# LINEAR ALGEBRA

# AND

# APPLICATIONS

# LINEAR ALGEBRA

# AND

# APPLICATIONS

# MAT2041 Notebook


## Instructor

*Ruoyu Sun*

*The Chinese University of Hongkong, Shenzhen*

## Instructor

*Tongxin Li*

*The Chinese University of Hongkong, Shenzhen*


香港中文大學（深圳）

The Chinese University of Hong Kong, Shenzhen

# Contents

# Acknowledgments

This book is from the MAT2041 in the Fall semester, 2022. It is based on the lecture notes of MAT2040 and revised in 2022 to include more data science application, correct some mistakes, and revise some proofs to give readers better insights on linear algebra.

CUHK(SZ)

# Notations and Conventions

| | |
|---|---|
| $\mathbb{R}^n$ | $n$-dimensional real space |
| $\mathbb{C}^n$ | $n$-dimensional complex space |
| $\mathbb{R}^{m \times n}$ | set of all $m \times n$ real-valued matrices |
| $\mathbb{C}^{m \times n}$ | set of all $m \times n$ complex-valued matrices |
| $x_i$ | $i$th entry of column vector $\boldsymbol{x}$ |
| $a_{ij}$ | $(i, j)$th entry of matrix $\boldsymbol{A}$ |
| $\boldsymbol{a}_i$ | $i$th column of matrix $\boldsymbol{A}$ |
| $\boldsymbol{a}_i^{\mathrm{T}}$ | $i$th row of matrix $\boldsymbol{A}$ |
| $\mathbb{S}^n$ | set of all $n \times n$ real symmetric matrices, i.e., $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ and $a_{ij} = a_{ji}$ for all $i, j$ |
| $\mathbb{H}^n$ | set of all $n \times n$ complex Hermitian matrices, i.e., $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ and $\bar{a}_{ij} = a_{ji}$ for all $i, j$ |
| $\boldsymbol{A}^{\mathrm{T}}$ | transpose of $\boldsymbol{A}$, i.e, $\boldsymbol{B} = \boldsymbol{A}^{\mathrm{T}}$ means $b_{ji} = a_{ij}$ for all $i, j$ |
| $\boldsymbol{A}^{\mathrm{H}}$ | Hermitian transpose of $\boldsymbol{A}$, i.e, $\boldsymbol{B} = \boldsymbol{A}^{\mathrm{H}}$ means $b_{ji} = \bar{a}_{ij}$ for all $i, j$ |
| $\mathrm{trace}(\boldsymbol{A})$ | sum of diagonal entries of square matrix $\boldsymbol{A}$ |
| $\mathbf{1}$ | A vector with all 1 entries |
| $\mathbf{0}$ | either a vector of all zeros, or a matrix of all zeros |
| $\boldsymbol{e}_i$ | a unit vector with the nonzero element at the $i$th entry |
| $\mathcal{C}(\boldsymbol{A})$ | the column space of $\boldsymbol{A}$ |
| $\mathcal{R}(\boldsymbol{A})$ | the row space of $\boldsymbol{A}$ |
| $\mathcal{N}(\boldsymbol{A})$ | the null space of $\boldsymbol{A}$ |
| $\mathrm{Proj}_{\mathcal{M}}(\boldsymbol{A})$ | the projection of $\boldsymbol{A}$ onto the set $\mathcal{M}$ |

# Chapter 1

# Week1

## 1.1. Lecture 1: Linear Algebra in Data Science and Intro to Vectors

### 1.1.1. Introduction

#### 1.1.1.1. *Why do you learn Linear Algebra?*

So, we raise the question again, why do we learn LA?

- Baisis of AI/ML/SP/etc.
  In information age, *artificial intelligence, machine learning, structured programming*, and otherwise gains great popularity among researchers. LA is the basis of them, so in order to explore science in modern age, you should learn LA well.

- Solving linear system of equations.
  How to solve linear system of equations efficiently and correctly is the **key** question for mathematicians.

- Internal grace.
  LA is very beautiful, hope you enjoy the beauty of math.

- Interview questions.
  LA is often used for interview questions for phd. The interviewer usually ask difficult questions about LA.

### 1.1.1.2. Preview of LA

**Important: LA + Calculus + Probability.** The main branches of Mathematics are given below:

$$\text{mathematics}\begin{cases}\text{Analysis + Calculus}\\ \text{Algebra: foucs on structure}\\ \text{Geometry}\end{cases}$$

All parts of math are based on **axiom systems**. And **LA** is the significant part of *Algebra*, which focus on the linear structure. Every SSE student should learn **Linear Algebra**, **Calculus**, and **Probability** to build strong fundation.

**Practical: Computation.** Linear Algebra is more widely used than Calculus since we could use this **powerful** tool to do discrete computation. (As we know, we can use calculus to deal with something continuous. But how do we do integration when facing lots of **discrete data**? But linear algebra can help us deal with these data.)

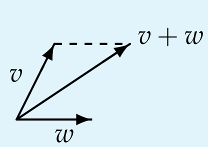
**Visualize.** Conncect between **Geometry** and **Algebra**.

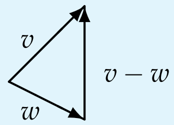
Let's take an easy example:

**Example 1.1** Let $v$ and $w$ donate two vectors as below:

$$v = \begin{bmatrix}1\\2\end{bmatrix}, \quad w = \begin{bmatrix}3\\4\end{bmatrix}.$$

Then we can donate these two vectors in the graph:




And we can also add two vectors to get $v+w$. Additionally, we can change the coefficients in front of $v$ and $w$ to get $v-w$.

In two dimension space, we can visualize the vector in the coordinate. Then let's watch the **three** dimension space.There are four vectors $u$,$v$,$w$ and $b$. We can also denote it in coordinate.

Here we raise a question: Can we denote vector $b$ as a linear combination with the three vectors $u$, $v$, and $w$? That is to say, Is there exists coefficients $x_1$, $x_2$, $x_3$ such that

$$x_1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + x_2 \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + x_3 \begin{pmatrix} 1 \\ 3 \\ 4 \end{pmatrix} = \begin{pmatrix} 2 \\ 5 \\ 7 \end{pmatrix}?$$

Then we only need to solve the system of equations

$$\begin{cases} x_1 + x_2 + x_3 = 2 \\ x_1 + 2x_2 + 3x_3 = 5 \\ x_1 + 3x_2 + 4x_3 = 7 \end{cases} \implies \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

**Abstract: Broad Applications.** Don't worry, broad doesn't mean boring. Instead, it means Linear Algebra can applied to lots of applications.

For example, if we denote a sequence of infinite numbers as a tuple that contains infinite numbers, and we denote this tuple as a vector, then we could build **an infinite banach space.** Moreover, Given a function $f : \mathbb{R} \to \mathbb{R}$, we can describe a set of functions as a tuple, then we could build a **function space**. These abstract knowledge may be not covered in this course. We will learn it in future courses.

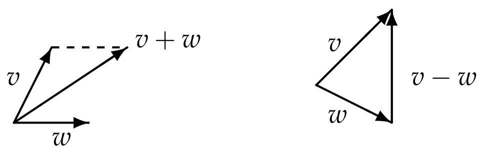
#### 1.1.1.3. What is Linear Algebra?

### 1.1.2. Vector, Addition

Let $v$ and $w$ donate two vectors as below:

$$v = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad w = \begin{bmatrix} 3 \\ 4 \end{bmatrix}.$$

Then we can donate these two vectors in the graph:




And we can also add two vectors to get $v + w$. Additionally, we can change the coefficients in front of $v$ and $w$ to get $v - w$.

In two dimension space, we can visualize the vector in the coordinate.

## 1.2. Lecture 2: Vector II: Norms and Inner Products

### 1.2.1. Vector Norm

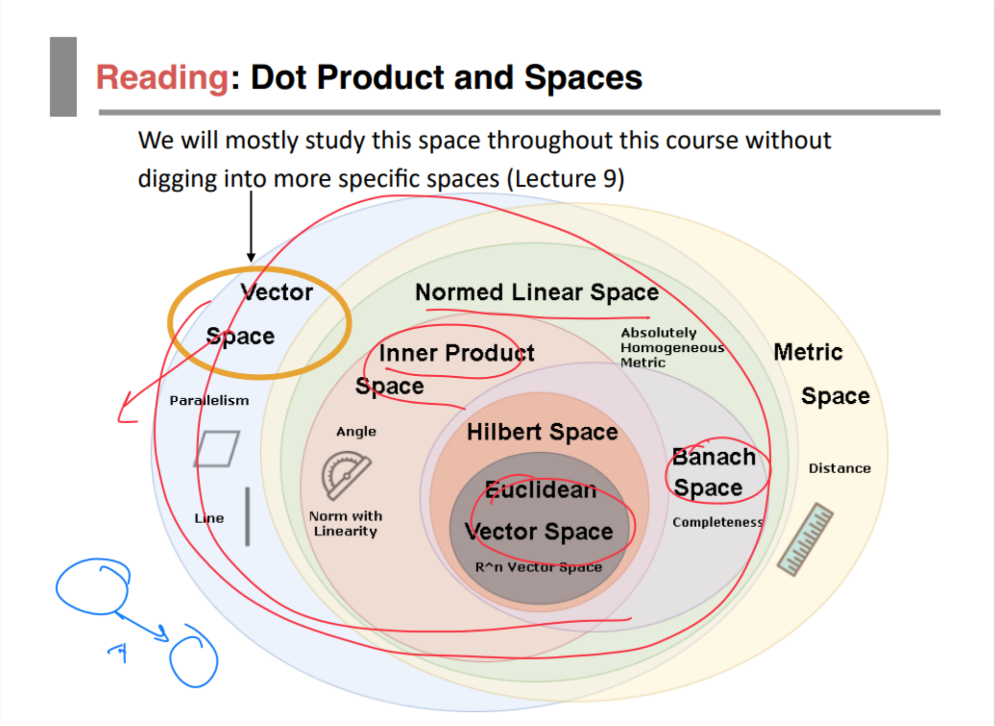
**Definition 1.1** [vector norm] Let $\boldsymbol{v} = (v_1, \cdots, v_n)$ be a $n-$length vector, the norm of $\boldsymbol{v}$, denoted by $\|\boldsymbol{v}\|_2$, is defined by

$$\|\boldsymbol{v}\|_2 = \sqrt{v_1^2 + \cdots + v_n^2}$$

**Definition 1.2** [unit vector] $\boldsymbol{v}$ is a unit vector if $\|\boldsymbol{v}\|_2 = 1$.

Note that for any vector $\boldsymbol{v}$, $\frac{v}{\|\boldsymbol{v}\|_2}$ is a unit vector.

**Proposition 1.1**
1. $\|\boldsymbol{v}\|_2 \geq 0$, and $\|\boldsymbol{v}\|_2 = 1$ iff $\boldsymbol{v} = 0$
2. $\|c\boldsymbol{v}\|_2 = |c|\|\boldsymbol{v}\|_2$, $\forall c \in \mathbb{R}$
3. $\|\boldsymbol{v} + \boldsymbol{w}\|_2 \leq \|\boldsymbol{v}\|_2 + \|\boldsymbol{w}\|_2$, $\forall \boldsymbol{v}, \boldsymbol{w}$



### 1.2.2. Inner Product

Here we introduce the definition for inner product of vector:

**Definition 1.3** [inner product] Given two vectors $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$, the inner product between $x$ and $y$ is given by

$$\langle x, y\rangle = x_1y_1 + x_2y_2 + \cdots + x_ny_n$$

The notation of inner product can also be written as $x^{\mathrm{T}}y$ or $x \cdot y$.

Pro. Tom Luo highly recommends you to write *inner product* as $\langle x, y\rangle$. For myself, I also try to avoid using notation $x \cdot y$ to avoid misunderstanding.

Note that $\langle \boldsymbol{v}, \boldsymbol{v}\rangle = \|\boldsymbol{v}\|_2^2$, we denote $\|\boldsymbol{v}\| = \sqrt{\langle \boldsymbol{v}, \boldsymbol{v}\rangle}$.

**Proposition 1.2** (1). Linearity $\langle a\boldsymbol{u} + b\boldsymbol{v}, \boldsymbol{w}\rangle$ =a$\langle \boldsymbol{u}, \boldsymbol{w}\rangle$ +b$\langle \boldsymbol{v}, \boldsymbol{w}\rangle$

(2). Symmetry $\langle \boldsymbol{u}, \boldsymbol{v}\rangle$ = $\langle \boldsymbol{v}, \boldsymbol{u}\rangle$

(3). Positivity $\langle \boldsymbol{v}, \boldsymbol{v}\rangle \geq 0, \langle \boldsymbol{v}, \boldsymbol{v}\rangle = 0 iff \boldsymbol{v} = 0$

### 1.2.3. Properties of Inner Product and Norm

**Theorem 1.1 — Cauchy–Schwarz Inequality.** For any vectors $\boldsymbol{v}$ and $\boldsymbol{w}$,

$$\|\boldsymbol{v}\| \cdot \|\boldsymbol{w}\| \geq \langle \boldsymbol{v}, \boldsymbol{w}\rangle$$

We can define the angle $\theta$ between vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ by

$$\cos\theta = \frac{\langle \boldsymbol{v}, \boldsymbol{w}\rangle}{\|\boldsymbol{v}\| \cdot \|\boldsymbol{w}\|}$$

**Theorem 1.2** For any vectors $\boldsymbol{v}$ and $\boldsymbol{w}$,

$$\|\boldsymbol{v}\| + \|\boldsymbol{w}\| \geq \|\boldsymbol{v} + \boldsymbol{w}\|$$

*Proof.*

$$\begin{aligned}
(\|\boldsymbol{v}\| + \|\boldsymbol{w}\|)^2 &= \|\boldsymbol{v}\|^2 + \|\boldsymbol{w}\|^2 + 2\|\boldsymbol{v}\| \cdot \|\boldsymbol{w}\| \\
&\geq \|\boldsymbol{v}\|^2 + \|\boldsymbol{w}\|^2 + 2\langle \boldsymbol{v}, \boldsymbol{w} \rangle \\
&= (\|\boldsymbol{v} + \boldsymbol{w}\|)^2
\end{aligned}$$

■

# Chapter 2

# Week 2: System of Linear Equations

## 2.1. Lecture 3: Solving Systems of Linear Equations

We have learned vectors, norm and inner product, and provided a few examples of the inner product. You shall be able to compute the inner product of the two vectors now. In this lecture, we will learn an "inverse" problem of how to compute one vector based on inner products: solving a system of linear equations.

**Learning goals**. After this lecture, you should be able to:

- Describe the general form of linear system of equations;
- Solve a small linear system of equations by Gaussian elimination;
- Utilize the linear system to model a simple recommendation problem.
- Draw plots to show when the system has no solution, one solution and infinitely many solutions

### 2.1.1. Motivating Applications

You probably have learned simple examples of linear systems of equations during middle school.

**Example 1.** (a middle-school example) Imagine telling a friend about your favorite fruit store. Your friend asks how much it charges for a kilogram of oranges and a kilogram of apples. While you cannot remember how much each one costs individually, you do remember that the last time you ordered one kilogram of oranges and two kilograms of apples, your bill was 8 RMB. And today, when you bought two two kilograms of oranges and three kilograms of apples, your bill came to 13 RMB. Now you have all the information you need to answer your friend's question.

Assume the price of one kilogram of oranges is $x_1$ RMB, and the the price of one kilogram of apples is $x_2$ RMB. Then according to the descriptions above, we have

$$x_1 + 2x_2 = 8, \tag{2.1a}$$

$$2x_1 + 3x_2 = 13. \tag{2.1b}$$

In middle school, we call it "system of two-variable first-order equations", or "system of linear equations in two variables".

How to solve this *system of linear equations* in two variables? You can try yourself.

This kind of examples are well studied during middle school. Then what is new in this course? A simple answer is: "large". All examples in middle school contain few variables, often two or three. When there are many more variables (say, tens, hundreds of even millions), things become much more challenging and interesting. In this course, we will learn how to solve a system of linear equations with an arbitrary number of variables.

Before diving into the solution methods, let us briefly discuss the motivation. Will we really encounter system of linear equations that involve more than, say, 10 variables? Yes, there are many large-scale examples in science and engineering. Below we provide an example from recommendation systems.

**Example 2.** (recommendation) Alice is a frequent user of a restaurant rating website and has provided ratings of hundreds of movies. The website wants to recommend new restaurants based on the ratings. The platform proposed the following method. For each restaurant $i \in \{1, 2, \ldots, m\}$, they compute a feature vector [1] $\mathbf{a}_i = (a_{i1}, \ldots, a_{in})$ which consists of the taste score, the freshness score, the diversity score, the environment score, the parking convenience score, etc. They make the following assumption:

**Assumption 1**: Alice assigns (unknown) weight $x_j$ to the $j$-th feature, and computes the rating of restaurant $i$ as $b_i = \sum_{j=1}^{n} x_j a_{i,j} = \langle \mathbf{a}_i, \mathbf{x} \rangle$, where $\mathbf{x} = (x_1, \ldots, x_n)$ is the weight vector.

If the website knows $\mathbf{x}$, then they can predict the rating of Alice for a new restaurant with

[1] How to compute this feature vector is also a nontrivial task, but for now, let us assume it is given.

feature vector $\hat{\mathbf{a}}$ as $r = \langle \hat{\mathbf{a}}, \mathbf{x} \rangle$. If the predicted rating is high, then the website can recommend the restaurant to Alice.

**Assumption 2**: We assume that the unknown vector $\mathbf{x}$ and the given feature vector $\mathbf{a}$ follow a linear model. i.e. $\langle \mathbf{x}, \mathbf{a}_i \rangle = b_i, \forall i = 1, 2, \cdots$ This amounts to solving a system of linear equations, which we formally define next.

### 2.1.2. Formal Definition of Linear System of Equations.

**Definition 2.1** [Linear Equations] A linear equation in $n$ unknowns is the equation of the form

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

where $a_1, a_2, \ldots, a_n, b$ are real numbers and $x_1, x_2, \ldots, x_n$ are variables.

**Definition 2.2** [Linear System of Equations] Linear system of $m$ equations in $n$ unknowns is the system of the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2, \\ &\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m, \end{aligned} \tag{2.2}$$

where $a_{ij}$ and the $b_i$ are all real numbers. We refer to (2.2) as an $m \times n$ **linear system of equations**; or sometimes simply as an $m \times n$ linear system.

**Non-example**: Let's look an example of nonlinear equations below:

$$\begin{cases} 3x_1x_2 + 5x_1^2 + 6x_2 = 9 \\ x_1x_2^2 + 5x_1 + 7x_2^2 = 10 \end{cases}$$

This system of equations looks complicated. Indeed, it is a system of nonlinear equations. We will NOT study this kind of equations in this course. In future (e.g. algebraic geometry courses in math departments), it is possible that you will encounter such nonlinear equations.

Let us come back to linear systems, and discuss how to solve it. We start by showing a $2 \times 2$ example, and then extend to general linear sytems.

### 2.1.3. Toy Example: How to Solve $2 \times 2$ Linear System

Let's review how to solve a $2 \times 2$ system of linear equations.

**Example 2.1**

$$1x_1 + 2x_2 = 5 \tag{2.3}$$

$$4x_1 + 5x_2 = 14. \tag{2.4}$$

Firstly, by adding $(-4) \times (2.3)$ and $(2.4)$, we obtain:

$$1x_1 + 2x_2 = 5 \tag{2.5}$$

$$0x_1 + (-3)x_2 = -6 \tag{2.6}$$

Secondly, by multiplying $-(1/3)$ of $(2.6)$, we obtain:

$$1x_1 + 2x_2 = 5 \tag{2.7}$$

$$1x_2 = 2 \tag{2.8}$$

Thirdly, by adding $(-2) \times (2.8)$ into $(2.7)$, we obtain:

$$1x_1 + 0x_2 = 1 \tag{2.9}$$

$$1x_2 = 2 \tag{2.10}$$

Here we get the solution $(x_1 = 1, x_2 = 2)$.

■

The method shown above is called **Gaussian Elimination** (GE). Here we just showed a very simple version of GE, and we will generalize it to an arbitrary linear system later.

You may notice that in the above process, the calculation is executed on the "numbers" 1, 2, 4, 5, 14, etc., but not the variables "$x_1, x_2$". The variables $x_1, x_2$ mainly provide indications of the places to put these numbers. With this understanding, we can remove the symbols $x_1, x_2$ in the above process, and only keep the numbers like $1,2,5,etc$. This leads to a cleaner process shown below.

**Example 2.2** (revisiting Example 2.1, using matrix form) Removing the variables $x_1, x_2$, we can rewrite the system in a form of so-called (**augmented matrix**):

$$\left[\begin{array}{cc|c} 1 & 2 & 5 \\ 4 & 5 & 14 \end{array}\right]$$

We can then rewrite the above process using the augmented matrix form:

$$\left[\begin{array}{cc|c} 1 & 2 & 5 \\ 4 & 5 & 14 \end{array}\right] \Longrightarrow \left[\begin{array}{cc|c} 1 & 2 & 5 \\ 0 & -3 & -6 \end{array}\right] \Longrightarrow \left[\begin{array}{cc|c} 1 & 2 & 5 \\ 0 & 1 & 2 \end{array}\right] \Longrightarrow \left[\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 2 \end{array}\right].$$

■

There are a few benefits of using the matrix form:

For now, you can simply view "matrix" as a specific object that collects a bunch of numbers. We will dive much deeper into "matrix" later.

Next, we give a formal definition of augmented matrix:

**Definition 2.3** [Augmented matrix] For a system of linear equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m, \end{aligned} \tag{2.11}$$

the corresponding augmented matrix is defined as

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_n \end{array}\right].$$

We give the definition for a new term **pivot**:

**Definition 2.4** [pivot] Returning to the example, we find after third step the matrix is given by

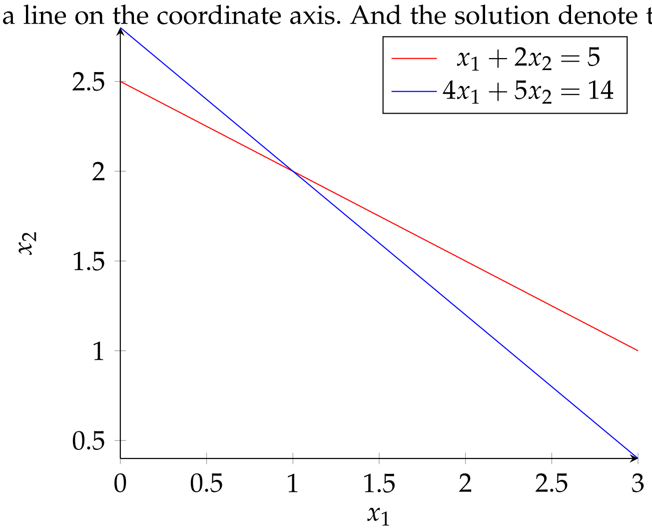
$$\left[\begin{array}{cc|c} 1 & 2 & 5 \\ 0 & 1 & 2 \end{array}\right].$$

We find that the second row will be used to eliminate the element in the second column of the first row. Here we refer to the second row as the **pivot row**. The first nonzero entry in the pivotal row is called the **pivot**. For this example, the element "1" in the intersection of the second row and the second column is the pivot.

#### 2.1.3.1. How to visualize the system of equation?

Here we try to visualize the system of equation $\begin{cases} 1x_1 + 2x_2 = 5 \\ 4x_1 + 5x_2 = 14 \end{cases}$:

**Row Picture.** Focusing on the row of the system of equation, we can denote each equation as a line on the coordinate axis. And the solution denote the coordinate.

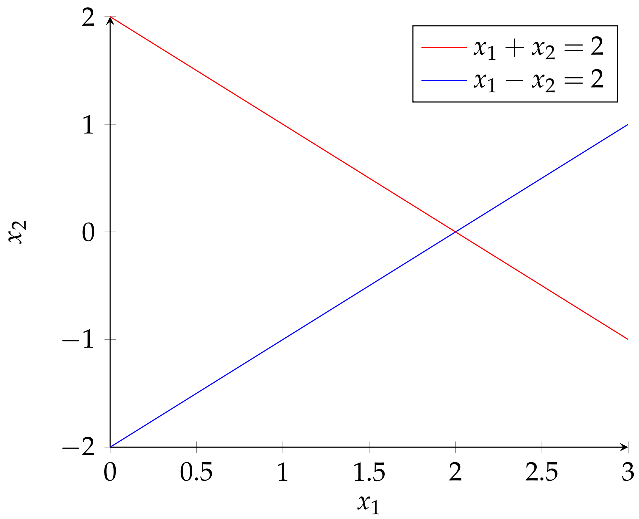


**Column Picture.** Focusing on the column of the system of equation, we can denote $\begin{bmatrix} 1 \\ 4 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 5 \end{bmatrix}$ as vectors in coordinate axis. Could the linear combinations of these two vectors form the vector $\begin{bmatrix} 5 \\ 14 \end{bmatrix}$? If we denote $x_1$ and $x_2$ as coefficients, it suffices to solve the equation $x_1 \begin{bmatrix} 1 \\ 4 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 5 \end{bmatrix} = \begin{bmatrix} 5 \\ 14 \end{bmatrix}$.

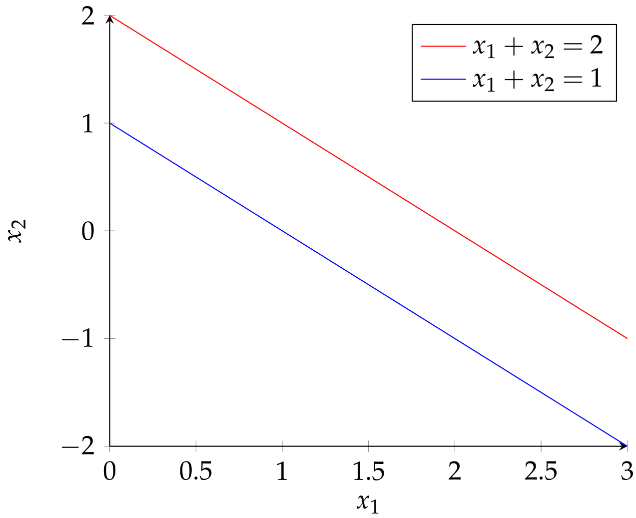
### 2.1.3.2. The solutions of the Linear System of Equations

The solution to linear system equation could only be **unique**, **infinite**, or **empty**. Let's talk about it case by case in graphic way:

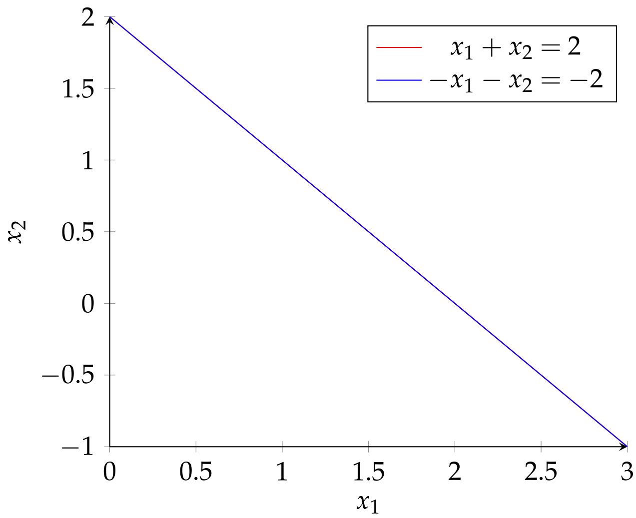
**Case 1: unique solution.** If two lines intersect at one point, then there is unique solution.




**Case2: no solution.** If two lines are parallel, then there is no solution.




**Case 3: infinite number of solutions.** If both equations represent the same line, then there are infinite number of solutions.




## 2.2. Lecture 4: Solving $n \times n$ System

In this lecture, we discuss how to solve an $n \times n$ system. We will discuss a general $m \times n$ (where $m$ can be different or the same as $n$) system in the next lecture.

## 2.2.1. Matrix and Triangular Matrix

We start by defining a matrix, then define coefficient matrix, triangular matrix and diagonal matrix.

### Matrix.

**Definition 2.5** [Matrix] An $m \times n$ matrix $A$ is a rectangular array of numbers with $m$ rows and $n$ columns in the following form:

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

where all $a_{ij}$ are scalars.

**Remark:**.

- For a matrix $A$, $a_{ij}$ is called the $(i, j)$-th entry of $A$, and we often denote $A$ as $(a_{ij})_{m \times n}$.
- Matrices are often denoted by capital letter $A$, $B$, $C$, ...
- When $m = n$, $A$ is called a square matrix. e.g. $\begin{bmatrix} 3 & 4 & 1 \\ -1 & -5 & 0 \\ -1 & 5 & 2 \end{bmatrix}$ is a $3 \times 3$ square matrix; $\begin{bmatrix} 2 & 5 & 3 \\ 6 & -5 & 2 \end{bmatrix}$ is not a square matrix.
- When all entries are zero, $A$ is called a zero matrix. e.g. $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ are zero matrices.
- when $m = 1$, $A$ becomes a row vector. e.g. $\begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix}$
- when $n = 1$, $A$ becomes a column vector. e.g. $\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$
- when $m = n = 1$, $A$ can be considered as a scalar. e.g. $\begin{bmatrix} 0 \end{bmatrix}$, $\begin{bmatrix} 2 \end{bmatrix}$.

### Coefficient Matrix.

**Definition 2.6** [Coefficient matrix] For a system of linear equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m, \end{aligned} \tag{2.12}$$

the coefficient matrix of the system is an $m \times n$ matrix

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

For example, if the linear system is given by $\begin{cases} 3x_1 + 5x_1 + 6x_3 + x_4 = 9 \\ x_1 + 4x_2 + 7x_3 + 2x_4 = 5 \\ 2x_1 + 3x_2 + x_3 + 4x_4 = 3 \end{cases}$, then it's coefficient matrix is $\begin{bmatrix} 3 & 5 & 6 & 1 \\ 1 & 4 & 7 & 2 \\ 2 & 3 & 1 & 4 \end{bmatrix}$

### Triangular Matrix.

**Definition 2.7** [Upper triangular matrix] $U = (u_{ij})_{n\times n}$ is called an upper triangular matrix if $u_{ij} = 0$ for all $1 \leq j < i \leq n$.

$$U = \begin{bmatrix} u_{1,1} & u_{1,2} & u_{1,3} & \cdots & u_{1,n} \\ & u_{2,2} & u_{2,3} & \cdots & u_{2,n} \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & u_{n-1,n} \\ 0 & & & & u_{n,n} \end{bmatrix} \tag{2.13}$$

**Example:** $\begin{bmatrix} 2 & 5 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 8 \end{bmatrix}$, $\begin{bmatrix} 1 & 4 & 2 & 3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 5 & 1 \\ 0 & 0 & 0 & 2 \end{bmatrix}$ are upper triangular matrices.

**Definition 2.8** [Lower triangular matrix] $L = (l_{ij})_{n\times n}$ is called a lower triangular matrix if $l_{ij} = 0$ for all $1 \le i < j \le n$.

$$L = \begin{bmatrix} l_{1,1} & & & & 0 \\ l_{2,1} & l_{2,2} & & & \\ l_{3,1} & l_{3,2} & \ddots & & \\ \vdots & \vdots & \ddots & \ddots & \\ l_{n,1} & l_{n,2} & l_{n,3} & \dots & l_{n,n} \end{bmatrix} \tag{2.14}$$

**Example:** $\begin{bmatrix} 4 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 5 & 6 \end{bmatrix}$, $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 \\ 4 & 3 & 2 & 1 \end{bmatrix}$ are lower triangular matrices.

### Diagonal Matrix.

**Definition 2.9** [Diagonal matrix] $D = (d_{ij})_{n\times n}$ is called a diagonal matrix if $d_{ij} = 0$ for all $i \neq j$.

$$D = \begin{bmatrix} d_{1,1} & 0 & 0 & 0 & 0 \\ 0 & d_{2,2} & 0 & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \\ 0 & 0 & 0 & \dots & d_{n,n} \end{bmatrix} \tag{2.15}$$

denote $D$ as diag$\{d_1, d_2, \dots, d_n\}$.

## 2.2.2. Allowable Operations

When solving a linear system of equations, a certain operation on the equations (equation-operation) correspond to a certain operation on the augmented matrix (aug-matrix-operation).

We first analyze allowable equations-operations, and then discuss allowable aug-matrix-operations.

### Allowable equation-operations..

What are the allowable equation-operations?

- (EO1) **Multiplication**: Multiply an equation by a nonzero constant. e.g.

$$\begin{cases} x_1 + \frac{5}{3}x_2 = 3 \\ x_1 + 4x_2 = 5 \end{cases} \xrightarrow{3\times \text{ row } 1} \begin{cases} 3x_1 + 5x_2 = 9 \\ x_1 + 4x_2 = 5 \end{cases}$$

- (EO2) **Addition-then-Multiplication**. Add to an equation by a constant multiple of another equation. e.g.

$$\begin{cases} 3x_1 + 5x_2 = 9 \\ x_1 + 4x_2 = 5 \end{cases} \xrightarrow{\text{Add } (-3)\times \text{ row 2 to row 1}} \begin{cases} -7x_2 = -6 \\ x_1 + 4x_2 = 5 \end{cases}$$

- (EO3) **Interchange**. Interchange two equations. e.g.

$$\begin{cases} -7x_2 = -6 \\ x_1 + 4x_2 = 5 \end{cases} \xrightarrow{\text{Interchange row 1 and row 2}} \begin{cases} x_1 + 4x_2 = 5 \\ -7x_2 = -6 \end{cases}$$

Why are these allowable? It is not hard to verify that the solutions do not change after the operations. A more formal statement is given in the above exercise.

**Exercise 4.1:** Denote the original system as $\mathcal{A}$ and the new system as $\hat{\mathcal{A}}$. Verify: (a) If $\mathbf{x}$ is a solution of $\mathcal{A}$, then $\mathbf{x}$ is a solution of $\hat{\mathcal{A}}$. (b) If $\mathbf{x}$ is a solution of $\hat{\mathcal{A}}$, then $\mathbf{x}$ is a solution of $\mathcal{A}$.

### Non-allowable equation-operations..

To better understand the allowable equation-operations, it will be good to know some non-allowable operations. Let us list a few examples (which is surely not an exhaustive list):

- (EO4) Multiplying an equation by zero;
- (EO5) Exchanging variables (e.g., change the system $x_1 + 2x_2 = 8; 2x_1 + 3x_2 = 13$ to $x_2 + 2x_1 = 8; 3x_1 + 2x_2 = 13$).
- (EO6) Multiplying the coefficients of two equations to get a new equation;

**Exercise 4.2:** Show that (EO4) not allowable. (Hint: Provide an example to show that (EO4) violates (b) of Exercise 4.1. )

**Allowable aug-matrix-operations..** The three allowable equation-operations correspond to three allowable aug-matrix-operations as below.

- **Multiplication** Multiply a row by a nonzero constant.
- **Addition-then-Multiplication**. Add to a row by a constant multiple of another row.
- **Interchange** Interchange two rows.

**Definition 4.1** (Elementary row operations) The three operations above are called elementary row operations.

**Non-allowable aug-matrix-operations..** The three non-allowable equation-operations correspond to three non-allowable aug-matrix-operations. We do not list all of them, but highlight just (EO5): exchanging variables corresponds to exchanging columns. Thus when manipulating the augmented matrix for solving linear systems, *exchanging columns is not allowed*.

**Remark**: If the purpose is not solving linear systems, then it is possible that exchanging columns is allowed. The term "allowable" is tied to a specific purpose. For different purposes, the meaning of "allowable" can be different.

**Understanding allowable operations by examples..** Why are some operations allowable and others are not? Though we have provided mathematical explanations, it is still nice to relate the allowable operations to a concrete example. Consider the example of fruit buying.

Think: What are the physical meanings of "adding one row to another", "multiplying a row", "exchanging varriables"?

## 2.2.3. Solving a $3 \times 3$ System

We discuss how to solve a $3 \times 3$ system equations.

$$\begin{cases} 2x_1 + x_2 + x_3 = 5 & (1) \\ 4x_1 + (-6)x_2 = -2 & (2) \\ -2x_2 + 7x_2 + 2x_3 = 9 & (3) \end{cases} \tag{2.16}$$

We first provide an overview of Gaussian elimination for solving this system using equations, which you probably learned in middle school before.

Firstly, we eliminate $x_1$ in equation (2) and (3). This is accomplished by adding $-2 \times (1)$ to (2) and adding (1) to (3). we get

$$\begin{cases} -8x_2 - 2x_3 = -12 & (4) \\ 8x_2 + 3x_3 = 14 & (5) \end{cases} \tag{2.17}$$

then, we eliminate $x_2$ in equation (5) by adding equation (4). We get $x_3 = 2$. Next, we plug in the value of $x_3$ into the equations to solve out $x_1$ and $x_2$. In particular, taking $x_3 = 2$ in equation (5), we get $8x_2 + 6 = 14$. Thus, $x_2 = 1$. Taking $x_2 = 1$,$x_3 = 2$ into equation (1), we get $2x_1 + 1 + 2 = 5$. Thus, $x_1 = 1$. Finally, we conclude that the solution for this linear system is $x_1 = 1$, $x_2 = 1$, $x_3 = 2$.

We will use the augmented matrix form to implement Gaussian elimination. We write the linear system in the **augmented matrix** form:

$$\begin{cases} 2x_1 + x_2 + x_3 = 5 \\ 4x_1 + (-6)x_2 = -2 \\ -2x_2 + 7x_2 + 2x_3 = 9 \end{cases} \Longrightarrow \left[\begin{array}{ccc|c} 2 & 1 & 1 & 5 \\ 4 & -6 & 0 & -2 \\ -2 & 7 & 2 & 9 \end{array}\right]$$

**Step 1**: **Elimination to get the value of one variable.**

$$\left[\begin{array}{ccc|c} 2 & 1 & 1 & 5 \\ 4 & -6 & 0 & -2 \\ -2 & 7 & 2 & 9 \end{array}\right] \xRightarrow[\text{Add row 1 to row 3}]{\text{Add } (-2)\times \text{ row 1 to row 2}} \left[\begin{array}{ccc|c} 2 & 1 & 1 & 5 \\ 0 & -8 & -2 & -12 \\ 0 & 8 & 3 & 14 \end{array}\right]$$

$$\xRightarrow{\text{Add row 2 to row 3}} \left[\begin{array}{ccc|c} 2 & 1 & 1 & 5 \\ 0 & -8 & -2 & -12 \\ 0 & 0 & 1 & 2 \end{array}\right]$$

The coefficient matrix is an upper triangular matrix [2].

**Step 2**: Back substitution.

[2]Note that the augmented matrix is not a square matrix, and thus we cannot say "the augmented matrix is an upper triangular matrix".

$$\xrightarrow[\text{Add } 2\times \text{ row 3 to row 2}]{} \left[\begin{array}{ccc|c} 2 & 1 & 1 & 5 \\ 0 & -8 & 0 & -8 \\ 0 & 0 & 1 & 2 \end{array}\right] \xrightarrow[\text{Add } (-1)\times \text{ (row 3) to (row 1)}]{} \left[\begin{array}{ccc|c} 2 & 1 & 0 & 3 \\ 0 & -8 & 0 & -8 \\ 0 & 0 & 1 & 2 \end{array}\right].$$

$$\xrightarrow[\text{Multiply row 2 by (-1/8)}]{} \left[\begin{array}{ccc|c} 2 & 1 & 0 & 3 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 2 \end{array}\right] \xrightarrow[\text{Add } (-1)\times \text{ row 2 to row 1}]{} \left[\begin{array}{ccc|c} 2 & 0 & 0 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 2 \end{array}\right]$$

$$\xrightarrow[\text{Multiply row 2 by (-1/8)}]{} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 2 \end{array}\right]$$

This augmented matrix can be transformed back to a linear system:

$$\begin{cases} 1 \times x_1 + 0 \times x_2 + 0 \times x_3 = 1 \\ 0 \times x_1 + 1 \times x_2 + 0 \times x_3 = 1 \\ 0 \times x_1 + 0 \times x_2 + 1 \times x_3 = 2 \end{cases}$$

This is equivalent to

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}.$$

This is the solution of the original linear system of equations given in 2.16.

## 2.2.4. How to Solve an $n \times n$ System?

In the last section, we demonstrate how to solve an $3 \times 3$ system by performing Gaussian elimination. In this section, We describe the general Gaussian elimination (GE) for solving an $n \times n$ system.

Given an $n \times n$ System of the form:

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & b_n \end{array}\right] \tag{2.18}$$

The basic idea of Gaussian elimination is to reduce the coefficient matrix to an upper triangular matrix, and then to a diagonal matrix. The outline of GE is provided as follows:

- **Step 1: Forward Elimination**.

  Perform elementary row operations in forward direction (from top rows to bottom rows) and try to get a "simpler" matrix (the upper triangular matrix)

- **Step 2: Backward substitution.**

  Perform elementary row operations in backward direction (from bottom rows to top rows) and simplify the system to a diagonal system.

**Explanation of GE for general system..** Now we explain the two steps below.

**Step 1: Forward Elimination**.

**Assumption 4.1**: The **diagonal entries** are always *nonzero* during our operation.

Add row 1 that multiplied by a constant to other $n-1$ row to ensure the first entry of other $n-1$ rows are all *zero* (one can eliminate the first entry of $i$th row by adding $(-\frac{a_{i1}}{a_{11}})\times$ row 1 to row $i$ ):

$$\Longrightarrow \left[\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ 0 & \times & \dots & \times & \times \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \times & \cdots & \times & \times \end{array}\right] \tag{2.19}$$

Then we proceed this way $n-1$ times to obtain:

$$\left[\begin{array}{ccccc|c} \times & \times & \times & \times & \times & \times \\ & \times & \times & \times & \times & \times \\ & & \times & \times & \times & \vdots \\ & \mathbf{0} & & \times & \times & \vdots \\ & & & & \times & \times \end{array}\right] \tag{2.20}$$

After forward elimination, the coefficient matrix is the **upper triangular form**.

**Step 2: Backward substitution.**

Add row n that multiplied by a constant to other $n-1$ row to ensure the last entry of other $n-1$ rows are all *zero*.

$$\left[\begin{array}{ccccc|c} \times & \times & \times & \times & 0 & \times \\ & \times & \times & \times & 0 & \times \\ & & \times & \times & 0 & \vdots \\ & \mathbf{0} & & \times & 0 & \vdots \\ & & & & \times & \times \end{array}\right] \tag{2.21}$$

Then we proceed this way $n-1$ times to obtain:

$$\left[\begin{array}{ccccc|c} \times & & & & & \times \\ & \times & & & \mathbf{0} & \times \\ & & \times & & & \vdots \\ & \mathbf{0} & & \times & & \vdots \\ & & & & \times & \times \end{array}\right] \tag{2.22}$$

After backward substitution, the coefficient matrix is the **diagonal form**.

Finally by multiplying every row by a nonzero constant to ensure its **diagnoal entries** are all 1:

$$\left[\begin{array}{ccccc|c} 1 & & & & & \times \\ & 1 & & & \mathbf{0} & \times \\ & & \ddots & & & \vdots \\ & \mathbf{0} & & 1 & & \vdots \\ & & & & 1 & \times \end{array}\right] \tag{2.23}$$

Can this method really solve the original linear system? We present a result here (skip the proof) which states that under a certain assumption, the method works.

**Claim:** For a given linear system with a coefficient matrix $A$, denote $\widetilde{A} = (\widetilde{a}_{ij})_{n\times n}$ as the matrix after performing forward elimination on $A$, if $\widetilde{A}$ is a upper triangular matrix with nonzero diagonal entries (i.e. $\widetilde{a}_{ii} \neq 0$ for all $1 \leq i \leq n$), then the linear system has a unique solution.

**Preview of the next lecture.** Note that this assumption may or may not hold in practice, and we will discuss a more general case next time. Another important topic is how to solve a general rectangular system

# Chapter 3

# Week3

## 3.1. Lecture 5: Solving Rectangular Systems: Row-Echelon Forms

In last lecture, we learned how to solving $n \times n$ linear system by Gaussian Elimination. More precisely, we perform elementary row operations to get an upper triangular matrix and later a diagonal matrix. We will discuss how to solve a general system in this lecture.

### 3.1.1. Identity Matrix and Pivot

We start by defining an identity matrix which appeared in the GE process.

**Identity Matrix.**

**Definition 3.1** [Identity Matrix] The identity matrix of size $n$ is the $n \times n$ square matrix with ones on the main diagonal and zeros elsewhere:

$$I_1 = \begin{bmatrix} 1 \end{bmatrix}.$$

$$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

$$I_n = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

Equivalently, $A = (a_{ij})$ is an identity matrix $\leftrightarrow a_{ii} = 1, \forall i;\ a_{ij} = 0, \forall i \neq j$

The identity matrix can also be written as $I_n = diag(1, 1, \cdots, 1)$

We then introduce the definition of *pivot*, which plays an essential role in Gaussian-Elimination.

### Pivot.

**Definition 3.2** [Pivot]

- The first nonzero in the row that does elimination.
- The row contains the pivot is called a *pivot row*.

**Example 3.1** We introduce the concept of pivots by raising simple examples

$$\left[\begin{array}{cc|c} \underline{4} & -8 & 4 \\ 3 & 2 & 11 \end{array}\right] \Longrightarrow \left[\begin{array}{cc|c} 4 & -8 & 4 \\ 0 & \underline{8} & 8 \end{array}\right] \tag{3.1}$$

The underlined elements are called the pivots, which are used to eliminate other rows.

**Remark:**

- The pivots are on the diagonal of the triangle after elimination.
- It is not necessary that each column has a pivot.
- Each column has at most one pivot.

## 3.1.2. Row Echelon Form and Reduced Row Echelon Form

Let's discuss an example to introduce the concept for row-echelon form.

**Example 3.2** We apply Gaussian Elimination to try to transform a Augmented matrix:

- In step one we choose the first row as pivot row (the first nonzero entry is the pivot):

$$\left(\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & 0 & 0 & 1 & -1 \\ -2 & -2 & 0 & 0 & 3 & 1 \\ 0 & 0 & 1 & 1 & 3 & -1 \\ 1 & 1 & 2 & 2 & 4 & 1 \end{array}\right) \xRightarrow[\text{Add } 1\times \text{ row 1 to row 2; Add } 2\times \text{ row 1 to row 3}]{\text{Add } (-1)\times \text{ row 1 to row 5}}$$

$$\left[\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 2 & 0 \\ 0 & 0 & 2 & 2 & 5 & 3 \\ 0 & 0 & 1 & 1 & 3 & -1 \\ 0 & 0 & 1 & 1 & 3 & 0 \end{array}\right]$$

- Then we choose second row as pivot row to continue elimination:

$$\xRightarrow[\text{Add } (-2)\times \text{ row 2 to row 3; Add } (-1)\times \text{ row 2 to row 4}]{\text{Add } (-1)\times \text{ row 2 to row 5}} \left[\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{array}\right]$$

- Next, we choose the third row as pivot row to continue elimination:

$$\xRightarrow[\text{Add } (-1)\times \text{ row 3 to row 5}]{\text{Add } (-1)\times \text{ row 3 to row 1; Add } (-1)\times \text{ row 3 to row 4}} \left[\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & 0 & -2 \\ 0 & 0 & 1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 & -4 \\ 0 & 0 & 0 & 0 & 0 & -3 \end{array}\right] \tag{3.2}$$

Note that the matrix (6.2) is said to be the **Row Echlon form**.

- Finally, we set second row as pivot row then set third row as pivot row to do elimination:

$$\xrightarrow[\text{Add } 2\times \text{ row 3 to row 1; Add } (-2)\times \text{ row 3 to row 2}]{\text{Add } (-1)\times \text{ row 2 to row 1}} \left[\begin{array}{ccccc|c} 1 & 1 & 0 & 0 & 0 & 4 \\ 0 & 0 & 1 & 1 & 0 & -6 \\ 0 & 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 & -4 \\ 0 & 0 & 0 & 0 & 0 & -3 \end{array}\right] \tag{3.3}$$

The matrix (3.3) is said to be the **Reduced Row Echelon form**. Or equivalently, it is said to be the *singular matrix*. (Don't worry, we will introduce these concepts in future.)

You may find there exist many solutions to this system of equation, which means Gaussian Elimination **doesn't** always derive **unique** solution.

**Definition 3.3** [Row Echelon Form] A matrix is said to be in **row echelon form** if

- **(i)** The first nonzero entry in each nonzero row is 1.
- **(ii)** If row $k$ does not consist entirely of zeros, the number of leading zero entries in row $k+1$ is greater than the number of leading zero entries in row $k$.
- **(iii)** If there are rows whose entries are all zero, they are below the rows having nonzero entries.

For example, the following matrices are in REF:

$$\begin{bmatrix} 1 & \times & \times & \times & \times \\ & 1 & \times & \times & \times \\ & & 1 & \times & \times \\ & \text{\Large 0} & & 1 & \times \\ & & & & 1 \end{bmatrix}. \tag{3.4}$$

More examples: $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & -1 \end{bmatrix}$

We can conclude that there is a "ladder" in the matrix in REF, which satisfies:

- Below the ladder, all entries are zeros
- On the right side, all entries are ones

**Definition 3.4** [Reduced Row Echelon Form]

A matrix is said to be in **Reduced row echelon form** if

- **(i)** The matrix is in *row echelon form*.
- **(ii)** The first nonzero entry in each row is the only nonzero entry in its column.

For example, the following matrix is in RREF:

$$\begin{bmatrix} \boxed{1} & * & \cdots & * & 0 & * & \cdots & * & 0 & * & \cdots \\ 0 & 0 & \cdots & 0 & \boxed{1} & * & \cdots & * & 0 & * & \cdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \boxed{1} & * & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \tag{3.5}$$

where * can be any number. For example, the matrix $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ is also of *Row Echelon Form*! Moreover, it is of *Reduced Row Echelon Form*.

## General properties of a matrix in row echelon form.

- Each non-zero row has a strictly increasing number of leading zeros than the previous row.
- Each non-zero row has one pivot/leading entry.
- Each column has at most one pivot/leading entry.

## 3.1.3. Solving the General Rectangular Systems

In conclusion, we process a general system using the following steps:

1. Form the augmented matrix of the system.
2. Perform elementary row operations to get a row echelon form(REF).
   (a) Find a pivot (swap rows if no pivot is found in this row)
   (b) Use pivot to eliminate entries below
3. Perform elementary row operations to get a reduce row echelon form(RREF).
   (a) Use pivot to eliminate entries above (skip the columns with no pivot)

**Example 3.3** Solve the rectangular system:

$$\begin{cases} 2x_1 + x_2 + 7x_3 - 7x_4 = 2 \\ -3x_1 + 4x_2 - 5x_3 - 6x_4 = 3 \\ x_1 + x_2 + 4x_3 - 5x_4 = 2 \end{cases}$$

- *Step 1:* We transform the system into augmented matrix:

$$\left[\begin{array}{cccc|c} 2 & 1 & 7 & -7 & 2 \\ -3 & 4 & -5 & -6 & 3 \\ 1 & 1 & 4 & -5 & 2 \end{array}\right]$$

- *Step 2:* Perform elementary row operations to get a REF

  Choose the first row as pivot row (the first nonzero entry is the pivot):

$$\left[\begin{array}{cccc|c} 2 & 1 & 7 & -7 & 2 \\ -3 & 4 & -5 & -6 & 3 \\ 1 & 1 & 4 & -5 & 2 \end{array}\right] \xrightarrow[R_2 \to 3R_1 + R_2,\ R_3 \to -2R_1 + R_3]{R_1 \leftrightarrow R_3}$$

$$\left[\begin{array}{cccc|c} 1 & 1 & 4 & -5 & 2 \\ 0 & 7 & 7 & -21 & 9 \\ 0 & -1 & -1 & 3 & -2 \end{array}\right]$$

Then we choose second row as pivot row to continue elimination:

$$\xrightarrow[R_2 \to -R_3,\ R_3 \to -7R_2 + R_3]{R_2 \leftrightarrow R_3} \left[\begin{array}{cccc|c} 1 & 1 & 4 & -5 & 2 \\ 0 & 1 & 1 & -3 & 2 \\ 0 & 0 & 0 & 0 & -5 \end{array}\right]$$

- *Step 3:* Perform elementary row operations to make it RREF (though not necessary for this example)

$$\left[\begin{array}{cccc|c} 1 & 0 & 3 & -2 & 0 \\ 0 & 1 & 1 & -3 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right] \tag{3.6}$$

You may find there is no solutions to this system of equation, which means Gaussian Elimination **doesn't** always ensure the system has a solution. ▪

We raise another example to explain further. Assume that a system of equations has an augmented matrix in reduced row echelon form as below:

$$\left[\begin{array}{ccc|c} 1 & 0 & 1 & 3 \\ 0 & 1 & -1 & 2 \\ 0 & 0 & 0 & 0 \end{array}\right] \tag{3.7}$$

We transform it back to a linear system:

$$\begin{cases} x_1 + x_3 = 3 \\ x_2 = 2 + x_3 \end{cases}$$

The last row is omitted. Then the solution set is: s = $\left\{ \begin{bmatrix} 3-t \\ 2+t \\ t \end{bmatrix}, t \in \boldsymbol{R} \right\}$ ,which has infinitely many solutions.

### Summary of Success/Failure for Square Systems.

- Check the last non-zero row of the augmented matrix in reduced row echelon form:
  - If it is of the form $(0, 0, \cdots, 0 \mid c)$, where $c$ is nonzero, then the system has no solution.
  - Otherwise, the system has at least one solution.
    * If it is of the form $(0, 0, \cdots, 0 \mid 0)$, then the system has infinitely many solutions.

**Remark**: Gaussian-Elimination is not only useful in solving linear systems. Other applications will be discussed further later in the course.

### Summary of Lecture 5.

## 3.2. Lecture 6: Matrix and Matrix Operations I

In the last lecture, we learned how to solve a general $m \times n$ linear system. Firstly, we form an augmented matrix. Secondly, we transform it into the row echelon form and the reduced row echelon form in turn by using elementary row operations. Finally, we turn it back to the linear system and get the final solution. Notice that the number of the solution can be unique, infinite or even zero depending on different situations.

R **Time Complexity Analysis for GE**

Gaussian Elimination is a **systematic** approach to linear systems. For relatively large linear systems, we often solve them by computers. Normally, for a rectangular $n \times n$ linear system:

- The computer use GE to solve the linear system.
- The time complexity of GE is $O(n^3)$, i.e. the GE algorithm uses $cn^3$ $(c > 0)$ arithmetic operations(e.g., division, multiplication, addition, subtraction) to reduce an augmented matrix to RREF.

More generally, for an $m \times n$ linear system, we choose $\bar{n} = \max(m, n)$ and replace $n$ by $\bar{n}$ in the expressions above.

### 3.2.1. Basic Matrices Operations

Firstly, we recall some conventions we have learned in 2.2.1.:

- For a matrix $\boldsymbol{A}$, $a_{ij}$ is called the $(i,j)$-th entry of $\boldsymbol{A}$, and we often denote $\boldsymbol{A}$ as $(a_{ij})_{m\times n}$.
- Matrices are often denoted by capital letter $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, ...
- When $m = n$, $\boldsymbol{A}$ is called a square matrix.
- When all entries are zero, $\boldsymbol{A}$ is called a zero matrix.
- when $m = 1$, $\boldsymbol{A}$ becomes a row vector.
- when $n = 1$, $\boldsymbol{A}$ becomes a column vector.

- when $m = n = 1$, $\boldsymbol{A}$ can be considered as a scalar.

Here we add something new to notate certain conventions:

- Column of a matrix $a_j = \begin{bmatrix} \boldsymbol{A}_{1j} \\ \vdots \\ \boldsymbol{A}_{mj} \end{bmatrix}$.
- Row of a matrix $a^{(i)} = \begin{bmatrix} \boldsymbol{A}_{i1} & \cdots & \boldsymbol{A}_{in} \end{bmatrix}$.

These two conventions are similar to the conventions for an augmented matrix mentioned before. We now give new definitions to some simple operations.

**Definition 3.5** [Matrix Equality] Let $\boldsymbol{A}_{m \times n}$ and $\boldsymbol{B}_{m \times n}$. $\boldsymbol{A} = \boldsymbol{B}$ means that $a_{ij} = b_{ij}$, for every $i = 1, \ldots, m$, $j = 1, \ldots, n$. ▪

**Definition 3.6** [Matrix Addition] Let $\boldsymbol{A}_{m \times n}$ and $\boldsymbol{B}_{m \times n}$. $\boldsymbol{C}_{m \times n} \triangleq \boldsymbol{A} + \boldsymbol{B}$ (addition of two matrices), where $\boldsymbol{C} = (c_{ij})_{m \times n}$ with entries $c_{ij} = a_{ij} + b_{ij}$, for every $i = 1, 2, \ldots, m$. ▪

**Definition 3.7** [Scalar Multiplication] Let $\boldsymbol{A}_{m \times n}$, and $\alpha$ be any real or complex number ($\alpha$ in $\mathbb{R}$ or $\mathbb{C}$). $\boldsymbol{D}_{m \times n} \triangleq \alpha \boldsymbol{A}$ (called scalar multiplication) with entries $d_{ij} = \alpha a_{ij}$, for every $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, m$. ▪

We have given definitions to some specific matrices. It should be noted that some matrices definitions focus on $n \times n$ matrices. For instance, an $m \times n$ zero matrix may not be a diagonal matrix if $m \neq n$. Instead, an $n \times n$ zero matrix is a diagonal matrix though all entries on the diagonal are zero.

**Definition 3.8** [Binary Matrix] Let $\boldsymbol{A}_{m \times n}$. The matrix $\boldsymbol{A}$ is a binary matrix if $a_{ij} = 0$ or $1$ for every $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, m$. ▪

**Definition 3.9** [Real / Complex Matrix] Let $\boldsymbol{A}_{m \times n}$. The matrix $\boldsymbol{A}$ is a real matrix if $a_{ij} \in \mathbb{R}$ for every $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, m$. The matrix $\boldsymbol{A}$ is a complex matrix if $a_{ij} \in \mathbb{C}$ for every $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, m$. ▪

**Definition 3.10** [Set of Matrices]

$$\mathbb{R}^{m\times n} = \{m \times n \ Matrix| \ entries \in \mathbb{R}\}$$

$$\mathbb{C}^{m\times n} = \{m \times n \ Matrix| \ entries \in \mathbb{C}\}$$

Given an $n \times n$ matrix $\boldsymbol{A}$, $\boldsymbol{A}$ is called a **square matrix.**

**Definition 3.11** [Set of Column Vectors]

$$\mathbb{R} = \mathbb{R}^{m\times 1} = \{m \times 1 \ Matrix| \ entries \in \mathbb{R}\}$$

$$\mathbb{C} = \mathbb{C}^{m\times 1} = \{m \times 1 \ Matrix| \ entries \in \mathbb{C}\}$$

R **Examples of Real-World Matrices**

- **Binary Image**

  A binary image can be represented by an $m \times n$ binary matrix. The values of black pixels are 0 (false), while the values of white pixels are 1 (true).

- **Grayscale**

  A grayscale image m pixels tall and n pixels wide is represented as a $m \times n$ matrix of double datatype. Element values denote the pixel grayscale intensities in $[0,1]$ with $0 =$ black and $1 =$ white.

- **Truecolor RGB**

  A truecolor red-green-blue (RGB) image is represented as a three-dimensional $m \times n \times 3$ matrix of double datatype. Each pixel has red, green, blue components along the third dimension with values in $[0,1]$.

- **Labeled Graph**

  A labeled graph means a graph having finite numbers of vertices with edges between some of them, representing certain kinds of relationships. We could use adjacency matrix to refer such relationships.

- **Power Network**

  A power network use admittance matrix $Y$ with the vector of the voltage $V$ to calculate the current $I = YV$ in the form of vector.

**Theorem 3.1 — Properties of Matrices Operations.** Let $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C} \in \mathbb{R}^{m \times n}$, $\alpha, \beta \in \mathbb{R}$.

(1) $\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{B} + \boldsymbol{A}$.

(2) $(\boldsymbol{A} + \boldsymbol{B}) + \boldsymbol{C} = \boldsymbol{A} + (\boldsymbol{B} + \boldsymbol{C})$.

(We can therefore use the notation $\boldsymbol{A} + \boldsymbol{B} + \boldsymbol{C}$.)

(3) $(\alpha\beta)\boldsymbol{A} = \alpha(\beta\boldsymbol{A})$.

(4) $\alpha(\boldsymbol{A} + \boldsymbol{B}) = \alpha\boldsymbol{A} + \alpha\boldsymbol{B}$.

(5) $(\alpha + \beta)\boldsymbol{A} = \alpha\boldsymbol{A} + \beta\boldsymbol{A}$

We give a brief proof for (4) as an example. We have

$$\alpha(\boldsymbol{A} + \boldsymbol{B}) = \begin{bmatrix} \alpha(a_{11} + b_{11}) & \cdots & \alpha(a_{1n} + b_{1n}) \\ \vdots & \ddots & \vdots \\ \alpha(a_{m1} + b_{m1}) & \cdots & \alpha(a_{mn} + b_{mn}) \end{bmatrix}.$$

We also have $\alpha\boldsymbol{A} = \begin{bmatrix} \alpha a_{11} & \cdots & \alpha a_{1n} \\ \vdots & \ddots & \vdots \\ \alpha a_{m1} & \cdots & \alpha a_{mn} \end{bmatrix}$ and $\alpha\boldsymbol{B} = \begin{bmatrix} \alpha b_{11} & \cdots & \alpha b_{1n} \\ \vdots & \ddots & \vdots \\ \alpha b_{m1} & \cdots & \alpha b_{mn} \end{bmatrix}$. According to the definition of matrix addition, we infer that $\alpha(\boldsymbol{A} + \boldsymbol{B}) = \alpha\boldsymbol{A} + \alpha\boldsymbol{B}$.

## 3.2.2. Matrix Multiplied by Vector

We start the learning by a movie preference example. Here is a table of movie preference for three movies filled by three members.

| | **Action Film** | **Hollywood** | **Comedy** | **Total Score** |
|---|---|---|---|---|
| **Weight** | $w_1$ | $w_2$ | $w_3$ | |
| **Value (M1)** | 5 | 5 | 5 | 5.0 |
| **Value (M2)** | 10 | 8 | 2 | 8.8 |
| **Value (M3)** | 3 | 7 | 8 | 4.3 |

The total score is obtained by adding the products of weight and its relevant value together. For instance, The total score of value (M1) $5.0 = 5w_1 + 5w_2 + 5w_3$. To solve the unknown variables **weight**, We can easily create a linear system of the table. Instead of writing a group of linear equations, can we express those equations in a more compact way? Here we introduce another way.

Firstly, we set a matrix $\boldsymbol{F}$ to represent features in the table, $\boldsymbol{F} = \begin{bmatrix} 5 & 5 & 5 \\ 10 & 8 & 2 \\ 3 & 7 & 8 \end{bmatrix}$. The variables and scores can be written in vectors, thus we let $\boldsymbol{w} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}$ and $\boldsymbol{s} = \begin{bmatrix} 5.0 \\ 8.8 \\ 4.3 \end{bmatrix}$. If we define

$$\boldsymbol{F}\boldsymbol{w} = \begin{bmatrix} 5 & 5 & 5 \\ 10 & 8 & 2 \\ 3 & 7 & 8 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 5w_1 + 5w_2 + 5w_3 \\ 10w_1 + 8w_2 + 2w_3 \\ 3w_1 + 7w_2 + 8w_3 \end{bmatrix},$$

we could figure out the equation $\boldsymbol{F}\boldsymbol{w} = \boldsymbol{s}$ holds. Hence, this matrix equation is equivalent to the linear system. We use such kind of matrix-vector product to express a linear system expediently. Alternatively, this equation can be written in an inner product way:

$$\boldsymbol{F}\boldsymbol{w} = \begin{bmatrix} \langle f^{(1)}, \boldsymbol{w} \rangle \\ \langle f^{(2)}, \boldsymbol{w} \rangle \\ \langle f^{(3)}, \boldsymbol{w} \rangle \end{bmatrix}.$$

Equivalently, we can also use the linear combination of matrix columns to denote the linear system.

$$\boldsymbol{s} = w_1 \begin{bmatrix} 5 \\ 10 \\ 3 \end{bmatrix} + w_2 \begin{bmatrix} 5 \\ 8 \\ 7 \end{bmatrix} + w_3 \begin{bmatrix} 5 \\ 2 \\ 8 \end{bmatrix}.$$

Let's study another example for matrix multiplied by a vector:

**Example 3.4** For the system of equations $\begin{cases} 2x_1 + x_2 + x_3 = 5 \\ 4x_1 - 6x_2 = -2 \\ -2x_2 + 7x_2 + 2x_3 = 9 \end{cases}$, we define

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad \boldsymbol{A} = \begin{pmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{pmatrix} = \begin{pmatrix} a^{(1)} \\ a^{(2)} \\ a^{(3)} \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} 5 \\ -2 \\ 9 \end{pmatrix}.$$

Here $\boldsymbol{x}$ and $a_1, a_2, a_3$ are all vectors. More specifically,

$$a^{(1)} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}, \quad a^{(2)} = \begin{pmatrix} 4 \\ -6 \\ 0 \end{pmatrix}, \quad a^{(3)} = \begin{pmatrix} -2 \\ 7 \\ 2 \end{pmatrix}.$$

Then we multiply matrix $\boldsymbol{A}$ with vector $\boldsymbol{x}$:

$$\boldsymbol{A}\boldsymbol{x} = \begin{pmatrix} 2x_1 + x_2 + x_3 \\ 4x_1 - 6x_2 \\ -2x_1 + 7x_2 + 2x_3 \end{pmatrix} = \begin{pmatrix} \langle a^{(1)}, \boldsymbol{x} \rangle \\ \langle a^{(2)}, \boldsymbol{x} \rangle \\ \langle a^{(3)}, \boldsymbol{x} \rangle \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

Hence we finally write the system equation as:

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \quad \text{Compact Matrix Form}$$

Also, if we regard $\boldsymbol{x}$ as a scalar, we can also write:

$$\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x} = \begin{pmatrix} a^{(1)} \\ a^{(2)} \\ a^{(3)} \end{pmatrix} \boldsymbol{x} = \begin{pmatrix} a^{(1)}\boldsymbol{x} \\ a^{(2)}\boldsymbol{x} \\ a^{(3)}\boldsymbol{x} \end{pmatrix}$$

Consider a general linear system:

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

The matrix representation of it is $\boldsymbol{Ax} = \boldsymbol{b}$, where we define $\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$, called the coefficient matrix, $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$, called the variable vector, and $\boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$, which is the solution set of the system. Equivalently, it can be denoted as a linear combination of matrix columns:

$$\boldsymbol{b} = x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix} = \begin{bmatrix} \langle a^{(1)}, \boldsymbol{x} \rangle \\ \langle a^{(2)}, \boldsymbol{x} \rangle \\ \vdots \\ \langle a^{(m)}, \boldsymbol{x} \rangle \end{bmatrix}$$

### 3.2.3. Matrix Multiply Matrix

We have already learned matrix-vector product. But what if each row in the matrix has different vector to multiply? Again, we use the movie preference example above to show such situation. If there is more than one set of weights, the total score is impossible to express in only one equation so far. Hence we use several matrix-vector products to show the result:

$$\begin{bmatrix} f_{11} & f_{12} & \cdots & f_{1n} \\ f_{21} & f_{22} & \cdots & f_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ f_{m1} & f_{m2} & \cdots & f_{mn} \end{bmatrix} \begin{bmatrix} w_{1k} \\ w_{2k} \\ \vdots \\ w_{nk} \end{bmatrix} = \begin{bmatrix} s_{1k} \\ s_{2k} \\ \vdots \\ s_{nk} \end{bmatrix},$$

where k is every positive integer that is smaller or equal to the number of sets of weights. It is a tedious and time-consuming process to write down every equations. Here we focus on a new definition of matrix product, which is a combination of matrix multiplied by several vectors.

**Definition 3.12** [Matrix Product] Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{B} = [\boldsymbol{b}_1, \boldsymbol{b}_2, \dots, \boldsymbol{b}_r] \in \mathbb{R}^{n\times r}$, then the matrix product of $\boldsymbol{A}$ by $\boldsymbol{B}$ is a $m \times r$ matrix defined by

$$\boldsymbol{AB} = [\boldsymbol{A}b_1, \boldsymbol{A}b_2, \dots, \boldsymbol{A}b_r].$$

It can also be represented as $\boldsymbol{C} := (c_{ij})_{m\times r}$, where $c_{ij} = a^{(i)} b_j = \sum_{k=1}^{n} a_{ik} b_{kj}$.

**Example 3.5** We want to calculate the result for $m \times n$ matrix $\boldsymbol{A}$ multiply $n \times k$ matrix $\boldsymbol{B}$, which is written as

$$\boldsymbol{AB} = \boldsymbol{C} = \begin{pmatrix} \boldsymbol{A}b_1 & \boldsymbol{A}b_2 & \dots & \boldsymbol{A}b_k \end{pmatrix}$$

Hence the $i$th row, $j$th column of $\boldsymbol{C}$ is given by

$$c_{ij} = \sum_{l=1}^{n} a_{il} b_{lj} = \langle a_i, b_j \rangle$$

Notice that the $ij$th entry of $\boldsymbol{C} = \boldsymbol{AB}$ is the inner product of the $i$th row of $\boldsymbol{A}$ and the $j$th row of $\boldsymbol{B}$.

(R) Here are a few notes about matrix product:

- An $m \times n$ matrix $\boldsymbol{A}$ can be written as $\begin{bmatrix} a_{ij} \end{bmatrix}$, where $a_{ij}$ denotes the entry of $i$th row, $j$th column of $\boldsymbol{A}$.
- Matrix $\boldsymbol{A}$ and $\boldsymbol{B}$ can do multiplication operator if and only if **the # for column of $\boldsymbol{A}$ equal to the # for row of $\boldsymbol{B}$.** Moreover, for $m \times n$ matrix $\boldsymbol{A}$ and $n \times k$ matrix $\boldsymbol{B}$, we can do multiplication as follows:

$$\boldsymbol{AB} = \boldsymbol{A} \begin{pmatrix} b_1 & b_2 & \dots & b_k \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}b_1 & \boldsymbol{A}b_2 & \dots & \boldsymbol{A}b_k \end{pmatrix}$$

  The result is a $m \times k$ matrix. Thus for matrix multiplication, it suffices to calculate matrix multiplied by vectors.
- Matrix product is a natural generalization of the matrix-vector product.

**R** **Time Complexity Analysis for matrix product**

- To Calculate the single entry of $\boldsymbol{C}$, you need to do $n$ times multiplication.
- There exists $n^2$ entries in $\boldsymbol{C}$
- Hence it takes $n \times n^2 \sim O(n^3)$ operations to compute $\boldsymbol{C}$. (Moreover, using more advanced algorithm, the time complexity could be reduced.)

**Preview of the next lecture.** In the next lecture, we will show more details about matrix product, as well as its properties and applications.

# Chapter 4

# Week4

## 4.1. Lecture 7: Matrix Operations 2

In the last lecture, we learned some basic knowledge of matrix operations.In this lecture, we will learn the transpose matrix, get deeper into the properties of matrix multiplication, and study block partition.

### 4.1.1. Transpose

**Definition 4.1** [Transpose] Let $\boldsymbol{A} = (a_{ij})_{m\times n}$,then the transpose of $\boldsymbol{A}$ is the matrix $\boldsymbol{B} = (b_{ij})_{n\times m}$,where

$$b_{ji} = a_{ij}(i = 1,\cdots,m, j = 1,\cdots,n)$$

Notation: $\boldsymbol{B}$=$\boldsymbol{A}^\top$

#### 4.1.1.1. Properties of Transpose Matrix

Let $\boldsymbol{A}$,$\boldsymbol{B} \in R^{m\times n}$, $\alpha \in$ R, then

1. $(\boldsymbol{A}+\boldsymbol{B})^\top = \boldsymbol{A}^\top + \boldsymbol{B}^\top$
2. $(\alpha\boldsymbol{A})^\top = \alpha\boldsymbol{A}^\top$
3. $(\boldsymbol{A}^\top)^\top$=$\boldsymbol{A}$
4. $(\boldsymbol{A}\boldsymbol{B})^\top = \boldsymbol{B}^\top\boldsymbol{A}^\top$

R Generalize:$(\boldsymbol{A}_1\boldsymbol{A}_2\cdots\boldsymbol{A}_n)^\top = \boldsymbol{A_n}^\top\boldsymbol{A_{n-1}}^\top\cdots\boldsymbol{A_1}^\top$

### 4.1.1.2. Symmetric Matrix

**Definition 4.2** [Symmetric Matrix] If a matrix $\boldsymbol{A} \in R^{m\times n}$ satisfies $\boldsymbol{A}=\boldsymbol{A}^\top$ we call $\boldsymbol{A}$ is symmetric.

## 4.1.2. Matrix Multiplication

### 4.1.2.1. vector-vector products

1. Vector Outer Product

**Definition 4.3** [Outer Product of Two Vectors]

Let $\boldsymbol{x}$ be the column vectors (the length of x is m,and $\overrightarrow{\boldsymbol{y}}$ is the row vector with the length n,the product $\boldsymbol{x}\ \overrightarrow{\boldsymbol{y}}$ (called ***outerproduct*** will result in a matrix.)

$$\boldsymbol{x}\overrightarrow{\boldsymbol{y}} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} [y_1, y_2, \cdots, y_n] = \begin{bmatrix} x_1y_1 & x_1y_2 & \cdots & x1y_n \\ x_2y_1 & x_2y_2 & \cdots & x2y_n \\ \vdots & & & \\ x_my_1 & x_my_2 & \cdots & x_my_n \end{bmatrix}$$

2. Vector Inner Product

   Recall: three forms of inner product.

   - Form-1: $< \boldsymbol{u}, \boldsymbol{v} >$
   - Form-2: $\boldsymbol{u}^\top \boldsymbol{v}$
   - Form-3: $\boldsymbol{u} \cdot \boldsymbol{v}$

R

(a) Attention : Only Form-1 and Form-2 are consistent with matrix multiplication.

(b) Convention : We often suggest column form of vectors.

(c) Principle : Dimension Match. For example:$\boldsymbol{u}^\top \boldsymbol{v}$ is valid, while $\boldsymbol{v}\boldsymbol{u}$ is invalid.

### 4.1.2.2. How to compute matrix multiplication quickly?

Given $m \times n$ matrix $\boldsymbol{A}$ and $n \times k$ matrix $\boldsymbol{B}$, then the result of $\boldsymbol{AB}$ should be a $m \times k$ matrix.

Let's show a specific example:

**Example 4.1** Given $4 \times 3$ matrix $\boldsymbol{A}$ and $3 \times 2$ matrix $\boldsymbol{B}$, then the result of $\boldsymbol{AB}$ should be a $4 \times 2$ matrix:

$$\boldsymbol{AB} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \times & \times \\ \times & \times \\ \times & \times \\ \times & \times \end{bmatrix}_{4\times 2}.$$

- The $(i,j)$th entry of the result should be the **inner product** between the $i$th row of $\boldsymbol{A}$ and the $j$th column of $\boldsymbol{B}$.

  Since the result has $4 \times 2$ entries, we have to process such progress $4 \times 2$ times to obtain the final result.

- But we can try a more effecient method. We can calculate the *entire row* of the result more easily.

  – For example, note that

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

  **The first row of the result is the linear combination of the row of matrix $\boldsymbol{B}$, and the coefficients are entries of the first row of matrix $\boldsymbol{A}$:**

  $$\begin{bmatrix} 1 \end{bmatrix} \times \begin{bmatrix} 1 & 1 \end{bmatrix} + \begin{bmatrix} 2 \end{bmatrix} \times \begin{bmatrix} 1 & 1 \end{bmatrix} + \begin{bmatrix} 3 \end{bmatrix} \times \begin{bmatrix} 1 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 1 \end{bmatrix}.$$

  – On the other hand, we can also calculate the *entire column* of the result quickly:

  $$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 6 & \times \\ 15 & \times \\ 24 & \times \\ 33 & \times \end{bmatrix}.$$

  **The first column of the result is the linear combination of the column of matrix**

$A$, and the coefficients are entries of the first column of matrix $B$:

$$\begin{bmatrix} 1 \\ 4 \\ 7 \\ 10 \end{bmatrix} \begin{bmatrix} 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 5 \\ 8 \\ 11 \end{bmatrix} \begin{bmatrix} 1 \end{bmatrix} + \begin{bmatrix} 3 \\ 6 \\ 9 \\ 12 \end{bmatrix} \begin{bmatrix} 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 15 \\ 24 \\ 33 \end{bmatrix}.$$

You can do the remaining calculation by yourself, and the final result is given by:

$$AB = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{bmatrix}_{4\times 3} \times \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 0 \end{bmatrix}_{3\times 2} = \begin{bmatrix} 6 & 1 \\ 15 & 4 \\ 24 & 7 \\ 33 & 10 \end{bmatrix}_{4\times 2}.$$

### 4.1.2.3. Properties of Matrix Multiplication

**Definition 4.4** [Identity Matrix] The $n \times n$ identity matrix is the matrix $I = [m_{ij}]$, where

$$m_{ij} = \begin{cases} 1, & \text{if } i = j; \\ 0, & \text{if } i \neq j. \end{cases}$$

**Proposition 4.1** Identity Matrix has the following properties:

$$IB = B, \quad AI = A,$$

where $A$ and $B$ could be any size-suitable matrix.

Suppose A,B,C are matrices with proper dimensions, $\alpha \in R$.Operations on matrix has the following properties:

1. $A(B + C) = AB + AC$.
2. $(B + C)A = BA + CA$.
3. $AB \neq BA$, i.e., $AB$ doesn't *necessarily* equal to $BA$.
4. $(AB)C = A(BC)$.

5. $\alpha(\boldsymbol{AB}) = (\alpha\boldsymbol{A})\boldsymbol{B} = \boldsymbol{A}(\alpha\boldsymbol{B})$
6. $\boldsymbol{A}^p = \boldsymbol{AAA}\cdots\ \boldsymbol{A}(p factors)$

   $(\boldsymbol{A}^p)(\boldsymbol{A}^q) = \boldsymbol{A}^{p+q}$

   $(\boldsymbol{A}^p)^q = (\boldsymbol{A}^{pq})$

   R $A^2$ is NOT a valid expression for non-square matrix A.
   Valid products: $(\mathbf{A}\mathbf{A}^\top)^n$ and $(\mathbf{A}^\top\mathbf{A})^n$.
   Useful in:Optimization, machine learning,etc.

R In some special cases, $\boldsymbol{AB}$ may equal to $\boldsymbol{BA}$. For example, for elementary matrix, we have $\boldsymbol{E}_{21}\boldsymbol{E}_{31} = \boldsymbol{E}_{31}\boldsymbol{E}_{21}$, this means the order of row operation can be changed sometimes.

However, for most cases the equality is not satisfied. given row vector $\boldsymbol{a} = \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix}$ and column vector $\boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$, the result of $\boldsymbol{ab}$ and $\boldsymbol{ba}$ is given by:

$$\boldsymbol{ab} = \begin{pmatrix} a_1 & a_2 & a_3 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = a_1b_1 + a_2b_2 + a_3b_3$$

$$\boldsymbol{ba} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \begin{pmatrix} a_1 & a_2 & a_3 \end{pmatrix} = \begin{pmatrix} b_1a_1 & b_1a_2 & b_1a_3 \\ b_2a_1 & b_2a_2 & b_2a_3 \\ b_3a_1 & b_3a_2 & b_3a_3 \end{pmatrix}.$$

### 4.1.3. Block Partition

.

#### 4.1.3.1. Matrix Partition

**Definition 4.5** [Matrix Partition] The matrix A=

$$\begin{bmatrix} \boldsymbol{A}_{11} & \cdots & \boldsymbol{A}_{1t} \\ \vdots & \ddots & \vdots \\ \boldsymbol{A}_{s1} & \cdots & \boldsymbol{A}_{st} \end{bmatrix}$$

is a partition of matrix with s× t blocks if the matrices $\boldsymbol{A}_{ij}$ satisfies

(1) For each fixed i,the number of rows of all $\boldsymbol{A}_{ij}$ are equal.

(2) For each fixed j,the number of colomns of all $\boldsymbol{A}_{ij}$ are equal.

The matrix $\boldsymbol{A}_{ij}$ is called the (i,j)-block of $\boldsymbol{A}$.

The general application of block matrix:

(1) Augumented matrices: [A | b]

(2) Express a matrix by vectors:

$$\begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_m \end{bmatrix} or \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \end{bmatrix}$$

(3) Sometimes, matrix partition makes finding the inverse(which will be mentioned in future lectures) and transpose of matrix easy.For example,

**Example 4.2** Block-Diagonal Matrix.

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & 0 & \cdots & 0 \\ 0 & \boldsymbol{A}_{22} & & \\ \vdots & & \ddots & \\ 0 & & & \boldsymbol{A}_{nn} \end{bmatrix}$$

The diagonal blocks are square matrices,the transpose of $\boldsymbol{A}$ is

$$\begin{bmatrix} \boldsymbol{A_{11}}^\top & 0 & \cdots & 0 \\ 0 & \boldsymbol{A_{22}}^\top & & \\ \vdots & & \ddots & \\ 0 & & & \boldsymbol{A_{nn}}^\top \end{bmatrix}$$

## 4.1.3.2. Block Multiplication

. We use an example to show the process of block multiplicaion:

**Example 4.3** Given two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, we want to compute $\boldsymbol{C} := \boldsymbol{A} \times \boldsymbol{B}$, which can be done by **block multiplication**. We can partition $\boldsymbol{A}$ and $\boldsymbol{B}$ with appropriate sizes. For example,

$$\boldsymbol{A} = \left[\begin{array}{cc|c} 4 & 0 & 4 \\ 6 & 6 & 8 \\ \hline -9 & 5 & -8 \end{array}\right] = \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{A}_2 \\ \boldsymbol{A}_3 & \boldsymbol{A}_4 \end{bmatrix}, \quad \boldsymbol{B} = \left[\begin{array}{cc|c} 8 & -3 & -7 \\ 3 & -7 & -4 \\ \hline 4 & -4 & 1 \end{array}\right] = \begin{bmatrix} \boldsymbol{B}_1 & \boldsymbol{B}_2 \\ \boldsymbol{B}_3 & \boldsymbol{B}_4 \end{bmatrix}.$$

Then $\boldsymbol{A}$ and $\boldsymbol{B}$ could be considered as $2 \times 2$ **block matrices**. As a result, $\boldsymbol{C}$ have $2 \times 2$ blocks:

$$\boldsymbol{AB} = \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{A}_2 \\ \boldsymbol{A}_3 & \boldsymbol{A}_4 \end{bmatrix} \begin{bmatrix} \boldsymbol{B}_1 & \boldsymbol{B}_2 \\ \boldsymbol{B}_3 & \boldsymbol{B}_4 \end{bmatrix} = \begin{bmatrix} \boldsymbol{A_1B_1} + \boldsymbol{A_2B_3} & \boldsymbol{A_1B_2} + \boldsymbol{A_2B_4} \\ \boldsymbol{A_3B_1} + \boldsymbol{A_4B_3} & \boldsymbol{A_3B_2} + \boldsymbol{A_4B_4} \end{bmatrix} = \begin{bmatrix} \boldsymbol{C}_1 & \boldsymbol{C}_2 \\ \boldsymbol{C}_3 & \boldsymbol{C}_4 \end{bmatrix}$$

As a result, there is an effective way to calculate $\boldsymbol{C}_1$, that is the block multiplication method shown below:

$$\boldsymbol{C}_1 = \boldsymbol{A_1B_1} + \boldsymbol{A_2B_3} = \begin{bmatrix} 4 & 0 \\ 6 & 6 \end{bmatrix} \begin{bmatrix} 8 & -3 \\ 3 & -7 \end{bmatrix} + \begin{bmatrix} 4 \\ -8 \end{bmatrix} \begin{bmatrix} 4 & -4 \end{bmatrix} = \begin{bmatrix} 48 & -28 \\ 34 & -28 \end{bmatrix}.$$

You can do the remaining calculation to get result of $\boldsymbol{AB}$:

$$\boldsymbol{AB} = \boldsymbol{C} = \left[\begin{array}{cc|c} 48 & -28 & -24 \\ 34 & -28 & -74 \\ \hline -89 & 24 & 35 \end{array}\right].$$

Trick:

- Diagonal matrix left-multiply a matrix.

$$A = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix}, B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}, AB = \begin{bmatrix} A_{11}B_{11} & A_{11}B_{12} \\ A_{22}B_{21} & A_{22}B_{21} \end{bmatrix}$$

  conclusion:$A_{11}$ times 1st row of $B$. $A_{22}$ times 2nd row of $B$.

- Diagonal matrix right-multiply a matrix.

$$A = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix}, B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}, BA = \begin{bmatrix} B_{11}A_{11} & B_{12}A_{22} \\ B_{21}A_{11} & B_{21}A_{22} \end{bmatrix}$$

  conclusion:$A_{11}$ times 1st column of $B$. $A_{22}$ times 2nd column of $B$.

R When $A$,$B$ expand to nxn matrix,the conclusion will also establish.

There are also two useful ways to compute $AB$:

- If $B$ has $k$ columns, we can partition $B$ into $k$ blocks to compute $AB$:

$$AB = A \times \left[ B_1 \mid B_2 \mid \dots \mid B_k \right] = \left[ AB_1 \mid AB_2 \mid \dots \mid AB_k \right].$$

- If $A$ has $m$ rows, we can partition $A$ into $m$ blocks to compute $AB$:

$$AB = \begin{bmatrix} A_1 \\ \hline A_2 \\ \hline \dots \\ \hline A_m \end{bmatrix} \times B = \begin{bmatrix} A_1B \\ \hline A_2B \\ \hline \dots \\ \hline A_mB \end{bmatrix}$$

# 4.2. Lecture 8: LU decomposition

## 4.2.1. Inverse Matrices

**Definition 4.6** [Inverse Matrix] Let $A$ be an $n \times n$ matrix, an $n \times n$ matrix $B$ is called a inverse of $A$ if $AB = BA = I_n$. A square matrix is called invertible if it has an inverse.

**Proposition 4.2** If $A$ is invertible, then $A$ has a unique inverse matrix.

*Proof.* Suppose $A$ has two inverse matrices $B$ and $B'$, then by definition we have $AB = I_n$, $B'A = I_n$. Thus, $B' = B'I_n = B'(AB) = (B'A)B = I_nB = B$. Therefore, $B = B'$, $A$ has a unique inverse. ∎

Since the inverse matrix is unique, we can denote the inverse matrix for $A$ as $A^{-1}$.

**Example**

1. $A = 2$ is a saclar matrix , $A^{-1} = \frac{1}{2}$.
2. $A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$ is invertible, $A = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{bmatrix}$.
3. Zero matrices is not invertible. Because every matrix multiplies zero matrix is zero matrix.

**Proposition 4.3** A diagonal marix $D$ =diag$\{d_1, d_2, ..., d_n\}$ is invertible if and only if $d_i \neq 0$ for $i = 1, 2, ..., n$. The inverse matrix (if exists) is $D^{-1}$ =diag$\{\frac{1}{d_1}, \frac{1}{d_2}, ..., \frac{1}{d_n}\}$

**Proposition 4.4** A triangular matrix $A = (a_{ij})_{n\times n}$ is invertible if and only if $a_{ii} \neq 0$ for $i = 1, 2, ..., n$. Moreover,

- The inverse of an upper triangular matrix (if exists) is upper triangular matrix.
- The inverse of a lower triangular matrix (if exists) is lower triangular matrix.

A proof for $2 \times 2$ upper triangular matrix:

$A = \begin{bmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{bmatrix}$ is invertible if and only if there exists a matrix $B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$ such that

$$\begin{bmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

this is equivalent to solve the linear system below,

$$\begin{cases} a_{11}b_{11} + a_{12}b_{21} = 1 & (1) \\ a_{11}b_{12} + a_{12}b_{22} = 0 & (2) \\ a_{22}b_{21} = 0 & (3) \\ a_{22}b_{22} = 1 & (4) \end{cases} \tag{4.1}$$

Assume the linear system is solvable, we can deduce $a_{22} \neq 0$ from equation (4). Equation (3) deduces $b_{21} = 0$. Take $b_{21} = 0$ into the equation (1) we deduce $a_{11}b_{11} = 1$. Thus $a_{11} \neq 0$. Conversely, if $a_{11}$ and $a_{22}$ are not zero, then this system has solution

$$\begin{cases} b_{11} = \dfrac{1}{a_{11}} \\ b_{12} = -\dfrac{a_{12}}{a_{11}a_{22}} \\ b_{21} = 0 \\ b_{22} = \dfrac{1}{a_{12}} \end{cases} \tag{4.2}$$

Thus, $A$ has an inverse $\begin{bmatrix} \frac{1}{a_{11}} & -\frac{a_{12}}{a_{11}a_{22}} \\ 0 & \frac{1}{a_{12}} \end{bmatrix}$, which is upper triangular.

**Proposition 4.5** Let $A$ be an invertible matrices, $\alpha$ is a non zero scalar,

- $(\alpha A)^{-1} = \frac{1}{\alpha}A^{-1}$.
- $(A^T)^{-1} = (A^{-1})^T$
- $(A^{-1})^{-1} = A$

**Theorem 4.1 — Inverse of Matrices Product.** Let $A$ and $B$ be $n \times n$ invertible matrices, then $AB$ is invertible and

$$(AB)^{-1} = B^{-1}A^{-1}$$

*Proof.* By assumption, $A^{=1}$ and $B^{-1}$ exist. Thus,

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

$$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}B = I$$

Hence, $(AB)^{-1} = B^{-1}A^{-1}$. ■

By induction, we prove the theorem below.

**Theorem 4.2 — Inverse of Matrices Product.** Let $A_1$, $A_2$,...,$A_k$ be $n \times n$ invertible matrices, then $A_1A_2...A_k$ is invertible and

$$(A_1A_2...A_k)^{-1} = A_k^{-1}A_{k-1}^{-1}...A_1^{-1}$$

### 4.2.2. LU Decomposition

After learning matrix multiplication, we should be familiar some basic results of matrix multiplication:

1. **Product of upper triangular matries is also an upper triangular matrix.**

$$\begin{bmatrix} \times & \times & \times & \times & \times \\ & \times & \times & \times & \times \\ & & \times & \times & \times \\ & 0 & & \times & \times \\ & & & & \times \end{bmatrix} \begin{bmatrix} \times & \times & \times & \times & \times \\ & \times & \times & \times & \times \\ & & \times & \times & \times \\ & 0 & & \times & \times \\ & & & & \times \end{bmatrix} = \begin{bmatrix} \times & \times & \times & \times & \times \\ & \times & \times & \times & \times \\ & & \times & \times & \times \\ & 0 & & \times & \times \\ & & & & \times \end{bmatrix}.$$

2. **Product of diagonal matrices is also a diagonal matrix.**

$$\begin{bmatrix} d_1 & & & \\ & d_2 & & 0 \\ & & \ddots & \\ 0 & & & \ddots & \\ & & & & d_n \end{bmatrix} \begin{bmatrix} e_1 & & & \\ & e_2 & & 0 \\ & & \ddots & \\ 0 & & & \ddots & \\ & & & & e_n \end{bmatrix} = \begin{bmatrix} d_1e_1 & & & \\ & d_2e_2 & & 0 \\ & & \ddots & \\ 0 & & & \ddots & \\ & & & & d_ne_n \end{bmatrix}.$$

Just like permutation matrix, there are also some intersting properties of elementary matrix:

**Proposition 4.6** The inverse of an elementary matrix is also an elementary matrix.

This proposition can be interpreted easily if you combine the elementary matrix with elementary row operations on matrices. For non zero scalar $a$:

- $\begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & a & & \\ & & & \ddots & \\ & & & & 1 \end{bmatrix}$ has inverse $\begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & \frac{1}{a} & & \\ & & & \ddots & \\ & & & & 1 \end{bmatrix}$.
  The reverse operation of multiplying a row with $a$ is multiplying a row with $\frac{1}{a}$.

- $\begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & \ddots & & & \\ & & a & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}$ has inverse $\begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & \ddots & & & \\ & & -a & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}$
  The reverse operation of adding $a\times$ row $i$ to row $j$ is adding $(-a)\times$ row $i$ to row $j$.

- $\begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & & 1 & & \\ & & & \ddots & & & \\ & & 1 & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}$ has inverse $\begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & & 1 & & \\ & & & \ddots & & & \\ & & 1 & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}$
  The reverse operation of interchanging row $i$ and row $j$ is interchanging row $i$ and row $j$ again.

**Example 4.4** $\boldsymbol{E}_{21} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ is an elementary matrix, the result of postmultiplying $\boldsymbol{E}_{21}$ for identity matrix is given by:

$$\boldsymbol{E}_{21}\boldsymbol{I} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

which has the same effect of adding $(-2)\times$ row 1 to row 2 of $\boldsymbol{I}$. How to get the identity matrix again? We just need to add $2\times$ row 1 to row 2 of $\boldsymbol{I}$, which could be viewed as postmultiply another elementary matrix for $\boldsymbol{I}$:

$$\overline{\boldsymbol{E}_{21}}(\boldsymbol{E}_{21}\boldsymbol{I}) = \overline{\boldsymbol{E}_{21}}\boldsymbol{E}_{21} = \overline{\boldsymbol{E}_{21}} \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \boldsymbol{I}.$$

Hence, $\overline{\boldsymbol{E}_{21}}$ is the inverse matrix of $\boldsymbol{E}_{21}$, which is also an elementary matrix. ■

The elementary matrix $\boldsymbol{E}_{ij}(i<j)$ is a lower triangular matrix; and $\boldsymbol{E}_{ij}(i>j)$ is an upper triangular matrix. Let's look at an example:

**Example 4.5** Let's try Gaussian Elimination for a matrix that is nonsingular. Here we use elementary matrix to describe row operation above the arrow (without row exchange):

$$\boldsymbol{A} = \begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix} \xRightarrow{\boldsymbol{E}_{21}} \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ -2 & 7 & 2 \end{bmatrix} \xRightarrow{\boldsymbol{E}_{31}} \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 8 & 3 \end{bmatrix} \xRightarrow{\boldsymbol{E}_{32}} \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 0 & 1 \end{bmatrix} = \boldsymbol{U}$$

In this process we have

$$\boldsymbol{E}_{21} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{E}_{31} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{E}_{32} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}.$$

Finally we convert $\boldsymbol{A}$ into an upper triangular matrix $\boldsymbol{U}$. Let's do the reverse of this process to find some interesting results:

$$\begin{gathered} \boldsymbol{E}_{32}\boldsymbol{E}_{31}\boldsymbol{E}_{21}\boldsymbol{A} = \boldsymbol{U} \\ \implies \boldsymbol{E}_{32}^{-1}\boldsymbol{E}_{32}\boldsymbol{E}_{31}\boldsymbol{E}_{21}\boldsymbol{A} = \boldsymbol{E}_{32}^{-1}\boldsymbol{U} \implies \boldsymbol{E}_{31}\boldsymbol{E}_{21}\boldsymbol{A} = \boldsymbol{E}_{32}^{-1}\boldsymbol{U} \\ \cdots \implies \boldsymbol{A} = \boldsymbol{E}_{21}^{-1}\boldsymbol{E}_{31}^{-1}\boldsymbol{E}_{32}^{-1}\boldsymbol{U} := \boldsymbol{L}\boldsymbol{U}, \end{gathered}$$

where $\boldsymbol{L} = \boldsymbol{E}_{21}^{-1}\boldsymbol{E}_{31}^{-1}\boldsymbol{E}_{32}^{-1}$, which is lower triangular matrix.

Hence, we successfully decompose matrix $\boldsymbol{A}$ into the multiplication of a lower triangular matrix $\boldsymbol{L}$ and a upper triangular matrix $\boldsymbol{U}$. ■

For general case, Let $\boldsymbol{A}$ be a nonsingular matrix, we can perform Gaussian Elimination on $\boldsymbol{A}$ to get a upper triangular matrix $\boldsymbol{U}$.

$$\boldsymbol{A} \longrightarrow \boldsymbol{A}_1 \longrightarrow \boldsymbol{A}_2 \longrightarrow \cdots \longrightarrow \boldsymbol{U}$$

Suppose $\boldsymbol{A}$ does not require the row exchange during the Gaussian Elimination (i.e. the Gaussian Elimination only involves Multiplication and Addition-then-Multiplication operations). Expressing the Gaussian Elimination as matrix multiplication, we get

$$\boldsymbol{E}_k\boldsymbol{E}_{k-1}\cdots\boldsymbol{E}_1\boldsymbol{A} = \boldsymbol{U}$$

$$\boldsymbol{A} = (\boldsymbol{E}_k\boldsymbol{E}_{k-1}\cdots\boldsymbol{E}_1)^{-1}\boldsymbol{U}$$

By assumption, $\boldsymbol{E}_1$, $\boldsymbol{E}_2$,..., $\boldsymbol{E}_k$ are all lower triangular matrices, so$(\boldsymbol{E}_k\boldsymbol{E}_{k-1}\cdots\boldsymbol{E}_1)^{-1}$ is also a lower triangular matrix. Denote $\boldsymbol{L} = (\boldsymbol{E}_k\boldsymbol{E}_{k-1}\cdots\boldsymbol{E}_1)^{-1}$, then

$$\boldsymbol{A} = \boldsymbol{L}\boldsymbol{U}$$

where $\boldsymbol{L}$ is a lower triangular matrix, and $\boldsymbol{U}$ is an upper triangular matrix.

**Definition 4.7** [LU Decomposition] In conclusion, we decompose matrix $\boldsymbol{A}$ into the form:

$$\begin{aligned} &\boldsymbol{A} = \boldsymbol{L}\boldsymbol{U} \\ &\text{where:} \quad \boldsymbol{L} \text{ is lower triangular matrix} \\ &\qquad\qquad\ \boldsymbol{U} \text{ is upper triangular matrix} \end{aligned}$$

This decomposition is called **LU decomposition**.

### 4.2.2.1. One Square System = Two Triangular Systems

When considering the *nonsingular* case without row exchanges, recall what we have done before this lecture:

> we are working on $\boldsymbol{A}$ and $\boldsymbol{b}$ in **one** equation $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$.

To somplify computation, we aim to deal with $\boldsymbol{A}$ and $\boldsymbol{b}$ in **separate** equations. The LU decomposition can help us do that:

1. **Decomposition:** By Gaussian elimination on matrix $\boldsymbol{A}$, we can decompose $\boldsymbol{A}$ into matrix multiplications: $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{U}$.
2. **Solve:** forward elimination on $\boldsymbol{b}$ using $\boldsymbol{L}$, then back substitution for $\boldsymbol{x}$ using $\boldsymbol{U}$.

R

**The detail of Solve process.**

(a) First, we apply forward elimination on $\boldsymbol{b}$. In other words, we are actually solving $\boldsymbol{L}\boldsymbol{y} = \boldsymbol{b}$ for $\boldsymbol{y}$.

(b) After getting $\boldsymbol{y}$, we then do back substitution for $\boldsymbol{x}$. In other words, we are actually solving $\boldsymbol{U}\boldsymbol{x} = \boldsymbol{y}$ for $\boldsymbol{x}$.

**One square system = Two triangular systems.** During this process, the original system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ is converted into two triangular systems:

**Forward and Backward** Solve $\boldsymbol{L}\boldsymbol{y} = \boldsymbol{b}$ and then solve $\boldsymbol{U}\boldsymbol{x} = \boldsymbol{y}$.

There is nothing new about those steps. This is exactly what we have done all the time. We are really solving the triangular system $\boldsymbol{L}\boldsymbol{y} = \boldsymbol{b}$ as elimination went forward. Then we use back substitution to produce $\boldsymbol{x}$. An example shows what we actually did:

**Example 4.6** Forward elimination on $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ will result in equation $\boldsymbol{U}\boldsymbol{x} = \boldsymbol{y}$:

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \iff \begin{cases} u + 2v = 5 \\ 4u + 9v = 21 \end{cases} \text{forward elimination implies} \begin{cases} u + 2v = 5 \\ v = 1 \end{cases} \iff \boldsymbol{U}\boldsymbol{x} = \boldsymbol{y}.$$

We could express such process into matrix form:

**LU Decomposition.** : We could decompose $\boldsymbol{A}$ into product of $\boldsymbol{L}$ and $\boldsymbol{U}$:

$$\boldsymbol{L} = \begin{bmatrix} 1 & 0 \\ 4 & 1 \end{bmatrix}, \quad \boldsymbol{U} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$$

$\boldsymbol{L}\boldsymbol{y} = \boldsymbol{b}$. In this system of equation, in oder to solve $\boldsymbol{y}$, we only need to multiply the inverse of $\boldsymbol{L}$ both sides:

$$\begin{bmatrix}1 & 0\\4 & 1\end{bmatrix} \times \boldsymbol{y} = \begin{bmatrix}5\\21\end{bmatrix} \implies \boldsymbol{y} = \boldsymbol{L}^{-1}\begin{bmatrix}5\\21\end{bmatrix} = \begin{bmatrix}1 & 0\\-4 & 1\end{bmatrix}\begin{bmatrix}5\\21\end{bmatrix} = \begin{bmatrix}5\\1\end{bmatrix}.$$

$\boldsymbol{U}\boldsymbol{x} = \boldsymbol{y}$. In this system of equation, in oder to solve $\boldsymbol{x}$, we only need to multiply the inverse of $\boldsymbol{U}$ both sides:

$$\begin{bmatrix}1 & 2\\0 & 1\end{bmatrix} \times \boldsymbol{x} = \begin{bmatrix}5\\1\end{bmatrix} \implies \boldsymbol{x} = \boldsymbol{U}^{-1}\begin{bmatrix}5\\1\end{bmatrix} = \begin{bmatrix}1 & -2\\0 & 1\end{bmatrix}\begin{bmatrix}5\\1\end{bmatrix} = \begin{bmatrix}3\\1\end{bmatrix}.$$

Both **Forward** and **Back substitution** has $O(n^2)$ time complexity.

### 4.2.3. LDU Decomposition

The aim of LDU decomposition is to let the diagonal entries of $\boldsymbol{U}$ and $\boldsymbol{L}$ to be **one**.

Suppose we have decomposed $\boldsymbol{A}$ into $\boldsymbol{L}\boldsymbol{U}$, where the upper triangular matrix $\boldsymbol{U}$ is given by:

$$\begin{bmatrix}d_1 & \times & \times & \times & \times\\ & d_2 & \times & \times & \times\\ & & d_3 & \times & \times\\ & 0 & & d_4 & \times\\ & & & & d_5\end{bmatrix}$$

If we want to set its diagonal entries of $\boldsymbol{U}$ to be all **one**, we just need to multiply a matrix $\boldsymbol{D}^{-1}$ that is given by:

$$\boldsymbol{D}^{-1} := \begin{bmatrix}d_1^{-1} & & & & \\ & d_2^{-1} & & 0 & \\ & & d_3^{-1} & & \\ & 0 & & d_4^{-1} & \\ & & & & d_5^{-1}\end{bmatrix} \implies \boldsymbol{D}^{-1}\boldsymbol{U} = \begin{bmatrix}1 & \times & \times & \times & \times\\ & 1 & \times & \times & \times\\ & & 1 & \times & \times\\ & 0 & & 1 & \times\\ & & & & 1\end{bmatrix}.$$

We can convert LU decomposition into LDU decomposition by simply adding the multiplying factor $DD^{-1}$:

$$A = LU = LDD^{-1}U = LD(D^{-1}U) = LD\hat{U},$$

where $\hat{U} = D^{-1}U$ is also an upper triangular matix.

Here $D$ is the inverse matrix of $D^{-1}$:

$$D = \begin{bmatrix} d_1 & & & & \\ & d_2 & & \text{0} & \\ & & d_3 & & \\ & \text{0} & & d_4 & \\ & & & & d_5 \end{bmatrix}$$

Note that the *diagonal* entries of $D$ are all **pivots values** of $U$.

Similarly, we can also proceed this step again to let *diagonal* entries of $L$ to be **one**.

**Definition 4.8** [LDU Decomposition] In conclusion, we decompose matrix $A$ into the form:

$$A = LDU$$

where: $L$ is lower triangular matrix with unit entries in diagonal

$D$ is diagonal matrix

$U$ is upper triangular matrix with unit entries in diagonal

This decomposition is called **LDU decomposition**.

Here is a property of LDU decomposition, the proof of which is omitted.

**Proposition 4.7** **LDU decomposition is unique to any matrix**. Let $L, L_1$ denote a lower triangular matrix, $D, D_1$ diagonal, and $U, U_1$ upper triangular.

If $A = LDU$, and also, $A = L_1D_1U_1$, then we have $L = L_1, D = D_1, U = U_1$.

## 4.2.4. LU Decomposition with Row Exchanges

How can we handle row exchange in our $LU$ decomposition?

Assume we are going to do Gaussian Elimination with matrix $A$ with row exchange.

- At first We can postmultiply some elementary matrices $E$ to get $EEEA$.
- Sometimes we need to multiply by $P_{ij}$ to do *row exchange* to continue Gaussian Elimina-

tion.

- So we may end our elimination with something like $PEEEEPEEEEPEEEEA$.
- If we can get all the elementary matrix $L$ together, we could convert them into one single $L$ that has the same effect as before.
- The key problem is that how can we get all the row exchange matrix $P$ out from the elementary matrices?

**Theorem 4.3** If $A$ is *nonsingular*, then there exists a permutation matrix $P$ such that $PA = LU$.

The proof is omitted.

R For the nonsingular matrix $A$ **without** row exchange, we can always decompose it as $A = LU$; but for the row exchange case, we have to postmultiply a specific permutation matrix to obtain such LU decomposition.

# Chapter 5

# Week5

## 5.1. Lecture 9: LU decomposition and Inverse

### 5.1.1. Review

#### 5.1.1.1. Elementary Matrix

**Elementary row operations.** There are three kinds of elementary row operations we learned before:

1. **[Interchange]** Swap the positions of two rows
2. **[Multiplication]** Multiply a row by a **non-zero** scalar
3. **[Addition]** Add to on row a scalar multiple of another

Note that the operations preserve solutions.

**Elementary Matrix.**

**Definition 5.1** [Elementary Matrix] The matrices corresponding to a single elementary row operation are called elementary matrices. ■

For a given matrix $A$, performing elementary row operation for $A$ is equivalent to premultiplying $A$ by the corresponding elementary matrix. Thus, elementary matrices are essential in performing Gaussian elimination.

We also introduced permutation matrix:

**Definition 5.2** [Permutation Matrix] A permutation matrix is a square matrix that exactly one entry of 1 in each row and each column and 0s elsewhere.

### 5.1.1.2. Matrix Inverse

The concept of matrix inverse is based on the identity matrix:

**Definition 5.3** [Inverse] Suppose $A \in \mathbb{R}^{n\times n}$. If a matrix $B \in \mathbb{R}^{n\times n}$ satisfies

$$AB = BA = I_n$$

then we say $B$ is an inverse of $A$; denoted as $B = A^{-1}$

Based on the definition of inverse, we also introduced the concepts of **Invertible** and **Singular** matrices. There are some related useful theorems:

1. **(Matrix inverse is unique)** Suppose the square matrix $A$ has an inverse, then $A^{-1}$ is unique.
2. **(Solutions of Linear Systems)** If $A$ is invertible, then the linear system has a unique solution $x = A^{-1}b$.

In this lecture, we will continue to learn the application of elementary matrices and matrix inverse.

## 5.1.2. Properties of Inverse

### Inverse of Diagonal Matrix.

1. A diagonal matrix is invertile iff $d_{ii} \neq 0$, $\forall i \in \{1,2,\cdots,n\}$
2. The inverse of the diagonal matrix (if exists) is

$$D^{-1} = \begin{pmatrix} \frac{1}{d_{11}} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{d_{nn}} \end{pmatrix} \tag{5.1}$$

**Remark:** consider $2\times 2$ matrix: $D = \begin{pmatrix} d_{11} & 0 \\ 0 & d_{22} \end{pmatrix}$, then $D$ is invertible $\iff d_{11}, d_{22} \neq 0$.

In this case, $D^{-1} = \begin{pmatrix} \frac{1}{d_{11}} & 0 \\ 0 & \frac{1}{d_{22}} \end{pmatrix}$

## Inverse of Triangular Matrices.

1. A triangular matrix is invertible iff $a_{ii} \neq 0$, $\forall i \in \{1, 2, \cdots, n\}$
2. The inverse of an upper triangular matrix (if exists) is upper triangular matrix
3. The inverse of a lower triangular matrix (if exists) is lower triangular matrix

## Inverse and some operations.

1. Inverse of a Matrix Multiplied by a Scalar:

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

2. Inverse of transpose:

$$(A^T)^{-1} = (A^{-1})^T$$

3. Inverse of inverse:

$$(A^{-1})^{-1} = A$$

## Inverse of a Matrix Products.

Suppose $A, B$ are $n \times n$ invertible matrices, then:

$$(AB)^{-1} = B^{-1}A^{-1}$$

*Proof.* The property can be verified by definition of matrix inverse.

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AA^{-1} = I_n$$
$$(B^{-1}A^{-1})AB = B^{-1}(A^{-1}A)B = B^{-1}B = I_n$$

∎

**Remark:** The property can be extended to $(A_1A_2 \cdots A_n)$ case:

$$(A_1A_2 \cdots A_k)^{-1} = A_k^{-1} \cdots A_2^{-1}A_1^{-1}$$

Here are some concrete examples:

- Type-I $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

- Type-II $\begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{3} \end{bmatrix}$

- Type-III $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 1 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix}$

### Inverse of Elementary Matrix.

**Theorem 5.1 — Inverse of Elementary Matrices.** The inverse of the elementary matrices corresponding to the reverse row operations and belong to the same type of elementary matrices.

1. $E^{-1}_{R_iR_j} = E_{R_iR_j}$, corresponding to the reverse row operation 1, i.e $R_i \leftrightarrow R_j$
2. $E^{-1}_{\alpha R_i} = E_{\frac{1}{\alpha}R_i}(\alpha \neq 0)$,corresponding to the reverse row operation 2, i.e $R \to \frac{1}{\alpha}R_i$
3. $E^{-1}_{\beta R_i + R_j} = E_{-\beta R_i + R_j}$, corresponding to the reverse row operation 3, i.e $R_j \to -\beta R_i + R_j$

**Remark:**

- Any elementary matrix is a square matrix
- Any elementary matrix is invertible

## 5.1.3. LU Decomposition

### 5.1.3.1. A = LU

**Claim:.** Suppose $A$ is square matrix. If there is **no row exchange** in the process of GE, then the matrix $A$ can be decomposed as

$$A = LU$$

where $L$ is lower triangular, $U$ is upper triangular.

**Example 5.1** Let's try Gaussian Elimination for a matrix that is nonsingular. Here we use elementary matrix to describile row operation above the arrow (without row exchange):

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{bmatrix} \xrightarrow{E_{21}} \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ -2 & 7 & 2 \end{bmatrix} \xrightarrow{E_{31}} \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 8 & 3 \end{bmatrix} \xrightarrow{E_{32}} \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 0 & 1 \end{bmatrix} = U$$

In this process we have

$$E_{21} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad E_{31} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \quad E_{32} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}.$$

Finally we convert $A$ into an upper triangular matrix $U$. Let's do the reverse of this process to find some interesting results:

$$\begin{aligned} E_{32}E_{31}E_{21}A &= U \\ \Longrightarrow E_{32}^{-1}E_{32}E_{31}E_{21}A = E_{32}^{-1}U &\Longrightarrow E_{31}E_{21}A = E_{32}^{-1}U \\ \cdots \Longrightarrow A &= E_{21}^{-1}E_{31}^{-1}E_{32}^{-1}U := LU, \end{aligned}$$

where $L = E_{21}^{-1}E_{31}^{-1}E_{32}^{-1}$, which is lower triangular matrix.

Hence, we successfully decompose matrix $A$ into the multiplication of a lower triangular matrix $L$ and a upper triangular matrix $U$.

For general case, Let $A$ be a nonsingular matrix, we can perform Gaussian Elimination on $A$ to get a upper triangular matrix $U$.

$$A \longrightarrow A_1 \longrightarrow A_2 \longrightarrow \cdots \longrightarrow U$$

Suppose $A$ **does not require the row exchang** during the Gaussian Elimination (i.e. the Gaussian Elimination only involves Multiplication and Addition-then-Multiplication operations). Expressing the Gaussian Elimination as matrix multiplication, we get

$$E_kE_{k-1}\cdots E_1A = U$$

$$A = (E_k E_{k-1} \cdots E_1)^{-1} U$$

By assumption, $E_1$, $E_2$,..., $E_k$ are all lower triangular matrices, so$(E_k E_{k-1} \cdots E_1)^{-1}$ is also a lower triangular matrix. Denote $L = (E_k E_{k-1} \cdots E_1)^{-1}$, then

$$A = LU$$

where $L$ is a lower triangular matrix, and $U$ is an upper triangular matrix.

### 5.1.3.2. PA = LU

Next, we extend the LU decomposition to general $A$.

How can we handle row exchange in our $LU$ decomposition?

Assume we are going to do Gaussian Elimination with matrix $A$ with row exchange.

1. At first We can postmultiply some elementary matrices $E$ to get $EEEA$.
2. Sometimes we need to multiply by $P_{ij}$ to do *row exchange* to continue Gaussian Elimination.
3. So we may end our elimination with something like $PEEEEPEEEPEEEEA$.
4. If we can get all the elementary matrix $L$ together, we could convert them into one single $L$ that has the same effect as before.
5. The key point is to get all the row exchange matrix $P$ out from the elementary matrices, we define $P = P_k P_{k-1} \cdots P_1$
6. We then conduct GE on a new matrix $PA$

**Theorem 5.2 — LU decomposition.** Any square matrix $A$ can be written as

$$PA = LU,$$

where $L$ is lower triangular, $U$ is upper triangular, $P$ is a certain permutation matrix.

The proof is omitted.

The following question is: how do we know what row exchanges are needed in the beginning? The answer is to conduct GE twice. Precisely,

- Fisrt time.

$$GE: \quad E_5 P_{12} E_4 E_3 P_{32} E_2 E_1 P_{14} A = U$$

- Second time.

$$\hat{\boldsymbol{E}}_5\hat{\boldsymbol{E}}_4\hat{\boldsymbol{E}}_3\hat{\boldsymbol{E}}_2\hat{\boldsymbol{E}}_1(\boldsymbol{P}_{12}\boldsymbol{P}_{32}\boldsymbol{P}_{14}\boldsymbol{A}) = \boldsymbol{U}$$

  Then we have $\boldsymbol{PA} = \boldsymbol{LU}$

### 5.1.4. Summary

In this lecture, We cover properties of inverse and LU decomposition.

- Properties of Inverse
  - inverse of diagonal, triangular and elementary matrices
  - inverse of matrix products
- LU decomposition : $\boldsymbol{A} = \boldsymbol{LU}$ for some $\boldsymbol{A}$ (when there is no exchange in GE process)
- PLU decomposition: $\boldsymbol{PA} = \boldsymbol{LU}$ for any square matrix $\boldsymbol{A}$, where

  $\boldsymbol{P}$ is permutation matrix

  $\boldsymbol{L}$ is lower triangular matrix

  $\boldsymbol{U}$ is upper triangular matrix

## 5.2. Lecture 10: LU Decomposition II: Application

### 5.2.1. Review

#### 5.2.1.1. LU Decomposition

**LU Decomposition without row changes.** We have learned that, for a given real square matrix $A$, if no row change in the process of Gaussian Elimination, then $A$ can be decomposed as

$$A = LU,$$

where $L$ is **lower triangular** and $U$ is **upper triangular**.

**General LU Decomposition.** More generally, for any $n \times n$ real matrix $A$, due to the row changes, it has a factorization in the form

$$PA = LU,$$

where $L$ is **lower triangular**, $U$ is **upper triangular**, and $P$ is a **permutation matrix** representing the product of some elementary matrices for particular row changing.

Permutation matrix $P$ is needed because

1. **Matrix multiplication is not a commutative operation.**
2. **Type I elementary matrix (for row changing) is not a lower triangular matrix.**

**Why study LU Decomposition.** It's useful in but not limited to the cases below:

1. **Computing matrix inverse.**
2. **Speed up Gaussian Elimination.**
3. **Prove theorems.**

In this lecture, we focus more on the definition of inverse matrix, when a matrix would be invertible, and how to compute the inverse of a matrix if exists. In future lectures, we will also learn that whether $AB = I$ or $BA = I$ is sufficient enough for matrix A to be invertible.

## 5.2.2. Expression and Existence of Inverse

### 5.2.2.1. Two Cases of RREF for Square System

Consider we are given a coefficient matrix A which is a square matrix. We do the Gaussian Elimination by using both forward elimination and backward substitution.

$$A \to A_1 \to A_2 \to \cdots \to U \to \cdots \to B_1 \to \cdots \to D$$

Note that D is in RREF, but is not always a diagonal matrix.

- **case1:** All diagonal entries of $U$ are nonzero. After forward elimination it becomes an $n$ pivots upper triangular matrix. And after backward substitution we have an identity matrix.

- **case2:** Some diagonal entries of $U$ are zeros. Then the upper triangular matrix must have less than $n$ pivots after forward elimination. Hence we cannot get an identity matrix but it is in non-diagonal RREF after backward substitution.

### 5.2.2.2. Existence of Inverse

Here we use the general LU Decomposition to answer the question.

- **Lemma 1** For a General LU Decomposition $PA = LU$, $A$ is invertible if and only if $U$ is invertible.

In order to prove the lemma, we use the properties and the facts that we have already known. In the earlier lecture, we have shown that the product of invertible matrices are invertible, and for invertible matrices $A$ and $B$, $(AB)^{-1} = B^{-1}A^{-1}$. Additionally, we have the facts that $PA = LU$, and the permutation matrix $P$ as well as the lower triangular matrix $L$ are invertible ($L$ is the product of elementary matrices).

To prove the "if" part, if $U$ is invertible, then

$$PA = LU \to A = P^{-1}LU \to A \textit{ is invertible.}$$

We omit the "Only if" part because it is similar to the proof above. Having this lemma, we only need to show $U$ is invertible, then we can easily conclude $A$ is invertible as well.

- **Lemma 2** An upper triangular matrix $U$ is invertible if and only if $u_{ii} \neq 0, \forall i \in \{1,2,\cdots,n\}$.

This lemma holds due to the property of upper triangular matrix in the last lecture. By combining those two lemmas, we retrieve the following theorem.

**Theorem 5.3** Suppose $PA = LU$ is a General LU Decomposition, then $A$ is invertible if and only if all diagonal entries of $U$ are nonzeros.

Recall that non-zero diagonal entries of $U$ are pivots of $A$, so we have the first answer to the question of the existence of inverse so far.

- **Answer 1** A matrix $A \in \mathbb{R}^{n\times n}$ is invertible if and only if $A$ has $n$ pivots.

Besides the equivalent statement mentioned above, are there any other answers? Here we consider the following lemma from lecture 9:

- **Lemma 3 (Solution Uniqueness)** Consider a square real linear system $A\mathbf{x} = \mathbf{b}$. If $A$ is invertible, then the linear system has a **unique** solution $\mathbf{x} = A^{-1}\mathbf{b}$.

The proof is trivial. Suppose the system has another solution $\mathbf{y}$ that $\mathbf{x} \neq \mathbf{y}$, it's easy to show such solution does not exist by showing $\mathbf{x} - \mathbf{y} = \mathbf{0}$, which is a contradiction to the assumption.

Here we also give a more general version of the statement between the existence of matrix inverse and the solution of a linear system containing such matrix: For any $b \in \mathbb{R}^n$, $A \in \mathbb{R}^{n\times n}$, $A\mathbf{x} = \mathbf{b}$ has a unique solution is equivalent to the condition that $A$ is invertible.

Consider a $3 \times 3$ square system $A\mathbf{x} = \mathbf{0}$. Again, let's recall the second case of RREF for square systems. After doing the forward elimination, let $u$ be the entries of the upper triangular matrix $U$. Then we have the claim that if $u_{ii} = 0$ for some i, then there are infinitely many solutions. Remind that $\mathbf{x} = \mathbf{0}$ is always a solution for the linear system $A\mathbf{x} = \mathbf{0}$, then the case for no solution does not exist.

Finally we get the RREF of $A$. For instance,

$$\left[\begin{array}{ccc|c} 1 & 0 & \alpha_1 & 0 \\ 0 & 1 & \alpha_2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right],$$

Since we are considering $A\mathbf{x}=\mathbf{0}$, the corresponding linear system is:

$$\begin{cases} x_1+\alpha_1x_3=0 \\ x_2+\alpha_2x_3=0 \\ \qquad 0=0 \end{cases}$$

Then the solution set can be expressed as $\{(-\alpha_1t,-\alpha_2t,t)\,|t\in\mathbb{R}\}$

So here we give the second answer to the question.

- **Answer 2** A matrix $A\in\mathbb{R}^{n\times n}$ is invertible if and only if $A\mathbf{x}=\mathbf{0}$ has a unique solution $\mathbf{x}=\mathbf{0}$.

By lemma 3, we can easily prove the statement from left to right due to the solution uniqueness. In the other direction, If $A\mathbf{x}=\mathbf{0}$ has a unique solution, then $u_{ii}\neq 0$ for any $i$ inferred from the claim above. Thus we have the conclusion that $A$ is invertible.

Next, we will show how to express certain inverse of an invertible matrix in detail.

### 5.2.2.3. Expression of Inverse

Now let's consider the second question. How to express the inverse of a square matrix if it exists? Here we make the assumption that $A$ is invertible. Then $A$ has $n$ pivots.

**R** If A is not invertible, no need to discuss "how to compute" $A^{-1}$. So we will discuss this question only under the assumption above.

By the process of LU Decomposition, we have the expression below:

$$PA=LU\Rightarrow A^{-1}=U^{-1}L^{-1}P. \tag{5.2}$$

However, it's not clear how to compute $U^{-1}$ and $L^{-1}$. Therefore, we will utilize the GE process to give a computable expression.

$$A\to A_1\to A_2\to\cdots\to U\to\cdots\to B_1\to\cdots\to D$$

Suppose the corresponding elementary matrices are $E_1,E_2,\cdots,E_p$, then the matrix representation of the whole GE process is $E_pE_{p-1}\cdots E_1A=I_n$. Denote $E_pE_{p-1}\cdots E_1$ as the matrix $M$.

- **Lemma 4** If $MA = I_n$, and $M$ is invertible, then $A^{-1} = M$.

To prove it, we need to show $AM = I_n$ as well. Then by the given lemma, we have $A^{-1} = M$. Though there exists other approaches, the point here is to emphasize the importance of checking both directions. It then leads to the further conclusion:

- **Answer** Suppose $E_p E_{p-1} \cdots E_1$ are the elementary matrices corresponding to the operations in the GE to get an identity matrix (assuming we indeed get an identity matrix). Then,

$$A^{-1} = E_p \cdots E_2 E_1. \tag{5.3}$$

Hence, we give a summary of all the equivalent conditions for invertible so far.

**Theorem 5.4 — Equivalent Conditions for Invertibility.** Let $A \in \mathbb{R}^{n \times n}$. The following statements are equivalent:

1. $A$ is invertible.
2. The linear system $A\mathbf{x} = 0$ has a unique solution $\mathbf{x} = \mathbf{0}$.
3. $A$ is a product of elementary matrices.
4. The RREF of $A$ is $I_n$ (or $A$ is row-equivalent to $I_n$).
5. $A\mathbf{x} = \mathbf{0}$has a unique solution for any vector $\mathbf{b} \in \mathbb{R}$
6. More ...

The full proof will be given in the next lecture.

## 5.2.3. Computing Inverse and LU Decomposition

### 5.2.3.1. Computing Inverse

**To express a result**, we use a clear formula containing specific symbols with clear meanings, such as the quadratic formula. Meanwhile, **to compute a certain answer**, we use a procedure or algorithm to obtain it. For instance, We apply the quadratic formulas to compute and express the answer of a quadratic equation. However, we usually do not use the quadratic formula directly. Instead, theorems like computing square can somehow simplify the procedure. Similarly, we can use other algorithms to compute the inverse.

R Relation between expression and computation:

- Expression can be used to compute, if each symbol can be computed.
- But expressions do not have to be computable. e.g., if it contains symbols that are not easy to compute. (e.g.$U^{-1}$)
- Algorithms can help derive expressions sometimes.

To compute the matrix inverse, we give several algorithms.

**Algorithm 1.** Compute $A^{-1}$ using the expression.

- **Step 1: Forward elimination.**

  Run forward elimination, till get upper triangular matrix $U$.

- **Step 2: Backward substitution.**

  Run backward substitution, till get identity matrix $I_n$.

- **Step 3: Record elementary matrices.** Record elementary matrices $E_1, \cdots, E_k$ in Step 1. Record elementary matrices $E_{k+1}, E_{k+2}, \cdots, E_p$ in Step 2.

- **Step 4: Compute the inverses.**

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

Here we give an example of a $2 \times 2$ matrix $\begin{bmatrix} 2 & 1 \\ 6 & 8 \end{bmatrix}$. We follow the steps of GE.

**Step 1 & 2: GE.**

$$\begin{bmatrix} 2 & 1 \\ 6 & 8 \end{bmatrix} \xrightarrow{-3R_1+R_2} \begin{bmatrix} 2 & 1 \\ 0 & 5 \end{bmatrix} \xrightarrow{-\frac{1}{5}R_2+R_1} \begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix} \xrightarrow{\frac{1}{2}R_1, \frac{1}{5}R_2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

**Step 3** We record elementary matrices in Step 1 & 2 as $E_1, E_2, E_3, E_4$ in order.

**Step 4** We write the inverse by formula (5.3).

$$A^{-1} = E_4 E_3 E_2 E_1.$$

However, this way is relatively slow because we need to do a series of multiplications among those elementary matrices, and it is hard to think about what those elementary matrices are.

By observation, we can figure out that doing multiplications among those elementary matrices is equivalent to applying the same row operations to $I_n$. Hence we give the second algorithm.

**Algorithm 2.** Apply GE to $[A, I_n]$:

$$[A, I_n] \xrightarrow{operation1} \square \xrightarrow{operation2} \square \cdots \xrightarrow{operationk} [I_n, A^{-1}].$$

For example, to find the inverse of $A = \begin{bmatrix} 1 & 4 & 3 \\ -1 & -2 & 0 \\ 2 & 2 & 3 \end{bmatrix}$, we apply the second algorithm.

$$[A|I] = \left[\begin{array}{ccc|ccc} 1 & 4 & 3 & 1 & 0 & 0 \\ -1 & -2 & 0 & 0 & 1 & 0 \\ 2 & 2 & 3 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 4 & 3 & 1 & 0 & 0 \\ 0 & 2 & 3 & 1 & 1 & 0 \\ 0 & -6 & -3 & -2 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 4 & 3 & 1 & 0 & 0 \\ 0 & 2 & 3 & 1 & 1 & 0 \\ 0 & 0 & 6 & 1 & 3 & 1 \end{array}\right] \to$$

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ 0 & 2 & 0 & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & 6 & 1 & 3 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ 0 & 1 & 0 & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\ 0 & 0 & 1 & \frac{1}{6} & \frac{1}{2} & \frac{1}{6} \end{array}\right].$$

Thus the inverse of $A$ is $A^{-1} = \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{6} \end{bmatrix}$.

## 5.2.3.2. Computing LU Decomposition

Similar to the above algorithms, we can also use Gaussian Elimination to form LU Decomposition. Assume there is no row exchange operation during the forward elimination, then we will have the following algorithm.

**Algorithm 3.** Compute LU Decomposition.

**Step 1: Forward elimination.**

Run forward elimination, till get upper triangular matrix $U$.

**Step 2: Forward elimination.**

Record elementary matrices $E_1, \cdots, E_k$ in Step 1.

**Step 3: Compute L.**

Compute $L = E_1^{-1} \cdots E_k^{-1}$.

**Step 4: Conclusion.**

$A$ has an LU Decomposition $A = LU$, where $L$ is obtained in Step 3 and $U$ is recorded in Step 1.

R What if there is no LU Decomposition? (i.e. There are row exchange matrices in $E_1 \cdots E_k$.)

- **In homework/exam:**

  —First, it probably will not happen if the problem asks you compute LU decomposition.

  —Second, if it really happens in homework/exam, just say: there is no LU decomposition.

- **In practice:**

  —People can swap rows to perform PLU decomposition.

  —There may be other ways (beyond the lecture).

### 5.2.4. Time Complexity

To solve a linear system $A\mathbf{x} = \mathbf{b}_k$, where $k = 1, 2, 3, \cdots, 10^{10}$, it's hard to use hand to calculation anymore. Thus, we use computer to help us do the calculation. Here we have two methods.

- **Method 1**

  After computing and saving $M = A^{-1}$, then we can solve by computing $M\mathbf{b_k}$ for any $k$.

- **Method 2**

  Directly use GE to solve each problem.

But how do we analyze which method is better? Here we need a notion of "time" that is objective: **Time Complexity**. We say each scalar multiplication takes 1 unit time, and each scalar sum takes 1 unit time. We also say an algorithm takes time $K$ if the algorithm requires $K$ unit operations. Then, how shall we compare different algorithms? We follow the steps bellow:

- **First**, fix "input size".

- **Second**, consider the time as a function of "input size".

For example, if $\mathbf{u}, \mathbf{v} \in \mathbb{R}^{n \times 1}$, then $\mathbf{u} + \mathbf{v}$ takes time $n$ due to $n$ times of the scalar addition. Normally, while computing the time complexity, there is no need to obtain the concrete coefficient of each $n^k$ term. For instance, if $A, B \in \mathbb{R}^{n \times n}$, then $AB$ takes time $cn^3$ for some constant $c > 0$.

Generally, we use the **Big O notation** to represent the time complexity. In such notation, we only consider the leading term and ignore the different between terms having difference coefficient.

**Examples:**

1. $n$, $n+1$, $3n$, $100n$ are all $O(n)$.

2. $n^2$, $3n^2$, $7n^2 - 100n + 1$, $100n^2$ are all $O(n^2)$.

3. $n^3$, $0.5n^3$, $7n^3 + 108n^2 - 9n + 100$ are all $O(n^3)$.

Considering the method 2, to compute time complexity of forward elimination, we need following steps.

- **Step 1.1:** One scalar-vector multiplication, we need $(n-1)$ vector additions. **Time:** $n^2$

$$\begin{bmatrix} * & * & * & * & * \\ * & * & * & * & * \\ \vdots & \vdots & \ddots & \ddots & * \\ * & * & \cdots & * & * \\ * & * & \cdots & * & * \end{bmatrix} \rightarrow \begin{bmatrix} * & * & * & * & * \\ 0 & * & * & * & * \\ \vdots & \vdots & \ddots & \ddots & * \\ 0 & * & \cdots & * & * \\ 0 & * & \cdots & * & * \end{bmatrix}$$

- **Step 1.2:** Multiply size $(n-1)$ vector by scalar, we need $(n-2)$ additions of size $(n-1)$ vectors. **Time:** $(n-1)^2$

$$\vdots$$

- **Step 1.(n-1):** Multiply size 2 vector by scalar, we need 1 addition of size 2 vectors. **Time:** $2^2$

By adding those time together, we retrieve the time complexity $O(n^3)$. Then we perform the backward substitution similarly:

- **Step 2.1:** One multiplication, $(n-1)$ additions. **Time:** $3n$

- **Step 2.2:** One multiplication, $(n-2)$ additions. **Time:** $3(n-1)$

$$\vdots$$

- **Step 2.(n-1):** One multiplication, one addition. **Time:** 6

- **Step 2.n:** One multiplication. **Time:** 1

The total time complexity is $O(n^2)$. Thus, we figure out GE runs in $O(n^3)$ time due to the dominating term $O(n^3)$.

So for method 1, after computing and $M = A^{-1}$, for each $b_k$, the time complexity for computing $M\mathbf{b_k}$ is $O(n^2)$. For method 2, the time complexity for directly using GE to solve $A\mathbf{x} = \mathbf{b_k}$ for each $b_k$.

However, the time complexity of computing matrix inverse is also $O(n^3)$ in method 1, so the time complexity is similar to GE. Actually, the time complexity can be much smaller than $O(n^3)$, but we will not have further discussion on it.

### 5.2.5. Summary

In this lecture, we learned matrix inverse and computation of LU Decomposition.

- Existence of matrix inverse

  —Algorithm test: $n$ pivots.

  —Equation test: $A\mathbf{x} = \mathbf{0}$ has a unique solution.

  —Can be written as product of elementary matrices.

- Expressions and computation of inverse

  —Expression $A^1 = E_p \cdots E_2 E_1$.

  —Use expression or GE to get inverse.

- Time complexity

  —Gaussian elimination: $O(n^3)$

# Chapter 6

# Lecture 11:Linear Space and Solution Set

In the last chapter, we study the vectors from $R^m$, we show that vectors satisfy some nice algebraic properties.

**But there are multiple mathematical objects that share the same properties of vectors in $R^m$, we will develop a theory for all these different objects in this chapter.**

## 6.0.1. Motivation : Solution of Rectangular System

Before this lecture: previous theory guarantees to solve "good" square linear systems

After this lecture: develop new theory to solve all rectangular linear systems.

R "Solve" doesn't mean finding a solution. It means finding the solution set

**Rectangular Cases**

[1] No Solution.

**Example 6.1**

$$\left[\begin{array}{cccc|c} \boxed{1} & 0 & 3 & -2 & 0 \\ 0 & \boxed{1} & 1 & -3 & 0 \\ 0 & 0 & 0 & 0 & \boxed{1} \end{array}\right] \tag{6.1}$$

You may find there is no solutions to this system of equation, which means Gaussian Elimination **doesn't** always ensure the system has a solution.

**Solution set:** $\emptyset$

[2] Infinitely many solutions.

**Example 6.2**

$$\left[\begin{array}{ccccc|c} \boxed{1} & 1 & 0 & 0 & 0 & 3 \\ 0 & 0 & \boxed{1} & 1 & 0 & -4 \\ 0 & 0 & 0 & 0 & \boxed{1} & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \tag{6.2}$$

$$x_1 = 3 - x_2$$

$$x_3 = -4 - x_4$$

$$x_5 = 2$$

**Solution set:**

$$\left\{ \begin{pmatrix} 3-s \\ s \\ -4-t \\ t \\ 2 \end{pmatrix} | s,t \in R \right\}$$

**Question:** Can $\left\{ \begin{pmatrix} t \\ t^2 \end{pmatrix} | t \in R \right\}$ be a solution set?

No, but how to prove it?

**Need theory: a proper way of expressing the solution set**

## 6.0.2. Vector Space

We move to a new topic: vector spaces.

**From Numbers to Vectors.** We know matrix calculation(such as $\boldsymbol{A}x = \boldsymbol{b}$) involves many numbers, but they are just linear combinations of $n$ vectors.

**Third Level Undetstanding.** This topic moves from numbers and vectors to a third level of understanding (the highest level). Instead of individual column vectos, we look at "*spaces*" of vectors. And this topic will end with the "*Fundamental Theorem of Linear Algebra*".

$$\text{Matrix Calculation:} \quad \text{Numbers} \implies \text{Vectors} \implies \textbf{Spaces}$$

We begin with the typical vector space, which is denoted as $\mathbb{R}^n$.

**Definition 6.1** [Real Space] The space $\mathbb{R}^n$ contains all column vectors $v$ such that $v$ has $n$ real number entries.

**Notation.** We denote vectors as *a column between brackets*, or *along a line using commas and parentheses:*

$$\begin{bmatrix} 4 \\ \pi \end{bmatrix} \text{ is in } \mathbb{R}^2 \quad (1,1,1) \text{ is in } \mathbb{R}^3.$$

## 6.0.2.1. Definition

**Definition 6.2** [vector space] A **vector space** $\boldsymbol{V}$ is a set of vectors such that these vectors satisfy *vector addition* and *scalar multiplication*:

- **vector addition:**If vector $v$ and $w$ is in $\boldsymbol{V}$, then $v + w \in \boldsymbol{V}$.(**Additive Closure**)
- **scalar multiplication:**If vector $v \in \boldsymbol{V}$, then $cv \in \boldsymbol{V}$ for any real numbers $c$(**Scalar Closure**).

Then V is called a $\boldsymbol{vectorspaceoverR}$ if the following eight axioms are satisfied:

A1. $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}, \forall \boldsymbol{u}, \boldsymbol{v} \in V$.

A2. $\boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w}) = (\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w} = \boldsymbol{u} + \boldsymbol{v} + \boldsymbol{w} \in W$.

A3. There exists an element $\mathbf{0}$ s.t. $\boldsymbol{u} + \mathbf{0} = \boldsymbol{u}, \forall \boldsymbol{u} \in V$.

A4. If $.u \in V$, then there exists -$\boldsymbol{u} = (-1)\boldsymbol{u}, s.t. \boldsymbol{u} + (-\boldsymbol{u}) = 0$.

A5. $\alpha(\boldsymbol{u} + \boldsymbol{v}) = \alpha\boldsymbol{u} + \alpha\boldsymbol{v}, \forall \alpha \in R, \boldsymbol{u}, \boldsymbol{v} \in V$

A6. $(\alpha + \beta)\boldsymbol{u} = \alpha\boldsymbol{u} + \beta\boldsymbol{u}, \forall \alpha, \beta \in R, \boldsymbol{u} \in$ V.

A7. $\alpha(\beta\boldsymbol{u}) = (\alpha\beta)\boldsymbol{u}, \forall \alpha, \beta \in R, \boldsymbol{u} \in V$.

A8. $1\boldsymbol{u} = \boldsymbol{u}$.

And we call the elements in the set V "$\boldsymbol{vectors}$".

In other words, the set of vectors is **closed** under *addition* $v + w$ and *multiplication* $cv$. In other words,

**any linear combination is closed under vector space.**

R

- The definition of the vector space is **"abstract"**, which means that V may not be $R^n$.
- The definition of vector space is a generalization of $R^n$.
- **Remark:** Linear space is also called "Vector space"

**Proposition 6.1** Every vector space must contain the zero vector.

*Proof.* Given $v \in \boldsymbol{V} \implies -v \in \boldsymbol{V} \implies v + (-v) = \boldsymbol{0} \in \boldsymbol{V}$. ■

### ⋆ some common vector spaces.

R some examples about "span" will be introduced more deeply in "Null Space and Colum Space and Span" part.

**Example 6.3**

$$\boldsymbol{V} = \left\{ \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \\ \vdots \end{pmatrix} \middle| \{a_n\} \text{ is infinite length sequences.} \right\}$$

is a vector space.

This is because for any vector $v = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \\ \vdots \end{pmatrix}, w = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \\ \vdots \end{pmatrix}$, we can define vector addition and scalar multiplication as follows:

$$v + w = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \\ \vdots \end{pmatrix} \quad cv = \begin{pmatrix} ca_1 \\ ca_2 \\ \vdots \\ ca_n \\ \vdots \end{pmatrix} \quad \text{for any } c \in \mathbb{R}.$$

$$
\begin{aligned}
\boldsymbol{V} = \operatorname{span}&\left\{ v_1 = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{4} \\ \vdots \\ \frac{1}{2^n} \\ \vdots \end{pmatrix}, v_2 = \begin{pmatrix} \frac{1}{3} \\ \frac{1}{9} \\ \vdots \\ \frac{1}{3^n} \\ \vdots \end{pmatrix}, v_3 = \begin{pmatrix} \frac{1}{4} \\ \frac{1}{16} \\ \vdots \\ \frac{1}{4^n} \\ \vdots \end{pmatrix} \right\} \\
&= \{\alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 \mid \alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}\}
\end{aligned}
$$

is also vector space.

**Definition 6.3** [Span] The **span** of a collection of vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n \in \mathbb{R}^m$ is defined as:

$$
\operatorname{span}\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\} = \left\{ \boldsymbol{y} \in \mathbb{R}^m \,\middle|\, \boldsymbol{y} = \sum_{i=1}^{n} \alpha_i \boldsymbol{a}_i, \boldsymbol{\alpha} \in \mathbb{R}^n \right\},
$$

i.e., it is the set of all linear combinations of $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$.

How to check $\boldsymbol{V}$ is a vector space?

Given any two vectors $u, w$ in $\boldsymbol{V}$, suppose

$$
u = \alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3, \quad v = \beta_1 v_1 + \beta_2 v_2 + \beta_3 v_3,
$$

then we obtain:

$$
\begin{aligned}
\gamma_1 u + \gamma_2 v &= \gamma_1(\alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3) + \gamma_2(\beta_1 v_1 + \beta_2 v_2 + \beta_3 v_3) \\
&= (\gamma_1 \alpha_1 + \gamma_2 \beta_1) v_1 + (\gamma_1 \alpha_2 + \gamma_2 \beta_2) v_2 + (\gamma_1 \alpha_3 + \gamma_2 \beta_3) v_3
\end{aligned}
$$

where $\gamma_1, \gamma_2 \in \mathbb{R}$. Hence any linear combination of $u$ and $w$ are also in $\boldsymbol{V}$. Hence $\boldsymbol{V}$ is a vector space.

**Example 6.4** $\boldsymbol{F} = \{f(x) \mid f : [0,1] \mapsto \mathbb{R}\}$ is also a vector space. (verify it by yourself.)

This vector space $\boldsymbol{F}$ contains all real functions defined on $[0,1]$, an it is infinite dimensional.

Given two functions $f$ and $g$ in $\boldsymbol{F}$, the inner product of $f$ and $g$ is defined as:

$$
\langle f, g \rangle := \int_0^1 f(x) g(x) \, \mathrm{d}x
$$

Also, we can use the span to form a vector space:

$$\boldsymbol{F} = \operatorname{span}\{sinx, x^3, e^x\} = \{\alpha_1 sinx + \alpha_2 x^3 + \alpha_3 e^x \mid \alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}.\}$$

This set $\boldsymbol{F}$ is also a vector space. ■

**Example 6.5**

$$\boldsymbol{V} = \left\{ \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \middle| a_{ij} \in \mathbb{R} \text{ for } i = 1,2; j = 1,2,3. \right\}$$

is a vector space. Moreover, it is equivalent to the span of six basic vectors:

$$\boldsymbol{V} = \operatorname{span}\left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right\}$$

We say that $\boldsymbol{V}$ is 6-dimensional without introducing the definiton of dimension formally. ■

**Example 6.6**

$$\boldsymbol{V} = \left\{ \begin{bmatrix} a_{ij} \end{bmatrix}_{3\times 3} \middle| \text{any } 3 \times 3 \text{ matrices} \right\}$$

is also a vector space.

Obviously, it is 9-dimensional. We usually denote it as $\dim(\boldsymbol{V}) = 9$.

$$\boldsymbol{V}_1 = \left\{ \begin{bmatrix} a_{ij} \end{bmatrix}_{3\times 3} \middle| \text{any } 3 \times 3 \text{ symmetric matrices} \right\}$$

is a special vector space.

Notice that $\boldsymbol{V}_1 \subset \boldsymbol{V}$, so we say $\boldsymbol{V}_1$ is a *subspace* of $\boldsymbol{V}$. In the future we will know $\dim(\boldsymbol{V}_1) = 6 < 9$. ■

**Example 6.7** (**Vector Space**$P_n$) Let the set of all polynomials of degree $\leq$ n is

$$P_n = \left\{ \sum_{i=0}^{n} a_i x^i | a_i \in R \right\}$$

with

(1) Addition: $\sum_{i=0}^{n} a_i x^i + \sum_{i=0}^{n} b_i x^i = \sum_{i=0}^{n} (a_i + b_i) x^i \in P_n$, for any $\sum_{i=0}^{n} a_i x^i, \sum_{i=0}^{n} b_i x^i \in P_n$.

(2) Scalar multiplication:

$\alpha(\sum_{i=0}^{n} a_i x^i) = \sum_{i=0}^{n}(\alpha a_i)x^i \in P_n, \forall \alpha \in$ R , $\sum_{i=0}^{n} a_i x^i \in P_n$

Then one can show that $P_n$ is a vector space. Each polynomial in $P_n$ can be treated as a "vector".

Eight properties need to be verified.

1. For (A1):

   Pick any $\sum_{i=0}^{n} a_i x^i, \sum_{i=0}^{n} b_i x^i \in P_n$, then we have

$$\sum_{i=0}^{n} a_i x^i + \sum_{i=0}^{n} b_i x^i = \sum_{i=0}^{n}(a_i + b_i)x^i$$

$$\sum_{i=0}^{n} b_i x^i + \sum_{i=0}^{n} a_i x^i = \sum_{i=0}^{n}(b_i + a_i)x^i$$

   Then (A1) is satisfied.

2. Pick any $\sum_{i=0}^{n} a_i x^i, \sum_{i=0}^{n} b_i x^i, \sum_{i=0}^{n} c_i x^i \in P_n$,then we have

$$(\sum_{i=0}^{n} a_i x^i + \sum_{i=0}^{n} b_i x^i) + \sum_{i=0}^{n} c_i x^i = \sum_{i=0}^{n}(a_i + b_i)x^i + \sum_{i=0}^{n} c_i x^i = \sum_{i=0}^{n}(a_i + b_i + c_i)x^i$$

$$\sum_{i=0}^{n} a_i x^i + (\sum_{i=0}^{n} b_i x^i + \sum_{i=0}^{n} c_i x^i) = \sum_{i=0}^{n} a_i x^i + \sum_{i=0}^{n}(b_i + c_i)x^i = \sum_{i=0}^{n}(a_i + b_i + c_i)x^i$$

   Thus

$$(\sum_{i=0}^{n} a_i x^i + \sum_{i=0}^{n} b_i x^i) + \sum_{i=0}^{n} c_i x^i = \sum_{i=0}^{n} a_i x^i + (\sum_{i=0}^{n} b_i x^i + \sum_{i=0}^{n} c_i x^i)$$

   Axiom (A2) is satisfied.

3. The zero polynomial 0 = 0 $+0 * x + 0 * x^2 + \cdots + 0 * x^n$ as the zero vector , since

$$(a_0 + a_1 x + \cdots a_n x^n) + 0 + 0 * x + 0 * x^2 + \cdots + 0 * x^n = (a_0 + a_1 x + \cdots a_n x^n)$$

   (A3) is valid.

4. (A4)-(A8) can also be verified. (omit)

■

**Example 6.8** (**Vector Space** C[a,b]) $Let C[a,b] = \{f | f : [a,b] \Rightarrow R\}$ is the set of continuous functions defined in [a,b], and take $f, g \in C[a,b], f = g$ if and only if $f(x) = g(x)$, $\forall x \in$ [a,b].C[a,b] is associated with the following operations.

(i) Addition: for f,g $\in$ C[a,b], (f+g)(x) $\triangleq$ f(x) + g(x),x$\in$[a,b]. thus f+g $\in$ C[a.b].

(ii) Scalar multiplication: for f $\in C[a,b], \alpha \in R, (\alpha f)(x) \triangleq \alpha f(X), x \in [a,b]$,thus $\alpha f \in C[a,b]$.

One can check that C[a,b] is a vector space over R by checking the eight conditions. We can treat each function from C[a,b] as a "vector".

In fact

1. To show (A1), we need to show that (f+g)+h =f+(g+h) $\forall f,g,h \in C[a,b]$. That is we need to show that $(f+g)(x) = (g+f)(x)$ for each x $\in [a,b]$. Since $(f+g)(x) = f(x)+g(x) = g(x)+f(x) = (g+f)(x)$, $\forall x \in [a,b]$, thus (A1) is valid.

2. To show (A2), we need to show that $((f+g)+h)(x) = (f+(g+h))(x).\forall f,g,h \in C[a,b]$. That is we need to show that $((f+g)+h)(x) = (f+(g+h))(x)$ for each x $\in [a,b]$. Since $((f+g)+h)(x) = (f+g)(x)+h(x) = f(x)+g(x)+h(x)$,$\forall x \in [a,b]$ and $(f+(g+h))(x) = f(x)+(g+h)(x) = f(x)+g(x)+h(x)$. thus $(f+g)+h = f+(g+h)$.

3. Now set $z(x) = 0$,$\forall x \in [a,b]$,then z can be treated as the zero vector, that is $f+z = f$,i.e., $f(x)+z(x) = f(x)$,$\forall x \in [a,b]$,thus (A3) is valid.

4. (A4)-(A5) can also be verified, you can check them by yourself.

$\blacksquare$

Some other simple examples:

- Eucilidean space
  - n=1: $R^1$ is a line.

    Each element corresponds to a point on the line.
  - n=2: $R^2$ is a plane.

    Each element corresponds to a point on the plane.
  - $\vdots$

.

- $x \in R^2 : x_2 = 2x_1$

- Set of nxn upper triangular matrices.

**NonEg 4:** $R^1 \bigcup R^2$ .

### ⋆ some Non-examples.

1. **NonEg 1:** $\{1,2,3,4,\cdots\}$ is not a linear space.
2. **NonEg 2:** Set of non-negative real numbers $R_+$ is not a linear space.
3. **NonEg 3:** $x \in R^2 : x_1 = 1$
4. **NonEg 4:** Set of nxn elementary matrices.
5. **NonEg 5:** $R^1 \bigcup R^2$

   *Proof.* $x + \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} =?$
   Not defined in this chapter. Thus NOT a linear space.

   R In Matlab, $x + \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \triangleq \begin{pmatrix} x+y_1 \\ x+y_2 \end{pmatrix}$, but that is not a standard addition.

   ■

#### 6.0.2.2. More Properties of Vector Space

**Theorem 6.1** (More Properties of Vector Space) Let V be a vector space over R. then

1. Zero vector is unique.
2. $0\boldsymbol{x} = 0, \forall \boldsymbol{x} \in V$
3. $c\mathbf{0} = \mathbf{0}, \forall c \in$ R.
4. (The additive inverse if unique) For each $\boldsymbol{u} \in V$ there is a unique $\boldsymbol{x} \in V$ so that $\boldsymbol{x} + \boldsymbol{y} = \mathbf{0}$.(This element $\boldsymbol{y}$ is denoted by -$\boldsymbol{x}$.)

Is there an easier way to verify a linear space?

Key property: closed under linear combination

### 6.0.3. subspace

#### 6.0.3.1. Definition

**Definition 6.4** [subsection] Let V be a vector space over R. A subset W $\subset$ V is called a ***subspace*** if W is a vector space.

Instead of checking all conditions, it only needs to check three (because the other conditions follow from the fact that $W \subset V$ and V is a vector space).

**Definition 6.5** [subsection(alternative, equivalent definition)] Let V be a vector space over R. A subset W $\subset$ V is a subspace of V if the following are satisfied:

1. $\mathbf{0} \subset W$.
2. W is closed under vector addition: $\forall \boldsymbol{u}, \boldsymbol{v} \subset W$,we have $\boldsymbol{u} + \boldsymbol{v} \subset W$.
3. W is closed under scalar multiplication: $\forall \alpha \subset R, \boldsymbol{u} \subset W$, we have$\alpha \boldsymbol{u} \subset W$.

## 6.0.3.2. Some examples

**Example 6.9** $P_n$ (the set of all polynomials of degree at most n) is a subspace of $P_N$ for N > n ($P_n \subset P_N$), since

1. $0 = 0 + 0 * x + \cdots + 0 * x^n \subset P_n$
2. $\forall f(x) = a_0 + a_1 x + \cdots a_n x^n$, $g(x) = b_0 + b_1 x + \cdots b_n x^n \subset P_n$,then $f(x) + g(x) = (a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_n + b_n)x^n \subset P_n$
3. $\forall \alpha \subset R, \forall f(x) = a_0 + a_1 x + \cdots a_n x^n \subset P_n$, then $\alpha f(x) = \alpha a_0 + \alpha a_1 x + \cdots + \alpha a_n x^n \subset P_n$.

**Example 6.10**

$$\left\{ W = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in R^3 | x + y - 3z = 0 \right\}$$

is a subspace of $R^3$

(1) $\mathbf{0} \in W \Rightarrow W \neq \emptyset$

(2) Let $\begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} \in$ W and Let $\begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} \in W$ , then

$$x_1 + y_1 - 3z_1 = 0$$

and

$$x_2 + y_2 - 3z_2 = 0$$

Thus,

$$(x_1 + x_2) + (y_1 + y_2) - 3(z_1 + z_2) = (x_1 + y_1 - 3z_1) + x_2 + y_2 - 3z_2 = 0$$

Therefore, $\begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} + \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} \in W.$

(3) Let $\begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} \in W$, then x+y-3z=0,therefore $\alpha x + \alpha y - 3\alpha z = 0, \forall \alpha \in R$. Thus $\alpha \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \alpha x \\ \alpha y \\ \alpha z \end{bmatrix} \in W$

■

R $a_1 x_1 + \cdots + a_n x_n = b$ where b $\neq 0$ is NOT a linear space. when b=0, is a linear set

### 6.0.4. Null Space and Column Space and Span

R [Two ways to generate subspaces]

- Solution set of linear equation + Taking **intersection** ==> null space
- Span (linear combination)

**Span.**

**Definition 6.6** [span]

Suppose $\mathbf{V}$ is a linear space.

Suppose $\mathcal{U} = \boldsymbol{u_1}, \boldsymbol{u_2}, \cdots, \boldsymbol{u}_k$ is a subset of V.

The span of $\mathcal{U}$ is defined as span $\mathbf{U} \triangleq \{a_1\boldsymbol{u_1} + \cdots + a_k\boldsymbol{u_k} | a_1, \cdots, a_k \in R\}$

**In words, the span (of elements of a linear space) is the set of all linear combinations of these elements**

**Eg:** $W = \{s\boldsymbol{u} + t\boldsymbol{v} | s, t \in R\}$ is the span of $\{\boldsymbol{u}, \boldsymbol{v}\}$

R For simplicity, we can also say W is the span of $\boldsymbol{u}, \boldsymbol{v}$

**Definition 6.7** [spanning set]

Suppose $\mathbf{V}$ is a linear space.

Suppose $\mathcal{U} = \boldsymbol{u_1}, \boldsymbol{u_2}, \cdots, \boldsymbol{u}_k$ is a subset of V.

If $span(\mathcal{U}) = V$,then we say $\mathcal{U}$ is a spanning set of V , or $\mathcal{U}$ spans V.

R Spanning set is NOT unique. Cannot say "the spanning set"

**Theorem 6.2** (Span is a subspace) Let $\mathcal{U} = \boldsymbol{u_1}, \cdots, \boldsymbol{u_n} \in$ V ( V is a vector space), then $\boldsymbol{Span}(\mathcal{U})$ is a subspace of V.

*Proof.* 1. $\mathbf{0} = 0\boldsymbol{u_1} + \cdots + 0\boldsymbol{u_n} \subset \mathcal{U}$

2. Take $\boldsymbol{w}, \boldsymbol{v} \in \boldsymbol{Span}(\mathcal{U})$, then $\boldsymbol{w} = h_1\boldsymbol{u_1} + \cdots + h_n\boldsymbol{u_n}$, $\boldsymbol{v} = k_1\boldsymbol{u_1} + \cdots + k_n\boldsymbol{u_n}$ thus,$\boldsymbol{w} + \boldsymbol{v} = (h_1 + k_1)\boldsymbol{u_1} + \cdots + (h_n + k_n)\boldsymbol{u_n} \in \boldsymbol{Span}(\mathcal{U})$
3. Take $\boldsymbol{w} \in \boldsymbol{Span}(\mathcal{U})$ and $\alpha \in R$, then $\boldsymbol{w} = h_1\boldsymbol{u_1} + \cdots + h_n\boldsymbol{u_n}$ thus, $\alpha\boldsymbol{w} = \alpha h_1\boldsymbol{u_1} + \cdots + \alpha h_n\boldsymbol{u_n}$

■

R Now we have generalized the idea of a ***vector*** from a column vector in $R^n$ to elements of a "vector space".

Note that concepts like "linear combination" ,"span", "linearly dependent" and "linear independent" are all well defined even for abstract vectors (the addition and scalar multiplication operations are defined on these abstract vector spaces, eight axioms are satisfied for these vector spaces.)

### 6.0.4.1. The solution to $\boldsymbol{Ax} = \mathbf{0}$

We can use vector space to discuss the solution to system of equation. Firstly, let's introduce some definitions:

**Definition 6.8** [homogeneous equations] A system of linear equations is said to be **homogeneous** if the constants on the righthand side are all zero. In other words, $\boldsymbol{Ax} = \mathbf{0}$ is said to be **homogeneous**.

**Definition 6.9** The solution set of homogeneous linear system $Ax - 0$ is a linear space.

**Definition 6.10** [column space] The column space consists of all linear combinations of the columns of matrix $\boldsymbol{A}$. In other words, for the matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ given by $\boldsymbol{A} = \left[\begin{array}{c|c|c|c} a_1 & a_2 & \dots & a_n \end{array}\right]$, its column space is denoted as

$$\boldsymbol{C}(\boldsymbol{A}) := \operatorname{span}(a_1, a_2, \dots, a_n) \subset \mathbb{R}^m.$$

**Definition 6.11** [null space] The null space of a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ consists of all solutions to $\boldsymbol{Ax} = \mathbf{0}$, which can be denoted as

$$\boldsymbol{N}(\boldsymbol{A}) = \{\boldsymbol{x} \mid \boldsymbol{Ax} = \mathbf{0}\} \subset \mathbb{R}^n.$$

**Proposition 6.2** The null space $\boldsymbol{N}(\boldsymbol{A})$ is a vector space.

*Proofoutline.* For any two vectors $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{N}(\boldsymbol{A})$, we have $\boldsymbol{Ax} = \mathbf{0}, \boldsymbol{Ay} = \mathbf{0}$.

$$\implies \boldsymbol{A}(\alpha \boldsymbol{x} + \beta \boldsymbol{y}) = \alpha(\boldsymbol{Ax}) + \beta(\boldsymbol{Ay}) = \alpha \mathbf{0} + \beta \mathbf{0} = \mathbf{0} \quad \alpha, \beta \in \mathbb{R}.$$

Since the linear combination of $\boldsymbol{x}$ and $\boldsymbol{y}$ is also in $\boldsymbol{N}(\boldsymbol{A})$, $\boldsymbol{N}(\boldsymbol{A})$ is a vector space. ■

**Example 6.11** Describe the null space of $\boldsymbol{A} = \begin{bmatrix} 1 & 0 \\ 5 & 0 \\ 2 & 3 \end{bmatrix}$.

Obviously, converting matrix into linear system of equation we obtain:

$$\begin{cases} x_1 + 0x_2 = 0 \\ 5x_1 + 4x_2 = 0 \\ 2x_1 + 3x_2 = 0 \end{cases}$$

We can easily obtain the solution $\begin{cases} x_1 = 0 \\ x_2 = 0 \end{cases}$. Hence the null space is $\boldsymbol{N}(\boldsymbol{A}) = \boldsymbol{0}$.

**Example 6.12** Describe the null space of $\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 1 \\ 5 & 4 & 9 \\ 2 & 3 & 5 \end{bmatrix}$.

In the next lecture we will know its null space is a line.

We find that $\boldsymbol{A}\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} = \boldsymbol{0}$, so $\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$ is a special solution.

Note that *the null space contains all linear combinations of special solutions.* Hence the null space is $\boldsymbol{N}(\boldsymbol{A}) = \left\{ c \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} \middle| c \in \mathbb{R} \right\}$.

Some further knowledge.

#### 6.0.4.2. The complete solution to $\boldsymbol{Ax} = \boldsymbol{b}$

**"What" is the solution set of $\boldsymbol{Ax} = \boldsymbol{b}$? How to express/compute the solution set** In order to find all solutions of $\boldsymbol{Ax} = \boldsymbol{b}$, ($\boldsymbol{A}$ may not be square matrix), let's introduce two kinds of solutions:

**Definition 6.12** [Particular & Special Solution] For the system of equations $\boldsymbol{Ax} = \boldsymbol{b}$, there are two kinds of solutions:

$$x_{\text{particular}} \quad \textbf{The particular solution that solves } \boldsymbol{Ax} = \boldsymbol{b}$$

$$x_{\text{nullspace}} \quad \textbf{The special solutions that solves } \boldsymbol{Ax} = \mathbf{0}$$

There is a theorem that helps us to obtain the complete solution to $\boldsymbol{Ax} = \boldsymbol{b}$.

**Theorem 6.3** Any solution to $\boldsymbol{Ax} = \boldsymbol{b}$ can be represented as $\boldsymbol{x}_{complete} = \boldsymbol{x_p} + \boldsymbol{x_n}$.

*Proof.* *Sufficiency.* Given $\boldsymbol{x}_{complete} = \boldsymbol{x_p} + \boldsymbol{x_n}$, it suffices to show $\boldsymbol{x}_{complete}$ is the solution to $\boldsymbol{Ax} = \boldsymbol{b}$.

Note that

$$\boldsymbol{Ax}_{complete} = \boldsymbol{A}(\boldsymbol{x_p} + \boldsymbol{x_n}) = \boldsymbol{Ax_p} + \boldsymbol{Ax_n} = \boldsymbol{b} + \mathbf{0} = \boldsymbol{b}.$$

Hence $\boldsymbol{x}_{complete}$ is the solution to $\boldsymbol{Ax} = \boldsymbol{b}$. ■

*Necessity.* Suppose $\boldsymbol{x}^*$ is the solution to $\boldsymbol{Ax} = \boldsymbol{b}$, it suffices to show $\boldsymbol{x}^*$ could be represented as $\boldsymbol{x_p} + \boldsymbol{x_n}$.

It suffices to show $\boldsymbol{x}^* - \boldsymbol{x_p} \in \boldsymbol{N}(\boldsymbol{A})$.

Notice that $\boldsymbol{A}(\boldsymbol{x}^* - \boldsymbol{x_p}) = \boldsymbol{Ax}^* - \boldsymbol{Ax_p} = \boldsymbol{b} - \boldsymbol{b} = \mathbf{0} \implies \boldsymbol{x}^* - \boldsymbol{x_p} \in \boldsymbol{N}(\boldsymbol{A})$. ■

■

**Proposition 6.3** $Ax = b$ has a solution iff b ∈ C(A).

**R** $b = Ax$ means b is a linear combination of columns of A with coefficient $x_1, x_2, \cdots, x_n$. "Solvable" means such a linear combination exists.

**Example 6.13** Column Space of $A = \begin{bmatrix} 1 & 0 \\ 4 & 3 \\ 2 & 3 \end{bmatrix}$ is the set $\left\{ \alpha \begin{bmatrix} 1 \\ 4 \\ 2 \end{bmatrix} + \alpha \begin{bmatrix} 0 \\ 3 \\ 3 \end{bmatrix} | \alpha_1, \alpha_2 \in R \right\}$, or set

$$\left\{\begin{bmatrix} \alpha_1 \\ 4\alpha_1 + 3\alpha_2 \\ 2\alpha_1 + 3\alpha_2 \end{bmatrix} |\alpha_1, \alpha_2 \in R \right\}$$, or set $\{A\alpha | \alpha \in R^2\}$.

- **Matrix form:**

  $b \in C(A) \Longleftrightarrow \exists \alpha_1, \alpha_2$, s.t. $A\alpha = b \Longleftrightarrow Ax = b$ has at least one solution **x**.

- **Scalar form:** $b \in C(A) \Longleftrightarrow \exists \alpha_1, \alpha_2$, s.t. $\begin{bmatrix} \alpha_1 \\ 4\alpha_1 + 3\alpha_2 \\ 2\alpha_1 + 3\alpha_2 \end{bmatrix} = \begin{bmatrix} b1 \\ b2 \\ b3 \end{bmatrix} \Longleftrightarrow \begin{cases} l\alpha_1 = b_1 \\ 4\alpha_1 + 3\alpha_2 = b_2 \\ 2\alpha_1 + 3\alpha_2 = b_3 \end{cases}$

  has at least one solution $(\alpha_1, \alpha_2)$.

- **Column form:**

  $b \in C(A) \Longleftrightarrow \exists \alpha_1, \alpha_2$, s.t. $\alpha \begin{bmatrix} 1 \\ 4 \\ 2 \end{bmatrix} + \alpha \begin{bmatrix} 0 \\ 3 \\ 3 \end{bmatrix} = b \Longleftrightarrow Ax = b$ has at least one solution **x**.

■

**Example 6.14** Let's study a system that has $n = 2$ unknowns but only $m = 1$ equation:

$$x_1 + x_2 = 2.$$

It's easy to check that the particular solution is $\boldsymbol{x_p} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, the special solutions are $\boldsymbol{x_n} = c \begin{pmatrix} 1 \\ -1 \end{pmatrix}$, $c$ can be taken arbitararily.

Hence the complete solution for the equations could be written as

$$\boldsymbol{x}_{complete} = \boldsymbol{x_p} + \boldsymbol{x_n} = \begin{pmatrix} c+1 \\ -c+1 \end{pmatrix}.$$

So we summarize that if there are $n$ unknowns and $m$ equations such that $m < n$, then $\boldsymbol{Ax} = \boldsymbol{b}$ is **underdetermined** (It may have infinitely many solutions since the special solutions could be infinite).

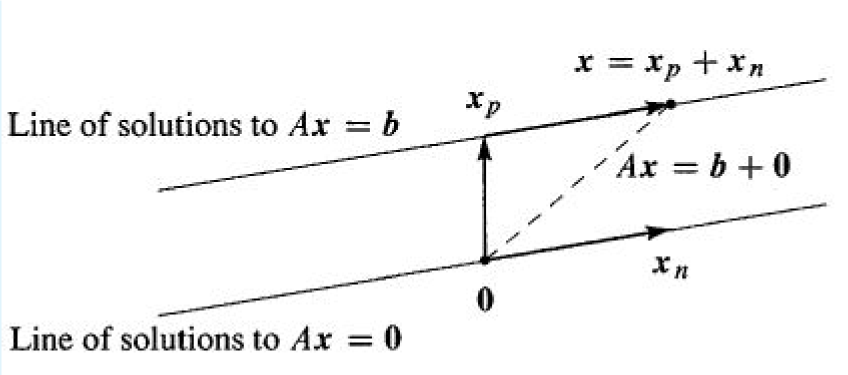



Figure 6.1: Complete solution = one particular solution + all nullspace solutions

### 6.0.4.3. Problem Size Analysis

When faced with $m \times n$ matrix $\boldsymbol{A}$, notice that $\boldsymbol{m}$ refers to the **number of equations**, $\boldsymbol{n}$ refers to the **number of variables**. Assume $\boldsymbol{r}$ denotes **number of pivots**, then we know $\boldsymbol{r}$ is also the **number of pivot variables**, $\boldsymbol{n} - \boldsymbol{r}$ is the **number of free variables**. Finally we have $\boldsymbol{m} - \boldsymbol{r}$ **redundant equations** and $\boldsymbol{r}$ **irredundant equations**. In next lecture, we will introduce the definition for $\boldsymbol{r}$ formally (rank).

# 6.1. Lecture 12: Vector Space II

## 6.1.1. Introduction

In this lecture, we aim to completely solve the linear singular system. That is, the following two questions will be answered.

- When does the system have no solution, when does the system have infinitely many solutions? (Note that singular system don't has unique solution.)
- If it has infinitely many solutions, how to find and express all these solutions?

If we express system into matrix form, the question turns into:

How to solve the rectangular $\mathbf{A}x = b$?

## 6.1.2. Examples of Solving Equations

- For square case, we often convert the system into $\boldsymbol{U}\boldsymbol{x} = \boldsymbol{c}$, where $\boldsymbol{U}$ is of *row echelon form*.
- However, for rectangular case, *row echelon form*(ref) is not enough, we must convert it into **reduced row echelon form**(rref):

$$\boldsymbol{U}(\text{ref}) = \begin{bmatrix} 1 & 0 & \times & \times & \times & 0 & \times \\ 0 & 1 & \times & \times & \times & 0 & \times \\ 0 & 0 & 0 & 0 & 0 & 1 & \times \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \implies \boldsymbol{R}(\text{rref}) = \begin{bmatrix} 1 & 0 & \times & \times & \times & 0 & \times \\ 0 & 1 & \times & \times & \times & 0 & \times \\ 0 & 0 & 0 & 0 & 0 & 1 & \times \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

**Example 6.15** We discuss how to solve **square** matrix of **rref**:

- If all rows have nonzero entry, we have:

$$\begin{bmatrix} 1 & & & 0 \\ & 1 & & \\ & & 1 & \\ 0 & & & 1 \end{bmatrix} \boldsymbol{x} = \boldsymbol{c} \implies \boldsymbol{x} = \boldsymbol{c}$$

- But note that *some rows could be all zero*:

$$\begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 0 \end{bmatrix} \boldsymbol{x} = \boldsymbol{c} \implies \begin{cases} x_1 = c_1 \\ x_2 = c_2 \\ x_3 = c_3 \\ 0 = c_4 \end{cases}$$

  So the solution results have two cases:

  - If $c_4 \neq 0$, we have no solution of this system.
  - If $c_4 = 0$, we have infinitely many solutions, which can be expressed as:

$$x_{\text{complete}} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

    where $x_4$ could be arbitarary number.

Hence, for square system, does Gaussian Elimination work?

Answer: Almost, except for the "pivot=0"case:

- All pivots $\neq 0 \implies$ the system has unique solution.
- Some pivots $= 0$ (The matrix is singular)

  1. No solution. (When LHS $\neq$ RHS)
  2. Infinitely many solutions.

#### 6.1.2.1. Example for solving rectangular system of rref

Recall the definition for rref:

**Definition 6.13** [reduced row echelon form] Suppose a matrix has $r$ *nonzero* rows, each row has leading 1 as pivots. If all columns with pivots (call it pivot column) are all zero entries apart from the pivot in this column, then this matrix is said to be **reduced row echelon form**(**rref**).

Next, we want to show how to solve a rectangular system of rref. Note that in last lecture we study the solution to a rectangular system is given by:

$$\boldsymbol{x}_{\text{complete}} = \boldsymbol{x}_p + \boldsymbol{x}_{\text{special}}.$$

**Example 6.16** Solve the system

$$\begin{bmatrix} 1 & 3 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \boldsymbol{x} = \boldsymbol{c}.$$

**Step 1: Find null space.** Firstly we solve for $\boldsymbol{R}\boldsymbol{x} = \boldsymbol{0}$:

$$\begin{bmatrix} 1 & 3 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \implies \begin{cases} x_1 + 3x_2 = 0 \\ x_3 + x_4 = 0 \end{cases}$$

Then we express the **pivot variables** in the form of **free variables.**

Note that the pivot columns in $\boldsymbol{R}$ are column 1 and 3, so the pivot variable is $x_1$ and $x_3$. The free variable is the remaining variable, say, $x_2$ and $x_4$.

The expressions for $x_1$ and $x_3$ are given by:

$$\begin{cases} x_1 = -3x_2 \\ x_3 = -x_4 \end{cases}$$

Hence, all solutions to $\boldsymbol{R}\boldsymbol{x} = \boldsymbol{0}$ are

$$\boldsymbol{x}_{\text{special}} = \begin{bmatrix} -3x_2 \\ x_2 \\ -x_4 \\ x_4 \end{bmatrix} = x_2 \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 0 \\ 0 \\ -1 \\ 1 \end{bmatrix}$$

where $x_2$ and $x_4$ can be taken arbitararily.

**Step 2: Find one particular solution to $\boldsymbol{R}\boldsymbol{x} = \boldsymbol{c}$.** The trick for this step is to set $x_2 = x_4 = 0$. (*set free variable to be zero and then derive the pivot variable.*):

$$\begin{bmatrix} 1 & 3 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ 0 \\ x_3 \\ 0 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} \Longrightarrow \begin{cases} x_1 = c_1 \\ x_3 = c_2 \\ 0 = c_3 \end{cases}$$

which follows that:

- if $c_3 = 0$, then exists particular solution $\boldsymbol{x}_p = \begin{bmatrix} c_1 \\ 0 \\ c_2 \\ 0 \end{bmatrix}$;
- if $c_3 \neq 0$, then $\boldsymbol{R}\boldsymbol{x} = \boldsymbol{c}$ has no solution.

**Final solution.** If assume $c_3 = 0$, then all solutions to $\boldsymbol{R}\boldsymbol{x} = \boldsymbol{c}$ are given by:

$$\boldsymbol{x}_{complete} = \boldsymbol{x}_p + \boldsymbol{x}_{\text{special}} = \begin{bmatrix} c_1 \\ 0 \\ c_2 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 0 \\ 0 \\ -1 \\ 1 \end{bmatrix}$$

Next we show how to solve a general rectangular:

## 6.1.3. How to Solve a General Rectangular?

For linear system $\boldsymbol{Ax} = \boldsymbol{b}$, where $\boldsymbol{A}$ is rectangular, we can solve this system as follows:

**Step 1: Gaussian Elimination.** With proper row permutaion (postmultiply $\boldsymbol{P}_{ij}$) and row transformation (postmultiply $\boldsymbol{E}_{ij}$), we convert $\boldsymbol{A}$ into $\boldsymbol{R}$(rref), then we only need to solve $\boldsymbol{Rx} = \boldsymbol{c}$.

**Example 6.17** The first example is a $3 \times 4$ matrix with two pivots:

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 2 & 3 \\ 2 & 2 & 8 & 10 \\ 3 & 3 & 10 & 13 \end{bmatrix}$$

Clearly $a_{11} = 1$ is the first pivot, then we clear row 2 and row 3 of this matrix:

$$\boldsymbol{A} \xrightarrow[\boldsymbol{E}_{31}=\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix}]{\boldsymbol{E}_{21}=\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}} \begin{bmatrix} 1 & 1 & 2 & 3 \\ 0 & 0 & 4 & 4 \\ 0 & 0 & 4 & 4 \end{bmatrix} \xrightarrow[\boldsymbol{E}_{32}=\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}]{\boldsymbol{E}_{12}=\begin{bmatrix} 1 & -\frac{1}{2} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}} \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 0 & 4 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

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

If we want to solve $\boldsymbol{Ax} = \boldsymbol{b}$, firstly we should convert $\boldsymbol{A}$ into $\begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ (rref).

Then we should identify **pivot variables** and **free variables**. we can follow the proceed below:

$$\text{pivots} \Longrightarrow \text{pivot columns} \Longrightarrow \text{pivot variables}$$

**Example 6.18** we want to identify **pivot variables** and **free variables** of $\boldsymbol{R}$:

$$\boldsymbol{R} = \begin{bmatrix} 1 & 0 & \times & \times & \times & 0 & \times \\ 0 & 1 & \times & \times & \times & 0 & \times \\ 0 & 0 & 0 & 0 & 0 & 1 & \times \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The pivot are $r_{11}, r_{22}, r_{36}$. So the pivot columns are column 1,2,6. So the *pivot variables* are $x_1, x_2, x_6$; the *free variables* are $x_3, x_4, x_5, x_7$. ▪

**Step2: Compute null space $N(\boldsymbol{A})$.** In order to find $N(\boldsymbol{A})$, it is equivalent to compute $N(\boldsymbol{R})$. The space $N(\boldsymbol{R})$ has $(n-r)$ dimensions, so it suffices to get $(n-r)$ special solutions first:

- For each of the $(n-r)$ free variables,
  - set the value of **it** to be 1;
  - set the value of other **free variables** to be 0;
  - Then solve $\boldsymbol{Rx} = \boldsymbol{0}$ (to get the value of pivot variables) to get the special solution.

**Example 6.19** Continue with $3 \times 4$ matrix example:

$$\boldsymbol{R} = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

We want to find special solutions to $\boldsymbol{Rx} = \boldsymbol{0}$:

1. Set $x_2 = 1$ and $x_4 = 0$. Solve $\boldsymbol{Rx} = \boldsymbol{0}$, then $x_1 = -1$ and $x_3 = 0$.

   Hence one special solution is $y_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$.
2. Set $x_2 = 0$ and $x_4 = 1$. Solve $\boldsymbol{Rx} = \boldsymbol{0}$, then $x_1 = -1$ and $x_3 = -1$.

Then another special solution is $y_2 = \begin{bmatrix} -1 \\ 0 \\ -1 \\ 1 \end{bmatrix}$.

- Then $N(\boldsymbol{A})$ is the collection of linear combinations of these special solutions:

$$N(\boldsymbol{A}) = \operatorname{span}(y_1, y_2, \dots, y_{n-r}).$$

**Example 6.20** We continue the example above, when we get all special solutions

$$y_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad y_2 = \begin{bmatrix} -1 \\ 0 \\ -1 \\ 1 \end{bmatrix},$$

**the null space contains all linear combinations of the special solutions**:

$$\boldsymbol{x}_{\text{special}} = \operatorname{span}(\begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ -1 \\ 1 \end{bmatrix}) = x_2 \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 0 \\ -1 \\ 1 \end{bmatrix}$$

where $x_2, x_4$ here could be arbitarary.

**Step3: Compute a particular solution $\boldsymbol{x}_p$.** The easiest way is to "read" from $\boldsymbol{R}\boldsymbol{x} = \boldsymbol{c}$:

- **Guarantee the existence of the solution**. Suppose $\boldsymbol{R} \in \mathbb{R}^{m \times n}$ has $r(\leq m)$ pivot variables, then it has $(m-r)$ zero rows and $(n-r)$ free variables. For the existence of solutions, the value of entries of $\boldsymbol{c}$ which correspond to zero rows in $\boldsymbol{R}$ must also be zero.

**Example 6.21** If $\boldsymbol{R}\boldsymbol{x} = \begin{bmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \boldsymbol{x} = \boldsymbol{c} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}$, then in order to have a solution, we must let $c_3 = 0$. ■

- If the condition above is not satisfied, then the system has no solution. Let's preassume the satisfaction of such a condition. To compute a particular solution $\boldsymbol{x}_p$, we set *the value for all **free variables** of $\boldsymbol{x}_p$ to be zero, and the value for the pivot variables are from $\boldsymbol{c}$.*

  More specifically, the first entry in $\boldsymbol{c}$ is exactly the value for the first pivot variable;the second entry in $\boldsymbol{c}$ is exactly the value for the second pivot variable......, and the remaining entries of $\boldsymbol{x}_p$ are set to be zero.

**Example 6.22** If $\boldsymbol{R}\boldsymbol{x} = \begin{bmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \boldsymbol{x} = \boldsymbol{c} = \begin{bmatrix} c_1 \\ c_2 \\ 0 \end{bmatrix}$, we want to compute particular solution

$$\boldsymbol{x}_p = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

As we know $x_2, x_4$ are free variable, $x_2 = x_4 = 0$; and $x_1, x_3$ are pivot variable, so we have $\begin{pmatrix} x_1 \\ x_3 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$.
Hence the solution for $\boldsymbol{R}\boldsymbol{x} = \boldsymbol{c}$ is

$$\boldsymbol{x}_p = \begin{bmatrix} c_1 \\ 0 \\ c_2 \\ 0 \end{bmatrix}.$$

■

**Final step: Obtain complete solutions.** All solution of $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ are

$$\boldsymbol{x}_{\text{complete}} = \boldsymbol{x}_p + \boldsymbol{x}_{\text{special}},$$

where $x_{\text{special}} \in N(\boldsymbol{A})$. Note that $\boldsymbol{x}_p$ is defined in step3, $\boldsymbol{x}_{\text{special}}$ is defined in step2.

The above algorithm is very practical in solving the linear system. However, by adding one more step in this algorithm, we can have a better view about the solution.

**Step1.5: Interchange columns.** After the step 1, we get a linear system $\boldsymbol{R}x = b$, where $\boldsymbol{R}_{m\times n}$ is rref with $r$ pivot columns. Now we interchange the columns of $\boldsymbol{R}$ so that the first $r$ rows and first $r$ columns form an $r \times r$ identity matrix (This can be accomplished by interchanging all pivot columns to the first $r$ colum). Remember when we interchanging the columns, the vecter $x$ should also interchange rows, but the vecter $b$ remain unchanged (This is because $b = \boldsymbol{R}x = \boldsymbol{R} \cdot P_{ij}P_{ij} \cdot x = (\boldsymbol{R} \cdot P_{ij})(P_{ij} \cdot x)$). The result should be like:

$$\begin{bmatrix} \boldsymbol{I_r} & \boldsymbol{F} \\ \mathbf{0}_{(m-r)\times r} & \mathbf{0}_{(m-r)\times(n-r)} \end{bmatrix} \begin{bmatrix} \boldsymbol{x_P} \\ \boldsymbol{x_F} \end{bmatrix} = \begin{bmatrix} \boldsymbol{b}_{r\times 1} \\ \boldsymbol{b}'_{(m-r)\times 1} \end{bmatrix}$$

Where $\boldsymbol{x}_P$ are pivot variables, and $\boldsymbol{x}_F$ are free variables. The following table lists all possible forms for the matrix:

| **Four types for R** | $[I]$ | $\begin{bmatrix} \boldsymbol{I} & \boldsymbol{F} \end{bmatrix}$ | $\begin{bmatrix} \boldsymbol{I} \\ \mathbf{0} \end{bmatrix}$ | $\begin{bmatrix} \boldsymbol{I} & \boldsymbol{F} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ |
|---|---|---|---|---|
| **Rank** | $r = m = n$ | $r = m < n$ | $r = n < m$ | $r < m, r < n$ |
| **# of solutions** | 1 | $\infty$ | 0 or 1 | 0 or $\infty$ |
| **# of solutions when $\boldsymbol{b}'_{(m-r)\times 1} = \mathbf{0}$** | 1 | $\infty$ | 1 | $\infty$ |

## 6.2. Assignment Four

1. Let
$$A = \begin{bmatrix} 1 & 2 & 3 & 1 & -3 \\ 2 & 5 & 5 & 4 & 9 \\ 3 & 7 & 8 & 5 & 6 \end{bmatrix}$$
   (a) Compute the *reduced row echelon form* $\boldsymbol{U}$ of $\boldsymbol{A}$.

   (b) Compute all solutions of $\boldsymbol{Ax} = \boldsymbol{b}$, where $\boldsymbol{b} = \begin{bmatrix} 1 & 1 & 2 \end{bmatrix}^{\mathrm{T}}$.

   (c) Compute all solutions of $\boldsymbol{Ax} = \boldsymbol{b}$, where $\boldsymbol{b} = \begin{bmatrix} b_1 & b_2 & b_3 \end{bmatrix}^{\mathrm{T}}$.

   Note:Identify when there is *no solution*, and when the solution *exists*, write down all solutions in terms of $b_1, b_2, b_3$.

2. In each of the following, determine the *dimension* of the space:

   (a) $\operatorname{span}\left\{ \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ -2 \\ 4 \end{pmatrix}, \begin{pmatrix} -3 \\ 3 \\ 6 \end{pmatrix} \right\}$;

   (b) $\operatorname{col}(\boldsymbol{A})$, where $\boldsymbol{A} = \begin{bmatrix} 1 & -2 & 3 & 2 \\ -1 & 2 & -2 & -1 \\ 2 & -4 & 5 & 3 \end{bmatrix}$;

   (c) $N(\boldsymbol{B})$, where $\boldsymbol{B} = \begin{bmatrix} 1 & 3 & 2 \\ 2 & 1 & 4 \\ 4 & 7 & 8 \end{bmatrix}$;

   (d) $\operatorname{span}\{(x-2)(x+2), x^2(x^4-2), x^6-8\}$;

   (e) $\operatorname{span}\{5, \cos 2x, \cos^2 x\}$ as a *subspace* of $C[-\pi, \pi]$.

   $C[-\pi, \pi]$ denotes the space of *continuous functions defined on the domain* $C[-\pi, \pi]$.

3. Let $\boldsymbol{A}$ be an $6 \times n$ matrix of rank $r$. For each pair of values of $r$ and $n$ below, how many solutions could one have for the linear system $\boldsymbol{Ax} = \boldsymbol{b}$? Explain your answers.

   (a) $n = 7, r = 5$;

   (b) $n = 7, r = 6$;

   (c) $n = 5, r = 5$.

4. Prove the following proposition:

   Let $\boldsymbol{V}$ be a vector space of dimension $n > 0$, then

(a) Any set of n *linearly independent* vectors in $\boldsymbol{V}$ form a basis.

(b) Any set of $n$ vectors that span $\boldsymbol{V}$ form a basis.

*Hint: refer to theorem(**??**)*

5. (a) Assume $\boldsymbol{U}.\boldsymbol{V}$ are subspaces of a vector space $\boldsymbol{W}$.

   Define $\boldsymbol{U}+\boldsymbol{V} = \{u+v | u \in \boldsymbol{U}, v \in \boldsymbol{V}\}$, i.e. each vector in $\boldsymbol{U}+\boldsymbol{V}$ is the sum of one vector in $\boldsymbol{U}$ and one vector in $\boldsymbol{V}$.

   Prove that $\boldsymbol{U}+\boldsymbol{V}$ is a subspace of $\boldsymbol{W}$.

   (b) Prove the intersection $\boldsymbol{U} \cap \boldsymbol{V} = \{x | x \in \boldsymbol{U} \text{ and } x \in \boldsymbol{V}\}$ is also a subspace of $\boldsymbol{W}$.

   (c) In $\mathbb{R}^4$, let $\boldsymbol{U}$ be the subspace of all vectors of the form $\begin{bmatrix} u_1 & u_2 & 0 & 0 \end{bmatrix}^{\mathrm{T}}$, and let $\boldsymbol{V}$ be the subspace of all vectors of the form $\begin{bmatrix} 0 & v_2 & v_3 & 0 \end{bmatrix}^{\mathrm{T}}$. What are the dimensions of $\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{U} \cap \boldsymbol{V}, \boldsymbol{U}+\boldsymbol{V}$?

   (d) If $\boldsymbol{U} \cap \boldsymbol{V} = \{\boldsymbol{0}\}$, prove that $\dim(\boldsymbol{U}+\boldsymbol{V}) = \dim(\boldsymbol{U}) + \dim(\boldsymbol{V})$.

6. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be $m \times n$ matrices. Prove that

$$\operatorname{rank}(\boldsymbol{A}+\boldsymbol{B}) \leq \operatorname{rank}(\boldsymbol{A}) + \operatorname{rank}(\boldsymbol{B}).$$

7. Let $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is an arbitrary matrix, $\boldsymbol{B} \in \mathbb{R}^{n \times n}$ is a square matrix. Prove that

   (a) $\operatorname{rank}(\boldsymbol{A}\boldsymbol{B}) \leq \operatorname{rank}(\boldsymbol{A})$;

   (b) If $\operatorname{rank}(\boldsymbol{B}) = n$, then $\operatorname{rank}(\boldsymbol{A}\boldsymbol{B}) = \operatorname{rank}(\boldsymbol{A})$.

8. Prove that any $(n-1)$ vectors in $\mathbb{R}^n$ cannot form a basis.

   Note: this is a corollary of theorem(**??**). You should prove it by assuming theorem(**??**) is unknown. You may check the proposition(7.5) as hint.

# Chapter 7

# Week7

## 7.1. Lecture13: Linear Independence, Basis and Solving Ax = b

### 7.1.1. Review

Last lecture, we talked about the definition of span, column space and how to solve a linear system in a faster and more efficient way. We have also known different kinds of spaces have the following relationships:

- Subspace is a special type of linear space.
- Span is a special type of subspace.
- Column space is a particular span composed by columns of a certain matrix.

Noted that spanning sets are not unique. There are many ways to form a spanning set.

#### 7.1.1.1. Summary of Steps for Solving Ax = 0

**Algorithm 1 to find the solution set of Ax = 0**

- **Input**: $m \times n$ matrix A.
- **Output**: matrix $M$, such that $C(M)$ is the solution set of $A\boldsymbol{x} = \mathbf{0}$.
- **Step1**:
  - Conduct elimination on $A$, to obtain RREF $R$.

- Suppose $R$ has $r$ pivot columns, $n-r$ free columns.
- Denote the indices of free columns as $i_1 < \cdots < i_{n-r} \in \{1, \cdots, n\}$

- **Step2: In $R$, Conduct 3 operations to obtain $-F$**
  - Step 2.1 Delete all pivot columns;
  - Step 2.2 Delete all non-pivot rows;
  - Step 2.3 Multiply by (-1).
- **Step3**: In $-F$, sequentially insert $\boldsymbol{e}_1^{\mathrm{T}}, \cdots, \boldsymbol{e}_{n-r}^{\mathrm{T}}$ into row $i_1, i_2, \cdots, i_r$. The resulting matrix is the output $M$.
- **Conclusion**:The solusion set is $\{M[\alpha_1, \cdots, \alpha_{n-r}]^{\mathrm{T}} : \alpha_1, \cdots, \alpha_{n-r} \in \mathbb{R}\}$

R **Free Columns and Number of Solutions**

- If there is no free column, then we get an empty matrix after Step 2.1. In this case, the system has unique solution.
- If there is at least one free column, then we get a non-empty empty matrix after Step 2.1. In this case, the system has infinitely many solution.

**Proposition 7.1 — Under-determined Homogeneous System.** Suppose $m < n$, then $A\boldsymbol{x} = \boldsymbol{0}$ has infinitely many solutions.

By observation, is $m < n$, so there is at least one free column. The rigorous prove set up on this observation. Because the matrix has at most $r$ pivots, so at most $r$ pivot columns. Thus the number of columns is equal to $n-r$ which is larger than $m-r \geq 0$, so at least one free column. Thus it has infinitely many solutions.

### 7.1.2. Solving Ax = b

Again, we use the inserting trick to solve a more generalized system $A\boldsymbol{x} = \boldsymbol{b}$

#### 7.1.2.1. Summary of steps for Solving Ax = b

**Algorithm 2 to find the solution set of Ax = b**

- **Input**: $m \times n$ matrix A, $m \times 1$ vector $\boldsymbol{b}$.
- **Desired output**: $n \times 1$ vector $\boldsymbol{x_p}$, matrix $M$, such that $\boldsymbol{x_p} + C(M)$ is the solution set of $A\boldsymbol{x} = \boldsymbol{b}$.
- **Step1**:
  - Conduct elimination on $[A|\boldsymbol{b}]$, to obtain RREF $[R|\hat{\boldsymbol{b}}]$.
  - Suppose $R$ has $r$ pivot columns, $n-r$ free columns.
  - Denote the indices of free columns as $i_1 < \cdots < i_{n-r} \in \{1, \cdots, n\}$
- **Step2: Judging solvability.**
  - Suppose the entries of $\hat{\boldsymbol{b}}$ in the $m-r$ non-pivot rows are $c_1, \cdots, c_{m-r}$.
  - **IF** any $c_t \neq 0$, **THEN**: STOP, and REPORT "No solution".
  - **ELSE**: continue.
- **Step3: In $[R|\hat{\boldsymbol{b}}]$, conduct 3 operations to obtain $[-F|\boldsymbol{p}]$**
  - Step 3.1 Delete all pivot columns;
  - Step 3.2 Delete all non-pivot rows, get $[F|\boldsymbol{p}]$;
  - Step 3.3 Multiply $F$ by (-1), to get $[-F|\boldsymbol{p}]$.
- **Step4:** In $[-F|\boldsymbol{p}]$, sequentially insert $[\boldsymbol{e}_1^{\mathrm{T}}, 0], \cdots, [\boldsymbol{e}_{n-r}^{\mathrm{T}}, 0]$ into row $i_1, i_2, \cdots, i_r$. The resulting matrix is the output $[M|\boldsymbol{x_p}]$.
- **Conclusion**:The solusion set is $\{\boldsymbol{x_p} + M[\alpha_1, \cdots, \alpha_{n-r}]^{\mathrm{T}} : \alpha_1, \cdots, \alpha_{n-r} \in \mathbb{R}\}$

### 7.1.2.2. Number of solutions of the linear system

So how does a linear system affect the solution number? Here we try to find the factors that determine the number of solutions.

Firstly, we change the augmented matrix $[A|b]$ into its reduced row echelon form R. Because R has already been simplified, thus we can rewrite it into the form like this:

$$\left[\begin{array}{cc|c} \boldsymbol{I} & \boldsymbol{F} & b \\ \boldsymbol{0} & \boldsymbol{0} & c \end{array}\right] \Longleftrightarrow \begin{bmatrix} \boldsymbol{I} & \boldsymbol{F} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} x_p \\ x_F \end{bmatrix} = \begin{bmatrix} b \\ c \end{bmatrix}$$

Denoted $\boldsymbol{I}$ an $r \times r$ identity matrix, then $r$ is equal to the number of pivots as well as rank($\boldsymbol{A}$). And the matrix remains $n-r$ non-pivot columns and $m-r$ zero rows. We can also transfer the matrix into the following equations:

$$\begin{cases} x_p + \boldsymbol{F}x_F = b \\ \vec{0}^{\mathrm{T}} = c \end{cases}.$$

The actual determinant of the number of the solution sets depends on the last $m-r$ rows. Here are total three types:

- If $m-r > 0$, and $c \neq \vec{0}_{m-r}$, then there is no solutions.
- If $c = \vec{0}_{m-r}$, there would be 1 or infinite solutions.
- If $m = r$, there would be 1 or infinite solutions/

The only possibility of "no solution" is when the type I occurs.

R Only "free rows" provide possibilities for a contradiction, i.e., it determines if there exist(s) solution(s).

Let's consider the columns of $\boldsymbol{R}$. Assume $c = \vec{0}_{m-r}$, or $m = r$ in order that the solution set is not empty.

- If $n-r = 0$, then the first linear equation becomes $x_p = b$. Then there would be only one unique solution.
- If $n-r > 0$, then the first equation becomes $x_p = b - \boldsymbol{F}x_F$. This provides some free variables ($x_F$). Thus it has infinitely many solutions.

R Columns of RREF provide possibilities for having infinitely many solutions.

## 7.1.3. Linear Independence

Let's consider a few examples here.

**Example 7.1** Simpler Way of Expressing Span.

**Observation:** $\text{span}(\{\begin{bmatrix}1\\1\\1\\1\end{bmatrix}, \begin{bmatrix}2\\2\\2\\2\end{bmatrix}\}) = \text{span}(\{\begin{bmatrix}1\\1\\1\\1\end{bmatrix}\})$.

This is true because the second column vector is a linear combination of the first column vector. So by definition, the second vector is in the span of the first vector. Those two spans are equivalent. The following simplification is similar.

**Observation:** $\text{span}(\{u,\ 2u\}) = \text{span}(\{u\})$.

It's a generalization of the first observation.

**Observation:** $\text{span}(\{u,\ v,\ u+v\}) = \text{span}(\{u,\ v\})$.

**Observation:** $\text{span}(\{u,\ v,\ 2u+v,\ 100u+v,\ u-25v,\ 4u+3v\}) = \text{span}(\{u,\ v\})$.

Because any of the vector inside the first span could be expressed by a linear combination of $u$ and $v$, thus two spans are equivalent.

According to the example above, we have the conclusion that if a vector can be expressed by other vectors inside that set, then we can simply ignore that vector while creating a span of this set. Therefore we have those important definitions below.

**Definition 7.1** [Linear Dependence]

Suppose $V$ is a linear space over $\mathbb{R}$. Suppose $\boldsymbol{u_1}, \boldsymbol{u_2}, \cdots, \boldsymbol{u_k} \in V$. We say $\boldsymbol{u_1}, \boldsymbol{u_2}, \cdots, \boldsymbol{u_k}$ are linearly independent if there exist real numbers $c_1, c_2, \cdots, c_k$ such that $(c_1, c_2, \cdots, c_k) \neq \mathbf{0}$ and $c_1\boldsymbol{u_1} + \cdots + c_k\boldsymbol{u_k} = 0$.

**Definition 7.2** [Linear Independence]

Suppose $V$ is a linear space over $\mathbb{R}$. Suppose $\boldsymbol{u_1}, \boldsymbol{u_2}, \cdots, \boldsymbol{u_k} \in V$. We say $\boldsymbol{u_1}, \boldsymbol{u_2}, \cdots, \boldsymbol{u_k}$ are linearly independent if for any real numbers $c_1, c_2, \cdots, c_k$ such that $(c_1, c_2, \cdots, c_k) \neq \mathbf{0}$ and $c_1\boldsymbol{u_1} + \cdots + c_k\boldsymbol{u_k} \neq 0$.

In short, $c_1\boldsymbol{u_1} + \cdots + c_k\boldsymbol{u_k} = \mathbf{0}$ only happens when $c_1 = \cdots = c_k = 0$.

**Example 7.2** Linear Dependence and Independence

- $\begin{bmatrix}1\\1\\1\\1\end{bmatrix}, \begin{bmatrix}2\\2\\2\\2\end{bmatrix}$ are linearly dependent.
- $\boldsymbol{u}$, $2\boldsymbol{u}$ are linearly dependent.
- $\begin{bmatrix}1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\\0\end{bmatrix}$ are linearly independent.
- $\begin{bmatrix}1\\1\\0\\0\end{bmatrix}, \begin{bmatrix}1\\1\\3\\0\end{bmatrix}, \begin{bmatrix}1\\1\\3\\2\end{bmatrix}$ are linearly independent.
- $\begin{bmatrix}1\\2\\1\\2\end{bmatrix}, \begin{bmatrix}3\\6\\-3\\-6\end{bmatrix}$ are linearly independent.
- $\boldsymbol{u}$, $2\boldsymbol{u}+2\boldsymbol{v}$, $2\boldsymbol{u}-5\boldsymbol{v}$ are linearly dependent. Because $-14\boldsymbol{u}+5(2\boldsymbol{u}+2\boldsymbol{v})+2(2\boldsymbol{u}-5\boldsymbol{v})=\mathbf{0}$
- $\boldsymbol{u}$, $\boldsymbol{v}$, $\mathbf{0}$ are linearly dependent. Because $0\boldsymbol{u}+0\boldsymbol{v}+10\times\mathbf{0}=\mathbf{0}$. (10 can be any non-zero term which is required according to the definition.)
- Columns of $\begin{bmatrix}1&0&a&0&c\\0&1&b&0&d\\0&0&0&1&e\\0&0&0&0&0\end{bmatrix}$ are linearly dependent. Because it contains non-pivot columns, which can be expressed by pivot columns.

By observation, we can figure out that the columns of a matrix $A$ are linearly dependent if and only if $A\boldsymbol{x}=\mathbf{0}$ has a non-zero solution. And they are linearly independent if and only if $A\boldsymbol{x}=\mathbf{0}$ has a unique solution $\boldsymbol{x}=\mathbf{0}$.

**Corollary 7.1** A square matrix $A$ is invertible if and only if the columns are linearly independent.

So how do we check whether a set of elements are linearly independent or not? If the elements are vectors in $\mathbb{R}^n$, then we just need to solve the linear system $A\boldsymbol{x} = \mathbf{0}$ as is showed above. However, if they are not in $\mathbb{R}^n$, then we need some extra tools.

**Theorem 7.1 — Linear Dependence and Span.** Suppose $V$ is a linear space over $\mathbb{R}$.
If $\boldsymbol{u_1}, \boldsymbol{u_2}, \cdots, \boldsymbol{u_k} \in V$ are linearly dependent, then there exists $t \in \{1, \cdots, k\}$ such that:
i) $\boldsymbol{u_t}$ is a linear combination of $\boldsymbol{u_1}, \cdots, \boldsymbol{u}_{t-1}, \boldsymbol{u}_{t+1}, \cdots, \boldsymbol{u_k}$.
ii) $\text{span}(\{\boldsymbol{u_1}, \boldsymbol{u_2}, \cdots, \boldsymbol{u_k}\}) = \text{span}(\{\boldsymbol{u_1}, \cdots, \boldsymbol{u}_{t-1}, \boldsymbol{u}_{t+1}, \cdots, \boldsymbol{u_k}\})$

In other words, if one element lies in the span of other elements, then the span of these elements can be further simplified.

**Corollary 7.2** If $\boldsymbol{u_1}, \boldsymbol{u_2}, \cdots, \boldsymbol{u_k} \in V$ are linearly independent, then $\text{span}(\{\boldsymbol{u_1}, \boldsymbol{u_2}, \cdots, \boldsymbol{u_k}\})$ can NOT be simplified. (i.e., expressed as the span of $k-1$ elements.)

Here we give the proof of the first part of the theorem. By the definition of span, the second part is easy to obtain. Suppose $\boldsymbol{u_1}, \boldsymbol{u_2}, \cdots, \boldsymbol{u_k} \in V$ are linearly dependent. Then it is telling exist a set of $c_1, c_2, \cdots, c_k$ such that

$$\begin{cases} c_1\boldsymbol{u_1} + \cdots + c_k\boldsymbol{u_k} = \mathbf{0} \\ c_t \neq 0, \text{ for some } t \in \{1, \cdots, k\} \end{cases}.$$

It means for a certain $t$, $c_t\boldsymbol{u_t} = -\sum_{\substack{j=1,\cdots,k \\ j\neq t}} c_j\boldsymbol{u_j}$, which can be further transformed to $\boldsymbol{u_t} = -\frac{1}{c_t}\sum_{\substack{j=1,\cdots,k \\ j\neq t}} c_j\boldsymbol{u_j}$. So it tells us $\boldsymbol{u_t}$ is a linear combination of other vectors, which gives us the first statement.

**R** **Linear Dependence / Independence in Geometry**

In a 2D plane:

- Two-dimensional vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are linearly dependent while they have the same or opposite directions.
- Two-dimensional vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are linearly independent while there is an angle (not equals to $\pi$) between those two vectors.

In a 3D space:

- Three-dimensional vectors $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$ are linearly dependent while they are within a plane.
- Three-dimensional vectors $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$ are linearly independent while they are not in the same plane.

### 7.1.4. Basis

It is clear that a certain linear space can be constructed by the span of a set of vectors. And that set is not unique. For instance, 4 unit vectors $\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$ span $\mathbb{R}^4$. But then we got a few questions:

**Question 1**: Can we use only 3 unit vectors to span $\mathbb{R}^4$?

Answer: No, because they cannot together form a spanning set. Thus, $\{\boldsymbol{e}_1, \cdots, \boldsymbol{e}_4\}$ is not just a spanning set, but also a minimal spanning set. It cannot be further simplified. We will give such a spanning set a name: **Basis**.

**Question 2**: Are there 3 vectors that span $\mathbb{R}^4$?

Answer: The answer is no. And after we learn the concept of basis, we will get familiar will it.

**Definition 7.3** [Basis]

Suppose $V$ is a linear space over $\mathbb{R}$. Suppose $\mathscr{U} \triangleq \{\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_k\} \subseteq V$. We say $\mathscr{U}$ is a basis if

(i) $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_k$ are linearly independent.

(ii) $\text{span}(\mathscr{U}) = V$.

Simply put, the basis elements are linearly independent and span the whole space.

R

(1) A basis is a **minimal spanning set** (deleting an element from the basis cannot span the whole space).

(2) A basis is also a **maximal independent set** (adding an element to the basis makes the set linearly dependent).

**Example 7.3** Basis Examples

1. $\{\boldsymbol{e_1}, \cdots, \boldsymbol{e_n}\}$ is a basis of $\mathbb{R}^n$. (called a "**standard basis**")

2. $\{\boldsymbol{u}\}$ is a basis of span$\{\boldsymbol{u}\}$, if $\boldsymbol{u} \neq 0$.

- claim: Any collection of linearly independent elements cannot contain 0. (Here 0 means zero element in the linear space.)

To check whether a set is a basis, we firstly check whether the elements in the candidate set are linearly independent. Then we check whether any element can be expressed as a linear combination of those elements in that set. if satisfied, we can thus claim it is a basis.

**Example 7.4** Find Basis

1. Find a basis of $\mathbb{R}^{2\times 2}$.

Let the set $B$ consist

$$B_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, B_{12} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, B_{21} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, B_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Then it is a basis of $\mathbb{R}^{2\times 2}$. $B$ is linearly independent. Also notice that $\forall A \in \mathbb{R}^{2\times 2}$,

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = aB_{11} + bB_{12} + cB_{21} + dB_{22}.$$

In general, $\{E_{ij}, i = 1, \cdots, m; j = 1, \cdots, n\}$ is a basis of $\mathbb{R}^{m\times n}$, where $E_{ij}$ is a matrix with only ($i$, $j$)-th entry being 1 and other entries being 0.

2. $V = P_1$ (polynomials of degree at most 1),

(1) $\boldsymbol{p_1}(x) = 1$, $\boldsymbol{p_2}(x) = x$, $\boldsymbol{p_3}(x) = 2 - 3x$. Is $U = \{\boldsymbol{p_1}, \boldsymbol{p_2}, \boldsymbol{p_3}\}$ a basis for $P_1$?

The answer is no. Because $U$ is linearly dependent since $\boldsymbol{p_3} = 2\boldsymbol{p_1} - 3\boldsymbol{p_2}$.

(2) $U = \{1, x\}$ is a basis for $P_1$.

3. Generally, for $V = P_n$ (polynomials of degree at most n),

$U = \{1, x, x^2, \cdots, x^n\}$ is a basis for $P_n$. It's a standard way of creating a basis for polynomials.

**Proposition 7.2** Let $\mathscr{U} = \{\boldsymbol{u_1}, \cdots, \boldsymbol{u_m}\}$ be a basis of a linear space $v$. Then any element $v \in V$ can be represented as a linear combination of $\boldsymbol{u_1}, \cdots, \boldsymbol{u_m}$.

By definition of basis, span($U$) = $V$. Then by definition of span, we can get the conclusion directly. However, this statement can be further strengthened.

**Proposition 7.3 — Unique Expression.** Let $\mathscr{U} = \{\boldsymbol{u_1}, \cdots, \boldsymbol{u_m}\}$ be a basis of a linear space $v$. Then any element $v \in V$ can be represented as a linear combination of $\boldsymbol{u_1}, \cdots, \boldsymbol{u_m}$.

If there exists $v \in V$, you have two different representations, such that

$$\begin{cases} v = \sum_{i=1}^{n} c_i \boldsymbol{u_i} \\ v = \sum_{i=1}^{n} b_i \boldsymbol{u_i} \\ (b_1, \cdots, b_n) \neq (c_1, \cdots, c_n) \end{cases}.$$

Then we take subtraction of the first two equations, we have $0 = \sum_{i=1}^{n} (c_i - b_i)\boldsymbol{u_i}$. Let $c_i - b_i$ be $\alpha_i$, then there exists a sequence of scalars $(\alpha_1, \cdots, \alpha_n)$ such that $\alpha_1 \boldsymbol{u_1} + \cdots + \alpha_n \boldsymbol{u_n} = 0$. Because $(\alpha_1, \cdots, \alpha_n)$ is not a zero vector, thus $\boldsymbol{u_1}, \cdots, \boldsymbol{u_n}$ are linearly dependent, contradicting to the definition of basis. Hence we prove the proposition by contradiction.

### 7.1.5. Summary

In this lecture, we study how to solve $A\boldsymbol{x} = \boldsymbol{b}$, linear independence and the concept of basis.

**1. Solving Ax = b**

- Procedure: Transform to RREF, then use inserting trick to write $\boldsymbol{x_p} + N(A)$.
- Different cases on the number of solutions.
- Under-determined Ax=0 has infinity many solutions.

**2. Linear dependence**

- Linear dependent elements: trivial linear combination gets 0.
- Related to $A\boldsymbol{x} = \boldsymbol{0}$ having infinitely many solutions.

**3. Basis**

- **Basis:** a spanning set that are linearly independent.

# 7.2. Lecture 14: Dimension and Rank

## 7.2.1. Review

Last lecture, we talked about the inserting trick of general system, the definition of linear independence, span, basis and how to solve a linear system in a faster and more efficient way. Noted that spanning sets are not unique. There are many ways to form a spanning set. In this lecture, we will introduce definitions of dimension and rank.

### 7.2.1.1. Linear Independence

By observation, we can figure out that the columns of a matrix $A$ are linearly dependent if and only if $A\boldsymbol{x} = \mathbf{0}$ has a non-zero solution. And they are linearly independent if and only if $A\boldsymbol{x} = \mathbf{0}$ has a unique solution $\boldsymbol{x} = \mathbf{0}$.

So how do we check whether a set of elements are linearly independent or not? If the elements are vectors in $\mathbb{R}^n$, then we just need to solve the linear system $A\boldsymbol{x} = \mathbf{0}$ as is showed above. However, if they are not in $\mathbb{R}^n$, then we need some extra tools.

**Theorem 7.2 — Linear Dependence and Span.** Suppose $V$ is a linear space over $\mathbb{R}$. If $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_k \in V$ are linearly dependent, then there exists $t \in \{1, \cdots, k\}$ such that:
i) $\boldsymbol{u}_t$ is a linear combination of $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{t-1}, \boldsymbol{u}_{t+1}, \cdots, \boldsymbol{u}_k$.
ii) $\text{span}(\{\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_k\}) = \text{span}(\{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_{t-1}, \boldsymbol{u}_{t+1}, \cdots, \boldsymbol{u}_k\})$

R **Linear Dependence / Independence in Geometry**

In a 2D plane:

- Two-dimensional vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are linearly dependent while they have the same or opposite directions.
- Two-dimensional vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are linearly independent while there is an angle (not equals to $\pi$) between those two vectors.

In a 3D space:

- Three-dimensional vectors $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$ are linearly dependent while they are within a plane.
- Three-dimensional vectors $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$ are linearly independent while they are not in the same plane.

### 7.2.1.2. Basis

Simply put, the basis elements are linearly independent and span the whole space.

**Example 7.5** Basis Examples

1. $\{\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n\}$ is a basis of $\mathbb{R}^n$. (called a "**standard basis**")
2. $\{\boldsymbol{u}\}$ is a basis of span$\{\boldsymbol{u}\}$, if $\boldsymbol{u} \neq 0$.

R Any collection of linearly independent elements cannot contain 0. (Here 0 means zero element in the linear space.)

To check whether a set is a basis, we firstly check whether the elements in the candidate set are linearly independent. Then we check whether any element can be expressed as a linear combination of those elements in that set. if satisfied, we can thus claim it is a basis.

**Proposition 7.4** Let $\mathscr{U} = \{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_m\}$ be a basis of a linear space $v$. Then any element $v \in V$ can be represented as a linear combination of $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_m$.

By definition of basis, span$(U) = V$. Then by definition of span, we can get the conclusion directly. However, this statement can be further strengthened.

**Proposition 7.5** Undetermined system $\boldsymbol{Ax} = \boldsymbol{b}$ with $m < n$, i.e., number of equations < number of unknowns, has **no solution** or **infinitely many solutions**.

**Proposition 7.6** Undetermined system $\boldsymbol{Ax} = \boldsymbol{b}$ with $m \geq n$, i.e., number of equations $\geq$ number of unknowns may have **no solution** or **unique solution** or **infinitely many solutions**.

The next interesting question is: What's the simplest/biggest basis?

Thus we introduce **dimension** to denote the *number of vectors in a basis.*

## 7.2.2. Dimension

First, consider the size of bases.

**Claim:** The set of four vectors $v_1$,$v_2$, $v_3$, $v_4 \subset \mathbb{R}^3$ is not a basis of $\mathbb{R}^3$

This is because the set of four vectors are impossible to be linearly independent (may or may not span $\mathbb{R}^3$).

**Example 7.6** Show that $\begin{bmatrix}1\\3\end{bmatrix}, \begin{bmatrix}5\\8\end{bmatrix}, \begin{bmatrix}15\\27\end{bmatrix}$ are linearly dependent.

Equivalently, the linear system $\begin{bmatrix}1 & 5 & 15\\3 & 8 & 27\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix} = \mathbf{0}$ has non-zero solution. Since there are more columns than pivots, the free variables exist and the system has infinitely many solutions.

We can extend the result to any other three vectors in $\mathbb{R}^2$, as there is at least one free variable, the three given vectors must be linearly dependent.

**Proposition 7.7** Columns of a wide matrix ($\mathbb{R}^{m\times n}, m > n$) are linearly **dependent**.

We states the key theorem in a formal way:

**Theorem 7.3** Suppose $m > n$.

If $\{\boldsymbol{v_1}, \boldsymbol{v_2}, \cdots, \boldsymbol{v_n}\}$ is a basis of a linear space $\mathbf{V}$, the nany $m$ elements $\boldsymbol{u_1}, \cdots, \boldsymbol{u_m}$ are linearly dependent.

We can also make some conclusions of the size of bases.

**Theorem 7.4 — bases have same size.** If $\{\boldsymbol{v_1}, \boldsymbol{v_2}, \cdots, \boldsymbol{v_n}\}$ and $\{\boldsymbol{u_1}, \boldsymbol{u_2}, \cdots, \boldsymbol{u_m}\}$ are bases of a linear space $\mathbf{V}$, then $m = n$.

*Proof.* Argue by contradiction, assume $m > n$. By theorem 7.3 $\{\boldsymbol{u_1}, \boldsymbol{u_2}, \cdots, \boldsymbol{u_m}\}$ are linearly dependent, contradiction to the definition of basis. Then we must have $m \leq n$ Similarly, $n \leq m$. Hence, $m = n$. ■

The big theorem states that every basis of a fixed linear space $\mathbf{V}$ has the same size, we then define the size as the dimension of $\mathbf{V}$.

**Definition 7.4** [dimension] Suppose $\mathbf{V}$ is a linear space.

If $\mathbf{V}$ has a basis $U$ with n elements, then we say the dimension of $\mathbf{V}$ is $n$, denoted as dim($\mathbf{V}$)$= n$, or $\mathbf{V}$ is n-dimensional.

(R) If there is no finite set of elements that can span $\mathbf{V}$, we say $\mathbf{V}$ is infinite dimensional.

**Example 7.7** dimension of some linear spaces

- $\dim(\{0\}) = 0$
- $\dim(\mathbb{R}^n) = n$
- $\dim(\mathbb{R}^{2\times 3}) = 6$
- $P \triangleq \{\text{All polynomials}\} = \text{span}\{1, x, x^2, \dots\} \implies \dim(P) = \infty.$
- $P_3 \triangleq \{\text{All polynomials with degree} \leq 3\} = \text{span}\{1, x, x^2, x^3\} \implies \dim(P) = 4.$
- $Q \triangleq \text{span}\{x^2, 1 + x^3 + x^{10}, x^{300}\} \implies \dim(Q) = 3.$

There are some abstract exercises for you:

**Example 7.8** dimension of common linear spaces:

1. $V_1$ : space of $n \times n$ diagonal matrices.

   As $\{\boldsymbol{e_1}, \boldsymbol{e_2}, \cdots, \boldsymbol{e_n}\}$ is a basis, then $\dim(V_1) = n$
2. $V_2$ : space of $n \times n$ upper triangular matrices.

   Since $\{\boldsymbol{e_ij} : i \leq j\}$ form a basis, then $\dim(V_2) = \sum_{k=1}^{n} k = \frac{1}{2}n(n+1)$
3. $P_n = \{\sum_{i=0}^{n} a_i x^i \mid a_i \in \mathbb{R}\}$.

   $\{1, x, x^2, \cdots, x^n\}$ form a basis, then $\dim(P_n) = n + 1$
4. C: Column space of $\begin{bmatrix} 1 & 0 & a & 0 & c \\ 0 & 1 & b & 0 & d \\ 0 & 0 & 0 & 1 & e \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

   Let $\boldsymbol{c_i}$ denote the $i^{th}$ column of the matrix. Since the $\boldsymbol{c_3}$ and $\boldsymbol{c_5}$ column are linear combination of $\boldsymbol{c_1}, \boldsymbol{c_2}, \boldsymbol{c_4}$, and $\boldsymbol{c_1}, \boldsymbol{c_2}, \boldsymbol{c_4}$ are linearly independent. Then $\text{Col}\{\boldsymbol{c_1}, \boldsymbol{c_2}, \boldsymbol{c_3}, \boldsymbol{c_4}, \boldsymbol{c_5}\} = \text{Col}\{\boldsymbol{c_1}, \boldsymbol{c_2}, \boldsymbol{c_4}\}$,and $\dim(C) = 3$.

We introduce a quicker way to check basis:

**Proposition 7.8** Consider a linear space V with dim(V) = n

- Any linearly independent set can be extended to a basis, by adding more linearly independent elements if necessary.

- Any spanning set in V can be reduced to a basis, by deleting some elements if necessary.

**Example 7.9** $v_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$ is not a basis of $\mathbb{R}^3$.

We can add $v_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, which is ind. of $v_1$. But $v_1, v_2$ still don't form a basis.

If we add one more vector $v_3 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, then $v_1, v_2, v_3$ form a basis of $\mathbb{R}^3$. ■

**Proposition 7.9** Consider a linear space $\mathbf{V}$ with dim($\mathbf{V}$) = n.

- If $\boldsymbol{u_1}, \cdots, \boldsymbol{u_n}$ are linearly independent, then they form a basis of $\mathbf{V}$.
- If $\boldsymbol{u_1}, \cdots, \boldsymbol{u_n} \in \mathbf{V}$ can span $\mathbf{V}$, then they form a basis of $\mathbf{V}$.

**Example 7.10** $\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}$ are ind. $\implies$ they span $\mathbb{R}^3$. ■

Recall that we can determine the invertibility of matrices by check whether the columns are linearly independent. In this lecture, we have more methods to check whether $\mathbf{A} \in \mathbb{R}^{n \times n}$ is invertible.

**Theorem 7.5** Equivalent Conditions for Invertibility
Let $\mathbf{A} \in \mathbb{R}^{n \times n}$. The following statements are equivalent:

1. $\mathbf{A}$ is invertible.
2. The linear system $Ax = 0$ has a unique solution $x = 0$.
3. $\mathbf{A}$ is a product of elementary matrices.
4. A has $n$ pivots; or equivalently: rank($\mathbf{A}$) $= n$.

5. The columns of $\mathbf{A}$ span $\mathbb{R}^n$.
6. The columns of $\mathbf{A}$ are linearly independent.
7. The columns of $\mathbf{A}$ form a basis of $\mathbb{R}^n$.
8. $\dim(c(\mathbf{A})) = n$.

R dim of space $\neq$ dim of the space it lives in.
For example, the line in $\mathbb{R}^{100}$ has dim 1.

## 7.2.3. Rank

### 7.2.3.1. Relation of Dimension and Rank

Theorem (7.5) says: $\text{rank}(\mathbf{A}) = n \iff \dim(c(\mathbf{A})) = n$. Moreover,

$$\begin{aligned} rank(\mathbf{A}) &\iff n \text{ pivots} \\ &\iff Ax = 0 \text{ has unique solution } 0 \\ &\iff n \text{ columns are linearly independent} \\ &\iff C(\mathbf{A}) \text{ has dimension } n \end{aligned}$$

**Definition 7.5** [Rank] The rank of matrix $A$ is defined as the **number of nonzero pivots of rref of** $A$.

**Example 7.11**

$$A = \begin{bmatrix} 1 & 3 & 3 & 4 \\ 2 & 6 & 9 & 7 \\ -1 & -3 & 3 & 4 \end{bmatrix} \xRightarrow{\text{row transform}} U = \begin{bmatrix} 1 & 3 & 0 & -1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$U$ has two pivots, hence $\text{rank}(A) = \text{rank}(U) = 2$.

However, the definition for rank is too complicated, can we define rank of $A$ directly?

**Key question: What quantity is not changed under row transformation?**

*Answer:* Dimension of row space.

**Definition 7.6** [column space] The **column space** of a matrix is the subspace of $\mathbb{R}^n$ spanned by the columns.

In other words, suppose $\boldsymbol{A} = \left[\begin{array}{c|c|c} a_1 & \dots & a_n \end{array}\right]$, the column space of $\boldsymbol{A}$ is given by

$$\mathcal{C}(\boldsymbol{A}) = \operatorname{span}\{a_1, a_2, \dots, a_n\}.$$

**Definition 7.7** [row space] The **row space** of a matrix is the subspace of $\mathbb{R}^n$ spanned by the rows.

Suppose $\boldsymbol{A} = \left[\begin{array}{c} a_1 \\ \hline \dots \\ \hline a_n \end{array}\right]$, the row space of $\boldsymbol{A}$ is given by

$$\mathcal{R}(\boldsymbol{A}) = \operatorname{span}\{a_1, a_2, \dots, a_n\}.$$

The **row space** of $\boldsymbol{A}$ is essentially $\mathcal{R}(\boldsymbol{A}) := \mathcal{C}(\boldsymbol{A}^{\mathrm{T}})$, i.e., the column space of $\boldsymbol{A}^{\mathrm{T}}$.

**Proposition 7.10** Row transforamtion doesn't change the row space

*Proof.* After row transformation, **new rows are linear combinations of old rows.**

Hence we have $\mathcal{R}(\text{new rows}) \subset \mathcal{R}(\text{old rows})$.

More specifically, assuming $\boldsymbol{A} \xrightarrow{\text{Row Transfom}} \boldsymbol{B}$, then we have $\mathcal{R}(\boldsymbol{B}) \subset \mathcal{R}(\boldsymbol{A})$.

Since row transformations are invertible, we also have $\boldsymbol{B} \xrightarrow{\text{Row Transfom}} \boldsymbol{A}$, thus we have $\mathcal{R}(\boldsymbol{A}) \subset \mathcal{R}(\boldsymbol{B})$.

In conclusion, we obtain $\mathcal{R}(\boldsymbol{B}) = \mathcal{R}(\boldsymbol{A})$. ∎

Hence $\operatorname{rank}(\boldsymbol{A}) = \text{pivots of } \boldsymbol{U} = \dim(\operatorname{row}(\boldsymbol{U})) = \dim(\operatorname{row}(\boldsymbol{A}))$.

Hence we have a much simpler definition for rank:

**Definition 7.8** [rank] The **dimension of the row space** is the **rank of a matrix**, i.e.,

$$\operatorname{rank}(\boldsymbol{A}) = \dim(\mathcal{R}(\boldsymbol{A})).$$

In the example (7.10), we find $\dim(\operatorname{row}(\boldsymbol{A})) = \dim(\operatorname{col}(\boldsymbol{A})) = 2$, is this a coincidence? *The* *fundamental theorem of linear algebra* gives this answer:

**Theorem 7.6** The row space and column space both have the **same** dimension $\boldsymbol{r}$.

We call $\dim(\mathcal{C}(\boldsymbol{A}))$ as *column rank*; $\dim(\mathcal{R}(\boldsymbol{A}))$ as *row rank*.

In brevity, **column rank=row rank= rank**, i.e.,

$$\dim(\mathcal{C}(\boldsymbol{A})) = \dim(\mathcal{R}(\boldsymbol{A})) = \operatorname{rank}(\boldsymbol{A}), \quad \text{for matrix } \boldsymbol{A}$$

Let's discuss an example to have an idea of proving it.

**Example 7.12**

$$\boldsymbol{A} = \begin{bmatrix} 1 & 3 & 3 & 4 \\ 2 & 6 & 9 & 7 \\ -1 & -3 & 3 & 4 \end{bmatrix} \xrightarrow{\text{row transform}} \boldsymbol{U} = \begin{bmatrix} 1 & 3 & 0 & -1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

We notice that **column rank of $\boldsymbol{A} = 2$** and **column rank of $\boldsymbol{U} = 2$**.

Why do they have the same **column space dimension**?

**Wrong reason: $\boldsymbol{A}$ and $\boldsymbol{U}$ has the same column space.** This is false. For example, the first column of $\boldsymbol{A}$ is $\begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} \notin \operatorname{col}(\boldsymbol{U})$. The column spaces of $\boldsymbol{A}$ and $\boldsymbol{U}$ are **different**, but the dimension of them are **equal**.

**Right reason: $\boldsymbol{Ax} = \boldsymbol{0}$ iff. $\boldsymbol{Ux} = \boldsymbol{0}$.** The same combinations of the columns are zero (or nonzero) for $\boldsymbol{A}$ and $\boldsymbol{U}$.

In other words, the $r$ pivot columns (for both $\boldsymbol{A}$ and $\boldsymbol{U}$) are independent; the $(n-r)$ free columns (for both $\boldsymbol{A}$ and $\boldsymbol{U}$) are dependent.

For example, for $\boldsymbol{U}$, column 1 and 3 are ind.(pivot columns); column 2 and 4 are dep.(free columns).

For $\boldsymbol{A}$, column 1 and 3 are also ind.(pivot columns); column 2 and 4 are also dep.(free columns).

■

This example shows that **Row transformation doesn't change independence relations of columns**. We give a formal proof below:

**Proposition 7.11** Suppose matrix $\boldsymbol{A}$ is converted into $\boldsymbol{B}$ by row transformation. If a set of columns of $\boldsymbol{A}$ are ind. then so are the corresponding columns of $\boldsymbol{B}$.

*Proof.* Assume $\boldsymbol{A} = \left[\begin{array}{c|c|c} a_1 & \dots & a_n \end{array}\right], \boldsymbol{B} = \left[\begin{array}{c|c|c} b_1 & \dots & b_n \end{array}\right]$.

Without loss of generality (We often denote it as "WLOG"), we assume $a_1, a_2, \dots, a_k$ are ind.(We can achieve it by switching columns.)

We define the sub-matrices $\hat{\boldsymbol{A}} = \left[\begin{array}{c|c|c} a_1 & \dots & a_k \end{array}\right]$ and $\hat{\boldsymbol{B}} = \left[\begin{array}{c|c|c} b_1 & \dots & b_k \end{array}\right]$.

1. Notice that $\hat{\boldsymbol{A}}$ could be converted into $\hat{\boldsymbol{B}}$ by row transformation.

   Hence $\hat{\boldsymbol{A}}\boldsymbol{x} = \boldsymbol{0}$ and $\hat{\boldsymbol{B}}\boldsymbol{x} = \boldsymbol{0}$ has the same solutions.

2. On the other hand, $a_1, a_2, \dots, a_k$ are ind. columns.

   Hence $\hat{\boldsymbol{A}}\boldsymbol{x} = \boldsymbol{0}$ has the only zero solution.

Combining (1) and (2), $\hat{\boldsymbol{B}}\boldsymbol{x} = \boldsymbol{0}$ has the only zero solution. Hence $b_1, b_2, \dots, b_k$ are ind. ■

We can answer why the coincidence shown in the example, i.e., $\boldsymbol{A}$ and $\boldsymbol{U}$ has the same column space dimension:

**Proposition 7.12** Row transformation doesn't change the column rank.

*Proof.* Assume $\boldsymbol{A} \xrightarrow{\text{row transform}} \boldsymbol{B}$.

Suppose $\dim(\mathcal{C}(\boldsymbol{A})) = r$, then we pick $r$ ind. columns of $\boldsymbol{A}$. After row transformation, they are still ind. Hence $\dim(\mathcal{C}(\boldsymbol{B})) \geq r = \dim(\text{col}(\boldsymbol{A}))$.

Since row transformations are invertible, we get $\boldsymbol{B} \xrightarrow{\text{row transform}} \boldsymbol{A}$. Similarly, $\dim(\mathcal{C}(\boldsymbol{A})) \geq \dim(\mathcal{C}(\boldsymbol{B}))$.

Hence $\dim(\mathcal{C}(\boldsymbol{A})) = \dim(\mathcal{C}(\boldsymbol{B}))$. ■

Combining proposition (7.10) and (7.12), we can proof theorem (7.6):

*Proof for theorem 7.6.* Assume $\boldsymbol{A} \xrightarrow{\text{row transform}} \boldsymbol{U}$(rref).

- Proposition (7.10) $\implies \dim(\mathcal{R}(\boldsymbol{A})) = \dim(\mathcal{R}(\boldsymbol{U}))$.

- Proposition (7.12) $\implies \dim(\mathcal{C}(\boldsymbol{A}) = \dim(\mathcal{C}(\boldsymbol{U}))$.

- Notice that $\dim(\mathcal{R}(\boldsymbol{U}))$ denotes the number of pivots, $\dim(\mathcal{C}(\boldsymbol{U}))$ denotes the number of pivot columns. Obviously, $\dim(\mathcal{R}(\boldsymbol{U})) = \dim(\mathcal{C}(\boldsymbol{U}))$.

Hence $\dim(\mathcal{R}(\boldsymbol{A})) = \dim(\mathcal{C}(\boldsymbol{A}))$. ■

**R** $\dim(\mathcal{R}(\boldsymbol{U}))$ essentially denotes the number of "real" equations. $\dim(\mathcal{C}(\boldsymbol{U}))$ denotes the number of "real" variables.

So Theorem 7.6 implies that the number of "real" equations should equal to the number of "real" variables.

# Chapter 8

# Week8

## 8.1. Lecture 15: Four Fundamental Subspaces

### 8.1.1. Full Rank and Rank-1 Matrices

#### 8.1.1.1. Full Rank Matrix

**Definition 8.1** (full rank matrix)

Let $\mathbf{A} \in R^{mxn}$.

If rank(A)=m, then we say A has full row rank.

If rank(A)=n, then we say A has full column rank.

If rank(A)=min{m,n}, then we say A has full rank.

From Lecture 13:

**Recall**: m-r, n-r are critical for # of solutions.

m-r shall be interpreted as (# of rows)-(row rank)

n-r shall be interpreted as (# of columns)-(col rank)

we can convert $\boldsymbol{Ax} = \boldsymbol{b}$ into $\boldsymbol{Rx} = \boldsymbol{c}$. WLOG, we switch columns of $\boldsymbol{R}$ to put pivot columns in the left-most:

$$\begin{bmatrix} 1 & & & \times & \times \\ & \ddots & & \times & \times \\ & & 1 & \times & \times \\ 0 & 0 & 0 & 0 & 0 \\ \cdots & & & & \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} c_1 \\ \vdots \\ c_r \\ c_{r+1} \\ \vdots \\ c_n \end{bmatrix},$$

where $x_1.x_2.\ldots,x_r$ are pivot variables. Hence, we have $(n-r)$ free variables, and $N(\boldsymbol{A})$ is spanned by $(n-r)$ special vectors $y_1,y_2,\ldots,y_{n-r}$.

**Q1: If m-r, how many solutions does Ax=b have?**

If m=r, then m-r=0, $\begin{bmatrix} 1 & & & \times & \times \\ & \ddots & & \times & \times \\ & & 1 & \times & \times \\ 0 & 0 & 0 & 0 & 0 \\ \cdots & & & & \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ becomes

$\begin{bmatrix} 1 & & & \times & \times \\ & \ddots & & \times & \times \\ & & 1 & \times & \times \end{bmatrix}$. The system becomes $\begin{bmatrix} 1 & & & \times & \times \\ & \ddots & & \times & \times \\ & & 1 & \times & \times \end{bmatrix}$x=$\begin{bmatrix} c_1 \\ \vdots \\ c_r \end{bmatrix}$, No "$0 = c_{r+1} \cdots c_n$", so no possibility of "zero solution"

**Q2: If n=r, how many solutions does $Ax = b$ have?** If $n = r$,then $n - r = 0$, then $\begin{bmatrix} 1 & & & \times & \times \\ & \ddots & & \times & \times \\ & & 1 & \times & \times \\ 0 & 0 & 0 & 0 & 0 \\ \cdots & & & & \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

becomes $\begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \\ 0 & 0 & 0 \\ \cdots & & \\ 0 & 0 & 0 \end{bmatrix}$. The system becomes $\begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \\ 0 & 0 & 0 \\ \cdots & & \\ 0 & 0 & 0 \end{bmatrix}$ x= $\begin{bmatrix} c_1 \\ \vdots \\ c_r \\ c_{r+1} \\ \vdots \\ c_n \end{bmatrix}$

The only possibility of "No solution" is $m > r$ and $\begin{bmatrix} c_{r+1} \\ \vdots \\ c_n \end{bmatrix} \neq 0$.

| Four types for R | $[I]$ | $\begin{bmatrix} I & F \end{bmatrix}$ | $\begin{bmatrix} I \\ 0 \end{bmatrix}$ | $\begin{bmatrix} I & F \\ 0 & 0 \end{bmatrix}$ |
|---|---|---|---|---|
| **Rank** | $r = m = n$ | $r = m < n$ | $r = n < m$ | $r < m, r < n$ |
| **# of solutions** | 1 | $\infty$ | 0 or 1 | 0 or $\infty$ |
| **# of solutions when $\boldsymbol{b}'_{(m-r)\times 1} = \boldsymbol{0}$** | 1 | $\infty$ | 1 | $\infty$ |

Only "free rows" provide possibility of contradiction.

To get infinite solutions, need "n>r".

Only "free variables" provide possibility of many solutions.

**Proposition 8.1** If rank(A)=n, i.e., full column rank, then $Ax = b$ has at most one solution.

**Proposition 8.2** If rank(A)=m, i.e., full row rank, then $Ax = b$ has at least one solution.

**Corollary 8.1** If rank(A)=m=n, then $Ax = b$ has exactly one solution.

### 8.1.1.2. Matrices of rank 1

**Example 8.1**

$$\boldsymbol{A} = \begin{bmatrix} 2 & 1 & 1 \\ 4 & 2 & 2 \\ 8 & 4 & 4 \\ -2 & -1 & -1 \end{bmatrix} \xlongequal{\boldsymbol{v}^{\mathrm{T}} = \begin{bmatrix} 2 & 1 & 1 \end{bmatrix}} \begin{bmatrix} \boldsymbol{v}^{\mathrm{T}} \\ 2\boldsymbol{v}^{\mathrm{T}} \\ 4\boldsymbol{v}^{\mathrm{T}} \\ -\boldsymbol{v}^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 4 \\ -1 \end{bmatrix} \boldsymbol{v}^{\mathrm{T}} \xlongequal{\boldsymbol{u} = \begin{bmatrix} 1 & 2 & 4 & -1 \end{bmatrix}^{\mathrm{T}}} \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$$

Here $\text{rank}(\boldsymbol{A}) = 1$.

**Proposition 8.3** Every rank 1 matrix $\boldsymbol{A}$ has the form $\boldsymbol{A} = \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}} =$ column vector $\times$ row vector.

You may prove it directly by SVD decomposition (we will learn it later, but note that most theorems or propositions could be proved by SVD). Alternatively, we have another proof:

$rank = 1$ means 1 pivot $\longrightarrow$ of of method 1

$rank = 1$ means dim(C(A))=1 $\longrightarrow$ of of method 2

*Proof.* method 1

"pivot" comes from Gaussian Elimination.

$$A \overset{row operations}{\longrightarrow} U = \begin{bmatrix} 1 & x & x & \cdots & x \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & & 0 \end{bmatrix} = \begin{bmatrix} \vec{u_1}^\top \\ \vec{0}^\top \\ \vdots \\ \vec{0}^\top \end{bmatrix}$$

Matrix representation:

$$A = P^{-1}LU = \begin{bmatrix} \vec{v_1} & \vec{v_2} & \cdots & \vec{v_n} \end{bmatrix} \begin{bmatrix} \vec{u_1}^\top \\ \vec{0}^\top \\ \vdots \\ \vec{0}^\top \end{bmatrix} = \vec{v_1}\vec{u_1}^\top$$ ■

*Proof.* method 2

By definition of dimension, dim(C(A))=1 $\Longleftrightarrow$ # of vectors in a basis of C(A) is 1.

Suppose $\vec{v} \in$ C(A), $\vec{v} \neq 0$ , then C(A) must be $span(\{\vec{v}\})$

C(A)=$span(\{\vec{v}\})$={$\alpha\vec{v} : \alpha \in R$ }, A=$\begin{bmatrix} \alpha_1\vec{v} & \alpha_2\vec{v} & \cdots & \alpha_n\vec{v} \end{bmatrix} = \vec{v}\begin{bmatrix} \alpha_1, \alpha_2, \cdots, \alpha_n \end{bmatrix} = \vec{v}\vec{u}^\top$(outer product) ■

Reverse: rank($\vec{v}\vec{u}^\top$)=1 is also true.

**Proposition 8.4** For any non-zero vectors $\vec{v} \in R^{mx1}, \vec{u} \in R^{nx1}$, the outer product $\vec{v}\vec{u}^\top \in R^{mxn}$ has rank 1.

#### 8.1.1.3. Rank-2 and Rank-r Matrices

Question: What about the form of rank 2? *Answer:* it has the form $\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + \boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}}$.

Question: What about the form of rank r? *Answer:* sum of r rank-1 matrices.

### 8.1.2. Nullity

**Definition 8.2** The dimension of the null space dim(N(A)) is called the "nullity" of A.

From Lecture 12: Solution of $AX = 0$

**Example 8.2** Solve the system

$$\begin{bmatrix} 1 & 3 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \boldsymbol{x} = \mathbf{0}.$$

$$\begin{bmatrix} 1 & 3 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \implies \begin{cases} x_1 + 3x_2 = 0 \\ x_3 + x_4 = 0 \end{cases}$$

Then we express the **pivot variables** in the form of **free variables.**

Note that the pivot columns in $\boldsymbol{R}$ are column 1 and 3, so the pivot variable is $x_1$ and $x_3$. The free variable is the remaining variable, say, $x_2$ and $x_4$.

The expressions for $x_1$ and $x_3$ are given by:

$$\begin{cases} x_1 = -3x_2 \\ x_3 = -x_4 \end{cases}$$

Hence, all solutions to $\boldsymbol{R}\boldsymbol{x} = \mathbf{0}$ are

$$\boldsymbol{x} = \begin{bmatrix} -3x_2 \\ x_2 \\ -x_4 \\ x_4 \end{bmatrix} = x_2 \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 0 \\ 0 \\ -1 \\ 1 \end{bmatrix}$$

where $x_2$ and $x_4$ can be taken arbitararily.

$$\text{Null(R)=span}\{ \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ -1 \\ 1 \end{bmatrix} \}$$

Nullity=dim(N(R))=2

"Nullity" is the "true size" of solution set.

Previous lecture shows: dim(N(A))=n-r.Equivalently:

**Theorem 8.1** (rank-nullity theorem)

Suppose A has n columns. Then

$$rank(A) + dim(N(A)) = n$$

**Corollary 8.2** If rank(A)=n,then dim(N(A))=0, i.e.,Ax=0 has a trivial solution **0**.

**Theorem 8.2** (Equivalent Conditions for Invertibility)

Let $A \in R^{nxn}$.

(1) A is invertible.

(2) The linear system $Ax = 0$ has a unique solution $x = 0$.

(3) A is a product of elementary matrices.

(4) A has n pivots; or equivalently: rank(A)=n.

(5) The columns of A span $R^n$.

(6) The columns of A are linearly independent;

(7) The columns of A form a basis;

(8) dim(C(A)) = n.

(9) dim(N(A)) = 0, or N(A)={0}

## 8.1.3. Orthogonality

To truly understand rank-nullity theorem, we need to introduce: Orthogonality.

**Definition 8.3** (Orthogonal)

Two vectors **u**,**v** $\in R^n$ are orthogonal if $u^\top v = 0$. Denote **u** $\perp$ **v**

$\cos\theta = \frac{x^\top y}{\|x\|\|y\|}$

Thus x and y are orthogonal $\iff x^\top y = 0 \iff \theta$ is the right angle.

**Example 8.3** Vectors$[3,2]^\top$ and $[-4,6]^\top$ are orthogonal in $R^2$.

Vectors$[2,-3,1]^\top$ and $[1,1,1]^\top$ are orthogonal in $R^3$.

**First of all, we will introduce some properties of norm and inner multiplication.**

An important case is the inner product of a vector with *itself*. The inner product $\langle \boldsymbol{x}, \boldsymbol{x} \rangle$ gives the *length of* $\boldsymbol{v}$ *squared*:

**Definition 8.4** [length/norm] The **length(norm)** $\|\boldsymbol{x}\|$ of a vector $\boldsymbol{x} \in \mathbb{R}^n$ is the square root of $\langle \boldsymbol{x}, \boldsymbol{x} \rangle$:

$$\textbf{length} = \|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle} = \sqrt{x_1^2 + \cdots + x_n^2}.$$

### 8.1.3.1. Function space

We can talk about inner product between functions under the function space. For example, if we define $V = \{f(t) \mid \int_0^1 f^2(t)dt < \infty\}$, then we can define inner product and norm under $V$:

**Definition 8.5** [Inner product; norm] The **inner product** and the **norm** of $f(x), g(x)$ under the function space $V = \{f(t) \mid \int_0^1 f^2(t)dt < \infty\}$, are defined as:

$$\langle f, g \rangle = \int_0^1 f(x)g(x)dx \qquad \text{and} \qquad \|f\|^2 = \sqrt{\int_0^1 f^2(x)dx}$$

Moreover, when $\langle f, g \rangle = 0$, we say two functions are **orthogonal** and denote it as $f \perp g$.

### 8.1.3.2. Cauchy-Schwarz Inequality

In $\mathbb{R}^2$, suppose $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$, then we set:

$$\begin{cases} x_1 = \|\boldsymbol{x}\| \cos\theta \\ x_2 = \|\boldsymbol{x}\| \sin\theta \end{cases} \quad \begin{cases} y_1 = \|\boldsymbol{y}\| \cos\varphi \\ y_2 = \|\boldsymbol{y}\| \sin\varphi \end{cases}$$

The inner product of $\boldsymbol{x}$ and $\boldsymbol{y}$ is given by:

$$\begin{aligned} < \boldsymbol{x}, \boldsymbol{y} > = \boldsymbol{x}^{\mathrm{T}} \boldsymbol{y} = x_1 x_2 + y_1 y_2 \\ &= \|\boldsymbol{x}\| \|\boldsymbol{y}\| (\cos\theta \cos\varphi + \sin\theta \sin\varphi) \\ &= \|\boldsymbol{x}\| \|\boldsymbol{y}\| \cos(\theta - \varphi) \end{aligned}$$

Since $|\cos(\theta - \varphi)|$ never exceeds 1, the cosine formula gives a great inequality:

**Theorem 8.3 — Cauchy Schwarz Inequality.**

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle \leq \|\boldsymbol{x}\| \|\boldsymbol{y}\|$$

holds for two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$.

*Proof.* Firstly, we want to find optimizer $t^*$ such that

$$\min \|\boldsymbol{x} - t\boldsymbol{y}\|^2 = \|\boldsymbol{x} - t^*\boldsymbol{y}\|^2.$$

Note that

$$\begin{aligned}
\|\boldsymbol{x} - t\boldsymbol{y}\|^2 = \langle \boldsymbol{x} - t\boldsymbol{y}, \boldsymbol{x} - t\boldsymbol{y} \rangle &= \langle \boldsymbol{x}, \boldsymbol{x} \rangle + \langle -t\boldsymbol{y}, \boldsymbol{x} \rangle + \langle \boldsymbol{x}, -t\boldsymbol{y} \rangle + \langle -t\boldsymbol{y}, -t\boldsymbol{y} \rangle \\
&= \|\boldsymbol{x}\|^2 - t\langle \boldsymbol{y}, \boldsymbol{x} \rangle - t\langle \boldsymbol{x}, \boldsymbol{y} \rangle + t^2\|\boldsymbol{y}\|^2 \\
&= \|\boldsymbol{x}\|^2 - 2t\langle \boldsymbol{x}, \boldsymbol{y} \rangle + t^2\|\boldsymbol{y}\|^2
\end{aligned}$$

Hence the minimizer $t*$ must satisfy

$$\Delta = 0 \implies t^* = \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{y}\|^2}$$

Hence we have

$$\begin{aligned}
\|\boldsymbol{x} - t\boldsymbol{y}\|^2_{\min} = \|\boldsymbol{x} - t^*\boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 - \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle^2}{\|\boldsymbol{y}\|^2} \\
= \frac{\|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2 - \langle \boldsymbol{x}, \boldsymbol{y} \rangle^2}{\|\boldsymbol{y}\|^2} \geq 0 \\
\implies \|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2 \geq \langle \boldsymbol{x}, \boldsymbol{y} \rangle^2
\end{aligned}$$

Or equivalently,

$$|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| \leq \|\boldsymbol{x}\| \|\boldsymbol{y}\|.$$

■

R

**Cauchy-Schwarz inequality also holds for functions.** If we consider functions $f,g$ as vectors, then

$$\left[\int_0^1 f(t)g(t)dt\right] \leq \int_0^1 f^2 dt \int_0^1 g^2 dt$$

**The normalization of inner product is bounded by 1.** Since $|\langle \boldsymbol{x}, \boldsymbol{y}\rangle| \leq \|\boldsymbol{x}\|\|\boldsymbol{y}\|$, we have

$$-1 \leq \frac{\langle \boldsymbol{x}, \boldsymbol{y}\rangle}{\|\boldsymbol{x}\|\|\boldsymbol{y}\|} \leq 1$$

If we define $\frac{\langle \boldsymbol{x}, \boldsymbol{y}\rangle}{\|\boldsymbol{x}\|\|\boldsymbol{y}\|} := \cos\theta$, then $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \|\boldsymbol{x}\|\|\boldsymbol{y}\|\cos\theta$, the angle $\theta$ is said to be the intersection angle between $\boldsymbol{x}$ and $\boldsymbol{y}$.

Cauchy-Schwarz equality holds for **Hilbert space**, which will be discussed in other courses.

**Proposition 8.5** If two vectors **u,v** $\in \mathrm{R}^n$ are orthogonal, then $\|u\|^2 + \|v\|^2 = \|u+v\|^2$

*Proof.* Recall:$\|\vec{a}\|^2 = \sum_{i=1}^n a_i^2 = \langle \vec{a}, \vec{a}\rangle$,for any$\vec{a} \in R^n$.
Thus $\|u+v\|^2 = \langle \vec{u}+\vec{v}, \vec{u}+\vec{v}\rangle = \langle \vec{u}, \vec{u}\rangle + \langle \vec{u}, \vec{v}\rangle + \langle \vec{v}, \vec{u}\rangle + \langle \vec{v}, \vec{v}\rangle = \|\vec{u}\|^2 + \|\vec{v}\|^2$

■

**Definition 8.6** [Orthogonal vectors] Two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ are orthogonal when their inner product is zero:

$$\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \sum_{i=1}^n x_i y_i = 0.$$

R Note that the inner product of two **vectors** satisfies the *commutative rule*. In other words, $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \langle \boldsymbol{y}, \boldsymbol{x}\rangle$ for vectors $\boldsymbol{x}$ and $\boldsymbol{y}$. The inner product defined for matrices may not satisfy the *commutative rule*. Generally, if the result of inner product is a scalar, then inner product satisfies commutative rule.

### 8.1.3.3. orthogonal subspaces

If $\boldsymbol{x}$ is in the solution set of $Ax = 0$, then $\begin{bmatrix} a_{(1)}^\top \\ a_{(2)}^\top \\ \vdots \\ a_{(n)}^\top \end{bmatrix} x = 0$ a$\perp$ x,$\forall a \in Row(A), x \in N(A)$.

After defining inner product, we can discuss the orthogonality for space:

**Definition 8.7** [Orthogonal subspaces] Two subspaces $\boldsymbol{U}$ and $\boldsymbol{V}$ of a vector space are **orthogonal** if every vector $\boldsymbol{u}$ in $\boldsymbol{U}$ is *perpendicular* to every vector $\boldsymbol{v}$ in $\boldsymbol{V}$:

$$\textbf{Orthogonal subspaces} \quad \langle \boldsymbol{u}, \boldsymbol{v} \rangle = 0 \text{ for all } \boldsymbol{u} \text{ in } \boldsymbol{U} \text{ and all } \boldsymbol{v} \text{ in } \boldsymbol{V}.$$

**Row(A) $\perp$ N(A)**

**Example 8.4** Eg1: Floor and vertical flag are orthogonal

Eg2: One line on the floor and vertical flag are orthogonal

Non-example:Floor and wall are not orthogonal. i.e., in 3D sapce,xy plane and xz plane are NOT orthogonal.

$span\{e_1, e_2\}$ is not perpendicular to$span\{e_2, e_3\}$

*Proof.* By definition, if $span\{e_1, e_2\}$ is perpendicular to$span\{e_2, e_3\}$,pick $e_1$ from $span\{e_1, e_2\}$ and pick $e_1$ from $span\{e_2, e_3\}$, $e_1^{1}{}_1^{=1 \neq 0}$. This violates the definition of orthogonal subspace, thus not orthogonal. ∎

### 8.1.3.4. orthogonal complement

**Definition 8.8** (orthogonal complement)

For any subspaces V$\in$ $R^n$, the set if vectors that are orthogonal to V

$$\{u \in R^n | u \perp v, \forall v \in V\}.$$

is called the orthogonal complement of V, denoted as $U^{\perp}$

Fact: $Row(A)^{\perp} = N(A)$

R A subspace has a unique orthogonal complement!

**Example 8.5** Eg1: Floor = $(vertical\ flag)^{\perp}$. In 3D space, $span(\{e_1,e_2\}) = span(\{e_3\})^{\perp}$

Eg2: in 2D, x-axis=$(y_a xis)^{\perp}$. $span(\{e_1\}) = span(\{e_2\})^{\top}$

Non-example: one line on the floor $\neq (vertical\ flag)^{\perp}$ in 3D,x-axis $\neq (y - axis)^{\perp}$ ■

**Example 8.6** (Dimension of Orthogonal Complement)

Suppose S is a subspace of $R^n$. Then $S^{\perp}$ is a subspace and

$$dim(S) + dim(S^{\perp}) = N$$

Furthermore, if $\{u_1, \cdots, u_r\}$ is a basis of S, and $\{u_{r+1}, \cdots, u_n\}$ is a basis of $S^{\perp}$, then $\{u_1, \cdots, u_r, u_{r+1}, \cdots, u_n\}$ is a basis of $R^n$. ■

Proof of theorem 8.6

*Proof.* (1) If $S = \emptyset$, then $S^{\perp} = R^n$, the statement is true.

(2) Assume that $S \neq \emptyset$, then let $\{u_1, \cdots, u_r\}$ be a basis for S, let A=$[u_1, \cdots, u_r]$, the S=Col(A), rank(A)=rank($A^{\top}$)=r and

$$S^{\perp} = Col(A)^{\perp} = Null(A^{\top})$$

By the Rank-Nullity theorem, we have $rank(A^{\top}) + dim(Null(A^{\top})) = n$,

$$dimS + dimS^{\perp} = n$$

Now suppose that the following linear combination is zero, i.e.,

$$\alpha_1 u_1 + \cdots + \alpha_r u_r + \alpha_{r+1} u_{r+1} + \cdots + \alpha_n u_n = 0$$

then

$$\alpha_1 u_1 + \cdots + \alpha_r u_r = -\alpha_{r+1} u_{r+1} - \cdots - \alpha_n u_n$$

The LHS is a vector in S and the RHS is a vector in $S^{\perp}$, since $S^{\perp} = \{0\}$,then

$$\alpha_1 u_1 + \cdots + \alpha_r u_r = 0 = -\alpha_{r+1} u_{r+1} - \cdots - \alpha_n u_n$$

Since $\{u_1, \cdots, u_r\}$ is a basis for S and $\{u_{r+1}, \cdots, u_n\}$ is a basis for $S^\perp$, thus

$$\alpha_1 = \cdots = \alpha_r = \alpha_{r+1} = \cdots = \alpha_n = 0$$

Thus,$\{u_1, \cdots, u_r, u_{r+1}, \cdots, u_n\}$ is a basis for $R^n$. ∎

**Question on Expressing Solution Set.**

The complete solution is

$$x_p + N(A) = x_p + C(M) = x_p + \alpha_1 v_1 + \cdots + \alpha_{n-r} v_{n-r}$$

We use n-r vectors.

Question: Is there a "simpler" way to express C(M)=N(A)?

Answer:

If "simpler" means fewer vectors,the answer is "no". Because **Rank-nullity theorem says:** For rank-r matrix A, need exactly (n-r) vectors to express solution set of $Ax = 0$.

Deeper understanding: null space is orthogonal complement to the row space, so dim is (n-r).

## 8.1.4. Four Fundamental Subspaces

### 8.1.4.1. Left Null space

**Definition 8.9** (left null space)

The left null space of a matrix A is defined as $N(A^\top)$.

*Proof.*

$$N(A^\top) = \{y : Ay^\top = 0\} = \{y : y^\top A = 0\}$$

$$[y_1, y_2, \cdots, y_m] \begin{bmatrix} a_1^\top \\ a_2^\top \\ \vdots \\ a_m^\top \end{bmatrix} = 0 \Longleftrightarrow \sum_{i=1}^{m} y_i a_i^\top = 0$$

linear combination of A's rows=0

∎

**Corollary 8.3** Suppose A has m rows. Then

$$rank(A) + dim(N(A^\top)) = m$$

Row space dim + Left null space dim=m

*Proof.* of coro.8.3

Let $B = A^\top$, according to rank-nullity theorem,

$$rank(B) + dim(N(B)) = m \longleftrightarrow rank(A^\top) + dim(N(A^\top)) = m \longleftrightarrow rank(A) + dim(N(A^\top)) = m$$

■

#### 8.1.4.2. Fundamental theorem

**Theorem 8.4** (Fundamental Theorem of Linear Algebra)

Suppose $A \in R^{mxn}$. Then

$$(1)N(A) = Row(A)^\perp, dim(N(A)) = n - rank(A)$$

$$(2)N(A^\top) = Column(A)^\perp, dim(N(A^\top)) = m - rank(A)$$

.

## 8.2. Lecture 16: Least Squares Approximations

### 8.2.1. Least Square Problem

When linear system $Ax = b$ has no solution, we try to find the best "approximation". For each $x$, the error is defined to be $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{Ax}$. We want to use *least square method* to minimize the error. In other words, our goal is to

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \boldsymbol{e}^2 := \min_{\boldsymbol{x}} \|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|^2 = \sum_{i=1}^{m}(a_i^{\mathrm{T}}\boldsymbol{x}-b_i)^2$$

where $\boldsymbol{A}\in\mathbb{R}^{m\times n}$, $a_i^{\mathrm{T}}$ are row vectors of $\boldsymbol{A}$, and $\boldsymbol{b}\in\mathbb{R}^m$. The minimizer $\boldsymbol{x}$ is called the **linear least squares solution**.

**Definition 8.10** A linear system $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ with a coefficient matrix $\boldsymbol{A}\in R^{m\times n}$ is **over-determined** if $m>n$.

In this section, we only discuss over-determined linear systems.

**Proposition 8.6** $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ has a solution if and only if $\boldsymbol{b}\in C(A)$.

**Definition 8.11** [Least Squares Problem] Given a linear system $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$, the least square problem is

$$\min_{x}\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|$$

Remark: $\|\cdot\|$ is the $L_2$-norm for vector space. i.e. $v=[v_1,\cdots v_n]^{\mathrm{T}}$, then

$$\|v\|=\sqrt{v_1^2+\cdots+v_n^2}$$

.

By definition, $y\in\mathbb{R}$ is a solution of the least square problem if

$$\|\boldsymbol{A}\boldsymbol{y}-\boldsymbol{b}\|\le\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|$$

for any $x\in\mathbb{R}$.

## 8.2.2. Solving Least Square Problem

**Theorem 8.5 — Orthogonal Projection.** Let $V$ be a vector space with subspace $S$, for any $b\in V$ and $p\in S$

(1) $\|b - s\| \geq \|b - p\|$ for any $s \in S$

(2) $\langle b - p, s \rangle = 0$ for any $s \in S$

(1) and (2) are equivalent.

*Proof.* Notice that:

$$\begin{aligned}
\|b - s\|^2 &= \|b - p + p - s\|^2 \\
&= \langle b - p + p - s, b - p + p - s \rangle \\
&= \langle b - p, b - p \rangle + 2\langle b - p, p - s \rangle + \langle p - s, p - s \rangle \\
&= \|b - p\|^2 + \|p - s\|^2 + 2\langle b - p, p - s \rangle
\end{aligned}$$

(1) $\Rightarrow$ (2) : suppose there exists $s' \in S$, such that $\langle b - p, s' \rangle = \epsilon \neq 0$, WLOG, assume $\epsilon > 0$. Take $s'' = -\frac{\epsilon}{\|s'\|^2} s'$, then $\langle b - p, s'' \rangle = -\frac{\epsilon^2}{\|s'\|^2}$. Take $s = p - s''$, then

$$\begin{aligned}
\|b - s\|^2 &= \|b - p\|^2 + \|p - s\|^2 + 2\langle b - p, p - s \rangle \\
&= \|b - p\|^2 + \|s''\|^2 + 2\langle b - p, s'' \rangle \\
&= \|b - p\|^2 + \frac{\epsilon^2}{\|s'\|^2} - 2\frac{\epsilon^2}{\|s'\|^2} \\
&< \|b - p\|^2
\end{aligned}$$

Contradiction! Thus, $\langle b - p, s \rangle = 0$ for any $s \in S$.

(2) $\Rightarrow$ (1) : Notice that $p - s \in S$,

$$\begin{aligned}
\|b - s\|^2 &= \|b - p\|^2 + \|p - s\|^2 + 2\langle b - p, p - s \rangle \\
&= \|b - p\|^2 + \|p - s\|^2 \\
&\geq \|b - p\|^2
\end{aligned}$$

■

**Theorem 8.6** For a least square problem, the following are equivalent:

(1) $\boldsymbol{y}$ minimizes $\|\boldsymbol{Ax} - \boldsymbol{b}\|$

(2) $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ay} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$

*Proof.* $\boldsymbol{y}$ minimizes $\|\boldsymbol{Ax} - \boldsymbol{b}\| \iff \|\boldsymbol{Ay} - \boldsymbol{b}\| \leq \|\boldsymbol{Ax} - \boldsymbol{b}\|$ for any $x \iff \|\boldsymbol{Ay} - \boldsymbol{b}\| \leq \|\boldsymbol{s} - \boldsymbol{b}\|$

for any $\boldsymbol{s} \in C(\boldsymbol{A})$. By the theorem of orthogonal projection, this is equivalent to $\langle \boldsymbol{Ay} - \boldsymbol{b}, \boldsymbol{s} \rangle = 0$ for any $\boldsymbol{s} \in C(\boldsymbol{A}) \iff \boldsymbol{Ay} - \boldsymbol{b} \in C(\boldsymbol{A})^{\perp} \iff \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{Ay} - \boldsymbol{b}) = 0 \iff \boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ay} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$

■

**Theorem 8.7** For a least square problem $\boldsymbol{Ax} = \boldsymbol{b}$, suppose $\boldsymbol{A}$ has linearly independent columns. Then the least square solution is

$$\boldsymbol{y} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$$

*Proof.* Since $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is over-determined, we have $m > n$. So, $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is an $n \times n$ matrix. If $\boldsymbol{A}$ has linearly independent columns, then $rank(\boldsymbol{A}) = n$. We claim that $rank(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = n$. Assume $\boldsymbol{x} \in N(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})$, then $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax} = 0$. This implies $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax} = 0$, $\|\boldsymbol{A}x\| = 0$, $\boldsymbol{Ax} = 0$. Since $\boldsymbol{A}$ has linearly independent columns, we must have $x = 0$. Thus $dim(N(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})) = 0$, $rank(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = n$. Thus, $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is an invertible matrix. The least square solution $\boldsymbol{y}$ should satisfy $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ay} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$, so $\boldsymbol{y} = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ ■

### 8.2.2.1. Proof via Calculus

Firstly, you should know some basic calculus knowledge for matrix:

**The Chian Rule.** Given two vectors $f(x)$,$g(x)$ of appropriate size,

$$\frac{\partial(f^{\mathrm{T}}g)}{\partial x} = \frac{\partial f(x)}{\partial x}g(x) + \frac{\partial g(x)}{\partial x}f(x)$$

**Examples of Matrix Derivative.**

$$\frac{\partial(a^{\mathrm{T}}\boldsymbol{x})}{\partial \boldsymbol{x}} = a \tag{8.1}$$

$$\frac{\partial(a^{\mathrm{T}}\boldsymbol{Ax})}{\partial \boldsymbol{x}} = \frac{\partial((\boldsymbol{A}^{\mathrm{T}}a)^{\mathrm{T}}\boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{A}^{\mathrm{T}}a \tag{8.2}$$

$$\frac{\partial(\boldsymbol{Ax})}{\partial \boldsymbol{x}} = \boldsymbol{A}^{\mathrm{T}} \tag{8.3}$$

$$\frac{\partial(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Ax})}{\partial \boldsymbol{x}} = \boldsymbol{Ax} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{x} \tag{8.4}$$

Thus, in order to minimize $\|\boldsymbol{Ax} - \boldsymbol{b}\|^2 = (\boldsymbol{Ax} - \boldsymbol{b})^{\mathrm{T}}(\boldsymbol{Ax} - \boldsymbol{b})$, it suffices to let its **derivative** with respect to $\boldsymbol{x}$ to be **zero.** Hence we have:

$$\begin{aligned}\frac{\partial(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})}{\partial\boldsymbol{x}} &= \frac{\partial(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})}{\partial\boldsymbol{x}}(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})+\frac{\partial(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})}{\partial\boldsymbol{x}}(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})\\ &= 2\frac{\partial(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})}{\partial\boldsymbol{x}}(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})\\ &= 2(\frac{\partial(\boldsymbol{A}\boldsymbol{x})}{\partial\boldsymbol{x}}-\frac{\partial(\boldsymbol{b})}{\partial\boldsymbol{x}})(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})\\ &= 2\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})=\mathbf{0}.\end{aligned}$$

Or equivalently,

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}.$$

#### 8.2.2.2. Application: Fit a stright line

Given a collection of data $(\boldsymbol{x}_i, y_i)$ for $i=1,\dots,m$, we can use a stright line to fit these points:

$$\begin{cases} y_1 = a_0 + a_1x_{1,1} + a_2x_{1,2} + \cdots + a_nx_{1,n} + \varepsilon_1 \\ y_2 = a_0 + a_1x_{2,1} + a_2x_{2,2} + \cdots + a_nx_{2,n} + \varepsilon_2 \\ \quad\vdots \\ y_m = a_0 + a_1x_{m,1} + a_2x_{m,2} + \cdots + a_nx_{m,n} + \varepsilon_m \end{cases}$$

Our fit line is

$$\hat{y} = a_0 + a_1x_1 + a_2x_2 + \cdots + a_nx_n$$

In *compact matrix form*, we have

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \dots & x_{1,n} \\ 1 & x_{2,1} & x_{2,2} & \dots & x_{2,n} \\ \vdots & \vdots & & & \\ 1 & x_{m,1} & x_{m,2} & \dots & x_{m,n} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix}$$

Or equivalently, we have

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{A}=\begin{bmatrix}1 & x_{1,1} & x_{1,2} & \dots & x_{1,n}\\ 1 & x_{2,1} & x_{2,2} & \dots & x_{2,n}\\ \vdots & \vdots & & & \\ 1 & x_{m,1} & x_{m,2} & \dots & x_{m,n}\end{bmatrix}_{m\times(n+1)}$, $\boldsymbol{x}=\begin{bmatrix}a_0\\ a_1\\ a_2\\ \vdots\\ a_n\end{bmatrix}_{(n+1)\times 1}$, $\boldsymbol{\varepsilon}=\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \vdots\\ \varepsilon_m\end{bmatrix}_{m\times 1}$.

Our goal is to minimize $\|\hat{\boldsymbol{y}}-\boldsymbol{y}\|^2=\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y}\|^2$. Then, it suffices to sovle $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}$.

# Chapter 9

# Week9

## 9.1. Lecture 17: Orthonormal Basis and Gram-Schmidt Process

### 9.1.1. Review: Orthogonal Complement and Orthogonal Subspaces

We have learned Orthogonal complement(8.1.3.4) and orthogonal subspaces(8.1.3.3) before. The Fundamental Theorem of Linear Algebra is the following:

**Theorem 9.1 — Fundamental Theorem of Linear Algebra.** Suppose $A \in \mathbb{R}^{m \times n}$. Then:

1. $N(A^\top) = C(A)^\perp, \quad dim(N(A)) = n - rank(A)$
2. $N(A) = R(A)^\perp, \quad dim(N(A^\top)) = m - rand(A)$

To prove the theorem, note that:

$$N(A^\top) = \{\boldsymbol{x} \in \mathbb{R}^m : A^\top \boldsymbol{x} = 0\}$$
$$C(A) = \{\boldsymbol{y} \in \mathbb{R}^m : \boldsymbol{y} = A \cdot \boldsymbol{z}, \boldsymbol{z} \in \mathbb{R}^n\}$$

In this lecture, we will first introduce orthogonal set and orthonormal basis based on the previous concept. The second important topic is Gram-Schmidt process.

## 9.1.2. Orthonormal Basis

Recall the definition of **Orthogonality** in 8.1.3, the next question is: Can we generalize this definition from $\mathbb{R}^n$ to any linear space?

We first need to define an operation called **inner product** and the corresponding **inner product space**in a linear space.

**Definition 9.1** [Inner Product Space] An inner product space $V$ over $\mathbb{R}$ is a **linear space**, together with an **inner product** $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ satisfying

1. $\langle v, v \rangle \geq 0$ with the equality holds iff $v$ is the zero element in $V$
2. $\langle v, w \rangle = \langle w, v \rangle \quad \forall v, w \in V$
3. $\langle \alpha v + \beta w, z \rangle = \alpha \langle v, z \rangle + \beta \langle w, z \rangle \quad \forall v, w, y \in V$ and $\alpha, \beta \in \mathbb{R}$

**Example 9.1** The common linear space V:

1. For $V = \mathbb{R}^n, \langle v, w \rangle = v^\top w$ (We have verified the properties in lecture 2)
2. For $V = \mathbb{R}^{n \times m}, \langle A, B \rangle = \sum_{i=1}^{m} \sum j = 1^n a_{ij} b_{ij}$

Now we can extend the property orthogonality for linear space:

**Definition 9.2** [Orthogonality for Linear Space] Let $V$ be an inner product space over $\mathbb{R}$. Two elements $u, v \in V$ are orthogonal if $\langle u, v \rangle = 0$, denoted by $u \perp v$.

With the defined inner product, we can also define norm in an inner product space.

**Theorem 9.2 — Norm in an Inner Product Space.** Let $V$ be an inner product space over $\mathbb{R}$. For any $v \in V, ||v|| := \sqrt{\langle v, v \rangle}$ defined a norm that satisfies:

1. $||v|| \leq 0$ with an equality holds iff $v$ is the zero element in $V$
2. $||\alpha v|| = \alpha ||v||, \quad \forall \alpha \in \mathbb{R}, v \in V$
3. $||v + w|| \leq ||v|| + ||w|| \quad \forall v, w \in V$

Then we define **Orthogonal Set** and **Orthonormal Set of Vectors**:

**Definition 9.3** [Orthogonal Set ]

Let $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$ be a set of nonzero elements in an inner product space $V$ over $\mathbb{R}$. If $\langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle = 0 \quad \forall i \neq j$ . This set is called an **orthogonal set**.

**Definition 9.4** [Orthonormal Set of Vectors] An orthonormal set $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$ is an orthogonal set of unit elements, i.e, $||v_i|| \forall i = 1, \cdots, n$

**Example 9.2** Consider $\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3$ in $\mathbb{R}^3$,
$\boldsymbol{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \boldsymbol{u}_2 = \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix}, \boldsymbol{u}_3 = \begin{bmatrix} 4 \\ -5 \\ 1 \end{bmatrix}$ then $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$ is an orthogonal set, the orthonormal set is

$$\left\{ \boldsymbol{v}_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \boldsymbol{v}_2 = \frac{1}{\sqrt{14}} \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix}, \boldsymbol{v}_3 = \frac{1}{\sqrt{42}} \begin{bmatrix} 4 \\ -5 \\ 1 \end{bmatrix} \right\}$$

Note that an orthonormal basis unique for a given inner product space.

Another question is: What is $dim(span(S))$ if $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$ is an orthogonal set?

The answer is not unique, both $dim(S) \leq n$ or $dim(S) = n$ are possible. Actually, there is no information about the dimension given a specific orthogonal set.

Then, what is Special about an Orthogonal Set?

**Proposition 9.1** Let $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$ be an orthogonal set. Then $\boldsymbol{v}_1, \cdots \boldsymbol{v}_n$ are linearly independent.

In words: Vectors in an orthogonal set are linearly independent

*Proof.* For any $i = 1, \cdots, n$, by definition of orthogonal set,

$$\begin{aligned}
0 &= \langle \boldsymbol{v}_i, c_1\boldsymbol{v}_1 + \cdots c_n\boldsymbol{v}_n \rangle \\
&= \langle \boldsymbol{v}_i, \sum_{j=1}^{n} c_j\boldsymbol{v}_j \rangle \\
&= \sum_{j=1}^{n} \langle \boldsymbol{v}_i, c_j\boldsymbol{v}_j \rangle \\
&= \sum_{j=1}^{n} c_j \langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle \\
&= c_i ||\boldsymbol{v}_i||
\end{aligned}$$

Moreover, $\boldsymbol{v}_i \neq 0$ as an orthogonal set has no zero element. We must have $c_i = 0 \quad i = 1, \cdots, n$
Hence, $\boldsymbol{v}_1, \cdots \boldsymbol{v}_n$ are linearly independent. ■

We can further define the orhonormal basis:

**Definition 9.5** A set of vectors $S = \boldsymbol{v}_1, \cdots \boldsymbol{v}_n$ is called an orthonormal basis of a linear space $V$ if:

1. $S$ is an orthonormal set
2. $span(S) = V$

**Example 9.3** [orthonormal basis]

1. Standard basis : $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$
2. $\left\{ \boldsymbol{v}_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \boldsymbol{v}_2 = \frac{1}{\sqrt{14}} \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix}, \boldsymbol{v}_3 = \frac{1}{\sqrt{42}} \begin{bmatrix} 4 \\ -5 \\ 1 \end{bmatrix} \right\}$

Recall that any element $v \in V$ can be uniquely represented as a linear combination of $u_1, \cdots, u_n$. What if a basis is orthonormal?

**Proposition 9.2 — Representation via Orthonormal Basis.** Let $\mathscr{U} = \{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n\}$ be an orthonormal basis of an inner product space $V$. Any element $v \in V$ can be uniquely represented as a linear combination

$$v = \sum_{i=1}^{n} \langle v, u_i \rangle u_i$$

*Proof.* We've already know that $v = \sum_{i=1}^{n} c_i u_i$ since $\mathscr{U}$ is a basis. We observe that:

$$\langle u_j, u_i \rangle = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

Then, consider $u_i$, $i = 1, \cdots, n$

$$\begin{aligned} \langle v, u_i \rangle &= \langle \sum_{j=1}^{n} c_j u_j, u_i \rangle \\ &= \sum_{j=1}^{n} c_j \langle u_j, u_i \rangle \\ &= c_i ||u_i|| \quad = c_i \end{aligned}$$

Plugging back to the original equation, we get $v = \sum_{i=1}^{n} \langle v, u_i \rangle u_i$ ■

**Example 9.4** $\left\{ \boldsymbol{v}_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \boldsymbol{v}_2 = \frac{1}{\sqrt{14}} \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix}, \boldsymbol{v}_3 = \frac{1}{\sqrt{42}} \begin{bmatrix} 4 \\ -5 \\ 1 \end{bmatrix} \right\}$ is an orthonormal basis of $\mathbb{R}^3$

For any $\boldsymbol{x} = [x, y, z]^\top \in \mathbb{R}^3$, one has

$$\begin{aligned} \boldsymbol{x} &= \langle \boldsymbol{x}, \boldsymbol{v}_1 \rangle + \langle \boldsymbol{x} \boldsymbol{v}_2 \rangle \boldsymbol{v}_2 + \langle \boldsymbol{x}, \boldsymbol{v}_3 \rangle \boldsymbol{v}_3 \\ &= \frac{x+y+z}{\sqrt{3}} \boldsymbol{v}_1 + \frac{2x+y-3z}{\sqrt{14}} \boldsymbol{v}_2 + \frac{4x-5y+z}{\sqrt{42}} \boldsymbol{v}_3 \end{aligned}$$

Orthonormal Basis is an important concepts, which is useful in:

- Representing any element in an inner product space by computing the coordinates
- eigen-theory

Then you may ask: How can we obtain an orthonormal basis?

### 9.1.3. Gram-Schmidt Process

Suppose $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n\}$ is a set of linearly independent elements. Based on S, we can find an orthonormal set $U$ such that

$$span(U) = span(S)$$

by a method called Gram-Schmidt Process.

Let's begin with projection.

**Proposition 9.3** For any subspace $W \in V$ and any $v \in V$, there is a **unique** p $\in W$ such that $(v - p) \perp W$

**Definition 9.6** [Projection] The unique vector **p** in Prop.9.3 is called the projection of **v** onto $W$, denoted by **p**$= proj_W(v)$

**Proposition 9.4 — Projection Representation.** Let $W$ be a subspace of an inner product space$V$ and denote by $S = \{\boldsymbol{w}_1, \cdots, \boldsymbol{w}_n\}$ an orthonormal basis of $W$. For any $v \in V$,if $\{v - p\} \perp W$ for some $p \in W$, then $p$ is uniquely determined by:

$$p = \sum_{i=1}^{n} \langle w_i, v \rangle w_i$$

**Note:** $\{v - p\} \perp W$ means that $p$ is the projection of $v$ onto $W$

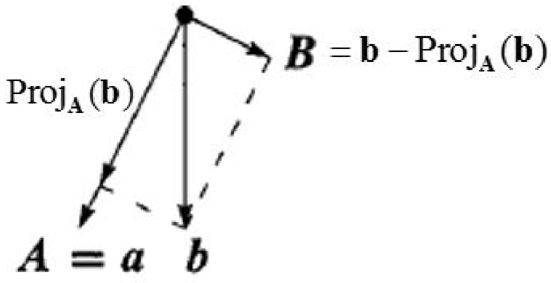
*Proof.* By orthonormal representation (Prop 9.4), $p \in W$,

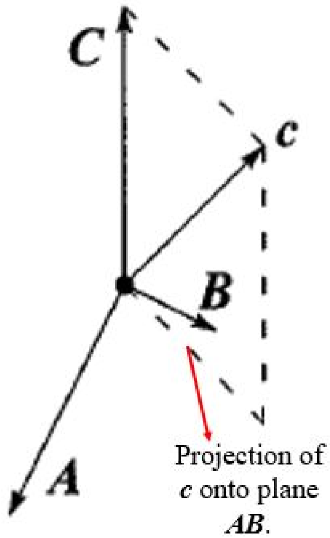
$$\begin{aligned}
p &= \sum_{i=1}^{n} \langle w_i, p \rangle w_i \\
&= \sum_{i=1}^{n} \langle w_i, p - v + v \rangle w_i \\
&= \sum_{i=1}^{n} \langle w_i, p - v \rangle w_i + \sum_{i=1}^{n} \langle w_i, v \rangle w_i \\
&= \sum_{i=1}^{n} \langle w_i, v \rangle w_i
\end{aligned}$$

■

We've proved the existence and uniqueness of $p$, but the original question is: how to find it?

Idea:




**[Gram-Schmidt Process]**

$$\text{Input}S = \{\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n\}$$

$$\text{For}i = 1, \cdots, n$$

$$\text{if } i = 1:$$

$$p_1 = 0$$

$$\text{Else:}$$

$$p_{i-1} = \sum_{j=1}^{i-1} \langle u_i, v_j \rangle v_j = proj_{span(u_1, \cdots, u_{i-1})}(u_i)$$

$$\text{Set } v_i = \frac{u_i - p_{i-1}}{||u_i - p_{i-1}||} \in span(u_1, \cdots, u_{i-1})^{\perp}$$

$$\text{Return } U = \{v_1, \cdots, v_n\}$$

**Proposition 9.5 — Gram-Schmidt Process.** The set $U = \{v_1, \cdots, v_n\}$ returned by the Gram-Schmidt process is an orthonormal basis of $span(S)$

*Proof.* Orthogonality: by definition $\langle v_i, v_j \rangle = 0$

Norm: As we normalize $v_1, \cdots, v_n \Rightarrow ||v_i|| = 1$

Linear independence: verified in Prop9.1

Finally, $dim(span(S)) = n$ implies the theorem

■

**Example 9.5** $S=\{u_1,u_2,u_3\}=\left\{\begin{bmatrix}1\\1\\1\\1\end{bmatrix},\begin{bmatrix}-1\\4\\4\\-1\end{bmatrix},\begin{bmatrix}4\\-2\\2\\0\end{bmatrix}\right\}\subseteq\mathbb{R}^4$ In $\mathbb{R}^n$, the standard inner product is the scalar product $\langle \boldsymbol{x},\boldsymbol{y}\rangle=\boldsymbol{y}^\top\boldsymbol{x}$. Now find the orthonormal basis for $span(u_1,u_2,u_3)$

- Step 1: $v_1=\frac{u_1}{||u_1||}=\frac{1}{2}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}=\begin{bmatrix}\frac{1}{2}\\\frac{1}{2}\\\frac{1}{2}\\\frac{1}{2}\end{bmatrix}$
- Step 2: calculate $u_2'=u_2-\langle u_2,v_1\rangle v_1=\begin{bmatrix}-1\\4\\4\\-1\end{bmatrix}-\begin{bmatrix}-1&4&4&-1\end{bmatrix}\begin{bmatrix}\frac{1}{2}\\\frac{1}{2}\\\frac{1}{2}\\\frac{1}{2}\end{bmatrix}\cdot\begin{bmatrix}\frac{1}{2}\\\frac{1}{2}\\\frac{1}{2}\\\frac{1}{2}\end{bmatrix}=\begin{bmatrix}-\frac{5}{2}\\\frac{5}{2}\\\frac{5}{2}\\--\frac{5}{2}\end{bmatrix}$

  Then, $v_2=\frac{u_2'}{||u_2||}=\begin{bmatrix}-\frac{2}{5}\\\frac{2}{5}\\\frac{2}{5}\\-\frac{2}{5}\end{bmatrix}=\begin{bmatrix}-\frac{1}{2}\\\frac{1}{2}\\\frac{1}{2}\\-\frac{1}{2}\end{bmatrix}$ Observe that $\{v_1,v_2\}$ form an orthonormal basis.
- Step 3: calculate

$$u_3'=u_3-\langle u_3,v_1\rangle v_1-\langle u_3v_2\rangle v_2$$

$$=\begin{bmatrix}4\\-2\\2\\0\end{bmatrix}-\begin{bmatrix}4&-2&2&0\end{bmatrix}\begin{bmatrix}\frac{1}{2}\\\frac{1}{2}\\\frac{1}{2}\\\frac{1}{2}\end{bmatrix}\cdot\begin{bmatrix}\frac{1}{2}\\\frac{1}{2}\\\frac{1}{2}\\\frac{1}{2}\end{bmatrix}-\begin{bmatrix}4&-2&2&0\end{bmatrix}\begin{bmatrix}-\frac{1}{2}\\\frac{1}{2}\\\frac{1}{2}\\-\frac{1}{2}\end{bmatrix}\cdot\begin{bmatrix}-\frac{1}{2}\\\frac{1}{2}\\\frac{1}{2}\\-\frac{1}{2}\end{bmatrix}=\begin{bmatrix}2\\-2\\2\\-2\end{bmatrix}$$

  Then, $v_3=\frac{u_3'}{||u_3||}=\frac{1}{4}\begin{bmatrix}2\\-2\\2\\-2\end{bmatrix}=\begin{bmatrix}\frac{1}{2}\\-\frac{1}{2}\\\frac{1}{2}\\-\frac{1}{2}\end{bmatrix}$ Observe that $\{v_1,v_2,v_3\}$ form an orthonormal basis.

## 9.1.4. Application of Orthonormal Basis

### 9.1.4.1. Compute Projection

- First, find an orthonormal basis $\{v_1, \cdots, v_k\}$ by Gram-Schmidt process.
- Second,use formula $p = \sum_{i=1}^{k} \langle u, v_i \rangle v_i$

### 9.1.4.2. Solve Least Squares

**Proposition 9.6 — special case S = C(A).** Consider a least squares problem. The following statements are equivalent:

1. $\boldsymbol{y}$ minimizes $||A\boldsymbol{x} - \boldsymbol{b}||$
2. $\boldsymbol{b} - A\boldsymbol{y} \perp S$, i.e. $A\boldsymbol{y} = Proj_{C(A)}(\boldsymbol{b})$

## 9.1.5. Orthogonal Matrix

**Definition 9.7** [Orthogonal Matrix] An orthogonal matrix $Q \in \mathbb{R}^{n \times n}$ is a real square matrix whose columns form an orthonormal set in $\mathbb{R}^n$ **Note:** The columns of $Q$ form an orthonormal basis of $\mathbb{R}^n$

**Proposition 9.7 — Orthogonal Matrix.** A matrix $Q \in \mathbb{R}^{n \times n}$ is an orthogonal matrix if and only if :

$$Q^{-1} = Q^{\top}$$

**Example 9.6** For any $\theta$, $Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$

$$Q^{-1} = Q^{\top} \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

Let's prove Prop 9.7

*Proof.* Recall the definition of inverse: $AB = BA = I$ (you can verify this by computing the rank), it is sufficient to show that $Q^\top Q = I$. Let $Q = \begin{bmatrix} | & & | \\ q_1 & \cdots & q_n \\ | & & | \end{bmatrix}, Q^\top = \begin{bmatrix} q_1^\top \\ \vdots \\ q_n^\top \end{bmatrix}$, and $\{q_1, \cdots, q_n\}$ is an orthonormal basis, we have

$$Q^\top Q = (q_i^\top q_j)_{n \times n} = I_n$$

■

We now introduce the great properties of orthogonal matrix:

**Proposition 9.8 — Properties of Orthogonal Matrix.** If $Q \in \mathbb{R}^{n \times n}$ is an orthogonal matrix, then:

- the column vectors of $Q$ form an orthonormal basis for $\mathbb{R}^n$
- $Q^{-1} = Q^\top$
- $Q^\top Q = I_n$
- $\langle Q\boldsymbol{x}, Q\boldsymbol{y} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle$
- $||Q\boldsymbol{x}|| = ||\boldsymbol{x}||$

## 9.1.6. Summary

We have studied Orthonormal basis and Gram-Schmidt Process.

- Orthonormal basis
  - Basis consisQng of unit vectors that are orthogonal to each other.
  - Prop 9.1: Representation of vector by ortho-basis of whole space.
  - Prop 9.2: Representation of projection by ortho-space of a subspace.
- Gram-Schmidt process
  - Goal: Construct orthonormal basis of a subspace, from a basis of the subspace.
  - Main trick: find a vector orthogonal to a subspace, sequentially.
  - Theoretical result: Gram-Schmidt process indeed returns an orthonormal basis.
- Orthogonal matrix
  - Square matrix whose columns form a orthonormal basis of $\mathbb{R}^n$

– Properties: $QQ^\top = I, ||Q\boldsymbol{x}|| = ||\boldsymbol{x}|| \cdots$.

## 9.2. Lecture 18: Determinants

### 9.2.1. Review

We have finished the first part of the course, which mainly focuses on how to solve linear system in Lecture 3 - 15, and how to solve least squares problem in Lecture 16 - 17. In the next part, we will learn two relatively independent concepts: determinant and linear transformation. They are not directly related to solving systems or problems, however, they are fundamental and useful tools.

In today's lecture, we will focus on the concept, properties and applications of determinant.

### 9.2.2. Motivation of determinants

Let's first recall the equivalent conditions for invertibility:

**Theorem 9.3** (Equivalent Conditions for Invertibility)
Let $A \in \mathbb{R}^{nxn}$.

(1) A is invertible.

(2) The linear system $Ax = 0$ has a unique solution $x = 0$.

(3) A is a product of elementary matrices.

(4) A has n pivots; or equivalently: rank(A)=n.

(5) The columns of A span $R^n$.

(6) The columns of A are linearly independent;

(7) The columns of A form a basis;

(8) dim(C(A)) = n.

(9) dim(N(A)) = 0, or N(A)={0}.

It gives us a numerical way to verify whether a matrix is invertible. Then we look at a homework problem: when is a $2 \times 2$ matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ invertible?

We can easily get the answer by the theorem above: If and only if $ad - bc \neq 0$. The advantage over the existing 9 conditions in Theorem 9.1 is, they all use closed-form expression to judge invertibility.

**R** We had an expression of an inverse, but:

- No expression to "numerically determine invertibility".
- Rely on an algorithm.

Just like matrix inverse, can we find a similar way to determine invertibility? Let $A \in \mathbb{R}^{n\times n}$ be a real square matrix. So here we have two questions to find an invertibility condition:

- **Question 18.1: [extension of invertibility condition]**

  Can we extend $(ad - bc)$ to $n \times n$ matrix, i.e. find a **determinant condition** $det(\cdot) : \mathbb{R}^{n\times n} \to \mathbb{R}$ (actually a function) such that $det(A) \neq 0$ if and only if $A$ is invertible?

- **Question 18.2: What is $ad - bc$?**

First let's take a look at Q2. This fundamental understanding can help answer Q1. Though it looks hard, we can use geometry to understand it.

- **Case 1:** If $b = c = 0$, then it's the area of a rectangle form by $a$ and $d$.
- **Case 2:** More generalized, if $b \neq 0$ or $c \neq 0$, then it's the area of a parallelogram formed by the column vectors. The coordinates of the four points are $(0, 0)$, $(a, c)$, $(a + b, c + d)$ and $(b, d)$.

According to the cases above, what does $ad - bc$ mean? It means the area of the parallelogram is 0; equivalently, the two column vectors are parallel and the matrix is not invertible.

## 9.2.3. Definitions and Properties

Then we look at Q1 again. Let's extend the geometrical meaning of "area" to higher dimensions. For example, in three dimensional space, we can infer that it's the **"volume"** of a parallelepiped also formed by the column vectors. For higher dimension, it's more likely to be a higher-dimensional "volume". But how to compute such "volume"? If you find it hard to think about the general case, then it's helpful to start with the simplest special case.

R

**We denote $|A| = det(A)$. Noted that it's not an absolute value!**

Let's start with a $2 \times 2$ matrix. For such matrix, we have the following properties:

(1) Multiply row 1 by any number $t$: $\begin{vmatrix} ta & tb \\ c & d \end{vmatrix} = t\begin{vmatrix} a & b \\ c & d \end{vmatrix}$.

(2) Add row 1 of the first matrix to row 1 of the second matrix: $\begin{vmatrix} a_1 + a_2 & b_1 + b_2 \\ c & d \end{vmatrix} =$ $\begin{vmatrix} a_1 & b_1 \\ c & d \end{vmatrix} + \begin{vmatrix} a_2 & b_2 \\ c & d \end{vmatrix}$.

Those properties can be easily verified by drawing the graphs of each expression separately and calculate the areas. It seems that determinants have similar operations as the linear product. Thus the determinant seems to be a "linear function" of each row separately.

Then we apply it to $n = 3$. For a $3 \times 3$ matrix $A$, we have the following steps to calculate determinant.

- **Step 1:** Reduce $A$ to simpler matrices $B_i$.

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{bmatrix} + \begin{bmatrix} 0 & a_{12} & 0 \\ a_{21} & 0 & a_{23} \\ a_{31} & 0 & a_{33} \end{bmatrix} + \begin{bmatrix} 0 & 0 & a_{13} \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & 0 \end{bmatrix}.$$

  Similar to the addition above, we desire a property that can relate $det(A)$ and $det(B_i)$ together: $|A| = |B_1| + |B_2| + |B_3|$. Such property is called Original-Unit Property.

- **Step 2: (Building block)** compute the "volume" of simpler matrices. Similar to the first property above, we have $\begin{vmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{vmatrix} = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$. However, for $\begin{vmatrix} 0 & a_{12} & 0 \\ a_{21} & 0 & a_{23} \\ a_{31} & 0 & a_{33} \end{vmatrix}$, the answer is not trivial due to a zero column in the middle that can't be eliminated after picking up $a_{12}$. Therefore, we need a second desired property for columns swapping. By checking a $2 \times 2$ matrix, we let $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = -\begin{vmatrix} b & a \\ d & c \end{vmatrix}$, because $ad - bc = -(bc - ad)$. Otherwise, it is contradicted to the definition of $n = 2$. Similarly, we apply it to $n = 3$:

$$\begin{vmatrix} 0 & a_{12} & 0 \\ a_{21} & 0 & a_{23} \\ a_{31} & 0 & a_{33} \end{vmatrix} = -\begin{vmatrix} a_{12} & 0 & 0 \\ 0 & a_{21} & a_{23} \\ 0 & a_{31} & a_{33} \end{vmatrix} = -a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix}.$$

- **Step 3:** Finally, we wrap things up and can get the following expression:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} 0 & a_{12} & 0 \\ a_{21} & 0 & a_{23} \\ a_{31} & 0 & a_{33} \end{vmatrix} + \begin{vmatrix} 0 & 0 & a_{13} \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & 0 \end{vmatrix}$$

$$= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}.$$

After considering it in a geometric way, we now give the formal definition of determinant.

**Definition 9.8** [Determinant]

- For a scalar $\alpha \in \mathbb{R}$, define $det(\alpha) = \alpha$.
- For any $A \in \mathbb{R}^{n \times n}$ with $n \geq 2$, define

$$det(A) = \sum_{j=1}^{n} (-1)^{1+j} det(M_{1j}) a_{1j}.$$

Where $M_{ij} \in \mathbb{R}^{(n-1)\times(n-1)}$ is a matrix formed by deleting the $i$-th row and $j$-th column of $A$.

**Example 9.7** (1) $|A| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$.

(2) $|A| = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & w \\ g & h \end{vmatrix} = aei + bfg + cdh - ceg - bdi - afh$.

Actually, we can expand the matrix along any row to get the determinant, and the values are all the same.

**Proposition 9.9 — Laplace Expansion of Determinants.** For any $A \in \mathbb{R}^{n\times n}$ with $n \geq 2$, for every $i = 1, \cdots, n$, we have

$$det(A) = \sum_{j=1}^{n} (-1)^{i+j} det(M_{ij}) a_{ij}.$$

The column is the same. Owing to this proposition, we can choose any row or column that we like. More specifically, we can choose those which have less amount of calculation to calculate the determinant.

**Example 9.8** (1) Evaluate $\begin{vmatrix} 0 & 2 & 3 & 0 \\ 0 & 4 & 5 & 0 \\ 0 & 1 & 0 & 3 \\ 2 & 0 & 1 & 3 \end{vmatrix}$.

We can expand it along the last row, then we get $(-1)^5 \cdot 2 \begin{vmatrix} 2 & 3 & 0 \\ 4 & 5 & 0 \\ 1 & 0 & 3 \end{vmatrix}$ because the rest of the components contain a zero column, which means the determinant is zero. Then it can be further expand along the last row to $-2 \cdot (-1)^6 \cdot 3 \begin{vmatrix} 2 & 3 \\ 4 & 5 \end{vmatrix} = 12$.

## 9.2.4. An Alternative Definition

Similarly, we start with the desired properties we want in order to form the determinant.

- **Desired Property 1**

  (Noted that this property is what we want and they are different from a standard property.)

  If $A \in \mathbb{R}^{n\times n}$ is a triangular matrix with diagonal entries $a_{11}, a_{22}, \cdots, a_{nn}$, then $det(A) = a_{11} \cdots a_{nn}$. If it's in the first definition, we can form it by Laplace expansion, but now it is served as a property.

  If a triangular matrix is inevitable, then the diagonal entries are all non-zeros, which means the determinant is a non-zero value as well.

Moreover, if given two matrices $A, B \in \mathbb{R}^{n\times n}$, $AB$ is invertible if and only if $A$ and $B$ are both inevitable. Thus, we also need the following property:

- **Desired Property 2**

  Given two matrices $A, B \in \mathbb{R}^{n\times n}$,

$$det(AB) = det(A)det(B).$$

According to the desired properties above, we now give the alternative definition of determinant.

**Definition 9.9** [Determinant]

A determinant of a matrix $A \in \mathbb{R}^{n\times n}$ is a map $det(\cdot) : \mathbb{R}^{n\times n} \to \mathbb{R}$ with

(1) if $A$ is a triangular matrix with diagonal entries $a_{11}, \cdots, a_{nn}$, then

$$det(A) = a_{11} \cdots_{nn}.$$

(2) Given two matrices $A, B \in \mathbb{R}^{n\times n}$, we have

$$det(AB) = det(A)det(B).$$

We can show that those two definitions of determinant are equivalent. Here we omit the proof.

**Theorem 9.4 — Properties of determinant.** Determinant has the following properties:

1. Swapping two columns (rows) changes the sign of $det(A)$.
2. The determinant is a linear function of each row and each column separately.
3. For any $A \in \mathbb{R}^{n\times n}$, $det(A^{\mathrm{T}}) = det(A)$.
4. For any invertible $A \in \mathbb{R}^{n\times n}$, $det(A^{-1}) = \frac{1}{det(A)}$.

Here we give the proof of 3 and 4. To prove property 3, we use the LU decomposition in order to get triangular matrices. By $PA = LU$, we have $A = P^{-1}LU$. Based on definition 2, we get the equations below:

$$\begin{aligned} det(A) &= det(P^{-1}LU) \\ &= det(P^{-1})det(L)det(U). \end{aligned} \tag{9.1}$$

$$
\begin{aligned}
det(A^T) &= det((P^{-1}LU)^T) \\
&= det(U^T L^T P) \\
&= det(P)det(U^T)det(L^T) \\
&= det(P)det(L)det(U).
\end{aligned} \tag{9.2}
$$

Because $P^{-1} = P^T$, we only need to prove $det(P) = det(P^T)$, which will be shown later.

For property 4, we have:

$$
\begin{aligned}
1 &= det(I) \\
&= det(A^{-1} \cdot A) = det(A^{-1})det(A).
\end{aligned} \tag{9.3}
$$

Thus we have:

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

**Example 9.9** 1. $\begin{vmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 2 & 1 & 2 & 2 \\ 1 & 2 & 1 & 2 & 3 \\ 1 & 1 & 1 & 3 & 2 \\ 1 & 1 & 1 & 1 & 4 \end{vmatrix} \begin{pmatrix} R_2 \to -R_1 + R_2 \\ R_3 \to -R_1 + R_3 \\ R_4 \to -R_1 + R_4 \\ R_5 \to -R_1 + R_5 \end{pmatrix} = \begin{vmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 2 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 3 \end{vmatrix}$ (expand along 3rd column) $= (-1)^{1+3} \begin{vmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 3 \end{vmatrix} = 6.$

Why can we do type 3 row operation here? Because by definition 2, we have $det(EA) = det(E)det(A)$, where $det(E)$ is 1 here. Thus we claim that type 3 elementary row operations do not change the determinant.

### 9.2.5. Verify $det(A) = 0$ iff $A$ is invertible

Here we still use the LU decomposition. FOr any matrix $A \in \mathbb{R}^{n \times n}$, $PA = LU$. By the definition of determinant, we have $det(P)det(A) = det(PA) = det(LU) = det(L)det(U)$. Recall that when

$A$ is invertible, all the diagonal entries of $U$ are non-zeroes.

- **Lemma 1** For LU decomposition, assume $det(P) \neq 0$ and $det(L) \neq 0$. Then $det(A) \neq 0$ if and only if $A$ is invertible.

Since $U$ is triangular, $det(U) = u_{11}u_{22} \cdots u_{nn} \neq 0$. $det(P)det(A) = det(L)det(U)$ implies the result. We can then derive the following theorem.

> **Theorem 9.5** For any $A \in \mathbb{R}^{n \times n}$, $det(A) \neq 0$ if and only if $A$ is invertible.

To prove the theorem, we only need to check $det(P) \neq 0$ and $det(L) \neq 0$ by using the lemma. Because $P$ is a permutation matrix which is the product of type 1 elementary matrices, and for $L$, it is the product of several type 2 and 3 elementary matrices. Thus we only need to prove those three elementary matrices are nonzero.

Type 2: Multiply a Row by a Nonzero Scalar ($R_i \to \alpha R_i (\alpha \neq 0)$).

$$
\begin{array}{cc}
E_{\alpha R_i} = \left[\begin{array}{ccc|c|ccc} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ \hline & & & \alpha & & & \\ \hline & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{array}\right] & i\text{th row.} \\
i\text{th column} &
\end{array}
$$

$$det(E_{\alpha R_i}) = 1 \times 1 \times \cdots \times \alpha \times \cdots \times 1 = \alpha \neq 0.$$

Type 3: Add a Scaled Row to Another ($R_j \to \beta R_i + R_j$).

$$
E_{\alpha R_i} = \begin{array}{cc}
\left[\begin{array}{ccc|c|ccc|c|ccc}
1 & & & & & & & & & & \\
 & \ddots & & & & & & & & & \\
 & & 1 & & & & & & & & \\
\hline
 & & & 1 & & & & & & & \\
\hline
 & & & & 1 & & & & & & \\
 & & & & & \ddots & & & & & \\
 & & & & & & 1 & & & & \\
\hline
 & & & \beta & & & & 1 & & & \\
\hline
 & & & & & & & & 1 & & \\
 & & & & & & & & & \ddots & \\
 & & & & & & & & & & 1
\end{array}\right] & i,j\text{th row.} \\
i,j\text{th column} &
\end{array}
$$

$$det(E_{\beta R_i + R_j}) = 1 \times 1 \times \cdots \times 1 = 1.$$

Type 1: Swap Two Rows ($R_i \leftrightarrow R_j$). It's not a triangular matrix, but we can represent it using the other two operations.

- **Fact:** $E_{R_iR_j} = E_{R_i+R_j}E_{-R_i+R_j}E_{R_i+R_j}E_{-R_i}$.

Because $det(E_{R_i+R_j}) = 1$, $det(E_{-R_i+R_j}) = 1$ and $det(E_{-R_i}) = -1$, therefore $det(E_{R_iR_j}) = -1$.

Look back to theorem 9.5, we can now conclude that $det(P) \neq 0$ and $det(L) \neq 0$ due to properties above. So we can expand the equivalent conditions mentioned in lecture 15.

**Theorem 9.6** (Equivalent Conditions for Invertibility)

Let $A \in \mathbb{R}^{nxn}$.

(1) A is invertible.

(2) The linear system $Ax = 0$ has a unique solution $x = 0$.

(3) A is a product of elementary matrices.

(4) A has n pivots; or equivalently: rank(A)=n.

(5) The columns of A span $R^n$.

(6) The columns of A are linearly independent;

(7) The columns of A form a basis;

(8) dim(C(A)) = n.

(9) dim(N(A)) = 0, or N(A)={0}

(10) **det(A) $\neq$ 0.**

### 9.2.6. Adjoint Matrix

**Definition 9.10** [Adjoint Matrix]

Given $A \in \mathbb{R}^{n\times n}$, let $C_{ij} := (-1)^{i+j} det(M_{ij})$ and define

$$adj(A) = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{bmatrix}^{\mathrm{T}} = \begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix}.$$

**Proposition 9.10** Let $A \in \mathbb{R}^{n\times n}$ be an invertible matrix. We have

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

To prove this, We add A to both sides, so it's equivalent to prove $A \cdot adj(A) = I \cdot det(A)$.

$$\begin{aligned} [A \cdot adj(A)]_{ij} &= \sum_{j=1}^{n} a_{ik} C_{jk} \\ &= \begin{cases} det(A) & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{9.4}$$

Thus those two are equivalent.

### 9.2.7. Summary

In this lecture, we learned determinants.

- Motivation:
  - $ad - bc \neq 0$ is a simple criterion for invertibility.
  - To extend it, we notice it indicates "area" of parallelogram.

- Motivating question: how to compute/define "volume" of polytope?

- Computing "volume":
  - Overall idea: "Simplification" framework.
  - Properties: linear over rows/columns; swapping columns changes sign.
  - Definition: Laplacian expansion over 1st row.

- Properties:
  - $det(AB) = det(A)det(B)$.
  - $det(A) \neq 0$ iff $A$ is invertible.

# Chapter 10

# Week10

## 10.1. Lecture 19: linear transformation 1

### 10.1.1. Linear transformation

We start with a matrix $\boldsymbol{A}$. When multiplying $\boldsymbol{A}$ with a vector $\boldsymbol{v}$, it essentially transforms $\boldsymbol{v}$ to another vector $\boldsymbol{A}\boldsymbol{v}$. Matrix multiplication $L(\boldsymbol{v}) = \boldsymbol{A}\boldsymbol{v}$ gives a **linear transformation**:

**Definition 10.1** [linear transformation] A transformation $L$ assigns an output $T(\boldsymbol{v})$ to each input vector $\boldsymbol{v}$ in $\boldsymbol{V}$. L: V$\longrightarrow W$

$The transformation$L$(\cdot)$ is said to be a **linear transformation** if it satisfies

$$L(\alpha\boldsymbol{v}_1 + \beta\boldsymbol{v}_2) = \alpha L(\boldsymbol{v}_1) + \beta L(\boldsymbol{v}_2)$$

for all vectors $v_1, v_2 \in V$ and scalars $\alpha, \beta$.

R V is called the domain of the linear transformation.
W is called the codomain of the linear transformation.

**Key Observation:** If the input is $\boldsymbol{v} = \boldsymbol{0}$, the output must be $L(\boldsymbol{v}) = \boldsymbol{0}$.

**Example 10.1**
1. L: $R^2 \longrightarrow R^2$. Stretching or shrinking: $L((x,y)^\top) = (\alpha x, \alpha y)^\top (\alpha > 0)$
2. L: $R^2 \longrightarrow R^2$.Rotation:

$$L((x,y)^\top) = (x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta)^\top.$$

(Rotation in anticlockwise by angle $\theta$)

3. Let L: $R^2 \longrightarrow R$ be a mapping defined as

$$L((x,y)^\top) = x - y$$

then L is a linear transformation.

4. Let L: $R^2 \longrightarrow R$ be a mapping defined as

$$L((x,y)^\top) = \sqrt{x^2 + y^2}$$

then L is not a linear transformation since

$$L(-(x,y)^\top) = L((-x,-y)^\top) = \sqrt{(-x)^2 + (-y)^2} = \sqrt{x^2 + y^2} \neq -L((x,y)^\top)$$

## Linear Transformations Properties.

**Theorem 10.1** (property of linear transformation) Let L be a linear transformation from V to W, then $\forall \alpha_i \in R, \forall u_i \in V$

1. $L(0_v) = 0_w$ ($0_v$ is the zero vector in V and $0_w$ is the zero vector in W)
2. $L(\alpha_1 u_1 + \cdots + \alpha_n u_n) = \alpha_1 L(u_1) + \cdots + \alpha_n L(u_n)$
3. L(-u)=-L(u)

*Proof.* 1. $L(\alpha u) = \alpha L(u)$,let $\alpha = 0$ , then $L(0_v) = 0_w$

2. It can be proved by mathematical induction. n=1 is valid, suppose it is valid for n=k, then

$$\begin{aligned} &L((\alpha_1 u_1 + \cdots + \alpha_k u_k) + \alpha_{k+1} u_{k+1}) \\ =&L(\alpha_1 u_1 + \cdots + \alpha_k u_k) + L(\alpha_{k+1} u_{k+1}) \\ =&\alpha_1 L(u_1) + \cdots + \alpha_k L(u_k) + L(\alpha_{k+1} u_{k+1}) \end{aligned} \tag{10.1}$$

Note that $0_w = L(0_v) = L(u + (-u)) = L(u) + L(-u)$, this gives $L(-u) = -L(u)$ ■

**Example 10.2** Let T:$R^3 \longrightarrow R^3$ be a mapping defined as

$$T([x,y,z]^\top) = [3x + 2y, 0, 2x + y - 1]^\top$$

T is not a linear transformation since

$$T((0,0,0)^{\top}) = [0,0,-1]^{\top} \neq (0,0,0)^{\top}$$

**Example 10.3** Let T:$P_3 \longrightarrow r^{2X2}$ be a mapping defined as:

$$T(ax^3 + bx^2 + cx + d) = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

L is a linear transformation

**Theorem 10.2** Let V be a vector space with a basis $\mathcal{U} = \{u_1, u_2, \cdots, u_n\}$, and let x,y $\in$ $V.Forany\alpha, \beta \in$R, one has

$$[\alpha x + \beta y]_{\mathcal{U}} = \alpha[x]_{\mathcal{U}} + \beta[y]_{\mathcal{U}}$$

Therefore, $[\cdot]_{\mathcal{U}}$ is a linear transformation from V to R, where $[\cdot]_{\mathcal{U}}$ is the operator of taking the coordinate w.r.t. basis $\mathcal{U}$.

## 10.1.2. Kernel, Image, and Range

Conversely, given $m \times n$ matrix $\boldsymbol{A}$, $L(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$ defines a linear mapping. This is because matrix multiplication is also a linear operator.

R Transformations have a new "language". For example, for *nonlinear* transformation, if there is **no matrix**, we cannot talk about **column space**. But this idea could be rescued. We know the *column space* consists of all outputs $\boldsymbol{A}\boldsymbol{v}$, the *null space* consists of all inputs for which $\boldsymbol{A}\boldsymbol{v} = \boldsymbol{0}$. We could generalize those terms into "range" and "kernel":

### Kernel.

**Definition 10.2** [kernel] The kernel of $L$ refers to the set of all inputs for which $L(\boldsymbol{v}) = \boldsymbol{0}$, which is denoted as:

$$\ker(L) = \{\boldsymbol{x} : L(\boldsymbol{x}) = \boldsymbol{0}\}$$

**Kernel corresponds to the null space**. If $L(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$, we have $\ker(L) = N(\boldsymbol{A})$.

R For linear transformation $L : \boldsymbol{V} \mapsto \boldsymbol{W}$, where $L(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$. We have two rules:

$$L(\cdot) : \begin{cases} N(\boldsymbol{A}) \mapsto \{\boldsymbol{0}\} \\ \boldsymbol{V} \mapsto \operatorname{col}(\boldsymbol{A}) \end{cases}$$

**Definition 10.3** (Kernel of Linear Transformation) Let L be a linear transformation from V to W, then the kernel of L, denoted by ker(L) is defined as

$$ker(L) = \{\mathbf{v} \in V | L(\mathbf{v}) = \mathbf{0}_w\}$$

## Image and Range.

**Definition 10.4** [Image] Let L be a linear transformation from V to W and let S be a subspace of V , the image of S , denoted by L(S), is defined by

$$L(S) = \{\mathbf{w} \in W | \exists \mathbf{v} \in S, s.t. L(v) = w\}$$

**Definition 10.5** [range] For a linear transformation $L : V \mapsto W$, the range (or image) of $L$ refers to the set of all outputs $T(\boldsymbol{v})$, which is denoted as:

$$\operatorname{Range}(L) = \{L(\boldsymbol{x}) : x \in \boldsymbol{V}\}$$

The image of the entire vector space V, i.e., L(V) is called the **range** of L.

**The range corresponds to the column space**. If $L(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$, we have $\operatorname{Range}(L) = \mathcal{C}(\boldsymbol{A})$.

**Theorem 10.3** (Kernel and Image are subspaces) Let L be a linear transformation from V to W, and let S be a subspace of V, then

1. Ker(L) is a subspace of V.
2. L(S) is a subspace of W.

### Injective and Surjective.

**Definition 10.6** (Injective and Surjective)Let L be a linear transformation from V to W, then

1. L is injective if L(u)=L(v), then u=v.
2. L is surjective if $\forall w \in$ W, there exists a u $\in$ V, such that

$$L(u) = w$$

**Example 10.4**
1. L: $R^2 \longrightarrow R^3$ is a linear transformation, $L([x,y]^\top) = [x,y,0]^\top, Ker(L) = \{[0,0]^\top\}, L(R^2) = \{[x,y,0]^\top | x,y \in R\}$
2. L:$R^3 \longrightarrow R^2$ is a linear transformation, $L([x,y,z]^\top) = [x,y]^\top, Ker(L) = \{[0,0,z]^\top | z \in R\}, L(R^3) = R^2$.
3. Identity transformation I from vector space V to V, $Ker(I) = \{0\}, I(V) = V$.

**Theorem 10.4** (Injective Equivalent Condition) Let L be a linear transformation from V to W, then L is injective if and only if Ker(L)={0}

*Proof.* $\Longrightarrow$ If L(u)=0=L(0), then u=0. Thus Ker(L)={0}.

$\Longleftarrow$ Assume L(u)=L(v). then L(u-v)=0,then

$$u - v - 0$$

thus u=v. ∎

**Example 10.5**
1. L: $R^2 \longrightarrow R^3$ is a linear transformation, $L([x,y]^\top) = [x,y,0]^\top$,(Injective but not surjective) $Ker(L) = \{[0,0]^\top\}$
2. L:$R^3 \longrightarrow R^2$ is a linear transformation, $L([x,y,z]^\top) = [x,y]^\top$(Surjective but not injective), $Ker(L) = \{[0,0,z]^\top | z \in R\}$.
3. Identity transformation I from vector space V to V(Surjective and injective), $Ker(I) = \{0\}$.

## 10.1.3. Matrix Representation

### 10.1.3.1. Transformation of $L:\mathbb{R}^n \mapsto \mathbb{R}^m$

Given the linear transformation $L:\mathbb{R}^n \mapsto \mathbb{R}^m$, let's show that in order to study the output, it suffices to start from the **basis** of our output:

Assume the basis of $\mathbb{R}^n$ is $\{e_1,e_2,\dots,e_n\}$, where $L(e_i)=a_i \in \mathbb{R}^m$ for $i=1,\dots,n$. **The linearity of transformation extends to the combinations of $n$ vectors.**

Hence given any vector $\boldsymbol{x}=x_1e_1+x_2e_2+\cdots+x_ne_n \in \mathbb{R}^n$, we can express its transformation in matrix multiplication form:

$$\begin{aligned} L(\boldsymbol{x}) &= L(x_1e_1+x_2e_2+\cdots+x_ne_n) \\ &= x_1L(e_1)+x_2L(e_2)+\cdots+x_nL(e_n) \\ &= x_1a_1+x_2a_2+\cdots+x_na_n = \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} \\ &= \boldsymbol{A}\boldsymbol{x} \end{aligned}$$

where $a_i := L(e_i)$, and $\boldsymbol{A}$ is a $m\times n$ matrix with columns $a_1,\dots,a_n$.

**Theorem 10.5** **(Matrix Representation for linear reansformation between vector spaces w.r.t. standard bases)** If L is a linear transformation from $R^n$ to $R^m$, there is a mxn matrix A such that

$$L(x)=Ax$$

for each **x** $\in$R$^n$. In fact, the jth column vector of A=$[a_1,\cdots,a_n]$ is given by

$$a_j = L(e_j), j=1,2,\cdots,n$$

where $\mathcal{E}=\{r_1,\cdots,e_n\}$ is the standard basis pf $R^n$.

R 1.Linear transformation and its matrix representation is completely characterized by its action on a basis of its domain.

2.The matrix A in this theorem is the matrix representation of the linear transformation L w.r.t. standard bases of $R^n$ and $R^m$

**Example 10.6** $\begin{bmatrix}1\\0\end{bmatrix},\begin{bmatrix}0\\1\end{bmatrix}$ is the standard basis of $R^2$. Now look for action on this standard basis.

Since L( $\begin{bmatrix}1\\0\end{bmatrix}$ )= $\begin{bmatrix}\cos\theta\\\sin\theta\end{bmatrix}$ and L( $\begin{bmatrix}0\\1\end{bmatrix}$ )= $\begin{bmatrix}-\sin\theta\\\cos\theta\end{bmatrix}$ The matrix representation of L w.r.t the standard bases will be

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

And the rotation linear transform is

$$L(\begin{bmatrix}x\\y\end{bmatrix})=A\begin{bmatrix}x\\y\end{bmatrix}=\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}=\begin{bmatrix}x\cos\theta-y\sin\theta\\ x\sin\theta+y\cos\theta\end{bmatrix}$$

■

**Example 10.7** Define the linear transformation L:$R^3\longrightarrow R^2$ by

$$L(\begin{bmatrix}x_1\\x_2\\x3\end{bmatrix})=\begin{bmatrix}x_1+x_2\\x_2+x_3\end{bmatrix}$$

L is a linear transformation from $R^3\longrightarrow R^2$, it is completes characterized by its action on the standard basis of $R^3$.

The matrix A can be constructed as follows: the first column is L( $\begin{bmatrix}1\\0\\0\end{bmatrix}$ )= $\begin{bmatrix}1\\0\end{bmatrix}$ ,the second column is L( $\begin{bmatrix}0\\1\\0\end{bmatrix}$ )= $\begin{bmatrix}1\\1\end{bmatrix}$ , the third column isL( $\begin{bmatrix}0\\0\\1\end{bmatrix}$ )= $\begin{bmatrix}0\\1\end{bmatrix}$

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

### 10.1.3.2. Example: Differentiation

**Key idea of this section:**

> ***Suppose we know $L(\boldsymbol{v}_1),\ldots,L(\boldsymbol{v}_n)$ for the basis vectors $\boldsymbol{v}_1,\ldots,\boldsymbol{v}_n$,Then the linearity property produces $L(\boldsymbol{v})$ for every other input vector $\boldsymbol{v}$***

**Reason:** Every $\boldsymbol{v}$ has a unique combination $c_1\boldsymbol{v}_1+\cdots+c_n\boldsymbol{v}_n$ of the basis vector $\boldsymbol{v}_i$. Suppose $L$ is a linear transformation, then $L(\boldsymbol{v})$ must be the **same combination** $c_1L(\boldsymbol{v}_1)+\cdots+c_nL(v_n)$ of the **known outputs** $L(\boldsymbol{v}_i)$.

**Derivative is a linear transformation.** The derivative of the functions $1,x,x^2,x^3$ are $0,1,2x,3x^2$. If we consider "**taking the derivative**" as a transformation, whose inputs and outputs are functions, then we claim that the **derivative transformation** is **linear**:

$$L(\boldsymbol{v})=\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}x}\qquad\text{obeys the linearity rule}\qquad\frac{\mathrm{d}}{\mathrm{d}x}(c\boldsymbol{v}+d\boldsymbol{w})=c\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}x}+d\frac{\mathrm{d}\boldsymbol{w}}{\mathrm{d}x}$$

If we consider $1,x,x^2,x^3$ as vectors instead of functions, we notice they form a basis for the space $\boldsymbol{V}:=\{\textit{polynomials with degree}\le 3\}$. Find derivatives of these four basis tells us all derivatives in $\boldsymbol{V}$:

**Example 10.8** Given any vector $\boldsymbol{v}$ in $\boldsymbol{V}$, it can be expressed as $\boldsymbol{v}=a+bx+cx^2+dx^3$. We want to find the derivative transformation output for $\boldsymbol{v}$:

$$\begin{aligned}L(\boldsymbol{v})&=aL(1)+bL(x)+cL(x^2)+dL(x^3)\\&=a\times(0)+b\times(1)+c\times(2x)+d\times(3x^2)\\&=b+2cx+3dx^2\end{aligned}$$

Can we express this linear transformation $L$ by a matrix $\boldsymbol{A}$? The answer is *Yes*:

The derivative transforms the space $\boldsymbol{V}$ of cubics to the space $\boldsymbol{W}$ of quadratics. The basis for $\boldsymbol{V}$ is $1,x,x^2,x^3$. The basis for $\boldsymbol{W}$ is $1,x,x^2$. It follows that *The derivative matrix is 3 by 4*:

$$\boldsymbol{A}:=\begin{bmatrix}0&1&0&0\\0&0&2&0\\0&0&0&3\end{bmatrix}=\textbf{matrix form of derivative } L.$$

Why do we define the derivative matrix? Because **multiplying by $A$ agrees with transforming by $L$**. The derivative of $v = a + bx + cx^2 + dx^3$ is $L(v) = b + 2cx + 3dx^2$. The same numbers $b, 2c, 3d$ appear when we multiply by matrix $A$:

$$\textbf{\textit{Take the derivative}} \quad \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} b \\ 2c \\ 3d \end{bmatrix}.$$

What does the matrix $\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$ and $\begin{bmatrix} b \\ 2c \\ 3d \end{bmatrix}$ mean?

It is the **coordinate vector** of $v$ and $L(v)$. If we consider $a + bx + cx^2 + dx^3$ as a vector, then it's better for us to study its corresponding coordinate vector $\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$.

Hence, taking derivative of $v$ is the same as multiplying matrix $A$ by its coordinate vector. ■

### 10.1.3.3. The inverse of the derivative.

**The integral is the inverse of the derivative. .** That is from the Fundamental Theorem of Calculus. We review it from the perspective of linear algebra. The integral transformation $L^{-1}$ that *takes the integral from 0 to x* is also linear! Applying $L^{-1}$ to $1, x, x^2$, which are $w_1, w_2, w_3$:

$$\textbf{Integration is } L^{-1} \qquad \int_0^x 1\,\mathrm{d}x = x, \quad \int_0^x x\,\mathrm{d}x = \frac{1}{2}x^2, \quad \int_0^x x^2\,\mathrm{d}x = \frac{1}{3}x^3.$$

By linearity, the integral of $w = B + Cx + Dx^2$ is $L^{-1}(w) = Bx + \frac{1}{2}Cx^2 + \frac{1}{3}Dx^3$. The integral of a quadratic is a cubic. The input space of $L^{-1}$ is the quadratics, the output space is the cubics. **Integration takes W back to V**. Integration matrix will be 4 by 3:

$$\textbf{Take the integral} \quad \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix} \begin{bmatrix} B \\ C \\ D \end{bmatrix} = \begin{bmatrix} 0 \\ B \\ \frac{1}{2}C \\ \frac{1}{3}D \end{bmatrix}.$$

If our input is $\boldsymbol{w} = B + Cx + Dx^2$, our output integral is $0 + Bx + \frac{1}{2}Cx^2 + \frac{1}{3}Dx^3$.

**The derivative and the integration are essentially matrix multiplication.** We have the corresponding derivative and integration matrix:

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} \qquad \boldsymbol{A}^{-1} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix}$$

I want to call this matrix $\boldsymbol{A}^{-1}$, though rectangular matrices don't have inverses. Note that $\boldsymbol{A}^{-1}$ is the **right inverse** of matrix $\boldsymbol{A}$!

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{but} \quad \boldsymbol{A}^{-1}\boldsymbol{A} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

This is reasonable. If you integrate a function and then differentiate, you get back to the start. Hence $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}$. But if you differentiate before integrating, the constant term is lost.

**The integral of the derivative of 1 is zero.**

$$L^{-1}L(1) = \text{integral of zero function} = 0.$$

**Summary:.** In this example, we want to take the derivative. Then we let $\boldsymbol{V}$ be a vector space of polynomials with degree $\leq 3$. Its basis is given by $E = \{1, x, x^2, x^3\}$. Any $v \in \boldsymbol{V}$ there is a unique linear combination of the basis vectors that equals to $v$:

$$v = a + bx + cx^2 + dx^3$$

We write the coordinate vector of $v$ w.r.t. to $E$:

$$[v]_E = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$$

Then we postmultiply $\boldsymbol{A}$ by $[v]_E$ to get the corresponding coordinate vector of output space:

$$[L(v)]_F = \boldsymbol{A}[v]_E$$

where $F = \{1, x, x^2\}$.

Here we give the formal definition for the coordinate vector:

**Definition 10.7** [coordinate vector] Let $\boldsymbol{V}$ be a vector space of dimension $n$ and let $B = \{v_1, v_2, \dots, v_n\}$ be an **ordered** basis for $\boldsymbol{V}$. Then for any $v \in \boldsymbol{V}$ there is a unique linear combination of the basis vectors such that

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n$$

where $\alpha_1, \dots, \alpha_n$ are scalars.

The **coordinate vector** of $v$ w.r.t. to $B$ is defined by

$$[v]_B = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}$$

Hence, vector $v$ could be expressed as: $v = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix} \times [v]_B$. ■

More specifically, the linear transformation of vectors is essentially the matrix multiplication of the corresponding coordinate vectors:

**Theorem 10.6** Let $E = \{v_1, \dots, v_n\}$ be a basis for $\boldsymbol{V}$; $F = \{w_1, \dots, w_m\}$ be a basis for $\boldsymbol{W}$. Given linear transformation $L : \boldsymbol{V} \mapsto \boldsymbol{W}$, for any vector $v \in \boldsymbol{V}$, there exists $m \times n$ matrix $\boldsymbol{A}$ such that

$$[L(v)]_F = \boldsymbol{A}[v]_E$$

If we let $\boldsymbol{W} = \boldsymbol{V}$, then we obtain a more commonly useful corollary:

**Corollary 10.1** Given linear transformation $L : \boldsymbol{V} \mapsto \boldsymbol{V}$. We set $E = \{\alpha_1, \ldots, \alpha_n\}$ to be the basis of $\boldsymbol{V}$. Then given any vector $v$, there exists $n \times n$ matrix $\boldsymbol{A}$ such that

$$[L(v)]_E = \boldsymbol{A}[v]_E$$

# 10.2. Lecture 20: linear transformation

## 10.2.1. Orthogonality

Recall that two vectors are orthogonal if their inner product is zero:

$$\boldsymbol{u} \perp \boldsymbol{v} \Longleftrightarrow \langle \boldsymbol{u}, \boldsymbol{v} \rangle = 0$$

Orthogonality among vectors has an important property:

**Proposition 10.1** If **nonzero** vectors $v_1, \ldots, v_k$ are mutually orthogonal, i.e., $v_i \perp v_j$ for any $i \neq j$, then $\{v_1, \ldots, v_k\}$ must be ind.

*Proof.* It suffices to show that

$$\alpha_1 v_1 + \cdots + \alpha_k v_k = \mathbf{0} \implies \alpha_i = 0 \text{ for any } i \in \{1, 2, \ldots, k\}.$$

- We do inner product to show $\alpha_1$ must be zero:

$$\begin{aligned}
\langle v_1, \alpha_1 v_1 + \cdots + \alpha_k v_k \rangle &= \langle v_1, \mathbf{0} \rangle = 0 \\
&= \alpha_1 \langle v_1, v_1 \rangle + \alpha_2 \langle v_1, v_2 \rangle + \cdots + \alpha_k \langle v_1, v_k \rangle \\
&= \alpha_1 \langle v_1, v_1 \rangle = \alpha_1 \|v_1\|_2^2 \\
&= 0
\end{aligned}$$

  Since $v_1 \neq \mathbf{0}$, we have $\alpha_1 = 0$.

- Similarly, we have $\alpha_i = 0$ for $i = 1, \ldots, k$.

■

Now we can also talk about orthogonality among spaces:

**Definition 10.8** [Subspace Orthogonality] Two subspaces $\boldsymbol{U}$ and $\boldsymbol{V}$ of a vector space are **orthogonal** if every vector $\boldsymbol{u}$ in $\boldsymbol{U}$ is *perpendicular* to every vector $\boldsymbol{v}$ in $\boldsymbol{V}$:

$$\textbf{Orthogonal subspaces} \quad \boldsymbol{u} \perp \boldsymbol{v}, \quad \forall \boldsymbol{u} \in \boldsymbol{U}, \boldsymbol{v} \in \boldsymbol{V}.$$

**Example 10.9** Two walls look *perpendicular* but they are not orthogonal subspaces! The meeting line is in both $\boldsymbol{U}$ and $\boldsymbol{V}$-and this line is not perpendicular to itself. Hence, two planes (both with dimension 2 in $\mathbb{R}^3$) cannot be orthogonal subspaces.

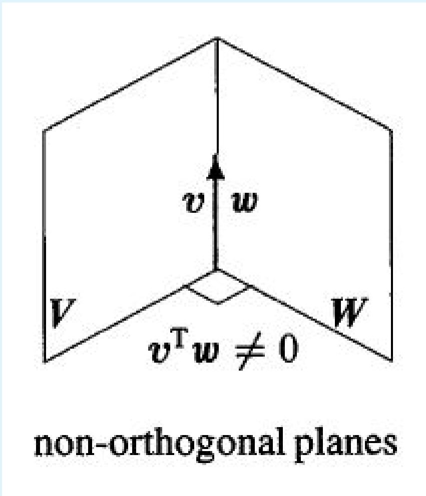



Figure 10.1: Orthogonality is impossible when $\dim \boldsymbol{U} + \dim \boldsymbol{V} > \dim(\boldsymbol{U} \cup \boldsymbol{V})$

■

R When a vector is in two orthogonal subspaces, it *must* be zero. It is **perpendicular** to itself.

The reason is clear: this vector $\boldsymbol{u} \in \boldsymbol{U}$ and $\boldsymbol{u} \in \boldsymbol{V}$, so $\langle \boldsymbol{u}, \boldsymbol{u} \rangle = 0$. It has to be zero vector.

If two subspaces are perpendicular, their basis must be ind.

**Theorem 10.7** Assume $\{u_1, \ldots, u_k\}$ is the basis for $\boldsymbol{U}$, $\{v_1, \ldots, v_l\}$ is the basis for $\boldsymbol{V}$. If $\boldsymbol{U} \perp \boldsymbol{V}$ ($u_i \perp v_j$ for $\forall i, j$), then $u_1, u_2, \ldots, u_k, v_1, v_2, \ldots, v_l$ must be ind.

*Proof.* Suppose there exists $\{\alpha_1, \ldots, \alpha_k\}$ and $\{\beta_1, \ldots, \beta_l\}$ such that

$$\alpha_1 u_1 + \cdots + \alpha_k u_k + \beta_1 v_1 + \cdots + \beta_l v_l = \mathbf{0}$$

then equivalently,

$$\alpha_1 u_1 + \cdots + \alpha_k u_k = -(\beta_1 v_1 + \cdots + \beta_l v_l).$$

Then we set $\boldsymbol{w} = \alpha_1 u_1 + \cdots + \alpha_k u_k$, obviously, $\boldsymbol{w} \in \boldsymbol{U}$ and $\boldsymbol{w} \in \boldsymbol{V}$.

Hence it must be zero (This is due to remark above). Thus we have

$$\alpha_1 u_1 + \cdots + \alpha_k u_k = \mathbf{0}$$
$$\beta_1 v_1 + \cdots + \beta_l v_l = \mathbf{0}.$$

Due to the independence, we have $\alpha_i = 0$ and $\beta_j = 0$ for $\forall i, j$. ■

**Corollary 10.2** For subspaces $\boldsymbol{U}$ and $\boldsymbol{V}$, we obtain

$$\dim(\boldsymbol{U} \cup \boldsymbol{V}) \leq \dim(\boldsymbol{U}) + \dim(\boldsymbol{V}).$$

For subspaces $\boldsymbol{U}$ and $\boldsymbol{V} \in \mathbb{R}^n$, if $\mathbb{R}^n = \boldsymbol{U} \cup \boldsymbol{V}$, and moreover, $n = \dim(\boldsymbol{U}) + \dim(\boldsymbol{V})$, then we say $\boldsymbol{V}$ is the **orthogonal complement** of $\boldsymbol{U}$.

**Definition 10.9** [orthogonal complement] For subspaces $\boldsymbol{U}$ and $\boldsymbol{V} \in \mathbb{R}^n$, if $\dim(\boldsymbol{U}) + \dim(\boldsymbol{V}) = n$ and $\boldsymbol{U} \perp \boldsymbol{V}$, then we say $\boldsymbol{V}$ is the **orthogonal complement** of $\boldsymbol{U}$. We denote $\boldsymbol{V}$ as $\boldsymbol{U}^\perp$.

Moreover, $\boldsymbol{V} = \boldsymbol{U}^\perp$ iff $\boldsymbol{V}^\perp = \boldsymbol{U}$.

**Example 10.10** Suppose $\boldsymbol{U} \cup \boldsymbol{V} = \mathbb{R}^3$, $\boldsymbol{U} = \text{span}\{\boldsymbol{e}_1, \boldsymbol{e}_2\}$. If $\boldsymbol{V}$ is the orthogonal complement of $\boldsymbol{U}$, then $\boldsymbol{V} = \text{span}\{\boldsymbol{e}_3\}$.

Next we study the relationship between the null space and the row space in $\mathbb{R}^n$.

**Theorem 10.8 — Fundamental theorem for linear alegbra, part 2.** Given $\boldsymbol{A} \in \mathbb{R}^{m \times n}$,

**$N(\boldsymbol{A})$ is the orthogonal complement of the row space of $\boldsymbol{A}$, $\mathcal{C}(\boldsymbol{A}^{\mathrm{T}})$ (in $\mathbb{R}^n$).**

**$N(\boldsymbol{A}^{\mathrm{T}})$ is the orthogonal complement of the column space $\mathcal{C}(\boldsymbol{A})$ (in $\mathbb{R}^m$).**

*Proof.*
- Firstly, we show $\dim(N(\boldsymbol{A})) + \dim(\mathcal{C}(\boldsymbol{A}^{\mathrm{T}})) = n$:

  We know that $\dim(N(\boldsymbol{A})) = n - r$ and $\dim(\mathcal{C}(\boldsymbol{A}^{\mathrm{T}})) = r$, where $r = \text{rank}(\boldsymbol{A})$.

  Hence $\dim(N(\boldsymbol{A})) + \dim(\mathcal{C}(\boldsymbol{A}^{\mathrm{T}})) = n$.

- Then we show $N(\boldsymbol{A}) \perp \mathcal{C}(\boldsymbol{A}^{\mathrm{T}})$:

For any $x \in N(\boldsymbol{A})$, if we set $\boldsymbol{A} = \begin{bmatrix} a_1^{\mathrm{T}} \\ a_2^{\mathrm{T}} \\ \vdots \\ a_m^{\mathrm{T}} \end{bmatrix}$, then we obtain:

$$\boldsymbol{A}\boldsymbol{x} = \begin{bmatrix} a_1^{\mathrm{T}} \\ a_2^{\mathrm{T}} \\ \vdots \\ a_m^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Hence *every row has a zero product with* $\boldsymbol{x}$, i.e., $\langle a_i, \boldsymbol{x} \rangle = 0$ for $\forall i \in \{1, 2, \dots, m\}$.

For any $y = \sum_{i=1}^m \alpha_i a_i \in \mathcal{C}(\boldsymbol{A}^{\mathrm{T}})$, we obtain:

$$\begin{aligned} \langle \boldsymbol{x}, y \rangle = \langle y, \boldsymbol{x} \rangle &= \langle \sum_{i=1}^m \alpha_i a_i, \boldsymbol{x} \rangle \\ &= \sum_{i=1}^m \alpha_i \langle a_i, \boldsymbol{x} \rangle = 0. \end{aligned}$$

Hence $\boldsymbol{x} \perp y$ for $\forall \boldsymbol{x} \in N(\boldsymbol{A})$ and $y \in \mathcal{C}(\boldsymbol{A}^{\mathrm{T}})$.

Hence $N(\boldsymbol{A})^{\perp} = \mathcal{C}(\boldsymbol{A}^{\mathrm{T}})$. Similarly, we have $N(\boldsymbol{A}^{\mathrm{T}})^{\perp} = \mathcal{C}(\boldsymbol{A})$. ■

**Corollary 10.3** $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ is solvable if and only if $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{0}$ implies $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{b}$=0.

*Proof.* The following statements are equivalent:

- $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ is solvable.
- $\boldsymbol{b} \in \mathcal{C}(\boldsymbol{A})$.
- $\boldsymbol{b} \in N(\boldsymbol{A}^{\mathrm{T}})^{\perp}$
- $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{b} = 0$ for $\forall y \in N(\boldsymbol{A}^{\mathrm{T}})$
- Given $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{0}$, i.e., $y \in N(\boldsymbol{A}^{\mathrm{T}})$, it implies $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{b} = 0$.

■

The **Inverse Negative Proposition** is more commonly useful:

**Corollary 10.4** $\boldsymbol{Ax} = \boldsymbol{b}$ has no solution if and and only if $\exists \boldsymbol{y}$ s.t. $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A} = 0$ and $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{b} \neq 0$.

We could extend this corollary into general case:

R

**Theorem 10.9** $\boldsymbol{Ax} \geq \boldsymbol{b}$ has no solution if and only if $\exists \boldsymbol{y} \geq \boldsymbol{0}$ such that $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{0}$ and $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{b} \geq \boldsymbol{0}$.

$\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A} = 0$ requires that there exists one linear combination of the row space to be zero.

The complete proof for this theorem is not required in this course. We only show the necessity case.

*Necessity case.* Suppose $\exists \boldsymbol{y} \geq \boldsymbol{0}$ such that $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{0}$ and $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{b} \geq \boldsymbol{0}$. Assume there exists $x^*$ such that $\boldsymbol{A}x^* \geq \boldsymbol{b}$. By postmultiplying $\boldsymbol{y}^{\mathrm{T}}$ we have

$$\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}x^* \geq \boldsymbol{y}^{\mathrm{T}}\boldsymbol{b} > \boldsymbol{0} \implies \boldsymbol{0} > \boldsymbol{0}.$$

which is a contradiction! ■

**Example 10.11** Given the system

$$x_1 + x_2 \geq 1 \tag{10.2}$$

$$-x_1 \geq -1 \tag{10.3}$$

$$-x_2 \geq 2 \tag{10.4}$$

Eq.(10.2)×1+Eq(10.3)×1+Eq.(10.4)×1 gives

$$0 \geq 2$$

which is a contradiction!

So the key idea of theorem (10.9) is to construct a linear combination of row space to let it become zero. If the right hand is larger than zero, then this system has no solution.

R

**Corollary 10.5** If $\boldsymbol{A} = \boldsymbol{A}^{\mathrm{T}}$, then $N(\boldsymbol{A}^{\mathrm{T}})^{\perp} = \mathcal{C}(A) = \mathcal{C}(\boldsymbol{A}^{\mathrm{T}}) = N(\boldsymbol{A})$.

**Corollary 10.6** The system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ may not have a solution, but $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ always have at least one solution for $\forall \boldsymbol{b}$.

*Proof.* Since $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is symmetric, we have $\mathcal{C}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = \mathcal{C}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})$. Show by yourself that $\mathcal{C}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}) = \mathcal{C}(\boldsymbol{A}^{\mathrm{T}})$, hence $\mathcal{C}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = \mathcal{C}(\boldsymbol{A}^{\mathrm{T}})$.

For any vector $\boldsymbol{b}$, we have $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} \in \mathcal{C}(\boldsymbol{A}^{\mathrm{T}}) \implies \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} \in \mathcal{C}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})$, which means there exists a linear combination of the columns of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ that equals to $\boldsymbol{b}$.

Or equivalently, there exists a solution to $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$. ■

**Corollary 10.7** $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is invertible if and only if $\boldsymbol{A}$ is full column rank, i.e., columns of $\boldsymbol{A}$ are ind.

*Proof.* We have shown that $C(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = C(\boldsymbol{A}^{\mathrm{T}})$.

Hence $C(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{\perp} = C(\boldsymbol{A}^{\mathrm{T}})^{\perp} \implies N(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = N(\boldsymbol{A})$.

Thus, the following statements are equivalent:

- $\boldsymbol{A}$ has ind. columns
- $N(\boldsymbol{A}) = \{\boldsymbol{0}\}$
- $N(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = \{\boldsymbol{0}\}$
- $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is invertible.

■

### 10.2.2. Projections

In corollary (10.7), we know that if $\boldsymbol{A}$ has ind. columns, then $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is invertible. On this condition, the normal equation $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ has the unique solution $\boldsymbol{x}^* = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$, which follows that the error $\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^*$ is minimized. Note that $\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ is **approximately** equal to $\boldsymbol{b}$.

- If $\boldsymbol{b}$ and $\boldsymbol{A}\boldsymbol{x}^*$ are exactly in the same space, i.e., $\boldsymbol{b} \in \mathcal{C}(\boldsymbol{A})$, then $\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{b}$. The error is equal to zero.

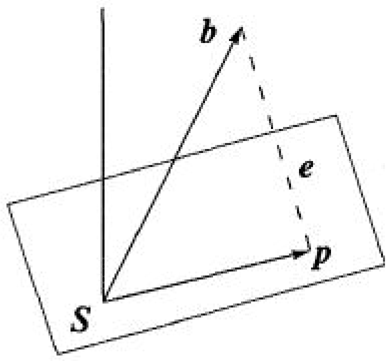



Figure 10.2: The projection of $\boldsymbol{b}$ onto a subspace $\boldsymbol{S} := \mathcal{C}(\boldsymbol{A})$.

- Otherwise, just as the Figure (10.2) shown, $\boldsymbol{A}\boldsymbol{x}^*$ is the projection of $\boldsymbol{b}$ to subspace $\mathcal{C}(\boldsymbol{A})$.

**Definition 10.10** [Projection] Let $\boldsymbol{S} \in \mathbb{R}^m$ be a non-empty closed set and $\boldsymbol{b} \in \mathbb{R}^m$ be given. Then the projection of $\boldsymbol{b}$ onto the set $\boldsymbol{S}$ is the solution to

$$\min_{\boldsymbol{z} \in \boldsymbol{S}} \|\boldsymbol{z} - \boldsymbol{b}\|_2^2,$$

where we use notation $\text{Proj}_{\boldsymbol{S}}(\boldsymbol{b})$ to denote the projection of $\boldsymbol{b}$ onto $\boldsymbol{S}$.

By definition, the projection of $\boldsymbol{b}$ onto the subspace $\mathcal{C}(\boldsymbol{A})$ is given by

$$\text{Proj}_{\mathcal{C}(\boldsymbol{A})}(\boldsymbol{b}) := \boldsymbol{A}\boldsymbol{x}^*, \quad \text{where } \boldsymbol{x}^* = \arg\min_{\boldsymbol{x} \in \mathbb{R}^n} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|.$$

**Definition 10.11** [Projection matrix] Given the projection

$$\text{Proj}_{\mathcal{C}(\boldsymbol{A})}(\boldsymbol{b}) := \boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b},$$

since $[\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}]\boldsymbol{b}$, we call the projection operator $\boldsymbol{P} := \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}$ as the **projection matrix** of $\boldsymbol{A}$.

**Definition 10.12** [Idempotent] Let $\boldsymbol{A}$ be a **square** matrix that satisfies $\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}$, then $\boldsymbol{A}$ is called an **idempotent** matrix.

Let's show that the projection matrix is *idempotent*:

$$\begin{aligned} \boldsymbol{P}^2 &= \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}} \\ &= \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}} \\ &= \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{P}. \end{aligned}$$

### 10.2.2.1. Observations

- Suppose $\boldsymbol{b} \in \mathcal{C}(\boldsymbol{A})$, i.e., $\exists \boldsymbol{x}$ s.t. $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$. Then the projection of $\boldsymbol{b}$ is exactly $\boldsymbol{b}$:

$$\begin{aligned} \boldsymbol{P}\boldsymbol{b} &= \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{b}) \\ &= \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}) \\ &= \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})\boldsymbol{x} \\ &= \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}. \end{aligned}$$

- Assume $\boldsymbol{A}$ has only one column, say, $\boldsymbol{a}$. Then we have

$$\begin{aligned} \boldsymbol{x}^* &= (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} = \frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}} \\ \boldsymbol{A}\boldsymbol{x}^* &= \boldsymbol{P}\boldsymbol{b} = \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{b}) = \frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}} \times \boldsymbol{a} = \frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\|\boldsymbol{a}\|^2} \times \boldsymbol{a} \end{aligned}$$

  More interestingly,

$$\frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\|\boldsymbol{a}\|^2} \times \boldsymbol{a} = \frac{\|\boldsymbol{a}\|\|\boldsymbol{b}\|\cos\theta}{\|\boldsymbol{a}\|^2} \times \boldsymbol{a} = \|\boldsymbol{b}\|\cos\theta \times \frac{\boldsymbol{a}}{\|\boldsymbol{a}\|}$$

  which is the projection of $\boldsymbol{b}$ onto a line $\operatorname{span}\{\boldsymbol{a}\}$. (Shown in figure (10.3).)

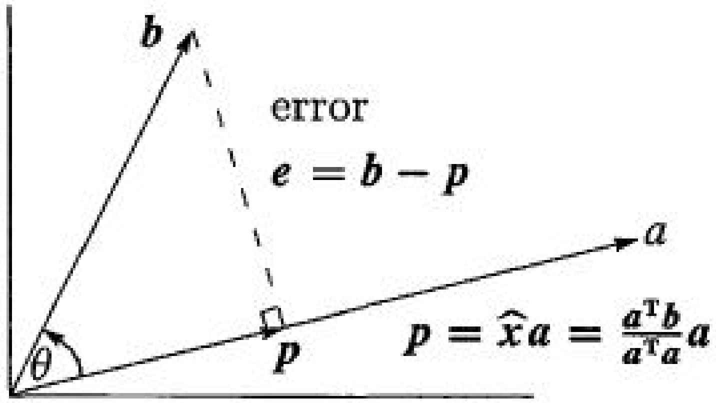



Figure 10.3: The projection of $\boldsymbol{b}$ onto a line $\boldsymbol{a}$.

More generally, we can write the projection of $\boldsymbol{b}$ onto the line span$\{\boldsymbol{a}\}$ as:

$$\text{Proj}_{\text{span}\{\boldsymbol{a}\}}(\boldsymbol{b}) = \frac{\langle \boldsymbol{a}, \boldsymbol{b} \rangle}{\langle \boldsymbol{a}, \boldsymbol{a} \rangle} \boldsymbol{a}$$

**Changing an Orthogonal Basis.** Note that the error $\boldsymbol{b} - \text{Proj}_{\text{span}\{\boldsymbol{a}\}}(\boldsymbol{b})$ is perpendicular to $\boldsymbol{a}$, and $\boldsymbol{b} - \text{Proj}_{\text{span}\{\boldsymbol{a}\}}(\boldsymbol{b}) \in \text{span}\{\boldsymbol{a}, \boldsymbol{b}\}$.

If we define $\boldsymbol{b}' = \boldsymbol{b} - \text{Proj}_{\text{span}\{\boldsymbol{a}\}}(\boldsymbol{b})$, then it's easy to check that $\text{span}\{\boldsymbol{a}, \boldsymbol{b}'\} = \text{span}\{\boldsymbol{a}, \boldsymbol{b}\}$ and $\boldsymbol{a} \perp \boldsymbol{b}'$.

Hence, we convert the basis $\{\boldsymbol{a}, \boldsymbol{b}\}$ into another basis $\{\boldsymbol{a}, \boldsymbol{b}'\}$ such that the elements are orthogonal to each other. For general subspace we could also use this approach to obtain an orthogonal basis, which will be discussed in next lecture.

## 10.3. Lecture 21: Eigenvalues and Eigenvectors

### 10.3.1. Why do we study eigenvalues and eigenvectors?

- **Motivation 1:** If we consider matrices as the *movements* (linear transformation) for *vectors* in vector space. Then roughly speaking, *eigenvalues* are the *speed* of the movements, *eigenvectors* are the *direction* of the movements
- **Motivation 2:** We know that linear transformation has different matrix representation for different basis. But which representation is **simplest** for a linear transformation? This topic gives us answer to this question.

When vectors are multiplied by $\boldsymbol{A}$, almost all vectors change direction. If $\boldsymbol{x}$ has the same direction as $\boldsymbol{A}\boldsymbol{x}$, they are called **eigenvectors**.

**The key equation is** $\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}$, **The number $\lambda$ is the eigenvalue of $\boldsymbol{A}$.**

**Definition 10.13** [Eigenvectors and Eigenvalues] Given a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ (or $\mathbb{C}^{n\times n}$), if there exist a vector $\boldsymbol{v} \in \mathbb{C}^n$ with $\boldsymbol{v} \neq \boldsymbol{0}$ such that

$$\boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v}, \quad \text{for some } \lambda \in \mathbb{C} \tag{10.5}$$

then $\lambda$ is called an **eigenvalue** of $\boldsymbol{A}$; $\boldsymbol{v}$ is called an **eigenvector** of $\boldsymbol{A}$ associated with $\lambda$.

Eigenvalues can be complex numbers, even for real matrices!

We illustrate an example of an eigenvalue problem:

**Example 10.12** Consider $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ and $x = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. $Ax = \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 1$. Thus, 1 is an eigenvalue of $A$

associated with eigenvector $x = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

**Example 10.13** Consider an eigenvalue problem $\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}$, where

$$\boldsymbol{A} = \begin{bmatrix} 4 & -2 \\ 1 & 1 \end{bmatrix}, \quad \boldsymbol{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

We can verify that

$$\boldsymbol{A}\boldsymbol{x} = \begin{bmatrix} 6 \\ 3 \end{bmatrix} = 3 \begin{bmatrix} 2 \\ 1 \end{bmatrix} = 3\boldsymbol{x}$$

Therefore, $\lambda = 3$ is the eigenvalue of $\boldsymbol{A}$; $\boldsymbol{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ is the eigenvector of $\boldsymbol{A}$ associated with $\lambda = 3$.

**Proposition 10.2** If $(\boldsymbol{v}, \lambda)$ is an eigen-pair of $\boldsymbol{A}$, then $(\alpha \boldsymbol{v}, \lambda)$ is also an eigen-pair for any $\alpha \in \mathbb{C}, \alpha \neq 0$.

#### 10.3.1.1. Calculation for eigen-pairs

How to find eigen-pairs $(\lambda, \boldsymbol{x})$? In other words, how to solve the nonlinear equation $\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}$, where $\lambda$ and $\boldsymbol{x}$ are unknowns? Consider a simpler case. If we can know the eigenvalues $\lambda$, then we can solve the linear system $(\lambda \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$ to get the corresponding eigenvectors.

But how to find eigenvalues? $\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}$ has a nonzero solution $\Longleftrightarrow (\lambda \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$ has a nonzero solution $\Longleftrightarrow$ $(\lambda \boldsymbol{I} - \boldsymbol{A})$ is singular $\Longleftrightarrow \det(\lambda \boldsymbol{I} - \boldsymbol{A}) = 0$.

Therefore, solving the determinant equation gives a way to find eigenvalues:

**Proposition 10.3** The number $\lambda$ is the eigenvalue of $\boldsymbol{A}$ if and only if $\lambda \boldsymbol{I} - \boldsymbol{A}$ is singular.

$$\textbf{Equation for the eigenvalues} \quad \det(\lambda \boldsymbol{I} - \boldsymbol{A}) = 0. \tag{10.6}$$

If $A = (a_{ij})_{n\times n}$, then

$$\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \dots & -a_{nn} \\ -a_{21} & \lambda - a_{22} & \dots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & \dots & \dots & \lambda - a_{nn} \end{vmatrix}$$

this is a monic polynomial of degree $n$.

**Definition 10.14** [characteristic polynomial] Define $P_{\boldsymbol{A}}(\lambda) := \det(\lambda \boldsymbol{I} - \boldsymbol{A})$.

$P_{\boldsymbol{A}}(\lambda)$ is called the **characteristic polynomial** for the matrix $\boldsymbol{A}$; the equation $\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = 0$ is called the **characteristic equation** for the matrix $\boldsymbol{A}$; the set $Null(\lambda \boldsymbol{I} - \boldsymbol{A})$ is called the **eigenspace** with respect to $\lambda$.

If $P_{\boldsymbol{A}}(\lambda^*) = 0$, then we say $\lambda^*$ is the root of $P_{\boldsymbol{A}}(\lambda)$.

**Proposition 10.4**

$$Null(\lambda I_n - A) = \{\text{the set of all eigenvectors of A with respect to } \lambda\} \cup \{0\}$$

**Example 10.14** Find the eigenvalues and eigenvectors of $\boldsymbol{A} = \begin{bmatrix} 3 & 2 \\ 3 & -2 \end{bmatrix}$.

$$\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = \begin{bmatrix} \lambda - 3 & -2 \\ -3 & \lambda + 2 \end{bmatrix} = 0.$$

$$\implies (\lambda + 3)(\lambda - 2) - 6 = 0. \implies \lambda^2 - \lambda - 12 = 0. \implies \lambda_1 = 4 \quad \lambda_2 = -3.$$

Eigenvalues of $\boldsymbol{A}$ are $\lambda_1 = 4$ and $\lambda_2 = -3$.

In order to get eigenvectors, we solve $(\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{x} = \boldsymbol{0}$:

- For $\lambda_1$, $(\boldsymbol{A} - \lambda_1 \boldsymbol{I})\boldsymbol{x} = \begin{bmatrix} -1 & 2 \\ 3 & -6 \end{bmatrix} = \boldsymbol{0}$.

$$\implies \boldsymbol{x} = \begin{bmatrix} 2x_2 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

Hence any $\alpha \begin{bmatrix} 2 & 1 \end{bmatrix}^{\mathrm{T}}$ $(\alpha \neq 0)$ is the eigenvector of $\boldsymbol{A}$ associated with $\lambda_1 = 4$.

- For $\lambda_2$, similarly, we derive

$$\boldsymbol{x} = \begin{bmatrix} -x_2 \\ 3x_2 \end{bmatrix} = x_2 \begin{bmatrix} -1 \\ 3 \end{bmatrix}$$

Hence any $\beta \begin{bmatrix} -1 & 3 \end{bmatrix}^{\mathrm{T}}$ $(\beta \neq 0)$ is the eigenvector of $\boldsymbol{A}$ associated with $\lambda_2 = -3$.

**Proposition 10.5** Let $A \in \mathbb{R}^{n\times n}$ (or $\mathbb{C}^{n\times n}$) be a square matrix. If $\lambda$ is an eigenvalue of $A$, then $\lambda$ is also an eigenvalue of $A^T$.

*Proof.*

$$0 = \det(\lambda I_n - A) = \det\left((\lambda I_n - A)^T\right) = \det(\lambda I_n - A^T)$$

Thus, $\lambda$ is also an eigenvalue of $A^T$. ■

### 10.3.1.2. Possible difficulty: how to solve $\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = 0$?

$P_{\boldsymbol{A}}(\lambda)$ is a characteristic polynomial with degree $n$. Actually, we can write $P_{\boldsymbol{A}}(\lambda)$ as:

$$P_{\boldsymbol{A}}(\lambda) = \lambda^n - a_1\lambda^{n-1} + a_2\lambda^{n-2} - \cdots + (-1)^n a_n$$

where $a_i$'s depend on matrix $\boldsymbol{A}$.

When $n$ increases, it's hard to find its roots:

- When $n = 2,3,4$, solution to $P_{\boldsymbol{A}}(\lambda) = 0$ has the *closed form*, which has been proved in 15th century.
- However, when $n \geq 5$, the characteristic equation has *no closed form* solution.

Although we cannot find closed form solution for large $n$, we want to study whether this characteristic polynomial with degree $n$ has exactly $n$ solutions. Gauss gives us the answer:

**Theorem 10.10 — Fundamental theorem of algebra.** Every nonzero, single variable, degree $n$ polynomial with *complex coefficients* has *exactly* $n$ complex roots. (Counted with multiplicity.)

What's the meaning of *multiplicity*? For example, the polynomial $(x-1)^2$ has one root 1 with multiplicity 2.

**Implication.** Hence, every polynomial $f(x)$ could be written as

$$\begin{aligned}f(x) &= a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x_1 + a_0 \\ &= a_n(x-x_1)(x-x_2)\dots(x-x_n)\end{aligned}$$

where $x_i$'s are roots for $f(x)$.

Moreover, $P_\lambda(\boldsymbol{A})$ **has exactly** $n$ **roots, i.e.,** $\boldsymbol{A}$ **has** $n$ **eigenvalues.(counted with multiplicity.)**

R Exact roots are almost impossible to find. But approximate roots (eigenvalues) can be find easily by numerical algorithm.

### 10.3.2. Products and Sums of Eigenvalue

#### The coefficient of the highest order for the characteristic polynomial is 1.

Suppose $P_{\boldsymbol{A}}(\lambda) = \det(\lambda \boldsymbol{I} - \boldsymbol{A})$ has $n$ roots $\lambda_1,\dots,\lambda_n$, then we obtain:

$$P_{\boldsymbol{A}}(\lambda) = \det(\lambda \boldsymbol{I} - \boldsymbol{A}) = (\lambda-\lambda_1)\dots(\lambda-\lambda_n) \tag{10.7}$$

Why the coefficient for $\lambda^n$ is 1 in equation (10.7)? If we expand $\det(\lambda \boldsymbol{I} - \boldsymbol{A})$, we find

$$\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \dots & -a_{nn} \\ -a_{21} & \lambda - a_{22} & \dots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & \dots & \dots & \lambda - a_{nn} \end{vmatrix}, \tag{10.8}$$

in which the variable $\lambda$ only appears in diagonal. By expaning the determinant, the coefficient of highest order is obviously 1.

#### The sum of eigenvalues equals to the sum of the $n$ diagonal entries of $\boldsymbol{A}$.

In (10.7), the coefficient of $\lambda^{n-1}$ is

$$-(\lambda_1 + \lambda_2 + \cdots + \lambda_n)$$

In (10.8), $\lambda^{n-1}$ only appears among $(\lambda - a_{11})(\lambda - a_{22})\ldots(\lambda - a_{nn})$, i.e., the coefficient of $\lambda^{n-1}$ is

$$-(a_{11} + a_{22} + \cdots + a_{nn})$$

Consequently, as $(10.7) = (10.8)$, we obtain

$$\sum \lambda_i = \textbf{trace} = \sum a_{ii}$$

The sum of the entries on the main diagonal is called the **trace** of $\boldsymbol{A}$, denoted by $\text{trace}(\boldsymbol{A})$.

**The product of the eigenvalues equals to the determinant of $\boldsymbol{A}$.** If let $\lambda = 0$ in (10.7), then we obtain $\det(-\boldsymbol{A}) = (-1)^n \lambda_1 \lambda_2 \ldots \lambda_n$. Obviously, $\det(-\boldsymbol{A}) = (-1)^n \det(\boldsymbol{A})$.

Hence $(-1)^n \det(\boldsymbol{A}) = (-1)^n \lambda_1 \lambda_2 \ldots \lambda_n \implies \det(\boldsymbol{A}) = \lambda_1 \lambda_2 \ldots \lambda_n$.

**Theorem 10.11** ***The product of the $n$ eigenvalues equals the determinant of $\boldsymbol{A}$.***
***The sum of the $n$ eigenvalues equals the sum of the $n$ diagonal entries of $\boldsymbol{A}$.***

### 10.3.3. Applications of Eigenvalues and Eigenvectors

- (Control) Stability in linear control theory
- (Optimization and learning) Gradient descent as a Krylov subspace method
- (Image processing) Image compression
- (Computer science) PageRank (PCA and SVD)
- (Statistics) Limit states of Markov chains
- (Physics) Solutions of linear PDEs
- (Physics) Cascading failure analysis

**Page Rank and Web Search**

Google is the largest web search engine in the world. When you enter a keyworld, the *PageRank* algorithm is used by Google to rank the search results of your keyworld.

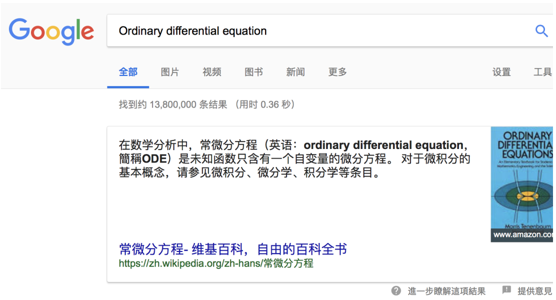



Figure 10.4: Google interface

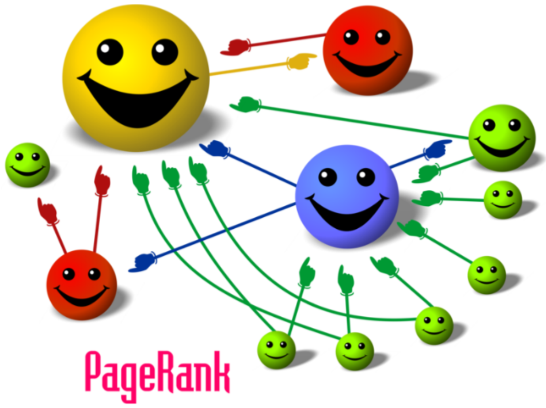



Figure 10.5: PageRank Diagram, source: Wiki

To rank the pages with respect to its importance, the idea is to use counts of links of other pages, i.e., if a page is referenced by many many other pages, it must be very important.

**PageRank Model.** The PageRank model is given as follows:

$$\sum_{j \in \mathcal{L}_i} \frac{v_j}{c_j} = v_i, \quad i = 1, \ldots, n, \tag{10.9}$$

where $c_j$ is the number of outgoing links from page $j$; $\mathcal{L}_i$ is the set of pages with a link to page $i$; $v_i$ is the importance score of page $i$. (We skip the procedure for how to construct this model)

> **Example 10.15** If we assume that there are only four pages in the world, and the diagram below shows the reference situations:

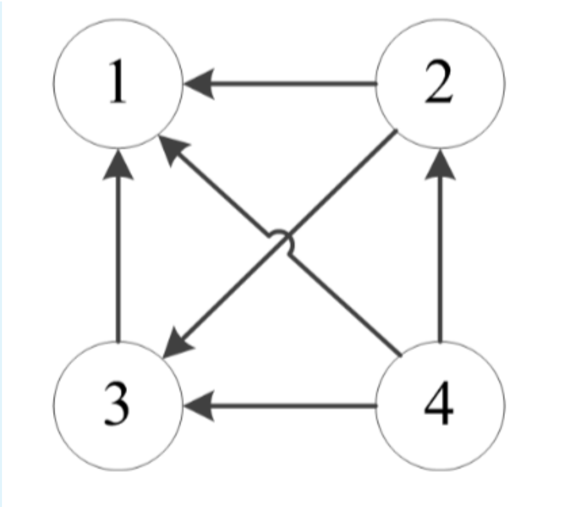



Figure 10.6: Reference situation of these four pages

Let's consider the $i = 3$ case of Eq.(10.9). The set of pages with a link to page 3 is

$$\mathcal{L}_3 := \{2,4\}$$

Next, we find that the number of outgoing links from page 2,4 are 2,3 respectively. Hence we build a equation for $i = 3$ case:

$$\frac{v_2}{2} + \frac{v_4}{3} = v_3$$

Similarly, we could use this procedure to obtain the $i = 1,2,3,4$ cases of Eq.(10.9):

$$\begin{aligned}
\frac{1}{2}v_2 + v_3 + \frac{1}{3}v_4 &= v_1 \\
\frac{1}{3}v_4 &= v_2 \\
\frac{1}{2}v_2 + \frac{1}{3}v_4 &= v_3 \\
0 &= v_4
\end{aligned}$$

Or equailently, we write the equations above into matrix form:

$$\underbrace{\begin{bmatrix} 0 & \frac{1}{2} & 1 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 \end{bmatrix}}_{\boldsymbol{A}} \underbrace{\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix}}_{\boldsymbol{v}} = \underbrace{\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix}}_{\boldsymbol{v}}$$

**PageRank Problem.** Our goal is to find the importance score $v_i$, i.e., find a **non-negative** $\boldsymbol{v}$ such that $\boldsymbol{A}\boldsymbol{v} = \boldsymbol{v}$.

In practical, $\boldsymbol{A}$ is extremely large and sparse. To solve such a eigenvalue problem, we want to use the numerical method (power method). The further reading is recommended:

> K. Bryan and L. Tanya, "The 25, 000, 000, 000 eigenvector: The linear algebra behind Google," SIAM Review, vol. 48, no. 3, pp. 569–581, 2006.

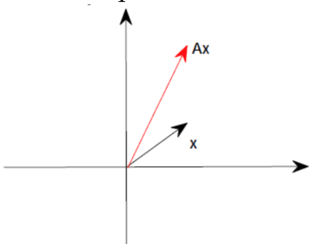
# Chapter 11

# week11

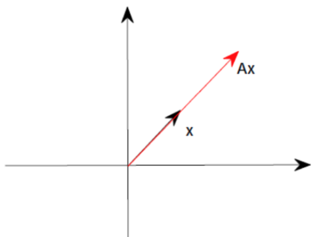
## 11.1. Lecture 22:Eigenvalues II: Diagonalization

### 11.1.1. Review

In the last lecture, we've learned the definitions of eigenvalue and eigenvectors. (definition) The geometric interpretation is the following:




$x$ is **not** an eigenvector




$x$ is an eigenvector

We also introduced the general steps to find eigenvalues and eigenvectors:

- Step1: Solve $det(A - \lambda I) = 0$ to get $n$ roots $\lambda_1, \cdots, \lambda_2$
- Step2: For each $\lambda_i$, find the eigenspace $Null(\lambda_i I - A)$
- Step3: Notice that any nonzero vector in $Null(\lambda_i I - A)$ is an eigenvector corresponding to $\lambda_i$

There is an important fact that any $n \times n$ matrix $A$ has $n$ complex eigenvalues(counting with multiplicity), since the characteristic polynomial has $n$ roots.

## 11.1.2. Motivation: Power of A

Calculate power of $A$ is important in many cases. We will illustrate that by the following examples.

### 11.1.2.1. Application 1: Fibonacci sequence

The Fibonacci sequence is defined in recursion:

$$F_0 = 0,\ F_1 = 1,\ F_{k+1} = F_k + F_{k-1}$$

The problem is : how to compute the closed-form expression? We can solve the problem by matrix-form computation:

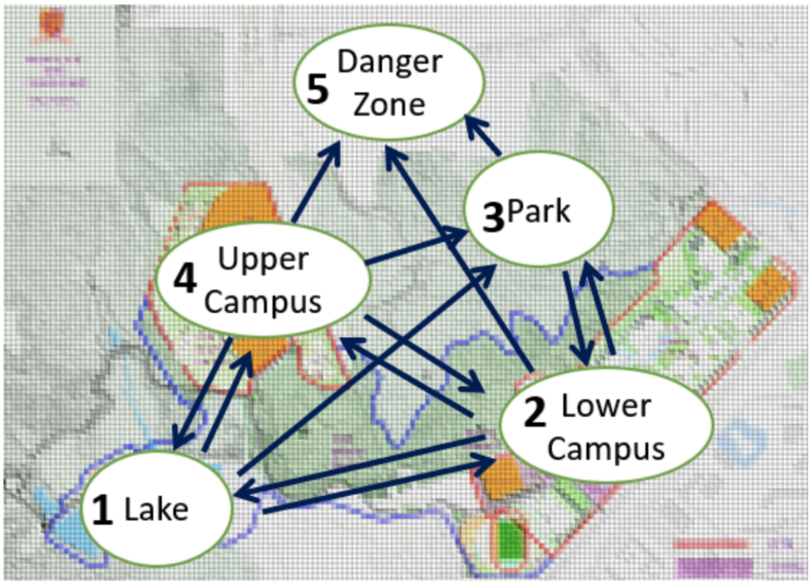
$$\begin{cases} F_{k+1} = F_k + F_{k-1} \\ F_k = F_k \end{cases} \implies \begin{bmatrix} F_k \\ F_{k+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} F_{k-1} \\ F_k \end{bmatrix}$$

$$\implies \begin{bmatrix} F_k \\ F_{k+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}^k \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

The remaining question is to compute the k-th power of the matrix .

#### 11.1.2.2. Application 2:Number of Walks with Length k

Consider the following problem:




Can we find the number of different walks in from node $i$ to node $j$ with length $k = 2022$ ? Definitely yes. If we denote the adjacency matrix to be $A$, the problem is reduced to compute $A^{2022}$.

From the two examples above, the general question is: how do we compute $A^k$ ?

### 11.1.3. Deriving Diagonalization

From the early discussion, we know that a matrix can be viewed as a linear transformation. The idea is to change the basis and find a simple representation of $A$.

- Standard basis $\{e_1, \cdots, e_n\}$

$$\begin{aligned} x &= \sum_{i=1}^{n} x_i e_i \\ f(x) &= f(\sum_{i=1}^{n} x_i e_i) \\ f(x) &= \sum_{i=1}^{n} f(e_i) x_i = Ax \end{aligned}$$

- New basis $\{v_1, \cdots, v_n\}$

$$x = \sum_{i=1}^{n} \alpha_i v_i$$
$$f(x) = f(\sum_{i=1}^{n} \alpha_i v_i)$$
$$f(x) = \sum_{i=1}^{n} f(v_i)\alpha_i = ?$$

We can write $\sum_{i=1}^{n} \alpha_i v_i$ into matrix form, which is

$$x = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}$$

Equivalently, $x = V\alpha$. Then, $\alpha = V^{-1}x$ Then we have:

$$f(x) = \sum_{i=1}^{n} f(v_i)\alpha_i = A'V^{-1}x$$

$$A' = \begin{bmatrix} \vdots & \vdots & \vdots \\ f(v_1) & \cdots & f(v_n) \\ \vdots & \vdots & \vdots \end{bmatrix}$$

The goal is to make $A'$ as simple as possible. Notice that

$$A' = \begin{bmatrix} | & & | \\ f(v_1) & \cdots & f(v_n) \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ Av_1 & \cdots & Av_n \\ | & & | \end{bmatrix}$$

If we find some $\lambda_i$ s.t. $Av_i = \lambda_i v_i$, $\lambda_i$ is the eigenvalue of $A$, then we can write

$$A' = \begin{bmatrix} \vdots & \vdots & \vdots \\ f(v_1) & \cdots & f(v_n) \\ \vdots & \vdots & \vdots \end{bmatrix} = \begin{bmatrix} \vdots & \vdots & \vdots \\ \lambda_1 v_1 & \cdots & \lambda_n v_n \\ \vdots & \vdots & \vdots \end{bmatrix} = V \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} = VD$$

$D$ represents the diagonal matrix. Combining the results:

$$f(x) = \sum_{i=1}^{n} f(e_i)x_i = Ax$$
$$f(x) = \sum_{i=1}^{n} f(v_i)\alpha_i = VDV^{-1}x$$

Hence,

$$A = VDV^{-1}$$

This is called diagonalization of $A$.

**Proposition 11.1** If an $n \times n$ matrix $A$ has $n$ linearly independent eigenvectors $v_1, \cdots, v_n$ corresponding to eigenvalues $\lambda_1, \cdots, \lambda_n$, then

$$A = VDV^{-1}$$

where $V = [v_1, \cdots, v_n] \in \mathbb{R}^{n \times n}$ is invertible and $D = diag(\lambda_1, \cdots, \lambda_n$

**Proposition 11.2** If an $n \times n$ matrix $A$ can be written as $A = VDV^{-1}$ where $D$ is a diagonal matrix, then it has $n$ linearly independent eigenvectors.

Moreover, we can define the similar matrix that represent same linear transformation by changing bases.

**Definition 11.1** [Similar matrix] If two $n \times n$ matrices $A$ and $B$ satisfy $A = S^{-1}BS$ for some invertible $S \in \mathbb{R}^{n \times n}$, the we say $A$ and $B$ are similar.

**Definition 11.2** [Diagonalizable] If a square matrix $A$ is similar to a diagonal matrix, then we say $A$ is diagonalizable

Combining previous definitions and Propositions:

**Theorem 11.1** An $n \times n$ matrix $A$ is diagonalizable $\iff$ eigenvectors of $A$ can form a basis.

The basis is called an 'eigenbasis' corresponding to $A$.

### 11.1.3.1. Revisit Applications

Let's solve the problem of Fibonacci sequence in 11.1.2.1.

**Example 11.1** Derive the closed-form expression from

$$\begin{bmatrix} F_k \\ F_{k+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}^k \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

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

- Step 1: Find eigenvalues and linearly independent eigenvectors of $A$.

  $det(A - \lambda I) = (-\lambda)(1-\lambda) - 1 = \lambda^2 - \lambda - 1$

  Eigenvalues: $\begin{cases} \lambda_1 = \frac{1+\sqrt{5}}{2} \\ \lambda_2 = \frac{1-\sqrt{5}}{2} \end{cases}$

  By solving linear systems, the linearly independent eigenvectors are:

$$v_1 = \begin{bmatrix} 1 \\ \frac{1+\sqrt{5}}{2} \end{bmatrix}, \quad v_2 = \begin{bmatrix} 1 \\ \frac{1-\sqrt{5}}{2} \end{bmatrix}$$

- Step 2: Change basis and perform diagonalization.

$$A = VDV^{-1}$$

$$D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \begin{bmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{bmatrix}$$

$$V = \begin{bmatrix} | & | \\ v_1 & v_2 \\ | & | \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} \end{bmatrix}$$

$$V^{-1} = \begin{bmatrix} \frac{1-\sqrt{5}}{2} & -1 \\ -\frac{1+\sqrt{5}}{2} & 1 \end{bmatrix} \cdot (-\frac{1}{\sqrt{5}})$$

Let $B = A^k = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}^k = VD^kV^{-1}$, plugging in $V, D, V^{-1}$:

$$
\begin{aligned}
B &= VD^kV^{-1} \\
&= \begin{bmatrix} 1 & 1 \\ \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} \end{bmatrix} \begin{bmatrix} (\frac{1+\sqrt{5}}{2})^k & 0 \\ 0 & (\frac{1-\sqrt{5}}{2})^k \end{bmatrix} \begin{bmatrix} \frac{1-\sqrt{5}}{2} & -1 \\ -\frac{1+\sqrt{5}}{2} & 1 \end{bmatrix} \cdot (-\frac{1}{\sqrt{5}}) \\
&= \begin{bmatrix} (\frac{1+\sqrt{5}}{2})^k & (\frac{1-\sqrt{5}}{2})^k \\ (\frac{1+\sqrt{5}}{2})^k & (\frac{1-\sqrt{5}}{2})^k \end{bmatrix} \begin{bmatrix} \frac{1-\sqrt{5}}{2} & -1 \\ -\frac{1+\sqrt{5}}{2} & 1 \end{bmatrix} \cdot (-\frac{1}{\sqrt{5}})
\end{aligned}
$$

- Step 3: Compute $F_k$ Note that $F_k = B_{12}$, then

$$
\begin{aligned}
F_k &= (-(\frac{1+\sqrt{5}}{2})^k) + (\frac{1-\sqrt{5}}{2})^k \cdot (-\frac{1}{\sqrt{5}} \\
&= \frac{1}{\sqrt{5}} \cdot (\frac{1+\sqrt{5}}{2})^k - (\frac{1-\sqrt{5}}{2})^k)
\end{aligned}
$$

## 11.1.4. Summary

In this lecture, we learned diagonalization of square matrix and eigenbasis. To compute $A^k$, we introduce:

- Eigenbasis
- Diagonalization $A = SDS^{-1}$
- Theorem 11.1: Diagonalizable $\iff$ eigenbasis exists
- Prop 11.2: Distinct eigenvalues $\implies$ diagonalizable

# Chapter 12

# Week12

## 12.1. Lecture 23: Spectral Decomposition

### 12.1.1. Complex Space and Inner Products

Recall the definition of real vector space, A linear space over $\mathbb{C}$ is defined in a similar way: just replace the "field" $\mathbb{R}$ by $\mathbb{C}$.

**Definition 12.1** [linear space over $\mathbb{C}$] Suppose V is a set associated with two operations:

- Addition $"+"$: $\forall v, w \in V,\ v + w \in V$
- Scalar multiplication: $\forall \alpha \in V,\ \forall v \in V,\ \alpha v \in V$

V is called a linear space over $\mathbb{C}$ if the 8 axiom axioms hold:

- (A1) $\forall v, w \in V,\ v + w \in V$
- (A2) $\forall u, v, w \in V,\ (u + v) + w = u + (v + w)$
- (A3) There exists an element $\mathbf{0}$, s.t. $v + \mathbf{0} = v$, $\forall v \in V$
- (A4) $\forall v \in V,\ \exists - v \in V$, s.t. $v + (-v) = \mathbf{0}$
- (A5) $\alpha(v + w) = \alpha v + \alpha w$, $\forall \alpha \in \mathbb{C}, v, w \in V$
- (A6) $(\alpha + \beta)v = \alpha v + \beta v$, $\forall \alpha,\ \beta \in \mathbb{C}, v \in V$
- (A7) $\alpha(\beta v) = (\alpha\beta)v$, $\forall \alpha,\ \beta \in \mathbb{C}, v \in V$
- (A8) $1v = v$, $\forall v \in V$

**Definition 12.2** [Length (norm) for complex] Given a complex-valued $n$-dimension column vector

$$z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \in \mathbb{C}^n,$$

its **length (norm)** is defined as

$$\|z\| = \sqrt{|z_1|^2 + |z_2|^2 + \cdots + |z_n|^2} = \sqrt{\langle z, z \rangle} = \sqrt{z_1\bar{z}_1 + z_2\bar{z}_2 + \cdots + z_n\bar{z}_n}.$$

Before introducing the definition of inner product for complex, let's introduce the **Hermitian transpose** for a complex-valued vector:

**Definition 12.3** [Hermitian transpose] Given $\boldsymbol{z} \in \mathbb{C}^n$, we use $\mathbf{z}^{\mathrm{H}}$ denote its **Hermitian transpose**:

$$\boldsymbol{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} \implies \mathbf{z}^{\mathrm{H}} = \bar{\mathbf{z}}^{\mathrm{T}} = \begin{bmatrix} \bar{z}_1 & \dots & \bar{z}_n \end{bmatrix}.$$

where $\bar{z}_i$ denotes the complex conjugate of $z_i$.

**Definition 12.4** [Inner product] The inner product of complex-valued vectors $\boldsymbol{z}$ and $\boldsymbol{w}$ is defined as

$$\langle \boldsymbol{z}, \boldsymbol{w} \rangle = \boldsymbol{w}^{\mathrm{H}}\boldsymbol{z} = \begin{bmatrix} \bar{\boldsymbol{w}}_1 & \dots & \bar{\boldsymbol{w}}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{z}_1 \\ \vdots \\ \boldsymbol{z}_n \end{bmatrix} = \bar{\boldsymbol{w}}_1\boldsymbol{z}_1 + \cdots + \bar{\boldsymbol{w}}_n\boldsymbol{z}_n.$$

R Note that with complex-valued vectors, $\boldsymbol{w}^{\mathrm{H}}\boldsymbol{z}$ is different from $\boldsymbol{z}^{\mathrm{H}}\boldsymbol{w}$. **The order of the vectors is now important!** In fact, $\boldsymbol{z}^{\mathrm{H}}\boldsymbol{w} = \bar{\boldsymbol{z}}_1\boldsymbol{w}_1 + \cdots + \bar{\boldsymbol{z}}_n\boldsymbol{w}_n$ is the complex conjugate of $\boldsymbol{w}^{\mathrm{H}}\boldsymbol{z}$.

**Definition 12.5** [Orthogonal] Two complex-valued vectors are *orthogonal* if their **inner product** is zero:

$$\boldsymbol{z} \perp \boldsymbol{w} \implies \langle \boldsymbol{z}, \boldsymbol{w} \rangle = \boldsymbol{w}^{\mathrm{H}} \boldsymbol{z} = 0$$

**Example 12.1** Given complex-valued vectors $\boldsymbol{z} = \begin{pmatrix} 1 \\ i \end{pmatrix}$ and $\boldsymbol{w} = \begin{pmatrix} -i \\ 1 \end{pmatrix}$, although we have $\boldsymbol{z}^{\mathrm{T}} \boldsymbol{w} = 0$, these two vectors are not perpendicular.

This is because $\langle \boldsymbol{z}, \boldsymbol{w} \rangle = \boldsymbol{w}^{\mathrm{H}} \boldsymbol{z} = \begin{bmatrix} i & 1 \end{bmatrix} \begin{bmatrix} 1 \\ i \end{bmatrix} = 2i \neq 0$.

**Example 12.2** The inner product of $\boldsymbol{u} = \begin{bmatrix} 1 \\ i \end{bmatrix}$ and $\boldsymbol{v} = \begin{bmatrix} i \\ 1 \end{bmatrix}$ is

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \begin{bmatrix} -i & 1 \end{bmatrix} \begin{bmatrix} 1 \\ i \end{bmatrix} = 0.$$

Although these vectors $(1, i)$ and $(i, 1)$ don't look perpendicular, actually they are!

**Proposition 12.1 — Conjugate symmetry.**
For two vectors $\boldsymbol{z}$ and $\boldsymbol{w} \in \mathbb{C}^n$, we have $\overline{\langle \boldsymbol{z}, \boldsymbol{w} \rangle} = \langle \boldsymbol{w}, \boldsymbol{z} \rangle$.

*Verify:*

$$\langle \boldsymbol{z}, \boldsymbol{w} \rangle = \boldsymbol{w}^{\mathrm{H}} \boldsymbol{z} = \bar{\boldsymbol{w}}^{\mathrm{T}} \boldsymbol{z} = \bar{w}_1 z_1 + \cdots + \bar{w}_n z_n$$
$$\langle \boldsymbol{w}, \boldsymbol{z} \rangle = \mathrm{z}^{\mathrm{H}} \boldsymbol{w} = \bar{\mathrm{z}}^{\mathrm{T}} \boldsymbol{w} = \bar{z}_1 w_1 + \cdots + \bar{z}_n w_n$$

Since we have $\overline{wv} = \bar{w}\bar{v}$ and $\overline{w + v} = \bar{w} + \bar{v}$, it's easy to find that

$$\overline{\bar{w}_1 z_1 + \cdots + \bar{w}_n z_n} = w_1 \bar{z}_1 + \cdots + w_n \bar{z}_n. = \bar{z}_1 w_1 + \cdots + \bar{z}_n w_n.$$

Hence $\overline{\langle \boldsymbol{z}, \boldsymbol{w} \rangle} = \langle \boldsymbol{w}, \boldsymbol{z} \rangle$. ■

**Proposition 12.2 — Sesquilinear.** For two vectors $\boldsymbol{z}$ and $\boldsymbol{w} \in \mathbb{C}^n$, we have

$$\langle \alpha \boldsymbol{z}, \boldsymbol{w} \rangle = \alpha \langle \boldsymbol{z}, \boldsymbol{w} \rangle \tag{12.1}$$

$$\langle \boldsymbol{z}, \beta \boldsymbol{w} \rangle = \bar{\beta} \langle \boldsymbol{z}, \boldsymbol{w} \rangle \tag{12.2}$$

for scalars $\alpha$ and $\beta$.

*Verify:*

$$\begin{aligned} \langle \alpha \boldsymbol{z}, \boldsymbol{w} \rangle &= \boldsymbol{w}^{\mathrm{H}}(\alpha \boldsymbol{z}) \\ &= \alpha(\boldsymbol{w}^{\mathrm{H}} \boldsymbol{z}) \\ &= \alpha \langle \boldsymbol{z}, \boldsymbol{w} \rangle. \end{aligned}$$

To show the equation (12.2), due to the conjugate symmetry, we derive

$$\langle \boldsymbol{z}, \beta \boldsymbol{w} \rangle = \overline{\langle \beta \boldsymbol{w}, \boldsymbol{z} \rangle}$$

Since $\langle \beta \boldsymbol{w}, \boldsymbol{z} \rangle = \beta \langle \boldsymbol{w}, \boldsymbol{z} \rangle = \beta \overline{\langle \boldsymbol{z}, \boldsymbol{w} \rangle}$, we obtain

$$\langle \boldsymbol{z}, \beta \boldsymbol{w} \rangle = \overline{\beta \overline{\langle \boldsymbol{z}, \boldsymbol{w} \rangle}} = \bar{\beta} \langle \boldsymbol{z}, \boldsymbol{w} \rangle.$$

■

#### 12.1.1.1. Hermitian transpose for matrix

Similarly, the **Hermitian transpose** of a complex-valued matrix $\boldsymbol{A}$ is given by

$$\boldsymbol{A}^{\mathrm{H}} := \bar{\boldsymbol{A}}^{\mathrm{T}}$$

The rules for Hermitian transpose usually comes from transpose. For example, the Hermitian transpose for matrics has the property

- $(\boldsymbol{A}\boldsymbol{B})^{\mathrm{H}} = \boldsymbol{B}^{\mathrm{H}} \boldsymbol{A}^{\mathrm{H}}$.
- $(\boldsymbol{A}^{\mathrm{H}})^{\mathrm{H}} = \boldsymbol{A}$.
- $(\boldsymbol{A} + \boldsymbol{B})^{\mathrm{H}} = \boldsymbol{A}^{\mathrm{H}} + \boldsymbol{B}^{\mathrm{H}}$.

The rules for Hermitian transpose of complex-valued vectors might be slightly different from the transpose of real-valued vectors:

| $\mathbb{R}^n$ | $\mathbb{C}^n$ |
|---|---|
| $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$ | $\langle \boldsymbol{z}, \boldsymbol{w} \rangle = \boldsymbol{w}^{\mathrm{H}}\boldsymbol{z}$ |
| $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}$ | $\boldsymbol{z}^{\mathrm{H}}\boldsymbol{w} = \overline{\boldsymbol{w}^{\mathrm{H}}\boldsymbol{z}}$ |
| $\|\boldsymbol{x}\|^2 = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$ | $\|\boldsymbol{z}\|^2 = \boldsymbol{z}^{\mathrm{H}}\boldsymbol{z}$ |
| $\boldsymbol{x} \perp \boldsymbol{y} \Longleftrightarrow \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = 0$ | $\boldsymbol{z} \perp \boldsymbol{w} \Longleftrightarrow \boldsymbol{w}^{\mathrm{H}}\boldsymbol{z} = 0$ |

R *What aspects of eigenvalues/eigenvectors are not nice?*

- Some matrix are *non-diagonalizable*. (or equivalently, eigenvectors aren't independent.)
- Eigenvalues can be *complex* even for a real-valued matrix.

We are curious about what kind of matrix has all real eigenvalues? Let's focus on real-valued matrix first. The answer is the real-valued symmetric matrix.

You should remember the proposition(12.3) below carefully, they are very important!

**Proposition 12.3** For a *real symmetric* matrix $\boldsymbol{A}$,

- All eigenvalues are real numbers.
- The eigenvectors associated with distinct eigenvalues are orthogonal.

Before the proof, let's introduce a useful formula: $\langle \boldsymbol{A}\boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{x}, \boldsymbol{A}^{\mathrm{H}}\boldsymbol{y} \rangle$.

$$\textit{Verify: } \langle \boldsymbol{A}\boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{y}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x} = (\boldsymbol{A}^{\mathrm{H}}\boldsymbol{y})^{\mathrm{H}}\boldsymbol{x} = \langle \boldsymbol{x}, \boldsymbol{A}^{\mathrm{H}}\boldsymbol{y} \rangle$$

*Proof.*
- For the first part, given any eigen-pair $(\lambda, \boldsymbol{x})$, we we obtain

$$\langle \boldsymbol{A}\boldsymbol{x}, \boldsymbol{x} \rangle = \langle \boldsymbol{x}, \boldsymbol{A}^{\mathrm{H}}\boldsymbol{x} \rangle$$

  – For the LHS, $\langle \boldsymbol{A}\boldsymbol{x}, \boldsymbol{x} \rangle = \langle \lambda\boldsymbol{x}, \boldsymbol{x} \rangle = \lambda \langle \boldsymbol{x}, \boldsymbol{x} \rangle$.

  – For the RHS, since $\boldsymbol{A}$ is a real symmetric matrix, we have

$$\boldsymbol{A}^{\mathrm{H}} = \bar{\boldsymbol{A}}^{\mathrm{T}} = \boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A} \implies \langle \boldsymbol{x}, \boldsymbol{A}^{\mathrm{H}}\boldsymbol{x} \rangle = \langle \boldsymbol{x}, \boldsymbol{A}\boldsymbol{x} \rangle$$

Moreover, $\langle \boldsymbol{x}, \boldsymbol{A}\boldsymbol{x} \rangle = \langle \boldsymbol{x}, \lambda \boldsymbol{x} \rangle = \bar{\lambda} \langle \boldsymbol{x}, \boldsymbol{x} \rangle$. Hence, $\langle \boldsymbol{x}, \boldsymbol{A}^{\mathrm{H}}\boldsymbol{x} \rangle = \bar{\lambda} \langle \boldsymbol{x}, \boldsymbol{x} \rangle$.

Finally we have $\lambda \langle \boldsymbol{x}, \boldsymbol{x} \rangle = \bar{\lambda} \langle \boldsymbol{x}, \boldsymbol{x} \rangle$. Since $\boldsymbol{x} \neq \boldsymbol{0}$, $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \neq 0$. Hence $\lambda = \bar{\lambda}$, i.e, $\lambda$ is real.

- For the second part, suppose $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are two eigenvectors corresponding to two **distinct** eigenvalues $\lambda_1$ and $\lambda_2$ respectively. Our goal is to show $\boldsymbol{x}_1 \perp \boldsymbol{x}_2$. We find that

$$\langle \boldsymbol{A}\boldsymbol{x}_1, \boldsymbol{x}_2 \rangle = \langle \boldsymbol{x}_1, \boldsymbol{A}^{\mathrm{H}}\boldsymbol{x}_2 \rangle$$

  - For LHS, $\langle \boldsymbol{A}\boldsymbol{x}_1, \boldsymbol{x}_2 \rangle = \langle \lambda_1 \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle = \lambda_1 \langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle$.
  - For RHS, $\langle \boldsymbol{x}_1, \boldsymbol{A}^{\mathrm{H}}\boldsymbol{x}_2 \rangle = \langle \boldsymbol{x}_1, \boldsymbol{A}\boldsymbol{x}_2 \rangle = \langle \boldsymbol{x}_1, \lambda_2 \boldsymbol{x}_2 \rangle = \bar{\lambda}_2 \langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle$. From part one we derive that $\langle \boldsymbol{x}_1, \boldsymbol{A}^{\mathrm{H}}\boldsymbol{x}_2 \rangle = \lambda_2 \langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle$.

  Hence $\lambda_1 \langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle = \lambda_2 \langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle$.

  Since $\lambda_1 \neq \lambda_2$, we obtain $\langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle = 0$, i.e., $\boldsymbol{x}_1 \perp \boldsymbol{x}_2$.

■

**Proposition 12.4** $\boldsymbol{A}$ is a real symmetric matrix, which also can be viewed as a linear transformation of $\mathbb{R}^n$. Suppose $W \subseteq \mathbb{R}^n$ is a $\boldsymbol{A}$ invariant space, i.e. $\boldsymbol{A}W \subseteq W$, then $W^{\perp}$ is also an invariant space.

*Proof.* For $w \in W^{\perp}$ and $\forall v \in W$, $\boldsymbol{A}v \in W$, so $\langle \boldsymbol{A}w, v \rangle = \langle w, \boldsymbol{A}v \rangle = 0$. Thus, $W^{\perp}$ is $\boldsymbol{A}$ invariant. ■

### 12.1.2. Spectral Decomposition

Suppose $A$ is a diagonalizable matrix, $A = P^{-1}DP$, then $D$ consists of eigenvalues, $P$ consists of linearly independent basis. We call this basis an "eigenbasis".

**Thought**: what if "eigenbasis" is an orthonormal basis. In this case $P$ is an orthogonal matrix, hence $A = P^{-1}DP = P^{T}DP$ is a symmetric matrix

**Theorem 12.1 — Spectral Theorem.** Any real symmetric matrix $\boldsymbol{A}$ has the factorization

$$\boldsymbol{A} = \boldsymbol{Q}\Lambda\boldsymbol{Q}^{\mathrm{T}}, \tag{12.3}$$

where $\Lambda \in \mathbb{R}^{n \times n}$ is diagonal matrix, $\boldsymbol{Q} \in \mathbb{R}^{n \times n}$ is orthogonal.

*Proof.* the only proposition we need to prove is that $\boldsymbol{A}$ is diagonalizable. This can be prove by induction on $\dim A$. ∎

R

1. Since $\boldsymbol{A} = \boldsymbol{Q}\Lambda\boldsymbol{Q}^{\mathrm{T}} = \boldsymbol{Q}\Lambda\boldsymbol{Q}^{-1}$, $\boldsymbol{A}$ could be diagonalized by an orthogonal matrix.
2. Suppose $\boldsymbol{Q} = \begin{bmatrix} q_1 & \dots & q_n \end{bmatrix}$, $\Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$, then $\boldsymbol{A}$ could be rewritten as:

$$\boldsymbol{A} = \begin{bmatrix} q_1 & \dots & q_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \begin{bmatrix} q_1^{\mathrm{T}} \\ \vdots \\ q_n^{\mathrm{T}} \end{bmatrix}$$

Or equivalently,

$$\boldsymbol{A} = \lambda_1 q_1 q_1^{\mathrm{T}} + \lambda_2 q_2 q_2^{\mathrm{T}} + \dots + \lambda_n q_n q_n^{\mathrm{T}} \tag{12.4}$$

Note that each term $q_i q_i^{\mathrm{T}}$ is the **projection matrix** for $q_i$. Hence spectral theorem says that a real symmetric matrix is a linear combination of projection matrices.

**Example 12.3** If we write $\boldsymbol{A}$ as a linear combination of projection matrices, we can have a deep understanding for the linear transformation $\boldsymbol{Ax}$:

$$\boldsymbol{A} = \sum_{j=1}^{n} \lambda_j q_j q_j^{\mathrm{T}} \implies \boldsymbol{Ax} = \sum_{j=1}^{n} \lambda_j q_j q_j^{\mathrm{T}} \boldsymbol{x} = \sum_{j=1}^{n} \lambda_j (q_j q_j^{\mathrm{T}} \boldsymbol{x}).$$

For the case $n = 2$, it's clear to find that

$$\boldsymbol{x} = c_1 q_1 + c_2 q_2 \implies \boldsymbol{Ax} = \lambda_1 c_1 q_1 + \lambda_2 c_2 q_2$$

Showing in graph, we have

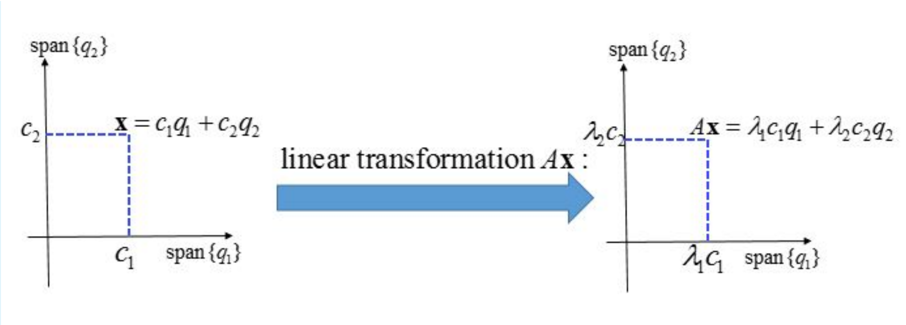



Figure 12.1: Linear transformation of $\boldsymbol{A}$.

R The formula

$$\boldsymbol{A} = \sum_{j=1}^{n} \lambda_j q_j q_j^{\mathrm{T}} \text{ or } \boldsymbol{A} = \boldsymbol{Q}\Lambda\boldsymbol{Q}^{\mathrm{T}}$$

are called the **eigen-decomposition** or **eigenvalue decomposition** of $\boldsymbol{A}$.

$\{\lambda_1, \ldots, \lambda_n\}$ are called the **spectum** of $\boldsymbol{A}$.

Also, we can extend our result from real symmetric matrix into complex-valued.

### 12.1.3. Hermitian Matrix

**Definition 12.6** [Symmetric and Hermitian]

- Recall that a square matrix $\boldsymbol{A}$ is said to be **symmetric** if $a_{ij} = a_{ji}$ for all $i, j$, or equivalently, if $\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A}$
- For complex-valued case, a square matrix $\boldsymbol{A}$ is said to be **Hermitian** if $a_{ij} = \bar{a}_{ji}$ for all $i, j$, or equivalently, if $\boldsymbol{A}^{\mathrm{H}} = \boldsymbol{A}$.

we denote the set of all $n \times n$ real symmetric matrices by $\mathbb{S}^n$; and we denote the set of all $n \times n$ complex Hermitian matrices by $\mathbb{H}^n$.

**Example:** $\boldsymbol{M} = \begin{bmatrix} 3 & 2-i \\ 2+i & 4 \end{bmatrix} \in \mathbb{H}^2$ since $\boldsymbol{M} = \boldsymbol{M}^{\mathrm{H}}$.

If $\boldsymbol{M}$ is a real matrix, then $\boldsymbol{M} = \boldsymbol{M}^{\mathrm{H}} \Longleftrightarrow \boldsymbol{M} = \boldsymbol{M}^{\mathrm{T}}$. So if the real matrix is a Hermitian matrix, it is equivalent to say it is real symmetric matrix.

Hermitian matrix has many interesting properties:

**Proposition 12.5** If $M \in \mathbb{H}^n$, then $x^{\mathrm{H}} M x \in \mathbb{R}$ for any complex-valued vectors $x$.

*Proof.* We set $\alpha := x^{\mathrm{H}} M x$. Since $\alpha$ is a scalar (easy to check), we obtain $\alpha^{\mathrm{T}} = \alpha$.

It follows that $\bar{\alpha} = \alpha^{\mathrm{H}} = (x^{\mathrm{H}} M x)^{\mathrm{H}} = x^{\mathrm{H}} M x = \alpha$. Hence $\alpha$ is real. ■

**Proposition 12.6** If $M \in \mathbb{H}^n$, then $\langle x, My \rangle = \langle Mx, y \rangle$.

*Proof.* By definition,

$$\langle x, My \rangle = (My)^{\mathrm{H}} x = y^{\mathrm{H}} M^{\mathrm{H}} x = y^{\mathrm{H}} M x = \langle Mx, y \rangle.$$

■

We have the general orthogonal matrices for complex-valued matrices:

**Definition 12.7** [Unitary] A complex-valued matrix having **orthonormal columns** is said to be unitary. In other words, $U$ is unitary if $U^{\mathrm{H}} U = I$.

The spectral theorem can also apply for Hermitian matrix:

**Theorem 12.2 — Spectral Theorem.** Any Hermitian matrix $M$ can be factorized into

$$M = U \Lambda U^{\mathrm{H}}$$

where $\Lambda$ is a real diagonal matrix, $U$ is a complex-valued unitary matrix.

R What good points does Hermitian matrix have?

- It is diagonalizable.
- Its eigenvectors form the orthogonal basis.
- Its eigenvalues are all real.

# 12.2. Lecture 24: Eigenvalues IV: Properties and Applications

## 12.2.1. Properties of Spectral Decomposition

Recall the process of spectral decomposition we learned in the last lecture, we can derive the following properties:

**Proposition 12.7** Let $A \in \mathbb{R}^{n\times n}$ and $A = A^\top$. Then all eigenvalues of $A$ are real.

It can be directly derived from Proposition 12.3.

**Proposition 12.8** Let $A \in \mathbb{R}^{n\times n}$ and $A = A^\top$. Then the eigenvectors corresponding to distinct eigenvalues of $A$ are orthogonal.

*Proof.* Suppose we have two distinct eigenvalues $\lambda_1 \neq \lambda_2$. Consider computing $X_1^H A X_2$ in two ways:

$$X_1^H X_2 = \lambda_2 X_1^H X_2$$
$$(X_1^H A)X_2 = (X_1^H A^H)X_2 = (AX_1)^H X_2 = \bar{\lambda_1} X_1^H X_2$$

Notice that $\bar{\lambda_1} = \lambda_1$ since $\lambda_1 \in \mathbb{R}$. Then we have:

$$\lambda_2 X_1^H X_2 = \lambda_1 X_1^H X_2 \iff (\lambda_2 - \lambda_1) X_1^H X_2 = 0 \quad \Rightarrow$$
$$X_1^H X_2 = 0 \iff \langle X_1, X_2 \rangle = 0 \Rightarrow X_1 \perp X_2$$

■

## 12.2.2. Further Properties of Eigenvalues

### 12.2.2.1. Sum and Product of Eigenvalues

Given a matrix, what can we say about the sum of its eigenvalues?

**Proposition 12.9 — Sum of Eigenvalues.** Let $A \in \mathbb{R}^{n\times n}$ (or $\mathbb{C}^{n\times n}$) be a square matrix. Then,

$$\sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} a_{ii}$$

The right hand side $\sum_{i=1}^{n} a_{ii}$ is called the **trace** of $A$, denoted by $tr(A)$.

The next question is: what can we say about the product of eigenvalues?

**Proposition 12.10 — Product of Eigenvalues.** Let $A \in \mathbb{R}^{n\times n}$ (or $\mathbb{C}^{n\times n}$) be a square matrix. Then,

$$\prod_{i=1}^{n} \lambda_i = det(A)$$

*Proof.* Recall that $det(A - \lambda I) = (\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n)$. Take $\lambda = 0$ we get $det(A) = \prod_{i=1}^{n} \lambda_i$ ■

**Example 12.4** Verify the propositions when $A = \begin{bmatrix} 5 & -18 \\ 1 & -1 \end{bmatrix}$.

We can compute that $det(A) = 13$ and $tr(A) = 4$.

- the characteristic polynomials $p_A(\lambda) = det(A - \lambda I) = (5 - \lambda)(-1 - \lambda) + 18 = \lambda^2 - 4\lambda + 13$.
- Let $p_A(\lambda) = 0$ and we get $\lambda_1 = 2 - 3i,\ \lambda_2 = 2 + 3i$

One can check that $det(A) = \lambda_1\lambda_2 = 4 + 9 = 13$ and $tr(A) = \lambda_1 + \lambda_2 = 4$ .

## 12.2.2.2. Eigenvalues of Similar Matrices

**Proposition 12.11 — First version; Imprecise.** Suppose $A, B \in \mathbb{R}^{n\times n}$ and are similar, i.e., there exist an invertible matrix $S$ such that $A = SBS^{-1}$, then $A$ and $B$ have the same eigenvalues.

Before we prove the proposition, we first introduce another version defined in multiset.

**Definition 12.8** [Multiset] A multiset is a modification of the concept of a set that, unlike a set, allows for multiple instances for each of it. One can use $\#\{a_1, a_2, \cdots\}$ to denote a multiset. Given a multiset, if an element appears $k$ times in the multiset, then the multiplicity of the element in the multiset is $k$

**Example 12.5** [multiset]

1. In the multiset $\#\{a,a,b\}$, the element $a$ has multiplicity 2, and $b$ has multiplicity 1.
2. In the multiset $\#\{a,a,a,b,b,b\}$, the element $a$ has multiplicity 3, and $b$ has multiplicity 3.
3. In the set $\{a,b\}$ only contains elements $a$ and $b$, each having multiplicity 1.

**Proposition 12.12 — Precise version. Eigenvalues of Similar matrices.** Suppose $A,B \in \mathbb{R}^{n\times n}$ and are similar, i.e., there exist an invertible matrix $S$ such that $A = SBS^{-1}$, then $A$ and $B$ have the same multiset of eigenvalues, denoted by $EIG(A) = EIG(B)$

*Proof.* We want to check the structure of eigenvalues. The characteristic polynomial: $p_\lambda(A) = det(A - \lambda I)$. Plugging in $A = SBS^{-1}$, we get:

$$\begin{aligned} p_\lambda(A) &= det(SBS^{-1} - \lambda I) \\ &= det(S^{-1}) \cdot det(SBS^{-1} - \lambda I) \cdot det(S) \\ &= det(B - S^{-1}(\lambda I)S) \\ &= det(B - \lambda I) \\ &= p_\lambda(B) \end{aligned}$$

■

**Example 12.6** [Judgement]

1. If $A,B \in \mathbb{R}^{n\times n}$ are similar, then $A$ and $B$ have the same eigenvalues.

   False, $A$ and $B$ have the same multiset, i.e., $EIG(A) = EIG(B)$.
2. If $A,B \in \mathbb{R}^{n\times n}$ have the same eigenvalues, then they are similar.

   False, one can find a counterexample and the converse is not true.

With the summation and product properties of eigenvalues, we can extend equivalent conditions for invertibility.

**Theorem 12.3 — Equivalent Conditions for Invertibility.** Let $A \in \mathbb{R}^{n\times n}$

The following statements are equivalent:

1. $A$ is invertible
2. The linear system $A\boldsymbol{x} = 0$ has a unique solution $\boldsymbol{x} = 0$
3. $A$ is a product of elementary matrices
4. $A$ has $n$ pivots; or equivalently: $rank(A) = n$
5. The columns of $A$ span $\mathbb{R}^n$
6. The columns of $A$ are linearly independent
7. The columns of $A$ form a basis of $\mathbb{R}^n$
8. $dim(C(A)) = n$
9. $dim(N(A)) = 0$ or $N(A) = \{\mathbf{0}\}$
10. $det(A) \neq 0$
11. The eigenvalues of $A$ satisfy $\lambda_i \neq 0$ for all $i = 1, \cdots, n$

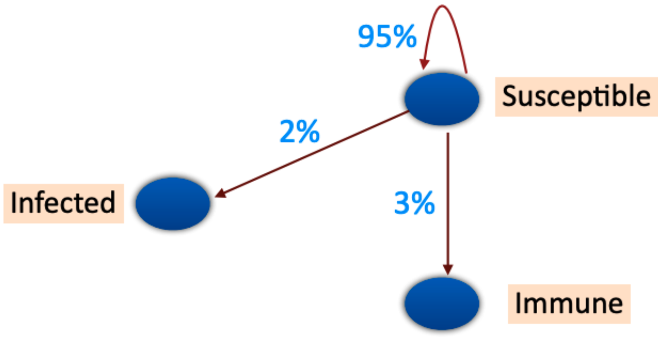
## 12.2.3. Applications

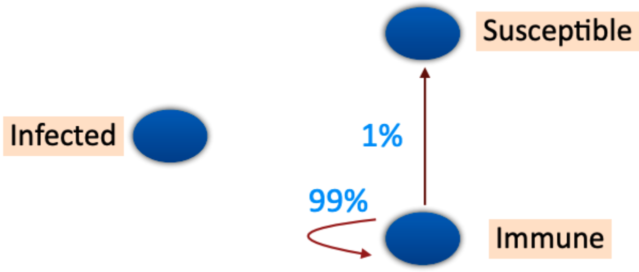
### 12.2.3.1. Epidemic Dynamics

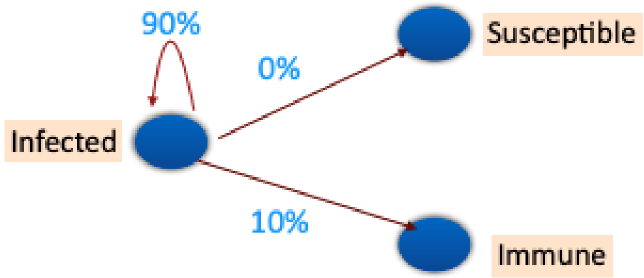
Assume a disease is introduced to population.

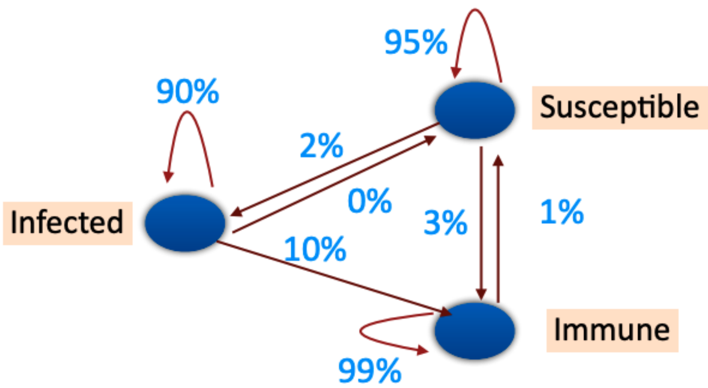
| Type of Person | Number in day $t$ |
|---|---|
| Infected | $x_t$ |
| Susceptible | $y_t$ |
| Recovered(Immune) | $z_t$ |

To be more specific, susceptible person can get infected next day and recovered people won't be infected next day. Numbers $x_t, y_t, z_t$ change day to day. There are many ways to model the dynamics, we discuss a simple one. The transition probability is showed as follows:


95%
Susceptible
2%
Infected
3%
Immune


Susceptible
Infected
1%
99%
Immune


90%
0%
Susceptible
Infected
10%
Immune


95%
90%
Susceptible
2%
0%
3%
1%
Infected
10%
Immune
99%

By assumption, we can formulate a linear system:

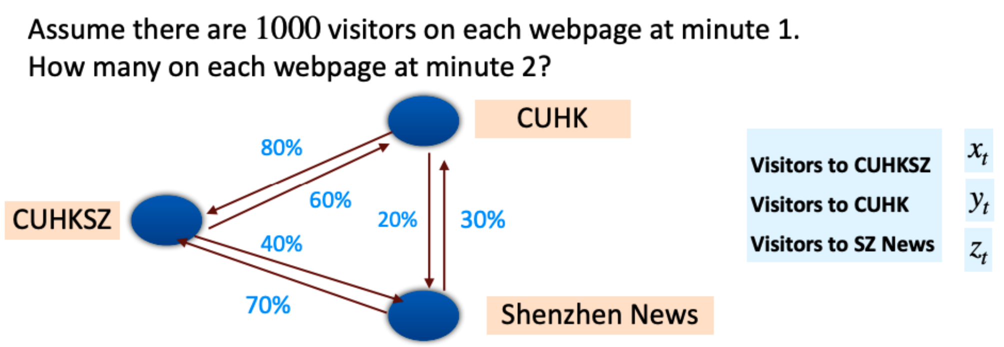
$$x_{t+1} = 0.90x_t + 0.02y_t + 0z_t$$
$$y_{t+1} = 0x_t + 0.95y_t + 0.01z_t$$
$$z_{t+1} = 0.1x_t + 0.03y_t + 0.99z_t$$

It can be formulated in matrix form with $A = \begin{bmatrix} 0.9 & 0 & 0.1 \\ 0.02 & 0.95 & 0.03 \\ 0 & 0.01 & 0.99 \end{bmatrix}$, where $a_{ij}$ denotes the transition probability from different groups of people. Let $\boldsymbol{u}_{t+1} = \begin{bmatrix} x_{t+1} \\ y_{t+1} \\ z_{t+1} \end{bmatrix}$ Note that $\boldsymbol{u}_{t+1} = A^\top \boldsymbol{u}_t$, we can calculate $\boldsymbol{u}_t$ by diagonalizing $A^\top$. In real cases, $A^\top$ is likely to be diagonalizable. Let $A^\top = SDS^{-1}$, then $\boldsymbol{u}_t = SD^tS^{-1}\boldsymbol{u}_0$

## 12.2.3.2. PageRank(Google's 1st Algorithm)

When we search the internet, google always displays many webpages as response. Here comes one of the key IT problems: how to pick good webpages?

In fact, there are well-developed algorithms to rank webpages. We now introduce the brief ideas. Consider the following example:




We can solve the problem by formulating a linear system.

**Formulation**

Assume we have $n$ webpages in total, each webpage may have links to other webpages. At time $t$, there are $v_1(t), v_2(t), \cdots, v_n(t)$ visitors at page $1, 2, \cdots, n$

**Assumptions:**

If there are $m$ links of page at webpage $j$, then a visitor at webpage $j$ will click one of the $m$ pages randomly. Denote the probability of a visitor at webpage $j$ to visit webpage $k$ by $a_{j,k}$. Let $A = (a_{jk})_{1 \leq j,k \leq n}$

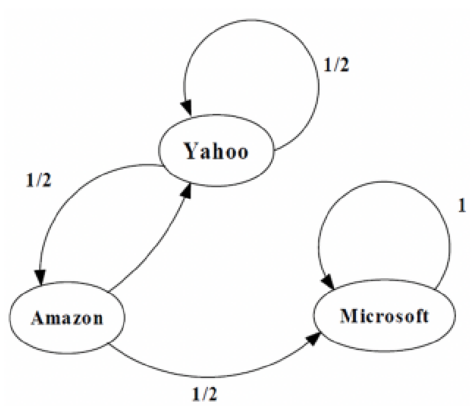
**Mathematical Model:**

$$\begin{bmatrix} v_1(t+1) \\ v_2(t+1) \\ \vdots \\ v_n(t+1) \end{bmatrix} = A \begin{bmatrix} v_1(t) \\ v_2(t) \\ \vdots \\ v_n(t) \end{bmatrix}$$

**Algorithm:**

Compute $\boldsymbol{v}(\infty) := \lim_{t \to \infty} \boldsymbol{v}(t)$, and rank webpages by entries of $\boldsymbol{v}(\infty)$. By Perron's theorem, the $\boldsymbol{v}(\infty)$ exists (not cover in our lecture).

Consider the following example with three webpages:




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

$$\begin{bmatrix} \text{yahoo} \\ \text{Amazon} \\ \text{Microsoft} \end{bmatrix} = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}$$

We can compute the vector of visitors at each webpage:

- Time 1: $\begin{bmatrix} \frac{1}{3} \\ \frac{1}{6} \\ \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix}$

- Time 2: $\begin{bmatrix} \frac{1}{4} \\ \frac{1}{6} \\ \frac{7}{12} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{3} \\ \frac{1}{6} \\ \frac{1}{2} \end{bmatrix}$

  ...

  We can repeat the process to get $\boldsymbol{v}(t)$:

$$\boldsymbol{v}_1 = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{bmatrix} \quad \boldsymbol{v}_2 = \begin{bmatrix} \frac{1}{4} \\ \frac{1}{6} \\ \frac{7}{12} \end{bmatrix} \quad \boldsymbol{v}_3 = \begin{bmatrix} \frac{5}{24} \\ \frac{1}{8} \\ \frac{2}{3} \end{bmatrix} \quad \boldsymbol{v}_4 = \begin{bmatrix} \frac{1}{6} \\ \frac{5}{48} \\ \frac{35}{48} \end{bmatrix} \quad \cdots$$

Observe that $\boldsymbol{v}(\infty) = M\boldsymbol{v}(\infty)$, which means $\boldsymbol{v}(\infty)$ is an eigenvector of $M$. We can easily compute the eigenvector to get the answer.

# Index