Matrix Algebra

Hao Zhuang

Summer 2022

Contents

1 Matrices and Equations **1**

1.1 The Motivation and Definition 1

1.2 Matrices Acting on Vectors 7

1.3 Solving Systems of Homogeneous Linear Equations 14

1.4 Recitation 26

1.5 Solving Non-homogeneous Systems 30

2 Matrix Algebra **35**

2.1 Matrix Operations 35

2.2 Partitioned Matrices 40

2.3 The Inverse of a Matrix 43

3 Linear Spaces **49**

3.1 Linear Spaces and Linear Maps 49

3.2 Linear Independence and Bases 53

3.3 Bases and Dimensions 59

3.4 Change of Basis 63

3.5 Coordinates 67

3.6 Images, Kernels and Ranks 75

4 Determinants **89**

4.1 Expansion Along Rows 89

4.2 Expansion Along Columns 96

4.3 A Nonrigorous Approach 100

5 Eigenspaces **107**

5.1 Eigenvalues and Eigenvectors 107

5.2 Diagonalization 115

5.3 Complex Eigenvalues 119

6 Inner Product **123**
6.1 Gram-Schmidt Process 123
6.2 Projections and Least-Squares Method 128

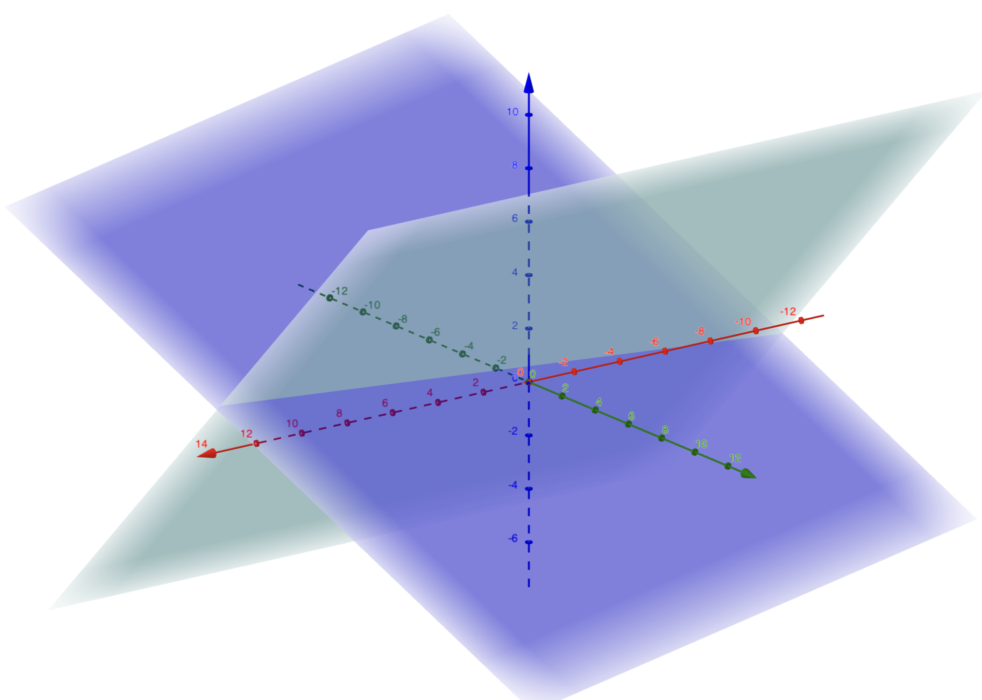
Chapter 1

Matrices and Equations

1.1 The Motivation and Definition

What is a matrix? Why do we need it? Well, before we plunge into its definition, let's recall what we learned in Calculus III.

Example 1.1.1. Given two planes $x_1 - 3x_2 - 5x_3 = 0$ and $x_2 - x_3 = -1$, find their intersection line. (See [3], page 21, practice problem 1.)



Solution. From $x_2 - x_3 = -1$ we get

$$x_2 = x_3 - 1.$$

Then we plug it into $x_1 - 3x_2 - 5x_3 = 0$ and get

$$x_1 = 8x_3 - 3.$$

Let $t = x_3$, we get the parameterized expression

$$(8t - 3, t - 1, t).$$

Another way is to use the cross product of normal vectors. The two normal vectors are $(1, -3, -5)$ and $(0, 1, -1)$. Their cross product is

$$\begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & -3 & -5 \\ 0 & 1 & -1 \end{vmatrix} = 8\mathbf{i} + \mathbf{j} + \mathbf{k} = (8, 1, 1).$$

Then we find a point $(-3, -1, 0)$ on the intersection and get the line

$$(-3, -1, 0) + t \cdot (8, 1, 1),$$

where t is a parameter. □

We can generalize this type of problems to "planes" of higher dimensions (they are called hyperplanes):

Remark 1.1.2. Given three hyperplanes

$$\begin{aligned} x_1 + x_2 + 3x_3 + 4x_4 - 7x_5 &= 0 \\ 2x_1 - 4x_2 + 2x_3 + 6x_4 - 3x_5 &= 1 \\ 15x_1 - 8x_2 + 3x_3 + 7x_4 - 10x_5 &= 9, \end{aligned}$$

find their intersection. What if the dimension is higher than 5? What if we have more than three hyperplanes?

As the dimension in the problem becomes higher and higher, it is certain that the cross product method is not applicable. Also, the elimination method gets more and more laborious. This is because when we try to express one of the unknowns with other unknowns and then eliminate it, all the coefficients and unknowns are entwined.

For example, in Remark 1.1.2, we get

$$x_1 = -x_2 - 3x_3 - 4x_4 - 7x_5$$

from the first equation. Then we plug it into the second and third equations... I believe everyone can imagine the workload.

However, we can separate the unknowns and the coefficients. Then we only need to deal with the coefficients. In Example 1.1.1, we have two equations

$$\begin{aligned} x_1 - 3x_2 - 5x_3 &= 0 \\ x_2 - x_3 &= -1. \end{aligned}$$

Then on the right-hand side, we get a list

$$\begin{bmatrix} 0 \\ -1 \end{bmatrix}.$$

On the left hand side, we take out coefficients and write them into a 2×3 array

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

Here you need to keep the position of each coefficient and be careful with their signs. Then we attach the list

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

of unknowns to the 2×3 array and get

$$\begin{bmatrix} 1 & -3 & -5 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}. \tag{1.1.1}$$

Example 1.1.3. Write the equations in Remark 1.1.2 like the equation (1.1.1).

Solution. We write the equation system into

$$\begin{bmatrix} 1 & 1 & 3 & 4 & -7 \\ 2 & -4 & 2 & 6 & -3 \\ 15 & -8 & 3 & 7 & -10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 9 \end{bmatrix}.$$

□

In general, we call equation systems like those in Example 1.1.1 and Remark 1.1.2 linear equation systems. We write them into arrays and lists in the following way:

Definition 1.1.4. For a linear equation system

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \cdots + a_{mn}x_n &= b_m, \end{aligned}$$

where a_{ij}'s and b_k's are numbers and x_l's are unknowns, we can write it into

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_m \end{bmatrix}$$

Now we are able to give the definition of matrix:

Definition 1.1.5. An $m \times n$ matrix is an $m \times n$ array

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix}$$

of real or complex numbers. Here a_{ij} means the entry at the i-th row and the j-th column.

Homework 1.1.6. Write the following two linear equation systems as in Definition 1.1.4.

(a) Let x_1, x_2, x_3, x_4 be unknowns:

$$\begin{aligned} 3x_1 + 5x_2 + 5x_3 + 7x_4 &= 8 \\ 8x_1 + 7x_2 + 3x_3 + 8x_4 &= 1 \\ 6x_2 + 7x_4 &= 4; \end{aligned}$$

(b) Let x, y, z, w be unknowns:

$$\begin{aligned} 2x + 3y + 5z + 7w &= 1 \\ 4x + 7y + 6z + 8w &= 2 \\ 9y + 2z + 10w &= 0. \end{aligned}$$

Remark 1.1.7. In Definition 1.1.4, if we switch the order of two unknowns, for example, x_2 and x_3, how do we adjust the $m \times n$ matrix? What if we switch the order of three unknowns?

Solution. If I switch x_2 and x_3, then the list of unknowns becomes

$$\begin{bmatrix} x_1 \\ x_3 \\ x_2 \\ x_4 \\ \vdots \\ x_n \end{bmatrix}.$$

In addition, the equation system becomes

$$\begin{aligned} a_{11}x_1 + a_{13}x_3 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{23}x_3 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m3}x_3 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m, \end{aligned}$$

So, according to the new positions of coefficients in the equations, we get

$$\begin{bmatrix} a_{11} & a_{13} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{23} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m3} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

In other words, if I switch x_α and x_β, I only need to switch column α and column β to get the new matrix.

If I switch three variables, for example, I replace x_1, x_2, x_3 by x_3, x_1, x_2. Then I rewrite the equation system into

$$\begin{aligned}
a_{13}x_1 + a_{11}x_3 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\
a_{23}x_1 + a_{21}x_3 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\
&\vdots \\
a_{m3}x_1 + a_{m1}x_3 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m,
\end{aligned}$$

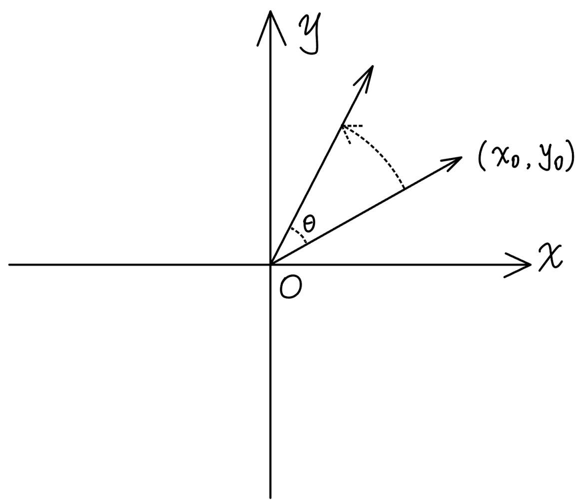
replace $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ \vdots \\ x_n \end{bmatrix}$ by $\begin{bmatrix} x_3 \\ x_1 \\ x_2 \\ x_4 \\ \vdots \\ x_n \end{bmatrix}$ and get the new matrix $\begin{bmatrix} a_{13} & a_{11} & a_{12} & \cdots & a_{1n} \\ a_{23} & a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m3} & a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$. We can keep going to switch more unknowns, but it is not necessary now. $\square$

Homework 1.1.8. For each linear equation system in Homework 1.1.6, choose two variables you like, switch their orders and write down the new matrix.

1.2 Matrices Acting on Vectors

Now, you may think that a matrix represents the information describing a linear equation system. However, the most interesting thing of matrix is that you can view it as an action, a transformation, or a function on vectors.

Example 1.2.1. Given any vector (x_0, y_0) starting from the origin on the Oxy plane, if we rotate it counterclockwise about the origin by angle θ, what is its coordinate after the rotation?



You may need the following formulas:

$$\cos(A+B) = \cos(A)\cos(B) - \sin(A)\sin(B);$$

$$\sin(A+B) = \sin(A)\cos(B) + \cos(A)\sin(B).$$

Solution. Using polar coordinate, we can assume that

$$x_0 = r\cos\alpha \tag{1.2.1}$$
$$y_0 = r\sin\alpha \tag{1.2.2}$$

where r is the length of vector (x_0, y_0), and α is the angle between vector (x_0, y_0) and the x-axis.

After the rotation, we see that the angle between the new vector and the x-axis is now $\theta + \alpha$, while the length remains unchanged. Therefore, the coordinate (x_1, y_1) of the new vector is

$$x_1 = r\cos(\alpha+\theta) = r\cos\alpha\cos\theta - r\sin\alpha\sin\theta$$
$$y_1 = r\sin(\alpha+\theta) = r\sin\alpha\cos\theta + r\cos\alpha\sin\theta$$

Now plug (1.2.1) and (1.2.2) into x_1 and y_1, we get

$$\begin{aligned} x_1 &= x_0 \cos\theta - y_0 \sin\theta \\ y_1 &= x_0 \sin\theta + y_0 \cos\theta. \end{aligned}$$

□

Now, let's align what we get in the example:

$$\begin{aligned} x_1 &= x_0 \cos\theta - y_0 \sin\theta \\ y_1 &= x_0 \sin\theta + y_0 \cos\theta. \end{aligned}$$

We see that on the left-hand side, we have a list

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$$

that is the new coordinate, while on the right-hand side, we have a mix of the old coordinate and the information of the rotation.

Besides the expression of the new coordinate, we are also interested in the rotation transformation itself. So, why don't we take out the information of the rotation and then use it to study the rotation?

A natural way to express the coordinate is to take out the old coordinate

$$\begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$$

and then write down each coefficient according to their positions:

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}.$$

Then we get a 2×2 matrix

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

We study the rotation transformation by studying this type of matrices (We will see this later).

There are also many other interesting transformations:

Example 1.2.2. Let (x_0, y_0, z_0) be a vector in the $Oxyz$ space starting from the origin. Project it to the yz plane and find the matrix as in Example 1.2.1.

Solution. Let the new y-coordinate and the new z-coordinate be y_1 and z_1 respectively, then we see

$$\begin{aligned} y_1 &= y_0 = 0 \cdot x_0 + 1 \cdot y_0 + 0 \cdot z_0 \\ z_1 &= z_0 = 0 \cdot x_0 + 0 \cdot y_0 + 1 \cdot z_0. \end{aligned}$$

So the matrix form is

$$\begin{bmatrix} y_1 \\ z_1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix}.$$

□

From Examples 1.2.1 and 1.2.2, we have a matrix for each transformation. However, if we are given a matrix, how do we define the action of this matrix on vectors?

In general, we do not restrict ourselves in 2 or 3 dimensional cases. We expand the concept of vectors to higher dimensional cases.

Definition 1.2.3. An $n \times 1$ matrix

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

is called an n-dimensional column vector.

Remark 1.2.4. Here are the operation rules for row or column vectors: Let r and s be scalars and $\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$ and $\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$ be column vectors:

(a) scalar multiplication:

$$r \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} ra_1 \\ ra_2 \\ \vdots \\ ra_n \end{bmatrix};$$

(b) linearity:
$$r\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} + s\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} ra_1 + sb_1 \\ ra_2 + sb_2 \\ \vdots \\ ra_n + sb_n \end{bmatrix}.$$

Remark 1.2.5. Why do we focus on column vectors instead of row vectors here? This is because we want to express things in a more natural way. If you use row vectors, then you must put matrices on the right-hand side of row vectors.

Now, recall the transformation in Example 1.2.1, it is like

$$\begin{aligned} x_1 &= ax_0 + by_0 \\ y_1 &= cx_0 + dy_0. \end{aligned}$$

So, if we have a 2×2 matrix
$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$
and a column vector $\begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$, we define the action to be

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} := \begin{bmatrix} ax_0 + by_0 \\ cx_0 + dy_0 \end{bmatrix}.$$

In Example 1.2.2, things are like

$$\begin{aligned} x_1 &= a_{11}x_0 + a_{12}y_0 + a_{13}z_0 \\ y_1 &= a_{21}x_0 + a_{22}y_0 + a_{23}z_0. \end{aligned}$$

So, for 2×3 matrix
$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix},$$
the action is
$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}\begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} := \begin{bmatrix} a_{11}x_0 + a_{12}y_0 + a_{13}z_0 \\ a_{21}x_0 + a_{22}y_0 + a_{23}z_0 \end{bmatrix}.$$

Now we give the general definition of the action of matrix on column vectors.

Definition 1.2.6. Given an $m \times n$ matrix

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{n} \end{bmatrix}$$

and an n-dimensional column vector

$$\begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_m \end{bmatrix},$$

we define the following multiplication

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix} = \begin{bmatrix} a_{11}t_1 + a_{12}t_2 + a_{13}t_3 + \cdots & a_{1n}t_n \\ a_{21}t_1 + a_{22}t_2 + a_{23}t_3 + \cdots & a_{2n}t_n \\ \vdots & \\ a_{m1}t_1 + a_{m2}t_2 + a_{m3}t_3 + \cdots & a_{mn}t_n \end{bmatrix}.$$

With Definition 1.2.6, we see that each matrix is an action on vectors. In other words, matrix is a "functions" of which the domain consists of vectors.

Example 1.2.7. Calculate $\begin{bmatrix} 1 & \sqrt{3} & 0.5 \\ 2 & \pi & \dfrac{1}{4} \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 + 2\sqrt{3} \\ \frac{5}{2} + 2\pi \end{bmatrix}.$

Remark 1.2.8. In Definition 1.2.6, the number of columns of the big matrix

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix}$$

should be equal to the dimension of vector $\begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_m \end{bmatrix}$ and they are both n. However, after the multiplication, you get an m-dimensional new vector.

Example 1.2.9. Special matrices acting on vectors:

(a) $n \times n$ identity matrix: $\begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix} = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix}$

(b) $n \times n$ zero matrix, i.e., each entry is 0: You will get a zero vector.

(c) $n \times n$ diagonal matrix $\begin{bmatrix} a_{11} & & & \\ & a_{22} & & \\ & & \ddots & \\ & & & a_{nn} \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix} = \begin{bmatrix} a_{11}t_1 \\ a_{22}t_2 \\ \vdots \\ a_{nn}t_n \end{bmatrix}$

Remark 1.2.10. If the matrix is a $1 \times n$ matrix, what do we get? Think about operations between vectors.

Solution. It is a generalized "dot product":

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix} = a_{11}t_1 + a_{12}t_2 + \cdots + a_{1n}t_n.$$

□

Remark 1.2.11. According to Definition 1.2.6, Definition 1.1.4 gives a multiplication of the matrix of coefficients and the column vector of unknowns.

Homework 1.2.12. Calculate the following multiplications. Some of them are not well-defined.

(a) $\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 5 & 10 & -1 \\ 2 & 2 & 1 & 9 \end{bmatrix} \begin{bmatrix} -1 \\ -1 \\ 4 \\ 5 \end{bmatrix} =$

(b) $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} =$

(c) $\begin{bmatrix} 3 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} =$

(d) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} 3 \\ 4 \\ 8 \\ 9 \end{bmatrix} =$

(e) $\begin{bmatrix} -1 & \sqrt{-1} & 1 \\ 0.618 & 1.414 & 3.1416 \\ 2.718 & 10^{17} & \ln 2 \\ 0 & \sin\left(\frac{\pi}{8}\right) & \cos(1+\pi) \end{bmatrix} \begin{bmatrix} \pi \\ e \\ \pi^e \\ e^\pi \end{bmatrix} =$

Homework 1.2.13. Let matrix $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ act on the following vectors:

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

What can you see? Figure out the multiplications, sketch the new vectors in Oxy space after transformation and tell the effect of this matrix.

Let's mention the following distribution law at the end of this section.

Proposition 1.2.14. (**distribution law**) Let A be an $m\times n$ matrix, $\mathbf{v}$ and $\mathbf{w}$ be two $n\times 1$ column vector and r and s be two scalars, then

$$A(r\mathbf{v} + s\mathbf{w}) = A(r\mathbf{v}) + A(s\mathbf{w}) = r(A\mathbf{v}) + s(A\mathbf{w}).$$

1.3 Solving Systems of Homogeneous Linear Equations

Given a linear equation system

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \cdots + a_{mn}x_n &= b_m, \end{aligned}$$

we can write it into

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_m \end{bmatrix}.$$

Now, how do we manipulate the matrix to find the solution?

Before we go to the general case, we talk about the homogeneous ones.

Definition 1.3.1. When $b_1 = b_2 = \cdots = b_m = 0$, we call

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

a system of homogeneous linear equations.

Now, let's look at the following homogeneous example

$$\begin{bmatrix} 1 & 1 & 1 & 8 & 1 \\ 2 & 3 & 4 & 8 & 1 \\ 3 & 2 & 1 & 5 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{1.3.1}$$

First, in these equations

$$x_1 + x_2 + x_3 + 8x_4 + x_5 = 0 \tag{1.3.2}$$
$$2x_1 + 3x_2 + 4x_3 + 8x_4 + x_5 = 0 \tag{1.3.3}$$
$$3x_1 + 2x_2 + x_3 + 5x_4 + 2x_5 = 0 \tag{1.3.4}$$

we want to eliminate x_1 in (1.3.3) and (1.3.4). You may want to use

$$x_1 = -x_2 - x_3 - 8x_4 - x_5$$

to replace x_1's, but you know the heavy workload ...

However, we have a trick to reduce the workload. That is, $(-2)\times$ (1.3.2) + (1.3.3) and $(-3)\times$ (1.3.2) + (1.3.4). Then we get

$$x_1 + x_2 + x_3 + 8x_4 + x_5 = 0 \tag{1.3.5}$$
$$x_2 + 2x_3 - 8x_4 - x_5 = 0 \tag{1.3.6}$$
$$-x_2 - 2x_3 - 19x_4 - x_5 = 0. \tag{1.3.7}$$

If we look at the matrix, the operation is to replace row 2 by $(-2)\times$row 1 + row 2, and replace row 3 by $(-3)\times$row 1 + row 3. This operation is shown as follows:

$$\begin{bmatrix} 1 & 1 & 1 & 8 & 1 \\ 2 & 3 & 4 & 8 & 1 \\ 3 & 2 & 1 & 5 & 2 \end{bmatrix}$$
$$\xrightarrow[(-2)\times\text{row } 1 + \text{row } 2]{\text{replace row 2 by}} \begin{bmatrix} 1 & 1 & 1 & 8 & 1 \\ 0 & 1 & 2 & -8 & -1 \\ 3 & 2 & 1 & 5 & 2 \end{bmatrix}$$
$$\xrightarrow[(-3)\times\text{row } 1 + \text{row } 3]{\text{replace row 3 by}} \begin{bmatrix} 1 & 1 & 1 & 8 & 1 \\ 0 & 1 & 2 & -8 & -1 \\ 0 & -1 & -2 & -19 & -1 \end{bmatrix}.$$

Let's pause here and define a row operation.

Definition 1.3.2. (**row reduction**) For matrix

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & a_{i3} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{j1} & a_{j2} & a_{j3} & \cdots & a_{jn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix}$$

and a scalar r, there is a row operation

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & a_{i3} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{j1} & a_{j2} & a_{j3} & \cdots & a_{jn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix} \xrightarrow[r \times \text{row } i + \text{row } j]{\text{replace row } j \text{ by}} \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & a_{i3} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{j1} + ra_{i1} & a_{j2} + ra_{i2} & a_{j3} + ra_{i3} & \cdots & a_{jn} + ra_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix}.$$

Now we go back to problem of equation (1.3.1). Currently, the matrix has been transformed into

$$\begin{bmatrix} 1 & 1 & 1 & 8 & 1 \\ 0 & 1 & 2 & -8 & -1 \\ 0 & -1 & -2 & -19 & -1 \end{bmatrix}$$

and the equation is now

$$\begin{bmatrix} 1 & 1 & 1 & 8 & 1 \\ 0 & 1 & 2 & -8 & -1 \\ 0 & -1 & -2 & -19 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Now, let's see if we can eliminate more unknowns. It is inappropriate to use row 1 now (Why? Because we don't want to destroy those zeros).

Let's use row 2 to eliminate x_2 in row 3:

$$\begin{bmatrix} 1 & 1 & 1 & 8 & 1 \\ 0 & 1 & 2 & -8 & -1 \\ 0 & -1 & -2 & -19 & -1 \end{bmatrix} \xrightarrow[\text{row 2 + row 3}]{\text{replace row 3 by}} \begin{bmatrix} 1 & 1 & 1 & 8 & 1 \\ 0 & 1 & 2 & -8 & -1 \\ 0 & 0 & 0 & -27 & -2 \end{bmatrix}.$$

Now we use the row 3 to eliminate x_3 in row 4. Oh, there is no row 4! So, we should move up by one step, i.e., use row 3 to eliminate x_3 in row 2. Wait! We cannot do it! The coefficient of x_3 in row 3 is 0.

So, we look at x_4 and x_5. Can we use row 3 to eliminate x_4 in row 2? Yes, it is a choice. However, the coefficient of x_4 in row 3 is now 28. This number is a little bit annoying. However, the coefficient of x_5 in row 5 is -2, so I choose to deal with x_5.

Now, I am going to switch the positions of x_3 and x_5. Then the matrix becomes

$$\begin{bmatrix} 1 & 1 & 1 & 8 & 1 \\ 0 & 1 & -1 & -8 & 2 \\ 0 & 0 & -2 & -27 & 0 \end{bmatrix}$$

and the column vector of unknowns becomes

$$\begin{bmatrix} x_1 \\ x_2 \\ x_5 \\ x_4 \\ x_3 \end{bmatrix}.$$

Here we define a column operation:

Definition 1.3.3. (**column switching**) For matrix

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1i} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2i} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mi} & \cdots & a_{mj} & \cdots & a_{mn} \end{bmatrix},$$

the column switching is defined to be

$$
\begin{aligned}
&\begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1i} & \cdots & a_{1j} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2i} & \cdots & a_{2j} & \cdots & a_{2n} \\
\vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \cdots & \vdots \\
a_{m1} & a_{m2} & \cdots & a_{mi} & \cdots & a_{mj} & \cdots & a_{mn}
\end{bmatrix} \\
\xrightarrow{\text{switch column } i \text{ and column } j}
&\begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1i} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2i} & \cdots & a_{2n} \\
\vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \cdots & \vdots \\
a_{m1} & a_{m2} & \cdots & a_{mj} & \cdots & a_{mi} & \cdots & a_{mn}
\end{bmatrix}.
\end{aligned}
$$

Now we go back to the problem. We replace row 2 by $-\dfrac{1}{2}\times$row 3 + row 2, and replace row 1 by $\dfrac{1}{2}\times$row 3 + row 1:

$$
\begin{aligned}
&\begin{bmatrix}
1 & 1 & 1 & 8 & 1 \\
0 & 1 & -1 & -8 & 2 \\
0 & 0 & -2 & -27 & 0
\end{bmatrix} \\
\xrightarrow[-\frac{1}{2}\times \text{row } 3 + \text{row } 2]{\text{replace row } 2 \text{ by}}
&\begin{bmatrix}
1 & 1 & 1 & 8 & 1 \\
0 & 1 & 0 & \frac{11}{2} & 2 \\
0 & 0 & -2 & -27 & 0
\end{bmatrix} \\
\xrightarrow[\frac{1}{2}\times \text{row } 3 + \text{row } 1]{\text{replace row } 1 \text{ by}}
&\begin{bmatrix}
1 & 1 & 0 & -\frac{11}{2} & 1 \\
0 & 1 & 0 & \frac{11}{2} & 2 \\
0 & 0 & -2 & -27 & 0
\end{bmatrix}
\end{aligned}
$$

And then we move up again, using row 2 to eliminate x_2 in row 1:

$$
\begin{bmatrix}
1 & 1 & 0 & -\frac{11}{2} & 1 \\
0 & 1 & 0 & \frac{11}{2} & 2 \\
0 & 0 & -2 & -27 & 0
\end{bmatrix}
\xrightarrow[\text{row } 1 - \text{row } 2]{\text{replace row } 1 \text{ by}}
\begin{bmatrix}
1 & 0 & 0 & -11 & -1 \\
0 & 1 & 0 & \frac{11}{2} & 2 \\
0 & 0 & -2 & -27 & 0
\end{bmatrix}
$$

Now is the time to introduce row scalar multiplication:

Definition 1.3.4. (**row scalar multiplication**)

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & a_{i3} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix} \xrightarrow[\text{by a scalar } r]{\text{multiply each entry in row } i} \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ ra_{i1} & ra_{i2} & ra_{i3} & \cdots & ra_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix}.$$

Now we do row scalar multiplication to the matrix in our problem:

$$\begin{bmatrix} 1 & 0 & 0 & -11 & -1 \\ 0 & 1 & 0 & \frac{11}{2} & 2 \\ 0 & 0 & -2 & -27 & 0 \end{bmatrix} \xrightarrow[\text{by a scalar } \left(-\frac{1}{2}\right)]{\text{multiply row 3}} \begin{bmatrix} 1 & 0 & 0 & -11 & -1 \\ 0 & 1 & 0 & \frac{11}{2} & 2 \\ 0 & 0 & 1 & \frac{27}{2} & 0 \end{bmatrix}$$

Remember that the order of unknowns is now x_1, x_2, x_5, x_4, x_3. So, this scalar multiplication is equivalent to transform

$$-2x_5 - 27x_4 = 0$$

into

$$x_5 + \frac{27}{2}x_4 = 0.$$

Now we are able to give the solution set to the equation system. Write down the matrix after so many operations and don't forget the order of unknowns:

$$\begin{bmatrix} 1 & 0 & 0 & -11 & -1 \\ 0 & 1 & 0 & \frac{11}{2} & 2 \\ 0 & 0 & 1 & \frac{27}{2} & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_5 \\ x_4 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

This is equivalent to

$$\begin{array}{lllllll} x_1 & & & -11x_4 & -x_3 & = 0 \\ & x_2 & & +\frac{11}{2}x_4 & +2x_3 & = 0. \\ & & x_5 & +\frac{27}{2}x_4 & & = 0 \end{array}$$

Then we see we have two free variables x_4 and x_3, while x_1, x_2, x_5 can be expressed as

$$\begin{aligned} x_1 &= 11x_4 + x_3 \\ x_2 &= -\frac{11}{2}x_4 - 2x_3 \\ x_5 &= -\frac{27}{2}x_4. \end{aligned}$$

Write them into column vectors:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = x_3 \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 11 \\ -\frac{11}{2} \\ 0 \\ 1 \\ -\frac{27}{2} \end{bmatrix},$$

or you can write them into

$$\begin{bmatrix} x_1 \\ x_2 \\ x_5 \\ x_4 \\ x_3 \end{bmatrix} = x_4 \begin{bmatrix} 11 \\ -\frac{11}{2} \\ -\frac{27}{2} \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ -2 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Example 1.3.5. Solve the following homogeneous equation system:

$$\begin{bmatrix} 0 & 1 & -1 \\ 1 & -3 & -5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

There seems to be no way to get started because the (1,1)-entry is 0. How do you deal with it?

Solution. The order of unknowns is important, but the order of equations is not important. The equation here is

$$\begin{aligned} x_2 - x_3 &= 0 \\ x_1 - 3x_2 - 5x_3 &= 0, \end{aligned}$$

but you can deal with

$$\begin{aligned} x_1 - 3x_2 - 5x_3 &= 0 \\ x_2 - x_3 &= 0. \end{aligned}$$

This time the matrix becomes

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

In other words, you swap the two rows. □

We give the last matrix operation that we need to solve equation systems:

Definition 1.3.6. (**row switching**)

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & a_{i3} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{j1} & a_{j2} & a_{j3} & \cdots & a_{jn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix} \xrightarrow{\text{switch row } i \text{ and row } j} \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{j1} & a_{j2} & a_{j3} & \cdots & a_{jn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & a_{i3} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix}.$$

Notice that the order of unknowns has nothing to do with row operations (row reduction, row scalar multiplication, row switching). However, please be careful when you do column switchings.

Remark 1.3.7. When you switch column i and column j, please also switch x_i and x_j.

Example 1.3.8. For the following matrix

$$\begin{bmatrix} 1 & 2 & 4 \\ 0 & 4 & 8 \end{bmatrix},$$

you may first transform it into

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 8 \end{bmatrix},$$

or first transform it into

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

Now we summarize the algorithm for solving homogeneous equations:

Theorem 1.3.9. (**Solving Homogeneous Linear Equation System**) For any homogeneous linear equation system

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

Step 1 Make sure that a_{11} is non-zero.

(a) If a_{11} is zero, then switch rows or columns to make the entry at (1,1) position non-zero. (I suggest you use row switching if possible, because when switching columns you must switch unknowns as well.)

(b) After making a_{11} non-zero, use row 1 to eliminate $a_{21}, a_{31}, \cdots, a_{m1}$ by row reduction.

Step 2 Now you get a matrix which looks like

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & b_{22} & b_{23} & \cdots & b_{2n} \\ 0 & b_{32} & b_{33} & \cdots & b_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & b_{m2} & b_{m3} & \cdots & b_{mn} \end{bmatrix}.$$

You should make sure that b_{22} is non-zero.

(a) If b_{22} is zero, then switch rows or columns to make it non-zero. However, this time row 1 does not involve in the procedure. You can only switch row 2,3,...,m, or column 2,3,...,n. Again, you'd better consider row switching first.

(b) Now, b_{22} is non-zero, use row reduction to eliminate $b_{32}, b_{42}, \cdots, b_{m2}$ and get

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & b_{22} & b_{23} & \cdots & b_{2n} \\ 0 & 0 & c_{33} & \cdots & c_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & c_{m3} & \cdots & c_{mn} \end{bmatrix}.$$

Step 3 Repeat the above procedures to row $3, 4, \cdots, m$. Remember that every time when you make an (i,i)-entry into a non-zero number, you can only switch row i with rows $i+1, i+2, \cdots, m$, or switch column j with $j+1, j+2, \cdots, n$.

Step 4 Finally, you see a matrix like

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1r} & a_{1\,r+1} & a_{1\,r+2} & \cdots & a_{1n} \\ 0 & b_{22} & b_{23} & \cdots & b_{2r} & b_{2\,r+1} & b_{2\,r+2} & \cdots & b_{2n} \\ 0 & 0 & c_{33} & \cdots & c_{3r} & c_{3\,r+1} & c_{3\,r+2} & \cdots & c_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & d_{rr} & d_{r\,r+1} & d_{r\,r+2} & \cdots & d_{rn} \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}.$$

(a) You should make sure that d_{rr} is non-zero. Otherwise, do column switching to make it non-zero. This time, column switching is the only choice and you should only switch column r with columns $r+1, \cdots, n$. Also, don't forget to switch the order of unknowns.

(b) Use row r to eliminate entries in column r above d_{rr}.

(c) Repeat the row reduction, finally, you will use c_{33} to eliminate a_{13} and b_{23}, and use b_{22} to eliminate a_{12}. You will get

$$\begin{bmatrix} a_{11} & 0 & 0 & \cdots & 0 & \hat{a}_{1\,r+1} & \hat{a}_{1\,r+2} & \cdots & \hat{a}_{1n} \\ 0 & b_{22} & 0 & \cdots & 0 & \hat{b}_{2\,r+1} & \hat{b}_{2\,r+2} & \cdots & \hat{b}_{2n} \\ 0 & 0 & c_{33} & \cdots & 0 & \hat{c}_{3\,r+1} & \hat{c}_{3\,r+2} & \cdots & \hat{c}_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & d_{rr} & d_{r\,r+1} & d_{r\,r+2} & \cdots & d_{rn} \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}.$$

Step 5 Divide each row by $a_{11}, b_{22}, c_{33}, \cdots, d_{rr}$ respectively and get

$$\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & \tilde{a}_{1\ r+1} & \tilde{a}_{1\ r+2} & \cdots & \tilde{a}_{1n} \\ 0 & 1 & 0 & \cdots & 0 & \tilde{b}_{2\ r+1} & \tilde{b}_{2\ r+2} & \cdots & \tilde{b}_{2n} \\ 0 & 0 & 1 & \cdots & 0 & \tilde{c}_{3\ r+1} & \tilde{c}_{3\ r+2} & \cdots & \tilde{c}_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 & \tilde{d}_{r\ r+1} & \tilde{d}_{r\ r+2} & \cdots & \tilde{d}_{rn} \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}.$$

The column vector of unknowns is now

$$\begin{bmatrix} x_{k_1} \\ x_{k_2} \\ x_{k_3} \\ \vdots \\ x_{k_r} \\ x_{k_{r+1}} \\ x_{k_{r+2}} \\ \vdots \\ x_{k_n} \end{bmatrix}$$

and the solution is

$$\begin{bmatrix} x_{k_1} \\ x_{k_2} \\ x_{k_3} \\ \vdots \\ x_{k_r} \\ x_{k_{r+1}} \\ x_{k_{r+2}} \\ \vdots \\ x_{k_n} \end{bmatrix} = x_{k_{r+1}} \begin{bmatrix} -\tilde{a}_{1\ r+1} \\ -\tilde{b}_{2\ r+1} \\ -\tilde{c}_{3\ r+1} \\ \vdots \\ -\tilde{d}_{r\ r+1} \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + x_{k_{r+2}} \begin{bmatrix} -\tilde{a}_{1\ r+2} \\ -\tilde{b}_{2\ r+2} \\ -\tilde{c}_{3\ r+2} \\ \vdots \\ -\tilde{d}_{r\ r+2} \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \cdots + x_{k_n} \begin{bmatrix} -\tilde{a}_{1\ n} \\ -\tilde{b}_{2\ n} \\ -\tilde{c}_{3\ n} \\ \vdots \\ -\tilde{d}_{r\ n} \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix},$$

where $x_{k_{r+1}}, x_{k_{r+2}}, \cdots, x_{k_n}$ are free.

Homework 1.3.10. This problem is adapted from [3]: Example 1 on page 57.

$$\begin{bmatrix} 1 & 4 & 2 \\ 2 & 5 & 1 \\ 3 & 6 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Homework 1.3.11. This problem is adapted from [3]: Example 1 on page 65.

$$\begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Remark 1.3.12. There is in fact no specific order for the four matrix operations in this section. For example, for matrix

$$\begin{bmatrix} 1 & 2 & 3 \\ 12 & 3 & 6 \end{bmatrix},$$

you may divide the second row by 3 first.

1.4 Recitation

In this section, we practice solving homogeneous linear equation systems.

Example 1.4.1. This problem is adapted from [3]: Exercise 12 on page 10.

$$\begin{bmatrix} 1 & -3 & 4 \\ 3 & -7 & 7 \\ -4 & 6 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Solution.

$$\begin{aligned} & \begin{bmatrix} 1 & -3 & 4 \\ 3 & -7 & 7 \\ -4 & 6 & -1 \end{bmatrix} \\ \xrightarrow{\text{row } 2 + (-3)\times \text{row } 1} & \begin{bmatrix} 1 & -3 & 4 \\ 0 & 2 & -5 \\ -4 & 6 & -1 \end{bmatrix} \\ \xrightarrow{\text{row } 3 + 4\times \text{row } 1} & \begin{bmatrix} 1 & -3 & 4 \\ 0 & 2 & -5 \\ 0 & -6 & 15 \end{bmatrix} \\ \xrightarrow{\text{row } 3 + 3\times \text{row } 2} & \begin{bmatrix} 1 & -3 & 4 \\ 0 & 2 & -5 \\ 0 & 0 & 0 \end{bmatrix} \\ \xrightarrow{\text{row } 1 + \frac{3}{2}\times \text{row } 2} & \begin{bmatrix} 1 & 0 & -\frac{7}{2} \\ 0 & 2 & -5 \\ 0 & 0 & 0 \end{bmatrix} \\ \xrightarrow{\text{multiply row } 2 \text{ by } \frac{1}{2}} & \begin{bmatrix} 1 & 0 & -\frac{7}{2} \\ 0 & 1 & -\frac{5}{2} \\ 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

The solution is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} \frac{7}{2} \\ \frac{5}{2} \\ 1 \end{bmatrix},$$

where x_3 is free. □

Example 1.4.2. This problem is adapted from [3]: Exercise 9 on page 32.

$$\begin{bmatrix} 0 & 1 & 5 & 1 \\ 4 & 6 & -1 & 2 \\ -1 & 3 & -8 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Solution.

$$\begin{aligned}
& \begin{bmatrix} 0 & 1 & 5 & 1 \\ 4 & 6 & -1 & 2 \\ -1 & 3 & -8 & 0 \end{bmatrix} \\
\xrightarrow{\text{swap row 3 and row 1}} & \begin{bmatrix} -1 & 3 & -8 & 0 \\ 4 & 6 & -1 & 2 \\ 0 & 1 & 5 & 1 \end{bmatrix} \\
\xrightarrow{\text{row 2 } + 4\times \text{row 1}} & \begin{bmatrix} -1 & 3 & -8 & 0 \\ 0 & 18 & -33 & 2 \\ 0 & 1 & 5 & 1 \end{bmatrix} \\
\xrightarrow{\text{row 3 } + (-\frac{1}{18})\times \text{row 2}} & \begin{bmatrix} -1 & 3 & -8 & 0 \\ 0 & 18 & -33 & 2 \\ 0 & 0 & \frac{41}{6} & \frac{8}{9} \end{bmatrix} \\
\xrightarrow{\text{row 2 } + \frac{198}{41}\times \text{row 3}} & \begin{bmatrix} -1 & 3 & -8 & 0 \\ 0 & 18 & 0 & \frac{258}{41} \\ 0 & 0 & \frac{41}{6} & \frac{8}{9} \end{bmatrix} \\
\xrightarrow{\text{row 1 } + \frac{48}{41}\times \text{row 3}} & \begin{bmatrix} -1 & 3 & 0 & \frac{128}{123} \\ 0 & 18 & 0 & \frac{258}{41} \\ 0 & 0 & \frac{41}{6} & \frac{8}{9} \end{bmatrix} \\
\xrightarrow{\text{row 1 } + (-\frac{1}{6})\times \text{row 2}} & \begin{bmatrix} -1 & 0 & 0 & -\frac{1}{123} \\ 0 & 18 & 0 & \frac{258}{41} \\ 0 & 0 & \frac{41}{6} & \frac{8}{9} \end{bmatrix} \\
\xrightarrow{\text{multiply rows by } -1, \frac{1}{18}, \frac{6}{41} \text{ respectively}} & \begin{bmatrix} 1 & 0 & 0 & \frac{1}{123} \\ 0 & 1 & 0 & \frac{43}{123} \\ 0 & 0 & 1 & \frac{16}{123} \end{bmatrix}.
\end{aligned}$$

So the solution is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = x_4 \begin{bmatrix} -\frac{1}{123} \\ -\frac{43}{123} \\ -\frac{16}{123} \\ 1 \end{bmatrix},$$

where x_4 is free. □

Example 1.4.3. This problem is adapted from [3]: Exercise 31 on page 62.

$$\begin{bmatrix} 2 & 3 & 5 \\ -5 & 1 & -4 \\ -3 & -1 & -4 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Solution.

$$\begin{aligned} & \begin{bmatrix} 2 & 3 & 5 \\ -5 & 1 & -4 \\ -3 & -1 & -4 \\ 1 & 0 & 1 \end{bmatrix} \\ \xrightarrow{\text{row } 2 + (\frac{5}{2})\times\text{row } 1} & \begin{bmatrix} 2 & 3 & 5 \\ 0 & \frac{17}{2} & \frac{17}{2} \\ -3 & -1 & -4 \\ 1 & 0 & 1 \end{bmatrix} \\ \xrightarrow{\text{row } 3 + (\frac{3}{2})\times\text{row } 1} & \begin{bmatrix} 2 & 3 & 5 \\ 0 & \frac{17}{2} & \frac{17}{2} \\ 0 & \frac{7}{2} & \frac{7}{2} \\ 1 & 0 & 1 \end{bmatrix} \\ \xrightarrow{\text{row } 4 + (-\frac{1}{2})\times\text{row } 1} & \begin{bmatrix} 2 & 3 & 5 \\ 0 & \frac{17}{2} & \frac{17}{2} \\ 0 & \frac{7}{2} & \frac{7}{2} \\ 0 & -\frac{3}{2} & -\frac{3}{2} \end{bmatrix} \\ \xrightarrow{\text{use row 2 to eliminate row 3 and row 4}} & \begin{bmatrix} 2 & 3 & 5 \\ 0 & \frac{17}{2} & \frac{17}{2} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{aligned}$$

$$\xrightarrow{\text{divide row 2 by } \frac{17}{2}} \begin{bmatrix} 2 & 3 & 5 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\xrightarrow{\text{row 1} + (-3)\times \text{row 2}} \begin{bmatrix} 2 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\xrightarrow{\text{multiply row 1 by } \frac{1}{2}} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

So, the solution is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$$

□

Homework 1.4.4. This problem is adapted from [3]: Exercise 7 on page 61.

$$\begin{bmatrix} 0 & 4 & -3 & 0 \\ -2 & -7 & 5 & 1 \\ -4 & -5 & 7 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Homework 1.4.5. This problem is adapted from [3]: Example 4 on page 77.

$$\begin{bmatrix} 1 & -4 & 8 & 1 \\ 0 & 2 & -1 & 3 \\ 0 & 0 & 0 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Homework 1.4.6. This problem is adapted from [3]: Exercise 34 on page 49.

$$\begin{bmatrix} 4 & -6 \\ -8 & 12 \\ 6 & -9 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

1.5 Solving Non-homogeneous Systems

What is a non-homogeneous system?

Definition 1.5.1. If in the following equation system

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_m \end{bmatrix}, \tag{1.5.1}$$

$b_1, \cdots b_m$ are not all zeros, then we say this is a non-homogeneous system.

This time we deal with the following augmented matrix (Because when you do row reductions for non-homogeneous equations, $b_1, \cdots, b_m$ are also engaged.).

Definition 1.5.2. The augmented matrix of non-homogeneous system (1.5.1) is

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} & b_m \end{bmatrix}.$$

Based on what we've learned in homogeneous cases, we have this result:

Theorem 1.5.3. Apply the algorithm in Theorem 1.3.9 (but avoid column switching with the column consisting of $b_1, b_2, \cdots, b_m$) to the augmented matrix, we get a matrix like

$$\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & \tilde{a}_{1\ r+1} & \tilde{a}_{1\ r+2} & \cdots & \tilde{a}_{1n} & \varepsilon_1 \\ 0 & 1 & 0 & \cdots & 0 & \tilde{b}_{2\ r+1} & \tilde{b}_{2\ r+2} & \cdots & \tilde{b}_{2n} & \varepsilon_2 \\ 0 & 0 & 1 & \cdots & 0 & \tilde{c}_{3\ r+1} & \tilde{c}_{3\ r+2} & \cdots & \tilde{c}_{3n} & \varepsilon_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & \tilde{d}_{r\ r+1} & \tilde{d}_{r\ r+2} & \cdots & \tilde{d}_{rn} & \varepsilon_r \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \varepsilon_{r+1} \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \varepsilon_{r+2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \varepsilon_m \end{bmatrix}. \tag{1.5.2}$$

Remark 1.5.4. Homogeneous systems always have solutions (at least they have zero vectors as their solutions). However, non-homogeneous systems do not always have solutions.

Example 1.5.5. If we get an augmented matrix

$$\begin{bmatrix} 1 & 0 & 0 & 4 & 2 \\ 0 & 1 & 0 & 5 & 1 \\ 0 & 0 & 1 & 2 & 7 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

by matrix operations, can you write down the solution?

Solution. The equation system is

$$\begin{array}{cccccc} x_1 & & & +4x_4 & = 2 \\ & x_2 & & +5x_4 & = 1 \\ & & x_3 & +2x_4 & = 7. \end{array}$$

Therefore, x_4 is free and

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = x_4 \begin{bmatrix} -4 \\ -5 \\ -2 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \\ 7 \\ 0 \end{bmatrix}$$

□

Example 1.5.6. If we get an augmented matrix

$$\begin{bmatrix} 1 & 0 & 0 & 4 & 2 \\ 0 & 1 & 0 & 5 & 1 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 7 \end{bmatrix}$$

by matrix operations, can you write down the solution?

Solution. The equation system is

$$\begin{array}{cccccc} x_1 & & & +4x_4 & = 2 \\ & x_2 & & +5x_4 & = 1 \\ & & x_3 & +2x_4 & = 0 \\ & & & 0 & = 7. \end{array}$$

The last row leads to a contradiction. So, this system has no solution. □

Proposition 1.5.7. When $\varepsilon_{r+1} = \varepsilon_{r+2} = \cdots = \cdots = \varepsilon_{r+m} = 0$ in matrix 1.5.2, the system has solutions. Otherwise, it does not have solutions.

Let's look at some examples:

Example 1.5.8. This problem is adapted from [3]: Exercise 25 on page 42:

$$\begin{bmatrix} 4 & -2 & 1 \\ 8 & -2 & 4 \\ -12 & 2 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -7 \\ -3 \\ 10 \end{bmatrix}.$$

Solution.

$$\begin{aligned}
& \begin{bmatrix} 4 & -2 & 1 & -7 \\ 8 & -2 & 4 & -3 \\ -12 & 2 & -3 & 10 \end{bmatrix} \\
\xrightarrow{\text{row } 2 + (-2)\times \text{row } 1} & \begin{bmatrix} 4 & -2 & 1 & -7 \\ 0 & 2 & 2 & 11 \\ -12 & 2 & -3 & 10 \end{bmatrix} \\
\xrightarrow{\text{row } 3 + 3\times \text{row } 1} & \begin{bmatrix} 4 & -2 & 1 & -7 \\ 0 & 2 & 2 & 11 \\ 0 & -4 & 0 & -11 \end{bmatrix} \\
\xrightarrow{\text{row } 3 + 2\times \text{row } 2} & \begin{bmatrix} 4 & -2 & 1 & -7 \\ 0 & 2 & 2 & 11 \\ 0 & 0 & 4 & 11 \end{bmatrix} \\
\xrightarrow{\text{row } 2 + (-\frac{1}{2})\times \text{row } 3} & \begin{bmatrix} 4 & -2 & 1 & -7 \\ 0 & 2 & 0 & \frac{11}{2} \\ 0 & 0 & 4 & 11 \end{bmatrix} \\
\xrightarrow{\text{row } 1 + (-\frac{1}{4})\times \text{row } 3} & \begin{bmatrix} 4 & -2 & 0 & -\frac{39}{4} \\ 0 & 2 & 0 & \frac{11}{2} \\ 0 & 0 & 4 & 11 \end{bmatrix} \\
\xrightarrow{\text{row } 1 + 1\times \text{row } 2} & \begin{bmatrix} 4 & 0 & 0 & -\frac{17}{4} \\ 0 & 2 & 0 & \frac{11}{2} \\ 0 & 0 & 4 & 11 \end{bmatrix} \\
\xrightarrow{\text{divide rows by 4, 2, 4 respectively}} & \begin{bmatrix} 1 & 0 & 0 & -\frac{17}{16} \\ 0 & 1 & 0 & \frac{11}{4} \\ 0 & 0 & 1 & \frac{11}{4} \end{bmatrix}.
\end{aligned}$$

Thus we get

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -\frac{17}{16} \\ \frac{11}{4} \\ \frac{11}{4} \end{bmatrix}.$$

□

Example 1.5.9. This problem is adapted from [3]: Example 3 on page 45:

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

What if we replace the matrix by

$$\begin{bmatrix} 3 & 5 & -4 \\ -3 & -2 & 4 \\ -3 & -5 & 4 \end{bmatrix} ?$$

Solution.

$$\begin{aligned}
&\begin{bmatrix} 3 & 5 & -4 & 7 \\ -3 & -2 & 4 & -1 \\ 6 & 1 & -8 & -4 \end{bmatrix} \\
\xrightarrow{\text{row } 2 + 1\times\text{row } 1} &\begin{bmatrix} 3 & 5 & -4 & 7 \\ 0 & 3 & 0 & 6 \\ 6 & 1 & -8 & -4 \end{bmatrix} \\
\xrightarrow{\text{row } 3 + (-2)\times\text{row } 1} &\begin{bmatrix} 3 & 5 & -4 & 7 \\ 0 & 3 & 0 & 6 \\ 0 & -9 & 0 & -18 \end{bmatrix} \\
\xrightarrow{\text{row } 3 + 3\times\text{row } 2} &\begin{bmatrix} 3 & 5 & -4 & 7 \\ 0 & 3 & 0 & 6 \\ 0 & 0 & 0 & 0 \end{bmatrix} \\
\xrightarrow{\text{row } 1 + (-\frac{5}{3})\times\text{row } 2} &\begin{bmatrix} 3 & 0 & -4 & -3 \\ 0 & 3 & 0 & 6 \\ 0 & 0 & 0 & 0 \end{bmatrix} \\
\xrightarrow[\text{divide row 2 by 3}]{\text{divide row 1 by 3}} &\begin{bmatrix} 1 & 0 & -\frac{4}{3} & -1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}.
\end{aligned}$$

Thus we get

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{4}{3}x_3 - 1 \\ 2 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} \frac{4}{3} \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}.$$

If the matrix is replaced by

$$\begin{bmatrix} 3 & 5 & -4 \\ -3 & -2 & 4 \\ -3 & -5 & 4 \end{bmatrix},$$

then by row reductions, the augmented matrix becomes

$$\begin{bmatrix} 3 & 0 & -4 & -3 \\ 0 & 3 & 0 & 6 \\ 0 & 0 & 0 & 3 \end{bmatrix}.$$

The last row gives us $0 = 3$. So, there is no solution in this case. □

Homework 1.5.10. This problem is adapted from [3]: Exercise 15 on page 48.

$$\begin{bmatrix} 1 & 3 & 1 \\ -4 & -9 & 2 \\ 0 & -3 & -6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ -3 \end{bmatrix}$$

Homework 1.5.11. This problem is adapted from [3]: Exercise 34 on page 49.

$$\begin{bmatrix} 4 & -6 \\ -8 & 12 \\ 6 & -9 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix}$$

Chapter 2

Matrix Algebra

2.1 Matrix Operations

As we said before, a matrix can be viewed as an action on vectors. In other words, a matrix is a "function" defined on the collection of vectors.

Recall what we have seen for functions:

Proposition 2.1.1. Let $f : A \to B$ and $g : B \to C$ be two functions, then for any $x \in A$, we have

$$(g \circ f)(x) = g((f(x)).$$

We hope that for any two matrices A and B, and any column vector $\mathbf{v}$, we have

$$(AB)\mathbf{v} = A(B\mathbf{v}). \tag{2.1.1}$$

Here AB is the product of A and B.

First of all, if A is an $m \times n$ matrix while B is an $k \times l$ matrix, to make $B\mathbf{v}$ in 2.1.1 well-defined, $\mathbf{v}$ should be an $l \times 1$ column vector. Then, $B\mathbf{v}$ is a $k \times 1$ column vector. So, to make $A(B\mathbf{v})$ well-defined, we must let $n = k$.

Proposition 2.1.2. To defined the product AB, the amount of columns of A must be the same as the amount of rows of B, i.e., A is an $m \times n$ matrix, while B is an $n \times l$ matrix.

Now we write matrix B into

$$\begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_l \end{bmatrix},$$

where $\mathbf{b}_i$ is an $n \times 1$ column vector and it means the i-th column of B (This idea is from [3]:

page 96.). Then we write the $l \times 1$ column vector $\mathbf{v}$ into

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_l \end{bmatrix}.$$

Then recall the rule of matrix acting on column vectors, we see that

$$B\mathbf{v} = x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + \cdots x_l\mathbf{b}_l. \text{ (Why?)}$$

Therefore, if we want

$$(AB)\mathbf{v} = A(B\mathbf{v}),$$

we should have

$$(AB)\mathbf{v} = A(B\mathbf{v}) = A(x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + \cdots x_l\mathbf{b}_l).$$

Again, recall the rule of matrix acting on column vectors, we get

$$(AB)\mathbf{v} = x_1(A\mathbf{b}_1) + x_2(A\mathbf{b}_2) + \cdots x_l(A\mathbf{b}_l). \text{ (Why?)}$$

Can you see the columns of AB now?

Definition 2.1.3. (**matrix multiplication**) Let A and B be $m \times n$ and $n \times l$ matrices respectively, and write B into

$$\begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_l \end{bmatrix}.$$

Then their product AB is given by

$$AB = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & \cdots & A\mathbf{b}_l \end{bmatrix}.$$

Remark 2.1.4. The product of a matrix and a column vector is a special case of matrix multiplication.

Remark 2.1.5. Let $A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$ and $B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1l} \\ b_{21} & b_{22} & \cdots & b_{2l} \\ \vdots & \vdots & & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nl} \end{bmatrix}$. The (i,j)-entry of AB is equal to

$$a_{i1}b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j} + \cdots + a_{in}b_{nj}.$$

Example 2.1.6. Let $A = \begin{bmatrix} 2 & 0 & 2 & 0 \\ 2 & 0 & 2 & 1 \\ 2 & 0 & 2 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \\ 7 & 8 \end{bmatrix}$. Find AB.

Solution. The first column of AB is

$$\begin{bmatrix} 2 & 0 & 2 & 0 \\ 2 & 0 & 2 & 1 \\ 2 & 0 & 2 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 5 \\ 7 \end{bmatrix} = \begin{bmatrix} 2\times 1 + 0\times 3 + 2\times 5 + 0\times 7 \\ 2\times 1 + 0\times 3 + 2\times 5 + 1\times 7 \\ 2\times 1 + 0\times 3 + 2\times 5 + 2\times 7 \end{bmatrix} = \begin{bmatrix} 12 \\ 19 \\ 26 \end{bmatrix},$$

and the second column is

$$\begin{bmatrix} 2 & 0 & 2 & 0 \\ 2 & 0 & 2 & 1 \\ 2 & 0 & 2 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \\ 6 \\ 8 \end{bmatrix} = \begin{bmatrix} 2\times 2 + 0\times 4 + 2\times 6 + 0\times 8 \\ 2\times 2 + 0\times 4 + 2\times 6 + 1\times 8 \\ 2\times 2 + 0\times 4 + 2\times 6 + 2\times 8 \end{bmatrix} = \begin{bmatrix} 16 \\ 24 \\ 32 \end{bmatrix}.$$

So, $AB = \begin{bmatrix} 12 & 16 \\ 19 & 24 \\ 26 & 32 \end{bmatrix}$. □

Using similar techniques (matrices acting on vectors), we define the following operations.

Definition 2.1.7. (**matrix addition and subtraction**) For two $m \times n$ matrices

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}$$

and

$$B = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{bmatrix},$$

we define

$$A + B = \begin{bmatrix} \mathbf{a}_1 + \mathbf{b}_1 & \mathbf{a}_2 + \mathbf{b}_2 & \cdots & \mathbf{a}_n + \mathbf{b}_n \end{bmatrix},$$

$$A - B = \begin{bmatrix} \mathbf{a}_1 - \mathbf{b}_1 & \mathbf{a}_2 - \mathbf{b}_2 & \cdots & \mathbf{a}_n - \mathbf{b}_n \end{bmatrix}.$$

Example 2.1.8. $\begin{bmatrix} 2 & 1 & 2 \\ 2 & 0 & 1 \\ 4 & 3 & 2 \end{bmatrix} + \begin{bmatrix} 4 & 5 & 2 \\ 3 & 7 & 1 \\ -1 & 0 & -2 \end{bmatrix} = \begin{bmatrix} 2+4 & 1+5 & 2+2 \\ 2+3 & 0+7 & 1+1 \\ 4-1 & 3+0 & 2-2 \end{bmatrix} = \begin{bmatrix} 6 & 6 & 4 \\ 5 & 7 & 2 \\ 3 & 3 & 0 \end{bmatrix}.$

Definition 2.1.9. (**scalar multiplication**) Let $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}$ and r be a scalar, we define

$$rA = \begin{bmatrix} r\mathbf{a}_1 & r\mathbf{a}_2 & \cdots & r\mathbf{a}_n \end{bmatrix}.$$

In particular, we define $-A = (-1)A$.

Example 2.1.10. Let $A = \begin{bmatrix} 4 & 5 & 2 \\ 3 & 7 & 1 \\ -1 & 0 & -2 \end{bmatrix}$. Then $2A = \begin{bmatrix} 8 & 10 & 4 \\ 6 & 14 & 2 \\ -2 & 0 & -4 \end{bmatrix}$.

We give the following properties of matrix multiplication, matrix addition and subtraction and scalar multiplication without proof:

Proposition 2.1.11. Let A, B, C be matrices, r, s be scalars and v, w be column vectors. We assume that all the operations in this proposition are well-defined.

(a) $(A+B)+C = A+(B+C)$;

(b) $A(BC) = (AB)C$;

(c) $(A+B)C = AC+BC$;

(d) $A(B+C) = AB+AC$;

(e) $(rs)A = r(sA)$;

(f) $r(AB) = (rA)B = A(rB)$;

(g) $r(A+B) = rA+rB$;

(h) $(r+s)A = rA+sA$.

These properties can sometimes simplify our calculations (see Homework 2.1.15).

Remark 2.1.12. In general $AB \neq BA$. For example, if A is a 3×4 matrix while B is a 4×3 matrix, then AB is a 3×3 matrix while BA is a 4×4 matrix.

Remark 2.1.13. Let I be the identity matrix and O be the zero matrix. For any other matrix A, if each multiplication IA, AI, OA, AO is well-defined, then

$$IA = A, AI = A, OA = O, AO = O.$$

Homework 2.1.14. For the following multiplications, point out those that are **NOT** well-defined.

(a) $$\begin{bmatrix} -1 & \sqrt{-1} & 1 \\ 0.618 & 1.414 & 3.1416 \\ 2.718 & 10^{17} & \ln 2 \\ 0 & \sin\left(\frac{\pi}{8}\right) & \cos\left(1+\pi\right) \end{bmatrix} \begin{bmatrix} -1 & \sqrt{-1} & 1 \\ 0.618 & 1.414 & 3.1416 \\ 2.718 & 10^{17} & \ln 2 \\ 0 & \sin\left(\frac{\pi}{8}\right) & \cos\left(1+\pi\right) \end{bmatrix};$$

(b) $$\begin{bmatrix} -1 & \sqrt{-1} \\ 0.618 & 3.1416 \\ 2.718 & \ln 2 \\ 0 & \sin\left(\frac{\pi}{8}\right) \end{bmatrix} \begin{bmatrix} 2.718 & 10^{17} & \ln 2 \\ 0 & \sin\left(\frac{\pi}{8}\right) & \cos\left(\frac{\pi}{12}\right) \end{bmatrix};$$

(c) $$\begin{bmatrix} -1 & \sqrt{-1} & 1 \\ 0.618 & 1.414 & 3.1416 \\ 2.718 & 10^{17} & \ln 2 \\ 0 & \sin\left(\frac{\pi}{8}\right) & \cos\left(\frac{\pi}{12}\right) \end{bmatrix} \begin{bmatrix} -1 & \sqrt{-1} & 1 \\ 0.618 & 1.414 & 3.1416 \\ 2.718 & 10^{17} & \ln 2 \end{bmatrix};$$

(d) $$\begin{bmatrix} -1 & \sqrt{-1} & 1 \\ 0.618 & 1.414 & 3.1416 \\ 2.718 & 10^{17} & \ln 2 \end{bmatrix} \begin{bmatrix} -1 & \sqrt{-1} & 1 \\ 0.618 & 1.414 & 3.1416 \\ 2.718 & 10^{17} & \ln 2 \\ 0 & \sin\left(\frac{\pi}{8}\right) & \cos\left(\frac{\pi}{12}\right) \end{bmatrix}.$$

Homework 2.1.15. Let $A = \begin{bmatrix} 3 & -4 \\ 4 & 3 \end{bmatrix}$, $B = \begin{bmatrix} 3 & 4 \\ -4 & 3 \end{bmatrix}$, $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

(a) Find AB and BA;

(b) Find $\left(I - \frac{1}{16}A\right)B$.

Hint: Use the distributive law (Proposition 2.1.11(c)) to simplify the calculation.

(c) Find A^2B^2, i.e., $AABB$.

Hint: Use the associative law (Proposition 2.1.11(b)) to simplify the calculation. Also, don't forget the property of identity matrix that $AI = A$ and $IA = A$.

2.2 Partitioned Matrices

This section will be useful when we describe theories and proofs. And, definitely, sometimes it provides us with some convenience in calculations.

For example, let

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

we have several ways to divide it into smaller matrices:

(a) $A = \left[\begin{array}{cc|cc} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ \hline 9 & 10 & 11 & 12 \end{array}\right]$

(b) $A = \left[\begin{array}{c|cc|c} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ \hline 9 & 10 & 11 & 12 \end{array}\right]$

(c) $A = \left[\begin{array}{c|cc|c} 1 & 2 & 3 & 4 \\ \hline 5 & 6 & 7 & 8 \\ \hline 9 & 10 & 11 & 12 \end{array}\right]$

(d) ...

Now for two partitioned matrices

$$A = \begin{bmatrix} M_{11} & M_{12} & \cdots & M_{1n} \\ M_{21} & M_{22} & \cdots & M_{2n} \\ \vdots & \vdots & & \vdots \\ M_{m1} & M_{m2} & \cdots & M_{mn} \end{bmatrix}, \quad B = \begin{bmatrix} N_{11} & N_{12} & \cdots & N_{1l} \\ N_{21} & N_{22} & \cdots & N_{2l} \\ \vdots & \vdots & & \vdots \\ N_{n1} & N_{n2} & \cdots & N_{nl} \end{bmatrix},$$

where M_{ij}'s and N_{ij}'s are matrices, we have the following propositions coming from the definitions of matrix operations in the previous section.

Proposition 2.2.1. Let

$$P_{ij} = M_{i1}N_{1j} + M_{i2}N_{2j} + M_{i3}N_{3j} + \cdots + M_{in}N_{nj}$$

and suppose each of these matrix multiplications is well-defined, then

$$AB = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1l} \\ P_{21} & P_{22} & \cdots & P_{2l} \\ \vdots & \vdots & & \vdots \\ P_{m1} & P_{m2} & \cdots & P_{ml} \end{bmatrix}.$$

Proposition 2.2.2. (**addition and subtraction**) For two partitioned matrices

$$A = \begin{bmatrix} M_{11} & M_{12} & \cdots & M_{1n} \\ M_{21} & M_{22} & \cdots & M_{2n} \\ \vdots & \vdots & & \vdots \\ M_{m1} & M_{m2} & \cdots & M_{mn} \end{bmatrix}, \quad B = \begin{bmatrix} N_{11} & N_{12} & \cdots & N_{1n} \\ N_{21} & N_{22} & \cdots & N_{2n} \\ \vdots & \vdots & & \vdots \\ N_{m1} & N_{m2} & \cdots & N_{mn} \end{bmatrix},$$

we have

$$A \pm B = \begin{bmatrix} M_{11} \pm N_{11} & M_{12} \pm N_{12} & \cdots & M_{1n} \pm N_{1n} \\ M_{21} \pm N_{21} & M_{22} \pm N_{22} & \cdots & M_{2n} \pm N_{2n} \\ \vdots & \vdots & & \vdots \\ M_{m1} \pm N_{m1} & M_{m2} \pm N_{m2} & \cdots & M_{mn} \pm N_{mn} \end{bmatrix}$$

if all additions and subtractions are well-defined.

Proposition 2.2.3. (**scalar multiplication**) For scalar r and

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

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

Homework 2.2.4. Use the formula

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

to figure out

$$\begin{bmatrix} \cos\theta_1 & -\sin\theta_1 & 1 & 0 \\ \sin\theta_1 & \cos\theta_1 & 0 & 1 \\ 0 & 0 & \cos\alpha_1 & -\sin\alpha_1 \\ 0 & 0 & \sin\alpha_1 & \cos\alpha_1 \end{bmatrix} \begin{bmatrix} \cos\theta_2 & -\sin\theta_2 & 1 & 0 \\ \sin\theta_2 & \cos\theta_2 & 0 & 1 \\ 0 & 0 & \cos\alpha_2 & -\sin\alpha_2 \\ 0 & 0 & \sin\alpha_2 & \cos\alpha_2 \end{bmatrix}.$$

2.3 The Inverse of a Matrix

An important problem of matrix actions on vectors is how to find a way to reverse the action. In other words, can we find our way home?

Definition 2.3.1. Let A and B be two $n \times n$ matrices, if we have $BA = AB = I$, where I is the identity matrix, then we say that A is the inverse of B, or B is the inverse of A. The notation is $B = A^{-1}$ or $A = B^{-1}$.

Remark 2.3.2. Why do we only focus on square matrices? For example, if $A = \begin{bmatrix} a \\ b \end{bmatrix}$ and $B = \begin{bmatrix} c & d \end{bmatrix}$. Let $BA = \begin{bmatrix} 1 \end{bmatrix}$, we get

$$ac + bd = 1.$$

Let $AB = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, we get

$$ac = bd = 1, ad = bc = 0.$$

These equations have no common solution.

How do we find the inverse? Well, for 2-dimensional case, we have a formula which you may see in other courses:

Proposition 2.3.3. For 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, we have

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

However, this is not enough for larger matrices. We hope to find an algorithm to calculate the inverse for the $n \times n$ case.

If we write B into

$$\begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{bmatrix}$$

and I into

$$\begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{bmatrix},$$

where the i-th entry of column vector $\mathbf{e}_i$ is 1 while other entries of $\mathbf{e}_i$ are 0. Then we get

$$\begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{bmatrix} = AB = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & \cdots & A\mathbf{b}_n \end{bmatrix}.$$

In other words, if we can solve equations $A\mathbf{b}_1 = \mathbf{e}_1, \cdots, A\mathbf{b}_n = \mathbf{e}_n$, then we find the inverse $B = A^{-1}$. However, this is too time consuming to be an algorithm.

In fact, what we need for solving linear equations are often row operations. We give another description of row operations using matrix multiplication, then we are able to give the algorithm.

Suppose that after some row operations on matrix A, we get another matrix $\widetilde{A}$.

Proposition 2.3.4. There is a matrix E such that $\widetilde{A} = EA$.

Example 2.3.5. Verify Proposition 2.3.4 in the following cases: Let $E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix}$, $E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ and $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$.

(a) Multiply the second row of A by 3, then compare the matrix you get with E_1A.

(b) Replace the third row of A by $2 \times$ row 1 + row 3, and compare the matrix you get with E_2A.

(c) Swap row 2 and row 3 of A, then compare the matrix you get with E_3A.

Solution. (a) They are the same.

(b) They are the same.

(c) They are the same.

□

Remark 2.3.6. Matrices like those in Example 2.3.5 are called elementary matrices or row operation matrices. The fact is, **a composition of row operations on A is equivalent to multiplying A by a series of elementary matrices from the left**.

Therefore, if we can do row reductions on A as in solving equations, then we get a new matrix EA. If $EA = I$, then E is the inverse that we need.

Now, we construct a partitioned matrix

$$\begin{bmatrix} A & I \end{bmatrix},$$

where I is the identity matrix. Applying row reductions to it, we get

$$\begin{bmatrix} EA & EI \end{bmatrix} = \begin{bmatrix} I & E \end{bmatrix}.$$

Now we can see the algorithm to find the inverse. It is almost the same as that of solving equations and we illustrate it by two examples.

Example 2.3.7. (This example is [3]: Example 7 on page 110.)

Find the inverse of matrix $A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 4 & -3 & 8 \end{bmatrix}$.

Solution.

$$\begin{bmatrix} 0 & 1 & 2 & 1 & 0 & 0 \\ 1 & 0 & 3 & 0 & 1 & 0 \\ 4 & -3 & 8 & 0 & 0 & 1 \end{bmatrix}$$

$$\xrightarrow{\text{swap row 1 and row 2}} \begin{bmatrix} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 4 & -3 & 8 & 0 & 0 & 1 \end{bmatrix}$$

$$\xrightarrow{\text{row 3} + (-4)\times\text{row 1}} \begin{bmatrix} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & -3 & -4 & 0 & -4 & 1 \end{bmatrix}$$

$$\xrightarrow{\text{row 3} + 3\times\text{row 2}} \begin{bmatrix} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 2 & 3 & -4 & 1 \end{bmatrix}$$

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

$$\xrightarrow{\text{row 1} + (-\frac{3}{2})\times\text{row 3}} \begin{bmatrix} 1 & 0 & 0 & -\frac{9}{2} & 7 & -\frac{3}{2} \\ 0 & 1 & 0 & -2 & 4 & -1 \\ 0 & 0 & 2 & 3 & -4 & 1 \end{bmatrix}$$

$$\xrightarrow{\text{divide row 3 by 2}} \begin{bmatrix} 1 & 0 & 0 & -\frac{9}{2} & 7 & -\frac{3}{2} \\ 0 & 1 & 0 & -2 & 4 & -1 \\ 0 & 0 & 1 & \frac{3}{2} & -2 & \frac{1}{2} \end{bmatrix}.$$

So, the inverse is $A^{-1} = \begin{bmatrix} -\frac{9}{2} & 7 & -\frac{3}{2} \\ -2 & 4 & -1 \\ \frac{3}{2} & -2 & \frac{1}{2} \end{bmatrix}$. □

Example 2.3.8. Find the inverse of matrix $A = \begin{bmatrix} 3 & 5 & 1 \\ 6 & 10 & 3 \\ 9 & 10 & 5 \end{bmatrix}$.

Solution.

$$
\begin{aligned}
&\begin{bmatrix} 3 & 5 & 1 & 1 & 0 & 0 \\ 6 & 10 & 3 & 0 & 1 & 0 \\ 9 & 10 & 5 & 0 & 0 & 1 \end{bmatrix} \\
\xrightarrow{\text{row 2 + }(-2)\times\text{row 1}} &\begin{bmatrix} 3 & 5 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & 0 \\ 9 & 10 & 5 & 0 & 0 & 1 \end{bmatrix} \\
\xrightarrow{\text{row 3 + }(-3)\times\text{row 1}} &\begin{bmatrix} 3 & 5 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & 0 \\ 0 & -5 & 2 & -3 & 0 & 1 \end{bmatrix} \\
\xrightarrow{\text{swap row 3 and row 2}} &\begin{bmatrix} 3 & 5 & 1 & 1 & 0 & 0 \\ 0 & -5 & 2 & -3 & 0 & 1 \\ 0 & 0 & 1 & -2 & 1 & 0 \end{bmatrix} \\
\xrightarrow{\text{row 2 + }(-2)\times\text{row 3}} &\begin{bmatrix} 3 & 5 & 1 & 1 & 0 & 0 \\ 0 & -5 & 0 & 1 & -2 & 1 \\ 0 & 0 & 1 & -2 & 1 & 0 \end{bmatrix} \\
\xrightarrow{\text{row 1 - row 3}} &\begin{bmatrix} 3 & 5 & 0 & 3 & -1 & 0 \\ 0 & -5 & 0 & 1 & -2 & 1 \\ 0 & 0 & 1 & -2 & 1 & 0 \end{bmatrix} \\
\xrightarrow{\text{row 1 + row 2}} &\begin{bmatrix} 3 & 0 & 0 & 4 & -3 & 1 \\ 0 & -5 & 0 & 1 & -2 & 1 \\ 0 & 0 & 1 & -2 & 1 & 0 \end{bmatrix} \\
\xrightarrow{\text{divide rows by 3, -5, 1 respectively}} &\begin{bmatrix} 1 & 0 & 0 & 4/3 & -1 & 1/3 \\ 0 & 1 & 0 & -1/5 & 2/5 & -1/5 \\ 0 & 0 & 1 & -2 & 1 & 0 \end{bmatrix}.
\end{aligned}
$$

So, the inverse is $A^{-1} = \begin{bmatrix} 4/3 & -1 & 1/3 \\ -1/5 & 2/5 & -1/5 \\ -2 & 1 & 0 \end{bmatrix}$. □

Remark 2.3.9. What about column operations? What if we have column switching? In fact, we can define column operations in a similar way like that in row operations. This time, the matrix you deal with becomes $\begin{bmatrix} A \\ I \end{bmatrix}$. However, we do not necessarily need it because the complexity of this algorithm is the same as that applied on $\begin{bmatrix} A & I \end{bmatrix}$.

Remark 2.3.10. What we have done is only $EA = I$. Do we necessarily have $AE = I$?

Solution. The answer is Yes.

We know that E is a product

$$E_1 E_2 \cdots E_k$$

of elementary matrices. For each elementary matrix, its inverse can be figured out without using the algorithm. In fact, you only need to reverse each row operation.

Therefore, if we have

$$E_1 E_2 \cdots E_k A = I,$$

then we apply the reversed row operations to the left-hand side and get

$$A = E_k^{-1} \cdots E_2^{-1} E_1^{-1}.$$

Then we have

$$AE_1 E_2 \cdots E_k = (E_k^{-1} \cdots E_2^{-1} E_1^{-1})(E_1 E_2 \cdots E_k) = I.$$

□

Remark 2.3.11. For two $n \times n$ matrices A and B, we can show that $(AB)^{-1} = B^{-1}A^{-1}$.

Homework 2.3.12. For the following 6 matrices, choose 5 of them and find their inverse matrices.

(a) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$

(d) $\begin{bmatrix} 2 & 4 & 8 \\ -2 & -4 & -7 \\ -4 & -11 & -16 \end{bmatrix}$

(e) $\begin{bmatrix} 0 & 7 & 1 \\ 1 & 5 & 1 \\ 3 & 15 & 5 \end{bmatrix}$

(f) $\begin{bmatrix} \cos(29°) & -\sin(29°) & 0 \\ \sin(29°) & \cos(29°) & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

Hint: Think about the formula

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

Chapter 3

Linear Spaces

3.1 Linear Spaces and Linear Maps

Starting from this chapter, we use $\mathbb{R}^n$ to denote the space of column vectors:

$$\left\{ \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} : x_1, x_2, \cdots, x_n \in \mathbb{R} \right\}.$$

(1) For the space of column vectors, we have defined an addition and this addition satisfies

(1.a) For any $\mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^n$, the summation $\mathbf{v} + \mathbf{w}$ is also in $\mathbb{R}^n$.

(1.b) There is a zero vector $\mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ satisfying that $\mathbf{0} + \mathbf{v} = \mathbf{v} + \mathbf{0} = \mathbf{v}$ for all $\mathbf{v} \in \mathbb{R}^n$;

(1.c) Associative law: For any $\mathbf{v}, \mathbf{w}, \mathbf{u} \in \mathbb{R}^n$, we have $(\mathbf{v} + \mathbf{w}) + \mathbf{u} = \mathbf{v} + (\mathbf{w} + \mathbf{u})$;

(1.d) Commutative law: For any $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, we have $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$;

(1.e) For every $\mathbf{v} \in \mathbb{R}^n$, there is a vector $-\mathbf{v}$ such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$.

(2) We also have a scalar multiplication which satisfies

(2.a) For any $r \in \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^n$, $r\mathbf{v}$ is still in $\mathbb{R}^n$;

(2.b) We have number 1 satisfying $1\mathbf{v} = \mathbf{v}$.

(2.c) Associative law: For any $r, s \in \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^n$, we have $(rs)\mathbf{v} = r(s\mathbf{v})$;

(2.d) Distributive law: For any $r, s \in \mathbb{R}$ and $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, we have $(r+s)\mathbf{v} = r\mathbf{v} + s\mathbf{v}$ and $r(\mathbf{v}+\mathbf{w}) = r\mathbf{v} + r\mathbf{w}$.

Generalize all the above properties, we give the definition of linear space.

Definition 3.1.1. (See [2]: Definition 1.1 on page 5) We call set V a linear space over $\mathbb{R}$ if:

(1) There is an addition defined on V and this addition satisfies

(1.a) For any elements $\mathbf{v}$ and $\mathbf{w}$ in V, the summation $\mathbf{v}+\mathbf{w}$ is also in V.

(1.b) There is a "zero" element $\mathbf{0}$ satisfying that $\mathbf{0}+\mathbf{v} = \mathbf{v}+\mathbf{0} = \mathbf{v}$ for all $\mathbf{v} \in V$;

(1.c) Associative law: For any $\mathbf{v}, \mathbf{w}, \mathbf{u} \in V$, we have $(\mathbf{v}+\mathbf{w})+\mathbf{u} = \mathbf{v}+(\mathbf{w}+\mathbf{u})$;

(1.d) Commutative law: For any $\mathbf{v}, \mathbf{w} \in V$, we have $\mathbf{v}+\mathbf{w} = \mathbf{w}+\mathbf{v}$;

(1.e) For every $\mathbf{v} \in V$, there is another element $-\mathbf{v}$ such that $\mathbf{v}+(-\mathbf{v}) = \mathbf{0}$.

(2) There is also a scalar multiplication on V which satisfies

(2.a) For any $r \in \mathbb{R}$ and $\mathbf{v} \in V$, $r\mathbf{v}$ is still in V;

(2.b) We have number $1 \in \mathbb{R}$ satisfying $1\mathbf{v} = \mathbf{v}$ for any $\mathbf{v} \in V$.

(2.c) Associative law: For any $r, s \in \mathbb{R}$ and $\mathbf{v} \in V$, we have $(rs)\mathbf{v} = r(s\mathbf{v})$;

(2.d) Distributive law: For any $r, s \in \mathbb{R}$ and $\mathbf{v}, \mathbf{w} \in V$, we have $(r+s)\mathbf{v} = r\mathbf{v} + s\mathbf{v}$ and $r(\mathbf{v}+\mathbf{w}) = r\mathbf{v} + r\mathbf{w}$.

Remark 3.1.2. We can also define a linear space on $\mathbb{C}$ just by replacing $\mathbb{R}$ with $\mathbb{C}$ in Definition 3.1.1

Remark 3.1.3. Elements in a linear space behave like column vectors. Therefore, we also call the elements in a linear space vectors. Some people also use the terminology "vector space" instead of our "linear space".

The verification of the following linear spaces is straightforward.

Example 3.1.4. $\mathbb{R}^n$ is already a linear space over $\mathbb{R}$.

Example 3.1.5. The collection $M_n(\mathbb{R})$ of all $n \times n$ real matrices is a linear space over $\mathbb{R}$.

Example 3.1.6. The collection of real polynomials is a linear space over $\mathbb{R}$.

Example 3.1.7. Any plane defined by equation $Ax + By + Cz = 0$ in $Oxyz$-space is a linear space over $\mathbb{R}$.

Example 3.1.8. The collection $C^\infty(\mathbb{R})$ of smooth function on $\mathbb{R}$ is a linear space over $\mathbb{R}$.

Example 3.1.9. Any plane defined by equation $Ax + By + Cz = D$ ($D \neq 0$) in $Oxyz$-space is NOT a linear space over $\mathbb{R}$.

Example 3.1.10. The space $\mathbb{R}_{\geq 0}$ is NOT a linear space over $\mathbb{R}$.

Homework 3.1.11. Show that the collection

$$\left\{ \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix} : a, b, c \in \mathbb{R} \right\}$$

of 3×3 real diagonal matrices is a linear space over $\mathbb{R}$.

Homework 3.1.12. Show that the collection

$$\left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x, y \geq 0 \right\}$$

of "positive" vectors is NOT a linear space over $\mathbb{R}$.

The most important maps between linear spaces are linear maps.

Definition 3.1.13. For two linear spaces V and W over $\mathbb{R}$ (or $\mathbb{C}$), a map

$$f : V \to W$$

is called a linear map if for any $r, s \in \mathbb{R}$ (or $\mathbb{C}$) and any element $\mathbf{v}, \mathbf{w} \in V$, we have

$$f(r\mathbf{v} + s\mathbf{w}) = rf(\mathbf{v}) + sf(\mathbf{w}).$$

Example 3.1.14. Let A be an $m \times n$ matrix, then the map

$$\begin{aligned} f : \mathbb{R}^n &\to \mathbb{R}^m \\ \mathbf{v} &\mapsto A\mathbf{v} \end{aligned}$$

is a linear map.

Example 3.1.15. Let $\mathbb{R}[x]$ be the collection of polynomials with real coefficients, then the map

$$\begin{aligned} f : \mathbb{R}[x] &\to \mathbb{R}[x] \\ a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 &\mapsto (a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0)^2 \end{aligned}$$

is NOT a linear map.

Example 3.1.16. Let $C^\infty(\mathbb{R})$ be the collection of smooth functions defined on $\mathbb{R}$, then the derivative

$$\frac{d}{dx} : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$$
$$g \mapsto \frac{dg}{dx}$$

is a linear map.

Homework 3.1.17. Show that the integral from 0 to 1

$$\int_0^1 : C^\infty(\mathbb{R}) \to \mathbb{R}$$
$$g \mapsto \int_0^1 g\,dx$$

is a linear map.

A quick and interesting observation about linear maps is as follows.

Proposition 3.1.18. Any linear map satisfies $f(\mathbf{0}) = \mathbf{0}$.

Proof. According to the definition of linear map, we have

$$f(r\mathbf{v} + s\mathbf{w}) = rf(\mathbf{v}) + sf(\mathbf{w}).$$

Let $r = s = 1$ and let $\mathbf{v} = \mathbf{w} = \mathbf{0}$, we get

$$f(\mathbf{0} + \mathbf{0}) = f(\mathbf{0}) + f(\mathbf{0}).$$

The left-hand side equals $f(\mathbf{0})$, so we get

$$f(\mathbf{0}) = f(\mathbf{0}) + f(\mathbf{0}).$$

Since the additive inverser $(-f(\mathbf{0}))$ always exists, we then have

$$f(\mathbf{0}) + (-f(\mathbf{0})) = f(\mathbf{0}) + f(\mathbf{0}) + (-f(\mathbf{0})),$$

i.e., $\mathbf{0} = f(\mathbf{0})$. □

3.2 Linear Independence and Bases

A linear space may have infinitely many elements. However, in the following examples, we do have ways to express all elements.

Example 3.2.1. In $\mathbb{R}^3$, we let $\mathbf{e}_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}, \mathbf{e}_2 = \begin{bmatrix}0\\1\\0\end{bmatrix}, \mathbf{e}_3 = \begin{bmatrix}0\\0\\1\end{bmatrix}$.

Then we see any vector

$$\mathbf{v} = \begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}$$

in $\mathbb{R}^3$ can be written as $x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3$.

Another way to express vectors in $\mathbb{R}^3$ is as follows. Let

$$\mathbf{v}_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix}0\\\frac{2}{3}\\0\end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix}0\\\frac{1}{3}\\-1\end{bmatrix}, \mathbf{v}_4 = \begin{bmatrix}0\\0\\1\end{bmatrix}.$$

Then $\mathbf{v} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_2\mathbf{v}_3 + (x_3 + x_2)\mathbf{v}_4$. So, we can say $\mathbb{R}^3$ is totally determined by

$$\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$$

or

$$\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}.$$

Example 3.2.2. In the linear space $\mathbb{R}[x]$ of real polynomials, any polynomial can be written into

$$a_0 + a_1x + a_2x^2 + \cdots + a_kx^k.$$

In other words, the set

$$\{1, x, x^2, x^3, \cdots, x^n, x^{n+1}, \cdots\}$$

fully determines $\mathbb{R}[x]$.

Example 3.2.3. In the linear space $M_2(\mathbb{R})$ of real 2×2 matrices, any matrix

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

is equal to

$$a\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}+b\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}+c\begin{bmatrix}0 & 0\\ 1 & 0\end{bmatrix}+d\begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}.$$

So, $M_2(\mathbb{R})$ is totally determined by

$$\left\{\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix},\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix},\begin{bmatrix}0 & 0\\ 1 & 0\end{bmatrix},\begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}\right\}.$$

The following result is proved by Zorn's lemma (see [2]: section 0.6 on page 4).

Theorem 3.2.4. For any linear space V over $\mathbb{R}$ or $\mathbb{C}$, we can find a subset $S \subseteq V$ so that each element $\mathbf{v} \in V$ is equal to

$$r_1\mathbf{v}_1 + r_2\mathbf{v}_2 + \cdots + r_k\mathbf{v}_k,$$

where $r_i \in \mathbb{R}$ and $\mathbf{v}_j \in S$.

Definition 3.2.5. (**generating set**) We call the set S in Theorem 3.2.4 a generating set of V.

Remark 3.2.6. From now on, we assume that generating sets are always finite. Therefore, examples like Examples 3.2.2 will not be taken into consideration any more.

Remark 3.2.7. According to Example 3.2.1, the generating set is not unique. For $\mathbb{R}^3$, the generating set can be

$$\left\{\begin{bmatrix}1\\ 0\\ 0\end{bmatrix},\begin{bmatrix}0\\ 1\\ 0\end{bmatrix},\begin{bmatrix}0\\ 0\\ 1\end{bmatrix}\right\},$$

or

$$\left\{\begin{bmatrix}1\\ 0\\ 0\end{bmatrix},\begin{bmatrix}0\\ \frac{2}{3}\\ 0\end{bmatrix},\begin{bmatrix}0\\ \frac{1}{3}\\ -1\end{bmatrix},\begin{bmatrix}0\\ 0\\ 1\end{bmatrix}\right\}.$$

In fact, you can construct many other generating sets.

The generating set is not unique. We hope to find the "best" generating set to study the linear space.

Example 3.2.8. In Example 3.2.1, the vector $\mathbf{v}_2$ and $\mathbf{v}_3$ has the relation that $\mathbf{v}_2 = 2\mathbf{v}_3 + 2\mathbf{v}_4$. So in fact, we can remove $\mathbf{v}_2$ from the generating set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$. This is because for any $\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + a_4\mathbf{v}_4$, we have

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2(2\mathbf{v}_3 + 2\mathbf{v}_4) + a_3\mathbf{v}_3 + a_4\mathbf{v}_4.$$

Therefore, we only need 3 vectors to generate all vectors in $\mathbb{R}^3$.

The ideas is, if one element $\mathbf{u}$ in S can be written as a linear combination of other elements in S, then we should remove $\mathbf{u}$ from S to reduce the number of elements in S.

Definition 3.2.9. Let V be a linear space over $\mathbb{R}$ (or $\mathbb{C}$). An element $\mathbf{u} \in V$ is called a linear combination of $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n \in V$ if there are real (or complex) numbers $a_1, a_2, \cdots, a_n$ satisfying

$$\mathbf{u} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + a_n\mathbf{u}_n.$$

Now, move $\mathbf{u}$ to the right-hand side, we get

$$-\mathbf{u} + a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + a_n\mathbf{u}_n = 0. \tag{3.2.1}$$

Immediately we see, at least one coefficient in equation (3.2.1) is now non-zero.

Definition 3.2.10. (**Linear dependence**) A set of elements $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ is linearly dependent if you can find scalars $a_1, a_2, \cdots, a_n$ such that

(a) At least one of $a_1, a_1, \cdots, a_n$ is non-zero;

(b) $a_1\mathbf{u}_1 + \cdots + a_n\mathbf{u}_n = 0$.

Meanwhile, we can define linear independence.

Definition 3.2.11. (**Linear independence**) A set of elements $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ is linearly independent if it is not linearly dependent.

Example 3.2.12. In Example 3.2.1, $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is linearly independent, while $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4$ is linearly dependent.

Solution. Suppose we have $a_1 \begin{bmatrix}1\\0\\0\end{bmatrix} + a_2 \begin{bmatrix}0\\1\\0\end{bmatrix} + a_3 \begin{bmatrix}0\\0\\1\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix}$. Then we get a homogeneous equation system

$$\begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\end{bmatrix}\begin{bmatrix}a_1\\a_2\\a_3\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix}.$$

Solve this system we get $a_1 = a_2 = a_3 = 0$. Similarly, suppose we have

$$a_1 \begin{bmatrix}1\\0\\0\end{bmatrix} + a_2 \begin{bmatrix}0\\\frac{2}{3}\\0\end{bmatrix} + a_3 \begin{bmatrix}0\\\frac{1}{3}\\-1\end{bmatrix} + a_4 \begin{bmatrix}0\\0\\1\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix},$$

then we get a homogeneous equation system

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{2}{3} & \frac{1}{3} & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Solve this equation we get

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{1}{2}a_4 \\ a_4 \\ a_4 \end{bmatrix},$$

where a_4 is free. If we choose $a_4 = 1$, then we have at least one non-zero coefficient. □

Example 3.2.13. What about the generating set in Example 3.2.3? Independent or dependent?

Solution. Suppose we have

$$a_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + a_2 \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + a_3 \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + a_4 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Add up everything on the left-hand side, we get

$$\begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Therefore $a_1 = a_2 = a_3 = a_4 = 0$ and the generating set is independent. □

In general, there is no unified way to prove that a set of vectors is independent. However, for column vectors, we have already had a way to do it.

Suppose that $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n$ are column vectors in $\mathbb{R}^m$. Then

$$a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + a_n\mathbf{u}_n = 0$$

can be written into

$$\begin{bmatrix} \mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = 0.$$

Then everything is the same as solving linear equations. If the only solution is

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

then $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n$ form a linearly independent set. Otherwise, they form a linearly dependent set.

Example 3.2.14. Show that vectors $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$ form a linearly independent set in $\mathbb{R}^3$.

Solution. Write down equation system

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

The only solution is $a_1 = a_2 = a_3 = 0$. □

Homework 3.2.15. Show that vectors $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$ are linearly dependent in $\mathbb{R}^3$.

Homework 3.2.16. Find a generating set for the space $M_3(\mathbb{R})$ of real 3×3 matrices.

Homework 3.2.17. Independent or dependent: $\begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 6 \\ 2 \end{bmatrix}$.

Homework 3.2.18. Independent or dependent: $\begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$.

Remark 3.2.19. If $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ is linearly independent, then none of them can be written as a linear combination of others. This is because if some

$$\mathbf{u}_i = a_1\mathbf{u}_1 + \cdots a_{i-1}\mathbf{u}_{i-1} + a_{i+1}\mathbf{u}_{i+1} + \cdots + a_n\mathbf{u}_n,$$

then we have

$$a_1\mathbf{u}_1 + \cdots a_{i-1}\mathbf{u}_{i-1} - \mathbf{u}_i + a_{i+1}\mathbf{u}_{i+1} + \cdots + a_n\mathbf{u}_n = \mathbf{0}.$$

In other words, at least one of these coefficients is nonzero. This contradicts to the condition of independence.

By Remark 3.2.19, we see that when elements in a generating set are independent, then none of them can be written as a combination of others. In other words, none of them can be removed.

Definition 3.2.20. Let $S = \{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ be a generating set of a linear space V. Then we call S a basis of V if $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ is linearly independent.

Proposition 3.2.21. If $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ is linearly independent, then

$$a_1\mathbf{u}_1 + \cdots a_n\mathbf{u}_n = b_1\mathbf{u}_1 + \cdots b_n\mathbf{u}_n$$

implies $a_1 = b_1, a_2 = b_2, \cdots, a_n = b_n$.

Proof. If some $a_i \neq b_i$, then at least one coefficient in

$$(a_1 - b_1)\mathbf{v}_1 + \cdots + (a_n - b_n)\mathbf{v}_n = 0$$

is non-zero. This contradicts to that condition of independence. □

Proposition 3.2.21 shows that any element in a linear space has the unique expression under a basis.

Homework 3.2.22. In Example 3.2.1 and Example 3.2.3, point out which generating set is a basis.

Homework 3.2.23. Find a basis of the space $M_3(\mathbb{R})$ of real 3×3 matrices.

We end this section by another result proved by Zorn's lemma.

Theorem 3.2.24. If we are given a linearly independent list

$$\{\mathbf{u}_1, \cdots, \mathbf{u}_m\},$$

then we can put more elements $\mathbf{u}_{m+1}, \mathbf{u}_{m+2}, \cdots$ into it to make it a basis.

3.3 Bases and Dimensions

In fact, bases are still not unique.

Example 3.3.1. The space $\mathbb{R}^3$. We can either use $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$, where

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

or we can use $\{-\mathbf{e}_1, -\mathbf{e}_2, -\mathbf{e}_3\}$ instead.

Example 3.3.2. The space $M_2(\mathbb{R})$ of 2×2 real matrices. Both

$$\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

and

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

are bases of $M_2(\mathbb{R})$.

Example 3.3.3. The Oxy plane. This space can be viewed as $\mathbb{R}^2$. Like $\mathbb{R}^3$, the bases of $\mathbb{R}^2$ are also not unique.

Fortunately, each basis contains the same amount of elements:

Theorem 3.3.4. If $S_1 = \{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n\}$ and $S_2 = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m\}$ are both bases of the same linear space V, then $m = n$.

Proof. We illustrate the idea of the proof in the simplest case: Suppose $S_1 = \{\mathbf{u}_1\}$ and $S_2 = \{\mathbf{v}_1, \mathbf{v}_2\}$ are two bases. Then we can write $\mathbf{v}_1 = r_1\mathbf{u}_1$ and $\mathbf{v}_2 = r_2\mathbf{u}_2$, where u_1 and u_2 are real numbers. It is clear that at least one of r_1 and r_2 is non-zero (Why?) This means $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly dependent. Contradiction! $\square$

Since the number of elements in bases are the same, we can define:

Definition 3.3.5. (**dimension**) The number of elements in a basis S of a linear space V is called the dimension of V, or we say that V is an n-dimensional linear space. The notation is $\dim V = n$.

Example 3.3.6. $\mathbb{R}^3$ is a 3-dimensional linear space.

In general, $\mathbb{R}^n$ is an n-dimensional linear space.

Example 3.3.7. $M_2(\mathbb{R})$ is a 4-dimensional linear space.

In general, $M_n(\mathbb{R})$ is an n^2-dimensional linear space.

Example 3.3.8. The Oxy plane is a 2-dimensional linear space.

In this book, we stay in finitely dimensional cases. For infinitely dimensional topics, you should learn functional analysis.

We here introduce another technique to see if a set of elements is a basis. Originally, to verify a set

$$\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n\}$$

is a basis, we should verify two things:

(1) It is a generating set;

(2) It is linearly independent.

Now, suppose that we have already known $\dim V = n$. Then by Theorem 3.2.24 and Theorem 3.3.4, to verify that

$$\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n\}$$

is a basis of V, you only need to verify that it is linearly independent.

Example 3.3.9. Given three column vectors

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

they form a basis of $\mathbb{R}^3$. However, vectors

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

do not form a basis of $\mathbb{R}^3$.

Solution. This is because the solution of

$$\begin{bmatrix} 1 & 2 & 0 \\ 0 & 3 & 0 \\ 2 & 4 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

is $a_1 = a_2 = a_3 = 0$. These three vectors form an independent set and then must form a basis of $\mathbb{R}^3$ since a basis of $\mathbb{R}^3$ has three vectors,

However, the solution of

$$\begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 3 \\ 1 & -2 & -1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

is $\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} -a_3 \\ -a_3 \\ a_3 \end{bmatrix} = a_3 \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$. They do not form an independent set, so they do not form a basis. □

Example 3.3.10. Given four column vectors

$$\begin{bmatrix} 1.414 \\ 3.14 \\ 0.618 \end{bmatrix}, \begin{bmatrix} 2.718 \\ \ln 2 \\ \sqrt{2} \end{bmatrix}, \begin{bmatrix} e^2 \\ \cos 29^\circ \\ \sin 29^\circ \end{bmatrix}, \begin{bmatrix} \pi^2 \\ -1 \\ \tan 77^\circ \end{bmatrix}$$

in $\mathbb{R}^3$, no doubt they form a linearly dependent family. So, they do not form a basis.

What about two of them? Do any two of them form a basis of $\mathbb{R}^3$? The answer is NO since any basis of $\mathbb{R}^3$ should have three vectors.

We summarize everything we get about dependence, independence and basis.

Remark 3.3.11. Given a linear space V and a family $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\} \subseteq V$:

1. If $\dim V \geq k$ and the family is linearly independent:
   (a) If $\dim V = k$, then the family must be a basis.
   (b) If $\dim V > k$, then at present we can do nothing.
2. If $\dim V \geq k$ and the family is linearly dependent, then at present we can do nothing.
3. If $\dim V < k$, the family must be linearly dependent. At present we can do nothing.
4. If we do not know the value of $\dim V$, then we can only verify if the family is dependent or independent. If it is independent, we can try to verify if it is a generating set to see if it is a basis.

Homework 3.3.12. Show that vectors $\begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix}$ also form a basis of $\mathbb{R}^3$.

Homework 3.3.13. Let $\mathbf{v}_0 = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \mathbf{v}_1 = \begin{bmatrix} 1.5 \\ 2.5 \\ 3.5 \\ 4.5 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 2 \\ 3 \\ 4 \\ 5 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 2.5 \\ 3.5 \\ 4.5 \\ 5.5 \end{bmatrix}, \mathbf{v}_4 = \begin{bmatrix} 3 \\ 4 \\ 5 \\ 6 \end{bmatrix}$, do they form a basis of $\mathbb{R}^4$?

3.4 Change of Basis

Since bases are not unique, we need to talk about change of basis. In other words, the relation between two different bases of the same linear space V.

Suppose we have two basis

$$\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$$

and

$$\{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_n\}$$

of an n-dimensional space V. Then by the definition of basis, we can write each element in $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ into a linear combination of elements in $\{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_n\}$. In other words, we have

$$\begin{aligned}
\mathbf{v}_1 &= a_{11}\mathbf{w}_1 + a_{21}\mathbf{w}_2 + a_{31}\mathbf{w}_3 + \cdots + a_{n1}\mathbf{w}_n \\
\mathbf{v}_2 &= a_{12}\mathbf{w}_1 + a_{22}\mathbf{w}_2 + a_{32}\mathbf{w}_3 + \cdots + a_{n2}\mathbf{w}_n \\
\mathbf{v}_3 &= a_{13}\mathbf{w}_1 + a_{23}\mathbf{w}_2 + a_{33}\mathbf{w}_3 + \cdots + a_{n3}\mathbf{w}_n \\
&\vdots \\
\mathbf{v}_n &= a_{1n}\mathbf{w}_1 + a_{2n}\mathbf{w}_2 + a_{3n}\mathbf{w}_3 + \cdots + a_{nn}\mathbf{w}_n.
\end{aligned}$$

Write these relations into a matrix equation, we get

$$\begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n\end{bmatrix} = \begin{bmatrix}\mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_n\end{bmatrix}\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn}\end{bmatrix}. \tag{3.4.1}$$

We want to find that coefficient matrix in (3.4.1). How do we get it?

In fact, we do not have a unified algorithm to find the matrix for all linear spaces. However, for $\mathbb{R}^n$, the space of column vectors, the two arrays

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

and

$$\begin{bmatrix}\mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_n\end{bmatrix}$$

are two $n \times n$ matrices. Then the answer becomes quite straightforward: Find the inverse of $\begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_n \end{bmatrix}$, then the coefficient matrix in (3.4.1) is equal to

$$\begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_n \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}.$$

In fact, when we find the inverse of a matrix, we only use row operations. Here, when we really do calculations, we only need row operations as well.

Let $A = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}$, $B = \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_n \end{bmatrix}$ and

$$C = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{bmatrix}.$$

First, we write down a partitioned matrix

$$\begin{bmatrix} B & | & A \end{bmatrix}.$$

Then we apply the same row operation steps as those in finding the inverse of a matrix and turn B into the identity matrix I. Then we get

$$\begin{bmatrix} I & | & B^{-1}A \end{bmatrix}.$$

This $B^{-1}A$ is the coefficient matrix C that we need. We call this C a change of basis matrix.

Example 3.4.1. In $\mathbb{R}^2$, Let

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 4 \\ 2 \end{bmatrix}, \mathbf{w}_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \mathbf{w}_2 = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

(a) Show that $\{\mathbf{v}_1, \mathbf{v}_2\}$ and $\{\mathbf{w}_1, \mathbf{w}_2\}$ are both bases of $\mathbb{R}^2$.

(b) Find the change of basis matrix C satisfying

$$\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 \end{bmatrix} C.$$

Solution. (a) Show that they are both linearly independent sets.

(b) Write down matrix

$$\begin{bmatrix}\mathbf{w}_1 & \mathbf{w}_2 & | & \mathbf{v}_1 & \mathbf{v}_2\end{bmatrix} = \left[\begin{array}{cc|cc} 3 & 3 & 1 & 4 \\ 2 & 3 & 2 & 2 \end{array}\right].$$

Use row operations to turn the left side into an identity matrix, we get

$$\left[\begin{array}{cc|cc} 1 & 0 & -1 & 2 \\ 0 & 1 & 4/3 & -2/3 \end{array}\right].$$

Then $C = \begin{bmatrix} -1 & 2 \\ 4/3 & -2/3 \end{bmatrix}$. □

Example 3.4.2. Let

$$\mathbf{v}_1 = \begin{bmatrix}0\\2\\0\end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix}2\\0\\4\end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix}0\\0\\1\end{bmatrix}, \mathbf{w}_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}, \mathbf{w}_2 = \begin{bmatrix}1\\1\\0\end{bmatrix}, \mathbf{w}_3 = \begin{bmatrix}1\\1\\-1\end{bmatrix}.$$

We have verified that $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ is a basis of $\mathbb{R}^3$.

(a) Verify that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is also basis of $\mathbb{R}^3$.

(b) Find the matrix C satisfying $\begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3\end{bmatrix} = \begin{bmatrix}\mathbf{w}_1 & \mathbf{w}_2 & \mathbf{w}_3\end{bmatrix} C$.

Solution. (a) Show that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly independent.

(b) Write down matrix

$$\left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 0 & 2 & 0 \\ 0 & 1 & 1 & 2 & 0 & 0 \\ 0 & 0 & -1 & 0 & 4 & 1 \end{array}\right].$$

Then row operations give us

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -2 & 2 & 0 \\ 0 & 1 & 0 & 2 & 4 & 1 \\ 0 & 0 & 1 & 0 & -4 & -1 \end{array}\right].$$

Thus $C = \begin{bmatrix} -2 & 2 & 0 \\ 2 & 4 & 1 \\ 0 & -4 & -1 \end{bmatrix}$. □

Homework 3.4.3. Let

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 2 \\ 4 \\ 4 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 0 \\ 4 \\ 4 \end{bmatrix}, \mathbf{w}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{w}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \mathbf{w}_3 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}.$$

We have verified that $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ is a basis of $\mathbb{R}^3$.

(a) Verify that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is also a basis of $\mathbb{R}^3$.

(b) Find the matrix C satisfying

$$\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \mathbf{w}_3 \end{bmatrix} C.$$

3.5 Coordinates

For a linear space V with $\dim V = n$, if we fix a basis $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$, then any $\mathbf{v}$ can be expressed uniquely (see Proposition 3.2.21)as

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n.$$

Definition 3.5.1. We call the array

$$(a_1, a_2, \cdots, a_n)$$

the coordinate of $\mathbf{v}$ relative to basis $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$. In addition, we can write it into a column vector

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}.$$

Example 3.5.2. Let $\mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ be a column vector in $\mathbb{R}^3$. Then relative to the basis formed by

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

$\mathbf{v}$ has coordinate $(1, 2, 3)$ or $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. Therefore, column vectors in $\mathbb{R}^3$ (more generally, in $\mathbb{R}^n$) is equal to its coordinate relative to the standard basis.

Example 3.5.3. In the space $M_2(\mathbb{R})$ of 2×2 real matrices, the coordinate of each matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ relative to the standard basis formed by

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

is (a, b, c, d) or $\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$.

Homework 3.5.4. Find a basis of space $\{\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} : a, b \in \mathbb{R}\}$ and then write down the coordinate of $\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$ relative to the basis you find.

However, if we change the basis, then the coordinate of $\mathbf{v}$ will also change. Suppose we have two basis

$$\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$$

and

$$\{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_n\},$$

then we have a matrix C such that

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

Suppose the coordinate of $\mathbf{v}$ relative to $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ is $\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$, then we have

$$\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_n \end{bmatrix} C \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}.$$

Therefore, the coordinate of $\mathbf{v}$ relative to basis $\{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_n\}$ is column vector

$$C \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}.$$

Example 3.5.5. In $\mathbb{R}^2$, Let

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 4 \\ 2 \end{bmatrix}, \mathbf{w}_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \mathbf{w}_2 = \begin{bmatrix} 3 \\ 3 \end{bmatrix}.$$

We've known that $\{\mathbf{e}_1, \mathbf{e}_2\}$, $\{\mathbf{v}_1, \mathbf{v}_2\}$, $\{\mathbf{w}_1, \mathbf{w}_2\}$ are bases of $\mathbb{R}^2$.

(a) Find the matrices C and D satisfying

$$\begin{bmatrix} \mathbf{e}_1, \mathbf{e}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1, \mathbf{v}_2 \end{bmatrix} C$$

and

$$\begin{bmatrix} \mathbf{e}_1, \mathbf{e}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{w}_1, \mathbf{w}_2 \end{bmatrix} D.$$

(b) Find the coordinates of $\mathbf{u} = \begin{bmatrix} 2 \\ 6 \end{bmatrix}$ relative to $\{\mathbf{v}_1, \mathbf{v}_2\}$ and $\{\mathbf{w}_1, \mathbf{w}_2\}$ respectively.

(c) Suppose we have a vector $\mathbf{v} = 2\mathbf{v}_1 + 3\mathbf{v}_2$. Find the coordinate of $\mathbf{v}$ relative to $\{\mathbf{w}_1, \mathbf{w}_2\}$.

Solution. (a) Write down matrix

$$\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & | & \mathbf{e}_1 & \mathbf{e}_2 \end{bmatrix} = \left[\begin{array}{cc|cc} 1 & 4 & 1 & 0 \\ 2 & 2 & 0 & 1 \end{array}\right].$$

Use row operations to turn the left side into an identity matrix, we get

$$\left[\begin{array}{cc|cc} 1 & 0 & -1/3 & 2/3 \\ 0 & 1 & 1/3 & -1/6 \end{array}\right].$$

Then $C = \begin{bmatrix} -1/3 & 2/3 \\ 1/3 & -1/6 \end{bmatrix}$. Similarly, $D = \begin{bmatrix} 1 & -1 \\ -2/3 & 1 \end{bmatrix}$.

(b) $\mathbf{u} = 2\mathbf{e}_1 + 6\mathbf{e}_2$. Thus the new coordinate of $\mathbf{u}$ relative to basis $\{\mathbf{v}_1, \mathbf{v}_2\}$ is

$$C \begin{bmatrix} 2 \\ 6 \end{bmatrix} = \begin{bmatrix} 10/3 \\ -1/3 \end{bmatrix}.$$

Similarly, the new coordinate of $\mathbf{u}$ relative to basis $\{\mathbf{w}_1, \mathbf{w}_2\}$ is

$$D \begin{bmatrix} 2 \\ 6 \end{bmatrix} = \begin{bmatrix} -4 \\ 14/3 \end{bmatrix}.$$

(c) In Example 3.4.1, we find that the matrix M satisfying $\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 \end{bmatrix} M$ is

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

Thus the coordinate of vector $\mathbf{v} = 2\mathbf{v}_1 + 3\mathbf{v}_2$ relative to $\{\mathbf{w}_1, \mathbf{w}_2\}$ is

$$M \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 4 \\ 2/3 \end{bmatrix}$$

□

Example 3.5.6. Let

$$\mathbf{v}_1 = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 2 \\ 0 \\ 4 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \mathbf{w}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{w}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \mathbf{w}_3 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}.$$

We have verified that both $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ and $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ are bases of $\mathbb{R}^3$.

(a) Find the coordinates of $\mathbf{u} = \begin{bmatrix} 2 \\ 6 \\ 1 \end{bmatrix}$ relative to $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ and $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ respectively.

(b) Let $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2 + 2\mathbf{v}_3$. Find the coordinate of $\mathbf{v}$ relative to $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$.

Solution. (a) We want to first figure out matrices C and D that satisfies

$$\begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} C.$$

Write down matrix

$$\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & | & \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \end{bmatrix} = \left[\begin{array}{ccc|ccc} 0 & 2 & 0 & 1 & 0 & 0 \\ 2 & 0 & 0 & 0 & 1 & 0 \\ 0 & 4 & 1 & 0 & 0 & 1 \end{array}\right].$$

Row operations gives us

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

Thus we get

$$C = \begin{bmatrix} 0 & 1/2 & 0 \\ 1/2 & 0 & 0 \\ -2 & 0 & 1 \end{bmatrix}.$$

Similarly,

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

Notice that

$$\mathbf{u} = 2\mathbf{e}_1 + 6\mathbf{e}_2 + \mathbf{e}_3 = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \end{bmatrix} \begin{bmatrix} 2 \\ 6 \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} C \begin{bmatrix} 2 \\ 6 \\ 1 \end{bmatrix},$$

we see that the coordinate of $\mathbf{u}$ relative to $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is

$$C \begin{bmatrix} 2 \\ 6 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ -3 \end{bmatrix}.$$

Similarly, the coordinate of $\mathbf{u}$ relative to $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ is

$$D \begin{bmatrix} 2 \\ 6 \\ 1 \end{bmatrix} = \begin{bmatrix} -4 \\ 7 \\ -1 \end{bmatrix}.$$

(b) Write down matrix

$$\begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \mathbf{w}_3 & | & \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} = \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 0 & 2 & 0 \\ 0 & 1 & 1 & 2 & 0 & 0 \\ 0 & 0 & -1 & 0 & 4 & 1 \end{array}\right]$$

and then row operations give us

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -2 & 2 & 0 \\ 0 & 1 & 0 & 2 & 4 & 1 \\ 0 & 0 & 1 & 0 & -4 & -1 \end{array}\right].$$

Thus we have matrix $M = \begin{bmatrix} -2 & 2 & 0 \\ 2 & 4 & 1 \\ 0 & -4 & -1 \end{bmatrix}$ satisfying $\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \mathbf{w}_3 \end{bmatrix} M$. Thus we have

$$\mathbf{v} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \mathbf{w}_3 \end{bmatrix} M \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}.$$

Thus the coordinate of $\mathbf{v}$ relative to $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ is

$$M \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 8 \\ -6 \end{bmatrix}.$$

□

Homework 3.5.7. Let

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 2 \\ 4 \\ 4 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 0 \\ 4 \\ 4 \end{bmatrix}, \mathbf{w}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{w}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \mathbf{w}_3 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}.$$

We have verified that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ and $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ are both bases of $\mathbb{R}^3$.

(a) Find the coordinates of $\mathbf{u} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$ relative to $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ and $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ respectively.

(b) Let $\mathbf{v} = \mathbf{v}_1 + 2\mathbf{v}_2 + \mathbf{v}_3$. Find the coordinate of $\mathbf{v}$ relative to $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$.

An interesting observation is that when we fix bases, every linear map is a matrix acting on coordinates.

Suppose $f : V \to W$ is a linear map. Let $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ and $\{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_m\}$ be bases of V and W respectively. Then by the definition of basis, we can write $f(\mathbf{v}_i)$'s into

$$\begin{aligned}
f(\mathbf{v}_1) &= b_{11}\mathbf{w}_1 + b_{21}\mathbf{w}_2 + b_{31}\mathbf{w}_3 + \cdots + b_{m1}\mathbf{w}_n \\
f(\mathbf{v}_2) &= b_{12}\mathbf{w}_1 + b_{22}\mathbf{w}_2 + b_{32}\mathbf{w}_3 + \cdots + b_{m2}\mathbf{w}_n \\
f(\mathbf{v}_3) &= b_{13}\mathbf{w}_1 + b_{23}\mathbf{w}_2 + b_{33}\mathbf{w}_3 + \cdots + b_{m3}\mathbf{w}_n \\
&\vdots \\
f(\mathbf{v}_n) &= b_{1n}\mathbf{w}_1 + b_{2n}\mathbf{w}_2 + b_{3n}\mathbf{w}_3 + \cdots + b_{mn}\mathbf{w}_n.
\end{aligned}$$

Then we rewrite them into matrices and get

$$\begin{bmatrix} f(\mathbf{v}_1) & f(\mathbf{v}_2) & \cdots & f(\mathbf{v}_n) \end{bmatrix} = \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_n \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & b_{13} & \cdots & b_{1n} \\ b_{21} & b_{22} & b_{23} & \cdots & b_{2n} \\ b_{31} & b_{32} & b_{33} & \cdots & b_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ b_{m1} & b_{m2} & b_{m3} & \cdots & b_{mn} \end{bmatrix}. \tag{3.5.1}$$

Now, for any $\mathbf{v} \in V$, if its coordinate relative to $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is $(x_1, x_2, \cdots, x_n)$, then the linear map gives us

$$\begin{aligned} f(\mathbf{v}) &= f(x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n) && (3.5.2) \\ &= x_1 f(\mathbf{v}_1) + x_2 f(\mathbf{v}_2) + \cdots + x_n f(\mathbf{v}_n). && (3.5.3) \\ &= \begin{bmatrix} f(\mathbf{v}_1) & f(\mathbf{v}_2) & \cdots & f(\mathbf{v}_n) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} && (3.5.4) \end{aligned}$$

Combine (3.5.1) and (3.5.4), we get

$$\begin{aligned} f(\mathbf{v}) &= \begin{bmatrix} f(\mathbf{v}_1) & f(\mathbf{v}_2) & \cdots & f(\mathbf{v}_n) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} && (3.5.5) \\ &= \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_n \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & b_{13} & \cdots & b_{1n} \\ b_{21} & b_{22} & b_{23} & \cdots & b_{2n} \\ b_{31} & b_{32} & b_{33} & \cdots & b_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ b_{m1} & b_{m2} & b_{m3} & \cdots & b_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}. && (3.5.6) \end{aligned}$$

So, we see that the coordinate $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ of $\mathbf{v}$ relative to $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ is transformed to another coordinate

$$\begin{bmatrix} b_{11} & b_{12} & b_{13} & \cdots & b_{1n} \\ b_{21} & b_{22} & b_{23} & \cdots & b_{2n} \\ b_{31} & b_{32} & b_{33} & \cdots & b_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ b_{m1} & b_{m2} & b_{m3} & \cdots & b_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

relative to $\{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_m\}$. Therefore, to study a linear map between finitely dimensional spaces V and W, we usually:

(a) Find suitable bases of V and W respectively.

(b) Determine the corresponding matrix in (3.5.1).

(c) For every vector in V, find its coordinate relative to the basis of V and then multiply it by the matrix.

With this idea in mind, we know that finitely dimensional linear spaces are no more than $\mathbb{R}^n$ and linear maps between them are no more than a matrix.

3.6 Images, Kernels and Ranks

To study the action of a linear map

$$f : V \to W,$$

we need to study what $f(\mathbf{v})$ looks like in W for each $\mathbf{v} \in V$. This is the image space of f.

Definition 3.6.1. The image space of a linear map $f : V \to W$ is defined to be

$$\{f(\mathbf{v}) : \mathbf{v} \in V\}.$$

The notation is $\operatorname{im}(f)$.

Remark 3.6.2. The image space of a linear map is a linear space.

Let $n = \dim V$ and $m = \dim W$. Since we know that each linear map corresponds to a matrix A, the map f is equivalent to

$$\begin{aligned} \mathbb{R}^n &\to \mathbb{R}^m \\ \mathbf{v} &\mapsto A\mathbf{v}, \end{aligned}$$

where A is the $m \times n$ matrix corresponding to f when bases of V and W are chosen. Therefore, we only need to study how matrices transform vectors.

Definition 3.6.3. The image space of an $m \times n$ matrix A is the space of vectors like $A\mathbf{v}$, where $\mathbf{v} \in \mathbb{R}^n$. More precisely,

$$\operatorname{im}(A) = \{A\mathbf{v} : \mathbf{v} \in \mathbb{R}^n\}.$$

Using the multiplication rule of matrices, we can find a generating set of $\operatorname{im}(A)$.

Proposition 3.6.4. A generating set of an $m \times n$ matrix A is the collection of its columns.

Proof. Write $A = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{bmatrix}$. Since column vectors in $\mathbb{R}^n$ are all like $\mathbf{v} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$, where $x_1, x_2, \cdots, x_n$ are any real numbers, we get

$$A\mathbf{v} = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + \cdots + x_n\mathbf{b}_n.$$

□

Since $\text{im}(A)$ is a linear space, we would like to know its dimension or even directly find a basis. The generating set $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ may not be a basis, but we can manipulate it to get a basis.

Homework 3.6.5. Show that the columns of matrix

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

is linearly dependent. Therefore, the columns of A do not form a basis of $\text{im}(A)$.

Recall Example 3.2.1 and Example 3.2.8, at that time when we find $\mathbf{v}_2 = 2\mathbf{v}_3 + 2\mathbf{v}_4$, we throw $\mathbf{v}_2$ away from the generating set. Therefore, the idea is, if we add up the linear combination of some vectors and then find that the summation is another vector $\mathbf{v}$, then $\mathbf{v}$ is eliminated.

Since here we are doing scalar multiplications and additions on column vectors, we get the definition of column operations:

Definition 3.6.6. There are three column operations on matrices:

(1) column reduction: Replace column i by "$r\times$ column j + column i".

(2) column switching: Swap column i and column j.

(3) column scalar multiplication: Multiply column i by a scalar r.

Example 3.6.7. Let $E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix}$, $E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ and $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$.

(a) Multiply column 2 of A by 3 and compare the result with AE_1.

(b) Replace column 1 by "2$\times$ column 3 + column 1", then compare the result with AE_2.

(c) Swap column 2 and column 3 and then compare the result with AE_3.

Solution. (a) They are the same.

(b) They are the same.

(c) They are the same.

□

Remark 3.6.8. Column operations are the same as multiplying the matrix by elementary matrices from the right side.

Proposition 3.6.9. Column operations do not change $\mathrm{im}(A)$.

Proof. Let E be a matrix representing column operations. For any vector $A\mathbf{v} \in \mathrm{im}(A)$, where $\mathbf{v}$ is a column vector, we see that

$$A\mathbf{v} = AE(E^{-1}\mathbf{v}) \in \mathrm{im}(AE).$$

Therefore, $\mathrm{im}(A) \subseteq \mathrm{im}(AE)$. Similarly, we can prove $\mathrm{im}(AE) \subseteq \mathrm{im}(A)$. □

With Proposition 3.6.9 in mind, we can use column operations to find a basis of $\mathrm{im}(A)$. We use three examples to show the algorithm.

Example 3.6.10. Find a basis of $\mathrm{im}(A)$ for the following matrices A.

(a) $A = \begin{bmatrix} 1 & 2 & 0 & -3 \\ 4 & 6 & 2 & -8 \\ 4 & 7 & 6 & -5 \\ 3 & 4 & 1 & -6 \end{bmatrix}$;

(b) $A = \begin{bmatrix} 1 & 2 & -1 & -3 \\ 4 & 6 & -4 & -8 \\ 4 & 7 & -4 & -5 \\ 3 & 4 & -2 & -6 \end{bmatrix}$;

(c) $A = \begin{bmatrix} 1 & 2 & -1 & -3 & -2 \\ 4 & 6 & -4 & -8 & -8 \\ 4 & 7 & -4 & -10 & -8 \\ 3 & 4 & -2 & -6 & -7 \end{bmatrix}$.

Solution. (a)

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

$$\xrightarrow[\text{column 2 }+(-2)\times\text{column 1}]{\text{replace column 2 by}}\begin{bmatrix}1&0&0&-3\\4&-2&2&-8\\4&-1&6&-5\\3&-2&1&-6\end{bmatrix}$$

$$\xrightarrow[\text{column 4 }+3\times\text{column 1}]{\text{replace column 4 by}}\begin{bmatrix}1&0&0&0\\4&-2&2&4\\4&-1&6&7\\3&-2&1&3\end{bmatrix}$$

$$\xrightarrow[\text{column 3 }+\text{column 2}]{\text{replace column 3 by}}\begin{bmatrix}1&0&0&0\\4&-2&0&4\\4&-1&5&7\\3&-2&-1&3\end{bmatrix}$$

$$\xrightarrow[\text{column 4 }+2\times\text{column 2}]{\text{replace column 4 by}}\begin{bmatrix}1&0&0&0\\4&-2&0&0\\4&-1&5&5\\3&-2&-1&-1\end{bmatrix}$$

$$\xrightarrow[\text{column 4 }+(-1)\times\text{column 3}]{\text{replace column 4 by}}\begin{bmatrix}1&0&0&0\\4&-2&0&0\\4&-1&5&0\\3&-2&-1&0\end{bmatrix}.$$

Then we get a basis

$$\left\{\begin{bmatrix}1\\4\\4\\3\end{bmatrix},\begin{bmatrix}0\\-2\\-1\\-2\end{bmatrix},\begin{bmatrix}0\\0\\5\\-1\end{bmatrix}\right\}.$$

(b)

$$\begin{bmatrix}1&2&-1&-3\\4&6&-4&-8\\4&7&-4&-5\\3&4&-2&-6\end{bmatrix}$$

$$\xrightarrow[\text{column 2 +(−2)×column 1}]{\text{replace column 2 by}}\begin{bmatrix}1&0&-1&-3\\4&-2&-4&-8\\4&-1&-4&-5\\3&-2&-2&-6\end{bmatrix}$$

$$\xrightarrow[\text{column 3 +column 1}]{\text{replace column 3 by}}\begin{bmatrix}1&0&0&-3\\4&-2&0&-8\\4&-1&0&-5\\3&-2&1&-6\end{bmatrix}$$

$$\xrightarrow[\text{column 4 +3×column 1}]{\text{replace column 4 by}}\begin{bmatrix}1&0&0&0\\4&-2&0&4\\4&-1&0&7\\3&-2&1&3\end{bmatrix}$$

$$\xrightarrow[\text{column 4 +2×column 2}]{\text{replace column 4 by}}\begin{bmatrix}1&0&0&0\\4&-2&0&0\\4&-1&0&5\\3&-2&1&-1\end{bmatrix}$$

$$\xrightarrow{\text{swap column 3 and column 4}}\begin{bmatrix}1&0&0&0\\4&-2&0&0\\4&-1&5&0\\3&-2&-1&1\end{bmatrix}.$$

Thus we get a basis

$$\left\{\begin{bmatrix}1\\4\\4\\3\end{bmatrix},\begin{bmatrix}0\\-2\\-1\\-2\end{bmatrix},\begin{bmatrix}0\\0\\5\\-1\end{bmatrix},\begin{bmatrix}0\\0\\0\\1\end{bmatrix}\right\}.$$

(c)

$$
\begin{aligned}
&\begin{bmatrix} 1 & 2 & -1 & -3 & -2 \\ 4 & 6 & -4 & -8 & -8 \\ 4 & 7 & -4 & -10 & -8 \\ 3 & 4 & -2 & -6 & -7 \end{bmatrix} \\
\xrightarrow[\text{column 2 +(−2)×column 1}]{\text{replace column 2 by}}
&\begin{bmatrix} 1 & 0 & -1 & -3 & -2 \\ 4 & -2 & -4 & -8 & -8 \\ 4 & -1 & -4 & -10 & -8 \\ 3 & -2 & -2 & -6 & -7 \end{bmatrix} \\
\xrightarrow[\text{column 3 +column 1}]{\text{replace column 3 by}}
&\begin{bmatrix} 1 & 0 & 0 & -3 & -2 \\ 4 & -2 & 0 & -8 & -8 \\ 4 & -1 & 0 & -10 & -8 \\ 3 & -2 & 1 & -6 & -7 \end{bmatrix} \\
\xrightarrow[\text{column 4 +3×column 1}]{\text{replace column 4 by}}
&\begin{bmatrix} 1 & 0 & 0 & 0 & -2 \\ 4 & -2 & 0 & 4 & -8 \\ 4 & -1 & 0 & 2 & -8 \\ 3 & -2 & 1 & 3 & -7 \end{bmatrix} \\
\xrightarrow[\text{column 5 +2×column 1}]{\text{replace column 5 by}}
&\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 4 & -2 & 0 & 4 & 0 \\ 4 & -1 & 0 & 2 & 0 \\ 3 & -2 & 1 & 3 & -1 \end{bmatrix} \\
\xrightarrow[\text{column 4 +2×column 2}]{\text{replace column 4 by}}
&\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 4 & -2 & 0 & 0 & 0 \\ 4 & -1 & 0 & 0 & 0 \\ 3 & -2 & 1 & -1 & -1 \end{bmatrix} \\
\xrightarrow[\text{column 4 +column 3}]{\text{replace column 4 by}}
&\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 4 & -2 & 0 & 0 & 0 \\ 4 & -1 & 0 & 0 & 0 \\ 3 & -2 & 1 & 0 & -1 \end{bmatrix} \\
\xrightarrow[\text{column 5 +column 3}]{\text{replace column 5 by}}
&\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 4 & -2 & 0 & 0 & 0 \\ 4 & -1 & 0 & 0 & 0 \\ 3 & -2 & 1 & 0 & 0 \end{bmatrix}.
\end{aligned}
$$

Thus we find a basis

$$\left\{\begin{bmatrix}1\\4\\4\\3\end{bmatrix},\begin{bmatrix}0\\-2\\-1\\-2\end{bmatrix},\begin{bmatrix}0\\0\\0\\1\end{bmatrix}\right\}.$$

□

Remark 3.6.11. In general, the idea is to create stairs. Every time we do column reductions, we check if a_{ii} is nonzero. If it is zero, swap columns to make it nonzero. However, if this still cannot be realized by swapping columns, go to the next element $a_{i+1\,j}$.

Let's practice more.

Example 3.6.12. Find a basis of $\mathrm{im}(A)$ for the following matrices A.

(a) $A=\begin{bmatrix}-1&-2&3&0\\-4&-6&8&-2\\-4&-7&5&-6\\-3&-4&6&-1\end{bmatrix}$;

(b) $A=\begin{bmatrix}1&2&-1&-3&-2\\4&8&-4&-12&-7\\1&2&-1&-3&-2\\4&8&-4&-12&-8\\3&5&-2&-8&-7\end{bmatrix}$.

Solution. (a) A is transformed to

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

by column operations. The basis is

$$\left\{\begin{bmatrix}-1\\-4\\-4\\-3\end{bmatrix},\begin{bmatrix}0\\2\\1\\2\end{bmatrix},\begin{bmatrix}0\\0\\-5\\1\end{bmatrix}\right\}.$$

(b) A is transformed to

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

The basis is

$$\left\{ \begin{bmatrix} 1 \\ 4 \\ 1 \\ 4 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -1 \end{bmatrix} \right\}.$$

□

Definition 3.6.13. We call the dimension of the image space $\mathrm{im}(A)$ the rank of matrix A. The notation is $\mathrm{rk}(A)$.

Then Proposition 3.6.9 has a corollary:

Corollary 3.6.14. All column operations do not change $\mathrm{rk}(A)$.

Homework 3.6.15. Find a basis of $\mathrm{im}(A)$ and point out $\mathrm{rk}(A)$ for the following two matrices A.

(a) $A = \begin{bmatrix} -1 & -2 & 1 & 3 \\ -4 & -8 & 5 & 8 \\ -4 & -7 & 4 & 10 \\ -3 & -4 & 2 & 6 \end{bmatrix}$;

(b) $A = \begin{bmatrix} -1 & 2 & 3 & 3 & 2 \\ -4 & 8 & 11 & 12 & 7 \\ -1 & 2 & 3 & 3 & 2 \\ -4 & 8 & 12 & 12 & 8 \\ -3 & 5 & 9 & 8 & 7 \end{bmatrix}$.

Similar to Proposition 3.6.9, we can prove that row operations do not affect $\mathrm{rk}(A)$.

Proposition 3.6.16. Row operations do not change the rank of a matrix.

Proof. Let A be a matrix and E be a row operation matrix. Recall that row operations transform A to EA. The following maps

$$\begin{aligned} F : \operatorname{im}(A) &\longrightarrow \operatorname{im}(EA) \\ A\mathbf{v} &\longmapsto EA\mathbf{v} \end{aligned}$$

and

$$\begin{aligned} G : \operatorname{im}(EA) &\longrightarrow \operatorname{im}(A) \\ EA\mathbf{v} &\longmapsto E^{-1}(EA\mathbf{v}) = A\mathbf{v} \end{aligned}$$

satisfy

$$F \circ G = \text{identity map}$$

and

$$G \circ F = \text{identity map}.$$

Thus F (and also G) is a one-to-one correspondence between the two image spaces. Therefore, $\dim \operatorname{im}(A) = \dim \operatorname{im}(EA)$. □

Remark 3.6.17. Column operations do not change $\operatorname{im}(A)$. However, although row operations do not change $\operatorname{rk}(A)$, they sometimes change $\operatorname{im}(A)$. This is why when we try to find a basis of $\operatorname{im}(A)$, we do not apply row operations.

Example 3.6.18. Let $A = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Think about the following matrix action

$$\begin{aligned} f : \mathbb{R} &\to \mathbb{R}^2 \\ t &\mapsto At. \end{aligned}$$

This is equivalent to transform the real field to the line $y = x$ in Oxy plane.

However, if we do row reduction, we get another matrix $B = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. This matrix B maps the real field to the x-axis in Oxy plane. So, the image space is changed when we do row operations.

However, for matrix A, we do have a space that remains unchanged under row operations. It is called the kernel of A.

Definition 3.6.19. The kernel of an $m \times n$ matrix A is the space

$$\{\mathbf{v} \in \mathbb{R}^n : A\mathbf{v} = \mathbf{0}\}.$$

The notation is $\ker(A)$.

From the definition, we know that $\ker(A)$ is a linear space and finding $\ker(A)$ is the same as solving homogeneous equations.

Proposition 3.6.20. Row operations do not change $\ker(A)$.

Proof. Row operations are equivalent to transform A to EA. For any $\mathbf{v} \in \ker(A)$, we have

$$(EA)\mathbf{v} = E(A\mathbf{v}) = E \cdot \mathbf{0} = \mathbf{0},$$

so $\ker(A) \subseteq \ker(EA)$. Conversely, for any $\mathbf{w} \in \ker(EA)$, we have

$$A\mathbf{w} = E^{-1}(EA)\mathbf{w} = E^{-1} \cdot \mathbf{0} = \mathbf{0},$$

so $\ker(EA) \subseteq \ker(A)$. $\square$

Now, suppose that when we solve a homogeneous system

$$A \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

after row operations and column switching, we get an equation

$$\begin{bmatrix} 1 & 0 & 0 & 3 & 4 & -1 \\ 0 & 1 & 0 & 4 & 2 & -7 \\ 0 & 0 & 1 & -1 & -3 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_4 \\ x_3 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \tag{3.6.1}$$

Then we see x_3, x_5, x_6 are free. We immediately write down the solution

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} -3x_3 - 4x_5 + x_6 \\ -4x_3 - 2x_5 + 7x_6 \\ x_3 \\ x_3 + 3x_5 - 3x_6 \\ x_5 \\ x_6 \end{bmatrix} = x_3 \begin{bmatrix} -3 \\ -4 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -4 \\ -2 \\ 0 \\ 3 \\ 1 \\ 0 \end{bmatrix} + x_6 \begin{bmatrix} 1 \\ 7 \\ 0 \\ -3 \\ 0 \\ 1 \end{bmatrix}.$$

(Why do I go back to the original order of x_3 and x_4 here?) This shows that $\ker(A)$ has a basis $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, where

$$\mathbf{v}_1 = \begin{bmatrix} -3 \\ -4 \\ 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -4 \\ -2 \\ 0 \\ 3 \\ 1 \\ 0 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 1 \\ 7 \\ 0 \\ -3 \\ 0 \\ 1 \end{bmatrix}.$$

Another observation is that, since row operations do not change the dimension of $\operatorname{im}(A)$, we immediately see that the number of diagonal 1's in the matrix in (3.6.1) is equal to the $\dim \operatorname{im}(A)$.

Corollary 3.6.21. After using up row operations and column switching on matrix A to solve the homogeneous system, the number of diagonal 1's is equal to $\dim(\operatorname{im}(A))$.

Also, from our previous experience in solving equations, we deduce that

$$\begin{aligned} &\text{the number of columns of } A - \text{the number of diagonal 1's after row operations and column switching} \\ &= \text{the number of variables} - \text{the number of non-free variables} \\ &= \text{the number of free variables} \\ &= \dim(\ker(A)). \end{aligned}$$

Theorem 3.6.22. Let A be an $m \times n$ matrix, then we have

$$\dim(\ker(A)) + \dim(\operatorname{im}(A)) = \dim(\ker(A)) + \dim(\operatorname{rk}(A)) = n.$$

Corollary 3.6.23. If $n \times n$ matrix A has an inverse A^{-1}, then $\ker(A) = \{\mathbf{0}\}$ and $\operatorname{rk}(A) = n$. In other words, its columns form an independent set.

Proof. Suppose $A\mathbf{v} = \mathbf{0}$. Then $\mathbf{v} = A^{-1}\mathbf{0} = \mathbf{0}$. Therefore, $\operatorname{rk}(A) = n - \dim(\ker(A)) = n$. □

Example 3.6.24. For the following matrices

$$A = \begin{bmatrix} 4 & -6 & 6 \\ -8 & 11 & -12 \end{bmatrix}, B = \begin{bmatrix} 1 & 3 & 3 \\ 4 & 12 & 12 \\ 1 & 3 & 6 \end{bmatrix}.$$

(a) Find bases of $\ker(A)$ and $\ker(B)$;

(b) Find $\mathrm{rk}(A)$ and $\mathrm{rk}(B)$.

Solution. (a) Solve equation system

$$\begin{bmatrix} 4 & -6 & 6 \\ -8 & 11 & -12 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

we get

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} -3/2 \\ 0 \\ 1 \end{bmatrix}$$

where x_3 is free. Thus we find a basis of $\ker(A)$ consisting of only one vector $\begin{bmatrix} -3/2 \\ 0 \\ 1 \end{bmatrix}$.

Similarly, if we solve equation system

$$\begin{bmatrix} 1 & 3 & 3 \\ 4 & 12 & 12 \\ 1 & 3 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

we get

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix}.$$

Thus $\ker(B)$ has a basis consisting of $\begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix}$.

(b) $\mathrm{rk}(A) = 3 - 1 = 2$ and $\mathrm{rk}(B) = 3 - 1 = 2$.

□

Homework 3.6.25. (This problem is adapted from [3]: Exercise 9 on page 160.) For the following matrices

$$A = \begin{bmatrix} 1 & -3 & 2 & -4 \\ -3 & 9 & -1 & 5 \\ 2 & -6 & 4 & -3 \\ -4 & 12 & 2 & 7 \end{bmatrix}, B = \begin{bmatrix} 1 & 3 & 3 \\ 4 & 6 & 7 \\ 1 & 2 & 3 \end{bmatrix}.$$

(a) Find bases of $\ker(A)$ and $\ker(B)$;

(b) Find $\mathrm{rk}(A)$ and $\mathrm{rk}(B)$.

Chapter 4

Determinants

4.1 Expansion Along Rows

Why do we always have an inverse for the change of basis problem? This will be answered by determinants.

First of all, let's recall the definition of cross product. Let (x_1, x_2, x_3) and (y_1, y_2, y_3) be two vectors in $Oxyz$ space. Then we see

$$(x_1, x_2, x_3) \times (y_1, y_2, y_3) = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix} = \mathbf{i}(x_2y_3 - x_3y_2) - \mathbf{j}(x_1y_3 - x_3y_1) + \mathbf{k}(x_1y_2 - x_2y_1).$$

For $a, b, c, d \in \mathbb{R}$, let

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc.$$

Some of you may know this is the determinant of a 2×2 matrix. Then we get

$$(x_1, x_2, x_3) \times (y_1, y_2, y_3) = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix} = \mathbf{i} \cdot \det \begin{bmatrix} x_2 & x_3 \\ y_2 & y_3 \end{bmatrix} - \mathbf{j} \cdot \det \begin{bmatrix} x_1 & x_3 \\ y_1 & y_3 \end{bmatrix} + \mathbf{k} \cdot \det \begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \end{bmatrix}.$$

Now recall the definition of triple product. Let (z_1, z_2, z_3) be another vector in $Oxyz$ space. Then we have

$$(z_1, z_2, z_3) \cdot ((x_1, x_2, x_3) \times (y_1, y_2, y_3)) = z_1 \cdot \det \begin{bmatrix} x_2 & x_3 \\ y_2 & y_3 \end{bmatrix} - z_2 \cdot \det \begin{bmatrix} x_1 & x_3 \\ y_1 & y_3 \end{bmatrix} + z_3 \cdot \det \begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \end{bmatrix}.$$

Therefore, for any 3×3 matrix

$$\begin{bmatrix} z_1 & z_2 & z_3 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{bmatrix},$$

we let

$$\det \begin{bmatrix} z_1 & z_2 & z_3 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{bmatrix} = z_1 \cdot \det \begin{bmatrix} x_2 & x_3 \\ y_2 & y_3 \end{bmatrix} - z_2 \cdot \det \begin{bmatrix} x_1 & x_3 \\ y_1 & y_3 \end{bmatrix} + z_3 \cdot \det \begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \end{bmatrix}.$$

Generalizing this pattern, we get the definition of determinant.

Definition 4.1.1. (**Determinant**) For an $n \times n$ matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{bmatrix},$$

its determinant $\det(A)$ also equals

$$\sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det(A_{1j}),$$

where A_{1j} is an $(n-1) \times (n-1)$ matrix given by removing the 1st row and the j-th column of A (The notations in this definition are the same as [3]: page 167.).

Example 4.1.2. For the following matrices, find their determinants:

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

The most important property of determinant is its product rule.

Theorem 4.1.3. For any two $n \times n$ matrices A and B, we have

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

This proposition is hard to prove, so we separate it into several parts.

Lemma 4.1.4. Row operations have the following influences on the original $\det(A)$.

(a) After multiplying any row by a scalar r, the determinant becomes $r\det(A)$.

(b) After swapping any two rows, the determinant becomes $-\det(A)$.

(c) After replacing row i by "row $i + r\times$row j", the determinant remains unchanged.

Proof. Use mathematical induction method. We omit details here and leave them to readers. □

Example 4.1.5. We check Lemma 4.1.4 in this example. Let $A = \begin{bmatrix} 1 & 2 & 0 \\ 4 & 1 & 1 \\ 2 & 2 & 1 \end{bmatrix}$. We already know that $\det(A) = -5$.

(a) Multiply the second row of A by 3 and calculate the new determinant. Compare it with the previous determinant.

(b) Swap the first and the third rows of a and calculate the new determinant. Compare it with the previous determinant.

(c) Replace row 1 of A by "row $1 + 2\times$row 3" and calculate the new determinant. Compare it with the previous determinant.

Lemma 4.1.4 (b) tells us that we can expand determinants along any row.

Corollary 4.1.6. Given matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{k1} & a_{k2} & a_{k3} & \cdots & a_{kn} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{bmatrix},$$

its determinant $\det(A)$ also equals

$$\sum_{j=1}^{n} (-1)^{k+j} a_{kj} \det(A_{kj}),$$

where A_{kj} is an $(n-1)\times(n-1)$ matrix given by removing the k-th row and the j-th column of A.

Example 4.1.7. For the following matrices, find their determinants:

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

It is very straightforward to find the determinants of elementary row operation matrices.

Example 4.1.8. Let $E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $E_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$, $E_3 = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ and $A = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 2 & 2 & 1 \end{bmatrix}$. We already know that $\det(A) = -5$.

(a) Find determinants $\det(E_1)$, $\det(E_2)$, $\det(E_3)$.

(b) Without calculation, find $\det(E_1A)$, $\det(E_2A)$, $\det(E_3A)$ from Example 4.1.5.

(c) Find $\det(E_1)\det(A)$, $\det(E_2)\det(A)$, $\det(E_3)\det(A)$. What can you see?

Considering Lemma 4.1.4 together with Example 4.1.8, we get the following corollary which is very close to Proposition 4.1.3.

Corollary 4.1.9. Let E be an $n \times n$ matrix representing a composition of row operations. Then for any $n \times n$ matrix A,

$$\det(EA) = \det(E)\det(A).$$

Recall that an invertible matrix A is exactly a matrix representing a composition of row operations. We immediately get the following.

Corollary 4.1.10. Let A and B be two $n \times n$ matrices. If A is invertible, then

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

In particular,

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

However, what if A does not have an inverse?

Example 4.1.11. For matrix $A = \begin{bmatrix} 1 & -1 & 2 \\ 2 & -2 & 4 \\ 0 & 2 & 0 \end{bmatrix}$, can you use row operations to find an inverse?

Definition 4.1.12. We call $1 \times n$ matrices like

$$\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$$

an n-dimensional row vector. Here $x_1, \cdots, x_n$ are scalars.

It is straightforward to see that a matrix is invertible if and only if we do not see a zero row vector after row operations. If we get a zero row, it means some row is a linear combination of other rows. Then we get the following criterion.

Proposition 4.1.13. The criterion of the existence of inverse matrices:

(a) "A is not invertible" $\Longleftrightarrow$ "The rows of A form a linearly dependent set".

(b) "A is invertible" $\Longleftrightarrow$ "The rows of A form a linearly independent set".

Remark 4.1.14. Denote the space of all $1 \times n$ real row vectors by $\mathbb{R}_n$. Then, if $n \times n$ matrix A is invertible, its rows form a basis of $\mathbb{R}_n$.

Considering the effect of row operations on determinants, we get

Proposition 4.1.15. Another criterion for the existence of inverse matrices:

(a) "A is not invertible" $\Longleftrightarrow$ "$\det(A) = 0$".

(b) "A is invertible" $\Longleftrightarrow$ "$\det(A) \neq 0$".

Proof. Suppose after row operations, we get a matrix EA which has a zero row. Expand $\det(EA)$ across that zero row, we see $\det(A) = 0/\det(E) = 0$.

Conversely, if $\det(A) = 0$ but A^{-1} exists, then $1 = \det(I) = \det(AA^{-1}) = \det(A)\det(A^{-1}) = 0$, inducing a contradiction. □

If A is not invertible, we see that rows of A form a dependent set. Then the rows of AB also form a dependent set (Why?). This means AB is also not invertible.

Proposition 4.1.16. If A is not invertible, then AB is not invertible either.

Therefore, when A is not invertible, we see that

$$0 = 0 \cdot \det(B) = \det(A)\det(B) = \det(AB) = 0.$$

Thus we finish the proof of Theorem 4.1.3.

Remark 4.1.17. In fact, AB is invertible if and only if both A and B are invertible.

Homework 4.1.18. Let A, B and C be $n \times n$ matrices. True or false and briefly write down your reasons:

(a) If $\det(A) = 0$, then $\det(AB) = 0$.

(b) If $\det(AB) = 0$, then $\det(A)$ must be 0 as well.

(c) If $\det(ABC) = 0$, then at least one of $\det(A)$, $\det(B)$ and $\det(C)$ must be 0.

(d) If $\det(ABC) \neq 0$, then none of $\det(A)$, $\det(B)$ and $\det(C)$ is 0.

4.2 Expansion Along Columns

Another way to define determinant is the expansion along the 1st column.

Definition 4.2.1. (**Determinant by Column Expansion**) For an $n \times n$ matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{bmatrix},$$

its determinant by column expansion $\Delta(A)$ equals

$$\sum_{i=1}^{n} (-1)^{i+1} a_{i1} \Delta(A_{i1}),$$

where A_{i1} is an $(n-1) \times (n-1)$ matrix given by removing the 1st column and the i-th row of A.

Example 4.2.2. For the following matrices

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

(a) Find their determinants by expanding along the 1st column.

(b) Find their determinants by expanding along rows.

(c) Compare the above results.

It is reasonable to guess that $\det(A) = \Delta(A)$. One way to verify this requires some knowledge on signs of permutations (see [1]: Chapter 1). Another way is given as follows.

Imitate all the things we did for row expansion, we get the following conclusions for column expansion.

Lemma 4.2.3. Column operations have the following influences on the original $\Delta(A)$.

(a) After multiplying any column by a scalar r, the original $\Delta(A)$ becomes $r\Delta(A)$.

(b) After swapping any two columns, the original $\Delta(A)$ becomes $-\Delta(A)$.

(c) After replacing column i by "column i + $r\times$column j", the original $\Delta(A)$ remains unchanged.

Lemma 4.2.3 (b) guarantees expansion along any column.

Corollary 4.2.4. Given matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2k} & \cdots & a_{2n} \\ a_{31} & a_{32} & \cdots & a_{3k} & \cdots & a_{3n} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nk} & \cdots & a_{nn} \end{bmatrix},$$

its determinant by column expansion $\Delta(A)$ equals

$$\sum_{i=1}^{n}(-1)^{i+k}a_{ik}\Delta(A_{ik}),$$

where A_{ik} is an $(n-1)\times(n-1)$ matrix given by removing the k-th column and the i-th row of A.

Considering the effects of column operations in Lemma 4.2.3, we get the following product rule.

Corollary 4.2.5. Let E be an $n\times n$ matrix representing a composition of column operations. Then for any $n\times n$ matrix A,

$$\Delta(AE) = \Delta(A)\Delta(E).$$

If we can use row operations to find an inverse for A, then we have $EA = I$. Recall that this can always be written into $A = E^{-1}$ because row operations are always reversible, and therefore $AE = I$. In other words, column operations can also give us an inverse. If $AE = I$, then Corollary 4.2.5 tells us

$$\Delta(A) = \frac{1}{\Delta(E)}.$$

Proposition 4.2.6. If A is invertible, then $\det(A) = \Delta(A)$.

Proof. As we know, when $EA = AE = I$, we have

$$\det(A) = \frac{1}{\det(E)} \text{ and } \Delta(A) = \frac{1}{\Delta(E)}.$$

In addition, for matrices E representing row (or column) operations, we have

$$\det(E) = \Delta(E).$$

The verification is given in Homework 4.2.7. □

Homework 4.2.7. Let $E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $E_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$, $E_3 = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Check that

$$\Delta(E_i) = \det(E_i),$$

where $i = 1, 2, 3$.

However, what if A is not invertible? For this A, if we apply column operations to partitioned matrix

$$\begin{bmatrix} A \\ I \end{bmatrix},$$

we will get a zero column. Therefore, $\Delta(A) = 0$. Immediately, $\det(A) = 0 = \Delta(A)$.

Theorem 4.2.8. The two definitions of determinant are exactly the same. Therefore, for any $n \times n$ matrices,

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

Remark 4.2.9. From now on, we unify the two notations and only use notation "det".

It is straightforward to generalize results verified under row expansion to column expansion:

Proposition 4.2.10. Let A be an $n \times n$ matrix. Here is the criterion of the existence of A^{-1}:

(a) "$\det(A) = 0$" $\Longleftrightarrow$ "A is not invertible" $\Longleftrightarrow$ "The columns of A form a linearly dependent set" $\Longleftrightarrow$ "The rows of A form a linearly dependent set" $\Longleftrightarrow$ "$\mathrm{rk}(A) < n$".

(b) "$\det(A) \neq 0$" $\Longleftrightarrow$ "A is invertible" $\Longleftrightarrow$ "The columns of A form a linearly independent set" $\Longleftrightarrow$ "The rows of A form a linearly independent set" $\Longleftrightarrow$ "$\mathrm{rk}(A) = n$".

Homework 4.2.11. Let $A = \begin{bmatrix} 1 & 3 & 4 \\ 2 & 6 & 2 \\ 3 & 10 & 14 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 3 & 1 \\ 2 & 6 & 2 \\ 3 & 4 & -2 \end{bmatrix}$.

(a) Find bases of $\operatorname{im}(A)$ and $\operatorname{im}(B)$ respectively.

(b) Find $\operatorname{rk}(A)$ and $\operatorname{rk}(B)$.

(c) Suppose C is the one among A and B with smaller rank. Find $\det(C^{1000})$.

4.3 A Nonrigorous Approach

This section provides instructors with another approach to present determinants and related important results.

Why do we always have an inverse for the change of basis problem? This will be answered by determinants.

First of all, let's recall the definition of cross product. Let (x_1, x_2, x_3) and (y_1, y_2, y_3) be two vectors in $Oxyz$ space. Then we see

$$(x_1, x_2, x_3) \times (y_1, y_2, y_3) = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix} = \mathbf{i}(x_2y_3 - x_3y_2) - \mathbf{j}(x_1y_3 - x_3y_1) + \mathbf{k}(x_1y_2 - x_2y_1).$$

For $a, b, c, d \in \mathbb{R}$, let

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc.$$

Some of you may know this is the determinant of a 2×2 matrix. Then we get

$$(x_1, x_2, x_3) \times (y_1, y_2, y_3) = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix} = \mathbf{i} \cdot \det \begin{bmatrix} x_2 & x_3 \\ y_2 & y_3 \end{bmatrix} - \mathbf{j} \cdot \det \begin{bmatrix} x_1 & x_3 \\ y_1 & y_3 \end{bmatrix} + \mathbf{k} \cdot \det \begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \end{bmatrix}.$$

Now recall the definition of triple product. Let (z_1, z_2, z_3) be another vector in $Oxyz$ space. Then we have

$$(z_1, z_2, z_3) \cdot ((x_1, x_2, x_3) \times (y_1, y_2, y_3)) = z_1 \cdot \det \begin{bmatrix} x_2 & x_3 \\ y_2 & y_3 \end{bmatrix} - z_2 \cdot \det \begin{bmatrix} x_1 & x_3 \\ y_1 & y_3 \end{bmatrix} + z_3 \cdot \det \begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \end{bmatrix}.$$

Therefore, for any 3×3 matrix

$$\begin{bmatrix} z_1 & z_2 & z_3 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{bmatrix},$$

we let

$$\det \begin{bmatrix} z_1 & z_2 & z_3 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{bmatrix} = z_1 \cdot \det \begin{bmatrix} x_2 & x_3 \\ y_2 & y_3 \end{bmatrix} - z_2 \cdot \det \begin{bmatrix} x_1 & x_3 \\ y_1 & y_3 \end{bmatrix} + z_3 \cdot \det \begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \end{bmatrix}.$$

Generalizing this pattern, we get the definition of determinant.

Definition 4.3.1. (**Determinant**) For an $n \times n$ matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{bmatrix},$$

its determinant $\det(A)$ equals

$$\sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det(A_{1j}),$$

where A_{1j} is an $(n-1) \times (n-1)$ matrix given by removing the 1st row and the j-th column of A (The notations in this definition are the same as [3]: page 167.).

We acknowledge the following property of expansion without proving it.

Theorem 4.3.2. We can find determinants by expansion along any row or even any column. To be more precise, let A_{ij} be the matrix given by removing the i-th row and the j-th column. Then we can choose the k-th row and find

$$\det(A) = \sum_{j=1}^{n} (-1)^{k+j} a_{kj} \det(A_{kj}).$$

Or, we can choose the k-th column and find

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+k} a_{ik} \det(A_{ik}).$$

Example 4.3.3. For the following matrices

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, A = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 2 & 2 & 1 \end{bmatrix}, B = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 4 & 1 \\ 0 & 1 & 0 \end{bmatrix}, C = \begin{bmatrix} 1 & -1 & 2 & 0 \\ 0 & 2 & 0 & 0 \\ 1 & 4 & 1 & 1 \\ 1 & 2 & 2 & 1 \end{bmatrix},$$

use any way you like to figure out their determinants.

Solution. (1) $\det(I) = 1$.

(2) Expand $\det(A)$ along the second row and get $\det(A) = 2$.

(3) Expand $\det(B)$ along the third row and get $\det(B) = 2$.

(4) Expand $\det(C)$ along the second row and get $\det(C) = -2$.

□

We also acknowledge the following product rule without proving it.

Theorem 4.3.4. For any two $n \times n$ matrices A and B, we have

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

Theorem 4.3.2 and Theorem 4.3.4 can also be proved by expanding everything (you can imagine the workload) without using elementary row or column operation matrices (the trick in the previous two sections to avoid a lot of calculations). This is why I am confident to put Theorem 4.3.2 and Theorem 4.3.4 here without any proof.

Theorem 4.3.4 tells us how $\det(A)$ changes when we do row or column operations.

Example 4.3.5. Let $E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $E_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$, $E_3 = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ and $A = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 2 & 2 & 1 \end{bmatrix}$.

(a) Check that $\det(A) = 2$.

(b) What are the meanings of $E_1A, E_2A, E_3A, AE_1, AE_2, AE_3$ respectively? Think about row and column operations.

(c) Find determinants $\det(E_1), \det(E_2), \det(E_3)$.

(d) Find $\det(E_1A), \det(E_2A), \det(E_3A), \det(AE_1), \det(AE_2), \det(AE_3)$ using Theorem 4.3.4.

(e) Compare $\det(E_1A), \det(E_2A), \det(E_3A), \det(AE_1), \det(AE_2), \det(AE_3)$ with $\det(A)$.

Solution. (a) It has been checked.

(b) E_1 on the left means multiplying row 2 by 3.

E_2 on the left means swapping row 1 and row 3.

E_3 on the left means replacing row 1 by 2×row 3 + row 1.

E_1 on the right means multiplying column 2 by 3.

E_2 on the right means swapping column 1 and column 3.

E_3 on the right means replacing column 3 by 2×column 1 + column 3.

(c) $\det(E_1) = 3, \det(E_2) = -1, \det(E_3) = 1$.

(d) $\det(E_1A) = 6$, $\det(E_2A) = -2$, $\det(E_3A) = 2$, $\det(AE_1) = 6$, $\det(AE_2) = -2$, $\det(AE_3) = 2$.

(e) Row/column scalar multiplication puts a scalar on $\det(A)$. Row/column switching puts a minus sign on $\det(A)$. Row/column reduction does not change $\det(A)$.

□

With the same idea as that in Example 4.3.5, we can use Theorem 4.3.4 to prove the following proposition.

Proposition 4.3.6. Row operations have the following influences on the original $\det(A)$:

(a) After multiplying any row by a scalar r, the determinant becomes $r\det(A)$.

(b) After swapping any two rows, the determinant becomes $-\det(A)$.

(c) After replacing row i by "row i + $r\times$row j", the determinant remains unchanged.

Similarly, column operations have the following influences on the original $\det(A)$:

(a) After multiplying any column by a scalar r, the original $\det(A)$ becomes $r\det(A)$.

(b) After swapping any two columns, the original $\det(A)$ becomes $-\det(A)$.

(c) After replacing column i by "column i + $r\times$column j", the original $\det(A)$ remains unchanged.

Proposition 4.3.6 are sometimes used to simplify the calculation of determinants. However, these tricks are not so useful today since we already have computers.

Example 4.3.7. Let $A = \begin{bmatrix} 1 & -1 & 2 & 2 \\ 2 & 1 & 3 & 3 \\ -1 & 4 & -3 & -3 \\ 1 & 2 & 2 & 1 \end{bmatrix}$. Find $\det(A)$. (Hint: $(-3)\times$row 1+ row 2 = row 3.)

Solution. Replacing row 2 by row 2 + $(-3)\times$row 1, we get

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

Replacing row 3 by row 3 − row 2, we get

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

These row reductions do not change $\det(A)$, so now expand the determinant along the 3rd row, we get $\det(A) = 0$. □

Theorem 4.3.4 can also give us the relations between linear independence (or dependence), the existence of inverse matrices and determinants. First, let's look at an example.

Example 4.3.8. For matrix $A = \begin{bmatrix} 1 & -1 & 2 \\ 2 & -2 & 4 \\ 0 & 2 & 0 \end{bmatrix}$, you can never find an inverse using row operations because you will find a row with all entries equal to zero.

Definition 4.3.9. We call $1 \times n$ matrices like

$$\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$$

an n-dimensional row vector. Here $x_1, \cdots, x_n$ are scalars. The space of all n-dimensional row vectors is denoted by $\mathbb{R}_n$.

If a matrix A is not invertible, then we can always find a zero row after some row operations. If the row operation is represented by a matrix E, by Theorem 4.3.2, we get that

$$\det(EA) = 0.$$

Then by Theorem 4.3.4, we see that either $\det(E)$ or $\det(A)$ is 0. Which one will happen? Think about Example 4.3.5.

Proposition 4.3.10. Let A be an $n \times n$ matrix. Here is the criterion of the existence of A^{-1}:

(a) "$\det(A) = 0$" $\Longleftrightarrow$ "A is not invertible" $\Longleftrightarrow$ "The columns of A form a linearly dependent set" $\Longleftrightarrow$ "The rows of A form a linearly dependent set" $\Longleftrightarrow$ "$\mathrm{rk}(A) < n$".

(b) "$\det(A) \neq 0$" $\Longleftrightarrow$ "A is invertible" $\Longleftrightarrow$ "The columns of A form a linearly independent set" $\Longleftrightarrow$ "The rows of A form a linearly independent set" $\Longleftrightarrow$ "$\mathrm{rk}(A) = n$".

Homework 4.3.11. Let A, B and C be $n \times n$ matrices. True or false and briefly write down your reasons:

(a) If $\det(A) = 0$, then $\det(AB) = 0$.

(b) If $\det(AB) = 0$, then $\det(A)$ must be 0 as well.

(c) If $\det(ABC) = 0$, then at least one of $\det(A)$, $\det(B)$ and $\det(C)$ must be 0.

(d) If $\det(ABC) \neq 0$, then none of $\det(A)$, $\det(B)$ and $\det(C)$ is 0.

Homework 4.3.12. Let $A = \begin{bmatrix} 1 & 3 & 4 \\ 2 & 6 & 2 \\ 3 & 10 & 14 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 3 & 1 \\ 2 & 6 & 2 \\ 3 & 4 & -2 \end{bmatrix}$.

(a) Find bases of $\operatorname{im}(A)$ and $\operatorname{im}(B)$ respectively.

(b) Find $\operatorname{rk}(A)$ and $\operatorname{rk}(B)$.

(c) Find $\det(B^{1000})$.

Chapter 5

Eigenspaces

5.1 Eigenvalues and Eigenvectors

For a linear map $f : V \to V$, we hope to find a basis $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ of V such that the corresponding matrix is diagonal.

Now, suppose we already have a basis $\{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_n\}$ of V, and relative to this basis, the matrix corresponding to f is A. In other words,

$$\begin{bmatrix} f(\mathbf{w}_1) & f(\mathbf{w}_2) & \cdots & f(\mathbf{w}_n) \end{bmatrix} = \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_n \end{bmatrix} A.$$

Let P be the change of basis matrix P satisfying

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

Since f is a linear map, we get that

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

Combining these three relations, we get

$$\begin{bmatrix} f(\mathbf{v}_1) & f(\mathbf{v}_2) & \cdots & f(\mathbf{v}_n) \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix} P^{-1}AP.$$

Thus the problem becomes how to find an appropriate matrix P to make $P^{-1}AP$ a diagonal matrix.

Suppose that we can finally get a diagonal matrix like

$$\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}.$$

In other words,

$$P^{-1}AP = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}.$$

Let $\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_n$ be columns of P, we get

$$A \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}.$$

This is the same as

$$\begin{aligned} A\mathbf{b}_1 &= \lambda_1 \mathbf{b}_1, \\ A\mathbf{b}_2 &= \lambda_2 \mathbf{b}_2, \\ &\vdots \\ A\mathbf{b}_n &= \lambda_2 \mathbf{b}_n. \end{aligned}$$

Take $A\mathbf{b}_1 = \lambda_1 \mathbf{b}_1$ as an example, this equation is the same as

$$(A - \lambda_1 I)\mathbf{b}_1 = \mathbf{0}.$$

If matrix $A - \lambda_1 I$ is invertible, then $\mathbf{b}_1$ must only be $\mathbf{0}$. However, it is impossible for an invertible P to have a zero column. Otherwise, its determinant is 0.

Therefore, $A - \lambda_1 I$ is not invertible and thus we have

$$\det(A - \lambda_1 I) = 0.$$

Now we are able to introduce the definition of eigenvalues and eigenvectors.

Definition 5.1.1. Any number λ that satisfies equation

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

is an eigenvalue of A.

Definition 5.1.2. Suppose λ is already an eigenvalue, then any column vector $\mathbf{v}$ satisfying

$$(A - \lambda I)\mathbf{v} = \mathbf{0}$$

is called an eigenvector of A corresponding to eigenvalue λ. We call space

$$E_\lambda := \{\mathbf{v} : (A - \lambda I)\mathbf{v} = \mathbf{0}\}$$

the eigenspace of A corresponding to eigenvalue λ.

As you can see, finding eigenvalues is no more than solving a higher-order equations in one variable, while finding eigenvectors is no more than solving linear equations.

Example 5.1.3. For each of the following matrices, find all its eigenvalues and eigenvectors. Then point out the dimensions of eigenspaces.

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

Solution. (1) Let $\det(I - \lambda I) = 0$, we get that $\lambda = 1$. Solve equation

$$(I - 1 \cdot I)\mathbf{x} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

we get that $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$. Thus all eigenvectors with respect to $\lambda = 1$ are like

$$x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

where x_1, x_2, x_3 are free. Also, $\dim E_1 = 3$.

(2) Let $\det(A-\lambda I)=0$, we get that $\lambda=-1,1,2$. Solve equation

$$(A+I)\mathbf{x}=\begin{bmatrix}2&2&0\\0&1&1\\0&2&2\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix},$$

we get that $\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=x_3\begin{bmatrix}1\\-1\\1\end{bmatrix}$. Thus all eigenvectors with respect to $\lambda=-1$ are like

$$x_3\begin{bmatrix}1\\-1\\1\end{bmatrix}$$

where x_3 is free. Also, $\dim E_{-1}=1$.

Solve equation

$$(A-I)\mathbf{x}=\begin{bmatrix}0&2&0\\0&-1&1\\0&2&0\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix},$$

we get that $\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=x_1\begin{bmatrix}1\\0\\0\end{bmatrix}$. Thus all eigenvectors with respect to $\lambda=1$ are like

$$x_1\begin{bmatrix}1\\0\\0\end{bmatrix}$$

where x_1 is free. Also, $\dim E_1=1$.

Solve equation

$$(A-2I)\mathbf{x}=\begin{bmatrix}-1&2&0\\0&-2&1\\0&2&-1\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix},$$

we get that $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} 1 \\ 1/2 \\ 1 \end{bmatrix}$. Thus all eigenvectors with respect to $\lambda = 2$ are like

$$x_3 \begin{bmatrix} 1 \\ 1/2 \\ 1 \end{bmatrix}$$

where x_3 is free. Also, $\dim E_2 = 1$.

(3) Let $\det(B - \lambda I) = 0$, we get that $\lambda = 1, 2$. Solve equation

$$(B - I)\mathbf{x} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

we get that $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$. Thus all eigenvectors with respect to $\lambda = 1$ are like

$$x_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

where x_3 is free. Also, $\dim E_1 = 1$.

Solve equation

$$(B - 2I)\mathbf{x} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

we get that $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$. Thus all eigenvectors with respect to $\lambda = 2$ are like

$$x_3 \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

where x_2 and x_3 are free. Also, $\dim E_2 = 2$.

(4) Let $\det(C-\lambda I)=0$, we get that $\lambda=-2,2,3$. Solve equation

$$(C+2I)\mathbf{x}=\begin{bmatrix}4&0&0\\0&5&0\\0&0&0\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix},$$

we get that $\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=x_3\begin{bmatrix}0\\0\\1\end{bmatrix}$. Thus all eigenvectors with respect to $\lambda=-2$ are like

$$x_3\begin{bmatrix}0\\0\\1\end{bmatrix}$$

where x_3 is free. Also, $\dim E_{-2}=1$.

Solve equation

$$(C-2I)\mathbf{x}=\begin{bmatrix}0&0&0\\0&1&0\\0&0&-4\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix},$$

we get that $\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=x_1\begin{bmatrix}1\\0\\0\end{bmatrix}$. Thus all eigenvectors with respect to $\lambda=2$ are like

$$x_1\begin{bmatrix}1\\0\\0\end{bmatrix}$$

where x_1 is free. Also, $\dim E_2=1$.

Solve equation

$$(C-3I)\mathbf{x}=\begin{bmatrix}-1&0&0\\0&0&0\\0&0&-5\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix},$$

we get that $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$. Thus all eigenvectors with respect to $\lambda = 3$ are like

$$x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

where x_2 is free. Also, $\dim E_3 = 1$.

(5) Let $\det(D - \lambda I) = 0$, we get that $\lambda = 2, -4$. Solve equation

$$(D - 2I)\mathbf{x} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

we get that $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$. Thus all eigenvectors with respect to $\lambda = 2$ are like

$$x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

where x_1 is free. Also, $\dim E_2 = 1$.

Solve equation

$$(D + 4I)\mathbf{x} = \begin{bmatrix} 6 & 1 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

we get that $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$. Thus all eigenvectors with respect to $\lambda = -4$ are like

$$x_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

where x_3 is free. Also, $\dim E_{-4} = 1$. $\square$

Homework 5.1.4. For the following matrices, choose **four** of them to find all eigenvalues and all eigenvectors. Then point out the dimensions of eigenspaces.

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}, B = \begin{bmatrix} 3 & 2 & 0 \\ -3 & -2 & 1 \\ 2 & 2 & 1 \end{bmatrix}, C = \begin{bmatrix} \ln 3 & 0 & 0 \\ 0 & \sin(29^\circ) & 0 \\ 0 & 0 & \pi \end{bmatrix}, D = \begin{bmatrix} 2 & 2 & 8 \\ 0 & 2 & 4 \\ 0 & 0 & -4 \end{bmatrix}, E = \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}.$$

5.2 Diagonalization

Let A be an $n \times n$ matrix. If we have a matrix P satisfying that

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

then we let $\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_n$ be columns of P and get

$$A\begin{bmatrix}\mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n\end{bmatrix} = \begin{bmatrix}\mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n\end{bmatrix}\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{bmatrix}.$$

These column vectors are eigenvectors because

$$\begin{aligned} A\mathbf{b}_1 &= \lambda_1\mathbf{b}_1, \\ A\mathbf{b}_2 &= \lambda_2\mathbf{b}_2, \\ &\vdots \\ A\mathbf{b}_n &= \lambda_2\mathbf{b}_n. \end{aligned}$$

Since P is invertible, we must require that all these $\mathbf{b}_i$'s form an independent set. Conversely, if we can find eigenvectors that form an independent set, then we can let matrix P be the matrix consisting of these eigenvectors.

Theorem 5.2.1. Suppose $\lambda_1, \lambda_2, \cdots, \lambda_m$ be all eigenvalues of an $n \times n$ matrix A (Here we should remove duplicated eigenvalues, so $m \leq n$ and $\lambda_i \neq \lambda_j$ if $i \neq j$.). If we have

$$\dim E_{\lambda_1} + \dim E_{\lambda_2} + \cdots + \dim E_{\lambda_m} = n,$$

then we can find matrix P to make $P^{-1}AP$ diagonal.

To prove this theorem, we point out a lemma first.

Lemma 5.2.2. Let A be an $n \times n$ matrix and λ and μ be two eigenvalues of A. Then

$$(A - \lambda I)(A - \mu I) = (A - \mu I)(A - \lambda I).$$

Proof. Expand both sides by distributive law, then proof complete. □

Now we can prove Theorem 5.2.1.

Proof. (Proof of Theorem 5.2.1) Suppose that for each λ_i, we have a basis

$$\{\mathbf{b}_1^i, \cdots, \mathbf{b}_{k_i}^i\}$$

of eigenspace E_{λ_i}. Now, we prove that

$$\{\mathbf{b}_1^1, \cdots, \mathbf{b}_{k_1}^1, \mathbf{b}_1^2, \cdots, \mathbf{b}_{k_2}^2, \cdots, \mathbf{b}_1^m, \cdots, \mathbf{b}_{k_m}^m\}$$

is linearly independent. Suppose that we have coefficients

$$a_1^1, \cdots, a_{k_1}^1, a_1^2, \cdots, a_{k_2}^2 \cdots a_1^m, \cdots, a_{k_m}^m$$

such that

$$a_1^1\mathbf{b}_1^1 + \cdots + a_{k_1}^1\mathbf{b}_{k_1}^1 + a_1^2\mathbf{b}_1^2 + \cdots + a_{k_2}^2\mathbf{b}_{k_2}^2 + \cdots + a_1^m\mathbf{b}_1^m + \cdots + a_{k_m}^m\mathbf{b}_{k_m}^m = \mathbf{0}.$$

Then apply $(A - \lambda_2 I)(A - \lambda_3 I)\cdots(A - \lambda_m I)$ to both sides, by Lemma 5.2.2, we find that

$$(\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3)\cdots(\lambda_1 - \lambda_m)(a_1^1\mathbf{b}_1^1 + \cdots + a_{k_1}^1\mathbf{b}_{k_1}^1) = \mathbf{0}.$$

Since $\lambda_i \neq \lambda_j$ if $i \neq j$, we get

$$a_1^1\mathbf{b}_1^1 + \cdots + a_{k_1}^1\mathbf{b}_{k_1}^1 = \mathbf{0}.$$

Since $\{\mathbf{b}_1^i, \cdots, \mathbf{b}_{k_i}^i\}$ is a basis of E_{λ_1}, we see that $a_1^1 = \cdots = a_{k_1}^1 = 0$. Therefore,

$$\{\mathbf{b}_1^1, \cdots, \mathbf{b}_{k_1}^1, \mathbf{b}_1^2, \cdots, \mathbf{b}_{k_2}^2, \cdots, \mathbf{b}_1^m, \cdots, \mathbf{b}_{k_m}^m\}$$

is linearly independent. Since $k_1 + k_2 + \cdots + k_m = n$, we see that

$$P := \begin{bmatrix} \mathbf{b}_1^1 & \cdots & \mathbf{b}_{k_1}^1 & \mathbf{b}_1^2 & \cdots & \mathbf{b}_{k_2}^2 & \cdots & \mathbf{b}_1^m & \cdots & \mathbf{b}_{k_m}^m \end{bmatrix}$$

is the $n \times n$ matrix we want. □

Remark 5.2.3. With the formula $\det(A)\det(B) = \det(AB)$, can you see the relation between $\det(A)$ and all the eigenvalues of A?

Remark 5.2.4. What if we have

$$\dim E_{\lambda_1} + \dim E_{\lambda_2} + \cdots + \dim E_{\lambda_m} < n \quad ?$$

Is it possible to have

$$\dim E_{\lambda_1} + \dim E_{\lambda_2} + \cdots + \dim E_{\lambda_m} > n \quad ?$$

Example 5.2.5. For each of the following matrices, point out whether they can be diagonalized or not. If they are diagonalizable, write down the corresponding matrix P and $P^{-1}AP$.

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

Here we use the information in Example 5.1.3.

(1) Since $\dim(E_1) = 3$, I can be diagonalized. Let

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

we see that

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

(2) Since $\dim(E_1) + \dim(E_{-1}) + \dim(E_2) = 3$, A can be diagonalized. Let

$$P = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1/2 \\ 1 & 0 & 1 \end{bmatrix},$$

we see that

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

(3) Since $\dim(E_2) + \dim(E_1) = 3$, we see that B can be diagonalized. Let

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

we get

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

(4) Since $\dim(E_2)+\dim(E_3)+\dim(E_{-2})=3$, we see that C can be diagonalized. Let

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

we get

$$P^{-1}CP=\begin{bmatrix}2&0&0\\0&3&0\\0&0&-2\end{bmatrix}.$$

(5) Since $\dim(E_2)+\dim(E_{-4})=2$, we see that D cannot be diagonalized.

Homework 5.2.6. For **all** the following matrices, point out whether they can be diagonalized or not. If they are diagonalizable, write down the corresponding matrix P and $P^{-1}AP$.

$$A=\begin{bmatrix}2&0&0\\0&2&0\\0&0&2\end{bmatrix},B=\begin{bmatrix}3&2&0\\-3&-2&1\\2&2&1\end{bmatrix},C=\begin{bmatrix}\ln 3&0&0\\0&\sin(29^\circ)&0\\0&0&\pi\end{bmatrix},D=\begin{bmatrix}2&2&8\\0&2&4\\0&0&-4\end{bmatrix},E=\begin{bmatrix}3&1\\0&3\end{bmatrix}.$$

5.3 Complex Eigenvalues

There is another way to deal with complex eigenvalues. In fact, given a real matrix A, sometimes what we really want is to make the change of basis matrix P into a real matrix. This time $P^{-1}AP$ will not be a diagonal matrix, but it will become an anti-symmetric matrix, which is still convenient for further usage.

Let A be an $n \times n$ real matrix. Suppose $\lambda = a + bi$ ($a, b \in \mathbb{R}, i = \sqrt{-1}$) is a complex eigenvalue of A, then we can see that its conjugation $\bar{\lambda} = a - bi$ is also an eigenvalue.

Lemma 5.3.1. Let A be an $n \times n$ real matrix. If $\lambda \in \mathbb{C}$ is a complex eigenvalue of A, then $\bar{\lambda}$ is also an eigenvalue of A.

Proof. Expand $\det(A - \lambda I) = 0$ and suppose we get

$$a_n \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_0 = 0.$$

Then take conjugation on both sides, we get

$$a_n \bar{\lambda}^n + a_{n-1}\bar{\lambda}^{n-1} + \cdots + a_0 = 0.$$

This means $\det(A - \bar{\lambda} I) = 0$. Therefore, $\bar{\lambda}$ is also an eigenvalue. □

Furthermore, we have a similar property of eigenvectors of A corresponding to λ and $\bar{\lambda}$. The verification of the following lemma is straightforward.

Lemma 5.3.2. Let A be an $n \times n$ real matrix and suppose it has two complex eigenvalues λ and $\bar{\lambda}$. If we have $A\mathbf{v} = \lambda \mathbf{v}$, then we must have $A\bar{\mathbf{v}} = \bar{\lambda}\bar{\mathbf{v}}$. Furthermore, there is a one-to-one correspondence

$$\begin{aligned} E_\lambda &\to E_{\bar{\lambda}} \\ \mathbf{v} &\mapsto \bar{\mathbf{v}} \end{aligned}$$

between E_λ and $E_{\bar{\lambda}}$.

Now let A be an $n \times n$ real matrix with $\lambda_1 = a + bi$ and $\bar{\lambda}_1 = a - bi$ be two of its complex eigenvalues. If we can diagonalize it using complex eigenvalues, then the change of basis matrix P looks like

$$P = \begin{bmatrix} \mathbf{b}_1 & \overline{\mathbf{b}}_1 & \cdots \end{bmatrix}.$$

Then we get

$$A\begin{bmatrix}\mathbf{b}_1 & \overline{\mathbf{b}}_1 & \cdots\end{bmatrix} = \begin{bmatrix}\mathbf{b}_1 & \overline{\mathbf{b}}_1 & \cdots\end{bmatrix}\begin{bmatrix}\lambda_1 & & \\ & \bar{\lambda}_1 & \\ & & \ddots\end{bmatrix}.$$

Plug $\lambda_1 = a + bi$, $\bar{\lambda}_1 = a - bi$ into this equation and write $\mathbf{b}_1 = \mathbf{v}_1 - i\mathbf{w}_1$, $\bar{b}_1 = \mathbf{v}_1 - i\mathbf{w}_1$, where $\mathbf{v}_1$ and $\mathbf{w}_1$ are real vectors, we get

$$A\begin{bmatrix}\mathbf{v}_1 + i\mathbf{w}_1 & \mathbf{v}_1 - i\mathbf{w}_1 & \cdots\end{bmatrix} = \begin{bmatrix}a\mathbf{v}_1 - b\mathbf{w}_1 + i(a\mathbf{w}_1 + b\mathbf{v}_1) & a\mathbf{v}_1 - b\mathbf{w}_1 - i(a\mathbf{w}_1 + b\mathbf{v}_1) & \cdots\end{bmatrix}.$$

Now we apply column operation matrix

$$E = \begin{bmatrix}\frac{1}{2} & 0 & 0 & \cdots & 0\\ \frac{1}{2} & 1 & 0 & \cdots & 0\\ 0 & 0 & 1 & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \cdots & 1\end{bmatrix}$$

to both sides, we get

$$A\begin{bmatrix}\mathbf{v}_1 + i\mathbf{w}_1 & \mathbf{v}_1 - i\mathbf{w}_1 & \cdots\end{bmatrix}E = \begin{bmatrix}a\mathbf{v}_1 - b\mathbf{w}_1 + i(a\mathbf{w}_1 + b\mathbf{v}_1) & a\mathbf{v}_1 - b\mathbf{w}_1 - i(a\mathbf{w}_1 + b\mathbf{v}_1) & \cdots\end{bmatrix}E,$$

which is in fact

$$A\begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_1 - i\mathbf{w}_1 & \cdots\end{bmatrix} = \begin{bmatrix}a\mathbf{v}_1 - b\mathbf{w}_1 & a\mathbf{v}_1 - b\mathbf{w}_1 - i(a\mathbf{w}_1 + b\mathbf{v}_1) & \cdots\end{bmatrix}.$$

Now, apply

$$\tilde{E} = \begin{bmatrix}1 & -i & 0 & \cdots & 0\\ 0 & i & 0 & \cdots & 0\\ 0 & 0 & 1 & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \cdots & 1\end{bmatrix}$$

to both sides, we get

$$A\begin{bmatrix}\mathbf{v}_1 & \mathbf{w}_1 & \cdots\end{bmatrix} = \begin{bmatrix}a\mathbf{v}_1 - b\mathbf{w}_1 & a\mathbf{w}_1 + b\mathbf{v}_1 & \cdots\end{bmatrix}.$$

Rewrite the right-hand side into a product of two matrices, we see

$$A \begin{bmatrix} \mathbf{v}_1 & \mathbf{w}_1 & \cdots \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{w}_1 & \cdots \end{bmatrix} \begin{bmatrix} a & b & \\ -b & a & \\ & & \ddots \end{bmatrix}.$$

Chapter 6

Inner Product

6.1 Gram-Schmidt Process

First, we define the transpose of column vector.

Definition 6.1.1. Let $\mathbf{v} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ be an $n \times 1$ column vector. Then the transpose $\mathbf{v}^T$ of $\mathbf{v}$ is a $1 \times n$ row vector

$$\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}.$$

Similarly, we can define the transpose of matrix A.

Definition 6.1.2. Let A be an $m \times n$ matrix and write A into column vectors:

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

Then the transpose A^T of A is defined to be

$$A = \begin{bmatrix} \mathbf{b}_1^T \\ \mathbf{b}_2^T \\ \vdots \\ \mathbf{b}_n^T \end{bmatrix}.$$

Example 6.1.3. Let $\mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$. Then $\mathbf{v}^T = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$ and $A^T = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}$.

Here are some quick properties of transpose. The verification is straightforward.

Proposition 6.1.4. Let A be an $m \times n$ matrix and B be an $n \times k$ matrix.

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

(b) If $m = n$, then $\det(A) = \det(A^T)$.

(c) If $m = n$, then A^T and A has the same eigenvalues.

(d) If $m = n$ and suppose A^{-1} exists, then $(A^T)^{-1} = (A^{-1})^T$.

(e) $\text{rk}(A) = \text{rk}(A^T)$.

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

With the definition of transpose, we can define the inner product of two column vectors

Definition 6.1.5. Let $\mathbf{u}$ and $\mathbf{v}$ be two column vectors in $\mathbf{R}^n$, then their inner product is

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

We say that $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if $\mathbf{u} \cdot \mathbf{v} = 0$.

Remark 6.1.6. It is straightforward to see

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

(b) $\mathbf{u} \cdot \mathbf{u} \geq 0$ and we have "=" if and only if $\mathbf{u} = \mathbf{0}$.

(c) For scalar r, $(r\mathbf{u}) \cdot \mathbf{v} = r(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot (r\mathbf{v})$.

(d) $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$.

The remark guarantees the definition of length.

Definition 6.1.7. The length of a real column vector $\mathbf{u}$ is

$$|\mathbf{u}| = \sqrt{\mathbf{u} \cdot \mathbf{u}}.$$

Length and inner product has the well-known Cauchy-Schwarz inequality.

Theorem 6.1.8. (**Cauchy-Schwarz**) Let $\mathbf{u}$ and $\mathbf{v}$ be two column vectors in $\mathbf{R}^n$, then

$$|\mathbf{u}||\mathbf{v}| \geq \mathbf{u} \cdot \mathbf{v}.$$

Proof. For any real number α, we have

$$0 \le |\mathbf{u} - \alpha\mathbf{v}|^2 = |\mathbf{u}|^2 - 2\mathbf{u}\cdot\mathbf{v}\alpha + |\mathbf{v}|^2\alpha^2.$$

Thus the equation

$$|\mathbf{u}|^2 - 2\mathbf{u}\cdot\mathbf{v}\alpha + |\mathbf{v}|^2\alpha^2 = 0$$

with respect to α will satisfy

$$4(\mathbf{u}\cdot\mathbf{v})^2 - 4|\mathbf{u}|^2|\mathbf{v}|^2 \le 0.$$

□

In $Oxyz$ space, we have the following formula

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

which shows the angle between **u** and **v**. With this formula, we find that the projection of **u** to **v** is

$$\mathrm{proj}_{\mathbf{v}}\mathbf{u} = (|\mathbf{u}|\cos\alpha)\,\frac{\mathbf{v}}{|\mathbf{v}|} = \frac{\mathbf{u}\cdot\mathbf{v}}{|\mathbf{v}|^2}\mathbf{v}.$$

We can generalize this definition to any dimension.

Definition 6.1.9. Given two column vectors **u** and **v** in $\mathbb{R}^n$, the projection of **u** to **v** is

$$\mathrm{proj}_{\mathbf{v}}\mathbf{u} = \frac{\mathbf{u}\cdot\mathbf{v}}{|\mathbf{v}|^2}\mathbf{v} = \frac{\mathbf{u}^T\mathbf{v}}{|\mathbf{v}|^2}\mathbf{v}.$$

Here comes the main question in this section: Given an independent set

$$\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n\}$$

of column vectors, how do we turn it into another independent set

$$\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$$

satisfying

$$\mathbf{v}_i\cdot\mathbf{v}_j = 0$$

when $i \ne j$? In other words, we hope to get an orthogonal set from the original set.

The idea is, for each $\mathbf{u}_k$, we subtract the projection of it to $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_{k-1}$. The algorithm is given as follows. This is the Gram-Schmidt process.

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{u}_1 \\
\mathbf{v}_2 &= \mathbf{u}_2 - \operatorname{proj}_{\mathbf{v}_1}\mathbf{u}_2 \\
\mathbf{v}_3 &= \mathbf{u}_3 - \operatorname{proj}_{\mathbf{v}_1}\mathbf{u}_3 - \operatorname{proj}_{\mathbf{v}_2}\mathbf{u}_3 \\
&\vdots \\
\mathbf{v}_n &= \mathbf{u}_n - \operatorname{proj}_{\mathbf{v}_1}\mathbf{u}_n - \operatorname{proj}_{\mathbf{v}_2}\mathbf{u}_n - \cdots - \operatorname{proj}_{\mathbf{v}_{n-1}}\mathbf{u}_n.
\end{aligned}$$

If furthermore, we let

$$\mathbf{w}_1 = \frac{\mathbf{v}_1}{|\mathbf{v}_1|}, \mathbf{w}_2 = \frac{\mathbf{v}_2}{|\mathbf{v}_2|}, \cdots, \mathbf{w}_n = \frac{\mathbf{v}_n}{|\mathbf{v}_n|},$$

then we get the so-called Gram-Schmidt normalization.

Example 6.1.10. Let $A = \begin{bmatrix} 1 & 0 & 3 \\ 3 & 3 & -1 \\ 0 & 0 & 1 \end{bmatrix}$ and let $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ be its three columns respectively from left to right.

(a) Find $\mathbf{b}_1 \cdot \mathbf{b}_3$ and show that they are orthogonal.

(b) Show that $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ form a basis of $\mathbb{R}^3$.

(c) Apply Gram-Schmidt process to $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$.

(d) Normalize the orthogonal vectors you get.

Solution. (a) $\mathbf{b}_1 \cdot \mathbf{b}_3 = \begin{bmatrix} 1 & 3 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix} = 0$. Thus they are orthogonal.

(b) Equation $\begin{bmatrix} 1 & 0 & 3 \\ 3 & 3 & -1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ only has solution $x = y = z = 0$.

(c) After Gram-Schmidt process, we get

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -9/10 \\ 3/10 \\ 0 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

(d) Normalize them to get

$$\mathbf{w}_1 = \frac{1}{\sqrt{10}}\begin{bmatrix}1\\3\\0\end{bmatrix}, \mathbf{w}_2 = \frac{\sqrt{10}}{\sqrt{9}}\begin{bmatrix}-9/10\\3/10\\0\end{bmatrix}, \mathbf{w}_3 = \begin{bmatrix}0\\0\\1\end{bmatrix}.$$

□

Homework 6.1.11. Let $A = \begin{bmatrix}1 & 0 & 2\\0 & 2 & -1\\0 & 0 & 1\end{bmatrix}$ and let $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ be its three columns respectively from left to right.

(a) Find $\mathbf{b}_1 \cdot \mathbf{b}_2$ and show that they are orthogonal.

(b) Show that $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ form a basis of $\mathbb{R}^3$.

(c) Apply Gram-Schmidt process to $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$.

(d) Normalize the orthogonal vectors you get.

6.2 Projections and Least-Squares Method

Given a set of independent vectors $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ in $\mathbb{R}^n$ $(k \leq n)$, and given any vector $\mathbf{b} \in \mathbb{R}^n$, we have the following decomposition theorem.

Theorem 6.2.1. We can write $\mathbf{b}$ uniquely into $\mathbf{b}_1 + \mathbf{b}_2$, where $\mathbf{b}_1$ and $\mathbf{b}_2$ satisfy:

(a) $\mathbf{b}_1 \cdot \mathbf{v}_1 = \mathbf{b}_1 \cdot \mathbf{v}_2 = \cdots = \mathbf{b}_1 \cdot \mathbf{v}_k = 0$.

(b) $\mathbf{b}_2$ is a linear combination of $\mathbf{v}_1, \cdots, \mathbf{v}_k$.

Definition 6.2.2. The vector $\mathbf{b}_2$ is called the (orthogonal) projection of $\mathbf{b}$ to the hyperplane generated by $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$.

Proof. (Proof of Theorem 6.2.1) First, we look at a homogeneous equation system

$$\begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_k^T \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Since $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ is independent, we know that we do not have zero rows during the process of solving equations. Therefore, the dimension of the solution space is $n - k$.

Let $\{\mathbf{w}_1, \cdots, \mathbf{w}_{n-k}\}$ be an orthogonal basis (Why?) of the solution space. We claim that

$$\{\mathbf{v}_1, \cdots, \mathbf{v}_k, \mathbf{w}_1, \cdots, \mathbf{w}_{n-k}\}$$

is independent and hence is a basis of $\mathbb{R}^n$. In fact, suppose we have

$$s_1\mathbf{v}_1 + \cdots + s_k\mathbf{v}_k + t_1\mathbf{w}_1 + \cdots + t_{n-k}\mathbf{w}_{n-k} = \mathbf{0},$$

then we have

$$\mathbf{w}_1 \cdot (s_1\mathbf{v}_1 + \cdots + s_k\mathbf{v}_k + t_1\mathbf{w}_1 + \cdots + t_{n-k}\mathbf{w}_{n-k}) = 0.$$

Simplify it we get

$$t_1\mathbf{w}_1 \cdot \mathbf{w}_1 = 0.$$

Therefore $t_1 = 0$. Similarly, $t_2 = \cdots = t_{n-k} = 0$. Therefore,

$$s_1\mathbf{v}_1 + \cdots + s_k\mathbf{v}_k = \mathbf{0}.$$

Because $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ is independent, we see that $s_1 = \cdots = s_k = 0$.

Now that

$$\{\mathbf{v}_1, \cdots, \mathbf{v}_k, \mathbf{w}_1, \cdots, \mathbf{w}_{n-k}\}$$

is a basis of $\mathbb{R}^n$, we can uniquely write $\mathbf{b}$ into linear combination

$$p_1\mathbf{v}_1 + \cdots + p_k\mathbf{v}_k + q_1\mathbf{w}_1 + \cdots + q_{n-k}\mathbf{w}_{n-k}$$

Let $\mathbf{b}_2 = p_1\mathbf{v}_1 + \cdots + p_k\mathbf{v}_k$ and $\mathbf{b}_1 = q_1\mathbf{w}_1 + \cdots + q_{n-k}\mathbf{w}_{n-k}$, we finish the proof. □

We can also show that the projection $\mathbf{b}_2$ is the closest vector to $\mathbf{b}$ in the hyperplane.

Proposition 6.2.3. For any vector $\mathbf{v}$ in the hyperplane generated by $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$, we have

$$|\mathbf{b} - \mathbf{b}_2| \leq |\mathbf{b} - \mathbf{v}|.$$

Proof. This is equivalent to prove

$$(\mathbf{b} - \mathbf{b}_2) \cdot (\mathbf{b} - \mathbf{b}_2) \leq (\mathbf{b} - \mathbf{v}) \cdot (\mathbf{b} - \mathbf{v}).$$

Plug in $\mathbf{b} = \mathbf{b}_1 + \mathbf{b}_2$ and notice that $\mathbf{b}_1 \cdot \mathbf{v} = \mathbf{b}_1 \cdot \mathbf{b}_2 = 0$, we see it is equivalent to show that

$$\mathbf{b}_2 \cdot \mathbf{b}_2 - 2\mathbf{b}_2 \cdot \mathbf{v} + \mathbf{v} \cdot \mathbf{v} \geq 0.$$

By Cauchy-Schwarz inequality, we notice that

$$\mathbf{b}_2 \cdot \mathbf{b}_2 - 2\mathbf{b}_2 \cdot \mathbf{v} + \mathbf{v} \cdot \mathbf{v} \geq \mathbf{b}_2 \cdot \mathbf{b}_2 - 2|\mathbf{b}_2||\mathbf{v}| + \mathbf{v} \cdot \mathbf{v} = |\mathbf{b}_2 - \mathbf{v}|^2 \geq 0.$$

Then the proof is done. □

Let A be an $m \times n$ matrix. If you still remember the problem of solving non-homogeneous equations, you know that some non-homogeneous equations do not have solution. However, even if we do not have a solution for

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

we can at least find a column vector $\mathbf{v}$ so that

$$|A\mathbf{x} - \mathbf{b}|$$

reaches its minimum (This idea is from [3]: section 6.5). In other words,

$$|A\mathbf{v} - \mathbf{b}| = \min_{\mathbf{x} \in \mathbb{R}^n} |A\mathbf{x} - \mathbf{b}|.$$

According to Proposition 6.2.3, to obtain the minimum, we can project $\mathbf{b}$ to the hyperplane generated by the columns of A and get a projection $\mathbf{b}_2$. Since $\mathbf{b}_2$ is a linear combination of columns of A, the equation

$$A\mathbf{y} = \mathbf{b}_2$$

with respect to $\mathbf{y}$ must have solutions. We use this solution $\mathbf{y}$ as an approximation of $\mathbf{x}$ in the original $A\mathbf{x} = \mathbf{b}$.

However, as you can see from the Gram-Schmidt process, the calculation of projection is very time consuming. Therefore, we need another way to figure out this $\mathbf{y}$. If we have $A\mathbf{y} = \mathbf{b}_2$, then we apply A^T to each side from the left and get

$$A^T A\mathbf{y} = A^T\mathbf{b}_2 = A^T(\mathbf{b}_2 - \mathbf{b} + \mathbf{b}).$$

Since we know that $\mathbf{b} - \mathbf{b}_2$ is orthogonal to the hyperplane generated by the columns of A, we immediately have $A^T(\mathbf{b}_2 - \mathbf{b}) = 0$. Thus we see $\mathbf{y}$ satisfies

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

Conversely, if we find a solution $\mathbf{y}$ that satisfies

$$A^T A\mathbf{y} = A^T\mathbf{b},$$

we get another orthogonal projection $A\mathbf{y}$ of $\mathbf{b}$ to the hyperplane generated by the columns of A. Now we verify this fact. First, we see that

$$A^T(\mathbf{b} - A\mathbf{y}) = A^T\mathbf{b} - A^T A\mathbf{y} = 0.$$

Second, it is clear that $A\mathbf{y}$ is a linear combination of the columns of A. Therefore, we get another orthogonal projection. However, orthogonal projection is unique, so $A\mathbf{y} = \mathbf{b}_2$.

Theorem 6.2.4. (**Least-squares Method**) For equation $A\mathbf{x} = \mathbf{b}$, we can solve

$$A^T A\mathbf{y} = A^T\mathbf{b}$$

instead and use the solution $\mathbf{y}$ as an approximate solution.

Example 6.2.5. We can use least-squares method to find the best fit.

(a) Given two points $(1,2)$ and $(3,4)$, find a function $y = kx + b$ to fit these points.

(b) Given three points $(1,2)$, $(3,4)$ and $(-1,-1)$, can you find a function $y = kx + b$ that exactly fits all three points?

(c) Use least-squares method to find the best fit $y = kx + b$ for part (b).

Solution. (a) Plug those two points into $y = kx + b$ and get

$$k + b = 2$$
$$3k + b = 4.$$

Write it into matrix form

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

and then we find $k = 1, b = 1$.

(b) No, because the matrix form of the equation is

$$\begin{bmatrix} 1 & 1 \\ 3 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} k \\ b \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ -1 \end{bmatrix}$$

which does not have a solution (use row operations to see this).

(c) We solve equation

$$\begin{bmatrix} 1 & 3 & -1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 3 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} k \\ b \end{bmatrix} = \begin{bmatrix} 1 & 3 & -1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \\ -1 \end{bmatrix}$$

and get $k = 5/4, b = 5/12$. □

Example 6.2.6. We can also use functions of higher degrees to fit points.

(a) Given three points $(1, 0)$, $(2, 0)$ and $(0, 1)$, find a function $y = ax^2 + bx + c$ to fit these points.

(b) Given four points $(1, 0)$, $(2, 0)$, $(0, 1)$ and $(-1, -1)$, can you find a function $y = ax^2+bx+c$ that exactly fits all four points?

(c) Use least-squares method to find the best fit $y = ax^2 + bx + c$ for part (b).

Solution. (a) The parabola is $y = \frac{1}{2}x^2 - \frac{3}{2}x + 1$.

(b) No. Because the equation

$$\begin{bmatrix} 1 & 1 & 1 \\ 4 & 2 & 1 \\ 0 & 0 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}$$

does not have a solution (use row operations to see this.).

(c) We solve equation

$$\begin{bmatrix} 1 & 4 & 0 & 1 \\ 1 & 2 & 0 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 4 & 2 & 1 \\ 0 & 0 & 1 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 1 & 4 & 0 & 1 \\ 1 & 2 & 0 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}$$

and get $a = 1/2, b = 7/10, 4/10$. □

Homework 6.2.7. We can use least-squares method in higher dimensions.

(a) Given three points $(1,0,0)$, $(0,0,2)$ and $(0,3,0)$, find a function $z = ax + by + c$ to fit these points.

(b) Given four points $(1,0,0)$, $(0,0,2)$, $(0,3,0)$ and $(0,0,0)$, can you find a function $z = ax + by + c$ that exactly fits all four points?

(c) Use least-squares method to find the best fit $z = ax + by + c$ for part (b).

Bibliography

[1] M. Artin. *Algebra (2nd edition)*. Pearson Education, 2014.

[2] W. Greub. *Linear Algebra*. Graduate Texts in Mathematics. Springer New York, 1981.

[3] D. Lay, S. Lay, and J. McDonald. *Linear Algebra and Its Applications (5th edition)*. Pearson Education, 2015.