vec{v}_2\ \cdots\ \vec{v}_n)$. Then the following are equivalent.*

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

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, we have

$$f(x_1) = f(x_2) \implies x_1 = x_2.$$

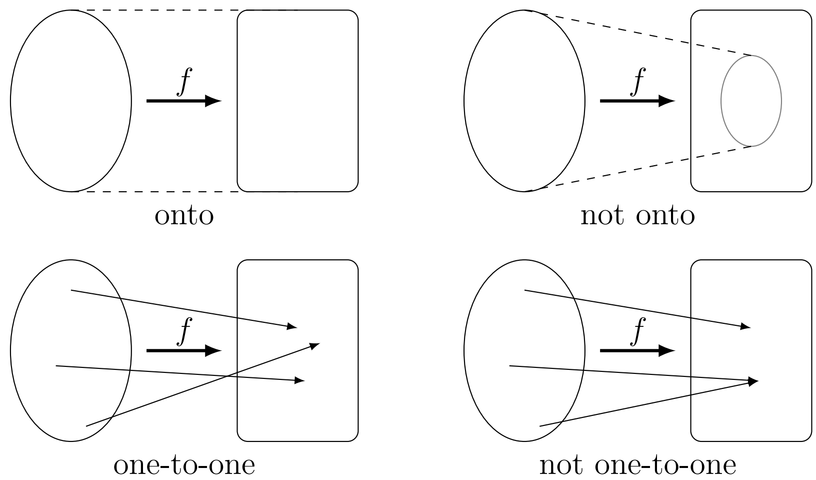

Figure 3.3.2: Onto and one-to-one.

This is equivalent to

$$x_1 \neq x_2 \implies f(x_1) \neq f(x_2).$$

The one-to-one property for a linear transformation $L(\vec{x}) = A\vec{x} \colon \mathbb{R}^n \to \mathbb{R}^m$ means that the solution of $A\vec{x} = \vec{b}$ is unique. By Theorems 1.3.2 and 1.4.2, we get a "dictionary" between different viewpoints.

Proposition 3.3.2. *Suppose a linear transformation $L \colon \mathbb{R}^n \to \mathbb{R}^m$ has corresponding matrix $A = (\vec{v}_1 \ \vec{v}_2 \ \cdots \ \vec{v}_n)$. Then the following are equivalent.*

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

Example 3.3.3. The Instructor map in Example 3.3.1 is onto if every professor teaches some course. The map is one-to-one if each professor teaches at most one course.

Exercise 3.30. Which maps in Exercise 3.27 are onto? Which are one-to-one?

Example 3.3.4. The identity, antipode, 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.

3.4 Inverse

3.4.1 Inverse Map

Many maps can be reversed. For example, the reverse of the rotation of $\mathbb{R}^2$ by angle θ is 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.*

The theorem says that f is invertible if and only if $f(x) = y$ has unique solution $x \in X$ for all the right side $x \in Y$.

Since onto and one-to-one are respectively called surjective and injective, we also call an invertible map *bijective*.

Example 3.4.1. The map "Instructor: Courses $\to$ Professors" in Example 3.3.1 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

$$\begin{aligned} K(a\vec{x}+b\vec{y}) &= K(aL(K(\vec{x}))+bL(K(\vec{y}))) && (L\circ K = I)\\ &= K(L(aK(\vec{x})+bK(\vec{y}))) && (L \text{ is linear})\\ &= aK(\vec{x})+bK(\vec{y}). && (K\circ L = I)\end{aligned}$$

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

Proposition 3.4.2. *Suppose a linear transformation $L\colon \mathbb{R}^n \to \mathbb{R}^m$ has corresponding matrix $A = (\vec{v}_1\ \vec{v}_2\ \cdots\ \vec{v}_n)$. Then the following are equivalent.*

1. *L is invertible.*

2. *$A\vec{x} = \vec{b}$ has unique solution for all $\vec{b}$.*

3. *$\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ is a basis of $\mathbb{R}^m$.*

4. *All rows and columns of A are pivot.*

The following is the linear transformation version of Theorem 2.4.4.

Theorem 3.4.3. *Suppose $L\colon \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation.*

1. *If L is onto, then $n \geq m$.*

2. *If L is one-to-one, then $n \leq m$.*

Moreover, in both cases, $m = n$ implies L is invertible.

3.4.2 Inverse Matrix

A matrix is *invertible* if it is the matrix of an invertible linear transformation. This means there is a matrix B satisfying $AB = I$ and $BA = I$. We denote the inverse matrix by $B = A^{-1}$.

The last statement in Proposition 3.4.2 implies A is a square matrix, i.e., $m = n$, and the reduced row echelon form of an invertible matrix is I.

The following is the matrix version of Theorem 1.4.4.

Proposition 3.4.4. *An invertible matrix is a square matrix. 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 matrix.

We need to explain the relation between $AB = I$ and the existence of solution, and between $BA = I$ and the uniqueness of solution.

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

If $BA = I$, then the following shows the uniqueness of the solution

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

If A is a square matrix, then the existence and the uniqueness are equivalent, and both $AB = I$ and $BA = I$ imply the invertiblity of A. It remains to show that $AB = I = B'A$ implies $B = B'$. The following is the argument

$$AB = I = B'A \implies B = BI = BAB' = IB' = B'.$$

Table 3.1 is the dictionary between equivalent concepts:

- system of m linear equations $A\vec{x} = \vec{b}$ in n variables.
- $m \times n$ matrix A.
- set α of n vectors in $\mathbb{R}^m$
- linear transformation $L\colon \mathbb{R}^n \to \mathbb{R}^m$.

The fifth column gives the calculation criterion. The identity I means A can be changed to the identity matrix (as the reduced row echelon form) by either row operations or by column operations.

system $A\vec{x} = \vec{b}$	matrix A	vector set α	linear L	row echelon form
existence	Col$A$	Spanα	Ran$L$	$\vec{b}$ not pivot
existence for all $\vec{b}$	Col$A = \mathbb{R}^m$	span $\mathbb{R}^m$	onto	all rows pivot
variation of solution	Nul$A$		Ker$L$	non-pivot columns
uniqueness	Nul$A = \{\vec{0}\}$	independence	one-to-one	all columns pivot
uniquely exist for all	invertible	basis of $\mathbb{R}^m$	invertible	identity I

Table 3.1: Key linear algebra dictionary.

Example 3.4.2. The 1×1 matrix (a) is invertible if and only if $a \neq 0$, and the inverse is $(a)^{-1} = (a^{-1})$.

The 2×2 matrix $\begin{pmatrix} a & c \\ b & d \end{pmatrix}$ is invertible if and only if the two columns are linearly independent. By Example 4.2.4, this means the two columns are not parallel. Algebraically, this means $ad \neq bc$. Then we may directly verify (i.e., $AA^{-1} = I = A^{-1}A$) that

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -c \\ -b & a \end{pmatrix}.$$

Example 3.4.3. In Example 3.1.8, 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 $\cos^2\theta + \sin^2\theta = 1$.

Exercise 3.31. The inverse of the flip of $\mathbb{R}^2$ is the flip itself. Derive a trigonometric equality from this fact.

Example 3.4.4. We try to find the inverse of the matrix in Example 3.1.9

$$A = (\vec{v}_1 \; \vec{v}_2) = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}.$$

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 matrix of L^{-1} is A^{-1}.

The inverse L^{-1} here is the linear transformation L in Example 3.1.9. 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

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

Example 3.4.4 gives the method for calculating the inverse of a square matrix A. Apply row operations to the matrix $(A \; I)$, until the A part becomes I

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

Then $B = A^{-1}$. Note that the row operations can change A to I, such that the method works, if and only if I is the reduced row echelon form of A. This means A is invertible.

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. In Example 3.1.12, we obtained the matrix of the orthogonal projection $P\colon \mathbb{R}^3 \to \mathbb{R}^3$ onto the subspace H given by $x+y+z=0$. The linear transformation is characterized by

$$P(\vec{v}_1)=\vec{v}_1,\quad P(\vec{v}_2)=\vec{v}_2,\quad P(\vec{v}_3)=\vec{0},\quad \vec{v}_1=\begin{pmatrix}1\\-1\\0\end{pmatrix},\quad \vec{v}_2=\begin{pmatrix}1\\0\\-1\end{pmatrix},\quad \vec{v}_3=\begin{pmatrix}1\\1\\1\end{pmatrix}.$$

In Example 3.1.12, we find $[P]$ by using the method of Example 3.1.10. Here we use the method of Example 3.1.9, which involves the calculation of inverse matrix.

We express $\vec{e}_1$ as a linear combination of $\vec{v}_1,\vec{v}_2,\vec{v}_3$. This means solving $A\vec{x}=x_1\vec{v}_1+x_2\vec{v}_2+x_3\vec{v}_3=\vec{e}_1$ 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 also need to do the same for $\vec{e}_2$ and $\vec{e}_3$. Therefore we solve three systems of linear equations together by the row operations

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

The solution of $A\vec{x}=\vec{e}_i$ is the i-th columns of A^{-1}, and we get

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

Then

$$P(\vec{e}_1)=\tfrac{1}{3}(P(\vec{v}_1)+P(\vec{v}_2)+P(\vec{v}_3))=\tfrac{1}{3}(\vec{v}_1+\vec{v}_2+\vec{0})=\frac{1}{3}\begin{pmatrix}2\\-1\\-1\end{pmatrix}.$$

Similarly, we get

$$P(\vec{e}_2)=\tfrac{1}{3}(-2\vec{v}_1+\vec{v}_2+\vec{0})=\frac{1}{3}\begin{pmatrix}-1\\2\\-1\end{pmatrix},\quad P(\vec{e}_3)=\tfrac{1}{3}(\vec{v}_1-2\vec{v}_2+\vec{0})=\frac{1}{3}\begin{pmatrix}-1\\-1\\2\end{pmatrix}.$$

Then

$$[P]=(P(\vec{e}_1)\ P(\vec{e}_2)\ P(\vec{e}_3))=\frac{1}{3}\begin{pmatrix}2&-1&-1\\-1&2&-1\\-1&-1&2\end{pmatrix}.$$

Exercise 3.32. Find the matrix of the orthogonal projection of $\mathbb{R}^2$ onto $ax + by = 0$.

Exercise 3.33. Find the matrix of the orthogonal flipping of $\mathbb{R}^3$ with respect to $x+y+z = 0$.

Exercise 3.34. Find the matrix of the orthogonal projection of $\mathbb{R}^4$ onto $x_1+x_2+x_3+x_4 = 0$.

Exercise 3.35. The orthogonal projection onto Col$A$ is characterized by fixing vectors in Col$A$ and sending vectors in NulA^T to $\vec{0}$. Find the matrix of the orthogonal projection onto Col$A$ for A in Exercise 2.35.

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 flip 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

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

In general, consider a linear transformation $L\colon \mathbb{R}^4 \to \mathbb{R}^4$ given by a matrix A. The restriction of L to the first 2-coordinates of the domain $\mathbb{R}^4$ 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 the range $\mathbb{R}^4$) 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 of the domain $\mathbb{R}^4$ to get a linear transformation L_2, and then taking the first two and last two coordinates of L_2, we get

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

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, if A_{ij} is the matrix of L_{ij}, then the matrix of L is a *block matrix*

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}.$$

In general, for a linear transformation $L\colon \mathbb{R}^n \to \mathbb{R}^m$, it may be convenient to have a combination in the domain $\mathbb{R}^n$ and another combination in the range $\mathbb{R}^m$

$$\begin{aligned} \vec{x} &= (\vec{x}_1, \vec{x}_2, \dots, \vec{x}_k) \in \mathbb{R}^n, & \vec{x}_i &\in \mathbb{R}^{n_i}, & n_1 + n_2 + \cdots + n_k &= n; \\ \vec{y} &= (\vec{y}_1, \vec{y}_2, \dots, \vec{y}_l) \in \mathbb{R}^m, & \vec{y}_j &\in \mathbb{R}^{m_j}, & m_1 + m_2 + \cdots + m_l &= m. \end{aligned}$$

Then we have the block form of the 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, for the $m_j \times n_i$ matrix A_{ji} of L_{ji}, we have the block matrix

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1k} \\ A_{21} & A_{22} & \cdots & A_{2k} \\ \vdots & \vdots & & \vdots \\ A_{l1} & A_{l2} & \cdots & A_{lk} \end{pmatrix}.$$

The operations of block matrices are the same as the operations of usual matrices that have numbers as entries

$$\begin{aligned} a\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} + b\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} &= \begin{pmatrix} aA_{11} + bB_{11} & aA_{12} + bB_{12} \\ aA_{21} + bB_{21} & aA_{22} + bB_{22} \end{pmatrix}, \\ \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} &= \begin{pmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{pmatrix}. \end{aligned}$$

The only thing we need to be careful is that blocks should have matching size, and matrix multiplication 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 3.36. Find the inverse of $\begin{pmatrix} I & O & O \\ A & I & O \\ B & C & I \end{pmatrix}$.

Exercise 3.37. If A and C are invertible, show that $\begin{pmatrix} A & B \\ O & C \end{pmatrix}$ is invertible.

3.6 *LU*-Decomposition

The LU-decomposition is the matrix interpretation of Gaussian elimination. The Gaussian elimination is carried out by row operations. Examples 3.1.4 and 3.2.3 show that row operations on a matrix is the same as multiplying elementary matrices to the left of the matrix.

Example 3.6.1. We may interpret the row operations in Example 1.2.1
$$\begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{pmatrix} \xrightarrow[\text{Row}_3 - 3\text{Row}_1]{\text{Row}_2 - 2\text{Row}_1} \begin{pmatrix} 1 & 4 & 7 & 10 \\ 0 & -3 & -6 & -9 \\ 0 & -6 & -12 & -18 \end{pmatrix} \xrightarrow{\text{Row}_3 - 2\text{Row}_2} \begin{pmatrix} 1 & 4 & 7 & 10 \\ 0 & -3 & -6 & -9 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

as successively multiplying on the left by elementary matrices

$$E_1 = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad E_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{pmatrix}, \quad E_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{pmatrix}$$

We note that in $E_3E_2E_1A$, E_1 is multiplied to A first

$$E_3E_2E_1A = E_3E_2E_1 \begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{pmatrix} = \begin{pmatrix} 1 & 4 & 7 & 10 \\ 0 & -3 & -6 & -9 \\ 0 & 0 & 0 & 0 \end{pmatrix} = R.$$

The row echelon form R is *upper triangular*, and the elementary matrices E_i are *lower triangular*. The multiplication $E_3E_2E_1$ of the lower triangular matrices is still lower triangular, and the inverse

$$E_1^{-1}E_2^{-1}E_3^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 2 & 1 \end{pmatrix}$$

is still lower triangular. Then we get

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 4 & 7 & 10 \\ 0 & -3 & -6 & -9 \\ 0 & 0 & 0 & 0 \end{pmatrix} = LU,$$

where L is lower triangular and $U = R$ is upper triangular. This is the LU-decomposition of A.

We remark that the numbers $2, 3, 2$ in the matrix L are exactly the numbers used in the row operations $\text{Row}_2 - 2\text{Row}_1$, $\text{Row}_3 - 3\text{Row}_1$, $\text{Row}_3 - 2\text{Row}_2$. There are two reasons behind the observation.

First, the inverse of the operation $\text{Row}_i - c\text{Row}_j$ is $\text{Row}_i + c\text{Row}_j$. Therefore -2 in E_1 becomes 2 in E_1^{-1}, and similarly for the other entries.

Second, if the row operations are performed from left columns to the right columns, and within each column, we only use the upper nonzero entries to eliminate lower nonzero entries, then the product of the inverses of the corresponding third type elementary matrices is very easy

$$\begin{pmatrix} 1 & 0 & 0 \\ c_{21} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ c_{31} & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & c_{32} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ c_{21} & 1 & 0 \\ c_{31} & c_{32} & 1 \end{pmatrix}.$$

The equality can be interpreted as successively applying the row operations $\text{Row}_3 + c_{32}\text{Row}_2$, $\text{Row}_3 + c_{31}\text{Row}_1$, $\text{Row}_2 + c_{21}\text{Row}_1$ to the identity matrix. If the order is changed, then the product may become more complicated

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & c_{32} & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ c_{21} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ c_{31} & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ c_{21} & 1 & 0 \\ c_{31} + c_{32}c_{21} & c_{32} & 1 \end{pmatrix}.$$

We conclude that, if we carry out row operations in the right order, such as the following

$$A = \begin{pmatrix} \bullet & * & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{pmatrix} \to \begin{pmatrix} \bullet & * & * & * \\ 0 & \bullet & * & * \\ 0 & * & * & * \\ 0 & * & * & * \end{pmatrix} \to \begin{pmatrix} \bullet & * & * & * \\ 0 & \bullet & * & * \\ 0 & 0 & \bullet & * \\ 0 & 0 & * & * \end{pmatrix} \to \begin{pmatrix} \bullet & * & * & * \\ 0 & \bullet & * & * \\ 0 & 0 & \bullet & * \\ 0 & 0 & 0 & \bullet \end{pmatrix} = U, \tag{3.6.1}$$

then we can easily get the decomposition $A = LU$ without further calculation.

Example 3.6.2. The row operations

$$\begin{pmatrix} 1 & 1 & 0 & 3 \\ 2 & 1 & -1 & 1 \\ 3 & -1 & -1 & 2 \\ -1 & 2 & 3 & -1 \end{pmatrix} \xrightarrow[\text{Row}_4+1\text{Row}_1]{\substack{\text{Row}_2-2\text{Row}_1 \\ \text{Row}_3-3\text{Row}_1}} \begin{pmatrix} 1 & 1 & 0 & 3 \\ 0 & -1 & -1 & -5 \\ 0 & -4 & -1 & -7 \\ 0 & 3 & 3 & 2 \end{pmatrix} \xrightarrow[\text{Row}_4+3\text{Row}_2]{\text{Row}_3-4\text{Row}_2} \begin{pmatrix} 1 & 1 & 0 & 3 \\ 0 & -1 & -1 & -5 \\ 0 & 0 & 3 & 13 \\ 0 & 0 & 0 & -13 \end{pmatrix}$$

give the LU-decomposition

$$\begin{pmatrix} 1 & 1 & 0 & 3 \\ 2 & 1 & -1 & 1 \\ 3 & -1 & -1 & 2 \\ -1 & 2 & 3 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 4 & 1 & 0 \\ -1 & -3 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & 3 \\ 0 & -1 & -1 & -5 \\ 0 & 0 & 3 & 13 \\ 0 & 0 & 0 & -13 \end{pmatrix}.$$

Exercise 3.38. Find LU-decompositions.

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

The row operations (3.6.1) may not always happen, because we may find that some expected $\bullet$-entries become zero. In this case, we use the exchange of rows $\text{Row}_i \leftrightarrow \text{Row}_j$ to make higher entries to become nonzero. This means that, we may need to exchange rows of A by multiplying a suitable *permutation matrix* P to the left of A. Then PA has LU-decomposition.

Example 3.6.3. We carry out the elimination in the first column

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 1 \\ 3 & 4 & 5 \end{pmatrix} \xrightarrow[\text{Row}_3-3\text{Row}_1]{\text{Row}_2-2\text{Row}_1} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & -1 \\ 0 & 1 & 2 \end{pmatrix}.$$

Since the $(2,2)$-entry becomes 0, we need to do $\text{Row}_2 \leftrightarrow \text{Row}_3$. Therefore we multiply the corresponding elementary matrix to the left, and then carry out the row operations

$$PA = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 1 \\ 3 & 4 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 3 & 4 & 5 \\ 2 & 2 & 1 \end{pmatrix} \xrightarrow[\text{Row}_3 - 2\text{Row}_1]{\text{Row}_2 - 3\text{Row}_1} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & -1 \end{pmatrix} = U.$$

This gives the LU-decomposition up to permutation

$$PA = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 1 \\ 3 & 4 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & -1 \end{pmatrix} = LU.$$

Exercise 3.39. Find LU-decompositions up to permutation.

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

Chapter 4

Vector Space

The goal of this chapter is to introduce non-Euclidean vector spaces, and translate the earlier linear algebra results in Euclidean spaces to general vector spaces.

4.1 Vector Space

4.1.1 Axioms of Vector Space

The essence of linear algebra is linear combination. Vector space is any setting where linear combination makes sense and satisfies the usual properties. Then we can carry out linear algebra in any vector space just like the Euclidean space.

Definition 4.1.1. A *(real) vector space* is a set V, together with *addition* and *scalar multiplication* operations

$$\vec{x}+\vec{y}\colon V\times V\to V,\quad a\vec{x}\colon \mathbb{R}\times V\to V,$$

such that the following are satisfied:

1. Commutativity: $\vec{x}+\vec{y}=\vec{y}+\vec{x}$.
2. Additive associativity: $(\vec{x}+\vec{y})+\vec{z}=\vec{x}+(\vec{y}+\vec{z})$.
3. Zero: There is an element $\vec{0}\in V$ satisfying $\vec{x}+\vec{0}=\vec{x}=\vec{0}+\vec{x}$.
4. Negative: For any $\vec{x}$, there is $\vec{y}$ (to be denoted $-\vec{x}$), such that $\vec{x}+\vec{y}=\vec{0}=\vec{y}+\vec{x}$.
5. One: $1\vec{x}=\vec{x}$.
6. Multiplicative associativity: $(ab)\vec{x}=a(b\vec{x})$.
7. Scalar distributivity: $(a+b)\vec{x}=a\vec{x}+b\vec{x}$.
8. Vector distributivity: $a(\vec{x}+\vec{y})=a\vec{x}+a\vec{y}$.

The properties imply that the linear combination $a_1\vec{x}_1 + a_2\vec{x}_2 + \cdots + a_n\vec{x}_n$ is not ambiguous.

Example 4.1.1. The *zero vector space* $\{\vec{0}\}$ consists of a single element $\vec{0}$. This leaves no choice for the two operations: $\vec{0}+\vec{0}=\vec{0}$, $a\vec{0}=\vec{0}$. It can be easily verified that all eight axioms are satisfied.

Example 4.1.2. Consider all polynomials of degree $\le n$

$$P_n = \{x_0 + x_1t + x_2t^2 + \cdots + x_nt^n\}.$$

We know how to add two polynomials together

$$\begin{aligned}&(x_0 + x_1t + x_2t^2 + \cdots + x_nt^n) + (y_0 + y_1t + y_2t^2 + \cdots + y_nt^n)\\ =&(x_0+y_0) + (x_1+y_1)t + (x_2+y_2)t^2 + \cdots + (x_n+y_n)t^n,\end{aligned}$$

and how to multiply a number to a polynomial

$$a(x_0 + x_1t + x_2t^2 + \cdots + x_nt^n) = ax_0 + ax_1t + ax_2t^2 + \cdots + ax_nt^n.$$

It is easy to verify that all eight axioms are satisfied. Therefore P_n is a vector space.

The coefficients in polynomials provide a one-to-one correspondence

$$x_0 + x_1t + x_2t^2 + \cdots + x_nt^n \in P_n \longleftrightarrow (x_0, x_1, x_2, \ldots, x_n) \in \mathbb{R}^{n+1}.$$

Since the one-to-one correspondence preserves the addition and scalar multiplication, it identifies the linear algebra in the polynomial vector space P_n with the linear algebra in the Euclidean vector space $\mathbb{R}^{n+1}$. In particular, since the addition and scalar multiplication in $\mathbb{R}^{n+1}$ satisfy the eight axioms, the addition and scalar multiplication in P_n also satisfy the eight axioms.

A one-to-one correspondence that preserves the addition and scalar multiplication is called an *isomorphism*. An isomorphism preserves linear algebra problems and solutions.

Example 4.1.3. All $m \times n$ matrices form a vector space $M(m,n)$ with the addition and scalar multiplication in Section 3.2. The operations are completely similar to the Euclidean space, and satisfy the eight axioms. See the discussion before Exercise 3.18.

Alternatively, there is a one-to-one correspondence that identifies matrices with Euclidean vectors, and preserves the addition and scalar multiplication

$$\begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \\ x_3 & y_3 \end{pmatrix} \in M(3,2) \longleftrightarrow (x_1, x_2, x_3, y_1, y_2, y_3) \in \mathbb{R}^6.$$

Then we may use the fact that $\mathbb{R}^6$ satisfies eight axioms to argue that $M(m,n)$ is a vector space.

Example 4.1.4. All infinite sequences $(x_n)_{n=1}^{\infty}$ of real numbers form a vector space, with the addition and scalar multiplications given by

$$(x_n) + (y_n) = (x_n + y_n), \quad a(x_n) = (ax_n).$$

In fact, a sequence is the same as a function $x(n) = x_n$ on the set $\mathbb{N}$ of natural numbers. Moreover, we have the usual addition and scalar multiplication of functions

$$(f + g)(x) = f(x) + g(x), \quad (af)(x) = a\ f(x).$$

This makes all functions on a set X to become a vector space $\text{Fun}(X)$. The vector space is infinite dimensional (therefore cannot be identified with a Euclidean space) if X is infinite.

If f and g are smooth functions on $\mathbb{R}$, then $f + g$ and af are also smooth. Therefore all smooth functions on $\mathbb{R}$ form a vector space C^{∞}. The real line $\mathbb{R}$ can be changed to other intervals.

If f and g are continuous functions on $[0, 1]$, then $f+g$ and af are also continuous. Therefore all continuous functions on $[0, 1]$ form a vector space $C[0, 1]$. The interval $[0, 1]$ can be changed to other intervals.

Exercise 4.1. For the following addition and scalar multiplication in $\mathbb{R}^2$, verify which axioms of vector space are satisfied, and which are not satisfied.

1. $(x_1, x_2) + (y_1, y_2) = (x_1 + y_2, x_2 + y_1)$, $a(x_1, x_2) = (ax_1, ax_2)$.

2. $(x_1, x_2) + (y_1, y_2) = (x_1 + y_1, 0)$, $a(x_1, x_2) = (ax_1, 0)$.

3. $(x_1, x_2) + (y_1, y_2) = (x_1 + ky_1, x_2 + ly_2)$, $a(x_1, x_2) = (ax_1, ax_2)$.

In the last problem, find condition on k, l, such that $\mathbb{R}^2$ is a vector space.

The eight properties imply many other properties that we already know in Euclidean space. The following are some examples:

- The zero vector is unique.

- If $\vec{x} + \vec{y} = \vec{x}$, then $\vec{y} = \vec{0}$.

- $a\vec{x} = \vec{0}$ if and only if $a = 0$ or $\vec{x} = \vec{0}$.

Here we provide only the rigorous proof of the first property.

Suppose $\vec{0}_1$ and $\vec{0}_2$ are two zero vectors. By applying the first equality in Axiom 3 to $\vec{x} = \vec{0}_1$ and $\vec{0} = \vec{0}_2$, we get $\vec{0}_1 + \vec{0}_2 = \vec{0}_1$. By applying the second equality in Axiom 3 to $\vec{0} = \vec{0}_1$ and $\vec{x} = \vec{0}_2$, we get $\vec{0}_2 = \vec{0}_1 + \vec{0}_2$. Combining the two equalities, we get $\vec{0}_2 = \vec{0}_1 + \vec{0}_2 = \vec{0}_1$.

4.1.2 Subspace

Definition 4.1.2. A subset $H \subset V$ of a vector space V is a *subspace* if

$$\vec{x}, \vec{y} \in H,\ a, b \in \mathbb{R} \implies a\vec{x} + b\vec{y} \in H.$$

With the addition and scalar multiplication inherited from V, the subset H is still a vector space. For example, a matrix A gives four vector spaces Col$A$, Row$A$, Nul$A$, NulA^T.

The *span* of a collection of vectors $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\} \subset V$ was defined in Section 2.2. The concept can be extended to general vector space

$$\begin{aligned}\text{Span}\alpha &= \{x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n \colon x_i \in \mathbb{R}\} \\ &= \mathbb{R}\vec{v}_1 + \mathbb{R}\vec{v}_2 + \cdots + \mathbb{R}\vec{v}_n \subset V.\end{aligned}$$

In Section 2.3, we showed that a linear combination of linear combinations is still a linear combination. Therefore Spanα is a subspace of V, similar to the case of Euclidean space.

The vector spaces C^∞ is a subspace of Fun($\mathbb{R}$), and the vector space $C[0,1]$ is a subspace of Fun($[0,1]$).

Exercise 4.2. Extend Exercise 2.20 to general vector space: Suppose H_1, H_2 are subspaces of a vector space, show that the intersection $H_1 \cap H_2$ is also a subspace of V.

Example 4.1.5. A linear combination of even functions is still linear. Therefore even smooth functions form a subspace of C^∞. Similarly, odd smooth functions also form a subspace of C^∞.

If $f(1) = 0$ and $g(1) = 0$, then

$$(af + bg)(1) = af(1) + bg(1) = a0 + b0 = 0.$$

Therefore smooth functions satisfying $f(1) = 0$ form a subspace.

By taking $a = b = 0$ in the definition of subspace, we kn ow a subspace contains $\vec{0}$. This implies that smooth functions satisfying $f(1) = 1$ do not form a subspace.

Exercise 4.3. Determine whether the conditions give subspaces of C^∞.

1. $f(1) = 0$.
2. $f(0) = 1$.
3. $f(0) = f(1)$.
4. $f'(0) = f(1)$.
5. $f(t) = f(t+1)$.
6. $f''(t) + f(t) = 0$.
7. $f''(t) + f(t) = 1$.
8. $\lim_{t\to\infty} f(t) = 0$.
9. $\lim_{t\to\infty} f(t) = 1$.
10. $\lim_{t\to\infty} f(t)$ diverges.
11. $\int_0^1 f(t)dt = 0$.
12. $\int_0^1 tf(t)dt = 0$.

Exercise 4.4. Determine whether the subsets are subspaces of the space of all sequences (x_n).

1. x_n converges.
2. $|x_n|$ converges.
3. x_n diverges.
4. $|x_n|$ diverges.
5. $\sum x_n$ converges.
6. $\sum x_n$ diverges.
7. $\sum x_n$ absolutely converges.
8. $\sum x_n$ conditionally converges.

Example 4.1.6. A matrix A is *symmetric*, if is satisfies $A^T = A$. A symmetric 3×3 matrix is of the form

$$\begin{pmatrix} a_1 & b_1 & b_2 \\ b_1 & a_2 & b_3 \\ b_2 & b_3 & a_3 \end{pmatrix}.$$

If $A^T = A$ and $B^T = B$, then we have (see Exercise 3.16)

$$(aA + bB)^T = aA^T + bB^T = aA + bB.$$

This shows the linear combination $aA + bB$ is also symmetric. Therefore all $n \times n$ symmetric matrices form a subspace of the vector space $M(n, n)$ of all $n \times n$ matrices.

Exercise 4.5. A square matrix A is *skew-symmetric* if $A^T = -A$. Show that all $n \times n$ skew-symmetric matrices form a subspace of $M(n, n)$.

4.2 Basis

4.2.1 Coordinates with Respect to Ordered Basis

In Example 4.1.2, we identify the polynomial vector space with Euclidean space

$$x_0 + x_1 t + x_2 t^2 \in P_2 \longleftrightarrow (x_0, x_1, x_2) \in \mathbb{R}^3.$$

We regard the $x_0 + x_1 t + x_2 t^2$ as a linear combination of the monomials $1, t, t^2$, and take the coefficients to get a vector in $\mathbb{R}^3$.

In Example 4.1.3, we identify the matrix vector space with Euclidean space

$$\begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \end{pmatrix} \in M(2, 2) \longleftrightarrow (x_1, x_2, y_1, y_2) \in \mathbb{R}^4.$$

We express a matrix as a linear combination of basic matrices

$$\begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \end{pmatrix} = x_1 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + y_1 \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + y_2 \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},$$

and take the coefficients to get a vector in $\mathbb{R}^4$.

For a vector space V, we find an *ordered* set of vectors $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$. We express any vector $\vec{x} \in V$ as a linear combination of α

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

Then the coefficients in the expression form a Euclidean vector, called the *coordinates* of $\vec{x}$ with respect to α

$$[\vec{x}]_\alpha = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n.$$

For the process to be successful, the vector set α needs to satisfy two conditions:

1. Any vector $\vec{x} \in V$ is a linear combination of α.

2. The coefficients in the linear combination expression is unique.

The first property means α spans V

$$\text{Span}\alpha = \mathbb{R}\vec{v}_1 + \mathbb{R}\vec{v}_2 + \cdots + \mathbb{R}\vec{v}_n = V.$$

The second property is the following paraphrase of Definition 2.2.3.

Definition 4.2.1. Vectors $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ in a vector space are *linearly independent* if

$$x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n = y_1\vec{v}_1 + y_2\vec{v}_2 + \cdots + y_n\vec{v}_n \implies x_1 = y_1, x_2 = y_2, \dots, x_n = y_n.$$

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

We have Propositions 2.2.4 and 2.2.5 that give the criteria for the linear independence and dependence in Euclidean spaces. Since the proofs did not use the Euclidean space, they are still valid in general vector space.

In a Euclidean space, the basis of a subspace is defined as the equivalent properties in Theorem 2.4.1. The first property in the theorem is exactly the two conditions above, and can be used to define basis in general.

Definition 4.2.2. A (finite) collection of vectors in a vector space V is a *basis* of V, if it spans V and is linearly independent.

The proof of Theorem 2.4.1 did not use Euclidean spaces. Therefore it remains true in general vector spaces.

Theorem 4.2.3. *The following are equivalent for a finite set of vectors α in V:*

1. *α is a basis of V.*

2. *α is a minimal spanning set of V.*

3. *α is a maximal linearly independent set of V.*

A consequence of the theorem is the following existence of basis.

Proposition 4.2.4. *If a vector space is spanned by finitely many vectors, then it has a basis.*

The assumption of being spanned by finitely many vectors is clearly necessary. We call such vector spaces *finite dimensional*.

Here is the argument for Proposition 4.2.4. Suppose α is a finite set of vectors spanning V. If α is linearly independent, then α is a basis of V. If α is linearly dependent, then a vector $\vec{v} \in \alpha$ is a linear combination of the remaining vectors $\alpha' = \alpha - \{\vec{v}\}$. By Exercise 2.12, we know $\text{Span}\alpha' = \text{Span}\alpha = V$. Then we get a strictly smaller spanning set $\alpha' \subset \alpha$ of V. Then we ask whether α' is linearly independent and repeat the process. Since α is finite, the process eventually stops, and we get a minimal spanning set. By Theorem 4.2.3, we get a basis.

Example 4.2.1. In Example 2.4.1, we get the standard basis of $\mathbb{R}^n$

$$\epsilon = \{\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n\}.$$

The following shows $[\vec{x}]_\epsilon = \vec{x}$ in $\mathbb{R}^n$ for the case $n = 3$

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = x_1\vec{e}_1 + x_2\vec{e}_2 + x_3\vec{e}_3, \quad [\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}]_\epsilon = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

If we change the order in the standard basis, then we should also change the order of coordinates

$$[\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}]_{\{\vec{e}_2, \vec{e}_1, \vec{e}_3\}} = \begin{pmatrix} x_2 \\ x_1 \\ x_3 \end{pmatrix}, \quad [\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}]_{\{\vec{e}_3, \vec{e}_1, \vec{e}_2\}} = \begin{pmatrix} x_3 \\ x_1 \\ x_2 \end{pmatrix}.$$

This shows the importance of order in a basis in getting the coordinates.

Example 4.2.2. The set $\alpha = \{1, t, t^2\}$ spans P_2 because any polynomial $x_0 + x_1t + x_2t^2$ in P_2 is a linear combination of $1, t, t^2$.

The set is also linearly independent. If a linear combination $x_0 + x_1t + x_2t^2 = 0$, then we take $t = -1, 0, 1$ to get

$$x_0 + x_1 + x_2 = 0, \quad x_0 = 0, \quad x_0 - x_1 + x_2 = 0.$$

Then we can easily get $x_0 = x_1 = x_2 = 0$. By Proposition 2.2.4, this means α is linearly independent. Therefore $\alpha = \{1, t, t^2\}$ is a basis of P_2.

In general, $1, t, t^2, \dots, t^n$ form a basis of P_n, and we have

$$[x_0 + x_1t + x_2t^2 + \cdots + x_nt^n]_{\{1,t,t^2,\dots,t^n\}} = (x_0, x_1, \dots, x_n).$$

Exercise 4.6. Inductively explain that $1, t, t^2, \dots, t^n$ are linearly independent, and therefore form a basis of P_n.

Exercise 4.7. Explain that

$$\alpha = \{\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\}$$

is a basis of $M(2,2)$.

Exercise 4.8. Show that

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

form a basis of the subspace of 3×3 symmetric matrices in Example 4.1.6.

Exercise 4.9. Consider a vector set $\alpha = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ in V. The following operations change α to another vector set β

1. Exchange: $\beta = \{\vec{v}_2, \vec{v}_1, \vec{v}_3\}$.
2. Scale: $\beta = \{c\vec{v}_1, \vec{v}_2, \vec{v}_3\}$, $c \neq 0$.
3. Add a scale multiple: $\beta = \{\vec{v}_1 + c\vec{v}_2, \vec{v}_2, \vec{v}_3\}$.

Prove that α spans V if and only if β spans V. Moreover, the same holds for the linear independence and basis properties.

The general changes of α are given by Proposition 2.4.6. The changes preserve the span, linear independence, and basis properties.

4.2.2 Calculation of Coordinates

An ordered basis $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ of V gives a well defined coordinate map

$$\vec{x} = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n \in V \mapsto [\vec{x}]_\alpha = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n.$$

The map is invertible, with the inverse given by the linear combination map

$$(x_1, x_2, \dots, x_n) \in \mathbb{R}^n \mapsto x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n \in V.$$

The following shows that the coordinate map preserves linear combinations. Therefore linear algebra problems in V can be translated into linear algebra problems in $\mathbb{R}^n$.

Proposition 4.2.5. $[a\vec{x} + b\vec{y}]_\alpha = a[\vec{x}]_\alpha + b[\vec{y}]_\alpha$.

Suppose

$$\alpha = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\}, \quad [\vec{x}]_\alpha = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad [\vec{y}]_\alpha = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}.$$

Then

$$\vec{x} = x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{v}_3, \quad \vec{y} = y_1\vec{v}_1 + y_2\vec{v}_2 + y_3\vec{v}_3,$$

and we get

$$\begin{aligned} a\vec{x} + a\vec{y} &= a(x_1\vec{v}_1 + x_2\vec{v}_2 + x_3\vec{v}_3) + b(y_1\vec{v}_1 + y_2\vec{v}_2 + y_3\vec{v}_3) \\ &= (ax_1 + by_1)\vec{v}_1 + (ax_2 + by_2)\vec{v}_2 + (ax_3 + by_3)\vec{v}_3. \end{aligned}$$

Therefore

$$[a\vec{x} + b\vec{y}]_\alpha = \begin{pmatrix} ax_1 + by_1 \\ ax_2 + by_2 \\ ax_3 + by_3 \end{pmatrix} = a \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + b \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = a[\vec{x}]_\alpha + b[\vec{y}]_\alpha.$$

Example 4.2.3. Consider the basis $\alpha = \{\vec{v}_1, \vec{v}_2\} = \{(1,2),(3,4)\}$ of $\mathbb{R}^2$. The α-coordinate of $(5,6) \in \mathbb{R}^2$

$$[\begin{pmatrix} 5 \\ 6 \end{pmatrix}]_\alpha = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

means the system of linear equations

$$\begin{pmatrix} 5 \\ 6 \end{pmatrix} = y_1\vec{v}_1 + y_2\vec{v}_2 = y_1 \begin{pmatrix} 1 \\ 2 \end{pmatrix} + y_2 \begin{pmatrix} 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

The coefficient matrix is $(\alpha) = (\vec{v}_1\ \vec{v}_2)$, and the right side is $(5,6)$. We may use row operations to find the solution (see Example 1.1.1)

$$\begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix} \to \begin{pmatrix} 1 & 3 & 5 \\ 0 & -2 & -4 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{pmatrix}, \quad [\begin{pmatrix} 5 \\ 6 \end{pmatrix}]_\alpha = \begin{pmatrix} -1 \\ 2 \end{pmatrix}.$$

In general, a basis $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ of $\mathbb{R}^n$ corresponds to an invertible matrix

$$A = (\alpha) = (\vec{v}_1\ \vec{v}_2\ \cdots\ \vec{v}_n).$$

The α-coordinate of $\vec{x} \in \mathbb{R}^n$ is the solution of the system $A\vec{y} = \vec{x}$ with $\vec{x}$ on the right side, and we get

$$[\vec{x}]_\alpha = A^{-1}\vec{x} = (\alpha)^{-1}\vec{x}.$$

For example, the three vectors in Example 3.1.12 form a basis of $\mathbb{R}^3$

$$\alpha = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\} = \{(1,-1,0),(1,0,-1),(1,1,1)\}.$$

In Example 3.4.6, we calculated the inverse matrix

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

Then

$$[\vec{x}]_\alpha = \frac{1}{3} \begin{pmatrix} 1 & -2 & 1 \\ 1 & 1 & -2 \\ 1 & 1 & 1 \end{pmatrix} \vec{x} = \frac{1}{3} \begin{pmatrix} x_1 - 2x_2 + x_3 \\ x_1 + x_2 - 2x_3 \\ x_1 + x_2 + x_3 \end{pmatrix}.$$

Exercise 4.10. Find the coordinates of two vectors with respect to the given basis.

1. $(3,4),(4,3)$, $\alpha=\{(2,1),(1,2)\}$.
2. $(1,1,1),(-1,3,-2)$, $\alpha=\{(1,1,0),(1,0,1),(0,1,1)\}$.
3. $(1,-1,-1),(3,4,3)$, $\alpha=\{(1,2,3),(0,1,2),(0,0,1)\}$.
4. $(-1,1,-1,-3),(1,-1,0,1)$, $\alpha=\{(0,-1,2,1),(2,3,2,1),(-1,0,3,2),(4,1,2,3)\}$.

Exercise 4.11. Find the coordinates of a general vector in Euclidean space with respect to the given basis.

1. $(0,1),(1,0)$.
2. $(1,2),(3,4)$.
3. $(a,0),(0,b)$, $a,b\neq 0$.
4. $(\cos\theta,\sin\theta),(-\sin\theta,\cos\theta)$.
5. $(1,1,0),(1,0,1),(0,1,1)$.
6. $(1,2,3),(0,1,2),(0,0,1)$.
7. $(0,1,2),(0,0,1),(1,2,3)$.
8. $(0,-1,2,1),(2,3,2,1),(-1,0,3,2),(4,1,2,3)$.

Exercise 4.12. Suppose $ad\neq bc$. Find the coordinate of a vector in $\mathbb{R}^2$ with respect to the basis $(a,b),(c,d)$. See Example 3.4.2.

Example 4.2.4. In P_2, we consider the monomials $1,t-1,(t-1)^2$ at 1. The following change of variable shows any vector in P_2 is a linear combination of $1,t-1,(t-1)^2$

$$\begin{aligned}a_0+a_1t+a_2t^2&=a_0+a_1[1+(t-1)]+a_2[1+(t-1)]^2\\&=(a_0+a_1+a_2)+(a_1+2a_2)(t-1)+a_2(t-1)^2.\end{aligned}$$

Moreover, if two linear combinations are equal

$$a_0+a_1(t-1)+a_2(t-1)^2=b_0+b_1(t-1)+b_2(t-1)^2,$$

then substituting t by $t+1$ gives the equality

$$a_0+a_1t+a_2t^2=b_0+b_1t+b_2t^2.$$

By Example 4.2.2, we know $1,t,t^2$ are linearly independent. Therefore $a_0=b_0$, $a_1=b_1$, $a_2=b_2$.

Therefore $1,t-1,(t-1)^2$ form a basis of P_2, and we have

$$[a_0+a_1t+a_2t^2]_{\{1,t-1,(t-1)^2\}}=\begin{pmatrix}a_0+a_1+a_2\\a_1+2a_2\\a_2\end{pmatrix}.$$

In general, $1,t-t_0,(t-t_0)^2,\dots,(t-t_0)^n$ form a basis of P_n.

Example 4.2.5. Example 4.2.4 uses special features of vectors. If there is no special features we can use, then we may always translate the basis problem to Euclidean space.

We would like to know the condition on a, such that

$$\begin{pmatrix}1 & 3\\ 2 & 4\end{pmatrix}, \begin{pmatrix}1 & 2\\ 3 & 4\end{pmatrix}, \begin{pmatrix}3 & 1\\ 4 & 2\end{pmatrix}, \begin{pmatrix}2 & 1\\ 4 & a\end{pmatrix}$$

form a basis of $M(2,2)$. We translate the problem into $\mathbb{R}^4$, and carry out row operations

$$\begin{aligned}\begin{pmatrix}1 & 1 & 3 & 2\\ 2 & 3 & 4 & 4\\ 3 & 2 & 1 & 1\\ 4 & 4 & 2 & a\end{pmatrix} &\to \begin{pmatrix}1 & 1 & 3 & 2\\ 0 & 1 & -2 & 0\\ 0 & -1 & -8 & -5\\ 0 & 0 & -10 & a-8\end{pmatrix}\\ &\to \begin{pmatrix}1 & 1 & 3 & 2\\ 0 & 1 & -2 & 0\\ 0 & 0 & -10 & -5\\ 0 & 0 & -10 & a-8\end{pmatrix} \to \begin{pmatrix}1 & 1 & 3 & 2\\ 0 & 1 & -2 & 0\\ 0 & 0 & 2 & 1\\ 0 & 0 & 0 & a-3\end{pmatrix}\end{aligned}$$

The four matrices form a basis of $M(2,2)$ if and only if $a \neq 3$.

Exercise 4.13. Find conditions on a, such that the vectors form bases.

1. $1+t, 1+t^2, t+at^2$ in P_2.
2. $\begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}, \begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}, \begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix}, \begin{pmatrix}0 & 1\\ 1 & a\end{pmatrix}$ in $M(2,2)$.

Exercise 4.14. Determine whether polynomials form a basis of P_3. Moreover, find the coordinates with respect to the basis.

1. $1-t$, $t-t^2$, t^2-t^3, t^3-1.
2. $1+t$, $t+t^2$, t^2+t^3, t^3+1.
3. 1, $1+t$, $1+t^2$, $1+t^3$.
4. 1, $t+2$, $(t+2)^2$, $(t+2)^3$.

4.2.3 Linear Algebra Problem in General Vector Space

We may always translate linear algebra problems in finite dimensional vector spaces to problems in Euclidean spaces.

Example 4.2.6. We study the span and linear independence of the vectors in P_2

$$p_1(t) = 1+2t+3t^2, \quad p_2(t) = 4+5t+6t^2, \quad p_3(t) = 7+8t+at^2, \quad p_4(t) = 10+11t+bt^2$$

We use coordinates with respect to the basis $\{1, t, t^2\}$ to translate the problem to Euclidean vectors

$$\vec{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad \vec{v}_2 = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}, \quad \vec{v}_3 = \begin{pmatrix} 7 \\ 8 \\ a \end{pmatrix}, \quad \vec{v}_4 = \begin{pmatrix} 10 \\ 11 \\ b \end{pmatrix}.$$

In Example 1.3.4, we get row operations

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

The vectors $p_1(t), p_2(t), p_3(t), p_4(t)$ span P_2 if and only if the Euclidean vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4$ span $\mathbb{R}^3$. This means $a \neq 9$ or $b \neq 12$.

By restricting the row operations to $(\vec{v}_1\ \vec{v}_2\ \vec{v}_3)$ or $(\vec{v}_1\ \vec{v}_2\ \vec{v}_4)$, we know $\vec{v}_1, \vec{v}_2, \vec{v}_3$ span $\mathbb{R}^3$ for $a \neq 9$, and $\vec{v}_1, \vec{v}_2, \vec{v}_4$ span $\mathbb{R}^3$ for $b \neq 12$. Correspondingly, $p_1(t), p_2(t), p_3(t)$ span P_2 for $a \neq 9$, and $p_1(t), p_2(t), p_4(t)$ span P_2 for $b \neq 12$.

Since the four vectors in $\mathbb{R}^3$ are linearly dependent, we know $p_1(t), p_2(t), p_3(t), p_4(t)$ are linearly dependent. Moreover, by restricting the row operations to $(\vec{v}_1\ \vec{v}_2\ \vec{v}_3)$ or $(\vec{v}_1\ \vec{v}_2\ \vec{v}_4)$, we know $\vec{v}_1, \vec{v}_2, \vec{v}_3$ are linearly independent for $a \neq 9$, and $\vec{v}_1, \vec{v}_2, \vec{v}_4$ are linearly independent for $b \neq 12$. Correspondingly, $p_1(t), p_2(t), p_3(t)$ are linearly independent for $a \neq 9$, and $p_1(t), p_2(t), p_4(t)$ are linearly independent for $b \neq 12$.

Combining the span and linear independence properties of the polynomials, we know $p_1(t), p_2(t), p_3(t)$ form a basis of P_2 for $a \neq 9$, and $p_1(t), p_2(t), p_4(t)$ form a basis of P_2 for $b \neq 12$.

Example 4.2.7. The coordinates of the 2×2 matrices

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

with respect to the basis in Exercise 4.7 are the vectors in $\mathbb{R}^4$

$$\vec{v}_1 = \begin{pmatrix} 1 \\ 4 \\ 7 \\ 10 \end{pmatrix}, \quad \vec{v}_2 = \begin{pmatrix} 2 \\ 5 \\ 8 \\ 11 \end{pmatrix}, \quad \vec{v}_3 = \begin{pmatrix} 3 \\ 6 \\ a \\ b \end{pmatrix}.$$

Since the three vectors cannot span $\mathbb{R}^4$, we know the three matrices cannot span the vector space $M(2, 2)$.

For the linear independence, we carry out the row operations

$$A = (\vec{v}_1\ \vec{v}_2\ \vec{v}_3) = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & a \\ 10 & 11 & b \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & -6 & a-21 \\ 0 & -9 & b-30 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & 0 & a-9 \\ 0 & 0 & b-12 \end{pmatrix}.$$

The Euclidean vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3$ are linearly independent if and only if $a \neq 9$ or $b \neq 12$. Correspondingly, the three matrices are linearly independent if and only if $a \neq 9$ or $b \neq 12$.

Exercise 4.15. Find the condition on a, such that the last vector can be expressed as a linear combination of the previous ones.

1. $1+2t+3t^2$, $7+at+9t^2$, $10+11t+12t^2$.
2. t^2+2t+3, $7t^2+at+9$, $10t^2+11t+12$.
3. $\begin{pmatrix}1 & 2\\ 2 & 3\end{pmatrix}$, $\begin{pmatrix}4 & 5\\ 5 & 6\end{pmatrix}$, $\begin{pmatrix}7 & a\\ a & 9\end{pmatrix}$, $\begin{pmatrix}10 & 11\\ 11 & 12\end{pmatrix}$.
4. $\begin{pmatrix}1 & 2\\ 3 & 3\end{pmatrix}$, $\begin{pmatrix}7 & a\\ 9 & 9\end{pmatrix}$, $\begin{pmatrix}10 & 11\\ 12 & 12\end{pmatrix}$.

Exercise 4.16. Determine whether the vectors span P_2 or $M(2,2)$, and whether they are linearly independent.

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

Linear algebra problems in infinite dimensional vector spaces cannot be translated into problems in Euclidean spaces. We need to use special features of the vectors to solve problems. For functions, one commonly used technique is the evaluation at particular places.

Example 4.2.8. The equalities $1 = \cos^2 t + \sin^2 t$ and $\cos 2t = \cos^2 t - \sin^2 t$ shows that the constant function 1 and the function $\cos 2t$ are linear combinations of $\cos^2 t$ and $\sin^2 t$. This implies

$$\mathbb{R}1 + \mathbb{R}\cos^2 t + \mathbb{R}\sin^2 t = \mathbb{R}\cos 2t + \mathbb{R}\cos^2 t + \mathbb{R}\sin^2 t = \mathbb{R}\cos^2 t + \mathbb{R}\sin^2 t.$$

Are the remaining functions $\cos^2 t$ and $\sin^2 t$ linearly independent, and therefore form a basis of the span? Consider $x_1 \cos^2 t + x_2 \sin^2 t = 0$. By taking two values of t, we get

$$\begin{aligned} t = 0\colon x_1 &= x_1 \cdot 1^2 + x_2 \cdot 0^2 = 0, \\ t = \tfrac{\pi}{2}\colon x_2 &= x_1 \cdot 0^2 + x_2 \cdot 1^2 = 0. \end{aligned}$$

This shows that $\cos^2 t, \sin^2 t$ are linearly independent.

The function t is not a linear combination of $\cos^2 t$ and $\sin^2 t$. One reason is that any linear combination $x_1 \cos^2 t + x_2 \sin^2 t$ is bounded, while the function t is not bounded. Alternatively, if $t = x_1 \cos^2 t + x_2 \sin^2 t$, then by taking $t = \frac{\pi}{2}, -\frac{\pi}{2}$, we get

$$\tfrac{\pi}{2} = x_1 0^2 + x_2 1^2 = x_2, \quad -\tfrac{\pi}{2} = x_1 0^2 + x_2(-1)^2 = x_2.$$

This is a contradiction.

We may also directly show that $\cos^2 t, \sin^2 t, t$ are linearly independent. Consider $x_1 \cos t + x_2 \sin t + x_3 t = 0$. By evaluating at $0, \frac{\pi}{2}, \pi$, we get

$$\begin{aligned} t = 0 &\colon x_1 = 0, \\ t = \tfrac{\pi}{2} &\colon x_2 + \tfrac{\pi}{2} x_3 = 0, \\ t = \pi &\colon -x_1 + \pi x_3 = 0. \end{aligned}$$

Then we can derive $x_1 = x_2 = x_3 = 0$. This shows the three functions are linearly independent.

Exercise 4.17. Determine whether $f(t)$, $g(t)$ are linear combinations of α.

1. $\alpha = \{\cos^2 t, \sin^2 t\}$. $f(t) = \cos 2t$, $g(t) = \sin 2t$.
2. $\alpha = \{\cos 2t, \sin 2t\}$. $f(t) = 1$, $g(t) = \frac{1}{1+|t|}$.
3. $\alpha = \{e^{t+1}, e^{-t}\}$. $f(t) = e^{t-1}$, $g(t) = e^t + e^{-t}$.
4. $\alpha = \{1, t, e^t, te^t\}$. $f(t) = (1+t)e^t$, $g(t) = f'(t)$.
5. $\alpha = \{\cos^2 t, \cos 2t\}$. $f(t) = a$, $g(t) = a + \sin^2 t$.

Exercise 4.18. Find a smooth function that is not a linear combinations of $\cos t, \sin t, t$.

Exercise 4.19. Determine the linear independence.

1. $\cos^2 t, \sin^2 t$.
2. $\cos^2 t, \sin^2 t, \cos 2t$.
3. $\cos^2 t, \sin^2 t, \sin 2t$.
4. $1, t, e^t, te^t$.
5. e^t, e^{t+1}, e^{2t}.
6. $t, e^t, \cos t$.

4.2.4 Dimension

Definition 4.2.6. The *dimension* of a finite dimensional vector space is the number of vectors in a basis.

In Section 2.4, we showed that the dimension is well defined by considering the subspaces of $\mathbb{R}^n$ as column spaces. Here we can be more direct, by using the coordinates with respect to basis.

Suppose $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_m\}$ and $\beta = \{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_n\}$ are two bases of a finite dimensional vector space V. Then the α-coordinate

$$[\cdot]_\alpha \colon V \longleftrightarrow \mathbb{R}^m$$

translates the linear algebra in V to the linear algebra in $\mathbb{R}^m$. In particular, the basis β of V corresponds to the basis $[\beta]_\alpha = \{[\vec{w}_1]_\alpha, [\vec{w}_2]_\alpha, \dots, [\vec{w}_n]_\alpha\}$ of $\mathbb{R}^m$. This means that the $m \times n$ matrix $B = ([\vec{w}_1]_\alpha \ [\vec{w}_2]_\alpha \ \cdots \ [\vec{w}_n]_\alpha)$ has the property that $B\vec{x} = \vec{b}$ has unique solution for all the right side $\vec{b}$. By Theorem 1.4.4, this implies $m = n$, and completes the proof that the dimension is well defined.

By Examples 4.2.2, we get $\dim P_n = n + 1$. We also have $\dim M(m, n) = mn$.

Exercise 4.20. Use Exercise 4.8 to determine the dimension of the subspace of all $n \times n$ symmetric matrices.

Exercise 4.21. Exercise 4.5 shows that all the $n \times n$ skew-symmetric matrices form a subspace. What is the dimension?

We may use coordinates to easily translate many results about dimensions from the Euclidean spaces to general vector spaces. The following is the extension of Theorem 2.4.3 to general subspaces.

Theorem 4.2.7. *If H is a subspace of a vector space V, then $\dim H \le \dim V$. Moreover, if $\dim H = \dim V$, then $H = V$.*

The following extends Theorem 2.4.4.

Theorem 4.2.8. *Suppose α is a set of vectors in a vector space V.*

1. *If α spans V, then $\#\alpha \ge \dim V$.*

2. *If α is linearly independent, then $\#\alpha \le \dim V$.*

Moreover, in both cases, $\#\alpha = \dim V$ implies α is a basis of V.

Example 4.2.9. In Example 4.2.6, we first find $1+2t+3t^2$, $4+5t+6t^2$, $7+8t+at^2$ span P_2 in case $a \ne 9$, and later find the polynomials are linearly independent in case $a \ne 9$. In fact, since the number of polynomials is $3 = \dim P_2$, we know the two facts imply each other. We do not need to make the argument twice.

Example 4.2.10. We claim $t(t-1), t(t-2), (t-1)(t-2)$ is a basis of P_2. Since $\dim P_2 = 3$, we only need to show thatthat are linearly independent. Consider

$$x_1 t(t-1) + x_2 t(t-2) + x_3 (t-1)(t-2) = 0.$$

Taking $t=0$, we get $x_3(-1)(-2)=0$, which implies $x_3=0$. Similarly, by taking $t=1$ and $t=2$, we get $x_2=0$ and $x_1=0$. Therefore the three polynomials are linearly independent.

In general, suppose $t_0, t_1, \dots, t_n$ are distinct, and[1]

$$p_i(t) = \prod_{j\neq i}(t-t_j) = (t-t_0)(t-t_1)\cdots\widehat{(t-t_i)}\cdots(t-t_n)$$

is the product of all $t - t_*$ except $t-t_i$. Then $p_0(t), p_1(t), \dots, p_n(t)$ form a basis of P_n.

Exercise 4.22. Explain that the vectors do not span the vector space.

1. $3+\sqrt{2}t-\pi t^2-3t^3$, $e+100t+2\sqrt{3}t^2$, $4\pi t-15.2t^2+t^3$.
2. $\begin{pmatrix}3 & 8\\ 4 & 9\end{pmatrix}, \begin{pmatrix}2 & 8\\ 6 & 5\end{pmatrix}, \begin{pmatrix}4 & 7\\ 5 & 0\end{pmatrix}$.
3. $\begin{pmatrix}\pi & \sqrt{3}\\ 1 & 2\pi\end{pmatrix}, \begin{pmatrix}\sqrt{2} & \pi\\ -10 & 2\sqrt{2}\end{pmatrix}, \begin{pmatrix}3 & 100\\ -77 & 6\end{pmatrix}, \begin{pmatrix}\sin 2 & \pi\\ \sqrt{2}\pi & 2\sin 2\end{pmatrix}$.

Exercise 4.23. Explain that the vectors are linearly dependent.

1. $3+\sqrt{2}t-\pi t^2$, $e+100t+2\sqrt{3}t^2$, $4\pi t-15.2t^2$, $\sqrt{\pi}+e^2t^2$.
2. $\begin{pmatrix}3 & 8\\ 4 & 9\end{pmatrix}, \begin{pmatrix}2 & 8\\ 6 & 5\end{pmatrix}, \begin{pmatrix}1 & 0\\ -2 & 4\end{pmatrix}, \begin{pmatrix}5 & -1\\ 3 & 0\end{pmatrix}, \begin{pmatrix}0 & 3\\ -4 & 1\end{pmatrix}$.
3. $\begin{pmatrix}\pi & \sqrt{3}\\ 1 & 2\pi\end{pmatrix}, \begin{pmatrix}\sqrt{2} & \pi\\ -10 & 2\sqrt{2}\end{pmatrix}, \begin{pmatrix}3 & 100\\ -77 & 6\end{pmatrix}, \begin{pmatrix}\sin 2 & \pi\\ \sqrt{2}\pi & 2\sin 2\end{pmatrix}$.

Exercise 4.24. Suppose $\text{Span}\alpha = \text{Span}\beta$, and α is linearly independent. Prove that β is also linearly independent if and only if $\#\alpha = \#\beta$.

4.2.5 Sum and Direct Sum

All the discussions in Section 2.5 remain valid for general vector space. The *sum* of subspaces $H_1, H_2, \dots, H_n$ of a vector space V is

$$H_1+H_2+\cdots+H_n = \{\vec{x}_1+\vec{x}_2+\cdots+\vec{x}_n \colon \vec{x}_i \in H_i\}.$$

This is still a subspace of V.

The sum is *direct*, in which case we denote the sum by $H_1\oplus H_2\oplus\cdots\oplus H_n$, if for $\vec{x}_i, \vec{y}_i \in H_i$, we have

$$\vec{x}_1+\vec{x}_2+\cdots+\vec{x}_n = \vec{y}_1+\vec{y}_2+\cdots+\vec{y}_n \implies \vec{x}_i = \vec{y}_i.$$

As argued in Example 2.5.1, a single vector space, considered as a sum of one subspace, is always direct. Moreover, H_1+H_2 is direct if and only if $H_1\cap H_2 = \{\vec{0}\}$.

[1]The notation $\hat{?}$ is the mathematical convention that the term ? is missing.

Example 4.2.11. Let H and H' be the subspaces of symmetric and skew-symmetric $n \times n$ matrices in Example 4.1.6 and Exercise 4.5. We claim $M(n,n) = H \oplus H'$.

We wish to write

$$X = A + B, \quad A^T = A\ (A \in H), \quad B^T = -B\ (B \in H').$$

This implies

$$X^T = A^T + B^T = A - B, \quad A = \tfrac{1}{2}(X + X^T), \quad B = \tfrac{1}{2}(X - X^T).$$

Therefore the decomposition $X = A + B$ is unique. Moreover, we have

$$\begin{aligned} A \in H &\colon A^T = \tfrac{1}{2}(X + X^T)^T = \tfrac{1}{2}(X^T + (X^T)^T) = \tfrac{1}{2}(X^T + X) = A, \\ B \in H' &\colon B^T = \tfrac{1}{2}(X - X^T)^T = \tfrac{1}{2}(X^T - (X^T)^T) = \tfrac{1}{2}(X^T - X) = -B. \end{aligned}$$

Therefore $M(n,n) = H \oplus H'$.

Exercise 4.25. Let H and H' be the subspaces of even and odd functions in C^∞, in Example 4.1.5. Prove that $C^\infty = H \oplus H'$.

Example 4.2.12 (Abstract Direct Sum)**.** Let V and W be vector spaces. We introduce addition and scalar multiplication on the cartesian product $V \times W = \{(\vec{x}, \vec{y}) \colon \vec{x} \in V,\ \vec{y} \in W\}$

$$(\vec{x}_1, \vec{y}_1) + (\vec{x}_2, \vec{y}_2) = (\vec{x}_1 + \vec{x}_2, \vec{y}_1 + \vec{y}_2), \quad a(\vec{x}, \vec{y}) = (a\vec{x}, a\vec{y}).$$

It is easy to verify that $V \times W$ is a vector space. Moreover, V and W are isomorphic to the following subspaces of $V \times W$

$$V \leftrightarrow V \times \vec{0} = \{(\vec{x}, \vec{0}_W) \colon \vec{x} \in V\}, \quad W \leftrightarrow \vec{0} \times W = \{(\vec{0}_V, \vec{y}) \colon \vec{y} \in W\}.$$

Since any vector in $V \times W$ can be uniquely expressed as $(\vec{x}, \vec{y}) = (\vec{x}, \vec{0}_W) + (\vec{0}_V, \vec{y})$, we get the direct sum

$$V \times W = (V \times \vec{0}) \oplus (\vec{0} \times W).$$

Since the right side is essentially $V \oplus W$, we call the vector space $V \times W$ the *abstract direct sum* and denote it by $V \oplus W$.

The construction allows us to write $\mathbb{R}^m \oplus \mathbb{R}^n = \mathbb{R}^{m+n}$. Strictly speaking, $\mathbb{R}^m$ and $\mathbb{R}^n$ are not subspaces of $\mathbb{R}^{m+n}$. The equality means

1. $\mathbb{R}^m$ is isomorphic to the subspace of vectors in $\mathbb{R}^{m+n}$ with the last n coordinates vanishing.

2. $\mathbb{R}^n$ is isomorphic to the subspace of vectors in $\mathbb{R}^{m+n}$ with the first m coordinates vanishing.

3. $\mathbb{R}^{m+n}$ is the direct sum of these two subspaces.

4.3 Linear Transformation

4.3.1 Linear Transformation between General Vector Spaces

The definition of linear transformation between Euclidean spaces in Section 3.1.1 can be directly extended to general vector spaces.

Definition 4.3.1. A transformation $L\colon V \to W$ between vector spaces is *linear* if

$$L(\vec{x}+\vec{y}) = L(\vec{x}) + L(\vec{y}), \quad L(a\vec{x}) = aL(\vec{x}).$$

A linear transformation preserves linear combinations. We call a linear transformation $L\colon V \to V$ a *linear operator* on V.

Example 4.3.1. The identity map $I(\vec{x}) = \vec{x}\colon V \to V$ is a linear operator.
The scaling map $L(\vec{x}) = a\vec{x}\colon V \to V$ is a linear operator.
The zero map $O(\vec{x}) = \vec{0}\colon V \to W$ is also a linear transformation.

Example 4.3.2. Proposition 4.2.5 says that the α-coordinate map in Section 4.2.2 is a linear transformation.

Example 4.3.3. Let A be a 2×3 matrix. The multiplication on the left by A is a linear transformation

$$L_A(X) = AX\colon M(3,5) \to M(2,5).$$

The linear transformation means

$$L_A(aX+bY) = A(aX+bY) = aAX + bAY = aL_A(X) + bL_A(Y).$$

The reason for the equality is the corresponding property for the composition of linear transformations between Euclidean spaces in Exercise 3.23.

Similarly, the right multiplication by a matrix is also a linear transformation (A is 5×2)

$$R_A(X) = XA\colon M(3,5) \to M(3,2).$$

A special case is that A is a fixed $n \times 1$ matrix, which is a vertical vector $\vec{v} \in \mathbb{R}^n$. Then the right multiplication is the evaluation map (A below is X above)

$$A \mapsto A\vec{v}\colon M(m,n) \to \mathbb{R}^m.$$

This is is a linear transformation.

Behind the discussion above is that the matrix multiplication

$$\mu(X,Y) = XY\colon M(m,n) \times M(n,k) \to M(m,k)$$

is a *bilinear map*: If we fix Y, then XY is linear in X. If we fix X, then XY is linear in Y.

Example 4.3.4. Exercise 3.16 shows $(A+B)^T = A^T + B^T$ and $(aA)^T = aA^T$. Therefore the transpose is a linear transformation between vector spaces of matrices

$$A \mapsto A^T \colon M(m,n) \to M(n,m).$$

Example 4.3.5. The *trace* of a square matrix is the sum of its diagonal entries

$$\mathrm{tr}\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = a_{11} + a_{22} + a_{33}.$$

The trace is a linear transformation

$$\mathrm{tr}\colon M(n,n) \to \mathbb{R}.$$

The trace clearly satisfies $\mathrm{tr}A^T = \mathrm{tr}A$.

For a 2×3 matrix A and a 3×2 matrix B, we have

$$\begin{aligned} AB &= \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{pmatrix} \\ &= \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \end{pmatrix} \\ BA &= \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \\ &= \begin{pmatrix} b_{11}a_{11} + b_{12}a_{21} & b_{11}a_{12} + b_{12}a_{22} & b_{11}a_{13} + b_{12}a_{23} \\ b_{21}a_{11} + b_{22}a_{21} & b_{21}a_{12} + b_{22}a_{22} & b_{21}a_{13} + b_{22}a_{23} \\ b_{31}a_{11} + b_{32}a_{21} & b_{31}a_{12} + b_{32}a_{22} & b_{31}a_{13} + b_{32}a_{23} \end{pmatrix}. \end{aligned}$$

This shows

$$\mathrm{tr}AB = \sum_{i,j} a_{ij}b_{ji} = \mathrm{tr}BA.$$

Exercise 4.26. Extending Example 3.1.5, a linear transformation $l\colon V \to \mathbb{R}$ is called a *linear functional*. The trace is a linear functional on the vector space $M(n,n)$.

For an ordered basis $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ of V, show that linear functionals on V are

$$l(\vec{x}) = a_1x_2 + a_2x_2 + \cdots + a_nx_n, \quad [\vec{x}]_\alpha = (x_1, x_2, \dots, x_n).$$

Exercise 4.27. Explain that a linear transformation $L\colon \mathbb{R} \to V$ is the scaling of a vector.

Example 4.3.6. The evaluation of functions at t_0 is a map

$$E_{t_0}(f) = f(t_0)\colon \mathrm{Fun}(\mathbb{R}) \to \mathbb{R}.$$

The following shows the evaluation map is a linear transformation

$$E_{t_0}(af+bg) = (af+bg)(t_0) = af(t_0) + bg(t_0) = aE_{t_0}(f) + bE_{t_0}(g).$$

We may combine several evaluations and still get a linear transformation

$$L(f) = (f(0), f(1), f(2)) \colon \text{Fun}(\mathbb{R}) \to \mathbb{R}^3.$$

In the reverse direction, the linear combination of several functions is a linear transformation (see Exercise 4.28)

$$L(x_1, x_2, x_3) = x_1 \cos t + x_2 \sin t + x_3 e^t \colon \mathbb{R}^3 \to \text{Fun}(\mathbb{R}).$$

Exercise 4.28. Show that the linear combination map is a linear transformation (enough to show for $n = 3$)

$$(x_1, x_2, \dots, x_n) \in \mathbb{R}^n \mapsto x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n \in V.$$

Here we do not require $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ to be a basis.

Example 4.3.7. Taking the derivatives of smooth functions is a linear operator

$$f \mapsto f' \colon C^\infty \to C^\infty.$$

The integration is a linear operator

$$f(t) \mapsto \int_0^t f(\tau)d\tau \colon C^\infty \to C^\infty.$$

Multiplying a fixed smooth function $a(t)$ is a linear operator

$$f(t) \mapsto a(t)f(t) \colon C^\infty \to C^\infty.$$

Substituting the variable by a smooth function $u(t)$ is a linear operator

$$f(t) \mapsto f(u(t)) \colon C^\infty \to C^\infty.$$

Exercise 4.29. Determine whether maps are linear.

1. $f \mapsto f^2 \colon C^\infty \to C^\infty$.
2. $f(t) \mapsto f(t^2) \colon C^\infty \to C^\infty$.
3. $f \mapsto f'' \colon C^\infty \to C^\infty$.
4. $f(t) \mapsto f(t-2) \colon C^\infty \to C^\infty$.
5. $f(t) \mapsto f(2t) \colon C^\infty \to C^\infty$.
6. $f \mapsto f' + 2f \colon C^\infty \to C^\infty$.
7. $f \mapsto (f(0) + f(1), f(2)) \colon C^\infty \to \mathbb{R}^2$.
8. $f \mapsto f(0)f(1) \colon C^\infty \to \mathbb{R}$.
9. $f \mapsto \int_0^1 f(t)dt \colon C^\infty \to \mathbb{R}$.
10. $f \mapsto \int_0^t \tau f(\tau)d\tau \colon C^\infty \to C^\infty$.

4.3.2 Operations of Linear Transformations

In Section 3.2, we introduced the addition $L+K$, scalar multiplication aL, and composition $L \circ K$ of linear transformations between Euclidean spaces. The operations can be extended to linear transformations in general

$$(L+K)(\vec{x}) = L(\vec{x}) + K(\vec{x}), \quad (aL)(\vec{x}) = a(L(\vec{x})), \quad (L \circ K)(\vec{x}) = L(K(\vec{x})).$$

By the same argument as before, we know $L+K, aL, L \circ K$ are still linear transformations, and the operations satisfy many properties such as the following (see Exercises 3.18 and 3.23)

$$\begin{aligned}
L+K &= K+L, & (L+K)+M &= L+(K+M), \\
a(L+K) &= aL+aK, & (a+b)L &= aL+bL, \\
(ab)L &= a(bL), & (L \circ K) \circ M &= L \circ (K \circ M), \\
(aL+bK) \circ M &= aL \circ M + bK \circ M, & M \circ (aL+bK) &= aM \circ L + bM \circ K.
\end{aligned}$$

Exercise 4.30. The set of all linear transformations $V \to W$ is denoted $\mathrm{Hom}(V, W)$. With the operations $L+K$ and aL, shows that $\mathrm{Hom}(V, W)$ is a vector space.

Example 4.3.8. Consider the differential equation

$$(1+t^2)f'' + (1+t)f' - f = t + 2t^3.$$

The left is the addition of three transformations $f \mapsto (1+t^2)f''$, $f \mapsto (1+t)f'$, $f \mapsto -f$.

Let $D(f) = f'$ be the derivative linear transformation, and let $M_a(f) = af$ be the linear transformation of multiplying a function $a(t)$, in Example 4.3.7. Then $(1+t^2)f''$ is the composition $D^2 = M_{1+t^2} \circ D \circ D$, $(1+t)f'$ is the composition $M_{1+t} \circ D$, and $f \mapsto -f$ is the linear transformation M_{-1}. Therefore the left side of the differential equation is the linear transformation

$$L = M_{1+t^2} \circ D \circ D + M_{1+t} \circ D + M_{-1}.$$

The differential equation can be expressed as $L(f(t)) = b(t)$ with $b(t) = t+2t^3 \in C^\infty$.

In general, a *linear differential equation* of order n is

$$a_0(t)\frac{d^n f}{dt^n} + a_1(t)\frac{d^{n-1} f}{dt^{n-1}} + a_2(t)\frac{d^{n-2} f}{dt^{n-2}} + \cdots + a_{n-1}(t)\frac{df}{dt} + a_n(t)f = b(t).$$

If the coefficient functions $a_0(t), a_1(t), \dots, a_n(t)$ are smooth, then the left side is a linear transformation $C^\infty \to C^\infty$.

Exercise 4.31. Interpret the Newton-Leibniz formula $f(t) = f(0) + \int_0^t f'(\tau)d\tau$ as an equality of linear transformations.

4.3.3 Range and Kernel

In Section 3.3.1, for a linear transformation between Euclidean spaces, the range and kernel are introduced as the counterpart of the column and null spaces. The concepts can be extended to a linear transformation $L\colon V \to W$ in general

$$\begin{aligned}\text{Ran}L &= \{L(\vec{x})\colon \vec{x} \in V\} = L(V) \subset W,\\ \text{Ker}L &= \{\vec{x} \in V\colon L(\vec{x}) = \vec{0}\} = L^{-1}(\vec{0}) \subset V.\end{aligned}$$

These are subspaces.

The range corresponds to the column space, by Proposition 2.4.2, we define the rank of linear transformation

$$\text{rank}L = \dim \text{Ran}L.$$

Then L is onto if and only if $\text{Ran}L = W$, which is equivalent to $\text{rank}L = \dim W$.

The kernel corresponds to the null space. Then Proposition 2.4.5 becomes

$$\dim \text{Ker}L = \dim V - \text{rank}L, \quad \dim \text{Ran}L + \dim \text{Ker}L = \dim V.$$

L is one-to-one if and only if $\text{Ker}L = \{\vec{0}\}$.

More generally, for a subspaces $H \subset V$ or $H' \subset W$, we have the image and preimage

$$\begin{aligned}L(H) &= \{L(\vec{x})\colon \vec{x} \in H\} \subset W,\\ L^{-1}(H') &= \{\vec{x} \in V\colon L(\vec{x}) \in H'\} \subset V.\end{aligned}$$

These are also subspaces.

Example 4.3.9. Consider the linear transformation in Example 4.3.3

$$L_A(X) = AX\colon M(3,5) \to M(2,5).$$

We think of $X = (\vec{x}_1\ \vec{x}_2\ \vec{x}_3\ \vec{x}_4\ \vec{x}_5)$ as five vectors $\vec{x}_i \in \mathbb{R}^3$, and regard L_A as

$$L_A(X) = A(\vec{x}_1\ \vec{x}_2\ \vec{x}_3\ \vec{x}_4\ \vec{x}_5) = (A\vec{x}_1\ A\vec{x}_2\ A\vec{x}_3\ A\vec{x}_4\ A\vec{x}_5)\colon (\mathbb{R}^3)^5 \to (\mathbb{R}^2)^5.$$

This shows that

$$\begin{aligned}\text{Ran}L_A &= \{(A\vec{x}_1\ A\vec{x}_2\ A\vec{x}_3\ A\vec{x}_4\ A\vec{x}_5)\colon \vec{x}_i \in \mathbb{R}^3\} = (\text{Col}A)^5,\\ \text{Ker}L_A &= \{(\vec{x}_1\ \vec{x}_2\ \vec{x}_3\ \vec{x}_4\ \vec{x}_5)\colon \vec{x}_i \in \mathbb{R}^3,\ A\vec{x}_i = \vec{0}\} = (\text{Nul}A)^5.\end{aligned}$$

We conclude

$$\text{rank}L_A = \dim \text{Ran}L_A = 5 \dim \text{Col}A = 5\, \text{rank}A.$$

In general, If A is an $m \times n$ matrix, then we have linear transformation $L_A(X) = AX\colon M(n,k) \to M(m,k)$, and $\text{rank}L_A = k\, \text{rank}A$.

Exercise 4.32. Find the range, kernel, and rank of the trace in Example 4.3.5. You may work on 3×3 matrices.

Exercise 4.33. Find the range, kernel, and rank of the following linear operators on P_3.

1. $f'(t)$.
2. $tf'(t)$.
3. $2f(t) - tf'(t)$.
4. $2f(t) + tf'(t)$.

Example 4.3.10. Since the identity and the transpose are linear transformations, we know

$$L(X) = X - X^T\colon M(n,n) \to M(n,n)$$

is a linear transformation. We have

$$\mathrm{Ker}L = \{X\colon L(X) = X - X^T = O\} = \{X\colon X = X^T\} = \{\text{symmetric matrices}\}.$$

By

$$L(X)^T = X^T - (X^T)^T = X^T - X = -L(X),$$

we know $\mathrm{Ran}L \subset \{\text{skew-symmetric matrices}\}$. Conversely, suppose $A^T = -A$, then for $X = \frac{1}{2}A$, we get

$$L(X) = \tfrac{1}{2}A - \tfrac{1}{2}A^T = \tfrac{1}{2}A + \tfrac{1}{2}A = A.$$

Therefore we get

$$\mathrm{Ran}L = \{\text{skew-symmetric matrices}\}.$$

Exercise 4.34. Find the range and kernel of $K(X) = X + X^T$.

Exercise 4.35. Let $L(f(t)) = f(t) - f(-t)$ and $K(f(t)) = f(t) + f(-t)$.

1. Explain that L and K are linear operators on C^∞.
2. Show that $\mathrm{Ker}L$ and $\mathrm{Ran}K$ are the subspace of even functions.
3. Show that $\mathrm{Ran}L$ and $\mathrm{Ker}K$ are the subspace of odd functions.

Exercise 4.36. For the subspaces in Exercise 4.3, can you interpret them as the kernel or the range of suitable linear transformations?

4.3.4 Isomorphism

Definition 4.3.2. A linear transformation is an *isomorphism*, if it is invertible. Two vector spaces are *isomorphic* if there is an isomorphism between them.

We denote isomorphism by $V \cong W$.

Using bases of V and W, we get $V \cong \mathbb{R}^n$ and $W \cong \mathbb{R}^m$, where $n = \dim V$ and $m = \dim W$. Then $V \cong W$ is the same as $\mathbb{R}^n \cong \mathbb{R}^m$. An isomorphism $\mathbb{R}^n \cong \mathbb{R}^m$ is given by an invertible $m \times n$ matrix. Therefore the condition for $V \cong W$ is $m = n$.

Theorem 4.3.3. *Two finite dimensional vector spaces are isomorphic if and only if they have the same dimension.*

Example 4.3.11. The identity transformation $I: V \to V$ is an isomorphism. More generally, for any $a \neq 0$, the scalar multiplication $aI: V \to V$ is an isomorphism.

Example 4.3.12. The transpose of matrices $A \in M(m,n) \mapsto A^T \in M(n,m)$ is an isomorphism. The inverse is the transpose $M(n,m) \to M(m,n)$.

Example 4.3.13. As explained in Section 4.2.2, the coordinate map

$$\vec{x} \in V \mapsto [\vec{x}]_\alpha \in \mathbb{R}^n$$

is an isomorphism.

Example 4.3.14. In Section 4.3.2, we introduced the vector space $\mathrm{Hom}(V,W)$ of all linear transformations $V \to W$. Then Section 3.1.2 gives an invertible map

$$[L(\vec{x}) = A\vec{x}] \in \mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m) \longleftrightarrow A = [L] = (L(\vec{e}_1)\ L(\vec{e}_2)\ \cdots\ L(\vec{e}_n)) \in M(m,n).$$

In Section 3.2.2, we introduce the addition and scalar multiplication of matrices on the right as corresponding to the addition and scalar multiplication of linear transformations on the left. This means exactly that $\mathrm{Hom}(\mathbb{R}^n, \mathbb{R}^m) \longleftrightarrow M(m,n)$ is a linear transformation. Therefore the map is an isomorphism.

Since general vector spaces are isomorphic to Euclidean spaces, all the results about linear transformations between Euclidean spaces remain valid for general linear transformations. As an example, Theorem 3.4.4 says that, if $\dim V = \dim W$, then the following are equivalent for a linear transformation $L: V \to W$.

1. L is an isomorphism.
2. L has right inverse: $L \circ K = I$ for some $K: W \to V$.
3. L has left inverse: $K \circ L = I$ for some $K: W \to V$.

The two K are equal and are the inverse L^{-1}.

Example 4.3.15 (Lagrange interpolation)**.** In Example 4.2.10, we considered the evaluation of quadratic polynomials at three distinct locations t_0, t_1, t_2

$$L(f(t)) = (f(t_0), f(t_1), f(t_2)) \colon P_2 \to \mathbb{R}^3.$$

We express vectors on the right as $(x_0, x_1, x_2) \in \mathbb{R}^3$, which has the standard basis

$$\vec{e}_0 = (1,0,0), \quad \vec{e}_1 = (0,1,0), \quad \vec{e}_2 = (0,0,1).$$

We claim that L is onto by finding polynomials $p_i(t)$ satisfying $L(p_i(t)) = \vec{e}_i$. The quadratic polynomial $p_0(t)$ satisfies

$$p_0(t_0) = 1, \quad p_0(t_1) = 0, \quad p_0(t_2) = 0.$$

The second and third equalities are easily satisfied by the quadratic polynomial $(t-t_1)(t-t_2)$. Then we may multiply a suitable constant to further satisfy the first equality. This leads to

$$p_0(t) = \frac{(t-t_1)(t-t_2)}{(t_0-t_1)(t_0-t_2)}.$$

By the similar idea, we get

$$p_1(t) = \frac{(t-t_0)(t-t_2)}{(t_1-t_0)(t_1-t_2)}, \quad p_2(t) = \frac{(t-t_0)(t-t_1)}{(t_2-t_0)(t_2-t_1)}.$$

The polynomials $p_i(t)$ are constructed to have special values $\vec{e}_i$ at t_0, t_1, t_2. To get a quadratic polynomial $f(t)$ with general values x_0, x_1, x_2

$$L(f(t)) = (f(t_0), f(t_1), f(t_2)) = (x_0, x_1, x_2) = x_0\vec{e}_0 + x_1\vec{e}_1 + x_2\vec{e}_2,$$

we may simply take the corresponding linear combination

$$\begin{aligned} f(t) &= x_0p_0(t) + x_1p_1(t) + x_2p_2(t) \\ &= x_0\frac{(t-t_1)(t-t_2)}{(t_0-t_1)(t_0-t_2)} + x_1\frac{(t-t_0)(t-t_2)}{(t_1-t_0)(t_1-t_2)} + x_2\frac{(t-t_0)(t-t_1)}{(t_2-t_0)(t_2-t_1)}. \end{aligned}$$

For example, the quadratic polynomial

$$f(t) = 5\frac{t(t-1)}{(-1)\cdot(-2)} + 6\frac{(t+1)(t-1)}{1\cdot(-1)} + 7\frac{(t+1)t}{2\cdot 1} = t+6$$

satisfies $f(-1) = 5, f(0) = 6, f(1) = 7$.

We have shown that L is onto. Then by $\dim P_2 = 3 = \dim \mathbb{R}^3$ and Theorem 3.4.3, we know L is an isomorphism. The inverse of L is $L^{-1}(x_0, x_1, x_2) = x_0p_0(t) + x_1p_1(t) + x_2p_2(t)$, given by the *Lagrange interpolation* formula above.

The discussion is of course not restricted to quadratic polynomials. The evaluation of degree n polynomials at $n+1$ distinct locations $t_0, t_1, \dots, t_n$

$$L(f(t)) = (f(t_0), f(t_1), \dots, f(t_n)) \colon P_n \to \mathbb{R}^{n+1}$$

is also an isomorphism. The inverse linear transformation $L^{-1} \colon \mathbb{R}^{n+1} \to P_n$ is given by

$$L^{-1}(x_0, x_1, \dots, x_n) = \sum_{i=0}^{n} x_i \prod_{0\le j\le n,\ j\ne i} \frac{t-t_j}{t_i-t_j}.$$

Exercise 4.37. Find polynomial with prescribed values.

1. Quadratic $f(t)$ satisfying $f(0) = 2, f(1) = 1, f(2) = 0$.

2. Cubic $f(t)$ satisfying $f(-1) = 2, f(0) = -2, f(1) = 1, f(2) = -1$.

Exercise 4.38. Explain that the linear transformation

$$f \in P_n \mapsto (f(t_0), f'(t_0), \dots, f^{(n)}(t_0)) \in \mathbb{R}^{n+1}$$

is an isomorphism. What is the inverse isomorphism?

Exercise 4.39. Explain that the linear transformation (the right side has obvious vector space structure)

$$f \in C^\infty \mapsto (f', f(t_0)) \in C^\infty \times \mathbb{R}$$

is an isomorphism. This restricts to the isomorphism $L(f(t)) = (f'(t), f(0)) \colon P_n \to P_{n-1} \oplus \mathbb{R}$.

4.3.5 Matrix of Linear Transformation

Suppose $\alpha = \{\vec{v}_1, \vec{v}_2\}$ and $\beta = \{\vec{w}_1, \vec{w}_2, \vec{w}_3\}$ are (ordered) bases of vector spaces V and W. Then we can use the bases to translate a linear transformation $L \colon V \to W$ to a linear transformation[2] $L_{\beta\alpha} \colon \mathbb{R}^2 \to \mathbb{R}^3$.

$$\begin{tikzcd} V \arrow[r, "L"] \arrow[d, "{[\cdot]_\alpha}"', "\cong"] & W \arrow[d, "{[\cdot]_\beta}", "\cong"'] \\ \mathbb{R}^2 \arrow[r, "L_{\beta\alpha}"] & \mathbb{R}^3 \end{tikzcd} \qquad L_{\beta\alpha}([\vec{x}]_\alpha) = [L(\vec{x})]_\beta \text{ for } \vec{x} \in V.$$

We denote the matrix of the linear transformation $L_{\beta\alpha}$ by $[L]_{\beta\alpha}$, and call it the *matrix of L with respect to bases α and β*.

The linear transformation L is determined by its values on α, which can be written as linear combinations of β. Suppose

$$L(\vec{v}_1) = 1\vec{w}_1 + 2\vec{w}_2 + 3\vec{w}_3,$$
$$L(\vec{v}_2) = 4\vec{w}_1 + 5\vec{w}_2 + 6\vec{w}_3.$$

Applying $[\cdot]_\alpha$ to the left side and $[\cdot]_\beta$ to the right side, we get the corresponding formula for the Euclidean spaces

$$L_{\beta\alpha}(\vec{e}_1) = 1\vec{e}_1 + 2\vec{e}_2 + 3\vec{e}_3 = (1, 2, 3),$$
$$L_{\beta\alpha}(\vec{e}_2) = 4\vec{e}_1 + 5\vec{e}_2 + 6\vec{e}_3 = (4, 5, 6).$$

Here we remark that $\vec{e}_1 = (1, 0)$ and $\vec{e}_2 = (0, 1)$ on the left is the standard basis of $\mathbb{R}^2$. Then the matrix of $L_{\beta\alpha}$ is

$$[L]_{\beta\alpha} = (L_{\beta\alpha}(\vec{e}_1)\ L_{\beta\alpha}(\vec{e}_2)) = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix} = ([L(\vec{v}_1)]_\beta\ [L(\vec{v}_2)]_\beta) = [L(\alpha)]_\beta.$$

[2]We emphasis that $L_{\beta\alpha}$ is the corresponding linear transformation from α to β (order is reversed).

Therefore $[L]_{\beta\alpha}$ is obtained by applying L to α and then taking the the β-coordinates. For example,

$$L(\vec{v}_1) = a_{11}\vec{w}_1 + a_{21}\vec{w}_2 + a_{31}\vec{w}_3,$$
$$L(\vec{v}_2) = a_{12}\vec{w}_1 + a_{22}\vec{w}_2 + a_{32}\vec{w}_3,$$

gives

$$[L(\vec{v}_1)]_\beta = \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix}, \quad [L(\vec{v}_2)]_\beta = \begin{pmatrix} a_{12} \\ a_{22} \\ a_{32} \end{pmatrix}, \quad [L]_{\beta\alpha} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}.$$

Note that the matrix $[L]_{\beta\alpha}$ is obtained by combining all the coefficients in $L(\alpha)$ and then take the transpose.

Example 4.3.16. The orthogonal projection P of $\mathbb{R}^3$ to the plane $x + y + z = 0$ in Example 3.1.12 satisfies

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

for vectors in the basis

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

This means that

$$[P]_{\alpha\alpha} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

This is much simpler than the matrix $[P]_{\epsilon\epsilon}$ with respect to the standard basis ϵ obtained in Example 3.1.12.

Example 4.3.17. Let α be a basis of a subspace H of $\mathbb{R}^n$. The inclusion $\mathrm{incl}\colon H \to \mathbb{R}^n$ is a linear transformation. Then

$$[\mathrm{incl}]_{\epsilon\alpha} = [\mathrm{incl}(\alpha)]_\epsilon = [\alpha]_\epsilon = (\alpha).$$

Example 4.3.18. The evaluation of quadratic polynomials in Example 4.3.15 at three locations t_0, t_1, t_2 is a linear transformation

$$L(f(t)) = \begin{pmatrix} f(t_0) \\ f(t_1) \\ f(t_2) \end{pmatrix} : P_2 \to \mathbb{R}^3.$$

If we take the standard monomial bases $\alpha = \{1, t, t^2\}$ of P_2 and the standard basis $\epsilon = \{\vec{e}_0, \vec{e}_1, \vec{e}_2\}$ of $\mathbb{R}^3$, then the matrix of L with respect to the bases is

$$[L]_{\epsilon\{1,t,t^2\}} = [L(1), L(t), L(t^2)]_\epsilon = (L(1)\ L(t)\ L(t^2)) = \begin{pmatrix} 1 & t_0 & t_0^2 \\ 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \end{pmatrix}.$$

We call this the *Vandermonde matrix* and denote $V(t_0, t_1, t_2)$.

Exercise 4.40. Find the Vandermonde matrix for the evaluation of degree n polynomial at $n+1$ locations.

Exercise 4.41. Find the matrix of the evaluation of quadratic polynomials at four locations t_0, t_1, t_2, t_3 with respect to the usual basis of P_2 and the standard basis of $\mathbb{R}^4$. What about the evaluation at two locations t_0, t_1?

Example 4.3.19. With respect to the standard monomial bases $\alpha = \{1, t, t^2, t^3\}$ and $\beta = \{1, t, t^2\}$ of P_3 and P_2, the matrix of the derivative linear transformation $D\colon P_3 \to P_2$ is

$$[D]_{\beta\alpha} = [(1)', (t)', (t^2)', (t^3)']_{\{1,t,t^2\}} = [0, 1, 2t, 3t^2]_{\{1,t,t^2\}} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

For example, the derivative $(1 + 2t + 3t^2 + 4t^3)' = 2 + 6t + 12t^2$ fits into

$$\begin{pmatrix} 2 \\ 6 \\ 12 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}.$$

If we modify the basis β to $\gamma = \{1, t-1, (t-1)^2\}$ in Example 4.2.4, then

$$\begin{aligned} [D]_{\gamma\alpha} &= [0, 1, 2t, 3t^2]_{\{1,t-1,(t-1)^2\}} \\ &= [0, 1, 2(1 + (t-1)), 3(1 + (t-1))^2]_{\{1,t-1,(t-1)^2\}} \\ &= [0, 1, 2 + 2(t-1), 3 + 6(t-1) + 3(t-1)^2]_{\{1,t-1,(t-1)^2\}} \\ &= \begin{pmatrix} 0 & 1 & 2 & 3 \\ 0 & 0 & 2 & 6 \\ 0 & 0 & 0 & 3 \end{pmatrix}. \end{aligned}$$

Example 4.3.20. The linear transformation in Example 4.3.8

$$L(f) = (1 + t^2)f'' + (1 + t)f' - f\colon P_3 \to P_3$$

satisfies

$$L(1) = -1, \quad L(t) = 1, \quad L(t^2) = 2 + 2t + 3t^2, \quad L(t^3) = 6t + 3t^2 + 8t^3.$$

Therefore

$$[L]_{\{1,t,t^2,t^3\}\{1,t,t^2,t^3\}} = \begin{pmatrix} -1 & 1 & 2 & 0 \\ 0 & 0 & 2 & 6 \\ 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 8 \end{pmatrix}.$$

To solve the equation $L(f) = t + 2t^3$ in Example 4.3.8, we do row operations

$$\begin{pmatrix} -1 & 1 & 2 & 0 & 0 \\ 0 & 0 & 2 & 6 & 1 \\ 0 & 0 & 3 & 3 & 0 \\ 0 & 0 & 0 & 8 & 2 \end{pmatrix} \to \begin{pmatrix} 1 & -1 & -2 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 2 & 6 & 1 \\ 0 & 0 & 0 & 8 & 2 \end{pmatrix}$$
$$\to \begin{pmatrix} 1 & -1 & -2 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 4 & 1 \\ 0 & 0 & 0 & 8 & 2 \end{pmatrix} \to \begin{pmatrix} 1 & -1 & 0 & 0 & -\frac{1}{2} \\ 0 & 0 & 1 & 0 & -\frac{1}{4} \\ 0 & 0 & 0 & 1 & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

This shows that L is neither one-to-one nor onto. Moreover, the solution of the differential equation is given by

$$a_3 = \tfrac{1}{4}, \quad a_2 = -\tfrac{1}{4}, \quad a_0 = -\tfrac{1}{2} + a_1.$$

In other words, the solution is

$$f = -\tfrac{1}{2} + a_1 + a_1 t - \tfrac{1}{4}t^2 + \tfrac{1}{4}t^3 = \tfrac{1}{4}(-2 - t^2 + t^3) + a_1(1 + t).$$

Exercise 4.42. In Example 4.3.19, what is the matrix of the derivative linear transformation if α is changed to $\{1, t+1, (t+1)^2, (t+1)^3\}$?

Exercise 4.43. Find the matrix of linear transformation with respect to the usual basis.

1. Integral $f(t) \mapsto \int_0^t f(\tau)d\tau \colon P_2 \to P_3$.
2. Integral $f(t) \mapsto \int_1^t f(\tau)d\tau \colon P_2 \to P_3$.
3. Multiplying a factor $f(t) \mapsto (2t-1)f(t) \colon P_2 \to P_3$.
4. Change of variable $f(t) \mapsto f(2t-1) \colon P_2 \to P_2$.
5. Change of variable $f(t) \mapsto f(2t^2-1) \colon P_2 \to P_4$.

Example 4.3.21. Consider the basis of $M(2,2)$

$$\sigma = \{S_1, S_2, S_3, S_4\}, \quad S_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \ S_2 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \ S_3 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \ S_4 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

The matrix of the transpose linear transformation $\cdot^T \colon M(2,2) \to M(2,2)$ with respect to σ is

$$\begin{aligned} [\cdot^T]_{\sigma\sigma} = [\sigma^T]_\sigma &= [S_1^T, S_2^T, S_3^T, S_4^T]_{\{S_1,S_2,S_3,S_4\}} \\ &= [S_1, S_3, S_2, S_4]_{\{S_1,S_2,S_3,S_4\}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \end{aligned}$$

Moreover, the matrix of the linear transformation $A\cdot\colon M(2,2) \to M(2,2)$ of the left multiplication by $A = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$ is

$$\begin{aligned}
[A\cdot]_{\sigma\sigma} = [A\sigma]_\sigma &= [AS_1, AS_2, AS_3, AS_4]_{\{S_1,S_2,S_3,S_4\}} \\
&= [\begin{pmatrix} 1 & 0 \\ 2 & 0 \end{pmatrix}, \begin{pmatrix} 3 & 0 \\ 4 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 2 \end{pmatrix}, \begin{pmatrix} 0 & 3 \\ 0 & 4 \end{pmatrix}]_{\{S_1,S_2,S_3,S_4\}} \\
&= \begin{pmatrix} 1 & 3 & 0 & 0 \\ 2 & 4 & 0 & 0 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 2 & 4 \end{pmatrix}.
\end{aligned}$$

Exercise 4.44. In Example 4.3.21, find the matrix of the right multiplication by A. What about the matrix of the right multiplication by A^T?

Exercise 4.45. Repeating a vector is a linear transformation

$$\vec{x} \in \mathbb{R}^3 \mapsto (\vec{x}\ \vec{x}) \in M(3,2).$$

Find the matrix of the linear transformation with respect to the standard basis of $\mathbb{R}^3$ and the basis of $M(3,2)$

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

Exercise 4.46. Fixing a vector $\vec{v} = \begin{pmatrix} a \\ b \end{pmatrix}$, we get a linear transformation

$$A \in M(3,2) \mapsto A\vec{v} \in \mathbb{R}^3.$$

Find the matrix of the linear transformation with respect to the standard basis of $\mathbb{R}^3$ and the basis of $M(3,2)$ in Exercise 4.45.

We remark that, for $\vec{v} = \vec{e}_i$, the linear transformation means picking the i-th column of A. What about the linear transformation of picking the i-th row of A?

Exercise 4.47. The row operations are linear transformations. Find the matrix of the following row operations on 3×2 matrices with respect to the basis of $M(3,2)$ in Exercise 4.45.

1. $\text{Row}_1 \leftrightarrow \text{Row}_3$.
2. $\text{Row}_2 \to -3\text{Row}_2$.
3. $\text{Row}_2 + 3\text{Row}_1$.

Proposition 4.3.4. *The matrix of linear transformation has the following properties*

$$[I]_{\alpha\alpha} = I, \quad [L+K]_{\beta\alpha} = [L]_{\beta\alpha} + [K]_{\beta\alpha}, \quad [aL]_{\beta\alpha} = a[L]_{\beta\alpha},$$

$$[L \circ K]_{\gamma\alpha} = [L]_{\gamma\beta}[K]_{\beta\alpha}, \quad [L^{-1}]_{\alpha\beta} = [L]_{\beta\alpha}^{-1}.$$

The equalities are the consequences of the equalities for the corresponding linear transformations between Euclidean spaces

$$I_{\alpha\alpha} = I, \quad (L+K)_{\beta\alpha} = L_{\beta\alpha} + K_{\beta\alpha}, \quad (aL)_{\beta\alpha} = aL_{\beta\alpha},$$
$$(L\circ K)_{\gamma\alpha} = L_{\gamma\beta}\circ K_{\beta\alpha}, \quad (L^{-1})_{\alpha\beta} = (L_{\beta\alpha})^{-1}.$$

To verify the equalities, we use the definition

$$L_{\beta\alpha}([\vec{x}]_\alpha) = [L(\vec{x})]_\beta.$$

We have

$$I_{\alpha\alpha}([\vec{x}]_\alpha) = [I(\vec{x})]_\alpha = [\vec{x}]_\alpha.$$

Therefore $I_{\alpha\alpha}$ is the identity transformation. We also have

$$\begin{aligned}(L\circ K)_{\gamma\alpha}([\vec{x}]_\alpha) &= [(L\circ K)(\vec{x})]_\gamma,\\ (L_{\gamma\beta}\circ K_{\beta\alpha})([\vec{x}]_\alpha) &= L_{\gamma\beta}(K_{\beta\alpha}([\vec{x}]_\alpha)) = L_{\gamma\beta}([K(\vec{x})]_\beta) = [L(K(\vec{x}))]_\gamma.\end{aligned}$$

Since the right sides are equal, we get $(L\circ K)_{\gamma\alpha} = L_{\gamma\beta}\circ K_{\beta\alpha}$.

Example 4.3.22. In Example 4.3.15, we know the evaluation of quadratic polynomials at three distinct locations $L(f(t)) = (f(t_0), f(t_1), f(t_2))\colon P_2 \to \mathbb{R}^3$ is invertible. In Example 4.3.18, we find the matrix of the linear transformation with respect to the standard bases is the Vandermonde matrix

$$[L]_{\epsilon\{1,t,t^2\}} = V(t_0,t_1,t_2) = \begin{pmatrix} 1 & t_0 & t_0^2 \\ 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \end{pmatrix}.$$

The inverse of the Vandermonde matrix is

$$\begin{aligned}V(t_0,t_1,t_2)^{-1} &= [L]^{-1}_{\epsilon\{1,t,t^2\}} = [L^{-1}]_{\{1,t,t^2\}\epsilon}\\ &= [L^{-1}(\vec{e}_0), L^{-1}(\vec{e}_1), L^{-1}(\vec{e}_2)]_{\{1,t,t^2\}} = [p_0(t), p_1(t), p_2(t)]_{\{1,t,t^2\}}.\end{aligned}$$

Here $p_i(t)$ is obtained in Example 4.3.15. In particular, we have

$$p_0(t) = \frac{(t-t_1)(t-t_2)}{(t_0-t_1)(t_0-t_2)} = \frac{t_1t_2 - (t_1+t_2)t + t^2}{(t_0-t_1)(t_0-t_2)},$$

and

$$[p_0(t)]_{\{1,t,t^2\}} = \frac{1}{(t_0-t_1)(t_0-t_2)} \begin{pmatrix} t_1t_2 \\ -(t_1+t_2) \\ 1 \end{pmatrix}.$$

We similarly get $[p_1(t)]$ and $[p_2(t)]$, and then get

$$\begin{pmatrix} 1 & t_0 & t_0^2 \\ 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{t_1t_2}{(t_0-t_1)(t_0-t_2)} & \frac{t_0t_2}{(t_1-t_0)(t_1-t_2)} & \frac{t_0t_1}{(t_2-t_0)(t_2-t_1)} \\ \frac{-(t_1+t_2)}{(t_0-t_1)(t_0-t_2)} & \frac{-(t_0+t_2)}{(t_1-t_0)(t_1-t_2)} & \frac{-(t_0+t_1)}{(t_2-t_0)(t_2-t_1)} \\ \frac{1}{(t_0-t_1)(t_0-t_2)} & \frac{1}{(t_1-t_0)(t_1-t_2)} & \frac{1}{(t_2-t_0)(t_2-t_1)} \end{pmatrix}.$$

Exercise 4.48. Verify $(L+K)_{\beta\alpha} = L_{\beta\alpha} + K_{\beta\alpha}$, $(aL)_{\beta\alpha} = aL_{\beta\alpha}$, and $(L^{-1})_{\alpha\beta} = (L_{\beta\alpha})^{-1}$.

Exercise 4.49. Find the matrices of the derivative linear transformations $D_4\colon P_4 \to P_3$, $D_3\colon P_3 \to P_2$, and the second order derivative linear transformation $D_3D_4\colon P_4 \to P_2$, with respect to the bases $\{1, t, t^2, \dots, t^n\}$. Then verify $[D_3D_4] = [D_3][D_4]$.

Exercise 4.50. The derivative and integral linear transformations in Example 4.3.7 give linear transformations of polynomials

$$D(f) = f'\colon P_3 \to P_2, \quad J(f) = \int_0^t f(\tau)d\tau\colon P_2 \to P_3.$$

1. What does the fundamental theorem of calculus tell you about $D \circ J\colon P_2 \to P_2$ and $J \circ D\colon P_3 \to P_3$?

2. Interpret the first part as two matrix equalities.

Exercise 4.51. For the matrix A in Example 4.3.21, we have the linear transformations $X \mapsto AXA$ and $X \mapsto AXA^T$. Find the matrices of the linear transformations in two ways and compare.

1. Direct calculation like in Example 4.3.21.

2. Use the matrix of left multiplication in Example 4.3.21 and the matrix of right multiplication in Exercise 4.44.

4.3.6 Change of Basis

The matrix $[L]_{\beta\alpha}$ depends on the choice of (ordered) bases α and β. If α' and β' are also bases, then by Proposition 4.3.4, the matrix of L with respect to the new choice is

$$[L]_{\beta'\alpha'} = [I \circ L \circ I]_{\beta'\alpha'} = [I]_{\beta'\beta}[L]_{\beta\alpha}[I]_{\alpha\alpha'}.$$

Therefore the matrix of linear transformation is modified by multiplying matrices $[I]_{\alpha\alpha'}$ and $[I]_{\beta'\beta}$ of the identity operator with respect to difference choices of bases. The two matrices are the *matrices for the change of basis*. The following calculation shows it is the coordinates of vectors in one basis with respect to the other basis

$$[I]_{\alpha\alpha'} = [I(\alpha')]_{\alpha} = [\alpha']_{\alpha}.$$

Proposition 4.3.4 implies the following properties.

Proposition 4.3.5. *The matrix for the change of basis has the following properties*

$$[I]_{\alpha\alpha} = I, \quad [I]_{\beta\alpha} = [I]_{\alpha\beta}^{-1}, \quad [I]_{\gamma\alpha} = [I]_{\gamma\beta}[I]_{\beta\alpha}.$$

Example 4.3.23. Let ϵ be the standard basis of $\mathbb{R}^n$, and let $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ be another basis. Then the matrix for changing from α to ϵ is

$$[I]_{\epsilon\alpha} = [\alpha]_\epsilon = (\vec{v}_1 \ \vec{v}_2 \ \cdots \ \vec{v}_n) = (\alpha).$$

In general, the matrix for changing from α to β is

$$[I]_{\beta\alpha} = [I]_{\beta\epsilon}[I]_{\epsilon\alpha} = [I]_{\epsilon\beta}^{-1}[I]_{\epsilon\alpha} = [\beta]_\epsilon^{-1}[\alpha]_\epsilon = (\beta)^{-1}(\alpha).$$

For example, we have the following bases from Examples 2.4.2, 3.1.12, 3.4.6

$$\alpha = \{(1,2,3),(4,5,6),(7,8,10)\}, \quad \beta = \{(1,-1,0),(1,0,-1),(1,1,1)\}.$$

The matrix for changing from α to β is

$$\begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 10 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 1 & -2 & 1 \\ 1 & 1 & -2 \\ 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 10 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 0 & 0 & 1 \\ -3 & -3 & -5 \\ 6 & 15 & 25 \end{pmatrix}.$$

Example 4.3.24. The matrix for the change from the basis $\alpha = \{1, t, t^2, t^3\}$ of P_3 to another basis $\beta = \{1, t-1, (t-1)^2, (t-1)^3\}$ is

$$\begin{aligned} [I]_{\beta\alpha} &= [1, t, t^2, t^3]_{\{1,t-1,(t-1)^2,(t-1)^3\}} \\ &= [1, 1+(t-1), 1+2(t-1)+(t-1)^2, \\ &\quad\ 1+3(t-1)+3(t-1)^2+(t-1)^3]_{\{1,t-1,(t-1)^2,(t-1)^3\}} \\ &= \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \end{aligned}$$

The inverse of this matrix is

$$\begin{aligned} [I]_{\alpha\beta} &= [1, t-1, (t-1)^2, (t-1)^3]_{\{1,t,t^2,t^3\}} \\ &= [1, -1+t, 1-2t+t^2, -1+3t-3t^2+t^3]_{\{1,t,t^2,t^3\}} \\ &= \begin{pmatrix} 1 & -1 & 1 & -1 \\ 0 & 1 & -2 & 3 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \end{aligned}$$

The equality

$$(t+1)^3 = 1+3t+3t^2+t^3 = ((t-1)+2)^3 = 8+12(t-1)+6(t-1)^2+(t-1)^3$$

gives the coordinates

$$[(t+1)^3]_\alpha = (1,3,3,1), \quad [(t+1)^3]_\beta = (8,12,6,1).$$

The two coordinates are related by the matrices for the change of basis

$$\begin{pmatrix}8\\12\\6\\1\end{pmatrix}=\begin{pmatrix}1&1&1&1\\0&1&2&3\\0&0&1&3\\0&0&0&1\end{pmatrix}\begin{pmatrix}1\\3\\3\\1\end{pmatrix},\qquad \begin{pmatrix}1\\3\\3\\1\end{pmatrix}=\begin{pmatrix}1&-1&1&-1\\0&1&-2&3\\0&0&1&-3\\0&0&0&1\end{pmatrix}\begin{pmatrix}8\\12\\6\\1\end{pmatrix}.$$

Exercise 4.52. Use matrices for the change of basis in Example 4.3.24 to find the matrix $[L]_{\{1,t,t^2,t^3\}\{1,t-1,(t-1)^2,(t-1)^3\}}$ of the linear transformation L in Example 4.3.20.

Exercise 4.53. Suppose $\alpha=\{\vec{v}_1,\vec{v}_2,\vec{v}_3\}$ is changed to α' as follows (see Proposition 2.4.6). How is the matrix of linear transformation changed?

1. $\alpha'=\{\vec{v}_2,\vec{v}_1,\vec{v}_3\}$.
2. $\alpha'=\{c\vec{v}_1,\vec{v}_2,\vec{v}_3\}$, $c\neq 0$.
3. $\alpha'=\{\vec{v}_1+c\vec{v}_2,\vec{v}_2,\vec{v}_3\}$.

You may describe the change by multiplying a matrix for changing basis, or as some row or column operation on the matrix.

For a linear operator $L\colon V\to V$, we usually choose the same basis α for V on both sides. The matrix of the linear operator with respect to the basis α is $[L]_{\alpha\alpha}$. The matrices with respect to different bases are related by

$$[L]_{\beta\beta}=[I]_{\beta\alpha}[L]_{\alpha\alpha}[I]_{\alpha\beta}=[I]_{\alpha\beta}^{-1}[L]_{\alpha\alpha}[I]_{\alpha\beta}=[I]_{\beta\alpha}[L]_{\alpha\alpha}[I]_{\beta\alpha}^{-1}.$$

We say the two matrices $A=[L]_{\alpha\alpha}$ and $B=[L]_{\beta\beta}$ are *similar* in the sense that they are related by

$$B=PAP^{-1}=Q^{-1}AQ,$$

where P (matrix for changing from α to β) is an invertible matrix, and $Q=P^{-1}$ (matrix for changing from β to α).

Exercise 4.54. Explain that if A is similar to B, then B is similar to A, and A^{-1} is similar to B^{-1}, and A^T is similar to B^T.

Exercise 4.55. Explain that if A is similar to B, and B is similar to C, then A is similar to C.

Example 4.3.25. The trace of a square matrix is defined in Example 4.3.5. For any linear operator $L\colon V\to V$, we take any basis α of V and define the trace of the linear operator

$$\mathrm{tr}L=\mathrm{tr}[L]_{\alpha\alpha}.$$

If β is another basis of V, then by the property $\mathrm{tr}AB = \mathrm{tr}BA$ established in Example 4.3.5, we have

$$\mathrm{tr}[L]_{\beta\beta} = \mathrm{tr}(P[L]_{\alpha\alpha}P^{-1}) = \mathrm{tr}([L]_{\alpha\alpha}P^{-1}P) = \mathrm{tr}[L]_{\alpha\alpha}.$$

Therefore the trace is well defined. It is a linear transformation from $M(n,n)$ to $\mathbb{R}$.

Example 4.3.26. Let P be the orthogonal projection of $\mathbb{R}^3$ to the plane $x+y+z=0$ in Example 3.1.12. In Example 3.4.6, we showed that the matrix of P with respect to $\alpha = \{(1,-1,0),(1,0,-1),(1,1,1)\}$ is very simple

$$[P]_{\alpha\alpha} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

For the standard basis ϵ, by Example 4.3.23, we get

$$[I]_{\epsilon\alpha} = (\alpha) = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix}.$$

Then by Example 3.4.6, we get

$$[I]_{\epsilon\alpha}^{-1} = \frac{1}{3}\begin{pmatrix} 1 & -2 & 1 \\ 1 & 1 & -2 \\ 1 & 1 & 1 \end{pmatrix}.$$

Then we get the matrix of P with respect the the standard basis

$$\begin{aligned} [P]_{\epsilon\epsilon} &= [I]_{\epsilon\alpha}[P]_{\alpha\alpha}[I]_{\epsilon\alpha}^{-1} \\ &= \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix}^{-1} = \frac{1}{3}\begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}. \end{aligned}$$

The matrix is the same as the one obtained in Examples 3.1.12 and 3.4.6 by different methods.

By the way, we have $\mathrm{tr}[P]_{\alpha\alpha} = 1+1+0 = 2$, and $\mathrm{tr}[P]_{\epsilon\epsilon} = \frac{1}{3}(2+2+2) = 2$. We showed in Example 4.3.25 that the two traces must be equal.

Example 4.3.27. Consider the linear operator $L(f(t)) = (tf(t))' = tf'(t)+f(t)\colon P_3 \to P_3$. Applying the operator to the basis $\alpha = \{1, t, t^2, t^3\}$, we get

$$L(1) = 1, \quad L(t) = 2t, \quad L(t^2) = 3t^2, \quad L(t^3) = 4t^3.$$

Therefore

$$[L]_{\alpha\alpha} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}.$$

Consider another basis $\beta = \{1, t-1, (t-1)^2, (t-1)^3\}$ of P_2. By Example 4.3.24, we have

$$[I]_{\beta\alpha} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad [I]_{\alpha\beta} = \begin{pmatrix} 1 & -1 & 1 & -1 \\ 0 & 1 & -2 & 3 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Therefore

$$\begin{aligned} [L]_{\beta\beta} &= [I]_{\beta\alpha}[L]_{\alpha\alpha}[I]_{\alpha\beta} \\ &= \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 1 & -1 & 1 & -1 \\ 0 & 1 & -2 & 3 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 2 & 2 & 0 \\ 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 4 \end{pmatrix}. \end{aligned}$$

We can verify the result by directly applying $L(f) = (tf(t))'$ to vectors in β

$$\begin{aligned} L(1) &= 1, \\ L(t-1) &= [(t-1) + (t-1)^2]' = 1 + 2(t-1), \\ L((t-1)^2) &= [(t-1)^2 + (t-1)^3]' = 2(t-1) + 3(t-1)^2, \\ L((t-1)^3) &= [(t-1)^3 + (t-1)^4]' = 3(t-1)^2 + 4(t-1)^3. \end{aligned}$$

Exercise 4.56. Suppose a linear operator L on $\mathbb{R}^n$ has matrix A. Find the matrix of L with respect to the given basis α.

1. $A = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$, $\alpha = \{(1,2), (2,3)\}$.

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

3. $A = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 0 & 2 \\ 1 & 2 & 0 \end{pmatrix}$, $\alpha = \{(1,-1,0), (0,1,-1), (0,0,1)\}$.

4. $A = \begin{pmatrix} 0 & -1 & 1 \\ 2 & 0 & 0 \\ -1 & 2 & 0 \end{pmatrix}$, $\alpha = \{(2,2,-1), (-1,-2,1), (-1,-1,1)\}$.

Exercise 4.57. Find the matrix of the linear operator.

1. In $\mathbb{R}^2$, sending $\vec{v}_1 = (1,2)$ and $\vec{v}_2 = (3,4)$ to $2\vec{v}_1$ and $3\vec{v}_2$.

2. In $\mathbb{R}^2$, sending $\vec{v}_1 = (1,2)$ and $\vec{v}_2 = (3,4)$ to $\vec{v}_2$ and $\vec{v}_1$.

3. Reflection of $\mathbb{R}^3$ with respect to the plane $x+y+z=0$: preserving the plane, and flip the orthogonal vector $(1,1,1)$ to the opposite $(-1,-1,-1)$.

4. In $\mathbb{R}^3$, circularly sending the vectors $(1,-1,0) \mapsto (1,0,-1) \mapsto (1,1,1) \mapsto (1,-1,0)$.

Chapter 5

Orthogonality

Orthogonality provides the most linearly independent vectors. It also gives the best approximations to given data, via the least square method. The basic tool for the calculation is the Gram-Schmidt process that produces an orthogonal basis from any basis.

5.1 Dot Product

5.1.1 Algebra

The *dot product* of two vectors in $\mathbb{R}^n$ 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 dot product is a map $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ with numerical output. If we regard vectors as $n \times 1$ matrices, then

$$\vec{x}\cdot\vec{y} = (x_1\ x_2\ \cdots\ x_n)\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \vec{x}^T\vec{y}.$$

The right side is the multiplication of the $1 \times n$ matrix $\vec{x}^T$ and the $n \times 1$ matrix $\vec{y}$.

The following is the matrix multiplication example in Section 3.2.2

$$\begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}\begin{pmatrix} 5 & 7 \\ 6 & 8 \end{pmatrix} = \begin{pmatrix} 1\cdot 5 + 3\cdot 6 & 1\cdot 7 + 3\cdot 8 \\ 2\cdot 5 + 4\cdot 6 & 2\cdot 7 + 4\cdot 8 \end{pmatrix} = \begin{pmatrix} (1,3)\cdot(5,6) & (1,3)\cdot(7,8) \\ (2,4)\cdot(5,6) & (2,4)\cdot(7,8) \end{pmatrix}.$$

We see that the multiplication AB of matrices A and B consists of the dot products of the rows of A and column of B. The following is the general formula of this observation

$$(\vec{v}_1\ \vec{v}_2\ \cdots\ \vec{v}_m)^T(\vec{w}_1\ \vec{w}_2\ \cdots\ \vec{w}_n) = \begin{pmatrix} \vec{v}_1^T \\ \vec{v}_2^T \\ \vdots \\ \vec{v}_m^T \end{pmatrix} (\vec{w}_1\ \vec{w}_2\ \cdots\ \vec{w}_n)$$
$$= \begin{pmatrix} \vec{v}_1 \cdot \vec{w}_1 & \vec{v}_1 \cdot \vec{w}_2 & \cdots & \vec{v}_1 \cdot \vec{w}_n \\ \vec{v}_2 \cdot \vec{w}_1 & \vec{v}_2 \cdot \vec{w}_2 & \cdots & \vec{v}_2 \cdot \vec{w}_n \\ \vdots & \vdots & & \vdots \\ \vec{v}_m \cdot \vec{w}_1 & \vec{v}_m \cdot \vec{w}_2 & \cdots & \vec{v}_m \cdot \vec{w}_n \end{pmatrix}.$$

The dot product has the following properties:

- Bilinearity: $(a\vec{x} + b\vec{y}) \cdot \vec{z} = a\vec{x} \cdot \vec{z} + b\vec{y} \cdot \vec{z}$, $\vec{x} \cdot (a\vec{y} + b\vec{z}) = a\vec{x} \cdot \vec{y} + b\vec{x} \cdot \vec{z}$.
- Symmetry: $\vec{x} \cdot \vec{y} = \vec{y} \cdot \vec{x}$.
- Positivity: $\vec{x} \cdot \vec{x} \geq 0$, and $\vec{x} \cdot \vec{x} = 0$ if and only if $\vec{x} = \vec{0}$.

Exercise 5.1. Verify the three properties of the dot product in $\mathbb{R}^3$.

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

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

The left side is the dot product of the coefficient vector $\vec{a}$ and the variable vector $\vec{x}$. In Example 3.1.5, we see that the left side $l(\vec{x}) = \vec{a} \cdot \vec{x} = \vec{a}^T\vec{x}$ is actually a linear transformation $\mathbb{R}^n \to \mathbb{R}$, called linear functional.

If $\vec{x} = \vec{y}$, then $l(\vec{x}) = l(\vec{y})$. Conversely, if $l(\vec{x}) = l(\vec{y})$ for all linear functional l, then the following says that $\vec{x} = \vec{y}$.

Proposition 5.1.1. *$\vec{x} = \vec{y}$ if and only if $\vec{a} \cdot \vec{x} = \vec{a} \cdot \vec{y}$ for all $\vec{a}$.*

The equality $\vec{x} = \vec{y}$ clearly implies $\vec{a} \cdot \vec{x} = \vec{a} \cdot \vec{y}$ for all $\vec{a}$. Conversely, by $\vec{a} \cdot \vec{x} = \vec{a} \cdot \vec{y}$ and the bilinear property, we get

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

Since $\vec{a}$ is arbitrary, we may take $\vec{a} = \vec{x} - \vec{y}$ and get

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

Then by the positive property, we get $\vec{x} - \vec{y} = \vec{0}$.

5.1.2 Bilinear Function

A function $f(\vec{x},\vec{y})\colon \mathbb{R}^m\times\mathbb{R}^n\to\mathbb{R}$ is *bilinear*, if it satisfies

$$f(a\vec{x}+b\vec{y},\vec{z})=af(\vec{x},\vec{z})+bf(\vec{y},\vec{z}),\quad f(\vec{x},a\vec{y}+b\vec{z})=af(\vec{x},\vec{y})+bf(\vec{x},\vec{z}).$$

The dot product is bilinear.

Consider a bilinear function $f(\vec{x},\vec{y})\colon \mathbb{R}^2\times\mathbb{R}^3\to\mathbb{R}$. Let $\alpha=\{\vec{v}_1,\vec{v}_2\}$ and $\beta=\{\vec{w}_1,\vec{w}_2,\vec{w}_3\}$ be bases (standard bases, for example) of $\mathbb{R}^2$ and $\mathbb{R}^3$. Then

$$\vec{x}=x_1\vec{v}_1+x_2\vec{v}_2,\quad [\vec{x}]_\alpha=\begin{pmatrix}x_1\\x_2\end{pmatrix};\quad \vec{y}=y_1\vec{w}_1+y_2\vec{w}_2+y_3\vec{w}_3,\quad [\vec{y}]_\beta=\begin{pmatrix}y_1\\y_2\\y_3\end{pmatrix}.$$

Let $a_{ij}=f(\vec{v}_i,\vec{w}_j)$. Then the bilinear function is

$$\begin{aligned}
f(\vec{x},\vec{y}) &= f(x_1\vec{v}_1+x_2\vec{v}_2, y_1\vec{w}_1+y_2\vec{w}_2+y_3\vec{w}_3)\\
&= x_1f(\vec{v}_1,y_1\vec{w}_1+y_2\vec{w}_2+y_3\vec{w}_3)+x_2f(\vec{v}_2,y_1\vec{w}_1+y_2\vec{w}_2+y_3\vec{w}_3)\\
&= x_1y_1f(\vec{v}_1,\vec{w}_1)+x_1y_2f(\vec{v}_1,\vec{w}_2)+x_1y_3f(\vec{v}_1,\vec{w}_3)\\
&+ x_2y_1f(\vec{v}_2,\vec{w}_1)+x_2y_2f(\vec{v}_2,\vec{w}_2)+x_2y_3f(\vec{v}_2,\vec{w}_3)\\
&= a_{11}x_1y_1+a_{12}x_1y_2+a_{13}x_1y_3+a_{21}x_2y_1+a_{22}x_2y_2+a_{23}x_2y_3\\
&= x_1(a_{11}y_1+a_{12}y_2+a_{13}y_3)+x_2(a_{21}y_1+a_{22}y_2+a_{23}y_3)\\
&= \begin{pmatrix}x_1\\x_2\end{pmatrix}\cdot\begin{pmatrix}a_{11}y_1+a_{12}y_2+a_{13}y_3\\a_{21}y_1+a_{22}y_2+a_{23}y_3\end{pmatrix}=\begin{pmatrix}x_1\\x_2\end{pmatrix}\cdot\begin{pmatrix}a_{11}&a_{12}&a_{13}\\a_{21}&a_{22}&a_{23}\end{pmatrix}\begin{pmatrix}y_1\\y_2\\y_3\end{pmatrix}\\
&= [\vec{x}]_\alpha\cdot A[\vec{y}]_\beta.
\end{aligned}$$

Here

$$A=\begin{pmatrix}f(\vec{v}_1,\vec{w}_1)&f(\vec{v}_1,\vec{w}_2)&f(\vec{v}_1,\vec{w}_3)\\f(\vec{v}_2,\vec{w}_1)&f(\vec{v}_2,\vec{w}_2)&f(\vec{v}_2,\vec{w}_3)\end{pmatrix}=[f]_{\alpha\beta}$$

is the matrix of f with respect to the bases α and β.

A consequence of the calculation is the following result (in fact, we only need α and β to span $\mathbb{R}^m$ and $\mathbb{R}^n$).

Proposition 5.1.2. *Suppose $f(\vec{x},\vec{y})$ and $g(\vec{x},\vec{y})$ are bilinear functions on $\mathbb{R}^m\times\mathbb{R}^n$, and $\alpha=\{\vec{v}_1,\vec{v}_2,\dots,\vec{v}_m\}$ is a basis of $\mathbb{R}^m$, and $\beta=\{\vec{w}_1,\vec{w}_2,\dots,\vec{w}_n\}$ is a basis of $\mathbb{R}^n$. If*

$$f(\vec{v}_i,\vec{w}_j)=g(\vec{v}_i,\vec{w}_j)\quad \text{for all } \vec{v}_i\in\alpha,\vec{w}_j\in\beta,$$

then $f(\vec{x},\vec{y})=g(\vec{x},\vec{y})$.

If α and β are standard bases of $\mathbb{R}^m$ and $\mathbb{R}^n$, then $[\vec{x}]_\alpha=\vec{x}$, $[\vec{y}]_\beta=\vec{y}$, and the bilinear function is

$$f(\vec{x},\vec{y})=\vec{x}\cdot A\vec{y},\quad A=(a_{ij})=(f(\vec{e}_i,\vec{e}_j)).$$

We remark that $\vec{e}_i\in\mathbb{R}^m$ and $\vec{e}_j\in\mathbb{R}^n$.

Example 5.1.1. Using the standard bases of Euclidean spaces, the matrix

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

gives the bilinear function on $\mathbb{R}^3 \times \mathbb{R}^3$

$$\begin{aligned} &\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \\ &= x_1y_1 + 2x_1y_2 + 3x_1y_3 + 4x_2y_1 + 5x_2y_2 + 6x_2y_3 + 7x_3y_1 + 8x_3y_2 + 9x_3y_3. \end{aligned}$$

In the reverse direction, we have

$$2x_1y_2 + 3x_2y_1 - 4x_2y_2 + 5x_2y_3 - 6x_3y_2 = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \cdot \begin{pmatrix} 0 & 2 & 0 \\ 3 & -4 & 5 \\ 0 & -6 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}.$$

Exercise 5.2. For the matries, write down the corresponding bilinear functions with respect the standard bases of the Euclidean spaces.

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

Exercise 5.3. Find the matrices of the bilinear functions with respect to the standard bases of the Euclidean spaces.

1. $5x_1y_1 + 3x_2y_2 - 4x_3y_3$.
2. $5x_1y_1 + 3x_2y_2 - 4x_3y_3 + 3x_1y_2 + 4x_1y_3 + x_2y_1 - 4x_2y_3 - 3x_3y_1 + 8x_3y_2$.
3. $5x_1y_1 - 4x_3y_3 + 3x_1y_2 + 4x_1y_3 + x_2y_1 - 4x_2y_3$.

Exercise 5.4. For any $m \times n$ matrix A, show that $f(\vec{x}, \vec{y}) = \vec{x} \cdot A\vec{y}$ is a bilinear function on $\mathbb{R}^m \times \mathbb{R}^n$.

Exercise 5.5. Use Proposition 5.1.1 to prove that $A = B$ if and only if $\vec{x} \cdot A\vec{y} = \vec{x} \cdot B\vec{y}$ for all $\vec{x}$ and $\vec{y}$. Equivalently, two linear transformations $L, K : \mathbb{R}^n \to \mathbb{R}^m$ are equal if and only if $\vec{x} \cdot L(\vec{y}) = \vec{x} \cdot K(\vec{y})$ for all $\vec{x}$ and $\vec{y}$.

Exercise 5.6. A bilinear function on $\mathbb{R}^n \times \mathbb{R}^n$ is *symmetric*, if $f(\vec{x}, \vec{y}) = f(\vec{y}, \vec{x})$. For example, the dot product is a symmetric bilinear function. What kind of matrices correspond to symmetric bilinear functions.

Exercise 5.7. A bilinear function on $\mathbb{R}^n \times \mathbb{R}^n$ is *skew-symmetric*, if $f(\vec{x},\vec{y}) = -f(\vec{y},\vec{x})$. What kind of matrices correspond to skew-symmetric bilinear functions.

For
$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix},$$
and $\vec{x} \in \mathbb{R}^2$, $\vec{y} \in \mathbb{R}^3$, we have

$$\begin{aligned}
\vec{x} \cdot A\vec{y} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} &= \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \cdot \begin{pmatrix} 1y_1 + 2y_2 + 3y_3 \\ 4y_1 + 5y_2 + 6y_3 \end{pmatrix} \\
&= x_1(1y_1 + 2y_2 + 3y_3) + x_2(4y_1 + 5y_2 + 6y_3) \\
&= (1x_1 + 4x_2)y_1 + (2x_1 + 5x_2)y_2 + (3x_1 + 6x_2)y_3 \\
&= \begin{pmatrix} 1x_1 + 4x_2 \\ 2x_1 + 5x_2 \\ 3x_1 + 6x_2 \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = A^T\vec{x} \cdot \vec{y}.
\end{aligned}$$

In general, we have
$$\vec{x} \cdot A\vec{y} = A^T\vec{x} \cdot \vec{y}.$$
By the symmetric property of the dot product, we also get
$$A\vec{x} \cdot \vec{y} = \vec{x} \cdot A^T\vec{y}.$$
We see that moving a matrix in an inner product from one vector to another vector requires the transpose of the matrix.

Proposition 5.1.3. $\text{Nul}A^TA = \text{Nul}A$.

The following shows $\text{Nul}A \subset \text{Nul}A^TA$
$$\vec{x} \in \text{Nul}A \implies A\vec{x} = \vec{0} \implies A^TA\vec{x} = \vec{0} \implies \vec{x} \in \text{Nul}A^TA.$$
The following shows $\text{Nul}A^TA \subset \text{Nul}A$
$$\begin{aligned}
\vec{x} \in \text{Nul}A^TA &\implies A^TA\vec{x} = \vec{0} \implies \|A\vec{x}\|^2 = A\vec{x} \cdot A\vec{x} = \vec{x} \cdot A^TA\vec{x} = 0 \\
&\implies A\vec{x} = \vec{0} \implies \vec{x} \in \text{Nul}A.
\end{aligned}$$

If $\text{Nul}A = \{\vec{0}\}$, i.e., the columns of A are linearly independent, then we get $\text{Nul}A^TA = \{\vec{0}\}$, i.e., the columns of A^TA are also linearly independent. Since A^TA is a square matrix, we actually know A^TA is invertible.

If we consider the linear transformation associated to matrix, then we get the following definition.

Definition 5.1.4. The *adjoint* of a linear transformation $L\colon \mathbb{R}^n \to \mathbb{R}^m$ is the linear transformation $L^*\colon \mathbb{R}^m \to \mathbb{R}^n$ satisfying

$$\vec{x} \cdot L(\vec{y}) = L^*(\vec{x}) \cdot \vec{y} \quad \text{for all } \vec{x} \in \mathbb{R}^m,\ \vec{y} \in \mathbb{R}^n.$$

Since the dot product is symmetric, the adjoint is also characterised by

$$L(\vec{x}) \cdot \vec{y} = \vec{x} \cdot L^*(\vec{y}) \quad \text{for all } \vec{x} \in \mathbb{R}^n,\ \vec{y} \in \mathbb{R}^m.$$

In Section 3.2, we defined the addition and multiplication of matrices as the operations corresponding to the addition and composition of linear transformations. Along the same spirit, we may define the transpose as the matrix of the adjoint linear transformation

$$[L]^T = [L^*].$$

The following calculates the adjoint of the composition

$$\begin{aligned}
&\vec{x} \cdot (L \circ K)(\vec{y}) = (L \circ K)^*(\vec{x}) \cdot \vec{y}, \\
&\vec{x} \cdot (L \circ K)(\vec{y}) = \vec{x} \cdot L(K(\vec{y})) = L^*(\vec{x}) \cdot K(\vec{y}) = K^*(L^*(\vec{x})) \cdot \vec{y} = (K^* \circ L^*)(\vec{x}) \cdot \vec{y}.
\end{aligned}$$

By Exercise 5.5 and comparing the right sides, we get

$$(L \circ K)^* = K^* \circ L^*.$$

Translated to the transpose of matrices, we get

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

Exercise 5.8. Show that $(aL + bK)^* = aL^* + bK^*$. This means $(aA + bB)^T = aA^T + bB^T$.

Exercise 5.9. Show that $(L^*)^* = L$. This means $(A^T)^T = A$.

5.1.3 Geometry

The dot product induces various sizes in the Euclidean space. 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 angle between the plane vectors $(1, 0)$ and $(\cos\theta, \sin\theta)$ is θ

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

On the other hand, by $|\cos\theta| \le 1$, the definition of angle must be justified by the following result.

Theorem 5.1.5 (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:

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

We will give a general argument in Section 5.5.1, after Theorem 5.5.2.

Exercise 5.10. Verify the Cauchy-Schwartz inequality for $\mathbb{R}^3$.

Exercise 5.11 (Triangle Inequality). Use the Cauchy-Schwartz inequality to prove $\|\vec{x}+\vec{y}\| \le \|\vec{x}\| + \|\vec{y}\|$.

The parallelogram spanned by $\vec{x}$ and $\vec{y}$ has base $\|\vec{x}\|$ and height $\|\vec{y}\|\sin\theta$. Therefore the area of the parallelogram is

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

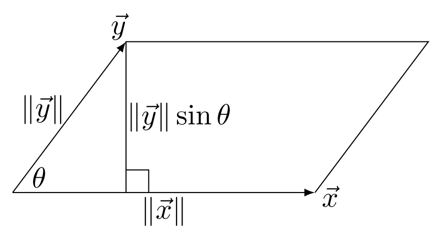

Figure 5.1.1: Angle and area of parallelogram.

Example 5.1.2. Let $\vec{x} = (1, 2)$ and $\vec{y} = (3, 4)$. Then

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

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 5.1.3. 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

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

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

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

Exercise 5.12. Calculate the area of the triangle in Example 5.1.3 in another way, for example, by using $\vec{x} - \vec{y}$ and $\vec{z} - \vec{y}$.

Exercise 5.13. Find the angle between $(1,1,1,1)$ and $(1,1,1,-1)$. Then find the area of the parallelogram spanned by the two vectors.

Exercise 5.14. Find the areas of the triangles with the three given points as vertices.

1. $(1,2),(3,4),(5,6)$.
2. $(1,2),(3,3),(5,6)$.
3. $(1,2,3),(2,3,4),(1,1,1)$.
4. $(1,2,3),(2,2,2),(1,1,1)$.
5. $(1,1,1,1),(1,1,-1,-1),(1,-1,1,-1)$.
6. $(1,1,1,1),(1,1,-1,-1),(-1,-1,1,1)$.

In case the area is zero, can you give geometrical interpretation?

Example 5.1.4 (Polarisation Identity)**.** The length is defined by the dot product. Conversely, the dot product can also be expressed in terms of the length. We have

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

Here the first equality is by the bilinear property, and the third equality is by the symmetric property. Then we have

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

This is called the *polarisation identity*.

Exercise 5.15. Prove another polarisation identity $\vec{x}\cdot\vec{y} = -\frac{1}{2}(\|\vec{x}-\vec{y}\|^2 - \|\vec{x}\|^2 - \|\vec{y}\|^2)$.

Exercise 5.16. Prove yet another polarisation identity $\vec{x}\cdot\vec{y} = \frac{1}{4}(\|\vec{x}+\vec{y}\|^2 - \|\vec{x}-\vec{y}\|^2)$.

5.2 Orthogonal Basis

5.2.1 Orthogonal Vectors

Two directions are orthogonal (or perpendicular) if the angle between them is $90°$. By the formula $\cos\theta = \frac{\vec{x}\cdot\vec{y}}{\|\vec{x}\|\|\vec{y}\|}$ for the angle θ between two vectors $\vec{x}$ and $\vec{y}$, we get the following definition.

Definition 5.2.1. Two vectors $\vec{x}$ and $\vec{y}$ are *orthogonal*, and denoted $\vec{x}\perp\vec{y}$, if $\vec{x}\cdot\vec{y}=0$. A vector $\vec{x}$ is orthogonal to a subspace H, and denoted $\vec{x}\perp H$, if $\vec{x}\perp\vec{h}$ for all $\vec{h}\in H$.

The most famous theorem about orthogonal vectors is the following. The proof follows from the calculation in Example 5.1.4.

Theorem 5.2.2 (Pythagorean Theorem)**.** *$\vec{x}\perp\vec{y}$ if and only if $\|\vec{x}+\vec{y}\|^2 = \|\vec{x}\|^2+\|\vec{y}\|^2$.*

Exercise 5.17. Prove properties of orthogonal vectors.

1. $\vec{x}\perp\vec{x}$ if and only if $\vec{x}=\vec{0}$.
2. $\vec{x}\perp\vec{y}$ implies $\vec{y}\perp\vec{x}$.
3. $\vec{x}\perp\vec{y}$ and $\vec{x}\perp\vec{z}$ implies $\vec{x}\perp(a\vec{y}+b\vec{z})$.
4. If $\vec{x},\vec{y},\vec{z}$ are pairwise orthogonal, then $\|\vec{x}+\vec{y}+\vec{z}\|^2 = \|\vec{x}\|^2+\|\vec{y}\|^2+\|\vec{z}\|^2$.

Example 5.2.1. In Example 5.1.3, we find the vectors $(1,2,3)$ and $(1,-2,1)$ are orthogonal. To find the third vector orthogonal to both, we solve the system of linear equations

$$\begin{aligned}(1,2,3)\cdot\vec{x} &= x_1+2x_2+3x_3=0,\\ (1,-2,1)\cdot\vec{x} &= x_1-2x_2+x_3=0.\end{aligned}$$

The general solution to the system is

$$x_1=-2x_3,\quad x_2=-\tfrac{1}{2}x_3,\quad x_3 \text{ arbitrary}.$$

We may choose $x_3=-2$ and get pairwise orthogonal vectors $(1,2,3)$, $(1,-2,1)$, $(4,1,-2)$.

Exercise 5.18. Find a, such that the vectors are orthogonal.

1. $(1,2,3),(4,5,a)$.
2. $(1,1,a),(1,a,1)$.
3. $(1,2,-1,-2),(2,-2,a,-a)$.
4. $(1,3,a,5,-2),(2,1,2,-1,a)$.

Exercise 5.19. Find the condition for the vectors to be pairwise orthogonal.

1. $(1,2,3)$, $(1,a,1)$.
2. $(a,1,1)$, $(1,a,1)$.
3. $(1,1,1,1)$, $(1,1,-1,a)$, $(1,b,1,-1)$.
4. $(a,1,1,1)$, $(1,a,1,1)$, $(1,1,b,1)$.

Then extend to bigger pairwise orthogonal set by adding one more vector.

In Exercise 5.17, we know that, if $\vec{x}$ is orthogonal to some vectors, then it is orthogonal to the linear combinations of the vectors. In particular, if

$$H=\mathbb{R}\vec{v}_1+\mathbb{R}\vec{v}_2+\cdots+\mathbb{R}\vec{v}_n,$$

then

$$\vec{x}\perp H \iff \vec{x}\perp\vec{v}_1,\ \vec{x}\perp\vec{v}_2,\ \dots,\ \vec{x}\perp\vec{v}_n.$$

Example 5.2.2. Suppose $\vec{x}_0$ is one solution of the linear equation $l(\vec{x})=\vec{a}\cdot\vec{x}=b$. Then we have $l(\vec{x}_0)=\vec{a}\cdot\vec{x}_0=b$. Now for any other solution $\vec{x}$ of the same linear equation, the difference $\vec{v}=\vec{x}-\vec{x}_0$ satisfies

$$\vec{a}\cdot\vec{v}=l(\vec{v})=l(\vec{x}-\vec{x}_0)=l(\vec{x})-l(\vec{x}_0)=b-b=0.$$

Therefore the solutions of $l(\vec{x})=b$ are

$$\vec{x}=\vec{x}_0+\vec{v} \text{ for arbitrary } \vec{v}\perp\vec{a}.$$

This explains that a linear equation $l(\vec{x})=\vec{a}\cdot\vec{x}=b$ is a hyperplane H orthogonal to the coefficient vector $\vec{a}$.

In terms of the discussion in Section 2.3.2, the null space of the $1 \times n$ matrix $\vec{a}^T$ is all the vectors orthogonal to $\vec{a}$. We will express the statement as $\text{Nul}\vec{a}^T = (\mathbb{R}\vec{a})^{\perp}$.

Next, we find the distance from a point $\vec{v}$ to the hyperplane H. This means finding $\vec{h} \in H$, such that $\vec{v} - \vec{h}$ is orthogonal to H. Then the distance is the length $\|\vec{v} - \vec{h}\|$.

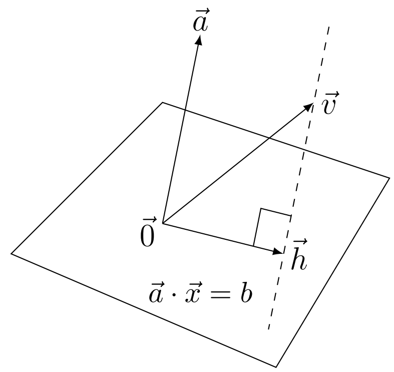

Figure 5.2.1: Distance from a point $\vec{v}$ to the hyperplane given by $\vec{a} \cdot \vec{x} = b$.

The direction orthogonal to H is the coefficient vector $\vec{a}$. Therefore $\vec{v} - \vec{h} = t\vec{a}$. Then $\vec{h} \in H$ means it satisfies the equation

$$b = \vec{a} \cdot \vec{h} = \vec{a} \cdot (\vec{v} - t\vec{a}) = \vec{a} \cdot \vec{v} - t\vec{a} \cdot \vec{a}.$$

Therefore

$$t = \frac{\vec{a} \cdot \vec{v} - b}{\vec{a} \cdot \vec{a}},$$

and the distance from $\vec{v}$ to the hyperplane $\vec{a} \cdot \vec{x} = b$ is

$$\|\vec{v} - \vec{h}\| = |t|\|\vec{a}\| = \frac{|\vec{a} \cdot \vec{v} - b|}{\vec{a} \cdot \vec{a}}\|\vec{a}\| = \frac{|\vec{a} \cdot \vec{v} - b|}{\|\vec{a}\|}.$$

For example, the distance from $(1, 1, 1, 1)$ to the hyperplane $x_1 + 2x_2 + 3x_3 + 4x_4 = 5$ is

$$\frac{|1 \cdot 1 + 2 \cdot 1 + 3 \cdot 1 + 4 \cdot 1 - 5|}{\sqrt{1^2 + 2^2 + 3^2 + 4^2}} = \sqrt{\frac{5}{6}}.$$

Exercise 5.20. Find the distance from the vector to the hyperplane.

1. (u, v) to $ax + by = c$.
2. $(1, 2, 3)$ to $x + y + z = 3$.
3. $(1, 1, 1, 1)$ to $x_1 + 2x_2 + 3x_3 + 4x_4 = 0$.
4. $(1, 2, 1, 2, 1)$ to $x_1 + 2x_3 + x_5 = 2$.

5.2.2 Orthogonal Basis

Definition 5.2.3. A set of vectors is an *orthogonal set*, if the vectors are nonzero and pairwise orthogonal. Moreover, if all vectors have length 1, then we say the set is *orthonormal*.

A vector set $\alpha = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ in $\mathbb{R}^n$ is orthogonal if

$$\vec{v}_1 \cdot \vec{v}_2 = \vec{v}_1 \cdot \vec{v}_3 = \vec{v}_2 \cdot \vec{v}_3 = 0.$$

The set is orthonormal if it further satisfies

$$\vec{v}_1 \cdot \vec{v}_1 = \vec{v}_2 \cdot \vec{v}_2 = \vec{v}_3 \cdot \vec{v}_3 = 1.$$

The standard basis $\epsilon = \{\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n\}$ of $\mathbb{R}^n$ is an orthonormal set.

Let

$$Q = (\vec{v}_1\ \vec{v}_2\ \vec{v}_3).$$

Then the orthogonal and orthonormal properties mean

$$\begin{aligned}
Q^TQ = \begin{pmatrix} \vec{v}_1^T \\ \vec{v}_2^T \\ \vec{v}_3^T \end{pmatrix} (\vec{v}_1\ \vec{v}_2\ \vec{v}_3) &= \begin{pmatrix} \vec{v}_1 \cdot \vec{v}_1 & \vec{v}_1 \cdot \vec{v}_2 & \vec{v}_1 \cdot \vec{v}_3 \\ \vec{v}_2 \cdot \vec{v}_1 & \vec{v}_2 \cdot \vec{v}_2 & \vec{v}_2 \cdot \vec{v}_3 \\ \vec{v}_3 \cdot \vec{v}_1 & \vec{v}_3 \cdot \vec{v}_2 & \vec{v}_3 \cdot \vec{v}_3 \end{pmatrix} \\
&= \begin{pmatrix} \vec{v}_1 \cdot \vec{v}_1 & 0 & 0 \\ 0 & \vec{v}_2 \cdot \vec{v}_2 & 0 \\ 0 & 0 & \vec{v}_3 \cdot \vec{v}_3 \end{pmatrix} \quad \text{(if α orthogonal)} \\
&= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = I. \quad \text{(if α orthonormal)}
\end{aligned}$$

Example 5.2.3. In Example 5.2.3, we get an orthogonal set consisting of $(1, 2, 3)$, $(1, -2, 1)$, $(4, 1, -2)$. We have

$$Q = \begin{pmatrix} 1 & 1 & 4 \\ 2 & -2 & 1 \\ 3 & 1 & -2 \end{pmatrix}, \quad Q^TQ = \begin{pmatrix} 1 & 2 & 3 \\ 1 & -2 & 1 \\ 4 & 1 & -2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 4 \\ 2 & -2 & 1 \\ 3 & 1 & -2 \end{pmatrix} = \begin{pmatrix} 14 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 21 \end{pmatrix}.$$

The orthogonality is not changed is we multiply scalars to vectors. Therefore we may change the length to 1 by dividing the length, and get an orthonormal set

$$\frac{(1,2,3)}{\|(1,2,3)\|} = \frac{(1,2,3)}{\sqrt{14}}, \quad \frac{(1,-2,1)}{\|(1,-2,1)\|} = \frac{(1,-2,1)}{\sqrt{6}}, \quad \frac{(4,1,-2)}{\|(4,1,-2)\|} = \frac{(4,1,-2)}{\sqrt{21}}.$$

Then the corresponding matrix

$$R = \begin{pmatrix} \frac{1}{\sqrt{14}} & \frac{1}{\sqrt{6}} & \frac{4}{\sqrt{21}} \\ \frac{2}{\sqrt{14}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{21}} \\ \frac{3}{\sqrt{14}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{21}} \end{pmatrix}$$

satisfies $R^TR = I$.

Proposition 5.2.4. *Suppose $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ and $\vec{x} \in \text{Span}\alpha$. If α is an orthogonal set, then*

$$\vec{x} = \frac{\vec{x} \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1 + \frac{\vec{x} \cdot \vec{v}_2}{\vec{v}_2 \cdot \vec{v}_2}\vec{v}_2 + \cdots + \frac{\vec{x} \cdot \vec{v}_n}{\vec{v}_n \cdot \vec{v}_n}\vec{v}_n.$$

In particular, if α is orthonormal, then

$$\vec{x} = (\vec{x} \cdot \vec{v}_1)\vec{v}_1 + (\vec{x} \cdot \vec{v}_2)\vec{v}_2 + \cdots + (\vec{x} \cdot \vec{v}_n)\vec{v}_n.$$

The assumption $\vec{x} \in \text{Span}\alpha$ means

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

Then

$$\begin{aligned} \vec{x} \cdot \vec{v}_1 &= x_1\vec{v}_1 \cdot \vec{v}_1 + x_2\vec{v}_2 \cdot \vec{v}_1 + \cdots + x_n\vec{v}_n \cdot \vec{v}_1 \\ &= x_1\vec{v}_1 \cdot \vec{v}_1 + x_2 0 + \cdots + x_n 0 = x_1\vec{v}_1 \cdot \vec{v}_1. \end{aligned}$$

This gives the formula for x_1. The formula for x_i is similar.

The explicit formula for the coefficients shows the uniqueness of the coefficients, which means linear independence.

Proposition 5.2.5. *Any orthogonal set is linearly independent.*

By Example 4.2.4, two vectors are linearly dependent if they are parallel, i.e., the angle between the two vectors is $0°$ or $180°$. Although the vectors become linearly independent if the angle becomes $1°$ or $179°$, we still feel the two vectors are almost dependent. As the angle become further and further away from $0°$ and $180°$, we feel the two vectors become more and more independent, and they are the most independent when the angle is $90°$. Therefore orthogonal intuitively means most independent.

Proposition 5.2.5 implies that, if an orthogonal set α spans H, then α is an *orthogonal basis* of H (see Exercise 2.30). If α is orthonormal, then we say α is an *orthonormal basis*.

Example 5.2.4. In Examples 5.2.1 and 5.2.3, we have an orthogonal set $\alpha = \{(1, 2, 3), (1, -2, 1), (4, 1, -2)\}$. By Proposition 5.2.5, the vectors are linearly independent. Then three vectors in $\mathbb{R}^3$ form a basis, and any vector in $\mathbb{R}^3$ is a linear combination of α

$$(x_1, x_2, x_3) = y_1(1, 2, 3) + y_2(1, -2, 1) + y_3(4, 1, -2).$$

Finding the coefficients y_1, y_2, y_3 usually means row operations on the augmented matrix

$$(Q\ \vec{x}) = \begin{pmatrix} 1 & 1 & 4 & x_1 \\ 2 & -2 & 1 & x_2 \\ 3 & 1 & -2 & x_3 \end{pmatrix}.$$

However, since the vectors are orthogonal, we have direct formula for the coefficients

$$\begin{aligned}
y_1 &= \frac{(x_1,x_2,x_3)\cdot(1,2,3)}{(1,2,3)\cdot(1,2,3)} = \frac{x_1+2x_2+3x_3}{14}, \\
y_2 &= \frac{(x_1,x_2,x_3)\cdot(1,-2,1)}{(1,-2,1)\cdot(1,-2,1)} = \frac{x_1-2x_2+x_3}{6}, \\
y_3 &= \frac{(x_1,x_2,x_3)\cdot(4,1,-2)}{(4,1,-2)\cdot(4,1,-2)} = \frac{4x_1+x_2-2x_3}{21}.
\end{aligned}$$

Exercise 5.21. Express (x, y) as a linear combination of $(\cos\theta, \sin\theta)$ and $(-\sin\theta, \cos\theta)$.

Exercise 5.22. The extensions of the orthogonal sets in Exercise 5.19 are orthogonal bases of the Euclidean spaces.

1. Improve the orthogonal bases to orthonormal bases.

2. Find the coordinates of any vector with respect to the orthogonal bases.

Example 5.2.5. The subspace $H \subset \mathbb{R}^3$ given by $x + y + z = 0$ contains a vector $\vec{v}_1 = (1, -1, 0)$. By $\dim H = 2$, to extend $\vec{v}_1$ to an orthogonal basis of H, we need to find $\vec{v}_2 = (x, y, z) \in H$ satisfying $\vec{v}_1 \cdot \vec{v}_2 = 0$. This means

$$x + y + z = 0, \quad x - y = 0.$$

The solution is $x = y$, $z = -2y$, y arbitrary. We choose $y = 1$ and get an orthogonal basis $\vec{v}_1 = (1, -1, 0)$, $\vec{v}_2 = (1, 1, -2)$ of H.

To get an orthogonal basis of $\mathbb{R}^3$, we need to add one more vector $\vec{v}_3$ orthogonal to $\vec{v}_1$ and $\vec{v}_2$. Since both $\vec{v}_1$ and $\vec{v}_2$ are in H, and we know the coefficient vector $\vec{v}_3 = (1, 1, 1)$ is orthogonal to H, we conclude that $\vec{v}_1 = (1, -1, 0)$, $\vec{v}_2 = (1, 1, -2)$, $\vec{v}_3 = (1, 1, 1)$ form an orthogonal basis α of $\mathbb{R}^3$. By dividing the length, we get the corresponding orthonormal basis

$$\frac{\vec{v}_1}{\|\vec{v}_1\|} = \frac{(1,-1,0)}{\sqrt{2}}, \quad \frac{\vec{v}_2}{\|\vec{v}_2\|} = \frac{(1,1,-2)}{\sqrt{6}}, \quad \frac{\vec{v}_3}{\|\vec{v}_3\|} = \frac{(1,1,1)}{\sqrt{3}}.$$

Usually we need to solve a system of linear equations to express a vector as a linear combination of basis. Proposition 5.2.4 provides a more straightforward way of calculating this

$$\begin{aligned}
(x,y,z) &= \frac{(x,y,z)\cdot(1,-1,0)}{(1,-1,0)\cdot(1,-1,0)}\vec{v}_1 + \frac{(x,y,z)\cdot(1,1,-2)}{(1,1,-2)\cdot(1,1,-2)}\vec{v}_2 + \frac{(x,y,z)\cdot(1,1,1)}{(1,1,1)\cdot(1,1,1)}\vec{v}_3 \\
&= \frac{x-y}{2}\vec{v}_1 + \frac{x+y-2z}{6}\vec{v}_2 + \frac{x+y+z}{3}\vec{v}_3.
\end{aligned}$$

Then the orthogonal projection P onto H is given by (see Example 3.1.12)

$$\begin{aligned}
P(\vec{x}) &= \frac{x-y}{2}P(\vec{v}_1) + \frac{x+y-2z}{6}P(\vec{v}_2) + \frac{x+y+z}{3}P(\vec{v}_3) \\
&= \frac{x-y}{2}\vec{v}_1 + \frac{x+y-2z}{6}\vec{v}_2 \\
&= \frac{x-y}{2}\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + \frac{x+y-2z}{6}\begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}.
\end{aligned}$$

5.2.3 Isometry

Definition 5.2.6. A linear transformation $L\colon \mathbb{R}^n \to \mathbb{R}^m$ is an *isometry* if it preserves the dot product:

$$L(\vec{x}) \cdot L(\vec{y}) = \vec{x} \cdot \vec{y} \quad \text{for all } \vec{x}, \vec{y} \in \mathbb{R}^n.$$

Both $L(\vec{x}) \cdot L(\vec{y})$ and $\vec{x} \cdot \vec{y}$ are bilinear functions. By Proposition 5.1.2, for any basis $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ of $\mathbb{R}^n$, L is an isometry if and only if it preserves the dot products of the basis vectors

$$L(\vec{v}_i) \cdot L(\vec{v}_j) = \vec{v}_i \cdot \vec{v}_j \text{ for all } i, j.$$

For the special case α is an orthonormal basis of $\mathbb{R}^n$, L is an isometry if and only if L takes an orthonormal basis to an orthonormal set.

An isometry preserves all the geometric quantities derived from the dot product. For example, L preserves the length

$$\|L(\vec{x})\| = \|\vec{x}\|.$$

Conversely, the polarisation identity shows that, if L preserves the length, then it preserves the dot product, and is an isometry.

An isometry also preserves the distance

$$\|L(\vec{x}) - L(\vec{y})\| = \|L(\vec{x} - \vec{y})\| = \|\vec{x} - \vec{y}\|.$$

This implies that an isometry is always one-to-one[1].

Using the adjoint in Section 5.1.2, the isometry means

$$\vec{x} \cdot \vec{y} = L(\vec{x}) \cdot L(\vec{y}) = \vec{x} \cdot L^*L(\vec{y}).$$

By Exercise 5.5 (essentially Proposition 5.1.1), this is the same as $L^*L = I$. For the matrix $Q = [L]$ of L, this means

$$Q^TQ = I.$$

[1]It is a general fact that, if a (not necessarily linear) map $F\colon \mathbb{R}^n \to \mathbb{R}^m$ preserves the Euclidean distance: $\|F(\vec{x}) - F(\vec{y})\| = \|\vec{x} - \vec{y}\|$, then $F(\vec{x}) = \vec{a} + L(\vec{x})$ for a fixed vector $\vec{a} = F(\vec{0})$ and an isometry L. The key message here is that L is linear.

In Section 5.2.2, we know this means that the columns of Q form an orthonormal set in $\mathbb{R}^m$. This has a geometric interpretation. The matrix Q is the application of L to the standard basis $\epsilon = \{\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n\}$ of $\mathbb{R}^n$

$$Q = (L(\vec{e}_1)\ L(\vec{e}_2)\ \cdots\ L(\vec{e}_n)).$$

Since ϵ is orthonormal, the set $L(\epsilon)$ of columns of Q is also orthonormal.

Example 5.2.6. Suppose $L\colon \mathbb{R}^3 \to \mathbb{R}^3$ is an isometry satisfying $L(1,-1,0) = (0,1,a)$. Then we should have

$$1 + a^2 = \|(0,1,a)\|^2 = \|(1,-1,0)\|^2 = 2.$$

This implies $a = \pm 1$. We choose $a = 1$ and get $L(1,-1,0) = (0,1,1)$.

Next we assume $L(3,0,-2) = (b,0,c)$. Then L preserves the dot products between $(3,0,-2)$ and $(3,0,-2), (1,-1,0)$

$$\begin{aligned} b^2 + c^2 &= \|(b,0,c)\|^2 = \|(3,0,-2)\|^2 = 13, \\ c &= (b,0,c)\cdot(0,1,1) = (3,0,-2)\cdot(1,-1,0) = 3. \end{aligned}$$

We get $b = \pm 2$ and $c = 3$. We choose $b = -2$ and get $L(3,0,2) = (-2,0,3)$.

To get the whole formula for L, we need to find the value of L on a vector independent of $(1,-1,0)$ and $(3,0,-2)$. Consider $L(1,0,0) = (d,e,f)$. Then L preserves the dot products between $(1,0,0)$ and $(1,0,0), (1,-1,0), (3,0,-2)$

$$\begin{aligned} d^2 + e^2 + f^2 &= \|(d,e,f)\|^2 = \|(1,0,0)\|^2 = 1, \\ e + f &= (d,e,f)\cdot(0,1,1) = (1,0,0)\cdot(1,-1,0) = 1, \\ -2d + 3f &= (d,e,f)\cdot(-2,0,3) = (1,0,0)\cdot(3,0,2) = 3. \end{aligned}$$

From the second and third equations, we get $e = -(f-1)$ and $d = \frac{3}{2}(f-1)$. Substituting into the first equation, we get

$$\tfrac{3^2}{2^2}(f-1)^2 + (f-1)^2 + f^2 = 1.$$

Using $f^2 - 1 = (f-1)(f+1)$, the equation is

$$(f-1)[\tfrac{3^2}{2^2}(f-1) + (f-1) + (f+1)] = \tfrac{1}{4}(f-1)(17f-9) = 0.$$

We have two possible solutions $(d,e,f) = (0,0,1)$ and $(d,e,f) = (-\frac{12}{17}, \frac{8}{17}, \frac{9}{17})$.

We may obtain the matrix of L by the method in Example 3.1.10, as the lower half of the following column operations

$$\begin{pmatrix} 1 & 3 & 1 \\ -1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & -2 & 0 \\ 1 & 0 & 0 \\ 1 & 3 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 3 & 1 \\ -1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & -2 & -\frac{12}{17} \\ 1 & 0 & \frac{8}{17} \\ 1 & 3 & \frac{9}{17} \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -\frac{12}{17} & -\frac{12}{17} & \frac{1}{17} \\ \frac{8}{17} & -\frac{9}{17} & -\frac{12}{17} \\ \frac{9}{17} & -\frac{8}{17} & \frac{12}{17} \end{pmatrix}.$$

Exercise 5.23. Find a, b, c, such that a linear transformation $L\colon \mathbb{R}^2 \to \mathbb{R}^3$ satisfying $L(1,2) = (a,0,2)$ and $L(3,4) = (1,b,c)$ is an isometry.

5.2.4 Orthogonal Matrix

Suppose $m = n$. Then by the one-to-one property of an isometry L, and Theorem 3.4.3, we know L is invertible. In this case, we call L an *orthogonal operator* on $\mathbb{R}^n$, and call the corresponding matrix Q an *orthogonal matrix*. We know the columns of Q is orthonormal. The invertibility of Q means the columns of Q is a basis of the Euclidean space.

Proposition 5.2.7. *An $n \times n$ matrix Q is an orthogonal matrix, if and only if its column vectors form an orthonormal basis of $\mathbb{R}^n$.*

Since an orthogonal matrix Q is invertible, the equality $Q^TQ = I$ implies $Q^{-1} = Q^T$, and $QQ^T = I$. In particular, we know the columns of Q is an orthonormal basis of $\mathbb{R}^n$ if and only if the rows of Q is an orthonormal basis of $\mathbb{R}^n$.

Example 5.2.7. The orthonormal basis in Example 5.2.5 gives an orthogonal matrix

$$Q = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{pmatrix}.$$

We have

$$Q^{-1} = Q^T = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix}.$$

Exercise 5.24. What are 1×1 orthogonal matrices? What are 2×2 orthogonal matrices?

Exercise 5.25. For the orthogonal matrices in Examples 5.2.3 and 5.2.6, find the inverse matrices.

Exercise 5.26. Find suitable parameters, such that the matrices are orthogonal matrices. The find the inverse matrices.

1. $\begin{pmatrix} \frac{3}{5} & a \\ b & c \end{pmatrix}$.
2. $\begin{pmatrix} \frac{1}{3} & \frac{2}{3} & d \\ a & b & e \\ \frac{2}{3} & c & f \end{pmatrix}$.
3. $\begin{pmatrix} a & b & e \\ \frac{1}{3} & c & \frac{2}{3} \\ \frac{2}{3} & d & f \end{pmatrix}$.

5.3 Orthogonal Projection

5.3.1 Orthogonal Projection

Definition 5.3.1. The *orthogonal projection* of $\vec{x} \in \mathbb{R}^m$ onto a subspace $H \subset \mathbb{R}^m$ is $\vec{y} = \text{proj}_H\vec{x} \in H$, such that $\vec{x} - \vec{y} \perp H$.

We decompose $\vec{x}$ into two directions, one in H and one orthogonal to H

$$\vec{x} = \vec{y} + \vec{z}, \quad \vec{y} \in H, \ \vec{z} \perp H.$$

Then $\text{proj}_H \vec{x} = \vec{y}$ is the component in H. The component $\vec{z}$ will be the component $\text{proj}_{H^\perp} \vec{x}$ in the orthogonal complement $H^\perp$.

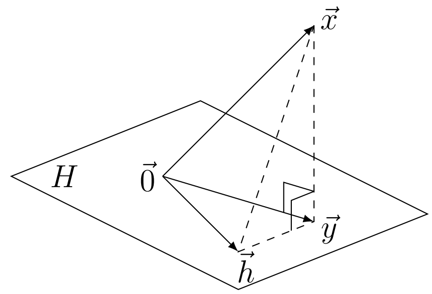

Figure 5.3.1: Orthogonal projection.

The following says the orthogonal projection of $\vec{x}$ onto H is the point on H that has the shortest distance from $\vec{x}$.

Proposition 5.3.2. *If $\vec{y} = \text{proj}_H \vec{x}$ and $\vec{h} \in H$, then $\|\vec{x} - \vec{h}\| \geq \|\vec{x} - \vec{y}\|$. Moreover, the equality happens if and only if $\vec{h} = \vec{y}$.*

Since $\vec{x} - \vec{y} \perp H$ and $\vec{y} - \vec{h} \in H$, we have $\vec{x} - \vec{y} \perp \vec{y} - \vec{h}$. Then by Theorem 5.2.2 (Pythagorean Theorem), we have

$$\|\vec{x} - \vec{h}\|^2 = \|(\vec{x} - \vec{y}) + (\vec{y} - \vec{h})\|^2 = \|\vec{x} - \vec{y}\|^2 + \|\vec{y} - \vec{h}\|^2 \geq \|\vec{x} - \vec{y}\|^2.$$

Moreover, the inequality $\geq$ becomes equality if and only if $\|\vec{y} - \vec{h}\| = 0$, which means $\vec{y} = \vec{h}$.

Proposition 5.3.3. *If $\alpha = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is an orthogonal basis of a subspace $H \subset \mathbb{R}^m$, then the orthogonal projection of $\vec{x}$ on H is*

$$\text{proj}_H \vec{x} = \frac{\vec{x} \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1} \vec{v}_1 + \frac{\vec{x} \cdot \vec{v}_2}{\vec{v}_2 \cdot \vec{v}_2} \vec{v}_2 + \cdots + \frac{\vec{x} \cdot \vec{v}_n}{\vec{v}_n \cdot \vec{v}_n} \vec{v}_n.$$

By the definition, it is clear that $\vec{x} = \text{proj}_H \vec{x}$ if and only if $\vec{x} \in H$. In this case, the proposition becomes Proposition 5.2.4.

Since $\vec{y} \in H$ is a linear combination of α, by Proposition 5.2.4, we get

$$\text{proj}_H \vec{x} = \vec{y} = \frac{\vec{y} \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1} \vec{v}_1 + \frac{\vec{y} \cdot \vec{v}_2}{\vec{v}_2 \cdot \vec{v}_2} \vec{v}_2 + \cdots + \frac{\vec{y} \cdot \vec{v}_n}{\vec{v}_n \cdot \vec{v}_n} \vec{v}_n.$$

By $\vec{x} - \vec{y} \perp H$, we get $(\vec{x} - \vec{y}) \cdot \vec{v}_i = 0$. Then $\vec{y} \cdot \vec{v}_i = \vec{x} \cdot \vec{v}_i$, and the formula above becomes the formula in Proposition 5.3.3.

Example 5.3.1. The orthogonal projection of $(1,2,3)$ onto the direction $\mathbb{R}(1,1,1)$ is

$$\text{proj}_{\mathbb{R}(1,1,1)}(1,2,3) = \frac{(1,2,3)\cdot(1,1,1)}{(1,1,1)\cdot(1,1,1)}(1,1,1) = (2,2,2).$$

In general, we have

$$\text{proj}_{\mathbb{R}(a_1,a_2,a_3)}(x_1,x_2,x_3) = \frac{a_1x_2 + a_2x_2 + a_3x_3}{a_1^2 + a_2^2 + a_3^2}(a_1,a_2,a_3).$$

Example 5.3.2. In Example 5.2.5, we get the orthogonal basis $\{(1,-1,0),(1,-12)\}$ of the subspace $H \subset \mathbb{R}^3$ given by $x+y+z=0$. Then the orthogonal projection onto H is given by the formula

$$\begin{aligned}
\text{proj}_H \begin{pmatrix} x \\ y \\ z \end{pmatrix} &= \frac{(x,y,z)\cdot(1,-1,0)}{(1,-1,0)\cdot(1,-1,0)} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + \frac{(x,y,z)\cdot(1,1,-2)}{(1,1,-2)\cdot(1,1,-2)} \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} \\
&= \frac{x-y}{2} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + \frac{x+y-2z}{6} \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} \\
&= \frac{1}{3} \begin{pmatrix} 2x-y-z \\ -x+2y-z \\ -x-y+2z \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.
\end{aligned}$$

Exercise 5.27. Find orthogonal projections.

1. $(1,2)$ onto $\mathbb{R}(3,4)$.
2. $(1,2,3,4)$ onto $\mathbb{R}(1,1,1,1)$.

Exercise 5.28. Verify that $(1,2,3)$ and $(1,-2,1)$ are orthogonal. Calculate the orthogonal projections onto $\mathbb{R}(1,2,3)$, $\mathbb{R}(1,-2,1)$, $\mathbb{R}(1,2,3)+\mathbb{R}(1,-2,1)$. How are the three projections related?

5.3.2 Orthogonalization

The *Gram-Schmidt process* converts any basis to an orthogonal basis.

Proposition 5.3.4. *Suppose $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ is a basis of a subspace $H \subset \mathbb{R}^m$. Then there is an orthogonal basis $\beta = \{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_n\}$ of H satisfying*

$$\mathbb{R}\vec{w}_1 + \mathbb{R}\vec{w}_2 + \cdots + \mathbb{R}\vec{w}_i = \mathbb{R}\vec{v}_1 + \mathbb{R}\vec{v}_2 + \cdots + \mathbb{R}\vec{v}_i, \quad i = 1,2,\dots,n.$$

The process is given by the formula

$$\vec{w}_{i+1} = \vec{v}_{i+1} - \text{proj}_{H_i}\vec{v}_{i+1}, \quad H_i = \mathbb{R}\vec{v}_1 + \mathbb{R}\vec{v}_2 + \cdots + \mathbb{R}\vec{v}_i.$$

By the definition of the orthogonal projection, we get $\vec{w}_{i+1} \perp H_i$.

Since $\text{proj}_{H_i}\vec{v}_{i+1}$ is a linear combination of $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_i\}$, the formula implies $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_i\}$ and $\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_i\}$ are linear combinations of each other. This further implies

$$H_i = \mathbb{R}\vec{w}_1 + \mathbb{R}\vec{w}_2 + \cdots + \mathbb{R}\vec{w}_i.$$

Then $\vec{w}_{i+1} \perp H_i$ means $\vec{w}_{i+1}$ is orthogonal to each of $\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_i$.

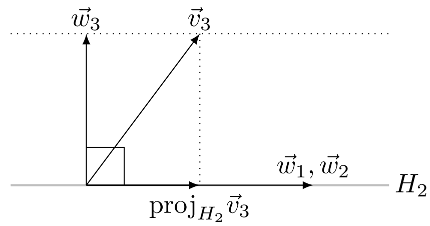

Figure 5.3.2: Gram-Schmidt process: $\vec{w}_3 = \vec{v}_3 - \text{proj}_{H_2}\vec{v}_3$.

The key trick is that, after obtaining the orthogonal set $\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_i\}$, we may use the formula in Proposition 5.3.3 to calculate $\text{proj}_{H_2}\vec{v}_3$

$$\vec{w}_{i+1} = \vec{v}_{i+1} - \text{proj}_{H_i}\vec{v}_{i+1} = \vec{v}_{i+1} - \frac{\vec{x}\cdot\vec{w}_1}{\vec{w}_1\cdot\vec{w}_1}\vec{w}_1 - \frac{\vec{x}\cdot\vec{w}_2}{\vec{w}_2\cdot\vec{w}_2}\vec{w}_2 - \cdots - \frac{\vec{x}\cdot\vec{w}_i}{\vec{w}_i\cdot\vec{w}_i}\vec{w}_i.$$

Example 5.3.3. In Example 3.4.6, the subspace H of $\mathbb{R}^3$ given by the equation $x+y+z=0$ has basis $\vec{v}_1 = (1,-1,0)$, $\vec{v}_2 = (1,0,-1)$. Then we derive an orthogonal basis of H

$$\begin{aligned}
\vec{w}_1 &= \vec{v}_1 = (1,-1,0),\\
\vec{w}_2' &= \vec{v}_2 - \frac{\vec{v}_2\cdot\vec{w}_1}{\vec{w}_1\cdot\vec{w}_1}\vec{w}_1 = (1,0,-1) - \frac{1+0+0}{1+1+0}(1,-1,0) = \frac{1}{2}(1,1,-2),\\
\vec{w}_2 &= 2\vec{w}_2' = (1,1,-2).
\end{aligned}$$

This is the orthogonal basis of H used for calculating the orthogonal projection onto H in Example 5.2.5.

Example 5.3.4. The subspace $x_1+2x_2-x_3-2x_4=0$ of $\mathbb{R}^4$ has basis $\vec{v}_1 = (1,0,1,0)$, $\vec{v}_2 = (2,-1,0,0)$, $\vec{v}_3 = (0,1,0,1)$. Applying the Gram-Schmidt process, we get

$$\begin{aligned}
\vec{w}_1 &= \vec{v}_1 = (1,0,1,0),\\
\vec{w}_2 &= \vec{v}_2 - \frac{\vec{v}_2\cdot\vec{w}_1}{\vec{w}_1\cdot\vec{w}_1}\vec{w}_1 = (2,-1,0,0) - \frac{2}{2}(1,0,1,0) = (1,-1,-1,0),\\
\vec{w}_3' &= \vec{v}_3 - \frac{\vec{v}_3\cdot\vec{w}_1}{\vec{w}_1\cdot\vec{w}_1}\vec{w}_1 - \frac{\vec{v}_3\cdot\vec{w}_2}{\vec{w}_2\cdot\vec{w}_2}\vec{w}_2\\
&= (0,1,0,1) - 0(1,0,1,0) - \frac{-1}{3}(1,-1,-1,0) = \frac{1}{3}(1,2,-1,3),\\
\vec{w}_3 &= 3\vec{w}_3' = (1,2,-1,3).
\end{aligned}$$

Exercise 5.29. Apply the Gram-Schmidt process to change to orthogonal bases.

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

Example 5.3.5. In Example 2.4.5, we get a basis $\{(1,4,7,10),(0,-3,-6,-9)\}$ of the row space of the matrix

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

We modify the basis to $\{(0,1,2,3),(1,4,7,10)\}$. Then the following produces an orthogonal basis $\{(0,1,2,3),(7,4,1,-2)\}$ of Row$A$.

$$(1,4,7,10)-\frac{(1,4,7,10)\cdot(0,1,2,3)}{(0,1,2,3)\cdot(0,1,2,3)}(0,1,2,3)=\frac{1}{7}(7,4,1,-2).$$

Then we get the formula for the orthogonal projection onto the row space

$$\begin{aligned}\mathrm{proj}_{\mathrm{Row}A}\vec{x}&=\frac{(x_1,x_2,x_3,x_4)\cdot(0,1,2,3)}{(0,1,2,3)\cdot(0,1,2,3)}\begin{pmatrix}0\\1\\2\\3\end{pmatrix}+\frac{(x_1,x_2,x_3,x_4)\cdot(7,4,1,-2)}{(7,4,1,-2)\cdot(7,4,1,-2)}\begin{pmatrix}7\\4\\1\\-2\end{pmatrix}\\&=\frac{x_2+2x_3+3x_4}{14}\begin{pmatrix}0\\1\\2\\3\end{pmatrix}+\frac{7x_1+4x_2+x_3-2x_4}{70}\begin{pmatrix}7\\4\\1\\-2\end{pmatrix}\\&=\frac{1}{10}\begin{pmatrix}7&4&1&-2\\4&3&2&1\\1&2&3&4\\-2&1&4&7\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\\x_4\end{pmatrix}.\end{aligned}$$

In the same example, we also get a basis $\{(1,1,1),(0,1,2)\}$ of the column space of the matrix

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

The following produces an orthogonal basis $\{(1,1,1),(-1,0,1)\}$ of Col$A$.

$$(0,1,2)-\frac{(0,1,2)\cdot(1,1,1)}{(1,1,1)\cdot(1,1,1)}(1,1,1)=(-1,0,1).$$

Then we get

$$
\begin{aligned}
\mathrm{proj}_{\mathrm{Col}A}\vec{x} &= \frac{(x_1,x_2,x_3)\cdot(1,1,1)}{(1,1,1)\cdot(1,1,1)}\begin{pmatrix}1\\1\\1\end{pmatrix} + \frac{(x_1,x_2,x_3)\cdot(-1,0,1)}{(-1,0,1)\cdot(-1,0,1)}\begin{pmatrix}-1\\0\\1\end{pmatrix}\\
&= \frac{x_1+x_2+x_3}{3}\begin{pmatrix}1\\1\\1\end{pmatrix} + \frac{-x_1+x_3}{2}\begin{pmatrix}-1\\0\\1\end{pmatrix} = \frac{1}{6}\begin{pmatrix}5&2&-1\\2&2&2\\-1&2&5\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}.
\end{aligned}
$$

Exercise 5.30. Find orthogonal bases of the null spaces of the matrix A and A^T in Example 5.3.5. Then find the formula of the orthogonal projections onto the subspaces.

Exercise 5.31. Find the orthogonal projections onto the subspaces spanned by the vectors in Exercise 5.29.

Exercise 5.32. Find the orthogonal projections onto the four subspaces associated to the matrix

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

5.3.3 QR-Decomposition

The vectors obtained by the Gram-Schmidt process are related in "triangular" way, with $a_{ii} \neq 0$

$$
\begin{aligned}
\vec{v}_1 &= a_{11}\vec{w}_1,\\
\vec{v}_2 &= a_{12}\vec{w}_1 + a_{22}\vec{w}_2,\\
&\vdots\\
\vec{v}_n &= a_{1n}\vec{w}_1 + a_{2n}\vec{w}_2 + \cdots + a_{nn}\vec{w}_n.
\end{aligned}
$$

The vectors $\vec{w}_i$ can also be expressed in $\vec{v}_i$ in similar triangular way. The relation can be rephrased in the matrix form

$$A = QR,\quad A = (\vec{v}_1\ \vec{v}_2\ \cdots\ \vec{v}_n),\ Q = (\vec{w}_1\ \vec{w}_2\ \cdots\ \vec{w}_n),\ R = \begin{pmatrix}a_{11}&a_{12}&\cdots&a_{1n}\\0&a_{22}&\cdots&a_{2n}\\\vdots&\vdots&&\vdots\\0&0&\cdots&a_{nn}\end{pmatrix}.$$

We may further divide the lengths of $\vec{w}_i$. Then $\vec{w}_1, \vec{w}_2, \dots, \vec{w}_n$ is orthonormal, and we get $Q^TQ = I$. In this case, we call $A = QR$ the *QR-decomposition* of A.

Proposition 5.3.5. *If the columns of a matrix A are linearly independent, then $A = QR$ for a unique matrix Q with orthonormal column and a unique upper triangular matrix R.*

The QR-decomposition is the matrix interpretation of the Gram-Schmidt process. To find the decomposition, we first use the Gram-Schmidt process to get orthogonal basis. Then we divide the length to calculate Q. Then we calculate the triangular matrix $Q^T A = Q^T QR = IR = R$.

Example 5.3.6. By dividing the lengths, we get an orthonormal basis for the subspace H in Example 5.3.3

$$\frac{\vec{w}_1}{\|\vec{w}_1\|} = \frac{(1,-1,0)}{\sqrt{2}}, \quad \frac{\vec{w}_2}{\|\vec{w}_2\|} = \frac{(1,1,-2)}{\sqrt{6}}.$$

Translated into matrix, we have

$$A = \begin{pmatrix} 1 & 1 \\ -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad Q = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ 0 & -\frac{\sqrt{2}}{\sqrt{3}} \end{pmatrix}.$$

Then

$$R = Q^T A = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{\sqrt{2}}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} \\ 0 & \frac{\sqrt{3}}{\sqrt{2}} \end{pmatrix},$$

and we get the QR-decomposition

$$\begin{pmatrix} 1 & 1 \\ -1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ 0 & -\frac{\sqrt{2}}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} \\ 0 & \frac{\sqrt{3}}{\sqrt{2}} \end{pmatrix}.$$

Example 5.3.7. In Example 3.4.6, we further extended the basis $\vec{v}_1, \vec{v}_2$ of H into a basis of $\mathbb{R}^3$ by adding $\vec{v}_3 = (1,1,1)$. The basis form the matrix in Example 3.4.6

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

Since $\vec{v}_3$ is already orthogonal to H, we may use the result of the Gram-Schmidt process in Example 5.3.3 to get an orthogonal basis

$$\vec{w}_1 = (1,-1,0), \quad \vec{w}_2 = (1,1,-2), \quad \vec{v}_3 = (1,1,1).$$

By dividing the lengths, we get the orthonormal basis in Example 5.2.5, and the corresponding matrix Q in Example 5.2.7. Then

$$R = Q^T A = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{pmatrix}^T \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix} = \begin{pmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} & 0 \\ 0 & \frac{\sqrt{3}}{\sqrt{2}} & 0 \\ 0 & 0 & \sqrt{3} \end{pmatrix}.$$

Exercise 5.33. Find the QR-decomposition based on Example 5.3.4.

Exercise 5.34. Find the QR-decompositions corresponding to Exercise 5.29.

5.4 Orthogonal Sum

5.4.1 Orthogonal Complement

Definition 5.4.1. Two subspaces H and H' are orthogonal, if

$$\vec{h} \in H,\ \vec{h}' \in H' \implies \vec{h} \perp \vec{h}'.$$

If $H_1, H_2, \dots, H_n$ are mutually orthogonal, then we call $H_1 + H_2 + \cdots + H_n$ an *orthogonal sum* and denote by $H_1 \perp H_2 \perp \cdots \perp H_n$.

A sum of lines $\mathbb{R}\vec{v}_1 + \mathbb{R}\vec{v}_2 + \cdots + \mathbb{R}\vec{v}_n$ is orthogonal if and only if $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ is an orthogonal set. The following extends Proposition 5.2.5.

Proposition 5.4.2. *Any orthogonal sum is direct.*

Given $H_1 \perp H_2 \perp H_3$ and $\vec{h}_i \in H_i$, we have

$$\vec{h}_1 + \vec{h}_2 + \vec{h}_3 = \vec{0} \implies \vec{h}_1 \cdot \vec{h}_1 = \vec{h}_1 \cdot \vec{h}_1 + \vec{h}_1 \cdot \vec{h}_2 + \vec{h}_1 \cdot \vec{h}_3 = 0 \implies \vec{h}_1 = \vec{0}.$$

By similar reason, we also get $\vec{h}_2 = \vec{h}_3 = \vec{0}$.

Exercise 5.35. Explain that the subspaces $H_1 + H_2 + \cdots + H_m$ and $H_1' + H_2' + \cdots + H_n'$ are orthogonal if and only if H_i and H_j' are all orthogonal for each i, j. For the special case H_i, H_j' are lines, what does this mean?

Exercise 5.36. Explain that the following are equivalent

1. orthogonal sums $H_1 \perp H_2$, $H_3 \perp H_4 \perp H_5$, $(H_1 + H_2) \perp (H_3 + H_4 + H_5)$.
2. orthogonal sum $H_1 \perp H_2 \perp H_3 \perp H_4 \perp H_5$.

Comparing with the discussion in Section 2.5.2, what can you say about layers of orthogonal sums.

A special case of orthogonal sum is the following.

Definition 5.4.3. The *orthogonal complement* of a subspace H is

$$H^{\perp} = \{\vec{v} \colon \vec{v} \perp \vec{h} \text{ for all } \vec{h} \in H\}.$$

The subspaces H and $H^\perp$ are automatically orthogonal to each other.

The orthogonal complement of

$$H = \mathbb{R}(1,2,3) + \mathbb{R}(4,5,6) = \text{Row}\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

consists of (x_1, x_2, x_3) satisfying

$$\begin{aligned} (1,2,3) \cdot (x_1, x_2, x_3) &= x_1 + 2x_2 + 3x_3 = 0, \\ (4,5,6) \cdot (x_1, x_2, x_3) &= 4x_1 + 5x_2 + 6x_3 = 0. \end{aligned}$$

This is exactly the definition of the null space. Therefore

$$\left(\text{Row}\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}\right)^\perp = \text{Nul}\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}.$$

In general, we get

$$(\text{Row}A)^\perp = \text{Nul}A, \quad (\text{Col}A)^\perp = \text{Nul}A^T.$$

For an $m \times n$ matrix A, we have the orthogonal sum subspace $\text{Row}A \perp \text{Nul}A \subset \mathbb{R}^n$. By Proposition 5.4.2, this is a direct sum. Then by Theorems 2.4.7 and 2.5.5, we get

$$\dim(\text{Row}A \perp \text{Nul}A) = \dim \text{Row}A + \dim \text{Nul}A = \text{rank}A + \dim \text{Nul}A = n.$$

Therefore

$$\text{Row}A \perp \text{Nul}A = \mathbb{R}^n.$$

Since any subspace of $\mathbb{R}^n$ is of the form $H = \text{Row}A$ for suitable A, the equality above means

$$H \perp H^\perp = \mathbb{R}^n.$$

This is a direct sum, and implies $\dim H + \dim H^\perp = n$.

Theorem 5.4.4. *A subspace H' is the orthogonal complement of a subspace $H \subset \mathbb{R}^n$, if and only if*

$$\mathbb{R}^n = H \perp H'.$$

Moreover, we have

$$(H^\perp)^\perp = H.$$

We already know that, if $H' = H^\perp$, then $\mathbb{R}^n = H \perp H'$. Conversely, if $\mathbb{R}^n = H \perp H'$, then the direct sum implies $\dim H + \dim H' = n$, and $H' \perp H$ implies $H' \subset H^\perp$. Then by

$$\dim H' = n - \dim H = \dim H^\perp,$$

we conclude $H' = H^{\perp}$.

To explain the second equality $(H^{\perp})^{\perp} = H$, we note that $\mathbb{R}^n = H \perp H^{\perp} = H^{\perp} \perp H$. Applying the first part of the theorem to the case H and H' are $H^{\perp}$ and H, we conclude $H = (H^{\perp})^{\perp}$.

Applying $(H^{\perp})^{\perp} = H$ to the equalities $(\text{Row}A)^{\perp} = \text{Nul}A$ and $(\text{Col}A)^{\perp} = \text{Nul}A^T$ that we already know, we get

$$(\text{Nul}A)^{\perp} = \text{Row}A, \quad (\text{Nul}A^T)^{\perp} = \text{Col}A.$$

Then we get

$$\text{Col}A \perp \text{Nul}A^T = \mathbb{R}^m, \quad \text{Row}A \perp \text{Nul}A = \mathbb{R}^n.$$

Example 5.4.1. Let

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

The column space $\text{Col}A$ has basis $\{(1,2,3),(4,5,6)\}$ from Example 2.4.3, and basis $\{(1,1,1),(0,1,2)\}$ from Example 2.4.5. In Example 1.3.5, we also know

$$\text{Col}A = \{(b_1,b_2,b_3) \colon b_3 - 2b_2 + b_1 = 0\} = (\mathbb{R}(1,-2,1))^{\perp}.$$

Therefore we get

$$\begin{aligned} \mathbb{R}^3 &= (\mathbb{R}(1,2,3) + \mathbb{R}(4,5,6)) \perp \mathbb{R}(1,-2,1) \\ &= (\mathbb{R}(1,1,1) + \mathbb{R}(0,1,2)) \perp \mathbb{R}(1,-2,1). \end{aligned}$$

The row space $\text{Row}A = \text{Col}A^T$ has basis $\{(1,4,7,10),(0,-3,-6,-9)\}$ from Example 2.4.3, and basis $\{(1,0,-1,-2),(0,1,2,3)\}$ from the row echelon form of A. We also get a basis $\{(1,-2,1,0),(2,-3,0,1)\}$ of $\text{Nul}A$ from Example 1.1.3. Then we get

$$\begin{aligned} \mathbb{R}^4 &= (\mathbb{R}(1,4,7,10) + \mathbb{R}(0,-3,-6,-9)) \perp (\mathbb{R}(1,-2,1,0) + \mathbb{R}(2,-3,0,1)) \\ &= (\mathbb{R}(1,0,-1,-2) + \mathbb{R}(0,1,2,3)) \perp (\mathbb{R}(1,-2,1,0) + \mathbb{R}(2,-3,0,1)). \end{aligned}$$

5.4.2 Pseudoinverse

Let $L \colon \mathbb{R}^n \to \mathbb{R}^m$ have $m \times n$ matrix A. We decompose the two Euclidean spaces into orthogonal (therefore direct) sums of the subspaces associated to A. Correspondingly, L has four components

$$L = \begin{pmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{pmatrix} \colon \mathbb{R}^n = \text{Row}A \perp \text{Nul}A \longrightarrow \mathbb{R}^m = \text{Col}A \perp \text{Nul}A^T,$$

where

$$\text{Row}A \xrightarrow{L_{11}} \text{Col}A, \qquad \text{Nul}A \xrightarrow{L_{12}} \text{Col}A,$$
$$\text{Row}A \xrightarrow{L_{21}} \text{Nul}A^T, \qquad \text{Nul}A \xrightarrow{L_{22}} \text{Nul}A^T.$$

First, we observe that the range $\text{Ran}L = \text{Col}A$, and does not overlap the orthogonal complement $\text{Nul}A^T$. This means both L_{21} and L_{22} are O.

Second, we observe that $L(\vec{x}) = A\vec{x} = \vec{0}$ if and only if $\vec{x} \in \text{Nul}A$. Since $\vec{x} \in \text{Nul}A$ implies $L(\vec{x}) = \vec{0}$, we get $L_{21} = O$. Then

$$L = \begin{pmatrix} L_{11} & O \\ O & O \end{pmatrix}, \quad L(\vec{x}+\vec{y}) = L_{11}(\vec{x}) \text{ for } \vec{x} \in \text{Row}A,\ \vec{y} \in \text{Nul}A.$$

This means L_{11} is the "core" of L. We also get

$$\begin{aligned} \text{Ran}L_{11} &= \text{Ran}L = \text{Col}A, \\ \text{Ker}L_{11} &= \{\vec{x} \in \text{Row}A \colon L(\vec{x}) = \vec{0}\} = \text{Nul}A \cap \text{Row}A = \{\vec{0}\}. \end{aligned}$$

This means L_{11} is onto and one-to-one. In other words, L_{11} is invertible.

The isomorphism L_{11} between $\text{Row}A$ and $\text{Col}A$ gives the material reason that $\text{rank}A^T = \text{rank}A$.

Since the core part L_{11} is invertible, we may invert this part and keep rest to be O. This gives the *pseudoinverse* of L

$$\begin{pmatrix} L_{11}^{-1} & O \\ O & O \end{pmatrix} : \text{Col}A \perp \text{Nul}A^T \to \text{Col}A^T \perp \text{Nul}A.$$

The matrix of the pseudoinverse of L is the pseudoinverse of the matrix A.

Example 5.4.2. The matrix

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

gives the linear transformation $L \colon \mathbb{R}^3 = \text{Row}A \perp \text{Nul}A \to \mathbb{R}^2 = \text{Col}A$. We have

$$L_{11} \colon \text{Row}A \to \text{Col}A, \quad L_{11}\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad L_{11}\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Therefore the pseudoinverse B of A is determined by

$$B\begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \quad B\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \quad B\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -1 & 0 \\ 0 & -1 \end{pmatrix} = A^T.$$

The matrix B can be calculated by the column operations

$$\begin{pmatrix} 2 & 1 \\ 1 & 2 \\ 1 & 1 \\ -1 & 0 \\ 0 & -1 \end{pmatrix} \to \begin{pmatrix} 1 & 2 \\ 2 & 1 \\ 1 & 1 \\ 0 & -1 \\ -1 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 0 \\ 2 & -3 \\ 1 & -1 \\ 0 & -1 \\ -1 & 2 \end{pmatrix} \to \begin{pmatrix} 1 & 0 \\ 2 & 1 \\ 1 & \frac{1}{3} \\ 0 & \frac{1}{3} \\ -1 & -\frac{2}{3} \end{pmatrix} \to \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \frac{1}{3} & \frac{1}{3} \\ -\frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{2}{3} \end{pmatrix}.$$

We conclude

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

In general, suppose and $m \times n$ matrix A has full rank in rows, i.e., rank$A = m$, or all rows of A are pivot. Then the corresponding linear transformation is $L\colon \mathbb{R}^n = \text{Row}A \perp \text{Nul}A \to \mathbb{R}^m = \text{Col}A$. The linear transformation $L_{11}\colon \text{Row}A \to \text{Col}A$ takes rows of A to the columns of AA^T. The pseudoinverse B takes the columns of AA^T back to rows of A, or columns of A^T. Therefore $BAA^T = A^T$. By rank$A = m$, we know AA^T is invertible. Then we get

$$B = A^T(AA^T)^{-1}.$$

In the example above, we have

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

Example 5.4.3. The matrix

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

gives the linear transformation $L\colon \mathbb{R}^2 = \text{Row}A \to \text{Col}A \perp \text{Nul}A^T$. We have

$$L_{11}\colon \text{Row}A \to \text{Col}A, \quad L_{11}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \quad L_{11}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix},$$

and

$$\text{Nul}A^T = \mathbb{R}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Therefore the pseudoinverse B of A is determined by

$$B\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad B\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad B\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

The matrix B can be calculated by the column operations

$$\begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 3 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{3} & -1 & 0 \\ \frac{1}{3} & 0 & -1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{1}{3} & -\frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & -\frac{2}{3} \end{pmatrix}.$$

We conclude

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

Compared with Example 5.4.2, this example suggests that, if B is the pseudoinverse of A, then B^T is the pseudoinverse of A^T.

In general, suppose and $m \times n$ matrix A has full rank in columns, i.e., $\text{rank}A = n$, or all columns of A are pivot. Then the corresponding linear transformation is $L\colon \mathbb{R}^n = \text{Row}A \to \mathbb{R}^m = \text{Col}A \perp \text{Nul}A^T$. The linear transformation $L_{11}\colon \text{Row}A \to \text{Col}A$ takes the standard basis of $\mathbb{R}^n$ to the columns of A. Therefore the pseudoinverse B of A takes the columns of A back to the standard basis. This means

$$BA = I.$$

Let a basis of $\text{Nul}A^T$ form the columns of a matrix C. Then the columns of $(A\ C)$ form a basis of $\mathbb{R}^m$. Therefore the matrix $(A\ C)$ is invertible. Moreover, the orthogonal property between $\text{Col}A$ and $\text{Nul}A^T$, i.e., the columns of A and C, imply

$$\begin{pmatrix} A^T \\ C^T \end{pmatrix} (A\ C) = \begin{pmatrix} A^TA & A^TC \\ C^TA & C^TC \end{pmatrix} = \begin{pmatrix} A^TA & O \\ O & C^TC \end{pmatrix}.$$

Then we get

$$(A\ C)^{-1} = \begin{pmatrix} (A^TA)^{-1} & O \\ O & (C^TC)^{-1} \end{pmatrix} \begin{pmatrix} A^T \\ C^T \end{pmatrix} = \begin{pmatrix} (A^TA)^{-1}A^T \\ (C^TC)^{-1}C^T \end{pmatrix}.$$

The pseudoinverse B of A satisfies $BC = O$. Therefore

$$B(A\ C) = (BA\ BC) = (I\ O),$$

and

$$B = (I\ O)(A\ C)^{-1} = (I\ O) \begin{pmatrix} (A^TA)^{-1}A^T \\ (C^TC)^{-1}C^T \end{pmatrix} = (A^TA)^{-1}A^T.$$

This explains the relation (at least for the case of full rank) between Examples 5.4.2 and 5.4.3, where we observe the pseudoinverse of transpose is the same as the transpose of the pseudoinverse.

Exercise 5.37. Consider a nonzero matrix $\vec{a}$ is an $n \times 1$ matrix. Show that pseudoinverse of the matrix $\vec{a}$ is $\frac{\vec{a}^T}{\|\vec{a}\|^2}$.

What is the pseudoinverse of the $1 \times n$ matrix $\vec{a}^T$.

Exercise 5.38. Find the pseudoinverses of the matrices.

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

Example 5.4.4. Examples 5.4.2 and 5.4.3 calculate the pseudoinverses of full rank matrixes. Here we consider a matrix that is not full rank

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

We have a basis $\{(1,0,-1,-2),(0,1,2,3)\}$ of Row$A$, and a basis $\{(1,-2,1)\}$ of NulA^T. Then we get

$$L_{11}\begin{pmatrix} 1 \\ 0 \\ -1 \\ -2 \end{pmatrix} = \begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \\ -1 \\ -2 \end{pmatrix} = \begin{pmatrix} -26 \\ -28 \\ -30 \end{pmatrix},$$

$$L_{11}\begin{pmatrix} 0 \\ 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 48 \\ 54 \\ 60 \end{pmatrix}.$$

The pseudoinverse B of A is determined by

$$B\begin{pmatrix} -26 \\ -28 \\ -30 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ -1 \\ -2 \end{pmatrix}, \quad B\begin{pmatrix} 48 \\ 54 \\ 60 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 2 \\ 3 \end{pmatrix}, \quad B\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Then we carry out the column operations

$$\begin{pmatrix} -26 & 48 & 1 \\ -28 & 54 & -2 \\ -30 & 60 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 2 & 0 \\ -2 & 3 & 0 \end{pmatrix} \to \begin{pmatrix} -26 & -4 & 1 \\ -28 & -2 & -2 \\ -30 & 0 & 1 \\ 1 & 2 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \\ -2 & -1 & 0 \end{pmatrix} \to \begin{pmatrix} 4 & 2 & 1 \\ -88 & 1 & -2 \\ 0 & 0 & 1 \\ 1 & -1 & 0 \\ 0 & -\frac{1}{2} & 0 \\ -1 & 0 & 0 \\ -2 & \frac{1}{2} & 0 \end{pmatrix}$$

$$\to \begin{pmatrix} 180 & 2 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \\ -87 & -1 & 0 \\ -44 & -\frac{1}{2} & 0 \\ -1 & 0 & 0 \\ 42 & \frac{1}{2} & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \\ -\frac{87}{180} & -\frac{1}{30} & 0 \\ -\frac{44}{180} & -\frac{1}{90} & 0 \\ -\frac{1}{180} & \frac{1}{90} & 0 \\ \frac{42}{180} & \frac{1}{30} & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -\frac{87}{180} & -\frac{1}{30} & \frac{5}{12} \\ -\frac{11}{45} & -\frac{1}{90} & \frac{2}{9} \\ -\frac{1}{180} & \frac{1}{90} & \frac{1}{36} \\ \frac{7}{20} & \frac{1}{30} & -\frac{1}{20} \end{pmatrix}$$

Therefore we get

$$B = \begin{pmatrix} -\frac{87}{180} & -\frac{1}{30} & \frac{5}{12} \\ -\frac{11}{45} & -\frac{1}{90} & \frac{2}{9} \\ -\frac{1}{180} & \frac{1}{90} & \frac{1}{36} \\ \frac{7}{20} & \frac{1}{30} & -\frac{1}{20} \end{pmatrix}.$$

Finally, we remark that the pseudoinverse of an $m\times n$ matrix A is usually defined as the $n \times m$ matrix B satisfying the following *Moore-Penrose conditions*

$$ABA = A, \quad BAB = B, \quad (AB)^T = AB, \quad (BA)^T = BA.$$

The definition does not give insights about the pseudoinverse.

Pseudoinverse is closely related to singular value decomposition, and can be calculated by the singular value decomposition.

5.4.3 Orthogonal Projection to Orthogonal Sum

Let $H = H_1 \perp H_2$. Then

$$\mathbb{R}^n = H \perp H^\perp = H_1 \perp H_2 \perp H^\perp.$$

Let

$$\vec{x} = \vec{h}_1 + \vec{h}_2 + \vec{h}', \quad \vec{h}_1 \in H_1,\ \vec{h}_2 \in H_2,\ \vec{h}' \in H^\perp.$$

By Theorem 5.4.4, we have $H_2 \perp H^\perp = H_1^\perp$ and $H_1 \perp H^\perp = H_2^\perp$. Then by the definition of orthogonal projection, we get

$$\vec{h}_1 = \text{proj}_{H_1}\vec{x}, \quad \vec{h}_2 = \text{proj}_{H_2}\vec{x}, \quad \text{proj}_H\vec{x} = \vec{h}_1 + \vec{h}_2 = \text{proj}_{H_1}\vec{x} + \text{proj}_{H_2}\vec{x}.$$

Proposition 5.4.5. $\text{proj}_{H_1\perp H_2\perp\cdots\perp H_k} = \text{proj}_{H_1} + \text{proj}_{H_2} + \cdots + \text{proj}_{H_k}$.

A special case is $H_1 = H$, $H_2 = H^\perp$, $H_1 \perp H_2 = \mathbb{R}^n$, and

$$\vec{x} = \text{proj}_{\mathbb{R}^n}\vec{x} = \text{proj}_H\vec{x} + \text{proj}_{H^\perp}\vec{x}.$$

Therefore calculating $\text{proj}_H\vec{x}$ is the same as calculating $\text{proj}_{H^\perp}\vec{x}$, and often one calculation is easier than the other.

Example 5.4.5. We calculated the matrix of the orthogonal projection onto the plane H given by $x+y+z=0$ using many different ways. Using Proposition 5.4.5, we may derive the matrix in the simplest way. By $H^\perp = \mathbb{R}(1,1,1)$, we get

$$\begin{aligned}\text{proj}_H\vec{x} = \vec{x} - \text{proj}_{\mathbb{R}(1,1,1)}\vec{x} &= \begin{pmatrix} x \\ y \\ z \end{pmatrix} - \frac{(x,y,z)\cdot(1,1,1)}{(1,1,1)\cdot(1,1,1)} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} x \\ y \\ z \end{pmatrix} - \frac{x+y+z}{3} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 2x-y-z \\ -x+2y-z \\ -x-y+2z \end{pmatrix}.\end{aligned}$$

In general, to find the matrix of the orthogonal projection onto the subspace $H \subset \mathbb{R}^n$ given by $\vec{a} \cdot \vec{x} = a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0$, we may first rescale $\vec{a}$, such that $\|\vec{a}\|^2 = a_1^2 + a_2^2 + \cdots + a_n^2 = 1$. Then

$$\begin{aligned}
\mathrm{proj}_H\vec{x} &= \vec{x} - \mathrm{proj}_{\mathbb{R}\vec{a}}\vec{x} = \vec{x} - (\vec{a}\cdot\vec{x})\vec{a} \\
&= \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} - (a_1x_1 + a_2x_2 + \cdots + a_nx_n)\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \\
&= \begin{pmatrix} 1-a_1^2 & -a_1a_2 & \cdots & -a_1a_n \\ -a_2a_1 & 1-a_2^2 & \cdots & -a_2a_n \\ \vdots & \vdots & & \vdots \\ -a_na_1 & a_na_2 & \cdots & 1-a_n^2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.
\end{aligned}$$

Example 5.4.6. In Example 5.3.5, we obtained the matrices for the orthogonal projections to Row$A$ and Col$A$ for

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

Then we get the matrices for the orthogonal projections to Nul$A$ and NulA^T

$$\begin{aligned}
[\mathrm{proj}_{\mathrm{Nul}A}] = I - [\mathrm{proj}_{\mathrm{Row}A}] &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} - \frac{1}{10}\begin{pmatrix} 7 & 4 & 1 & -2 \\ 4 & 3 & 2 & 1 \\ 1 & 2 & 3 & 4 \\ -2 & 1 & 4 & 7 \end{pmatrix} \\
&= \frac{1}{10}\begin{pmatrix} 3 & -4 & -1 & 2 \\ -4 & 7 & -2 & -1 \\ -1 & -2 & 7 & -4 \\ 2 & -1 & -4 & 3 \end{pmatrix}, \\
[\mathrm{proj}_{\mathrm{Nul}A^T}] = I - [\mathrm{proj}_{\mathrm{Col}A}] &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \frac{1}{6}\begin{pmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{pmatrix} = \frac{1}{6}\begin{pmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{pmatrix}.
\end{aligned}$$

In fact, it is easier to first calculate the projection to NulA^T, which is one dimensional, and then calculate the projection to Col$A$.

Exercise 5.39. Find the orthogonal projections onto the four subspaces associated to the matrix

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

5.4.4 Least Square Solution

A system of linear equations $A\vec{x} = \vec{b}$ has no solution if $\vec{b} \notin \mathrm{Col}A$. In this case, the next best thing we can do is to solve $A\hat{\vec{x}} = \hat{\vec{b}} = \mathrm{proj}_{\mathrm{Col}A}\vec{b}$, where $\hat{\vec{b}}$ is, by Proposition 5.3.2, the vector in $\mathrm{Col}A$ that is closest to $\vec{b}$. We call the solution $\hat{\vec{x}}$ the *least square solution*.

Example 5.4.7. The system of linear equations

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

has no solution. The possible right side to allow solutions form the subspace $H = \mathbb{R}(1,1,1) + \mathbb{R}(2,2,2) = \mathbb{R}(1,1,1)$. Then we have the orthogonal projection

$$\mathrm{proj}_H \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \frac{(1,2,3)\cdot(1,1,1)}{(1,1,1)\cdot(1,1,1)} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}.$$

Therefore the system for least square solution is

$$\begin{aligned} \hat{x} + 2\hat{y} &= 2, \\ \hat{x} + 2\hat{y} &= 2, \\ \hat{x} + 2\hat{y} &= 2. \end{aligned}$$

The solution is $\hat{x} = 2 - 2\hat{y}$, or $(\hat{x}, \hat{y}) = (2,0) + \hat{y}(-2,1)$, $\hat{y}$ arbitrary.

Calculating the orthogonal projection $\hat{\vec{b}}$ could be complicated. We have

$$\vec{b} - \hat{\vec{b}} \in (\mathrm{Col}A)^{\perp} = \mathrm{Nul}A^T.$$

Therefore

$$\vec{0} = A^T(\vec{b} - \hat{\vec{b}}) = A^T(\vec{b} - A\hat{\vec{x}}) = A^T\vec{b} - A^TA\hat{\vec{x}}.$$

In other words, we may solve

$$A^TA\hat{\vec{x}} = A^T\vec{b},$$

which does not require the calculation of $\hat{\vec{b}}$.

Example 5.4.8. To get the least square solution of the system in Example 5.4.7, we calculate A^TA and $A^T\vec{b}$

$$\begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 6 \\ 6 & 12 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 6 \\ 12 \end{pmatrix}.$$

Then the system $A^TA\hat{\vec{x}} = A^T\vec{b}$ is

$$\begin{aligned} 3\hat{x} + 6\hat{y} &= 6, \\ 6\hat{x} + 12\hat{y} &= 12. \end{aligned}$$

We get the same general solution $\hat{x} = 2 - 2\hat{y}$.

Exercise 5.40. Find the least square solutions.

1. $\begin{aligned} x_1 + x_2 &= 9, \\ 2x_1 - 2x_2 &= 7, \\ 3x_1 + x_2 &= 11. \end{aligned}$

2. $\begin{aligned} x_1 - 2x_2 &= 1, \\ 0 &= 1, \\ 2x_1 - x_2 &= 2. \end{aligned}$

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

Example 5.4.9. Suppose an experiment on the relations between two related quantities x and y gives data $(x, y) = (-6, -1), (-2, 2), (1, 1), (7, 6)$. We wish to find a linear relation $y = a + bx$ that best fits the data.

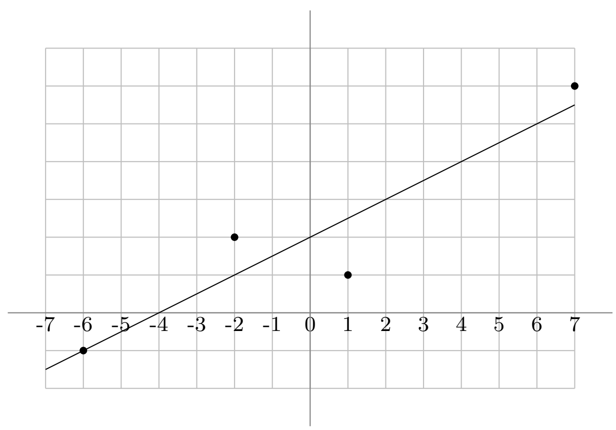

Figure 5.4.1: Fitting a straight line to data.

If the relation $y = a + bx$ is a perfect fit, then we would have

$$\begin{aligned} a - 6b &= -1, \\ a - 2b &= 2, \\ a + b &= 1, \\ a + 7b &= 6. \end{aligned}$$

Of course the system has not solution. The best fit means finding a and b, such that the left and right are the closest. This means finding the least square solution of the system.

We have

$$A = \begin{pmatrix} 1 & -6 \\ 1 & -2 \\ 1 & 1 \\ 1 & 7 \end{pmatrix}, \quad \vec{b} = \begin{pmatrix} -1 \\ 2 \\ 1 \\ 6 \end{pmatrix}, \quad A^T A = \begin{pmatrix} 4 & 0 \\ 0 & 90 \end{pmatrix}, \quad A^T \vec{b} = \begin{pmatrix} 8 \\ 45 \end{pmatrix}.$$

Then the least square system becomes

$$\begin{pmatrix} 4 & 0 \\ 0 & 90 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 8 \\ 45 \end{pmatrix}.$$

We get $a = 2$ and $b = \frac{1}{2}$, and the best fit is $y = 2 + \frac{1}{2}x$.

If $\text{Nul}A = \{\vec{0}\}$, i.e., solution of $A\vec{x} = \vec{b}$ is unique, then by Proposition 5.1.3 and the subsequent discussion, we know $A^T A$ is invertible. Then the solution of the least square system $A^T A\hat{\vec{x}} = A^T\vec{b}$ is

$$\hat{\vec{x}} = (A^T A)^{-1} A^T \vec{b}.$$

Moreover, since the columns of A are linearly independent, by Proposition 5.3.5, we have $A = QR$, where $Q^T Q = I$, and R is upper triangular and invertible. Then

$$\begin{aligned} A^T A &= R^T Q^T Q R = R^T R, \\ (A^T A)^{-1} A^T &= R^{-1} (R^T)^{-1} R^T Q^T = R^{-1} Q^T, \end{aligned}$$

and the least square solution

$$\hat{\vec{x}} = R^{-1} Q^T \vec{b}.$$

We remark that the inverse R^{-1} of upper triangular matrix is very easy to calculate.

5.5 Inner Product

5.5.1 Inner Product and Geometry

The dot product can be extended to general vector spaces.

Definition 5.5.1. An *inner product* in a real vector space V is a function

$$\langle \vec{x}, \vec{y} \rangle \colon V \times V \to \mathbb{R},$$

such that the following are satisfied:

1. Bilinearity: $\langle a\vec{x} + b\vec{y}, \vec{z} \rangle = a\langle \vec{x}, \vec{z} \rangle + b\langle \vec{y}, \vec{z} \rangle$, $\langle \vec{x}, a\vec{y} + b\vec{z} \rangle = a\langle \vec{x}, \vec{y} \rangle + b\langle \vec{x}, \vec{z} \rangle$.
2. Symmetry: $\langle \vec{x}, \vec{y} \rangle = \langle \vec{y}, \vec{x} \rangle$.

3. Positivity: $\langle\vec{x},\vec{x}\rangle \geq 0$, and $\langle\vec{x},\vec{x}\rangle = 0$ if and only if $\vec{x}=\vec{0}$.

An *inner product space* is a vector space equipped with an inner product.

Example 5.5.1. The dot product is not the only inner product in the Euclidean space. The following is also an inner product in $\mathbb{R}^n$

$$\langle\vec{x},\vec{y}\rangle = x_1y_1 + 2x_2y_2 + \cdots + nx_ny_n.$$

Example 5.5.2. In the vector space $C[0,1]$ of all continuous functions on $[0,1]$, we may introduce the inner product

$$\langle f,g\rangle = \int_0^1 f(t)g(t)dt.$$

This is also an inner product in the vector space P_n of polynomials of degree $\leq n$, or the vector space of continuous periodic functions on $\mathbb{R}$ of period 1.

The interval $[0,1]$ can be changed to any bounded interval $[a,b]$, and we get an inner product in $C[a,b]$

$$\langle f,g\rangle_{[a,b]} = \int_a^b f(t)g(t)dt.$$

The following is also an inner product in $C[0,1]$

$$\langle f,g\rangle = \int_0^1 tf(t)g(t)dt.$$

The function t can be replaced by any positive continuous function.

Exercise 5.41. Extend Proposition 5.1.1 to inner product space: If $\langle\vec{x},\vec{y}\rangle = \langle\vec{x},\vec{z}\rangle$ for all $\vec{x}$, prove that $\vec{y}=\vec{z}$.

Exercise 5.42. Extend Exercise 5.5 to inner product space: If $\langle\vec{x},L(\vec{y})\rangle = \langle\vec{x},K(\vec{y})\rangle$ for all $\vec{x}$ and $\vec{y}$, prove that $L=K$.

Similar to the dot product, inner product induces geometry.

Example 5.5.3. With respect to the first inner product $\langle f,g\rangle_{[0,1]}$ in Example 5.5.2, the lengths of functions 1 and t are

$$\|1\| = \sqrt{\int_0^1 1^2dt} = 1, \quad \|t\| = \sqrt{\int_0^1 t^2dt} = \frac{1}{\sqrt{3}}.$$

The angle between 1 and t is given by

$$\cos\theta = \frac{1}{\|1\|\,\|t\|}\int_0^1 tdt = \frac{\sqrt{3}}{2}, \quad \theta = \frac{\pi}{6}.$$

The area of the parallelogram spanned by 1 and t is

$$\sqrt{\int_0^1 dt \int_0^1 t^2 dt - \left(\int_0^1 t dt\right)^2} = \frac{1}{2\sqrt{3}}.$$

Exercise 5.43. Find the lengths, angles and the areas of the parallelograms with respect to the inner product in Example 5.5.1.

1. $(1,0)$, $(0,1)$.
2. $(1,2)$, $(2,-1)$.
3. $(1,1,0)$, $(0,1,1)$.
4. $(1,2,3)$, $(2,3,4)$.

Exercise 5.44. Find the areas of the triangles with given vertices, with respect to the three inner products in Example 5.5.2.

1. 1, t, t^2.
2. 0, $\sin \pi t$, $\cos \pi t$.
3. 1, $1+t$, $1+t+t^2$.
4. $1-t$, $t-t^2$, t^2-1.

The Cauchy-Schwartz Inequality (Theorem 5.1.5) can be extended.

Theorem 5.5.2 (Cauchy-Schwartz Inequality). *For any $\vec{x}$ and $\vec{y}$, we have*

$$|\langle \vec{x}, \vec{y}\rangle| \le \|\vec{x}\|\|\vec{y}\|.$$

Here we give a general proof. First, the inequality holds for $\vec{x} = \vec{0}$. For $\vec{x} \ne \vec{0}$, by the positivity, we have $\|\vec{x}\|^2 = \langle \vec{x}, \vec{x}\rangle > 0$. Then for any t, we have

$$\begin{aligned}
\langle t\vec{x} + \vec{y}, t\vec{x} + \vec{y}\rangle &= t^2\langle \vec{x}, \vec{x}\rangle + 2t\langle \vec{x}, \vec{y}\rangle + \langle \vec{y}, \vec{y}\rangle = t^2\|\vec{x}\|^2 + 2t\langle \vec{x}, \vec{y}\rangle + \|\vec{y}\|^2 \\
&= \left(t\|\vec{x}\| + \frac{\langle \vec{x}, \vec{y}\rangle}{\|\vec{x}\|}\right)^2 + \frac{1}{\|\vec{x}\|^2}\left(\|\vec{x}\|^2\|\vec{y}\|^2 - \langle \vec{x}, \vec{y}\rangle^2\right).
\end{aligned}$$

The last equality is obtained by completing the square. By the positivity, the left side ≥ 0 for all t. Then for the right side to be always ≥ 0, we need

$$\frac{1}{\|\vec{x}\|^2}\left(\|\vec{x}\|^2\|\vec{y}\|^2 - \langle \vec{x}, \vec{y}\rangle^2\right) \ge 0.$$

This proves Theorem 5.5.2.

Exercise 5.45. Prove the polarisation identities in Example 5.1.4 and Exercises 5.15, 5.16.

$$\begin{aligned}
\vec{x} \cdot \vec{y} &= \tfrac{1}{2}(\|\vec{x} + \vec{y}\|^2 - \|\vec{x}\|^2 - \|\vec{y}\|^2) \\
&= -\tfrac{1}{2}(\|\vec{x} - \vec{y}\|^2 - \|\vec{x}\|^2 - \|\vec{y}\|^2) \\
&= \tfrac{1}{4}(\|\vec{x} + \vec{y}\|^2 - \|\vec{x} - \vec{y}\|^2).
\end{aligned}$$

Exercise 5.46 (Parallelogram Law). Prove that $\|\vec{x} + \vec{y}\|^2 + \|\vec{x} - \vec{y}\|^2 = 2(\|\vec{x}\|^2 + \|\vec{y}\|^2)$.

5.5.2 Positive Definite Matrix

The first property of inner product is bilinearity. Let $f\colon V \times V \to \mathbb{R}$ be a bilinear function. Let $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ be a basis of V. Then by the discussion in Section 5.1.2, we know

$$f(\vec{x}, \vec{y}) = [\vec{x}]_\alpha^T A[\vec{y}]_\alpha = [\vec{x}]_\alpha \cdot A[\vec{y}]_\alpha, \quad A = [f]_{\alpha\alpha} = (f(\vec{v}_i, \vec{v}_j)).$$

The second property of inner product is symmetry. This means $f(\vec{v}_i, \vec{v}_j) = f(\vec{v}_j, \vec{v}_i)$. In other words, A is a symmetric matrix.

Example 5.5.4. For the inner product $\langle f, g\rangle_{[0,1]}$ in Example 5.5.2, the matrix with respect to the basis $\{1, t, t^2\}$ of P_2 has entries

$$a_{ij} = \int_0^1 t^i t^j dt = \frac{1}{i+j+1}, \quad i, j = 0, 1, 2.$$

Therefore the matrix is

$$[\langle \cdot, \cdot \rangle_{[0,1]}]_{\{1,t,t^2\}\,\{1,t,t^2\}} = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{pmatrix}.$$

Exercise 5.47. Find the matrix of the dot products with respect to bases.

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

In general, if the basis form the columns of a square matrix B, what is the matrix?

Exercise 5.48. Find the matrices of the inner products $\langle \cdot, \cdot \rangle_{[0,1]}$, $\langle \cdot, \cdot \rangle_{[-1,1]}$ in Example 5.5.2, with respect to the following bases.

1. P_2: $t^2, t, 1$.
2. P_3: $1, t, t^2, t^3$.
3. P_2: $1 - 2t, 2 + t^2, 1 + 2t - t^2$.
4. P_2: $t(t-1), t(t-2), (t-1)(t-2)$.

The third property of inner product is positivity. This leads to the following.

Definition 5.5.3. A symmetric matrix A is *positive definite* if $\vec{x}^T A\vec{x} = \vec{x} \cdot A\vec{x} > 0$ for all $\vec{x} \neq \vec{0}$. The matrix is *negative definite* if $\vec{x}^T A\vec{x} < 0$ for all $\vec{x} \neq \vec{0}$. It is *indefinite* if $\vec{x}^T A\vec{x} > 0$ for some $\vec{x}$ and $\vec{y}^T A\vec{y} < 0$ for some other $\vec{y}$.

We also call A *positive semi-definite* if $\vec{x}^T A\vec{x} \geq 0$ for all $\vec{x}$, and *negative semi-definite* if $\vec{x}^T A\vec{x} \leq 0$ for all $\vec{x}$.

Proposition 5.5.4. *Any inner product is given by $\langle \vec{x}, \vec{y} \rangle = [\vec{x}]_\alpha^T A[\vec{y}]_\alpha = [\vec{x}]_\alpha \cdot A[\vec{y}]_\alpha$, with a positive definite (symmetric) matrix A.*

In the Euclidean space, we denote the inner product given by a positive definite matrix A by

$$\langle \vec{x}, \vec{y} \rangle_A = \vec{x}^T A\vec{y} = \vec{x} \cdot A\vec{y}.$$

Example 5.5.5. For a diagonal matrix, we have

$$A = \begin{pmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a_n \end{pmatrix}, \quad \langle \vec{x}, \vec{y} \rangle_A = a_1x_1y_1 + a_2x_2y_2 + \cdots + a_nx_ny_n.$$

The matrix is positive definite, and gives an inner product, if and only if all $a_i > 0$. Example 5.5.1 is the special case $a_i = i$.

Example 5.5.6. For $A = \begin{pmatrix} 1 & 2 \\ 2 & a \end{pmatrix}$, we have

$$\vec{x}^T A\vec{x} = x_1^2 + 4x_1x_2 + ax_2^2 = (x_1 + 2x_2)^2 + (a - 4)x_2^2.$$

The right side is positive for all $(x_1, x_2) \neq (0, 0)$ if and only if $a > 4$. This is the condition for

$$\langle \vec{x}, \vec{y} \rangle_A = x_1y_1 + 2x_1y_2 + 2x_2y_1 + ax_2y_2$$

to be an inner product in $\mathbb{R}^2$.

Example 5.5.7. In Example 5.5.4, we get the matrix of the first inner product in Example 5.5.2 with respect to the basis $1, t, t^2$

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

Then

$$\begin{aligned} \vec{x}^T A\vec{x} &= x_1^2 + \tfrac{1}{3}x_2^2 + \tfrac{1}{5}x_3^2 + x_1x_2 + \tfrac{2}{3}x_1x_3 + \tfrac{1}{2}x_2x_3 \\ &= [x_1^2 + 2(\tfrac{1}{2}x_2 + \tfrac{1}{3}x_3) + (\tfrac{1}{2}x_2 + \tfrac{1}{3}x_3)^2] + \tfrac{1}{3}x_2^2 + \tfrac{1}{5}x_3^2 + \tfrac{1}{2}x_2x_3 - (\tfrac{1}{2}x_2 + \tfrac{1}{3}x_3)^2 \\ &= (x_1 + \tfrac{1}{2}x_2 + \tfrac{1}{3}x_3)^2 + \tfrac{1}{12}x_2^2 + \tfrac{4}{45}x_3^2 + \tfrac{1}{6}x_2x_3 \\ &= (x_1 + \tfrac{1}{2}x_2 + \tfrac{1}{3}x_3)^2 + \tfrac{1}{12}[x_2^2 + 2x_1x_2 + x_3^2] + \tfrac{4}{45}x_3^2 - \tfrac{1}{12}x_3^2 \\ &= (x_1 + \tfrac{1}{2}x_2 + \tfrac{1}{3}x_3)^2 + \tfrac{1}{12}(x_2 + x_3)^2 + \tfrac{1}{180}x_3^2 \\ &= y_1^2 + \tfrac{1}{12}y_2^2 + \tfrac{1}{180}y_3^2. \end{aligned}$$

We eliminate the cross terms x_ix_j, $i \neq j$, by the following change of variables

$$\begin{aligned} x_1 + \tfrac{1}{2}x_2 + \tfrac{1}{3}x_3 &= y_1, \\ x_2 + x_3 &= y_2, \\ x_3 &= y_3. \end{aligned}$$

All the coefficients in the final results are positive, because the inner product is positive definite.

The method in Examples 5.5.6 and 5.5.7 is called *completing the square*. The result is the elimination of cross terms. The symmetric matrix is positive definite if and only if all the coefficients in the final result are positive.

Completing the square is actually equivalent to row operations. For the symmetric matrix in Example 5.5.6, we use row operations to eliminate the entries below the diagonal

$$\begin{pmatrix} 1 & 2 \\ 2 & a \end{pmatrix} \xrightarrow{\text{Row}_2 - 2\text{Row}_1} \begin{pmatrix} 1 & 2 \\ 0 & a-4 \end{pmatrix}.$$

This produces a non-symmetric matrix. To get back the symmetric matrix, we need to carry out the same column operation

$$\begin{pmatrix} 1 & 2 \\ 0 & a-4 \end{pmatrix} \xrightarrow{\text{Col}_2 - 2\text{Col}_1} \begin{pmatrix} 1 & 0 \\ 0 & a-4 \end{pmatrix}.$$

This the the matrix for the final result in Example 5.5.6.

For the matrix in Example 5.5.7, we use row operations to eliminate entries below the diagonal[2]

$$\begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \end{pmatrix} \xrightarrow[\text{Row}_3 - \frac{1}{3}\text{Row}_1]{\text{Row}_2 - \frac{1}{2}\text{Row}_1} \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ 0 & \frac{1}{12} & \frac{1}{12} \\ 0 & \frac{1}{12} & \frac{4}{45} \end{pmatrix} \xrightarrow{\text{Row}_3 - \frac{1}{2}\text{Row}_1} \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ 0 & \frac{1}{12} & \frac{1}{12} \\ 0 & 0 & \frac{1}{180} \end{pmatrix}$$

Then exactly the same column operations eliminate the entries above the diagonal, and gives the diagonal matrix with the same diagonal entries $1, \frac{1}{12}, \frac{1}{180}$.

Exercise 5.49. Determine positive definite property.

1. $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 2 & 4 \\ 3 & 4 & 5 \end{pmatrix}$.
2. $\begin{pmatrix} 1 & 3 & 1 \\ 3 & 13 & 9 \\ 1 & 9 & 14 \end{pmatrix}$.
3. $\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{pmatrix}$.

[2] If you examine the process, you find the row operations exactly correspond to the calculation in Example 5.5.7.

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

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

6. $\begin{pmatrix} 1 & -1 & 1 \\ -1 & 0 & -3 \\ 1 & -3 & 3 \end{pmatrix}$.

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

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

Exercise 5.50. Find the condition for the symmetric matrix to be positive definite.

1. $\begin{pmatrix} 1 & a \\ a & b \end{pmatrix}$.

2. $\begin{pmatrix} a & 1 \\ 1 & b \end{pmatrix}$.

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

Exercise 5.51. Prove that $\begin{pmatrix} a & b \\ b & c \end{pmatrix}$ is positive definite if and only if $a > 0$ and $ac > b^2$.

Exercise 5.52. Prove that positive definite matrices are invertible.

Exercise 5.53. Prove that all the diagonal terms in a positive definite matrix must be positive.

Exercise 5.54. Suppose A and B are positive definite, and $a, b > 0$. Prove that $aA + bB$ is positive definite, and $-A$ is negative definite.

Exercise 5.55. For symmetric matrices A and B, prove that $\begin{pmatrix} A & O \\ O & B \end{pmatrix}$ is positive definite if and only if both A and B are positive definite.

Exercise 5.56. Suppose columns of A (not necessarily a square matrix) are linearly independent. Explain that A^TA (this is a symmetric matrix) is positive definite.

Exercise 5.57. Suppose the matrix of an inner product with respect to a basis is A. Show that the matrix with respect to another basis is P^TAP for a suitable invertible matrix P (P is actually the matrix for changing basis). In particular, this shows that, if A is positive definite, then P^TAP is also positive definite.

5.5.3 Orthogonality

Two vectors $\vec{x}$ and $\vec{y}$ are *orthogonal* with respect to an inner product $\langle \cdot, \cdot \rangle$, and denoted $\vec{x} \perp \vec{y}$, if $\langle \vec{x}, \vec{y} \rangle = 0$. A vector $\vec{x}$ is orthogonal to a subspace H, and denoted $\vec{x} \perp H$, if $\vec{x} \perp \vec{h}$ for all $\vec{h} \in H$.

Proposition 5.2.2 holds in general, with the same proof.

Theorem 5.5.5 (Pythagorean Theorem). *$\vec{x} \perp \vec{y}$ if and only if $\|\vec{x}+\vec{y}\|^2 = \|\vec{x}\|^2 + \|\vec{y}\|^2$.*

All the discussion in Section 5.2.1 about orthogonal projection and orthogonalisation is still valid for inner product. For a subspace $H \subset V$, we express any $\vec{x} \in V$ as follows

$$\vec{x} = \vec{y} + \vec{z}, \quad \vec{y} \in H,\ \vec{z} \perp H.$$

then we define the orthogonal projection of $\vec{x}$ onto H

$$\mathrm{proj}_H \vec{x} = \vec{y}.$$

If $\alpha = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is an orthogonal basis of H, then the orthogonal projection can still be calculated by the formula in Proposition 5.3.3

$$\mathrm{proj}_H \vec{x} = \frac{\langle \vec{x}, \vec{v}_1 \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle}\vec{v}_1 + \frac{\langle \vec{x}, \vec{v}_2 \rangle}{\langle \vec{v}_2, \vec{v}_2 \rangle}\vec{v}_2 + \cdots + \frac{\langle \vec{x}, \vec{v}_n \rangle}{\langle \vec{v}_n, \vec{v}_n \rangle}\vec{v}_n.$$

Based on the formula, we have the Gram-Schmidt process that modifies any basis of H to an orthogonal basis of H.

Example 5.5.8. We wish to calculate the orthogonal projection onto the subspace H of $\mathbb{R}^3$ given by $x_1 + x_2 + x_3 = 0$, with respect to the inner product $\langle \vec{x}, \vec{y} \rangle = x_1y_1 + 2x_2y_2 + 3x_3y_3$ in Example 5.5.1.

The subspace has basis given by

$$\vec{v}_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \quad \vec{v}_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

We apply the orthogonalisation process, with respect to the new inner product. We take $\vec{w}_1 = \vec{v}_1$. Then

$$\begin{aligned}
\vec{v}_2 - \mathrm{proj}_{\mathbb{R}\vec{w}_1} \vec{v}_2 &= \vec{v}_2 - \frac{\langle \vec{v}_2, \vec{w}_1 \rangle}{\langle \vec{w}_1, \vec{w}_1 \rangle}\vec{w}_1 \\
&= \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} - \frac{1 \cdot 1 + 2(0 \cdot (-1)) + 3((-1) \cdot 0)}{1 \cdot 1 + 2((-1) \cdot (-1)) + 3(0 \cdot 0)} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 2 \\ 1 \\ -3 \end{pmatrix}.
\end{aligned}$$

Then we multiply the vector by 3 to get $\vec{w}_1$, and

$$\begin{aligned}
\mathrm{proj}_H \vec{x} &= \frac{\langle \vec{x}, \vec{w}_1 \rangle}{\langle \vec{w}_1, \vec{w}_1 \rangle}\vec{w}_1 + \frac{\langle \vec{x}, \vec{w}_2 \rangle}{\langle \vec{w}_2, \vec{w}_2 \rangle}\vec{w}_2 \\
&= \frac{x_1 \cdot 1 + 2(x_2 \cdot (-1)) + 3(x_3 \cdot 0)}{1 \cdot 1 + 2((-1) \cdot (-1)) + 3(0 \cdot 0)} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \\
&\quad + \frac{x_1 \cdot 2 + 2(x_2 \cdot 1) + 3(x_3 \cdot (-3))}{2 \cdot 2 + 2(1 \cdot 1) + 3((-3) \cdot (-3))} \begin{pmatrix} 2 \\ 1 \\ -3 \end{pmatrix} \\
&= \frac{1}{11}\begin{pmatrix} 5x_1 - 6x_2 - 6x_3 \\ -3x_1 + 8x_2 - 3x_3 \\ 2x_1 + 2x_2 - 9x_3 \end{pmatrix} = \frac{1}{11}\begin{pmatrix} 5 & -6 & -6 \\ -3 & 8 & -3 \\ -2 & -2 & 9 \end{pmatrix}\vec{x}.
\end{aligned}$$

Exercise 5.58. Find the orthogonal projections with respect to the inner products.

1. Orthogonal projection of (x_1, x_2) onto $\mathbb{R}(2,1)$, with respect to $\langle \vec{x}, \vec{y} \rangle = x_1y_1 + x_1y_2 + x_2y_1 + 2x_2y_2$.
2. Orthogonal projection of $(1,2,3)$ onto $\mathbb{R}(1,1,1)$, with respect to $\langle \vec{x}, \vec{y} \rangle = x_1y_1 + 2x_2y_2 + 3x_3x_3$.
3. Orthogonal projection of $(1,1,1)$ onto H given by $x_1 + x_2 + x_3 = 0$, with respect to $\langle \vec{x}, \vec{y} \rangle = x_1y_1 + 2x_2y_2 + 3x_3y_3 - x_1y_2 - x_2y_1$.

Example 5.5.9. With respect to the inner product $\langle \cdot, \cdot \rangle_{[0,1]}$ in Example 5.5.2, we have

$$\langle t, t-a \rangle = \int_0^1 t(t-a)dt = \frac{1}{3} - \frac{1}{2}a.$$

Therefore $t \perp t - a$ if and only if $a = \frac{2}{3}$. By dividing the lengths

$$\|t\| = \sqrt{\int_0^1 t^2 dt} = \frac{1}{\sqrt{3}}, \quad \|t - \tfrac{2}{3}\| = \sqrt{\int_0^1 \left(t - \frac{2}{3}\right)^2 dt} = \frac{1}{3},$$

we get an orthonormal basis $\{\sqrt{3}t, 3t-2\}$ of P_1.

Using the orthonormal basis, the orthogonal projection of $t^2 \in P_2$ onto the subspace P_1 is

$$\begin{aligned}\operatorname{proj}_{P_1} t^2 &= \langle t^2, \sqrt{3}t \rangle \sqrt{3}t + \langle t^2, 3t-2 \rangle (3t-2) \\ &= \sqrt{3}t \int_0^1 \sqrt{3}t^3 dt + (3t-2) \int_0^1 t^2(3t-2)dt = t - \frac{1}{6}.\end{aligned}$$

Example 5.5.10. The natural basis $\{1, t, t^2\}$ of P_2 is not orthogonal with respect to the inner product $\langle f, g \rangle_{[0,1]}$ in Example 5.5.2. We apply the Gram-Schmidt process

$$\begin{aligned} f_1 &= 1, \\ f_2 &= t - \frac{\int_0^1 t \cdot 1 dt}{\int_0^1 1^2 dt} 1 = t - \frac{1}{2}, \\ f_3 &= t^2 - \frac{\int_0^1 t^2 \cdot 1 dt}{\int_0^1 1^2 dt} 1 - \frac{\int_0^1 t^2 \left(t - \frac{1}{2}\right) dt}{\int_0^1 \left(t - \frac{1}{2}\right)^2 dt} \left(t - \frac{1}{2}\right) = t^2 - t + \frac{1}{6}. \end{aligned}$$

By rescaling, we find that $1, 2t-1, 6t^2 - 6t + 1$ is an orthogonal basis of P_2.

Exercise 5.59. Apply the Gram-Schmidt process.

1. $1, t, t^2$, with respect to $\langle f, g \rangle_{[0,2]} = \int_0^2 f(t)g(t)dt$.
2. $t^2, t, 1$, with respect to $\langle f, g \rangle_{[0,1]} = \int_0^1 f(t)g(t)dt$.

3. $1, t, t^2$, with respect to $\langle f, g\rangle_{[0,1]} = \int_0^1 tf(t)g(t)dt$.

Example 5.5.11 (Fourier series). In the vector space V of periodic functions of period 2π, we use the inner product

$$\langle f(t), g(t)\rangle = \frac{1}{2\pi}\int_0^{2\pi} f(t)g(t)dt.$$

Basic examples of vectors in V are the trigonometric functions $\cos nt$ and $\sin nt$, for integers. By $\cos(-\theta) = \cos\theta$, $\sin(-\theta) = -\sin\theta$, and $\sin 0 = 0$, we only need consider $n = 0, 1, 2, \dots$ for $\cos nt$, and $n = 1, 2, \dots$ for $\sin nt$. The *Fourier series* of a function $f(t) \in V$ is

$$\begin{aligned} f(t) &= a_0 + a_1\cos t + b_1\sin t + a_2\cos 2t + b_2\sin 2t + \cdots + a_n\cos nt + b_n\sin nt + \cdots \\ &= a_0 + \sum_{n=1}^{\infty}(a_n\cos nt + b_n\sin nt). \end{aligned}$$

We regard the trigonometric functions as a basis of V, and $a_n, b)n$ are the coordinates.

For integers $m \neq n$, we have

$$\begin{aligned} \langle\cos mt, \cos nt\rangle &= \frac{1}{2\pi}\int_0^{2\pi}\cos mt\cos ntdt \\ &= \frac{1}{4\pi}\int_0^{2\pi}(\cos(m+n)t + \cos(m-n)t)dt \\ &= \frac{1}{4\pi}\left(\frac{\sin(m+n)t}{m+n} + \frac{\sin(m-n)t}{m-n}\right)_0^{\pi} = 0. \end{aligned}$$

We may similarly find $\langle\sin mt, \sin nt\rangle = 0$ for $m \neq n$, and $\langle\cos mt, \sin nt\rangle = 0$. Therefore the trigonometric functions $\cos nt$, $\sin nt$ form an orthogonal set.

When $m = n$, we have

$$\langle\cos nt, \cos nt\rangle = \frac{1}{2\pi}\int_0^{2\pi}\cos^2 ntdt = \frac{1}{4\pi}\int_0^{2\pi}(1+\cos 2nt)dt = \frac{1}{2}.$$

Similarly, we have $\langle\sin nt, \sin nt\rangle = \frac{1}{2}$, and $\langle 1, 1\rangle = 1$. Assuming Proposition 5.3.3 still works for infinite sum, we get

$$\begin{aligned} a_0 &= \frac{\langle f(t), 1\rangle}{\langle 1, 1\rangle} = \langle f(t), 1\rangle = \frac{1}{2\pi}\int_0^{2\pi} f(t)dt, \\ a_n &= \frac{\langle f(t), \cos nt\rangle}{\langle\cos nt, \cos nt\rangle} = 2\langle f(t), \cos nt\rangle = \frac{1}{\pi}\int_0^{2\pi} f(t)\cos ntdt, \\ b_n &= \frac{\langle f(t), \sin nt\rangle}{\langle\sin nt, \sin nt\rangle} = 2\langle f(t), \sin nt\rangle = \frac{1}{\pi}\int_0^{2\pi} f(t)\sin ntdt. \end{aligned}$$

5.5.4 Isometry

Definition 5.5.6. A linear transformation $L\colon V \to W$ between inner product spaces is an *isometry* if it preserves the inner product

$$\langle L(\vec{x}), L(\vec{y})\rangle_W = \langle \vec{x}, \vec{y}\rangle_V \quad \text{for all } \vec{x}, \vec{y} \in V.$$

If the isometry is also invertible, then it is an *isometric isomorphism*.

Both sides of the equality $\langle L(\vec{x}), L(\vec{y})\rangle_W = \langle \vec{x}, \vec{y}\rangle_V$ are bilinear functions. By Proposition 5.1.2 (the argument works for general inner product spaces), this means that, if $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ is an orthonormal basis of V, then

$$\langle L(\vec{v}_i), L(\vec{v}_j)\rangle_W = \langle \vec{v}_i, \vec{v}_j\rangle_V = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}.$$

Therefore L is an isometry if it takes an orthonormal basis α of V to an orthonormal set $L(\alpha)$ of W. If we further know $\dim V = \dim W$, then $L(\alpha)$ is an orthonormal basis, and L is an isomorphism.

By the polarisation identity in Exercise 5.45, we also have the following.

Proposition 5.5.7. *A linear transformation between inner product spaces is an isometry if and only if it preserves the length: $\|L(\vec{x})\| = \|\vec{x}\|$.*

Example 5.5.12. By Example 5.5.6, we know $A = \begin{pmatrix} 1 & 2 \\ 2 & 8 \end{pmatrix}$ is a positive definite matrix. Therefore

$$\langle \vec{x}, \vec{y}\rangle_A = \vec{x}^T A \vec{y} = x_1y_1 + 2x_1y_2 + 2x_2y_1 + 8x_2y_2$$

is an inner product in $\mathbb{R}^2$. We get the positivity by completing the square

$$\langle \vec{x}, \vec{x}\rangle_A = x_1^2 + 4x_1x_2 + 8x_2^2 = (x_1 + 2x_2)^2 + (2x_2)^2.$$

This shows that

$$L(\vec{x}) = \begin{pmatrix} x_1 + 2x_2 \\ 2x_2 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 0 & 2 \end{pmatrix} \vec{x}$$

satisfies $\langle \vec{x}, \vec{x}\rangle = L(\vec{x}) \cdot L(\vec{x})$. By Proposition 5.5.7, this implies $L\colon (\mathbb{R}^2, \langle \vec{x}, \vec{y}\rangle_A) \to (\mathbb{R}^2, \vec{x} \cdot \vec{y})$ is an isometric isomorphism.

The standard basis $\{\vec{e}_1, \vec{e}_2\}$ is an orthonormal basis of $(\mathbb{R}^2, \vec{x} \cdot \vec{y})$. Then

$$\{L^{-1}(\vec{e}_1), L^{-1}(\vec{e}_2)\} = \{(1, 0), (-1, \tfrac{1}{2})\}$$

is an orthonormal basis of $(\mathbb{R}^2, \langle \vec{x}, \vec{y}\rangle_A)$.

Exercise 5.60. A positive definite matrix in Exercise 5.49 gives an inner product in the Euclidean space. Find an isometric isomorphism between the inner prodoct and the dot product.

Exercise 5.61. The matrix $B = \begin{pmatrix} 1 & 2 \\ 0 & 2 \end{pmatrix}$ obtained by completing the square in Example 5.5.12 satisfies $\langle \vec{x}, \vec{x} \rangle_A = \vec{x}^T A \vec{x} = B\vec{x} \cdot B\vec{x}$. What is the relation between A and B?

Example 5.5.13. We try to find an isometric isomorphism between $\langle \cdot, \cdot \rangle_{[-1,1]}$ in $C[-1, 1]$ and $\langle \cdot, \cdot \rangle_{[0,1]}$ in $C[0, 1]$. The two intervals are related by

$$t \in [0, 1] \mapsto 2t - 1 \in [-1, 1].$$

Then we have the corresponding relation between the two inner products

$$\begin{aligned} \langle f(t), g(t) \rangle_{[-1,1]} &= \int_{-1}^{1} f(t)g(t)dt = \int_{0}^{1} f(2t-1)g(2t-1)d(2t-1) \\ &= 2\int_{0}^{1} f(2t-1)g(2t-1)dt = \langle \sqrt{2}f(2t-1), \sqrt{2}g(2t-1) \rangle_{[0,1]}. \end{aligned}$$

This implies that

$$L(f(t)) = \sqrt{2}f(2t-1) \colon (C[-1, 1], \langle \cdot, \cdot \rangle_{[-1,1]}) \to (C[0, 1], \langle \cdot, \cdot \rangle_{[0,1]})$$

is an isometric isomorphism between the inner product spaces.

Exercise 5.62. Find an isometric isomorphism between $\langle \cdot, \cdot \rangle_{[0,1]}$ and $\langle \cdot, \cdot \rangle_{[0,2]}$.

Since the Gram-Schmidt precess works for general inner product space, we can modify any basis to an orthogonal basis. Therefore any inner product space has an orthonormal basis α. Then the α-coordinate map $[\cdot]_\alpha \colon V \to \mathbb{R}^n$ takes the orthonormal basis α to the standard basis ϵ of $\mathbb{R}^n$, which is orthonormal with respect to the dot product in $\mathbb{R}^n$. Therefore the α-coordinate map is an isometric isomorphism between the inner product space V and the dot product space $\mathbb{R}^n$.

Theorem 5.5.8. *Any finite dimensional inner product space is isometrically isomorphic to the Euclidean space with dot product.*

We have used the isomorphism between general (finite dimensional) vector space and Euclidean space to translate results from Euclidean space to general vector space. We may also use the isometric isomorphism to translate all the results from the dot product to general inner product.

The orthogonal complement of any subspace $H \subset V$ is

$$H^\perp = \{\vec{x} \colon \langle \vec{x}, \vec{h} \rangle = 0 \text{ for all } \vec{h} \in H\}.$$

We have

$$V = H \oplus H^{\perp}, \quad (H^{\perp})^{\perp} = H, \quad \vec{x} = \text{proj}_H\vec{x} + \text{proj}_{H^{\perp}}\vec{x}.$$

For mutually orthogonal subspaces $H_1, H_2, \dots, H_k$, we still have the orthogonal sum $H_1 \perp H_2 \perp \cdots \perp H_k$. Proposition 5.4.5 can be translated to inner product space

$$\text{proj}_{H_1 \perp H_2 \perp \cdots \perp H_k} = \text{proj}_{H_1} + \text{proj}_{H_2} + \cdots + \text{proj}_{H_k}.$$

Example 5.5.14. In Example 5.5.8, we calculate the the orthogonal projection onto the subspace H of $\mathbb{R}^3$ given by $x_1 + x_2 + x_3 = 0$, with respect to the inner product $\langle \vec{x}, \vec{y} \rangle = x_1y_1 + 2x_2y_2 + 3x_3y_3$. We may use the formula $\text{proj}_H\vec{x} = \vec{x} - \text{proj}_{H^{\perp}}\vec{x}$ to get another calculation.

First, the subspace H is characterised by

$$0 = x_1 + x_2 + x_3 = x_1 \cdot 1 + 2(x_2 \cdot \tfrac{1}{2}) + 3(x_3 \cdot \tfrac{1}{3}) = \langle (x_1, x_2, x_3), (1, \tfrac{1}{2}, \tfrac{1}{3}) \rangle.$$

Therefore the orthogonal complement

$$H^{\perp} = \mathbb{R}(1, \tfrac{1}{2}, \tfrac{1}{3}) = \mathbb{R}\vec{v}, \quad \vec{v} = (6, 3, 2).$$

Then

$$\begin{aligned} \text{proj}_H\vec{x} = \vec{x} - \frac{\langle \vec{x}, \vec{v} \rangle}{\langle \vec{v}, \vec{v} \rangle}\vec{v} &= \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} - \frac{6x_1 + 2 \cdot 3x_2 + 3 \cdot 2x_3}{36 + 2 \cdot 9 + 3 \cdot 4} \begin{pmatrix} 6 \\ 3 \\ 2 \end{pmatrix} \\ &= \frac{1}{11} \begin{pmatrix} 5 & -6 & -6 \\ -3 & 8 & -3 \\ -2 & -2 & 9 \end{pmatrix} \vec{x}. \end{aligned}$$

You may compare with the calculation in Example 5.5.8.

Exercise 5.63. Find the orthogonal projection onto the subspace given by $a_1x_2 + a_2x_2 + \cdots + a_nx_n = 0$, with respect to the inner product $\langle \vec{x}, \vec{y} \rangle = x_1y_1 + 2x_2y_2 + \cdots + nx_nx_n$.

Exercise 5.64. Find the orthogonal projection of P_2 onto the subspace given by $\int_0^1 tf(t)dt = 0$, with respect to the inner product $\langle f, g \rangle_{[0,1]} = \int_0^1 f(t)g(t)dt$.

Exercise 5.65. Find the orthogonal projection of P_2 onto the subspace given by $\int_0^1 (t^2 - 1)f(t)dt = 0$, with respect to the inner product $\langle f, g \rangle = \int_0^1 (t+1)f(t)g(t)dt$.

Exercise 5.66. Suppose α, β are two orthonormal bases of V. Prove that the matrix $[I]_{\beta\alpha}$ for changing the basis is an orthogonal matrix.

Exercise 5.67. Suppose α, β are orthonormal bases of V, W. Prove that $L \colon V \to W$ is an isometry if and only if $Q = [L]_{\beta\alpha}$ satisfies $Q^TQ = I$.

Exercise 5.68 (Parseval's Identity). Let $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ be an orthonormal basis of an inner product space. Use the formula for the α-coordinate (which is an isometric isomorphism to dot product) to prove that

$$\langle \vec{x}, \vec{y} \rangle = \langle \vec{x}, \vec{v}_1 \rangle \langle \vec{y}, \vec{v}_1 \rangle + \langle \vec{x}, \vec{v}_2 \rangle \langle \vec{y}, \vec{v}_2 \rangle + \cdots + \langle \vec{x}, \vec{v}_n \rangle \langle \vec{y}, \vec{v}_n \rangle.$$

In particular, we have

$$\|\vec{x}\|^2 = \langle \vec{x}, \vec{v}_1 \rangle^2 + \langle \vec{x}, \vec{v}_2 \rangle^2 + \cdots + \langle \vec{x}, \vec{v}_n \rangle^2.$$

Example 5.5.15. Consider P_3 with the inner product $\langle \cdot, \cdot \rangle_{[-1,1]}$ in Example 5.5.2. The inner product has the property that, if $f(t)g(t)$ is an odd function, then $\langle f, g \rangle = \int_{-1}^1 f(t)g(t)dt = 0$. Therefore the even polynomials and odd polynomials form orthogonally complement subspaces

$$P_3 = (\mathbb{R}1 + \mathbb{R}t^2) \perp (\mathbb{R}t + \mathbb{R}t^3).$$

We may use the Gram-Schmidt process to separately find orthogonal bases of the two subspaces, and then combine them to get an orthogonal basis of P_3.

Applying the Gram-Schmidt process to $1, t^2$, we get an orthogonal basis of even polynomial subspace $\mathbb{R}1 + \mathbb{R}t^2$

$$f_0(t) = 1, \quad f_2(t) = t^2 - \frac{\int_{-1}^1 t^2 \cdot 1dt}{\int_{-1}^1 1^2 dt} 1 = t^2 - \frac{\int_0^1 t^2 dt}{\int_0^1 dt} = t^2 - \frac{1}{3}.$$

Applying the Gram-Schmidt process to t, t^3, we get an orthogonal basis of odd polynomial subspace $\mathbb{R}t + \mathbb{R}t^3$

$$f_1(t) = t, \quad f_3(t) = t^3 - \frac{\int_{-1}^1 t^3 \cdot tdt}{\int_{-1}^1 t^2 dt} t = t^3 - \frac{\int_0^1 t^4 dt}{\int_0^1 t^2 dt} t = t^3 - \frac{3}{5}t.$$

We further calculate the lengths of the four vectors

$$\|f_0\|^2 = \int_{-1}^1 1^2 dt = 2, \qquad \|f_1\|^2 = \int_{-1}^1 t^2 dt = \frac{2}{3},$$
$$\|f_2\|^2 = \int_{-1}^1 \left(t^2 - \frac{1}{3}\right)^2 dt = \frac{8}{3^2 \cdot 5}, \qquad \|f_3\|^2 = \int_{-1}^1 \left(t^3 - \frac{3}{5}t\right)^2 dt = \frac{8}{5^2 \cdot 7}.$$

We divide the lengths and get an orthonormal basis

$$g_0(t) = \tfrac{1}{\sqrt{2}},\ g_1(t) = \tfrac{\sqrt{3}}{\sqrt{2}}t,\ g_2(t) = \tfrac{\sqrt{5}}{2\sqrt{2}}(3t^2 - 1),\ g_3(t) = \tfrac{\sqrt{7}}{2\sqrt{2}}(5t^3 - 3t).$$

The orthonormal basis means that the map

$$(x_0, x_1, x_2, x_3) \in \mathbb{R}^4 \mapsto \tfrac{1}{\sqrt{2}}x_0 + \tfrac{\sqrt{3}}{\sqrt{2}}x_1 t + \tfrac{\sqrt{5}}{2\sqrt{2}}x_2(3t^2 - 1) + \tfrac{\sqrt{7}}{2\sqrt{2}}x_3(5t^3 - 3t) \in P_3$$

is an isometric isomorphism between $\mathbb{R}^4$ with the dot product and P_3 with the inner product $\langle\cdot,\cdot\rangle_{[-1,1]}$.

Now we may use the isometric isomorphism L in Example 5.6.5 to translate the orthonormal basis $\{g_0, g_1, g_2, g_3\}$ for the inner product $\langle\cdot,\cdot\rangle_{[-1,1]}$ to an orthonormal basis for the inner product $\langle\cdot,\cdot\rangle_{[0,1]}$.

$$\begin{aligned}
L(g_0(t)) &= \sqrt{2}\tfrac{1}{\sqrt{2}} = 1,\\
L(g_1(t)) &= \sqrt{2}\tfrac{\sqrt{3}}{\sqrt{2}}(2t-1) = \sqrt{3}(2t-1),\\
L(g_2(t)) &= \sqrt{2}\tfrac{\sqrt{5}}{2\sqrt{2}}(3(2t-1)^2-1) = \sqrt{5}(6t^2-6t+1),\\
L(g_3(t)) &= \sqrt{2}\tfrac{\sqrt{7}}{2\sqrt{2}}(5(2t-1)^3-3(2t-1)) = \sqrt{7}(2t-1)(10t^2-10t+1).
\end{aligned}$$

The first three polynomials are the polynomials in Example 5.5.10 divided by their lengths. In the earlier example, it would be more complicated to calculate the fourth polynomial $(2t-1)(10t^2-10t+1)$ by further Gram-Schmidt process.

Exercise 5.69. Find an isometric isomorphism between $(P_2, \langle\cdot,\cdot\rangle_{[0,1]})$ and $\mathbb{R}^3$ with the dot product, by two ways.

1. Use the orthonormal basis from Example 5.5.10.

2. Use the calculation in Example 5.5.15.

Exercise 5.70. Find an isometric isomorphism between P_2 with the inner product $\langle f,g\rangle = \int_0^1 tf(t)g(t)dt$ and $\mathbb{R}^3$ with the dot product.

5.5.5 Adjoint

Let $L\colon V \to W$ be a linear transformation between inner product spaces. The *adjoint* of L is the linear transformation $L^*\colon W \to V$ defined by

$$\langle \vec{x}, L(\vec{y})\rangle_W = \langle L^*(\vec{x}), \vec{y}\rangle_V \quad \text{for all } \vec{x}\in W,\ \vec{y}\in V.$$

By the symmetry property of the inner products, this is the same as

$$\langle L(\vec{x}), \vec{y}\rangle_W = \langle \vec{x}, L^*(\vec{y})\rangle_V \quad \text{for all } \vec{x}\in V,\ \vec{y}\in W.$$

Consider the case $L(\vec{y}) = A\vec{y}\colon \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation between Euclidean spaces with dot products. We have

$$\vec{x}\cdot L(\vec{y}) = \vec{x}\cdot A\vec{y} = \vec{x}^T A\vec{y} = (A^T\vec{x})^T\vec{y} = A^T\vec{x}\cdot\vec{y}.$$

The right side should be $L^*(\vec{x})\cdot\vec{y}$. Therefore

$$[L^*] = A^T.$$

The equality already appeared in Section 5.1.2.

In general, we use orthonormal bases α and β to get isometric isomorphisms between inner product spaces V and W and Euclidean spaces with dot products. This translates the linear transformation L and its adjoint L^* to $L_{\beta\alpha}\colon \mathbb{R}^n \to \mathbb{R}^m$ and its adjoint $L^*_{\alpha\beta}\colon \mathbb{R}^m \to \mathbb{R}^n$, with respect to the dot products.

$$\begin{tikzcd}
(V, \langle , \rangle_V) \arrow[r, "L"] \arrow[d, "{[\cdot]_\alpha}"', "\cong"] & (W, \langle , \rangle_W) \arrow[d, "\cong"', "{[\cdot]_\beta}"] \\
(\mathbb{R}^m, \cdot) \arrow[r, "L_{\beta\alpha}"] & (\mathbb{R}^n, \cdot)
\end{tikzcd}, \qquad
\begin{tikzcd}
(V, \langle , \rangle_V) \arrow[d, "{[\cdot]_\alpha}"', "\cong"] & (W, \langle , \rangle_W) \arrow[l, "L^*"'] \arrow[d, "\cong"', "{[\cdot]_\beta}"] \\
(\mathbb{R}^m, \cdot) & (\mathbb{R}^n, \cdot) \arrow[l, "L^*_{\alpha\beta}"']
\end{tikzcd}$$

Then we get

$$[L^*]_{\alpha\beta} = [L]^*_{\beta\alpha}.$$

A linear transformation $L\colon V \to W$ induces four subspaces

$$\mathrm{Ran}L \subset W, \quad \mathrm{Ran}L^* \subset V, \quad \mathrm{Ker}L \subset V, \quad \mathrm{Ker}L^* \subset W.$$

These correspond respectively to $\mathrm{Col}A$, $\mathrm{Row}A = \mathrm{Col}A^T$, $\mathrm{Nul}A$, $\mathrm{Nul}A^T$. The orthogonal sum (or complement) relations in Section 5.4.1 correspond to

$$V = \mathrm{Ran}L^* \perp \mathrm{Ker}L, \quad W = \mathrm{Ran}L \perp \mathrm{Ker}L^*.$$

Moreover, Proposition 5.1.3 corresponds to

$$\mathrm{Ker}L^*L = \mathrm{Ker}L.$$

Example 5.5.16. Let V be the vector space of smooth periodic functions of period 1: $f(t+1) = f(t)$. The vector space has the inner product

$$\langle f, g \rangle_{[0,1]} = \int_0^1 f(t)g(t)dt.$$

Consider the derivative linear operator $L(f(t)) = f'(t)\colon V \to V$.

5.6 Complex Linear Algebra

5.6.1 Complex Number

A complex number is of the form $a + ib$, with $a, b \in \mathbb{R}$ and $i = \sqrt{-1}$ satisfying $i^2 = -1$. A complex number has the *real* and *imaginary* parts

$$\mathrm{Re}(a + ib) = a, \quad \mathrm{Im}(a + ib) = b.$$

The arithmetic operations of complex numbers are

$$\begin{aligned}
(a+ib)+(c+id) &= (a+c)+i(b+d), \\
(a+ib)-(c+id) &= (a-c)+i(b-d), \\
(a+ib)(c+id) &= (ac-bd)+i(ad+bc), \\
\frac{a+ib}{c+id} &= \frac{(a+ib)(c-id)}{(c+id)(c-id)} = \frac{(ac+bd)+i(-ad+bc)}{c^2+d^2}.
\end{aligned}$$

Example 5.6.1. To solve the system of linear equations

$$\begin{aligned}
(2+i)z_1 + (1-3i)z_2 &= -1-i, \\
(2-3i)z_1 + \quad (1+i)z_2 &= 7-i.
\end{aligned}$$

We eliminate z_1 by $(2-3i)\mathrm{Eq}_1 - (2+i)\mathrm{Eq}_2$

$$[(2-3i)(1-3i)-(2+i)(1+i)]z_2 = (2-3i)(-1-i)-(2+i)(7-i).$$

By

$$\begin{aligned}
(2-3i)(1-3i)-(2+i)(1+i) &= (-7-9i)-(1+3i) = -8-12i, \\
(2-3i)(-1-i)-(2+i)(7-i) &= (-5+i)-(15+5i) = -20-4i,
\end{aligned}$$

we get

$$z_2 = \frac{-20-4i}{-8-12i} = \frac{5+i}{2+3i} = \frac{(5+i)(2-3i)}{(2+3i)(2-3i)} = \frac{13-13i}{2^2+3^2} = 1-i.$$

Substituting into the first equation, we get

$$(2+i)z_1 = (-1-i)-(1-3i)(1-i) = (-1-i)-(-2-4i) = 1+3i.$$

Then

$$z_1 = \frac{1+3i}{2+i} = \frac{(1+3i)(2-i)}{(2+i)(2-i)} = \frac{5+5i}{2^2+1^2} = 1+i.$$

We may also solve the system by row operations on the augmented matrix

$$\begin{aligned}
\begin{pmatrix} 2+i & 1-3i & -1-i \\ 2-3i & 1+i & 7-i \end{pmatrix}
&\xrightarrow{\mathrm{Row}_1-\mathrm{Row}_2}
\begin{pmatrix} 4i & -4i & -8 \\ 2-3i & 1+i & 7-i \end{pmatrix} \\
&\xrightarrow{\frac{1}{4i}\mathrm{Row}_1}
\begin{pmatrix} 1 & -1 & 2i \\ 2-3i & 1+i & 7-i \end{pmatrix} \\
&\xrightarrow{\mathrm{Row}_2-(2-3i)\mathrm{Row}_1}
\begin{pmatrix} 1 & -1 & 2i \\ 0 & 3-2i & 1-5i \end{pmatrix} \\
&\xrightarrow{\frac{1}{3-2i}\mathrm{Row}_2}
\begin{pmatrix} 1 & -1 & 2i \\ 0 & 1 & 1-i \end{pmatrix} \\
&\xrightarrow{\mathrm{Row}_1+\mathrm{Row}_2}
\begin{pmatrix} 1 & 0 & 1+i \\ 0 & 1 & 1-i \end{pmatrix}.
\end{aligned}$$

The example shows that the content of Chapter 1, including row echelon forms, pivots, free variables, and rank, are valid over complex numbers.

Exercise 5.71. Find row echelon forms of the augmented matrices, and solve systems of linear equations.

1. $\begin{aligned} z_1 + iz_2 &= 2-3i, \\ iz_1 + z_2 &= 1. \end{aligned}$

2. $\begin{aligned} z_1 + iz_2 &= 2-3i, \\ iz_1 - z_2 &= 1. \end{aligned}$

3. $\begin{aligned} z_1 + iz_2 &= 2-3i, \\ iz_1 - z_2 &= 3+2i. \end{aligned}$

4. $\begin{aligned} z_1 + iz_2 &= 2-3i, \\ iz_1 + z_2 &= 1, \\ iz_1 + iz_2 &= 2-i. \end{aligned}$

5. $\begin{aligned} z_1 + iz_2 &= 2-3i, \\ iz_1 + z_2 &= 1, \\ z_1 - z_2 &= 3+i. \end{aligned}$

6. $\begin{aligned} z_1 + iz_2 + (1+i)z_3 &= 2-3i, \\ iz_1 + z_2 + (1-i)z_3 &= 1. \end{aligned}$

7. $\begin{aligned} (2+i)z_1 + (1-2i)z_2 &= 0, \\ (1-2i)z_1 + (2+i)z_2 &= 2+i. \end{aligned}$

8. $\begin{aligned} (2+i)z_1 + (1-2i)z_2 &= 3+4i, \\ (1-2i)z_1 + (2+i)z_2 &= 2+i. \end{aligned}$

Complex numbers $x+iy \in \mathbb{C}$ can be identified with points $(x,y) \in \mathbb{R}^2$ of the Euclidean plane. The corresponding real vector (x,y) has length r and angle θ (i.e., polar coordinate), and we have

$$x + iy = r\cos\theta + ir\sin\theta = re^{i\theta}.$$

The first equality is trigonometry, and the second equality uses the expansion (the theoretical explanation is the complex analytic continuation of the exponential function of real numbers)

$$\begin{aligned} e^{i\theta} &= 1 + \frac{1}{1!}i\theta + \frac{1}{2!}(i\theta)^2 + \cdots + \frac{1}{n!}(i\theta)^n + \cdots \\ &= \left(1 - \frac{1}{2!}\theta^2 + \frac{1}{4!}\theta^4 + \cdots\right) + i\left(\theta - \frac{1}{3!}\theta^3 + \frac{1}{5!}\theta^5 + \cdots\right) \\ &= \cos\theta + i\sin\theta. \end{aligned}$$

The complex exponential has the usual properties of the real exponential (because of complex analytic continuation), and we can easily get the multiplication and division of complex numbers in polar expressions

$$(re^{i\theta})(r'e^{i\theta'}) = rr'e^{i(\theta+\theta')}, \quad \frac{re^{i\theta}}{r'e^{i\theta'}} = \frac{r}{r'}e^{i(\theta-\theta')}.$$

The polar viewpoint easily shows that multiplying $re^{i\theta}$ means scaling by r and rotation by θ.

The *complex conjugation* $\overline{x+yi} = x - yi$ preserves the four arithmetic operations

$$\overline{z_1+z_2} = \bar{z}_1 + \bar{z}_2, \quad \overline{z_1-z_2} = \bar{z}_1 - \bar{z}_2, \quad \overline{z_1z_2} = \bar{z}_1\bar{z}_2, \quad \overline{\left(\frac{z_1}{z_2}\right)} = \frac{\bar{z}_1}{\bar{z}_2}.$$

Therefore the conjugation an *automorphism* (self-isomorphism) of $\mathbb{C}$. Geometrically, the conjugation means reflection with respect to the x-axis. This gives the conjugation in polar coordinates

$$\overline{re^{i\theta}} = re^{-i\theta}.$$

The length can also be expressed in terms of the conjugation

$$z\bar{z} = x^2 + y^2 = r^2, \quad |z| = r = \sqrt{z\bar{z}}.$$

This suggests that the *complex dot product* in $\mathbb{C}^n$ should be defined as

$$(z_1, z_2, \dots, z_n) \cdot (z_1', z_2', \dots, z_n') = z_1\overline{z_1'} + z_2\overline{z_2'} + \cdots + z_n\overline{z_n'}.$$

A major difference between $\mathbb{R}$ and $\mathbb{C}$ is that the polynomial $t^2 + 1$ has no root in $\mathbb{R}$ but has a pair of roots $\pm i$ in $\mathbb{C}$. In fact, complex numbers has the following so called *algebraically closed* property.

Theorem 5.6.1 (Fundamental Theorem of Algebra). *Any non-constant complex polynomial has roots.*

The real number $\mathbb{R}$ is not algebraically closed.

5.6.2 Complex Vector Space

By replacing $\mathbb{R}$ with $\mathbb{C}$ in the definition of vector spaces, we get the defintion of complex vector spaces.

Definition 5.6.2. A *(complex) vector space* is a set V, together with the operations of addition and (complex) scalar multiplication

$$\vec{x} + \vec{y} \colon V \times V \to V, \quad a\vec{x} \colon \mathbb{C} \times V \to V,$$

such that the following are satisfied:

1. Commutativity: $\vec{x} + \vec{y} = \vec{y} + \vec{x}$.
2. Associativity for addition: $(\vec{x} + \vec{y}) + \vec{z} = \vec{x} + (\vec{y} + \vec{z})$.
3. Zero: There is an element $\vec{0} \in V$ satisfying $\vec{x} + \vec{0} = \vec{x} = \vec{0} + \vec{x}$.
4. Negative: For any $\vec{x}$, there is $\vec{y}$ (to be denoted $-\vec{x}$), such that $\vec{x}+\vec{y} = \vec{0} = \vec{y}+\vec{x}$.
5. One: $1\vec{x} = \vec{x}$.
6. Associativity for scalar multiplication: $(ab)\vec{x} = a(b\vec{x})$.

7. Distributivity in the scalar: $(a+b)\vec{x} = a\vec{x} + b\vec{x}$.

8. Distributivity in the vector: $a(\vec{x}+\vec{y}) = a\vec{x} + a\vec{y}$.

All the usual concepts of real vector spaces have the corresponding complex version. Span, linear independence, basis still make sense. Properties about these concepts remain valid.

A complex vector space is also a real vector space, by restricting the scalars to $\mathbb{R}$. If $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ is a $\mathbb{C}$-basis of V, then $\{\vec{v}_1, i\vec{v}_1, \vec{v}_2, i\vec{v}_2, \dots, \vec{v}_n, i\vec{v}_n\}$ is an $\mathbb{R}$-basis of V. Therefore

$$\dim_{\mathbb{R}} V = 2\dim_{\mathbb{C}} V.$$

A subset $H \subset V$ is a *complex subspace*, if

$$\vec{x}, \vec{y} \in H \implies a\vec{x} + b\vec{y} \in H \text{ for all } a, b \in \mathbb{C}.$$

Complex subspaces are real subspaces. However, real subspaces (for example, odd dimensional real subspaces) may not be complex subspaces.

A map $L\colon V \to W$ of complex vector spaces is *complex linear* if

$$L(a\vec{x} + b\vec{y}) = aL(\vec{x}) + bL(\vec{y}).$$

It is *complex conjugate linear* if

$$L(a\vec{x} + b\vec{y}) = \bar{a}L(\vec{x}) + \bar{b}L(\vec{y}).$$

For both types of linear transformations, we have complex subspaces $\text{Ran}L \subset W$ and $\text{Ker}L \subset V$.

Let W be a real vector space. Consider the vector space

$$W^{\mathbb{C}} = W \oplus iW,$$

in which vectors are form

$$\vec{z} = \vec{x} + i\vec{y}, \quad \vec{x}, \vec{y} \in W.$$

The vectors can be added in the obvious way. We may also multiply complex numbers to the vectors in the obvious way ($a, b \in \mathbb{R}$)

$$(a+ib)(\vec{x} + i\vec{y}) = (a\vec{x} - b\vec{y}) + i(b\vec{x} + a\vec{y}).$$

This makes $W^{\mathbb{C}}$ into a complex vector space. We call $W^{\mathbb{C}} = W \oplus iW$ the *complexification* of W. The complexification of the real Euclidean space $\mathbb{R}^n$ is the complex Euclidean space $\mathbb{C}^n$.

In a complexification, we can take the *conjugation of vectors*

$$\vec{z} = \vec{x} + i\vec{y} \implies \bar{\vec{z}} = \overline{\vec{x} + i\vec{y}} = \vec{x} - i\vec{y}.$$

Conversely, a *conjugation operator* on a complex vector space V is a conjugate linear transformation $C\colon V \to V$ satisfying $C^2 = I$. Given a conjugation operator C, we introduce two real subspaces of V

$$\mathrm{Re}V = \{\vec{x} \in V : C(\vec{x}) = \vec{x}\}, \quad \mathrm{Im}V = \{\vec{x} \in V : C(\vec{x}) = -\vec{x}\}.$$

Any $\vec{z} \in V$ can be written as

$$\vec{z} = \vec{x} + \vec{y}, \quad \vec{x} = \tfrac{1}{2}(\vec{z} + C(\vec{z})) \in \mathrm{Re}V, \quad \vec{y} = \tfrac{1}{2}(\vec{z} - C(\vec{z})) \in \mathrm{Im}V.$$

It is easy to see $\mathrm{Re}V \cap \mathrm{Im}V = \{\vec{0}\}$. Therefore we get a direct sum $V = \mathrm{Re}V \oplus \mathrm{Im}V$ of $\mathbb{R}$-vector spaces. Moreover, it is also easy to see that $\vec{x} \mapsto i\vec{x}$ is an isomorphism of real vector spaces $W = \mathrm{Re}V$ and $iW = \mathrm{Im}V$. Then we get $V = W^{\mathbb{C}}$.

Proposition 5.6.3. *A complex vector space is a complexification of a real vector space if and only if it has a conjugation operator.*

A complex vector space V can be the complexification of many real subspaces W. The complexifications are in fact in one-to-one correspondence with the conjugation operators C on V.

Example 5.6.2. Let $C\colon \mathbb{C} \to \mathbb{C}$ be a conjugation operator. Then we have a fixed complex number $c = C(1)$, and $C(z) = C(z1) = \bar{z}C(1) = c\bar{z}$. By $z = C^2(z) = C(c\bar{z}) = c\overline{c\bar{z}} = |c|^2 z$, we get $|c| = 1$. Therefore conjugations on $\mathbb{C}$ are in one-to-one correspondence with points $c = e^{i\theta}$ on the unit circle.

For $C_\theta(z) = e^{i\theta}\bar{z}$, the real part is the real line of angle $\frac{\theta}{2}$

$$\begin{aligned}\mathrm{Re}_\theta\mathbb{C} &= \{z : e^{i\theta}\bar{z} = z\} = \{re^{\rho} : e^{i\theta}e^{-\rho} = e^{\rho},\ r \geq 0\} \\ &= \{re^{\rho} : 2\rho = \theta \bmod 2\pi,\ r \geq 0\} = \mathbb{R}e^{i\frac{\theta}{2}}.\end{aligned}$$

The imaginary part is the real line of angle $\frac{\theta+\pi}{2}$

$$\mathrm{Im}_\theta\mathbb{C} = i\mathbb{R}e^{i\frac{\theta}{2}} = e^{\frac{\pi}{2}}\mathbb{R}e^{i\frac{\theta}{2}} = \mathbb{R}e^{i\frac{\theta+\pi}{2}}.$$

The two parts are orthogonal.

Consider a complex linear transformation between complexifications of real vector spaces

$$L\colon W_1^{\mathbb{C}} = W_1 \oplus iW_1 \to W_2^{\mathbb{C}} = W_2 \oplus iW_2.$$

The restriction to W_1 is given by two $\mathbb{R}$-linear transformations $L_1, L_2\colon W_1 \to W_2$

$$L(\vec{x}) = L_1(\vec{x}) + iL_2(\vec{x}).$$

The whole L is determined by L_1 and L_2

$$\begin{aligned} L(\vec{x}+i\vec{y}) &= L(\vec{x})+iL(\vec{y}) \\ &= L_1(\vec{x})+iL_2(\vec{x})+i(L_1(\vec{y})+iL_2(\vec{y})) \\ &= L_1(\vec{x})-L_2(\vec{y})+i(L_2(\vec{x})+L_1(\vec{y})). \end{aligned}$$

This means that, under the obvious real isomorphisms $W_1^{\mathbb{C}} \cong W_1 \oplus W_1$ and $W_2^{\mathbb{C}} \cong W_2 \oplus W_2$, we have the block form of L

$$L = \begin{pmatrix} L_1 & -L_2 \\ L_2 & L_1 \end{pmatrix} : W_1 \oplus (i)W_1 \to W_2 \oplus (i)W_2.$$

5.6.3 Complex Inner Product

As explained near the end of Section 5.6.1, the *complex dot product* in the complex Euclidean space $\mathbb{C}^n$ is

$$(x_1, x_2, \dots, x_n)\cdot(y_1, y_2, \dots, y_n) = x_1\overline{y_1} + x_2\overline{y_2} + \cdots + x_n\overline{y_n}.$$

If we express vectors in vertical way, then

$$\vec{x}\cdot\vec{y} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = (x_1\ x_2\ \cdots\ x_n)\begin{pmatrix} \overline{y_1} \\ \overline{y_2} \\ \vdots \\ \overline{y_n} \end{pmatrix} = \vec{x}^T\overline{\vec{y}}.$$

The definition is consistent with the length of complex number $|z| = \sqrt{z\bar{z}}$ and satisfies

$$\vec{x}\cdot\vec{x} = |x_1|^2 + |x_2|^2 + \cdots + |x_n|^2 \ge 0.$$

However, if we take the scalar multiplication in the complex dot product, then we have

$$(a\vec{x})\cdot\vec{y} = a(\vec{x}\cdot\vec{y}), \quad \vec{x}\cdot(a\vec{y}) = \bar{a}(\vec{x}\cdot\vec{y}).$$

We also have the conjugate symmetry property

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

Exercise 5.72. Verify $\vec{x}\cdot(a\vec{y}) = \bar{a}(\vec{x}\cdot\vec{y})$ and $\vec{y}\cdot\vec{x} = \overline{\vec{x}\cdot\vec{y}}$ in $\mathbb{C}^n$.

The properties of the dot product are summarised in the following definition of general complex inner product.

Definition 5.6.4. An *inner product* in a complex vector space V is a function

$$\langle\vec{x},\vec{y}\rangle : V \times V \to \mathbb{C},$$

such that the following are satisfied:

1. Sesquilinearity: $\langle a\vec{x}+b\vec{y},\vec{z}\rangle = a\langle\vec{x},\vec{z}\rangle + b\langle\vec{y},\vec{z}\rangle$, $\langle\vec{x},a\vec{y}+b\vec{z}\rangle = \bar{a}\langle\vec{x},\vec{y}\rangle + \bar{a}\langle\vec{x},\vec{z}\rangle$.
2. Conjugate symmetry: $\langle\vec{y},\vec{x}\rangle = \overline{\langle\vec{x},\vec{y}\rangle}$.
3. Positivity: $\langle\vec{x},\vec{x}\rangle \geq 0$, and $\langle\vec{x},\vec{x}\rangle = 0$ if and only if $\vec{x}=\vec{0}$.

The sesquilinear property means linear in the first vector and conjugate linear in the second vector. In fact, under the conjugate symmetry property, linear in the first vector is equivalent to conjugate linear in the second vector.

Example 5.6.3. For complex valued continuous functions on $[a,b]$, the most common inner product is

$$\langle f,g\rangle_{[a,b]} = \int_a^b f(t)\overline{g(t)}dt.$$

The inner product restricts to become inner products on the space P_n of complex polynomials of degree $\leq n$.

If $\kappa(x)$ is a real valued function on $[a,b]$ that is positive except at finitely many places, then

$$\langle f,g\rangle = \int_a^b \kappa(t)f(t)\overline{g(t)}dt.$$

The function κ is the *kernel* of the inner product.

Example 5.6.4. In the vector space $M(m,n)$ of $m\times n$ complex matrices, we may use the trace (see Example 4.3.5) to define the inner product

$$\langle A,B\rangle = \mathrm{tr}A^T\bar{B} = \sum_{i,j} a_{ij}\overline{b_{ij}}, \quad A=(a_{ij}), B=(b_{ij}).$$

We still have the length of complex vectors

$$\|\vec{x}\| = \sqrt{\vec{x}\cdot\vec{x}}.$$

Since the value of the dot product may not be a real number, we can no longer define the angle between two complex vectors. Consequently, we do not have the area or volume in terms of the complex dot product. However, we may still define the complex orthogonality $\vec{x}\perp\vec{y}$ by $\vec{x}\cdot\vec{y}=0$.

Exercise 5.73 (Cauchy-Schwartz Inequality). Prove the Cauchy-Schwartz inequality

$$|\langle\vec{x},\vec{y}\rangle| \leq \|\vec{x}\|\|\vec{y}\|$$

by the method after Theorem 5.5.2. Actually you first get $|\mathrm{Re}\langle\vec{x},\vec{y}\rangle| \leq \|\vec{x}\|\|\vec{y}\|$. Then you get the Cauchy-Schwartz inequality by multiplying suitable $e^{i\theta}$ to $\vec{x}$.

Exercise 5.74. Prove the polarisation identities for complex inner product

$$\begin{aligned}
\mathrm{Re}(\langle\vec{x},\vec{y}\rangle) &= \frac{1}{2}(\|\vec{x}+\vec{y}\|^2-\|\vec{x}\|^2-\|\vec{y}\|^2) = \frac{1}{4}(\|\vec{x}+\vec{y}\|^2-\|\vec{x}-\vec{y}\|^2),\\
\mathrm{Im}(\langle\vec{x},\vec{y}\rangle) &= \frac{1}{2}(\|\vec{x}+i\vec{y}\|^2-\|\vec{x}\|^2-\|\vec{y}\|^2) = \frac{1}{4}(\|\vec{x}+i\vec{y}\|^2-\|\vec{x}-i\vec{y}\|^2)\\
\langle\vec{x},\vec{y}\rangle &= \frac{1}{4}(\|\vec{x}+\vec{y}\|^2-\|\vec{x}-\vec{y}\|^2+i\|\vec{x}+i\vec{y}\|^2-i\|\vec{x}-i\vec{y}\|^2).
\end{aligned}$$

Exercise 5.75. Prove the parallelogram identity for complex inner product

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

Exercise 5.76 (Pythagorean Theorem). Prove that

$$\|\vec{x}+\vec{y}\|^2 = \|\vec{x}\|^2+\|\vec{y}\|^2$$

if and only if $\mathrm{Re}\langle\vec{x},\vec{y}\rangle = 0$. In particular, if $\vec{x}\perp\vec{y}$, then we have the equality above.

Due to the sesquilinear property, $\vec{x}\cdot\vec{a} = b$ is a linear equation, and $\vec{a}\cdot\vec{x} = b$ is a conjugate linear (therefore not linear) equation. For example,

$$\begin{aligned}
(x,y)\cdot(-i,1+i) = 2+i &\quad\text{is}\quad ix+(1-i)y = 2+i,\\
(-i,1+i)\cdot(x,y) = 2+i &\quad\text{is}\quad -i\bar{x}+(1+i)\bar{y} = 2+i.
\end{aligned}$$

By taking the conjugation, the second equation is also the same as

$$ix+(1-i)y = 2-i.$$

This means $\vec{a}\cdot\vec{x} = b$ and $\vec{x}\cdot\vec{a} = \bar{b}$ are the same.

Therefore we need to pay attention to the order of two vectors in complex inner product. For example, for an orthogonal basis $\alpha = \{\vec{v}_1,\vec{v}_2,\dots,\vec{v}_n\}$ of a complex subspace $H\subset\mathbb{C}^m$, the orthogonal projection of $\vec{x}$ on H is

$$\mathrm{proj}_H\vec{x} = \frac{\langle\vec{x},\vec{v}_1\rangle}{\langle\vec{v}_1,\vec{v}_1\rangle}\vec{v}_1+\frac{\langle\vec{x},\vec{v}_2\rangle}{\langle\vec{v}_2,\vec{v}_2\rangle}\vec{v}_2+\cdots+\frac{\langle\vec{x},\vec{v}_n\rangle}{\langle\vec{v}_n,\vec{v}_n\rangle}\vec{v}_n.$$

Here we cannot use $\frac{\langle\vec{v}_i,\vec{x}\rangle}{\langle\vec{v}_i,\vec{v}_i\rangle}$ as the coefficient.

Exercise 5.77. Let $\vec{v}_1 = (1,i,1+i)$ and $\vec{v}_2 = (i,-1,a)$.

1. Find a, such that $\vec{v}_1\perp\vec{v}_2$.
2. Find $\vec{v}_3$, such that $\{\vec{v}_1,\vec{v}_2,\vec{v}_3\}$ is an orthogonal basis of $\mathbb{C}^3$.
3. Find the orthogonal projections to the subspaces $\mathbb{C}\vec{v}_1+\mathbb{C}\vec{v}_2$ and $\mathbb{C}\vec{v}_1+\mathbb{C}\vec{v}_3$.

Exercise 5.78. Let $\vec{v}_1 = (1,i,0)$ and $\vec{v}_2 = (1,0,i)$. Let $A = (\vec{v}_1\ \vec{v}_2)$ and $H = \mathrm{Col\ A} = \mathbb{C}\vec{v}_1+\mathbb{C}\vec{v}_2$.

1. Find an orthogonal basis of H.
2. Find $H^{\perp}$.
3. Find the orthogonal projection onto H.
4. Find the QR-decomposition of A.

Using the formula for the orthogonal projection, we have the Gram-Schmidt process to change any complex basis to a complex orthogonal basis. By further dividing the lengths, this implies any finite dimensional complex inner product space V has an orthonormal basis

$$\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}, \quad \langle \vec{v}_i, \vec{v}_j \rangle = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}.$$

Then $[\cdot]_\alpha \colon V \to \mathbb{C}^n$ is an isometric isomorphism of complex inner product spaces.

5.6.4 Complex Adjoint

The following shows moving a matrix from one side of complex dot product to the other side requires conjugate transpose

$$\vec{x} \cdot A\vec{y} = \vec{x}^T \overline{A\vec{y}} = \vec{x}^T \overline{A}\,\overline{\vec{y}} = (\overline{A}^T \vec{x})^T \overline{\vec{y}} = \overline{A}^T \vec{x} \cdot \vec{y}.$$

We denote the conjugate transpose matrix

$$A^* = \overline{A}^T,$$

and call it the *adjoint* of A. The terminology is due to the complex adjoint linear transformation similar to Section 5.5.5.

The *adjoint* of a linear transformation $L \colon V \to W$ be between complex inner product spaces is the linear transformation $L^* \colon W \to V$ defined by

$$\langle \vec{x}, L(\vec{y}) \rangle_W = \langle L^*(\vec{x}), \vec{y} \rangle_V \quad \text{for all } \vec{x} \in W,\ \vec{y} \in V.$$

By the conjugate symmetry property, this is the same as

$$\langle L(\vec{x}), \vec{y} \rangle_W = \langle \vec{x}, L^*(\vec{y}) \rangle_V \quad \text{for all } \vec{x} \in V,\ \vec{y} \in W.$$

With respect to dot product, we have

$$[L^*] = [L]^*.$$

In general, suppose α and β are orthonormal bases of V and W. Then we have isometric isomorphisms

$$[\cdot]_\alpha \colon V \to \mathbb{C}^n, \quad [\cdot]_\beta \colon W \to \mathbb{C}^m,$$

that translate L and its adjoint L^* to $L_{\beta\alpha} \colon \mathbb{C}^n \to \mathbb{C}^m$ and its adjoint $L^*_{\alpha\beta} \colon \mathbb{C}^m \to \mathbb{C}^n$. Therefore we get

$$[L^*]_{\alpha\beta} = [L]^*_{\beta\alpha}.$$

Exercise 5.79. Prove the properties of the adjoint of linear transformation

$$(L^*)^* = L, \quad (L+K)^* = L^* + K^*, \quad (aL)^* = \bar{a}L^*, \quad (L \circ K)^* = K^* \circ L^*.$$

What does this tell you about the adjoint matrix?

Exercise 5.80. Explain the the four basic subspaces for $L\colon V \to W$ are related by orthogonal complements

$$V = \mathrm{Ran}L^* \perp \mathrm{Ker}L, \quad W = \mathrm{Ran}L \perp \mathrm{Ker}L^*.$$

Similar to Section 5.5.5, a linear transformation $L\colon V \to W$ is an isometry

$$\langle L(\vec{x}), L(\vec{y})\rangle_W = \langle \vec{x}, \vec{y}\rangle_V \quad \text{for all } \vec{x}, \vec{y} \in V,$$

if and only if $L^*L = I$. In case $\dim V = \dim W$, this implies L is invertible, $L^{-1} = L^*$, and $LL^* = I$.

A matrix U gives an isometry if and only if $U^*U = I$. This is equivalent to that the columns of U form an orthonormal set. The complex version of the Gram-Schmidt process gives the complex QR-decomposition.

If U is a square matrix, then U is invertible, with $U^{-1} = U^*$. We call U a *unitary matrix*. This is the complex counterpart of the orthogonal matrix.

5.6.5 Hermitian Matrix

An inner product is sesquilinear, conjugate symmetric, and positive definite. Similar to the discussion in Section 5.1.2, for a complex sesquilinear function $f(\vec{x}, \vec{y})\colon V \times W \to \mathbb{C}$, we assume bases $\alpha = \{\vec{v}_1, \vec{v}_2\}$ and $\beta = \{\vec{w}_1, \vec{w}_2, \vec{w}_3\}$ of V and W (say complex dimensions of 2 and 3). Then for

$$\vec{x} = x_1\vec{v}_1 + x_2\vec{v}_2, \quad [\vec{x}]_\alpha = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}; \quad \vec{y} = y_1\vec{w}_1 + y_2\vec{w}_2 + y_3\vec{w}_3, \quad [\vec{y}]_\beta = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix},$$

and $a_{ij} = f(\vec{v}_i, \vec{w}_j)$, we have

$$\begin{aligned} f(\vec{x}, \vec{y}) &= f(x_1\vec{v}_1 + x_2\vec{v}_2, y_1\vec{w}_1 + y_2\vec{w}_2 + y_3\vec{w}_3) \\ &= x_1\overline{y_1}f(\vec{v}_1, \vec{w}_1) + x_1\overline{y_2}f(\vec{v}_1, \vec{w}_2) + x_1\overline{y_3}f(\vec{v}_1, \vec{w}_3) \\ &+ x_2\overline{y_1}f(\vec{v}_2, \vec{w}_1) + x_2\overline{y_2}f(\vec{v}_2, \vec{w}_2) + x_2\overline{y_3}f(\vec{v}_2, \vec{w}_3) \\ &= (x_1 \; x_2) \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} \overline{y_1} \\ \overline{y_2} \\ \overline{y_3} \end{pmatrix} = [\vec{x}]_\alpha^T A\overline{[\vec{y}]_\beta} = [\vec{x}]_\alpha \cdot \overline{A}[\vec{y}]_\beta. \end{aligned}$$

Let $L\colon W \to V$ be the linear transformation with $[L]_{\beta\alpha} = \bar{A}$, then the calculation shows

$$f(\vec{x}, \vec{y}) = \langle \vec{x}, L(\vec{y})\rangle_V = \langle L^*(\vec{x}), \vec{y}\rangle_W.$$

Conversely, for any linear transformation L, $\langle \vec{x}, L(\vec{y})\rangle$ is a sesquilinear function. Therefore there is a one-to-one correspondence between sesquilinear functions on $V \times W$ and linear transformations $W \to V$.

For f to be inner product on V, we require $V = W$. Then a sesquilinear function on $V \times V$ is $\langle \vec{x}, L(\vec{y})\rangle$ for a linear operator L on V. For the function to be conjugate symmetric, we require

$$f(\vec{x},\vec{y}) = \langle \vec{x}, L(\vec{y})\rangle = \overline{f(\vec{y},\vec{x})} = \overline{\langle \vec{y}, L(\vec{x})\rangle} = \langle L(\vec{x}), \vec{y}\rangle = \langle \vec{x}, L^*(\vec{y})\rangle.$$

Therefore the conjugate symmetric property means exactly $L =^* = L$. We call such L a *Hermitian operator*. The corresponding concept is Hermitian matrix.

Definition 5.6.5. A complex square matrix A is *Hermitian* if $A = A^*$. It is *positive definite* if $\vec{x} \cdot A\vec{x} > 0$ for all $\vec{x} \neq \vec{0}$. The matrix is *negative definite* if $\vec{x} \cdot A\vec{x} < 0$ for all $\vec{x} \neq \vec{0}$. It is *indefinite* if $\vec{x} \cdot A\vec{x} > 0$ for some $\vec{x}$ and $\vec{x} \cdot A\vec{x} < 0$ for some other $\vec{y}$.

An inner product on $\mathbb{C}^n$ is

$$\langle \vec{x}, \vec{y}\rangle_A = \vec{x} \cdot A\vec{y}$$

for a positive Hermitian matrix A. There is similar definition for Hermitian operators, and an inner product on a complex vector space corresponds to a positive Hermitian operator.

We note that the diagonal entries in a Hemitian matrix are real numbers.

Example 5.6.5. For $A = \begin{pmatrix} 1 & 1+i \\ 1-i & a \end{pmatrix}$, we have

$$\begin{aligned}
\vec{x}^T A\overline{\vec{x}} &= x_1\overline{x}_1 + (1+i)x_1\overline{x_2} + (1-i)x_2\overline{x_1} + ax_2\overline{x_2} \\
&= x_1\overline{x_1} + x_1\overline{(1-i)x_2} + (1-i)x_2\overline{x_1} + ax_2\overline{x_2} \\
&= (x_1 + (1-i)x_2)\overline{(x_1 + (1-i)x_2)} + (a - |1-i|^2)x_2\overline{x_2} \\
&= |x_1 + (1-i)x_2|^2 + (a-2)|x_2|^2.
\end{aligned}$$

The right side is positive for all $(x_1, x_2) \neq (0,0)$ if and only if $a > 2$. This is the condition for A to be positive definite.

The example shows that the process of completing the square can also be applied to Hermitian symmetric matrices. The process in Example 5.6.5 is the same as the following. We first use row operations to eliminate the entries below the diagonal

$$\begin{pmatrix} 1 & 1+i \\ 1-i & a \end{pmatrix} \xrightarrow{\text{Row}_2 - (1-i)\text{Row}_1} \begin{pmatrix} 1 & 1+i \\ 0 & a-2 \end{pmatrix}.$$

Then we carry out the conjugate version of the same column operation

$$\begin{pmatrix} 1 & 1+i \\ 0 & a-2 \end{pmatrix} \xrightarrow{\text{Col}_2 - \overline{(1-i)}\text{Col}_1} \begin{pmatrix} 1 & 0 \\ 0 & a-2 \end{pmatrix}.$$

Exercise 5.81. For a linear operator L on an inner product space V, prove the equality

$$\langle L(\vec{x}),\vec{y}\rangle + \langle L(\vec{y}),\vec{x}\rangle = \langle L(\vec{x}+\vec{y}),\vec{x}+\vec{y}\rangle - \langle L(\vec{x}),\vec{x}\rangle - \langle L(\vec{y}),\vec{y}\rangle.$$

Then prove that the following are equivalent

1. $\langle L(\vec{y}),\vec{y}\rangle$ is real for all $\vec{y}$.
2. $\langle L(\vec{x}),\vec{y}\rangle + \langle L(\vec{y}),\vec{x}\rangle$ is real for all $\vec{x},\vec{y}$.
3. $L = L^*$.

Exercise 5.82. Prove that $\langle L(\vec{v}),\vec{v}\rangle$ is imaginary for all $\vec{v}$ if and only if $L^* = -L$.

Exercise 5.83. Suppose A and B are $n\times n$ and $m\times m$ positive definite Hermitian matrices. For a linear transformation $L(\vec{x}) = M\vec{x}\colon (\mathbb{C}^n,\langle\cdot,\cdot\rangle_A) \to (\mathbb{C}^m,\langle\cdot,\cdot\rangle_B)$ given by matrix M, find the matrix for the adjoint L^*.

Exercise 5.84. Find the condition for a Hermitian matrix $\begin{pmatrix} a & \bar{b} \\ b & c \end{pmatrix}$ to be positive definite.

Exercise 5.85. Find condition on a, such that the Hermitian matrix $\begin{pmatrix} 1 & 1+i & i \\ 1-i & 4 & 1 \\ -i & 1 & 2a \end{pmatrix}$ is positive definite.

Chapter 6

Determinant

The determinant was originally invented for solving systems of linear equations. For two equations in two variables

$$\begin{aligned} a_1x + b_1y &= c_1, \\ a_2x + b_2y &= c_2, \end{aligned}$$

the solution is (provided the denominator is not zero)

$$x = \frac{b_2c_1 - b_1c_2}{a_1b_2 - a_2b_1}, \quad y = \frac{a_1c_2 - a_2c_1}{a_1b_2 - a_2b_1}.$$

The denominator and numerators are determinants of 2×2 matrices. The formulae are called *Cramer's rule*. This is a very inefficient method compared with the row operations. Moreover, the formulae hold only when the coefficient matrix is invertible.

The modern way of understanding the determinant has two aspects. The algebraic aspect is the multilinear and alternating properties, which can be further extended to exterior algebra. The geometric aspect is the signed volume. You need to know both aspects in order to have a complete understanding of the determinant.

On important property of the determinant is the relation to the invertibility of the matrix. For a system of linear equations, this means the system has unique solution for all the right sides. This is exactly the case Cramer's rule is valid. The property is used in calculating eigenvalues in Chapter 7.

6.1 Algebra of Determinant

6.1.1 Multilinear and Alternating Function

Definition 6.1.1. The determinant of $n \times n$ matrix $A = (\vec{x}_1 \; \vec{x}_2 \; \cdots \; \vec{x}_n)$ is a function $\det A$ satisfying the following properties

1. Multilinear: Linear in each column vector
$$\det(\dots, a\vec{x}_{1i} + b\vec{x}_{2i}, \dots) = a\det(\dots, \vec{x}_{1i}, \dots) + b\det(\dots, \vec{x}_{2i}, \dots).$$

2. Alternating: Switching two columns changes sign
$$\det(\dots, \vec{x}_j, \dots, \vec{x}_i, \dots) = -\det(\dots, \vec{x}_i, \dots, \vec{x}_j, \dots).$$

3. Normal: $\det I = \det(\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n) = 1$.

If two columns are equal, then the alternating property implies
$$\det(\dots, \vec{x}, \dots, \vec{x}, \dots) = -\det(\dots, \vec{x}, \dots, \vec{x}, \dots).$$
Therefore
$$\det(\dots, \vec{x}, \dots, \vec{x}, \dots) = 0. \tag{6.1.1}$$

The determinant of a 1×1 matrix is $\det(x) = x$.

For a 2×2 matrix, we let $\vec{e}_1 = (1, 0)$, $\vec{e}_2 = (0, 1)$ be the standard basis. Then
$$\begin{aligned}
\det\begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} &= \det(x_{11}\vec{e}_1 + x_{21}\vec{e}_2, x_{12}\vec{e}_1 + x_{22}\vec{e}_2) \\
&= x_{11}\det(\vec{e}_1, x_{12}\vec{e}_1 + x_{22}\vec{e}_2) + x_{21}\det(\vec{e}_2, x_{12}\vec{e}_1 + x_{22}\vec{e}_2) \\
&= x_{11}x_{12}\det(\vec{e}_1, \vec{e}_1) + x_{11}x_{22}\det(\vec{e}_1, \vec{e}_2) \\
&\quad + x_{21}x_{12}\det(\vec{e}_2, \vec{e}_1) + x_{21}x_{22}\det(\vec{e}_2, \vec{e}_2) \\
&= x_{11}x_{12}0 + x_{11}x_{22}\det(\vec{e}_1, \vec{e}_2) - x_{21}x_{12}\det(\vec{e}_1, \vec{e}_2) + x_{21}x_{22}0 \\
&= (x_{11}x_{22} - x_{21}x_{12})\det(\vec{e}_1, \vec{e}_2) = x_{11}x_{22} - x_{21}x_{12}.
\end{aligned}$$
The second equality uses the linearity in the first column. The third equality uses the linearity in the second column. The fourth equality uses the alternating property and (6.1.1). Only the final property uses the normal property $\det(\vec{e}_1, \vec{e}_2) = 1$.

For a 3×3 matrix, we have
$$\begin{aligned}
&\det\begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix} \\
&= \det(x_{11}\vec{e}_1 + x_{21}\vec{e}_2 + x_{31}\vec{e}_3, x_{12}\vec{e}_1 + x_{22}\vec{e}_2 + x_{32}\vec{e}_3, x_{13}\vec{e}_1 + x_{23}\vec{e}_2 + x_{33}\vec{e}_3) \\
&= x_{11}x_{22}x_{33}\det(\vec{e}_1, \vec{e}_2, \vec{e}_3) + x_{11}x_{32}x_{23}\det(\vec{e}_1, \vec{e}_3, \vec{e}_2) \\
&+ x_{21}x_{12}x_{33}\det(\vec{e}_2, \vec{e}_1, \vec{e}_3) + x_{21}x_{32}x_{13}\det(\vec{e}_2, \vec{e}_3, \vec{e}_1) \\
&+ x_{31}x_{12}x_{23}\det(\vec{e}_3, \vec{e}_1, \vec{e}_2) + x_{31}x_{22}x_{13}\det(\vec{e}_3, \vec{e}_2, \vec{e}_1) \\
&= (x_{11}x_{22}x_{33} - x_{11}x_{32}x_{23} - x_{21}x_{12}x_{33} \\
&+ x_{21}x_{32}x_{13} + x_{31}x_{12}x_{23} - x_{31}x_{22}x_{13})\det(\vec{e}_1, \vec{e}_2, \vec{e}_3) \\
&= x_{11}x_{22}x_{33} - x_{11}x_{32}x_{23} - x_{21}x_{12}x_{33} + x_{21}x_{32}x_{13} + x_{31}x_{12}x_{23} - x_{31}x_{22}x_{13}.
\end{aligned}$$

In the second equality, we use (6.1.1) to get $\det(\vec{e}_1, \vec{e}_1, \vec{e}_2) = \det(\vec{e}_2, \vec{e}_2, \vec{e}_3) = 0$, and so on. In the third equality, we use the alternating property to get

$$\det(\vec{e}_3, \vec{e}_1, \vec{e}_2) = -\det(\vec{e}_1, \vec{e}_3, \vec{e}_2) = \det(\vec{e}_1, \vec{e}_2, \vec{e}_3),$$

and so on. Only the last equality uses the normal property $\det(\vec{e}_1, \vec{e}_2, \vec{e}_3) = 1$.

The calculation shows that, in general, if a function $D(A)$ satisfies the multilinear and alternating properties, then

$$D\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{pmatrix} = D(I) \sum \operatorname{sign}(i_1, i_2, \dots, i_n) x_{i_1 1} x_{i_2 2} \dots x_{i_n n}.$$

Here $i_1, i_2, \dots, i_n$ are distinct because otherwise $D(\vec{e}_{i_1}, \vec{e}_{i_2}, \dots, \vec{e}_{i_n}) = 0$ by (6.1.1). This means $(i_1, i_2, \dots, i_n)$ is a *permutation* of $(1, 2, \dots, n)$. Then the sign comes from

$$D(\vec{e}_{i_1}, \vec{e}_{i_2}, \dots, \vec{e}_{i_n}) = \pm D(\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n) = \pm D(I).$$

By the alternating property, any exchange of two terms in $(i_1, i_2, \dots, i_n)$ introduces a negative sign. Using several exchanges, we may change $(i_1, i_2, \dots, i_n)$ to $(1, 2, \dots, n)$. The sign is determined by the number of exchanges in the process

$$\operatorname{sign}(i_1, i_2, \dots, i_n) = \begin{cases} 1, & \text{even number of exchanges} \\ -1, & \text{odd number of exchanges} \end{cases}.$$

For example, the exchanges

$$\begin{aligned} &(2,4,1,5,3) \to (2,1,4,5,3) \to (1,2,4,5,3) \to (1,2,4,3,5) \to (1,2,3,4,5) \\ &(4,3,2,5,1) \to (1,3,2,5,4) \to (1,3,2,4,5) \to (1,2,3,4,5) \end{aligned}$$

imply $\operatorname{sign}(2,4,1,5,3) = 1$ and $\operatorname{sign}(4,3,2,5,1) = -1$.

Exercise 6.1. Find the signs of the permutations.

1. $(3,4,1,2)$.
2. $(4,1,3,2)$.
3. $(3,2,5,1,4)$.
4. $(2,3,5,1,6,4)$.

Exercise 6.2. Find the signs of the permutations $(n, n-1, \dots, 2, 1)$ and $(k+1, \dots, n, 1, 2, \dots, k)$.

Exercise 6.3. Find the determinants.

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

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

Exercise 6.4. For a 3×3 matrix A, How are $\det(-5A)$ and $\det A$ related?

Exercise 6.5. Is it true that $\det(A+B) = \det A + \det B$? Try with 2×2 matrices.

Since the determinant satisfies the additional normal condition $\det I = 1$, we get the explicit formula for the determinant

$$\det \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{pmatrix} = \sum \operatorname{sign}(i_1, i_2, \dots, i_n) x_{i_1 1} x_{i_2 2} \dots x_{i_n n}.$$

The earlier calculation means the following.

Lemma 6.1.2. *If $D(A)$ be a multilinear and alternating function of the columns of $n \times n$ matrix A, then $D(A) = c \det A$, where $c = D(I)$ is a constant.*

Example 6.1.1. The determinant of 4×4 matrix has 24 terms

$$\begin{aligned} &\det \begin{pmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ x_{31} & x_{32} & x_{33} & x_{34} \\ x_{41} & x_{42} & x_{43} & x_{44} \end{pmatrix} \\ &= x_{11}x_{22}x_{33}x_{44} + x_{21}x_{32}x_{43}x_{14} + x_{31}x_{42}x_{13}x_{24} + x_{41}x_{12}x_{23}x_{34} \\ &+ x_{11}x_{32}x_{43}x_{24} + x_{31}x_{42}x_{23}x_{14} + x_{41}x_{22}x_{13}x_{34} + x_{21}x_{12}x_{33}x_{44} \\ &+ x_{11}x_{42}x_{23}x_{34} + x_{41}x_{22}x_{33}x_{14} + x_{21}x_{32}x_{13}x_{44} + x_{31}x_{12}x_{43}x_{24} \\ &- x_{11}x_{22}x_{43}x_{34} - x_{21}x_{42}x_{33}x_{14} - x_{41}x_{32}x_{13}x_{24} - x_{31}x_{12}x_{23}x_{44} \\ &- x_{11}x_{42}x_{33}x_{24} - x_{41}x_{32}x_{23}x_{14} - x_{31}x_{22}x_{13}x_{44} - x_{21}x_{12}x_{43}x_{34} \\ &- x_{11}x_{32}x_{23}x_{44} - x_{31}x_{22}x_{43}x_{14} - x_{21}x_{42}x_{13}x_{34} - x_{41}x_{12}x_{33}x_{24}. \end{aligned}$$

6.1.2 Properties of Determinant

Theorem 6.1.3. $\det AB = \det A \det B$.

We fix A and consider $D(B) = \det AB$. We express B in column vectors (assuming B has three columns, for example) $B = (\vec{x}_1\ \vec{x}_2\ \vec{x}_3)$. Then $D(\vec{x}_1, \vec{x}_2, \vec{x}_3) = \det(A\vec{x}_1, A\vec{x}_2, A\vec{x}_3)$.

The following shows D is linear in the first column of B

$$\begin{aligned}D(a\vec{x}_1+b\vec{x}_1',\vec{x}_2,\vec{x}_3)&=\det(A(a\vec{x}_1+b\vec{x}_1'),A\vec{x}_2,A\vec{x}_3)\\&=\det(aA\vec{x}_1+bA\vec{x}_1',A\vec{x}_2,A\vec{x}_3)\\&=a\det(A\vec{x}_1,A\vec{x}_2,A\vec{x}_3)+b\det(A\vec{x}_1',A\vec{x}_2,A\vec{x}_3)\\&=aD(\vec{x}_1,\vec{x}_2,\vec{x}_3)+bD(\vec{x}_1',\vec{x}_2,\vec{x}_3).\end{aligned}$$

Here the third equality uses the linearity of the determinant in the first column. By similar argument, we know D is linear in other columns of B.

The following shows D is alternating for the exchange of the first and second columns of B

$$D(\vec{x}_2,\vec{x}_1,\vec{x}_3)=\det(A\vec{x}_2,A\vec{x}_1,A\vec{x}_3)=-\det(A\vec{x}_1,A\vec{x}_2,A\vec{x}_3)=-D(\vec{x}_1,\vec{x}_2,\vec{x}_3).$$

Here the second equality uses the alternating property of the determinant. By similar argument, we know D is alternating for the exchanges of the other columns of B.

By Lemma 6.1.2, we get $\det AB=D(B)=D(I)\det B$. Since $D(I)=\det AI=\det A$, we get $\det AB=\det A\det B$.

The following is the change of the determinant under the column operations.

Proposition 6.1.4. *For square matrix $A=(\vec{x}_1\ \vec{x}_2\ \cdots\ \vec{x}_n)$, the determinant has the following properties.*

1. $\det(\vec{x}_2,\vec{x}_1,\vec{x}_3,\dots)=-\det A$.

2. $\det(c\vec{x}_1,\vec{x}_2,\vec{x}_3,\dots)=c\det A$.

3. $\det(\vec{x}_1+c\vec{x}_2,\vec{x}_2,\vec{x}_3,\dots)=\det A$.

The first is the alternating property

$$\det(\vec{x}_2,\vec{x}_1,\vec{x}_3,\dots)=-\det(\vec{x}_1,\vec{x}_2,\vec{x}_3,\dots).$$

The second follows from the linearity in the first column

$$\det(c\vec{x}_1,\vec{x}_2,\vec{x}_3,\dots)=c\det(\vec{x}_1,\vec{x}_2,\vec{x}_3,\dots).$$

The third is a consequence of linear and alternating properties

$$\begin{aligned}\det(\vec{x}_1+c\vec{x}_2,\vec{x}_2,\vec{x}_3,\dots)&=\det(\vec{x}_1,\vec{x}_2,\vec{x}_3,\dots)+c\det(\vec{x}_2,\vec{x}_2,\vec{x}_3,\dots)\\&=\det(\vec{x}_1,\vec{x}_2,\vec{x}_3,\dots)+c0=\det(\vec{x}_1,\vec{x}_2,\vec{x}_3,\dots).\end{aligned}$$

Proposition 6.1.5. *For row operations, the determinant has the same properties as the column operations.*

By Example 3.2.3, a row operation on A gives EA, where E is the elementary matrix for the row operation. Then by Theorem 6.1.3, we get $\det EA = \det E \det A$. Then the proposition means $\det E = -1, c, 1$ for the three types of row operations. We show the values by applying the column operations to E and using Proposition 6.1.4. We use 3×3 matrix to illustrate this.

For the exchange $\mathrm{Row}_1 \leftrightarrow \mathrm{Row}_2$, we have

$$\det\begin{pmatrix}0&1&0\\1&0&0\\0&0&1\end{pmatrix} \xlongequal{\mathrm{Col}_1\leftrightarrow\mathrm{Col}_2} -\det\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix} = -1.$$

In general, we find the determinant of the elementary matrix for $\mathrm{Row}_i \leftrightarrow \mathrm{Row}_j$ is -1, by using $\mathrm{Col}_i \leftrightarrow \mathrm{Col}_j$.

For the scaling $c\mathrm{Row}_1$, we have

$$\det\begin{pmatrix}c&0&0\\0&1&0\\0&0&1\end{pmatrix} \xlongequal{c^{-1}\mathrm{Col}_1} c\det\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix} = c.$$

In general, we find the determinant of the elementary matrix for $c\mathrm{Row}_i$ is c, by using $c^{-1}\mathrm{Col}_i$.

For the operation $\mathrm{Row}_1 + c\mathrm{Row}_2$, we have

$$\det\begin{pmatrix}1&c&0\\0&1&0\\0&0&1\end{pmatrix} \xlongequal{\mathrm{Col}_2-c\mathrm{Col}_1} \det\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix} = 1.$$

In general, we find the determinant of the elementary matrix for $\mathrm{Row}_i + c\mathrm{Row}_j$ is 1, by using $\mathrm{Col}_j - c\mathrm{Col}_i$.

Theorem 6.1.6. *If A and B are square matrices, then*

$$\det\begin{pmatrix}A&X\\O&B\end{pmatrix} = \det A\det B = \det\begin{pmatrix}A&O\\X&B\end{pmatrix}.$$

We fix B, X, and consider the function $D(A) = \det\begin{pmatrix}A&X\\O&B\end{pmatrix}$. A linear combination within a column of A corresponds to the same linear combination within the same column of $\begin{pmatrix}A&X\\O&B\end{pmatrix}$. The exchange of two columns of A corresponds to the exchange of the same two columns of $\begin{pmatrix}A&X\\O&B\end{pmatrix}$. Therefore D is multilinear and alternating in the columns of A. By Lemma 6.1.2, we get

$$\det\begin{pmatrix}A&X\\O&B\end{pmatrix} = D(A) = D(I)\det A = \det A\det\begin{pmatrix}I&X\\O&B\end{pmatrix}.$$

Using the third type column operations, we may use I to eliminate X

$$\begin{pmatrix} I & X \\ O & B \end{pmatrix} \to \begin{pmatrix} I & O \\ O & B \end{pmatrix}.$$

Then by Proposition 6.1.4, we get

$$\det \begin{pmatrix} I & X \\ O & B \end{pmatrix} = \det \begin{pmatrix} I & O \\ O & B \end{pmatrix}, \quad \det \begin{pmatrix} A & X \\ O & B \end{pmatrix} = \det A \det \begin{pmatrix} I & O \\ O & B \end{pmatrix}.$$

Now we may repeat the multilinear and alternating argument for the columns of B, and get

$$\det \begin{pmatrix} I & O \\ O & B \end{pmatrix} = \det \begin{pmatrix} I & O \\ O & I \end{pmatrix} \det B = \det B.$$

Combining all the equalities, we get $\det \begin{pmatrix} A & X \\ O & B \end{pmatrix} = \det A \det B$.

The argument for $\det \begin{pmatrix} A & O \\ X & B \end{pmatrix} = \det A \det B$ is similar. We first do the multilinear and alternating argument for the columns of B, followed by the similar argument for A.

We may easily extend Theorem 6.1.6 to 3×3 block matrices, assuming A, B, C are square matrices

$$\det \begin{pmatrix} A & X & Y \\ O & B & Z \\ O & O & C \end{pmatrix} = \det \begin{pmatrix} A & X \\ O & B \end{pmatrix} \det C = \det A \det B \det C,$$
$$\det \begin{pmatrix} A & O & Y \\ X & B & Z \\ O & O & C \end{pmatrix} = \det \begin{pmatrix} A & O \\ X & B \end{pmatrix} \det C = \det A \det B \det C.$$

Of course the property can be further extended to larger block matrices. Eventually, we get the determinant of triangular matrices.

Proposition 6.1.7.

$$\det \begin{pmatrix} x_1 & * & \cdots & * \\ 0 & x_2 & \cdots & * \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & x_n \end{pmatrix} = x_1 x_2 \cdots x_n = \det \begin{pmatrix} x_1 & 0 & \cdots & 0 \\ * & x_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ * & * & \cdots & x_n \end{pmatrix}.$$

Propositions 6.1.4 and 6.1.5 suggest the following, similar to Theorem 2.4.7.

Proposition 6.1.8. $\det A^T = \det A$.

A consequence of the proposition is that $\det A$ is also multilinear and alternating in rows of A.

If A is invertible, then we may use row operations to change A to the reduced row echelon form I

$$A \xrightarrow{\text{row op}} I.$$

Taking the transpose, we get the column operations on A^T

$$A^T \xrightarrow{\text{col op}} I^T = I.$$

The changes of the determinant under the row operations are combined to give $\det A = d \det I = d$, where d is inverse of the product of $\det E$ for the elementary matrices E corresponding to the row operations. Then by Proposition 6.1.5, we also get $\det A^T = d \det I = d$, for the same d. This proves $\det A^T = \det A$ in case A is invertible. Moreover, since d is a product of $-1, c^{-1}(\neq 0), 1$, we get $\det A = d \neq 0$.

If A is not invertible, then we may use row operations to change A to its row echelon form (again using a 3×3 row echelon form as example)

$$A \xrightarrow{\text{row op}} R = \begin{pmatrix} * & * & * \\ 0 & * & * \\ 0 & 0 & 0 \end{pmatrix}, \quad A^T \xrightarrow{\text{col op}} R^T = \begin{pmatrix} * & 0 & 0 \\ * & * & 0 \\ * & * & 0 \end{pmatrix}.$$

Then we get $\det A = d \det R$ and $\det A^T = d \det R^T$, for the same $d \neq 0$. Then by Proposition 6.1.7, we get $\det R = 0 = \det R^T$. Therefore $\det A = 0 = \det A^T$.

The argument above also implies the following.

Theorem 6.1.9. *A square matrix A is invertible if and only of* $\det A \neq 0$.

Exercise 6.6. Suppose

$$\det \begin{pmatrix} a & b & c \\ x & y & z \\ k & l & m \end{pmatrix} = 3.$$

Find the determinants.

1. $\begin{pmatrix} x+a & y+b & z+c \\ a & b & c \\ k+2x & l+2y & m+2z \end{pmatrix}$. 2. $\begin{pmatrix} m & l+k-2m & k \\ c & b+a-2c & a \\ z & y+x-2z & x \end{pmatrix}$. 3. $\begin{pmatrix} a & -b & c \\ -2x & 2y & -2z \\ 3k & -3l & 3m \end{pmatrix}$.

Exercise 6.7. Use row and column exchanges to change to upper or lower triangular matrices. Then use Proposition 6.1.7 to calculate the determinants.

1. $\begin{pmatrix} * & * & a \\ * & b & 0 \\ c & 0 & 0 \end{pmatrix}$. 2. $\begin{pmatrix} 0 & b & 0 \\ a & 1 & 3 \\ 0 & 2 & c \end{pmatrix}$. 3. $\begin{pmatrix} 0 & 0 & 0 & a_1 \\ 0 & 0 & a_2 & * \\ 0 & a_3 & * & * \\ a_4 & * & * & * \end{pmatrix}$. 4. $\begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 2 & 3 & 4 \\ 5 & 6 & 0 & 7 \\ 8 & 0 & 0 & 9 \end{pmatrix}$.

Exercise 6.8. Explain that $\det A^n = (\det A)^n$ for any integer n.

Exercise 6.9. Use Theorem 6.1.9 to prove that, if A and B are square matrices, then AB is invertible if and only if A and B are invertible.

Exercise 6.10. Prove that any orthogonal matrix has determinant ± 1.

Exercise 6.11. Suppose A and B are 3×3 matrices, with $\det A = 2$ and $\det B = 3$. What are the following determinants?

1. $\det BA$.
2. $\det A^TB$.
3. $\det ABA^{-1}$.
4. $\det A^TBA$.
5. $\det A^{-1}$.
6. $\det B^TB^{-2}$.

Exercise 6.12. Suppose A is a 3×3 matrix and B is a 4×4 matrix, with $\det A = 2$ and $\det B = 3$. What are the following determinants?

1. $\det\begin{pmatrix} B & O \\ X & A \end{pmatrix}$.
2. $\det\begin{pmatrix} A^TA & X \\ O & BB^T \end{pmatrix}$.
3. $\det\begin{pmatrix} O & A \\ B & X \end{pmatrix}$.
4. $\det\begin{pmatrix} X & B \\ A & O \end{pmatrix}$.
5. $\det\begin{pmatrix} B & X & O \\ O & A^T & O \\ O & Z & B^{-1} \end{pmatrix}$.
6. $\det\begin{pmatrix} X & O & B \\ A^T & O & O \\ Z & B^{-1} & O \end{pmatrix}$.

For the determinant of a 3×3 matrix, we express the first column as a linear combination of the standard basis

$$A = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix}, \quad \begin{pmatrix} x_{11} \\ x_{21} \\ x_{31} \end{pmatrix} = x_{11}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + x_{21}\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + x_{31}\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

By the linearity in the first column, we get the first equality below

$$\begin{aligned}
&\det\begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix} \\
&= x_{11}\det\begin{pmatrix} 1 & x_{12} & x_{13} \\ 0 & x_{22} & x_{23} \\ 0 & x_{32} & x_{33} \end{pmatrix} + x_{21}\det\begin{pmatrix} 0 & x_{12} & x_{13} \\ 1 & x_{22} & x_{23} \\ 0 & x_{32} & x_{33} \end{pmatrix} + x_{31}\det\begin{pmatrix} 0 & x_{12} & x_{13} \\ 0 & x_{22} & x_{23} \\ 1 & x_{32} & x_{33} \end{pmatrix} \\
&= x_{11}\det\begin{pmatrix} 1 & x_{12} & x_{13} \\ 0 & x_{22} & x_{23} \\ 0 & x_{32} & x_{33} \end{pmatrix} - x_{21}\det\begin{pmatrix} 1 & x_{22} & x_{23} \\ 0 & x_{12} & x_{13} \\ 0 & x_{32} & x_{33} \end{pmatrix} + x_{31}\det\begin{pmatrix} 1 & x_{32} & x_{33} \\ 0 & x_{12} & x_{13} \\ 0 & x_{22} & x_{23} \end{pmatrix} \\
&= x_{11}\det\begin{pmatrix} x_{22} & x_{23} \\ x_{32} & x_{33} \end{pmatrix} - x_{21}\det\begin{pmatrix} x_{12} & x_{13} \\ x_{32} & x_{33} \end{pmatrix} + x_{31}\det\begin{pmatrix} x_{12} & x_{13} \\ x_{22} & x_{23} \end{pmatrix}.
\end{aligned}$$

The second equality uses the alternating property for the exchange of rows. The third equality uses Theorem 6.1.6.

The formula uses the linearity in the first column to reduce a 3×3 determinant to several 2×2 determinants. We may also use the linearity in any other column or any row to get similar reductions. Here are some examples

$$\begin{aligned}
&\det \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{pmatrix} \\
&= -x_{12} \det \begin{pmatrix} x_{21} & x_{23} \\ x_{31} & x_{33} \end{pmatrix} + x_{22} \det \begin{pmatrix} x_{11} & x_{13} \\ x_{31} & x_{33} \end{pmatrix} - x_{32} \det \begin{pmatrix} x_{11} & x_{13} \\ x_{21} & x_{23} \end{pmatrix} \\
&= x_{11} \det \begin{pmatrix} x_{22} & x_{23} \\ x_{32} & x_{33} \end{pmatrix} - x_{12} \det \begin{pmatrix} x_{21} & x_{23} \\ x_{31} & x_{33} \end{pmatrix} + x_{13} \det \begin{pmatrix} x_{21} & x_{22} \\ x_{31} & x_{32} \end{pmatrix}.
\end{aligned}$$

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 the *cofactor expansion* along the j-th column

$$\det A = (-1)^{1+j} x_{1j} \det A_{1j} + (-1)^{2+j} x_{2j} \det A_{2j} + \cdots + (-1)^{n+j} x_{nj} \det A_{nj},$$

and the cofactor expansion along the 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 6.1.2. Cofactor expansion is more efficient if we expand along a row or column with lots of 0.

$$\begin{aligned}
\det \begin{pmatrix} 0 & 1 & 0 & 2 \\ 3 & 4 & 5 & 6 \\ 7 & 8 & 0 & 9 \\ 0 & 10 & 0 & 11 \end{pmatrix} &= (-1)^{3+2} 5 \det \begin{pmatrix} 0 & 1 & 2 \\ 7 & 8 & 9 \\ 0 & 10 & 11 \end{pmatrix} = -5(-1)^{2+1} 7 \det \begin{pmatrix} 1 & 2 \\ 10 & 11 \end{pmatrix} \\
&= 5 \cdot 7(11 - 2 \cdot 10) = -315.
\end{aligned}$$

Exercise 6.13. Use cofactor expansions to calculate the determinants in Exercise 6.7.

Exercise 6.14. Write down the cofactor expansion along the third row of a 4×4 matrix.

6.1.3 Calculation of Determinant

Example 6.1.3. By column operations, we have

$$\begin{aligned}
\det \begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & a \end{pmatrix} &\xlongequal[\text{Col}_3 - 7\text{Col}_1]{\text{Col}_2 - 4\text{Col}_1} \det \begin{pmatrix} 1 & 0 & 0 \\ 2 & -3 & -6 \\ 3 & -6 & a - 21 \end{pmatrix} \\
&\xlongequal[3\text{Col}_2]{\text{Col}_3 - 2\text{Col}_2} -3 \det \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 2 & a - 9 \end{pmatrix} = -3(a - 9).
\end{aligned}$$

The last equality uses Proposition 6.1.7. By Theorem 6.1.9, the matrix is invertible if and only if $a \neq 9$.

Example 6.1.4. We have

$$\begin{aligned}
\det\begin{pmatrix}1 & 0 & 2 & 0\\ 5 & 6 & 7 & 8\\ 3 & 0 & 4 & 0\\ 9 & 10 & 11 & 12\end{pmatrix} &\overset{\text{Row}_2\leftrightarrow\text{Row}_3}{=\!=\!=\!=} -\det\begin{pmatrix}1 & 0 & 2 & 0\\ 3 & 0 & 4 & 0\\ 5 & 6 & 7 & 8\\ 9 & 10 & 11 & 12\end{pmatrix}\\
&\overset{\text{Col}_2\leftrightarrow\text{Col}_3}{=\!=\!=\!=} \det\begin{pmatrix}1 & 2 & 0 & 0\\ 3 & 4 & 0 & 0\\ 5 & 7 & 6 & 8\\ 9 & 11 & 10 & 12\end{pmatrix}\\
&= \det\begin{pmatrix}1 & 2\\ 3 & 4\end{pmatrix}\det\begin{pmatrix}6 & 8\\ 10 & 12\end{pmatrix}\\
&= (1\cdot 4 - 2\cdot 3)(6\cdot 12 - 8\cdot 10) = 16.
\end{aligned}$$

The third equality uses Proposition 6.1.6.

Exercise 6.15. Calculate the determinants.

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

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

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

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

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

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

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

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

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

Example 6.1.5.

$$
\begin{aligned}
&\det\begin{pmatrix} t & 0 & 0 & a_0 \\ -1 & t & 0 & a_1 \\ 0 & -1 & t & a_2 \\ 0 & 0 & -1 & t+a_3 \end{pmatrix} \\
\xlongequal{\text{Row}_3+t\text{Row}_4}\ &\det\begin{pmatrix} t & 0 & 0 & a_0 \\ -1 & t & 0 & a_1 \\ 0 & -1 & 0 & t^2+a_3t+a_2 \\ 0 & 0 & -1 & t+a_3 \end{pmatrix} \\
\xlongequal{\text{Row}_2+t\text{Row}_3}\ &\det\begin{pmatrix} t & 0 & 0 & a_0 \\ -1 & 0 & 0 & t^3+a_3t^2+a_2t+a_1 \\ 0 & -1 & 0 & t^2+a_3t+a_2 \\ 0 & 0 & -1 & t+a_3 \end{pmatrix} \\
\xlongequal{\text{Row}_1+t\text{Row}_2}\ &\det\begin{pmatrix} 0 & 0 & 0 & t^4+a_3t^3+a_2t^2+a_1t+a_0 \\ -1 & 0 & 0 & t^3+a_3t^2+a_2t+a_1 \\ 0 & -1 & 0 & t^2+a_3t+a_2 \\ 0 & 0 & -1 & t+a_3 \end{pmatrix} \\
\xlongequal[\text{Row}_3\leftrightarrow\text{Row}_4]{\substack{\text{Row}_1\leftrightarrow\text{Row}_2\\ \text{Row}_2\leftrightarrow\text{Row}_3}}\ &-\det\begin{pmatrix} -1 & 0 & 0 & t^3+a_3t^2+a_2t+a_1 \\ 0 & -1 & 0 & t^2+a_3t+a_2 \\ 0 & 0 & -1 & t+a_3 \\ 0 & 0 & 0 & t^4+a_3t^3+a_2t^2+a_1t+a_0 \end{pmatrix} \\
&= t^4+a_3t^3+a_2t^2+a_1t+a_0.
\end{aligned}
$$

Exercise 6.16. Calculate the determinants.

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

Example 6.1.6. We use row and column operations to make a row or column to contain only one nonzero entry. Then we do cofactor expansion along the row or column.

For

$$A = \begin{pmatrix} 1 & -2 & -4 \\ -2 & 4 & -2 \\ -4 & -2 & 1 \end{pmatrix},$$

we have

$$\det(tI-A)=\det\begin{pmatrix} t-1 & 2 & 4\\ 2 & t-4 & 2\\ 4 & 2 & t-1\end{pmatrix} \xlongequal{\mathrm{Col}_1-\mathrm{Col}_3} \det\begin{pmatrix} t-5 & 2 & 4\\ 0 & t-4 & 2\\ -t+5 & 2 & t-1\end{pmatrix}$$
$$\xlongequal{\mathrm{Row}_3+\mathrm{Row}_1} \det\begin{pmatrix} t-5 & 2 & 4\\ 0 & t-4 & 2\\ 0 & 4 & t+3\end{pmatrix} = (t-5)\det\begin{pmatrix} t-4 & 2\\ 4 & t+3\end{pmatrix}$$
$$=(t-5)(t^2-t-20)=(t-5)^2(t+4).$$

Exercise 6.17. Calculate $\det(tI-A)$.

1. $\begin{pmatrix} 2 & 1\\ 3 & 4\end{pmatrix}$.
2. $\begin{pmatrix} 2 & 0\\ 3 & 2\end{pmatrix}$.
3. $\begin{pmatrix} 2 & -1 & -1\\ -1 & 2 & -1\\ -1 & -1 & 2\end{pmatrix}$.
4. $\begin{pmatrix} 2 & -1 & -1\\ 0 & 2 & -1\\ 0 & 0 & 0\end{pmatrix}$.
5. $\begin{pmatrix} 0 & 1 & 1\\ -1 & -2 & -1\\ 1 & 1 & 0\end{pmatrix}$.
6. $\begin{pmatrix} 2 & 4 & 3\\ -4 & -6 & -3\\ 3 & 3 & 1\end{pmatrix}$.
7. $\begin{pmatrix} 2 & 2 & 1\\ 1 & 3 & 1\\ 1 & 2 & 2\end{pmatrix}$.
8. $\begin{pmatrix} 7 & 4 & 16\\ 2 & 5 & 8\\ -2 & -2 & -5\end{pmatrix}$.
9. $\begin{pmatrix} -7 & -16 & 4\\ 6 & 13 & -2\\ 12 & 16 & 1\end{pmatrix}$.

Example 6.1.7. We calculate the determinant of the 4×4 Vandermonde matrix in Example 4.3.22

$$\det\begin{pmatrix} 1 & t_0 & t_0^2 & t_0^3\\ 1 & t_1 & t_1^2 & t_1^3\\ 1 & t_2 & t_2^2 & t_2^3\\ 1 & t_3 & t_3^2 & t_3^3\end{pmatrix} \xlongequal[\mathrm{Col}_2-t_0\mathrm{Col}_1]{\substack{\mathrm{Col}_4-t_0\mathrm{Col}_3\\ \mathrm{Col}_3-t_0\mathrm{Col}_2}} \det\begin{pmatrix} 1 & 0 & 0 & 0\\ 1 & t_1-t_0 & t_1(t_1-t_0) & t_1^2(t_1-t_0)\\ 1 & t_2-t_0 & t_2(t_2-t_0) & t_2^2(t_2-t_0)\\ 1 & t_3-t_0 & t_3(t_3-t_0) & t_3^2(t_3-t_0)\end{pmatrix}$$
$$=\det\begin{pmatrix} t_1-t_0 & t_1(t_1-t_0) & t_1^2(t_1-t_0)\\ t_2-t_0 & t_2(t_2-t_0) & t_2^2(t_2-t_0)\\ t_3-t_0 & t_3(t_3-t_0) & t_3^2(t_3-t_0)\end{pmatrix}$$
$$\xlongequal[(t_3-t_0)\mathrm{Row}_3]{\substack{(t_1-t_0)\mathrm{Row}_1\\ (t_2-t_0)\mathrm{Row}_2}} (t_1-t_0)(t_2-t_0)(t_3-t_0)\det\begin{pmatrix} 1 & t_1 & t_1^2\\ 1 & t_2 & t_2^2\\ 1 & t_3 & t_3^2\end{pmatrix}.$$

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 6.18. Calculate the determinants.

1. $\begin{pmatrix} 1 & a & bc \\ 1 & b & ac \\ 1 & c & ab \end{pmatrix}$.
2. $\begin{pmatrix} 1 & t_0 & t_0^3 \\ 1 & t_1 & t_1^3 \\ 1 & t_2 & t_2^3 \end{pmatrix}$.
3. $\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & a \\ 1 & 1 & a & a \\ 1 & a & a & a \end{pmatrix}$.
4. $\begin{pmatrix} t & 1 & 1 & 1 \\ 1 & t & 1 & a \\ 1 & 1 & t & a \\ 1 & a & 1 & t \end{pmatrix}$.

6.2 Geometry of Determinant

A *parallelogram* in $\mathbb{R}^2$ is spanned by two vectors

$$P(\vec{v}, \vec{w}) = \{x\vec{v} + y\vec{w} \colon 0 \le x, y \le 1\}.$$

The parallelogram $P(\vec{v}, \vec{w})$ has four vertices $\vec{0}$, $\vec{v}$, $\vec{w}$ and $\vec{v} + \vec{w}$.

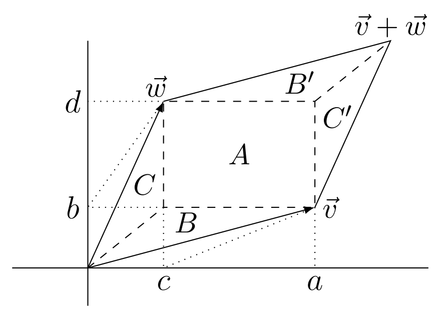

Figure 6.2.1: Area of parallelogram.

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. If $\vec{v} = (a, b)$ and $\vec{w} = (c, d)$, then 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 the dot product, in Section 5.1.3

$$\begin{aligned} \text{Area}(P(\vec{v}, \vec{w})) &= \sqrt{(\vec{v} \cdot \vec{v})(\vec{w} \cdot \vec{w}) - (\vec{v} \cdot \vec{w})^2} \\ &= \sqrt{(a^2 + b^2)(c^2 + d^2) - (ac + bd)^2} \\ &= \sqrt{a^2 d^2 + b^2 c^2 - 2abcd} = |ad - bc|. \end{aligned}$$

The area is non-negative, and only gives the absolute value of the determinant $ad - bc$. To uniquely fix the value of the determinant, we still need to know its sign. This is actually 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*.

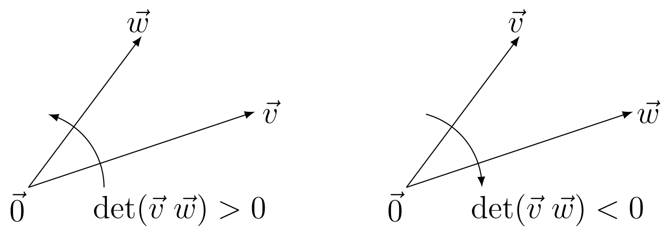

Figure 6.2.2: Sign of the determinant of a 2×2 matrix.

In general, we regard an $m \times n$ matrix $A = (\vec{v}_1\ \vec{v}_2\ \cdots\ \vec{v}_n)$ as an ordered vector set $\alpha = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ in $\mathbb{R}^m$. The *parallelotope* spanned by n vectors is

$$P(A) = P(\alpha) = \{x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n \colon 0 \leq x_i \leq 1\}.$$

For the special case $m = n$, the parallelotope is generally an n-dimensional body inside $\mathbb{R}^n$, and has n-dimensional volume. This volume should be the absolute value of $\det A$. The sign of $\det A$ is determined by the orientation of α. Of course, the sign is needed only if $|\det A| \neq 0$, which means α is an ordered basis of $\mathbb{R}^n$ (Theorem 6.1.9).

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 general, an ordered basis α of $\mathbb{R}^n$ can be continuously deformed to the standard basis $\epsilon = \{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$, or the basis $\epsilon' = \{-\vec{e}_1, \vec{e}_2, \vec{e}_3, \ldots, \vec{e}_n\}$. The deformation means a continuous function $\alpha(t)$ for $t \in [0, 1]$, such that $\alpha(t)$ is always a basis of $\mathbb{R}^n$, and $\alpha(0) = \alpha$, and $\alpha(1) = \epsilon$ or ϵ'.

Definition 6.2.1. An ordered basis α of $\mathbb{R}^n$ is *positively oriented*, if α can be continuously deformed to $\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$. It is *negatively oriented*, if it can be continuously deformed to $\{-\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$.

We will not further discuss the orientation. We only remark that switching two vectors in α changes the orientation, and the sign of the orientation is the sign of the determinant.

Next, we try to rigorously argue the geometric meaning of the determinant. For an $n \times n$ matrix A, let $d(A)$ be the *signed volume* of $P(A)$, i.e., $|d(A)|$ is the volume of $P(A)$, and the sign of $d(A)$ is the orientation of the column vectors of A. We note that the orientation makes sense only if the columns form a basis, i.e., the columns are linearly independent. If the columns are linearly dependent, then all the columns lie inside an $(n-1)$-dimensional subspace, and the n-dimensional volume of $P(A)$ is 0. In this case there is no need to have sign, and $|d(A)| = 0$ implies $d(A) = 0$.

Theorem 6.2.2. *The signed volume* $d(A) = \det A$.

We first remark that, if the columns of A do not form a basis of $\mathbb{R}^n$, then $d(A) = 0$. By Theorem 6.1.9, we also have $\det A = 0$. Therefore we only need to argue $d(A) = \det A$ for the case the columns of A are linearly independent. In this case, we may use column operations to change A to I

$$A \xrightarrow{\text{col op}} I.$$

Similar to the proof of $\det A^T = \det A$, we only need to show that the change of $d(A)$ under column operations on A is the same as the change of $\det A$ in Proposition 6.1.4, and $d(I) = \det I$. Moreover, since each column operation involves only two columns, we only argue for two vectors in $\mathbb{R}^2$. The spirit of the argument applies to the general case.

The first of Figure 6.2.3 compares $d(\vec{v}_1, \vec{v}_2)$ and $d(\vec{v}_2, \vec{v}_1)$. Since the two parallelograms are the same: $P(\vec{v}_1, \vec{v}_2) = P(\vec{v}_2, \vec{v}_1)$, we get the same volume: $|d(\vec{v}_1, \vec{v}_2)| = |d(\vec{v}_2, \vec{v}_1)|$. Moreover, since going from $\vec{v}_1$ to $\vec{v}_2$ is opposite to going from $\vec{v}_2$ to $\vec{v}_1$, we know $d(\vec{v}_1, \vec{v}_2)$ and $d(\vec{v}_2, \vec{v}_1)$ have opposite signs. The combination of volume and sign gives $d(\vec{v}_1, \vec{v}_2) = -d(\vec{v}_2, \vec{v}_1)$.

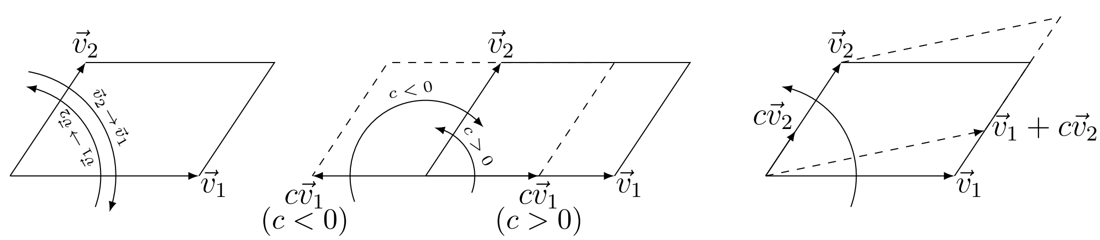

Figure 6.2.3: Column operations on parallelogram.

The second of Figure 6.2.3 compares $d(c\vec{v}_1, \vec{v}_2)$ and $d(\vec{v}_1, \vec{v}_2)$. The parallelogram $P(c\vec{v}_1, \vec{v}_2)$ is the stretch of $P(\vec{v}_1, \vec{v}_2)$ in $\vec{v}_1$ direction by c. Therefore the volume is multiplied by $|c|$: $|d(c\vec{v}_1, \vec{v}_2)| = |c||d(\vec{v}_1, \vec{v}_2)|$. Moreover, going from $c\vec{v}_1$ to $\vec{v}_2$ is the same as going from $\vec{v}_1$ to $\vec{v}_2$ in case $c > 0$, and is the the opposite of going from $\vec{v}_1$

to $\vec{v}_2$ in case $c < 0$. Therefore $d(c\vec{v}_1, \vec{v}_2)$ and $d(\vec{v}_1, \vec{v}_2)$ have the same signs if $c > 0$, and have the opposite signs if $c < 0$. The combination of volume and sign gives $d(c\vec{v}_1, \vec{v}_2) = cd(\vec{v}_1, \vec{v}_2)$.

The third of Figure 6.2.3 compares $d(\vec{v}_1 + c\vec{v}_2, \vec{v}_2)$ and $d(\vec{v}_1, \vec{v}_2)$. The parallelograms $P(\vec{v}_1 + c\vec{v}_2, \vec{v}_2)$ and $P(\vec{v}_1, \vec{v}_2)$ have the same "base" $\vec{v}_2$, and the other side $\vec{v}_1+c\vec{v}_2$ is the shift of $\vec{v}_1$ along the direction of the base. The shift does not change the distance to the base, and therefore preserves the volume: $|d(\vec{v}_1+c\vec{v}_2, \vec{v}_2)| = |d(\vec{v}_1, \vec{v}_2)|$. Moreover, since going from $\vec{v}_1 + c\vec{v}_2$ to $\vec{v}_2$ is the same as going from $\vec{v}_1$ to $\vec{v}_2$, we know $d(\vec{v}_1 + c\vec{v}_2, \vec{v}_2)$ and $d(\vec{v}_1, \vec{v}_2)$ have the same sign. The combination of volume and sign gives $d(\vec{v}_1 + c\vec{v}_2, \vec{v}_2) = d(\vec{v}_1, \vec{v}_2)$.

Since both $\det A$ and $d(A)$ change the same way under column operations, we get $d(A) = ad(I)$ and $\det A = a \det I$, for the same $a \neq 0$. We know $\det I = 1$. The parallelogram $P(I)$ is the unit square and has volume 1. Moreover, we have $I = (\vec{e}_1\ \vec{e}_2)$, and the standard basis $\{\vec{e}_1, \vec{e}_2\}$ is positively oriented. Therefore $d(I) = 1$. We conclude $\det A = a = d(A)$.

We remark that, if $n \leq m$, then an $m \times n$ matrix A gives parallelotope $P(A)$ that is n-dimensional when $n = \text{rank}A$, and is strictly less than n-dimensional when $n > \text{rank}A$. The parallelotope still has n-dimensional volume, given by the formula $\sqrt{\det A^T A}$. For example, if $A = (\vec{x}\ \vec{y})$ is $m \times 2$ matrix, then the area (2-dimensional volume) of $P(A) = P(\vec{x}, \vec{y})$ is

$$\sqrt{\det A^T A} = \sqrt{\det \begin{pmatrix} \vec{x}\cdot\vec{x} & \vec{x}\cdot\vec{y} \\ \vec{y}\cdot\vec{x} & \vec{y}\cdot\vec{y} \end{pmatrix}} = \sqrt{(\vec{x}\cdot\vec{x})(\vec{y}\cdot\vec{y}) - (\vec{x}\cdot\vec{y})^2}.$$

The formula already appeared in Section 5.1.3.

Although $P(A)$ has n-dimensional volume, for $n < m$, we cannot define an orientation for the ordered set of column vectors of A, which presumably leads to the "determinant" of non-square matrix A. The proper way of generalizing the determinant to A is the exterior algebra $\wedge^n \mathbb{R}^m$. For example, the "determinant" of a 3×2 matrix $(\vec{x}\ \vec{y})$ is the cross product $\vec{x} \times \vec{y}$.

What about the case $n > m$? You can show that, in this case, any function that is multilinear and alternating in columns of A is 0!

Exercise 6.19. Use geometric meaning of determinant to explain that the determinant of an orthogonal matrix is ± 1?

Exercise 6.20. Use geometric meaning of determinant to explain that the matrix of any orthogonal projection of $\mathbb{R}^n$ onto a proper subspace (i.e., a subspace that is not $\mathbb{R}^n$) is 0.

Finally, we state the following geometrical characterisation of the determinant.

Theorem 6.2.3. *The determinant is the function defined for all square matrices (all $n \times n$ matrices, for all $n \in \mathbb{N}$), satisfying 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 first equality has the following geometrical interpretation. Let A be $n_1 \times n_1$ and let B be $n_2 \times n_2$. Figure 6.2.4 shows the parallelotope $P\begin{pmatrix} A & O \\ O & B \end{pmatrix}$ can be regarded as a "super-rectangle" with $P(A)$ as one side and $P(B)$ as the other side. Therefore the $(n_1 + n_2)$-dimensional volume of $P\begin{pmatrix} A & O \\ O & B \end{pmatrix}$ is the product of the n_1-dimensional volume of $P(A)$ and the n_2-dimensional volume of $P(B)$. As for the sign, 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 $\mathbb{R}^{n_1+n_2}$.

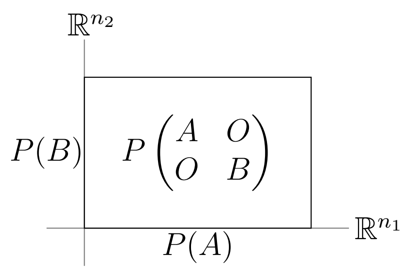

Figure 6.2.4: Geometric property of determinant.

The second equality has the following geometrical interpretation. A linear transformation $L(\vec{x}) = A\vec{x} \colon \mathbb{R}^n \to \mathbb{R}^n$ multiplies the volume by $|\det A|$[1]: $\mathrm{vol}(L(X)) = |\det A| \mathrm{vol}(X)$. Since L takes column vectors of B to column vectors of AB, it takes $P(B)$ to $P(AB)$: $L(P(B)) = P(AB)$. Then

$$\mathrm{vol}(P(AB)) = \mathrm{vol}(L(P(B))) = |\det A| \mathrm{vol}(P(B)) = |\det A| |\det B|.$$

This is the second equality up to the sign.

The proof of Theorem 6.2.3 can be found in any textbook on algebraic K-theory.

[1] This requires further investigation on the true meaning of volume, a topic in measure theory.

Chapter 7

Diagonalisation

7.1 Eigenvalue and Eigenvector

7.1.1 Diagonalisation

The famous Fibonacci numbers $0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, \dots$ is defined through the recursive relation

$$F_0 = 0, \quad F_1 = 1, \quad F_n = F_{n-1} + F_{n-2}.$$

Given a specific number, say 100, we can certainly calculate F_{100} by repeatedly applying the recursive relation. However, it is not obvious what the general formula for F_n should be.

The difficulty in finding the general formula is essentially due to the lack of understanding of the *structure* of the recursion process. The Fibonacci numbers is a linear system because it is governed by a linear equation $F_n = F_{n-1} + F_{n-2}$. Many differential equations such as Newton's second law $F = m\vec{x}''$ are also linear. Understanding the structure of linear operators inherent in linear systems helps us solving problems about the system.

We will obtain the general formula for the Fibonacci numbers in Example 7.1.12. Next we illustrate some simpler examples.

Example 7.1.1. Suppose a pair of numbers x_n, y_n is defined through the recursive relation

$$x_0 = 1, \quad y_0 = 0, \quad x_n = x_{n-1} - y_{n-1}, \quad y_n = x_{n-1} + y_{n-1}.$$

To find the general formula for x_n and y_n, we rewrite the recursive relation as a linear transformation

$$\vec{x}_n = \begin{pmatrix} x_n \\ y_n \end{pmatrix} = \begin{pmatrix} x_{n-1} - y_{n-1} \\ x_{n-1} + y_{n-1} \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_{n-1} \\ y_{n-1} \end{pmatrix} = A\vec{x}_{n-1}, \quad \vec{x}_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \vec{e}_1.$$

Since

$$A = \sqrt{2} \begin{pmatrix} \cos\frac{\pi}{4} & -\sin\frac{\pi}{4} \\ \sin\frac{\pi}{4} & \cos\frac{\pi}{4} \end{pmatrix},$$

the linear operator is the rotation by $\frac{\pi}{4}$ and the scalar multiplication by $\sqrt{2}$. Therefore

$$A^n = 2^{\frac{n}{2}} \begin{pmatrix} \cos\frac{n\pi}{4} & -\sin\frac{n\pi}{4} \\ \sin\frac{n\pi}{4} & \cos\frac{n\pi}{4} \end{pmatrix},$$

and $\vec{x}_n$ is obtained by rotating $\vec{e}_1$ by $\frac{n\pi}{4}$ and has length $(\sqrt{2})^n = 2^{\frac{n}{2}}$. We conclude that

$$x_n = 2^{\frac{n}{2}} \cos\frac{n\pi}{4}, \quad y_n = 2^{\frac{n}{2}} \sin\frac{n\pi}{4}.$$

For example, we have $(x_{8k}, y_{8k}) = (2^{4k}, 0)$ and $(x_{8k+3}, y_{8k+3}) = (-2^{4k+1}, 2^{4k+1})$.

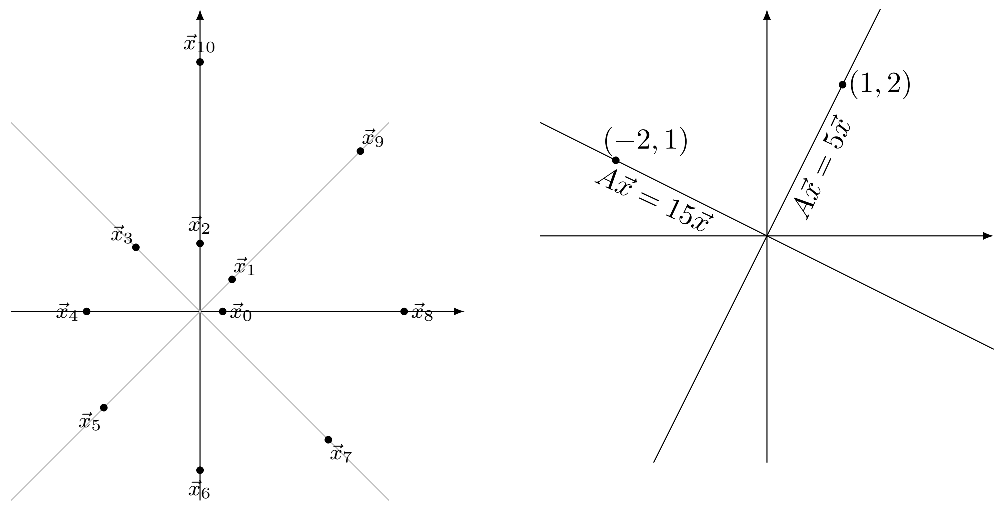

Figure 7.1.1: Geometric meaning of linear operators.

Example 7.1.2. Suppose a linear system is obtained by repeatedly applying the matrix

$$A = \begin{pmatrix} 13 & -4 \\ -4 & 7 \end{pmatrix}.$$

To find A^n, we note that

$$A\begin{pmatrix} 1 \\ 2 \end{pmatrix} = 5\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad A\begin{pmatrix} -2 \\ 1 \end{pmatrix} = 15\begin{pmatrix} -2 \\ 1 \end{pmatrix}.$$

This means that, with respect to the basis $\vec{v}_1 = (1,2), \vec{v}_2 = (-2,1)$, the linear operator simply multiplies 5 in the $\vec{v}_1$ direction and multiplies 15 in the $\vec{v}_2$ direction. In other words, with respect to the basis $\alpha = \{(1,2), (-2,1)\}$, we have

$$[A]_{\alpha\alpha} = \begin{pmatrix} 5 & 0 \\ 0 & 15 \end{pmatrix}.$$

The geometric meaning immediately implies $A^n\vec{v}_1 = 5^n\vec{v}_1$ and $A^n\vec{v}_2 = 15^n\vec{v}_2$. To get $A^n\vec{x}$ for other vector $\vec{x}$, we may express $\vec{x}$ as a linear combinations of $\vec{v}_1, \vec{v}_2$, and then apply A^n. For example, by

$$\begin{pmatrix}0\\1\end{pmatrix} = \frac{2}{5}\begin{pmatrix}1\\2\end{pmatrix} + \frac{1}{5}\begin{pmatrix}-2\\1\end{pmatrix},$$

we get

$$A^n\begin{pmatrix}0\\1\end{pmatrix} = \frac{2}{5}A^n\begin{pmatrix}1\\2\end{pmatrix} + \frac{1}{5}A^n\begin{pmatrix}-2\\1\end{pmatrix} = \frac{2}{5}5^n\begin{pmatrix}1\\2\end{pmatrix} + \frac{1}{5}15^n\begin{pmatrix}-2\\1\end{pmatrix} = 5^{n-1}\begin{pmatrix}2-2\cdot 3^n\\4+3^n\end{pmatrix}.$$

Exercise 7.1. In Example 7.1.1, what do you get if you start with $x_0 = 0$ and $y_0 = 1$?

Exercise 7.2. In Example 7.1.2, find the matrix A^n.

The key idea of Example 7.1.2 is the following. Suppose A is an $n \times n$ matrix. If we can find a basis $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ of $\mathbb{R}^n$ (a big if), such that

$$A\vec{v}_i = d_i\vec{v}_i, \quad d_i \in \mathbb{R},$$

then we may express any vector in $\mathbb{R}^n$ in terms of the basis and get

$$\vec{x} = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n \implies A\vec{x} = d_1x_1\vec{v}_1 + d_2x_2\vec{v}_2 + \cdots + d_nx_n\vec{v}_n.$$

By repeatedly applying A, we even get

$$A^k\vec{x} = d_1^kx_1\vec{v}_1 + d_2^kx_2\vec{v}_2 + \cdots + d_n^kx_n\vec{v}_n.$$

The equalities $A\vec{v}_i = d_i\vec{v}_i$ mean that the matrix of the linear operator $L(\vec{x}) = A\vec{x}$ with respect to the basis is a *diagonal matrix*

$$[L]_{\alpha\alpha} = \begin{pmatrix} d_1 & 0 & \cdots & 0\\ 0 & d_2 & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & d_n \end{pmatrix} = D.$$

Then we may use the change of basis formula to get the matrix A of L with respect to the standard basis ϵ

$$A = [L]_{\epsilon\epsilon} = [I]_{\epsilon\alpha}[L]_{\alpha\alpha}[I]_{\alpha\epsilon} = PDP^{-1}, \quad P = (\alpha) = (\vec{v}_1\ \vec{v}_2\ \cdots\ \vec{v}_n).$$

For example, Example 7.1.2 gives

$$\begin{pmatrix}13 & -4\\ -4 & 7\end{pmatrix} = \begin{pmatrix}1 & -2\\ 2 & 1\end{pmatrix}\begin{pmatrix}5 & 0\\ 0 & 15\end{pmatrix}\begin{pmatrix}1 & -2\\ 2 & 1\end{pmatrix}^{-1}.$$

We emphasis that the expression PDP^{-1} is not unique. By changing the vectors representing directions or exchanging the directions, we may get other PDP^{-1}

$$\begin{aligned}\begin{pmatrix}13 & -4\\ -4 & 7\end{pmatrix} &= \begin{pmatrix}3 & 2\\ 6 & -1\end{pmatrix}\begin{pmatrix}5 & 0\\ 0 & 15\end{pmatrix}\begin{pmatrix}3 & 2\\ 6 & -1\end{pmatrix}^{-1}\\ &= \begin{pmatrix}-2 & 1\\ 1 & 2\end{pmatrix}\begin{pmatrix}15 & 0\\ 0 & 5\end{pmatrix}\begin{pmatrix}-2 & 1\\ 1 & 2\end{pmatrix}^{-1}.\end{aligned}$$

Definition 7.1.1. A square matrix A is *diagonalisable* if $A = PDP^{-1}$ for an invertible matrix P and a diagonal matrix D.

The expression $A = PDP^{-1}$ is called a *diagonalisation*. It is very easy to use a diagonalisation to calculate the power of a matrix

$$A^k = PD^kP^{-1}, \quad D^k = \begin{pmatrix}d_1^k & 0 & \cdots & 0\\ 0 & d_2^k & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & d_n^k\end{pmatrix}.$$

Example 7.1.3. The orthogonal projection L of $\mathbb{R}^3$ onto the subspace H given by $x+y+z=0$ in Example 3.1.12 is characterised by

$$L\begin{pmatrix}1\\ -1\\ 0\end{pmatrix} = 1\begin{pmatrix}1\\ -1\\ 0\end{pmatrix}, \quad L\begin{pmatrix}1\\ 0\\ -1\end{pmatrix} = 1\begin{pmatrix}1\\ 0\\ -1\end{pmatrix}, \quad L\begin{pmatrix}1\\ 1\\ 1\end{pmatrix} = 0\begin{pmatrix}1\\ 1\\ 1\end{pmatrix}.$$

This leads to the diagonalisation

$$[L] = \begin{pmatrix}1 & 1 & 1\\ -1 & 0 & 1\\ 0 & -1 & 1\end{pmatrix}\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}1 & 1 & 1\\ -1 & 0 & 1\\ 0 & -1 & 1\end{pmatrix}^{-1}.$$

In Example 4.3.26, we used the diagonalisation to calculate the matrix of L.

We also remark that, if we exchange the second and third vectors, then we get another diagonalisation

$$[L] = \begin{pmatrix}1 & 1 & 1\\ -1 & 1 & 0\\ 0 & 1 & -1\end{pmatrix}\begin{pmatrix}1 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}1 & 1 & 1\\ -1 & 1 & 0\\ 0 & 1 & -1\end{pmatrix}^{-1}.$$

Exercise 7.3. Write down a diagonalisation of the flip of $\mathbb{R}^3$ with respect to the subspace $x+y+z=0$.

Exercise 7.4. Find diagonalisations and matrices of L.

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

Exercise 7.5. Find bases $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$ satisfying $A\vec{v}_i = d_i\vec{v}_i$.

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

Exercise 7.6. If A is diagonalisable, explain that the following are also diagonalisable: A^T, A^2, $3I + 2A$, A^{-1} (if A is invertible).

7.1.2 Eigenvector

The key for diagonalisation is to find vectors in the following definition.

Definition 7.1.2. Suppose A is a square matrix. If $A\vec{v} = \lambda\vec{v}$ for a number λ and a nonzero vector $\vec{v}$, then we say $\vec{v}$ is an *eigenvector* of A with *eigenvalue* λ.

Exercise 7.7. Determine whether the vectors are eigenvectors. For eigenvectors, also determine the corresponding eigenvalues.

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

Exercise 7.8. Suppose $\vec{u}$ and $\vec{v}$ are eigenvectors of A. If they also have the same eigenvalue, explain that $a\vec{u} + b\vec{v}$ (as long as this is nonzero) is also an eigenvector. What if they have different eigenvalues?

Exercise 7.9. Suppose A is diagonalisable, and has only one eigenvalue λ. What is A?

Exercise 7.10. Explain that 1 is the only eigenvalue of I, and 0 is the only eigenvalue of O. In general, c is the only eigenvalue of cI.

Exercise 7.11. Explain that a square matrix is invertible if and only if 0 is not an eigenvalue.

Exercise 7.12. Suppose λ and μ are eigenvalues of A. Explain that $2\lambda, \lambda\mu, \lambda+\mu$ may not be eigenvalues of A?

A diagonalisation is then a basis of eigenvectors. Example 7.1.3 shows that the same eigenvalue may be repeated for different vectors in a basis. The numbers d_i in a diagonalisation are repeated eigenvalues.

If $A\vec{v} = \lambda\vec{v}$, then we can easily get $A^k\vec{v} = \lambda^k\vec{v}$, and

$$(A^2 + 3A + 5I)\vec{v} = A^2\vec{v} + 3A\vec{v} + 5I\vec{v} = \lambda^2\vec{v} + 3\lambda\vec{v} + 5\vec{v} = (\lambda^2 + 3\lambda + 5)\vec{v}.$$

In general, for a polynomial

$$p(t) = a_n t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0,$$

we define

$$p(A) = a_n A^n + a_{n-1}A^{n-1} + \cdots + a_1 A + a_0 I.$$

Then

$$A\vec{v} = \lambda\vec{v} \implies p(A)\vec{v} = p(\lambda)\vec{v}.$$

Proposition 7.1.3. *If $\vec{v}$ is an* eigenvector *of A with eigenvalue λ, then for any polynomial $p(t)$, $\vec{v}$ is an* eigenvector *of $p(A)$ with eigenvalue $p(\lambda)$.*

Example 7.1.4. The linear operator in Example 7.1.3 has a basis of eigenvectors with two eigenvalues $\lambda = 1, 0$. The "minimal" polynomial vanishing at the two eigenvalues is $t(t-1) = t^2 - t$. Then for a basis $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ of eigenvectors, we get

$$(L^2 - L)(\vec{v}_i) = (\lambda^2 - \lambda)\vec{v}_i = 0\vec{v}_i = \vec{0}.$$

This implies $L^2 - L = O$, or $L^2 = L$. Of course, we already know this for the projection L.

Example 7.1.5. Suppose a matrix A satisfies $p(A) = A^3 - 3A^2 + 2A = O$. If λ is an eigenvalue of A, then $p(\lambda) = \lambda^3 - 3\lambda^2 + 2\lambda = \lambda(\lambda-1)(\lambda-2)$ is an eigenvalue of $p(A) = O$. Therefore $p(\lambda) = 0$ (see Exercise 7.10), and we get $\lambda = 0, 1$ or 2.

If A is invertible, then by $p(A) = (A^2 - 3A + 2I)A = O$, we get $A^2 - 3A + 2I = O$. By the same idea, this implies $\lambda = 1$ or 2. Moreover, $A^2 - 3A + 2I = O$ implies $A(A-3) = -2I$. Therefore we get a formula $A^{-1} = -\frac{1}{2}(A-3)$ for the inverse.

Exercise 7.13. Suppose $\vec{v}$ is an eigenvector of A with eigenvalue λ, and A is invertible. Explain that $\lambda \neq 0$, and $\vec{v}$ is an eigenvector of A with eigenvalue λ^{-1}.

Exercise 7.14. Use the diagonalisations in Exercise 7.4 to find the matrices of L^3 and $L^2 + L - 6I$.

Exercise 7.15. Suppose $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is a basis of $\mathbb{R}^3$, and $A\vec{v}_1 = 2\vec{v}_1$, $A\vec{v}_2 = 2\vec{v}_2$, $A\vec{v}_3 = -3\vec{v}_3$.

1. Show that $A^2 + A - 6I = O$.
2. Express A^3 as $a_2A^2 + a_1A + a_0I$.
3. Show that A is invertible, and express A^{-1} as $aL + bI$.

Exercise 7.16. Find the relation between the eigenvalues and eigenvectors of similar matrices A and $B = PAP^{-1}$.

7.1.3 Eigenspace

The essence of Example 7.1.3 is to decompose the Euclidean space $\mathbb{R}^3$ into a direct sum of special subspaces

$$\mathbb{R}^3 = E_1 \oplus E_0, \quad E_1 = \{(x, y, z) : x + y + z = 0\}, \quad E_0 = \mathbb{R}(1, 1, 1),$$

such that $A\vec{x} = 1\vec{x}$ on E_1 and $A\vec{x} = \vec{0} = 0\vec{x}$ on E_0. If we take any basis $\{\vec{v}_1, \vec{v}_2\}$ of E_1 and any basis $\{\vec{v}_3\}$ of E_0, then we get the diagonalisation

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

Similarly, the essence of Example 7.1.2 is

$$\mathbb{R}^2 = E_5 \oplus E_{15}, \quad E_5 = \mathbb{R}(1, 2), \quad E_{15} = \mathbb{R}(-2, 1).$$

The essence of diagonalisation is therefore finding subspaces E_λ consisting of all $\vec{x}$ satisfying $A\vec{x} = \lambda\vec{x}$. The equality is the same as $(\lambda I - A)\vec{x} = \vec{0}$.

Definition 7.1.4. If λ is an eigenvalue of A, then the *eigenspace* of A with eigenvalue λ is the null space $E_\lambda = \text{Nul}(\lambda I - A)$.

The definition of eigenvalue λ requires a nonzero eigenvector $\vec{v} \in \text{Nul}(\lambda I - A)$. Therefore the eigenspace is not the zero subspace.

Proposition 7.1.5. *The sum of eigenspaces of distinct eigenvalues is direct.*

Suppose $\lambda_1, \lambda_2, \lambda_3$ are distinct eigenvalues. Suppose $\vec{x}_i \in E_{\lambda_i}$ satisfy

$$\vec{x}_1 + \vec{x}_2 + \vec{x}_3 = \vec{0}.$$

We need to show $\vec{x}_1 = \vec{x}_2 = \vec{x}_3 = \vec{0}$.

By Proposition 7.1.3, for any polynomial $p(t)$, we may apply $p(A)$ to get

$$\vec{0} = p(A)(\vec{x}_1 + \vec{x}_2 + \vec{x}_3) = p(\lambda_1)\vec{x}_1 + p(\lambda_2)\vec{x}_2 + p(\lambda_3)\vec{x}_3.$$

Taking $p(t) = (t - \lambda_2)(t - \lambda_3)$, we get $p(\lambda_2) = p(\lambda_3) = 0$, and the equality above becomes

$$(\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3)\vec{x}_1 = \vec{0}.$$

Since $\lambda_1, \lambda_2, \lambda_3$ are distinct, we have $(\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3) \neq 0$. Therefore $\vec{x}_1 = \vec{0}$. By similar argument, we get $\vec{x}_2 = \vec{0}$ and $\vec{x}_3 = \vec{0}$.

Exercise 7.17. If $\lambda_1, \lambda_2, \lambda_3$ are distinct eigenvalues of A, and $\mathbb{R}^n = E_{\lambda_1} + E_{\lambda_2} + E_{\lambda_3}$. Explain that $\lambda_1, \lambda_2, \lambda_3$ are the only eigenvalues of A.

Let $\lambda_1, \lambda_2, \dots, \lambda_k$ be all the distinct eigenvalues of A. We have the corresponding eigenspaces $E_{\lambda_i} = \mathrm{Nul}(\lambda_i I - A)$. By Proposition 7.1.5, we get a direct sum

$$E_{\lambda_1} \oplus E_{\lambda_2} \oplus \cdots \oplus E_{\lambda_k} \subset \mathbb{R}^n.$$

The direct sum is the whole space $\mathbb{R}^n$

$$E_{\lambda_1} \oplus E_{\lambda_2} \oplus \cdots \oplus E_{\lambda_k} = \mathbb{R}^n,$$

if and only if

$$\dim E_{\lambda_1} + \dim E_{\lambda_2} + \cdots + \dim E_{\lambda_k} = n.$$

In this case, we take a basis α_i of E_{λ_i}. Then the union $\alpha_1 \cup \alpha_2 \cup \cdots \cup \alpha_k$ is a basis of $\mathbb{R}^n$ consisting of eigenvectors. This means exactly that A is diagonalisable.

Theorem 7.1.6. *Suppose $\lambda_1, \lambda_2, \dots, \lambda_k$ are all the distinct eigenvalues of A, and*

$$p(t) = (t - \lambda_1)(t - \lambda_2) \cdots (t - \lambda_k).$$

Then A is diagonalisable if and only if

$$p(A) = (A - \lambda_1 I)(A - \lambda_2 I) \cdots (A - \lambda_k I) = O.$$

Let us consider the case $k = 3$. If A is diagonalisable, then any vector $\vec{x} \in \mathbb{R}^n$ can be written as

$$\vec{x} = \vec{x}_1 + \vec{x}_2 + \vec{x}_3, \quad A\vec{x} = \lambda_i \vec{x}.$$

Then for any polynomial $p(t)$, we have

$$p(A)\vec{x} = p(\lambda_1)\vec{x}_1 + p(\lambda_2)\vec{x}_2 + p(\lambda_3)\vec{x}_3.$$

In particular, since the polynomial

$$p(t) = (t - \lambda_1)(t - \lambda_2)(t - \lambda_3)$$

satisfies $p(\lambda_1) = p(\lambda_2) = p(\lambda_3) = 0$, we get $p(A)\vec{x} = \vec{0}$ for any $\vec{x}$. This proves $p(A) = O$.

For the converse, we assume

$$p(A) = (A - \lambda_1 I)(A - \lambda_2 I)(A - \lambda_3 I) = O.$$

We wish to express any $\vec{x} \in \mathbb{R}^n$ as $\vec{x} = \vec{x}_1 + \vec{x}_2 + \vec{x}_3$, with $\vec{x}_i \in E_{\lambda_i}$. Let

$$p_1(t) = (t - \lambda_2)(t - \lambda_3), \quad p_2(t) = (t - \lambda_1)(t - \lambda_3), \quad p_3(t) = (t - \lambda_1)(t - \lambda_2).$$

By the Lagrange interpolation in Example 4.3.15, we can find suitable coefficients a_1, a_2, a_3, such that (in fact, $a_1 = \frac{1}{(\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3)}$, and so on)

$$1 = a_1 p_1(t) + a_2 p_2(t) + a_3 p_3(t).$$

Then we get

$$I = a_1 p_1(A) + a_2 p_2(A) + a_3 p_3(A).$$

Applying to any $\vec{x} \in \mathbb{R}^n$, we get

$$\vec{x} = a_1 p_1(A)\vec{x} + a_2 p_2(A)\vec{x} + a_3 p_3(A)\vec{x}.$$

We have

$$(A - \lambda_1 I)(a_1 p_1(A)\vec{x}) = a_1 (A - \lambda_1 I)(A - \lambda_2 I)(A - \lambda_3 I)\vec{x} = a_1 O\vec{x} = \vec{0}.$$

Therefore $a_1 p_1(A)\vec{x} \in E_{\lambda_1}$. Similarly, we get $a_2 p_2(A)\vec{x} \in E_{\lambda_2}$ and $a_3 p_3(A)\vec{x} \in E_{\lambda_3}$.

7.1.4 Characteristic Polynomial

To find diagonalisation of a matrix A, we need to first find all the eigenvalues of A. The definition of an eigenvalue λ of A means $(\lambda I - A)\vec{x} = \vec{0}$ has nonzero solution. Since $\lambda I - A$ is a square matrix, this means $\lambda I - A$ is not invertible. By Theorem 6.1.9, this means $\det(\lambda I - A) = 0$.

Definition 7.1.7. The *characteristic polynomial* of a square matrix A is $\det(tI - A)$.

If A is $n \times n$, then $\det(tI - A)$ is a polynomial of degree n. The eigenvalues of A are the roots of the characteristic polynomial.

Exercise 7.18. Explain that A and A^T have the same eigenvalues. Do they have the same eigenvectors? (You may use Exercise 7.7 to test this question).

Example 7.1.6. For the matrix in Example 7.1.2, the characteristic polynomial

$$\begin{aligned}\det(tI - A) &= \det\begin{pmatrix} t - 13 & 4 \\ 4 & t - 7 \end{pmatrix} \\ &= (t - 13)(t - 7) - 16 = t^2 - 20t + 75 = (t - 5)(t - 15).\end{aligned}$$

We get two eigenvalues 5 and 15.

For the eigenvalue 5, we solve

$$(5I - A)\vec{x} = \begin{pmatrix} 5-13 & 4 \\ 4 & 5-7 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -8 & 4 \\ 4 & -2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \vec{0}.$$

The eigenspace $E_5 = \mathrm{Nul}(5I - A) = \mathbb{R}(1,2)$. We may choose a basis $\vec{v}_1 = (1,2)$.

For the eigenvalue 15, we solve

$$(15I - A)\vec{x} = \begin{pmatrix} 15-13 & 4 \\ 4 & 15-7 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 & 4 \\ 4 & 8 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \vec{0}.$$

The eigenspace $E_{15} = \mathrm{Nul}(15I - A) = \mathbb{R}(-2,1)$. We may choose a basis $\vec{v}_2 = (-2,1)$.

The union of bases $\{\vec{v}_1, \vec{v}_2\}$ consists of two vectors, which is the same as $\dim \mathbb{R}^2$. Therefore the matrix is diagonalisable, with the diagonalisation

$$\begin{pmatrix} 13 & -4 \\ -4 & 7 \end{pmatrix} = \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} 5 & 0 \\ 0 & 15 \end{pmatrix}\begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix}^{-1}.$$

Example 7.1.7. By Example 6.1.6, the characteristic polynomial of the matrix

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

is $(t-5)^2(t+4)$. We get two eigenvalues 5 and -4, and

$$5I - A = \begin{pmatrix} 4 & 2 & 4 \\ 2 & 1 & 2 \\ 4 & 2 & 4 \end{pmatrix}, \qquad E_5 = \mathrm{Nul}(5I - A) = \mathbb{R}\begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix} \oplus \mathbb{R}\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix},$$

$$-4I - A = \begin{pmatrix} -5 & 2 & 4 \\ 2 & -9 & 2 \\ 4 & 2 & -5 \end{pmatrix}, \quad E_{-4} = \mathrm{Nul}(-4I - A) = \mathbb{R}\begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}.$$

The union of bases consists of three vectors, and we get the corresponding diagonalisation

$$\begin{pmatrix} 1 & -2 & -4 \\ -2 & 4 & -2 \\ -4 & -2 & 1 \end{pmatrix} = \begin{pmatrix} -1 & -1 & 2 \\ 2 & 0 & 1 \\ 0 & 1 & 2 \end{pmatrix}\begin{pmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & -4 \end{pmatrix}\begin{pmatrix} -1 & -1 & 2 \\ 2 & 0 & 1 \\ 0 & 1 & 2 \end{pmatrix}^{-1}.$$

Example 7.1.8. The characteristic polynomial of the matrix

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

is

$$\det(tI-A)=\det\begin{pmatrix}t-3 & -1 & 3\\ 1 & t-5 & 3\\ 6 & -6 & t+2\end{pmatrix}=\det\begin{pmatrix}t-4 & -1 & 3\\ t-4 & t-5 & 3\\ 0 & -6 & t+2\end{pmatrix}$$
$$=\det\begin{pmatrix}t-4 & -1 & 3\\ 0 & t-4 & 0\\ 0 & -6 & t+2\end{pmatrix}=(t-4)^2(t+2).$$

We get eigenvalues 4 and -2, and

$$4I-A=\begin{pmatrix}1 & -1 & 3\\ 1 & -1 & 3\\ 6 & -6 & 6\end{pmatrix}, \qquad E_4=\mathrm{Nul}(4I-A)=\mathbb{R}\begin{pmatrix}1\\1\\0\end{pmatrix},$$
$$-2I-A=\begin{pmatrix}-5 & -1 & 3\\ 1 & -7 & 3\\ 6 & -6 & 0\end{pmatrix}, \qquad E_{-2}=\mathrm{Nul}(-2I-A)=\mathbb{R}\begin{pmatrix}1\\1\\2\end{pmatrix}.$$

Since the sum of the eigenspaces $E_4\oplus E_{-2}$ is (two dimensional and) not $\mathbb{R}^3$, the matrix is *not diagonalisable*.

We emphasis that not all matrices are diagonalisable. In fact, the matrix in Example 7.1.8 fails Theorem 7.1.6

$$(A-4I)(A+2I)=\begin{pmatrix}1 & -1 & 3\\ 1 & -1 & 3\\ 6 & -6 & 6\end{pmatrix}\begin{pmatrix}-5 & -1 & 3\\ 1 & -7 & 3\\ 6 & -6 & 0\end{pmatrix}=\begin{pmatrix}12 & -12 & 0\\ 12 & -12 & 0\\ 0 & 0 & 0\end{pmatrix}\neq O.$$

However, if we substitute A into the characteristic polynomial $(t-4)^2(t+2)$, then we get

$$(A-4I)^2(A+2I)=(A-4I)(A-4I)(A+2I)$$
$$=\begin{pmatrix}1 & -1 & 3\\ 1 & -1 & 3\\ 6 & -6 & 6\end{pmatrix}\begin{pmatrix}12 & -12 & 0\\ 12 & -12 & 0\\ 0 & 0 & 0\end{pmatrix}=\begin{pmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{pmatrix}.$$

Theorem 7.1.8 (Caylay-Hamilton Theorem)**.** *Let $p(t)=\det(tI-A)$ be the characteristic polynomial of a matrix A. Then $p(A)=O$.*

For a 2×2 matrix, the theorem says

$$\begin{pmatrix}a & b\\ c & d\end{pmatrix}^2-(a+d)\begin{pmatrix}a & b\\ c & d\end{pmatrix}+(ad-bc)\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}=\begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix}.$$

One could imagine that a direct computational proof of the theorem, especially for large size matrix, would be very complicated. We skip the proof of the theorem.

In general, the characteristic polynomial is

$$p(t) = \det(tI - A) = (t-\lambda_1)^{n_1}(t-\lambda_2)^{n_2}\cdots(t-\lambda_k)^{n_k},$$

where λ_i are all the distinct roots of $p(t)$, and n_i is the *algebraic multiplicity* of λ_i, and $n_1 + n_2 + \cdots + n_k = n$.

Theorem 7.1.8 says that the following is always true

$$(A-\lambda_1 I)^{n_1}(A-\lambda_2 I)^{n_2}\cdots(A-\lambda_k I)^{n_k} = O.$$

Theorem 7.1.6 says that, if A is diagonalisable, then

$$(A-\lambda_1 I)(A-\lambda_2 I)\cdots(A-\lambda_k I) = O.$$

In general, we have

$$(A-\lambda_1 I)^{m_1}(A-\lambda_2 I)^{m_2}\cdots(A-\lambda_k I)^{m_k} = O$$

for some minimal $1 \le m_i \le n_i$. Then

$$m(t) = (t-\lambda_1)^{m_1}(t-\lambda_2)^{m_2}\cdots(t-\lambda_k)^{m_k}$$

is the polynomial of smallest degree, such that $m(A) = O$. We call $m(t)$ the *minimal polynomial* of A, and call m_i the *geometric multiplicity* of λ_i.

By Theorem 7.1.6, the minimal polynomials for Examples 7.1.6 and 7.1.7 are $(t-15)(t-5)$ and $(t-5)(t+4)$. The minimal polynomial for Example 7.1.8 is $(t-4)^2(t+2)$.

In the special case of all $n_i = 1$, we have all $m_i = 1$, and A satisfies Theorem 7.1.6. Therefore we get the following result.

Proposition 7.1.9. *If an $n \times n$ matrix has n distinct eigenvalues, then the matrix is diagonalisable.*

We give a direct explanation of the proposition, again for a 3×3 matrix A. Let $\lambda_1, \lambda_2, \lambda_3$ be three distinct eigenvalues of A. Then $\dim E_{\lambda_i} \ge 1$ for each i. This implies

$$\dim(E_{\lambda_1} \oplus E_{\lambda_2} \oplus E_{\lambda_3}) = \dim E_{\lambda_1} + \dim E_{\lambda_2} + \dim E_{\lambda_3} \ge 3.$$

Combined with $E_{\lambda_1} \oplus E_{\lambda_2} \oplus E_{\lambda_3} \subset \mathbb{R}^3$, we get $\dim E_{\lambda_i} = 1$ for each i, and $E_{\lambda_1} \oplus E_{\lambda_2} \oplus E_{\lambda_3} = \mathbb{R}^3$. In particular, we know A is diagonalisable.

Example 7.1.9. Consider

$$A = \begin{pmatrix} 1 & 0 & 0 \\ \sin 2 & \sqrt{2} & 0 \\ \pi & \sin 3 & e^3 \end{pmatrix}.$$

By Proposition 6.1.7, we have

$$\det(tI - A) = \det \begin{pmatrix} t-1 & 0 & 0 \\ -\sin 2 & t-\sqrt{2} & 0 \\ -\pi & -\sin 3 & t-e^3 \end{pmatrix} = (t-1)(t-\sqrt{2})(t-e^3).$$

We get three distinct eigenvalues $1, \sqrt{2}, e^3$. Then we know the matrix is diagonalisable even without calculating the eigenspaces.

Exercise 7.19. Diagonalise the matrices. If you cannot, explain why.

1. $\begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix}$.
2. $\begin{pmatrix} 2 & 0 \\ 3 & 2 \end{pmatrix}$.
3. $\begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}$.
4. $\begin{pmatrix} 2 & -1 & -1 \\ 0 & 2 & -1 \\ 0 & 0 & 0 \end{pmatrix}$.
5. $\begin{pmatrix} 0 & 1 & 1 \\ -1 & -2 & -1 \\ 1 & 1 & 0 \end{pmatrix}$.
6. $\begin{pmatrix} 2 & 4 & 3 \\ -4 & -6 & -3 \\ 3 & 3 & 1 \end{pmatrix}$.
7. $\begin{pmatrix} 2 & 2 & 1 \\ 1 & 3 & 1 \\ 1 & 2 & 2 \end{pmatrix}$.
8. $\begin{pmatrix} 7 & 4 & 16 \\ 2 & 5 & 8 \\ -2 & -2 & -5 \end{pmatrix}$.
9. $\begin{pmatrix} -7 & -16 & 4 \\ 6 & 13 & -2 \\ 12 & 16 & 1 \end{pmatrix}$.

Exercise 7.20. Without finding the eigenvectors, explain that the matrices are diagonalisale

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

Exercise 7.21. Find the condition on a, b, such that the matrices are diagonalisable

1. $\begin{pmatrix} 2 & 1 \\ a & 4 \end{pmatrix}$.
2. $\begin{pmatrix} a & b \\ 0 & a \end{pmatrix}$.

7.1.5 Complex Eigenvalue

For a real matrix A. the characteristic polynomial $\det(tI - A)$ is a polynomial with real coefficients. The roots of polynomial are either real, or complex conjugate pairs.

Example 7.1.10. The characteristic polynomial of the matrix in Example 7.1.1

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

is

$$\det(tI - A) = \det \begin{pmatrix} t-1 & 1 \\ -1 & t-1 \end{pmatrix} = (t-1)^2 + 1 = (t-1-i)(t-1+i).$$

We get a conjugate pair of eigenvalues $\lambda = 1 + i$ and $\bar{\lambda} = 1 - i$. Due to complex eigenvalues, we need to view the real matrix A as a linear operator on the complex vector space $\mathbb{C}^2$.

The (complex) eigenspace $E^{\mathbb{C}}_{1+i} = \mathrm{Nul}^{\mathbb{C}}((1+i)I - A)$ is the solution of

$$((1+i)I - A)\vec{x} = \begin{pmatrix} (1+i)-1 & 1 \\ -1 & (1+i)-1 \end{pmatrix}\vec{x} = \begin{pmatrix} i & 1 \\ -1 & i \end{pmatrix}\vec{x} = \vec{0}.$$

We get $E^{\mathbb{C}}_{1+i} = \mathbb{C}(1, -i)$, and the complex eigenvector $\vec{v} = (1, -i)$ with complex eigenvalue $\lambda = 1 + i$.

Taking the complex conjugation of $A\vec{v} = \lambda\vec{v}$, we get

$$A\bar{\vec{v}} = \bar{A}\bar{\vec{v}} = \bar{\lambda}\bar{\vec{v}}.$$

Therefore $\bar{\vec{v}} = (\bar{1}, \overline{-i}) = (1, i)$ is an eigenvector of A with eigenvalue $\bar{\lambda} = \overline{1+i} = 1-i$. The corresponding complex eigenspace $E^{\mathbb{C}}_{1-i} = \mathrm{Nul}^{\mathbb{C}}((1-i)I - A) = \mathbb{C}(1, i)$ is obtained by applying the complex conjugation to $E^{\mathbb{C}}_{1+i} = \mathbb{C}(1, -i)$.

Propositions 7.1.5 and 7.1.9 remain valid for complex vector spaces, and we get $E^{\mathbb{C}}_{1+i} \oplus E^{\mathbb{C}}_{1-i} = \mathbb{C}^2$. Therefore A is diagonalisable, with diagonalisation

$$\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix}\begin{pmatrix} 1+i & 0 \\ 0 & 1-i \end{pmatrix}\begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix}^{-1}.$$

Next we study the meaning of conjugate pairs of complex eigenvalues for real matrices, in the real vector space.

Let $\lambda = \mu + i\nu$ be an eigenvalue of a real matrix A, with $\mu, \nu \in \mathbb{R}$ and $\nu \neq 0$. An eigenvector of A with eigenvalue λ is also a complex eigenvector $\vec{v} = \vec{u} - i\vec{w}$, with $\vec{u}, \vec{w} \in \mathbb{R}^n$. We emphasis that $\vec{v}$ is *not* expressed as $\vec{u} + i\vec{w}$ here.

The equality $A\vec{v} = \lambda\vec{v}$ is

$$A\vec{u} - iA\vec{w} = A(\vec{u} - i\vec{w}) = (\mu + i\nu)(\vec{u} - i\vec{w}) = (\mu\vec{u} + \nu\vec{w}) - i(-\nu\vec{u} + \mu\vec{w}).$$

Comparing the real and imaginary parts of both sides, we get

$$\begin{aligned} A\vec{u} &= \mu\vec{u} + \nu\vec{w}, \\ A\vec{w} &= -\nu\vec{u} + \mu\vec{w}. \end{aligned}$$

The equality $A\bar{\vec{v}} = \bar{\lambda}\bar{\vec{v}}$ for the conjugate eigenvalue and eigenvector yields the same equalities.

We claim that $\vec{u}, \vec{w} \in \mathbb{R}^n$ are linearly independent. The reason is that, since λ and $\bar{\lambda}$ are distinct, by Proposition 7.1.5, we know $\vec{v} = \vec{u} - i\vec{w}$ and $\bar{\vec{v}} = \vec{u} + i\vec{w}$ are complex linearly independent. Since the matrix $\begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix}$ relating the pairs $\{\vec{v}, \bar{\vec{v}}\}$ and $\{\vec{u}, \vec{w}\}$ is invertible, we know $\vec{u}, \vec{w}$ are also complex linearly independent. Then for the real vectors $\vec{u}, \vec{w}$, complex linearly independent is equivalent to real linearly independent.

Now we have a 2-dimensional subspace $\mathbb{R}\vec{u} \oplus \mathbb{R}\vec{w} \subset \mathbb{R}^n$, and the restriction of A on the subspace has the matrix

$$[A|_{\mathbb{R}\vec{u}\oplus\mathbb{R}\vec{w}}] = \begin{pmatrix} \mu & -\nu \\ \nu & \mu \end{pmatrix}.$$

Let $\lambda = re^{i\theta}$, which means the polar coordinate $(\mu, \nu) = r(\cos\theta, \sin\theta)$. Then

$$[A|_{\mathbb{R}\vec{u}\oplus\mathbb{R}\vec{w}}] = r \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = rR_\theta$$

is the combination of rotation by θ and scaling by r.

Now we apply the interpretation to Example 7.1.10. The equality $A(1, -i) = (1+i)(1, -i)$ means

$$A(\vec{e}_1 - i\vec{e}_2) = (1+i)(\vec{e}_1 - i\vec{e}_2), \quad \mu = 1,\ \nu = 1,\ \vec{u} = \vec{e}_1,\ \vec{w} = \vec{e}_2.$$

Then $1 + i = \sqrt{2}e^{\frac{\pi}{4}}$, and A is the rotation by $\frac{\pi}{4}$ and scaling by $\sqrt{2}$, with respect to $\vec{u} = \vec{e}_1$ and $\vec{w} = \vec{e}_2$.

Example 7.1.11. Consider

$$A = \begin{pmatrix} 1 & -2 & 4 \\ 2 & 2 & -1 \\ 0 & 3 & 0 \end{pmatrix}, \quad \det(tI - A) = (t-3)(t^2+9).$$

We get eigenvalues $3, 3i, -3i$, and

$$3I - A = \begin{pmatrix} 2 & 2 & -4 \\ -2 & 1 & 1 \\ 0 & -3 & 3 \end{pmatrix}, \quad 3iI - A = \begin{pmatrix} -1+3i & 2 & -4 \\ -2 & -2+3i & 1 \\ 0 & -3 & 3i \end{pmatrix}.$$

For $\vec{x} \in E_{3i}^{\mathbb{C}} = \mathrm{Nul}^{\mathbb{C}}(3iI - A)$, the last equation $-3x_2 + 3ix_3 = 0$ gives $x_2 = ix_3$. Substituting into the second equation, we get

$$0 = -2x_1 + (-2+3i)x_2 + x_3 = -2x_1 + (-2+3i)ix_3 + x_3 = -2x_1 - (2+2i)x_3.$$

Therefore $x_1 = -(1+i)x_3$, and we get $E_{3i}^{\mathbb{C}} = \mathbb{C}(-1-i, i, 1)$. We get

$$E_3 = \mathbb{R}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad E_{3i}^{\mathbb{C}} = \mathbb{C}\begin{pmatrix} -1-i \\ i \\ 1 \end{pmatrix}, \quad E_{-3i}^{\mathbb{C}} = \mathbb{C}\begin{pmatrix} -1+i \\ -i \\ 1 \end{pmatrix},$$

and the complex diagonalisation

$$A = \begin{pmatrix} 1 & -1-i & -1+i \\ 1 & i & -i \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 3 & 0 & 0 \\ 0 & 3i & 0 \\ 0 & 0 & -3i \end{pmatrix} \begin{pmatrix} 1 & -1-i & -1+i \\ 1 & i & -i \\ 1 & 1 & 1 \end{pmatrix}^{-1}$$

To translate the complex diagonalisation to the real version, we notice $\lambda = 3i$ means $\mu = 0$ and $\nu = 3$, and we separate the real and imaginary parts of the eigenvector

$$\vec{v} = \begin{pmatrix} -1-i \\ i \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} - i \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \quad \vec{u} = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \quad \vec{w} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}.$$

Correspondingly, we get the real diagonalisation

$$A = \begin{pmatrix} 1 & -1 & 1 \\ 1 & 0 & -1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 3 & 0 & 0 \\ 0 & 0 & -3 \\ 0 & 3 & 0 \end{pmatrix} \begin{pmatrix} 1 & -1 & 1 \\ 1 & 0 & -1 \\ 1 & 1 & 0 \end{pmatrix}^{-1}.$$

Geometrically, the linear operator fixes $(1,1,1)$, “rotate” $\{(-1,0,1),(1,-1,0)\}$ by 90°, and then scales the whole space by 3.

We remark that the rotation is an authentic rotation only if $\vec{u} \perp \vec{w}$ and $\|\vec{u}\| = \|\vec{w}\|$. In general, it is a rotation by pretending $\{\vec{u}, \vec{w}\}$ to be an orthonormal set. In fact, the angle between $(-1,0,1)$ and $(1,-1,0)$ is $\frac{2\pi}{3}$, and the “rotation” by 90° with respect to $\{(-1,0,1),(1,-1,0)\}$ looks like the right of Figure 7.1.2.

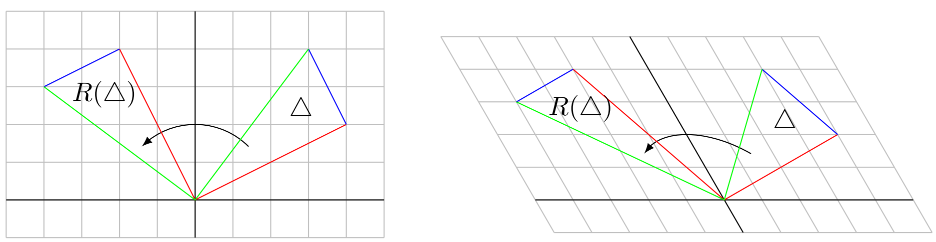

Figure 7.1.2: Authentic and fake rotations by 90°.

Exercise 7.22. Suppose A is a 2×2 real matrix, such that $(1+2i, 3-4i)$ is an eigenvector of eigenvalue $5+6i$. Find A.

Exercise 7.23. Convert complex diagonalisations to real diagonalisations.

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

2. $\begin{pmatrix} 1+2i & 5 & 1-2i \\ 3 & 6 & 3 \\ 4i & 7 & -4i \end{pmatrix} \begin{pmatrix} 1+i & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1-i \end{pmatrix} \begin{pmatrix} 1+2i & 5 & 1-2i \\ 3 & 6 & 3 \\ 4i & 7 & -4i \end{pmatrix}^{-1}$.

需要把u,v全部找出来

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

Exercise 7.24. Convert real diagonalisations to complex diagonalisations.

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

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

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

7.1.6 Application

Example 7.1.12. To find the general formula for the Fibonacci numbers, we introduce

$$\vec{x}_n = \begin{pmatrix} F_n \\ F_{n+1} \end{pmatrix}, \quad \vec{x}_0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \vec{x}_{n+1} = \begin{pmatrix} F_{n+1} \\ F_{n+1} + F_n \end{pmatrix} = A\vec{x}_n, \quad A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}.$$

Then $\vec{x}_n = A^n\vec{x}_0$ is obtained by repeatedly applying A to $\vec{x}_0$, and the n-th Fibonacci number F_n is the first coordinate of $\vec{x}_n$.

By solving $\det(tI - A) = t^2 - t - 1 = 0$, we get two eigenvalues $\frac{1\pm\sqrt{5}}{2}$. The first eigenspace

$$E_{\frac{1+\sqrt{5}}{2}} = \mathrm{Nul}(\tfrac{1+\sqrt{5}}{2}I - A) = \mathrm{Nul}\begin{pmatrix} \frac{1+\sqrt{5}}{2} & -1 \\ -1 & -\frac{1-\sqrt{5}}{2} \end{pmatrix} = \mathbb{R}\begin{pmatrix} 1 \\ \frac{1+\sqrt{5}}{2} \end{pmatrix}.$$

Substituting $\sqrt{5}$ by $-\sqrt{5}$, we get the second eigenspace

$$E_{\frac{1-\sqrt{5}}{2}} = \mathbb{R}\begin{pmatrix} 1 \\ \frac{1-\sqrt{5}}{2} \end{pmatrix}.$$

To find $\vec{x}_n$, we decompose $\vec{x}_0$ according to $\mathbb{R}^2 = E_{\frac{1+\sqrt{5}}{2}} \oplus E_{\frac{1-\sqrt{5}}{2}}$,

$$\vec{x}_0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \tfrac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ \frac{1+\sqrt{5}}{2} \end{pmatrix} - \tfrac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ \frac{1-\sqrt{5}}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{\sqrt{5}+1}{2\sqrt{5}} \end{pmatrix} + \begin{pmatrix} -\frac{1}{\sqrt{5}} \\ \frac{\sqrt{5}-1}{2\sqrt{5}} \end{pmatrix}.$$

Then we get

$$\vec{x}_n = A^n\vec{x}_0 = \left(\tfrac{1+\sqrt{5}}{2}\right)^n \begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{\sqrt{5}+1}{2\sqrt{5}} \end{pmatrix} + \left(\tfrac{1-\sqrt{5}}{2}\right)^n \begin{pmatrix} -\frac{1}{\sqrt{5}} \\ \frac{\sqrt{5}-1}{2\sqrt{5}} \end{pmatrix}.$$

Taking the first coordinate, we get

$$F_n = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right].$$

Exercise 7.25. Find the general formula for the Fibonacci numbers that start with $F_0 = 1$, $F_1 = 0$.

Exercise 7.26. Given the recursive relations and initial values. Find the general formula.

1. $x_n = x_{n-1} + 2y_{n-1}$, $y_n = 2x_{n-1} + 3y_{n-1}$, $x_0 = 1$, $y_0 = 0$.

2. $x_n = x_{n-1} + 2y_{n-1}$, $y_n = 2x_{n-1} + 3y_{n-1}$, $x_0 = 0$, $y_0 = 1$.

3. $x_n = y_{n-1} + z_{n-1}$, $y_n = -x_{n-1} - 2y_{n-1} - z_{n-1}$, $z_n = x_{n-1} + y_{n-1}$, $x_0 = 1$, $y_0 = 2$, $z_0 = 3$.

Example 7.1.13. Let L be the rotation of $\mathbb{R}^3$ around $\vec{v} = (1, 1, 1)$, by the right hand rule. In Example 5.3.3, we obtained an orthogonal basis $\vec{u} = (1, -1, 0), \vec{w} = (1, 1, -2)$ of $(\mathbb{R}\vec{v})^{\perp}$. By

$$\det(\vec{u}\ \vec{w}\ \vec{v}) = \det\begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & 1 \\ 0 & -2 & 1 \end{pmatrix} = 6 > 0,$$

the rotation from $\vec{u}$ direction to $\vec{w}$ direction is compatible with $\vec{v}$ by the right hand rule. Since $\|\vec{v}\| = \sqrt{2}$, $\|\vec{w}\| = \sqrt{6}$, and the rotating preserves the length, we know that L actually satisfies

$$L(\sqrt{3}\vec{u}) = \vec{w}, \quad L(\vec{w}) = -\sqrt{3}\vec{u}.$$

Combined with $L(\vec{v}) = \vec{v}$, we get the matrix of L (taking $P = (\sqrt{3}\vec{u}\ \vec{w}\ \vec{v})$)

$$\begin{aligned} &\begin{pmatrix} \sqrt{3} & 1 & 1 \\ -\sqrt{3} & 1 & 1 \\ 0 & -2 & 1 \end{pmatrix} \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{3} & 1 & 1 \\ -\sqrt{3} & 1 & 1 \\ 0 & -2 & 1 \end{pmatrix}^{-1} \\ &= \frac{1}{3}\begin{pmatrix} 1 & 1-\sqrt{3} & 1+\sqrt{3} \\ 1+\sqrt{3} & 1 & 1-\sqrt{3} \\ 1-\sqrt{3} & 1+\sqrt{3} & 1 \end{pmatrix}. \end{aligned}$$

From the meaning of the linear operator, we also know that the 4-th power of the matrix is the identity.

Exercise 7.27. What is the complex diagonalisation of the linear operator in Example 7.1.13?

Exercise 7.28. In Example 7.1.13, we change the rotation angle to $\frac{2\pi}{3}$. Writing your answer as complex and real diagonalisations is enough (although the matrix is extremely simple).

7.2 Symmetric Matrix

7.2.1 Orthogonal Diagonalisation

In the presence of dot product, we may wish to get better diagonalisation, by requiring the eigenspaces to be orthogonal to each other

$$E_{\lambda_1} \perp E_{\lambda_2} \perp \cdots \perp E_{\lambda_k} = \mathbb{R}^n.$$

By combining orthogonal bases of E_{λ_i}, this means that the matrix has an orthogonal (or even orthonormal) basis of eigenvectors. In this case, the matrix is *orthogonally diagonalisable*.

If A has an orthogonal basis of eigenvectors, then $A = QDQ^{-1}$, where the columns of Q form an orthonormal basis. Therefore Q is an orthogonal matrix, and satisfies $Q^{-1} = Q^T$. The expression $A = QDQ^{-1} = QDQ^T$ is an *orthogonal diagonalisation*, and we have

$$A^T = (QDQ^T)^T = QD^TQ^T = QDQ^T = A.$$

We get the necessary part of the next result.

Theorem 7.2.1. *A (real) matrix has orthogonal diagonalisation if and only if it is symmetric.*

Example 7.2.1. The matrix $\begin{pmatrix} 13 & -4 \\ -4 & 7 \end{pmatrix}$ in Examples 7.1.2 and 7.1.6 is a real symmetric matrix. Even without any calculation, we know the matrix has an orthogonal basis of eigenvectors. Indeed, the calculation in Example 7.1.6 gives the orthogonal diagonalisation

$$\begin{pmatrix} 13 & -4 \\ -4 & 7 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} 5 & 0 \\ 0 & 15 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}^{-1}.$$

Example 7.2.2. The matrix in Example 7.1.7 is symmetric. However, the basis of eigenvectors obtained in the earlier example is not orthogonal, because we did not try to find an orthogonal basis of the eigenspace $E_5 = \mathbb{R}(-1,2,0) \oplus \mathbb{R}(-1,0,1)$. If we choose an orthogonal basis $E_5 = \mathbb{R}(1,-4,1) \perp \mathbb{R}(-1,0,1)$, then we get an orthogonal basis of eignvectors $(1,-4,1), (-1,0,1), (2,1,2)$, and the orthogonal diagonalisation

$$\begin{pmatrix} 1 & -2 & -4 \\ -2 & 4 & -2 \\ -4 & -2 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{18}} & -\frac{1}{\sqrt{2}} & \frac{2}{3} \\ -\frac{4}{\sqrt{18}} & 0 & \frac{1}{3} \\ \frac{1}{\sqrt{18}} & \frac{1}{\sqrt{2}} & \frac{2}{3} \end{pmatrix} \begin{pmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & -4 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{18}} & -\frac{1}{\sqrt{2}} & \frac{2}{3} \\ -\frac{4}{\sqrt{18}} & 0 & \frac{1}{3} \\ \frac{1}{\sqrt{18}} & \frac{1}{\sqrt{2}} & \frac{2}{3} \end{pmatrix}^{-1}.$$

The orthogonality between eigenspaces of symmetric matrices is automatic, similar to Proposition 7.1.5.

Proposition 7.2.2. *All eigenvalues of real symmetric matrix are real, and eigenspaces of distinct eigenvalues are orthogonal to each other.*

The argument about real eigenvalue requires complex dot product, and is a special case of Proposition 7.3.3. We argue that $E_\lambda \perp E_\mu$ for distinct eigenvalues λ and μ.

Let $\vec{x} \in E_\lambda$ and $\vec{y} \in E_\mu$. Then $A\vec{x} = \lambda\vec{x}$, $A\vec{y} = \mu\vec{y}$, and

$$\lambda(\vec{x} \cdot \vec{y}) = A\vec{x} \cdot \vec{y} = \vec{x} \cdot A\vec{y} = \mu(\vec{x} \cdot \vec{y}).$$

Since $\lambda \neq \mu$, this implies $\vec{x} \cdot \vec{y} = 0$.

Exercise 7.29. The eigenvalues of real symmetric matrices are given. Find orthogonal diagonalisations.

1. $\begin{pmatrix} 7 & -2 \\ -2 & 4 \end{pmatrix}$, $\lambda_1 = 3$, $\lambda_2 = 8$.
2. $\begin{pmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{pmatrix}$, $\lambda_1 = 7$, $\lambda_2 = -2$.
3. $\begin{pmatrix} 9 & 4 \\ 4 & 3 \end{pmatrix}$, $\lambda_1 = 1$, $\lambda_2 = 11$.
4. $\begin{pmatrix} 1 & 1 & 5 \\ 1 & 5 & 1 \\ 5 & 1 & 1 \end{pmatrix}$, $\lambda_1 = -4$, $\lambda_2 = 4$, $\lambda_3 = 7$.

7.2.2 Quadratic Form

A *quadratic form* is a purely second order function:

$$\begin{aligned} q(x) &= ax^2, \\ q(x, y) &= ax^2 + by^2 + 2cxy = (x\ y) \begin{pmatrix} a & c \\ c & b \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \\ q(x_1, x_2, x_3) &= a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2 + 2a_{12}x_1x_2 + 2a_{13}x_1x_3 + 2a_{23}x_2x_3 \\ &= (x_1\ x_2\ x_3) \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}. \end{aligned}$$

Here we have $a_{ij} = a_{ji}$ in the 3×3 matrix. In general, a quadratic form corresponds to a symmetric matrix

$$q(\vec{x}) = \sum_{1 \le i,j \le n} a_{ij}x_ix_j = \vec{x}^T A\vec{x}, \quad a_{ij} = a_{ji},\ A^T = A.$$

Exercise 7.30. Write down the quadratic forms corresponding to the symmetric matrices in Exercise 5.49.

The symmetric matrix has an orthogonal diagonalisation $A = QDQ^{-1} = QDQ^T$. If we introduce the change of variable $\vec{x} = Q\vec{y}$, then $\vec{y} = Q^{-1}\vec{x} = Q^T\vec{x}$, and we get

$$\begin{aligned} q(\vec{x}) &= \vec{x}^T A\vec{x} = \vec{x}^T QDQ^T\vec{x} = (Q^T\vec{x})^T D(Q^T\vec{x}) \\ &= \vec{y}^T D\vec{y} = d_1y_1^2 + d_2y_2^2 + \cdots + d_ny_n^2. \end{aligned}$$

The result has no cross terms. You may compare with the process of the completion of square in Section 5.5.2, where we also eliminate the cross terms by change of variables given by upper triangular matrices.

Example 7.2.3. The symmetric matrix in Examples 7.1.2, 7.1.6, 7.2.1 gives a quadratic form

$$q(x_1, x_2) = 13x_1^2 + 7x_2^2 - 8x_1x_2.$$

By introducing

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \quad \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

we get

$$q = 5y_1^2 + 15y_2^2.$$

Example 7.2.4. The symmetric matrix in Examples 7.1.7, 7.3.1 gives a quadratic form

$$q(x_1, x_2, x_3) = x_1^2 + 4x_2^2 + x_3^2 - 4x_1x_2 - 8x_1x_3 - 4x_2x_3.$$

By introducing

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{18}} & -\frac{1}{\sqrt{2}} & \frac{2}{3} \\ -\frac{4}{\sqrt{18}} & 0 & \frac{1}{3} \\ \frac{1}{\sqrt{18}} & \frac{1}{\sqrt{2}} & \frac{2}{3} \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}, \quad \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{18}} & -\frac{4}{\sqrt{18}} & \frac{1}{\sqrt{18}} \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{2}{3} & \frac{1}{3} & \frac{2}{3} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix},$$

we get

$$q = 5y_1^2 + 5y_2^2 - 4y_3^2.$$

In Section 5.5.2, we determine whether a symmetric matrix is positive definite after eliminating the cross terms. If we use orthogonal diagonalisation to eliminate the cross terms, then the coefficients of the square terms are exactly the eigenvalues.

Proposition 7.2.3. *Suppose A is a symmetric matrix.*

1. *A is positive definite if and only if all the eigenvalues are > 0.*

2. *A is negative definite if and only if all the eigenvalues are < 0.*

3. *A is indefinite if and only if some eigenvalues is > 0 and some eigenvalues is < 0.*

Exercise 7.31. Use the orthogonal diagonalisations in Exercise 7.29 to simplify quadratic forms.

Exercise 7.32. Use Proposition 7.2.3 to determine the conditions on a, such that $A = \begin{pmatrix} 1 & 2 \\ 2 & a \end{pmatrix}$ is positive definite, negative definite, or indefinite.

Exercise 7.33. Suppose A is positive definite, explain that there is a positive definite B, such that $A = B^2 = B^T B$. This means that we can take "square root" of positive definite matrices. Then find the square roots of positive definite matrices in Exercise 7.29.

7.2.3 Singular Value Decomposition

In Section 5.4.2, we saw that the essential part of any linear transformation $L\colon \mathbb{R}^n \to \mathbb{R}^m$ is the isomorphism $L_{11}\colon \text{Ran}L^* \to \text{Ran}L$. The isomorphism is the restriction of L to the range of of the adjoint linear transformation, or the orthogonal complement $(\text{Ker}L)^\perp$ of the kernel of L. The *singular value decomposition* is a further refinement of this isomorphism. It produces orthonormal bases $\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_r\}$ and $\{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_r\}$, such that

$$L(\vec{v}_i) = \sigma_i \vec{w}_i, \quad \sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_r > 0.$$

To get the matrix of the whole L, we may extend the orthonormal bases to orthogonal bases $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_r, \vec{v}_{r+1}, \dots, \vec{v}_n\}$ and $\beta = \{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_r, \vec{w}_{r+1}, \dots, \vec{w}_m\}$ of $\mathbb{R}^n$ and $\mathbb{R}^m$. The extensions are actually obtained by adding orthonormal basis $\{\vec{v}_{r+1}, \dots, \vec{v}_n\}$ of $\text{Ker}L$ and orthonormal basis $\{\vec{w}_{r+1}, \dots, \vec{w}_m\}$ of $\text{Ker}L^*$. Then the matrix of L with respect to the orthonormal bases is

$$[L]_{\beta\alpha} = \begin{pmatrix} \Sigma & O \\ O & O \end{pmatrix}_{m\times n}, \quad \Sigma = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{pmatrix}.$$

If A is the standard matrix of L, i.e., with respect to the standard bases ϵ, then

$$A = [L]_{\epsilon\epsilon} = [I]_{\epsilon\beta}[L]_{\beta\alpha}[I]_{\alpha\epsilon} = (\beta) \begin{pmatrix} \Sigma & O \\ O & O \end{pmatrix} (\alpha)^{-1} = W \begin{pmatrix} \Sigma & O \\ O & O \end{pmatrix}_{m\times n} V^{-1}.$$

Here

$$V = (\alpha) = (\vec{v}_1 \; \vec{v}_2 \; \cdots \; \vec{v}_n), \quad W = (\beta) = (\vec{w}_1 \; \vec{w}_2 \; \cdots \; \vec{w}_m)$$

are orthogonal matrices. The expression above is the singular value decomposition of A. It easily gives the pseudoinverse of A

$$V \begin{pmatrix} \Sigma^{-1} & O \\ O & O \end{pmatrix}_{n\times m} W^{-1}.$$

Next we construct the singular value decomposition. Let A be an $m \times n$ matrix. Then $A^T A$ is a symmetric $n \times n$ matrix. By Theorem 7.2.1, $A^T A$ has an orthonormal basis of eigenvectors $\alpha = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$. By rearranging the orders of the vectors, we may assume the corresponding eigenvalues are

$$d_1, d_2, \dots, d_r \neq 0, \qquad \text{and } d_{r+1} = d_{r+2} = \cdots = d_n = 0.$$

For $i \neq j$, we have

$$\begin{aligned} A\vec{v}_i \cdot A\vec{v}_j &= \vec{v}_i \cdot A^T A\vec{v}_j = \vec{v}_i \cdot d_j \vec{v}_j = d_j(\vec{v}_i \cdot \vec{v}_j) = 0, \\ \|A\vec{v}_i\|^2 = A\vec{v}_i \cdot A\vec{v}_i &= \vec{v}_i \cdot A^T A\vec{v}_i = \vec{v}_i \cdot d_i \vec{v}_i = d_i \|\vec{v}_i\|^2 = d_i. \end{aligned}$$

The first equality means $A\vec{v}_i \perp A\vec{v}_j$. The second equality implies

$$\|A\vec{v}_i\| = \begin{cases} \sigma_i = \sqrt{d_i} \neq 0, & \text{if } i \leq r, \\ 0, & \text{if } i > r. \end{cases}$$

Then

$$A\vec{v}_{r+1} = A\vec{v}_{r+2} = \cdots = A\vec{v}_n = \vec{0},$$

and $\mathrm{Col}A$ is spanned by nonzero vectors $A\vec{v}_1, A\vec{v}_2, \dots, A\vec{v}_r$. By $A\vec{v}_i \perp A\vec{v}_j$, we know

$$\vec{w}_1 = \frac{A\vec{v}_1}{\sigma_1},\ \vec{w}_2 = \frac{A\vec{v}_2}{\sigma_2},\ \dots,\ \vec{w}_r = \frac{A\vec{v}_r}{\sigma_r}$$

is an orthonormal basis of $\mathrm{Col}A$. Therefore $r = \dim \mathrm{Col}A = \mathrm{rank}A$.

Let $\{\vec{w}_{r+1}, \dots, \vec{w}_m\}$ be an orthonormal basis of $\mathrm{Nul}A^T$. Then we get an orthonormal basis $\beta = \{\vec{w}_1, \vec{w}_2, \dots, \vec{w}_r, \vec{w}_{r+1}, \dots, \vec{w}_m\}$ of $\mathrm{Col}A \perp \mathrm{Nul}A^T = \mathbb{R}^m$. We conclude

$$A\vec{v}_1 = \sigma_1 \vec{w}_1,\ A\vec{v}_2 = \sigma_2 \vec{w}_2,\ \dots,\ A\vec{v}_r = \sigma_r \vec{w}_r, \quad A\vec{v}_{r+1} = A\vec{v}_{r+2} = \cdots = A\vec{v}_n = \vec{0}.$$

This further gives the singular value decomposition $A = W \begin{pmatrix} \Sigma & O \\ O & O \end{pmatrix} V^{-1}$.

Example 7.2.5. Consider the matrix in Example 5.4.2

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

We have

$$A^T A = \begin{pmatrix} 2 & -1 & -1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix},$$

and

$$\begin{aligned}\det(tI - A^TA) &= \det\begin{pmatrix} t-2 & 1 & 1 \\ 1 & t-1 & 0 \\ 1 & 0 & t-1 \end{pmatrix} = \det\begin{pmatrix} 0 & 1 & 1-(t-1)(t-2) \\ 0 & t-1 & -t+1 \\ 1 & 0 & t-1 \end{pmatrix} \\ &= \det\begin{pmatrix} 1 & 1-(t-1)(t-2) \\ t-1 & -t+1 \end{pmatrix} = (t-1)\det\begin{pmatrix} 1 & -t^2+3t-1 \\ 1 & -1 \end{pmatrix} \\ &= (t-3)(t-1)t.\end{aligned}$$

Then we get eigenspaces

$$\mathrm{Nul}(3I - A^TA) = \mathbb{R}\begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix}, \quad \mathrm{Nul}(1I - A^TA) = \mathbb{R}\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \quad \mathrm{Nul}(0I - A^TA) = \mathbb{R}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

This leads to

$$\vec{v}_1 = \frac{1}{\sqrt{6}}\begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix}, \quad \vec{v}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \quad \vec{v}_3 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix},$$

and

$$\vec{w}_1 = \frac{A\vec{v}_1}{\sqrt{3}} = \frac{1}{\sqrt{3}\sqrt{6}}\begin{pmatrix} 3 \\ 3 \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \vec{w}_2 = \frac{A\vec{v}_2}{\sqrt{1}} = \frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Then we get the singular value decomposition

$$\begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}\begin{pmatrix} \sqrt{3} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{pmatrix}^{-1}.$$

Example 7.2.6. Consider the matrix in Example 5.4.3

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

We have

$$A^TA = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \quad \det(t - A^TA) = \det\begin{pmatrix} t-2 & -1 \\ -1 & t-2 \end{pmatrix} = (t-1)(t-3).$$

Then we get eigenspaces

$$\mathrm{Nul}(3I - A^TA) = \mathbb{R}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \mathrm{Nul}(1I - A^TA) = \mathbb{R}\begin{pmatrix} -1 \\ 1 \end{pmatrix},$$

and

$$\vec{v}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix},\ \vec{v}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix}-1\\1\end{pmatrix},\quad \vec{w}_1 = \frac{A\vec{v}_1}{\sqrt{3}} = \frac{1}{\sqrt{6}}\begin{pmatrix}2\\-1\\-1\end{pmatrix},\ \vec{w}_2 = \frac{A\vec{v}_1}{\sqrt{1}} = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\1\\-1\end{pmatrix}.$$

We extend $\{\vec{w}_1, \vec{w}_2\}$ to an orthonormal basis of $\mathbb{R}^3$ by taking $\vec{w}_3 = \frac{1}{\sqrt{3}}(1,1,1)$. Then we get the singular value decomposition

$$\begin{pmatrix}1 & 1\\ -1 & 0\\ 0 & -1\end{pmatrix} = \begin{pmatrix}\frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}}\\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}}\\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}}\end{pmatrix}\begin{pmatrix}\sqrt{3} & 0\\ 0 & 1\\ 0 & 0\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{pmatrix}^{-1}.$$

Exercise 7.34. Consider a nonzero matrix $\vec{a}$ as an $n \times 1$ matrix. What are the singular values decompositions of the matrix $\vec{a}$ and the matrix $\vec{a}^T$?

Exercise 7.35. Find the singular value decompositions of the matrices.

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

Example 7.2.7. The singular value decomposition of the matrix in Example 5.4.4 is quite complicated. We have

$$A = \begin{pmatrix}1 & 4 & 7 & 10\\ 2 & 5 & 8 & 11\\ 3 & 6 & 9 & 12\end{pmatrix},\quad A^TA = \begin{pmatrix}14 & 32 & 50 & 68\\ 32 & 77 & 122 & 167\\ 50 & 122 & 194 & 266\\ 68 & 167 & 266 & 365\end{pmatrix}.$$

The eigenvalues of A^TA are

$$d_1 = 325 + \sqrt{104545},\ d_2 = 325 - \sqrt{104545},\ d_3 = d_4 = 0.$$

Then

$$\Sigma = \begin{pmatrix}\sqrt{325+\sqrt{104545}} & 0\\ 0 & \sqrt{325-\sqrt{104545}}\end{pmatrix} \approx \begin{pmatrix}25.462407 & 0\\ 0 & 1.290662\end{pmatrix}.$$

The singular value decomposition has

$$V \approx \begin{pmatrix}0.140877 & 0.824714 & 0.534522 & -0.119523\\ 0.343946 & 0.426264 & -0.801784 & -0.239046\\ 0.547016 & 0.027814 & 0 & 0.83666\\ 0.750086 & -0.370637 & 0.267261 & -0.478091\end{pmatrix},$$

$$W \approx \begin{pmatrix}0.504533 & -0.760776 & 0.408248\\ 0.574516 & -0.057141 & -0.816497\\ 0.64498 & 0.646495 & 0.408248\end{pmatrix}.$$

7.3 Complex Orthogonal Diagonalisation

7.3.1 Normal Matrix

For a complex $n \times n$ matrix A, we wish to have complex orthogonal diagonalisation

$$E^{\mathbb{C}}_{\lambda_1} \perp E^{\mathbb{C}}_{\lambda_2} \perp \cdots \perp E^{\mathbb{C}}_{\lambda_k} = \mathbb{C}^n.$$

Here $E^{\mathbb{C}}_{\lambda} = \text{Nul}^{\mathbb{C}}(\lambda I - A)$ is the complex eigenspace. By combining orthonormal bases of $E^{\mathbb{C}}_{\lambda_i}$, we get an orthonormal basis of eigenvectors. The eigenvectors form the columns of a unitary matrix U, and we have the *complex orthogonal diagonalisation* $A = UDU^{-1}$. Then $U^{-1} = U^*$, and

$$\begin{aligned} A^* &= (UDU^*)^* = UD^*U^*, \\ A^*A &= UD^*U^*UDU^* = UD^*DU^*, \\ AA^* &= UDU^*UD^*U^* = UDD^*U^*. \end{aligned}$$

Since D and D^* are diagonal, we have $D^*D = DD^*$. Then we get the necessary part of the next result.

Theorem 7.3.1. *A (complex) matrix A has complex orthogonal diagonalisation if and only if it satisfies $A^*A = AA^*$.*

We call a matrix satisfying $A^*A = AA^*$ a *normal matrix*. Similarly, a linear operator L on a complex inner product space is normal if and only if $L^*L = LL^*$, and this is the necessary and sufficient condition for L to have an (complex) orthogonal basis of eigenvectors.

A real matrix A is normal if $A^TA = AA^T$. It has orthogonal basis of complex eigenvectors, say

$$\vec{v}_1,\ \vec{v}_2,\ \vec{u}_1 - i\vec{w}_1,\ \vec{u}_1 + i\vec{w}_1,\ \vec{u}_2 - i\vec{w}_2,\ \vec{u}_2 + i\vec{w}_2.$$

Here $\vec{v}_1, \vec{v}_2$ have real eigenvalues d_1, d_2, and $\vec{u}_1 \mp i\vec{w}_1, \vec{u}_2 \mp i\vec{w}_2$ have conjugate pairs of complex eigenvalues $a_1 \pm ib_1, a_2 \pm ib_2$. By the discussion in Section 7.1.5, the conjugate pairs give $(j = 1, 2)$

$$\begin{aligned} A\vec{u}_j &= a\vec{u}_j + b\vec{w}_j, \\ A\vec{w}_j &= -b\vec{u}_j + a\vec{w}_j. \end{aligned}$$

Then the real matrix A is decomposed according to

$$\mathbb{R}^6 = \mathbb{R}\vec{v}_1 \oplus \mathbb{R}\vec{v}_2 \oplus (\mathbb{R}\vec{u}_1 \oplus \mathbb{R}\vec{w}_1) \oplus (\mathbb{R}\vec{u}_2 \oplus \mathbb{R}\vec{w}_2).$$

On $\mathbb{R}\vec{v}_1$ and $\mathbb{R}\vec{v}_2$, $A\vec{x}$ is multiplying a real number to $\vec{x}$. On $\mathbb{R}\vec{u}_1 \oplus \mathbb{R}\vec{w}_1$ and $\mathbb{R}\vec{u}_2 \oplus \mathbb{R}\vec{w}_2$, $A\vec{x}$ is rotation of $\vec{x}$ and then multiplying a real number.

Next we discussion the meaning of complex orthogonality

$$\mathbb{C}^6 = \mathbb{C}\vec{v}_1 \perp \mathbb{C}\vec{v}_2 \perp \mathbb{C}(\vec{u}_1 - i\vec{w}_1) \perp \mathbb{C}(\vec{u}_1 + i\vec{w}_1) \perp \mathbb{C}(\vec{u}_2 - i\vec{w}_2) \perp \mathbb{C}(\vec{u}_2 + i\vec{w}_2).$$

By

$$\mathbb{C}(\vec{u}_1 - i\vec{w}_1) \oplus \mathbb{C}(\vec{u}_1 + i\vec{w}_1) = \mathbb{C}\vec{u}_1 \oplus \mathbb{C}\vec{w}_1,$$

we get complex orthogonality among complex spans of real vectors

$$\mathbb{C}^6 = \mathbb{C}\vec{v}_1 \perp \mathbb{C}\vec{v}_2 \perp (\mathbb{C}\vec{u}_1 \oplus \mathbb{C}\vec{w}_1) \perp (\mathbb{C}\vec{u}_2 \oplus \mathbb{C}\vec{w}_2).$$

Since the vectors $\vec{v}_j, \vec{u}_j, \vec{w}_j$ are real, and the complex dot product of real vectors is the same as the real dot product of the same real vectors, the complex orthogonality above is actually real orthogonality

$$\mathbb{R}^6 = \mathbb{R}\vec{v}_1 \perp \mathbb{R}\vec{v}_2 \perp (\mathbb{R}\vec{u}_1 \oplus \mathbb{R}\vec{w}_1) \perp (\mathbb{R}\vec{u}_2 \oplus \mathbb{R}\vec{w}_2).$$

Within the subspace $\mathbb{R}\vec{u}_1 \oplus \mathbb{R}\vec{w}_1$, the complex orthogonality $\vec{u}_j - i\vec{w}_j \perp \vec{u}_j + i\vec{w}_j$ means

$$\begin{aligned} 0 &= (\vec{u}_j - i\vec{w}_j) \cdot (\vec{u}_j + i\vec{w}_j) \\ &= \vec{u}_j \cdot \vec{u}_j - i\vec{w}_j \cdot \vec{u}_j + \vec{u}_j \cdot i\vec{w}_j - i\vec{w}_j \cdot i\vec{w}_j \\ &= \vec{u}_j \cdot \vec{u}_j - i(-i)\vec{w}_j \cdot \vec{w}_j - i\vec{w}_j \cdot \vec{u}_j - i\vec{u}_j \cdot i\vec{w}_j \\ &= \|\vec{u}_j\|^2 - \|\vec{w}_j\|^2 - 2i\vec{u}_j \cdot \vec{w}_j. \end{aligned}$$

This is the same as

$$\|\vec{u}_j\| = \|\vec{w}_j\| \text{ and } \vec{u}_j \cdot \vec{w}_j = 0.$$

This implies that

$$[A|_{\mathbb{R}\vec{u}_j \perp \mathbb{R}\vec{w}_j}] = \begin{pmatrix} a_j & -b_j \\ b_j & a_j \end{pmatrix} = r_j R_{\theta_j}$$

is the combination of authentic rotation by θ_j and scaling by r_j. Moreover, by dividing the lengths, we may assume the vectors $\vec{v}_1, \vec{v}_2, \vec{u}_1, \vec{w}_1, \vec{u}_2, \vec{w}_2$ form an orthonormal basis of $\mathbb{R}^6$, and the formula for $[A|_{\mathbb{R}\vec{u}_j \perp \mathbb{R}\vec{w}_j}]$ remain valid. In the result below, $Q = (\vec{v}_1\ \vec{v}_2\ \vec{u}_1\ \vec{w}_1\ \vec{u}_2\ \vec{w}_2)$ is an orthogonal matrix.

Proposition 7.3.2. *A real matrix A is normal: $A^TA = AA^T$, if and only if $A = QDQ^{-1}$ for an orthogonal matrix Q and a block diagonal matrix*

$$D = \begin{pmatrix} d_1 & & & & & & & \\ & \ddots & & & & & O & \\ & & d_s & & & & & \\ & & & a_1 & -b_1 & & & \\ & & & b_1 & a_1 & & & \\ & & & & & \ddots & & \\ & O & & & & & a_t & -b_t \\ & & & & & & b_t & a_t \end{pmatrix}.$$

7.3.2 Orthogonal Diagonalisation of Hermitian Matrix

A (complex) Hermitian matrix satisfies $A^* = A$. It is a normal matrix with further special property.

Theorem 7.3.3. *A square matrix is Hermitian if and only if it is orthogonally diagonalisable, and all eigenvalues are real.*

Let $A\vec{x} = \lambda\vec{x}$ for some $\vec{x} \neq \vec{0}$. Then

$$\begin{aligned} \vec{x} \cdot A\vec{x} &= \vec{x} \cdot \lambda\vec{x} = \bar{\lambda}(\vec{x} \cdot \vec{x}) = \bar{\lambda}\|\vec{x}\|^2, \\ A\vec{x} \cdot \vec{x} &= \lambda\vec{x} \cdot \vec{x} = \lambda(\vec{x} \cdot \vec{x}) = \lambda\|\vec{x}\|^2. \end{aligned}$$

If $A^* = A$, then the left sides are equal, and we get $\bar{\lambda}\|\vec{x}\|^2 = \lambda\|\vec{x}\|^2$. Since $\vec{x} \neq \vec{0}$, this implies $\bar{\lambda} = \lambda$, which means the eigenvalue λ is a real number.

Conversely, suppose $A = UDU^{-1} = UDU^*$, with all the eigenvalues being real. Then D is a real matrix, and we get $D^* = \overline{D^T} = \bar{D} = D$. This implies

$$A^* = UD^*U^* = UDU^* = A.$$

Example 7.3.1. The Hermitian matrix

$$A = \begin{pmatrix} 2 & 1+i \\ 1-i & 3 \end{pmatrix}$$

has characteristic polynomial

$$\det \begin{pmatrix} t-2 & -1-i \\ -1+i & t-3 \end{pmatrix} = (t-2)(t-3) - (1^2+1^2) = t^2 - 5t + 4 = (t-1)(t-4).$$

We have

$$\begin{aligned} 4I - A &= \begin{pmatrix} 2 & -1-i \\ -1+i & 1 \end{pmatrix}, & E_4^{\mathbb{C}} &= \mathrm{Nul}^{\mathbb{C}}(4I - A) = \mathbb{C}\begin{pmatrix} 1 \\ 1-i \end{pmatrix}, \\ I - A &= \begin{pmatrix} -1 & -1-i \\ -1-i & -2 \end{pmatrix}, & E_1^{\mathbb{C}} &= \mathrm{Nul}^{\mathbb{C}}(I - A) = \mathbb{C}\begin{pmatrix} 1+i \\ -1 \end{pmatrix}. \end{aligned}$$

By $\|(1+i, -1)\| = \sqrt{3} = \|(1, 1-i)\|$, we get the unitary diagonalisation

$$\begin{pmatrix} 2 & 1+i \\ 1-i & 3 \end{pmatrix} = \begin{pmatrix} \frac{1+i}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} & \frac{1-i}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} \frac{1+i}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} & \frac{1-i}{\sqrt{3}} \end{pmatrix}^{-1}.$$

7.3.3 Orthogonal Diagonalisation of Unitary Matrix

A unitary matrix U satisfies $U^*U = I = UU^*$. Therefore U is normal, and orthogonally diagonalisable. Moreover, if $U\vec{x} = \lambda\vec{x}$, then by U preserving the length, we get

$$\|\vec{x}\| = \|U\vec{x}\| = |\lambda|\|\vec{x}\|.$$

Therefore we get $|\lambda| = 1$.

Conversely, if A is orthogonally diagonalisable, and all eigenvalues of A satisfy $|\lambda| = 1$, then we get $A = VDV^{-1} = VDV^*$, where V is unitary. The matrix D is diagonal, with all the diagonal entries λ satisfying $|\lambda| = 1$. Then D^*D are diagonal, with diagonal entries $\bar{\lambda}\lambda = |\lambda|^2 = 1$. Therefore $D^*D = I$, and

$$A^*A = (VDV^*)^*(VDV^*) = (V^*)^*D^*V^*VDV^* = VD^*DV^* = VV^* = I.$$

This shows A is a unitary matrix.

Proposition 7.3.4. *A square matrix is unitary if and only if it is orthogonally diagonalisable, and all eigenvalues satisfy* $|\lambda| = 1$.

Example 7.3.2. In Example 7.3.1, we get the unitary matrix

$$U = \begin{pmatrix} \frac{1+i}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} & \frac{1-i}{\sqrt{3}} \end{pmatrix}.$$

We have

$$\det(tI - U) = t^2 - \tfrac{2}{\sqrt{3}}t + 1 = (t - \tfrac{1+\sqrt{2}i}{\sqrt{3}})(t - \tfrac{1-\sqrt{2}i}{\sqrt{3}}).$$

Then

$$E^{\mathbb{C}}_{\frac{1+\sqrt{2}i}{\sqrt{3}}} = \mathrm{Nul}^{\mathbb{C}}(\tfrac{1+\sqrt{2}i}{\sqrt{3}}I - U) = \mathbb{C}\begin{pmatrix} 1+\sqrt{2} \\ i \end{pmatrix},$$
$$E^{\mathbb{C}}_{\frac{1-\sqrt{2}i}{\sqrt{3}}} = \mathrm{Nul}^{\mathbb{C}}(\tfrac{1-\sqrt{2}i}{\sqrt{3}}I - U) = \mathbb{C}\begin{pmatrix} i \\ 1+\sqrt{2} \end{pmatrix}.$$

The two eigenvectors have length $\sqrt{(1+\sqrt{2})^2 + 1^2} = \sqrt{4+2\sqrt{2}}$. We have $\frac{1+\sqrt{2}}{\sqrt{4+2\sqrt{2}}} = \frac{1}{\sqrt{4-2\sqrt{2}}}$, and the orthogonal diagonalisation

$$U = \begin{pmatrix} \frac{1}{\sqrt{4-2\sqrt{2}}} & \frac{i}{\sqrt{4+2\sqrt{2}}} \\ \frac{i}{\sqrt{4+2\sqrt{2}}} & \frac{1}{\sqrt{4-2\sqrt{2}}} \end{pmatrix} \begin{pmatrix} \frac{1+\sqrt{2}i}{\sqrt{3}} & 0 \\ 0 & \frac{1-\sqrt{2}i}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{4-2\sqrt{2}}} & \frac{i}{\sqrt{4+2\sqrt{2}}} \\ \frac{i}{\sqrt{4+2\sqrt{2}}} & \frac{1}{\sqrt{4-2\sqrt{2}}} \end{pmatrix}^{-1}.$$

Exercise 7.36. For real number a, find the orthogonal diagonalisation of $\frac{1}{\sqrt{a^2+1}}\begin{pmatrix} a & i \\ i & a \end{pmatrix}$.

A (real) orthogonal matrix Q is also a unitary matrix. Its eigenvalues satisfy $|\lambda| = 1$. If λ is real, then $\lambda = \pm 1$. If λ is not real, then we actually have a conjugate pair $\lambda = e^{\pm i\theta}$ of non-real complex numbers of length 1. By Proposition 7.3.2, we have $Q = PDP^{-1}$, where P is orthogonal, and

$$D = \begin{pmatrix} \ddots & & & & & & & \\ & 1 & & & & & & \\ & & \ddots & & & & O & \\ & & & -1 & & & & \\ & & & & \ddots & & & \\ & & & & & \cos\theta & -\sin\theta & \\ & O & & & & \sin\theta & \cos\theta & \\ & & & & & & & \ddots \end{pmatrix}$$

$$= \begin{pmatrix} I_{H_+} & & & & \\ & -I_{H_-} & & O & \\ & & R_{\theta_1} & & \\ & O & & \ddots & \\ & & & & R_{\theta_t} \end{pmatrix}, \quad R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Geometrically, we have the orthogonal sum decomposition

$$\mathbb{R}^n = H_+ \perp H_- \perp H_1 \perp H_2 \perp \cdots \perp H_t,$$

and

- Q is the identity: $Q\vec{x} = \vec{x}$, for $\vec{x} \in H_+$.
- Q is the flip: $Q\vec{x} = -\vec{x}$ for $\vec{x} \in H_-$.
- Q is a rotation on each of the 2-dimensional subspaces H_j.

Proposition 7.3.5. *An orthogonal matrix is the orthogonal sum of identities, reflections, and rotations (on planes).*

An orthogonal operator has the determinant $\det Q = (-1)^{\dim H_-}$. If Q preserves the orientation, then $\det Q = 1$. This means $\dim H_-$ is even, and $-I_{H_-}$ in the diagonal matrix D can be grouped into pairs and form rotations by π

$$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} \cos\pi & -\sin\pi \\ \sin\pi & \cos\pi \end{pmatrix}.$$

Therefore an orientation preserving orthogonal matrix is an orthogonal sum of identities and rotations.

Example 7.3.3. A 2×2 orthogonal matrix is either a rotation or a flip

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \quad F_\rho = \begin{pmatrix} \cos 2\rho & \sin 2\rho \\ -\sin 2\rho & \cos 2\rho \end{pmatrix}.$$

Example 7.3.4. Let Q be a 3×3 orthogonal matrix. If Q has non-real eigenvalue, then Q is ± 1 along a 1-dimensional subspace $\mathbb{R}\vec{v}$, and is a rotation on the 2-dimensional subspace $(\mathbb{R}\vec{v})^\perp$.

If $Q\vec{v} = \vec{v}$, then Q is the rotation of $\mathbb{R}^3$ around the axis $\mathbb{R}\vec{v}$. If $Q\vec{v} = -\vec{v}$, then Q is the rotation around the axis $\mathbb{R}\vec{v}$, followed by the flip with respect to the plane $(\mathbb{R}\vec{v})^\perp$.

If all eigenvalues of Q are real, then the eigenvalues are ± 1. If the eigenvalues are $1, 1, 1$, then $Q = I$. If the eigenvalues are $-1, -1, -1$, then $Q = -I$. If the eigenvalues are $1, -1, -1$, then $\mathrm{Nul}(1I - Q) = \mathbb{R}\vec{v}$ is a line, and Q is the rotation around the eigenspace by angle π. If the eigenvalues are $1, 1, -1$, then $\mathrm{Nul}(1I - Q)$ is a plane, and Q is the flip with respect to the eigenspace.

If Q preserves the orientation, then it is always a rotation around a line, called the *axis* of rotation.

Exercise 7.37. Find the 3×3 orthogonal matrix that rotates around $\mathbb{R}(-1, -1, -1)$ by $90°$, such that the direction of rotation is compatible with the direction $(-1, -1, -1)$ by the right hand rule.

Exercise 7.38. Suppose a 4×4 orthogonal matrix exchanges $(1, 1, 1, 1)$ and $(1, 1, -1, -1)$, and fixes $(1, -1, 1, -1)$. What can the orthogonal matrix be?