

---

# Matrix Algebra and Applications

---

## MATH 2111 Lecture Notes

LI, Hongrui

May 6, 2025

## Contents

# 1 Systems of Linear Equations

## 1.1 Systems of Linear Equations

At the start of this course, let's take a look at the word "Linear". Among all the types of single-variable functions and equations we have learnt before, the most simple and basic one is exactly the linear function, which is a function that can be written in the form  $f(x) = mx + b$ . The graph of a linear function is a straight line, and the slope of the line is  $m$ . In general, when we have more than one variable, a linear equation is an equation that can be written in the form

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

, where  $a_1, a_2, \dots, a_n$  and  $b$  are constants (which we call coefficients), and  $x_1, x_2, \dots, x_n$  are variables.

A system of linear equations is a set of two or more linear equations with the same variables. The solution to a system of linear equations is the set of all values of the variables that satisfy all the equations in the system. And a system of linear equations can have

1. No solution, or
2. Exactly one solution, or
3. Infinitely many solutions.

If there is no solution, we say the system is **inconsistent**; otherwise, we say the system is **consistent**. If there is exactly one solution, we say the system is **consistent and independent**. If there are infinitely many solutions, we say the system is **consistent and dependent**.

Previously, we have learnt how to solve a system of linear equations by using the method of substitution or elimination, which can be quite tedious and time-consuming when the number of equations and variables is large. Hence, we need to find a systematic and efficient way to solve a system of linear equations

In this course, we will use the language of matrices and vectors to represent and solve systems of linear equations. In particular, we will use the following notation:

- A system of linear equations will be represented by a matrix equation  $A\mathbf{x} = \mathbf{b}$ , where  $A$  is a matrix,  $\mathbf{x}$  is a vector of variables, and  $\mathbf{b}$  is a vector of constants.
- The solution to a system of linear equations will be represented by a set of vectors.
- The solution set of a system of linear equations will be represented by a subspace of  $\mathbb{R}^n$ , which can be denoted by a linear combination of vectors, a parametric vector form, or a set of vectors.

We will cover all the topics above in this chapter, and let's start with the matrix notation of a system of linear equations, coefficient matrix and augmented matrix.

### Example 1.1. Matrix Notation of a System of Linear Equations

Consider the following system of linear equations:

$$2x_1 + 3x_2 = 5$$

$$4x_1 + 5x_2 = 6$$

We can write this system of linear equations in matrix notation as

$$\begin{pmatrix} 2 & 3 \\ 4 & 5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 5 \\ 6 \end{pmatrix}$$

where the matrix on the left side is called the **coefficient matrix**, the matrix on the right side is called the **constant matrix**. Furthermore, we can write the **augmented matrix** of this system of linear equations as

$$\left( \begin{array}{cc|c} 2 & 3 & 5 \\ 4 & 5 & 6 \end{array} \right)$$

**Definition 1.1. Size of a Matrix**

The size of a matrix is denoted by the number of rows and columns in the matrix. For example, a matrix with  $m$  rows and  $n$  columns is called an  $m \times n$  matrix.

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

Recall how we solve a system of linear equations by using the method of substitution or elimination before. Representing the rules we follow previously in the language of matrices, we have the **Elementary Row Operations** as follows:

1. **Interchange** Interchange the  $i$ -th row and the  $j$ -th row.

$$\begin{pmatrix} \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{j1} & a_{j2} & \cdots & a_{jn} \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \rightarrow \begin{pmatrix} \vdots & \vdots & \vdots & \vdots \\ a_{j1} & a_{j2} & \cdots & a_{jn} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

2. **Scaling** Multiply the  $i$ -th row by a nonzero constant  $c$ .

$$\begin{pmatrix} \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \rightarrow \begin{pmatrix} \vdots & \vdots & \vdots & \vdots \\ ca_{i1} & ca_{i2} & \cdots & ca_{in} \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

3. **Replacement** Replace the  $i$ -th row by the sum of the  $i$ -th row and a multiple of the  $j$ -th row.

$$\begin{pmatrix} \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{j1} & a_{j2} & \cdots & a_{jn} \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} \rightarrow \begin{pmatrix} \vdots & \vdots & \vdots & \vdots \\ a_{i1} + ca_{j1} & a_{i2} + ca_{j2} & \cdots & a_{in} + ca_{jn} \\ \vdots & \vdots & \vdots & \vdots \\ a_{j1} & a_{j2} & \cdots & a_{jn} \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

**Proposition 1.1.** The solution set of a system of linear equations is unchanged if we perform the following operations (**Elementary Row Operations**) on the augmented matrix of the system:

1. Interchange two rows.
2. Multiply a row by a nonzero constant.
3. Replace a row by the sum of that row and a multiple of another row.

## 1.2 Row Reduction and Echelon Forms

Previously, after solving a system of linear equations we may represent the solution set by

$$\begin{aligned} x &= m \\ y &= n \end{aligned}$$

, a simple form of linear equations where  $m$  and  $n$  are constants. The ultimate goal of all your efforts spent in solving the problem is to write  $x, y$  like something above.

Similarly, when solving by using the matrix notation, we want to get a matrix in a form that is as simple as possible.

### Definition 1.2. Row Echelon Form

A rectangular matrix is in **echelon form** or **row echelon form** if it has the following three properties:

1. All zero rows are at the bottom of the matrix.
2. Each leading entry of a row is in a column to the right of the leading entry of the previous row.
3. All entries in a column below a leading entry are zeros.

If a matrix in echelon form satisfies the following additional properties, it is said to be in **reduced row echelon form**:

1. The leading entry in each nonzero row is 1.
2. Each leading 1 is the only nonzero entry in its column.

Note that **leading entry** is the first nonzero entry in a row.

### Definition 1.3. Pivot Position

A **pivot position** in a matrix is a location in the matrix that corresponds to a leading 1 in the reduced row echelon form of the matrix. A **pivot column** is a column that contains a pivot position.

We have a theorem that each matrix is row equivalent to one and only one reduced echelon matrix. If an echelon matrix  $U$  can be achieved by row operations from a matrix  $A$ , we say  $U$  is called an echelon form of  $A$ .

So far, we have all the theoretical foundations needed for introducing the **Row Reduction Algorithm**, the key of solving a system of linear equations by using the matrix notation.

**Example 1.2.** Apply elementary row operations to transform the following matrix first into echelon form and then into reduced echelon form:

$$\begin{pmatrix} 0 & 3 & -6 & 6 & 4 & -5 \\ 3 & -7 & 8 & -5 & 8 & 9 \\ 3 & -9 & 12 & -9 & 6 & 15 \end{pmatrix}$$

**Solution.** Step 1: Begin with the leftmost nonzero column. This is a pivot column. The pivot position is at the top.

Step 2: Select a nonzero entry in the pivot column as a pivot. If necessary, interchange rows to move this entry into the pivot position.

$$\begin{pmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 3 & -7 & 8 & -5 & 8 & 9 \\ 0 & 3 & -6 & 6 & 4 & -5 \end{pmatrix}$$

Step 3: Use row replacement operations to create zeros in all positions below the pivot.

$$\begin{pmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 0 & 2 & -4 & 4 & 2 & -6 \\ 0 & 3 & -6 & 6 & 4 & -5 \end{pmatrix}$$

Step 4: Cover (or ignore) the row containing the pivot position and cover all rows, if any, above it. Apply steps 1-3 to the submatrix that remains. Repeat the process until there are no more nonzero rows to modify.

$$\begin{pmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 0 & 2 & -4 & 4 & 2 & -6 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{pmatrix}$$

Step 5: Begin with the rightmost pivot and create zeros above it.

$$\begin{pmatrix} 3 & -9 & 12 & -9 & 0 & -9 \\ 0 & 2 & -4 & 4 & 0 & -14 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{pmatrix} \rightarrow \cdots \rightarrow \begin{pmatrix} 1 & 0 & -2 & 3 & 0 & -24 \\ 0 & 1 & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{pmatrix}$$

*Remark.* Step 1-4 are the steps to transform the matrix into echelon form, which is called the forward phase of the row reduction algorithm, and step 5 is the step to transform the matrix into reduced echelon form, which is called the backward phase.

### 1.3 Vector Equations

As we know, Vectors in  $\mathbb{R}^n$  is an ordered list of  $n$  numbers. The handwritten notes spend a lot of time on the arithmetic of vectors, which is quite trivial. This note will skip this part and start from the Linear combinations.

**Definition 1.4.** In general, given vectors  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$  in  $\mathbb{R}^n$  and given scalars  $c_1, c_2, \dots, c_p$ , the vector  $\mathbf{w}$

$$\mathbf{w} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p$$

is called a **linear combination** of  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$  with weights coefficients  $c_1, c_2, \dots, c_p$ .

**Example 1.3.** Determine whether  $\mathbf{b}$  is a linear combination of  $\mathbf{a}_1$  and  $\mathbf{a}_2$ .

$$\mathbf{a}_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \mathbf{a}_2 = \begin{pmatrix} 3 \\ 4 \end{pmatrix}, \mathbf{b} = \begin{pmatrix} 5 \\ 6 \end{pmatrix}$$

**Solution.** We need to find  $c_1$  and  $c_2$  such that  $c_1\mathbf{a}_1 + c_2\mathbf{a}_2 = \mathbf{b}$ . This is equivalent to solving the following system of linear equations:

$$\begin{aligned} c_1 + 3c_2 &= 5 \\ 2c_1 + 4c_2 &= 6 \end{aligned}$$

The augmented matrix of this system is

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

We can row reduce this matrix to

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

So  $c_1 = -1$  and  $c_2 = 2$ . Therefore,  $\mathbf{b}$  is a linear combination of  $\mathbf{a}_1$  and  $\mathbf{a}_2$ .

In general, a vector equation

$$\mathbf{b} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_p\mathbf{a}_p \quad \text{or} \quad \mathbf{b} = \sum_{i=1}^p x_i\mathbf{a}_i$$

has the same solution set as the linear system whose augmented matrix is

$$(\mathbf{a}_1 \quad \mathbf{a}_2 \quad \cdots \quad \mathbf{a}_p \quad \mathbf{b})$$

In particular,  $\mathbf{b}$  can be generated by a linear combination of  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_p$  if and only if there exists a solution to the linear system corresponding to the augmented matrix above.

### Definition 1.5. Span

If  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$  are in  $\mathbb{R}^n$  and  $c_1, c_2, \dots, c_p$  are scalars, then the set of all possible linear combinations of  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$  is called the **span** of  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$  and is denoted by  $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}$ .

That is,  $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}$  is the set of all vectors that can be written in the form  $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p$  for some scalars  $c_1, c_2, \dots, c_p$ .

### Example 1.4.

$$\text{Span}\left\{\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \\ 0 \end{pmatrix}\right\} = \left\{\begin{pmatrix} x \\ y \\ 0 \end{pmatrix} \mid x, y \in \mathbb{R}\right\}$$

Next let's discuss a geometric interpretation of the span of vectors.

- Let  $\mathbf{v}$  be a nonzero vector in  $\mathbb{R}^n$ . The span of  $\mathbf{v}$  is the line through the origin in the direction of  $\mathbf{v}$ .
- Let  $\mathbf{v}_1$  and  $\mathbf{v}_2$  be two nonzero vectors in  $\mathbb{R}^n$  that are not scalar multiples of each other. The span of  $\mathbf{v}_1$  and  $\mathbf{v}_2$  is the plane through the origin that contains  $\mathbf{v}_1$  and  $\mathbf{v}_2$ .

**Example 1.5.** For what values of  $h$  will  $\mathbf{y}$  be in the Span  $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ , if

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 5 \\ -4 \\ -7 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} -3 \\ 1 \\ 0 \end{pmatrix}, \text{ and } \mathbf{y} = \begin{pmatrix} -4 \\ 3 \\ h \end{pmatrix}$$

**Solution.** The vector  $\mathbf{y}$  belongs to Span  $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$  if and only if there exist scalars  $x_1, x_2, x_3$  such that

$$x_1 \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix} + x_2 \begin{pmatrix} 5 \\ -4 \\ -7 \end{pmatrix} + x_3 \begin{pmatrix} -3 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -4 \\ 3 \\ h \end{pmatrix},$$

which indicates that the following system of linear equations must have a solution:

$$\begin{aligned} x_1 + 5x_2 - 3x_3 &= -4 \\ -x_1 - 4x_2 + x_3 &= 3 \\ 2x_1 - 7x_2 &= h \end{aligned}$$

The augmented matrix of this system is

$$\left( \begin{array}{ccc|c} 1 & 5 & -3 & -4 \\ -1 & -4 & 1 & 3 \\ 2 & -7 & 0 & h \end{array} \right)$$

We can row reduce this matrix to

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

So the system has a solution if and only if  $h - 5 = 0$ , i.e.  $h = 5$ .

**Example 1.6.** The handwritten notes also provide examples about the application in "reality scenarios", please refer to file1.1

## 1.4 The Matrix Equation

### Study of the Matrix Equation

#### Definition 1.6. Matrix-Vector Product

$A$  is a  $m \times n$  matrix with column  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n \in \mathbb{R}^m$ , i.e.  $A = [\mathbf{a}_1, \dots, \mathbf{a}_n]$ , and  $\mathbf{x}$  is a vector in  $\mathbb{R}^n$ . Then we have

$$A \cdot \mathbf{x} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1 \mathbf{a}_1 + x_2 \mathbf{a}_2 + \cdots + x_n \mathbf{a}_n.$$

i.e. the product  $A \cdot \mathbf{x}$  is a linear combination of the columns of  $A$  with weights given by the entries of  $\mathbf{x}$ .

The form  $A\mathbf{x} = \mathbf{b}$  is called a **matrix equation**.



**Example 1.7.** Let  $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}$  and  $\mathbf{x} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ . Then

$$A \cdot \mathbf{x} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix} + 1 \begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix} = \begin{pmatrix} 4 \\ 10 \\ 16 \end{pmatrix}$$

In the example above, we calculated the matrix-vector Multiplication by strictly following the definition, which is a good way to avoid making mistakes in calculations or losing marks due to the steps issue.

From the definition, the equivalence of the matrix-vector multiplication and the linear combination of the columns of the matrix is quite trivial, and it's quite clear that the equation  $A \cdot \mathbf{x} = \mathbf{b}$  has the same solution set as the vector equation

$$\mathbf{b} = x_1 \mathbf{a}_1 + x_2 \mathbf{a}_2 + \cdots + x_n \mathbf{a}_n,$$

which, in turn, has the same solution set as the linear system whose augmented matrix is

$$[\mathbf{a}_1 \quad \mathbf{a}_2 \quad \cdots \quad \mathbf{a}_n \quad \mathbf{b}].$$

Simply following this definition, we have the following theorem.

**Proposition 1.2.** Let  $A$  be an  $m \times n$  matrix,  $\mathbf{b}$  be a vector in  $\mathbb{R}^m$ , and  $\mathbf{x}$  be a vector in  $\mathbb{R}^n$ . Then the following statements are equivalent:

1. For each  $\mathbf{b}$  in  $\mathbb{R}^m$ , the equation  $A \cdot \mathbf{x} = \mathbf{b}$  has a solution.
2. Each  $\mathbf{b}$  in  $\mathbb{R}^m$  is a linear combination of the columns of  $A$ .
3. The columns of  $A$  span  $\mathbb{R}^m$ .
4.  $A$  has a pivot position in every row.

*Proof.* The proof of  $1. \iff 2. \iff 3.$  is quite trivial, and the proof of  $2. \iff 4.$  has already been included in previous chapters. The handwritten lecture notes present a proof of  $(1.) \iff (4.)$ , but it seems to have assumed that  $(1.) \iff (2.)$  and  $(2.) \iff (4.)$  are true, and is not very clear in the perspective of rigorous reasoning.

## Computation of Matrix-Vector Multiplication

So far, we have been regarding  $A\mathbf{x}$  as a linear combination of the columns of  $A$ . Following this rule, we can write down the matrix-vector multiplication concisely where the  $i$ -th row,  $j$ -th column number in  $A$  is denoted by  $a_{ij}$ :

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

Take a closer look at each element in the result vector, we can see that the  $i$ -th element in the result vector is the dot product of the  $i$ -th row of  $A$  and the vector  $\mathbf{x}$ . Therefore, we can also write the matrix-vector multiplication as

### Proposition 1.3. Row vector Rule for Matrix-Vector Product

$$A\mathbf{x} = \begin{pmatrix} \mathbf{a}_1 \cdot \mathbf{x} \\ \mathbf{a}_2 \cdot \mathbf{x} \\ \vdots \\ \mathbf{a}_m \cdot \mathbf{x} \end{pmatrix}, \text{ where } \mathbf{a}_i \text{ is the } i\text{-th row of } A.$$

As all other operational rules, in the end of this subchapter, we present the properties of matrix-vector product as follows, which can be easily proved by examining each element in the result vector.

### Proposition 1.4. Properties of Matrix-Vector Product

Let  $A$  and  $B$  be matrices, and let  $\mathbf{x}$  and  $\mathbf{y}$  be vectors in  $\mathbb{R}^n$ . Then the following properties hold:

1.  $A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y}$ .
2.  $A(c\mathbf{x}) = c(A\mathbf{x})$ .
3.  $(A + B)\mathbf{x} = A\mathbf{x} + B\mathbf{x}$ .
4.  $A(B\mathbf{x}) = (AB)\mathbf{x}$ .
5.  $I\mathbf{x} = \mathbf{x}$ , where  $I$  is the identity matrix.

Remark: Property 3-5 are to be explained after defining the matrix-matrix operational rules in the next chapter.

## 1.5 Solution Sets of Linear Systems

### Homogeneous Linear Systems

#### Definition 1.7. Homogeneous Linear System

A linear system  $A\mathbf{x} = \mathbf{0}$  is called **homogeneous** if the vector of constants on the right side of the equation is the zero vector.

The solution set of a homogeneous linear system is a subspace of  $\mathbb{R}^n$ . It's obvious that the Homogeneous Linear System always has a solution  $\mathbf{x} = \mathbf{0}$ , which is called the **trivial solution**.

The homogeneous linear system  $A\mathbf{x} = \mathbf{0}$  has a nontrivial solution *if and only if* the system has at least one free variable.

**Example 1.8.** Determine whether the following homogeneous linear system has a nontrivial solution:

$$\begin{aligned}3x_1 + 5x_2 - 4x_3 &= 0 \\ -3x_1 - 2x_2 + 4x_3 &= 0 \\ 6x_1 + x_2 - 8x_3 &= 0\end{aligned}$$

**Solution.** The augmented matrix of this system is

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

We can row reduce this matrix to

$$\begin{pmatrix} 1 & 0 & -\frac{4}{3} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Hence, we get

$$\begin{aligned}x_1 &= \frac{4}{3}x_3 \\ x_2 &= 0 \\ x_3 &\text{ is a free variable}\end{aligned}$$

Therefore, the system has a nontrivial solution, and the solution set is

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} \frac{4}{3}x_3 \\ 0 \\ x_3 \end{pmatrix} = x_3 \begin{pmatrix} \frac{4}{3} \\ 0 \\ 1 \end{pmatrix}.$$

**Note. Parametric Vector Form** of the solution set of a homogeneous linear system is the form of the solution set that is expressed as a linear combination of one or more vectors, like

$$\mathbf{x} = s\mathbf{v}_1 + t\mathbf{v}_2 + \cdots \quad (s, t, \cdots \in \mathbb{R})$$

### Solutions of Nonhomogeneous Linear Systems

**Example 1.9.** Change the right side of the previous example to  $\begin{pmatrix} 7 \\ -1 \\ -4 \end{pmatrix}$ , and determine whether the system has a solution.

**Solution.** Skip the row reduction part, we get

$$\begin{aligned}x_1 - \frac{4}{3}x_3 &= -1 \\ x_2 &= 2 \\ x_3 &\text{ is a free variable}\end{aligned}$$

Therefore, the system has a solution, and the solution set is

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

In this example, we wrote in the form of

$$\mathbf{x} = \mathbf{p} + t\mathbf{v} \text{ where } t = x_3,$$

in which:

- $\mathbf{p}$  is a particular solution of the non-homogeneous equation  $A\mathbf{x} = \mathbf{0}$ , and
- $\mathbf{v}$  is the general solution of the corresponding homogeneous equation  $A\mathbf{x} = \mathbf{0}$ .

Now we can generalize this to the following theorem.

**Proposition 1.5. General Solution of a Nonhomogeneous Linear System**

Suppose  $A\mathbf{x} = \mathbf{b}$  is consistent for some given vector  $\mathbf{b}$ , and let  $\mathbf{p}$  be a particular solution. Then the general solution set of  $A\mathbf{x} = \mathbf{b}$  is the set of all vectors of the form

$$\mathbf{p} + t\mathbf{v}_n$$

, where  $\mathbf{v}_n$  is any solution of the corresponding homogeneous equation  $A\mathbf{x} = \mathbf{0}$ .

Hence, we end this topic by showing the following summary given in the handwritten notes.

- Row reduce the augmented matrix of the system to reduced row-echelon form.
- Express each basic variable in terms of any free variables.
- Write a typical solution  $\mathbf{x}$  as a vector whose entries depend on the free variables, if any.
- Decompose  $\mathbf{x}$  into a linear combination of vectors (with numeric entries) using the free variables as parameters.

*P.S.* For situations with more than 1 free variables, the handwritten notes present the solution in this way

$$\cdots = x_2 \begin{pmatrix} 4 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + x_6 \begin{pmatrix} -5 \\ 0 \\ 1 \\ 0 \\ 4 \\ 1 \end{pmatrix}$$

## 1.6 Linear Independence

**Example 1.10.**

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}$$

If  $x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3 = \mathbf{0}$ , are the solutions unique?

**Solution.** The vector equation  $x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3 = \mathbf{0}$  is equivalent to the homogeneous linear system whose augmented matrix is

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

We can row reduce this matrix to

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

Hence, we have

$$x_1 = 2x_3$$

$$x_2 = -x_3$$

In particular, we can write  $\mathbf{v}_3 = -2\mathbf{v}_1 + \mathbf{v}_2$ , which means  $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2\} = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$  is a plane. Therefore,  $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$  are linearly dependent.

From the example above, we can see that the vectors  $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$  are linearly dependent, which means that at least one of the vectors can be written as a linear combination of the others. From this observation, we can generalize the following definition.

### Definition 1.8. Linear Independence

A set of vectors  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$  in  $\mathbb{R}^n$  is said to be **linearly independent** if the vector equation

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \dots + x_p\mathbf{v}_p = \mathbf{0}$$

has only the trivial solution  $x_1 = x_2 = \dots = x_p = 0$ .

A set of vectors that is said to be **linearly dependent** if there exist weights  $c_1, c_2, \dots, c_p$ , not all zero, such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_p\mathbf{v}_p = \mathbf{0}.$$

For particular situations of the general definition above, we'll discuss the sets of one or two vectors as follows, the *proof* of which just follows the definition.

- A set of one vector  $\mathbf{v}$  is linearly independent if and only if  $\mathbf{v} \neq \mathbf{0}$ .
- A set of two vectors  $\mathbf{v}_1$  and  $\mathbf{v}_2$  is linearly independent if and only if neither vector is a scalar multiple of the other. i.e.  $\{\mathbf{v}_1, \mathbf{v}_2\}$  are linearly dependent if at least one of them is a scalar multiple of the other.

Moreover, for a set of three or more vectors, we can use the following proposition to determine whether they are linearly independent.

**Proposition 1.6.** A set of three or more vectors  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$  is linearly dependent if and only if at least one of the vectors can be written as a linear combination of the others.

Given the deductions above, if we want to prove a set of vectors are linearly dependent, one simple way is to find one vector that can be written as a multiple of another (for system with 2 vectors) or a linear combination of the others (for system with 3 or more vectors).

After getting familiar with the basic concepts, let's move on to some specific situations.

**Proposition 1.7.** Theorems for determining linear dependence

- For  $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}_{p \geq 2}$ , where  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$  are vectors in  $\mathbb{R}^n$ , if  $p > n$ , then  $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$  must be linearly dependent.
- If a set of vectors  $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$  in  $\mathbb{R}^n$  contains the zero vector, then  $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$  is linearly dependent.

## 1.7 Introduction to Linear Transformations

**Definition 1.9. Transformation**

A transformation from  $\mathbb{R}^n$  to  $\mathbb{R}^m$ ,  $T : \mathbb{R}^n \mapsto \mathbb{R}^m$ , is a rule that assigns to each vector  $\mathbf{x}$  in  $\mathbb{R}^n$  to a vector  $T(\mathbf{x})$  in  $\mathbb{R}^m$ . We call

$\mathbb{R}^n$  the **domain** of  $T$ ,

$\mathbb{R}^m$  the **codomain** of  $T$ .

For  $\mathbf{x} \in \mathbb{R}^n$ ,  $T(\mathbf{x})$  is called the **image** of  $\mathbf{x}$  under  $T$ , and the set of all images  $T(\mathbf{x})$  is called the **range** of  $T$ .

For example,

$$A : \text{matrix } m \times n, \quad \mathbf{x} \in \mathbb{R}^n, \quad T : \mathbb{R}^n \mapsto \mathbb{R}^m, \quad T(\mathbf{x}) \triangleq A\mathbf{x}$$

*Note.* The handwritten notes provide an example of  $T : \mathbb{R}^2 \mapsto \mathbb{R}^2$  regarding the geometric interpretation of the transformation, please refer to file1.2. The key ideas are:

- $T$  deforms the unit square into a parallelogram.
- $T$  maps a line segment to a line segment.

**Definition 1.10. Linear Transformation**

A transformation  $T : \mathbb{R}^n \mapsto \mathbb{R}^m$  is called a **linear transformation** if

$$\begin{aligned} T(\mathbf{u} + \mathbf{v}) &= T(\mathbf{u}) + T(\mathbf{v}) \quad \text{for all } \mathbf{u}, \mathbf{v} \in \mathbb{R}^n, \\ T(c\mathbf{u}) &= cT(\mathbf{u}) \quad \text{for all } \mathbf{u} \in \mathbb{R}^n \text{ and all scalars } c. \end{aligned}$$

i.e. a transformation  $T$  is linear if it preserves vector addition and scalar multiplication.

*Deduction.* If  $T$  is a linear transformation, then

$$\begin{aligned} T(\mathbf{0}) &= \mathbf{0}, \\ T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) &= c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2), \\ T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_p\mathbf{v}_p) &= c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2) + \dots + c_pT(\mathbf{v}_p). \end{aligned}$$

**Example 1.11. Geometric meaning of linear transformation**

- Given a scalar, define  $T : \mathbb{R}^2 \mapsto \mathbb{R}^2$  by  $T(\mathbf{x}) = r\mathbf{x}$ .  $T$  is called contraction if  $0 < r < 1$ , and dilation if  $r > 1$ .
- Define  $T : \mathbb{R}^2 \mapsto \mathbb{R}^2$  by  $T(\mathbf{x}) = A\mathbf{x}$ , where  $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ . Here  $T$  is a rotation.

## 1.8 The Matrix of a Linear Transformation

### Proposition 1.8. Matrix of a Linear Transformation

Let  $T : \mathbb{R}^n \mapsto \mathbb{R}^m$  be a linear transformation. Then there exists a unique  $m \times n$  matrix  $A$  such that  $T(\mathbf{x}) = A\mathbf{x}$  for all  $\mathbf{x}$  in  $\mathbb{R}^n$ .

In fact,  $A = [T(\mathbf{e}_1), T(\mathbf{e}_2), \dots, T(\mathbf{e}_n)]$ , where  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$  are the standard unit vectors in  $\mathbb{R}^n$ , which means

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \dots, \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

*Proof.* The key of this proof is to make use of the linearity of  $T$  stated in the Definition of Linear Transformation. Denote  $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ , then we have

$$\begin{aligned} T(\mathbf{x}) &= T(x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \dots + x_n\mathbf{e}_n) \\ &= x_1T(\mathbf{e}_1) + x_2T(\mathbf{e}_2) + \dots + x_nT(\mathbf{e}_n) \\ &= \begin{pmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) & \dots & T(\mathbf{e}_n) \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \\ &= A\mathbf{x}. \end{aligned}$$

## 2 Matrix Algebra

### 2.1 Matrix Operations

#### Sums and Scalar Multiplication

##### Example 2.1. Matrix Addition

1.  $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ ,  $B = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}$ , then  $A + B = \begin{bmatrix} 6 & 8 \\ 10 & 12 \end{bmatrix}$
2.  $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ ,  $B = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}$ , then  $A - B = \begin{bmatrix} -4 & -4 \\ -4 & -4 \end{bmatrix}$

Given two matrices  $A$  and  $B$  of the same size, the sum of  $A$  and  $B$  is the matrix obtained by adding corresponding elements of  $A$  and  $B$ . The difference of  $A$  and  $B$  is the matrix obtained by subtracting corresponding elements of  $A$  and  $B$ .

##### Example 2.2. Scalar Multiplication

1.  $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ ,  $k = 2$ , then  $kA = \begin{bmatrix} 2 & 4 \\ 6 & 8 \end{bmatrix}$
2.  $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ ,  $k = -1$ , then  $kA = \begin{bmatrix} -1 & -2 \\ -3 & -4 \end{bmatrix}$

Given a matrix  $A$  and a scalar  $k$ , the product of  $k$  and  $A$  is the matrix whose columns are  $k$  times the corresponding columns of  $A$ .

##### Proposition 2.1. Properties of Matrix Addition and Scalar Multiplication

1.  $A + B = B + A$
2.  $(A + B) + C = A + (B + C)$
3.  $A + 0 = A$
4.  $A + (-A) = 0$
5.  $k(A + B) = kA + kB$
6.  $(k + l)A = kA + lA$
7.  $k(lA) = (kl)A$
8.  $1A = A$

*Proof.* To prove Prop. 2.1, just consider the definition of matrix addition and scalar multiplication, and apply the Principles of numeral Addition and Multiplication.

#### Matrix Multiplication

Given two matrices  $A$  and  $B$ , for any vector  $\mathbf{x}$ , we want to make sure that  $(AB)\mathbf{x} = A(B\mathbf{x})$ . Starting from this point of view, we can define the matrix multiplication. **Justification of this definition remains to be solved.**



**Definition 2.1. Matrix Multiplication**

Let  $A$  be an  $m \times n$  matrix and  $B$  be an  $n \times p$  matrix, we now define that

$$AB = C = (A\mathbf{b}_1, \dots, A\mathbf{b}_p).$$

**Remark:** The product  $AB$  is defined if and only if the number of columns of  $A$  is equal to the number of rows of  $B$ , and the size of the product is  $m \times p$ .

**Example 2.3.** Find the product of the  $AB$ , where

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

**Solution:** To avoid making mistakes or losing points, follow strictly this definition and write in steps, or write down the calculation details of each elements.

Note that this definition actually uses the definition of matrix-vector multiplication, which is defined as the linear combination of the columns of the matrix. In many textbooks, the definition of matrix multiplication is defined in the following way, which is equivalent to the definition above and we regard it as Row-Column Rule for Matrix Multiplication in the context of MATH2111. In general, we have the following properties of matrix multiplication.

**Proposition 2.2. Row-Column Rule for Matrix Multiplication**

Let  $A$  be an  $m \times n$  matrix and  $B$  be an  $n \times p$  matrix, we now define that the product  $AB$  is the  $m \times p$  matrix whose  $(i, j)$ -entry is the dot product of the  $i$ -th row of  $A$  and the  $j$ -th column of  $B$ , that is,

$$(AB)_{ij} = \sum_{k=1}^n a_{ik}b_{kj},$$

$$\text{where } A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}, B = \begin{pmatrix} b_{11} & \cdots & b_{1p} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{np} \end{pmatrix}.$$

**Proposition 2.3. Properties of Matrix Multiplication**

1.  $A(BC) = (AB)C$
2.  $A(B + C) = AB + AC$
3.  $(A + B)C = AC + BC$
4.  $k(AB) = (kA)B = A(kB)$
5.  $I_m A = A = A I_n$

**Warning:**

- (1) In general,  $AB \neq BA$ , that is, the multiplication of matrices is not commutative. Moreover, the product of two matrices may not be defined, that is,  $AB$  may not be defined even if  $BA$  is defined.
- (2) In general,  $AB = AC$  does not imply  $B = C$ .
- (3) In general,  $AB = 0$  does not imply  $A = 0$  or  $B = 0$ .

## Powers and Transpose of a Matrix

### Definition 2.2. Powers of a Matrix

Let  $A$  be a  $n \times n$  matrix, then we define the power of  $A$  as follows:

$$\begin{aligned}A^0 &= I_n \\A^1 &= A \\A^2 &= A \cdot A \\A^3 &= A \cdot A \cdot A \\&\dots\end{aligned}$$

### Definition 2.3. Transpose of a Matrix

Let  $A$  be a  $m \times n$  matrix, then the transpose of  $A$ , denoted by  $A^T$ , is the  $n \times m$  matrix whose  $(i, j)$ -entry is the  $(j, i)$ -entry of  $A$ , that is,

$$(A^T)_{ij} = A_{ji}.$$

**Example 2.4.** Find the transpose of the matrix  $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$ .

**Solution:** We have  $A^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$ .

**Proposition 2.4. Properties of Transpose of a Matrix** Let  $A$  and  $B$  be matrices whose sizes are appropriate for the following operations, and let  $k$  be a scalar, then we have

1.  $(A^T)^T = A$
2.  $(kA)^T = kA^T$
3.  $(A + B)^T = A^T + B^T$
4.  $(AB)^T = B^T A^T$
5.  $(A^k)^T = (A^T)^k$

*Proof.* For Prop. 2.4.4, we have  $(AB)_{ij} = \sum_{k=1}^n a_{ik}b_{kj}$ , then

$$(AB)_{ij}^T = (AB)_{ji} = \sum_{k=1}^n a_{jk}b_{ki} = \sum_{k=1}^n b_{ki}a_{jk} = (B^T A^T)_{ij}.$$

Note that in Prop. 2.4.4,  $A^T B^T$  may not even be defined if  $A$  and  $B$  are not square matrices.

## 2.2 Inverses of Matrices and Elementary Matrices

### Inverses of Matrices

In many situations, we need to solve the equation  $A\mathbf{x} = B$ , where  $A$  is a matrix and  $\mathbf{x}$  is a vector. For numeral calculation, it's natural to find the solution by multiplying both sides by  $A^{-1}$  to get  $\mathbf{x} = A^{-1}B$ . In matrices, in order to solve this equation, we introduce the concept of the inverse of a matrix.

**Definition 2.4. Identity Matrices**

The  $n \times n$  identity matrix, denoted by  $I_n$ , is the matrix whose  $(i, j)$ -entry is 1 if  $i = j$  and 0 otherwise, that is,

$$(I_n)_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}.$$

**Definition 2.5. Inverses of Matrices**

Let  $A$  be a square matrix. If there exists a matrix  $B$  such that

$$AB = BA = I_n$$

, then  $B$  is called the inverse of  $A$ , denoted by  $A^{-1}$ .

Otherwise, if  $A$  does not have an inverse, then  $A$  is called singular.

Note that the inverse of a matrix is unique if it exists, which can be proved by contradiction. Assume that  $B$  and  $C$  are both inverses of  $A$ , then we have

$$B = BI_n = B(AC) = (BA)C = I_n C = C,$$

which shows that  $B = C$ . To find the unique inverse of a matrix, we can use the following method.

**Example 2.5.** Find the inverse of the matrix  $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ .

**Solution:** We have  $A^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$ , where  $a = 1$ ,  $b = 2$ ,  $c = 3$ ,  $d = 4$ . Thus,  $A^{-1} = \frac{1}{1 \cdot 4 - 2 \cdot 3} \begin{pmatrix} 4 & -2 \\ -3 & 1 \end{pmatrix} = \begin{pmatrix} -2 & 1 \\ 1.5 & -0.5 \end{pmatrix}$ .

*Theorem.* From this example, we can derive a general formula for the inverse of a  $2 \times 2$  matrix. Let  $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ , then we have

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

If  $ad - bc = 0$ , then  $A$  is singular and does not have an inverse. In general, we can use the Gauss-Jordan method to find the inverse of a matrix, which will be explained later. No matter whether a matrix is  $2 \times 2$  or not, however, we have the following properties covered in the lectures.

**Proposition 2.5. Properties of Inverses of Matrices** Let  $A$  and  $B$  be invertible matrices of the same size, and let  $k$  be a scalar, then we have

1.  $A^{-1}$  is invertible, and  $(A^{-1})^{-1} = A$
2.  $(kA)^{-1} = \frac{1}{k}A^{-1}$
3.  $(A^T)^{-1} = (A^{-1})^T$
4.  $(AB)^{-1} = B^{-1}A^{-1}$

*Proof.* For Prop. 2.5.3, just consider using the Properties of Transpose. For Prop. 2.5.4, we have

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I.$$

In the end, let's recall the problem of solving the equation  $A\mathbf{x} = B$ . With the concept of invertible matrices, we can now give a general solution. If  $A$  is invertible, then we can multiply both sides by  $A^{-1}$  to get  $\mathbf{x} = A^{-1}B$ . If  $A$  is not invertible, then we can use the Gauss-Jordan method to solve the equation

(which will be presented right after this part).

**Proposition 2.6. Solution of the Equation  $A\mathbf{x} = \mathbf{b}$**

If  $A$  is an invertible matrix, then the equation  $A\mathbf{x} = \mathbf{b}$  has **unique** solution  $\mathbf{x} = A^{-1}\mathbf{b}$

### Elementary Matrices

Recall the elementary row operations in Chapter 1, which are the following three operations: (1) Interchange two rows, (2) Multiply a row by a nonzero scalar, (3) Add a multiple of one row to another row. If an elementary row operation is performed on the  $m \times n$  matrix  $A$ , the resulting matrix can be written as  $EA$ , where  $E$  is the  $m \times m$  elementary matrix created by performing the elementary row operation on the  $m \times m$  identity matrix  $I_m$ .

Each elementary matrix  $E$  is invertible, and its inverse is also an elementary matrix. The inverse of  $E$  is the elementary matrix of the same type that transforms  $E$  back to  $I_m$ .

**Proposition 2.7.** An  $n \times n$  matrix  $A$  is invertible if and only if  $A$  is row equivalent to the  $n \times n$  identity matrix  $I_n$ . In this case, any sequence of elementary row operations that reduces  $A$  to  $I_n$  will also transform  $I_n$  into  $A^{-1}$ .

*Proof.* All the following statements are equivalent:

- $A$  is invertible
- $\Leftrightarrow$  The equation  $A\mathbf{x} = \mathbf{b}$  has a unique solution for all vectors  $\mathbf{b}$
- $\Leftrightarrow$  The equation  $A\mathbf{x} = \mathbf{0}$  has only the trivial solution
- $\Leftrightarrow$  The reduced row echelon form of  $A$  is  $I_n$  ( $A \sim I_n$ ).

Conversely, if  $A \sim I_n$ , each row operation corresponds to multiplying  $A$  by an elementary matrix, so  $E_p \cdots E_2 E_1 A = I_n$ , where  $E_1, E_2, \dots, E_p$  are elementary matrices. Thus we have

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

**Proposition 2.8. An algorithm for finding the inverse of a matrix (Gauss-Jordan method)**

Row operation reduce the matrix  $(A \ I_n)$ . If  $A$  is row equivalent to  $I_n$ , then  $(A \ I_n)$  is row equivalent to  $(I_n \ A^{-1})$ . Otherwise,  $A$  is not invertible.

**Example 2.6.** Find the inverse of the matrix  $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ .

**Solution:** We can use the following method to find the inverse of a matrix.

**Step 1:** Write down the augmented matrix  $(A \ I_2) = \begin{pmatrix} 1 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{pmatrix}$ .

**Step 2:** Use the row operations to transform the left side of the augmented matrix into the identity matrix:  $\begin{pmatrix} 1 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 & 1 & 0 \\ 0 & -2 & -3 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 & 1 & 0 \\ 0 & 1 & 3/2 & -1/2 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & -1 & 1 \\ 0 & 1 & 3/2 & -1/2 \end{pmatrix}$ .

**Step 3:** The right side of the augmented matrix is the inverse of the matrix  $A$ .

*Proof.*  $E_p \cdots E_1 (A \ I_n) = (E_p \cdots E_1 A \ E_p \cdots E_1 I_n) = (I_n \ A^{-1})$

## 2.3 Characterization of Invertible Matrices

**Proposition 2.9.** Let  $A$  be an  $n \times n$  matrix. The following statements are equivalent:

1.  $A$  is invertible.
2.  $A$  is row equivalent to  $I_n$ .
3.  $A$  has  $n$  pivot positions.
4. The equation  $A\mathbf{x} = \mathbf{0}$  has only the trivial solution.
5. The equation  $A\mathbf{x} = \mathbf{b}$  has a unique solution for all vectors  $\mathbf{b}$ .
6. The columns of  $A$  form a linearly independent set.
7. The columns of  $A$  span  $\mathbb{R}^n$ .
8. The linear transformation  $\mathbf{x} \mapsto A\mathbf{x}$  is one-to-one.
9. The linear transformation  $\mathbf{x} \mapsto A\mathbf{x}$  is onto.
10. There exists an  $n \times n$  matrix  $C$  such that  $CA = I_n$ .
11. There exists an  $n \times n$  matrix  $D$  such that  $AD = I_n$ .
12.  $A^T$  is invertible.

*Proof.* The equivalence of the statements can be proved by the following chain of implications:

$$\begin{aligned} (1) &\Rightarrow (2) \Rightarrow (3) \Rightarrow (4) \Rightarrow (5) \Rightarrow (6) \Rightarrow (7) \Rightarrow (8) \Rightarrow (9) \Rightarrow (10) \Rightarrow (11) \Rightarrow (12) \Rightarrow (1) \\ (1) &\Rightarrow (12) \Rightarrow (11) \Rightarrow (10) \Rightarrow (9) \Rightarrow (8) \Rightarrow (7) \Rightarrow (6) \Rightarrow (5) \Rightarrow (4) \Rightarrow (3) \Rightarrow (2) \Rightarrow (1) \end{aligned}$$

### Invertible Linear transformations

A linear transformation  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is invertible if and only if there exists a map  $S : \mathbb{R}^n \rightarrow \mathbb{R}^n$  such that  $T \circ S(\mathbf{x}) = S \circ T(\mathbf{x}) = \mathbf{x}$  for all  $\mathbf{x} \in \mathbb{R}^n$ . In this case,  $S$  is called the inverse of  $T$ , denoted by  $T^{-1}$ .

Let  $A$  be the standard matrix of the linear transformation  $T$ , then  $T$  is invertible if and only if  $A$  is invertible. In this case, the standard matrix of  $T^{-1}$  is  $A^{-1}$ .

To bring this chapter to an end, we show the following proposition given by *Github Copilot*.

**Proposition 2.10.** Let  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be a linear transformation. The following statements are equivalent:

1.  $T$  is invertible.
2.  $T$  is one-to-one.
3.  $T$  is onto.

## 3 Determinants

### 3.1 Introduction to Determinants

At the end of Chapter 2, we have discussed the invertibility of a linear transformation  $T : \mathbb{R}^T \mapsto \mathbb{R}^T$ , whose standard matrix happens to be a square matrix. In this chapter, we will introduce the concept of determinant, which is a scalar value that can be computed from the elements of a square matrix and encodes certain properties of the linear transformation described by the matrix. The determinant of a matrix is a fundamental concept in linear algebra, and it is used in many other areas of mathematics. Before we introduce the definition and properties of determinants, we claim that the thing called "determinant of a square matrix" has the following property:

**Proposition 3.1. Invertibility and Determinants**

Let  $A$  be an  $n \times n$  matrix. Then  $A$  is invertible if and only if  $\det(A) \neq 0$ .

**Corollary**

$\det A = 0$  if and only if rows of  $A$  are linearly dependent.

The next step is to give a definition of determinants. It's easy to observe that linear dependence is obvious when two columns or two rows are the same, or a column or a row is zero. Starting from a  $2 \times 2$  matrix, we find the following definition of determinant will work perfectly for what we want to be, and then we extend it to  $n \times n$  matrix by recursion.

**Remark:** The handwritten lecture notes presents the proposition above explicitly after defining the determinant of a matrix and showing it's properties. However, I think it's better to present the proposition first, since it gives a clear picture of why we need to introduce the concept of determinant.

**Definition 3.1. Determinant of a Matrix**

Let  $A$  be a  $2 \times 2$  matrix, then the determinant of  $A$ , denoted by  $\det(A)$  or  $|A|$ , is defined as the following

$$\det(A) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

In general, the determinant of an  $n \times n$  matrix  $A$ , and  $A_{ij}$  is the  $(n-1) \times (n-1)$  matrix obtained by deleting the  $i$ -th row and  $j$ -th column of  $A$ , is defined as  $\det(A) = \sum_{j=1}^n (-1)^{i+j} a_{ij} \det(A_{ij})$ .

**Cofactor**

The  $(i, j)$ -cofactor of  $A$  is defined as  $C_{ij} = (-1)^{i+j} \det(A_{ij})$ .

*Proof.* Given this definition, we can easily verify Prop. 3.1. Suppose  $A$  has been reduced to an echelon form  $U$  by row replacement and row interchanges. If there are  $r$  interchanges, then

$$\det A = (-1)^r \det U = (-1)^r \prod_{i=1}^n u_{ii}.$$

This part to be reconsidered!

The definition of determinant is a recursive definition, which is based on the definition of determinant of  $2 \times 2$  matrix. The determinant of a matrix can be calculated by the following formula:

**Proposition 3.2. Calculation of Determinant**

Let  $A$  be an  $n \times n$  matrix, then the determinant of  $A$  can be computed by a cofactor expansion across any row or down any column, that is,

$$\begin{aligned}\det(A) &= \sum_{j=1}^n a_{ij} C_{ij} \text{ (expansion across the } i\text{-th row)} \\ &= \sum_{i=1}^n a_{ij} C_{ij} \text{ (expansion down the } j\text{-th column).}\end{aligned}$$

**Thm for triangular matrix**

If  $A$  is a triangular matrix, then  $\det(A)$  is the product of the main diagonal entries of  $A$ .

$$\begin{pmatrix} * & * & * & \cdots & * \\ 0 & * & * & \cdots & * \\ 0 & 0 & * & \cdots & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & * \end{pmatrix} \text{ or } \begin{pmatrix} * & 0 & 0 & \cdots & 0 \\ * & * & 0 & \cdots & 0 \\ * & * & * & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ * & * & * & \cdots & * \end{pmatrix}$$

**Warning:**

- Mind the sign of each new determinant - indexes changed after each recursive call !!
- Remember to multiply the coefficient of each  $a_{ij}$  in the final results !!

## 3.2 Properties of Determinants

### Row and Column Operations

**Proposition 3.3. Row Operations**

Let  $A$  be an  $n \times n$  matrix.

- If a multiple of one row of  $A$  is added to another row to produce a matrix  $B$ , then  $\det(B) = \det(A)$ .
- If two rows of  $A$  are interchanged to produce a matrix  $B$ , then  $\det(B) = -\det(A)$ .
- If one row of  $A$  is multiplied by  $k$  to produce a matrix  $B$ , then  $\det(B) = k \det(A)$ .

**Example 3.1. Example**

Let  $A = \begin{pmatrix} 2 & -8 & 6 & 8 \\ 3 & -9 & 5 & 10 \\ -3 & 0 & 1 & -2 \\ 1 & -4 & 0 & 6 \end{pmatrix}$ . Calculate  $\det(A)$ .

**Solution** The answer is  $-36$  (please take care of all the calculation details: every step of the row operation may cause mistakes).

Now that we have the theorems for row operations, it's natural to think about whether the same properties work for column operations. The answer is yes, and the justification just follows the following theorem:

If  $A$  is an  $n \times n$  matrix, then  $\det(A^T) = \det(A)$ .

## Determinants of Products

### Proposition 3.4. Determinants of Products

Let  $A$  and  $B$  be  $n \times n$  matrices. Then  $\det(AB) = \det(A) \det(B)$ .

**Remark:**  $\det(A + B)$  is not  $\det(A) + \det(B)$ .

Using this proposition, it's easy to show that if  $A$  is invertible, then

$$\det(A^{-1}) = \frac{1}{\det(A)}$$

## A Linearity Property of Determinant Function

We have the following theorem:

$$\begin{aligned} & \det(\mathbf{a}_1, \dots, \mathbf{a}_{j-1}, \mathbf{a}_j + \mathbf{b}_j, \mathbf{a}_{j+1}, \dots, \mathbf{a}_n) \\ &= \det(\mathbf{a}_1, \dots, \mathbf{a}_{j-1}, \mathbf{a}_j, \mathbf{a}_{j+1}, \dots, \mathbf{a}_n) + \det(\mathbf{a}_1, \dots, \mathbf{a}_{j-1}, \mathbf{b}_j, \mathbf{a}_{j+1}, \dots, \mathbf{a}_n). \end{aligned}$$

## 3.3 Cramer's Rule, Volume, and Linear Transformations

### Cramer's Rule

To present Cramer's Rule, we first introduce a notation: For any  $n \times n$  matrix  $A$  and any  $\mathbf{b}$  in  $\mathbb{R}^n$ , let  $A_i(\mathbf{b})$  be the matrix obtained by replacing the  $i$ -th column of  $A$  by the column vector  $\mathbf{b}$ . In other words,

$$A = (a_1, \dots, a_i, \dots, a_n) \text{ and } A_i(\mathbf{b}) = (a_1, \dots, \mathbf{b}, \dots, a_n).$$

### Proposition 3.5. Cramer's Rule

Let  $A$  be an  $n \times n$  matrix and  $\mathbf{b}$  be in  $\mathbb{R}^n$ . If  $\det(A) \neq 0$ , then the unique solution of  $A\mathbf{x} = \mathbf{b}$  is given by

$$x_i = \frac{\det(A_i(\mathbf{b}))}{\det(A)} \text{ for } i = 1, 2, \dots, n.$$

*Proof.* Since the determinant of an arbitrary matrix  $A$  is quite a complicated problem, we try to get as many "0" as possible in the matrix when calculating the determinant. Hence, we can make use of identity matrix  $I_n = (\mathbf{e}_1, \dots, \mathbf{e}_i, \dots, \mathbf{e}_n)$  and  $I_n(\mathbf{b}) = (\mathbf{e}_1, \dots, \mathbf{b}, \dots, \mathbf{e}_n)$ . Then we have

$$\begin{aligned} A \cdot I_n(\mathbf{b}) &= (A\mathbf{e}_1, \dots, A\mathbf{b}, \dots, A\mathbf{e}_n) \\ &= (\mathbf{a}_1, \dots, \mathbf{b}, \dots, \mathbf{a}_n) \\ &= A_i(\mathbf{b}). \end{aligned}$$

$$\det(A \cdot I_n(\mathbf{b})) = \det(A) \cdot \det(I_n(\mathbf{b}))$$

When calculating the of  $I_n(\mathbf{b})$ , simply look at the  $i$ -th row of  $I_n$ , which only have one non-zero entry 1 to be multiplied with  $x_i$ , then we have

$$\det(I_n(\mathbf{b})) = (-1)^{i+i} \cdot x_i \cdot \det(I_{n-1}) = x_i.$$

Hence, we have  $x_i = \frac{\det(A_i(\mathbf{b}))}{\det(A)}$ .

There are many exercises of solving linear systems in previous chapters, so we don't give more examples here. The readers can solve the previous examples using the Cramer's Rule and check the correctness of the results.



## A formula for $A^{-1}$

**Proposition 3.6.** Let  $A$  be an  $n \times n$  matrix. If  $\det(A) \neq 0$ , then  $A^{-1} = \frac{1}{\det(A)} \text{adj}(A)$ , where  $\text{adj}(A)$  is the adjugate matrix of  $A$ :

$$\text{adj}(A) = \begin{pmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{pmatrix}$$

, where  $C_{ij}$  is the  $(i, j)$ -cofactor of  $A$ , i.e.  $C_{ij} = (-1)^{i+j} \det(A_{ij})$ .

**Warning** DO NOT mistake the meaning of  $C_{ij}$ , which is  $(-1)^{i+j}$  times the determinant of the  $(n-1) \times (n-1)$  matrix obtained by deleting the  $i$ -th row and  $j$ -th column of  $A$ . (NOT simply  $(-1)^{i+j} a_{ij}$ )  
For an example of this rule, please refer to the end of FILE Lecture 10.

## Determinants as Area of Volume

**Proposition 3.7.** Determinants as Area or Volume

- If  $A$  is a  $2 \times 2$  matrix, then  $|\det(A)|$  is the area of the parallelogram spanned by the columns of  $A$ .
- If  $A$  is a  $3 \times 3$  matrix, then  $|\det(A)|$  is the volume of the parallelepiped spanned by the columns of  $A$ .

The sign of  $\det(A)$  is determined by the orientation of the parallelogram or parallelepiped.

*Proof.* The theorem is obviously true for any  $2 \times 2$  diagonal matrix, since  $\det A = a_{11} \times a_{22} =$  area of rectangle. Hence, it suffice to show that any  $2 \times 2$  matrix can be transformed into a diagonal matrix in a way that changes neither the area of the associated parallelogram nor  $|\det A|$ .

- Let  $\mathbf{a}_1$  and  $\mathbf{a}_2$  be nonzero vectors, Then for any scalar  $c$ , the area of the parallelogram determined by  $\mathbf{a}_1$  and  $\mathbf{a}_2$  equals the area of the parallelogram determined by  $\mathbf{a}_1$  and  $\mathbf{a}_2 + c\mathbf{a}_1$ . Column interchanges do not change the parallelogram at all.
- The absolute value of the determinant is unchanged when two columns are interchanged or a multiple of one column is added to another. Such operation suffice to transform  $A$  into a diagonal matrix.

**Example 3.2.** Calculate the area of the parallelogram determined by the points  $(-2, -2)$ ,  $(0, 3)$ ,  $(4, -1)$ ,  $(6, 4)$

**Solution** Standard steps below:

First translate the parallelogram to one having the origin as a vertex. We subtract the vertex  $(-2, -2)$  from each of the other vertices to get the new vertices  $(0, 0)$ ,  $(2, 5)$ ,  $(6, 1)$ ,  $(8, 6)$ . The new parallelogram has the same area, and is determined by the columns of

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

Since  $|\det A| = |2 \times 1 - 6 \times 5| = |-28|$ , the area of the parallelogram is 28.

## Linear Transformations

### Proposition 3.8. Linear Transformations and Area or Volume

- Let  $T : \mathbb{R}^2 \mapsto \mathbb{R}^2$  be a linear transformation with standard matrix  $A$ . If  $S$  is a parrallelogram in  $\mathbb{R}^2$ , then

$$\text{area of } T(S) = |\det(A)| \cdot \text{area of } S.$$

- Let  $T : \mathbb{R}^3 \mapsto \mathbb{R}^3$  be a linear transformation with standard matrix  $A$ . If  $S$  is a parallelepiped in  $\mathbb{R}^3$ , then

$$\text{volume of } T(S) = |\det(A)| \cdot \text{volume of } S.$$

*Proof.* Just follows the previous proposition.

**Example 3.3.** Let  $a$  and  $b$  be positive numbers. Find the area of the region  $E$  bounded by the ellipse whose equation is  $\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1$ .

## 4 Vector Spaces

### 4.1 Vector Spaces and Subspaces

By observing the vectors we have previously encountered, we may find that all vectors in  $\mathbb{R}^n$  have some common properties. For example, the sum of two vectors in  $\mathbb{R}^n$  is still a vector in  $\mathbb{R}^n$ . In this section, we will introduce the concept of vector spaces, which is a generalization of the properties of vectors.

#### Example 4.1. Vector Space of Polynomials

Let  $P_n$  be the set of all polynomials of degree at most  $n$ , where any element in  $P_n$  can be written as

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n, \quad (1)$$

where  $a_0, a_1, \dots, a_n$  are real numbers. For any two polynomials  $f(x)$  and  $g(x)$  in  $P_n$ , we can add them together to get another polynomial in  $P_n$ :

$$(f + g)(x) = (a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + \cdots + (a_n + b_n)x^n. \quad (2)$$

For any element  $f(x)$  in  $P_n$ , we can multiply it by a scalar  $\alpha$  to get another element in  $P_n$ :

$$(\alpha f)(x) = \alpha a_0 + \alpha a_1x + \alpha a_2x^2 + \cdots + \alpha a_nx^n. \quad (3)$$

Therefore,  $P_n$  has two operations "addition" and "scalar multiplication". These operations satisfy the same rules as those on vectors in  $\mathbb{R}^n$ .

#### Example 4.2. Vector Space of Matrices

Let  $M_{m \times n}$  be the set of all  $m \times n$  matrices, where any element in  $M_{m \times n}$  can be written as

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad (4)$$

where  $a_{ij}$  are real numbers. For any two matrices  $A$  and  $B$  in  $M_{m \times n}$ , we can add them together to get another matrix in  $M_{m \times n}$ :

$$A + B = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2n} + b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \cdots & a_{mn} + b_{mn} \end{bmatrix}. \quad (5)$$

For any element  $A$  in  $M_{m \times n}$ , we can multiply it by a scalar  $\alpha$  to get another element in  $M_{m \times n}$ :

$$\alpha A = \begin{bmatrix} \alpha a_{11} & \alpha a_{12} & \cdots & \alpha a_{1n} \\ \alpha a_{21} & \alpha a_{22} & \cdots & \alpha a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha a_{m1} & \alpha a_{m2} & \cdots & \alpha a_{mn} \end{bmatrix}. \quad (6)$$

Now we see that  $\mathbb{R}^n$ ,  $P_n$  and  $M_{m \times n}$  have some common properties, even if they look totally different. Regarding these shared properties, we give the definition of vector spaces.

**Definition 4.1.** A **vector space** is a nonempty set  $V$  of elements called **vectors**, on which are defined two operations, called **addition** and **scalar multiplication**, subject to the ten axioms listed below. The axioms must hold for all vectors  $u, v, w$  in  $V$  and all scalars  $\alpha, \beta$ .

1. **Closure under Addition:** The sum  $u + v$  is a vector in  $V$ .
2. **Commutative Law of Addition:**  $u + v = v + u$ .
3. **Associative Law of Addition:**  $(u + v) + w = u + (v + w)$ .
4. **Existence of Zero Vector:** There exists a vector  $0$  in  $V$  such that  $u + 0 = u$  for all  $u$  in  $V$ .
5. **Existence of Additive Inverse:** For each  $u$  in  $V$ , there exists a vector  $-u$  in  $V$  such that  $u + (-u) = 0$ .
6. **Closure under Scalar Multiplication:** The product  $\alpha u$  is a vector in  $V$ .
7. **Distributive Law of Scalar Multiplication over Vector Addition:**  $\alpha(u + v) = \alpha u + \alpha v$ .
8. **Distributive Law of Scalar Multiplication over Scalar Addition:**  $(\alpha + \beta)u = \alpha u + \beta u$ .
9. **Associative Law of Scalar Multiplication:**  $\alpha(\beta u) = (\alpha\beta)u$ .
10. **Identity Law of Scalar Multiplication:**  $1u = u$ .

Furthermore, we have the concept of **subspace** of a vector space.

**Definition 4.2.** A **subspace** of a vector space  $V$  is a nonempty subset  $H$  of  $V$  that has the following three properties:

1. The zero vector of  $V$  is in  $H$ .
2.  $H$  is closed under addition.
3.  $H$  is closed under scalar multiplication.

An equivalent definition,  $H$  is a subspace of  $V$  if and only if for all  $u, v$  in  $H$  and all scalars  $\alpha, \beta$ , we have  $\alpha u + \beta v$  in  $H$ .

For example, the set consisting of zero vector only in a vector space is a subspace of the vector space, called the **zero subspace**. In addition, we can observe that a subset of a vector space in the form of

$$[f_1(x), f_2(x), \dots, f_n(x)]$$

can never be a subspace of the vector space if any of the  $f_i(x)$  is a constant other than zero. To further determine whether a subset is a subspace of a vector space, we can use the following theorem, whose proof directly follows the definition above.

**Proposition 4.1.** If  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$  are vectors in a vector space  $V$ , then the set of all linear combinations of  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$ , i.e.  $\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}$  is a subspace of  $V$ . Here  $H$  is called the subspace of  $V$  spanned by  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$ .

**Example 4.3.** Let  $H = \{(a - 3b, b - a, a, b) | a, b \in \mathbb{R}\}$ , show that  $H$  is a subspace of  $\mathbb{R}^4$ .

**Solution.** We can rewrite the set  $H$  as

$$H = \{a(1, 0, 1, 0) + b(-3, -1, 0, 1) | a, b \in \mathbb{R}\}.$$

Therefore,  $H$  is the span of the vectors  $(1, 0, 1, 0)$  and  $(-3, -1, 0, 1)$ . According to the previous proposition,

$H$  is a subspace of  $\mathbb{R}^4$ .

**Example 4.4.** For what values of  $h$  will  $\mathbf{y}$  be in the subspaces of  $\mathbb{R}^3$  spanned by  $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ , if

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ -1 \\ -2 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 5 \\ -4 \\ -7 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} -3 \\ 1 \\ 0 \end{pmatrix}, \mathbf{y} = \begin{pmatrix} -4 \\ 3 \\ h \end{pmatrix}?$$

**Solution.**  $\mathbf{y} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3$  has a solution for  $(x_1, x_2, x_3)^T$  if and only if  $\mathbf{y}$  is in the subspace spanned by  $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ . Thus, we can simply solve the system of equations (omitted here) to get the result  $h = 5$ .

## 4.2 Null Space, Column Space and Linear Transformation

### The null space of a matrix

**Definition 4.3. Null Space of a Matrix** The null space of an  $m \times n$  matrix  $A$ , denoted by  $\text{Nul } A$ , is the set of all solutions to the homogeneous equation  $A\mathbf{x} = \mathbf{0}$ . In set notation, we have

$$\text{Nul } A = \{\mathbf{x} | \mathbf{x} \in \mathbb{R}^n, A\mathbf{x} = \mathbf{0}\}.$$

$\text{Nul } A$  is the set of all  $\mathbf{x} \in \mathbb{R}^n$  that are mapped into the zero vector of  $\mathbb{R}^m$  via the transformation  $\mathbf{x} \mapsto A\mathbf{x}$ .

**Proposition 4.2.** The null space of an  $m \times n$  matrix  $A$  is a subspace of  $\mathbb{R}^n$ .

*Proof.* We can see that  $\mathbf{0}$  is in  $\text{Nul } A$  since  $A\mathbf{0} = \mathbf{0}$ . For any  $\mathbf{x}, \mathbf{y}$  in  $\text{Nul } A$  and any scalar  $\alpha$ , we have  $A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y} = \mathbf{0} + \mathbf{0} = \mathbf{0}$  and  $A(\alpha\mathbf{x}) = \alpha A\mathbf{x} = \alpha\mathbf{0} = \mathbf{0}$ . Therefore,  $\text{Nul } A$  is a subspace of  $\mathbb{R}^n$ .

**Example 4.5.** Let  $H = \{(a, b, c, d) | a - 2b + 5c = d, c - a = b\}$ . Show that  $H$  is a subspace of  $\mathbb{R}^4$ .

**Solution.** Let  $A = \begin{bmatrix} 1 & -2 & 5 & -1 \\ -1 & 1 & 1 & 0 \end{bmatrix}$ . Then  $H = \text{Nul } A$ . Since  $\text{Nul } A$  is a subspace of  $\mathbb{R}^4$ ,  $H$  is a subspace of  $\mathbb{R}^4$ . Of course, this problem can also be solved by the proposition in the previous section.

From this example, we may find that there seem to exist an relationship between the null space of a matrix and the subspace spanned by the vectors in the matrix. Actually, for each proper matrix  $A$ , it's easy to find that its null space is exactly the solution set of the homogeneous equation  $A\mathbf{x} = \mathbf{0}$  by definition (another trivial result treated as a subsection in MATH2111). Hence, we can also conclude that when  $\text{Nul } A$  contains nonzero vectors, the number of vectors in the spanning set for  $\text{Nul } A$  equals the number of free variables in the solution set of  $A\mathbf{x} = \mathbf{0}$ .

**Remark.** The handwritten notes summarized this as a method called "An Explicit Description of Null  $A$ " and provided an example asking readers to find a spanning set for the null space of a given matrix, which is simply solving the homogeneous equation  $A\mathbf{x} = \mathbf{0}$  and writing the solution set in the form of a linear combination of vectors.

### The column space of a matrix

**Definition 4.4.** The **column space** of an  $m \times n$  matrix  $A$ , denoted by  $\text{Col } A$ , is the set of all linear combinations of the columns of  $A$ . In set notation, we have

$$\text{Col } A = \{\mathbf{b} | \mathbf{b} = A\mathbf{x} \text{ for some } \mathbf{x} \in \mathbb{R}^n\}.$$

$\text{Col } A$  is the set of all  $\mathbf{b} \in \mathbb{R}^m$  that can be expressed as  $A\mathbf{x}$  for some  $\mathbf{x} \in \mathbb{R}^n$ .

Following this definition, we can easily get the following proposition.

**Proposition 4.3.** The column space of an  $m \times n$  matrix  $A$  is a subspace of  $\mathbb{R}^m$ .

The handwritten notes provides an example of finding a matrix  $A$  such that its column space is a given subset of  $\mathbb{R}^m$ . To solve this kind of problems, the standart steps are to first write the subset as a linear combination of vectors, and then a span of  $n$  vectors. Finally, augment the vectors to form an  $m \times n$  matrix.

Furthermore, please note that the column space of an  $m \times n$  matrix  $A$  is all of  $\mathbb{R}^m$  if and only if the equation  $A\mathbf{x} = \mathbf{b}$  has a solution for each  $\mathbf{b}$  in  $\mathbb{R}^m$ .

### Kernel and Range of a Linear Transformation

Recall that a linear transformation  $T$  from  $\mathbb{R}^n$  to  $\mathbb{R}^m$  is a function that satisfies the following two properties:

1.  $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$  for all  $\mathbf{u}, \mathbf{v}$  in  $\mathbb{R}^n$ , and
2.  $T(\alpha\mathbf{u}) = \alpha T(\mathbf{u})$  for all  $\mathbf{u}$  in  $\mathbb{R}^n$  and all scalars  $\alpha$ .

Now we expand the concept of linear transformation to a more general form.

**Definition 4.5.** A **linear transformation**  $T$  from a vector space  $V$  to a vector space  $W$  is a function that satisfies the following two properties:

1.  $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$  for all  $\mathbf{u}, \mathbf{v}$  in  $V$ , and
2.  $T(\alpha\mathbf{u}) = \alpha T(\mathbf{u})$  for all  $\mathbf{u}$  in  $V$  and all scalars  $\alpha$ .

Equivalently,  $T$  is a linear transformation if and only if  $T(\alpha\mathbf{u} + \beta\mathbf{v}) = \alpha T(\mathbf{u}) + \beta T(\mathbf{v})$  for all  $\mathbf{u}, \mathbf{v}$  in  $V$  and all scalars  $\alpha, \beta$ .

Also, recall that for each linear transformation  $T$ , there exists a unique  $m \times n$  matrix  $A$  such that  $T(\mathbf{x}) = A\mathbf{x}$  for all  $\mathbf{x}$  in  $\mathbb{R}^n$ . In previous chapters, we have discussed some relationships between the linear transformation and its corresponding matrix. Similarly, for null spaces and column spaces of a matrix, we also have the corresponding concepts for a linear transformation (the expanded version).

**Definition 4.6.** The **kernel** (or Null space) of a linear transformation  $T$  is

$$\{\mathbf{x} | T(\mathbf{x}) = \mathbf{0}\}.$$

The **range** of a linear transformation  $T$  is

$$\{\mathbf{b} | \mathbf{b} = T(\mathbf{x}) \text{ for some } \mathbf{x}\}.$$

## 4.3 Linearly Independent Sets; Bases

### Linear Independence

**Definition 4.7.** An indexed set of vectors  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}$  in a vector space  $V$  is said to be **linearly independent** if the vector equation

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \dots + x_p\mathbf{v}_p = \mathbf{0}$$

has only the trivial solution  $x_1 = x_2 = \dots = x_p = 0$ .

**Remark.** The set  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}$  is linearly dependent if there exist weights  $c_1, c_2, \dots, c_p$ , not all zero, such that  $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_p\mathbf{v}_p = \mathbf{0}$ . In such a case,  $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_p\mathbf{v}_p$  is called a **linear**

**dependence relation** among  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p$ . Directly following this, we get the theorem similar to the version for linear transformation  $\mathbb{R}^n \mapsto \mathbb{R}^m$ .

**Proposition 4.4.** An indexed set of vectors  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}$  of two or more vectors in a vector space  $V$ , with  $\mathbf{v}_1 \neq \mathbf{0}$  is linearly dependent *if and only if* at least one of the vectors in the set is a linear combination of the others.

**Remark.** The proof of the threorem above is quite trivial and can be done by simply following the definition. However, when trying to prove the "only if" part, remember to consider the situation where all coefficients are zero (which is impossible since  $\mathbf{v}_1 \neq \mathbf{0}$ ), following which we can construct some  $\mathbf{v}_j$  with  $c_j \neq 0$  as a linear combination of the other vectors.

For example, let  $P_1(t) = 1$   $P_2(t) = t$   $P_3(t) = 4 - t$ , then  $P_1(t), P_2(t), P_3(t)$  are linearly dependent since  $P_3(t) = 4P_1(t) - P_2(t)$ .

## Basis

**Definition 4.8.** Let  $H$  be a subspace of a vector space  $V$ . A **basis** for  $H$  is a linearly independent set in  $H$  that spans  $H$ . In other words, a set of vectors  $B = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}$  in  $V$  is a basis for  $H$  if

1.  $B$  is linearly independent, and
2. The subspace spanned by  $B$  coincides with  $H$ , that is,  $\text{Span } B = H$ .

Remark: The basis for the zero subspace is the empty set. Every subspace of a vector space has a basis, **which remains to be proved**.

For example, an  $n \times n$  matrix,  $A = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n]$ . If  $A$  is invertible, then the columns of  $A$  are linearly independent and form a basis for  $\mathbb{R}^n$ .

Another example is that the set of all polynomials of degree at most  $n$ , denoted by  $P_n$ , has a basis  $S = \{1, x, x^2, \dots, x^n\}$ . This can be proved by the argument that any  $f(x) \in P_n$ ,  $f(t) = a_0 + a_1t + a_2t^2 + \dots + a_nt^n$  can be written as a linear combination of the elements in  $S$ , and that if  $a_0 + a_1t + a_2t^2 + \dots + a_nt^n = 0$  as a polynomial, then  $a_0 = a_1 = \dots = a_n = 0$ , which implies that  $S$  is linearly independent.

Although we presented the concept of basis explicitly, after reading the definition above, it may be natural to note that "a set is a basis of  $H$ " implies "the set spans  $H$ ", while the converse is not necessarily true. However, if you have already find a set that spans  $H$ , which is not linearly independent, you may also wonder whether we can get a basis from the set by "reducing" it. You may want to remove some vectors from the set to make it linearly independent while making sure it still spans  $H$ . The following proposition will help you to do this.

**Proposition 4.5. The Spanning Set Theorem** If  $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}$  spans a subspace  $H$  of a vector space  $V$ , then

1. If one of the vectors in  $S$  - say,  $\mathbf{v}_k$  - is a linear combination of the remaining vectors in  $S$ , then the set obtained by removing  $\mathbf{v}_k$  from  $S$  still spans  $H$ .
2. If  $H \neq \{0\}$ , some subset of  $S$  is a basis for  $H$ .
3. If  $H = \{0\}$ , then  $S = \emptyset$  is a basis for  $H$ . (The empty set is linearly independent by definition.)

Hence, once you get a set that spans  $H$ , you can gradually remove the vectors in the set that are linear combinations of the others, until the time when deleting any vector will make the set not span  $H$ . Then the remaining set is linearly independent and spans  $H$ , which is a basis for  $H$ . Using this fact, we can give a corollary of the proposition above, a property of basis which the handwritten notes point out from nowhere.

$S$  is a basis of vector space  $V$ . If we delete one vector from  $S$ , the remaining set  $T$  is still linearly independent, but cannot generate  $V$ . If we add a vector in  $V$  to  $S$  to get  $R$ , it generate  $V$  but is not linearly independent.

Now that it's clear we can get a basis from a spanning set, the last thing is to find an efficient way to do the "reduction". Again, we can use the row reduction method to do this.

**Proposition 4.6.** The pivot columns of a matrix  $A$  form a basis for  $\text{Col}A$ . **proof remains to be done (the essence of row equivalent).**

Finally, the homework and examinations may ask you to do the following things:

- Find a basis for a given subspace of a vector space.
- Determine whether a given set of vectors is a basis for a given subspace of a vector space.

To find a basis, first solve a corresponding linear system to get the vectors that span the subspace, then verify whether the vectors are linearly independent, and reduce it if not. To determine whether a set of vectors is a basis, first check whether the set spans the subspace by checking pivot position, and then verify whether the set is linearly independent by calculating the determinant of the matrix formed by the vectors. Remember that a bunch of exercises is needed.

**Example 4.6.** Find a basis for the subspace  $W$  spanned by  $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$  where

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ -3 \\ 4 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 6 \\ 2 \\ -1 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} 2 \\ -2 \\ 3 \end{pmatrix}, \mathbf{v}_4 = \begin{pmatrix} -4 \\ -8 \\ 9 \end{pmatrix}.$$

**Solution.**

$$A \dim \begin{pmatrix} 1 & 6 & 2 & -4 \\ -3 & 2 & -2 & -8 \\ 4 & -1 & 3 & 9 \end{pmatrix} \dim \begin{pmatrix} 1 & 6 & 2 & -4 \\ 0 & 20 & 4 & -20 \\ 0 & -25 & -5 & 25 \end{pmatrix} \dim \begin{pmatrix} 1 & 6 & 2 & -4 \\ 0 & 5 & 1 & -5 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

The first two columns of  $A$  are the pivot columns and hence form a basis of  $\text{Col}A = W$ . Hence  $\{\mathbf{v}_1, \mathbf{v}_2\}$  is a basis for  $W$ .

## 4.4 Coordinate Systems

### The existence of coordinate system for a vector space

Recall the Cartesian coordinate system in  $\mathbb{R}^2$  and  $\mathbb{R}^3$ . In these coordinate systems, each point in the plane or in space is represented by a unique ordered pair or an ordered triple of real numbers. We may interpret "cartesian coordinate system" as a bijective function that maps each point in the space to a unique ordered pair or triple of real numbers. Hence, if we want to build a coordinate system for a vector space, we need to find a way that maps each vector in the space to a **unique** ordered pair or triple of real numbers. Luckily, we have the following theorem.

**Proposition 4.7.** Let  $B = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}$  be a basis for a vector space  $V$ . Then for each  $\mathbf{v}$  in  $V$ , there exists a unique set of scalars  $c_1, c_2, \dots, c_p$  such that

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_p\mathbf{v}_p.$$

The scalars  $c_1, c_2, \dots, c_p$  are called the **coordinates** of  $\mathbf{v}$  relative to the basis  $B$ .

*Proof.* Since  $B$  spans  $V$ , there exists scalars such that the equation above holds. Suppose  $\mathbf{v}$  also has the coordinates  $d_1, d_2, \dots, d_p$  relative to  $B$ . Then we have

$$\mathbf{v} - \mathbf{v} = (c_1 - d_1)\mathbf{v}_1 + (c_2 - d_2)\mathbf{v}_2 + \dots + (c_p - d_p)\mathbf{v}_p = \mathbf{0}.$$



Since  $B$  is linearly independent, we have  $c_1 = d_1, c_2 = d_2, \dots, c_p = d_p$ . Hence the coordinates are unique.

Bearing this in mind, we can define the **coordinate system** for a vector space  $V$  as follows.

**Definition 4.9.** Coordinate System

Suppose  $B = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}$  is a basis for a vector space  $V$  and  $\mathbf{x}$  is in  $V$ . The **coordinates of  $\mathbf{x}$  relative to the basis  $B$**  (or the  $B$ -coordinates of  $\mathbf{x}$ ) are the scalars  $c_1, c_2, \dots, c_p$  such that

$$\mathbf{x} = c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \dots + c_p \mathbf{v}_p.$$

If  $c_1, c_2, \dots, c_n$  are  $B$ -coordinates of  $\mathbf{x}$ , the the vector in  $\mathbb{R}^n$

$$[\mathbf{x}]_B = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{pmatrix}$$

is the **coordinate vector of  $\mathbf{x}$  (relative to the basis  $B$ )**, or the  **$B$ -coordinate vector of  $\mathbf{x}$** . The mapping

$$x \mapsto [\mathbf{x}]_B$$

is called the **coordinate mapping** (determined by  $B$ ).

**Remark.** A coordinate system on a set consists of a one-to-one mapping of the points in the set into the points in  $\mathbb{R}^n$ .

The handwritten notes gives some examples of coordinate systems in  $\mathbb{R}^3$ ,  $P_2$ , and  $M_{2 \times 2}$ , which are quite trivial and skipped here.

### Change of coordinates matrix

Let  $B = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}$  be a basis of  $\mathbb{R}^n$ . Let  $P_B = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n]$ . Then  $\mathbf{v} = c_1 \mathbf{b}_1 + c_2 \mathbf{b}_2 + \dots + c_n \mathbf{b}_n$  if and only if

$$\mathbf{v} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n] \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} = P_B [\mathbf{c}]_B.$$

Here  $P_B$  is called the **change of coordinates matrix** from  $B$  to the standard basis in  $\mathbb{R}^n$ . Since  $P_B$  is invertible, we can also define the change of coordinates matrix from the standard basis to  $B$  as  $P_B^{-1}$ . Simply following this definition, we can get the following proposition.

**Proposition 4.8. The Coordinate Mapping Theorem** Let  $B = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}$  be a basis for a vector space  $V$ , and let  $\mathbf{x}$  be in  $V$ . Then the coordinate mapping  $\mathbf{x} \mapsto [\mathbf{x}]_B$  is a one-to-one linear transformation from  $V$  onto  $\mathbb{R}^p$ . Furthermore, for each  $\mathbf{y} = a\mathbf{u} + b\mathbf{v}$  in  $V$ , the equation

$$\mathbf{y} = a\mathbf{u} + b\mathbf{v}$$

Although I don't quite understand why the handwritten notes put the following definition explicitly after the proposition above, I still write it down here. The word "isomorphism" is not necessary in the reasoning process so far, but it may be useful in the homework and exams, so it's necessary to remember it. In the logic of understanding, it should be put in plain text as a supplementary terminology, but here I still write it as a definition to stress the importance.

**Definition 4.10.** A one-to-one linear transformation from a vector space  $V$  a vector space  $W$  is called an **isomorphism** of  $V$  with  $W$ .

**Example 4.7.** Use coordinate vectors to verify that the polynomials  $1 + 2t^2$ ,  $4 + t + 5t^2$  and  $3 + 2t$  are linearly independent in  $P_2$ .

**Solution.** Let  $B = \{1, t, t^2\}$  be the basis for  $P_2$ . Then

$$\begin{aligned} P_2 &\mapsto \mathbb{R}^3 \\ P(t) &\mapsto [P(t)]_B \end{aligned}$$

is an isomorphism. Hence, we have

$$[1 + 2t^2]_B = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} = \mathbf{v}_1, [4 + t + 5t^2]_B = \begin{pmatrix} 4 \\ 1 \\ 5 \end{pmatrix} = \mathbf{v}_2, [3 + 2t]_B = \begin{pmatrix} 3 \\ 2 \\ 0 \end{pmatrix} = \mathbf{v}_3.$$

Thus, the three polynomials are linearly independent if and only if the matrix  $A = (\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$  is invertible. After this, the problem can be solved in two ways: by calculating the determinant of  $A$  or by row reducing  $A$  to see if  $A\mathbf{x} = \mathbf{0}$  has only the trivial solution.

## 4.5 Dimensions of a Vector Space

Recall what we have learnt in section 4.3, each vector space has a basis. And for any set of vectors that spans the vector space, it is either a basis or can be reduced to a basis by removing some vectors. Under this view, it's natural to regard basis as some kind of "final stage" of the sets of vectors that spans the vector space. While different spanning sets may be reduced to different basis, there may be some underlying properties of the vector space itself independent from the choice of spanning vectors, that is, the number of vectors in the basis. This observation can be stated formally as the following theorem.

**Proposition 4.9.** If a vector space  $V$  has a basis with  $n$  vectors, then any set in  $V$  containing more than  $n$  vectors must be linearly dependent.

*Proof.* One simple idea is to prove by contradiction. Suppose  $B = \{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n\}$  is a basis for  $V$ . If  $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n, \mathbf{v}_{n+1}, \dots\}$  is linearly independent, then  $S' = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$  must be linearly independent. Hence, for any  $1 \leq i \leq n + 1$ , denote

$$\mathbf{v}_i = c_{1i}\mathbf{b}_1 + c_{2i}\mathbf{b}_2 + \dots + c_{ni}\mathbf{b}_n,$$

then, since  $S'$  is linearly independent, we must have the set  $\{\mathbf{c}_i : \mathbf{c}_i = (c_{1i}, c_{2i}, \dots, c_{ni})^T \text{ for any integer } i \in [1, n]\}$  to be linearly independent (otherwise, by the property of vector spaces, we can find one vector in  $S'$  to be a linear combination of the others). Now to determine whether  $\mathbf{v}_{n+1}$  can be written as a linear combination of the vectors in  $S'$ , we can simply solve the equation

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \dots + x_n\mathbf{v}_n = \mathbf{v}_{n+1} = (d_1, d_2, \dots, d_n)^T,$$

which is equivalent to

$$\begin{pmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{pmatrix},$$

which must have a solution since the set  $\{\mathbf{c}_i\}$  is linearly independent. Hence,  $\mathbf{v}_{n+1}$  can be written as a linear combination of the vectors in  $S'$ , which contradicts the assumption that  $S$  is linearly independent.

The handwritten notes also provides a proof as follows.

$\mathbf{u}_1, \dots, \mathbf{u}_p$  in  $V$  with  $p > n$

$[\mathbf{u}_1]_B, \dots, [\mathbf{u}_p]_B$  form a linearly dependent set in  $\mathbb{R}^n$  because one of the columns in the matrix  $[\mathbf{u}_1, \dots, \mathbf{u}_p]$  is not a pivot column.

Thus there exists  $c_1, \dots, c_p$  not all zeros such that  $c_1[\mathbf{u}_1]_B + \dots + c_p[\mathbf{u}_p]_B = \mathbf{0}$ .

Hence  $[c_1\mathbf{u}_1 + \dots + c_p\mathbf{u}_p]_B = \mathbf{0}$ , which implies that  $\mathbf{u}_1, \dots, \mathbf{u}_p$  are linearly dependent.

**Proposition 4.10.** If a vector space  $V$  has a basis with  $n$  vectors, then every basis for  $V$  must consist of exactly  $n$  vectors.

*Proof.* If there exists two basis  $B_1$  and  $B_2$  for  $V$  with  $n_1$  and  $n_2$  vectors respectively. Without loss of generality, we can assume  $n_1 < n_2$ . Using the previous theorem, we can find that  $B_2$  is linearly dependent, which contradicts the assumption that  $B_2$  is a basis.

**Definition 4.11.** If  $V$  is spanned by a finite set, then  $V$  is said to be **finite-dimensional**. The **dimension** of  $V$ , denoted by  $\dim V$ , is the number of vectors in a basis for  $V$ . The dimension of the zero subspace  $\{0\}$  is defined to be zero.

If  $V$  is not spanned by a finite set, then  $V$  is said to be **infinite-dimensional**.

For example, the dimension of  $P_n$  is  $n + 1$ , the dimension of  $M_{m \times n}$  is  $mn$ , and the dimension of  $P$ -all polynomials in one variable- is infinite.

**Example 4.8.** Find the dimension of the subspace

$$H = \left\{ \begin{pmatrix} a - 3b + 6c \\ 5a + 4d \\ b - 2c - d \\ 5d \end{pmatrix} : a, b, c, d \in \mathbb{R} \right\}$$

**Solution.**

$$\begin{pmatrix} a - 3b + 6c \\ 5a + 4d \\ b - 2c - d \\ 5d \end{pmatrix} = a \begin{pmatrix} 1 \\ 5 \\ 0 \\ 0 \end{pmatrix} + b \begin{pmatrix} -3 \\ 0 \\ 1 \\ 0 \end{pmatrix} + c \begin{pmatrix} 6 \\ 0 \\ -2 \\ 0 \end{pmatrix} + d \begin{pmatrix} 0 \\ 4 \\ -1 \\ 5 \end{pmatrix}$$

Hence,  $H$  is spanned by the set

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

Hence, we perform row reduction to the following matrix

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

Since column 1, 2, 4 are pivot columns, the set is linearly dependent, and the dimension of  $H$  is 3.

Note that the dimension of a vector space is a property of a vector space, so we may wonder the relationship of their dimensions if two vector spaces are related in some way, especially the relationship of subspaces.

**Proposition 4.11.** Let  $H$  be a subspace of a finite-dimensional vector space  $V$ . Then  $H$  is finite-dimensional and  $\dim H \leq \dim V$ .

**Proposition 4.12. The Dimension Theorem** Let  $V$  be a  $p$ -dimensional vector space ( $p \geq 1$ ), then any linearly independent set of exactly  $p$  elements in  $V$  is automatically a basis of  $V$ . Any set of exactly  $p$  elements that spans  $V$  is automatically a basis of  $V$ .

### The dimension of Null $A$ and Col $A$

Facts:

- The dimension of the null space of an  $m \times n$  matrix  $A$  is the number of free variables in the solution set of  $A\mathbf{x} = \mathbf{0}$ .
- The dimension of the column space of an  $m \times n$  matrix  $A$  is the number of pivot columns in  $A$ .

**Example 4.9.** Find the dimension of the null space and the column space of the matrix

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

Answer: 3 and 2.

**Example 4.10.** Let  $H$  and  $K$  be subspaces of a vector space  $V$ . Prove that

1.  $H \cap K$  is a subspace of  $V$ .
2.  $H \cup K$  is a subspace of  $V$  if and only if  $H \subset K$  or  $K \subset H$ .
3.  $\dim(H \cap K) \leq \dim H$ .

## 4.6 Rank

### Row Space

For each matrix  $A$ , we denote

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \triangleq \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_m \end{pmatrix},$$

**row space** of  $A$ , denoted by  $\text{row } A$ , is the subspace of  $\mathbb{R}^n$  spanned by the rows of  $A$ , that is,

$$\text{row } A = \text{Span}\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m\}.$$

If two matrices  $A$  and  $B$  are row equivalent, then they have the same row space. If  $B$  is in the echelon form, the nonzero rows of  $B$  form a basis for the row space of  $A$ .

$\subseteq$  *Proof.* Since  $B$  is obtained from  $A$  by row operations, the rows of  $B$  are linear combinations of the rows of  $A$ . Hence, any row  $B$  is in  $\text{row } A$ . Similarly, we can prove that  $\text{row } A \subset \text{row } B$ . Hence,  $\text{row } A = \text{row } B$ .

Hence, we have enough knowledge to find  $\text{Row } A$ ,  $\text{Col } A$ ,  $\text{Null } A$ , and  $\text{Nul } A^T$  for a given matrix  $A$ , all of which involves row reduction.

**Example 4.11.** Find basis for Row  $A$ , Col  $A$ , Null  $A$  for the matrix

$$A = \begin{pmatrix} -2 & -5 & 8 & 0 & -17 \\ 1 & 3 & -5 & 1 & 5 \\ 3 & 11 & -19 & 7 & 1 \\ 1 & 7 & -13 & 5 & -3 \end{pmatrix}.$$

**Solution.** We perform row reduction on  $A$  and get

$$\begin{pmatrix} -2 & -5 & 8 & 0 & -17 \\ 1 & 3 & -5 & 1 & 5 \\ 3 & 11 & -19 & 7 & 1 \\ 1 & 7 & -13 & 5 & -3 \end{pmatrix} \sim \begin{pmatrix} 1 & 3 & -5 & 1 & 5 \\ 0 & 1 & -2 & 2 & -7 \\ 0 & 0 & 0 & -4 & 20 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Hence, the basis for Row  $A$  is  $\{(1 \ 3 \ -5 \ 1 \ 5), (0 \ 1 \ -2 \ 2 \ -7), (0 \ 0 \ 0 \ -4 \ 20)\}$ ,

the basis for Col  $A$  is  $\left\{ \begin{pmatrix} -2 \\ 1 \\ 3 \\ 1 \end{pmatrix}, \begin{pmatrix} -5 \\ 3 \\ 11 \\ 7 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 7 \\ 5 \end{pmatrix} \right\}$ ,

and the basis for Null  $A$  is  $\left\{ \begin{pmatrix} 1 \\ 2 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ -3 \\ 0 \\ 5 \\ 1 \end{pmatrix} \right\}$ .

Remark: Row operations may change the linear dependence relations among the rows of a matrix, but not the dimension of the row space.

## Rank

**Definition 4.12.** The **rank** of a matrix  $A$ , denoted by  $\text{rank } A$ , is the dimension of the column space of  $A$ , i.e.,

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

**Proposition 4.13. The Rank Theorem** Let  $A$  be an  $m \times n$  matrix. Then

$$\text{rank } A + \dim \text{Null } A = n,$$

and

$$\dim \text{Col } A = \dim \text{Row } A = \text{rank } A = \text{number of pivot columns in } A.$$

*Proof.* Since the pivot columns of a matrix forms a basis for Col  $A$ , we have

$$\dim \text{Col } A = \text{number of pivot columns in } A = \text{rank } A$$

, and number of pivots + number of free variables =  $n$ .

## 5 Eigenvalues and Eigenvectors

### 5.1 Eigenvalues and Eigenvectors

**Definition 5.1.** Let  $A$  be an  $n \times n$  matrix. A scalar  $\lambda$  is called an **eigenvalue** of  $A$  if there exists a *nonzero* vector  $\mathbf{x}$  in  $\mathbb{R}^n$  such that

$$A\mathbf{x} = \lambda\mathbf{x} \quad (7)$$

The vector  $\mathbf{x}$  is called an **eigenvector** of  $A$  corresponding to  $\lambda$ .

**Example 5.1.** Show that 7 is an eigenvalue of the matrix

$$A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} \quad (8)$$

and find the corresponding eigenvectors. **Solution.** The scalar 7 is an eigenvalue of  $A$  if and only if the equation

$$A\mathbf{x} = 7\mathbf{x} = 7I_2\mathbf{x}$$

has a nontrivial solution, which is equivalent to

$$(A - 7I_2)\mathbf{x} = \mathbf{0}. \quad (9)$$

We perform the row reduction on the augmented matrix (details should be shown in the exam), and get the result that  $x_1 - x_2 = 0$ . Hence, the general solution is

$$\mathbf{x} = t \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (10)$$

So  $(1, 1)^T$  is an eigenvector of  $A$  corresponding to the eigenvalue 7.

Instead of giving the general solution, we first consider the special case when the matrix is triangular.

**Proposition 5.1.** Let  $A$  be an  $n \times n$  upper triangular matrix. Then the eigenvalues of  $A$  are the diagonal entries of  $A$ .

*Proof.* Let  $A$  be an  $n \times n$  upper triangular matrix. Then we have

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}, A - \lambda I_n = \begin{bmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} - \lambda & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} - \lambda \end{bmatrix}$$

Hence we have the following statements equivalent:

- $\lambda$  is an eigenvalue of  $A$ .
- The system  $(A - \lambda I_n)\mathbf{x} = \mathbf{0}$  has a nontrivial solution.
- The system  $(A - \lambda I_n)\mathbf{x} = \mathbf{0}$  has a free variable.
- At least one of the entries on the diagonal of  $A - \lambda I_n$  is zero.
- $\lambda = a_{ii}$  for some  $i$ .

Remark: For a lower triangular matrix, the proof is similar and we have the same conclusion. 0 can also be an eigenvalue.

**Proposition 5.2.** If  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$  are eigenvectors of a matrix  $A$  corresponding to distinct eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_k$ , then the set  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$  is linearly independent.

*Proof.* [This proof is given by *Github Copilot* to make the notes more elegant, but the correctness has not yet been verified...] Without loss of generality, we can assume that  $\lambda_1 \neq 0$ . Suppose that there exist scalars  $c_1, c_2, \dots, c_k$  such that

$$c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \dots + c_k \mathbf{v}_k = \mathbf{0},$$

and at least one of the  $c_i$  is nonzero. Then we have

$$A(c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \dots + c_k \mathbf{v}_k) = c_1 A\mathbf{v}_1 + c_2 A\mathbf{v}_2 + \dots + c_k A\mathbf{v}_k = \mathbf{0}$$

Since  $A\mathbf{v}_i = \lambda_i \mathbf{v}_i$ , we have

$$c_1 \lambda_1 \mathbf{v}_1 + c_2 \lambda_2 \mathbf{v}_2 + \dots + c_k \lambda_k \mathbf{v}_k = \mathbf{0}$$

Multiplying both sides by  $\mathbf{v}_1^T$ , we get

$$c_1 \lambda_1 \mathbf{v}_1 \mathbf{v}_1^T + c_2 \lambda_2 \mathbf{v}_2 \mathbf{v}_1^T + \dots + c_k \lambda_k \mathbf{v}_k \mathbf{v}_1^T = 0$$

Since  $\mathbf{v}_i^T \mathbf{v}_j = 0$  for  $i \neq j$ , we have  $c_1 \lambda_1 \mathbf{v}_1^T \mathbf{v}_1 = 0$ . Since  $\mathbf{v}_1 \neq \mathbf{0}$ , we have  $\mathbf{v}_1^T \mathbf{v}_1 \neq 0$ . Hence  $c_1 \lambda_1 = 0$ . Since  $\lambda_1 \neq 0$ , we have  $c_1 = 0$ . Similarly, we can show that  $c_2 = c_3 = \dots = c_k = 0$ . Hence the set  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$  is linearly independent.

**Example 5.2.** Suppose that  $\mathbf{b}_1, \mathbf{b}_2$  are eigenvectors corresponding to distinct eigenvalues  $\lambda_1, \lambda_2$  of a matrix  $A$ . And suppose that  $\mathbf{b}_3, \mathbf{b}_4$  are linearly independent vectors corresponding to a third distinct eigenvalue  $\lambda_3$ . Show that the set  $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \mathbf{b}_4\}$  is linearly independent.

Remark: Note that under a same eigenvalue, the eigenvectors are actually the solutions to a homogeneous system of linear equations, which can be independent or dependent. Remark: Think more about the example above!

## 5.2 Characteristic Equation

In the previous section, we have already tried to get the eigenvalues and eigenvectors of a matrix. Recall the ways we used to find the eigenvalues and eigenvectors, we can generalize the idea behind it as follows:

For any  $n \times n$  matrix  $A$ , according to the definition of eigenvalues and eigenvectors, we have

$$A\mathbf{x} = \lambda\mathbf{x} \Leftrightarrow (A - \lambda I_n)\mathbf{x} = \mathbf{0}.$$

This equation has a nontrivial solution if and only if the matrix  $A - \lambda I_n$  is singular, i.e.,  $\det(A - \lambda I_n) = 0$ . Hence, all possible eigenvalues of  $A$  are the roots of the equation  $\det(A - \lambda I_n) = 0$ , which is called the **characteristic equation** of  $A$ , and the eigenvectors are the solutions to the system  $(A - \lambda I_n)\mathbf{x} = \mathbf{0}$  with corresponding  $\lambda$ . Furthermore, we summarize this process to get a more systematic method by giving the following definition and proposition.

**Definition 5.2.** Let  $A$  be an  $n \times n$  matrix, then the equation

$$\det(A - \lambda I_n) = 0$$

is called the **characteristic equation** of  $A$ .

The polynomial

$$p_A(\lambda) = \det(A - \lambda I_n)$$

is called the **characteristic polynomial** of  $A$ , a polynomial of degree  $n$  in variable  $\lambda$ .

**Proposition 5.3.**  $\lambda$  is an eigenvalue of  $A$  if and only if  $\lambda$  is a root of the characteristic equation  $\det(A - \lambda I_n) = 0$ .

This proposition has connected the concept of eigenvalues with the determinant of a matrix, and consequently, the invertibility of a matrix. Therefore, we give some new insights into the invertibility of a matrix.

$A$  : a  $n \times n$  matrix. Then  $A$  is invertible if and only if

- The number 0 is not an eigenvalue of  $A$ .
- The determinant  $\det(A) \neq 0$ , and

$$\det A = (-1)^r \det u = \begin{cases} (-1)^r \cdot (\text{product of pivots of } u) & u \text{ invertible;} \\ 0 & u \text{ singular,} \end{cases}$$

where  $u$  is an echelon form of  $A$  by row replacements and row interchanges (without scaling) with  $r$  row interchanges.

**Example 5.3.** The characteristic polynomial of a  $6 \times 6$  matrix  $A$  is given by

$$\lambda^6 - 4\lambda^5 - 12\lambda^4$$

Find the eigenvalues and their multiplicities.

**Solution.**

- $\lambda = 0$  multiplicity 4.
- $\lambda = 6$  multiplicity 1.
- $\lambda = -2$  multiplicity 1.

Remark: The purpose of putting this example here is to show the concept of *multiplicity* of an eigenvalue, which is the number of times the eigenvalue appears as a root of the characteristic polynomial.

**Example 5.4.** Find the eigenvalues of the matrix

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

**Solution.** -1, 4, -3.



### 5.3 Diagonalization

**Example 5.5.** Let  $A$  be the matrix  $A = \begin{bmatrix} 7 & -2 \\ -4 & 1 \end{bmatrix}$ . Find a formula for  $A^n$  for any positive integer  $n$ , given that  $A = PDP^{-1}$ , where

$$P = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} \text{ and } D = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}$$

**Solution.**

Since  $A = PDP^{-1}$ , we have  $A^2 = PD(P^{-1}P)DP^{-1} = PD^2P^{-1}$ ,  $A^3 = PD^3P^{-1}$ , and so on. In general we have

$$A^n = PD^nP^{-1},$$

which can be proved by induction. Hence we have

$$A^n = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} \begin{bmatrix} 5^n & 0 \\ 0 & 3^n \end{bmatrix} \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} 2 \cdot 5^n - 3^n & 5^n - 3^n \\ -2 \cdot 5^n + 2 \cdot 3^n & -5^n + 2 \cdot 3^n \end{bmatrix}$$

**Definition 5.3.** A square matrix  $A$  is said to be **diagonalizable** if  $A$  is similar to a diagonal matrix. That is, if  $A = PDP^{-1}$  for some invertible matrix  $P$  and some diagonal matrix  $D$ .

**Proposition 5.4.** An  $n \times n$  matrix  $A$  is diagonalizable if and only if  $A$  has  $n$  linearly independent eigenvectors.

$A = PDP^{-1}$  with  $D$  a diagonal matrix if and only if the columns of  $P$  are  $n$  linearly independent eigenvectors of  $A$ . In this case, the diagonal entries of  $D$  are the eigenvalues of  $A$  that correspond, respectively, to the eigenvectors in  $P$ .

*Proof.* If  $A$  is diagonalizable, then  $A = PDP^{-1}$ . Denote  $P = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$ , where  $\mathbf{v}_i$  are column vectors. Then  $AP = A(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) = (A\mathbf{v}_1, A\mathbf{v}_2, \dots, A\mathbf{v}_n)$ . Since  $A = PDP^{-1}$ , we have

$$AP = PD = (A\mathbf{v}_1, A\mathbf{v}_2, \dots, A\mathbf{v}_n)(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} = (\lambda_1 \mathbf{v}_1, \lambda_2 \mathbf{v}_2, \dots, \lambda_n \mathbf{v}_n).$$

Hence we have  $A\mathbf{v}_i = \lambda_i \mathbf{v}_i$ , which means that  $\mathbf{v}_i$  is an eigenvector of  $A$  corresponding to  $\lambda_i$ . Since  $P$  is invertible, the columns of  $P$  are linearly independent, completing the proof that

If  $A$  is diagonalizable,  $A$  has  $n$  linearly independent eigenvectors.

Conversely, if  $A$  has  $n$  linearly independent eigenvectors such that  $A\mathbf{v}_i = \lambda_i \mathbf{v}_i$  for any  $1 \leq i \leq n$  with  $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$  linearly independent, then we can construct the matrix

$$P = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$$

which is invertible, and the matrix

$$D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

Then we have  $AP = PD$ , which indicates  $A = PDP^{-1}$ , completing the proof that

If  $A$  has  $n$  linearly independent eigenvectors, then  $A$  is diagonalizable.

**Example 5.6.** Diagonalizing a matrix, if possible.

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

From the previous example, we can see that the matrix is not diagonalizable, since the matrix has only two linearly independent eigenvectors, which is less than the dimension of the matrix. From this observation, we may find that whether a matrix is diagonalizable is related to the number of linearly independent eigenvectors, which is also related to *the sum of multiplicity of the eigenvalues*. **[Remark]** This is worth further exploration, although it is not covered in the handwritten notes. Specifically, we have the following proposition.

**Proposition 5.5.** An  $n \times n$  matrix  $A$  with  $n$  distinct eigenvalues is diagonalizable.

*Proof.* If  $A$  has  $n$  distinct eigenvalues, then  $A$  has at least  $n$  linearly independent eigenvectors. According to the previous proposition,  $A$  is diagonalizable.

In a more general case, let  $A$  be an  $n \times n$  matrix whose distinct eigenvalues are  $\lambda_1, \lambda_2, \dots, \lambda_p$  with multiplicities  $m_1, m_2, \dots, m_p$ . Then we have the following proposition whose title in the handwritten notes is **Matrices Whose Eigenvalues Are Not Distinct**.

- (a) For  $1 \leq k \leq p$ , the dimension of the eigenspace for  $\lambda_k$  is less than or equal to the multiplicity of the eigenvalue  $\lambda_k$ .
- (b) The matrix  $A$  is diagonalizable if and only if the sum of the dimensions of the eigenspaces is equal to  $n$ , which happens if and only if
  - the characteristic polynomial factors completely into linear factors, and
  - the dimension of the eigenspace for each  $\lambda_k$  is equal to the multiplicity of the eigenvalue.
- (c) If  $A$  is diagonalizable and  $B_k$  is a basis for the eigenspace for  $\lambda_k$ , then the total collection of vectors in the sets  $B_1, B_2, \dots, B_k$  forms an eigenvector basis for  $\mathbb{R}^n$ .

Finally, in the end of this section, we want to clarify the two systems of methods we have learnt: **(Please further clarify and add details when reviewing this note)**

- **Method 1:** Find the eigenvalues and eigenvectors of the matrix  $A$ .
- **Method 2:** Find the matrix  $P$  and the diagonal matrix  $D$  such that  $A = PDP^{-1}$ .

The handwritten notes provides some reasoning exercises for the two methods, i.e., the following examples.

**Example 5.7.** 1. Show that if  $A$  is both diagonalizable and invertible, then  $A^{-1}$  is also diagonalizable.

2. Show that if  $A$  has  $n$  distinct eigenvalues, then so does  $A^T$ .

3. Find the range of  $a$  where the matrix

$$A = \begin{bmatrix} 1 & a \\ 1 & 1 \end{bmatrix}$$

has no real eigenvalues. (Ans:  $a < 0$ )

**Example 5.8. Discrete Dynamical Systems.** We provides a "real-life" problem regarding the growth of owl and wood rad populations in a forest, denoted by  $O_k$  and  $R_k$  respectively, where  $k$  is the time in months. Suppose that

$$\mathbf{x}_k = \begin{bmatrix} O_k \\ R_k \end{bmatrix} = \begin{bmatrix} 0.5Q_{k-1} + 0.4R_{k-1} \\ -0.104Q_{k-1} + 1.1R_{k-1} \end{bmatrix}.$$

Determine the evolution of the system  $\mathbf{x}_k$ .

**Solution.**

Let  $A = \begin{bmatrix} 0.5 & 0.4 \\ -0.104 & 1.1 \end{bmatrix}$ . Then we have  $\mathbf{x}_{k+1} = A\mathbf{x}_k$ .

It's easy to find that the eigenvalues for  $A$  are  $\lambda_1 = 1.02$ ,  $\lambda_2 = 0.58$ . The corresponding eigenvectors are  $\mathbf{v}_1 = \begin{bmatrix} 10 \\ 13 \end{bmatrix}$  and  $\mathbf{v}_2 = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$ .

Obviously,  $\mathbf{v}_1$  and  $\mathbf{v}_2$  form a basis in  $\mathbb{R}^2$ , and  $A$  is diagonalizable. For any initial  $x_0$ , there exists  $c_1, c_2$  such that

$$x_0 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$$

Hence we have

$$x_k = A^k x_0 = A^k (c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = c_1\lambda_1^k \mathbf{v}_1 + c_2\lambda_2^k \mathbf{v}_2$$

After substituting the values of  $\lambda_1, \lambda_2$  and  $\mathbf{v}_1, \mathbf{v}_2$ , for very large  $k$ , we have

$$\mathbf{x}_{k+1} \approx 1.02\mathbf{x}_k,$$

which indicates that eventually both entries of  $\mathbf{x}_k$  will increase by 2% each month.

## 5.4 Applications to Differential Equations

### System of Linear Differential Equations

$$\begin{cases} x'_1(t) = a_{11}x_1(t) + a_{12}x_2(t) + \cdots + a_{1n}x_n(t) \\ x'_2(t) = a_{21}x_1(t) + a_{22}x_2(t) + \cdots + a_{2n}x_n(t) \\ \vdots \\ x'_n(t) = a_{n1}x_1(t) + a_{n2}x_2(t) + \cdots + a_{nn}x_n(t) \end{cases}$$

Let

$$\mathbf{x}(\mathbf{t}) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix}, A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

Then  $\mathbf{x}'(\mathbf{t}) = A\mathbf{x}(\mathbf{t})$ .

### When $A$ is a diagonal matrix

Since

$$\begin{bmatrix} x'_1(t) \\ x'_2(t) \\ \vdots \\ x'_n(t) \end{bmatrix} = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix},$$

it can be decoupled into  $n$  independent equations:

$$\begin{cases} x_1'(t) = a_{11}x_1(t) \\ x_2'(t) = a_{22}x_2(t) \\ \vdots \\ x_n'(t) = a_{nn}x_n(t) \end{cases}$$

The solution to each equation is of the form  $x_i(t) = c_i e^{a_{ii}t}$ , where  $c_i$  is a constant. Hence the general solution to the system is

$$\mathbf{x}(t) = c_1 e^{a_{11}t} \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + c_2 e^{a_{22}t} \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \cdots + c_n e^{a_{nn}t} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

**For general equation**  $x'(t) = Ax(t)$

The solution is

$$\mathbf{x}(t) = c_1 e^{\lambda_1 t} \mathbf{v}_1 + c_2 e^{\lambda_2 t} \mathbf{v}_2 + \cdots + c_n e^{\lambda_n t} \mathbf{v}_n$$

where  $\lambda_1, \lambda_2, \dots, \lambda_n$  are the eigenvalues of  $A$  and  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  are the corresponding eigenvectors. Here  $\mathbf{v}_i e^{\lambda_i t}$  are called eigenfunctions of the differential equation system.

**Example 5.9.**

$$\mathbf{x}'(t) = \begin{bmatrix} -1.5 & 0.5 \\ 1 & -1 \end{bmatrix} \mathbf{x}(t), \quad \mathbf{x}(0) = \begin{bmatrix} 5 \\ 4 \end{bmatrix}$$

**Solution.** The eigenvalues are  $\lambda_1 = -0.5, \lambda_2 = -2$ , and the corresponding eigenvectors are  $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$  and  $\mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ . The eigenfunctions  $x_1(t) = \mathbf{v}_1 e^{\lambda_1 t}$  and  $x_2(t) = \mathbf{v}_2 e^{\lambda_2 t}$  both satisfy  $\mathbf{x}'(t) = A\mathbf{x}(t)$ . Hence we write

$$\mathbf{x}(t) = c_1 e^{-0.5t} \begin{bmatrix} 1 \\ 2 \end{bmatrix} + c_2 e^{-2t} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Let  $t = 0$ , we have  $\mathbf{x}(0) = c_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + c_2 \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 4 \end{bmatrix}$ . Solving this system of equations, we get  $c_1 = 3, c_2 = -2$ . Hence the solution is

$$\mathbf{x}(t) = 3e^{-0.5t} \begin{bmatrix} 1 \\ 2 \end{bmatrix} - 2e^{-2t} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

*Proof.* [This proof is given by *Github Copilot*, and a proof may also be found in handwritten lecture 20] To prove the general solution to the system of linear differential equations  $\mathbf{x}'(t) = A\mathbf{x}(t)$ , we can use the method of decoupling the system by diagonalizing the matrix  $A$ .

Assume that  $A$  is diagonalizable, i.e., there exists an invertible matrix  $P$  and a diagonal matrix  $D$  such that  $A = PDP^{-1}$ . Let  $\mathbf{y}(t) = P^{-1}\mathbf{x}(t)$ . Then we have:

$$\mathbf{x}(t) = P\mathbf{y}(t)$$

Differentiating both sides with respect to  $t$ , we get:

$$\mathbf{x}'(t) = P\mathbf{y}'(t)$$

Since  $\mathbf{x}'(t) = A\mathbf{x}(t)$ , we have:

$$P\mathbf{y}'(t) = AP\mathbf{y}(t)$$

Substituting  $A = PDP^{-1}$ , we get:

$$P\mathbf{y}'(\mathbf{t}) = PDP^{-1}P\mathbf{y}(\mathbf{t}) = PD\mathbf{y}(\mathbf{t})$$

Multiplying both sides by  $P^{-1}$ , we obtain:

$$\mathbf{y}'(\mathbf{t}) = D\mathbf{y}(\mathbf{t})$$

Since  $D$  is a diagonal matrix, the system  $\mathbf{y}'(\mathbf{t}) = D\mathbf{y}(\mathbf{t})$  is decoupled into  $n$  independent differential equations:

$$\begin{cases} y_1'(t) = \lambda_1 y_1(t) \\ y_2'(t) = \lambda_2 y_2(t) \\ \vdots \\ y_n'(t) = \lambda_n y_n(t) \end{cases}$$

where  $\lambda_1, \lambda_2, \dots, \lambda_n$  are the diagonal entries of  $D$ , which are the eigenvalues of  $A$ . The solution to each equation is of the form:

$$y_i(t) = c_i e^{\lambda_i t}$$

where  $c_i$  is a constant. Hence, the general solution to the system  $\mathbf{y}'(\mathbf{t}) = D\mathbf{y}(\mathbf{t})$  is:

$$\mathbf{y}(\mathbf{t}) = \begin{bmatrix} c_1 e^{\lambda_1 t} \\ c_2 e^{\lambda_2 t} \\ \vdots \\ c_n e^{\lambda_n t} \end{bmatrix}$$

Finally, transforming back to the original variables, we get the general solution to the system  $\mathbf{x}'(\mathbf{t}) = A\mathbf{x}(\mathbf{t})$ :

$$\mathbf{x}(\mathbf{t}) = P\mathbf{y}(\mathbf{t}) = P \begin{bmatrix} c_1 e^{\lambda_1 t} \\ c_2 e^{\lambda_2 t} \\ \vdots \\ c_n e^{\lambda_n t} \end{bmatrix} = c_1 e^{\lambda_1 t} \mathbf{v}_1 + c_2 e^{\lambda_2 t} \mathbf{v}_2 + \dots + c_n e^{\lambda_n t} \mathbf{v}_n$$

where  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  are the columns of  $P$ , which are the eigenvectors of  $A$ .

## 6 Inner Product Spaces

### 6.1 Inner Product, Length, and Orthogonality

In  $\mathbb{R}^2, \mathbb{R}^3$ , we have already learnt the concepts of dot product, length/distance, perpendicularity, etc. In this section, we will generalize these concepts to  $\mathbb{R}^n$  and more general vector spaces.

- **Inner Product:** Let  $\mathbf{u} = (u_1, u_2, \dots, u_n)$  and  $\mathbf{v} = (v_1, v_2, \dots, v_n)$  be vectors in  $\mathbb{R}^n$ . The **inner product** of  $\mathbf{u}$  and  $\mathbf{v}$  is defined as

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \dots + u_n v_n.$$

- **Length:** The **length** (or **norm**) of a vector  $\mathbf{v}$  in  $\mathbb{R}^n$  is defined as

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{v_1^2 + v_2^2 + \dots + v_n^2}.$$

- **Distance:** The **distance** between two vectors  $\mathbf{u}$  and  $\mathbf{v}$  in  $\mathbb{R}^n$  is defined as

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + \dots + (u_n - v_n)^2}.$$

- **Orthogonality:** Two vectors  $\mathbf{u}$  and  $\mathbf{v}$  in  $\mathbb{R}^n$  are **orthogonal** if  $\mathbf{u} \cdot \mathbf{v} = 0$ .

#### Proposition 6.1. Pythagorean Theorem

Two vectors  $\mathbf{u}$  and  $\mathbf{v}$  in  $\mathbb{R}^n$  are orthogonal if and only if

$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2.$$

### Orthogonal Complements

Let  $\mathbf{z} \in \mathbb{R}^n$  and  $W$  is a subspace of  $\mathbb{R}^n$ . If  $\mathbf{z}$  is orthogonal to every vector in  $W$ , we say that  $\mathbf{z}$  is orthogonal to  $W$ , and write  $\mathbf{z} \perp W$ . The set of all vectors in  $\mathbb{R}^n$  that are orthogonal to  $W$  is called the **orthogonal complement** of  $W$ , denoted by  $W^\perp$ . By definition, we can easily see that  $W^\perp$  is also a subspace of  $\mathbb{R}^n$ . Furthermore, we will look at some facts:

- Let  $A$  be an  $m \times n$  matrix. The orthogonal complement of the row space of  $A$  is the null space of  $A$ , and the orthogonal complement of  $A$  is the null space of  $A^T$ , that is

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

- Let  $W$  be a subspace of  $\mathbb{R}^n$ . Then  $(W^\perp)^\perp$  is also a subspace of  $\mathbb{R}^n$ , and

$$\dim W + \dim W^\perp = n.$$

### 6.2 Orthogonal Sets

In applications in different areas, like the Cartesian coordinate system, we use a set of orthogonal vectors to represent the space, and use them as the basis to represent other instances like vectors, surfaces, etc. Here, we summarize such set of vectors as orthogonal sets and show some properties of them.

**Definition 6.1.** A set of vectors  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$  in  $\mathbb{R}^n$  is said to be an **orthogonal set** if each pair of distinct vectors in the set is orthogonal, that is,  $\mathbf{v}_i \cdot \mathbf{v}_j = 0$  for all  $i \neq j$ .

**Proposition 6.2.** If  $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$  is an orthogonal set of nonzero vectors in  $\mathbb{R}^n$ , then the set  $S$  is linearly independent and hence is a basis for the subspace  $V$  of  $\mathbb{R}^n$  spanned by  $S$ .

*Proof.* Suppose that  $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$ , then we have

$$\begin{aligned}\mathbf{0} &= \mathbf{v}_1 \cdot (c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n) \\ &= c_1\mathbf{v}_1 \cdot \mathbf{v}_1 + c_2\mathbf{v}_1 \cdot \mathbf{v}_2 + \cdots + c_n\mathbf{v}_1 \cdot \mathbf{v}_n \\ &= c_1\|\mathbf{v}_1\|^2,\end{aligned}$$

which implies that  $c_1 = 0$ . Similarly, we can show that  $c_2 = c_3 = \cdots = c_n = 0$ . Thus, by contradiction, we can show that the set  $S$  is linearly independent.  $\square$

**Definition 6.2.** An **orthogonal basis** for a subspace  $V$  of  $\mathbb{R}^n$  is a basis for  $V$  that is also an orthogonal set.

**Proposition 6.3.** If  $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$  is an orthogonal basis for a subspace  $V$  of  $\mathbb{R}^n$ . For each  $\mathbf{y} \in V$ , the weights in the linear combination

$$\mathbf{y} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$$

are given by

$$c_i = \frac{\mathbf{y} \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i}, \quad i = 1, 2, \dots, n.$$

The proof is trivial by multiplying  $\mathbf{v}_i$  on both sides of the equation above, but always keep in mind that if the basis were not orthogonal, it would be necessary to solve a system of linear equations in order to find the weights.

In the 2D Cartesian coordinate system we have learnt before, we covered the concept of projection vectors. Here, we discuss the similar idea in the context of orthogonal sets and expand it to  $\mathbb{R}^n$ .

**Definition 6.3.** Let  $\mathbf{u} \in \mathbb{R}^n$  and  $\mathbf{y} \in \mathbb{R}^n$ , and denote

$$L \equiv \text{span}\{\mathbf{u}\}, \text{ the line parallel to } \mathbf{u}.$$

Write  $\mathbf{y}$  as

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{w}$$

such that  $\hat{\mathbf{y}} = c\mathbf{u}$  (parallel to  $\mathbf{u}$ ) and  $\mathbf{w}$  is orthogonal to  $\mathbf{u}$ . Then  $\hat{\mathbf{y}}$  is called the **orthogonal projection** of  $\mathbf{y}$  onto  $\mathbf{u}$ , denoted by  $\text{proj}_{\mathbf{u}}\mathbf{y}$  or  $\text{proj}_L\mathbf{y}$ . and  $\mathbf{w}$  is called the component of  $\mathbf{y}$  orthogonal to  $\mathbf{u}$ .

Just like the projection vector in 2D, we can calculate the orthogonal projection by the following formula:

$$\hat{\mathbf{y}} = \text{proj}_{\mathbf{u}}\mathbf{y} = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\mathbf{u}, \quad \mathbf{w} = \mathbf{y} - \hat{\mathbf{y}}.$$

Recall that in the Cartesian coordinate system, we require the basis to be not only orthogonal, but also normalized. In the context of  $\mathbb{R}^n$ , we have the word of "orthonormality" to describe such basis.

**Definition 6.4.** A set of vectors  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$  in  $\mathbb{R}^n$  is said to be an **orthonormal set** if it is an orthogonal set and each vector in the set has unit length, that is,  $\|\mathbf{v}_i\| = 1$  for all  $i$ . If  $W$  is a subspace spanned by an orthonormal set, then the set is called an **orthonormal basis** for  $W$ .

**Proposition 6.4.** An  $n \times n$  matrix  $U$  has orthonormal columns if and only if  $U^T U = I$ .

The proof of this proposition follows directly from how we calculate the product of two matrices, and the readers can try to prove it by themselves. This reveals the relationship between orthonormal sets and square matrices, and next, we will discuss some properties of general  $m \times n$  matrices.

**Proposition 6.5.** Let  $U$  be an  $m \times n$  matrix with orthonormal columns, and let  $\mathbf{x}$  and  $\mathbf{y}$  be vectors in  $\mathbb{R}^n$ . Then

- $\|U\mathbf{x}\| = \|\mathbf{x}\|$
- $(U\mathbf{x}) \cdot (U\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$
- $(U\mathbf{x}) \cdot (U\mathbf{y}) = 0$  if and only if  $\mathbf{x} \cdot \mathbf{y} = 0$

Remark: Properties a and c say that the linear mapping  $\mathbf{x} \mapsto U\mathbf{x}$  preserves lengths and orthogonality.

To bring this section to an end, we introduce the following lemma of the propositions above, whose proof is left as an exercise to the readers.

Let  $U$  be an  $n \times n$  matrix with orthonormal columns, show that  $\det U = \pm 1$ .

### 6.3 Orthogonal Projections

In the previous section, we have introduced the concept of orthogonal projection onto a vector, as well as the orthonormal set. Now let's combine them together and introduce the orthogonal projection onto a subspace.

**Example 6.1.** Let  $W$  to be the  $xy$ -plane, and  $\mathbf{u}_1 \in W, \mathbf{u}_2 \in W^\perp$ , then we can decompose the vector  $\mathbf{u} = (a, b, c)$  into two parts:

$$\mathbf{u} = (a, b, c) = (a, b, 0) + (0, 0, c) = \mathbf{u}_1 + \mathbf{u}_2$$

In general, we want to find a decomposition of a vector  $\mathbf{y}$  in  $\mathbb{R}^n$  into two parts, one in the subspace  $W$  and the other in the orthogonal complement  $W^\perp$ .

**Proposition 6.6. Orthogonal Decomposition Theorem**

Let  $W$  be a subspace of  $\mathbb{R}^n$ , then each  $\mathbf{y} \in \mathbb{R}^n$  can be uniquely written as

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$$

where  $\hat{\mathbf{y}} \in W$  and  $\mathbf{z} \in W^\perp$ . In fact, if  $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$  is any orthogonal basis for  $W$ , then

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 + \dots + \frac{\mathbf{y} \cdot \mathbf{u}_p}{\mathbf{u}_p \cdot \mathbf{u}_p} \mathbf{u}_p$$

and  $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$ .

Here,  $\hat{\mathbf{y}}$  is called the **orthogonal projection** of  $\mathbf{y}$  onto  $W$ , denoted by  $\text{proj}_W \mathbf{y}$ .

We may interpret the orthogonal projection onto an orthogonal basis as a generalization of the orthogonal projection onto a vector. Hence, the proof of this proposition is similar to the latter. Since the correctness of a decomposition can be easily verified, by checking whether  $\mathbf{z}$  is orthogonal to each  $\mathbf{u}_i$ , we will not provide examples here. Readers can randomly choose a vector and a subspace as exercises and verify the correctness by themselves.

*Proof.* The proof of this proposition can be split into two parts, i.e., the existence and uniqueness, and here is an outline of the proof:

- **Existence:** First we observe that  $\hat{\mathbf{y}}$  is in  $W$ , since it's a linear combination of the basis vectors, and next we only need to show that  $\mathbf{z}$  is orthogonal to  $W$ . This can be done by calculating the dot product of  $\mathbf{z}$  and  $\mathbf{u}_i$ , and we can show that  $\mathbf{z}$  is orthogonal to each  $\mathbf{u}_i$ .
- **Uniqueness:** Suppose that there are two, then  $\hat{\mathbf{y}} = \hat{\mathbf{y}}_1 + \hat{\mathbf{z}}_1 = \hat{\mathbf{y}}_2 + \hat{\mathbf{z}}_2$ , and

$$(\hat{\mathbf{y}}_1 - \hat{\mathbf{y}}_2) = -(\hat{\mathbf{z}}_1 - \hat{\mathbf{z}}_2)$$



Multiplying both sides by  $\hat{\mathbf{y}}_1 - \hat{\mathbf{y}}_2$  and using the fact that  $\hat{\mathbf{y}}_1 - \hat{\mathbf{y}}_2 \perp \hat{\mathbf{z}}_1 - \hat{\mathbf{z}}_2$ , we can show that  $\|\hat{\mathbf{y}}_1 - \hat{\mathbf{y}}_2\|$  and therefore the result.

From these concepts, we can easily get the following lemma:

Let  $W = \text{span}\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p\}$  be an orthogonal basis for  $W$ . If  $\mathbf{y} \in W$ , then  $\text{proj}_W \mathbf{y} = \mathbf{y}$ .

**Proposition 6.7. The Best Approximation Theorem**

Let  $W$  be a subspace of  $\mathbb{R}^n$ . Let  $\mathbf{y}$  be any vector in  $\mathbb{R}^n$ , and let  $\hat{\mathbf{y}}$  be the orthogonal projection of  $\mathbf{y}$  onto  $W$ . Then  $\hat{\mathbf{y}}$  is the closest point in  $W$  to  $\mathbf{y}$ , that is, for any  $\mathbf{w} \in W$ ,

$$\|\mathbf{y} - \hat{\mathbf{y}}\| \leq \|\mathbf{y} - \mathbf{w}\|$$

Remark: The vector  $\hat{\mathbf{y}}$  is called **the best approximation** to  $\mathbf{y}$  from  $W$ .

*Proof.* We write  $\mathbf{y} - \mathbf{w} = (\mathbf{y} - \hat{\mathbf{y}}) + (\hat{\mathbf{y}} - \mathbf{w})$ , and the result follows from the Pythagorean theorem. Following this proposition, we are safe to define the distance between a vector and a subspace.

**Definition 6.5.** The **distance** from a vector  $\mathbf{y}$  to a subspace  $W$  is defined as the distance from  $\mathbf{y}$  to the closest point in  $W$ , that is,

$$\text{dist}(\mathbf{y}, W) = \|\mathbf{y} - \hat{\mathbf{y}}\|$$

where  $\hat{\mathbf{y}}$  is the orthogonal projection of  $\mathbf{y}$  onto  $W$ .

**Example 6.2.** Find the distance from  $\mathbf{y}$  to  $W = \text{span}\{\mathbf{u}_1, \mathbf{u}_2\}$ , where

$$\mathbf{y} = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix}, \quad \mathbf{u}_1 = \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}.$$

**Solution.**  $3\sqrt{5}$

Lastly, from the Orthogonal Decomposition Theorem, for the case when the set is an orthonormal basis, we can easily get the following lemma: (hint: when proving the second one try to write  $UU^T \mathbf{y} = U(U^T \mathbf{y})$ )

Let  $W$  be a subspace of  $\mathbb{R}^n$  spanned by an orthonormal basis  $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p\}$ . If  $\mathbf{y} \in \mathbb{R}^n$ , then

$$\text{proj}_W \mathbf{y} = \mathbf{y} \cdot \mathbf{u}_1 \mathbf{u}_1 + \dots + \mathbf{y} \cdot \mathbf{u}_p \mathbf{u}_p.$$

If  $U = [\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_p]$ , then

$$\text{proj}_W \mathbf{y} = UU^T \mathbf{y} \text{ for all } \mathbf{y} \in \mathbb{R}^n$$

Also, we may show the following lemma to be true, which involves the linear combination:

Let  $W$  be a subspace of  $\mathbb{R}^n$ , and  $\mathbf{x}, \mathbf{y}$  are vectors in  $\mathbb{R}^n$ . If  $\mathbf{z} = \mathbf{x} + \mathbf{y}$ , then

$$\text{proj}_W \mathbf{z} = \text{proj}_W \mathbf{x} + \text{proj}_W \mathbf{y}.$$

## 6.4 Gram-Schmidt Process

After having the concepts of orthogonal and orthonormal sets, the thing left is to find such sets. Although we may be provided with a basis of a subspace, it may not be orthogonal or orthonormal, which limits its applications. For example, try the following exercise:

**Example 6.3.** Let  $W$  be the subspace of  $\mathbb{R}^3$  spanned by the vectors

$$\mathbf{v}_1 = \begin{pmatrix} 3 \\ 6 \\ 0 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}.$$

Find an orthogonal basis for  $W$ .

In this example, it is quite easy to "guess out" an answer, but in general, we need a systematic way to find the orthogonal basis. The Gram-Schmidt process is such a simple algorithm to find an orthogonal basis for a nonzero subspace  $W$  of  $\mathbb{R}^n$  given a basis for  $W$ .

**Remark:** There should have been some reasoning behind the Gram-Schmidt process, which can be "felt" but hard to write in texts.

**Proposition 6.8. The Gram-Schmidt Process**

Given a basis  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p\}$  for a nonzero subspace  $W$  of  $\mathbb{R}^n$ , define

$$\mathbf{v}_i = \begin{cases} \mathbf{x}_1 & \text{if } i = 1, \\ \mathbf{x}_i - \sum_{j=1}^{i-1} \text{proj}_{\mathbf{v}_j} \mathbf{x}_i & \text{if } i > 1. \end{cases}$$

Then  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}$  is an orthogonal basis for  $W$ , and

$$\text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\} = \text{Span}\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p\}.$$

After getting an **orthogonal basis** through the algorithm above, we can easily construct a **orthonormal basis** by normalizing each vector in the basis.

**Proposition 6.9. The QR Factorization**

Let  $A$  be an  $m \times n$  matrix with linearly independent columns. Then there exists an  $m \times n$  matrix  $Q$  with orthonormal columns and an  $n \times n$  upper triangular matrix  $R$  such that  $A = QR$ , and The  $Q$ 's columns form an orthonormal basis for the column space of  $A$ , and  $R$  is an  $n \times n$  upper triangular invertible matrix with positive diagonal entries.

*Proof.* The columns of  $A$  form a basis for  $\text{Col}A$ . Construct an orthonormal basis  $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$  for  $\text{Col}A$  using the Gram-Schmidt process. Let  $Q = [\mathbf{u}_1, \dots, \mathbf{u}_n]$  and  $R$  be the  $n \times n$  matrix whose  $(i, j)$  entry is  $\mathbf{u}_i \cdot \mathbf{a}_j$ , where  $\mathbf{a}_j$  is the  $j$ th column of  $A$ . Then  $A = QR$ .  $\square$

Furthermore, from this proof, if columns of  $A$  are linearly dependent and  $A = QR$ , then  $R$  will have zero entries on its diagonal.

There are lots of questions related to QR factorization which involves terrible calculations. The following is a common algorithm to compute the QR factorization of a matrix.

1. Given a matrix  $A$ , let  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$  be the columns of  $A$ , and perform the Gram-Schmidt process to get an orthogonal basis  $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$  for  $\text{Col}A$ .
  2. Let  $Q = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n]$  and recall that  $Q^T Q = I$ , so  $R = Q^T Q R = Q^T A$ .
- The blank here is left for rough work -

## 6.5 Least Squares Problems

### Least Squares Solutions

In real application, sometimes  $A\mathbf{x} = \mathbf{b}$  has no solution, and we want to find  $\hat{\mathbf{x}}$  such that  $A\mathbf{x}$  is as close as possible to  $\mathbf{b}$ , which is called the **least squares problem**.

**Definition 6.6.** If  $A$  is  $m \times n$  and  $\mathbf{b}$  is in  $\mathbb{R}^m$ , then the **least squares solution** to the inconsistent system  $A\mathbf{x} = \mathbf{b}$  is the vector  $\hat{\mathbf{x}}$  in  $\mathbb{R}^n$  such that

$$\|A\hat{\mathbf{x}} - \mathbf{b}\| \leq \|A\mathbf{x} - \mathbf{b}\|$$

for all  $\mathbf{x}$  in  $\mathbb{R}^n$ .

#### Proposition 6.10. The Least Squares Solution

The set of least-squares solutions of  $A\mathbf{x} = \mathbf{b}$  coincides with the nonempty set of solutions of the equations

$$A^T A\mathbf{x} = A^T \mathbf{b},$$

which is called the **normal equation** for the system  $A\mathbf{x} = \mathbf{b}$ .

*Proof.* First we prove that such  $\hat{\mathbf{x}}$  exists. Denote  $\hat{\mathbf{b}} = \text{proj}_{\text{Col}A} \mathbf{b}$ , then since  $\hat{\mathbf{b}}$  is a combination of column vectors of  $A$ , the system  $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$  has a solution  $\hat{\mathbf{x}}$ , which satisfies

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| = \|\mathbf{b} - \hat{\mathbf{b}}\| \leq \|\mathbf{b} - A\mathbf{x}\|.$$

Next, we prove that the solution set of  $A^T A\mathbf{x} = A^T \mathbf{b}$  (1) is the same as the solution set of  $A\mathbf{x} = \hat{\mathbf{b}}$  (2). According to the definition of orthogonal projection,  $\mathbf{b} - \hat{\mathbf{b}} \perp \text{Col}A$ , which indicates that

$$A^T(\mathbf{b} - \hat{\mathbf{b}}) = \mathbf{0}.$$

By substituting  $\hat{\mathbf{b}} = A\hat{\mathbf{x}}$  into the equation above, we can get (1). Conversely, from (1) we can get  $\mathbf{0} = A^T(A\mathbf{x} - \mathbf{b})$ , which implies that  $(\mathbf{b} - A\mathbf{x}) \perp \text{Col}A$ , so  $\mathbf{b} - A\mathbf{x} \in (\text{Col}A)^\perp$ . Here by the uniqueness of orthogonal decomposition, we must have  $\mathbf{b} - A\mathbf{x} = \mathbf{b} - \hat{\mathbf{b}}$ , which directly leads to (2).  $\square$

From the reasoning arguments above, we can easily find that for an  $m \times n$  matrix  $A$  the following statements are logically equivalent:

- The equation  $A\mathbf{x} = \mathbf{b}$  has a unique least-squares solution for every  $\mathbf{b}$  in  $\mathbb{R}^m$ .
- The columns of  $A$  are linearly independent.
- The matrix  $A^T A$  is invertible.

### Alternative ways for finding the least squares solution

- (1) When the columns of  $A$  are orthogonal.

Since  $\text{Col}A$  is an orthogonal subspace, by definition we can write the orthogonal projection of  $\mathbf{b}$  onto  $\text{Col}A$  as

$$\hat{\mathbf{b}} = \text{proj}_{\text{Col}A} \mathbf{b} = \frac{\mathbf{b} \cdot \mathbf{a}_1}{\mathbf{a}_1 \cdot \mathbf{a}_1} \mathbf{a}_1 + \cdots + \frac{\mathbf{b} \cdot \mathbf{a}_n}{\mathbf{a}_n \cdot \mathbf{a}_n} \mathbf{a}_n = A \cdot (x_1, x_2, \dots, x_n)^T,$$

where  $x_i = \frac{\mathbf{b} \cdot \mathbf{a}_i}{\mathbf{a}_i \cdot \mathbf{a}_i}$ . Hence the least squares solution is  $\hat{\mathbf{x}} = (x_1, x_2, \dots, x_n)^T$ .

- (2) When columns of  $A$  are linearly independent.

Here the least-squares solution can often be computed more reliable through a QR factorization of  $A$ . Let  $A = QR$  be a QR factorization of  $A$ . Then, for each  $\mathbf{b}$  in  $\mathbb{R}^m$ , the least-squares solution of  $A\mathbf{x} = \mathbf{b}$  is given by

$$\hat{\mathbf{x}} = R^{-1}Q^T \mathbf{b}.$$

*Proof.* Let  $\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}$ , then

$$A\hat{\mathbf{x}} = QR\hat{\mathbf{x}} = QR R^{-1}Q^T\mathbf{b} = QQ^T\mathbf{b}.$$

Note that the columns of  $Q$  form an orthonormal basis for  $\text{Col}A$ , so  $QQ^T\mathbf{b}$  is the orthogonal projection of  $\mathbf{b}$  onto  $\text{Col}A$  - **Why?**. So  $\hat{\mathbf{x}}$  is the least-squares solution. The uniqueness follows from the fact that the columns of  $A$  are linearly independent.

These methods requires lots of exercise to get familiar with, and the readers are encouraged to try more exercises to get a better understanding of the least squares problem. For example, following is an example which involves different methods covered in these chapters. (Please calculate this!)

**Example 6.4.** Find the least square solution using QR factorization for the following system:

$$\begin{pmatrix} 1 & 3 & 5 \\ 1 & 1 & 0 \\ 1 & 1 & 2 \\ 1 & 3 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 3 \\ 5 \\ 7 \\ -3 \end{pmatrix}$$

- This blank is left for the rough work of the example above -

## 6.6 Applications

### Linear Regression (Least Squares lines)

In the context of linear regression, we are given a set of data points  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ , and we want to find the best-fitting line  $y = mx + b$  to the data, where we want to minimize

$$\sum_{i=1}^n (y_i - (mx_i + b))^2.$$

Given the description above, we can always write

$$y_i = mx_i + b + \epsilon_i,$$

where  $\epsilon_i$  is the error in the  $i$ th measurement. Hence, we can use the matrix form to represent the problem as

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} b \\ m \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}.$$

Now we want to minimize  $\|\mathbf{x}\beta - \mathbf{y}\|$ , where

$$\mathbf{x} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \beta = \begin{pmatrix} b \\ m \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix},$$

which is equivalent to find the least squares solution of the system  $\mathbf{x}\beta = \mathbf{y}$ .

### General Linear Model

Fit data by curves that have the form

$$y = \beta_0 f_0(x) + \beta_1 f_1(x) + \dots + \beta_n f_n(x),$$

where  $f_0(x), f_1(x), \dots, f_n(x)$  are known functions of  $x$ . The goal is to find the coefficients  $\beta_0, \beta_1, \dots, \beta_n$  that minimize the sum of the squares of the errors.

**Example 6.5.** Given the data  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  that comes from a company's total costs. The equation of the form

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$$

is an appropriate model for the data. Describe the linear model that gives a best fit to the data.

**Solution.**  $\mathbf{y} = \mathbf{x}\beta + \epsilon$ , where

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 \end{pmatrix}.$$

Now we can solve the least squares solution of  $\mathbf{x}\beta = \mathbf{y}$  to get  $\beta$ .

Note that when we have more than one independent variable, we can use the same method to find the best-fitting plane or hyperplane, such as

$$y = \beta_0 + \beta_1 u + \beta_2 v + \beta_3 w.$$