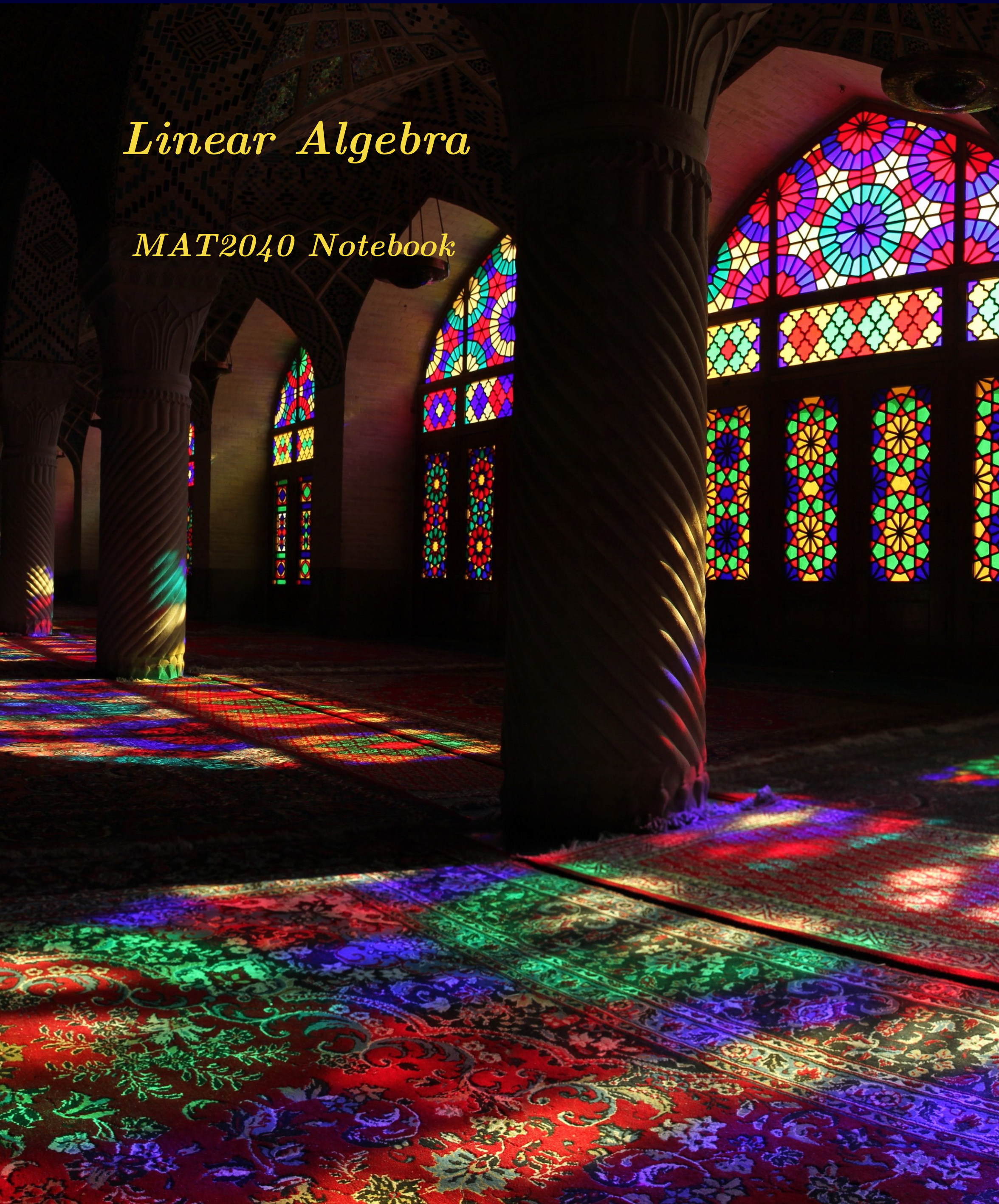
香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

Linear Algebra

MAT2040 Notebook



The Second Edition

A FIRST COURSE

IN

LINEAR ALGEBRA

A FIRST COURSE

IN

LINEAR ALGEBRA

MAT2040 Notebook

Lecturer

Prof. Tom Luo

The Chinese University of Hongkong, Shenzhen

Lecturer

Prof. Ruoyu Sun

University of Illinois Urbana-Champaign

Tex Written By

Mr. Jie Wang

The Chinese University of Hongkong, Shenzhen

香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

Contents

Preface xi

Acknowledgments xiii

Notations xv

1 Week1 1

1.1 Tuesday **1**

1.1.1 Introduction 1

1.1.2 Gaussian Elimination 3

1.1.3 Complexity Analysis 11

1.1.4 Brief Summary 12

1.2 Thursday **14**

1.2.1 Row-Echelon Form 14

1.2.2 Matrix Multiplication 16

1.2.3 Special Matrices 19

1.3 Friday **21**

1.3.1 Matrix Multiplication 21

1.3.2 Elementary Matrix 22

1.3.3 Properties of Matrix 24

1.3.4 Permutation Matrix 26

1.3.5 LU decomposition 29

1.3.6 LDU decomposition 33

1.3.7 LU Decomposition with row exchanges 35

1.4 Assignment One **36**

2 Week2 39

2.1 Tuesday 39

2.1.1 Review 39

2.1.2 Special matrix multiplication case 41

2.1.3 Inverse 44

2.2 Wednesday 49

2.2.1 Remarks on Gaussian Elimination 49

2.2.2 Properties of matrix 50

2.2.3 matrix transpose 53

2.3 Assignment Two 56

2.4 Friday 57

2.4.1 symmetric matrix 57

2.4.2 Interaction of inverse and transpose 58

2.4.3 Vector Space 59

2.5 Assignment Three 68

3 Week3 71

3.1 Tuesday 71

3.1.1 Introduction 71

3.1.2 Review of 2 weeks 72

3.1.3 Examples of solving equations 73

3.1.4 How to solve a general rectangular 78

3.2 Thursday 83

3.2.1 Review 83

3.2.2 Remarks on solving linear system equations 86

3.2.3 Linear dependence 88

3.2.4 Basis and dimension 90

3.3 Friday 96

3.3.1 Review 96

3.3.2 More on basis and dimension . . . 97
3.3.3 What is rank? . . . 99
3.4 Assignment Four 107
4 Midterm . . . 109
4.1 Sample Exam 109
4.2 Midterm Exam 116
5 Week4 . . . 123
5.1 Friday 123
5.1.1 Linear Transformation . . . 123
5.1.2 Example: differentiation . . . 125
5.1.3 Basis Change . . . 130
5.1.4 Determinant . . . 132
5.2 Assignment Five 138
6 Week5 . . . 141
6.1 Tuesday 141
6.1.1 Formulas for Determinant . . . 141
6.1.2 Determinant by Cofactors . . . 146
6.1.3 Determinant Applications . . . 147
6.1.4 Orthogonality . . . 150
6.2 Thursday 154
6.2.1 Orthogonality . . . 154
6.2.2 Least Squares Approximations . . . 160
6.2.3 Projections . . . 163
6.3 Friday 166
6.3.1 Orthonormal basis . . . 166
6.3.2 Gram-Schmidt Process . . . 170

6.3.3 The Factorization $A = QR$. 175
6.3.4 Function Space 177
6.3.5 Fourier Series 179
6.4 Assignment Six 180
7 Week6 181
7.1 Tuesday 181
7.1.1 Summary of previous weeks 181
7.1.2 Eigenvalues and eigenvectors 185
7.1.3 Products and Sums of Eigenvalue 189
7.1.4 Application: Page Rank and Web Search 190
7.2 Thursday 193
7.2.1 Review 193
7.2.2 Similarity 193
7.2.3 Diagonalization 195
7.2.4 Powers of A 200
7.2.5 Nondiagonalizable Matrices 201
7.3 Friday 205
7.3.1 Review 205
7.3.2 Fibonacci Numbers 205
7.3.3 Imaginary Eigenvalues 207
7.3.4 Complex Numbers and vectors 208
7.3.5 Spectral Theorem 214
7.3.6 Hermitian matrix 216
7.4 Assignment Seven 219
8 Week7 223
8.1 Tuesday 223
8.1.1 Quadratic form 223
8.1.2 Convex Optimization Preliminaries 225

8.1.3 Positive Definite Matrices 228

8.2 Thursday 237

8.2.1 SVD: Singular Value Decomposition 237

8.2.2 Remark on SVD decomposition 241

8.2.3 Best Low-Rank Approximation 246

8.3 Assignment Eight 250

9 Final Exam 253

9.1 Sample Exam 253

9.2 Final Exam 260

10 Solution 267

10.1 Assignment Solutions 267

10.1.1 Solution to Assignment One 267

10.1.2 Solution to Assignment Two 273

10.1.3 Solution to Assignment Three 277

10.1.4 Solution to Assignment Four 283

10.1.5 Solution to Assignment Five 294

10.1.6 Solution to Assignment Six 300

10.1.7 Solution to Assignment Seven 308

10.1.8 Solution to Assignment Eight 318

10.2 Midterm Exam Solutions 325

10.2.1 Sample Exam Solution 325

10.2.2 Midterm Exam Solution 335

10.3 Final Exam Solutions 343

10.3.1 Sample Exam Solution 343

10.3.2 Final Exam Solution 354

Preface

This book is intended for the foundation course MAT2040, which is the first course on the linear algebra. It aims to cover basic linear algebra knowledge and its simple applications. This book was first written in 2017, and it is reviewed and revised in 2018. We have corrected several mistakes shown in the previous book and modified some proofs a little bit to give readers better insights of linear algebra. During the modification, we also refer to many reading materials, which are also recommended for you:

- ENGG 5781 Course Notes by Prof. Wing-Kin (Ken) Ma, CUHK, Hongkong, China, http://www.ee.cuhk.edu.hk/~wkma/engg5781
- Roger A. Horn and Charles R. Johnson, Matrix Analysis (Second Edition), Cambridge University Press, 2012.
- S. Boyd and L. Vandenberghe, Introduction to Applied Linear Algebra (Vectors, Matrices, and Least Squares), Cambridge University Press, 2018.

The whole book can cover a semester course in a 14week, each section in which corresponds to a 2-hour lecture. If you read the whole book, and work some mini-exercises, you will learn a lot. We hope you will get the insights on linear algebra and apply them in your own subject.

CUHK(SZ)
October 27, 2018

Acknowledgments

This book is from the MAT2040 in summer semester, 2017. It is revised in 2018 to 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	ith entry of column vector $\boldsymbol{x}$
a_{ij}	(i, j)th entry of matrix $\boldsymbol{A}$
$\boldsymbol{a}_i$	ith column of matrix $\boldsymbol{A}$
$\boldsymbol{a}_i^{\mathrm{T}}$	ith 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 ith 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. Tuesday

1.1.1. Introduction

1.1.1.1. *Why do you learn Linear Algebra?*

Important: LA + Calculus + Probability. 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.)

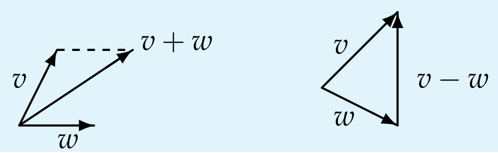
Visualize. Conncect between **Geometry** and **Algebra**.

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 describle 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.

1.1.1.2. What is Linear Algebra?

The central problem in math is to **solve equations**. And equations can be seperated into two parts, **nonlinear** and **linear** ones.

Let's look an example of Nonlinear equations below:

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

Well, it is a little bit complicated. We don't find a efficient algorithm to solve these equations. But in algebraic geometry course we will solve some nonlinear equations.

What you need to know about in this course is the linear equations and the methodology to solve it.

Definition 1.1 [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 $a_1, a_2, \dots, a_n, b$ are real numbers and $x_1, x_2, \dots, x_n$ are variables ■

Definition 1.2 [Linear System of Equations] Linear system of m equations in n unknowns is the system of the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{23}x_n &= b_2 \\ \cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{m3}x_n &= b_m, \end{aligned} \tag{1.1}$$

where a_{ij} and the b_i are all real numbers. We refer to (**??**)as $m \times n$ **linear systems**. ■

1.1.2. Gaussian Elimination

Here we mainly focus on $n \times n$ system of equations.

Example 1.2 Let's recall how to solve a 2×2 system equatons as below:

$$1x_1 + 2x_2 = 5 \tag{1.2}$$

$$4x_1 + 5x_2 = 14. \tag{1.3}$$

We can simplify the equation system above into the form (**Augmented matrix**):

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

Secondly, by adding $(-4) \times (1.2)$ into (1.3), we obtain:

$$1x_1 + 2x_2 = 5 \tag{1.4}$$

$$0x_1 + (-3)x_2 = -6 \tag{1.5}$$

Thirdly, by multiplying $-(1/3)$ of (1.5), we obtain:

$$1x_1 + 2x_2 = 5 \tag{1.6}$$

$$1x_2 = 2 \tag{1.7}$$

Fourthly, by adding $(-2) \times (1.7)$ into (1.6), we obtain:

$$1x_1 + 0x_2 = 1 \tag{1.8}$$

$$1x_2 = 2 \tag{1.9}$$

Here we get the solution $(x_1 = 1, x_2 = 2)$, and we could write the above process with augmented matrix form:

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

■

The method shown above is called **Gaussian Elimination**. Here we give a strict definition for Augmented matrix:

Definition 1.3 [Augmented matrix] For the system of equations

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1$$
$$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2$$
$$\cdots$$
$$a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m, \tag{1.10}$$

the corresponding augmented matrix is given by

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

■

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

Definition 1.4 [pivot] Returning to the example, we find after third step the matrix is given by

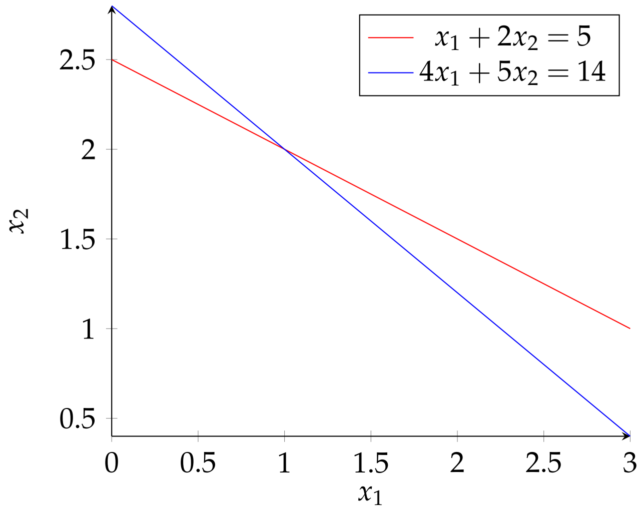
$$\left[\begin{array}{cc|c} 1 & 2 & 5 \\ 0 & 1 & 2 \end{array}\right].$$

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 the example case, the element in the second column of the second row is the pivot. ■

1.1.2.1. 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.

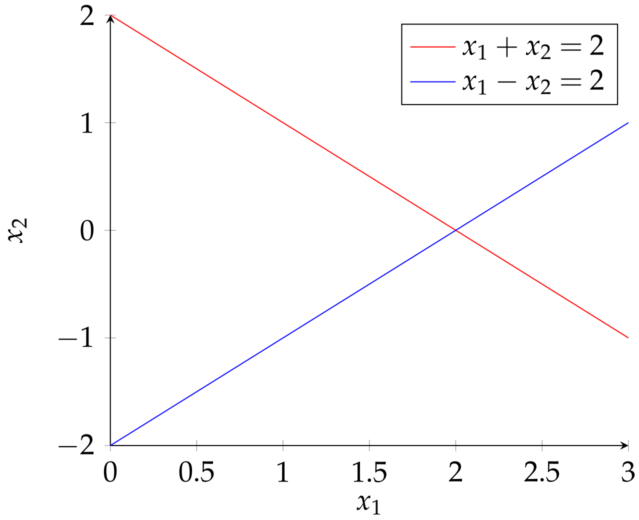


Column Picture. Focusing on the column of the system of equation, we can denote $\begin{bmatrix} 1 \\ 4 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 5 \end{bmatrix}$ as vectors in coordinate axis. Could the linear combinations of these two vectors form the vector $\begin{bmatrix} 5 \\ 14 \end{bmatrix}$?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

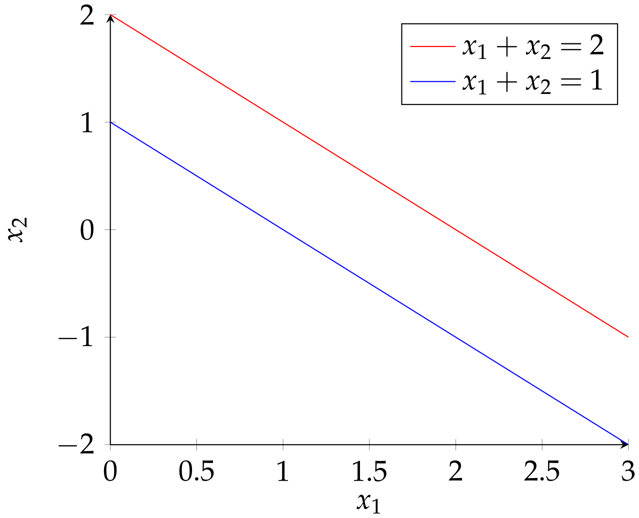
1.1.2.2. The solutions of the Linear System of Equations

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

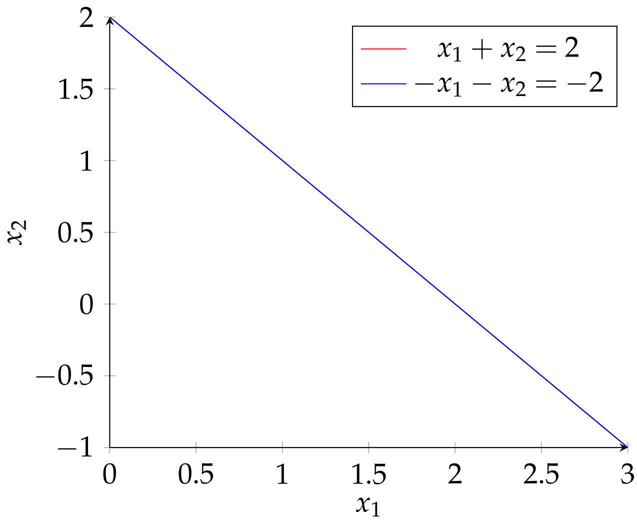
Case 1: unique solution. If two lines intersect at one point, then there is unique solution.



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.



1.1.2.3. How to solve 3×3 Systems?

■ **Example 1.3**

Let's recall how to solve a 3×3 system equations as below:

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

We can simplify the equation system above into the **Augmented matrix** form:

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

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

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

This augmented matrix is the **strictly triangular system**, and it's trial to get the final solution:

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

Here we give the definition for strictly triangular system:

Definition 1.5 [strictly triangular system] For the augmented matrix

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

if in the kth row, the first $(k-1)$th column entries are *all zero* and the kth column entries is nonzero, we say the augmented matrix(or corresponding system equation) is of **strictly triangular form**. This kind of matrix(or corresponding system equation) is called **strictly triangular system**. $(k = 1, ..., m)$.

1.1.2.4. How to solve $n \times n$ System?

We try to reduce an $n \times n$ System to strictly triangular form. Let's take a special example:

■ **Example 1.4** Given an $n \times n$ System of the form:

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & b_n \end{array}\right] \tag{1.11}$$

Assuming 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*:

$$\Longrightarrow \left[\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ 0 & \times & \dots & \times & \times \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \times & \cdots & \times & \times \end{array}\right] \tag{1.12}$$

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

$$\left[\begin{array}{ccccc|c} \times & \times & \times & \times & \times & \times \\ & \times & \times & \times & \times & \times \\ & & \times & \times & \times & \vdots \\ & 0 & & \times & \times & \vdots \\ & & & & \times & \times \end{array}\right] \tag{1.13}$$

This matrix is the **Row-echelon form**. And we do the back substitution again to obtain:

$$\left[\begin{array}{ccccc|c} \times & & & & & \times \\ & \times & & & 0 & \times \\ & & \times & & & \vdots \\ & 0 & & \times & & \vdots \\ & & & & \times & \times \end{array}\right] \tag{1.14}$$

This matrix is the **Reduced Row-Echelon Form**. Finally by multiplying every row by a nonzero constant to ensure its **diagnoal entries** are all 1:

$$\left[\begin{array}{ccccc|c} 1 & & & & & \times \\ & 1 & & 0 & & \times \\ & & \ddots & & & \vdots \\ & 0 & & 1 & & \vdots \\ & & & & 1 & \times \end{array}\right] \tag{1.15}$$

Then let's analysis the complexity of solving such a $n \times n$ system.

1.1.3. Complexity Analysis

1.1.3.1. Step1: Reduction from matrix (1.11) to matrix (1.12)

Proposition 1.1 The time complexity for Augmented matrix reduction using back-substitution algorithm is $\mathcal{O}(n^3)$.

Proof. The estimation for the time complexity requires us to estimate how many steps of **multiplication** we need. (The time for addition is so small that can be ignored).

- Reducing matrix (1.11) to matrix (1.12) we need to do $n(n-1)$times multiplications.

  This is because for each row (except first row) we have known the first entry is zero, while the remaining $(n-1)$ entries in each row should be computed by multiplying first row's entries and then add it to the row.

- Then it suffices to deal with the inner $(n-1) \times (n-1)$ matrix, which requires the $(n-1) \times (n-2)$ times multiplication.

- The back substitution for matrix (1.11) requires n times reduction.

Hence the total multiplication times for back substitution for matrix (1.11) is

$$\begin{aligned}\sum_{i=1}^{n} i(i-1) &= \sum_{i=1}^{n}(i^2-i)\\ &= \sum_{i=1}^{n} i^2 - \sum_{i=1}^{n} i\\ &= \frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)}{2}\\ &= \frac{n^3-2n}{3} \sim \frac{n^3}{3} = O(n^3)\end{aligned}$$

■

But we can always develop more advanced algorithm that have smaller time complexity.

1.1.3.2. Step2: Reduction from triangular system to diagonal system

In order to reducing matrix (1.13) to matrix (1.14) we need to do back-substitution again. The matrix (1.14) is diagonal system. Obviously, for this process the total multiplication times is given by

$$1+2+\cdots+n-1 = \frac{n(n-1)}{2} \sim O(n^2)$$

1.1.3.3. Step3: Get final solution

In the final step, we want to reduce matrix (1.14) to matrix (1.15), the only thing we need to do is to do one multiplication for each row to let the diagonal entries be 1. Hence the total multiplication times for this process is given by

$$\underbrace{1+1+\cdots+1}_{\text{totally } n \text{ terms}} = O(n)$$

1.1.4. Brief Summary

The reduction of $n \times n$ matrix requires three kinds of Row operations:

- **Addition and Multiplication**.

  Add to a row by a constant multiple of another row.

- **Multiplication**

  Multiply a row by a nonzero constant.

- **Interchange**

  Interchange two rows

1. agds

-

1.2. Thursday

1.2.1. Row-Echelon Form

1.2.1.1. Gaussian Elimination does't always work

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

■ **Example 1.5** We apply Gaussian Elimination to try to transfrom 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:

$$\xrightarrow[\text{Add } (-1)\times \text{ row 3 to row 5}]{\text{Add } (-1)\times \text{ row 3 to row 1; Add } (-1)\times \text{ row 3 to row 4}} \left[\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & 0 & -2 \\ 0 & 0 & 1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 & -4 \\ 0 & 0 & 0 & 0 & 0 & -3 \end{array}\right] \tag{1.16}$$

  Note that the matrix (1.16) 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:

$$\xrightarrow[\text{Add } 2\times \text{ row 3 to row 1; Add } (-2)\times \text{ row 3 to row 2}]{\text{Add } (-1)\times \text{ row 2 to row 1}} \left[\begin{array}{ccccc|c} 1 & 1 & 0 & 0 & 0 & 4 \\ 0 & 0 & 1 & 1 & 0 & -6 \\ 0 & 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 & -4 \\ 0 & 0 & 0 & 0 & 0 & -3 \end{array}\right] \tag{1.17}$$

The matrix (1.17) 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 1.6 [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.

Definition 1.7 [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 matrix $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ is also of *Row Echelon Form*! Moreover, it is of *Reduced Row Echelon Form*.

1.2.2. Matrix Multiplication

1.2.2.1. Matrix Multiplied by Vector

Here we introduce the definition for inner product of vector:

Definition 1.8 [inner product] Given two vectors $x = (x_1, x_2, \dots, x_n)$ and $y = (y_1, y_2, \dots, y_n)$, the inner product between x and y is given by

$$\langle x, y \rangle = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$$

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

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

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

■ **Example 1.6** 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

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad \boldsymbol{A} = \begin{pmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{pmatrix} = \begin{pmatrix} a_1^{\mathrm{T}} \\ a_2^{\mathrm{T}} \\ a_3^{\mathrm{T}} \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} 5 \\ -2 \\ 9 \end{pmatrix}.$$

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:

$$\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x} = \begin{pmatrix} a_1^{\mathrm{T}} \\ a_2^{\mathrm{T}} \\ a_3^{\mathrm{T}} \end{pmatrix} \boldsymbol{x} = \begin{pmatrix} a_1^{\mathrm{T}}\boldsymbol{x} \\ a_2^{\mathrm{T}}\boldsymbol{x} \\ a_3^{\mathrm{T}}\boldsymbol{x} \end{pmatrix}$$

■

1.2.2.2. Matrix Multiply Matrix

R Note that an $m \times n$ matrix $\boldsymbol{A}$ can be written as $\left[a_{ij}\right]$, where a_{ij} denotes the entry of ith row, jth column of $\boldsymbol{A}$.

Notice that 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:

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

The result is a $m \times k$ matrix. Thus for matrix multiplication, it suffices to calculate matrix multiplied by vectors.

Example 1.7 We want to calculate the result for $m \times n$ matrix $\boldsymbol{A}$ multiply $n \times k$ matrix $\boldsymbol{B}$, which is written as

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

Hence the ith row, jth column of $\boldsymbol{C}$ is given by

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

You should understand this result, this means the ith row, jth column entry of $\boldsymbol{C}$ is given by the ith row of $\boldsymbol{A}$ multiplying the jth column of $\boldsymbol{B}$. ■

R **Time Complexity Analysis**

- 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.

1.2.3. Special Matrices

Here we introduce several special matrices:

Definition 1.9 [Identity Matrix] The $n \times n$ identity matrix is the matrix $\boldsymbol{I} = [m_{ij}]$, where

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

Proposition 1.2 Identity Matrix has the following properties:

$$\boldsymbol{IB} = \boldsymbol{B}, \quad \boldsymbol{AI} = \boldsymbol{A},$$

where $\boldsymbol{A}$ and $\boldsymbol{B}$ coud be any size-suitable matrix.

Definition 1.10 [Elementary Matrix of type III] An elementary matrix $\boldsymbol{E}_{ij}$ of type III is a matrix such that

- its diagonal entries are all 1
- the ith row j th column is a scalar
- the remaining entries are all zero.

For example, the matrix $\boldsymbol{A} = \begin{pmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{pmatrix}$ is elementary matrix of type III.t

R If $\boldsymbol{A}$ is a matrix, then postmultiplying with $\boldsymbol{E}_{ij}$ has the same effect of performing row operation on $\boldsymbol{A}$.

For example, given an elementary matrix of type III and a matrix $\boldsymbol{A}$:

$$\boldsymbol{E}_{21} = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad A = \begin{pmatrix} 2 & 1 & 1 \\ 4 & -6 & 0 \\ -2 & 7 & 2 \end{pmatrix}$$

Then the effect of $\boldsymbol{EA}$ has the same effect of adding $(-2)\times$ row 1 to row 2:

$$\boldsymbol{E}_{21}A = \begin{pmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ -2 & 7 & 2 \end{pmatrix}$$

Moreover, if we define $\boldsymbol{E} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$, then continuing postmultiplying $\boldsymbol{E}_{31}$ is just like doing Gaussian Elimination:

$$\boldsymbol{E}_{31}\boldsymbol{E}_{21}A = \begin{pmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 8 & 3 \end{pmatrix}$$

1.3. Friday

1.3.1. Matrix Multiplication

1.3.1.1. 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 1.8** Given 4×3 matrix $\boldsymbol{A}$ and 3×2 matrix $\boldsymbol{B}$, then the result of $\boldsymbol{AB}$ should be a 4×2 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 ith row of $\boldsymbol{A}$ and the jth column of $\boldsymbol{B}$.

  SInce the result has 4×2 entries, we have to process such progress 4×2 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

B, and the coefficients are entries of the first row of matrix 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:

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

■

1.3.2. Elementary Matrix

So let's review the concept for elementary matrix by an example:

R In this course you can think there is **only one** type of elementary matrix. This may contradict what you see in the textbook.

Definition 1.11 [Elementary Matrix] An elementary matrix $\boldsymbol{E}_{ij}$ is a matrix that its *diagonal entries* are all 1 and the (i,j)th column is a *scalar*, and the remaining entries are all *zero*. ■

For example, the matrix $\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 4 & 0 & 1 \end{pmatrix}$ is elementary matrix.

■ **Example 1.9** Given vector $\boldsymbol{b} = \begin{bmatrix} b_1 & b_2 & b_3 \end{bmatrix}^{\mathrm{T}}$ and elementary matrix $\boldsymbol{E}_{31} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -l_{31} & 0 & 1 \end{bmatrix}$, the effct of *postmultiplying* $\boldsymbol{E}_{31}$ for $\boldsymbol{b}$ has the same effect of doing row operation:

$$\boldsymbol{E}_{31}\boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 - l_{31}b_1 \end{bmatrix}$$

Let's do more practice. Given matrix $\boldsymbol{E}_{21} = \begin{bmatrix} 1 & 0 & 0 \\ -l_{21} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, we can calculate the result of $\boldsymbol{E}_{21} \times (\boldsymbol{E}_{31}\boldsymbol{b})$ and $\boldsymbol{E}_{21}\boldsymbol{E}_{31}$:

$$\boldsymbol{E}_{21} \times (\boldsymbol{E}_{31}\boldsymbol{b}) = \begin{bmatrix} 1 & 0 & 0 \\ -l_{21} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} b_1 \\ b_2 \\ b_3 - l_{31}b_1 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 - l_{21}b_1 \\ b_3 - l_{31}b_1 \end{bmatrix}$$

$$\boldsymbol{E}_{21}\boldsymbol{E}_{31} = \begin{bmatrix} 1 & 0 & 0 \\ -l_{21} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -l_{31} & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -l_{21} & 1 & 0 \\ -l_{31} & 0 & 1 \end{bmatrix}$$

Additionally, we can use matrix multiplication to derive the result of $(\boldsymbol{E}_{21}\boldsymbol{E}_{31}) \times \boldsymbol{b}$:

$$(\boldsymbol{E}_{21}\boldsymbol{E}_{31}) \times \boldsymbol{b} = \begin{bmatrix} 1 & 0 & 0 \\ -l_{21} & 1 & 0 \\ -l_{31} & 0 & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 - l_{21}b_1 \\ b_3 - l_{31}b_1 \end{bmatrix}$$

Amazingly, we find that the result of $\boldsymbol{E}_{21} \times (\boldsymbol{E}_{31}\boldsymbol{b})$ is actually the same as $(\boldsymbol{E}_{21}\boldsymbol{E}_{31}) \times \boldsymbol{b}$, which is one of the properties of matrix. ∎

1.3.3. Properties of Matrix

Operations on matrix has the following properties:

1. $\boldsymbol{A}(\boldsymbol{B}+\boldsymbol{C}) = \boldsymbol{A}\boldsymbol{B} + \boldsymbol{A}\boldsymbol{C}$.
2. $\boldsymbol{A}\boldsymbol{B} \neq \boldsymbol{B}\boldsymbol{A}$, i.e., $\boldsymbol{A}\boldsymbol{B}$ doesn't *necessarily* equal to $\boldsymbol{B}\boldsymbol{A}$.

   (R) In some special cases, $\boldsymbol{A}\boldsymbol{B}$ may equal to $\boldsymbol{B}\boldsymbol{A}$. 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:

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

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

3. **Block Multiplication**. We use an example to show the process of block multiplicaion:

■ **Example 1.10** 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 2×2 **block matrices**. As a result, $\boldsymbol{C}$ have 2×2 blocks:

$$\boldsymbol{A}\boldsymbol{B} = \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{A}_2 \\ \boldsymbol{A}_3 & \boldsymbol{A}_4 \end{bmatrix} \begin{bmatrix} \boldsymbol{B}_1 & \boldsymbol{B}_2 \\ \boldsymbol{B}_3 & \boldsymbol{B}_4 \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_1\boldsymbol{B}_1 + \boldsymbol{A}_2\boldsymbol{B}_3 & \boldsymbol{A}_1\boldsymbol{B}_2 + \boldsymbol{A}_2\boldsymbol{B}_4 \\ \boldsymbol{A}_3\boldsymbol{B}_1 + \boldsymbol{A}_4\boldsymbol{B}_3 & \boldsymbol{A}_3\boldsymbol{B}_2 + \boldsymbol{A}_4\boldsymbol{B}_4 \end{bmatrix} = \begin{bmatrix} \boldsymbol{C}_1 & \boldsymbol{C}_2 \\ \boldsymbol{C}_3 & \boldsymbol{C}_4 \end{bmatrix}$$

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

$$\boldsymbol{C}_1 = \boldsymbol{A}_1\boldsymbol{B}_1 + \boldsymbol{A}_2\boldsymbol{B}_3 = \begin{bmatrix} 4 & 0 \\ 6 & 6 \end{bmatrix} \begin{bmatrix} 8 & -3 \\ 3 & -7 \end{bmatrix} + \begin{bmatrix} 4 \\ -8 \end{bmatrix} \begin{bmatrix} 4 & -4 \end{bmatrix} = \begin{bmatrix} 48 & -28 \\ 34 & -28 \end{bmatrix}.$$

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

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

■

There are also two useful ways to compute $\boldsymbol{A}\boldsymbol{B}$:

- If $\boldsymbol{B}$ has k columns, we can partition $\boldsymbol{B}$ into k blocks to compute $\boldsymbol{A}\boldsymbol{B}$:

$$\boldsymbol{A}\boldsymbol{B} = \boldsymbol{A} \times \left[\begin{array}{c|c|c|c} B_1 & B_2 & \dots & B_k \end{array}\right] = \left[\begin{array}{c|c|c|c} \boldsymbol{A}B_1 & \boldsymbol{A}B_2 & \dots & \boldsymbol{A}B_k \end{array}\right].$$

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

$$\boldsymbol{AB} = \begin{bmatrix} \boldsymbol{A}_1 \\ \hline \boldsymbol{A}_2 \\ \hline \dots \\ \hline \boldsymbol{A}_m \end{bmatrix} \times \boldsymbol{B} = \begin{bmatrix} \boldsymbol{A}_1\boldsymbol{B} \\ \hline \boldsymbol{A}_2\boldsymbol{B} \\ \hline \dots \\ \hline \boldsymbol{A}_m\boldsymbol{B} \end{bmatrix}$$

1.3.4. Permutation Matrix

Note that there also exists one kind of matrix $\boldsymbol{P}$ such that postmultiplying $\boldsymbol{P}$ for arbitararily matrix $\boldsymbol{A}$ has the same effect of interchanging two rows of $\boldsymbol{A}$.

For example, if $\boldsymbol{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ and $\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, then by postmultiplying $\boldsymbol{P}$ for $\boldsymbol{A}$ we obtain:

$$\boldsymbol{PA} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 3 & 4 \\ 1 & 2 \end{bmatrix}.$$

This progress has the same effect of interchanging the first row and the second row of $\boldsymbol{A}$.

This kind of matrix is called **permutaion matrix**:

Definition 1.12 [Permutation Matrix] $\boldsymbol{P}$ is a **permutation matrix** if postmultiplying $\boldsymbol{P}$ for matrix $\boldsymbol{A}$ has the same effect of interchanging rows of matrix $\boldsymbol{A}$. ■

Definition 1.13 [Row Exchange Matrix] $\boldsymbol{P}$ is a **row exchange matrix** if postmultiplying $\boldsymbol{P}$ for matrix $\boldsymbol{A}$ has the same effect of interchanging only two rows of matrix $\boldsymbol{A}$.

We use the notation $\boldsymbol{P}_{ij}$ to denote a matrix that has the effect of exchanging row i and row j of $\boldsymbol{A}$. ■

The way to obtain $\boldsymbol{P}_{ij}$ is simple. After an identity matrix's ith and jth row being exchanged, we could obtain the row exchange matrix $\boldsymbol{P}_{ij}$.

Let's raise some examples to show what is row exchange matrix:

■ **Example 1.11** P_{23} has the effect of exchanging 2th row and 3th row of arbitarary matrix. It is converted from an identity matrix:

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{\text{Interchange row 2 and 3}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = P_{23}.$$

$$I = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{\text{Interchange row 2 and 3}} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = P_{23}.$$

Postmultiplying by P_{23} exchanges row 2 and row 3 of any matrix:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 6 & 7 \\ 15 & 4 \\ 24 & 3 \end{bmatrix} = \begin{bmatrix} 6 & 7 \\ 24 & 3 \\ 15 & 4 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 6 \\ 24 \\ 15 \\ 4 \end{bmatrix} = \begin{bmatrix} 6 \\ 15 \\ 24 \\ 4 \end{bmatrix}$$

■

R You may be confused about the concept between *permutation matrix* and *row exchange matrix*. The row exchange matrix is a special case of permutation matrix, but permutation matrix could exchange several rows. For example, row 1,2,3,4 could be changed into row 4,3,2,1.

Before talking about the properties of permutation matrix, let's introduce the definition for nonsingular and inverse matrix:

Definition 1.14 [Nonsingular matrix] Let A be an $n \times n$ matrix, the following statements are equivalent:

1. A is **nonsingular** or **invertible**.
2. There exists a matrix B such that $AB = BA = I$. And the matrix B is said to be

the **inverse** of A, and we can write $B = A^{-1}$.

3. After multiplying finite numbers of **elementary matrix**, A can be converted to identity matrix I.
4. The system of equations $Ax = b$ has a unique solution.

If matrix A is **not** nonsingular, this matrix is called **singular**. ■

We are interested in the inverse of permutation matrix.

Proposition 1.3 1. For a permutation matrix P, it can always be decomposed into finite multiplications of row exchange matrices P_{ij}:

$$P = P_{i_1 j_1} P_{i_2 j_2} \dots P_{i_n j_n}$$

2. The inverse of a row exchange matrix is actually equal to itself:

$$P_{ij} P_{ij} = I \Longleftrightarrow P_{ij}^{-1} = P_{ij}$$

3. For a permutation matrix written as $P = P_{i_1 j_1} P_{i_2 j_2} \dots P_{i_n j_n}$, its inverse matrix is given by:

$$P^{-1} = P_{i_n j_n}^{-1} P_{i_{n-1} j_{n-1}}^{-1} \dots P_{i_1 j_1}^{-1} = P_{i_n j_n} P_{i_{n-1} j_{n-1}} \dots P_{i_1 j_1}$$

4. For a $n \times n$ permutation matrix P and a $n \times n$ matrix A given by:

$$P = \left[\begin{array}{c|ccc} 1 & 0 & 0 & 0 \\ \hline 0 & & & \\ \vdots & & P_{(n-1)\times(n-1)} & \\ 0 & & & \end{array}\right] \qquad A = \left[\begin{array}{c|ccc} a_{11} & a_{12} & \cdots & a_{1n} \\ \hline 0 & & & \\ \vdots & & A_{(n-1)\times(n-1)} & \\ 0 & & & \end{array}\right]$$

the multiplication result PA has the form:

$$PA = \left[\begin{array}{c|ccc} a_{11} & a_{12} & \cdots & a_{1n} \\ \hline 0 & & & \\ \vdots & & P_{(n-1)\times(n-1)} A_{(n-1)\times(n-1)} & \\ 0 & & & \end{array}\right]$$

Proofoutline.

- For proposition 2, it is because that if we exchange two rows of any matrix $\boldsymbol{A}$, and then we exchange the same rows again, the effect is cancelled out!
- For proposition 3, it is because that we just need to do the reverse order of our process in order to obtain the inverse matrix.

■

1.3.5. 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 1.4

The inverse of an elementary matrix is also an elementary matrix.

■ **Example 1.12** $E_{21} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ is an elementary matrix, the result of postmultiplying E_{21} for identity matrix is given by:

$$E_{21}I = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

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

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

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

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

■ **Example 1.13** Let's try Gaussian Elimination for a matrix that is nonsingular. Here we use elementary matrix to describle 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:

$$E_{32}E_{31}E_{21}A = U$$
$$\implies E_{32}^{-1}E_{32}E_{31}E_{21}A = E_{32}^{-1}U \implies E_{31}E_{21}A = E_{32}^{-1}U$$
$$\cdots \implies A = E_{21}^{-1}E_{31}^{-1}E_{32}^{-1}U := LU,$$

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. ■

Actually, any nonsingular matrix *without row exchanges, i.e., does not require the row exchange during the Gaussian Elimination*, could be decomposed as the multiplication of a lower triangular matrix with a upper triangular matrix U, which is called **LU decomposition**.

1.3.5.1. 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 A and b in **one** equation $Ax = b$.

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

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

Ⓡ

The detail of Solve process.

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

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

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

Forward and Backward Solve $\boldsymbol{Ly} = \boldsymbol{b}$ and then solve $\boldsymbol{Ux} = \boldsymbol{y}$.

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

■ **Example 1.14** Forward elimination on $\boldsymbol{Ax} = \boldsymbol{b}$ will result in equation $\boldsymbol{Ux} = \boldsymbol{y}$:

$$\boldsymbol{Ax} = \boldsymbol{b} \Longleftrightarrow \begin{cases} u + 2v = 5 \\ 4u + 9v = 21 \end{cases} \text{forward elimination implies} \begin{cases} u + 2v = 5 \\ v = 1 \end{cases} \Longleftrightarrow \boldsymbol{Ux} = \boldsymbol{y}.$$

We could express such process into matrix form:

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

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

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

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

$\boldsymbol{Ux} = \boldsymbol{y}$. 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.

1.3.6. 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 $\boldsymbol{LU}$, where the upper triangular matrix $\boldsymbol{U}$ is given by:

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

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:

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

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

$$\boldsymbol{A} = \boldsymbol{L}\boldsymbol{U} = \boldsymbol{L}\boldsymbol{D}\boldsymbol{D}^{-1}\boldsymbol{U} = \boldsymbol{L}\boldsymbol{D}(\boldsymbol{D}^{-1}\boldsymbol{U}) = \boldsymbol{L}\boldsymbol{D}\hat{\boldsymbol{U}},$$

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

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

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

Note that the *diagonal* entries of $\boldsymbol{D}$ are all **pivots values** of $\boldsymbol{U}$.

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

Definition 1.15 [LDU Decomposition] In conclusion, we decompose matrix $\boldsymbol{A}$ into the form:

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

where: $\boldsymbol{L}$ is lower triangular matrix with unit entries in diagonal

$\boldsymbol{D}$ is diagonal matrix

$\boldsymbol{U}$ 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 1.5 **LDU decomposition is unique to any matrix**. Let $\boldsymbol{L}, \boldsymbol{L}_1$ denote a lower triangular matrix, $\boldsymbol{D}, \boldsymbol{D}_1$ diagonal, and $\boldsymbol{U}, \boldsymbol{U}_1$ upper triangular.

If $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{D}\boldsymbol{U}$, and also, $\boldsymbol{A} = \boldsymbol{L}_1\boldsymbol{D}_1\boldsymbol{U}_1$, then we have $\boldsymbol{L} = \boldsymbol{L}_1, \boldsymbol{D} = \boldsymbol{D}_1, \boldsymbol{U} = \boldsymbol{U}_1$.

1.3.7. LU Decomposition with row exchanges

How can we handle row exchange in our $\boldsymbol{LU}$ decomposition?

Assume we are going to do Gaussian Elimination with matrix $\boldsymbol{A}$ with row exchange.

- At first We can postmultiply some elementary matrices $\boldsymbol{E}$ to get $\boldsymbol{EEEA}$.
- Sometimes we need to multiply by $\boldsymbol{P}_{ij}$ to do *row exchange* to continue Gaussian Elimination.
- So we may end our elimination with something like $\boldsymbol{PEEEEPEEEPEEEEA}$.
- If we can get all the elementary matrix $\boldsymbol{L}$ together, we could convert them into one single $\boldsymbol{L}$ that has the same effect as before.
- The key problem is that how can we get all the row exchange matrix $\boldsymbol{P}$ out from the elementary matrices?

Theorem 1.1 If $\boldsymbol{A}$ is *nonsingular*, then there exists a permutation matrix $\boldsymbol{P}$ such that $\boldsymbol{PA} = \boldsymbol{LU}$.

The proof is omitted.

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

1.4. Assignment One

1. Consider the system

$$ax + 2y + 3z = b_1$$
$$ax + ay + 4z = b_2$$
$$ax + ay + az = b_3$$

   For what three values of a will the *elimination* fail to give the pivots? (*Pivots means the first nonzero entry on rows.*)

2. It is impossible for a system of linear equations to have *exactly* two solutions? Explain your answers. And you may consider the following questions as intuitions to derive your final solution.

   (a) In $\mathbb{R}^3$ if (x, y, z) and (X, Y, Z) are two solutions, what is another one?

   (b) In $\mathbb{R}^3$ if 25 planes meet at two points, where else do they meet?

   (c) Extend the argument to $\mathbb{R}^n$.

3. In the following system

$$x + 4y - 2z = 1$$
$$x + 7y - 6z = 6$$
$$3y + qz = t$$

   (a) Which number q makes this system *singular*? Moreover, if this system is *singular*, which right-hand side t gives *infinitely* many solutions?

   (b) Find the solution that has $z = 1$.

4. By **trial** and **error**, find examples of 2×2 matrices such that:

   (a) $\boldsymbol{A}^2 = -\boldsymbol{I}$, where $\boldsymbol{A}$ has *real* entries.

(b) $\boldsymbol{B}^2 = \mathbf{0}$, where $\boldsymbol{B} \neq \mathbf{0}$.

(c) $\boldsymbol{CD} = -\boldsymbol{DC}$, where $\boldsymbol{C}D \neq \mathbf{0}$.

(d) $\boldsymbol{EF} = \mathbf{0}$, and no entries of $\boldsymbol{E}$ or $\boldsymbol{F}$ are zero.

5. For *real* matrices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ in *finite* field, prove the *associativity product rule:*

$$(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC}).$$

6. Matrices can be cut into blocks (which are smaller matrices). Here is a 4 by 6 matrix broken into blocks of size 2 by 2, in this example each block is just $\boldsymbol{I}$:

4 by 6 matrix

2 by 2 blocks

$$\boldsymbol{A} = \left[\begin{array}{cc|cc|cc} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ \hline 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{array}\right] = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{I} & \boldsymbol{I} \\ \boldsymbol{I} & \boldsymbol{I} & \boldsymbol{I} \end{bmatrix}.$$

We give the definition for *block multiplication*:

Definition 1.16 [Block Multiplication] If the cuts between columns of $\boldsymbol{A}$ match the cuts between rows of $\boldsymbol{B}$, then block multiplication of $\boldsymbol{AB}$ is allowed:

$$\begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} & \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{B}_{11} & \cdots \\ \boldsymbol{B}_{21} & \cdots \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_{11}\boldsymbol{B}_{11} + \boldsymbol{A}_{12}\boldsymbol{B}_{21} & \cdots \\ \boldsymbol{A}_{21}\boldsymbol{B}_{11} + \boldsymbol{A}_{22}\boldsymbol{B}_{21} & \cdots \end{bmatrix}.$$

If we have $\boldsymbol{A}, \boldsymbol{B}$ such that

$$\boldsymbol{AB} = \left[\begin{array}{cc|c} \times & \times & \times \\ \times & \times & \times \\ \hline \times & \times & \times \end{array}\right] \left[\begin{array}{cc|c} \times & \times & \times \\ \times & \times & \times \\ \hline \times & \times & \times \end{array}\right],$$

replace $\times$ by numbers to verify the block multiplication succeeds.

7.

$$\begin{array}{cc|c} 4 & 0 & 4 \\ 6 & 6 & 8 \\ \hline -9 & 5 & -8 \end{array} \boldsymbol{A} = \begin{bmatrix} a & a & a & a \\ a & b & b & b \\ a & b & c & c \\ a & b & c & d \end{bmatrix}.$$

Seperate $\boldsymbol{A}$ into $\boldsymbol{L}$ and $\boldsymbol{U}$. Moreover, Find four conditions on a, b, c, d to let $\boldsymbol{A}$ have *four* pivots.

Chapter 2

Week2

2.1. Tuesday

2.1.1. Review

2.1.1.1. Solving a system of linear Equations

Gaussian Elimination. For the system of equations $\boldsymbol{Ax} = \boldsymbol{b}$, it has three cases for its solutions:

$$\boldsymbol{Ax} = \boldsymbol{b} \begin{cases} \text{unique solution} \\ \text{no solution} \\ \text{infinitely many solutions} \end{cases}$$

We claim that

> if for this system of equation it has **infinitely** many solutions, then *its columns(or rows) could be linearly combined to zero nontrivially.*

Let's raise an example to explain this statement. Let's use an augmented matrix to represent $\boldsymbol{Ax} = \boldsymbol{b}$ (Assume $\boldsymbol{A}$ is a 3×3 matrix):

$$\boldsymbol{Ax} = \boldsymbol{b} \Longleftrightarrow \left[\begin{array}{ccc|c} a_{11} & a_{12} & a_{13} & b_1 \\ a_{21} & a_{22} & a_{23} & b_2 \\ a_{31} & a_{32} & a_{33} & b_3 \end{array}\right]$$

When focusing on the columns, we may have the question: in which case does its columns could be linearly combined to zero? That means we need to choose the coefficients c_1, c_2, c_3 such that

$$c_1 \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix} + c_2 \begin{pmatrix} a_{12} \\ a_{22} \\ a_{32} \end{pmatrix} + c_3 \begin{pmatrix} a_{13} \\ a_{23} \\ a_{33} \end{pmatrix} = 0$$

- It's obvious that when $c_1 = c_2 = c_3 = 0$ we can linearly combine the columns. So $c_1 = c_2 = c_3 = 0$ is the *trival* solution.

- But is there any **nontrival** solution? We claim that if this system of equation has *infinitely* many solutions, we could linearly combine the columns *nontrivally*. We will prove this statement in the end of this lecutre.

  If we focus on the rows, we may have the similar question and conclusion.

Matrix to describe Gaussian Elimination.

1. Firstly let's consider the nonsingular matrix $\boldsymbol{A}$ without row exchange case. We find that postmultiplying elementary matrix has the same effect as doing gaussian elimination. If we finally convert $\boldsymbol{A}$ into *upper triangular matrix* $\boldsymbol{U}$, we can write this process in matrix notation:

$$\boldsymbol{E}_n \dots \boldsymbol{E}_1 \boldsymbol{A} = \boldsymbol{U} \implies \boldsymbol{A} = (\boldsymbol{E}_n \dots \boldsymbol{E}_1)^{-1} \boldsymbol{U} \implies \boldsymbol{A} = \boldsymbol{E}_1^{-1} \dots \boldsymbol{E}_n^{-1} \boldsymbol{U}$$

   (a) If we define $\boldsymbol{L} := \boldsymbol{E}_1^{-1} \dots \boldsymbol{E}_n^{-1}$, which is a lower triangular matrix, then we finally decompose $\boldsymbol{A}$ into the product of two triangular matrix:

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

   (b) We can fuirther decompose $\boldsymbol{A}$ into product of three matrices to make the diagonal entries of $\boldsymbol{U}$ and $\boldsymbol{L}$ to be **one**:

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

Recall that the LDU decomposition is unique for any matrix.

2. If we have to do row exchange, the process for converting $\boldsymbol{A}$ into $\boldsymbol{U}$ may be like the form:

$$\boldsymbol{E}\cdots\boldsymbol{EPE}\cdots\boldsymbol{EPE}\cdots\boldsymbol{EA} = \boldsymbol{U},$$

but we can always do row exchange first to combine all elementary matrix together, which means we can convert this process into:

$$\boldsymbol{E}\cdots\boldsymbol{EPA} = \boldsymbol{U} \implies \boldsymbol{PA} = \boldsymbol{LU}$$

Also, we can do LDU decomposition to get $\boldsymbol{PA} = \boldsymbol{LDU}$.

2.1.2. Special matrix multiplication case

Firstly let's introduce a new type of vector named unit vector:

Definition 2.1 [unit vector] An ith unit vector is given by:

$$e_i = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Only in ith row its entry is 1, other entries of e_i are all 0. ■

Then let's discuss some interesting matrix multiplication cases:

1. (a) Given $m \times n$ matrix $\boldsymbol{A} = \begin{bmatrix} a_{ij} \end{bmatrix}_{m\times n}$, the product $\boldsymbol{A}e_i$ is given by:

$$\boldsymbol{A}e_i = \begin{bmatrix} a_{:i} \end{bmatrix},$$

where $\begin{bmatrix} a_{:i} \end{bmatrix}$ denotes the ith column of $\boldsymbol{A}$. (It is from the MATLAB or Julia language.)

(b) Also, given a row vector $e_j^{\mathrm{T}} := \begin{bmatrix} 0 & 0 & \dots & 1 & \dots & 0 \end{bmatrix}$, the product $e_j^{\mathrm{T}}\boldsymbol{A}$ is given by:

$$e_j^{\mathrm{T}}\boldsymbol{A} = \begin{bmatrix} a_{j:} \end{bmatrix},$$

where $\begin{bmatrix} a_{j:} \end{bmatrix}$ denotes the jth row of $\boldsymbol{A}$.

2. Secondly, we want to compute the product $\mathbf{1}^{\mathrm{T}}\boldsymbol{A}\mathbf{1}$, where $\mathbf{1}$ denotes a column vector that all entries of $\mathbf{1}$ are 1 and $\mathbf{1}^{\mathrm{T}}$ denotes the corresponding row vector. Let's first compute $\boldsymbol{A} \times \mathbf{1}$, where $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{1} \in \mathbb{R}^n$:

$$\boldsymbol{A} \times \mathbf{1} = \begin{pmatrix} \sum_{j=1}^n a_{1j} \\ \sum_{j=1}^n a_{2j} \\ \vdots \\ \sum_{j=1}^n a_{mj} \end{pmatrix}$$

It follows that

$$\mathbf{1}^{\mathrm{T}}\boldsymbol{A}\mathbf{1} = \mathbf{1}^{\mathrm{T}}(\boldsymbol{A}\mathbf{1}) = \mathbf{1}^{\mathrm{T}} \begin{pmatrix} \sum_{j=1}^n a_{1j} \\ \sum_{j=1}^n a_{2j} \\ \vdots \\ \sum_{j=1}^n a_{mj} \end{pmatrix} = \langle \mathbf{1}, \boldsymbol{A}\mathbf{1} \rangle = \sum_{i=1}^m \sum_{j=1}^n a_{ij},$$

3. For vectors $x \in \mathbb{R}^m$, $y \in \mathbb{R}^n$, we can compute $x^{\mathrm{T}}\boldsymbol{A}y$:

$$x^{\mathrm{T}}\boldsymbol{A}y = x^{\mathrm{T}} \begin{pmatrix} \sum_{j=1}^n a_{1j}y_j \\ \sum_{j=1}^n a_{2j}y_j \\ \vdots \\ \sum_{j=1}^n a_{mj}y_j \end{pmatrix} = \sum_{i=1}^m x_i(\sum_{i=1}^n a_{ij}y_j) = \sum_{i,j} a_{ij}x_iy_j$$

4. For vectors $x \in \mathbb{R}^n$, $y \in \mathbb{R}^n$, you should distinguish $x^{\mathrm{T}}y$ and xy^{T}:

$$x^{\mathrm{T}}y = \langle x, y \rangle = \sum_{i=1}^{n} x_i y_i$$

$$xy^{\mathrm{T}} = \begin{bmatrix} x_1y_1 & x_1y_2 & \dots & x_1y_n \\ x_2y_1 & x_2y_2 & \dots & x_2y_n \\ \vdots & & \vdots & \\ x_ny_1 & x_ny_2 & \dots & x_ny_n \end{bmatrix} = \begin{bmatrix} x_iy_j \end{bmatrix}_{n \times n}$$

5. For vectors $x \in \mathbb{R}^m$, $y \in \mathbb{R}^n$, we can compute $x^{\mathrm{T}}\boldsymbol{A}y$ by using block matrix:

   Firstly, We partition $\boldsymbol{A}$ into four parts:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix}_{(m_1+m_2)\times(n_1+n_2)}.$$

   Then we partition vector x and y respectively:

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}_{m_1+m_2}, \quad y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}_{n_1+n_2},$$

   where x_1 has m_1 rows, x_2 has m_2 rows, y_1 has n_1 rows, y_2 has n_2 rows.

   Then we can compute $x^{\mathrm{T}}\boldsymbol{A}y$:

$$x^{\mathrm{T}}\boldsymbol{A}y = \begin{bmatrix} x_1^{\mathrm{T}} & x_2^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \sum_{i=1}^{2}\sum_{j=1}^{2} x_i^{\mathrm{T}} \boldsymbol{A}_{ij} y_j.$$

6.

   Proposition 2.1 Postmultiplying $\boldsymbol{Q}$ for the vector $v = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ has the same effect of rotating v in the plane *anticlockwise* by the angle θ, where

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

Proof. We convert vector v into the form $v = \begin{bmatrix} \rho cos\varphi \\ \rho sin\varphi \end{bmatrix}$, where $\rho = \sqrt{x_1^2 + x_2^2}$, and $\varphi = arctan(\frac{x_2}{x_1})$. Hence we obtain the product of Q and v:

$$Qv = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix} \begin{bmatrix} \rho cos\varphi \\ \rho sin\varphi \end{bmatrix} = \begin{bmatrix} \rho cos\theta cos\varphi - \rho sin\theta sin\varphi \\ \rho cos\theta sin\varphi + \rho sin\theta cos\varphi \end{bmatrix} = \begin{bmatrix} \rho cos(\theta + \varphi) \\ \rho sin(\theta + \varphi) \end{bmatrix}$$

This is the form that this vector has been rotated anticlockwise by the angle θ. ■

7. Given $m \times n$ matrix $A = \begin{bmatrix} a_{ij} \end{bmatrix}$, how to flip this matrix vertically? We just need to postmultiply a special matrix:

$$\begin{bmatrix} 0 & & & 1 \\ & & 1 & \\ & \iddots & & \\ 1 & & & 0 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} = \begin{bmatrix} a_{m1} & a_{m2} & \dots & a_{mn} \\ a_{(m-1)1} & a_{(m-1)2} & \dots & a_{(m-1)n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{11} & a_{12} & \dots & a_{1n} \end{bmatrix}$$

If we aftermultiply this matrix for the matrix A, we can flip A horizontally:

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} 0 & & & 1 \\ & & 1 & \\ & \iddots & & \\ 1 & & & 0 \end{bmatrix} = \begin{bmatrix} a_{1n} & a_{1(n-1)} & \dots & a_{11} \\ a_{2n} & a_{2(n-1)} & \dots & a_{21} \\ \vdots & \vdots & \ddots & \vdots \\ a_{mn} & a_{m(n-1)} & \dots & a_{m1} \end{bmatrix}$$

2.1.3. Inverse

Let's introduce the definition for inverse matrix:

> **Definition 2.2** [Inverse matrix] For $n \times n$ matrix A, the matrix B is said to be the **inverse** of A if we have $AB = BA = I$. If such B exists, we say matrix A is **invertible** or **nonsingular**. ■

And inverse matrix has some interesting properties:

Proposition 2.2 **Matrix inverse is Unique**. In other words, if we have $AB_1 = B_1 A = I$

and $\boldsymbol{A}\boldsymbol{B}_2 = \boldsymbol{B}_2\boldsymbol{A} = \boldsymbol{I}$, then we obtain $\boldsymbol{B}_1 = \boldsymbol{B}_2$.

Proof.

$$\boldsymbol{A}\boldsymbol{B}_1 = \boldsymbol{I} \implies \boldsymbol{B}_2\boldsymbol{A}\boldsymbol{B}_1 = \boldsymbol{B}_2\boldsymbol{I} \implies \boldsymbol{B}_2\boldsymbol{A}\boldsymbol{B}_1 = \boldsymbol{B}_2$$

$$\implies (\boldsymbol{B}_2\boldsymbol{A})\boldsymbol{B}_1 = \boldsymbol{I}\boldsymbol{B}_1 = \boldsymbol{B}_1 = \boldsymbol{B}_2.$$

■

Proposition 2.3 If we have both $\boldsymbol{A}\boldsymbol{B} = \boldsymbol{I}$ and $\boldsymbol{C}\boldsymbol{A} = \boldsymbol{I}$, then we have $\boldsymbol{C} = \boldsymbol{B}$.

Proof. On the one hand, we have

$$\boldsymbol{C}\boldsymbol{A}\boldsymbol{B} = \boldsymbol{C}(\boldsymbol{A}\boldsymbol{B}) = \boldsymbol{C}\boldsymbol{I} = \boldsymbol{C}$$

On the other hand, we obtain:

$$\boldsymbol{C}\boldsymbol{A}\boldsymbol{B} = (\boldsymbol{C}\boldsymbol{A})\boldsymbol{B} = \boldsymbol{I}\boldsymbol{B} = \boldsymbol{B}$$

Hence we have $\boldsymbol{C} = \boldsymbol{B}$. ■

2.1.3.1. How to compute inverse? When does it exist?

Assuming the inverse of $n \times n$ matrix $\boldsymbol{A}$ exists, and we define it to be

$$\boldsymbol{A}^{-1} := \mathbf{X} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} = \begin{bmatrix} x_{ij} \end{bmatrix}$$

By definition, we have $\boldsymbol{A}\boldsymbol{X} = \boldsymbol{I}$. We write it into block columns:

$$\boldsymbol{A}\boldsymbol{X} = \boldsymbol{A}\begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} = \boldsymbol{I} = \begin{bmatrix} e_1 & e_2 & \dots & e_n \end{bmatrix},$$

where $e_1, e_2, \dots, e_n$ are all unit vectors.

Hence we obtain

$$\boldsymbol{A}\begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}x_1 & \boldsymbol{A}x_2 & \dots & \boldsymbol{A}x_n \end{bmatrix} = \begin{bmatrix} e_1 & e_2 & \dots & e_n \end{bmatrix}.$$

Thus we only need to compute n system of equations $Ax_i = e_i, i = 1, \dots, n$ to get the columns of the inverse matrix X. Or equivalently, we need to do Gaussian Elimination to convert the augmented matrix $\begin{bmatrix} A \mid I \end{bmatrix}$ into the form $\begin{bmatrix} I \mid X \end{bmatrix}$. Once we have done that, we get the inverse of A immediately. Let's discuss an example to show how to achieve it:

■ **Example 2.1** Assuming we have only 3 systems of equations to solve. And we put them altogehter into one Augmented matrix. And the right side of augmented matrix is an identity matrix

$$\begin{bmatrix} A \mid e_1 \mid e_2 \mid e_3 \end{bmatrix} = \left[\begin{array}{ccc|ccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 4 & -6 & 0 & 0 & 1 & 0 \\ -2 & 7 & 2 & 0 & 0 & 1 \end{array}\right] \xrightarrow[E_{21}=\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}]{E_{31}=\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}} \left[\begin{array}{ccc|ccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 0 & -8 & -2 & -2 & 1 & 0 \\ 0 & 8 & 3 & 1 & 0 & 1 \end{array}\right]$$

$$\xrightarrow{E_{32}=\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}} \left[\begin{array}{ccc|ccc} 2 & 1 & 1 & 1 & 0 & 0 \\ 0 & -8 & -2 & -2 & 1 & 0 \\ 0 & 0 & 1 & -1 & 1 & 1 \end{array}\right] \xrightarrow[E_{13}=\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}]{E_{23}=\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}} \left[\begin{array}{ccc|ccc} 2 & 1 & 0 & 2 & -1 & -1 \\ 0 & -8 & 0 & -4 & 3 & 2 \\ 0 & 0 & 1 & -1 & 1 & 1 \end{array}\right]$$

$$\implies \left[\begin{array}{ccc|ccc} 2 & 1 & 0 & 2 & -1 & -1 \\ 0 & 1 & 0 & \frac{1}{2} & -\frac{3}{8} & -\frac{1}{4} \\ 0 & 0 & 1 & -1 & 1 & 1 \end{array}\right] \xrightarrow{E_{12}=\begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}} \left[\begin{array}{ccc|ccc} 2 & 0 & 0 & \frac{12}{8} & -\frac{5}{8} & -\frac{6}{8} \\ 0 & 1 & 0 & \frac{1}{2} & -\frac{3}{8} & -\frac{1}{4} \\ 0 & 0 & 1 & -1 & 1 & 1 \end{array}\right]$$

$$\Longrightarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{12}{16} & -\frac{5}{16} & -\frac{6}{16} \\ 0 & 1 & 0 & \frac{1}{2} & -\frac{3}{8} & -\frac{1}{4} \\ 0 & 0 & 1 & -1 & 1 & 1 \end{array}\right]$$

The final augmented matrix is equivalent to the system $\boldsymbol{I}\boldsymbol{X} = \begin{bmatrix} \frac{12}{16} & -\frac{5}{16} & -\frac{6}{16} \\ \frac{1}{2} & -\frac{3}{8} & -\frac{1}{4} \\ -1 & 1 & 1 \end{bmatrix}$.

Hence we obtain the inverse: $\boldsymbol{A}^{-1} = \boldsymbol{X} = \begin{bmatrix} \frac{12}{16} & -\frac{5}{16} & -\frac{6}{16} \\ \frac{1}{2} & -\frac{3}{8} & -\frac{1}{4} \\ -1 & 1 & 1 \end{bmatrix}$. ■

Then let's study in which case does the inverse exist:

Theorem 2.1 The inverse of $n \times n$ matrix $\boldsymbol{A}$ exists if and only if $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ has a unique solution.

Proofoutline. The inverse of $n \times n$ matrix $\boldsymbol{A}$ exists

$\Leftrightarrow$ none pivot values of $\boldsymbol{A}$ is zero. $\Leftrightarrow$ $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ has a unique solution $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$. ■

At the end, let's prove the claim at the beginning of the lecture:

Theorem 2.2 Let $\boldsymbol{A}$ be $n \times n$ matrix, the following statements are equivalent:

1. Columns of $\boldsymbol{A}$ can be linearly combined to zero nontribally.
2. $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ has infinitely many solutions.
3. Row vectors of $\boldsymbol{A}$ can be linearly combined to zero nontrivally.

Proofoutline. The following statements are equivalent:

- Columns of $\boldsymbol{A}$ can be linearly combined to zero nontribally.
- Given $\boldsymbol{A} = \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix}$, then there exists x_i's that are not all zero such that

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0.$$

- $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ has a nonzero solution $\overline{\boldsymbol{x}}$.

- $2\overline{\boldsymbol{x}}, 3\overline{\boldsymbol{x}}, \dots$ are also solutions to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$.
- $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ has infinitely many solutions.
- $\boldsymbol{A}^{-1}$ does not exist. (otherwise we will only have unique solution $\boldsymbol{A}^{-1} \times \boldsymbol{0} = \boldsymbol{0}$.)
- Gaussian Elimination breaks down, i.e., there exists zero row in the row echelon form.
- Row vectors of $\boldsymbol{A}$ can be linearly combined to zero nontrivally.

■

2.2. Wednesday

2.2.1. Remarks on Gaussian Elimination

Gaussian Elimination to compute $\boldsymbol{A}^{-1}$ is equivalent to solving n linear systems $\boldsymbol{A}\boldsymbol{x}_i = e_i$, $i = 1,2,\ldots,n$.

Computing Complexity. For each i solving $\boldsymbol{A}\boldsymbol{x_i} = e_i$ takes $O(n^3)$ operations.

- Hence, solving these systems one by one take $O(n^4)$ time.
- However, if we solve $\boldsymbol{A}\boldsymbol{x_i} = e_i$ for $i = 1,2,\ldots,n$ simultaneously (that means we write all b_i at the right side of the Augmented matrix), by Gaussian Elimination, it only takes $O(n^3)$ operations.

Large Scale Inverse Computation. Gaussian Elimination is not a good job for large scale sparse matrix (**sparse matrix** is a matrix in which most of the elements are zero. If given a 1000×1000 sparse matrix, it is expensive to do Gaussian Elimination on this matrix).

Actually, for such matrix we use *iterative method* to solve it.

Gaussian Elimination is just a sequence of matrix multiplications. Given nonsingular matrix $\boldsymbol{A}$, Gaussian Elimination is really a sequence of multiplications by elementary matrices $\boldsymbol{E}$'s and permutation matrix $\boldsymbol{P}$:

$$\boldsymbol{E}\cdots\boldsymbol{E}\boldsymbol{P}\boldsymbol{A} = \boldsymbol{U},$$

where $\boldsymbol{U}$ is an upper triangular matrix.

By postmultiplying $\boldsymbol{U}^{-1}$ we obtain

$$\boldsymbol{U}^{-1}(\boldsymbol{E}\ldots\boldsymbol{E}\boldsymbol{P}\boldsymbol{A}) = \boldsymbol{I} \implies (\boldsymbol{U}^{-1}\boldsymbol{E}\ldots\boldsymbol{E}\boldsymbol{P})\boldsymbol{A} = \boldsymbol{I}.$$

Furthermore, we could decompose $\boldsymbol{A}$ as the product of a permutation matrix, a lower triangular matrix and an upper triangular matrix:

$$A = P^{-1}(E^{-1} \dots E^{-1})U$$

2.2.2. Properties of matrix

1. If A is a diagonal matrix which is given by

$$A = \begin{bmatrix} d_1 & & 0 \\ & \vdots & \\ 0 & & d_n \end{bmatrix},$$

   and $d_1 d_2 d_3 \dots d_n \neq 0$, then A^{-1} exists, and $A^{-1} = \begin{bmatrix} d_1^{-1} & & 0 \\ & \vdots & \\ 0 & & d_n^{-1} \end{bmatrix}$.

2. If D_1, D_2 are diagonal and their product exists, then we have

$$D_1 D_2 = D_2 D_1$$

3. If A, B are both invertible, then AB is also invertible. The inverse of product AB is

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

   Proofoutline. To see why the order is reversed, firstly multiply AB with $B^{-1}A^{-1}$:

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

   Similarly, $B^{-1}A^{-1}$ times AB leads to the same result. Hence we draw the conclusion: Inverse come in reverse order. ■

4. The same reverse order applies to three or more matrix:
   If A, B, C are nonsingular, then $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$.

5. It's hard to say whether $(\boldsymbol{A}+\boldsymbol{B})$ is invertible, but we have an interesting property:

When $\boldsymbol{A}$ is "small" (we will explain it later), we have $\quad (\boldsymbol{I}-\boldsymbol{A})^{-1} = \sum_{i=1}^{\infty} \boldsymbol{A}^i$

6. **A triangular matrix is invertible if and only if no diagonal entries are zero.**

   In order to explain it, let's discuss an example:

   ■ **Example 2.2**

   We want to find the inverse of a lower triangular matrix $\boldsymbol{A}$:

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

   Thus we do Gaussian Elimination to compute solution to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{I}$:

$$\left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \end{array}\right] \Longrightarrow \left[\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & -1 & 1 \end{array}\right]$$

   This result is obtained by three row operations:

   (a) "Add $(-1)\times$ row 3 to row 4";

   (b) "Add $(-1)\times$ row 2 to row 3";

   (c) "Add $(-1)\times$ row 1 to row 2". ■

   Proof. Only for a nonzero diagonal lower triangular matrix, we can continue the Gaussian Elimination to convert it into identity matrix. ■

7. Given an invertible lower triangular matrix $\boldsymbol{A}$, the inverse of $\boldsymbol{A}$ remains lower triangular.

8. The LDU decomposition is unique for an invertible matrix. (We assume the existence of the LDU decomposition).

Proof. • Assume the invertible matrix $\boldsymbol{A}$ could be decomposed as:

$$\boldsymbol{A} = \boldsymbol{L}_1\boldsymbol{D}_1\boldsymbol{U}_1 = \boldsymbol{L}_2\boldsymbol{D}_2\boldsymbol{U}_2$$

- By aftermultiplying $\boldsymbol{U}_1^{-1}$ and postmultiplying $\boldsymbol{L}_2^{-1}$ for the latter equation, we obtain:

$$\boldsymbol{L}_1\boldsymbol{D}_1\boldsymbol{U}_1 = \boldsymbol{L}_2\boldsymbol{D}_2\boldsymbol{U}_2 \implies \boldsymbol{L}_2^{-1}\boldsymbol{L}_1\boldsymbol{D}_1 = \boldsymbol{D}_2\boldsymbol{U}_2\boldsymbol{U}_1^{-1} \tag{2.1}$$

- Note that $\boldsymbol{L}_2^{-1}\boldsymbol{L}_1$ remains lower triangular with unit diagonal, thus $\boldsymbol{L}_2^{-1}\boldsymbol{L}_1\boldsymbol{D}_1$ must be lower triangular matrix. Similarly, $\boldsymbol{D}_2\boldsymbol{U}_2\boldsymbol{U}_1^{-1}$ must be upper triangular matrix. Hence $\boldsymbol{L}_2^{-1}\boldsymbol{L}_1\boldsymbol{D}_1$ and $\boldsymbol{D}_2\boldsymbol{U}_2\boldsymbol{U}_1^{-1}$ must be *diagonal* matrix due to equality (2.1).

- Note that the diagonal of $\boldsymbol{L}_2^{-1}\boldsymbol{L}_1\boldsymbol{D}_1$ is the same as the diagonal of $\boldsymbol{D}_1$ since $\boldsymbol{L}_2^{-1}\boldsymbol{L}_1$ has unit diagonal. Hence

$$\boldsymbol{L}_2^{-1}\boldsymbol{L}_1\boldsymbol{D}_1 = \boldsymbol{D}_1. \tag{2.2}$$

  Similarly,

$$\boldsymbol{D}_2\boldsymbol{U}_2\boldsymbol{U}_1^{-1} = \boldsymbol{D}_2. \tag{2.3}$$

  Combining (2.1) to (2.3), we derive $\boldsymbol{D}_1 = \boldsymbol{D}_2$.

- Furthermore,

$$\boldsymbol{L}_2^{-1}\boldsymbol{L}_1\boldsymbol{D}_1 = \boldsymbol{D}_1 \implies \boldsymbol{L}_2^{-1}\boldsymbol{L}_1 = \boldsymbol{I} \implies \boldsymbol{L}_1 = \boldsymbol{L}_2$$

  Similarly, $\boldsymbol{U}_1 = \boldsymbol{U}_2$.

■

2.2.3. matrix transpose

We introduce a new matrix, it is the **transpose** of $\boldsymbol{A}$:

Definition 2.3 [Transpose] The **transpose** of matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is denoted as $\boldsymbol{A}^{\mathrm{T}}$. The columns of $\boldsymbol{A}^{\mathrm{T}}$ are the rows of $\boldsymbol{A}$, i.e., $\boldsymbol{A}^{\mathrm{T}}$ means that

$$\boldsymbol{A}^{\mathrm{T}} = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix} \in \mathbb{R}^{n\times n}$$

For example,

- given a column vector $x \in \mathbb{R}^n$, the transpose $x^{\mathrm{T}} = (x_1, x_2, \dots, x_n)$ is row vector.
- When $\boldsymbol{A}$ is $m \times n$ matrix, the transpose is $n \times m$:

$$\boldsymbol{A} = \begin{bmatrix} 2 & 1 & 4 \\ 0 & 0 & 3 \end{bmatrix} \qquad \boldsymbol{A}^{\mathrm{T}} = \begin{bmatrix} 2 & 0 \\ 1 & 0 \\ 4 & 3 \end{bmatrix} \qquad (\boldsymbol{A}^{\mathrm{T}})^{\mathrm{T}} = \boldsymbol{A}$$

The entry in row i, column j of $\boldsymbol{A}^{\mathrm{T}}$ comes from row j, column i of the original matrix $\boldsymbol{A}$:

$$\textbf{Exchange rows and columns} \quad (\boldsymbol{A}^{\mathrm{T}})_{ij} = \boldsymbol{A}_{ji}$$

The rules for transposes are very direct:

Proposition 2.4

- **Sum** The transpose of $\boldsymbol{A} + \boldsymbol{B}$ is $\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{B}^{\mathrm{T}}$.
- **Product** The transpose of $\boldsymbol{AB}$ is $(\boldsymbol{AB})^{\mathrm{T}} = (\boldsymbol{B})^{\mathrm{T}}(\boldsymbol{A})^{\mathrm{T}}$.

Proofoutline of Product Rule.

- We start with $(\boldsymbol{A}x)^{\mathrm{T}} = x^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}$, where x refers to a vector:

  $\boldsymbol{A}x$ combines the columns of $\boldsymbol{A}$; while $x^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}$ combines the rows of $\boldsymbol{A}^{\mathrm{T}}$.

  Since they are the same combinations of the same vectors, we obtain $(\boldsymbol{A}x)^{\mathrm{T}} = x^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}$.

- Now we can prove the formula $(\boldsymbol{A}\boldsymbol{B})^{\mathrm{T}} = (\boldsymbol{B})^{\mathrm{T}}(\boldsymbol{A})^{\mathrm{T}}$, where $\boldsymbol{B}$ has several columns:

  Assuming $\boldsymbol{B} = \left[\begin{array}{c|c|c|c} b_1 & b_2 & \dots & b_k \end{array}\right]$, then Transposing $\boldsymbol{A}\boldsymbol{B} = \left[\begin{array}{c|c|c|c} \boldsymbol{A}b_1 & \boldsymbol{A}b_2 & \dots & \boldsymbol{A}b_k \end{array}\right]$ gives

$$(\boldsymbol{A}\boldsymbol{B})^{\mathrm{T}} = \begin{bmatrix} b_1^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}} \\ b_2^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}} \\ \vdots \\ b_k^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}} \end{bmatrix},$$

  which is actually $\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}$.

■

2.2.3.1. symmetric matrix

For a *symmetric matrix*, transposing $\boldsymbol{A}$ into $\boldsymbol{A}^{\mathrm{T}}$ makes no change.

Definition 2.4 [symmetric matrix] A matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is **symmetric matrix** if we have $\boldsymbol{A} = \boldsymbol{A}^{\mathrm{T}}$. This means that $a_{ij} = a_{ji}$ for all i, j. We usually denote it as $\boldsymbol{A} \in \mathbb{S}^{n\times n}$.

Choose any matrix $\boldsymbol{A}$ (probably rectangular), then postmultiplying $\boldsymbol{A}^{\mathrm{T}}$ for $\boldsymbol{A}$ automatically leads to a square symmetric matrix:

The transpose of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is $\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}})^{\mathrm{T}}$, which is $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$.

The matrix $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ is also symmetric. But note that $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ is a different matrix from $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$.

(R) For two vector x and y,

- **The dot product or inner product is denoted as** $x^{\mathrm{T}}y$
- **The rank one product or outer product is denoted as** xy^{T}

$x^{\mathrm{T}}y$ is a number while xy^{T} is a matrix.

We introduce a matrix that seems opposite to symmetric matrix:

Definition 2.5 [Skew-symmetric] For matrix $\boldsymbol{A}$, if we have $\boldsymbol{A}^{\mathrm{T}} = -\boldsymbol{A}$, then we say $\boldsymbol{A}$ is **skew-symmetric** or **anti-symmetric**. ■

Moreover, any $n \times n$ matrix can be decomposed as the summation of a symmetric and a skew-symmetric matrix. Let's prove it in the next lecture.

2.3. Assignment Two

1. Let $\boldsymbol{M} = \boldsymbol{ABC}$, where $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ are *square* matrices. Then show that $\boldsymbol{M}$ is *invertible* if and only if $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ are **all** invertible.

2. Find the inverses of

$$\begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{C} & \boldsymbol{I} \end{bmatrix} \quad \begin{bmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix} \quad \begin{bmatrix} \boldsymbol{0} & \boldsymbol{I} \\ \boldsymbol{I} & \boldsymbol{D} \end{bmatrix}.$$

3. For which values of c is the following matrix not *invertible*? Explain your answers.

$$\begin{bmatrix} 2 & c & c \\ c & c & c \\ 8 & 7 & c \end{bmatrix}.$$

4. Determine if the following statements are true or false. (with a counter example if false and a reason if true)

   (a) A 4×4 matrix with a row of **zeros** is not *invertible.*

   (b) A matrix with 1's down the *main diagonal* is *invertible.*

   (c) If $\boldsymbol{A}$ is invertible, then $\boldsymbol{A}^{-1}$ is *invertible.*

   (d) If $\boldsymbol{A}^{\mathrm{T}}$ is invertible, then $\boldsymbol{A}$ is *invertible.*

2.4. Friday

2.4.1. symmetric matrix

Definition 2.6 [symmetric matrix] A $n \times n$ matrix $\boldsymbol{A}$ is a **symmetric matrix** if we have $\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A}$, which means $a_{ij} = a_{ji}$ for all i, j.

For example, the matrix $\boldsymbol{A}$ shown below is a symmetric matrix:

$$\textbf{symmetric matrix} \quad \boldsymbol{A} = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} = \boldsymbol{A}^{\mathrm{T}}$$

Definition 2.7 [skew-symmetric matrix] A $n \times n$ matrix $\boldsymbol{A}$ is a **skew-symmetric matrix** or say, **anti-symmetric matrix** if we have $\boldsymbol{A} = -\boldsymbol{A}^{\mathrm{T}}$.

For example, matrix $\boldsymbol{B}$ shown below is a skew-symmetric matrix:

$$\textbf{skew-symmetric matrix} \quad \boldsymbol{B} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = -\boldsymbol{B}^{\mathrm{T}}$$

Theorem 2.3 Any $n \times n$ matrix can be decomposed as the sum of a *symmetric* and a *skew-symmetric* matrix.

Proofoutline. Given any $n \times n$ matrix $\boldsymbol{A}$, we can write $\boldsymbol{A}$ as:

$$\boldsymbol{A} = \underbrace{\frac{\boldsymbol{A} + \boldsymbol{A}^{\mathrm{T}}}{2}}_{\text{symmetric}} + \underbrace{\frac{\boldsymbol{A} - \boldsymbol{A}^{\mathrm{T}}}{2}}_{\text{skew-symmetric}}$$

■

2.4.2. Interaction of inverse and transpose

Proposition 2.5 If $\boldsymbol{A}$ exists, then $\boldsymbol{A}^{\mathrm{T}}$ also exists, and $(\boldsymbol{A}^{\mathrm{T}})^{-1} = (\boldsymbol{A}^{-1})^{\mathrm{T}}$.

Proof.

$$(\boldsymbol{A}^{-1}\boldsymbol{A})^{\mathrm{T}} = \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}^{-1})^{\mathrm{T}} = \boldsymbol{I} \implies (\boldsymbol{A}^{-1})^{\mathrm{T}} = (\boldsymbol{A}^{\mathrm{T}})^{-1}$$

■

Corollary 2.1 If matrix $\boldsymbol{A}$ is symmetric and invertible, then $\boldsymbol{A}^{-1}$ remains symmetric.

Proof.

$$(\boldsymbol{A}^{-1})^{\mathrm{T}} = (\boldsymbol{A}^{\mathrm{T}})^{-1} = \boldsymbol{A}^{-1} \implies \boldsymbol{A}^{-1} \text{ is symmetric.}$$

■

Proposition 2.6 If $\boldsymbol{M} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}$, then $\boldsymbol{M}^{\mathrm{T}} = \begin{bmatrix} \boldsymbol{A}^{\mathrm{T}} & \boldsymbol{C}^{\mathrm{T}} \\ \boldsymbol{B}^{\mathrm{T}} & \boldsymbol{D}^{\mathrm{T}} \end{bmatrix}$.

Corollary 2.2 Given matrix $\boldsymbol{M} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}$, matrix $\boldsymbol{M}$ is symmetric if and only if

$$\boldsymbol{A} = \boldsymbol{A}^{\mathrm{T}}, \boldsymbol{D} = \boldsymbol{D}^{\mathrm{T}}, \boldsymbol{B}^{\mathrm{T}} = \boldsymbol{C}.$$

Proposition 2.7 Suppose $\boldsymbol{A}$ is invertible and symmetric. When we do LDU decomposition such that $\boldsymbol{A} = \boldsymbol{LDU}$, $\boldsymbol{U}$ is exactly $\boldsymbol{L}^{\mathrm{T}}$.

Proofoutline. Note that

$$\boldsymbol{A}^{\mathrm{T}} = (\boldsymbol{LDU})^{\mathrm{T}} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{D}^{\mathrm{T}}\boldsymbol{L}^{\mathrm{T}} = \boldsymbol{A} = \boldsymbol{LDU}.$$

Since $\boldsymbol{D}$ is diagonal matrix, we have $\boldsymbol{D} = \boldsymbol{D}^{\mathrm{T}}$. It follows that

$$\boldsymbol{U}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{L}^{\mathrm{T}} = \boldsymbol{LDU} = \boldsymbol{A}.$$

Since U^{T} is also a lower triangular matrix, L^{T} is also an upper triangular matrix, $U^{\mathrm{T}}DL^{\mathrm{T}}$ is also the LDU decomposition of A.

Due to the uniqueness of LDU decomposition, we obtain $U^{\mathrm{T}} = L, L^{\mathrm{T}} = U$. ■

2.4.3. Vector Space

We move to a new topic: vector spaces.

From Numbers to Vectors. We know matrix calculation(such as $Ax = b$) 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*".

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

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

Definition 2.8 [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:*

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

Definition 2.9 [vector space] A **vector space** 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 V, then $v + w \in V$.
- **scalar multiplication:**If vector $v \in V$, then $cv \in V$ for any real numbers c.

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.

Proposition 2.8 Every vector space must contain the zero vector.

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

■ **Example 2.3**

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

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:

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

$$V = \text{span}\left\{ v_1 = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{4} \\ \vdots \\ \frac{1}{2^n} \\ \vdots \end{pmatrix}, v_2 = \begin{pmatrix} \frac{1}{3} \\ \frac{1}{9} \\ \vdots \\ \frac{1}{3^n} \\ \vdots \end{pmatrix}, v_3 = \begin{pmatrix} \frac{1}{4} \\ \frac{1}{16} \\ \vdots \\ \frac{1}{4^n} \\ \vdots \end{pmatrix} \right\}$$

$$= \{\alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 \mid \alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}\}$$

is also vector space.

Definition 2.10 [Span] The **span** of a collection of vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n \in \mathbb{R}^m$ is defined as:

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

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 2.4** $\boldsymbol{F} = \{f(x) \mid f : [0,1] \mapsto \mathbb{R}\}$ 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 2.5**

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

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 2.6**

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

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 $\dim(\boldsymbol{V}_1) = 6 < 9$. ▪

2.4.3.1. The solution to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$

We can use vector space to discuss the solution to system of equation. Firstly, let's introduce some definitions:

Definition 2.11 [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, $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ is said to be **homogeneous**.

Definition 2.12 [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 $\boldsymbol{A} = \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix}$, its column space is denoted as

$$\boldsymbol{C}(\boldsymbol{A}) := \operatorname{span}(a_1, a_2, \dots, a_n) \subset \mathbb{R}^m.$$

Definition 2.13 [null space] The null space of a matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ consists of all solutions to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$, which can be denoted as

$$\boldsymbol{N}(\boldsymbol{A}) = \{\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}\} \subset \mathbb{R}^n.$$

Proposition 2.9 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 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}, \boldsymbol{A}\boldsymbol{y} = \boldsymbol{0}$.

$$\implies \boldsymbol{A}(\alpha\boldsymbol{x} + \beta\boldsymbol{y}) = \alpha(\boldsymbol{A}\boldsymbol{x}) + \beta(\boldsymbol{A}\boldsymbol{y}) = \alpha\boldsymbol{0} + \beta\boldsymbol{0} = \boldsymbol{0} \quad \alpha, \beta \in \mathbb{R}.$$

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 2.7** 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 2.8** 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 $\boldsymbol{N}(\boldsymbol{A}) = \left\{ c \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} \middle| c \in \mathbb{R} \right\}$. ■

2.4.3.2. The complete solution to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$

In order to find all solutions of $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, ($\boldsymbol{A}$ may not be square matrix), let's introduce two kinds of solutions:

Definition 2.14 [Particular & Special Solution] For the system of equations $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, there are two kinds of solutions:

$\boldsymbol{x}_{\text{particular}}$ **The particular solution that solves $\boldsymbol{Ax} = \boldsymbol{b}$**

$\boldsymbol{x}_{\text{nullspace}}$ **The special solutions that solves $\boldsymbol{Ax} = \boldsymbol{0}$**

There is a theorem that helps us to obtain the complete solution to $\boldsymbol{Ax} = \boldsymbol{b}$.

Theorem 2.4 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

$$\boldsymbol{Ax}_{complete} = \boldsymbol{A}(\boldsymbol{x_p} + \boldsymbol{x_n}) = \boldsymbol{Ax_p} + \boldsymbol{Ax_n} = \boldsymbol{b} + \boldsymbol{0} = \boldsymbol{b}.$$

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 $\boldsymbol{x}^* - \boldsymbol{x_p} \in N(\boldsymbol{A})$.

Notice that $\boldsymbol{A}(\boldsymbol{x}^* - \boldsymbol{x_p}) = \boldsymbol{Ax}^* - \boldsymbol{Ax_p} = \boldsymbol{b} - \boldsymbol{b} = \boldsymbol{0} \implies \boldsymbol{x}^* - \boldsymbol{x_p} \in N(\boldsymbol{A})$. ∎

∎

▪ **Example 2.9** 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

$$x_{complete} = x_p + x_n = \begin{pmatrix} c+1 \\ -c+1 \end{pmatrix}.$$

So we summarize that if there are n unknowns and m equations such that $m < n$, then $Ax = b$ is **underdetermined** (It may have infinitely many solutions since the special solutions could be infinite).

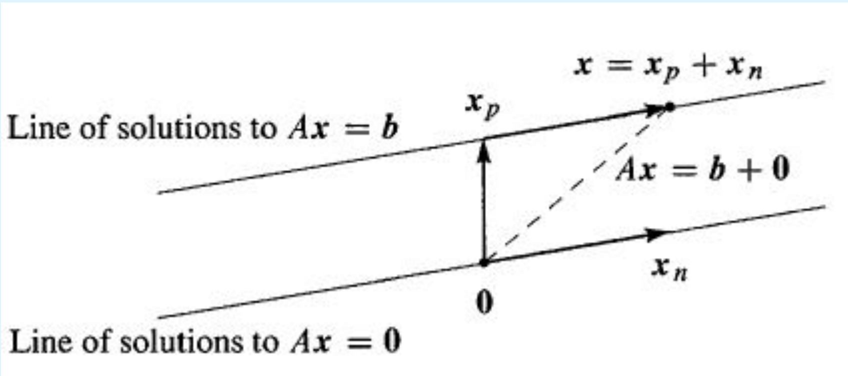


Figure 2.1: Complete solution = one particular solution + all nullspace solutions

■

2.4.3.3. Row-Echelon Matrices

Given $m \times n$ rectangular matrix A, we can still do Gaussian Elimination to convert A into U, where U is of **Row Echelon form**. The whole process could be expressed as:

$$PA = LDU.$$

where L is $m \times m$ **lower triangular** matrix, U is $m \times n$ matrix that is of **row echelon form**.

■ **Example 2.10** Here is a 4×7 row echelon matrix with the three pivots **1** highlighted in blue:

$$
\boldsymbol{U} = \begin{bmatrix} 1 & \times & \times & \times & \times & \times & \times \\ 0 & 1 & \times & \times & \times & \times & \times \\ 0 & 0 & 0 & 0 & 0 & 1 & \times \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}
$$

- Columns 3,4,5,7 have no pivots, and we say the free variables are x_3, x_4, x_5, x_7.
- Columns 1,2,6 have pivots, and we say the pivot variables are x_1, x_2, x_6.

Moreover, we can continue Gaussian Elimination to convert $\boldsymbol{U}$ into $\boldsymbol{R}$ that is of **reduced row echelon form**:

$$
\boldsymbol{R} = \begin{bmatrix} 1 & 0 & \times & \times & \times & 0 & \times \\ 0 & 1 & \times & \times & \times & 0 & \times \\ 0 & 0 & 0 & 0 & 0 & 1 & \times \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}
$$

The reduced row echelon matrix $\boldsymbol{R}$ has zeros above the pivots as well as below. Zeros above the pivots come from upward elimination.

R Remember the two steps (forward and back elimination) in solving $\boldsymbol{Ax} = \boldsymbol{b}$:

1. **Forward Elimination** takes $\boldsymbol{A}$ to $\boldsymbol{U}$. (or its reduced form $\boldsymbol{R}$)
2. **Back Elimination** in $\boldsymbol{Ux} = \boldsymbol{c}$ or $\boldsymbol{Rx} = \boldsymbol{d}$ produces $\boldsymbol{x}$.

2.4.3.4. 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).

2.5. Assignment Three

1. Check and verify the following:

   (a) If $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$, then

   $$\boldsymbol{M}^{-1} = \boldsymbol{I} + \frac{\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}}{1 - \boldsymbol{v}^{\mathrm{T}}\boldsymbol{u}}. \qquad (\boldsymbol{v}^{\mathrm{T}}\boldsymbol{u} \neq 1)$$

   (b) If $\boldsymbol{M} = \boldsymbol{A} - \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$, then

   $$\boldsymbol{M}^{-1} = \boldsymbol{A}^{-1} + \frac{\boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}}{1 - \boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{u}}. \qquad (\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{u} \neq 1)$$

   (c) If $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{U}\boldsymbol{V}$, where $\boldsymbol{U} \in \mathbb{R}^{n\times m}, \boldsymbol{V} \in \mathbb{R}^{m\times n}$, then

   $$\boldsymbol{M}^{-1} = \boldsymbol{I}_n + \boldsymbol{U}(\boldsymbol{I}_m - \boldsymbol{V}\boldsymbol{U})^{-1}\boldsymbol{V}.$$

   (d) If $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{U}\boldsymbol{W}^{-1}\boldsymbol{V}$, where $\boldsymbol{W} \in \mathbb{R}^{m\times m}, \boldsymbol{U} \in \mathbb{R}^{n\times m}, \boldsymbol{V} \in \mathbb{R}^{m\times n}$, then

   $$\boldsymbol{M}^{-1} = \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{U}(\boldsymbol{W} - \boldsymbol{V}\boldsymbol{A}^{-1}\boldsymbol{U})^{-1}\boldsymbol{V}\boldsymbol{A}^{-1}.$$

2. If $\boldsymbol{A} = \boldsymbol{A}^{\mathrm{T}}$ and $\boldsymbol{B} = \boldsymbol{B}^{\mathrm{T}}$, which of these matrices are certainly *symmetric*?

   (a) $\boldsymbol{A}^2 - \boldsymbol{B}^2$

   (b) $(\boldsymbol{A}+\boldsymbol{B})(\boldsymbol{A}-\boldsymbol{B})$

   (c) $\boldsymbol{A}\boldsymbol{B}\boldsymbol{A}$

   (d) $\boldsymbol{A}\boldsymbol{B}\boldsymbol{A}\boldsymbol{B}$

3. Strat from LDU decomposition, show that each $n \times n$ matrix $\boldsymbol{A}$ can be factorized into a *triangular* matrix times a *symmetric* matrix.

4. Let

   $$\boldsymbol{A} = \begin{bmatrix} 5 & 3 \\ 3 & 2 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 6 & 2 \\ 2 & 4 \end{bmatrix}, \quad \boldsymbol{C} = \begin{bmatrix} 4 & -2 \\ -6 & 3 \end{bmatrix}$$

   solve each of the following matrix equations:

   (a) $\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B} = \boldsymbol{C}$

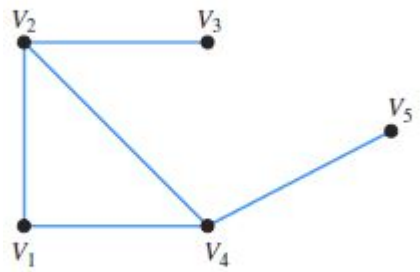
(b) $\boldsymbol{XA} + \boldsymbol{B} = \boldsymbol{C}$

(c) $\boldsymbol{AX} + \boldsymbol{B} = \boldsymbol{X}$

(d) $\boldsymbol{XA} + \boldsymbol{C} = \boldsymbol{X}$

5. Let $\boldsymbol{U}$ and $\boldsymbol{R}$ be $n \times n$ upper triangular matrices and $\boldsymbol{T} = \boldsymbol{UR}$, show that $\boldsymbol{T}$ is also *upper triangular* and that $t_{jj} = u_{jj} r_{jj}, j = 1, \ldots, n$.

6. Consider the graph



(a) Determine the adjacency matrix $\boldsymbol{A}$ of the graph.

(b) Compute $\boldsymbol{A}^2$. What do the entries in the first row of $\boldsymbol{A}^2$ tell you about walks of length 2 that start from V_1?

(c) Compute $\boldsymbol{A}^3$. How many walks of length 3 are there from V_2 to V_3? How many walks of length *less than* or *equal* to 3 are there from $\boldsymbol{V}_2$ to $\boldsymbol{V}_4$?

Chapter 3

Week3

3.1. Tuesday

3.1.1. Introduction

3.1.1.1. Motivation of 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.

3.1.1.2. Preview of LA

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.

3.1.2. Review of 2 weeks

How to solve linear system equations?. The basic method is **Gaussian Elimination**, and the main idea is *induction* to make simpler equations.

- Given one equation $ax = b$, we can easily sovle it:

$$\text{If } a = 0 \text{, there is no solution} \quad \text{otherwise } x = \frac{b}{a}.$$

- We could solve 1×1 system. By induction, if we could solve $n \times n$ systems, then we can solve $(n+1) \times (n+1)$ systems.

In the above process, math notations is needed:

- matrix multiplication
- matrix inverse
- transpose, symmetric matrices

So in first two weeks, we just learn two things:

- *linear system could be solved* ***almost*** *by G.E.*
- *Furthermore, Gaussian Elimination is (almost) LU decomposition.*

But there is a question remained to be solved:

How to solve linear singular system equations?.

- 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 these solutions?

If we express system into matrix form, the question turns into:

How to solve the rectangular?

3.1.3. 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 3.1** We discuss how to solve **square** matrix of **rref**:

- If all rows have nonzero entry, we have:

$$\begin{bmatrix} 1 & & & \text{\Large 0} \\ & 1 & & \\ & & 1 & \\ \text{\Large 0} & & & 1 \end{bmatrix} \boldsymbol{x} = \boldsymbol{c} \implies \boldsymbol{x} = \boldsymbol{c}$$

- 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.

3.1.3.1. Review of G.E. for Nonsingular case

We use matrix to represent system of equations:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{23}x_n = b_2 \\ \cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{m3}x_n = b_m \end{cases} \implies \boldsymbol{Ax} = \boldsymbol{b}$$

By postmultiplying $\boldsymbol{E}_{ij}$ or $\boldsymbol{P}_{ij}$, we are essentially doing one step of elimination:

$$\boldsymbol{E}_{ij}\boldsymbol{Ax} = \boldsymbol{E}_{ij}\boldsymbol{b} \quad \text{or} \quad \boldsymbol{P}_{ij}\boldsymbol{Ax} = \boldsymbol{E}_{ij}\boldsymbol{b}$$

By several steps of elimination, we obtain the final result:

$$\hat{\boldsymbol{L}}\boldsymbol{PAx} = \hat{\boldsymbol{L}}\boldsymbol{Pb}$$

where $\hat{\boldsymbol{L}}\boldsymbol{PA}$ represents an upper triangular matrix $\boldsymbol{U}$, $\hat{L}$ is the lower triangular matrix.

Equivalently, we obtain

$$\hat{\boldsymbol{L}}\boldsymbol{PA} = \boldsymbol{U} \implies \boldsymbol{PA} = \hat{\boldsymbol{L}}^{-1}\boldsymbol{U} \triangleq \boldsymbol{LU}$$

Hence, Gaussian Elimination is almost the LU decomposition.

3.1.3.2. Example for solving rectangular system of rref

Recall the definition for rref:

Definition 3.1 [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 3.2** 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 $\boldsymbol{Rx} = \boldsymbol{0}$:

$$\begin{bmatrix} 1 & 3 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \implies \begin{cases} x_1 + 3x_2 = 0 \\ x_3 + x_4 = 0 \end{cases}$$

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 $\boldsymbol{Rx} = \boldsymbol{0}$ 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 $\boldsymbol{Rx} = \boldsymbol{c}$. 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.*):

$$\begin{bmatrix} 1 & 3 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ 0 \\ x_3 \\ 0 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} \implies \begin{cases} x_1 = c_1 \\ x_3 = c_2 \\ 0 = c_3 \end{cases}$$

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 $\boldsymbol{Rx} = \boldsymbol{c}$ has no solution.

Final solution. If assume $c_3 = 0$, then all solutions to $\boldsymbol{Rx} = \boldsymbol{c}$ 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:

3.1.4. 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 3.3** The first example is a 3×4 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:

$$\text{pivots} \implies \text{pivot columns} \implies \text{pivot variables}$$

■ **Example 3.4** 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 suffices 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 $\boldsymbol{R}\boldsymbol{x} = \boldsymbol{0}$ (to get the value of pivot variables) to get the special solution.

■ **Example 3.5** Continue with 3×4 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 $\boldsymbol{R}\boldsymbol{x} = \boldsymbol{0}$:

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:

$$N(\boldsymbol{A}) = \text{span}(y_1, y_2, \dots, y_{n-r}).$$

■ **Example 3.6** 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:

$$\boldsymbol{x}_{\text{special}} = \text{span}(\begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ -1 \\ 1 \end{bmatrix}) = x_2 \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 0 \\ -1 \\ 1 \end{bmatrix}$$

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 3.7** 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 $c_3 \neq 0$. ■

- 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 3.8** 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 $\boldsymbol{Rx} = \boldsymbol{c}$ is

$$\boldsymbol{x}_p = \begin{bmatrix} c_1 \\ 0 \\ c_2 \\ 0 \end{bmatrix}.$$

Final step: Obtain complete solutions. All solution of $\boldsymbol{Ax} = \boldsymbol{b}$ 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.

However, where does the number r come? r denotes the **rank** of a matrix, which will be discussed in the next lecture.

3.2. Thursday

3.2.1. Review

The last lecture you may be confused about how to compute the null space $N(\boldsymbol{A})$, i.e., why we follow the proceed to compute special solutions y_i. Let's review the whole steps for solving rectangular by using block matrix form.

- After converting the matrix $\boldsymbol{A}$ into the rref form $\boldsymbol{R}$, without loss of generality, we could convert the rref into the form $\begin{bmatrix} \boldsymbol{I} & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}$ by switching columns.

■ **Example 3.9** Last time our rref is given by:

$$\boldsymbol{R} = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

We notice that column 3 is pivot column, so we can switch it into the second column. (By switching column 2 and column 3):

$$\boldsymbol{R} \implies \begin{bmatrix} 1 & 0 & 3 & -1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}$$

- Thus our system could be written into the form:

$$\boldsymbol{R}\boldsymbol{x} = \boldsymbol{c} \implies \begin{bmatrix} \boldsymbol{I} & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} \tag{3.1}$$

Since we have changed the columns of $\boldsymbol{R}$, so the row 2 and row 3 of $\boldsymbol{x}$ is also switched respectively. Thus x_1 and x_2 are pivot variables, and x_3 and x_4 are free variables of $\boldsymbol{x}$. From (3.1) we derive:

$$\begin{cases} \boldsymbol{I}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \boldsymbol{B}\begin{bmatrix} x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \\ 0 = c_3 \end{cases}$$

- If $c_3 \neq 0$, then there is no solution; so let's preassume $c_3 = 0$. Then *pivot variables* could be expressed as the form of *free variables*:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} - \boldsymbol{B}\begin{pmatrix} x_3 \\ x_4 \end{pmatrix}$$

Suppose $-\boldsymbol{B} = \begin{bmatrix} \hat{\boldsymbol{y}}_1 & \hat{\boldsymbol{y}}_2 \end{bmatrix}$, then pivot variables can be expressed as:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + x_3\hat{\boldsymbol{y}}_1 + x_4\hat{\boldsymbol{y}}_2$$

- Therefore, the complete solution to the system is given by

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} x_3\hat{\boldsymbol{y}}_1 + x_4\hat{\boldsymbol{y}}_2 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ x_3 \\ x_4 \end{pmatrix} \tag{3.2}$$

$$= \begin{pmatrix} c_1 \\ c_2 \\ 0 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} \hat{\boldsymbol{y}}_1 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} \hat{\boldsymbol{y}}_2 \\ 0 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \tag{3.3}$$

$$= \underbrace{\begin{pmatrix} c_1 \\ c_2 \\ 0 \\ 0 \end{pmatrix}}_{\boldsymbol{x}_p} + \underbrace{x_3 \begin{pmatrix} \hat{\boldsymbol{y}}_1 \\ 1 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} \hat{\boldsymbol{y}}_2 \\ 0 \\ 1 \end{pmatrix}}_{\boldsymbol{x}_{\text{special}}} \tag{3.4}$$

where x_3 and x_4 could be arbitarary.

- We can verify our computed special solutions is true by matrix multiplication:

$$\text{Special Solution Matrix: } \begin{pmatrix} \hat{\boldsymbol{y}}_1 & \hat{\boldsymbol{y}}_2 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -\boldsymbol{B} \\ \boldsymbol{I} \end{pmatrix}$$

$$\text{Verification: } \begin{pmatrix} \boldsymbol{I} & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \begin{bmatrix} -\boldsymbol{B} \\ \boldsymbol{I} \end{bmatrix} = \begin{bmatrix} -\boldsymbol{B} + \boldsymbol{B} \\ \boldsymbol{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}$$

Open Question: If our rectangular matrix is $m \times n (m > n)$, how to solve it?

Answer: Similarly, we do G.E. to get rref. After switching columns, it will be of the form:

$$\begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \\ 0 & \dots & 0 \end{bmatrix} \text{ or } \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & 0 & \\ & & & & 0 \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}$$

3.2.2. Remarks on solving linear system equations

There are two kinds of linear equations, and the classification criteria depends on m and n:

Theorem 3.1 Let m denotes the number of equations, n denotes the number of variables. For the number of solutions for $\boldsymbol{Ax} = \boldsymbol{b}$, where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, we obtain:

- $m < n$: either **no solution** or **infinitely many solutions**
- $m \geq n$: **no solution**; **unique solution** $(N(\boldsymbol{A}) = \boldsymbol{0})$; or **infinitely many solutions**.

We prove for the $m < n$ case first:

Proofoutline for $m < n$ case: Recall that 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:

$$\begin{bmatrix} 1 & & & \times & \times \\ & \ddots & & \times & \times \\ & & 1 & \times & \times \\ 0 & 0 & 0 & 0 & 0 \\ \dots & & & & \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} c_1 \\ \vdots \\ c_r \\ c_{r+1} \\ \vdots \\ c_n \end{bmatrix},$$

where $x_1.x_2.\dots,x_r$ are pivot variables. Hence, we have $(n-r)$ free variables, and $N(\boldsymbol{A})$ is spanned by $(n-r)$ special vectors $y_1, y_2, \dots, y_{n-r}$.

It suffices to show that the $m < n$ rectangular system does not have unique solution, i.e., $N(\boldsymbol{A}) > 0$. It suffices to show $n > r$.

Obviously, $r \le m$, and we have $n > m$, so we obtain $n > r$. ∎

Equivalently, we obtain the proposition and the corollary below:

Proposition 3.1 For system $\boldsymbol{Ax} = \boldsymbol{b}$, where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $m < n$, it either has no solution or infinitely many solutions.

Corollary 3.1 For system $\boldsymbol{Ax} = \boldsymbol{0}$, where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $m < n$, it always has infinitely many solutions.

3.2.2.1. What is r?

We ask the question again, what is r? Let's see some examples before answering this question.

■ **Example 3.10** If we want to solve system of equations of size 1000 as the following:

$$\begin{cases} x_1 + x_2 = 3 \\ 2x_1 + 2x_2 = 6 \\ \cdots \\ 1000x_1 + 1000x_2 = 3000 \end{cases}$$

It seems very difficult when hearding it has 1000 equations, but the remaining 999 equations could be redundant (They actually don't exist):

$$\begin{bmatrix} 1 & 1 \\ 2 & 2 \\ \vdots & \vdots \\ 1000 & 1000 \end{bmatrix} \implies \begin{bmatrix} 1 & 1 \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix}$$

■

Here we see that only one equation $x_1 + x_2 = 3$ is real, the remaining part is not real. So we claim that r is the number of "real" equations. But what is the definition for "real" equations? Let's discuss the definition for *linear dependence* first.

3.2.3. Linear dependence

Definition 3.2 [linear dependence] The vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n$ in linear space $\boldsymbol{V}$ are **linearly dependent** if there exists $c_1, c_2, \dots, c_n \in \mathbb{R}$ s.t.

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n = \mathbf{0}.$$

In other words, it means one of v_i could be expressed as the linear combination of others. Assume $c_n \neq 0$, we can express $\boldsymbol{v}_n$ as:

$$\boldsymbol{v}_n = -\frac{c_1}{c_n}\boldsymbol{v}_1 - \frac{c_2}{c_n}\boldsymbol{v}_2 - \cdots - \frac{c_{n-1}}{c_n}\boldsymbol{v}_{n-1}.$$

Definition 3.3 [linear independence] The vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n$ in linear space $\boldsymbol{V}$ are **linearly independent** if the equation

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n = \mathbf{0}$$

only has the trivial solution $c_1 = c_2 = \cdots = c_n = 0$.

In other words, if $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n$ are not **linearly dependent**, they must be **linearly independent**.

R Note that **only** in this course, if we say vectors are dependent, we mean they are **linearly** dependent. In other courses dependent may have other definitions. In the following lectures, we simplify the noun *dependent* as *dep.*; and the noun *independent* as *ind.*

Here we pick some examples to help you understand dep. and ind.:

■ **Example 3.11**

- $\boldsymbol{v}_1 = (1,1)$ and $\boldsymbol{v}_2 = (2,2)$ are **dep.** because

$$(-2) \times \boldsymbol{v}_1 + \boldsymbol{v}_2 = \mathbf{0}.$$

- The only one vector $\boldsymbol{v}_1 = 2$ is **ind.** because

$$c\boldsymbol{v}_1 = \mathbf{0} \Longleftrightarrow c = 0.$$

- The only one vector $\boldsymbol{v}_1 = 0$ is **dep.** because

$$2 \times \boldsymbol{v}_1 = \mathbf{0}$$

- $\boldsymbol{v}_1 = (1,2)$ and $\boldsymbol{v}_2 = (0,0)$ are **dep.** because

$$0 \times \boldsymbol{v}_1 + 1 \times \boldsymbol{v}_2 = \mathbf{0}.$$

- The upper triangular matrix $\boldsymbol{A} = \begin{bmatrix} 3 & 4 & 2 \\ 0 & 1 & 5 \\ 0 & 0 & 2 \end{bmatrix}$ has three column vectors:

$$\boldsymbol{v}_1 = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix}, \boldsymbol{v}_2 = \begin{bmatrix} 4 \\ 1 \\ 0 \end{bmatrix}, \boldsymbol{v}_3 = \begin{bmatrix} 2 \\ 5 \\ 2 \end{bmatrix}$$

  $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ are **ind.** because

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 = \mathbf{0} \Longleftrightarrow c_1 = c_2 = c_3 = 0.(\text{Why? because } \boldsymbol{A} \text{ is invertible})$$

■

3.2.3.1. Remarks

How many solutions meet the linear dependence criteria?. Recall that in last week we have studied that the following statements are equivalent: ()

- Vectors $a_1, a_2, \dots, a_n \in \mathbb{R}^m$ are dep.
- $\exists$ nonzero $\boldsymbol{c}$ s.t. $\sum_{i=1}^{n} c_i a_i = \mathbf{0}$.
- $\exists\ \boldsymbol{c} \neq \mathbf{0}$ s.t.

$$\boldsymbol{A}\boldsymbol{c} := \begin{bmatrix} a_1 & \Big| & \dots & \Big| & a_n \end{bmatrix} \boldsymbol{c} = \mathbf{0}$$

For the third statement, if we could choose one $\boldsymbol{c}$, then how many $\boldsymbol{c}$ can we choose? For the $m < n$ case, by corollary (3.1), we obtain:

Corollary 3.2 When vectors $a_1, a_2, \dots, a_n \in \mathbb{R}^m (m < n)$ are dependent, there exists infinitely solutions $c_1, c_2, \dots, c_n$ such that $\sum_{i=1}^{n} c_i a_i = \mathbf{0}$.

The real equations are essentially those linearly independent equations.

3.2.4. Basis and dimension

Definition 3.4 [Basis] The vectors $v_1, \dots, v_n$ form a **basis** for a vector space $\boldsymbol{V}$ if and only if:

1. $v_1, \dots, v_n$ are **linearly independent**.

   and
2. $v_1, \dots, v_n$ span $\boldsymbol{V}$.

■

■ **Example 3.12** In $\mathbb{R}^3$,

- $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ form a basis.

- $\begin{bmatrix}1\\0\\0\end{bmatrix}$ is not a basis, since it doesn't span $\mathbb{R}^3$.
- $\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}, \begin{bmatrix}2\\0\\0\end{bmatrix}$ don't form a basis, since they aren't linearly independent.
- $\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}1\\2\\3\end{bmatrix}, \begin{bmatrix}2\\1\\3\end{bmatrix}$ form a basis.

■

We find that the number of vectors for the basis of $\mathbb{R}^3$ is always 3, is this a coincidence? The theorem below gives the answer.

Theorem 3.2 If $v_1, v_2, \ldots, v_m$ is a basis; and $w_1, w_2, \ldots, w_n$ is another basis for the same vector space $\boldsymbol{V}$, then $n = m$.

In order to proof it, let's try simple case first:

proofoutline. 1. In order to proof it, let's try simple case first:

- Consider $\boldsymbol{V} = \mathbb{R}$ case first: For $\mathbb{R}$, the number 1 forms a basis. Let's show that 2 vectors in $\mathbb{R}$ cannot be a basis:
  - Given any two vectors x and y, they are not a basis for $\mathbb{R}$, since that
    * if $x = 0$ or $y = 0$, they are not ind.
    * otherwise, $y = \frac{y}{x} \times x \implies \frac{y}{x} \times x + (-1) \times y = 0$. So they are not ind.
- Then we consider $\boldsymbol{V} = \mathbb{R}^3$ case:

  For $\mathbb{R}^3$, $\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}$ is a basis. Our goal is to show that if $v_1, v_2, \ldots, v_m$ is a basis, then $m = 3$.
  - Let's show that $m = 4$ is impossible, i.e., 4 vectors in $\mathbb{R}^3$ cannot be a basis.):

    * It suffices to show that for $\forall a_1, a_2, a_3, a_4 \in \mathbb{R}^3$ they must be dep.
    * Or equivalently, $\boldsymbol{Ax} = \boldsymbol{0}$ has nonzero solutions, where $\boldsymbol{A} = \begin{bmatrix} a_1 & a_2 & \dots & a_4 \end{bmatrix} \in \mathbb{R}^{3\times 4}$, which is true by corollary (3.1).
  - Similarly, we could show any basis for $\mathbb{R}^3$ satisfies $m \le 3$ (i.e., m=4,5,... is impossible).
  - Then let's show that $m = 2$ is impossible, i.e., 2 vectors in $\mathbb{R}^2$ cannot be a basis:
    * It suffics to show that for $\forall a_1, a_2 \in \mathbb{R}^3$, they cannot span the whole space.
    * Otherwise, $\boldsymbol{Ax} = \boldsymbol{b}$ must have solution for arbitrary $\boldsymbol{b} \in \mathbb{R}^3$, where $\boldsymbol{A} = \begin{bmatrix} a_1 & a_2 \end{bmatrix} \in \mathbb{R}^{3\times 2}$.
    * However, this kind system may have no solution, which is a contradiction.
  - Similarly, we could show any basis for $\mathbb{R}^3$ satisfies $m \ge 3$.
- The same arugment could show any basis for $\mathbb{R}^n$ satisfies $m = n$.

2. Next, let's consider general vector space. We assume that $n < m$ (by contradiction method).
   Given that $v_1, \dots, v_n$ and $w_1, \dots, w_m$ are the basis of $\boldsymbol{V}$, our goal is to construct a contradiction that $w_1, \dots, w_m$ cannot form a basis.
   It suffices to show that $\exists$(construct) $\boldsymbol{c} = \begin{bmatrix} c_1 & c_2 & \dots & c_m \end{bmatrix}^{\mathrm{T}} \ne \boldsymbol{0}$ s.t.

$$c_1 w_1 + c_2 w_2 + \dots + c_m w_m = 0. \tag{3.5}$$

   Moreover, we can express $w_1, \dots, w_m$ in form of $v_1, \dots, v_n$:

$$\begin{cases} w_1 = a_{11} v_1 + \dots + a_{1n} v_n \\ \dots \\ w_m = a_{m1} v_1 + \dots + a_{mn} v_n \end{cases} \tag{3.6}$$

By (3.6), we can write (3.5) as:

$$\begin{aligned}
0 &= \sum_{j=1}^{m} c_j w_j \\
&= \sum_{j=1}^{m} c_j (\sum_{i=1}^{n} a_{ji} v_i) \\
&= \sum_{j=1}^{m} \sum_{i=1}^{n} c_j a_{ji} v_i \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m} c_j a_{ji} v_i \\
&= \sum_{i=1}^{n} v_i \times (\sum_{j=1}^{m} c_j a_{ji}) \\
&= v_1 \times (\sum_{j=1}^{m} c_j a_{j1}) + v_2 \times (\sum_{j=1}^{m} c_j a_{j2}) + \cdots + v_n \times (\sum_{j=1}^{m} c_j a_{jn})
\end{aligned}$$

So, in order to let LHS=0, we only need to let each of RHS=0, i.e.,

$$\sum_{j=1}^{m} c_j a_{j1} = \sum_{j=1}^{m} c_j a_{j2} = \cdots = \sum_{j=1}^{m} c_j a_{jn} = 0. \tag{3.7}$$

In order to construct c_j, we write (3.7) into matrix form:

$$\boldsymbol{A}^{\mathrm{T}} \boldsymbol{c} = \mathbf{0}, \text{ where } \quad \boldsymbol{A} = \begin{bmatrix} a_{ij} \end{bmatrix}_{1 \le i \le m; 1 \le j \le n}, \quad \boldsymbol{c} = \begin{bmatrix} c_1 & c_2 & \dots & c_m \end{bmatrix}^{\mathrm{T}}.$$

The system $\boldsymbol{A}^{\mathrm{T}} \boldsymbol{c} = \mathbf{0}$ has infinitie nonzero solutions by corollary (3.1). Hence we could construct infinitely such c_j.

■

During the proof, we face two difficulties:

1. For arbitararily $\boldsymbol{V}$, we write a concrete form to express $w_1, w_2, \dots, w_m$.
2. We write into matrix form to express $\sum_{j=1}^{m} c_j a_{j1} = \sum_{j=1}^{m} c_j a_{j2} = \cdots = \sum_{j=1}^{m} c_j a_{jn} = 0$.

Since any basis for $\boldsymbol{V}$ contains the same number of vectors, we can define the number of vectors to be dimension:

Definition 3.5 [Dimension] The **dimension** for a vector space is the number of vectors in a basis for it. ∎

(R) Remember that vector space $\{0\}$ has dimension 0.

In order to denote the dimension for a given vector space $\boldsymbol{V}$, we often write it as $\dim(\boldsymbol{V})$.

∎ **Example 3.13**

- $\mathbb{R}^n$ has dimension n.
- $\{$All $m \times n$ matrix$\}$ has dimension $m \cdot n$.
- $\{$All $n \times n$ symmetric matrix$\}$ has dimension $\frac{n(n+1)}{2}$.
- Let $\boldsymbol{P}$ denote the vector space of all polynomials $f(x) = a_0 + a_1 x + \cdots + a_n x^n$.

  $\dim(\boldsymbol{P}) \neq 3$ since $1, x, x^2, x^3$ are ind.

  The same argument can show $\dim(\boldsymbol{P})$ doesn't equal to any real number, so $\dim(\boldsymbol{P}) = \infty$

∎

Human beings often ask a question: for a line and a plane, which is bigger?

Does plane has more point than a line?. No, Cantor syas they have the same "number" of points by constructing one-to-one mapping.

Furthermore, $\mathbb{R}, \mathbb{R}^2, \ldots, \mathbb{R}^n$ has the same number of points.

Plane and line have different dimensions. However, a plane has more dimensions than a line. So from this point of view, a plane is bigger than a line.

You should know some common knowledge for dimension:

1. Programmer lives in **2** dimension world. (They only live with binary.)
2. Engineer lives in **3** dimension world. (They only live with enign.)
3. Physician lives in **4** dimension world. (They discuss time.)
4. String theories states that our world is **11** or **26** dimension, which has been proved by Qingshi Zhu.

What is rank?. Finally let's answer the question: What is rank?

rank = dimension of row space of a matrix.

We will discuss it in the next lecture.

3.3. Friday

3.3.1. Review

Proposition 3.2 Undetermined system $\boldsymbol{Ax} = \boldsymbol{b}$ with $m < n$, i.e., number of equations $<$ number of unknowns, has **no solution** or **infinitely many solutions**.

We want to understand the meaning of rank: number of "real" equations.

Then we introduce definition of *linearly independence* and *linearly dependence.*

The linear dependence has relation with the system:

Proposition 3.3 $\boldsymbol{Ax} = \boldsymbol{0}$ has nonzero solutions if and only if the column vectors of $\boldsymbol{A}$ are dep.

Combining proposition (3.3) with (3.2), we derive the corollary:

> **Corollary 3.3** Any $(n+1)$ vectors in $\mathbb{R}^n$ are dep.

Proposition 3.4 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**.

From this proposition we derive the corollary immediately:

> **Corollary 3.4** Any $(n-1)$ vectors in $\mathbb{R}^n$ cannot span the whole space.

Then we introduce the definition of basis:

> **Definition 3.6** [Basis] A set of ind. vectors that span this space is called the **basis** of this space. ■

Then we introduce a theorem saying that **All basis of a given vector space have the same size.**

Thus we introduce **dimension** to denote the *number of vectors in a basis.*

3.3.2. More on basis and dimension

The basis of a given vector space has to satisfy two conditions:

$$\underbrace{\text{linear independence}}_{\text{not too many}} + \underbrace{\text{span the space}}_{\text{not too few}}$$

The **ind.** constraint let the size of basis not too many. For example, if given 1000 vectors of $\mathbb{R}^3$, they are very likely to be dep.

Spanning the space let the size of basis not too few. For example, given only 3 vectors of $\mathbb{R}^{100}$, they cannot span the whole space obviously.

We claim that:

$$\begin{aligned} \text{A basis } &= \text{maximal ind. set} \\ &= \text{minimal spanning set} \end{aligned}$$

Definition 3.7 [spanning set] $v_1, v_2, \dots, v_n$ is said to be the spanning set of $\boldsymbol{V}$ if

$$\boldsymbol{V} = \text{span}\{v_1, v_2, \dots, v_n\}.$$

■ **Example 3.14** $v_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$ 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 $\mathbb{R}^3$. ▪

Theorem 3.3 Let $\boldsymbol{V}$ be a space of dimension $n > 0$, then

1. Any set of n ind. vectors span $\boldsymbol{V}$.
2. Any n vectors that span $\boldsymbol{V}$ are ind.

Here is the proof outline, but you should complete the proof in detail.

proofoutline.

1. Suppose $v_1, v_2, \dots, v_n$ are ind. and v is an arbitarary vector in $\boldsymbol{V}$. Firstly, show that $v_1, v_2, \dots, v_n, v$ is dep., thus derive the equation $c_1 v_1 + c_2 v_2 + \dots + c_n v_n + c_{n+1} v = \mathbf{0}$. Argue that the scalar $c_{n+1} \neq 0$. Then we can express v in form of $v_1, v_2, \dots, v_n$, i.e., $v_1, v_2, \dots, v_n$ span $\boldsymbol{V}$.
2. Suppose $v_1, v_2, \dots, v_n$ span $\boldsymbol{V}$. Assume $v_1, v_2, \dots, v_n$ are dep. Then show that v_n could be written as form of other $(n-1)$ vectors, it follows that $v_1, v_2, \dots, v_{n-1}$ still span $\boldsymbol{V}$. If $v_1, v_2, \dots, v_{n-1}$ are also dep, we can continue eliminating one vector. We continue this way until we get an ind. spanning set with $k < n$ elements, which contradicts dim($\boldsymbol{V}$)$= n$. Therefore, $v_1, v_2, \dots, v_n$ must be ind.

■

▪ **Example 3.15** $\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}$ are ind. $\implies$ they span $\mathbb{R}^3$. ▪

3.3.2.1. Clarification of dimension

Firstly, we need to understand "set":

1. $P \triangleq \{\text{All polynomials}\} = \text{span}\{1, x, x^2, \dots\} \implies \dim(P) = \infty$.
2. $P_3 \triangleq \{\text{All polynomials with degree} \leq 3\} = \text{span}\{1, x, x^2, x^3\} \implies \dim(P) = 4$.
3. $Q \triangleq \text{span}\{x^2, 1 + x^3 + x^{10}, x^{300}\} \implies \dim(Q) = 3$.

(R) dim of space $\neq$ dim of the space it lives in.

For example, the line in $\mathbb{R}^{100}$ has dim 1.

3.3.3. What is rank?

Definition 3.8 [Rank] The rank of matrix $\boldsymbol{A}$ is defined as the **number of nonzero pivots of rref of** $\boldsymbol{A}$.

■ **Example 3.16**

$$\boldsymbol{A} = \begin{bmatrix} 1 & 3 & 3 & 4 \\ 2 & 6 & 9 & 7 \\ -1 & -3 & 3 & 4 \end{bmatrix} \xRightarrow{\text{row transform}} \boldsymbol{U} = \begin{bmatrix} 1 & 3 & 0 & -1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$\boldsymbol{U}$ has two pivots, hence $\text{rank}(\boldsymbol{A}) = \text{rank}(\boldsymbol{U}) = 2$.

However, the definition for rank is too complicated, can we define rank of $\boldsymbol{A}$ directly?

Key question: What quantity is not changed under row transformation?

Answer: Dimension of row space.

Definition 3.9 [column space] The **column space** of a matrix is the subspace of $\mathbb{R}^n$ spanned by the columns.

In other words, suppose $\boldsymbol{A} = \begin{bmatrix} a_1 & \vert & \dots & \vert & a_n \end{bmatrix}$, the column space of $\boldsymbol{A}$ is given by

$$\mathcal{C}(\boldsymbol{A}) = \text{span}\{a_1, a_2, \dots, a_n\}.$$

Definition 3.10 [row space] The **row space** of a matrix is the subspace of $\mathbb{R}^n$ spanned by the rows.

Suppose $A = \begin{bmatrix} a_1 \\ \hline \dots \\ \hline a_n \end{bmatrix}$, the row space of A is given by

$$\mathcal{R}(A) = \text{span}\{a_1, a_2, \dots, a_n\}.$$

The **row space** of A is essentially $\mathcal{R}(A) := \mathcal{C}(A^{\mathrm{T}})$, i.e., the column space of A^{T}.

Proposition 3.5 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 $A \xrightarrow{\text{Row Transfom}} B$, then we have $\mathcal{R}(B) \subset \mathcal{R}(A)$.

Since row transformations are invertible, we also have $B \xrightarrow{\text{Row Transfom}} A$, thus we have $\mathcal{R}(A) \subset \mathcal{R}(B)$.

In conclusion, we obtain $\mathcal{R}(B) = \mathcal{R}(A)$. ∎

Hence $\text{rank}(A) = \text{pivots of } U = \dim(\text{row}(U)) = \dim(\text{row}(A))$.

Hence we have a much simpler definition for rank:

Definition 3.11 [rank] The **dimension of the row space** is the **rank of a matrix**, i.e.,

$$\text{rank}(A) = \dim(\mathcal{R}(A)).$$

In the example (3.15), we find $\dim(\text{row}(A)) = \dim(\text{col}(A)) = 2$, is this a coincidence? *The fundamental theorem of linear algebra* gives this answer:

Theorem 3.4 The row space and column space both have the **same** dimension r.

We call $\dim(\mathcal{C}(A))$ as *column rank*; $\dim(\mathcal{R}(A))$ as *row rank*.

In brevity, **column rank=row rank= rank**, i.e.,

$$\dim(\mathcal{C}(A)) = \dim(\mathcal{R}(A)) = \text{rank}(A), \quad \text{for matrix } A$$

Let's discuss an example to have an idea of proving it.

■ **Example 3.17**

$$A=\begin{bmatrix}1&3&3&4\\2&6&9&7\\-1&-3&3&4\end{bmatrix}\xrightarrow{\text{row transform}}U=\begin{bmatrix}1&3&0&-1\\0&0&1&1\\0&0&0&0\end{bmatrix}$$

We notice that **column rank of $A=2$** and **column rank of $U=2$**.

Why do they have the same **column space dimension**?

Wrong reason: A and U has the same column space. This is false. For example, the first column of A is $\begin{pmatrix}1\\2\\-1\end{pmatrix}\notin\text{col}(U)$. The column spaces of A and U are **different**, but the dimension of them are **equal**.

Right reason: $Ax=0$ iff. $Ux=0$. The same combinations of the columns are zero (or nonzero) for A and U.

In other words, the r pivot columns (for both A and U) are independent; the $(n-r)$ free columns (for both A and U) are dependent.

For example, for U, column 1 and 3 are ind.(pivot columns); column 2 and 4 are dep.(free columns).

For 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 3.6 Suppose matrix A is converted into B by row transformation. If a set of columns of A are ind. then so are the corresponding columns of B.

Proof. Assume $A=\begin{bmatrix}a_1 \mid \dots \mid a_n\end{bmatrix}, B=\begin{bmatrix}b_1 \mid \dots \mid b_n\end{bmatrix}$.

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 $\hat{A} = \begin{bmatrix} a_1 & \cdots & a_k \end{bmatrix}$ and $\hat{B} = \begin{bmatrix} b_1 & \cdots & b_k \end{bmatrix}$.

1. Notice that $\hat{A}$ could be converted into $\hat{B}$ by row transformation.

   Hence $\hat{A}x = 0$ and $\hat{B}x = 0$ has the same solutions.

2. On the other hand, $a_1, a_2, \dots, a_k$ are ind. columns.

   Hence $\hat{A}x = 0$ has the only zero solution.

Combining (1) and (2), $\hat{B}x = 0$ 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., A and U has the same column space dimension:

Proposition 3.7 Row transformation doesn't change the column rank.

Proof. Assume $A \xrightarrow{\text{row transform}} B$.

Suppose $\dim(\mathcal{C}(A)) = r$, then we pick r ind. columns of A. After row transformation, they are still ind. Hence $\dim(\mathcal{C}(B)) \geq r = \dim(\text{col}(A))$.

Since row transformations are invertible, we get $B \xrightarrow{\text{row transform}} A$. Similarly, $\dim(\mathcal{C}(A)) \geq \dim(\mathcal{C}(B))$.

Hence $\dim(\mathcal{C}(A)) = \dim(\mathcal{C}(B))$. ■

Combining proposition (3.5) and (3.7), we can proof theorem (3.4):

Proof for theorem 3.4. Assume $A \xrightarrow{\text{row transform}} U$(rref).

- Proposition (3.5) $\implies \dim(\mathcal{R}(A)) = \dim(\mathcal{R}(U))$.
- Proposition (3.7) $\implies \dim(\mathcal{C}(A) = \dim(\mathcal{C}(U))$.
- Notice that $\dim(\mathcal{R}(U))$ denotes the number of pivots, $\dim(\mathcal{C}(U))$ denotes the number of pivot columns. Obviously, $\dim(\mathcal{R}(U)) = \dim(\mathcal{C}(U))$.

Hence $\dim(\mathcal{R}(A)) = \dim(\mathcal{C}(A))$. ■

R $\dim(\mathcal{R}(U))$ essentially denotes the number of "real" equations. $\dim(\mathcal{C}(U))$ denotes the number of "real" variables.

So Theorem 3.4 implies that the number of "real" equations should equal to the number of "real" variables.

3.3.3.1. What is the null space dimension?

Assume the system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ has n variables.

Proposition 3.8 For matrix $\boldsymbol{A}$,

$$\text{rank}(\boldsymbol{A}) + \text{rank}(N(\boldsymbol{A})) = n.$$

Proof. **Number of pivot varibales + Number of free variables = $\boldsymbol{n}$.** ■

Note that $\boldsymbol{b} \in \text{col}(\boldsymbol{A})$ iff. $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ for some $\boldsymbol{x}$.

Hence $\mathcal{C}(\boldsymbol{A})$ denotes all possible vectors in the form $\boldsymbol{A}\boldsymbol{x}$. Hence we call $\mathcal{C}(\boldsymbol{A})$ as "**range space**" of $\boldsymbol{A}$, which is denoted as range($\boldsymbol{A}$).

Equivalently, we have $\dim(\text{range}(\boldsymbol{A})) + \dim(N(\boldsymbol{A})) = n$.

Proposition 3.9 If $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ has at least one solution, then $\text{rank}(\boldsymbol{A}) = \text{rank}(\begin{bmatrix} \boldsymbol{A} & \boldsymbol{b} \end{bmatrix})$.

■ **Example 3.18** Suppose $\boldsymbol{A} = \begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix}$. If $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ has at least one solution, then $\text{rank}(\begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix}) = \text{rank}(\begin{bmatrix} a_1 & a_2 & a_3 & b \end{bmatrix})$. ■

Proofoutline.

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \Longleftrightarrow \boldsymbol{b} \in \mathcal{C}(\boldsymbol{A})$$

Hence $\boldsymbol{b}$ is the linear combination of columns of $\boldsymbol{A}$. Adding one more column $\boldsymbol{b}$ into $\boldsymbol{A}$ doesn't change the dimension of $\mathcal{C}(\boldsymbol{A})$. Hence $\text{rank}(\boldsymbol{A}) = \text{rank}(\begin{bmatrix} \boldsymbol{A} & \boldsymbol{b} \end{bmatrix})$. ■

Proposition 3.10 If $\text{rank}(\boldsymbol{A}) \le n - 1$ for $m \times n$ matrix $\boldsymbol{A}$, then $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ has **no solution** or **infinitely many solutions**.

Proofoutline.

$$\dim(\mathcal{C}(\boldsymbol{A})) + \dim(N(\boldsymbol{A})) = n \implies \dim(N(\boldsymbol{A})) \geq 1$$

So we have special solutions for $\boldsymbol{Ax} = \boldsymbol{b}$. For the particular solution, if doesn't exist, then we have no solution, otherwise we have infinitely many solutions. ∎

Definition 3.12 [Full Rank] For $m \times n$ matrix $\boldsymbol{A}$, if $\text{rank}(\boldsymbol{A}) = \min(m,n)$, then we say $\boldsymbol{A}$ is **full rank**.

Theorem 3.5 For $n \times n$ matrix $\boldsymbol{A}$, it is invertible iff. $\text{rank}(\boldsymbol{A}) = n$.

Proof. Sufficiency. Assume $\text{rank}(\boldsymbol{A}) = r < n$, then by row transformation, we can convert $\boldsymbol{A}$ into $\boldsymbol{U} := \begin{bmatrix} \boldsymbol{I}_r & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}$(rref), where $\boldsymbol{B} \in \mathbb{R}^{r \times (n-r)}$. We can represent this process in matrix notation:

$$\boldsymbol{PA} = \boldsymbol{U} := \begin{bmatrix} \boldsymbol{I}_r & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix},$$

where $\boldsymbol{P}$ is the product of row transformation matrices, which is obviously invertible.

Since $\boldsymbol{A}$ is invertible, we let $\boldsymbol{A}^{-1} = \begin{bmatrix} \boldsymbol{C}_1 \\ \boldsymbol{C}_2 \end{bmatrix}_{(r+(n-r)) \times n}$. It follows that

$$\boldsymbol{P} = \boldsymbol{PI}_n = \boldsymbol{P}(\boldsymbol{AA}^{-1}) = (\boldsymbol{PA})\boldsymbol{A}^{-1} = \boldsymbol{UA}^{-1} = \begin{bmatrix} \boldsymbol{I}_r & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{C}_1 \\ \boldsymbol{C}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{C}_1 + \boldsymbol{BC}_2 \\ \boldsymbol{0} \end{bmatrix}.$$

Since $\boldsymbol{P}$ has $(n-r)$ zero rows as shown above, it is not invertible, which is a contradiction.

Necessity. If $\boldsymbol{A}$ is full rank, then it has n pivots, then by row transformation we can convert it into $\boldsymbol{I}$(rref). We can represent this process in matrix notation:

$$\boldsymbol{PA} = \boldsymbol{I}$$

where $\boldsymbol{P}$ is the product of row transformation matrix. Hence $\boldsymbol{P}$ is the left inverse of $\boldsymbol{A}$, $\boldsymbol{A}$ is invertible. ■

3.3.3.2. Matrices of rank 1

▪ **Example 3.19**

$$\boldsymbol{A} = \begin{bmatrix} 2 & 1 & 1 \\ 4 & 2 & 2 \\ 8 & 4 & 4 \\ -2 & -1 & -1 \end{bmatrix} \xlongequal{\boldsymbol{v}^{\mathrm{T}}=\begin{bmatrix}2 & 1 & 1\end{bmatrix}} \begin{bmatrix} \boldsymbol{v}^{\mathrm{T}} \\ 2\boldsymbol{v}^{\mathrm{T}} \\ 4\boldsymbol{v}^{\mathrm{T}} \\ -\boldsymbol{v}^{\mathrm{T}} \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 4 \\ -1 \end{bmatrix} \boldsymbol{v}^{\mathrm{T}} \xlongequal{\boldsymbol{u}=\begin{bmatrix}1 & 2 & 4 & -1\end{bmatrix}^{\mathrm{T}}} \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$$

Here $\operatorname{rank}(\boldsymbol{A}) = 1$.

Proposition 3.11 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:

Proof. We set

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{c}_1 \\ \boldsymbol{c}_2 \\ \vdots \\ \boldsymbol{c}_n \end{bmatrix},$$

where $\boldsymbol{c}_i$ is row vector. WLOG, we set $\boldsymbol{c}_1 \neq \boldsymbol{0}$ and $\boldsymbol{c}_1 = \begin{pmatrix} a_1 b_1 & a_1 b_2 & \dots & a_1 b_n \end{pmatrix}$, where $a_1 \neq 0$, and $b_i (i = 1, \dots, n)$ are not all zero.

Since $\operatorname{rank}(\boldsymbol{A}) = 1$, we have $\dim(\mathcal{R}(\boldsymbol{A})) = 1$. Hence other $\boldsymbol{c}_i$ are dep. with $\boldsymbol{c}_1$. So we set

$$b_i = \frac{a_i}{a_1} \text{ for } i = 1, 2, \dots, n.$$

Thus we construct the form of $\boldsymbol{A}$:

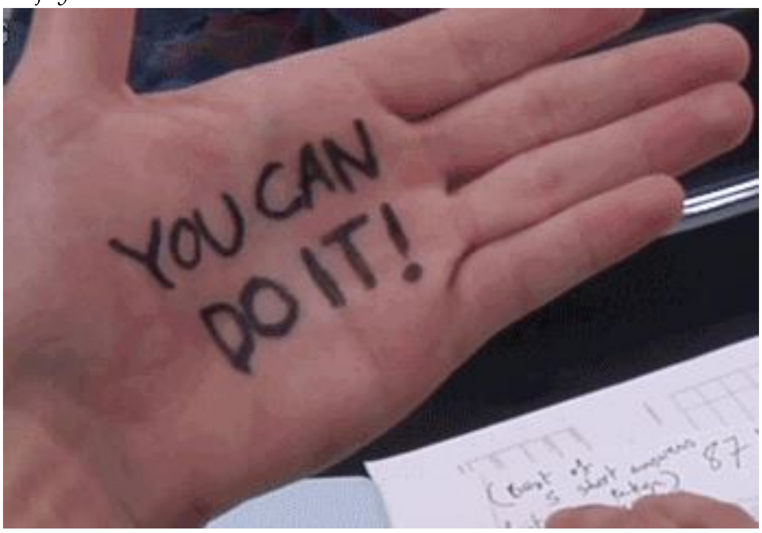
$$\boldsymbol{A} = \begin{bmatrix} a_1b_1 & a_1b_2 & \dots & a_1b_n \\ a_2b_1 & a_2b_2 & \dots & a_2b_n \\ \vdots & \vdots & & \vdots \\ a_nb_1 & a_nb_2 & \dots & a_nb_n \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \begin{bmatrix} b_1 & b_2 & \dots & b_n \end{bmatrix}$$

■

Question: What about the form of rank 2?

Answer: By SVD, it has the form $\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + \boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}}$.

Enjoy Your Midterm!



3.4. 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* U of A.

   (b) Compute all solutions of $Ax = b$, where $b = \begin{bmatrix} 1 & 1 & 2 \end{bmatrix}^{T}$.

   (c) Compute all solutions of $Ax = b$, where $b = \begin{bmatrix} b_1 & b_2 & b_3 \end{bmatrix}^{T}$.

   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) $\text{span}\left\{ \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ -2 \\ 4 \end{pmatrix}, \begin{pmatrix} -3 \\ 3 \\ 6 \end{pmatrix} \right\}$;

   (b) $\text{col}(A)$, where $A = \begin{bmatrix} 1 & -2 & 3 & 2 \\ -1 & 2 & -2 & -1 \\ 2 & -4 & 5 & 3 \end{bmatrix}$;

   (c) $N(B)$, where $B = \begin{bmatrix} 1 & 3 & 2 \\ 2 & 1 & 4 \\ 4 & 7 & 8 \end{bmatrix}$;

   (d) $\text{span}\{(x-2)(x+2), x^2(x^4-2), x^6-8\}$;

   (e) $\text{span}\{5, \cos 2x, \cos^2 x\}$ 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 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 $Ax = b$? 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(3.3)

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 $\boldsymbol{U} \cap \boldsymbol{V}=\{\mathbf{0}\}$, 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(3.2). You should prove it by assuming theorem(3.2) is unknown. You may check the proposition(3.2) as hint.

Chapter 4

Midterm

4.1. Sample Exam

DURATION OF EXAMINATION: 2 hours in-class

This examination paper includes 6 *pages and* 6 *problems. You are responsible for ensuring that your copy of the paper is complete. Bring any discrepancy to the attention of your invigilator.*

1. **(30 points)** *Solving a linear system of equations*

   For a real number c, consider the linear system:

$$x_1 + x_2 + cx_3 + x_4 = c \tag{4.1}$$

$$-x_2 + x_3 + 2x_4 = 0 \tag{4.2}$$

$$x_1 + 2x_2 + x_3 - x_4 = -c \tag{4.3}$$

   do the following:

   (a) Write out the *coefficient matrix* of the system.

   (b) Write out the *augmented matrix* for this system and calculate its *row-reduced echelon form.*

   (c) Write out the complete set of solutions in *vector form.*

   (d) What is the *rank* of the coefficient matrix $\boldsymbol{A}$? Justify your answer.

(e) Find a *basis* of the subspace of solutions when $c = 0$.

2. **(20 points)** *Vector space*

Find a *basis* for each of the following spaces.

- Space of $n \times n$ *skew symmetric matrices* (i.e. those matrix satisfying $\boldsymbol{A} = -\boldsymbol{A}^{\mathrm{T}}$)

- The space of all *polynomials* of the form $ax^2 + bx + 2a + 3b$, where $a, b \in \mathbb{R}$.

- $\operatorname{span}\{x - 1, x + 1, 2x^2 - 2\}$.

3. **(15 points)** *Matrix multiplications*

Prove the following statements:

(a) Define the set of $n \times n$ *diagonol matrices* to be κ. Prove that for a diagonal matrix $\boldsymbol{D}$ with *distinct* elements (i.e. $\boldsymbol{D}_{ii} \neq \boldsymbol{D}_{jj}, \forall i \neq j$), the set $\{\boldsymbol{A} \in \mathbb{R}^{n \times n} | \boldsymbol{AD} = \boldsymbol{DA}\}$ is exactly κ.

(b) If an $n \times n$ matrix $\boldsymbol{A}$ satisfies $\boldsymbol{AB} = \boldsymbol{BA}$ for any $n \times n$ matrix $\boldsymbol{B}$, then $\boldsymbol{A}$ must be of the form $c\boldsymbol{I}$, where c is a scalar.

4. **(10 points)** *Matrix Inverse*

(a) Compute the inverse of the matrix $\begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}$.

(b) Compute the inverse of the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ if exists. When does the inverse of the matrix exist?

5. **(15 points)** *Matrix rank*

(a) Suppose $\boldsymbol{u} \in \mathbb{R}^{n \times 1}$ satisfies $\|\boldsymbol{u}\| = 1$. What is the rank of the matrix $\boldsymbol{I} - \boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}$?

(b) Suppose $\boldsymbol{u} \in \mathbb{R}^{n \times 1}$ satisfies $\|\boldsymbol{u}\| = 1$. Define $\boldsymbol{P} = \boldsymbol{I} - \boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}$. What is the rank of $\boldsymbol{P}^2$? What about $\boldsymbol{P}^5$?

(c) Suppose $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{n \times 1}$. What is the rank of the matrix $\boldsymbol{I} - \boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}$?

6. **(20 points)**

State your answers. No justifications are required.

(a) We know $a^2 - b^2 = (a+b)(a-b)$, where $a, b \in \mathbb{R}$. When $\boldsymbol{A}, \boldsymbol{B}$ are square matrices, can we represent $\boldsymbol{A}^2 - \boldsymbol{B}^2$ by only $(\boldsymbol{A}+\boldsymbol{B})(\boldsymbol{A}-\boldsymbol{B})$?

(b) True or False: If $\boldsymbol{A}$ and $\boldsymbol{B}$ are *invertible*, then $\boldsymbol{A}+\boldsymbol{B}$ is also *invertible*.

(c) True or False: The set of all *real-valued* functions on $\mathbb{R}$ such that $f(1) = 0$ is a *vector space*.

(d) True or False: The product of two *invertible* $n \times n$ matrices is *invertible*

(e) True or False: If two matrices have the same *reduced row echelon form*, then they have the same *column space*.

(f) True or False: If two columns of the square $\boldsymbol{A}$ are the same, then $\boldsymbol{A}$ *cannot* be invertible.

(g) True or False: For an $m \times n$ matrice $\boldsymbol{A}$, $\text{rank}(\boldsymbol{A}) + \dim(\text{row}(\boldsymbol{A})) = n$.

4.2. Midterm Exam

DURATION OF EXAMINATION: 2 hours in-class

This examination paper includes 6 pages and 6 problems. You are responsible for ensuring that your copy of the paper is complete. Bring any discrepancy to the attention of your invigilator.

1. **(30 points)** *Solving a linear system of equations*

   For the system

$$x - y + 3z = 1 \tag{4.4}$$

$$y = -2x + 5 \tag{4.5}$$

$$9z - x - 5y + 7 = 0 \tag{4.6}$$

   do the following:

   (a) Write the system in the matrix form

$$\boldsymbol{Ax} = \boldsymbol{b} \text{ for } \boldsymbol{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

   (b) Write out the *augmented matrix* for this system and calculate its *row-reduced echelon form*.

   (c) Write out the complete set of solutions (if they exist) in *vector form* using parameters if needed.

   (d) Calculate the *inverse* of the *coefficient matrix* $\boldsymbol{A}$ you found in part (a), if it exists, or show that $\boldsymbol{A}^{-1}$ doesn't exist.

   (e) What is the rank of matrix $\boldsymbol{A}$? Justify your answer.

2. **(20 points)** *Vector space*

Let V be the subspace of $\mathbb{R}^4$ given by all solutions to the equation $2x_1 - x_2 + 3x_3 = 0$.

(a) Give the set of all solutions in terms of *free variables*.

(b) What is the dimension of V? Justify your answer.

(c) Find a 4 by 3 matrix $\boldsymbol{A}$ such that the *column space* of $\boldsymbol{A}$ is equal to V.

(d) Find a 1 by 4 matrix $\boldsymbol{B}$ such that the *null space* of $\boldsymbol{B}$ is equal to V.

3. **(15 points)** *Matrix multiplications*

   If possible, find 3 by 3 matrices $\boldsymbol{B}$ such that

   (a) $\boldsymbol{BA} = 2\boldsymbol{A}$ for every $\boldsymbol{A}$.

   (b) $\boldsymbol{BA} = 2\boldsymbol{B}$ for every $\boldsymbol{A}$.

   (c) $\boldsymbol{BA}$ has the *first* and *last* rows of $\boldsymbol{A}$ reversed.

   (d) $\boldsymbol{BA}$ has the *first* and *last* columns of $\boldsymbol{A}$ reversed.

4. **(10 points)** *Matrix Inverse*

For an $m \times n$ matrix $\boldsymbol{A}$, we say an $n \times m$ matrix $\boldsymbol{C}$ is a *right inverse* of $\boldsymbol{A}$ if $\boldsymbol{AC} = \boldsymbol{I}_m$, where $\boldsymbol{I}_m$ is the $m \times m$ identity matrix.

(a) Prove that $\boldsymbol{A}$ has a *right inverse* if and only if $\boldsymbol{Ax} = \boldsymbol{b}$ has at least one solution for any $\boldsymbol{b} \in \mathbb{R}^m$. Prove that the rank of such $\boldsymbol{A}$ must be m.

(b) Compute a *right inverse* of the following matrices (if exists):

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 7\pi \end{pmatrix}$$

$$\boldsymbol{B} = \begin{pmatrix} 1 \\ 2 \\ 7\pi \end{pmatrix}$$

5. **(15 points)** *Matrix rank*

(a) For a square matrix $\boldsymbol{A}$, is $\mathrm{rank}(\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{A}) = \mathrm{rank}(\boldsymbol{A})$ always true? Justify your answer.

(b) Prove that for any m by n matrix $\boldsymbol{A}$, the null space of $\boldsymbol{A}$ and the null space of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ are the same.

(c) Prove that for any m by n matrix $\boldsymbol{A}$, $\mathrm{rank}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = \mathrm{rank}(\boldsymbol{A})$.

6. **(20 points)**

State your answers. No justifications are required.

(a) If $\boldsymbol{A} = \boldsymbol{A}^{\mathrm{T}}$ and $\boldsymbol{B} = \boldsymbol{B}^{\mathrm{T}}$ which of these matrices are certainly *symmetric*?

i. $\boldsymbol{A}^2 - \boldsymbol{B}^2$

ii. $(\boldsymbol{A} + \boldsymbol{B})(\boldsymbol{A} - \boldsymbol{B})$

iii. $\boldsymbol{ABA}$

iv. $\boldsymbol{ABAB}$

(b) Let $\boldsymbol{A}$ be a 5×8 matrix with rank equal to 5 and let $\boldsymbol{b}$ be any vector in $\mathbb{R}^5$. How many solutions does this system have?

(c) True or False: If two $n \times n$ matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are both *singular*, then $\boldsymbol{A} + \boldsymbol{B}$ is also *singular*.

(d) True or False: The set of $n \times n$ matrices with rank no more than $r(r \leq n)$ is a vector space.

(e) True or False: The set of all *real-valued* functions on $\mathbb{R}$ such that $f(1) = 1$ is a vector space.

Chapter 5

Week4

5.1. Friday

5.1.1. 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 5.1 [linear transformation] A transformation L assigns an output $T(\boldsymbol{v})$ to each inpout vector $\boldsymbol{v}$ in $\boldsymbol{V}$.

The transformation $L(\cdot)$ is siad 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 and scalars α, β.

Key Observation: If the input is $\boldsymbol{v} = \boldsymbol{0}$, the output must be $L(\boldsymbol{v}) = \boldsymbol{0}$.

5.1.1.1. The idea of linear transformation

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 $\boldsymbol{n}$ 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 $a_1, \dots, a_n$.

5.1.1.2. Matrix defines linear transformation

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":

Definition 5.2 [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:

$$\text{Range}(L) = \{L(\boldsymbol{x}) : x \in \boldsymbol{V}\}$$

Sometimes we also use notation $\text{Im}(L)$ to express the same thing. ■

The range corresponds to the column space. If $L(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$, we have $\text{Range}(L) = \mathcal{C}(\boldsymbol{A})$.

Definition 5.3 [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:

$$L(\cdot) : \begin{cases} N(\boldsymbol{A}) \mapsto \{\boldsymbol{0}\} \\ \boldsymbol{V} \mapsto \text{col}(\boldsymbol{A}) \end{cases}$$

5.1.2. Example: differentiation

Key idea of this section:

Suppose we know $L(\boldsymbol{v}_1), \dots, L(\boldsymbol{v}_n)$ for the basis vectors $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$,Then the linearity property produces $L(\boldsymbol{v})$ for every other input vector $\boldsymbol{v}$

Reason: Every $\boldsymbol{v}$ has a unique combination $c_1\boldsymbol{v}_1 + \dots + c_n\boldsymbol{v}_n$ of the basis vector $\boldsymbol{v}_i$. Suppose L is a linear transformation, then $L(\boldsymbol{v})$ must be the **same combination** $c_1L(\boldsymbol{v}_1) + \dots + c_nL(v_n)$ 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 5.1** 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 $\boldsymbol{A}$ agrees with transforming by L**. The derivative of $\boldsymbol{v} = a + bx + cx^2 + dx^3$ is $L(\boldsymbol{v}) = b + 2cx + 3dx^2$. The same numbers $b, 2c, 3d$ appear when we multiply by matrix $\boldsymbol{A}$:

$$\textbf{\textit{Take the derivative}} \quad \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} b \\ 2c \\ 3d \end{bmatrix}.$$

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 $\boldsymbol{v}$ and $L(\boldsymbol{v})$. 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 $\boldsymbol{v}$ is the same as multiplying matrix $\boldsymbol{A}$ by its coordinate vector. ■

5.1.2.1. 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 $\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3$:

$$\textbf{Integration is } L^{-1} \qquad \int_0^x 1\,\mathrm{d}x = x, \quad \int_0^x x\,\mathrm{d}x = \frac{1}{2}x^2, \quad \int_0^x x^2\,\mathrm{d}x = \frac{1}{3}x^3.$$

By linearity, the integral of $\boldsymbol{w} = B + Cx + Dx^2$ is $L^{-1}(\boldsymbol{w}) = Bx + \frac{1}{2}Cx^2 + \frac{1}{3}Dx^3$. 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:

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} \qquad A^{-1} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix}$$

I want to call this matrix A^{-1}, though rectangular matrices don't have inverses. Note that A^{-1} is the **right inverse** of matrix A! (Do you remember the definition that shown in mid-term?)

$$AA^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{but} \quad A^{-1}A = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

This is reasonable. If you integrate a function and then differentiate, you get back to the start. Hence $AA^{-1} = I$. 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 V be a vector space of polynomials with degree ≤ 3. Its basis is given by $E = \{1, x, x^2, x^3\}$. Any $v \in V$ 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 5.4 [coordinate vector] Let $\boldsymbol{V}$ be a vector space of dimension n and let $B = \{v_1, v_2, \ldots, v_n\}$ 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 $\alpha_1, \ldots, \alpha_n$ 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: $v = \begin{bmatrix} v_1 & v_2 & \ldots & v_n \end{bmatrix} \times [v]_B$. ■

More specifically, the linear transformation of vectors is essentially the matrix multiplication of the corresponding coordinate vectors:

Theorem 5.1 Let $E = \{v_1, \ldots, v_n\}$ be a basis for $\boldsymbol{V}$; $F = \{w_1, \ldots, w_m\}$ 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 5.1 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

$$[L(v)]_E = \boldsymbol{A}[v]_E$$

5.1.3. Basis Change

Basis Change is essentially matrix multiplication. Suppose $L : \boldsymbol{V} \mapsto \boldsymbol{V}$. $E = \{v_1, \ldots, v_n\}$ is a basis for $\boldsymbol{V}$, $F = \{u_1, \ldots, u_n\}$ is another basis for $\boldsymbol{V}$. Then vector $u_1, \ldots, u_n$ could be expressed by vectors $v_1, \ldots, v_n$. So we set

$$\begin{aligned}
u_1 &= s_{11}v_1 + s_{12}v_2 + \cdots + s_{1n}v_n, \\
u_2 &= s_{21}v_1 + s_{22}v_2 + \cdots + s_{2n}v_n, \\
&\ldots \\
u_n &= s_{n1}v_1 + s_{n2}v_2 + \cdots + s_{nn}v_n.
\end{aligned}$$

We could write this system into matrix form:

$$(u_1, \ldots, u_n) = (v_1, \ldots, v_n) \begin{pmatrix} s_{11} & s_{12} & \ldots & s_{1n} \\ s_{21} & s_{22} & \ldots & s_{2n} \\ \vdots & \vdots & \ldots & \vdots \\ s_{n1} & s_{n2} & \ldots & s_{nn} \end{pmatrix}.$$

We set $\boldsymbol{S} = (s_{ij})$. Hence we obtain:

$$(u_1, \dots, u_n) = (v_1, \dots, v_n)\boldsymbol{S}. \tag{5.1}$$

You should **prove it by yourself** that $\boldsymbol{S}$ is invertible. Hence we have:

$$(u_1, \dots, u_n)\boldsymbol{S}^{-1} = (v_1, \dots, v_n). \tag{5.2}$$

We can express linear transformation in terms of different basis. Given any vector $x \in \boldsymbol{V}$, we want to study the relationship between $L(x)$ and $[x]_F$:

$$\begin{aligned} L(x) &= \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix} \times [L(x)]_E \\ &= \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix} \times (\boldsymbol{A}[x]_E) \qquad \leftarrow \text{due to corollary (5.1)} \\ &= \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix} \boldsymbol{S}^{-1} \times (\boldsymbol{A}[x]_E) \end{aligned} \tag{5.3}$$

- We claim that $[x]_E = \boldsymbol{S}[x]_F$:

  For any vector $x \in \boldsymbol{V}$, we obtain:

  $$\begin{aligned} x &= \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix} \times [x]_E \\ &= \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix} \times [x]_F \\ &= \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix} \times \boldsymbol{S}[x]_F \end{aligned}$$

  Hence $[x]_E = \boldsymbol{S}[x]_F$.

Substituting $[x]_E = \boldsymbol{S}[x]_F$ into Eq.(5.3), we obtain:

$$L(x) = \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix} \boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}[x]_F$$

What do the following process mean? We know that given basis $E = \{v_1, \dots, v_n\}$, performing linear transformation on any vector x is just the same as matrix multiplication:

$$L(x) = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix} \times \boldsymbol{A}[x]_E$$

In summary,

1. The linear transformation is essentially postmultiplying matrix for the coordiante vector:

$$x = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix} \times [x]_E \quad \Longrightarrow \quad L(x) = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix} \times \boldsymbol{A}[x]_E$$

2. If we change another basis $F = \{u_1, \dots, u_n\}$, we must change $\boldsymbol{A}$ into $\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}$:

$$x = \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix} \times [x]_F \quad \Longrightarrow \quad L(x) = \begin{bmatrix} u_1 & u_2 & \dots & u_n \end{bmatrix} \times \boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}[x]_F$$

It suffices to define $\boldsymbol{B} := \boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}$, The matrix $\boldsymbol{B}$ is said to be **similar** to $\boldsymbol{A}$.

Definition 5.5 [Similar] Let $\boldsymbol{A}, \boldsymbol{B}$ be $n \times n$ matrix. If there exists invertible $n \times n$ matrix $\boldsymbol{S}$ such that $\boldsymbol{B} = \boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}$, then we say that $\boldsymbol{A}$ is **similar** to $\boldsymbol{B}$. ■

5.1.4. Determinant

The determinat of a **square matrix** is a single number, which contains many amazing amount of information about the matrix. It has four major uses:

The determinant is zero if and only if the matrix has no inverse.

It can be used to calculate the area or volumn of a box. $|\det(\boldsymbol{A})|$ is the volume of the parallelepiped $\mathcal{P} = \{y = \sum_{i=1}^{m} \alpha_i \boldsymbol{a}_i \mid \alpha_i \in [0, 1]\}$:

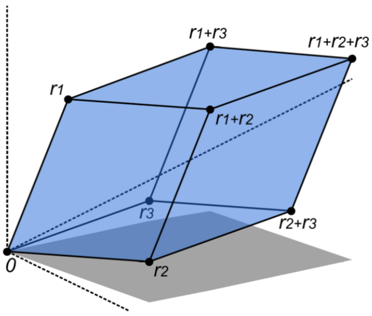


Figure 5.1: The parallelepiped $\mathcal{P} = \{y = \sum_{i=1}^{3} \alpha_i \boldsymbol{a}_i \mid \alpha_i \in [0,1]\}$, where r_1, r_2, r_3 are $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ on $\mathbb{R}^3$

The product of all the pivots $= (\pm 1) \times$ **the determinant.** For a 2 by 2 matrix $\boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, the pivots are a and $d - (\frac{c}{a})b$. The product of pivots is the determinant:

$$\textbf{Product of pivots} \quad a(d - \tfrac{c}{a}b) = ad - bc \quad \textbf{which is } \det \boldsymbol{A}$$

Compute determinants to find $\boldsymbol{A}^{-1}$ and $\boldsymbol{A}^{-1}\boldsymbol{b}$. (Cramer's Rule).

5.1.4.1. The properties of the Determinant

We don't intend to define the determinant directly by its formulas. It's better to start with its properties. These properties are simple, but they prepare for the formulas.

R Brackets for the matrix, straight bars for its determinant. For example,

$$\text{The determinant of } \begin{bmatrix} a & b \\ c & d \end{bmatrix} \text{ is } \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

The determinant is written in two ways, $\det \boldsymbol{A}$ or $|\boldsymbol{A}|$.

We will introduce three basic properties, then we will show how properties $1-3$ derive other properties.

1. **The determinant of the n by n identity matrix is 1.**

$$\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1 \qquad \text{and} \qquad \begin{vmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{vmatrix} = 1.$$

2. **The determianant changes sign when two rows are exchanged.** (sign reversal)

$$\text{Check: } \begin{vmatrix} c & d \\ a & b \end{vmatrix} = - \begin{vmatrix} a & b \\ c & d \end{vmatrix} \qquad \text{(both sides equal } bc - ad\text{).}$$

3. **The determinant is a linear function of each row separately.** (all other rows stay fixed).

$$\textbf{multiply row 1 by any number } \boldsymbol{t} \qquad \begin{vmatrix} ta & tb \\ c & d \end{vmatrix} = t \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

$$\textbf{Add row 1 of } \boldsymbol{A} \textbf{ to row 1 of } \boldsymbol{B}\textbf{:} \qquad \begin{vmatrix} a_1 + a_2 & b_1 + b_2 \\ c & d \end{vmatrix} = \begin{vmatrix} a_1 & b_1 \\ c & d \end{vmatrix} + \begin{vmatrix} a_2 & b_2 \\ c & d \end{vmatrix}$$

Note that this rule **deos not** mean $\det(\boldsymbol{A} + \boldsymbol{B}) = \det \boldsymbol{A} + \det \boldsymbol{B}$.

Note that this rule **does not** mean $\det(t\boldsymbol{A}) = t \det(\boldsymbol{A})$.

Actually, $\det(t\boldsymbol{A}) = t^n \det \boldsymbol{A}$. This is reasonable. Imagining that expanding a rectangle by 2, its area will increase by 4. Expand an $n-$dimensional box by t and its volumn will increase by t^n.

Pay special attention to property $1 \sim 3$. They completely determine the $\det \boldsymbol{A}$. We could stop here to find a formula for determinants. But before that we prefer to derive other properties that follow directly from the first three:

4. **If two rows of $\boldsymbol{A}$ are equal, then $\det \boldsymbol{A} = \boldsymbol{0}$.**

$$\text{Check 2 by 2: } \begin{vmatrix} a & b \\ a & b \end{vmatrix} = 0.$$

Property 4 follows from Property 2.

Proofoutline. Exchange the two equal row. The determinant $\boldsymbol{D}$ is supposed to change sign. But also the matrix is not changed, so we have $-\boldsymbol{D} = \boldsymbol{D} \implies \boldsymbol{D} = 0$. ■

5. **Adding a constant multiple of a row to another row doesn't change $\det \boldsymbol{A}$.**

$$\begin{vmatrix} a+lc & b+ld \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} lc & ld \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + l \begin{vmatrix} c & d \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = \det \boldsymbol{A}$$

Conclusion: *The determinant is not changed by the usual elimination step from $\boldsymbol{A}$ to* $\boldsymbol{U}$. Since every row exchange reverses the sign, we have $\det \boldsymbol{A} = \pm \det \boldsymbol{U}$.

6. **If $\boldsymbol{A}$ is triangular, then $\det \boldsymbol{A}$ = product of diagonal entries.**

$$\textbf{Triangular} \quad \begin{vmatrix} a & b \\ 0 & d \end{vmatrix} = ad \quad \text{and also} \quad \begin{vmatrix} a & 0 \\ c & d \end{vmatrix} = ad$$

Suppose all diagonal entries of $\boldsymbol{A}$ are nonzero. We do Gaussian Elimination to convert $\boldsymbol{A}$ into diagonal matrix:

$$\det \begin{bmatrix} a_{11} & & & 0 \\ & a_{22} & & \\ & & \ddots & \\ 0 & & & a_{nn} \end{bmatrix} = a_{11}a_{22}\dots a_{nn}.$$

Factor a_{11} from the first row by property 3; then factor a_{22} from the second row;....... Finally the determinant is $a_{11} \times a_{22} \times a_{33} \dots \times a_{nn} \times \det \boldsymbol{I} = a_{11} \times a_{22} \times a_{33} \dots \times a_{nn}$.

7. $\det(\boldsymbol{AB}) = \det(\boldsymbol{A})\det(\boldsymbol{B})$.

Proof.

- If $|\boldsymbol{B}|$ is zero, it's easy to verify that $\boldsymbol{B}$ is singular, then $\boldsymbol{AB}$ is singular. Thus $\det(\boldsymbol{AB}) = 0 = \det(\boldsymbol{A})\det(\boldsymbol{B})$.

- Suppose $|\boldsymbol{B}|$ is not zero, and $\boldsymbol{A}, \boldsymbol{B}$ is $n \times n$ matrix. Consider the ratio $D(\boldsymbol{A}) = \frac{|\boldsymbol{AB}|}{|\boldsymbol{B}|}$. *Check that this ratio has properties 1,2,3.* If so, $D(\boldsymbol{A})$ has to be the determinant, say, $|\boldsymbol{A}|$. Thus we have $|\boldsymbol{A}| = \frac{|\boldsymbol{AB}|}{\boldsymbol{B}}$:

  Property 1 *(Determinant of I)* If $\boldsymbol{A} = \boldsymbol{I}$, then the ratio becomes $D(\boldsymbol{A}) = \frac{|\boldsymbol{B}|}{|\boldsymbol{B}|} = 1$.

  Property 2 *(Sign reversal)* When two rows of $\boldsymbol{A}$ are exchanged, the same two rows of $\boldsymbol{AB}$ are also exchanged. Therefore $|\boldsymbol{AB}|$ changes sign and so does the ratio $\frac{|\boldsymbol{AB}|}{\boldsymbol{B}}$.

  Property 3 *(Linearity)* When row 1 of $\boldsymbol{A}$ is multiplied by t, so is row 1 of $\boldsymbol{AB}$. Thus the ratio is also increased by t. Thus we still have $|\boldsymbol{A}| = \frac{|\boldsymbol{AB}|}{\boldsymbol{B}}$. If we Add row 1 of $\boldsymbol{A}_1$ to row 1 of $\boldsymbol{A}_2$. Then row 1 of $\boldsymbol{A}_1\boldsymbol{B}$ also adds to row 1 of A_2B. By property three, determinants add. After dividing by $|\boldsymbol{B}|$, the ratios add. Hence we still have $|\boldsymbol{A}| = \frac{|\boldsymbol{AB}|}{\boldsymbol{B}}$.

  Conclusion: The ratio $D(\boldsymbol{A})$ has the same three properties that defines determinant, hence it equals $|\boldsymbol{A}|$. Hence we obtain the product rule $|\boldsymbol{AB}| = |\boldsymbol{A}||\boldsymbol{B}|$.

■

Immediately here follows a corollary:

Corollary 5.2

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

8. **The transpose $\boldsymbol{A}^{\mathrm{T}}$ has the same determinant as $\boldsymbol{A}$.**

$$\textbf{Transpose} \qquad \begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & c \\ b & d \end{vmatrix} \qquad \text{Both sides equal } ad - bc.$$

Proof.

- When $\boldsymbol{A}$ is singular, $\boldsymbol{A}^{\mathrm{T}}$ is also singular. Hence $|\boldsymbol{A}^{\mathrm{T}}| = |\boldsymbol{A}| = 0$.
- Otherwise $\boldsymbol{A}$ has LU decomposition $\boldsymbol{PA} = \boldsymbol{LU}$. Transposing both siders gives $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{P}^{\mathrm{T}} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{L}^{\mathrm{T}}$. By product rule we have

$$\det \boldsymbol{P} \det \boldsymbol{A} = \det \boldsymbol{L} \det \boldsymbol{U} \qquad \text{and} \qquad \det \boldsymbol{A}^{\mathrm{T}} \det \boldsymbol{P}^{\mathrm{T}} = \det \boldsymbol{U}^{\mathrm{T}} \det \boldsymbol{L}^{\mathrm{T}}.$$

  - Firstly, $\det \boldsymbol{L} = \det \boldsymbol{L}^{\mathrm{T}} = 1$. (By property 6, they both have 1's on the diagonal).
  - Secondly, $\det \boldsymbol{U} = \det \boldsymbol{U}^{\mathrm{T}}$. (By property 6, they have the same diagonal)
  - Thirdly, $\det \boldsymbol{P} = \det \boldsymbol{P}^{\mathrm{T}}$. (Verify by yourself that $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P} = \boldsymbol{I}$. Hence $|\boldsymbol{P}^{\mathrm{T}}||\boldsymbol{P}| = 1$. Since permutation matrix is obtained by exchanging rows of $\boldsymbol{I}$, the only possible value for determinant of permuation matrix is ± 1. Hence $\boldsymbol{P}$ and $\boldsymbol{P}^{\mathrm{T}}$ must both equal to 1 or both equal to -1).

  So $\boldsymbol{L}, \boldsymbol{U}, \boldsymbol{P}$ has the same determinants as $\boldsymbol{L}^{\mathrm{T}}, \boldsymbol{U}^{\mathrm{T}}, \boldsymbol{P}^{\mathrm{T}}$, Hence we have $\det \boldsymbol{A} = \det \boldsymbol{A}^{\mathrm{T}}$.

∎

5.2. Assignment Five

1. Prove the following properties of *similarity*:

   (a) Any square matrix $\boldsymbol{A}$ is *similar* to itself.

   (b) If $\boldsymbol{B}$ is *similar* to $\boldsymbol{A}$, then $\boldsymbol{A}$ is *similar* to $\boldsymbol{B}$.

   (c) If $\boldsymbol{A}$ is *similar* to $\boldsymbol{B}$ and $\boldsymbol{B}$ is *similar* to $\boldsymbol{C}$, then $\boldsymbol{A}$ is *similar* to $\boldsymbol{C}$.

2. Consider the linear operator

$$L\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 3x \\ x-y \end{bmatrix}$$

   on $\mathbb{R}^2$, use a *similarity transformation* to find the *matrix representation* with respect to the basis

$$B = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \end{bmatrix} \right\}$$

3. Let $\mathbb{R}[x]$ be the vector space of all real polynomials in x. Determine whether the following sets are subspaces of $\mathbb{R}[x]$. Justify your answer.

   (a) All polynomials $f(x)$ of degree ≥ 3.

   (b) All polynomials $f(x)$ satisfying $f(1) + 2f(2) = 1$.

   (c) All polynomials $f(x)$ satisfying $f(x) = f(1-x)$.

4. Let $V = \{a + bx + cy + dx^2 + exy + fy^2 | a,b,c,d,e,f \in \mathbb{R}\}$, where x,y are *variables*. Then V is just the set of all polynomials in x and y of degree two or less. One can show that V is a vector space in which the same way as we showed $\mathbb{P}_2$ is a vector space.

   Now consider the function

$$T : V \mapsto V \text{ by } T(f) = \frac{\partial f}{\partial x} - \frac{\partial f}{\partial y}$$

   where f denotes arbitrary vector in V.

   (a) Prove that T is a *linear transformation*.

(b) Find bases for $kernel(T)$.

5. Let S be the subspace of $C[a,b]$ spanned by e^x,xe^x and x^2e^x. Let D be the *differentiation operator* of S. Find the *matrix representation* of D with respect to $\{e^x, xe^x, x^2e^x\}$.

6. Suppose all vectors x in the unit square $0 \le x_1 \le 1, 0 \le x_2 \le 1$ are transformed to $\boldsymbol{Ax}$. ($\boldsymbol{A}$ is 2 by 2)

   (a) What's the shape of the *transformed* region (all $\boldsymbol{Ax}$)?

   (b) For which matrices $\boldsymbol{A}$ is that region a *square*?

7. (a) Show the column space of $\boldsymbol{AA}^{\mathrm{T}}$ and $\boldsymbol{A}$ are the same.

   (b) Show the rank of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}, \boldsymbol{AA}^{\mathrm{T}}, \boldsymbol{A}^{\mathrm{T}}, \boldsymbol{A}$ are the same.

Chapter 6

Week5

6.1. Tuesday

6.1.1. Formulas for Determinant

We want to use the **3 basic properties** to derive the formula for determinant:

1. **The determinant of the n by n identity matrix is 1.**

$$\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1 \qquad \text{and} \qquad \begin{vmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{vmatrix} = 1.$$

2. **The determianant changes sign when two rows are exchanged.** (sign reversal)

$$\text{Check: } \begin{vmatrix} c & d \\ a & b \end{vmatrix} = -\begin{vmatrix} a & b \\ c & d \end{vmatrix} \qquad \text{(both sides equal } bc - ad\text{)}.$$

3. **The determinant is a linear function of each row separately.** (all other rows stay fixed).

$$\textbf{multiply row 1 by any number } \boldsymbol{t} \qquad \begin{vmatrix} ta & tb \\ c & d \end{vmatrix} = t\begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

Add row 1 of A to row 1 of B: $$\begin{vmatrix} a_1+a_2 & b_1+b_2 \\ c & d \end{vmatrix} = \begin{vmatrix} a_1 & b_1 \\ c & d \end{vmatrix} + \begin{vmatrix} a_2 & b_2 \\ c & d \end{vmatrix}$$

Although we derive the formula for $\det \boldsymbol{A}$ is $\det \boldsymbol{A} = \pm \prod_i \text{pivots}_i$ (product of pivots), it is not **explicit**. We begin some example to show how to derive the explicit formula for determinant.

■ **Example 6.1** To derive the formula for determinant, let's start with $n = 2$.

Given $\boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, our goal is to get $\det(\boldsymbol{A}) = ad - bc$.

We can break each row into two simpler rows:

$$\begin{vmatrix} a & b \end{vmatrix} = \begin{vmatrix} a & 0 \end{vmatrix} + \begin{vmatrix} 0 & b \end{vmatrix} \quad \text{and} \quad \begin{vmatrix} c & d \end{vmatrix} = \begin{vmatrix} c & 0 \end{vmatrix} + \begin{vmatrix} 0 & d \end{vmatrix}$$

Now apply property 3, first in row 1(with row 2 fixed) and then in row 2(with row 1 fixed):

$$\begin{aligned} \begin{vmatrix} a & b \\ c & d \end{vmatrix} &= \begin{vmatrix} a & 0 \\ c & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & d \end{vmatrix} \\ &= \begin{vmatrix} a & 0 \\ c & 0 \end{vmatrix} + \begin{vmatrix} a & 0 \\ 0 & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & 0 \end{vmatrix} + \begin{vmatrix} 0 & b \\ 0 & d \end{vmatrix} \end{aligned}$$

The last line has $2^2 = 4$ determinants. The first and fourth are zero since their rows are **dep.** (one row is a multiple of the other row.) We left two terms to compute:

$$\begin{vmatrix} a & 0 \\ 0 & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & 0 \end{vmatrix} = ad \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} + bc \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = ad - bc$$

The permutation matrices $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ have determinant $+1$ or -1. ■

■ **Example 6.2** Now we try $n = 3$. Each row splits into 3 simpler rows such as $\begin{bmatrix} a_{11} & 0 & 0 \end{bmatrix}$.

Hence $\det \boldsymbol{A}$ will split into $3^3 = 27$ simple determinants. For simple determinant, if one column has two nonzero entries, (For example, $\begin{bmatrix} a_{11} & 0 & 0 \\ a_{21} & 0 & 0 \\ 0 & 0 & a_{33} \end{bmatrix}$), then its determinant will be zero.

Hence we only need to foucus on the matrix that **the nonzero terms come from defferent columns**:

$$\begin{aligned}
\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} &= \begin{vmatrix} a_{11} & & \\ & a_{22} & \\ & & a_{33} \end{vmatrix} + \begin{vmatrix} & a_{12} & \\ & & a_{23} \\ a_{31} & & \end{vmatrix} + \begin{vmatrix} & & a_{13} \\ a_{21} & & \\ & a_{32} & \end{vmatrix} \\
&+ \begin{vmatrix} a_{11} & & \\ & & a_{23} \\ & a_{32} & \end{vmatrix} + \begin{vmatrix} & a_{12} & \\ a_{21} & & \\ & & a_{33} \end{vmatrix} + \begin{vmatrix} & & a_{13} \\ & a_{22} & \\ a_{31} & & \end{vmatrix}
\end{aligned}$$

There are $3! = 6$ ways to permutate the three columns, so there leaves six determinants. The six permutations of $(1,2,3)$ is given by:

$$\textbf{Column numbers} = (1,2,3),(2,3,1),(3,1,2),(1,3,2),(2,1,3),(3,2,1).$$

The last three are *odd permutations* (One exchange from identity permutation $(1,2,3)$.) The first three are *even permutations*. (zero or two exchange from identity permutation $(1,2,3)$.) When the column number is (α,β,ω), we get the entries $a_{1\alpha},a_{2\beta},a_{3\omega}$. The permutation (α,β,ω) comes with a plus or minus sign. If you don't understand, look at example below:

$$\det \boldsymbol{A} = a_{11}a_{22}a_{33}\begin{vmatrix}1 & & \\ & 1 & \\ & & 1\end{vmatrix} + a_{12}a_{23}a_{31}\begin{vmatrix} & 1 & \\ & & 1 \\ 1 & & \end{vmatrix} + a_{13}a_{21}a_{32}\begin{vmatrix} & & 1 \\ 1 & & \\ & 1 & \end{vmatrix}$$

$$+ a_{11}a_{23}a_{32}\begin{vmatrix}1 & & \\ & & 1 \\ & 1 & \end{vmatrix} + a_{12}a_{21}a_{33}\begin{vmatrix} & 1 & \\ 1 & & \\ & & 1\end{vmatrix} + a_{13}a_{22}a_{31}\begin{vmatrix} & & 1 \\ & 1 & \\ 1 & & \end{vmatrix}$$

The first three (even) permutation matrices have $\det \boldsymbol{P} = +1$, the last three (odd) permutation matrices have $\det \boldsymbol{P} = -1$. Hence we have:

$$\begin{aligned}\det \boldsymbol{A} &= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31} \\ &= a_{11}(a_{22}a_{33} - a_{23}a_{32}) + a_{12}(a_{23}a_{31} - a_{21}a_{33}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})\end{aligned}$$

6.1.1.1. n by n formula of determinant

Now we can see n by n formula. There are $n!$ permutations of columns, so we have $n!$ terms for determinant.

Assuming $(\alpha,\beta,\ldots,\omega)$ is the permutation of $(1,2,\ldots,n)$. The coorsponding term is $a_{1\alpha}a_{2\beta}\ldots a_{n\omega}\det \boldsymbol{P}$, where $\boldsymbol{P}$ is the permutation matrix with column number $\alpha,\beta,\ldots,\omega$.

The complete determinant of $\boldsymbol{A}$ is the sum of these $n!$ simple determinants. $a_{1\alpha}a_{2\beta}\ldots a_{n\omega}$ is obtained by choosing **one entry from every row and every column:**

Definition 6.1 [Big formula for determinant]

$$\begin{aligned}\det \boldsymbol{A} &= \text{sum of all } n! \text{ column permutations} \\ &= \sum(\det \boldsymbol{P})a_{1\alpha}a_{2\beta}\ldots a_{n\omega} = \text{BIG FORMULA}\end{aligned}$$

where $\boldsymbol{P}$ is permutation matrix with column number $(\alpha,\beta,\ldots,\omega)$. And $\{\alpha,\beta,\ldots,\omega\}$ is a permutation of $\{1,2,\ldots,n\}$.

R

Complexity Analysis. However, if we want to use big formula to compute matrix, we need to do $n!(n-1)$ multiplications. If we use formula $\det \boldsymbol{A} = \pm \prod pivots$, we only need to do $O(n^3)$ multiplications. Hence the letter one is more efficient.

6.1.1.2. Verify property

We can also use the big formula to verify property 1 to property 3:

- $\det \boldsymbol{I} = 1$:

  Only when $(\alpha,\beta,\dots,\omega)$=$(1,2,\dots,n)$, there is no zero entries for $a_{1\alpha}a_{2\beta}\dots a_{n\omega}$. Hence $\det \boldsymbol{A} = a_{11}a_{22}\dots a_{nn} = 1$.

- **sign reversal**:

  If two rows are interchanged, then all determinant of permutation matrix will change its sign, hence the value for determinant $\boldsymbol{A}$ is opposite.

- **The determinant is a linear function of each row separately**.

  If we separate out the fator $a_{11}, a_{12}, \dots, a_{1\alpha}$ **that comes from the first row**, this property is easy to check. For 3 by 3 matrix, separate the usual 6 terms of the determinant into 3 pairs:

$$\det \boldsymbol{A} = a_{11}(a_{22}a_{33} - a_{23}a_{32}) + a_{12}(a_{23}a_{31} - a_{21}a_{33}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}).$$

  Those three quantities in parentheses are called **cofactors**. They are 2×2 determinant coming from matrices in row 2 and 3. The first row contributes the factors a_{11}, a_{12}, a_{13}. The lower rows contribute the cofactors $(a_{22}a_{33} - a_{23}a_{32}), (a_{23}a_{31} - a_{21}a_{33}), (a_{21}a_{32} - a_{22}a_{31})$. Certainly $\det \boldsymbol{A}$ depends **linearly** on a_{11}, a_{12}, a_{13}, which is property 3.

6.1.2. Determinant by Cofactors

We could write the determinant in this form:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & & \\ & a_{22} & a_{23} \\ & a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} & a_{12} & \\ a_{21} & & a_{23} \\ a_{31} & & a_{33} \end{vmatrix} + \begin{vmatrix} & & a_{13} \\ a_{21} & a_{22} & \\ a_{31} & a_{32} & \end{vmatrix}.$$

If we define $\boldsymbol{A}_{1j}$ to be **the submatrix obtained by removing row 1 and column j**, We could compute $\det \boldsymbol{A}$ in this way:

$$\text{The cofactors along row 1 are } C_{1j} = (-1)^{1+j} \det \boldsymbol{A}_{1j} \quad j = 1, 2, \ldots, n.$$

$$\textbf{The cofactor expansion is } \det \boldsymbol{A} = a_{11}C_{11} + a_{12}C_{12} + \cdots + a_{1n}C_{1n}.$$

More generally, we can cross row i to get the determinant:

Definition 6.2 [Determinant] The determinant is the **dot product** of any row i of $\boldsymbol{A}$ with its cofactors using other rows:

$$\textbf{Cofactor Formula} \qquad \det \boldsymbol{A} = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}.$$

Each cofactor C_{ij} is defined as:

$$\textbf{Cofactor} \qquad C_{ij} = (-1)^{i+j} \det \boldsymbol{A}_{ij}$$

where A_{ij} is the submatrix obtained by removing row i and column j. ■

Cofactors down a column. Since we have $\det \boldsymbol{A} = \det \boldsymbol{A}^{\mathrm{T}}$, we can expand the determinant in cofactors *down a column* instead of across a row. Down column j the entries are a_{1j} to a_{nj}, the cofactors are C_{1j} to C_{nj}. The determinant is given by:

$$\textbf{Cofactors down column } j\textbf{:} \quad \det \boldsymbol{A} = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}.$$

6.1.3. Determinant Applications

6.1.3.1. Inverse

It's easy to check that the inverse of 2 by 2 matrix $\boldsymbol{A}$ is

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

We could use determinant to compute inverse! Before that let's define **cofactor matrix**:

Definition 6.3 [cofactor matrix] The cofactor matrix of $n \times n$ matrix $\boldsymbol{A}$ is given by:

$$\boldsymbol{C} = \begin{bmatrix} C_{ij} \end{bmatrix}_{1 \le i,j \le n}$$

where C_{ij} is the cofactor of $\boldsymbol{A}$.

Then we try to derive the inverse of matrix $\boldsymbol{A}$.

For $n \times n$ matrix $\boldsymbol{A}$, the product of $\boldsymbol{A}$ and the **transpose** of *cofactor matrix* is given by:

$$\boldsymbol{A}\boldsymbol{C}^{\mathrm{T}} = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} C_{11} & \dots & C_{n1} \\ \vdots & \ddots & \vdots \\ C_{1n} & \dots & C_{nn} \end{bmatrix} = \begin{bmatrix} \det \boldsymbol{A} & & \\ & \det \boldsymbol{A} & \\ & & \det \boldsymbol{A} \end{bmatrix} \tag{6.1}$$

Proofoutline:

- Row 1 of $\boldsymbol{A}$ times the column 1 of $\boldsymbol{C}^{\mathrm{T}}$ yields the first $\det \boldsymbol{A}$ on the right:

$$a_{11}C_{11} + a_{12}C_{12} + \dots + a_{1n}C_{1n} = \det \boldsymbol{A}$$

  Similarly, row j of $\boldsymbol{A}$ times column j of $\boldsymbol{C}^{\mathrm{T}}$ yields the determinant.

- *How to explain the zeros off the main diagonal in equation (6.1)?* Rows of $\boldsymbol{A}$ are multiplying $\boldsymbol{C}^{\mathrm{T}}$ from **different** columns. Why is the answer zero? For example, the (2,1)th entry of the result is given by

$$\begin{array}{c}\textbf{Row 2 of } \boldsymbol{A} \\ \\ \textbf{Row 1 of } \boldsymbol{C}\end{array} \qquad a_{21}C_{11} + a_{22}C_{12} + \cdots + a_{2n}C_{1n} = 0. \tag{6.2}$$

Explaination for Eq.(6.2): If the second row of $\boldsymbol{A}$ is copied into its first row, we define this new matrix as $\boldsymbol{A}^*$. Thus the determinant of $\boldsymbol{A}^*$ is given by:

$$\begin{vmatrix} a_{21} & a_{22} & \dots & a_{2n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ a_{31} & a_{32} & \dots & a_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{21} & & & \\ & a_{22} & \dots & a_{2n} \\ & a_{32} & \dots & a_{3n} \\ & \vdots & \ddots & \vdots \\ & a_{n2} & \dots & a_{nn} \end{vmatrix} + \begin{vmatrix} & a_{22} & & \\ a_{21} & & \dots & a_{2n} \\ a_{31} & & \dots & a_{3n} \\ \vdots & & \ddots & \vdots \\ a_{n1} & & \dots & a_{nn} \end{vmatrix} + \cdots + \begin{vmatrix} & & & a_{2n} \\ a_{21} & a_{22} & a_{2(n-1)} & \\ a_{31} & a_{32} & a_{3(n-1)} & \\ \vdots & \vdots & \vdots & \\ a_{n1} & a_{n2} & a_{n(n-1)} & \end{vmatrix}$$

Or equivalently, we have

$$\det \boldsymbol{A}^* = \begin{vmatrix} a_{21} & a_{22} & \dots & a_{2n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ a_{31} & a_{32} & \dots & a_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} = a_{21}C_{11} + a_{22}C_{12} + \cdots + a_{2n}C_{1n}$$

Since $\boldsymbol{A}^*$ has two equal rows, the determinant must be zero. Hence $a_{21}C_{11} + a_{22}C_{12} + \cdots + a_{2n}C_{1n} = 0$.

Similarly, **all entries off the main diagonal in Eq.(6.1) are zero.**

Thus the equation (6.1) is correct:

$$\boldsymbol{A}\boldsymbol{C}^{\mathrm{T}} = \begin{bmatrix} \det \boldsymbol{A} & & \\ & \det \boldsymbol{A} & \\ & & \det \boldsymbol{A} \end{bmatrix} = \det(\boldsymbol{A})\boldsymbol{I} \implies \boldsymbol{A}^{-1} = \frac{1}{\det \boldsymbol{A}}\boldsymbol{C}^{\mathrm{T}}.$$

Hence we could compute the inverse by computing many determinants of submatrix:

Definition 6.4 [Inverse] **The (i,j)th entry of A^{-1} is the cofactor C_{ji} (not C_{ji}) divided by $\det A$:**

$$\text{Formula for } A^{-1} \qquad (A^{-1})_{ij} = \frac{C_{ji}}{\det A} \quad \text{and} \quad A^{-1} = \frac{C^{\mathrm{T}}}{\det A}.$$

6.1.3.2. Cramer's Rule

Cramer's Rule solves $Ax = b$. Assume A is a $n \times n$ matrix that is **nonsingular**. Then we can use determinant to solve this system:

Let's start with $n = 3$. We could multiply A with a new matrix C_1 to get B_1:

$$\textbf{Key idea:} \quad AC_1 = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 & 0 & 0 \\ x_2 & 1 & 0 \\ x_3 & 0 & 1 \end{bmatrix} = \begin{bmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{bmatrix} = B_1$$

Taking determinants both sides, then we have

$$\det(AC_1) = \det(A)\det(C_1) = \det(A)(x_1) = \det B_1 \implies x_1 = \frac{\det B_1}{\det A_1}.$$

The matrix B_1 is essentaily obtained by replacing the first column of A by the vector b.

Similarly, we could get all x_j in this way. $(i = 1, \ldots, n)$.

Definition 6.5 [Cramer's Rule] If $\det A$ is not zero, $Ax = b$ could be solved by determinants:

$$x_1 = \frac{\det B_1}{\det A} \qquad x_2 = \frac{\det B_2}{\det A} \qquad \ldots\ldots \qquad x_n = \frac{\det B_n}{\det A}$$

The matrix $\boldsymbol{B}_j$ has the jth column of $\boldsymbol{A}$ replaced by the vector $\boldsymbol{b}$. In other words,

$$\boldsymbol{B}_j = \begin{bmatrix} a_{11} & \dots & b_1 & \dots & a_{1n} \\ a_{21} & \dots & b_2 & \dots & a_{2n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{n1} & \dots & b_n & \dots & a_{nn} \end{bmatrix} \quad j = 1, \dots, n.$$

6.1.4. Orthogonality

Definition 6.6 [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.

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 6.7 [length/norm] The **length(norm)** $\|\boldsymbol{x}\|$ of a vector $\boldsymbol{x} \in \mathbb{R}^n$ is the square root of $\langle \boldsymbol{x}, \boldsymbol{x} \rangle$:

$$\textbf{length} = \|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle} = \sqrt{x_1^2 + \cdots + x_n^2}.$$

6.1.4.1. 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 6.8 [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:

$$\langle f, g \rangle = \int_0^1 f(x)g(x)dx \quad \text{and} \quad \|f\|^2 = \sqrt{\int_0^1 f^2(x)dx}$$

Moreover, when $\langle f, g \rangle = 0$, we say two functions are **orthogonal** and denote it as $f \perp g$.

6.1.4.2. 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:

$$\begin{aligned} <\boldsymbol{x}, \boldsymbol{y}> = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} &= x_1x_2 + y_1y_2 \\ &= \|\boldsymbol{x}\|\|\boldsymbol{y}\|(\cos\theta\cos\varphi + \sin\theta\sin\varphi) \\ &= \|\boldsymbol{x}\|\|\boldsymbol{y}\|\cos(\theta - \varphi) \end{aligned}$$

Since $|\cos(\theta - \varphi)|$ never exceeds 1, the cosine formula gives a great inequality:

Theorem 6.1 — Cauchy Schwarz Inequality.

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle \le \|\boldsymbol{x}\|\|\boldsymbol{y}\|$$

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 $|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| \le \|\boldsymbol{x}\| \|\boldsymbol{y}\|$, we have

$$-1 \le \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{x}\| \|\boldsymbol{y}\|} \le 1$$

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 θ 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.

6.1.4.3. Orthogonal for space

After defining inner product, we can discuss the orthogonality for space:

Definition 6.9 [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 $\boldsymbol{V}$:

Orthogonal subspaces $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = 0$ for all $\boldsymbol{u}$ in $\boldsymbol{U}$ and all $\boldsymbol{v}$ in $\boldsymbol{V}$.

6.2. Thursday

6.2.1. Orthogonality

Recall that two vectors are orthogonal if their inner product is zero:

$$\boldsymbol{u} \perp \boldsymbol{v} \iff \langle \boldsymbol{u}, \boldsymbol{v} \rangle = 0$$

Orthogonality among vectors has an important property:

Proposition 6.1 If **nonzero** vectors $v_1, \dots, v_k$ are mutually orthogonal, i.e., $v_i \perp v_j$ for any $i \neq j$, then $\{v_1, \dots, v_k\}$ must be ind.

Proof. It suffices to show that

$$\alpha_1 v_1 + \cdots + \alpha_k v_k = \mathbf{0} \implies \alpha_i = 0 \text{ for any } i \in \{1, 2, \dots, k\}.$$

- We do inner product to show α_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 $i = 1, \dots, k$.

■

Now we can also talk about orthogonality among spaces:

Definition 6.10 [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 $\boldsymbol{V}$:

$$\text{Orthogonal subspaces} \quad \boldsymbol{u} \perp \boldsymbol{v}, \quad \forall \boldsymbol{u} \in \boldsymbol{U}, \boldsymbol{v} \in \boldsymbol{V}.$$

■ **Example 6.3** 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.

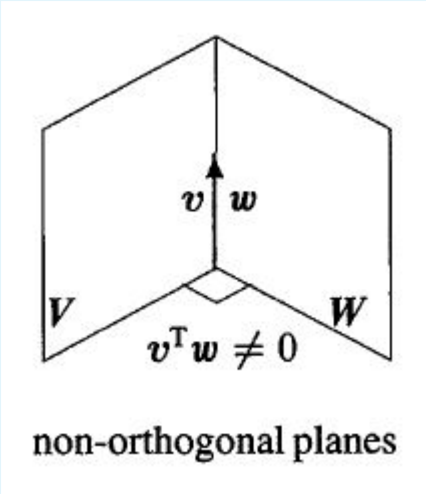


Figure 6.1: Orthogonality is impossible when $\dim \boldsymbol{U} + \dim \boldsymbol{V} > \dim(\boldsymbol{U} \cup \boldsymbol{V})$

Ⓡ 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 6.2 Assume $\{u_1, \dots, u_k\}$ is the basis for $\boldsymbol{U}$, $\{v_1, \dots, v_l\}$ is the basis for $\boldsymbol{V}$. If $\boldsymbol{U} \perp \boldsymbol{V}$ ($u_i \perp v_j$ for $\forall i, j$), then $u_1, u_2, \dots, u_k, v_1, v_2, \dots, v_l$ must be ind.

Proof. Suppose there exists $\{\alpha_1,\dots,\alpha_k\}$ and $\{\beta_1,\dots,\beta_l\}$ 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 6.1 For subspaces $\boldsymbol{U}$ and $\boldsymbol{V}$, we obtain

$$\dim(\boldsymbol{U} \cup \boldsymbol{V}) \le \dim(\boldsymbol{U}) + \dim(\boldsymbol{V}).$$

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 6.11 [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 6.4** 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 6.3 — 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 $r = \operatorname{rank}(\boldsymbol{A})$.

  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 6.2 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ 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 6.3 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ 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 6.4 $\boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b}$ 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 6.5** Given the system

$$x_1 + x_2 \geq 1 \tag{6.3}$$

$$-x_1 \geq -1 \tag{6.4}$$

$$-x_2 \geq 2 \tag{6.5}$$

Eq.(6.3)$\times$1+Eq(6.4)$\times$1+Eq.(6.5)$\times$1 gives

$$0 \geq 2$$

which is a contradiction!

So the key idea of theorem (6.4) 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 6.4 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 6.5 The system $\boldsymbol{Ax} = \boldsymbol{b}$ may not have a solution, but $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ 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 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$. ■

Corollary 6.6 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ 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.

■

6.2.2. Least Squares Approximations

The linear system $\boldsymbol{Ax} = \boldsymbol{b}$ often has no solution, if so, what should we do?

We cannot always get the error $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{Ax}$ down to zero, so we want to use *least square method* to minimize the error. In other words, our goal is to

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \boldsymbol{e}^2 := \min_{\boldsymbol{x}} \|\boldsymbol{Ax} - \boldsymbol{b}\|^2 = \sum_{i=1}^{m} (a_i^{\mathrm{T}}\boldsymbol{x} - b_i)^2$$

where $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$. The minimizer $\boldsymbol{x}$ is called the **linear least squares solution**.

6.2.2.1. Least Squares by Convex Optimization

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.

$$\frac{\partial(a^{\mathrm{T}}\boldsymbol{x})}{\partial\boldsymbol{x}} = a \tag{6.6}$$

$$\frac{\partial(a^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x})}{\partial\boldsymbol{x}} = \frac{\partial((\boldsymbol{A}^{\mathrm{T}}a)^{\mathrm{T}}\boldsymbol{x})}{\partial\boldsymbol{x}} = \boldsymbol{A}^{\mathrm{T}}a \tag{6.7}$$

$$\frac{\partial(\boldsymbol{A}\boldsymbol{x})}{\partial\boldsymbol{x}} = \boldsymbol{A}^{\mathrm{T}} \tag{6.8}$$

$$\frac{\partial(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x})}{\partial\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{x} \tag{6.9}$$

Thus, in order to minimize $\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|^2 = (\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})$, it suffices to let its **derivative** with respect to $\boldsymbol{x}$ to be **zero.** (Since $\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|^2$ is convex, which will be discussed in detail in other courses.) Hence we have:

$$\begin{aligned}
\frac{\partial(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})}{\partial\boldsymbol{x}} &= \frac{\partial(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})}{\partial\boldsymbol{x}}(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}) + \frac{\partial(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})}{\partial\boldsymbol{x}}(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}) \\
&= 2\frac{\partial(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})}{\partial\boldsymbol{x}}(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}) \\
&= 2(\frac{\partial(\boldsymbol{A}\boldsymbol{x})}{\partial\boldsymbol{x}} - \frac{\partial(\boldsymbol{b})}{\partial\boldsymbol{x}})(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}) \\
&= 2\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}) = \boldsymbol{0}.
\end{aligned}$$

Or equivalently,

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}.$$

According to corollary (6.5), this equation always exists a solution. This equation is called the **normal equation**.

Theorem 6.5 A vector $\boldsymbol{x}_{\mathrm{LS}}$ is an optimal solution to the least squares problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \|\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}\|_2^2 \tag{6.10a}$$

if and only if it satisfies

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}_{\mathrm{LS}} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}. \tag{6.10b}$$

6.2.2.2. Fit a stright line

Given a collection of data $(\boldsymbol{x}_i, y_i)$ for $i = 1, \ldots, m$, we can use a stright line to fit these points:

$$\begin{cases} y_1 = a_0 + a_1 x_{1,1} + a_2 x_{1,2} + \cdots + a_n x_{1,n} + \varepsilon_1 \\ y_2 = a_0 + a_1 x_{2,1} + a_2 x_{2,2} + \cdots + a_n x_{2,n} + \varepsilon_2 \\ \quad \vdots \\ y_m = a_0 + a_1 x_{m,1} + a_2 x_{m,2} + \cdots + a_n x_{m,n} + \varepsilon_m \end{cases}$$

Our fit line is

$$\hat{y} = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$$

In *compact matrix form*, we have

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \ldots & x_{1,n} \\ 1 & x_{2,1} & x_{2,2} & \ldots & x_{2,n} \\ \vdots & \vdots & & & \\ 1 & x_{m,1} & x_{m,2} & \ldots & x_{m,n} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix}$$

Or equivalently, we have

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{A} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \ldots & x_{1,n} \\ 1 & x_{2,1} & x_{2,2} & \ldots & x_{2,n} \\ \vdots & \vdots & & & \\ 1 & x_{m,1} & x_{m,2} & \ldots & x_{m,n} \end{bmatrix}_{m \times (n+1)}, \boldsymbol{x} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}_{(n+1) \times 1}, \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix}_{m \times 1}$.

Our goal is to minimize $\|\hat{\boldsymbol{y}} - \boldsymbol{y}\|^2 = \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|^2$. Then by theorem (6.5), it suffices to sovle $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}$.

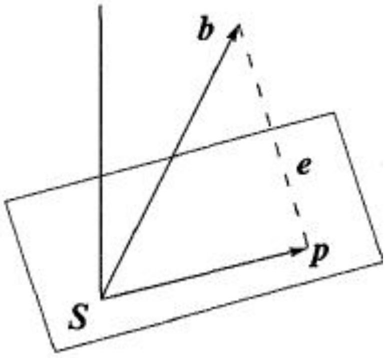


Figure 6.2: The projection of $\boldsymbol{b}$ onto a subspace $\boldsymbol{S} := \mathcal{C}(\boldsymbol{A})$.

6.2.3. Projections

In corollary (6.6), 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.
- Otherwise, just as the Figure (6.2) shown, $\boldsymbol{A}\boldsymbol{x}^*$ is the projection of $\boldsymbol{b}$ to subspace $\mathcal{C}(\boldsymbol{A})$.

Definition 6.12 [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 $\mathrm{Proj}_{\boldsymbol{S}}(\boldsymbol{b})$ 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

$$\mathrm{Proj}_{\mathcal{C}(\boldsymbol{A})}(\boldsymbol{b}) := \boldsymbol{A}\boldsymbol{x}^*, \quad \text{where } \boldsymbol{x}^* = \arg\min_{\boldsymbol{x} \in \mathbb{R}^n} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|.$$

Definition 6.13 [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 6.14 [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*:

$$\begin{aligned}\boldsymbol{P}^2 &= \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}} \\ &= \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}} \\ &= \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{P}.\end{aligned}$$

6.2.3.1. 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 span$\{\boldsymbol{a}\}$. (Shown in figure (6.3).)

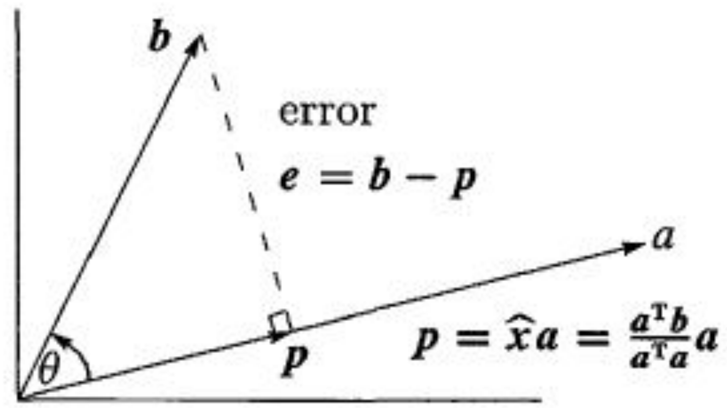


Figure 6.3: The projection of $\boldsymbol{b}$ onto a line $\boldsymbol{a}$.

More generally, we can write the projection of $\boldsymbol{b}$ onto the line span$\{\boldsymbol{a}\}$ as:

$$\mathrm{Proj}_{\mathrm{span}\{\boldsymbol{a}\}}(\boldsymbol{b}) = \frac{\langle \boldsymbol{a}, \boldsymbol{b} \rangle}{\langle \boldsymbol{a}, \boldsymbol{a} \rangle}\boldsymbol{a}$$

Changing an Orthogonal Basis. Note that the error $\boldsymbol{b} - \mathrm{Proj}_{\mathrm{span}\{\boldsymbol{a}\}}(\boldsymbol{b})$ is perpendicular to $\boldsymbol{a}$, and $\boldsymbol{b} - \mathrm{Proj}_{\mathrm{span}\{\boldsymbol{a}\}}(\boldsymbol{b}) \in \mathrm{span}\{\boldsymbol{a}, \boldsymbol{b}\}$.

If we define $\boldsymbol{b}' = \boldsymbol{b} - \mathrm{Proj}_{\mathrm{span}\{\boldsymbol{a}\}}(\boldsymbol{b})$, then it's easy to check that $\mathrm{span}\{\boldsymbol{a}, \boldsymbol{b}'\} = \mathrm{span}\{\boldsymbol{a}, \boldsymbol{b}\}$ 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.

6.3. Friday

This lecture has two goals. The first is to see **how orthogonality makes it easy to find the projection matrix $\boldsymbol{P}$ and the projection** $\text{Proj}_{\mathcal{C}(\boldsymbol{A})}\boldsymbol{b}$. The key idea is that *Orthogonality makes the product $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ a diagonal matrix*. The second goal is to **show how to construct orthogonal basis of $\mathcal{C}(\boldsymbol{A})$**. For matrix $\boldsymbol{A} = \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix}$, the columns may not be orthogonal. We intend to convert $a_1, \dots, a_n$ to orthogonal vectors, which will be the columns of a new matrix $\boldsymbol{Q}$.

6.3.1. Orthonormal basis

The vectors $\boldsymbol{q}_1, \dots, \boldsymbol{q}_n$ are **orthogonal** when their inner product $\langle \boldsymbol{q}_i, \boldsymbol{q}_j \rangle$ are zero. ($i \neq j$.) With one more step–each vector is just divided by its length, then the collection of vectors become **orthogonal unit vectors**. Their lengths are all 1. Then this basis is called **orthonormal**.

Definition 6.15 [orthonormal] The collection of vectors $\boldsymbol{q}_1, \dots, \boldsymbol{q}_n \in \mathbb{R}^m$ is said to be:

- **orthogonal** if $\langle \boldsymbol{q}_i, \boldsymbol{q}_j \rangle = 0$ for all i, j with $i \neq j$
- **orthonormal** if $\|\boldsymbol{q}_i\|_2 = 1$ for all i and $\langle \boldsymbol{q}_i, \boldsymbol{q}_j \rangle = 0$ for all i, j with $i \neq j$, or equivalently,

$$\langle \boldsymbol{q}_i, \boldsymbol{q}_j \rangle = \begin{cases} 0 & \text{when } i \neq j \qquad (\textbf{orthogonal} \text{ vectors}), \\ 1 & \text{when } i = j \qquad (\textbf{unit} \text{ vectors: } \|\boldsymbol{q}_i\| = 1). \end{cases}$$

Moreover, if $\boldsymbol{q}_1, \dots, \boldsymbol{q}_n$ are **orthonormal**, then the basis $\{\boldsymbol{q}_1, \dots, \boldsymbol{q}_n\}$ is called **orthonormal basis**. ▪

■ **Example 6.6** Given a collection of unit vectors

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

then $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ forms an *orthonormal basis* for $\mathbb{R}^n$. ■

If we want to express vector $\boldsymbol{b}$ as the linear combination of arbitrary basis (may not be orthogonal) $\{\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_n\}$, what should we do?

Answer: Solve the system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, where $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \cdots & \boldsymbol{q}_n \end{bmatrix}$

What if $\{\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_n\}$ is an **orthogonal** basis? How to find solution $\boldsymbol{x}$ s.t.

$$\boldsymbol{b} = x_1\boldsymbol{q}_1 + x_2\boldsymbol{q}_2 + \cdots + x_n\boldsymbol{q}_n? \tag{6.11}$$

Answer: We just do the inner product of each $\boldsymbol{q}_i$ with $\boldsymbol{b}$ to get the coefficient x_i:

$$\begin{aligned} \langle \boldsymbol{q}_i, \boldsymbol{b} \rangle &= x_1\langle \boldsymbol{q}_i, \boldsymbol{q}_1 \rangle + x_2\langle \boldsymbol{q}_i, \boldsymbol{q}_2 \rangle + \cdots + x_n\langle \boldsymbol{q}_i, \boldsymbol{q}_n \rangle \\ &= x_i\langle \boldsymbol{q}_i, \boldsymbol{q}_i \rangle = x_i \end{aligned} \tag{6.12}$$

By substituting Eq.(6.12) into Eq.(6.11), we could express $\boldsymbol{b}$ as:

$$\boldsymbol{b} = \sum_{i=1}^{n} \langle \boldsymbol{q}_i, \boldsymbol{b} \rangle \boldsymbol{q}_i.$$

In this case, from Eq.(6.12) we can see that if columns of $\boldsymbol{A}$ are orthogonal, we could easily obtain the solution to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$:

$$x_i = \langle \boldsymbol{q}_i, \boldsymbol{b} \rangle, \quad i = 1, 2, \dots, n.$$

Definition 6.16 [matrix with orthonormal columns] Given a collection of **orthonormal** vectors $\boldsymbol{q}_1,\dots,\boldsymbol{q}_n$, the matrix

$$\boldsymbol{Q} = \begin{bmatrix} q_1 & q_2 & \dots & q_n \end{bmatrix}$$

is said to be a matrix with **orthonormal** columns.

Note that a matrix with **orthonormal** columns is often denoted as $\boldsymbol{Q}$.

Or equivalently, a matrix $\boldsymbol{Q}$ is with **orthonormal** columns if and only if

$$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{q}_1^{\mathrm{T}} \\ \boldsymbol{q}_2^{\mathrm{T}} \\ \dots \\ \boldsymbol{q}_n^{\mathrm{T}} \end{pmatrix} \begin{pmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \dots & \boldsymbol{q}_n \end{pmatrix} = \begin{pmatrix} \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{q}_1 & & \\ & \ddots & \\ & & \boldsymbol{q}_n^{\mathrm{T}}\boldsymbol{q}_n \end{pmatrix} = \boldsymbol{I}. \tag{6.13}$$

(R) Note that a matrix $\boldsymbol{Q}$ with orthonormal columns is *not required to be square*! Moreover, $\{\boldsymbol{q}_1,\dots,\boldsymbol{q}_n\}$ in $\boldsymbol{Q}$ is *not required to form a basis.*

Definition 6.17 [orthogonal matrix] A matrix $\boldsymbol{Q}$ is said to be **orthogonal** if it is square and its columns are orthonormal.

Question: Why we call it an orthogonal matrix, but not an orthonormal matrix?

Answer: Orthogonal matrix usually transform an orthogonal basis into another orthogonal basis by matrix multiplication. This special property requires its column to be **orthonormal**.

Example 6.7 If $\boldsymbol{Q}$ is an orthogonal matrix, while $\hat{\boldsymbol{Q}}$ is a matrix with orthonormal columns that is **not square**. Do the products $\boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}}$ and $\hat{\boldsymbol{Q}}\hat{\boldsymbol{Q}}^{\mathrm{T}}$ always be *identity matrix*?

Answer:

- $\boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}}$ is always *identity matrix*. According to equation (6.13), we have $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{I}$.

Hence Q^{T} is the left inverse of square matrix Q, which implies

$$Q^{-1} = Q^{\mathrm{T}} \implies QQ^{\mathrm{T}} = QQ^{-1} = I.$$

Moreover, solving $Qx = b$ is equivalent to $x = Q^{-1}b = Q^{\mathrm{T}}b$, which is *exactly*

$$x = \begin{bmatrix} \langle q_1, b\rangle \\ \langle q_2, b\rangle \\ \vdots \\ \langle q_n, b\rangle \end{bmatrix}.$$

- Although $\hat{Q}^{\mathrm{T}}\hat{Q} = I$, the product $\hat{Q}\hat{Q}^{\mathrm{T}}$ will never be identity matrix for nonsquare $\hat{Q}$. We can verify it by the its rank:

  Assume $\hat{Q} \in \mathbb{R}^{m\times n}(m \neq n)$. Then it's easy to verify that $\operatorname{rank}(\hat{Q}\hat{Q}^{\mathrm{T}}) = \operatorname{rank}(\hat{Q})$. Since $\hat{Q}$ has orthonormal columns, the columns of $\hat{Q}$ are independent, i.e., $\operatorname{rank}(\hat{Q}) = n$. But $\operatorname{rank}(\hat{Q}\hat{Q}^{\mathrm{T}}) = \operatorname{rank}(\hat{Q}) = n \neq m = \operatorname{rank}(I_m)$.

  Moreover, if $\hat{Q}$ has only one column $\hat{q}$, then $\hat{Q}\hat{Q}^{\mathrm{T}} = \hat{q}\hat{q}^{\mathrm{T}} = \operatorname{rank}(1) \neq \operatorname{rank}(I_m)$.

Proposition 6.2

If Q has orthonormal columns, then it *leaves lengths unchanged*, in other words,

$$\textbf{Same length} \quad \|Qx\| = \|x\| \text{ for every vector } x.$$

Also, Q preserves inner products for vectors, i.e., :

$$\langle Qx, Qy\rangle = \langle x, y\rangle \quad \text{for every vectors } x \text{ and } y.$$

Proofoutline. $\|Qx\|^2 = \|x\|^2$ because

$$\begin{aligned} \langle Qx, Qx\rangle &= x^{\mathrm{T}}Q^{\mathrm{T}}Qx = x^{\mathrm{T}}(Q^{\mathrm{T}}Q)x \\ &= x^{\mathrm{T}}Ix = x^{\mathrm{T}}x \end{aligned}$$

Hence we have $\|\boldsymbol{Q}\boldsymbol{x}\| = \|\boldsymbol{x}\|$. Just using $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{I}$, we can derive $\langle \boldsymbol{Q}\boldsymbol{x}, \boldsymbol{Q}\boldsymbol{y}\rangle = \langle \boldsymbol{x}, \boldsymbol{y}\rangle$. ■

Orthogonal matrices are excellent for computations, since the inverse of matrices could usually be converted into transpose.

When Least Squares Meet Orthogonality. In particular, if $\boldsymbol{Q} \in \mathbb{R}^{m\times n}$ has orthonormal columns, the least square problem is easy:

Although $\boldsymbol{Q}\boldsymbol{x} = \boldsymbol{b}$ may not have a solution, but the normal equation

$$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}\hat{x} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}$$

must have the unique solution $\hat{x} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}$. Why? Since $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{I}$, we derive

$$\hat{x} = (\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q})^{-1}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}.$$

6.3.1.1. Summary

Hence the **least squares solution** to $\boldsymbol{Q}\boldsymbol{x} = \boldsymbol{b}$ is $\hat{x} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}$. In other words, $\boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b} \approx \boldsymbol{b}$. **The projection matrix is $\boldsymbol{P} = \boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}}$**. Note that the projection $\mathrm{Proj}_{C(\boldsymbol{Q})}(\boldsymbol{b}) = \boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}$ doesn't equal to $\boldsymbol{b}$ in general.

For general matrix $\boldsymbol{A}$, the projection matrix is more complicated:

$$\boldsymbol{P} = \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}.$$

6.3.2. Gram-Schmidt Process

"Orthogonal is good". So our goal for this section is: *Given a collection of independent vectors, how to make them orthonormal?*

We start with three independent vectors $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ in $\mathbb{R}^3$. In order to construct orthonormal vectors, firstly we construct three **orthogonal** vectors $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$. Secongly we divide $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ by their lengths to get three **orthonormal** vectors $\boldsymbol{q}_1 = \frac{\boldsymbol{A}}{\|\boldsymbol{A}\|}, \boldsymbol{q}_2 = \frac{\boldsymbol{B}}{\|\boldsymbol{B}\|}, \boldsymbol{q}_3 = \frac{\boldsymbol{C}}{\|\boldsymbol{C}\|}$.

- Firstly we set $\boldsymbol{A} = \boldsymbol{a}$.

- The next vector $\boldsymbol{B}$ must be perpendicular to $\boldsymbol{A}$. Look at the figure (6.4) below, We find that $\boldsymbol{B} = \boldsymbol{b} - \text{Proj}_{\boldsymbol{A}}(\boldsymbol{b})$. Or equivalently,

$$\textbf{First Gram-Schmidt step} \qquad \boldsymbol{B} = \boldsymbol{b} - \frac{\langle \boldsymbol{A}, \boldsymbol{b} \rangle}{\langle \boldsymbol{A}, \boldsymbol{A} \rangle} \boldsymbol{A}.$$

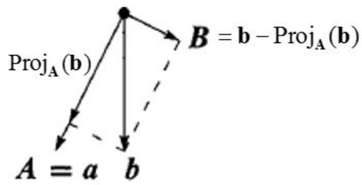


Figure 6.4: Subtract projection to get $\boldsymbol{B} = \boldsymbol{b} - \text{Proj}_{\boldsymbol{A}} \boldsymbol{b}$.

You can take inner product between $\boldsymbol{A}$ and $\boldsymbol{B}$ to verify that $\boldsymbol{A}$ and $\boldsymbol{B}$ are orthogonal in Figure (6.4). Note that $\boldsymbol{B}$ is not zero (otherwise $\boldsymbol{a}$ and $\boldsymbol{b}$ would be dependent. We will show it later.)

- Then we want to construct another vector $\boldsymbol{C}$. Most likely $\boldsymbol{c}$ is **not** perpendicular to $\boldsymbol{A}$ and $\boldsymbol{B}$. What we do is to **subtract $\boldsymbol{c}$ off its projections onto the column space of A and B to get C:**

$$\begin{aligned} \boldsymbol{C} &= \boldsymbol{c} - \text{Proj}_{\text{span}\{\boldsymbol{A},\boldsymbol{B}\}}(\boldsymbol{c}) \\ \textbf{Next Gram-Schmidt step} \qquad &= \boldsymbol{c} - \text{Proj}_{\boldsymbol{A}}(\boldsymbol{c}) - \text{Proj}_{\boldsymbol{B}}(\boldsymbol{c}) \\ &= \boldsymbol{c} - \frac{\langle \boldsymbol{A}, \boldsymbol{c} \rangle}{\langle \boldsymbol{A}, \boldsymbol{A} \rangle} \boldsymbol{A} - \frac{\langle \boldsymbol{B}, \boldsymbol{c} \rangle}{\langle \boldsymbol{B}, \boldsymbol{B} \rangle} \boldsymbol{B}. \end{aligned}$$

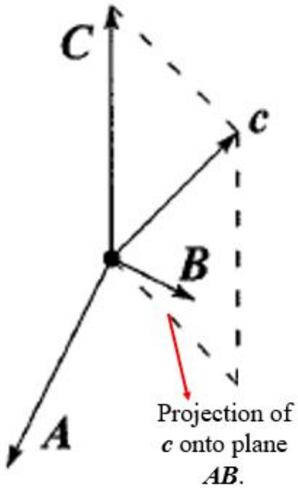


Figure 6.5: Subtract $\boldsymbol{c}$ off its projections onto the column space of $\boldsymbol{A}$ and $\boldsymbol{B}$ to get $\boldsymbol{C}$

- Finally we get orthogonal vectors $\boldsymbol{A},\boldsymbol{B},\boldsymbol{C}$. Orthonormal vectors $\boldsymbol{q}_1,\boldsymbol{q}_2,\boldsymbol{q}_3$ are obtained by dividing their lengths (shown in Figure (6.6)):

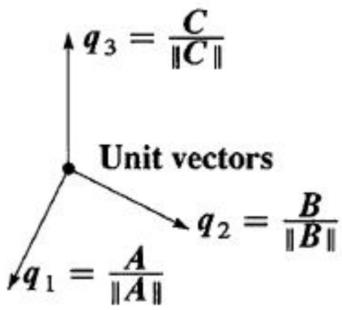


Figure 6.6: Final Gram-Schmidt step

Next we show an example of Gram-Schmidt step:

Example 6.8 How to construct orthonormal vectors from

$$\boldsymbol{a} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \boldsymbol{c} = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}?$$

- Firstly we set $\boldsymbol{A} = \boldsymbol{a} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$.

- $$\begin{aligned} \boldsymbol{B} &= \boldsymbol{b} - \mathrm{Proj}_{\boldsymbol{A}}(\boldsymbol{b}) = \boldsymbol{b} - \frac{\langle \boldsymbol{A}, \boldsymbol{b} \rangle}{\langle \boldsymbol{A}, \boldsymbol{A} \rangle} \boldsymbol{A} \\ &= \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} 2^{-1} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{2} \\ 0 \\ -\frac{1}{2} \end{pmatrix} \end{aligned}$$

- $$\begin{aligned} \boldsymbol{C} &= \boldsymbol{c} - \mathrm{Proj}_{\boldsymbol{A}}(\boldsymbol{c}) - \mathrm{Proj}_{\boldsymbol{B}}(\boldsymbol{c}) = \boldsymbol{c} - \frac{\langle \boldsymbol{A}, \boldsymbol{c} \rangle}{\langle \boldsymbol{A}, \boldsymbol{A} \rangle} \boldsymbol{A} - \frac{\langle \boldsymbol{B}, \boldsymbol{c} \rangle}{\langle \boldsymbol{B}, \boldsymbol{B} \rangle} \boldsymbol{B} \\ &= \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} 2^{-1} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} - \begin{pmatrix} \frac{1}{2} \\ 0 \\ -\frac{1}{2} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} (\frac{1}{2})^{-1} \begin{pmatrix} \frac{1}{2} \\ 0 \\ -\frac{1}{2} \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \end{aligned}$$

  Hence we obtain our orthonormal vectors:

  $$\boldsymbol{q}_1 = \frac{\boldsymbol{A}}{\|\boldsymbol{A}\|} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix}, , \boldsymbol{q}_2 = \frac{\boldsymbol{B}}{\|\boldsymbol{B}\|} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ -\frac{1}{\sqrt{2}} \end{pmatrix}, \boldsymbol{q}_3 = \frac{\boldsymbol{C}}{\|\boldsymbol{C}\|} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

And we derive the orthogonal matrix $\boldsymbol{Q}$:

$$Q = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{pmatrix}$$

■

When will the Gram-Schmidt process "fail"? Let's describe this process in general case first, then we answer this question.

6.3.2.1. Gram-Schmidt process in general case

Algorithm: Gram-Schmidt Process

Input: a collection of vectors $\boldsymbol{a}_1, \dots, \boldsymbol{a}_n$, presumably linear independent.

Firstly construct orthogonal vectors $\boldsymbol{A}_1, \dots, \boldsymbol{A}_n$.

$\boldsymbol{A}_1 = \boldsymbol{a}_1$.

To construct $\boldsymbol{A}_j$, $j \in \{2, \dots, n\}$, we compute $\boldsymbol{a}_j$ minus its projection in the column space spanned by $\{\boldsymbol{A}_1, \boldsymbol{A}_2, \dots, \boldsymbol{A}_{j-1}\}$:

$$\begin{aligned} \boldsymbol{A}_j &= \boldsymbol{a}_j - \text{Proj}_{\text{span}\{\boldsymbol{A}_1, \boldsymbol{A}_2, \dots, \boldsymbol{A}_{j-1}\}}(\boldsymbol{a}_j) \\ &= \boldsymbol{a}_j - \text{Proj}_{\boldsymbol{A}_1}(\boldsymbol{a}_j) - \text{Proj}_{\boldsymbol{A}_2}(\boldsymbol{a}_j) - \cdots - \text{Proj}_{\boldsymbol{A}_{j-1}}(\boldsymbol{a}_j) \\ &= \boldsymbol{a}_j - \frac{\langle \boldsymbol{A}_1, \boldsymbol{a}_j \rangle}{\langle \boldsymbol{A}_1, \boldsymbol{A}_1 \rangle} \boldsymbol{A}_1 - \frac{\langle \boldsymbol{A}_2, \boldsymbol{a}_j \rangle}{\langle \boldsymbol{A}_2, \boldsymbol{A}_2 \rangle} \boldsymbol{A}_2 - \cdots - \frac{\langle \boldsymbol{A}_{j-1}, \boldsymbol{a}_j \rangle}{\langle \boldsymbol{A}_{j-1}, \boldsymbol{A}_{j-1} \rangle} \boldsymbol{A}_{j-1} \end{aligned}$$

Secondly, after getting $\boldsymbol{A}_1, \dots, \boldsymbol{A}_n$, we can construct orthonormal vectors:

$$\boldsymbol{q}_j = \frac{\boldsymbol{A}_j}{\|\boldsymbol{A}_j\|} \quad \text{for } j = 1, 2, \dots, n.$$

So when do this process fail? When $\exists j$ such that $\boldsymbol{A}_j = \boldsymbol{0}$, we cannot continue this process anymore:

Proposition 6.3 $\boldsymbol{A}_j \neq \boldsymbol{0}$ for $\forall j$ if and only if $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_n$ are indendent.

Proofoutline. $\boldsymbol{A}_j = \boldsymbol{0} \Longleftrightarrow \boldsymbol{a}_j = \text{Proj}_{\text{span}\{\boldsymbol{A}_1, \dots, \boldsymbol{A}_{j-1}\}}(\boldsymbol{a}_j)$. It suffices to prove $\exists j$ s.t. $\boldsymbol{A}_j = \boldsymbol{0}$ if and only if $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_n$ are depependent.

Sufficiency. Given $\boldsymbol{A}_j = \boldsymbol{0}$, then $\boldsymbol{a}_j = \text{Proj}_{\text{span}\,\boldsymbol{A}_1,\dots,\boldsymbol{A}_{j-1}}(\boldsymbol{a}_j) \in \text{span}\{\boldsymbol{A}_1,\dots,\boldsymbol{A}_{j-1}\}$. It's easy to verify that $\text{span}\{\boldsymbol{A}_1,\dots,\boldsymbol{A}_{j-1}\} = \text{span}\{\boldsymbol{a}_1,\dots,\boldsymbol{a}_{j-1}\}$. Hence $\boldsymbol{a}_j \in \text{span}\{\boldsymbol{a}_1,\dots,\boldsymbol{a}_{j-1}\}$. Hence $\boldsymbol{a}_1,\dots,\boldsymbol{a}_j$ are dependent. Thus $\boldsymbol{a}_1,\dots,\boldsymbol{a}_n$ are dependent.

Necessity. Given dependent vectors $\boldsymbol{a}_1,\dots,\boldsymbol{a}_n$, obviously, $\boldsymbol{a}_n \in \text{span}\{\boldsymbol{a}_1,\dots,\boldsymbol{a}_{n-1}\}$. It's easy to verify that $\boldsymbol{a}_n = \text{Proj}_{\text{span}\{\boldsymbol{a}_1,\dots,\boldsymbol{a}_{n-1}\}}(\boldsymbol{a}_n)$. Thus $\boldsymbol{a}_n = \text{Proj}_{\text{span}\{\boldsymbol{A}_1,\dots,\boldsymbol{A}_{n-1}\}}(\boldsymbol{a}_n) \implies \boldsymbol{A}_n = \boldsymbol{0}$. ∎

6.3.3. The Factorization $A = QR$

We know that Gaussian Elimination leads to *LU decomposition*; in fact, Gram-Schmidt process leads to *QR factorization*. These two decomposition methods are quite important in Linear Algebra, let's discuss QR factorization briefly:

> Given a matrix $\boldsymbol{A} = \begin{bmatrix}\boldsymbol{a} & \boldsymbol{b} & \boldsymbol{c}\end{bmatrix}$, we finally end with a matrix $\boldsymbol{Q} = \begin{bmatrix}\boldsymbol{q}_1 & \boldsymbol{q}_2 & \boldsymbol{q}_3\end{bmatrix}$. How are these two matrices related?

Answer: Since the linear combination of $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}$ leads to $\boldsymbol{q}_1,\boldsymbol{q}_2,\boldsymbol{q}_3$ (vice versa), there must be a third matrix connecting $\boldsymbol{A}$ to $\boldsymbol{Q}$. This third matrix is the triangular $\boldsymbol{R}$ such taht $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$.

Let's discuss a specific example to show how to do QR factorization.

■ **Example 6.9** Given $\boldsymbol{A} = \begin{bmatrix}\boldsymbol{a} & \boldsymbol{b} & \boldsymbol{c}\end{bmatrix}$, whose columns are independent, then we can use Gram-Schmidt process to obtain the corresponding orthonormal vectors $\boldsymbol{q}_1,\boldsymbol{q}_2,\boldsymbol{q}_3$ from $\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}$. As a result, we can write $\boldsymbol{A}$ as:

$$\boldsymbol{A} = \begin{bmatrix}\boldsymbol{q}_1 & \boldsymbol{q}_2 & \boldsymbol{q}_3\end{bmatrix}\begin{bmatrix}\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{a} & \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b} & \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{c} \\ 0 & \boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{b} & \boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{c} \\ 0 & 0 & \boldsymbol{q}_3^{\mathrm{T}}\boldsymbol{c}\end{bmatrix}$$

We define $\boldsymbol{R} \triangleq \begin{bmatrix} \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{a} & \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b} & \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{c} \\ 0 & \boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{b} & \boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{c} \\ 0 & 0 & \boldsymbol{q}_3^{\mathrm{T}}\boldsymbol{c} \end{bmatrix}, \boldsymbol{Q} \triangleq \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \boldsymbol{q}_3 \end{bmatrix}$.
Hence $\boldsymbol{A}$ could be factorized into:

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$$

where $\boldsymbol{R}$ is upper triangular, $\boldsymbol{Q}$ is a matrix with orthonormal columns. ■

QR factorization holds for every matrix with independent columns:

Theorem 6.6 Every $m \times n$ matrix $\boldsymbol{A}$ with ind. columns can be factorized as

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$$

where $\boldsymbol{Q}$ is a matrix with *orthonormal columns*, $\boldsymbol{R}$ is an upper triangular matrix (always square).

We omit the proof of this theorem. Now we show that the inverse of $\boldsymbol{R}$ always exists:

Proof. suppose $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \dots & \boldsymbol{a}_n \end{bmatrix}, \boldsymbol{Q} = \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \dots & \boldsymbol{q}_n \end{bmatrix}$. Thus we derive

$$\boldsymbol{R} = \boldsymbol{Q}^{-1}\boldsymbol{A} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A} = \begin{bmatrix} \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{a}_1 & \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{a}_2 & \dots & \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{a}_n \\ 0 & \boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{a}_2 & \dots & \boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{a}_n \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \boldsymbol{q}_n^{\mathrm{T}}\boldsymbol{a}_n \end{bmatrix}$$

For every step j we have

$$\boldsymbol{A}_j = \boldsymbol{a}_j - \mathrm{Proj}_{\mathrm{span}\{a_1,\dots,a_{j-1}\}}(\boldsymbol{a}_j), \qquad \boldsymbol{q}_j = \frac{\boldsymbol{A}_j}{\|\boldsymbol{A}_j\|}.$$

Since $\langle \boldsymbol{A}_j, \boldsymbol{a}_j \rangle = \langle \boldsymbol{a}_j, \boldsymbol{a}_j \rangle - \langle \mathrm{Proj}_{\mathrm{span}\{a_1,\dots,a_{j-1}\}}(\boldsymbol{a}_j), \boldsymbol{a}_j \rangle = \|a_j\|^2 - \|\mathrm{Proj}_{\mathrm{span}\{a_1,\dots,a_{j-1}\}}(\boldsymbol{a}_j)\|^2 > 0$, we have $\langle \boldsymbol{q}_j, \boldsymbol{a}_j \rangle = \frac{\langle \boldsymbol{A}_j, \boldsymbol{a}_j \rangle}{\|\boldsymbol{A}_j\|} > 0$. Hence the diagonal of $\boldsymbol{R}$ are all positive. Hence this triangular matrix is *invertible.* ■

Proposition 6.4 If $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$, then the least squares solution is given by:

$$\boldsymbol{x} = (\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R})^{-1}\boldsymbol{R}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b} = \boldsymbol{R}^{-1}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}.$$

Explain: Since we have

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{R}\boldsymbol{x} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{x}$$
$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}$$

it's equivalent to solve $\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{x} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}$.

Sicne $\boldsymbol{R}$ is *invertible*, we solve by back substitution to get

$$\boldsymbol{x} = (\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R})^{-1}\boldsymbol{R}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b} = \boldsymbol{R}^{-1}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}.$$

■

6.3.4. Function Space

Sometimes we may also discuss orthonormal basis and Gram-Schmidt process on function space. There is a simple example:

■ **Example 6.10** For subspace $\mathrm{span}\{1, x, x^2\} \subset \mathcal{C}[-1,1]$, firstly, how to define orthogonal for the basis $\{1, x, x^2\}$?

Pre-requisite Knowledge: Inner product.

$$\langle f, g\rangle = \int_a^b fg \,\mathrm{d}x \text{ for } f, g \in C[a,b]. \qquad \|f\|^2 = \int_a^b f^2 \,\mathrm{d}x$$

If we have defined inner product, then we can talk about *orthogonality* for $\{1, x, x^2\}$. It's easy to verify that

$$\langle 1, x\rangle = 0 \quad \langle x, x^2\rangle = 0 \quad \langle 1, x^2\rangle = \frac{2}{3}.$$

If we do the Gram-Schmidt Process similarly, we obtain:

$$A = 1, \qquad B = x, \qquad C = x^2 - \frac{\langle 1, x^2 \rangle}{\langle 1, 1 \rangle} 1 - \frac{\langle x, x^2 \rangle}{\langle x, x \rangle} x = x^2 - \frac{1}{3}$$

where A, B, C are *orthogonal*. We can divide their length to obtain orthonormal basis:

$$\begin{aligned}
q_1 &= \frac{A}{\|A\|} = \frac{1}{\sqrt{\int_{-1}^{1} 1^2 \, \mathrm{d}x}} = \frac{1}{2} \\
q_2 &= \frac{B}{\|B\|} = \frac{x}{\sqrt{\int_{-1}^{1} x^2 \, \mathrm{d}x}} = \frac{x}{2/3} = \frac{3}{2} x \\
q_3 &= \frac{C}{\|C\|} = \frac{x^2 - \frac{1}{3}}{\sqrt{\int_{-1}^{1} (x^2 - \frac{1}{3})^2 \, \mathrm{d}x}} = \frac{x^2 - \frac{1}{3}}{\frac{8}{45}} = \frac{45x^2 - 15}{8}
\end{aligned}$$

Hence $\{q_1, q_2, q_3\}$ is the orthonormal basis for $\{1, x, x^2\}$. ■

■ **Example 6.11** Consider the collection $\mathcal{F}$ of functions defined on $[0, 2\pi]$, where

$$\mathcal{F} := \{1, \cos x, \sin x, \cos 2x, \sin 2x, \dots, \cos mx, \sin mx, \dots\}$$

Using various trigonometric identities, we can show that if f and g are **distinct**(different) functions in $\mathcal{F}$, we have $\int_0^{2\pi} fg \, \mathrm{d}x = 0$. For example,

$$\langle \sin x, \sin 2x \rangle = \int_0^{2\pi} \sin x \sin 2x \, \mathrm{d}x = \int_0^{2\pi} \frac{1}{2} (\cos x - \cos 3x) \, \mathrm{d}x = 0.$$

And moreover, if $f = g$, we have $\int_0^{2\pi} f^2 \, \mathrm{d}x = \pi$. For example,

$$\langle \sin 5x, \sin 5x \rangle = \int_0^{2\pi} \sin^2 5x \, \mathrm{d}x = \int_0^{2\pi} \frac{1}{2} (1 + \cos 10x) \, \mathrm{d}x = \pi.$$

In conclusion, the collection of functions $\{1, \sin mx, \cos mx\}$ for $k = 1, 2, \dots$ are *orthogonal* in $\mathcal{C}[0, 2\pi]$. Note that this set is **not orthonormal**. ■

This example gives a motivation of the fourier transformation:

6.3.5. Fourier Series

Since we have shown the orthogonality of $\mathcal{F}$ in Example.(6.11), our question is that what kind of function can be written as the linear combination of functions from $\mathcal{F}$.

The Fourier series of a function is its expansion into sines and cosines:

$$f(x) = a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + \dots$$

where $f(x) \in \mathcal{C}[0,2\pi]$. So our question turns into what kind of function could be expressed as fuourier series?

Theorem 6.7 If a function f have the finite length in its function space $\mathcal{C}[a,b]$, then it could be expressed as *fourier series*.

But how to compute the coefficients $a_i's$ and $b_j's$? The key is orthogonality! For example, in order to get a_1, we just do the inner product between $f(x)$ and $\cos x$:

Figure 6.7: Enjoy fourier series!

$$\langle f(x), \cos x\rangle = a_1\langle \cos x, \cos x\rangle + 0 \implies a_1 = \frac{\langle f(x), \cos x\rangle}{\langle \cos x, \cos x\rangle} = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos x \, dx$$

Similarly we derive

$$a_m = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos mx \, dx \qquad b_m = \frac{1}{\pi}\int_0^{2\pi} f(x)\sin mx \, dx.$$

6.4. Assignment Six

1. Find the *determinant* of the linear transformation $T(f(t)) = f(3t-2)$ from $\mathbb{P}_2$ to $\mathbb{P}_2$.
2. Suppose that $\boldsymbol{A}$ is a m by n real matrix. And suppose that $\boldsymbol{Ax} = \boldsymbol{0}$ and $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} = 2\boldsymbol{y}$. Show that $\boldsymbol{x}$ is *orthogonal* to $\boldsymbol{y}$.
3. State and justify whether the following three statements are True or False (give an example in either case):
   (a) $\boldsymbol{Q}^{-1}$ is an *orthogonal* matrix when $\boldsymbol{Q}$ is an *orthogonal* matrix.
   (b) If $\boldsymbol{Q}$ (a m by n matrix with $m > n$) has *orthonormal columns*, then $\|\boldsymbol{Qx}\| = \|\boldsymbol{x}\|$.
   (c) If $\boldsymbol{Q}$ (a m by n matrix with $m > n$) has *orthonormal columns*, then $\|\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{y}\| = \|\boldsymbol{y}\|$.
4. Let us make $P(\mathbb{R})$ into an *inner product space* using the inner product
$$\langle p, q\rangle = \int_{-1}^{1} p(x)q(x)\,\mathrm{d}x$$
Recall that we say a function is *even* if $\forall x$ we have $f(-x) = f(x)$ and *odd* if $\forall x$ we have $f(-x) = -f(x)$.
W_1 corresponds to the set of *odd polynomials* and W_2 the set of *even polynomials*. Show that $W_1 = W_2^{\perp}$.
5. Let $\boldsymbol{V} = \mathbb{R}^3$, $\boldsymbol{U}$ the *orthogonal complement* to span $\left\{\begin{pmatrix}1\\2\\-5\end{pmatrix}\right\}$. Find an *orthonormal basis* of $\boldsymbol{U}$.
6. Find the best line $C + Dt$ to fit $b = 4, 2, -1, 0, 0$ at times $t = -2, -1, 0, 1, 2$.

Chapter 7

Week6

7.1. Tuesday

7.1.1. Summary of previous weeks

In the first two weeks, we have learnt how to solve linear system of equations $\boldsymbol{Ax} = \boldsymbol{b}$. To understand this equation better, we learn the definition for matrices and vector space. The columns of matrix product $\boldsymbol{Ax}$ are the linear combination of columns of $\boldsymbol{A}$.

7.1.1.1. Determinants

Then we learnt how to describe the **quantity of a matrix**–determinant. The determinant of a square matrix is a single number. This number contains an amazing amount of information about the matrix. There are three main points about determinant:

- *Determinants is related to invertibility, rank, eigenvalue, PSD,...*
- $\det(\boldsymbol{AB}) = \det(\boldsymbol{A})\det(\boldsymbol{B})$.
- *The square matrix* $\boldsymbol{A}$ *is invertible* if and only if $\det(\boldsymbol{A}) \neq 0$.

7.1.1.2. Linear Transformation

Linear transfromation is another important topic. The matrix multiplication $T(\boldsymbol{v}) = \boldsymbol{Av}$ is essentially a linear transformation. If we consider a vector as a point in vector space, then *the linear transformation allows movements of point in the space*. It "transforms" vector $\boldsymbol{v}$ to another vector $\boldsymbol{Av}$.

In the view of linear transformation, we can understand $\det(\boldsymbol{AB}) = \det(\boldsymbol{A})\det(\boldsymbol{B})$ better:

$$\det(\boldsymbol{A}) = \text{Volumn of } \boldsymbol{Ak}, \text{ where } \boldsymbol{k} \text{ is a unit cube.}$$

If we transform the unit cube $\boldsymbol{k}$ by $\boldsymbol{A}$ secondly by $\boldsymbol{B}$, actually, it has the same effect of transforming $\boldsymbol{k}$ directly by the matrix $\boldsymbol{BA}$.

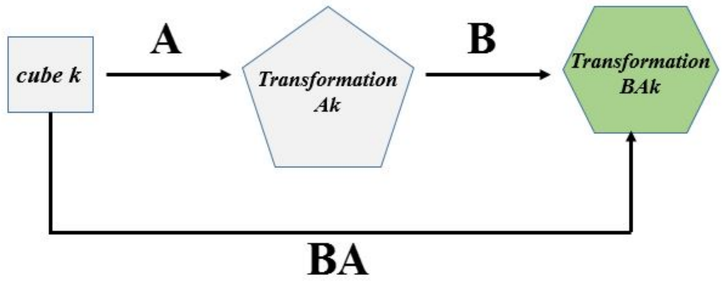


Figure 7.1: Transformation of a vector by $\boldsymbol{A}$, then by $\boldsymbol{B}$ has the same effect by $\boldsymbol{BA}$.

If we denote $\det(\cdot)$ as the volumn of a graph, since we find that the volumn of $\boldsymbol{B}(\boldsymbol{Ak})$ is exactly the same as $(\boldsymbol{BA})\boldsymbol{k}$, consequently $\det(\boldsymbol{B})\det(\boldsymbol{A}) = \det(\boldsymbol{BA})$.

Moreover, $\det(\boldsymbol{A}) = 0 \Longleftrightarrow \text{Volumn of } \boldsymbol{Ak} = 0 \Longleftrightarrow \dim(\boldsymbol{Ak}) = 0$.

Cramer's Rule also has geometric meaning, which will not be talked in this lecture. (In big data age, people will not use cramer's rule frequently due to its high computing complexity.)

Linear transformation has a matrix representation form under certain basis. **How to transform one basis into another basis?** We use *similar matrices* as the matrix representation, which will be studied in next lecture.

7.1.1.3. Orthogonality

Why we learn orthogonality? It has two motivations:

1. Linear independence between vectors $\Longleftrightarrow$ Angle $\neq 0°$.

   Similarly, we are interested in the case which the angle is 90 degrees:

$$\text{orthogonal} \Longleftrightarrow \text{Angle} = 90°$$

2. Solving least squares problem more efficiently.

   In pratical, suppose we are given two kinds of data, i.e., input: x =age of propellant and output: y =shear strength. Our data contains $S = \{(x_1, y_1), \dots, (x_n, y_n)\}$, $n = 20$ samples. Our goal is to find a best line that fit the data:

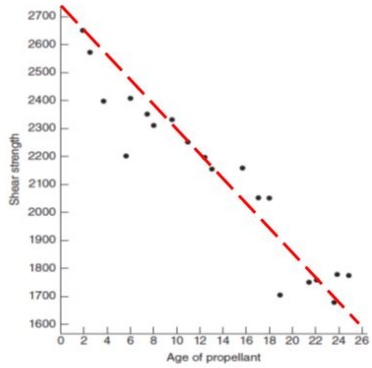


Figure 7.2: The relationship between x and y.

In other words, we want to find $\boldsymbol{x}$ s.t.

$$(\boldsymbol{A} \quad \boldsymbol{x} \approx \boldsymbol{b})$$

age coefficient strength

where

$$\boldsymbol{A} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \quad \boldsymbol{b} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

More generally, our goal is to solve the least square problem given by:

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2$$

where $\boldsymbol{b} \in \mathbb{R}^m, \boldsymbol{A} \in \mathbb{R}^{m \times n}$.

- If $\boldsymbol{b} \in \mathcal{C}(\boldsymbol{A})$, this optimization problem is converted into finding the solution to equation $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$.

- Otherwise, we want to find the least squares solution $\boldsymbol{x}^*$, which must satisfy

$$\frac{\partial}{\partial \boldsymbol{x}^*}\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|^2=\boldsymbol{0} \implies \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}^*=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}. \qquad \textbf{(normal equation.)}$$

This opotimization problem also has geometric meaning. We want to find a solution $\boldsymbol{x}^*$ such that $\boldsymbol{A}\boldsymbol{x}^*$ best approximates the vector $\boldsymbol{b}$, i.e., $\boldsymbol{A}\boldsymbol{x}^*=\text{Proj}_{\mathcal{C}(\boldsymbol{A})}(\boldsymbol{b})$.

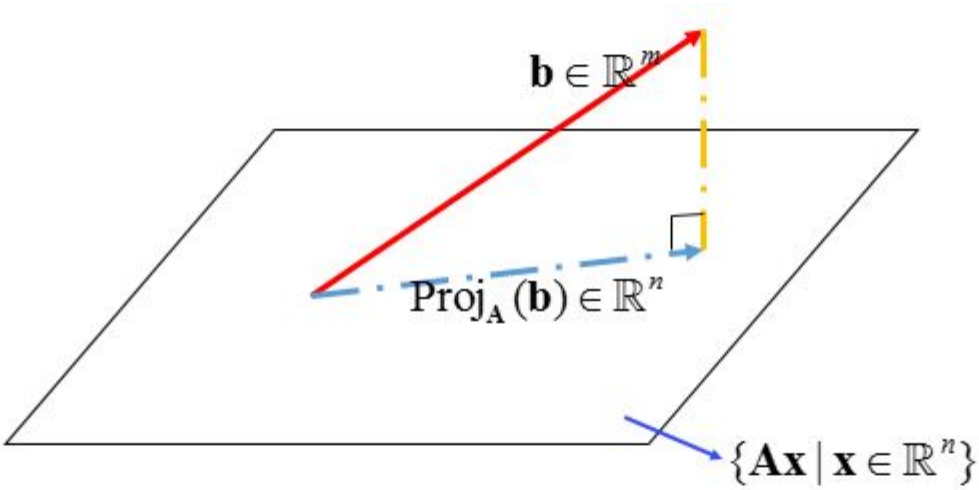


Figure 7.3: Least square problem: find $\boldsymbol{x}$ such that $\boldsymbol{A}\boldsymbol{x}=\text{Proj}_{\mathcal{C}(\boldsymbol{A})}(\boldsymbol{b})$.

The expression of the projection $\text{Proj}_{\mathcal{C}(\boldsymbol{A})}(\boldsymbol{b})$ is given by:

$$\text{Proj}_{\mathcal{C}(\boldsymbol{A})}(\boldsymbol{b})=\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}.$$

Therefore, one least squares solution is given by:

$$\boldsymbol{x}^*=(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}.$$

When $\boldsymbol{A}$ has full column rank, this solution is the unique least squares solution. (verify by yourself)

Moreover, when $\boldsymbol{A}$ is an **orthogonal matrix**, the least squares solution could be computed more efficiently:

$$\boldsymbol{x}^*=\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}.$$

7.1.2. Eigenvalues and eigenvectors

7.1.2.1. 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 $\boldsymbol{Ax}$, they are called **eigenvectors**.

$$\textbf{The key equation is}\quad \boldsymbol{Ax} = \lambda \boldsymbol{x},\quad \textbf{The number } \lambda \textbf{ is the eigenvalue of } \boldsymbol{A}.$$

Definition 7.1 [Eigenvectors and Eigenvalues] Given a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ (or $\mathbb{C}^{n\times n}$), our goalis to find a vector $\boldsymbol{v} \in \mathbb{C}^n$ with $\boldsymbol{v} \neq \boldsymbol{0}$ such that

$$\boldsymbol{Av} = \lambda \boldsymbol{v},\quad \text{for some } \lambda \in \mathbb{C} \tag{7.1}$$

- (7.1) is called an **eigenvalue problem** or **eigen-equation**
- Let $(\boldsymbol{v}, \lambda)$ be a solution to (7.1), we call
  - $(\boldsymbol{v}, \lambda)$ an **eigen-pair** of $\boldsymbol{A}$
  - λ an **eigenvalue** of $\boldsymbol{A}$; $\boldsymbol{v}$ an **eigenvector** of $\boldsymbol{A}$ associated with λ.

We illustrate an example of an eigenvalue problem:

■ **Example 7.1** 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 7.1 If $(\boldsymbol{v}, \lambda)$ is an eigen-pair of $\boldsymbol{A}$, then $(\alpha\boldsymbol{v}, \lambda)$ is also an eigen-pair for any $\alpha \in \mathbb{C}, \alpha \neq 0$.

7.1.2.2. 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 λ and $\boldsymbol{x}$ are unknowns? Consider a simpler case. If we can know the eigenvalues λ, 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 7.2 The number λ is the eigenvalue of $\boldsymbol{A}$ if and only if $\lambda\boldsymbol{I} - \boldsymbol{A}$ is singular.

$$\textbf{Equation for the eigenvalues} \quad \det(\lambda\boldsymbol{I} - \boldsymbol{A}) = 0. \tag{7.2}$$

Definition 7.2 [characteristic polynomial] Define $P_{\boldsymbol{A}}(\lambda) := \det(\lambda\boldsymbol{I} - \boldsymbol{A})$.

Then $P_{\boldsymbol{A}}(\lambda) = \det(\lambda\boldsymbol{I} - \boldsymbol{A})$ 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 $N(\lambda \boldsymbol{I} - \boldsymbol{A})$ is called the **eigenspace** associated with λ.

If $P_{\boldsymbol{A}}(\lambda^*) = 0$, then we say λ^* is the root of $P_{\boldsymbol{A}}(\lambda)$. ■

The roots of $P_{\boldsymbol{A}}(\lambda)$ are the **eigenvalues** of $\boldsymbol{A}$. $\forall \boldsymbol{x} \in N(\lambda \boldsymbol{I} - \boldsymbol{A})$ (*eigenspace*) is an eigenvector associated with λ.

■ **Example 7.2** 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 λ_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 λ_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 $\lambda_2 = -3$.

■

7.1.2.3. 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 7.1** — **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.

7.1.3. 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:

$$P_{\boldsymbol{A}}(\lambda) = \det(\lambda \boldsymbol{I} - \boldsymbol{A}) = (\lambda - \lambda_1)\dots(\lambda - \lambda_n) \tag{7.3}$$

Why the coefficient for λ^n is 1 in equation (7.3)? If we expand $\det(\lambda \boldsymbol{I} - \boldsymbol{A})$, we find

$$\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \dots & -a_{nn} \\ -a_{21} & \lambda - a_{22} & \dots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & \dots & \dots & \lambda - a_{nn} \end{vmatrix}, \tag{7.4}$$

in which the variable λ 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 (7.3), the coefficient of λ^{n-1} is

$$-(\lambda_1 + \lambda_2 + \dots + \lambda_n)$$

In (7.4), λ^{n-1} only appears among $(\lambda - a_{11})(\lambda - a_{22})\dots(\lambda - a_{nn})$, i.e., the coefficient of λ^{n-1} is

$$-(a_{11} + a_{22} + \dots + a_{nn})$$

Consequently, as (7.3) = (7.4), 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 (7.3), 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 7.2 *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 $\boldsymbol{A}$.*

7.1.4. Application: 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.

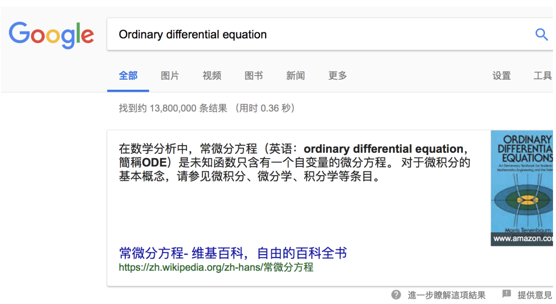


Figure 7.4: Google interface

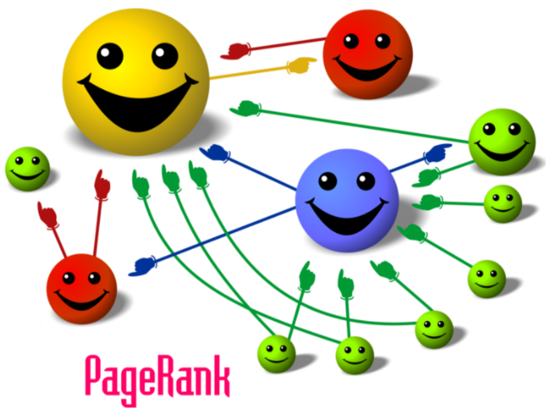


Figure 7.5: PageRank Diagram, source: Wiki

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:

$$\sum_{j \in \mathcal{L}_i} \frac{v_j}{c_j} = v_i, \quad i = 1, \ldots, n, \tag{7.5}$$

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 7.3** If we assume that there are only four pages in the world, and the diagram below shows the reference situations:

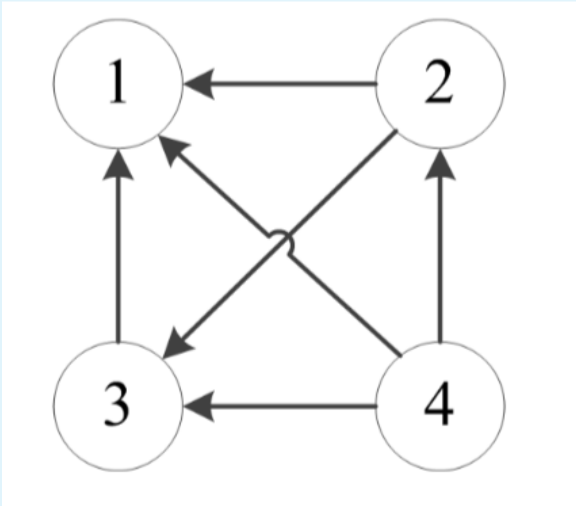


Figure 7.6: Reference situation of these four pages

Let's consider the $i = 3$ case of Eq.(7.5). 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.(7.5):

$$\begin{aligned} \frac{1}{2}v_2 + v_3 + \frac{1}{3}v_4 &= v_1 \\ \frac{1}{3}v_4 &= v_2 \\ \frac{1}{2}v_2 + \frac{1}{3}v_4 &= v_3 \\ 0 &= v_4 \end{aligned}$$

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.

7.2. Thursday

7.2.1. Review

- **Eigenvalue and eigenvectors**: If for **square** matrix $\boldsymbol{A}$ we have

$$\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}$$

where $\boldsymbol{x} \neq \boldsymbol{0}$, then we say λ is the *eigenvalue*, $\boldsymbol{x}$ is the *eigenvector* associated with λ.

- **How to compute eigenvalues and eigenvectors?** To solve the eigenvalue problem for matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, you should follow these steps:
  - *Compute the characteristic polynomial of* $\lambda \boldsymbol{I} - \boldsymbol{A}$. The determinant is a polynomial in λ of degree n.
  - *Find the roots of this polynomial*, by solving $\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = 0$. The n roots are the n eigenvalues of $\boldsymbol{A}$. They make $\boldsymbol{A} - \lambda \boldsymbol{I}$ singular.
  - For each eigenvalue λ, *solve* $(\lambda \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$ *to find a corresponding eigenvector* $\boldsymbol{x}$.

7.2.2. Similarity

The similar matrices have the same eigenvalues:

Definition 7.3 [Similar] If there exists a *nonsingular* matrix $\boldsymbol{S}$ such that

$$\boldsymbol{B} = \boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S},$$

then we say $\boldsymbol{A}$ is **similar** to $\boldsymbol{B}$. ∎

Proposition 7.3 Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be $n \times n$ matrices. If $\boldsymbol{B}$ is *similar* to $\boldsymbol{A}$, then $\boldsymbol{A}$ and $\boldsymbol{B}$ have the same eigenvalues.

Proofidea. Since eigenvalues are the roots of the *characteristic polynomial*, so it suffices to prove these two polynomials are the same.

Proof. The *characteristic polynomial* for $\boldsymbol{B}$ is given by

$$\begin{aligned} P_{\boldsymbol{B}}(\lambda) &= \det(\lambda \boldsymbol{I} - \boldsymbol{B}) \\ &= \det(\lambda \boldsymbol{I} - \boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}) = \det(\boldsymbol{S}^{-1}\lambda \boldsymbol{I}\boldsymbol{S} - \boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}) \\ &= \det(\boldsymbol{S}^{-1}(\lambda \boldsymbol{I} - \boldsymbol{A})\boldsymbol{S}) \\ &= \det(S^{-1})\det(\lambda \boldsymbol{I} - \boldsymbol{A})\det(\boldsymbol{S}) \end{aligned}$$

Since $\det(\boldsymbol{S}^{-1})\det(\boldsymbol{S}) = 1$, we obtain:

$$\begin{aligned} P_{\boldsymbol{B}}(\lambda) &= \det(\lambda \boldsymbol{I} - \boldsymbol{A}) \\ &= P_{\boldsymbol{A}}(\lambda). \end{aligned}$$

Since they have the same *characteristic polynomial*, the roots for *characteristic polynomials* of $\boldsymbol{A}$ and $\boldsymbol{B}$ must be same. Therefore they have the same eigenvalues. ■

R What is invarient? In other words, what is not changed during matrix transformation?

- **Rank** is invarient under *row transformation*.
- **Eigenvalues** is invarient undet *similar transformation*.
- Unluckily, similar matrices usually don't have the same eigenvectors. It's easy to raise a counterexample.

By using eigenvalues, we have a new proof for $\det(\boldsymbol{S}^{-1}) = \frac{1}{\det(\boldsymbol{S})}$:

Proof. Suppose $\det(\boldsymbol{S}) = \lambda_1\lambda_2 \dots \lambda_n$, where λ_i's are eigenvalues of $\boldsymbol{S}$. Then there exists $\boldsymbol{x}_i$ such that

$$\boldsymbol{S}\boldsymbol{x}_i = \lambda_i \boldsymbol{x}_i$$

for $i = 1, \dots, n$.

Since $\boldsymbol{S}$ is invertible and all λ_i's are nonzero, we imply that:

$$\boldsymbol{S}\boldsymbol{x}_i = \lambda_i \boldsymbol{x}_i \implies \boldsymbol{x}_i = \lambda_i \boldsymbol{S}^{-1}\boldsymbol{x}_i \implies \boldsymbol{S}^{-1}\boldsymbol{x}_i = \frac{1}{\lambda_i}\boldsymbol{x}_i$$

Hence, $\frac{1}{\lambda_i}$'s are eigenvalues of $\boldsymbol{S}^{-1}$. Since $S^{-1} \in \mathbb{R}^{n\times n}$, $\frac{1}{\lambda_i}$'s $(i = 1,\dots,n)$ are the only eigenvalues of $\boldsymbol{S}^{-1}$.

Hence the determinant of $\boldsymbol{S}^{-1}$ is the product of its eigenvalues:

$$\det(\boldsymbol{S}^{-1}) = \frac{1}{\lambda_1}\frac{1}{\lambda_2}\dots\frac{1}{\lambda_n} = \frac{1}{\det(\boldsymbol{S})}.$$

■

We can also use eigenvalue to proof the statement shown below:

Proposition 7.4 $\boldsymbol{A}$ is singular if and only if $\det(\boldsymbol{A}) = 0$.

Proof. Suppose $\det(\boldsymbol{A}) = \lambda_1\lambda_2\dots\lambda_n$, where λ_i's are eigenvalues of $\boldsymbol{A}$.

Thus

$$\det(\boldsymbol{A}) = 0 \iff \exists\lambda_i = 0 \iff \exists \text{ nonzero } \boldsymbol{x} \text{ s.t. } \boldsymbol{A}\boldsymbol{x} = \lambda_i\boldsymbol{x} = 0\boldsymbol{x} = \boldsymbol{0}.$$

Or equivalently, $\boldsymbol{A}$ is singular. ■

7.2.3. Diagonalization

Proposition (7.3) says if $\boldsymbol{A}$ is similar to $\boldsymbol{B}$, then they have the same eigenvalues.

Question 1. What about the reverse direction?

Question 2. We all approve that the simplest form of a matrix to have eigenvalues $\lambda_1,\dots,\lambda_n$ is the diagonal matrix $\text{diag}(\lambda_1,\dots,\lambda_n)$. Suppose $\boldsymbol{A}$ has eigenvalues $\lambda_1,\dots,\lambda_n$, is $\boldsymbol{A}$ similar to the diagonal matrix $\text{diag}(\lambda_1,\dots,\lambda_n)$?

R Why the matrix $\text{diag}(\lambda_1,\dots,\lambda_n)$ has eigenvalues $\lambda_1,\dots,\lambda_n$?

Answer: Let's explain it with $n = 2$:

$$\begin{pmatrix} \lambda_1 & \\ & \lambda_2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \lambda_1 \\ 0 \end{pmatrix} = \lambda_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \begin{pmatrix} \lambda_1 & \\ & \lambda_2 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ \lambda_2 \end{pmatrix} = \lambda_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

The case for general n is also easy to verify.

The answers to Question 1 and 2 are both **No**! Let's raise a counterexample to explain it:

■ **Example 7.4** We give a counterexample to show that two matrices with the same eigenvalues are not necessarily similar to each other; and $\boldsymbol{A}$ does not necessarily similar to the corresponding diagonal matrix.

Given $\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, then $P_{\boldsymbol{A}}(\lambda) = \det(\lambda \boldsymbol{I} - \boldsymbol{A}) = \begin{vmatrix} \lambda & -1 \\ 0 & \lambda \end{vmatrix}$. Hence its eigenvalues are $\lambda_1 = \lambda_2 = 0$.

Hence, $\boldsymbol{A}$ and $\boldsymbol{D} = \text{diag}(0,0)$ have the same eigenvalues. Then we show that $\boldsymbol{A}$ and $\boldsymbol{D}$ are not similar:

Assume they are similar, which means there exists invertible matrix $\boldsymbol{S}$ such that

$$\boldsymbol{A} = \boldsymbol{S}^{-1}\boldsymbol{D}\boldsymbol{S} = \boldsymbol{S}^{-1} \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \boldsymbol{S} = \boldsymbol{0} \implies \text{contradiction!}$$

■

Suppose $\boldsymbol{A}$ has eigenvalues $\lambda_1, \dots, \lambda_n$, but $\boldsymbol{A}$ and $\text{diag}(\lambda_1, \dots, \lambda_n)$ may not be similar! We are curious about what kind of matrix can be similar to a diagonal matrix:

Definition 7.4 [Diagonalizable] An $n \times n$ matrix $\boldsymbol{A}$ is **diagonalizable** if $\boldsymbol{A}$ is similar to a *diagonal matrix*, that is to say, $\exists$ nonsingular matrix $\boldsymbol{S}$ and diagonal matrix $\boldsymbol{D}$ such that

$$\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S} = \boldsymbol{D} \tag{7.6}$$

We say $\boldsymbol{S}$ *diagonalizes* $\boldsymbol{A}$. ■

(R) Note that Eq.(7.6) can be equivalently written as $\boldsymbol{AS}=\boldsymbol{SD}$, or in column-by-column form:

$$\boldsymbol{A}\boldsymbol{s}_i = d_i\boldsymbol{s}_i, \quad i=1,\dots,n, \tag{7.7}$$

where $\boldsymbol{s}_i$ denotes the ith column of $\boldsymbol{S}$, d_i denotes the (i,i)th entry of $\boldsymbol{D}$. The equivalent form Eq.(7.7) also implies that every $(\boldsymbol{s}_i, d_i)$ must be an eigen-pair of $\boldsymbol{A}$. (Proposition (7.5))

Proposition 7.5 Suppose that $\boldsymbol{A}$ is diagonalizable, then the column vectors of the diagonalizing matrix $\boldsymbol{S}$ are eigenvectors of $\boldsymbol{A}$; and the diagonal elements of $\boldsymbol{D}$ are the corresponding eigenvalues of $\boldsymbol{A}$.

Proposition 7.6 The diagonalizing matrix $\boldsymbol{S}$ is not unique.

Proof. Suppose there exists a diagonalizing matrix $\boldsymbol{S}$, verify by yourself that $\alpha\boldsymbol{S}$ is also a a diagonalizing matrix for any $\alpha \neq 0$. ■

(R) We know that the reverse of proposition (7.3) is not true. However, if we add one more constraint that all eigenvalues of $\boldsymbol{A}$ are distinct, the reverse is true. We will give a proof of it later.

1. If $\boldsymbol{A}$ is $n \times n$ and A has n distinct eigenvalues, then $\boldsymbol{A}$ is diagonalizable. If the eigenvalues are not distinct, then $\boldsymbol{A}$ may or may not be diagonalizable depending on whether $\boldsymbol{A}$ has n linearly independent eigenvectors.

Why is diagonalizable good?

Theorem 7.3 — Diagonalization. A $n \times n$ matrix $\boldsymbol{A}$ is *diagonalizable* iff $\boldsymbol{A}$ has n independent eigenvectors.

Proof. Necessity. For n eigen-pairs $(\lambda_i, \boldsymbol{x}_i)$ of $\boldsymbol{A}$, suppose that $\boldsymbol{x}_i$'s are independent.

We after-multiply $\boldsymbol{A}$ with $\boldsymbol{S} = \begin{bmatrix}\boldsymbol{x}_1 & \boldsymbol{x}_2 & \dots & \boldsymbol{x}_n\end{bmatrix}$. The first column of $\boldsymbol{AS}$ is $\boldsymbol{A}\boldsymbol{x}_1 = \lambda_1\boldsymbol{x}_1$. Hence we obtain the result for the product $\boldsymbol{AS}$:

$$\textbf{A times S} \quad \boldsymbol{AS} = \boldsymbol{A}\begin{bmatrix}\boldsymbol{x}_1 & \boldsymbol{x}_2 & \dots & \boldsymbol{x}_n\end{bmatrix} = \begin{bmatrix}\lambda_1\boldsymbol{x}_1 & \lambda_2\boldsymbol{x}_2 & \dots & \lambda_n\boldsymbol{x}_n\end{bmatrix}. \tag{7.8}$$

Note that the right side of Eq.(7.8) is essentially the product $\boldsymbol{SD}$:

$$\textbf{\textit{S} times \textit{D}} \quad \begin{bmatrix} \lambda_1\boldsymbol{x}_1 & \lambda_2\boldsymbol{x}_2 & \dots & \lambda_n\boldsymbol{x}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \dots & \boldsymbol{x}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} = \boldsymbol{SD}.$$

Hence we obtain $\boldsymbol{AS} = \boldsymbol{SD}$. Since $\boldsymbol{x}_i$'s are independent, there exists the inverse $\boldsymbol{S}^{-1}$.

Therefore, $\boldsymbol{D} = \boldsymbol{S}^{-1}\boldsymbol{AS}$.

Sufficiency. If $\boldsymbol{A}$ is diagonalizable, then there exists $\boldsymbol{S}$ and $\boldsymbol{D}$ such that

$$\boldsymbol{D} = \boldsymbol{S}^{-1}\boldsymbol{AS} \tag{7.9}$$

where $\boldsymbol{S}$ is nonsingular. Suppose $\boldsymbol{D} = \operatorname{diag}(\lambda_1,\dots,\lambda_n)$, and $\boldsymbol{S} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \dots & \boldsymbol{x}_n \end{bmatrix}$, where $\boldsymbol{x}_i$'s are independent.

The Eq.(7.9) can be equivalently written as $\boldsymbol{AS} = \boldsymbol{SD}$, i.e., $\boldsymbol{Ax}_i = \lambda_i\boldsymbol{x}_i$ for $i = 1,2,\dots,n$.

Hence $\boldsymbol{x}_i$'s are the independent eigenvectors of $\boldsymbol{A}$ associated with λ_i's. ■

Diagonalizable matrix is very useful. For diagonalizable matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, it follows that its eigenvectors $\{\boldsymbol{x}_1,\dots,\boldsymbol{x}_n\}$ are independent, i.e., form a basis for $\mathbb{R}^n$. Then for any $\boldsymbol{y} \in \mathbb{R}^n$, there exists $(c_1,c_2,\dots,c_n)$ such that

$$\boldsymbol{y} = c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 + \dots + c_n\boldsymbol{x}_n$$

If we consider matrix $\boldsymbol{A}$ as representation of linear transformation, we obtain

$$\begin{aligned} \boldsymbol{Ay} &= c_1\boldsymbol{Ax}_1 + \dots + c_n\boldsymbol{Ax}_n \\ &= c_1\lambda_1\boldsymbol{x}_1 + \dots + c_n\lambda_n\boldsymbol{x}_n \end{aligned}$$

Hence, the linear transformation from $\boldsymbol{y}$ into $\boldsymbol{A}\boldsymbol{y}$ is equivalent to transforming the coordinate coefficients from $(c_1,\dots,c_n)$ into $(c_1\lambda_1,\dots,c_n\lambda_n)$:

$$\boldsymbol{y} \overset{\boldsymbol{A}}{\Longrightarrow} \boldsymbol{A}\boldsymbol{y}$$

$$(c_1,\dots,c_n) \xrightarrow{\boldsymbol{D}=\text{diag}(\lambda_1,\dots,\lambda_n)} (c_1\lambda_1,\dots,c_n\lambda_n) = (c_1,\dots,c_n)\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

We are curious about whether there is an useful way to determine whether $\boldsymbol{A}$ is diagonalizable.

Theorem 7.4 If $\lambda_1,\dots,\lambda_k$ are *distinct* eigenvalues of a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}(n \geq k)$ with the corresponding eigenvectors $\boldsymbol{x}_1,\dots,\boldsymbol{x}_k$, then $\boldsymbol{x}_1,\dots,\boldsymbol{x}_k$ are linearly independent.

Proof.

- Let's start with the case $k = 2$. Assume that $\lambda_1 \neq \lambda_2$ but $\boldsymbol{x}_1,\boldsymbol{x}_2$ are dependent, i.e., $\exists(c_1,c_2) \neq \boldsymbol{0}$ s.t.

$$c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 = \boldsymbol{0}. \tag{7.10}$$

Postmultiplying $\boldsymbol{A}$ for Eq.(7.10) both sides results in

$$\boldsymbol{A}(c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2) = \boldsymbol{0} \implies c_1\lambda_1\boldsymbol{x}_1 + c_2\lambda_2\boldsymbol{x}_2 = \boldsymbol{0}. \tag{7.11}$$

Eq.(7.10)$\times\lambda_2-$Eq.(7.11) results in:

$$(c_1\lambda_2 - c_1\lambda_1)\boldsymbol{x} = \boldsymbol{0}. \implies c_1(\lambda_2 - \lambda_1)\boldsymbol{x} = \boldsymbol{0}.$$

Since $\lambda_1 \neq \lambda_2$ and $\boldsymbol{x} \neq \boldsymbol{0}$, we derive $c_2 = 0$. Similarly, if we let Eq.(7.10)$\times\lambda_1-$Eq.(7.11) to cancel c_2, then we get $c_1 = 0$.

Therefore, $(c_1,c_2) = \boldsymbol{0}$ leads to a contradiction!

- How to proof this statement for general k?

Assume there exists $(c_1,\dots,c_k) \neq \mathbf{0}$ s.t.

$$c_1\boldsymbol{x}_1 + \cdots + c_k\boldsymbol{x}_k = \mathbf{0} \tag{7.12}$$

Then we obtain two equations from Eq.(7.12):

$$\boldsymbol{A}(c_1\boldsymbol{x}_1 + \cdots + c_k\boldsymbol{x}_k) = c_1\lambda_1\boldsymbol{x}_1 + c_2\lambda_2\boldsymbol{x}_2 + \cdots + c_k\lambda_k\boldsymbol{x}_k = \mathbf{0}. \tag{7.13}$$

$$\lambda_k(c_1\boldsymbol{x}_1 + \cdots + c_k\boldsymbol{x}_k) = c_1\lambda_k\boldsymbol{x}_1 + c_2\lambda_k\boldsymbol{x}_2 + \cdots + c_k\lambda_k\boldsymbol{x}_k = \mathbf{0}. \tag{7.14}$$

We can let Eq.(7.13)−Eq.(7.14) to cancel $\boldsymbol{x}_k$:

$$c_1(\lambda_1 - \lambda_k)\boldsymbol{x}_1 + \cdots + c_k(\lambda_{k-1} - \lambda_k)\boldsymbol{x}_{k-1} = \mathbf{0}. \tag{7.15}$$

By repeatedly applying the trick from (7.12) to (7.15), we can show that

$$c_1(\lambda_1 - \lambda_k)\dots(\lambda_1 - \lambda_2)\boldsymbol{x}_1 = \mathbf{0} \qquad \text{which forces } c_1 = 0.$$

Similarly every $c_i = 0$ for $i = 1,\dots,n$. Here is the contradiction!

■

Corollary 7.1 If all eigenvalues of $\boldsymbol{A}$ are *distinct*, then $\boldsymbol{A}$ is *diagonalizable*

7.2.4. Powers of $\boldsymbol{A}$

Matrix Powers. If $\boldsymbol{A} = \boldsymbol{S}^{-1}\boldsymbol{D}\boldsymbol{S}$, then $\boldsymbol{A}^2 = (\boldsymbol{S}^{-1}\boldsymbol{D}\boldsymbol{S})(\boldsymbol{S}^{-1}\boldsymbol{D}\boldsymbol{S}) = \boldsymbol{S}^{-1}\boldsymbol{D}^2\boldsymbol{S}$.

In general, $\boldsymbol{A}^k = (\boldsymbol{S}^{-1}\boldsymbol{D}\boldsymbol{S})\dots(\boldsymbol{S}^{-1}\boldsymbol{D}\boldsymbol{S}) = \boldsymbol{S}^{-1}\boldsymbol{D}^k\boldsymbol{S}$.

Eigenvalues of matrix powers. We may ask if eigenvalues of $\boldsymbol{A}$ are $\lambda_1,\dots,\lambda_n$, then what is the eigenvalues of $\boldsymbol{A}^k$? The answer is intuitive, the eigenvalues of $\boldsymbol{A}^k$ are $\lambda_1^k,\dots,\lambda_n^k$. However, you may use the wrong way to prove this statement:

Proposition 7.7 If eigenvalues of $n \times n$ matrix $\boldsymbol{A}$ are $\lambda_1,\dots,\lambda_n$, then eigenvalues of $\boldsymbol{A}^k$ are $\lambda_1^k,\dots,\lambda_n^k$.

Wrong proof 1: Assume $\boldsymbol{A} = \boldsymbol{S}^{-1}\boldsymbol{D}\boldsymbol{S}$, then $\boldsymbol{A}^k = \boldsymbol{S}^{-1}\boldsymbol{D}^k\boldsymbol{S}$. Suppose $\boldsymbol{D} = \text{diag}(\lambda_1, \dots, \lambda_n)$, then $\boldsymbol{D}^k = \text{diag}(\lambda_1^k, \dots, \lambda_n^k)$. Hence eigenvalues of $\boldsymbol{A}^k$ are $\lambda_1^k, \dots, \lambda_n^k$.

This proof is wrong, because $\boldsymbol{A}$ may not be *diagonalizable*, which means $\boldsymbol{A}$ may not have the form $\boldsymbol{A} = \boldsymbol{S}^{-1}\boldsymbol{D}\boldsymbol{S}$. ∎

Wrong proof 2: If $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$, then $\boldsymbol{A}^2\boldsymbol{x} = \boldsymbol{A}(\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{A}(\lambda\boldsymbol{x}) = \lambda(\boldsymbol{A}\boldsymbol{x}) = \lambda^2\boldsymbol{x}$.
Hence for general k, $\boldsymbol{A}^k\boldsymbol{x} = \lambda^k\boldsymbol{x}$.

This proof only states that if λ is the eigenvalue of $\boldsymbol{A}$, then λ^k is the eigenvalues of $\boldsymbol{A}^k$. Unfortunately, it still cannot derive this proposition. Because it does not prove that if λ are the eigenvalues with multiplicity m, then λ^k are the eigenvalues of $\boldsymbol{A}^k$ with multiplicity m.
Let's raise a counterexample: Let eigenvalues of $\boldsymbol{A}$ be $\lambda_1 = 1, \lambda_2 = 1, \lambda_3 = 2$; the eigenvalues of $\boldsymbol{A}^2$ could be $1^2, 2^2, 2^2$. Hence $\boldsymbol{A}$ has the eigenvalues 1 with multiplicity 2; while $\boldsymbol{A}^2$ has the eigenvalue 1^2 with multiplicity 1. So this $\boldsymbol{A}$ and $\boldsymbol{A}^2$ is a contradiction for this proof. In other words, this proof fails to determine the multiplicity of eigenvalues. ∎

R The proposition(7.7) could be proved using **Jordan form**, i.e., for any matrix $\boldsymbol{A}$ there exists invertible matrix $\boldsymbol{S}$ such that $\boldsymbol{A} = \boldsymbol{S}^{-1}\boldsymbol{U}\boldsymbol{S}$, where $\boldsymbol{U}$ is an upper triangular matrix with diagonal entries $\lambda_1, \dots, \lambda_n$. Then $\boldsymbol{A}^k = \boldsymbol{S}^{-1}\boldsymbol{U}^k\boldsymbol{S}$, where $\boldsymbol{U}^k$ is an upper triangular matrix with diagonal entries $\lambda_1^k, \dots, \lambda_n^k$. Hence the eigenvalues of $\boldsymbol{A}^k$ are $\lambda_1^k, \dots, \lambda_n^k$.

7.2.5. Nondiagonalizable Matrices

Sometimes we face some matrices that have too few eigenvalues. (don't count with multiplicity)

For example, given $\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, it's easy to verify that its eigenvalue is $\lambda = 0$ and eigenvectors are of the form $\boldsymbol{x} = \begin{bmatrix} c \\ 0 \end{bmatrix}$.

This 2×2 matrix cannot be diagonalized. Why? Let's introduce the definition for multiplicity first:

Definition 7.5 [Multiplicity] Suppose matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ has k **distinct** eigenvalues λ_i for $i = 1, 2, \ldots, k$.

- The **algebraic multiplicity** of an eigenvalue $\lambda_i, i \in \{1, 2, \ldots, k\}$ is defined as the *number of times that* λ_i *appears as a root of the* $\det(\boldsymbol{A} - \lambda \boldsymbol{I})$. We denote the *algebraic multiplicity of* λ_i as m_i. In other words, we denote m_i as the number of repeated eigenvalues of λ_i.
- The **geometric multiplicity** of an eigenvalue $\lambda_i, i \in \{1, 2, \ldots, k\}$ is defined as the *maximal number of* ***linearly independent*** *eigenvectors associated with* λ_i. We denote the *geometric multiplicity of* λ_i as q_i. Note that $q_i = \dim(N(\boldsymbol{A} - \lambda_i \boldsymbol{I}))$.

■

Proposition 7.8 We have $m_i \geq q_i$ for $i = 1, 2, \ldots, k$.

The implication is that **the number of repeated eigenvalues of $\lambda_i \geq$ the number of linearly independent eigenvectors associated with λ_i.**

Note that $m_i > q_i$ is possible, let's raise an example:

■ **Example 7.5**

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

We can verify that the roots of $\det(\boldsymbol{A} - \lambda \boldsymbol{I})$ are $\lambda_1 = \lambda_2 = \lambda_3 = 0$. Thus we have $k = 1, m_1 = 3$.

However, we can also verify that

$$N(\lambda - \lambda_1 \boldsymbol{I}) = N(\boldsymbol{A}) = \operatorname{span}\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$$

Consequently, $q_1 = \dim(N(\boldsymbol{A} - \lambda_1 \boldsymbol{I})) = 2$. Thus $m_1 > q_1$. ■

Proof for proposition. For convenience, we let $\lambda_0 \in \{\lambda_1, \dots, \lambda_k\}$ be any eigenvalue of $\boldsymbol{A}$, and we denote $q = \dim(N(\boldsymbol{A} - \lambda_0 \boldsymbol{I}))$. We only need to show that $\det(\boldsymbol{A} - \lambda \boldsymbol{I})$ has at least q repeated roots for $\lambda = \lambda_0$.

Firstly, let's focus on **real** eigenvalues and **real** eigenvectors:

- From concepts for subspace, we can find a collection of **orthonormal** vectors $\boldsymbol{v}_1, \dots, \boldsymbol{v}_q \in N(\boldsymbol{A} - \lambda_0 \boldsymbol{I})$ and a collection of vectors $\boldsymbol{v}_{q+1}, \dots, \boldsymbol{v}_n \in \mathbb{R}^n$ such that

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix} \text{ is orthogonal.}$$

Let $\boldsymbol{V}_1 = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_q \end{bmatrix}$, $\boldsymbol{V}_2 = \begin{bmatrix} \boldsymbol{v}_{q+1} & \boldsymbol{v}_{q+2} & \cdots & \boldsymbol{v}_n \end{bmatrix}$ and note $\boldsymbol{V} = \begin{bmatrix} \boldsymbol{V}_1 & \boldsymbol{V}_2 \end{bmatrix}$. Thus we have

$$\boldsymbol{V}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{V} = \begin{bmatrix} \boldsymbol{V}_1^{\mathrm{T}} \\ \boldsymbol{V}_2^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \boldsymbol{A}\boldsymbol{V}_1 & \boldsymbol{A}\boldsymbol{V}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{V}_1^{\mathrm{T}} \boldsymbol{A} \boldsymbol{V}_1 & \boldsymbol{V}_1^{\mathrm{T}} \boldsymbol{A} \boldsymbol{V}_2 \\ \boldsymbol{V}_2^{\mathrm{T}} \boldsymbol{A} \boldsymbol{V}_1 & \boldsymbol{V}_2^{\mathrm{T}} \boldsymbol{A} \boldsymbol{V}_2 \end{bmatrix}$$

Since $\boldsymbol{A}\boldsymbol{v}_i = \lambda_0 \boldsymbol{v}_i$ for $i = 1, 2, \dots, q$, we get $\boldsymbol{A}\boldsymbol{V}_1 = \lambda_0 \boldsymbol{V}_1$. By also noting that $\boldsymbol{V}_1^{\mathrm{T}} \boldsymbol{V}_1 = \boldsymbol{I}$ and $\boldsymbol{V}_2^{\mathrm{T}} \boldsymbol{V}_1 = \boldsymbol{0}$, we can simplify the above matrix equation into:

$$\boldsymbol{V}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{V} = \begin{bmatrix} \lambda_0 \boldsymbol{I} & \boldsymbol{V}_1^{\mathrm{T}} \boldsymbol{A} \boldsymbol{V}_2 \\ \boldsymbol{0} & \boldsymbol{V}_2^{\mathrm{T}} \boldsymbol{A} \boldsymbol{V}_2 \end{bmatrix}$$

It follows that

$$\begin{aligned} \det(\boldsymbol{A} - \lambda \boldsymbol{I}) = \det(\boldsymbol{V}^{\mathrm{T}} (\boldsymbol{A} - \lambda \boldsymbol{I}) \boldsymbol{V}) &= \det(\boldsymbol{V}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{V} - \lambda \boldsymbol{I}) \\ &= \det \begin{pmatrix} (\lambda_0 - \lambda) \boldsymbol{I} & \boldsymbol{V}_1^{\mathrm{T}} \boldsymbol{A} \boldsymbol{V}_2 \\ \boldsymbol{0} & \boldsymbol{V}_2^{\mathrm{T}} \boldsymbol{A} \boldsymbol{V}_2 - \lambda \boldsymbol{I} \end{pmatrix} \\ &= (\lambda_0 - \lambda)^q \det(\boldsymbol{V}_2^{\mathrm{T}} \boldsymbol{A} \boldsymbol{V}_2 - \lambda \boldsymbol{I}) \end{aligned}$$

Here $\det(\boldsymbol{V}_2^{\mathrm{T}} \boldsymbol{A} \boldsymbol{V}_2 - \lambda \boldsymbol{I})$ is a polynomial of degree of $n - q$. From the above equation we see that $\det(\boldsymbol{A} - \lambda \boldsymbol{I})$ has at least q repeated roots for $\lambda = \lambda_0$.

Secondly, the complex eigenvalues and eigenvectors could be proved by extending **orthogonal** matrix into **unitary** matrix.

The proof is complete. ∎

Proposition 7.9 A matrix is not diagonalizable if and only if there exists an eigenvalue such that its corresponding algebraic multiplicity is strictly larger than the corresponding geometric multiplicity.

Proof. The following statements are equivalent:

- The matrix $\boldsymbol{A} \in \ \times$ is not diagonalizble
- Any n eigenvectors of $\boldsymbol{A}$ cannot be independent.
- The sum of the dimensions of all eigenspace of $\boldsymbol{A}$ is strictly less than n, i.e., the sum of the algebraic multiplicity of all eigenvalues of $\boldsymbol{A}$
- There exists an eigenvalue such that the corresponding geometric multiplicity is strictly less than the corresponding algebraic multiplicity.

∎

7.3. Friday

7.3.1. Review

- **Diagonalization:** Suppose the matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ is diagonalizable, it's equivalent to say it has n independent eigenvectors. These n independent eigenvectors form a basis for $\mathbb{R}^n$. $(*)$
- If all *eigenvalues of* $\boldsymbol{A}$ *are distinct*, then $(*)$ holds.

7.3.2. Fibonacci Numbers

We show a famous example, where the eigenvalues tell how to find the formula for Fibonacci Numbers.

Every new Fibonacci number come from two previous ones.

Fibonacci Number: $0, 1, 1, 2, 3, 5, 8, 13, \dots$

Fibonacci Equation: $\boldsymbol{F}_{k+2} = \boldsymbol{F}_{k+1} + \boldsymbol{F}_k, \quad \boldsymbol{F}_0 = 0, \boldsymbol{F}_1 = 1.$

How to compute $\boldsymbol{F}_{100}$ without computing $\boldsymbol{F}_2$ to $\boldsymbol{F}_{99}$?. The key is to begin with a matrix equation $\boldsymbol{u}_{k+1} = \boldsymbol{A}\boldsymbol{u}_k$. We put two Fibonacci number into a vector $\boldsymbol{u}_k$, then you will see the matrix $\boldsymbol{A}$:

Define $\boldsymbol{u}_k := \begin{bmatrix} \boldsymbol{F}_{k+1} \\ \boldsymbol{F}_k \end{bmatrix}$. The rule $\begin{cases} \boldsymbol{F}_{k+2} = \boldsymbol{F}_{k+1} + \boldsymbol{F}_k \\ \boldsymbol{F}_0 = 0, \boldsymbol{F}_1 = 1 \end{cases}$ implies that

$$\boldsymbol{u}_{k+1} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \boldsymbol{u}_k, \quad \boldsymbol{u}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Every step we mutliply $\boldsymbol{u}_0$ by $\boldsymbol{A}$. After 100 steps we obtain $\boldsymbol{u}_{100} = \boldsymbol{A}^{100}\boldsymbol{u}_0$:

$$\boldsymbol{u}_{100} = \begin{bmatrix} \boldsymbol{F}_{101} \\ \boldsymbol{F}_{100} \end{bmatrix} = \boldsymbol{A}^{100}\boldsymbol{u}_0 = \boldsymbol{A}^{100} \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

How to compute the matrix powers $\boldsymbol{A}^{100}$? Diagonalizing $\boldsymbol{A}$ is possible. It's easy to verify that the matrix $\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ can be decomposed into $\boldsymbol{A} = \boldsymbol{S}\boldsymbol{D}\boldsymbol{S}^{-1}$, where

$$\boldsymbol{D} = \text{diag}(\lambda_1, \lambda_2), \quad \boldsymbol{S} = \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix},$$
$$\left(\lambda_1, \begin{bmatrix} \lambda_1 \\ 1 \end{bmatrix}\right), \qquad \left(\lambda_2, \begin{bmatrix} \lambda_2 \\ 1 \end{bmatrix}\right) \text{ are two eigen-pairs of } \boldsymbol{A},$$

with $\lambda_1 = \frac{1+\sqrt{5}}{2}, \lambda_2 = \frac{1-\sqrt{5}}{2}$.

If follows that $\boldsymbol{A}^{100} = \boldsymbol{S}\boldsymbol{D}^{100}\boldsymbol{S}^{-1}$. Hence we can compute $\boldsymbol{u}_{100}$:

$$\begin{aligned} \boldsymbol{u}_{100} = \boldsymbol{A}^{100}\boldsymbol{u}_0 = \boldsymbol{S}\boldsymbol{D}^{100}\boldsymbol{S}^{-1}\boldsymbol{u}_0 &= \boldsymbol{S} \begin{pmatrix} \lambda_1^{100} & \\ & \lambda_2^{100} \end{pmatrix} \boldsymbol{S}^{-1}\boldsymbol{u}_0 \\ &= \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix} \begin{pmatrix} \lambda_1^{100} & \\ & \lambda_2^{100} \end{pmatrix} \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{F}_{101} \\ \boldsymbol{F}_{100} \end{bmatrix} \end{aligned}$$

After messy computation, we obtain $\boldsymbol{F}_{100}$:

$$\boldsymbol{F}_{100} = \frac{1}{\sqrt{5}} \left[\lambda_1^{100} - \lambda_2^{100}\right] = \frac{1}{\sqrt{5}} \left[\left(\frac{1+\sqrt{5}}{2}\right)^{100} - \left(\frac{1-\sqrt{5}}{2}\right)^{100}\right]$$

Another way to compute $\boldsymbol{F}_{100}$. As $\boldsymbol{u}_{k+1} = \boldsymbol{A}\boldsymbol{u}_k$, we apply a trick to simplify $\boldsymbol{u}_0$ at first:

We set $\boldsymbol{S} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 \end{bmatrix}$, where $\boldsymbol{x}_1 = \begin{bmatrix} \lambda_1 \\ 1 \end{bmatrix}, \boldsymbol{x}_2 = \begin{bmatrix} \lambda_2 \\ 1 \end{bmatrix}$. It follows that

$$\boldsymbol{u}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{1}{\lambda_1 - \lambda_2} \left(\begin{bmatrix} \lambda_1 \\ 1 \end{bmatrix} - \begin{bmatrix} \lambda_2 \\ 1 \end{bmatrix} \right) \implies \boldsymbol{u}_0 = \frac{\boldsymbol{x}_1 - \boldsymbol{x}_2}{\lambda_1 - \lambda_2}$$

Then we multiply $\boldsymbol{u}_0$ by $\boldsymbol{A}^{100}$ to get $\boldsymbol{u}_{100}$:

$$\begin{aligned}
\boldsymbol{u}_{100} = \boldsymbol{A}^{100}\boldsymbol{u}_0 &= \frac{\boldsymbol{A}^{100}\boldsymbol{x}_1 - \boldsymbol{A}^{100}\boldsymbol{x}_2}{\lambda_1 - \lambda_2} \\
&= \frac{\boldsymbol{A}^{99}(\boldsymbol{A}\boldsymbol{x}_1) - \boldsymbol{A}^{99}(\boldsymbol{A}\boldsymbol{x}_2)}{\lambda_1 - \lambda_2} = \frac{\lambda_1\boldsymbol{A}^{99}\boldsymbol{x}_1 - \lambda_2\boldsymbol{A}^{99}\boldsymbol{x}_2}{\lambda_1 - \lambda_2} = \frac{\lambda_1^2\boldsymbol{A}^{98}\boldsymbol{x}_1 - \lambda_2^2\boldsymbol{A}^{98}\boldsymbol{x}_2}{\lambda_1 - \lambda_2} = \dots \\
&= \frac{\lambda_1^{100}\boldsymbol{x}_1 - \lambda_2^{100}\boldsymbol{x}_2}{\lambda_1 - \lambda_2}
\end{aligned}$$

After simplification, finally we obtain the same result.

7.3.3. Imaginary Eigenvalues

The eigenvalues might not be real numbers sometimes.

■ **Example 7.6** Consider the rotation matrix given by $\boldsymbol{K} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. It rotates our vector by $90°$:

$$\boldsymbol{K}\begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}.$$

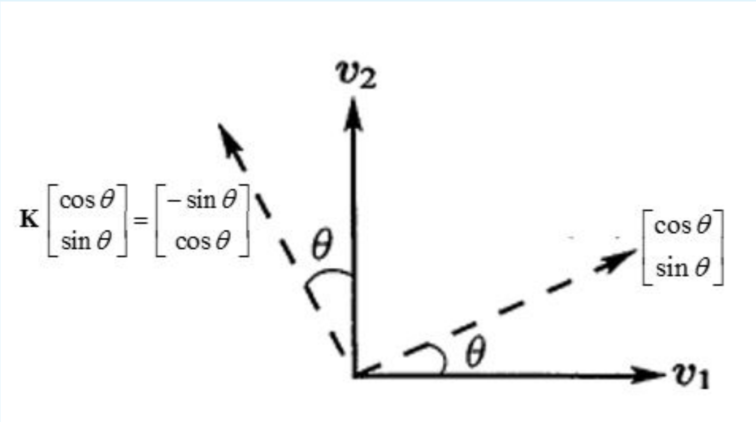


Figure 7.7: Rotate a vector by $90°$.

This rotation matrix exists eigenvector and eigenvalue, i.e., $\exists \boldsymbol{v} \neq \boldsymbol{0}$ and λ s.t.

$$\boldsymbol{K}\boldsymbol{v} = \lambda\boldsymbol{v}.$$

However, the equation above means the rotaion matrix doesn't change the direction of $\boldsymbol{v}$. In geometric meaning it rotates vector $\boldsymbol{v}$ by 90°. It seems a contradiction. **This phenomenon will not happen unless we go to imaginary eigenvectors.** Let's compute eigenvalues and eigenvectors for $\boldsymbol{K}$ first:

$$P_{\boldsymbol{K}}(\lambda) = \begin{vmatrix} \lambda & 1 \\ -1 & \lambda \end{vmatrix} = \lambda^2 + 1 \implies \lambda_1 = i, \quad \lambda_2 = -i.$$

$$(\lambda_1 \boldsymbol{I} - \boldsymbol{K})\boldsymbol{x} = \begin{pmatrix} i & 1 \\ -1 & i \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \boldsymbol{0} \implies \boldsymbol{x} = \alpha \begin{pmatrix} 1 \\ -i \end{pmatrix}.$$

$$(\lambda_2 \boldsymbol{I} - \boldsymbol{K})\boldsymbol{x} = \begin{pmatrix} -i & 1 \\ -1 & -i \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \boldsymbol{0} \implies \boldsymbol{x} = \beta \begin{pmatrix} 1 \\ i \end{pmatrix}.$$

Moverover, we can diagonalize $\boldsymbol{K}$:

$$\boldsymbol{D} = \boldsymbol{S}^{-1}\boldsymbol{K}\boldsymbol{S} = \begin{pmatrix} i & \\ & -i \end{pmatrix} \quad \text{where } \boldsymbol{S} = \begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix}.$$

■

(R) For motion in vector space, eigenvalues are "speed" and eigenvectors are "directions" under the basis $\boldsymbol{S} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \dots & \boldsymbol{x}_n \end{bmatrix}$.

$$\boldsymbol{v} = c_1\boldsymbol{x}_1 + \dots + c_n\boldsymbol{x}_n \xrightarrow{\text{postmultiply } \boldsymbol{A}} \boldsymbol{A}\boldsymbol{v} = c_1\lambda_1\boldsymbol{x}_1 + \dots + c_n\lambda_n\boldsymbol{x}_n.$$

$$\begin{pmatrix} c_1 & \dots & c_n \end{pmatrix} \xrightarrow{\text{rightmultiply } \boldsymbol{D} = \text{diag}(\lambda_1, \dots, \lambda_n)} \begin{pmatrix} c_1\lambda_1 & \dots & c_n\lambda_n \end{pmatrix}.$$

7.3.4. Complex Numbers and vectors

From Example(7.6) we can see that even for a real matrix, its eigenvaluesmay be complex numbers.

Definition 7.6 [Complex Numbers] A complex number $\boldsymbol{x} \in \mathbb{C}$ could be written as

$$\boldsymbol{x} = a + bi,$$

where $i^2 = -1$.

Its **complex conjugate** is defined as $\bar{\boldsymbol{x}} = a - bi$.

Its **modulus** is defined as $|\boldsymbol{x}| = \sqrt{a^2 + b^2} = \sqrt{\boldsymbol{x}\bar{\boldsymbol{x}}}$. ■

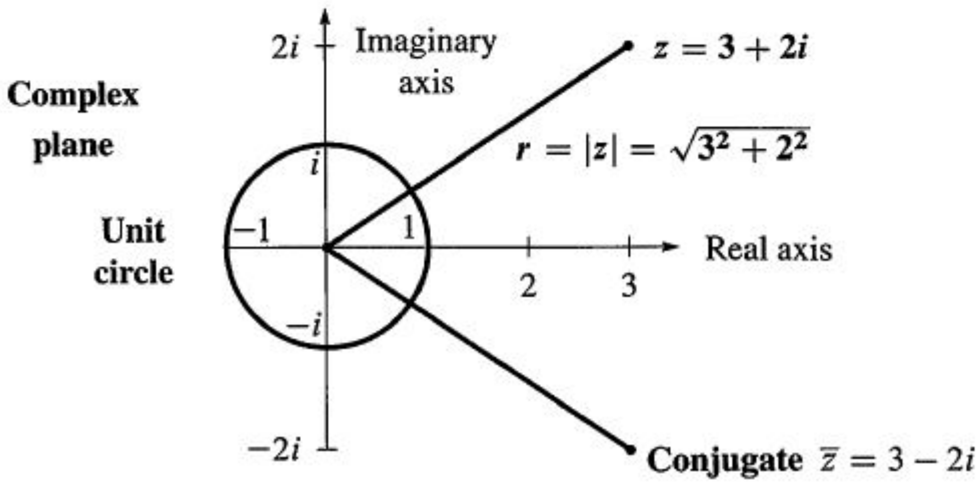


Figure 7.8: The number $z = a + bi$ corrsponds to the vector (a, b).

Definition 7.7 [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 7.8 [Hermitian transpose] Given $\boldsymbol{z} \in \mathbb{C}^n$, we use $\boldsymbol{z}^{\mathrm{H}}$ denote its **Hermitian transpose**:

$$\boldsymbol{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} \implies \boldsymbol{z}^{\mathrm{H}} = \bar{\boldsymbol{z}}^{\mathrm{T}} = \begin{bmatrix} \bar{z}_1 & \dots & \bar{z}_n \end{bmatrix}.$$

where $\bar{z}_i$ denotes the complex conjugate of z_i.

Definition 7.9 [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 7.10 [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 7.7** 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 7.8** 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 7.10 — 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:

$$\langle \boldsymbol{z}, \boldsymbol{w} \rangle = \boldsymbol{w}^{\mathrm{H}} \boldsymbol{z} = \bar{\boldsymbol{w}}^{\mathrm{T}} \boldsymbol{z} = \bar{w}_1 z_1 + \cdots + \bar{w}_n z_n$$

$$\langle \boldsymbol{w}, \boldsymbol{z} \rangle = \mathbf{z}^{\mathrm{H}} \boldsymbol{w} = \bar{\mathbf{z}}^{\mathrm{T}} \boldsymbol{w} = \bar{z}_1 w_1 + \cdots + \bar{z}_n w_n$$

Since we have $\overline{\boldsymbol{w}\boldsymbol{v}} = \bar{\boldsymbol{w}}\bar{\boldsymbol{v}}$ and $\overline{\boldsymbol{w} + \boldsymbol{v}} = \bar{\boldsymbol{w}} + \bar{\boldsymbol{v}}$, 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 7.11 — Sesquilinear. For two vectors $\boldsymbol{z}$ and $\boldsymbol{w} \in \mathbb{C}^n$, we have

$$\langle \alpha \boldsymbol{z}, \boldsymbol{w} \rangle = \alpha \langle \boldsymbol{z}, \boldsymbol{w} \rangle \tag{7.16}$$

$$\langle \boldsymbol{z}, \beta \boldsymbol{w} \rangle = \bar{\beta} \langle \boldsymbol{z}, \boldsymbol{w} \rangle \tag{7.17}$$

for scalars α and β.

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 (7.17), 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

$$\langle \boldsymbol{z}, \beta \boldsymbol{w} \rangle = \overline{\beta \overline{\langle \boldsymbol{z}, \boldsymbol{w} \rangle}} = \bar{\beta} \langle \boldsymbol{z}, \boldsymbol{w} \rangle.$$

■

7.3.4.1. 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(7.12) below carefully, they are very important!

Proposition 7.12 For a *real symmetric* matrix $\boldsymbol{A}$,

- All eigenvalues are real numbers.
- The eigenvectors associated with distinct eigenvalues are orthogonal.
- $\boldsymbol{A}$ is diagonalizable. More general, all eigenvectors of $\boldsymbol{A}$ 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, λ is real.

- For the second part, suppose $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are two eigenvectors corresponding to two **distinct** eigenvalues λ_1 and λ_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., $\boldsymbol{x}_1 \perp \boldsymbol{x}_2$.

- The proof for the third part is not required.

■

7.3.5. Spectral Theorem

We have a stronger version of the third part of proposition(7.12):

Theorem 7.5 — Spectral Theorem. Any real symmetric matrix $\boldsymbol{A}$ has the factorization

$$\boldsymbol{A} = \boldsymbol{Q}\Lambda\boldsymbol{Q}^{\mathrm{T}}, \tag{7.18}$$

where $\Lambda \in \mathbb{R}^{n\times n}$ is diagonal matrix, $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ is orthogonal.

Proof. From proposition (7.12) we know that $\boldsymbol{A}$ is *diagonalizable*, i.e., there exists invertible matrix $\boldsymbol{Q}$ and diagonal matrix Λ such that

$$\boldsymbol{A} = \boldsymbol{Q}\Lambda\boldsymbol{Q}^{-1}$$

From proposition(7.12), since all eigenvalues of $\boldsymbol{A}$ are real, Λ is a real matrix.

Since all eigenvectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ are orthogonal, from proposition(7.5), matrix $\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{x}_1 & \dots & \boldsymbol{x}_n \end{bmatrix}$, we imply $\boldsymbol{Q}$ is orthogonal. ■

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,
$$\boldsymbol{A} = \lambda_1 q_1 q_1^{\mathrm{T}} + \lambda_2 q_2 q_2^{\mathrm{T}} + \dots + \lambda_n q_n q_n^{\mathrm{T}} \tag{7.19}$$
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 7.9** If we write $\boldsymbol{A}$ as a linear combination of projection matrices, we can have a deep understanding for the linear transformation $\boldsymbol{A}\boldsymbol{x}$:

$$\boldsymbol{A} = \sum_{j=1}^{n} \lambda_j q_j q_j^{\mathrm{T}} \implies \boldsymbol{A}\boldsymbol{x} = \sum_{j=1}^{n} \lambda_j q_j q_j^{\mathrm{T}} \boldsymbol{x} = \sum_{j=1}^{n} \lambda_j (q_j q_j^{\mathrm{T}} \boldsymbol{x}).$$

For the case $n = 2$, it's clear to find that

$$x = c_1 q_1 + c_2 q_2 \implies Ax = \lambda_1 c_1 q_1 + \lambda_2 c_2 q_2$$

Showing in graph, we have

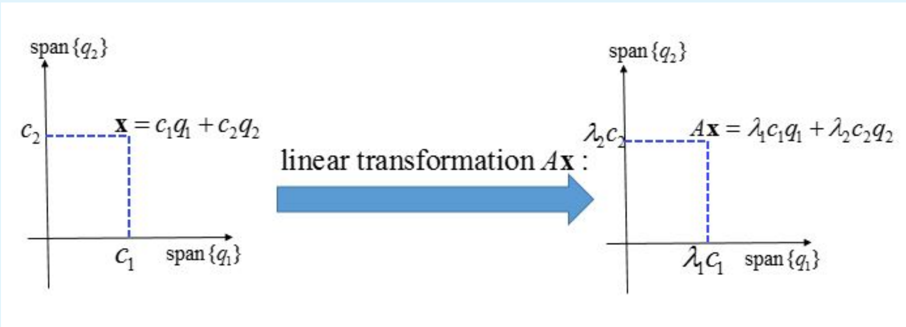


Figure 7.9: Linear transformation of A.

R The formula

$$A = \sum_{j=1}^{n} \lambda_j q_j q_j^{\mathrm{T}} \text{ or } A = Q \Lambda Q^{\mathrm{T}}$$

are called the **eigen-decomposition** or **eigenvalue decomposition** of A.

$\{\lambda_1, \dots, \lambda_n\}$ are called the **spectum** of A.

Also, we can extend our result from real symmetric matrix into complex-valued.

7.3.6. Hermitian matrix

Definition 7.11 [Symmetric and Hermitian]

- Recall that a square matrix A is said to be **symmetric** if $a_{ij} = a_{ji}$ for all i, j, or equivalently, if $A^{\mathrm{T}} = A$
- For complex-valued case, a square matrix A is said to be **Hermitian** if $a_{ij} = \bar{a}_{ji}$ for all i, j, or equivalently, if $A^{\mathrm{H}} = A$.

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: $M = \begin{bmatrix} 3 & 2-i \\ 2+i & 4 \end{bmatrix} \in \mathbb{H}^2$ since $M = M^{\mathrm{H}}$.

If M is a real matrix, then $M = M^{\mathrm{H}} \iff M = M^{\mathrm{T}}$. 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 7.13 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 α 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 α is real. ∎

Proposition 7.14 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 7.12 [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 7.6 — Spectral Theorem. Any Hermitian matrix M can be factorized into

$$M = U \Lambda U^{\mathrm{H}}$$

where Λ 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.

7.4. Assignment Seven

1. Here is a wrong "proof" that the *eigenvalues of all real matrices are real*:

$$\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x} \text{ gives } \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} \implies \lambda = \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}} \in \mathbb{R}.$$

   Find the flaw in this reasoning: a hidden assumption that is not justified.

2. Let $\boldsymbol{A}$ be an $n \times n$ matrix and let λ be an eigenvalue of $\boldsymbol{A}$ whose eigenspace has dimension k, where $1 < k < n$. Any basis $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_k\}$ for the eigenspace can be extended to a basis $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ for $\mathbb{R}^n$. Let $\boldsymbol{X} = \begin{bmatrix}\boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n\end{bmatrix}^{\mathrm{T}}$ and $\boldsymbol{B} = \boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{X}$.

   (a) Show that $\boldsymbol{B}$ is of the form

$$\begin{bmatrix}\lambda\boldsymbol{I} & \boldsymbol{B}_{12} \\ \boldsymbol{0} & \boldsymbol{B}_{22}\end{bmatrix}$$

   where $\boldsymbol{I}$ is the $k \times k$ identity matrix

   (b) Show that λ is an eigenvalue of $\boldsymbol{A}$ with multiplicity at least k.

3. Let $\boldsymbol{x}, \boldsymbol{y}$ be nonzero vectors in $\mathbb{R}^n$, $n \geq 2$, and let $\boldsymbol{A} = \boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}$. Show that

   (a) $\lambda = 0$ is an eigenvalue of $\boldsymbol{A}$ with $n-1$ linearly independent eigenvectors. Moreover, due to the conclusion of question 2, 0 is an eigenvalue of $\boldsymbol{A}$ with multiplicity at least $n-1$.

   (b) The remaining eigenvalue of $\boldsymbol{A}$ is

$$\lambda_n = \text{trace}(\boldsymbol{A}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$$

   and $\boldsymbol{x}$ is an eigenvector belonging to λ_n.

   (c) If $\lambda_n = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} \neq 0$, then $\boldsymbol{A}$ is *diagonalizable.*

4. Suppose an $n \times n$ matrix $\boldsymbol{A}$ has n distinct eigenvalues $\lambda_1, \dots, \lambda_n$. Consider the matrix $\boldsymbol{B} = (\boldsymbol{A} - \lambda_1\boldsymbol{I}) \dots (\boldsymbol{A} - \lambda_n\boldsymbol{I})$. Prove that $\boldsymbol{B}$ must be a *zero matrix*.
   Hint: How to do eigendecomposition for $\boldsymbol{A} - \lambda_i\boldsymbol{I}$?

5. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be $n \times n$ matrices. Show that

   (a) If λ is a **nonzero** eigenvalue of $\boldsymbol{A}\boldsymbol{B}$, then it is also an eigenvalue of $\boldsymbol{B}\boldsymbol{A}$.

(b) If $\lambda = 0$ is an eigenvalue of $\boldsymbol{AB}$, then $\lambda = 0$ is also an eigenvalue of $\boldsymbol{BA}$.

6. (a) The sequence a_k is defined as

$$a_0 = 4, a_1 = 5, a_{k+1} = 3a_k - 2a_{k-1}, k = 1, 2, \dots$$

What is the *general formula* for a_k?

(b) The sequence b_k is defined as

$$b_0 = \alpha, b_1 = \beta, b_{k+1} = 4b_k - 4b_{k-1}, k = 1, 2, \dots$$

What is the *general formula* for b_k?

Hint: Prove the corresponding matrix is similar to

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

In order to compute

$$\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}^k,$$

you need to use the fact that

$$\text{Given sequence } p_{k+1} = 2p_k + 2^k \implies p_k = (p_0 + \frac{k}{2}) \times 2^k.$$

7. State and justify whether the following three statements are True or False:

(a) If $\boldsymbol{A}$ is *real symmetric* matrix, then any 2 linearly independent eigenvectors of $\boldsymbol{A}$ are perpendicular.

(b) Any n by n complex matrix with n real eigenvalues and n orthonormal eigenvectors is a *Hermitian matrix*.

(c) If $\boldsymbol{A}$ is diagonalizable. then $e^{\boldsymbol{A}}$ is diagonalizable.

(We define $e^{\boldsymbol{A}} = \boldsymbol{I} + \boldsymbol{A} + \frac{1}{2!}\boldsymbol{A}^2 + \dots$)

(d) If $\boldsymbol{A}$ is Hermitian and $\boldsymbol{A}$ is invertible, then $\boldsymbol{A}^{-1}$ is also Hermitian.

8. (a) For a complex $\boldsymbol{A}$, is the left nullspace $N(\boldsymbol{A}^{\mathrm{T}})$ orthogonal to $C(\boldsymbol{A})$ under the old unconjugated inner product $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$ or new conjugated inner product $\boldsymbol{x}^{\mathrm{H}}\boldsymbol{y}$? What about $N(\boldsymbol{A}^{\mathrm{H}})$ and $C(\boldsymbol{A})$?

(b) For a real vector subspace V, the intersection of V and $V^{\perp}$ is only the single point $\{\mathbf{0}\}$. Now suppose V is a complex vector subspace. If we define $V^{\perp}$ as the set of vector $\boldsymbol{x}$ with $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{v} = 0$ for all $\boldsymbol{v} \in V$. Give an example of a V that intersects $V^{\perp}$ at a nonzero vector. What about if we use $\boldsymbol{x}^{\mathrm{H}}\boldsymbol{v} = 0$? Does V ever intersect $V^{\perp}$ at a nonzero vector using the conjugated definition of orthogonality?

Chapter 8

Week7

8.1. Tuesday

8.1.1. Quadratic form

The graphs of the following equations are easy to plot:

$$x^2 + y^2 = 1 \implies \text{Circle.} \tag{8.1}$$

$$\frac{x^2}{2} + \frac{y^2}{5} = 1 \implies \text{Elipse.} \tag{8.2}$$

$$\frac{x^2}{2} - \frac{y^2}{5} = 1 \implies \text{Hyperbola.} \tag{8.3}$$

$$\left.\begin{aligned} x^2 &= \alpha y \\ y^2 &= \alpha x \end{aligned}\right\} \implies \text{Parabola.} \tag{8.4}$$

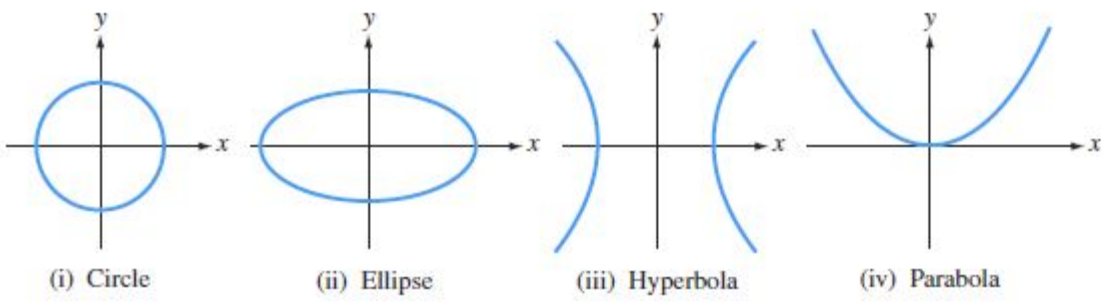


Figure 8.1: graph for quadratic form equations of two variables

The equations (8.1) – (8.4) is the *quadratic form equations of two variables*. Now we give the general form for quadratic equations:

Definition 8.1 [Quadratic form] The formula of **quadratic form** is given by

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$$

where $\boldsymbol{A} \in \mathbb{S}^n$ and $\boldsymbol{x} \in \mathbb{R}^n$.

Moreover, sometimes we write $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$ as:

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \sum_{i,j=1}^{n} x_i x_j a_{ij}$$

where x_i is the ith entry of $\boldsymbol{x}$ and a_{ij} are (i,j)th entry of $\boldsymbol{A}$. ■

Moverover, we say an equation is the **conic section of quadratic form** if it can be written as

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = 1.$$

You may be confused why the quadratic form requires the symmetric constraint. Now we give the reason:

- It is easy to verify $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}$.
- Hence given any matrix $\boldsymbol{A}$, we always have

$$\begin{aligned}
\boldsymbol{x}^{\mathrm{T}}\left(\frac{\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}}}{2}\right)\boldsymbol{x} &= \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} + \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x} \\
&= \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} + \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} \\
&= \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}.
\end{aligned}$$

Note that $\left(\frac{\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}}}{2}\right)$ is *symmetric*! Hence given any $\boldsymbol{A}$, since $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\left(\frac{\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}}}{2}\right)\boldsymbol{x}$, it suffices to consider a symmetric matrix.

■ **Example 8.1** Given the equation $3x^2 + 2xy + 3y^2 = 1$, how we transform it into the conic section of quadratic form? How can we determine its shape in view of matirx?

Actually, It can be written as

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 1. \qquad \textbf{conic section of quadratic form.} \tag{8.5}$$

We could obatin a simpler version of the conic section of quadratic form, i.e., the middle matrix should be diagonal. We define $\boldsymbol{A} = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$. Since $\boldsymbol{A} \in \mathbb{S}^2$, it admits the eigenvalue decomposition:

$$\boldsymbol{A} = \boldsymbol{Q}\Lambda\boldsymbol{Q}^{\mathrm{T}}$$

where $\Lambda = \begin{pmatrix} \lambda_1 & \\ & \lambda_2 \end{pmatrix}$, $\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 \end{bmatrix}$.

Thus we convert equation (8.5) into

$$\begin{pmatrix} x & y \end{pmatrix} \boldsymbol{Q}\Lambda\boldsymbol{Q}^{\mathrm{T}} \begin{pmatrix} x \\ y \end{pmatrix} = 1 \implies \tilde{\boldsymbol{x}}^{\mathrm{T}}\Lambda\tilde{\boldsymbol{x}} = 1. \tag{8.6}$$

where $\tilde{\boldsymbol{x}} = \boldsymbol{Q}^{\mathrm{T}} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{pmatrix}$.

Then how to determine the shape of this equation? We just do matrix multiplication of Eq.(8.6) to obtain:

$$\lambda_1\tilde{x}_1^2 + \lambda_2\tilde{x}_2^2 = 1.$$

After computation, we find $\lambda_1, \lambda_2 > 0$. Hence this equation is an **elipse.** ■

8.1.2. Convex Optimization Preliminaries

Now we recall how to compute derivative for matrix:

$$\frac{\partial(f^{\mathrm{T}}g)}{\partial x} = \frac{\partial f(x)}{\partial x}g(x) + \frac{\partial g(x)}{\partial x}f(x)$$

Examples of matrix derivatives:

$$
\begin{aligned}
\frac{\partial(a^{\mathrm{T}}\boldsymbol{x})}{\partial \boldsymbol{x}} &= a \\
\frac{\partial(a^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}} &= \frac{\partial((\boldsymbol{A}^{\mathrm{T}}a)^{\mathrm{T}}\boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{A}^{\mathrm{T}}a \\
\frac{\partial(\boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}} &= \boldsymbol{A}^{\mathrm{T}} \\
\frac{\partial(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}} &= \boldsymbol{A}\boldsymbol{x} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}
\end{aligned}
$$

■ **Example 8.2**

Given $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}$. We want to do the optimization:

$$
\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x})
$$

How to find the optimal solution? The direct idea is to take the first order derivative:

$$
\begin{aligned}
\frac{\partial f}{\partial \boldsymbol{x}} &= \frac{1}{2}\frac{\partial(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}} + \frac{\partial(\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x})}{\boldsymbol{x}} \\
&= \frac{1}{2}(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}) + \boldsymbol{b}.
\end{aligned}
$$

Since $\boldsymbol{A}$ is symmetric, we obtain

$$
\frac{\partial f}{\partial \boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}.
$$

If $\boldsymbol{x}^*$ is an optimal solution, then it must satisfy:

$$
\nabla f(\boldsymbol{x}^*) = \frac{\partial f(\boldsymbol{x}^*)}{\partial \boldsymbol{x}} = \mathbf{0} \implies \boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{b} = \mathbf{0}.
$$

There may follow these cases:

- If equation $\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} = \mathbf{0}$ has no solution, then $f(\boldsymbol{x})$ is unbounded. (We omit the proof of this statement)

- If equation $\boldsymbol{Ax}+\boldsymbol{b}=\boldsymbol{0}$ has a solution $\boldsymbol{x}^*$, it doesn't mean $\boldsymbol{x}^*$ is an optimal solution. (Note that the reverse is true.)

  Let's raise a counterexample: if we set

$$\boldsymbol{A}=\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix},\quad \boldsymbol{b}=\boldsymbol{0},\quad \boldsymbol{x}=\begin{pmatrix}x_1\\ x_2\end{pmatrix},$$

  then $f(\boldsymbol{x})=\frac{1}{2}(x_1^2-x_2^2)$. One solution to $\boldsymbol{Ax}+\boldsymbol{b}=\boldsymbol{0}$ is $\boldsymbol{x}^*=\begin{pmatrix}0\\ 0\end{pmatrix}$.

  Obviously, $\boldsymbol{x}^*$ is not a optimal solution. If $x_1=0, x_2\to\infty$, then $f(\boldsymbol{x})\to-\infty$!

8.1.2.1. Second optimality condition

If $\boldsymbol{x}^*$ is a optimal solution to $f(\boldsymbol{x})$, what else condition should $\boldsymbol{x}^*$ satisfy?

Let's take $f(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Ax}+\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}$ as an example, we want to find $\boldsymbol{x}^*$ s.t.

$$\min f(\boldsymbol{x})=f(\boldsymbol{x}^*).$$

Firstly, we write $f(\boldsymbol{x})$ into its *taylor expansion*:

$$f(\boldsymbol{x})=f(\boldsymbol{x}^*)+\langle\nabla f(\boldsymbol{x}^*),\boldsymbol{x}-\boldsymbol{x}^*\rangle+\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}^*)^{\mathrm{T}}\nabla^2 f(\boldsymbol{x}^*)(\boldsymbol{x}-\boldsymbol{x}^*). \tag{8.7}$$

Note that $\nabla^2 f(\boldsymbol{x}^*)$ is the Hessian matrix of $f(\boldsymbol{x}^*)$, which is defined as

$$\nabla^2 f(\boldsymbol{x}^*):=\left[\frac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_i\partial x_j}\right]=\nabla(\nabla f(\boldsymbol{x}^*)).$$

We compute $\triangledown f(\boldsymbol{x})$ and $\triangledown^2 f(\boldsymbol{x})$:

$$\begin{aligned}\triangledown f(\boldsymbol{x})&=\frac{1}{2}(\boldsymbol{Ax}+\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x})+\boldsymbol{b}.\\ \triangledown^2 f(\boldsymbol{x})&=\triangledown\left[\frac{1}{2}(\boldsymbol{Ax}+\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x})+\boldsymbol{b}\right]=\frac{1}{2}\triangledown(\boldsymbol{Ax})+\frac{1}{2}\triangledown(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x})=\frac{1}{2}(\boldsymbol{A}+\boldsymbol{A}^{\mathrm{T}}).\end{aligned}$$

If assume $\boldsymbol{A}$ is **symmetric**, then we have $\nabla f(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ and $\nabla^2 f(\boldsymbol{x}) = \boldsymbol{A}$.

Since the optimal solution $\boldsymbol{x}^*$ satisfies $\bigtriangledown f(\boldsymbol{x}^*) = \boldsymbol{0}$, we deive

$$\langle \bigtriangledown f(\boldsymbol{x}^*), \boldsymbol{x} - \boldsymbol{x}^* \rangle = 0.$$

Then substituting it into Eq.(8.7), we obtain:

$$f(\boldsymbol{x}) = f(\boldsymbol{x}^*) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^*)^{\mathrm{T}} \bigtriangledown^2 f(\boldsymbol{x}^*)(\boldsymbol{x} - \boldsymbol{x}^*).$$

Or equivalently, $f(\boldsymbol{x}) - f(\boldsymbol{x}^*) = \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{x}^*)$.

Since $\boldsymbol{x}^*$ is optimal that minimize $f(\boldsymbol{x})$, $LHS = f(\boldsymbol{x}) - f(\boldsymbol{x}^*) \geq 0$ for $\forall \boldsymbol{x}$. It follows that

$$\frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{x}^*) \geq 0, \text{ for } \forall \boldsymbol{x}.$$

Or equivalently,

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} \geq 0 \text{ for } \forall \boldsymbol{x}.$$

Our conclusion is that if there exists a optimal solution for $f(\boldsymbol{x})$, then the matrix $\boldsymbol{A}$ should satisfy $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} \geq 0$ for $\forall \boldsymbol{x}$. We have a specific name for such $\boldsymbol{A}$.

R The Hessian matrix $\nabla^2 f(\boldsymbol{x})$ is the **second order derivative** of $f(\boldsymbol{x})$. In scalar case we know that the second optimality condition to minimize the function $f(x)$ is to let its second order derivative no less than zero. In vector case, the second optimality condition is $\nabla^2 f(\boldsymbol{x}) \succeq 0$, where $\succeq 0$ denotes the **positive semi-definite**.

8.1.3. Positive Definite Matrices

Definition 8.2 [Positive-definite] A matrix $\boldsymbol{A} \in \mathbb{S}^n$ is said to be

- *positive-semi-definite* (PSD) if $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} \geq 0$ for $\forall \boldsymbol{x}$. We denote it as $\boldsymbol{A} \succeq 0$.
- *positive-definite* (PD) if $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} > 0$ for $\forall \boldsymbol{x} \neq \boldsymbol{0}$. We denote it as $\boldsymbol{A} \succ 0$.

- *indefinite* if there exist some $\boldsymbol{x}$ and $\boldsymbol{y}$ s.t.

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} < 0 < \boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y}.$$

Theorem 8.1 Given a matrix $\boldsymbol{A} \in \mathbb{S}^n$, the following statements are equivalent:

1. $\boldsymbol{A}$ is PD.
2. All eigenvalues of $\boldsymbol{A}$ are positive.
3. All n *upper left square submatrices* $\boldsymbol{A}_1, \dots, \boldsymbol{A}_n$ all have positive determinants.
4. $\boldsymbol{A}$ could be factorized as $\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}$, where $\boldsymbol{R}$ is nonsingular.

You may be confused about the "upper left submatrices". They are the 1 by 1, 2 by 2,…,n by n submatrices of $\boldsymbol{A}$ on the upper left. The n by n submatrix is exactly $\boldsymbol{A}$. Before we geive a detailed proof, let's show how to test some matrices for positive definiteness by using this theorem:

Example 8.3 Test these matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ for positive definiteness:

$$\boldsymbol{A} = \begin{bmatrix} 1 & & & \\ & 2 & & \\ & & 2 & \\ & & & 2 \end{bmatrix} \quad \text{and} \quad \boldsymbol{B} = \begin{bmatrix} 1 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 2 & -1 \\ & & -1 & 2 \end{bmatrix}$$

- For matrix $\boldsymbol{A}$, its eigenvalues are $\{1,2,2,2\}$. So all eigenvalues of $\boldsymbol{A}$ are positive, $\boldsymbol{A}$ is PD. Moverover, we can test its positive definiteness by definition:

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = x_1^2 + 2x_2^2 + 2x_3^2 + 2x_4^2 > 0.$$

for $\forall \boldsymbol{x} = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \end{pmatrix}^{\mathrm{T}} \neq \boldsymbol{0}$.

- For matrix $\boldsymbol{B}$, all *upper left square submatrices* are given by

$$\boldsymbol{B}_1 = \begin{bmatrix} 1 \end{bmatrix} \quad \boldsymbol{B}_2 = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \quad \boldsymbol{B}_3 = \begin{bmatrix} 1 & -1 & \\ -1 & 2 & -1 \\ & -1 & 2 \end{bmatrix} \quad \boldsymbol{B}_4 = \begin{bmatrix} 1 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 2 & -1 \\ & & -1 & 2 \end{bmatrix}$$

After messy computation, we obtain

$$\det(\boldsymbol{B}_1) = 1 \quad \det(\boldsymbol{B}_2) = 1 \quad \det(\boldsymbol{B}_3) = 1 \quad \det(\boldsymbol{B}_4) = 1.$$

Hence all *upper left square determinants* are positive, $\boldsymbol{B}$ is PD.

■

Then we begin to give a proof for this theorem:

Proof.

- $(1) \implies (2)$: Given any eigen-pair $(\lambda, \boldsymbol{x})$ of $\boldsymbol{A}$, we have

$$\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}, \text{ for } \forall \boldsymbol{x} \neq \boldsymbol{0}.$$

By postmutliplying $\boldsymbol{x}^{\mathrm{T}}$ both sides, we obtain:

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = \lambda\|\boldsymbol{x}\|^2 \implies \lambda = \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\|\boldsymbol{x}\|^2} > 0.$$

- $(2) \implies (1)$: Assume all eigenvalues $\lambda_i > 0$ for $i = 1, 2, \dots, n$. Our goal is to show $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} > 0$ for $\forall \boldsymbol{x} \neq \boldsymbol{0}$.

Since $\boldsymbol{A} \in \mathbb{S}^n$, it admits the eigen-decomposition:

$$\boldsymbol{A} = \boldsymbol{Q}\Lambda\boldsymbol{Q}^{\mathrm{T}} \qquad \boldsymbol{Q} \text{ is orthogonal matrix.}$$

It follows that

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\Lambda\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{x} = (\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{x})^{\mathrm{T}}\Lambda(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{x}).$$

By setting $\tilde{\boldsymbol{x}} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{x} = \begin{bmatrix} \tilde{x_1} & \dots & \tilde{x_n} \end{bmatrix}$, $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$ can be rewritten as

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \tilde{\boldsymbol{x}}^{\mathrm{T}}\Lambda\tilde{\boldsymbol{x}} = \sum_{i=1}^{n}\lambda_i\tilde{x_i}^2 \geq 0.$$

Then we aruge for $\sum_{i=1}^{n}\lambda_i\tilde{x_i}^2 \neq 0$. It suffices to show $\|\tilde{\boldsymbol{x}}\| \neq 0$.

You can verify by yourself that $\|\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{x}\| = \|\boldsymbol{x}\|$. Thus we obtain:

$$\|\tilde{\boldsymbol{x}}\| := \|\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{x}\| = \|\boldsymbol{x}\| \neq 0.$$

- $(1) \implies (3)$: We only to show $\det(\boldsymbol{A}_k) > 0$ for any upper left matrices $\boldsymbol{A}_k$.

  Given any nonzero vector $\tilde{\boldsymbol{x}} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix} \in \mathbb{R}^k$, we construct $\boldsymbol{x} = \begin{pmatrix} \tilde{\boldsymbol{x}} \\ \boldsymbol{0} \end{pmatrix} \in \mathbb{R}^n$.

  Since $\boldsymbol{A} \succ 0$, we find

$$\begin{aligned} \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} &= \begin{pmatrix} \tilde{\boldsymbol{x}}^{\mathrm{T}} & \boldsymbol{0} \end{pmatrix} \boldsymbol{A} \begin{pmatrix} \tilde{\boldsymbol{x}} \\ \boldsymbol{0} \end{pmatrix} \\ &= \tilde{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{A}_k\tilde{\boldsymbol{x}} > 0. \end{aligned}$$

  Since $\tilde{\boldsymbol{x}}$ is arbitrary nonzero vector in $\mathbb{R}^k$, we derive $\boldsymbol{A}_k \succ 0$. By (2) of this theorem, all eigenvalues of $\boldsymbol{A}_k$ are positive.

  Thus $\det(\boldsymbol{A}_k) =$ product of all eigenvalues of $\boldsymbol{A}_k > 0$.

- $(3) \implies (4)$:

  - We want to show that all pivots of $\boldsymbol{A}$ are positive first:

We do row transform to convert $\boldsymbol{A}$ into upper triangular matrix $\tilde{\boldsymbol{A}}$:

$$\begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{bmatrix} \Longrightarrow \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \end{bmatrix}$$

During row transformation, the determinant for the correponding *upper left submatrices* $\boldsymbol{A}_i$ doesn't change. In other words, we obtain

$$\det(\tilde{\boldsymbol{A}}_i) = \det(\boldsymbol{A}_i) \text{ for } i = 1, \dots, n.$$

Moreover, $\tilde{\boldsymbol{A}}_i$ always contains $\tilde{\boldsymbol{A}}_{i-1}$ on its upper left side:

$$\tilde{\boldsymbol{A}}_i = \begin{bmatrix} \tilde{\boldsymbol{A}}_{i-1} & \boldsymbol{B} \\ \mathbf{0} & \tilde{a}_{ii} \end{bmatrix}$$

Note that $\tilde{\boldsymbol{A}}_i$'s are also upper triangular matrices. The determinant of an upper triangular matrix is the product of its diagonal entries. Hence we obtain

$$\det(\tilde{\boldsymbol{A}}_i) = \tilde{a}_{ii} \det(\tilde{\boldsymbol{A}}_{i-1}) \text{ for } i = 2, \dots, n.$$

It follows that

$$\tilde{a}_{ii} = \frac{\det(\tilde{\boldsymbol{A}}_i)}{\det(\tilde{\boldsymbol{A}}_{i-1})} = \frac{\det(\boldsymbol{A}_i)}{\det(\boldsymbol{A}_{i-1})} \text{ for } i = 2, \dots, n.$$

Due to (3) of this theorem, $\tilde{a}_{ii} > 0$ for $i = 2, \dots, n$. Also, $\tilde{a_{11}} = \det(\tilde{\boldsymbol{A}}_1) = \det(\boldsymbol{A}_1) > 0$.

In conclusion, all pivots $\tilde{a}_{ii} > 0$ for $i = 1, \dots, n$.

– Then we apply the LDU composition for $\boldsymbol{A}$. Since $\boldsymbol{A} \in \mathbb{S}^n$, we obtain

$$\boldsymbol{A} = \boldsymbol{L}\boldsymbol{D}\boldsymbol{L}^{\mathrm{T}}$$

where $\boldsymbol{D} = \operatorname{diag}(d_1, \dots, d_n)$. The diagonal entries of $\boldsymbol{D}$ are pivots of $\boldsymbol{A}$. $\boldsymbol{L}$ is a lower triangular matrix with 1's on the diagonal entries.

Since all pivots of $\boldsymbol{A}$ are positive, we define $\sqrt{\boldsymbol{D}} := \text{diag}(\sqrt{d_1}, \dots, \sqrt{d_n})$.

Hence we rewrite $\boldsymbol{A}$ as:

$$\boldsymbol{A} = \boldsymbol{L}\begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix}\boldsymbol{L}^{\mathrm{T}} = \boldsymbol{L}\sqrt{\boldsymbol{D}}\sqrt{\boldsymbol{D}}\boldsymbol{L}^{\mathrm{T}} = (\sqrt{\boldsymbol{D}}\boldsymbol{L}^{\mathrm{T}})^{\mathrm{T}}(\sqrt{\boldsymbol{D}}\boldsymbol{L}^{\mathrm{T}}).$$

We define $\boldsymbol{R} := \sqrt{\boldsymbol{D}}\boldsymbol{L}^{\mathrm{T}}$. Since $\sqrt{\boldsymbol{D}}$ and $\boldsymbol{L}^{\mathrm{T}}$ are nonsingular, $\boldsymbol{D}$ is nonsingular.

Hence $\boldsymbol{A} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}$, where $\boldsymbol{R}$ is a nonsingular matrix.

- $(4) \implies (1)$: Suppose $\boldsymbol{A} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}$, where $\boldsymbol{R}$ is nonsingular. Then for any $\boldsymbol{x} \in \mathbb{R}^n$, we have

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{x} = \|\boldsymbol{R}\boldsymbol{x}\|^2 \geq 0.$$

Then it suffices to show that if $\boldsymbol{x} \neq \boldsymbol{0}$, then $\|\boldsymbol{R}\boldsymbol{x}\| \neq 0$.:

Since $\boldsymbol{R}$ is nonsinguar, when $\boldsymbol{x} \neq \boldsymbol{0}$, we obtain $\boldsymbol{R}\boldsymbol{x} \neq \boldsymbol{0}$. Hence $\|\boldsymbol{R}\boldsymbol{x}\| \neq 0$.

■

Is there any quick ways to determine the positive definiteness of a matrix? The answer is yes. Let's introduce some definitions first:

Definition 8.3 [Submatrix] If $\boldsymbol{A}$ is a $n \times n$ matrix, then a submatrix of $\boldsymbol{A}$ is obtained by keeping some collection of rows and columns.

■ **Example 8.4** For matrix $\boldsymbol{A} = \begin{bmatrix} 1 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 2 & -1 \\ & & -1 & 2 \end{bmatrix}$, if we keep the (1,3,4)th row and

(1,2)th column of $\boldsymbol{A}$, our submatrix is denoted as

$$\boldsymbol{A}_{(1,3,4),(1,2)} = \begin{bmatrix} 1 & -1 \\ 0 & -1 \\ 0 & 0 \end{bmatrix}$$

■

Definition 8.4 [principal submatrix] If $\boldsymbol{A}$ is a $n \times n$ matrix, then a principal submatrix of $\boldsymbol{A}$ is obtained by keeping the same collection of rows and columns. For example, if we want to keep the (5,7)th row of $\boldsymbol{A}$, in order to construct a principal submatrix, we must keep the (5,7)th column of $\boldsymbol{A}$ as well. ■

■ **Example 8.5** If $\boldsymbol{A} = \begin{bmatrix} 1 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 2 & -1 \\ & & -1 & 2 \end{bmatrix}$, then if we keep the (1,3,4)th row of $\boldsymbol{A}$, in order to construct a principal submatrix, we have to keep (1,3,4)th column of $\boldsymbol{A}$ as well. Our principal submatrix is denoted as

$$\boldsymbol{A}_{(1,3,4),(1,3,4)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

■

Definition 8.5 [leading principal submatrix] If $\boldsymbol{A}$ is a $n \times n$ matrix, then a leading principal submatrix of $\boldsymbol{A}$ is obtained by keeping the first k rows and columns of $\boldsymbol{A}$, where $k \in \{1,2,\ldots,n\}$. ■

Note that the leading principal submatrix is just the upper left submatrix we have mentioned before.

Corollary 8.1 Suppose $A \in \mathbb{S}^n$, if $A \succ 0$, then all principal submatrices of A are PD as well.

Proof. Our goal is to show that $A_{\alpha,\alpha} \succ 0$, where α contains the first k elements of $\{1,\dots,n\}$.

For any $x_\alpha \in \mathbb{R}^{|\alpha|}$, it suffices to show $x_\alpha^T A_{\alpha,\alpha} x_\alpha > 0$. Here $|\alpha|$ denotes the number of elements in set α.

We construct $x \in \mathbb{R}^n$ s.t. the ith entry of x is

$$x_i = \begin{cases} (x_\alpha)_i & i \in \alpha \\ 0 & i \notin \alpha \end{cases}$$

It's obvious that

$$\begin{aligned} x^T A x &= \sum_{i,j=1}^{n} x_i x_j A_{ij} \\ &= \sum_{i,j \in \alpha} (x_\alpha)_i (x_\alpha)_j (A_{\alpha,\alpha})_{ij} \\ &= x_\alpha^T A_{\alpha,\alpha} x_\alpha > 0. \end{aligned}$$

■

How to use this corollary to test the positive definiteness?

For example, given $A = \begin{bmatrix} 2 & -1 & 1 \\ -1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}$, immediately we find one principal matrix is $A_{2,2} = 0$. Hence it is not PD.

Also, there are many equivalent statements related to PSD. The proof is similar to the PSD case, so you may complete the proof by yourself.

Theorem 8.2 Let $A \in \mathbb{S}^n$, the following statements are equivalent:

1. A is PSD.
2. All eigenvalues of A are nonnegative.

3. $\boldsymbol{A}$ could be factorized as $\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}$, where $\boldsymbol{R}$ is square.

(R) Is $\boldsymbol{A} \succeq 0$ equivalent to $\boldsymbol{A}_{ij} \geq 0$ for all i, j? No. Let's raise a counterexample:

$$\boldsymbol{A} = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix} \succeq 0.$$

PSD has many interesting properties. Before we introduce one, let's extend the definiton of inner product into matrix form:

Definition 8.6 [Frobenius inner product] For two matrices $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{B} \in \mathbb{R}^{m\times n}$, the **Frobenius inner product** is given by

$$\langle \boldsymbol{A}, \boldsymbol{B} \rangle = \sum_{i,j=1}^{n} \boldsymbol{A}_{ij}\boldsymbol{B}_{ij}$$

Or equivalently, $\langle \boldsymbol{A}, \boldsymbol{B} \rangle = \mathrm{trace}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A})$. ■

Proposition 8.1 Given two matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{S}^n$, if $\boldsymbol{A} \succeq 0, \boldsymbol{B} \succeq 0$, then $\langle \boldsymbol{A}, \boldsymbol{B} \rangle \geq 0$.

Proof. Since $\boldsymbol{A} \succeq 0$, there exists square matrix $\boldsymbol{R} = \begin{bmatrix} \boldsymbol{r}_1 & \dots & \boldsymbol{r}_n \end{bmatrix}$ s.t.

$$\boldsymbol{A} = \boldsymbol{R}\boldsymbol{R}^{\mathrm{T}} = \sum_{k=1}^{n} \boldsymbol{r}_k\boldsymbol{r}_k^{\mathrm{T}}$$

Hence our inner product is given by

$$\begin{aligned}
\langle \boldsymbol{A}, \boldsymbol{B} \rangle &= \langle \sum_{k=1}^{n} \boldsymbol{r}_k\boldsymbol{r}_k^{\mathrm{T}}, \boldsymbol{B} \rangle \\
&= \sum_{k=1}^{n} \langle \boldsymbol{r}_k\boldsymbol{r}_k^{\mathrm{T}}, \boldsymbol{B} \rangle \\
&= \sum_{k=1}^{n} (\sum_{i,j=1}^{n} \boldsymbol{B}_{ij}\boldsymbol{R}_{ki}\boldsymbol{R}_{kj}) \\
&= \sum_{k=1}^{n} \boldsymbol{r}_k^{\mathrm{T}}\boldsymbol{B}\boldsymbol{r}_k
\end{aligned}$$

Since $\boldsymbol{B} \succeq 0$, we derive $\langle \boldsymbol{A}, \boldsymbol{B} \rangle = \sum_{k=1}^{n} \boldsymbol{r}_k^{\mathrm{T}}\boldsymbol{B}\boldsymbol{r}_k \geq 0$. ■

8.2. Thursday

Three ways for matrix decomposition are significant in linear alegbra:

$$\begin{cases} \text{LU (from } \textbf{Gaussian elimination}\text{)} \\ \text{QR (from } \textbf{Orthogonalization}\text{)} \\ \text{SVD (from } \textbf{eigenvalues and eigenvectors}\text{)} \end{cases}$$

We have learnt the first two decomposition. And the third way is increasingly significant in the information age.

In the last lecture we learnt that any real symmetric matrix adimits *diagonalization*, i.e., *eigendecomposition*. However, can we get some **universal** decomposition, i.e., Is there any decomposition that can be applied to all matrices?

The anwer is yes. The key idea behind is to do *symmetrization*. We have to consider $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ and $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$.

8.2.1. SVD: Singular Value Decomposition

Theorem 8.3 — SVD. Given any matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, there exists a 3-tuple $(\boldsymbol{U}, \Sigma, \boldsymbol{V}) \in \mathbb{R}^{m\times m} \times \mathbb{R}^{m\times n} \times \mathbb{R}^{n\times n}$ such that

$$\boldsymbol{A} = \boldsymbol{U}\Sigma\boldsymbol{V}^{\mathrm{T}},$$

where $\boldsymbol{U}, \boldsymbol{V}$ are **orthogonal**, and Σ takes the form

$$\Sigma_{ij} = \begin{cases} \sigma_i, & i = j \\ 0, & i \neq j \end{cases},$$

with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$ and with $p = \min\{m, n\}$.

R

- If $\boldsymbol{V} = \boldsymbol{U}$, this decomposition is exactly **eigen-decomposition**.

- Specifically speaking,
  - $\boldsymbol{U} \in \mathbb{R}^{m \times m}$ such that its columns are eigenvectors of $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$
  - $\boldsymbol{V} \in \mathbb{R}^{n \times n}$ such that its columns are eigenvectors of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$
  - $\Sigma \in \mathbb{R}^{m \times n}$ looks like a diagonal matrix, i.e., it has the form

$$\Sigma = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix} \text{ if } m \geq n$$

$$\Sigma = \begin{pmatrix} \sigma_1 & & & 0 & \dots & 0 \\ & \ddots & & \vdots & \ddots & \vdots \\ & & \sigma_m & 0 & \dots & 0 \end{pmatrix} \text{ if } m < n.$$

with $\sigma_i = \sqrt{\lambda_i}$ for $i = 1,2,\dots,\min\{m,n\}$, where λ_i*'s are eigenvalues of* $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ *(if $m < n$) or* $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$*. (if $m \geq n$))*

Definition 8.7 [SVD] The above decomposition is called the **singular value decomposition (SVD)**

- σ_i is called the ith **singular value**
- The columns of $\boldsymbol{U}$ and $\boldsymbol{V}$, $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ are called the ith **left and right singular vectors**, respectively.
- $(\sigma_i, \boldsymbol{u}_i)$ are the eigen-pairs of $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$; $(\sigma_i, \boldsymbol{v}_i)$ are the eigen-pairs of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ for $i = 1,2,\dots,\min\{m,n\}$.
- The following notations may be used to denote the singular values of $\boldsymbol{A}$:

$$\sigma_{\max}(\boldsymbol{A}) \geq \sigma_1(\boldsymbol{A}) \geq \sigma_2(\boldsymbol{A}) \geq \cdots \geq \sigma_p(\boldsymbol{A}) = \sigma_{\min}(\boldsymbol{A})$$

The proof for the SVD decomposition is constructive. To see the insights of the proof, let's study the case $m = n$ first, then we extend the proof for general case:

Proposition 8.2 SVD always exists for any **real square nonsingular** matrix.

Proof. For $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, you may verify that $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ is PD, thus it admits the eigen-decomposition:

$$\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{U}\Sigma\boldsymbol{U}^{\mathrm{T}}, \quad \text{with } \lambda_1 \geq \cdots \geq \lambda_n > 0. \tag{8.8}$$

We define $\Sigma := \operatorname{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_m})$ and $\boldsymbol{V} := \boldsymbol{A}^{\mathrm{T}}\boldsymbol{U}\Sigma^{-1}$.

You may verify that $\boldsymbol{U}\Sigma\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{A}$ and $\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V} = \boldsymbol{I}$, i.e., $\boldsymbol{V}$ is orthogonal. The proof is complete. ■

Proposition 8.3 SVD always exists for any **real** matrix.

Proof.

- Firstly, consider the matrix product $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$. Since $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} \in \mathbb{S}^m$ and $\boldsymbol{A} \succeq 0$, we can decompose $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ as

$$\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{U}\Sigma\boldsymbol{U}^{\mathrm{T}} = \begin{bmatrix} \boldsymbol{U}_1 & \boldsymbol{U}_2 \end{bmatrix} \begin{bmatrix} \tilde{\Sigma} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{U}_1^{\mathrm{T}} & \boldsymbol{U}_2^{\mathrm{T}} \end{bmatrix} = \boldsymbol{U}_1\tilde{\Sigma}\boldsymbol{U}_1^{\mathrm{T}} \tag{8.9}$$

  where:

  - we assume that the eigenvalues are ordered, i.e.,

$$\lambda_1 \geq \cdots \geq \lambda_r > 0, \quad \text{and} \quad \lambda_{r+1} = \cdots = \lambda_p = 0$$

    with r being the number of nonzero eigenvalues
  - $\boldsymbol{U} \in \mathbb{R}^{m\times m}$ denotes an orthogonal matrix, and its columns are the corresponding eigenvectors
  - We partition $\boldsymbol{U}$ as

$$\boldsymbol{U} = \begin{bmatrix} \boldsymbol{U}_1 & \boldsymbol{U}_2 \end{bmatrix}, \quad \boldsymbol{U}_1 \in \mathbb{R}^{m\times r}, \boldsymbol{U}_2 \in \mathbb{R}^{m\times(m-r)},$$

    and $\tilde{\Sigma} = \operatorname{diag}(\lambda_1, \dots, \lambda_r)$.

- Secondly, we show that

$$U_2^{\mathrm{T}} A = 0 \tag{8.10}$$

Since U is orthogonal, we obtain:

$$U^{\mathrm{T}} U = \begin{bmatrix} U_1^{\mathrm{T}} \\ U_2^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} U_1 & U_2 \end{bmatrix} = \begin{bmatrix} U_1^{\mathrm{T}} U_1 & U_1^{\mathrm{T}} U_2 \\ U_2^{\mathrm{T}} U_1 & U_2^{\mathrm{T}} U_2 \end{bmatrix} = I \implies U_2^{\mathrm{T}} U_1 = \mathbf{0}.$$

Substituting Eq.(8.9) into $U_2^{\mathrm{T}} A (U_2^{\mathrm{T}} A)^{\mathrm{T}}$, we obtain:

$$U_2^{\mathrm{T}} A (U_2^{\mathrm{T}} A)^{\mathrm{T}} = (U_2^{\mathrm{T}} U_1) \tilde{\Sigma} U_1^{\mathrm{T}} U_2 = \mathbf{0} \tag{8.11}$$

By Eq.(8.11) and the simple result that $BB^{\mathrm{T}} = \mathbf{0}$ implies $B = \mathbf{0}$ (write B into column vectors form to verify it), we conclude that $U_2 A = \mathbf{0}$

- Thirdly, we construct the following matrices:

$$\widehat{\Sigma} = \tilde{\Sigma}^{1/2} = \operatorname{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_r}), \quad V_1 = A^{\mathrm{T}} U_1 \widehat{\Sigma}^{-1} \in \mathbb{R}^{n \times r}.$$

Combining it with Eq.(8.9), we can verify that $V_1^{\mathrm{T}} V_1 = I$. Furthermore, there exists a matrix $V_2 \in \mathbb{R}^{n \times (n-r)}$ such that $V = \begin{bmatrix} V_1 & V_2 \end{bmatrix}$ is orthogonal. Moreover, we can verify that

$$U_1^{\mathrm{T}} A V_1 = \widehat{\Sigma}, \quad U_1^{\mathrm{T}} A V_2 = \mathbf{0} \tag{8.12}$$

- Fourthly, consider the matrix product $U^{\mathrm{T}} A V$. From Eq.(8.12) and Eq.(8.10), we have

$$\begin{aligned} U^{\mathrm{T}} A V &= \begin{bmatrix} U_1^{\mathrm{T}} A V_1 & U_1^{\mathrm{T}} A V_2 \\ U_2^{\mathrm{T}} A V_1 & U_2^{\mathrm{T}} A V_2 \end{bmatrix} \\ &= \begin{bmatrix} \widehat{\Sigma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} := \Sigma \end{aligned}$$

By multiplying the above equation on the left by U and on the right by V^{T}, we obtain the desired result $A = U\Sigma V^T$. The proof is complete.

■

8.2.2. Remark on SVD decomposition

8.2.2.1. Remark 1: Different Ways of Writing out SVD

Definition 8.8 [Paritioned form of SVD] let r be the number of nonzero singular values, and note that $\sigma_1 \geq \cdots \geq \sigma_r > 0, \sigma_{r+1} = \cdots = \sigma_p = 0$. Then from the standard form, we derive the **partitioned form of SVD**:

$$A = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \tilde{\Sigma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} \tag{8.13}$$

where:

- $\tilde{\Sigma} = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)$
- $U_1 = \begin{bmatrix} u_1 & \cdots & u_r \end{bmatrix} \in \mathbb{R}^{m\times r}, U_2 = \begin{bmatrix} u_{r+1} & \cdots & u_m \end{bmatrix} \in \mathbb{R}^{m\times(m-r)}$
- $V_1 = \begin{bmatrix} v_1 & \cdots & v_r \end{bmatrix} \in \mathbb{R}^{n\times r}, V_2 = \begin{bmatrix} v_{r+1} & \cdots & v_n \end{bmatrix} \in \mathbb{R}^{n\times(n-r)}$

Note that U_1, U_2, V_1, V_2 are semi-orthogonal, i.e., they all have orthonormal columns. ■

Definition 8.9 [Thin SVD] We can re-write Eq.(8.13) as the **thin form of SVD**:

$$A = U_1 \tilde{\Sigma} V_1^T \tag{8.14}$$

■

Definition 8.10 [Outer-product form] By expanding the Eq.(8.14), we derive the **outer-product form of SVD**:

$$A = \sum_{i=1}^{p} \sigma_i u_i v_i^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T \tag{8.15}$$

8.2.2.2. Remark 2: SVD and Eigen-decomposition

The eigenvalues for $A^{\mathrm{T}}A$ and AA^{T} are the same for first p terms.

Proposition 8.4 Suppose A admits the SVD $A = U\Sigma V^{\mathrm{T}}$, then we have:

$$AA^{\mathrm{T}} = UD_1U^{\mathrm{T}}, \qquad D_1 = \Sigma\Sigma^{\mathrm{T}} = \operatorname{diag}(\sigma_1^2,\dots,\sigma_p^2, \underbrace{0,\dots,0}_{m-p \text{ zeros}}) \tag{8.16}$$

$$A^{\mathrm{T}}A = VD_2V^{\mathrm{T}}, \qquad D_2 = \Sigma^{\mathrm{T}}\Sigma = \operatorname{diag}(\sigma_1^2,\dots,\sigma_p^2, \underbrace{0,\dots,0}_{n-p \text{ zeros}}) \tag{8.17}$$

Proof. Just apply the SVD form and the orthogonality of U and V. ∎

8.2.2.3. Remark 3: SVD and Subspace

We are curious about how many singular values of A are nonzero.

Proposition 8.5 The following properties hold:

1. $\mathcal{C}(A) = \mathcal{C}(U_1), \mathcal{C}(A)^{\perp} = \mathcal{C}(U_2)$;
2. $\mathcal{C}(A^{\mathrm{T}}) = \mathcal{C}(V_1), \mathcal{C}(A^{\mathrm{T}})^{\perp} = \mathcal{N}(A) = \mathcal{C}(V_2)$;
3. $\operatorname{rank}(A) = r$, i.e., the number of nonzero singular values.

Proof. The above properties are easily seen to be true using SVD. Also, you should apply the definition for column space and null space. You should verify these properties by yourself.

$$\mathcal{C}(A) = \{y \in \mathbb{R}^m \mid y = Ax, x \in \mathbb{R}^n\} \tag{8.18a}$$

$$\mathcal{N}(A) = \{x \in \mathbb{R}^n \mid Ax = 0\} \tag{8.18b}$$

For the third part of proposition(8.5), since $\operatorname{rank}(A) = \dim(\mathcal{C}(A)) = \dim(\mathcal{C}(U_1))$, and U_1 has r orthonormal columns, we derive that $\dim(\mathcal{C}(U_1)) = r = \operatorname{rank}(A)$. ∎

For the SVD decomposition

$$A = U\Sigma V^{\mathrm{T}},$$

we can convert it into the following two forms:

$$AV = U\Sigma V^{\mathrm{T}} V = U\Sigma$$

$$A = U\Sigma V^{\mathrm{T}} \implies A^{\mathrm{T}} = V\Sigma U^{\mathrm{T}} \implies A^{\mathrm{T}} U = V\Sigma U^{\mathrm{T}} U = V\Sigma.$$

If we write it into vector forms, we obtain:

$$\begin{cases} A v_j = \sigma_j u_j \\ A^{\mathrm{T}} u_j = \sigma_j v_j \end{cases}, \quad j = 1,2,\ldots,r. \tag{8.19}$$

The columns of U (u_j) are called the **left singular vector** of A; the columns of V (v_j) are called the **right singular vector** of A; σ_j is called the **singular value**.

We can easily understand the proposition(8.5) and Eq.(8.19) by the following graph:

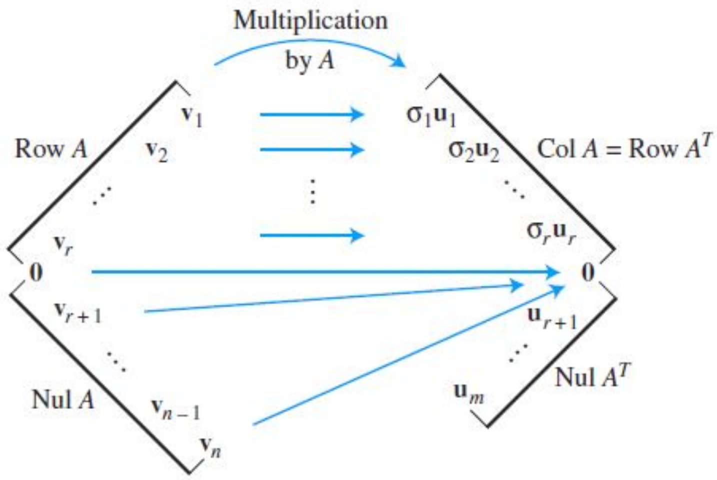


Figure 8.2: The fundamental spaces and the action of A.

Explanation:

- When $\{v_1, \ldots, v_r\}$ are multiplied by A, they are converted into $\{\sigma_1 u_1, \ldots, \sigma_r u_r\}$; when $\{v_{r+1}, \ldots, v_n\}$ are multiplied by A, they are converted into $\mathbf{0}$.
- The first r columns of V forms the basis for the row space of A, i.e., $\mathcal{C}(V_1) = \mathcal{C}(A^{\mathrm{T}})$.
- The last $n-r$ columns of V forms the basis for the null space of A, i.e., $\mathcal{C}(V_2) =$

$\mathcal{N}(\boldsymbol{A})$.

- The first r columns of $\boldsymbol{U}$ forms the basis for the column space of $\boldsymbol{A}$, i.e., $\mathcal{C}(\boldsymbol{U}_1) = \mathcal{C}(\boldsymbol{A})$.

- The last $m-r$ columns of $\boldsymbol{U}$ forms the basis for the null space of $\boldsymbol{A}^{\mathrm{T}}$, i.e., $\mathcal{C}(\boldsymbol{U}_2) = \mathcal{N}(\boldsymbol{A}^{\mathrm{T}})$

Recall the outer-product form of SVD,

$$\boldsymbol{A} = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \cdots + \sigma_r \boldsymbol{u}_r \boldsymbol{v}_r^{\mathrm{T}}$$

where $r = \operatorname{rank}(\boldsymbol{A}) =$ number of nonzero singular values, which is the third meaning for the rank:

R Up till now, $\operatorname{rank}(\boldsymbol{A})$ has three meanings:

- $\operatorname{rank}(\boldsymbol{A}) = \dim(\operatorname{row}(\boldsymbol{A}))$
- $\operatorname{rank}(\boldsymbol{A}) = \dim(\operatorname{col}(\boldsymbol{A}))$
- $\operatorname{rank}(\boldsymbol{A}) =$ number of nonzero singular values of $\boldsymbol{A}$.

R However, $\operatorname{rank}(\boldsymbol{A}) \neq$ number of nonzero eigenvalues. Let me raise a counterexample:

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

then eigenvalues are $\lambda_1 = \lambda_2 = 0$, and $\operatorname{rank}(\boldsymbol{A}) = 1$.

R Also, note that many properties can be easily proved by **thin** or **outer-product** form of SVD. For example, $\operatorname{rank}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{A})$. If you have no ideas of a proof in exam, you may try SVD.

8.2.2.4. Compact SVD

Due to the outer-product form of SVD, i.e., any matrix with rank r can be factorized into

$$\begin{aligned} A &= U\Sigma V^{\mathrm{T}} \\ &= \begin{bmatrix} u_1 & \dots & u_r \end{bmatrix} \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix} \begin{bmatrix} v_1^{\mathrm{T}} \\ \vdots \\ v_r^{\mathrm{T}} \end{bmatrix}, \end{aligned}$$

we obtain the following corollary:

Corollary 8.2 Every rank r matrix can be written as the sum of r rank 1 matrices. Moreover, these matrices could be perpendicular!

What's the meaning of perpendicular?

Definition 8.11 [perpendicular for matrix] For two real $n \times n$ matrix A and B, they are said to be **perpendicular** (**orthogonal**) if the inner product between A and B is zero:

$$\langle A, B \rangle = \operatorname{trace}(B^{\mathrm{T}} A) = \sum_{i,j=1}^{n} A_{ij} B_{ij} = 0.$$

Decompose $A := \sum_{i=1}^{r} \sigma_i u_i v_i^{\mathrm{T}}$. If we set $A_i = u_i v_i^{\mathrm{T}} \sigma_i$, let's show A_i's are perpendicular:

$$\begin{aligned} \langle A_i, A_j \rangle &= \operatorname{trace}(A_j^{\mathrm{T}} A_i) \\ &= \operatorname{trace}(\sigma_i \sigma_j v_j u_j^{\mathrm{T}} u_i v_i^{\mathrm{T}}) = \sigma_i \sigma_j \operatorname{trace}(v_j u_j^{\mathrm{T}} u_i v_i^{\mathrm{T}}) \\ &= \sigma_i \sigma_j \operatorname{trace}(v_j (u_j^{\mathrm{T}} u_i) v_i^{\mathrm{T}}) = \sigma_i \sigma_j \operatorname{trace}(v_j \mathbf{0} v_i^{\mathrm{T}}) \\ &= 0. \end{aligned}$$

How many rank 1 matrices do we need to pick to construct matrix A? In fact, this number has no upper bound. For example, if we obtain

$$A = u_1 v_1^T + u_2 v_2^T$$

Then we can always decompose any rank 1 matrix into 2 rank 1 matrices:

$$A = u_1 v_1^T + \frac{1}{2} u_2 v_2^T + \frac{1}{2} u_2 v_2^T.$$

But this number has a lower bound, that is rank. In other words, rank(A) = smallest number of rank 1 matrices with sum A.

8.2.3. Best Low-Rank Approximation

Given matrix A. What is the *best rank k approximation*? In other words, given matrix $A \in \mathbb{R}^{m \times n}$, what is the optimal solution for the optimization:

$$\begin{aligned} \min_{Z} \quad & \|A - Z\|_F^2 \\ \text{s.t.} \quad & \text{rank}(Z) = k \\ & Z \in \mathbb{R}^{m \times n} \end{aligned}$$

Firstly let's introduce the definition for Frobenius norm:

Definition 8.12 [Frobenius norm] The Frobenius norm for $m \times n$ matrix A is given by

$$\|A\|_F = \sqrt{\langle A, A \rangle} = \sqrt{\text{trace}(A^T A)}.$$

Theorem 8.4 Suppose the SVD for $A \in \mathbb{R}^{m \times n}$ is given by

$$A = \sigma_1 u_1 v_1^T + \cdots + \sigma_r u_r v_r^T.$$

with $\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$.

Then the best rank $k(k \leq r)$ approximation of $\boldsymbol{A}$ is

$$\boldsymbol{A}_k = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \cdots + \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^{\mathrm{T}}.$$

For example, $\sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}}$ is the best rank 1 approximation of $\boldsymbol{A}$.

8.2.3.1. Analogy with least square problem

For least squares problem, the key is to do approximation for $\boldsymbol{b} \in \mathbb{R}^m$. In other words, we just do a projection from $\boldsymbol{b}$ to the plane $\{\boldsymbol{A}\boldsymbol{x} | \boldsymbol{x} \in \mathbb{R}^n\}$:

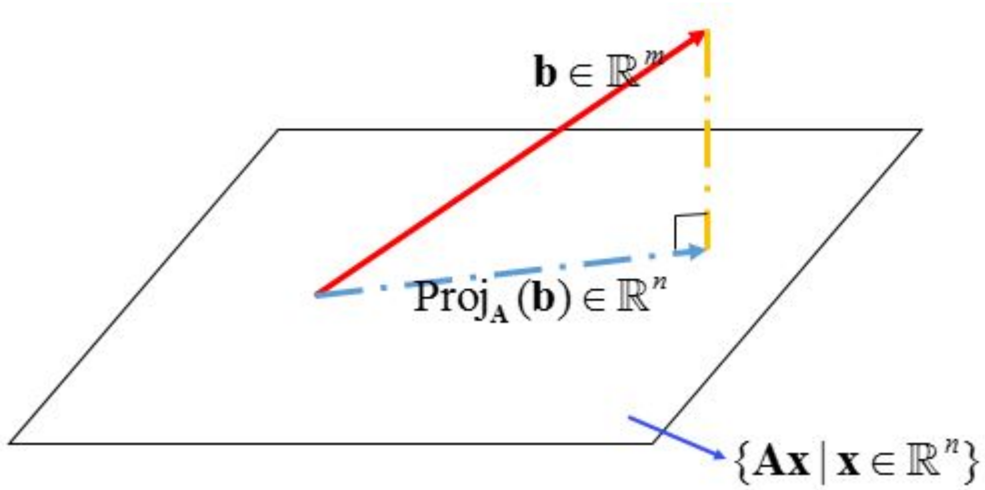


Figure 8.3: Least square problem: find $\boldsymbol{x}$ such that $\boldsymbol{A}\boldsymbol{x} = \mathrm{Proj}_{\mathcal{C}(\boldsymbol{A})}(\boldsymbol{b})$.

R For the least squares problem

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2 \\ \text{s.t.} \quad & \boldsymbol{x} \in \mathbb{R}^n \end{aligned}$$

with $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$, the key is to do the projection of $\boldsymbol{b}$ onto $\mathcal{C}(\boldsymbol{A})$, thus it suffices to solve the **equality**

$$\boldsymbol{A}\boldsymbol{x} = \mathrm{Proj}_{\mathcal{C}(\boldsymbol{A})}(\boldsymbol{b}).$$

Similarly, the beast rank k approximation could be viewed as a projction from $\boldsymbol{A}$ with rank r to the "plane" that contains all rank k matrices:

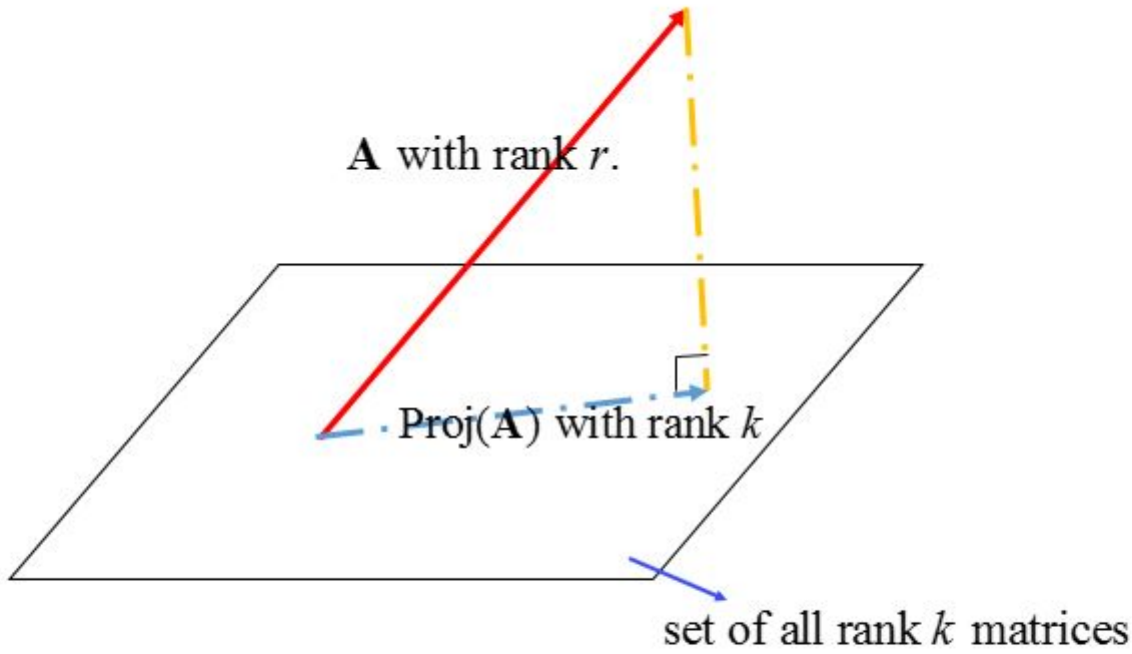


Figure 8.4: Best rank k approximation: find the projection from matrix $\boldsymbol{A}$ with rank r onto the plane that contains all rank k matrices

R Similarly, for the best rank k approximation problem

$$\begin{aligned} \min_{\boldsymbol{Z}} \quad & \|\boldsymbol{A}-\boldsymbol{Z}\|_F^2 \\ \text{s.t.} \quad & \operatorname{rank}(\boldsymbol{Z}) = k \\ & \boldsymbol{Z} \in \mathbb{R}^{m\times n} \end{aligned}$$

with $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, the key is to do the projection of $\boldsymbol{A}$ onto the set $\mathcal{M} = \{\boldsymbol{M} \in \mathbb{R}^{m\times n} \mid \operatorname{rank}(\boldsymbol{M}) = k\}$, thus it suffices to solve the **equality**

$$\boldsymbol{Z} = \operatorname{Proj}_{\mathcal{M}}(\boldsymbol{A}).$$

For some non-convex optimization problems, this idea is very useful. The further reading is recommended:

Jain, Prateek, and P. Kar. "Non-convex Optimization for Machine Learning." Foundations & Trends® in Machine Learning 10.3-4(2017):142-336.

8.3. Assignment Eight

1. Let $\boldsymbol{A}$ be an $n \times n$ matrix. Show that $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ and $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ are *similar*.
2. Let $\boldsymbol{A}$ be $m \times n$ $(m \geq n)$ matrix with singular value decomposition $\boldsymbol{U}\Sigma\boldsymbol{V}^{\mathrm{T}}$. Let Σ^{+} denote the $n \times m$ matrix

$$\begin{pmatrix} \frac{1}{\sigma_1} & & & 0 & \dots & 0 \\ & \ddots & & \vdots & \ddots & \vdots \\ & & \frac{1}{\sigma_n} & 0 & \dots & 0 \end{pmatrix}$$

   And we define $\boldsymbol{A}^{+} = \boldsymbol{V}\Sigma^{+}\boldsymbol{U}^{\mathrm{T}}$

   (a) Show that

$$\boldsymbol{A}\boldsymbol{A}^{+} = \begin{bmatrix} \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{A}^{+}\boldsymbol{A} = \boldsymbol{I}_n.$$

   (Note that $\boldsymbol{A}^{+}$ is called the **pseudo-inverse** of $\boldsymbol{A}$.)

   (b) If $\operatorname{rank}(\boldsymbol{A}) = n$, Show that $\hat{\boldsymbol{x}} = \boldsymbol{A}^{+}\boldsymbol{b}$ satisfies the normal equation $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$.

3. Suppose $\boldsymbol{A} \in \mathbb{R}^{m\times n}(m \geq n)$ has an SVD

$$\boldsymbol{A} = \sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + \cdots + \sigma_n\boldsymbol{u}_n\boldsymbol{v}_n^{\mathrm{T}}$$

   where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$.

   (a) Prove that $\|\boldsymbol{A}\|_F^2 = \sum_{i=1}^{n}\sigma_i^2$.

   (b) Let $\boldsymbol{A}_k$ be the *best rank-k approximation* of $\boldsymbol{A}$, what is $\|\boldsymbol{A} - \boldsymbol{A}_k\|_F$?

4. Suppose the *maximal singular value* of $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is σ_1, prove

$$\sigma_1 = \max_{\boldsymbol{x},\boldsymbol{y}} \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y}$$

   where $\boldsymbol{x} \in \mathbb{R}^m, \boldsymbol{y} \in \mathbb{R}^n, \|\boldsymbol{x}\| = \|\boldsymbol{y}\| = 1$.

5. Let $\boldsymbol{A}$ be a *symmetric positive definite* $n \times n$ matrix. Show that $\boldsymbol{A}$ can be factored into a product $\boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}}$, where $\boldsymbol{Q}$ is an $n \times n$ matrix whose columns are *mutually*

orthogonal.

Chapter 9

Final Exam

9.1. Sample Exam

DURATION OF EXAMINATION: 2 hours in-class

This examination paper includes 6 *pages and* 6 *problems. You are responsible for ensuring that your copy of the paper is complete. Bring any discrepancy to the attention of your invigilator.*

1. **(20 points)** *Matrix representation for linear transformation*

   Let D be defined as (*differentiate operator*):

   $$D(f) = \frac{\mathrm{d}f}{\mathrm{d}x}$$

   Consider the space $\text{span}\{\sin x, \cos x, \sin 2x, \cos 2x\}$.

   (a) Write down a *matrix representation* of T with respect to the basis $\{\sin x, \cos x, \sin 2x, \cos$

   (b) If a polynomial $f(x)$ satisfies

   $$T(f) = \lambda f,$$

   we say f is an *eigenvector* of T.

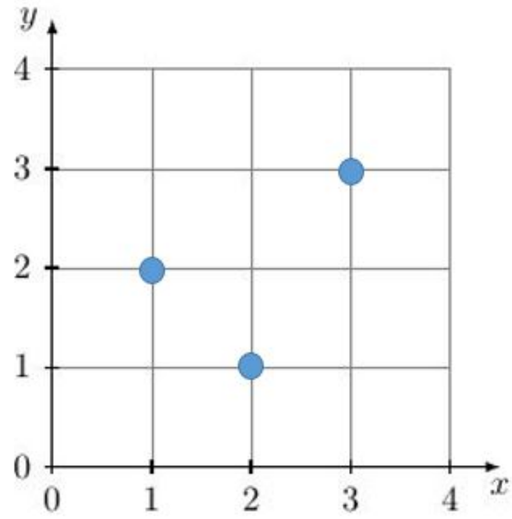
   Find 4 *linearly independent* eigenvectors of D^2. In other words, find f_k such that

   $$D^2(f_k) = \lambda_k f_k$$

for $k = 1, 2, 3, 4$.

2. **(20 points)** *Least Square Method*

(a) Find the *least squares fit line* $y = C + Dx$ to the following 3 data points:



(b) Let $\boldsymbol{A}$ be a matrix with *linearly independent columns* and consider the *projection matrix* $\boldsymbol{P} = \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}$. What are the possible eigenvalues for $\boldsymbol{P}$? Give your reasons.

3. **(20 points)**

True or False. No justifications are required.

(a) For real symmetric matrix $\boldsymbol{A}$, if $\boldsymbol{A} \succ 0$, then $\boldsymbol{A}^{-1}$ exists and $\boldsymbol{A}^{-1} \succ 0$.

(b) If $\boldsymbol{A}$ is a matrix, (Note that $\boldsymbol{A}$ may not be real) then any element of the *kernel* of $\boldsymbol{A}$ is *perpendicular* to any element of the *image* of $\boldsymbol{A}^{\mathrm{T}}$.

(c) The only $m \times n$ matrix of rank 0 is $\mathbf{0}$.

(d) Let $\boldsymbol{A}$ be real square matrix. If $\boldsymbol{x}$ is in $N(\boldsymbol{A})$ and $\boldsymbol{y}$ is in $C(\boldsymbol{A}^{\mathrm{T}})$, then $\boldsymbol{x}\boldsymbol{y}^{\mathrm{T}} = 0$.

(e) If $\boldsymbol{A}$ and $\boldsymbol{B}$ are *diagonalizable* matrices, then $\begin{bmatrix} \boldsymbol{A} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{B} \end{bmatrix}$ is *diagonalizable*.

4. **(20 points)** *SVD decomposition*

(a) Find the limiting values of y_k and z_k $(k \to \infty)$:

$$\begin{cases} y_{k+1} = 0.8y_k + 0.3z_k, \\ z_{k+1} = 0.2y_k + 0.7z_k, \end{cases}$$

And $y_0 = 0, z_0 = 5$.

Hint: Show that $\begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix}$ *is similar to* $\begin{bmatrix} 0.5 & 0 \\ 0 & 1 \end{bmatrix}$.

(b) Find the SVD of the matrix $\begin{pmatrix} 0.5 & 0 \\ 0 & 1 \end{pmatrix}$.

5. **(15+5 points)** *Eigenvalues and Eigenvectors*

Given a *real symmetric* matrix $\boldsymbol{A}$, the **Rayleigh quotient** $R(\boldsymbol{x})$ is defined as

$$R(\boldsymbol{x}) = \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}} \text{ for } \boldsymbol{x} \neq \boldsymbol{0}.$$

Suppose the *eigenvalues* of $\boldsymbol{A}$ are $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$.

(a) Prove that the minimum eigenvalue λ_1 is the minimal value of $R(\boldsymbol{x})$.

$$\text{i.e. } \lambda_1 = \min_{\boldsymbol{x} \in \mathbb{R}^n - \{\boldsymbol{0}\}} R(\boldsymbol{x}).$$

(b) Suppose $\boldsymbol{x}_1$ is the eigenvector associated with λ_1, i.e. $\boldsymbol{A}\boldsymbol{x}_1 = \lambda_1 \boldsymbol{x}_1$.

$$\text{Prove that } \lambda_2 = \min_{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}_1 = 0} R(\boldsymbol{y}).$$

(c) **(bonus question)**

Suppose $\boldsymbol{v} \in \mathbb{R}^n$ is an arbitrary given vector.

$$\text{Prove that } \lambda_2 \geq \min_{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{v} = 0} R(\boldsymbol{y}).$$

6. **(10 points)** *Positive semi-definite*

Definition 9.1 [diagonal dominant]

A matrix $\boldsymbol{M} \in \mathbb{R}^{n\times n}$ is called **diagonal dominant** if for $\forall i \in \{1,2,\dots,n\}$,

$$|\boldsymbol{M}_{ii}| \ge \sum_{j\neq i} |\boldsymbol{M}_{ij}|.$$

It is called **strictly diagonal dominant** if for $\forall i \in \{1,2,\dots,n\}$,

$$|\boldsymbol{M}_{ii}| > \sum_{j\neq i} |\boldsymbol{M}_{ij}|.$$

Prove the following statements:

(a) $\boldsymbol{Z} = \begin{pmatrix} 5 & 1 & 4 \\ 1 & 5 & 3 \\ 4 & 3 & 7 \end{pmatrix}$ is *positive semi-definite*.

(b) If $\boldsymbol{M}$ is *symmetric* and *diagonal dominant*, then $\boldsymbol{M} \succeq 0$.

9.2. Final Exam

DURATION OF EXAMINATION: 2 hours and 35 minutes in-class

This examination paper includes 6 pages and 6 problems. You are responsible for ensuring that your copy of the paper is complete. Bring any discrepancy to the attention of your invigilator.

1. **(20 points)** *Matrix representation for linear transformation*

   (a) Let T be the transformation

   $$T : \{\textbf{polynomials of degree} \le 4\} \mapsto \{\textbf{polynomials of degree} \le 4\}$$
   $$T(p) = (x-2)\frac{\mathrm{d}p}{\mathrm{d}x}$$

   Show that T is a *linear transformation* and write down a *matrix representation* of T with respect to basis $\{1, x, x^2, x^3, x^4\}$ for the input and output space.

   (b) If a polynomial $f(x)$ satisfies

   $$T(f) = \lambda f,$$

   we say f is an *eigenvector* of T. Find two *linearly independent* eigenvectors of T.

2. **(20 points)** *Least Square Method*

(a) Find the projection of $\boldsymbol{z} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ onto the column space of $\begin{bmatrix} 1 & -1 \\ 1 & -1 \\ -2 & 4 \end{bmatrix}$.

(b) Let $\mathcal{A} : \mathbb{R}^{2\times 1} \mapsto \mathbb{R}^{2\times 2}$ be a mapping defined as

$$\mathcal{A}\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a+b & a-b \\ -2a+4b & 0 \end{bmatrix}, \forall a, b \in \mathbb{R}.$$

Define $\kappa = \{\boldsymbol{Ax} | \boldsymbol{x} \in \mathbb{R}^{2\times 1}\}$.

Find the best approximation of $\boldsymbol{B} = \begin{bmatrix} 1 & 2 \\ 7 & 1 \end{bmatrix}$ in the space κ.

Hint: Consider $\begin{bmatrix} 1 & 2 \\ 7 & 1 \end{bmatrix}$ *and* $\begin{bmatrix} a+b & a-b \\ -2a+4b & 0 \end{bmatrix}$ *as* $\mathbb{R}^{4\times 1}$ *vector.*

Then you only need to find the best approximation of $\begin{pmatrix} 1 \\ 2 \\ 7 \\ 1 \end{pmatrix}$ *onto the set* $\{\boldsymbol{Ax} | \boldsymbol{x} \in \mathbb{R}^{2\times 1}\}$, *where* $\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ -2 & 4 \\ 0 & 0 \end{bmatrix}$.

3. **(20 points)**

True or False. No justifications are required.

(a) If all the entries of a square matrix $\boldsymbol{A}$ are *positive*, then $\boldsymbol{A}^{-1}$ exist.

(b) If $\boldsymbol{Q}$ is an *orthogonal matrix*, then $\det(\boldsymbol{Q}) = \pm 1$.

(c) If $\boldsymbol{A}$ is a *negative definite* matrix, then its singular values have the same absolute values as its eigenvalues.

Hint: Note that $\boldsymbol{A}$ is said to be negative definite when $-\boldsymbol{A}$ is positive definite.

(d) If $\boldsymbol{A}$ is an $n \times n$ matrix with *characteristic polynomial* $p_{\boldsymbol{A}}(t) = t^n$, then $\boldsymbol{A} = \boldsymbol{0}$.

(e) If $\boldsymbol{A}$ is the sum of 5 rank one matrices, then $\text{rank}(\boldsymbol{A}) \leq 5$.

4. **(20 points)** *SVD decomposition*

The question is about the matrix

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

(a) Find its eigenvalues and eigenvectors, write the vector $\boldsymbol{u} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$ as a combination of those eigenvectors.

(b) Do the SVD decomposition to derive $\boldsymbol{A} = \boldsymbol{U}\Sigma\boldsymbol{V}^{\mathrm{T}}$ in two steps:

- First, compute $\boldsymbol{V}$ and Σ using the matrix $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$.
- Second, find the (*orthonormal*) columns of $\boldsymbol{U}$.

5. **(15+5 points)** *Eigenvalues and Eigenvectors*

(a) Suppose $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n\times n}$ can can be *diagonalized* by the same matrix, prove that $\boldsymbol{A}\boldsymbol{B} = \boldsymbol{B}\boldsymbol{A}$.

Hint: Note that $\boldsymbol{A}$ *is said to be diagonalized by* $\boldsymbol{S}$ *if* $\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}$ *is diagonal.*

(b) Suppose $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n\times n}$ satisfy $\boldsymbol{A}\boldsymbol{B} = \boldsymbol{B}\boldsymbol{A}$, and both $\boldsymbol{A}$ and $\boldsymbol{B}$ are diagonalizable. $\boldsymbol{A}$ has *n distinct* eigenvalues. Prove that $\boldsymbol{A}, \boldsymbol{B}$ can can be *diagonalized* by the same matrix.

Hint: Suppose $\boldsymbol{A}$ *has eigenvectors* $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$*. You can express* $\boldsymbol{B}\boldsymbol{v}_i$ *as linear combination of* $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$*. Then you can express* $\boldsymbol{A}(\boldsymbol{B}\boldsymbol{v}_i)$ *and* $\boldsymbol{B}(\boldsymbol{A}\boldsymbol{v}_i)$*. Finally compute* $\boldsymbol{A}(\boldsymbol{B}\boldsymbol{v}_i) - \boldsymbol{B}(\boldsymbol{A}\boldsymbol{v}_i)$ *to derive something.*

(c) **(bonus question)**

Prove part (b) without the assumption that $\boldsymbol{A}$ has *n distinct* eigenvalues. (i.e. $\boldsymbol{A}$ might have **repeated** eigenvalues)

Hint: Since $\boldsymbol{A}$ *is diagonalizable, there exists* $\boldsymbol{Q}$ *such that* $\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{Q} = \boldsymbol{D}$*, where* $\boldsymbol{D}$ *is diagonal. Then you should express* $\boldsymbol{D}$*. Then you compute* $\boldsymbol{Q}^{-1}\boldsymbol{B}\boldsymbol{Q} = \boldsymbol{C}$*, i.e. partition* $\boldsymbol{C}$ *in the same way of* $\boldsymbol{D}$*. Next you should show us that* $\boldsymbol{C}$ *is block diagonal. Then you construct* ***diagonal*** *matrix* $\boldsymbol{T}_*$ *that diagonalize* $\boldsymbol{C}$*. Finally you construct* $\boldsymbol{P}$ *that diagonalize both* $\boldsymbol{A}$ *and* $\boldsymbol{B}$*.*

6. **(10 points)** *Positive definite*

Suppose $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n \times n}$, where $\boldsymbol{A} = \begin{bmatrix} a_{ij} \end{bmatrix}_{i,j=1}^{n}, \boldsymbol{B} = \begin{bmatrix} b_{ij} \end{bmatrix}_{i,j=1}^{n}$.
Define the **Hadamard product** $\boldsymbol{A} \circ \boldsymbol{B}$ as an $n \times n$ matrix with entries

$$\begin{bmatrix} \boldsymbol{A} \circ \boldsymbol{B} \end{bmatrix}_{ij} = a_{ij} b_{ij}.$$

For example, if $\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 3 & 7 \end{bmatrix}, \boldsymbol{B} = \begin{bmatrix} 0 & \pi \\ 1 & e \end{bmatrix}$, then $\boldsymbol{A} \circ \boldsymbol{B} = \begin{bmatrix} 0 & 2\pi \\ 3 & 7e \end{bmatrix}$.
Prove the following statements:

(a) $\text{rank}(\boldsymbol{A} \circ \boldsymbol{B}) \leq \text{rank}(\boldsymbol{A}) \text{rank}(\boldsymbol{B})$;

Hint: Extend Hadamard product into vector. Then it's easy to verify that $(\boldsymbol{A} \circ \boldsymbol{B}) \circ \boldsymbol{C} = \boldsymbol{A} \circ \boldsymbol{C} + \boldsymbol{B} \circ \boldsymbol{C}$ *and* $(\boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}}) \circ (\boldsymbol{u}_2 \circ \boldsymbol{v}_2^{\mathrm{T}}) = (\boldsymbol{u}_1 \circ \boldsymbol{u}_2) \times (\boldsymbol{v}_1 \circ \boldsymbol{v}_2)^{\mathrm{T}}$. *Then you can do SVD decomposition for* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *(vector form, related to rank.) Then you can express* $\boldsymbol{A} \circ \boldsymbol{B}$ *as the sum of some matrices with rank 1.*

(b) If $\boldsymbol{A} \succeq 0, \boldsymbol{B} \succeq 0$ and $\boldsymbol{A}, \boldsymbol{B}$ are *symmetric matrix*, prove that

$$\boldsymbol{A} \circ \boldsymbol{B} \succeq 0.$$

Hint: Note that $\boldsymbol{A} = \boldsymbol{R}^{\mathrm{T}} \boldsymbol{R}$, *where* $\boldsymbol{R}$ *is square. Then you should express* $\boldsymbol{R}^{\mathrm{T}} \boldsymbol{R}$ *into vector form. Similarly, you can express* $\boldsymbol{B}$ *into vector form. Then you compute* $\boldsymbol{A} \circ \boldsymbol{B}$ *and show it is PSD by definition.*

Chapter 10

Solution

10.1. Assignment Solutions

10.1.1. Solution to Assignment One

1. *Solution.* Firstly we do the elimination shown as below:

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

Here in order to give three pivots we need to let the diagonal be nonzero, which is to say:

$$a = 0 \qquad \text{or} \qquad a-2=0 \qquad \text{or} \qquad a-4=0$$

$$\Longrightarrow a = 0 \qquad \text{or} \qquad a=2 \qquad \text{or} \qquad a=4$$

■

2. let's solve this problem by answering the following questions first.

   (a) The other solution is given by:$(m_1x + m_2X, m_1y + m_2Y, m_1z + m_2Z)$, where $m_1 + m_2 = 1$.

   (b) They also meet the line that passes these two points

   (c) In $\mathbb{R}^n$ space we can also ensure every point on the line that determined by the two solutions is also a solution.

   Then let's proof the begining statement rigorously:

Proof. Assume the system of equation is given by

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1$$
$$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2$$
$$\cdots$$
$$a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \tag{10.1}$$

where it contains two solutions $(y_1, y_2, \ldots, y_n)$ and $(z_1, z_2, \ldots, z_n)$. Let's show that every point on the line that determined by the two solutions is also a solution. In other words, once the system has two solutions, it will contain infinitely many solutions.

Any point on the line that determined by the two solutions is given by

$$(m_1y_1 + m_2z_1, \ldots, m_1y_n + m_2z_n), \qquad \text{where } m_1 + m_2 = 1$$

And then we show that this point is also a solution to this system:

for the ith linear equation it satisfies that

$$\begin{cases} a_{i1}y_1 + a_{i2}y_2 + \cdots + a_{in}y_n = b_i \\ a_{i1}z_1 + a_{i2}z_2 + \cdots + a_{in}z_n = b_i \end{cases}$$

Hence we set $x_j = m_1y_j + m_2z_j$ for $j = 1, 2, \ldots, n$. Then we obtain:

$$\begin{aligned} &a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \\ &\quad = a_{i1}(m_1y_1 + m_2z_1) + a_{i2}(m_1y_2 + m_2z_2) + \cdots + a_{in}(m_1y_n + m_2z_n) \\ &\quad = m_1(a_{i1}y_1 + a_{i2}y_2 + \cdots + a_{in}y_n) + m_2(a_{i1}z_1 + a_{i2}z_2 + \cdots + a_{in}z_n) \\ &\quad = m_1b_i + m_2b_i = (m_1 + m_2)b_i = b_i. \end{aligned}$$

where $i = 1, 2, \ldots, m$

Since the choice of point on the line was arbitrary, we see that every point on the line determined by the two solutions is also a solution, so there are infinitely many solutions to the system ■

3. *Solution.* (a) We begin to do the elimination for the system:

$$\left[\begin{array}{ccc|c} 1 & 4 & -2 & 1 \\ 1 & 7 & -6 & 6 \\ 0 & 3 & q & t \end{array}\right] \xrightarrow{\text{Add } (-1)\times \text{ row 1 to row 2}} \left[\begin{array}{ccc|c} 1 & 4 & -2 & 1 \\ 0 & 3 & -4 & 5 \\ 0 & 3 & q & t \end{array}\right]$$

$$\xrightarrow{\text{Add } (-1)\times \text{ row 2 to row 3}} \left[\begin{array}{ccc|c} 1 & 4 & -2 & 1 \\ 0 & 3 & -4 & 5 \\ 0 & 0 & q+4 & t-5 \end{array}\right]$$

In order to make this system singular we need to make the third row has no pivot. $\implies q+4=0 \implies q=-4$. In order to give infinitely many solutions we have to let the third equation satisfies $0=0$. $\implies t-5=0 \implies t=5$.

(b) When $z=1$, the second equation $3y-4z=5$ gives $y=3$; the third equation $x+4y-2z=1$ gives $x=-9$.

■

4. *Solution.* (a)

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \implies \boldsymbol{A}^2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$$

(b)

$$\boldsymbol{B} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \implies \boldsymbol{B}^2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \boldsymbol{0}$$

(c)

$$\boldsymbol{C} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}; \boldsymbol{D} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \implies \boldsymbol{CD} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} = -\boldsymbol{DC}$$

(d)

$$\boldsymbol{E} = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}; \boldsymbol{F} = \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix} \implies \boldsymbol{EF} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \boldsymbol{0}$$

■

5. *Proof.* We assume $\boldsymbol{A}$ is a $m \times n$ matrix,$\boldsymbol{B}$ is a $n \times p$ matrix,$\boldsymbol{C}$ is a $p \times q$ matrix which is given by:

$$\boldsymbol{A} := \begin{bmatrix} a_{ij} \end{bmatrix}, \boldsymbol{B} := \begin{bmatrix} b_{ij} \end{bmatrix}, \boldsymbol{C} := \begin{bmatrix} c_{ij} \end{bmatrix}.$$

And we also define:

$$\boldsymbol{AB} := \boldsymbol{D} := \begin{bmatrix} d_{ij} \end{bmatrix}, \boldsymbol{BC} := \boldsymbol{E} := \begin{bmatrix} e_{ij} \end{bmatrix}.$$

Obviously, $\boldsymbol{AB}$ and $\boldsymbol{BC}$ are well-defined and they are all $m \times q$ matrix.

•According to the definition for multiplication, $d_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$. We define $(\boldsymbol{AB})\boldsymbol{C} := \boldsymbol{H} = \begin{bmatrix} h_{ij} \end{bmatrix}$, thus

$$h_{ij} = \sum_{l=1}^{p} d_{il} c_{lj} = \sum_{l=1}^{p} (\sum_{k=1}^{n} a_{ik} b_{kl}) c_{lj} = \sum_{k=1}^{n} \sum_{l=1}^{p} a_{ik} b_{kl} c_{lj}$$

where $i = 1, 2, \dots, m$ and $i = 1, 2, \dots, q$.

•On the other hand, $e_{ij} = \sum_{l=1}^{p} b_{il} c_{lj}$. We define $\boldsymbol{A}(\boldsymbol{BC}) := \boldsymbol{G} = \begin{bmatrix} g_{ij} \end{bmatrix}$, thus

$$g_{ij} = \sum_{k=1}^{n} a_{ik} e_{kj} = \sum_{k=1}^{n} (\sum_{l=1}^{p} b_{kl} c_{lj}) a_{ik} = \sum_{k=1}^{n} \sum_{l=1}^{p} a_{ik} b_{kl} c_{lj}$$

where $i = 1, 2, \dots, m$ and $i = 1, 2, \dots, q$.

Hence we have $h_{ij} = g_{ij}$, $i = 1, 2, \dots, m$ and $i = 1, 2, \dots, q$. Hence we have $\boldsymbol{H} = \boldsymbol{G} \implies (\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$. ■

6. *Solution.*

For matrix $\boldsymbol{A} = \begin{bmatrix} 4 & 0 & 4 \\ 6 & 6 & -8 \\ -9 & 5 & -8 \end{bmatrix}$, we can split $\boldsymbol{A}$ into blocks $\boldsymbol{A} = \left[\begin{array}{cc|c} 4 & 0 & 4 \\ 6 & 6 & -8 \\ \hline -9 & 5 & -8 \end{array}\right] =$

$\begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}$, where $A_1 = \begin{bmatrix} 4 & 0 \\ 6 & 6 \end{bmatrix}$, $A_2 = \begin{bmatrix} 4 \\ -8 \end{bmatrix}$, $A_3 = \begin{bmatrix} -9 & 5 \end{bmatrix}$, $A_4 = \begin{bmatrix} -8 \end{bmatrix}$.

For matrix $\boldsymbol{B} = \begin{bmatrix} 8 & -3 & -7 \\ 3 & -7 & -4 \\ 4 & -4 & 1 \end{bmatrix}$, we can split $\boldsymbol{B}$ into blocks $\boldsymbol{B} = \left[\begin{array}{cc|c} 8 & -3 & -7 \\ 3 & -7 & -4 \\ \hline 4 & -4 & 1 \end{array}\right] =$ $\begin{bmatrix} B_1 & B_2 \\ B_3 & B_4 \end{bmatrix}$, where $B_1 = \begin{bmatrix} 8 & -3 \\ 3 & -7 \end{bmatrix}$, $B_2 = \begin{bmatrix} -7 \\ -4 \end{bmatrix}$, $B_3 = \begin{bmatrix} 4 & -4 \end{bmatrix}$, $B_4 = \begin{bmatrix} 1 \end{bmatrix}$.

We let $\boldsymbol{C} = \boldsymbol{AB} = \begin{bmatrix} C_1 & C_2 \\ C_3 & C_4 \end{bmatrix}$,we can find C_1, C_2, C_3, C_4 in two different ways, if we get the same answers, we can verify the block multiplication succeeds.

(a) Multiply $\boldsymbol{A}$ times $\boldsymbol{B}$ to find $\boldsymbol{C} = \left[\begin{array}{cc|c} 48 & -28 & -24 \\ 34 & -28 & -74 \\ \hline -89 & 24 & 35 \end{array}\right]$,

Hence $C_1 = \begin{bmatrix} 48 & -28 \\ 34 & -28 \end{bmatrix}$, $C_2 = \begin{bmatrix} -24 \\ -74 \end{bmatrix}$, $C_3 = \begin{bmatrix} -89 & 24 \end{bmatrix}$, $C_4 = \begin{bmatrix} 35 \end{bmatrix}$.

(b) On the other hand, we have $\begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix} \begin{bmatrix} B_1 & B_2 \\ B_3 & B_4 \end{bmatrix} = \begin{bmatrix} A_1B_1 + A_2B_3 & A_1B_2 + A_2B_4 \\ A_3B_1 + A_4B_3 & A_3B_2 + A_4B_4 \end{bmatrix}$

Hence we find $C_1 = A_1B_1 + A_2B_3 = \begin{bmatrix} 4 & 0 \\ 6 & 6 \end{bmatrix} \begin{bmatrix} 8 & -3 \\ 3 & -7 \end{bmatrix} + \begin{bmatrix} 4 \\ -8 \end{bmatrix} \begin{bmatrix} 4 & -4 \end{bmatrix} =$ $\begin{bmatrix} 48 & -28 \\ 34 & -28 \end{bmatrix}$.

Similarly, we have

$$C_2 = A_1B_2 + A_2B_4 = \begin{bmatrix} -24 \\ -74 \end{bmatrix}$$

$$C_3 = A_3B_1 + A_4B_3 = \begin{bmatrix} -89 & 24 \end{bmatrix}$$

$$C_4 = A_3B_2 + A_4B_4 = \begin{bmatrix} 35 \end{bmatrix}.$$

■

7. *Solution.*

$$\boldsymbol{A} = \begin{bmatrix} a & a & a & a \\ a & b & b & b \\ a & b & c & c \\ a & b & c & d \end{bmatrix} \xRightarrow{\boldsymbol{E}_{41}\boldsymbol{E}_{31}\boldsymbol{E}_{21}} \begin{bmatrix} a & a & a & a \\ 0 & b-a & b-a & b-a \\ 0 & b-a & c-a & c-a \\ 0 & b-a & c-a & d-a \end{bmatrix}$$

$$\xRightarrow{\boldsymbol{E}_{42}\boldsymbol{E}_{32}} \begin{bmatrix} a & a & a & a \\ 0 & b-a & b-a & b-a \\ 0 & 0 & c-b & c-b \\ 0 & 0 & c-b & d-b \end{bmatrix} \xRightarrow{\boldsymbol{E}_{43}} \begin{bmatrix} a & a & a & a \\ 0 & b-a & b-a & b-a \\ 0 & 0 & c-b & c-b \\ 0 & 0 & 0 & d-c \end{bmatrix} = \boldsymbol{U}$$

$$\implies \boldsymbol{E}_{43}\boldsymbol{E}_{42}\boldsymbol{E}_{32}\boldsymbol{E}_{41}\boldsymbol{E}_{31}\boldsymbol{E}_{21}\boldsymbol{A} = \boldsymbol{U} \implies \boldsymbol{A} = \boldsymbol{E}_{21}^{-1}\boldsymbol{E}_{31}^{-1}\boldsymbol{E}_{41}^{-1}\boldsymbol{E}_{32}^{-1}\boldsymbol{E}_{42}^{-1}\boldsymbol{E}_{43}^{-1}\boldsymbol{U}$$

$$\implies \boldsymbol{A} = \begin{bmatrix} a & a & a & a \\ a & b & b & b \\ a & b & c & c \\ a & b & c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} a & a & a & a \\ 0 & b-a & b-a & b-a \\ 0 & 0 & c-b & c-b \\ 0 & 0 & 0 & d-c \end{bmatrix}$$

$$\implies \boldsymbol{L} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}; \qquad \boldsymbol{U} = \begin{bmatrix} a & a & a & a \\ 0 & b-a & b-a & b-a \\ 0 & 0 & c-b & c-b \\ 0 & 0 & 0 & d-c \end{bmatrix}$$

In order to get four pivots, we need to let the diagonal entries of $\boldsymbol{U}$ to be nonzero.

$$\implies a \neq 0 \qquad a \neq b \qquad b \neq c \qquad c \neq d$$

■

10.1.2. Solution to Assignment Two

1. *Proof. Sufficiency.*

   If $\boldsymbol{M}$ is invertible, then there exists matrix $\boldsymbol{N}$ such that $\boldsymbol{MN} = \boldsymbol{NM} = \boldsymbol{I}$.

   $$\implies (\boldsymbol{ABC})\boldsymbol{N} = \boldsymbol{I}, \boldsymbol{N}(\boldsymbol{ABC}) = \boldsymbol{I}. \implies \boldsymbol{A}(\boldsymbol{BCN}) = \boldsymbol{I}, (\boldsymbol{NAB})\boldsymbol{C} = \boldsymbol{I}.$$

   $\implies \boldsymbol{BCN}$ is the right inverse of $\boldsymbol{A}$, $\boldsymbol{NAB}$ is the left inverse of $\boldsymbol{C}$.

   Hence $\boldsymbol{A}$ and $\boldsymbol{C}$ is invertible.

   Moreover, $(\boldsymbol{ABC})\boldsymbol{N} = \boldsymbol{I} \implies (\boldsymbol{AB})\boldsymbol{CN} = \boldsymbol{I}$. Hence $\boldsymbol{CN}$ is the right inverse of $\boldsymbol{AB}$.

   Hence $\boldsymbol{AB}$ is invertible. Hence there exists $(\boldsymbol{AB})^{-1}$ such that $((\boldsymbol{AB})^{-1})(\boldsymbol{AB}) = \boldsymbol{I}$.

   $\implies ((\boldsymbol{AB})^{-1}\boldsymbol{A})\boldsymbol{B} = \boldsymbol{I}$. Hence $(\boldsymbol{AB})^{-1}\boldsymbol{A}$ is the left inverse of $\boldsymbol{B}$.

   Hence $\boldsymbol{B}$ is invertible.

   Necessity.

   If $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ is invertible, then there exist $\boldsymbol{A}^{-1}, \boldsymbol{B}^{-1}, \boldsymbol{C}^{-1}$ such that

   $$\boldsymbol{AA}^{-1} = \boldsymbol{I}, \boldsymbol{BB}^{-1} = \boldsymbol{I}, \boldsymbol{CC}^{-1} = \boldsymbol{I}.$$

   $$\begin{aligned}\implies \boldsymbol{ABC}(\boldsymbol{C}^{-1}\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}) &= \boldsymbol{AB}(\boldsymbol{CC}^{-1})(\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}) = \boldsymbol{ABI}(\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}) \\ &= \boldsymbol{AB}(\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}) = \boldsymbol{A}(\boldsymbol{BB}^{-1})\boldsymbol{A}^{-1} = \boldsymbol{AIA}^{-1} \\ &= \boldsymbol{AA}^{-1} = \boldsymbol{I}.\end{aligned}$$

   Hence $\boldsymbol{C}^{-1}\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$ is the right inverse of $\boldsymbol{ABC}$. Hence $\boldsymbol{ABC}$ is invertible. ■

   ■

   ■

2. *Solution.* The inverse are respectively given by

   $$\begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ -\boldsymbol{C} & \boldsymbol{I} \end{bmatrix}, \quad \begin{bmatrix} \boldsymbol{A}^{-1} & \boldsymbol{0} \\ -\boldsymbol{D}^{-1}\boldsymbol{CA}^{-1} & \boldsymbol{D}^{-1} \end{bmatrix}, \quad \begin{bmatrix} -\boldsymbol{D} & \boldsymbol{I} \\ \boldsymbol{I} & \boldsymbol{0} \end{bmatrix}.$$

- $$\begin{bmatrix} I & 0 \\ C & I \end{bmatrix} \begin{bmatrix} I & 0 \\ -C & I \end{bmatrix} = \begin{bmatrix} II + 0(-C) & I0 + 0I \\ CI + I(-C) & C0 + II \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$$

  Hence $\begin{bmatrix} I & 0 \\ -C & I \end{bmatrix}$ is the right inverse of $\begin{bmatrix} I & 0 \\ C & I \end{bmatrix}$, hence $\begin{bmatrix} I & 0 \\ -C & I \end{bmatrix}$ is the inverse of $\begin{bmatrix} I & 0 \\ C & I \end{bmatrix}$.

- $$\begin{bmatrix} A & 0 \\ C & D \end{bmatrix} \begin{bmatrix} A^{-1} & 0 \\ -D^{-1}CA^{-1} & D^{-1} \end{bmatrix} = \begin{bmatrix} AA^{-1} + 0(-D^{-1}CA^{-1}) & A0 + 0D^{-1} \\ CA^{-1} + D(-D^{-1}CA^{-1}) & C0 + DD^{-1} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$$

  Hence $\begin{bmatrix} A^{-1} & 0 \\ -D^{-1}CA^{-1} & D^{-1} \end{bmatrix}$ is the right inverse of $\begin{bmatrix} A & 0 \\ C & D \end{bmatrix}$, hence $\begin{bmatrix} A^{-1} & 0 \\ -D^{-1}CA^{-1} & D^{-1} \end{bmatrix}$ is the inverse of $\begin{bmatrix} A & 0 \\ C & D \end{bmatrix}$.

- $$\begin{bmatrix} 0 & I \\ I & D \end{bmatrix} \begin{bmatrix} -D & I \\ I & 0 \end{bmatrix} = \begin{bmatrix} 0(-D) + II & 0I + I0 \\ I(-D) + DI & II + D0 \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$$

  Hence $\begin{bmatrix} -D & I \\ I & 0 \end{bmatrix}$ is the right inverse of $\begin{bmatrix} 0 & I \\ I & D \end{bmatrix}$, hence $\begin{bmatrix} -D & I \\ I & 0 \end{bmatrix}$ is the inverse of $\begin{bmatrix} 0 & I \\ I & D \end{bmatrix}$.

■

3. *Solution.* Firstly, we do Elimination for this matrix:

$$\begin{bmatrix} 2 & c & c \\ c & c & c \\ 8 & 7 & c \end{bmatrix} \xRightarrow[E_{21}=\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{c}{2} & 0 & 1 \end{bmatrix}]{E_{31}=\begin{bmatrix} 1 & 0 & 0 \\ -4 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}} \begin{bmatrix} 2 & c & c \\ 0 & c-\frac{c^2}{2} & c-\frac{c^2}{2} \\ 0 & 7-4c & -3c \end{bmatrix}$$

Notice that $c - \frac{c^2}{2} \neq 0$, otherwise the second row has no nonzero entries, the Gaussian Elimination cannot continue.

$$\xRightarrow{E_{32}=\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{4c-7}{c-c^2/2} & 1 \end{bmatrix}} \begin{bmatrix} 2 & c & c \\ 0 & c-\frac{c^2}{2} & c-\frac{c^2}{2} \\ 0 & 0 & c-7 \end{bmatrix}$$

In order to continue the Gaussian Elimination, we have to let three pivots not equal to zero, hence we have $c - \frac{c^2}{2} \neq 0, c-7 \neq 0$.

Hence $c \neq 0, c \neq 2, c \neq 7$. ■

4. *Solution.*

   (a) True, because if the whole row has no nonzero entries, the pivot in this row doesn't exist, the Gaussian Elimination cannot continue, hence there doesn't exist the inverse.

   (b) False, for example, for matrix $\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, if we do elimination, we obtain

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

   so we cannot continue Gaussian Elimination as the second row has no pivot, hence $\boldsymbol{A}$ is not invertible.

(c) True, if $\boldsymbol{A}$ is invertible, we have $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}$. Hence $\boldsymbol{A}$ is the left inverse of $\boldsymbol{A}^{-1}$. Hence $\boldsymbol{A}$ is the inverse of $\boldsymbol{A}^{-1}$.

(d) True, if $\boldsymbol{A}^{\mathrm{T}}$ is invertible, there exists $\boldsymbol{B}$ such that $\boldsymbol{B}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{I}$.

$$\implies (\boldsymbol{B}\boldsymbol{A}^{\mathrm{T}})^{\mathrm{T}} = (\boldsymbol{A}^{\mathrm{T}})^{\mathrm{T}}(\boldsymbol{B})^{\mathrm{T}} = \boldsymbol{A}\boldsymbol{B}^{\mathrm{T}} = \boldsymbol{I}$$

Hence $\boldsymbol{B}^{\mathrm{T}}$ is the right inverse of $\boldsymbol{A}$. Hence $\boldsymbol{B}$ is the inverse of $\boldsymbol{A}$.

■

10.1.3. Solution to Assignment Three

1. *Solution.* (a)

$$
\begin{aligned}
\boldsymbol{M}\boldsymbol{M}^{-1} &= (\boldsymbol{I}-\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}})(\boldsymbol{I}+\frac{\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}}{1-\boldsymbol{v}^{\mathrm{T}}\boldsymbol{u}}) \\
&= \boldsymbol{I}+\frac{\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}}{1-\boldsymbol{v}^{\mathrm{T}}\boldsymbol{u}}-\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}-\frac{\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}}{1-\boldsymbol{v}^{\mathrm{T}}\boldsymbol{u}} \\
&= \boldsymbol{I}+\frac{\boldsymbol{u}\times\boldsymbol{v}^{\mathrm{T}}-(\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{u})\times\boldsymbol{v}^{\mathrm{T}}}{1-\boldsymbol{v}^{\mathrm{T}}\boldsymbol{u}}-\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}} \\
&= \boldsymbol{I}+\frac{\boldsymbol{u}\times(1-\boldsymbol{v}^{\mathrm{T}}\boldsymbol{u})\times\boldsymbol{v}^{\mathrm{T}}}{1-\boldsymbol{v}^{\mathrm{T}}\boldsymbol{u}}-\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}} \\
&= \boldsymbol{I}+\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}-\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}=\boldsymbol{I}
\end{aligned}
\tag{10.2}
$$

(b)

$$
\begin{aligned}
\boldsymbol{M}\boldsymbol{M}^{-1} &= (\boldsymbol{A}-\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}})(\boldsymbol{A}^{-1}+\frac{\boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}}{1-\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{u}}) \\
&= \boldsymbol{I}+\frac{\boldsymbol{A}\boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}}{1-\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{u}}-\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}-\frac{\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}}{1-\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{u}} \\
&= \boldsymbol{I}+\frac{\boldsymbol{I}\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}}{1-\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{u}}-\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}-\frac{\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}}{1-\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{u}} \\
&= \boldsymbol{I}+\frac{\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}-\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}}{1-\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{u}}-\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1} \\
&= \boldsymbol{I}+\frac{(\boldsymbol{u}-\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{u})\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}}{1-\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{u}}-\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1} \\
&= \boldsymbol{I}+\frac{\boldsymbol{u}(1-\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{u})\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}}{1-\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{u}}-\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1} \qquad \text{note } 1-\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{u} \text{ is scalar} \\
&= \boldsymbol{I}+\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}-\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}^{-1}=\boldsymbol{I}.
\end{aligned}
\tag{10.3}
$$

(c)

$$
\begin{aligned}
MM^{-1} &= (I_n - UV)(I_n + U(I_m - VU)^{-1}V) \\
&= I_n + U(I_m - VU)^{-1}V - UV - UVU(I_m - VU)^{-1}V \\
&= I_n + U \times (I_m - VU)^{-1}V - (UVU) \times (I_m - VU)^{-1}V - UV \\
&= I_n + (U - UVU)(I_m - VU)^{-1}V - UV \\
&= I_n + (UI_m - UVU)(I_m - VU)^{-1}V - UV \\
&= I_n + U(I_m - VU)(I_m - VU)^{-1}V - UV \\
&= I_n + UV - UV = I_n.
\end{aligned}
\tag{10.4}
$$

(d)

$$
\begin{aligned}
MM^{-1} &= (A - UW^{-1}V)(A^{-1} + A^{-1}U(W - VA^{-1}U)^{-1}VA^{-1}) \\
&= I_n + U(W - VA^{-1}U)^{-1}VA^{-1} - UW^{-1}VA^{-1} \\
&\quad - UW^{-1}VA^{-1}U(W - VA^{-1}U)^{-1}VA^{-1} \\
&= I_n + U\{(W - VA^{-1}U)^{-1} - W^{-1} - W^{-1}VA^{-1}U(W - VA^{-1}U)^{-1}\}VA^{-1} \\
&= I_n + U\{I_m(W - VA^{-1}U)^{-1} - W^{-1}(W - VA^{-1}U)(W - VA^{-1}U)^{-1} \\
&\quad - W^{-1}VA^{-1}U(W - VA^{-1}U)^{-1}\}VA^{-1} \\
&= I_n + U(I_m - W^{-1}(W - VA^{-1}U) - W^{-1}VA^{-1}U)(W - VA^{-1}U)^{-1}VA^{-1} \\
&= I_n + U(I_m - I_m + W^{-1}VA^{-1}U - W^{-1}VA^{-1}U)(W - VA^{-1}U)^{-1}VA^{-1} \\
&= I_n + U \times \mathbf{0} \times (W - VA^{-1}U)^{-1}VA^{-1} = I_n
\end{aligned}
\tag{10.5}
$$

■

2. *Solution.* (a) $A^2 - B^2$ is symmetric. The reason is that

$$(A^2 - B^2)^{\mathrm{T}} = (AA)^{\mathrm{T}} - (BB)^{\mathrm{T}} = A^{\mathrm{T}}A^{\mathrm{T}} - B^{\mathrm{T}}B^{\mathrm{T}} = AA - BB = A^2 - B^2.$$

(b) $(A + B)(A - B)$ may not be symmetric. Let me raise a counterexample to explain it:

Suppose $A = \begin{bmatrix} 1 & 7 \\ 7 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 2 & 5 \\ 5 & 1 \end{bmatrix}$. Then $A + B = \begin{bmatrix} 3 & 12 \\ 12 & 1 \end{bmatrix}$, $A - B = \begin{bmatrix} -1 & 2 \\ 2 & -1 \end{bmatrix}$. The product $(A + B)(A - B)$ is given by:

$$(A + B)(A - B) = \begin{bmatrix} 21 & -6 \\ -10 & 23 \end{bmatrix}$$

which is obviously *not symmetric.*

(c) ABA is symmetric. The reason is that

$$(ABA)^{\mathrm{T}} = A^{\mathrm{T}} B^{\mathrm{T}} A^{\mathrm{T}} = ABA$$

(d) $ABAB$ may not be symmetric, let me raise a counterexample to explain it:

Suppose $A = \begin{bmatrix} 1 & 7 \\ 7 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 2 & 5 \\ 5 & 1 \end{bmatrix}$. Then the product $ABAB$ is given by:

$$ABAB = \begin{bmatrix} 1537 & 864 \\ 1008 & 1393 \end{bmatrix}$$

which is obviously *not symmetric.*

■

3. *Solution.* Starting from $A = LDU$, then $A = L(U^{\mathrm{T}})^{-1} \times (U^{\mathrm{T}} DU)$.

- $L(U^{\mathrm{T}})^{-1}$ is lower triangular with unit diagonals.

  Reason: U is upper triangular, hence U^{T} is lower triangular, its inverse $(U^{\mathrm{T}})^{-1}$ is also lower triangular. And L is also lower triangular. Hence the product $L(U^{\mathrm{T}})^{-1}$ remains lower triangular. Since L and U has unit diagonals, their transformation $L(U^{\mathrm{T}})^{-1}$ also has unit diagonals.

- $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{U}$ is symmetric. The reason is that

$$(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{U})^{\mathrm{T}} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{D}^{\mathrm{T}}(\boldsymbol{U}^{\mathrm{T}})^{\mathrm{T}} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{U}$$

In conclusion, here lists a new factorization of $\boldsymbol{A}$ into *triangular* times *symmetric*.

■

4. *Solution.* (a)

$$\boldsymbol{A}\boldsymbol{X}+\boldsymbol{B}=\boldsymbol{C} \implies \boldsymbol{A}\boldsymbol{X}=\boldsymbol{C}-\boldsymbol{B} \implies \boldsymbol{X}=\boldsymbol{A}^{-1}(\boldsymbol{C}-\boldsymbol{B}).$$

Since $\boldsymbol{A} = \begin{bmatrix} 5 & 3 \\ 3 & 2 \end{bmatrix}$, we obtain $\boldsymbol{A}^{1} = \frac{1}{10-9}\begin{bmatrix} 2 & -3 \\ -3 & 5 \end{bmatrix} = \begin{bmatrix} 2 & -3 \\ -3 & 5 \end{bmatrix}$.

$$\implies \boldsymbol{X}=\boldsymbol{A}^{-1}(\boldsymbol{C}-\boldsymbol{B}) = \begin{bmatrix} 2 & -3 \\ -3 & 5 \end{bmatrix}\begin{bmatrix} 4-6 & -2-2 \\ -6-2 & 3-4 \end{bmatrix} = \begin{bmatrix} 20 & -5 \\ -34 & 7 \end{bmatrix}.$$

(b)

$$\boldsymbol{X}\boldsymbol{A}+\boldsymbol{B}=\boldsymbol{C} \implies \boldsymbol{X}\boldsymbol{A}=\boldsymbol{C}-\boldsymbol{B} \implies \boldsymbol{X}=(\boldsymbol{C}-\boldsymbol{B})\boldsymbol{A}^{-1}.$$

Hence the solution is given by

$$\boldsymbol{X}=(\boldsymbol{C}-\boldsymbol{B})\boldsymbol{A}^{-1} = \begin{bmatrix} -2 & -4 \\ -8 & -1 \end{bmatrix}\begin{bmatrix} 2 & -3 \\ -3 & 5 \end{bmatrix} = \begin{bmatrix} 8 & -14 \\ -13 & 19 \end{bmatrix}.$$

(c)

$$\boldsymbol{A}\boldsymbol{X}+\boldsymbol{B}=\boldsymbol{X} \implies (\boldsymbol{A}-\boldsymbol{I})\boldsymbol{X}=-\boldsymbol{B} \implies \boldsymbol{X}=-(\boldsymbol{A}-\boldsymbol{I})^{-1}\boldsymbol{B}$$

Hence the soluion is given by

$$\boldsymbol{X}=-(\boldsymbol{A}-\boldsymbol{I})^{-1}\boldsymbol{B}=-\begin{bmatrix} 5-1 & 3 \\ 3 & 2-1 \end{bmatrix}^{-1}\begin{bmatrix} 6 & 2 \\ 2 & 4 \end{bmatrix} = -\frac{1}{4-9}\begin{bmatrix} 1 & -3 \\ -3 & 4 \end{bmatrix}\begin{bmatrix} 6 & 2 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 0 & -2 \\ -2 & 2 \end{bmatrix}.$$

(d)

$$\boldsymbol{X}\boldsymbol{A}+\boldsymbol{C}=\boldsymbol{X} \implies \boldsymbol{X}(\boldsymbol{A}-\boldsymbol{I})=-\boldsymbol{C} \implies \boldsymbol{X}=-\boldsymbol{C}(\boldsymbol{A}-\boldsymbol{I})^{-1}$$

Hence the solution is given by

$$X = -C(A - I)^{-1} = -\begin{bmatrix} 4 & -2 \\ -6 & 3 \end{bmatrix} \begin{bmatrix} -0.2 & 0.6 \\ 0.6 & -0.8 \end{bmatrix} = \begin{bmatrix} 2 & -4 \\ -3 & 6 \end{bmatrix}.$$

■

5. *Solution.* Firstly, we show $t_{jj} = u_{jj}r_{jj}$ for $j = 1, \ldots, n$:

$$\begin{aligned} t_{jj} &= \sum_{k=1}^{n} u_{jk}r_{kj} \\ &= \sum_{k=1, j<k} u_{jk}r_{kj} + u_{jj}r_{jj} + \sum_{k=1, j>k} u_{jk}r_{kj} \\ &= \sum_{k=1, j<k} u_{jk} \times 0 + u_{jj}r_{jj} + \sum_{k=1, j>k} 0 \times r_{kj} \\ &= u_{jj}r_{jj} \end{aligned}$$

Secondly, we show that $t_{ij} = 0$ if $i > j$ for $i, j \in \{1, 2, \ldots, n\}$:

$$\begin{aligned} t_{ij} &= \sum_{k=1}^{n} u_{ik}r_{kj} \\ &= \sum_{k=1, k<i}^{n} u_{ik}r_{kj} + u_{ii}r_{ij} + \sum_{k=1, k>i}^{n} u_{ik}r_{kj} \\ &= \sum_{k=1, k<i}^{n} 0 \times r_{kj} + u_{ii} \times 0 + \sum_{k=1, k>i}^{n} u_{ik} \times 0 \\ &= 0 \end{aligned}$$

Hence $t_{ij} = 0$ for $i < j$. Hence T is upper triangular. ■

6. *Solution.* (a)

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

(b)

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

It tells us that there are 2 walks of length 2 that from v_1 to v_1; 1 walk of length 2 that from v_1 to v_2; 1 walk of length 2 that from v_1 to v_3; 1 walk of length 2 that from v_1 to v_4; 1 walk of length 2 that from v_1 to v_5.

(c)

$$A^3 = \begin{bmatrix} 2 & 4 & 1 & 4 & 1 \\ 4 & 2 & 3 & 5 & 1 \\ 1 & 3 & 0 & 1 & 1 \\ 4 & 5 & 1 & 2 & 3 \\ 1 & 1 & 1 & 3 & 0 \end{bmatrix}$$

There are $a_{23} = 3$ walks of length 3 from v_2 to v_3. There are $1 + 1 + 5 = 7$ walks of length 3 from v_2 to v_4.

■

10.1.4. Solution to Assignment Four

1. *Solution.* (a)

$$\begin{bmatrix} 1 & 2 & 3 & 1 & -3 \\ 2 & 5 & 5 & 4 & 9 \\ 3 & 7 & 8 & 5 & 6 \end{bmatrix} \xrightarrow[\text{Add } (-3)\times\text{Row 1 to Row 3}]{\text{Add } (-2)\times\text{Row 1 to Row 2}} \begin{bmatrix} 1 & 2 & 3 & 1 & -3 \\ 0 & 1 & -1 & 2 & 15 \\ 0 & 1 & -1 & 2 & 15 \end{bmatrix} \xrightarrow{\text{Add } (-1)\times\text{Row 2 to Row}}$$

$$\begin{bmatrix} 1 & 2 & 3 & 1 & -3 \\ 0 & 1 & -1 & 2 & 15 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow{\text{Add } (-2)\times\text{Row 2 to Row 1}} \begin{bmatrix} 1 & 0 & 5 & -3 & -33 \\ 0 & 1 & -1 & 2 & 15 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \text{(rref)}$$

(b) We write $\boldsymbol{Ax} = \boldsymbol{b}$ in argumented matrix form:

$$\left[\begin{array}{ccccc|c} 1 & 2 & 3 & 1 & -3 & 1 \\ 2 & 5 & 5 & 4 & 9 & 1 \\ 3 & 7 & 8 & 5 & 6 & 2 \end{array}\right]$$

We convert $\boldsymbol{A}$ into $\boldsymbol{U}$(rref):

$$\left[\begin{array}{ccccc|c} 1 & 0 & 5 & -3 & -33 & 3 \\ 0 & 1 & -1 & 2 & 15 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Hence we only need to solve

$$\begin{cases} x_1 + 5x_3 - 3x_4 - 33x_5 = 3 \\ x_2 - x_3 + 2x_4 + 15x_5 = -1 \end{cases} \implies \begin{cases} x_1 = 3 - 5x_3 + 3x_4 + 33x_5 \\ x_2 = -1 + x_3 - 2x_4 - 15x_5 \end{cases}$$

Hence all solutions is given by

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 3 - 5x_3 + 3x_4 + 33x_5 \\ -1 + x_3 - 2x_4 - 15x_5 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} -5 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 3 \\ -2 \\ 0 \\ 1 \\ 0 \end{pmatrix} + x_5 \begin{pmatrix} 33 \\ -15 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

where x_3, x_4, x_5 can be taken arbitrarily.

(c) We write $\boldsymbol{Ax} = \boldsymbol{b}$ in argumented matrix form:

$$\left[\begin{array}{ccccc|c} 1 & 2 & 3 & 1 & -3 & b_1 \\ 2 & 5 & 5 & 4 & 9 & b_2 \\ 3 & 7 & 8 & 5 & 6 & b_3 \end{array}\right]$$

We convert $\boldsymbol{A}$ into $\boldsymbol{U}$(rref):

$$\left[\begin{array}{ccccc|c} 1 & 0 & 5 & -3 & -33 & 4b_1 - b_2 \\ 0 & 1 & -1 & 2 & 15 & -2b_1 + b_2 \\ 0 & 0 & 0 & 0 & 0 & -b_1 - b_2 + b_3 \end{array}\right]$$

- When $-b_1 - b_2 + b_3 \neq 0$, there is **no solution.**

- When $-b_1 - b_2 + b_3 = 0$, we only need to solve

$$\begin{cases} x_1 + 5x_3 - 3x_4 - 33x_5 = 5b_1 - 2b_2 \\ x_2 - x_3 + 2x_4 + 15x_5 = -2b_1 + b_2 \end{cases} \implies \begin{cases} x_1 = 4b_1 - b_2 - 5x_3 + 3x_4 + 33x_5 \\ x_2 = -2b_1 + b_2 + x_3 - 2x_4 - 15x_5 \end{cases}$$

Hence all solutions is given by

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 4b_1 - b_2 - 5x_3 + 3x_4 + 33x_5 \\ -2b_1 + b_2 + x_3 - 2x_4 - 15x_5 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 4b_1 - b_2 \\ -2b_1 + b_2 \\ 0 \\ 0 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} -5 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + x_4$$

■

2. *Proof.* (a) We set $v_1 = \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}, v_2 = \begin{pmatrix} 2 \\ -2 \\ 4 \end{pmatrix}, v_3 = \begin{pmatrix} -3 \\ 3 \\ 6 \end{pmatrix}$. Then we claim that $\dim(\text{span}\{v_1, v_2, v_3\}) = 3$. Hence we only need to show that v_1, v_2, v_3 forms the basis for $\text{span}\{v_1, v_2, v_3\}$. Hence we only need to show they are ind. Thus we only need to show $\boldsymbol{Ax} = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} x = \boldsymbol{0}$ has unique solution. Thus we only need to show $\boldsymbol{A} = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix}$ is invertible:

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & -3 \\ -2 & -2 & 3 \\ 2 & 4 & 6 \end{bmatrix} \xrightarrow[\text{Add } (-2)\times\text{Row 1 to Row 3}]{\text{Add } 2\times\text{Row 1 to Row 2}} \begin{bmatrix} 1 & 2 & -3 \\ 0 & 2 & -3 \\ 0 & 0 & 12 \end{bmatrix} \xrightarrow[\text{Row } 3\times\frac{1}{12}]{\text{Row } 2\times\frac{1}{2}} \begin{bmatrix} 1 & 2 & -3 \\ 0 & 1 & -\frac{3}{2} \\ 0 & 0 & 1 \end{bmatrix} \text{(rre}$$

Hence $\text{rank}(\boldsymbol{A}) = 3$. Thus $\boldsymbol{A}$ is full rank, which means $\boldsymbol{A}$ is invertible.

(b) We do elimination to convert $\boldsymbol{A}$ into its rref form:

$$\begin{bmatrix} 1 & -2 & 3 & 2 \\ -1 & 2 & -2 & -1 \\ 2 & -4 & 5 & 3 \end{bmatrix} \xrightarrow[\text{Add } (-2)\times\text{Row 1 to Row 3}]{\text{Add } 1\times\text{Row 1 to Row 2}} \begin{bmatrix} 1 & -2 & 3 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & -1 & -1 \end{bmatrix}$$

$$\xrightarrow[\text{Add } (-3)\times\text{Row 2 to Row 3}]{\text{Add } 1\times\text{Row 2 to Row 3}} \begin{bmatrix} 1 & -2 & 0 & -1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \text{(rref)}$$

Hence $\text{rank}(\boldsymbol{A}) = \dim(\text{col}(\boldsymbol{A})) = 2$. Hence dimension of $\text{col}(\boldsymbol{A})$ is 2.

(c) We convert $\boldsymbol{B}$ into rref:

$$\boldsymbol{B} = \begin{bmatrix} 1 & 3 & 2 \\ 2 & 1 & 4 \\ 4 & 7 & 8 \end{bmatrix} \implies \boldsymbol{R} = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \text{(rref)}$$

Thus we only need to compute the solution to $\boldsymbol{Ux} = \boldsymbol{0}$.

If $x_3 = 1$, then $x_1 = -2, x_2 = 0$.

Hence the basis for $N(\boldsymbol{R})$ is $\begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$. Hence $\dim(N(\boldsymbol{B})) = \dim(N(\boldsymbol{R})) = 1$.

(d) The linear combination of $(x-2)(x+2), x^2(x^4-2), x^6-8$ is given by:

$$m_1(x-2)(x+2) + m_2x^2(x^4-2) + m_3(x^6-8) = (m_2+m_3)x^6 + (m_1-2m_2)x^2 + (-4m_1-8m_3)$$

where $m_1, m_2, m_3 \in \mathbb{R}$.

- Firstly we show $\{x^4-4, x^6-8\}$ span the space $\text{span}\{(x-2)(x+2), x^2(x^4-2), x^6-8\}$:

  Given any vector

  $$(m_2+m_3)x^6 + (m_1-2m_2)x^2 + (-4m_1-8m_3) \in \text{span}\{(x-2)(x+2), x^2(x^4-2), x^6-8\}$$

  for $\forall m_1, m_2, m_3 \in \mathbb{R}$,

  we construct $a_1 = m_2 + m_3, a_2 = m_1 - 2m_2$. Then the linear combination of $x^6 - 8$ and $x^4 - 4$ with coefficient a_1, a_2 is exactly

  $$a_2(x^4-4) + a_1(x^6-8) = (m_2+m_3)x^6 + (m_1-2m_2)x^2 + (-4m_1-8m_3)$$

  Hence

  $$(m_2+m_3)x^6 + (m_1-2m_2)x^2 + (-4m_1-8m_3) \in \text{span}\{x^4-4, x^6-8\}$$

$$\implies \text{span}\{(x-2)(x+2), x^2(x^4-2), x^6-8\} \subset \text{span}\{x^4-4, x^6-8\}$$

Conversely, by setting $m_1 = 2a_1 + a_2, m_2 = a_1, m_3 = 0$ we can show $\text{span}\{x^4-4, x^6-8\} \subset \text{span}\{(x-2)(x+2), x^2(x^4-2), x^6-8\}$.

Hence $\text{span}\{x^4-4, x^6-8\} = \text{span}\{(x-2)(x+2), x^2(x^4-2), x^6-8\}$

Then we show x^4-4, x^6-8 are ind.:

$$\text{Given } a_1(x^4-4) + a_2(x^6-8) = 0 \implies a_2x^6 + a_1x^4 + (-4a_1 - 8a_2) = 0$$

$$\implies \begin{cases} a_2 = 0 \\ a_1 = 0 \\ -4a_1 - 8a_2 = 0 \end{cases} \implies \begin{cases} a_1 = 0 \\ a_2 = 0 \end{cases}$$

Hence x^4-4, x^6-8 are ind. They form the basis for the space $\text{span}\{(x-2)(x+2), x^2(x^4-2), x^6-8\}$.

Hence $\dim(\text{span}\{(x-2)(x+2), x^2(x^4-2), x^6-8\}) = 2$.

(e) Firstly, it's easy to verify that 5 and $\cos^2 x$ are ind.

Next, let's show $\text{span}\{5, \cos^2 x\} = \text{span}\{5, \cos 2x, \cos^2 x\}$:

Any linear combination of $\{5, \cos 2x, \cos^2 x\}$ is given by:

$$5m_1 + m_2 \cos 2x + m_3 \cos^2 x = (2m_2 + m_3)\cos^2 x + (5m_1 - m_2)$$

where $m_1, m_2, m_3 \in \mathbb{R}$.

Any linear combination of $\{5, \cos^2 x\}$ is given by:

$$5n_1 + n_2 \cos^2 x$$

where $n_1, n_2 \in \mathbb{R}$.

- if we construct $n_1 = m_1 - \frac{1}{5}m_2, n_2 = 2m_2 + m_3$, then it means any linear combination of $\{5, \cos 2x, \cos^2 x\}$ can be expressed in form of $\{5, \cos^2 x\}$. Hence $\text{span}\{5, \cos 2x, \cos^2 x\} \subset \{5, \cos^2 x\}$.
- if wr construct $m_1 = n_1 + \frac{1}{10}n_2, m_2 = \frac{1}{2}n_2, m_3 = 0$, then it means any linear

combination of $\{5, \cos^2 x\}$ can be expressed in form of $\{5, \cos 2x, \cos^2 x\}$.

Hence $\text{span}\{5, \cos^2 x\} \subset \{5, \cos 2x, \cos^2 x\}$.

Hence $\text{span}\{5, \cos^2 x\} = \{5, \cos 2x, \cos^2 x\}$. $\{5, \cos^2 x\}$ is the basis for $\text{span}\{5, \cos 2x, \cos^2 x\}$. Hence $\dim(\text{span}\{5, \cos 2x, \cos^2 x\}) = 2$.

■

3. *Solution.* (a) It can have **no** or **infinitely many** solutions.

   Since $r < m$ and $r < n$, matrix $\boldsymbol{A}$ is not full rank. When reducing $\boldsymbol{A}$ into rref, there must exist row that contains all zero entries. For its augmented matrix which is rref, when the right hand side is zero for the zero row in the left, it has **infinitely many** solutions; when the right hand side is nonzero for the zero rwo in the left, it has **no** solutions.

   (b) It has **infinitely many** solutions.

   Since $r = m$ and $r < n$, $\boldsymbol{A}$ is full rank. Hence $\boldsymbol{Ax} = \boldsymbol{b}$ has at least one solutions. Since $\dim(N(\boldsymbol{A})) = n - r > 0$, there exists **infinitely many** solutions for $\boldsymbol{Ax} = \boldsymbol{0}$. Sicne $\boldsymbol{x}_{\text{complete}} = \boldsymbol{x}_p + \boldsymbol{x}_{\text{special}}$, $\boldsymbol{Ax} = \boldsymbol{b}$ has **infinitely many** solutions.

   (c) It has **no** or **unique** solution.

   Since $r < m$ and $r = n$, the rref of $\boldsymbol{A}$ must be of the form $\boldsymbol{R} = \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{0} \end{bmatrix}$. If $\boldsymbol{d}$ has nonzero entries for the zero rows in the left side equation, then $\boldsymbol{Rx} = \boldsymbol{d}$(And the orignal $\boldsymbol{Ax} = \boldsymbol{b}$) has no solution. If $\boldsymbol{d}$ has all zero entries for the zero rows in the left side equation, then $\boldsymbol{Rx} = \boldsymbol{d}$(And the orignal $\boldsymbol{Ax} = \boldsymbol{b}$) has unique solution.

■

4. *Proof.* (a) For any given ind. vectors $v_1, v_2, \dots, v_n$, suppose v is the any vector in $\boldsymbol{V}$.

   - Let's show $v_1, v_2, \dots, v_n, v$ must be dep:

     It suffices to show $c_1 v_1 + \cdots + c_n v_n + c_{n+1} v = \boldsymbol{0}$ has nontrival solution for $c_1,\dots,c_{n+1} \in \mathbb{R}$.

$$\iff \boldsymbol{Ax} = \mathbf{0} \text{ has nontrival solution, where } \boldsymbol{A} = \left[\begin{array}{c|c|c|c} v_1 & \dots & v_n & v \end{array}\right]$$

which is obviously true since $\boldsymbol{A}$ is a $n \times n+1$ matrix $(n < n+1)$

- Hence there exists $(c_1,c_2,\dots,c_{n+1}) \neq (0,0,\dots,0)$ such that

$$c_1v_1 + \dots + c_nv_n + c_{n+1}v = \mathbf{0}$$

If $c_{n+1} = 0$, then we have $(c_1,c_2,\dots,c_n) \neq (0,0,\dots,0)$ such that

$$c_1v_1 + \dots + c_nv_n = \mathbf{0},$$

which contradicts that $v_1,\dots,v_n$ are ind.

Hence $c_{n+1} \neq 0$. Then any $v \in \boldsymbol{V}$ could be expressed as:

$$v = -\frac{c_1}{c_{n+1}}v_1 - \frac{c_2}{c_{n+1}}v_2 - \dots - \frac{c_n}{c_{n+1}}v_n$$

which means $v_1,v_2,\dots,v_n$ spans $\boldsymbol{V}$. And they are ind.

So they form a basis for $\boldsymbol{V}$.

(b) Suppose $v_1 \dots,v_n$ spans $\boldsymbol{V}$. We assume that they are dep. Hence there exists $(c_1,c_2,\dots,c_n) \neq (0,0,\dots,0)$ such that

$$c_1v_1 + c_2v_2 + \dots + c_nv_n = \mathbf{0}$$

WLOG, we set $c_n \neq 0$. Hence we could express v_n as:

$$v_n = -\frac{c_1}{c_n}v_1 - \frac{c_2}{c_n}v_2 - \dots - \frac{c_{n-1}}{c_n}v_{n-1}$$

- We claim that $v_1,v_2,\dots,v_{n-1}$ still spans $\boldsymbol{V}$:

For any vector $v \in \boldsymbol{V}$, since $v_1,\dots,v_n$ spans $\boldsymbol{V}$, v could be expressed in form of $v_1, \ldots, v_n$:

$$v = m_1 v_1 + \cdots + m_n v_n$$

where $m_1, \ldots, m_n \in \mathbb{R}$.

Hence it could also be expressed in form of $v_1, \ldots, v_{n-1}$:

$$\begin{aligned} v &= m_1 v_1 + \cdots + m_n(-\frac{c_1}{c_n} v_1 - \frac{c_2}{c_n} v_2 - \cdots - \frac{c_{n-1}}{c_n} v_{n-1}) \\ &= (m_1 - \frac{m_n c_1}{c_n}) v_1 + (m_2 - \frac{m_n c_2}{c_n}) v_2 - \cdots - (m_{n-1} - \frac{m_n c_{n-1}}{c_n}) v_{n-1} \end{aligned}$$

Hence $v_1 . v_2, \ldots, v_{n-1}$ still spans $\boldsymbol{V}$.

- If $v-_1, v_2, \ldots, v_n$ still dep, we continue eliminating vectors until we get ind. vectors, say, $v_1, v_2, \ldots, v_k$. Hence $\dim(\boldsymbol{V}) = k < n$. which contradicts $\dim(\boldsymbol{V}) = n$.

■

5. *Proof.* (a) Suppose $u_1 + v_1$ is one vector in $\boldsymbol{U} + \boldsymbol{V}$ s.t. $u_1 \in \boldsymbol{U}, v_1 \in \boldsymbol{V}$; $u_2 + v_2$ is one vector in $\boldsymbol{U} + \boldsymbol{V}$ s.t. $u_2 \in \boldsymbol{U}, v_2 \in \boldsymbol{V}$.

Hence we claim addition and scalar multiplication is still closed under $\boldsymbol{U} + \boldsymbol{V}$:

$$(u_1 + v_1) + (u_2 + v_2) = (u_1 + u_2) + (v_1 + v_2) \qquad c(u_1 + v_1) = cu_1 + cv_1$$

where c is a scalar.

- Since $u_1, u_2 \in \boldsymbol{U}$, $u_1 + u_2 \in \boldsymbol{U}$. Similarly, $v_1 + v_2 \in \boldsymbol{V}$.
  Hence $(u_1 + u_2) + (v_1 + v_2) = (u_1 + v_1) + (u_2 + v_2) \in \boldsymbol{U} + \boldsymbol{V}$.

- Since $u_1 \in \boldsymbol{U}$, $cu_1 \in \boldsymbol{U}$. Similarly, $cv_1 \in \boldsymbol{U}$.
  Hence $cu_1 + cv_1 = c(u_1 + v_1) \in \boldsymbol{U} + \boldsymbol{V}$

Hence addition and scalar multiplication is still closed under $\boldsymbol{U} + \boldsymbol{V}$. Hence $\boldsymbol{U} + \boldsymbol{V}$ is still a subspace of W.

(b) If $w_1, w_2 \in \boldsymbol{U} \cap \boldsymbol{V}$, then $w_1, w_2 \in \boldsymbol{U}$ and $w_1, w_2 \in \boldsymbol{V}$. Thus the linear combintation of w_1, w_2 is still in $\boldsymbol{U}$ and $\boldsymbol{V}$:

$$a_1 w_1 + a_2 w_2 \in \boldsymbol{U} \qquad a_1 w_1 + a_2 w_2 \in \boldsymbol{V}$$

where a_1, a_2 is a scalar.

Hence $a_1 w_1 + a_2 w_2 \in \boldsymbol{U} \cap \boldsymbol{V}$. Hence $\boldsymbol{U} \cap \boldsymbol{V}$ is also a subspace of $\boldsymbol{W}$.

(c) $\dim(\boldsymbol{U}) = 2$. The set $\{\boldsymbol{e}_1, \boldsymbol{e}_2\}$ is a basis for $\boldsymbol{U}$.

$\dim(\boldsymbol{V}) = 2$. The set $\{\boldsymbol{e}_2, \boldsymbol{e}_3\}$ is a basis for $\boldsymbol{V}$.

$\dim(\boldsymbol{U} \cap \boldsymbol{V}) = 1$. The set $\{\boldsymbol{e}_2\}$ is a basis for $\boldsymbol{U} \cap \boldsymbol{V}$.

$\dim(\boldsymbol{U} + \boldsymbol{V}) = 3$. The set $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ is a basis for $\boldsymbol{U} + \boldsymbol{V}$.

(d) Let $\boldsymbol{U}$ and $\boldsymbol{V}$ be subspaces of $\mathbb{R}^n$ such that $\boldsymbol{U} \cap \boldsymbol{V} = \{\boldsymbol{0}\}$.

If either $\boldsymbol{U} = \{0\}$ or $\boldsymbol{U} = \{0\}$ the result is obvious.

Assume that both subspaces are nontrivial with $\dim(\boldsymbol{U}) = m > 0$ and $\dim(\boldsymbol{V}) = n > 0$.

Let $\{u_1, ..., u_m\}$ be a basis for $\boldsymbol{U}$ and let $\{v_1, ..., v_n\}$ be a basis for $\boldsymbol{V}$. These vectors $u_1, u_2, \ldots, u_m, v_1, v_2, \ldots, v_n$ spans $\boldsymbol{U} + \boldsymbol{V}$.

- We claim that these vectors form a basis for $\boldsymbol{U} + \boldsymbol{V}$. It suffices to show they are ind:

  If we have the condition

  $$c_1 u_1 + c_2 u_2 + \cdots + c_m u_m + c_{m+1} v_1 + \cdots + c_{m+n} v_n = \boldsymbol{0}$$

  where $c_1, \ldots, c_{m+n}$ are scalars,

  if we set $\boldsymbol{u} = c_1 u_1 + c_2 u_2 + \cdots + c_m u_m$ and $\boldsymbol{v} = c_{m+1} v_1 + \cdots + c_{m+n} v_n$, then we have

  $$\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{0}$$

  Hence $\boldsymbol{u} = -\boldsymbol{v}$. Then $\boldsymbol{u}, \boldsymbol{v} \in \boldsymbol{U}$ and $\boldsymbol{u}, \boldsymbol{v} \in \boldsymbol{V}$. Hence $\boldsymbol{u}, \boldsymbol{v} \in \boldsymbol{U} \cap \boldsymbol{V}$.

Hence $\boldsymbol{u}, \boldsymbol{v} = \boldsymbol{0}$ since $\boldsymbol{U} \cap \boldsymbol{V} = \{\boldsymbol{0}\}$. Thus we have

$$\begin{aligned} c_1 u_1 + c_2 u_2 + \cdots + c_m u_m &= \boldsymbol{0} \\ c_{m+1} v_1 + c_{m+2} v_2 + \cdots + c_{m+n} v_n &= \boldsymbol{0} \end{aligned}$$

By the independence of $u_1, \dots, u_m$ and the independence of $v_1, \dots, v_n$ it follows that

$$c_1 = c_2 = \cdots = c_{m+n} = 0$$

- Thus $\{u_1, u_2, \dots, u_m, v_1, v_2, \dots, v_n\}$ form a basis for $\boldsymbol{U} + \boldsymbol{V}$.

Hence $\dim(\boldsymbol{U} + \boldsymbol{V}) = m + n$.

■

6. *Proof.* For any vector $\boldsymbol{y} \in \text{range}(\boldsymbol{A} + \boldsymbol{B})$, there exists vector $\boldsymbol{x}$ such that

$$(\boldsymbol{A} + \boldsymbol{B})\boldsymbol{x} = \boldsymbol{y}$$

Also, we can express $\boldsymbol{y}$ as sum of vectors in range of $\boldsymbol{A}$ and $\boldsymbol{B}$:

$$\boldsymbol{y} = (\boldsymbol{A} + \boldsymbol{B})\boldsymbol{x} \quad = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{x}$$

Hence we obtain

$$\text{range}(\boldsymbol{A} + \boldsymbol{B}) \subset \text{range}(\boldsymbol{A}) + \text{range}(\boldsymbol{B})$$

Assume one basis for $\text{range}(\boldsymbol{A})$ is $\{a_1, \dots, a_s\}$; $\boldsymbol{B} = \begin{bmatrix} B_1 & \big| & \dots & \big| & B_n \end{bmatrix}$ one basis for $\text{range}(\boldsymbol{B})$ is $\{b_1, \dots, b_t\}$. Thus we obtain:

$$\begin{aligned} \dim(\text{range}(\boldsymbol{A}) + \text{range}(\boldsymbol{B})) &= \dim(a_1, \dots, a_s, b_1, \dots, b_t) \\ &\le s + t \\ &= \dim(\text{range}(\boldsymbol{A})) + \dim(\text{range}(\boldsymbol{B})) \\ &= \text{rank}(\boldsymbol{A}) + \text{rank}(\boldsymbol{B}) \end{aligned}$$

Hence we have

$$
\begin{aligned}
\operatorname{rank}(\boldsymbol{A}+\boldsymbol{B}) &= \dim(\operatorname{range}(\boldsymbol{A}+\boldsymbol{B})) \\
&\leq \dim(\operatorname{range}(\boldsymbol{A})+\operatorname{range}(\boldsymbol{B})) \\
&\leq \operatorname{rank}(\boldsymbol{A})+\operatorname{rank}(\boldsymbol{B})
\end{aligned}
$$

■

7. *Proof.* (a) We assume $\boldsymbol{A} = \begin{bmatrix} A_1 & \Big| & \dots & \Big| & A_n \end{bmatrix}$, $\boldsymbol{B} = \begin{bmatrix} B_1 & \Big| & \dots & \Big| & B_n \end{bmatrix}^{\mathrm{T}}$.

   Hence $\boldsymbol{AB}$ could be expressed as:

$$
\boldsymbol{AB} = A_1 B_1 + \cdots + A_n B_n
$$

   which means every column of $\boldsymbol{AB}$ is a linear combination of columns of $\boldsymbol{A}$. Assume one basis for $\operatorname{col}(\boldsymbol{A})$ is $a_1, \dots, a_s$. Then $\{a_1, \dots, a_s\}$ can also span $\operatorname{col}(\boldsymbol{AB})$.

   Hence $\operatorname{rank}(\boldsymbol{AB}) = dim(\operatorname{col}(\boldsymbol{AB})) \leq \dim(\operatorname{col}(\boldsymbol{A})) = \operatorname{rank}(\boldsymbol{A})$

   (b) We use the conclusion of part(a) to derive this statement:

   If $\operatorname{rank}(\boldsymbol{B}) = n$, then $\boldsymbol{B}$ is invertible, $\boldsymbol{A} = \boldsymbol{ABB}^{-1}$.

   Since product $\boldsymbol{AB}$ is a $m \times n$ matrix, $\boldsymbol{B}^{-1}$ is a $n \times n$ matrix, by part(a), $\operatorname{rank}(\boldsymbol{ABB}^{-1}) \leq \operatorname{rank}(\boldsymbol{AB})$.

   In conclusion,

$$
\operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{ABB}^{-1}) \leq \operatorname{rank}(\boldsymbol{AB}) \leq \operatorname{rank}(\boldsymbol{A})
$$

   The equality must be satisfied, hence we have $\operatorname{rank}(\boldsymbol{AB}) = \operatorname{rank}(\boldsymbol{A})$.

■

8. *Proof.* We assume $\{v_1, \dots, v_{n-1}\}$ form a basis for $\mathbb{R}^n$.

   It is equivalent to $\boldsymbol{Ax} = \boldsymbol{b}$ must have a solution $\forall \boldsymbol{b} \in \mathbb{R}^n$ and $\boldsymbol{A} = \begin{bmatrix} v_1 & \Big| & \dots & \Big| & v_{n-1} \end{bmatrix}$.

   However, since $\boldsymbol{A}$ is $n \times (n-1)$ matrix, the number of equations is greater than number of unknowns, this system may not have a solution, which forms a contradiction! ■

10.1.5. Solution to Assignment Five

1. *Proof.* (a) For square matrix $\boldsymbol{A}$, there exists identity matrix $\boldsymbol{I}$, such that $\boldsymbol{A} = \boldsymbol{I}^{-1}\boldsymbol{A}\boldsymbol{I}$. Hence $\boldsymbol{A}$ is *similar* to itself.

   (b) If $\boldsymbol{B}$ is similar to $\boldsymbol{A}$, then there exists invertible matrix $\boldsymbol{S}_1$ such that $\boldsymbol{B} = \boldsymbol{S}_1^{-1}\boldsymbol{A}\boldsymbol{S}_1$. Hence we obtain:

$$\boldsymbol{S}_1\boldsymbol{B} = \boldsymbol{A}\boldsymbol{S}_1 \implies \boldsymbol{A} = \boldsymbol{S}_1\boldsymbol{B}\boldsymbol{S}_1^{-1}$$

   If we set $\boldsymbol{S}_2 = \boldsymbol{S}_1^{-1}$, then we have

$$\boldsymbol{A} = \boldsymbol{S}_2^{-1}\boldsymbol{B}\boldsymbol{S}_2$$

   Thus $\boldsymbol{A}$ is **simialr** to $\boldsymbol{B}$.

   (c) Since $\boldsymbol{A}$ is similar to $\boldsymbol{B}$, $\boldsymbol{B}$ is similar to $\boldsymbol{C}$, there exists invertible matrices $\boldsymbol{S}_1, \boldsymbol{S}_2$ such that

$$\boldsymbol{A} = \boldsymbol{S}_1^{-1}\boldsymbol{B}\boldsymbol{S}_1 \quad \text{and} \quad \boldsymbol{B} = \boldsymbol{S}_2^{-1}\boldsymbol{C}\boldsymbol{S}_2$$

   It follows that

$$\begin{aligned}\boldsymbol{A} &= \boldsymbol{S}_1^{-1}(\boldsymbol{S}_2^{-1}\boldsymbol{C}\boldsymbol{S}_2)\boldsymbol{S}_1 \\ &= (\boldsymbol{S}_1^{-1}\boldsymbol{S}_2^{-1})\boldsymbol{C}(\boldsymbol{S}_2\boldsymbol{S}_1) \\ &= (\boldsymbol{S}_2\boldsymbol{S}_1)^{-1}\boldsymbol{C}(\boldsymbol{S}_2\boldsymbol{S}_1)\end{aligned}$$

   If we set $\boldsymbol{S}_3 = \boldsymbol{S}_2\boldsymbol{S}_1$, since $\boldsymbol{S}_1, \boldsymbol{S}_2$ are invertible, then $\boldsymbol{S}_3$ is invertible. Hence $\boldsymbol{A} = \boldsymbol{S}_3^{-1}\boldsymbol{C}\boldsymbol{S}_3$. Thus $\boldsymbol{A}$ is **simialr** to $\boldsymbol{C}$.

■

2. *Solution.* Obviously, L is a linear operator defined by $L(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$, where

$$\boldsymbol{A} = \begin{pmatrix} 3 & 0 \\ 1 & -1 \end{pmatrix}$$

   We set $\boldsymbol{S} = \begin{bmatrix} b_1 & b_2 \end{bmatrix}$, where b_1, b_2 are the ordered vector in basis $\boldsymbol{B}$.
   We use similartiy transformation to compute the matrix representation $\boldsymbol{D}$ with

respect to basis B:

$$\begin{aligned}\boldsymbol{D} &= \boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S} \\ &= \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}^{-1} \begin{pmatrix} 3 & 0 \\ 1 & -1 \end{pmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} -11 & -20 \\ 7 & 13 \end{bmatrix}\end{aligned}$$

■

3. *Solution.* (a) No, since the zero function $f(x) \equiv 0$ does not belong to this set.

(b) No, since the zero function $f(x) \equiv 0$ does not belong to this set.

(c) Yes.

- Firstly this set belongs to $\mathbb{R}[x]$.

- Secondly, given zero function $f(x) \equiv 0$, for any $x \in \mathbb{R}$, we have $f(x) = 0 = f(1-x)$. Hence this set contains zero function $f(x) \equiv 0$.

- Thirdly, given two function f, g in this set, we have

$$f(x) = f(1-x) \quad \text{and} \quad g(x) = g(1-x) \text{ for all } x \in \mathbb{R}.$$

Then we set any linear combination of f and g to be $T = \alpha_1 f + \alpha_2 g$, where α_1, α_2 are scalars.
For any $x \in \mathbb{R}$, we have

$$\begin{aligned}T(x) &= \alpha_1 f(x) + \alpha_2 g(x) \\ &= \alpha_1 f(1-x) + \alpha_2 g(1-x) \\ &= T(1-x)\end{aligned}$$

Hence $T = \alpha_1 f + \alpha_2 g$ also belongs to this set.

In conclusion, this set is **subspace** of $\mathbb{R}[x]$.

■

4. *Proof.* (a) Given $f, g \in \boldsymbol{V}$, we have

$$\begin{aligned} T(\alpha_1 f + \alpha_2 g) &= \frac{\partial}{\partial x}(\alpha_1 f + \alpha_2 g) - \frac{\partial}{\partial y}(\alpha_1 f + \alpha_2 g) \\ &= \alpha_1 \frac{\partial f}{\partial x} + \alpha_2 \frac{\partial g}{\partial x} - \alpha_1 \frac{\partial f}{\partial y} - \alpha_2 \frac{\partial g}{\partial y} \\ &= \alpha_1 (\frac{\partial f}{\partial x} - \frac{\partial f}{\partial y}) + \alpha_2 (\frac{\partial f}{\partial x} - \frac{\partial f}{\partial y}) \\ &= \alpha_1 T(f) + \alpha_2 T(g) \end{aligned}$$

where α_1, α_2 are scalars. It immediately follows that T is a transformation.

(b) Given any $f = a + bx + cy + dx^2 + exy + fy^2 \in \boldsymbol{V}$, $f \in \ker T$ if and only if $\frac{\partial}{\partial x} f - \frac{\partial}{\partial y} f = 0$. Thus $f \in \ker T$ if and only if $b + 2dx + ey - (c + ex + 2fy) = 0$. Hence $f \in \ker T$ if and only if

$$\begin{aligned} b - c &= 0 \\ 2d - e &= 0 \\ e - 2f &= 0 \end{aligned}$$

The general solution is given by

$$\begin{pmatrix} a \\ b \\ c \\ d \\ e \\ f \end{pmatrix} = m_1 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + m_2 \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + m_3 \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 2 \\ 1 \end{pmatrix}$$

where $m_1, m_2, m_3 \in \mathbb{R}$.

Therefore, $f \in \ker T$ if and only if for any $m_1, m_2, m_3 \in \mathbb{R}$,

$$\begin{aligned} f &= m_1 + m_2 x + m_2 y + m_3 x^2 + 2m_3 xy + m_3 y^2 \\ &= m_1 \times 1 + m_2(x + y) + m_3(x^2 + 2xy + y^2) \end{aligned}$$

Obviously, the set $\{1, x+y, x^2+2xy+y^2\}$ is ind. and it spans $\ker T$ by the above argument. Hence $\{1, x+y, x^2+2xy+y^2\}$ is a basis for $\ker T$.

■

5. *Solution.*

$$D(e^x) = 1 \cdot e^x + 0 \cdot xe^x + 0 \cdot x^2e^x$$
$$D(xe^x) = 1 \cdot e^x + 1 \cdot xe^x + 0 \cdot x^2e^x$$
$$D(x^2e^x) = 0 \cdot e^x + 2 \cdot xe^x + 1 \cdot x^2e^x$$

Thus, the matrix representation of D with respect to $\{e^x, xe^x, x^2e^x\}$ is given by

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

■

6. *Solution.* (a) The transformed region will be a **parallelogram**.

In order to find the shape we only need to focus on the corner point $O(0,0), A(1,0), B(1,1), C(0,1)$. Suppose the matrix $\boldsymbol{A}$ is $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$. By matrix multiplication we find $OABC$ is transformed into $O_1A_1B_1C_1$ such that

$$O_1 = (0,0) \qquad A_1 = (a,c) \qquad B_1 = (a+c, b+d) \qquad C_1 = (b,d)$$

Since vector $\overrightarrow{O_1B_1} = \overrightarrow{O_1A_1} + \overrightarrow{O_1C_1}$, we find area $O_1A_1B_1C_1$ is a **parallelogram**.

(b) In order to get a square, we have to let the inner product of two adjacent sides of the parallelogram to be zero:

$$\overrightarrow{O_1A_1} \cdots \overrightarrow{O_1C_1} = ab + cd = 0.$$

And then we have to let all sides to have the same length:

$$|\overrightarrow{O_1A_1}|^2 = |\overrightarrow{O_1C_1}|^2 \implies a^2 + c^2 = b^2 + d^2$$

Finally we derive $b = \pm c, a = \mp d$. Hence when matrix $\boldsymbol{A}$ is of this form:

$$\boldsymbol{A} = b \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + d \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \text{ or } \boldsymbol{A} = b \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} + d \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

where $b, d \in \mathbb{R}$, it will transform the unit square into another square.

■

7. *Proof.* (a)
   - Firstly we show $\text{col}(\boldsymbol{A}\boldsymbol{A}^T) \subset \text{col}(\boldsymbol{A})$:

     For any $\boldsymbol{b} \in \text{col}(\boldsymbol{A}\boldsymbol{A}^T)$, there exists $\boldsymbol{x}_0$ such that $\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{x}_0 = \boldsymbol{b}$, which implies $\boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{x}_0) = \boldsymbol{b}$. Hence there exists vector $(\boldsymbol{A}^T\boldsymbol{x}_0)$ such that

     $$\boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{x}_0) = \boldsymbol{b}$$

     Hence $\boldsymbol{b} \in \text{col}(\boldsymbol{A})$. Hence $\text{col}(\boldsymbol{A}\boldsymbol{A}^T) \subset \text{col}(\boldsymbol{A})$.
   - In part *b* we will show $\text{rank}(\boldsymbol{A}\boldsymbol{A}^T) = \text{rank}(\boldsymbol{A})$. Hence $\dim(\text{col}(\boldsymbol{A}\boldsymbol{A}^T)) = \dim(\text{col}(\boldsymbol{A}))$.
   - We assume $\dim(\text{col}(\boldsymbol{A}\boldsymbol{A}^T)) = \dim(\text{col}(\boldsymbol{A})) = n$, the basis for $\text{col}(\boldsymbol{A})$ is $\{v_1, v_2, \dots, v_n\}$. Thus since $\text{col}(\boldsymbol{A}\boldsymbol{A}^T) \subset \text{col}(\boldsymbol{A})$, basis $\{v_1, v_2, \dots, v_n\}$ must span $\text{col}(\boldsymbol{A}\boldsymbol{A}^T)$. Since $\dim(\text{col}(\boldsymbol{A}\boldsymbol{A}^T)) = n$, $\{v_1, v_2, \dots, v_n\}$ must be the basis for $\text{col}(\boldsymbol{A}\boldsymbol{A}^T)$.
   - Since $\text{col}(\boldsymbol{A}\boldsymbol{A}^T)$ and $\text{col}(\boldsymbol{A})$ have the same basis, we obtain $\text{col}(\boldsymbol{A}\boldsymbol{A}^T) = \text{col}(\boldsymbol{A})$.

   (b)
   - Firstly, we show $N(\boldsymbol{A}) \subset N(\boldsymbol{A}^T\boldsymbol{A})$:

     For any $\boldsymbol{x}_0 \in N(\boldsymbol{A})$, we have $\boldsymbol{A}\boldsymbol{x}_0 = \boldsymbol{0}$. Thus by postmultiplying $\boldsymbol{A}^T$ we have $\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x}_0 = \boldsymbol{0}$. Hence $\boldsymbol{x}_0 \in N(\boldsymbol{A}^T\boldsymbol{A})$.
   - Then we show $N(\boldsymbol{A}^T\boldsymbol{A}) \subset N(\boldsymbol{A})$:

     For any $\boldsymbol{x}_0 \in N(\boldsymbol{A}^T\boldsymbol{A})$, we have $\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x}_0 = \boldsymbol{0}$. Thus by postmultiplying $\boldsymbol{x}_0^T$

we have $\boldsymbol{x}_0^{\mathrm{T}} \boldsymbol{A}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{x}_0 = \mathbf{0}$, which implies $\|\boldsymbol{A}\boldsymbol{x}_0\|^2 = \boldsymbol{x}_0^{\mathrm{T}} \boldsymbol{A}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{x}_0 = \mathbf{0}$. Hence $\boldsymbol{A}\boldsymbol{x}_0 = \mathbf{0}$. Hence $\boldsymbol{x}_0 \in N(\boldsymbol{A})$.

Hence we obtain $N(\boldsymbol{A}) \subset N(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})$ and $N(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) \subset N(\boldsymbol{A})$, which implies $N(\boldsymbol{A}) = N(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})$.

If we assume $\boldsymbol{A}$ is $m \times n$ matrix, then $\operatorname{rank}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) + \dim(N(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})) = n = \operatorname{rank}(\boldsymbol{A}) + \dim(N(\boldsymbol{A}))$.

- Since $\dim(N(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})) = \dim(N(\boldsymbol{A}))$, we obtain $\operatorname{rank}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{A})$.
- Similarly, we obtain $\operatorname{rank}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}) = \operatorname{rank}(\boldsymbol{A}^{\mathrm{T}})$ by changing $\boldsymbol{A}$ into $\boldsymbol{A}^{\mathrm{T}}$.
- Obviously, $\operatorname{rank}(\boldsymbol{A}^{\mathrm{T}}) = \dim(\operatorname{row}(\boldsymbol{A}^{\mathrm{T}})) = \dim(\operatorname{col}(\boldsymbol{A})) = \operatorname{rank}(\boldsymbol{A})$.

In conclusion, $\operatorname{rank}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}) = \operatorname{rank}(\boldsymbol{A}^{\mathrm{T}}) = \operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})$.

■

10.1.6. Solution to Assignment Six

1. *Solution.* One basis for $\mathbb{P}_2$ is $\{t^2, t, 1\}$. And we obtain:

$$\begin{aligned} T(t^2) &= (3t-2)^2 && = 9t^2 - 6t + 4 \times 1 \\ T(t) &= 3t - 2 && = 0t^2 + 3t + (-2) \times 1 \\ T(1) &= 1 && = ot^2 + 0t + 1 \times 1 \end{aligned}$$

Hence the matrix representation is given by:

$$\boldsymbol{A} = \begin{bmatrix} 9 & -6 & 4 \\ 0 & 3 & -2 \\ 0 & 0 & 1 \end{bmatrix}$$

We croos the column 1 to compute determinant:

$$\det(\boldsymbol{A}) = 9 \begin{vmatrix} 3 & -2 \\ 0 & 1 \end{vmatrix} = 27.$$

■

2. *Proof.* We only need to show $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = 0$:

By *postmultiplying* $\boldsymbol{x}^{\mathrm{T}}$ for $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} = 2\boldsymbol{y}$ both sides we obtain:

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} = 2\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$$

Or equivalently,

$$(\boldsymbol{A}\boldsymbol{x})^{\mathrm{T}}\boldsymbol{y} = 2\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} \implies \boldsymbol{0}^{\mathrm{T}}\boldsymbol{y} = 2\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} \implies \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = 0.$$

■

3. *Solution.* (a) *True.*

Reason: Assume $\boldsymbol{Q}$ is a $n \times n$ matrix s.t.

$$\boldsymbol{Q} = \begin{bmatrix} q_1 & q_2 & \dots & q_n \end{bmatrix}$$

Then the product of $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}$ is

$$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \begin{bmatrix} q_1^{\mathrm{T}} \\ q_2^{\mathrm{T}} \\ \vdots \\ q_n^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} q_1 & q_2 & \dots & q_n \end{bmatrix} = \begin{bmatrix} q_1^{\mathrm{T}}q_1 & q_1^{\mathrm{T}}q_2 & \dots & q_1^{\mathrm{T}}q_n \\ q_2^{\mathrm{T}}q_1 & q_2^{\mathrm{T}}q_2 & \dots & q_2^{\mathrm{T}}q_n \\ \vdots & \vdots & \ddots & \vdots \\ q_n^{\mathrm{T}}q_1 & q_n^{\mathrm{T}}q_2 & \dots & q_n^{\mathrm{T}}q_n \end{bmatrix}$$

Due to the orthonormality of $q_1, \dots, q_n$, we obtain:

$$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{I}_n.$$

Hence $\boldsymbol{Q}^{-1} = \boldsymbol{Q}^{\mathrm{T}}$. If we define $\boldsymbol{Q}^{-1} = \begin{bmatrix} q_1^* & q_2^* & \dots & q_n^* \end{bmatrix}$, then we obtain:

$$(\boldsymbol{Q}^{-1})^{\mathrm{T}}\boldsymbol{Q}^{-1} = \begin{bmatrix} (q_1^*)^{\mathrm{T}} \\ (q_2^*)^{\mathrm{T}} \\ \vdots \\ (q_n^*)^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} q_1^* & q_2^* & \dots & q_n^* \end{bmatrix} = \begin{bmatrix} (q_1^*)^{\mathrm{T}}q_1^* & (q_1^*)^{\mathrm{T}}q_2^* & \dots & (q_1^*)^{\mathrm{T}}q_n^* \\ (q_2^*)^{\mathrm{T}}q_1^* & (q_2^*)^{\mathrm{T}}q_2^* & \dots & (q_2^*)^{\mathrm{T}}q_n^* \\ \vdots & \vdots & \ddots & \vdots \\ (q_n^*)^{\mathrm{T}}q_1^* & (q_n^*)^{\mathrm{T}}q_2^* & \dots & (q_n^*)^{\mathrm{T}}q_n^* \end{bmatrix} = \boldsymbol{I}$$

Hence for columns $q_1^*, q_2^*, \dots, q_n^*$ we have:

$$\langle \boldsymbol{q}_i^*, \boldsymbol{q}_j^* \rangle = \begin{cases} 0 & \text{when } i \neq j \qquad (\textbf{orthogonal} \text{ vectors}), \\ 1 & \text{when } i = j \qquad (\textbf{unit} \text{ vectors: } \|\boldsymbol{q}_i^*\| = 1). \end{cases}$$

for $i, j \in \{1, 2, \dots, n\}$.

By definition, $q_1^*, q_2^*, \dots, q_n^*$ are orthonormal. Hence $\boldsymbol{Q}^{-1}$ is a orthogonal matrix.

Example:

$$Q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \implies Q^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

which is obviously orthonormal.

(b) *True.*

Reason: Assume $Q = \begin{bmatrix} q_1 & q_2 & \dots & q_n \end{bmatrix}$, where $q_i \in \mathbb{R}^m$ for $i = 1, \dots, n$.

- Firstly we show $Q^T Q = I$:

$$Q^T Q = \begin{bmatrix} q_1^T \\ q_2^T \\ \vdots \\ q_n^T \end{bmatrix} \begin{bmatrix} q_1 & q_2 & \dots & q_n \end{bmatrix} = \begin{bmatrix} q_1^T q_1 & & & \\ & q_2^T q_2 & & \\ & & \ddots & \\ & & & q_n^T q_n \end{bmatrix} = I_n.$$

- Hence we derive

$$\begin{aligned} \|Qx\|^2 &= x^T Q^T Q x \\ &= x^T (Q^T Q) x = x^T I x \\ &= x^T x \\ &= \|x\|^2 \end{aligned}$$

Hence $\|Qx\| = \|x\|$.

Example:

If $Q = \begin{bmatrix} 1 \\ 0 \end{bmatrix}_{2 \times 1}$, then for any $x = \begin{bmatrix} \alpha \end{bmatrix}$ (α is a row vector),

$$\|Qx\| = \| \begin{bmatrix} \alpha \\ \mathbf{0} \end{bmatrix} \| = \sqrt{|\langle \alpha, \alpha \rangle| + \mathbf{0}^2} = \sqrt{|\langle \alpha, \alpha \rangle|} \tag{10.6}$$

$$\|x\| = \sqrt{|\langle \alpha, \alpha \rangle|}. \tag{10.7}$$

Hence we obtain $\|\boldsymbol{Q}\boldsymbol{x}\| = \|\boldsymbol{x}\|$ for $\forall \boldsymbol{x}$.

(c) *False.*

Example:

$\boldsymbol{Q} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}, \boldsymbol{y} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, then note that

$$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{y} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Thus $\|\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{y}\| = 0 \neq 1 = \|\boldsymbol{y}\|$.

■

4. *Solution.* • Firstly we show $\boldsymbol{W}_1 \subset \boldsymbol{W}_2^{\perp}$:

For $\forall p \in \boldsymbol{W}_1, \forall q \in \boldsymbol{W}_2$, we only need to show $\langle p, q \rangle = 0$:

– For $\forall f \in \boldsymbol{W}_2$, we have

$$\begin{aligned}
\int_{-1}^{1} f(x)\,\mathrm{d}x &= \int_{-1}^{0} f(x)\,\mathrm{d}x + \int_{0}^{1} f(x)\,\mathrm{d}x \\
&= \int_{-1}^{0} -f(-x)\,\mathrm{d}x + \int_{0}^{1} f(x)\,\mathrm{d}x \\
&= \int_{-1}^{0} f(-x)\,\mathrm{d}(-x) + \int_{0}^{1} f(x)\,\mathrm{d}x \\
&= \int_{1}^{0} f(x)\,\mathrm{d}(x) + \int_{0}^{1} f(x)\,\mathrm{d}x \\
&= 0.
\end{aligned}$$

– And the product $pq \in \boldsymbol{W}_2$, this is because:

$$\begin{aligned}
(pq)(x) = p(x)q(x) &= p(-x) - q(-x) \\
&= -p(-x)q(-x) \\
&= -(pq)(-x).
\end{aligned}$$

Hence the inner product $\langle p,q\rangle$ is given by:

$$\langle p,q\rangle = \int_{-1}^{1} p(x)q(x)\,\mathrm{d}x = \int_{-1}^{1} (pq)(x)\,\mathrm{d}x = 0$$

Hence $\boldsymbol{W}_1 \perp \boldsymbol{W}_2 \implies \boldsymbol{W}_1 \subset \boldsymbol{W}_2^{\perp}$.

- Then we show $\boldsymbol{W}_2^{\perp} \subset \boldsymbol{W}_1$:

  Suppose $p^* \notin \boldsymbol{W}_1$, then we want to show $\langle p^*,q\rangle \neq 0$ for some $q \in \boldsymbol{W}_2$:

  – We decompose p^* into

  $$p^*(x) = p_1(x) + p_2(x)$$

  where $p_1(x) = \frac{p^*(x)+p^*(-x)}{2}$ and $p_2(x) = \frac{p^*(x)-p^*(-x)}{2}$. Since we have

  $$p_1(-x) = \frac{p^*(-x)+p^*(x)}{2} = p_1(x)$$
  $$p_2(-x) = \frac{p^*(-x)-p^*(x)}{2} = -p_2(x),$$

  we derive $p_1(x) \in \boldsymbol{W}_1, p_2(x) \in \boldsymbol{W}_2$. ($p^* \notin \boldsymbol{W}_1 \implies p_2 \neq 0$.)

  – Thus the inner product for $\langle p^*, p_2\rangle$ is positive:

  $$\begin{aligned}\langle p^*, p_2\rangle &= \langle p_1 + p_2, p_2\rangle \\ &= \langle p_1, p_2\rangle + \langle p_2, p_2\rangle \\ &= 0 + \int_{-1}^{1} p_2^2(x)\,\mathrm{d}x > 0.\end{aligned}$$

  Hence given $\forall p^* \notin \boldsymbol{W}_1$, there exists $q = p_2 \in \boldsymbol{W}_2$ s.t. $\langle p^*,q\rangle \neq 0$.

  Thus $p^* \notin \boldsymbol{W}_2^{\perp} \implies \boldsymbol{W}_2^{\perp} \subset \boldsymbol{W}_1$.

Hence we obtain $\boldsymbol{W}_1 = \boldsymbol{W}_2^{\perp}$. ■

5. *Solution.* • Firstly we find a basis for $\boldsymbol{U}$:

The space span $\left\{\begin{bmatrix}1\\2\\-5\end{bmatrix}\right\}$ is the row space for matrix

$$\boldsymbol{A} = \begin{bmatrix}1 & 2 & -5\end{bmatrix}$$

Hence $\boldsymbol{U} = (C(\boldsymbol{A}))^{\perp} = N(\boldsymbol{A})$. We only need to find the basis for $N(\boldsymbol{A})$:

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0} \implies x_1 + 2x_2 - 5x_3 = 0.$$

Hence the solution to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ is

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

where x_2, x_3 are arbitrary scalars.

Hence $\boldsymbol{U}$ is spanned by $\left\{\begin{pmatrix}-2\\1\\0\end{pmatrix}, \begin{pmatrix}5\\0\\1\end{pmatrix}\right\}$. And obviously, $\begin{pmatrix}-2\\1\\0\end{pmatrix}$ and $\begin{pmatrix}5\\0\\1\end{pmatrix}$ are ind.

Hence one basis for $\boldsymbol{U}$ is $\left\{\begin{pmatrix}-2\\1\\0\end{pmatrix}, \begin{pmatrix}5\\0\\1\end{pmatrix}\right\}$.

- Let's do Gram-Schmidt Process to convert this basis into *orthonormal*:

  We set $\boldsymbol{a} = \begin{bmatrix}-2\\1\\0\end{bmatrix}$ and $\boldsymbol{b} = \begin{bmatrix}5\\0\\1\end{bmatrix}$.

  - Then we set $\boldsymbol{A} = \begin{bmatrix}-2\\1\\0\end{bmatrix}$.

– Next step, we compute

$$
\begin{aligned}
\boldsymbol{B} = \boldsymbol{b} - \mathrm{Proj}_{\boldsymbol{A}}(\boldsymbol{b}) &= \boldsymbol{b} - \frac{\langle \boldsymbol{A}, \boldsymbol{b} \rangle}{\langle \boldsymbol{A}, \boldsymbol{A} \rangle}\boldsymbol{A} \\
&= \begin{pmatrix} 5 \\ 0 \\ 1 \end{pmatrix} - \frac{-10}{5}\begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}
\end{aligned}
$$

– Then we convert orthogonal sets $\{\boldsymbol{A}, \boldsymbol{B}\}$ into orthonormal:

$$
\boldsymbol{q}_1 := \frac{\boldsymbol{A}}{\|\boldsymbol{A}\|} = \begin{pmatrix} -\frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \\ 0 \end{pmatrix} \qquad \boldsymbol{q}_2 := \frac{\boldsymbol{B}}{\|\boldsymbol{B}\|} = \begin{pmatrix} \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{pmatrix}
$$

In conclusion, one orthonormal basis for $\boldsymbol{U}$ is $\left\{ \begin{pmatrix} -\frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{pmatrix} \right\}$. ■

6. *Solution.* We only need to find *least squares solution* $\boldsymbol{x}^*$ to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, where

$$
\boldsymbol{A} = \begin{bmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \qquad \boldsymbol{x} = \begin{bmatrix} C \\ D \end{bmatrix} \qquad \boldsymbol{b} = \begin{bmatrix} 4 \\ 2 \\ -1 \\ 0 \\ 0 \end{bmatrix}.
$$

Take on trust that we only need to solve $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$.

- But before that, let's do QR factorization for $\boldsymbol{A}$:

  Define $\boldsymbol{A} := \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 \end{bmatrix}$ $\quad \langle \boldsymbol{a}_1, \boldsymbol{a}_2 \rangle = 0 \implies$ Columns of $\boldsymbol{A}$ are orthogonal.

  So we obtain orthonormal vectors:

$$\boldsymbol{q}_1 := \frac{\boldsymbol{a}_1}{\|\boldsymbol{a}_1\|} = \begin{bmatrix} \frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \end{bmatrix} \qquad \boldsymbol{q}_2 = \frac{\boldsymbol{a}_2}{\|\boldsymbol{a}_2\|} = \begin{bmatrix} -\frac{2}{\sqrt{10}} \\ -\frac{1}{\sqrt{10}} \\ 0 \\ \frac{1}{\sqrt{10}} \\ \frac{2}{\sqrt{10}} \end{bmatrix}$$

  Thus the factor is given by

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 \end{bmatrix} \qquad \boldsymbol{R} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A} = \begin{bmatrix} \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{a}_1 & \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{a}_2 \\ 0 & \boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{a}_2 \end{bmatrix} = \begin{bmatrix} \sqrt{5} & 0 \\ 0 & \sqrt{10} \end{bmatrix}.$$

- Hence we could compute the least squares solution more easily:

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} \iff \boldsymbol{R}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{R}\boldsymbol{x} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b} \iff \boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{x} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}$$

$$\begin{aligned} \implies \boldsymbol{x} = \boldsymbol{R}^{-1}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b} &= \frac{1}{5\sqrt{2}} \begin{bmatrix} \sqrt{10} & 0 \\ 0 & \sqrt{5} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ -\frac{2}{\sqrt{10}} & -\frac{1}{\sqrt{10}} & 0 & \frac{1}{\sqrt{10}} & \frac{2}{\sqrt{10}} \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ -1 \end{bmatrix} \end{aligned}$$

Hence we have $\begin{cases} C = 1 \\ D = -1. \end{cases}$ The best line is $\hat{y} = 1 - x$. ■

10.1.7. Solution to Assignment Seven

1. *Solution.* A hidden assumption is $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = \|\boldsymbol{x}\|^2 \neq 0$. But this is not always true, let me raise a counterexample:

   The eigenvectors of rotation matrix $\boldsymbol{K} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ are $\boldsymbol{x}_1 = \alpha \begin{pmatrix} 1 \\ i \end{pmatrix}$ associated with eigenvalue $\lambda_1 = i$ and $\boldsymbol{x}_2 = \beta \begin{bmatrix} 1 \\ -i \end{bmatrix}$ associated with eigenvalue $\lambda_2 = -i$. Foe each $\boldsymbol{x}_i$ we obtain

   $$\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_i = 1 + i^2 = 0.$$

   But the eigenvalues are all complex, which leads to a contradiction for the statement. ■

2. *Proof.* (a) The eigenspace for λ is given by

   $$\{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}\}.$$

   Firstly we investigate $\boldsymbol{A}\boldsymbol{X}$:

   $$\begin{aligned}\boldsymbol{A}\boldsymbol{X} &= \boldsymbol{A}\begin{bmatrix}\boldsymbol{x}_1 & \dots & \boldsymbol{x}_n\end{bmatrix} \\ &= \begin{bmatrix}\boldsymbol{A}\boldsymbol{x}_1 & \dots & \boldsymbol{A}\boldsymbol{x}_k & \boldsymbol{A}\boldsymbol{x}_{k+1}\dots & \boldsymbol{A}\boldsymbol{x}_n\end{bmatrix} \\ &= \begin{bmatrix}\lambda\boldsymbol{x}_1 & \dots & \lambda\boldsymbol{x}_k & \boldsymbol{A}\boldsymbol{x}_{k+1} & \dots & \boldsymbol{A}\boldsymbol{x}_n\end{bmatrix}\end{aligned}$$

   Then investigate $\boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{X}$:

   $$\begin{aligned}\boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{X} &= \boldsymbol{X}^{-1}\begin{bmatrix}\lambda\boldsymbol{x}_1 & \dots & \lambda\boldsymbol{x}_k & \boldsymbol{A}\boldsymbol{x}_{k+1} & \dots & \boldsymbol{A}\boldsymbol{x}_n\end{bmatrix} \\ &= \begin{bmatrix}\lambda\boldsymbol{X}^{-1}\boldsymbol{x}_1 & \dots & \lambda\boldsymbol{X}^{-1}\boldsymbol{x}_k & \boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{x}_{k+1} & \dots & \boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{x}_n\end{bmatrix}\end{aligned}$$

   Since $\boldsymbol{x}_1,\dots,\boldsymbol{x}_k$ are columns of $\boldsymbol{X}$, and $\boldsymbol{X}^{-1}\boldsymbol{X} = \boldsymbol{I}$, we obtain

   $$\boldsymbol{X}^{-1}\boldsymbol{x}_i = \boldsymbol{e}_i \text{ for } i = 1,\dots,k.$$

Hence

$$
\begin{aligned}
\boldsymbol{B} =& \boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{X} \\
=& \begin{bmatrix} \lambda \boldsymbol{X}^{-1}\boldsymbol{x}_1 & \dots & \lambda \boldsymbol{X}^{-1}\boldsymbol{x}_k & \boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{x}_{k+1} & \dots & \boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{x}_n \end{bmatrix} \\
=& \begin{bmatrix} \lambda \boldsymbol{e}_1 & \dots & \lambda \boldsymbol{X}^{-1}\boldsymbol{e}_k & \boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{x}_{k+1} & \dots & \boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{x}_n \end{bmatrix} \\
=& \begin{bmatrix} \lambda & 0 & \dots & 0 & b_{1(k+1)} & \dots & b_{1n} \\ 0 & \lambda & \dots & 0 & b_{2(k+1)} & \dots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda & b_{k(k+1)} & \dots & b_{kn} \\ 0 & 0 & \dots & 0 & b_{(k+1)(k+1)} & \dots & b_{(k+1)n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & b_{n(k+1)} & \dots & b_{nn} \end{bmatrix}
\end{aligned}
$$

If we write $\boldsymbol{B}$ in block matrix form, then we obtain:

$$
\boldsymbol{B} = \begin{bmatrix} \lambda \boldsymbol{I} & \boldsymbol{B}_{12} \\ \mathbf{0} & \boldsymbol{B}_{22} \end{bmatrix}.
$$

(b) For a fixed eigenvalue λ^*, $\boldsymbol{B}$ could be written as

$$
\boldsymbol{B} = \begin{bmatrix} \lambda^* & 0 & \dots & 0 & b_{1(k+1)} & \dots & b_{1n} \\ 0 & \lambda^* & \dots & 0 & b_{2(k+1)} & \dots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda^* & b_{k(k+1)} & \dots & b_{kn} \\ 0 & 0 & \dots & 0 & b_{(k+1)(k+1)} & \dots & b_{(k+1)n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & b_{n(k+1)} & \dots & b_{nn} \end{bmatrix}
$$

Hence the matrix for $\lambda\boldsymbol{I}-\boldsymbol{B}$ is given by:

$$\lambda\boldsymbol{I}-\boldsymbol{B}=\begin{bmatrix}\lambda-\lambda^* & 0 & \dots & 0 & -b_{1(k+1)} & \dots & -b_{1n}\\ 0 & \lambda-\lambda^* & \dots & 0 & -b_{2(k+1)} & \dots & -b_{2n}\\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & \lambda-\lambda^* & -b_{k(k+1)} & \dots & -b_{kn}\\ 0 & 0 & \dots & 0 & \lambda-b_{(k+1)(k+1)} & \dots & -b_{(k+1)n}\\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \dots & 0 & -b_{n(k+1)} & \dots & \lambda-b_{nn}\end{bmatrix}$$

In order to compute $\left|\lambda\boldsymbol{I}-\boldsymbol{B}\right|$, we cross the first k columns to get

$$\left|\lambda\boldsymbol{I}-\boldsymbol{B}\right|=(\lambda-\lambda^*)^k\begin{vmatrix}\lambda-b_{(k+1)(k+1)} & \dots & -b_{(k+1)n}\\ \vdots & \ddots & \vdots\\ -b_{n(k+1)} & \dots & \lambda-b_{nn}\end{vmatrix}$$

Hence the term $(\lambda-\lambda^*)$ appears at least k times in the characteristic polynomial of $\left|\lambda\boldsymbol{I}-\boldsymbol{B}\right|$.

Hence λ^* is an eigenvalue of $\boldsymbol{B}$ with multiplicity at least k.

Since $\boldsymbol{B}$ is similar to $\boldsymbol{A}$, they have the same eigenvalues. Hence λ^* is an eigenvalue of $\boldsymbol{A}$ with multiplicity at least k.

■

3. *Solution.* (a) $\boldsymbol{Ax}=\lambda\boldsymbol{x}\implies(\boldsymbol{A}-\lambda\boldsymbol{I})\boldsymbol{x}=\boldsymbol{0}$. Since $\lambda=0$, we only need to investigate the dimension for $\boldsymbol{x}$, where $\boldsymbol{Ax}=\boldsymbol{0}$.

   Since $\boldsymbol{A}=\boldsymbol{xy}^{\mathrm{T}}$, $\operatorname{rank}(\boldsymbol{A})=1$. Hence $\dim(N(\boldsymbol{A}))=n-1$. So the eigenspace for λ is $n-1$ dimension.

   Thus $\lambda=0$ is an eigenvalue of $\boldsymbol{A}$ with $n-1$ ind. eigenvectors.

   (b) By part (a),

$$\lambda_1=\lambda_2=\dots=\lambda_{n-1}=0.$$

The sum of the eigenvalues is the trace of $\boldsymbol{A}$ which equals to $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$. Thus

$$\sum_{i=1}^{n}\lambda_i = \lambda_n = \text{trace}(\boldsymbol{A}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}.$$

Hence the remaining eigenvalue of $\boldsymbol{A}$ is $\lambda_n = \text{trace}(\boldsymbol{A}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$.

(c) From part(a) $\lambda = 0$ has $n-1$ ind. eigenvectors.

Since $\lambda_n \neq 0$, the eigenvector associated to λ_n will be independent from the $n-1$ eigenvectors.(A theorem says if eigenvalues $\lambda_1,\dots,\lambda_k$ are distinct, their corresponding eigenvectors $\boldsymbol{x}_1,\dots,\boldsymbol{x}_k$ will be ind.)

Hence $\boldsymbol{A}$ has n ind. eigenvectors, $\boldsymbol{A}$ is diagonalizable.

∎

4. This question is the special case for Cayley-Hamilton theorem. It states that if the charactristic polynomial for $\boldsymbol{A}$ is $P_{\boldsymbol{A}}(\lambda) = (\lambda-\lambda_1)\dots(\lambda-\lambda_n)$, then

$$P_{\boldsymbol{A}}(\boldsymbol{A}) = (\boldsymbol{A}-\lambda_1\boldsymbol{I})\dots(\boldsymbol{A}-\lambda_n\boldsymbol{I}) = \boldsymbol{0}.$$

Proof. Obviously, $\boldsymbol{A}$ has n ind. eigenvectors. Hence $\boldsymbol{A}$ is *diagonalizable*. Hence we decompose $\boldsymbol{A}$ as

$$\boldsymbol{A} = \boldsymbol{S}\boldsymbol{D}\boldsymbol{S}^{-1}$$

where $\boldsymbol{D} = \text{diag}(\lambda_1,\dots,\lambda_n)$, and $\lambda_1,\dots,\lambda_n$ are n eigenvalues of $\boldsymbol{A}$.
Hence we write $\boldsymbol{B}$ as:

$$\begin{aligned}
\boldsymbol{B} &= (\boldsymbol{A}-\lambda_1\boldsymbol{I})\dots(\boldsymbol{A}-\lambda_n\boldsymbol{I}) \\
&= (\boldsymbol{S}\boldsymbol{D}\boldsymbol{S}^{-1}-\lambda_1\boldsymbol{I})\dots(\boldsymbol{S}\boldsymbol{D}\boldsymbol{S}^{-1}-\lambda_n\boldsymbol{I}) \\
&= (\boldsymbol{S}\boldsymbol{D}\boldsymbol{S}^{-1}-\lambda_1\boldsymbol{S}\boldsymbol{S}^{-1})\dots(\boldsymbol{S}\boldsymbol{D}\boldsymbol{S}^{-1}-\lambda_n\boldsymbol{S}\boldsymbol{S}^{-1}) \\
&= \left[\boldsymbol{S}(\boldsymbol{D}-\lambda_1\boldsymbol{I})\boldsymbol{S}^{-1}\right]\dots\left[\boldsymbol{S}(\boldsymbol{D}-\lambda_n\boldsymbol{I})\boldsymbol{S}^{-1}\right] = \boldsymbol{S}(\boldsymbol{D}-\lambda_1\boldsymbol{I})\dots(\boldsymbol{D}-\lambda_n\boldsymbol{I})\boldsymbol{S}^{-1}
\end{aligned}$$

For each term $(\boldsymbol{D}-\lambda_i\boldsymbol{I})$, $i\in\{1,2,\dots,n\}$, we find its ith row are all zero.
Hence the product $(\boldsymbol{D}-\lambda_1\boldsymbol{I})\dots(\boldsymbol{D}-\lambda_n\boldsymbol{I})$ must be zero matrix.
Hence $\boldsymbol{B} = \boldsymbol{S}(\boldsymbol{D}-\lambda_1\boldsymbol{I})\dots(\boldsymbol{D}-\lambda_n\boldsymbol{I})\boldsymbol{S}^{-1}$ is a *zero matrix*. ∎

5. *Solution.* (a) Since $\lambda \neq 0$ is a eigenvalue of $\boldsymbol{AB}$, there exists vector $\boldsymbol{x}$ s.t.

$$\boldsymbol{ABx} = \lambda \boldsymbol{x}$$

By postmultiplying $\boldsymbol{B}$ both sides we obtain

$$\boldsymbol{B}(\boldsymbol{ABx}) = \lambda \boldsymbol{Bx} \implies \boldsymbol{BA}(\boldsymbol{Bx}) = \lambda(\boldsymbol{Bx})$$

Hence we only need to show $\boldsymbol{Bx} \neq \boldsymbol{0}$:

Assume $\boldsymbol{Bx} = \boldsymbol{0}$, then $\boldsymbol{ABx} = \boldsymbol{A}(\boldsymbol{Bx}) = \boldsymbol{A0} = \boldsymbol{0} = \lambda \boldsymbol{x}$.

Hence $\lambda = 0$, which leads to a contradiction. Hence there exists eigenvector $\boldsymbol{Bx} \neq \boldsymbol{0}$ s.t.

$$\boldsymbol{BA}(\boldsymbol{Bx}) = \lambda(\boldsymbol{Bx})$$

Thus λ is also an eigenvalue of $\boldsymbol{BA}$.

(b) By definition, there exists vector $\boldsymbol{x} \neq \boldsymbol{0}$ s.t.

$$\boldsymbol{ABx} = \lambda \boldsymbol{x} = 0\boldsymbol{x} = \boldsymbol{0}.$$

Hence $\boldsymbol{AB}$ is *singular*, the determinant $\det(\boldsymbol{AB}) = 0$.

$$\det(\boldsymbol{AB}) = \det(\boldsymbol{A})\det(\boldsymbol{B}) = \det(\boldsymbol{B})\det(\boldsymbol{A}) = \det(\boldsymbol{BA}) = 0.$$

Hence $\boldsymbol{BA}$ is also *singular*. Thus there exists $\boldsymbol{y} \neq \boldsymbol{0}$ s.t.

$$\boldsymbol{BAy} = \boldsymbol{0} = 0\boldsymbol{y}$$

By definition, $\lambda = 0$ is also an eigenvalue of $\boldsymbol{BA}$.

■

6. *Proof.* (a) We set $\boldsymbol{u}_k = \begin{bmatrix} a_{k+1} \\ a_k \end{bmatrix}$. The rule

$$\begin{cases} a_{k+2} = 3a_{k+1} - 2a_k \\ a_{k+1} = a_{k+1} \end{cases}$$

can be written as $\boldsymbol{u}_{k+1} = \begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix} \boldsymbol{u}_k$. And $\boldsymbol{u}_0 = \begin{bmatrix} 5 \\ 4 \end{bmatrix}$.

After computation we derive $\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is eigenvector of $\boldsymbol{A}$ corresponding to eigenvalue $\lambda_1 = 1$; $\boldsymbol{x}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ is eigenvector of $\boldsymbol{A}$ corresponding to eigenvalue $\lambda_2 = 2$.

And then, we want to find the lienar combination of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ to get $\boldsymbol{u}_0 = \begin{bmatrix} 5 \\ 4 \end{bmatrix}$:

$$\begin{bmatrix} 5 \\ 4 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \end{bmatrix}. \qquad \text{Or } \boldsymbol{u}_0 = 3\boldsymbol{x}_1 + \boldsymbol{x}_2$$

Then we multiply $\boldsymbol{u}_0$ by $\boldsymbol{A}^k$ to get $\boldsymbol{u}_k$:

$$\begin{aligned} \boldsymbol{u}_k = \boldsymbol{A}^k u_0 &= 3\boldsymbol{A}^k \boldsymbol{x}_1 + \boldsymbol{A}^k \boldsymbol{x}_2 \\ &= 3\lambda_1^k \boldsymbol{x}_1 + \lambda_2^k \boldsymbol{x}_2 \\ &= 3\boldsymbol{x}_1 + 2^k \boldsymbol{x}_2 \\ &= \begin{bmatrix} 3 + 2^{k+1} \\ 3 + 2^k \end{bmatrix}. \end{aligned}$$

Hence the general formula is $\boldsymbol{a}_k = 3 + 2^k$.

(b) We set $\boldsymbol{u}_k = \begin{bmatrix} b_{k+1} \\ b_k \end{bmatrix}$. The rule

$$\begin{cases} b_{k+2} = 4b_{k+1} - 4b_k \\ b_{k+1} = b_{k+1} \end{cases}$$

can be written as $\boldsymbol{u}_{k+1} = \begin{bmatrix} 4 & -4 \\ 1 & 0 \end{bmatrix} \boldsymbol{u}_k$. And $\boldsymbol{u}_0 = \begin{bmatrix} \beta \\ \alpha \end{bmatrix}$.

We set $\boldsymbol{A} = \begin{bmatrix} 4 & -4 \\ 1 & 0 \end{bmatrix}$, then there exists nonsingular $\boldsymbol{S} = \begin{bmatrix} 2 & 3 \\ 1 & 1 \end{bmatrix}$ such that

$$\boldsymbol{SD} = \begin{bmatrix} 2 & 3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 4 & -4 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 1 & 1 \end{bmatrix} \implies \boldsymbol{D} = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix} = \boldsymbol{S}^{-1}\boldsymbol{AS}.$$

Hence $\boldsymbol{A}$ is similar to $\boldsymbol{D}$.

Then we compute $\boldsymbol{A}^k$:

$$\begin{aligned} \boldsymbol{A}^k &= (\boldsymbol{SDS}^{-1})^k \\ &= \boldsymbol{SD}^k\boldsymbol{S}^{-1} \end{aligned}$$

Hence we only need to compute $\boldsymbol{D}^k$:

- We have known $\boldsymbol{D}^1 = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$.
- If we assume $\boldsymbol{D}^k = \begin{bmatrix} p(k) & q(k) \\ s(k) & t(k) \end{bmatrix}$, then $\boldsymbol{D}^{k+1} = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} p & q \\ s & t \end{bmatrix} = \begin{bmatrix} 2p+s & 2q+t \\ 2s & 2t \end{bmatrix} = \begin{bmatrix} p(k+1) & q(k+1) \\ s(k+1) & t(k+1) \end{bmatrix}$.
- Hence by induction, $s = 0, t(k) = 2^k$. And $p(k+1) = 2p(k) + 0 \implies p(k) = 2^k$; $q(k+1) = 2q(k) + t = 2q(k) + 2^k \implies q(k) = 2^{k-1}[q(1) + k - 1] = k \times 2^{k-1}$

- Hence $\boldsymbol{D}^k = \begin{bmatrix} 2^k & k\times 2^{k-1} \\ 0 & 2^k \end{bmatrix}$.

Thus $\boldsymbol{A}^k = \boldsymbol{S}\boldsymbol{D}^k\boldsymbol{S}^{-1} = \begin{bmatrix} 2 & 3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2^k & k\times 2^{k-1} \\ 0 & 2^k \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 1 & 1 \end{bmatrix}^{-1} = 2^k \begin{bmatrix} k+1 & -2k \\ \frac{k}{2} & 1-k \end{bmatrix}$.

Hence $\boldsymbol{u}_k = \boldsymbol{A}^k\boldsymbol{u}_0 = 2^k \begin{bmatrix} k+1 & -2k \\ \frac{k}{2} & 1-k \end{bmatrix} \begin{bmatrix} \beta \\ \alpha \end{bmatrix} = 2^k \begin{bmatrix} \beta(k+1)-2k\alpha \\ \beta(\frac{k}{2})+(1-k)\alpha \end{bmatrix}$

Hence the general formula is $b_k = 2^k \left[(1-k)\times\alpha + \frac{k}{2}\times\beta\right]$.

■

7. *Solution.* (a) False.

Reason: For *real symmetric* matrix, we have shown that its eigenvectors corrsponding to *distinct* eigenvalues are *orthigonal*. However, ind. eigenvectors corresponding to the same eigenvalue may not be *orthogonal.*

Example: Let $\boldsymbol{A} = \boldsymbol{I}$. Any nonzero vector is eigenvector. But two different vectors may not have to be orthogonal.

(b) True.

Reason: We do the eigendecomposition for $\boldsymbol{A}$:

$$\boldsymbol{A} = \boldsymbol{S}\Lambda\boldsymbol{S}^{-1}$$

where $\Lambda = \text{diag}(\lambda_1,\dots,\lambda_n)$ and $\boldsymbol{S} = \begin{bmatrix} \boldsymbol{x}_1 & \dots & \boldsymbol{x}_n \end{bmatrix}$, where $\boldsymbol{x}_i$ is the eigenvector of $\boldsymbol{A}$ associated with eigenvalue λ_i for $i = 1,2,\dots,n$.

Since columns of $\boldsymbol{S}$ are orthonormal vectors, it is *unitary.* Hence $\boldsymbol{A} = \boldsymbol{S}\Lambda\boldsymbol{S}^{\mathrm{H}}$. Since $\Lambda^{\mathrm{H}} = \Lambda$, we obtain

$$\boldsymbol{A}^{\mathrm{H}} = (\boldsymbol{S}\Lambda\boldsymbol{S}^{\mathrm{H}})^{\mathrm{H}} = \boldsymbol{S}\Lambda^{\mathrm{H}}\boldsymbol{S}^{\mathrm{H}} = \boldsymbol{S}\Lambda\boldsymbol{S}^{\mathrm{H}} = \boldsymbol{A}$$

So $\boldsymbol{A}$ is Hermitian.

(c) True.

Reason: Suppose $\boldsymbol{A}$ has the eigendecomposition

$$\boldsymbol{A} = \boldsymbol{S}\Lambda\boldsymbol{S}^{-1}$$

Then for the series we obtain:

$$\begin{aligned}
\boldsymbol{I} + \boldsymbol{A} + \frac{1}{2!}\boldsymbol{A}^2 + \dots &= \boldsymbol{S}\boldsymbol{S}^{-1} + \boldsymbol{S}\Lambda\boldsymbol{S}^{-1} + \frac{1}{2!}\boldsymbol{S}\Lambda^2\boldsymbol{S}^{-1} + \dots \\
&= \boldsymbol{S}(\boldsymbol{I} + \Lambda + \frac{1}{2!}\Lambda^2 + \dots)\boldsymbol{S}^{-1}
\end{aligned}$$

If we define the series $\boldsymbol{I} + \boldsymbol{A} + \frac{1}{2!}\boldsymbol{A}^2 + \dots := e^{\boldsymbol{A}}$, then we obtain:

$$e^{\boldsymbol{A}} = \boldsymbol{S}e^{\Lambda}\boldsymbol{S}^{-1}$$

Since every term for the series e^{Λ} is *diagonal matrix*, the series e^{Λ} is consequently a *diagonal matrix*.

Hence $e^{\boldsymbol{A}}$ is diagonalizable.

(d) True.

Reason: Since $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}$, taking complex conjugate we obtain $\overline{\boldsymbol{A}\boldsymbol{A}^{-1}} = \boldsymbol{I}$. Taking transpose we get $(\boldsymbol{A}^{-1})^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}} = \boldsymbol{I}$.

And we have $\boldsymbol{A}^{\mathrm{H}} = \boldsymbol{A}$, so $(\boldsymbol{A}^{-1})^{\mathrm{H}}\boldsymbol{A} = \boldsymbol{I}$. That is to say $(\boldsymbol{A}^{-1})^{\mathrm{H}} = \boldsymbol{A}^{-1}$. Hence $\boldsymbol{A}^{-1}$ is Hermitian.

■

8. *Solution.* (a)

- $N(\boldsymbol{A}^{\mathrm{T}})$ is *orthogonal* to $C(\boldsymbol{A})$ under the **old unconjugated inner product.**

  In fact, for $\forall \boldsymbol{u} \in N(\boldsymbol{A}^{\mathrm{T}})$ and $\forall \boldsymbol{A}\boldsymbol{v} \in C(\boldsymbol{A})$,

  $$(\boldsymbol{A}\boldsymbol{v})^{\mathrm{T}}\boldsymbol{u} = \boldsymbol{v}^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}) = \boldsymbol{v}^{\mathrm{T}}\boldsymbol{0} = \boldsymbol{0}. \implies C(\boldsymbol{A}) \perp N(\boldsymbol{A}^{\mathrm{T}}) \iff N(\boldsymbol{A}^{\mathrm{T}}) \perp C(\boldsymbol{A}).$$

- However, $N(\boldsymbol{A}^{\mathrm{T}})$ is *not always orthogonal* to $C(\boldsymbol{A})$ under the **new unconjugated inner product.**

Example: If $\boldsymbol{A} = \begin{pmatrix} 1 & 1 \\ i & i \end{pmatrix}$, then $\boldsymbol{u} = \begin{pmatrix} 1 \\ i \end{pmatrix} \in C(\boldsymbol{A})$ and $\boldsymbol{u} \in N(\boldsymbol{A}^{\mathrm{T}})$.

But $\boldsymbol{u}^{\mathrm{H}}\boldsymbol{u} = 2 \neq 0$.

- $N(\boldsymbol{A}^{\mathrm{H}})$ is *orthogonal* to $C(\boldsymbol{A})$ under the **new unconjugated inner product.**

  In fact, for $\forall \boldsymbol{u} \in N(\boldsymbol{A}^{\mathrm{H}})$ and $\forall \boldsymbol{A}\boldsymbol{v} \in C(\boldsymbol{A})$,

  $$(\boldsymbol{A}\boldsymbol{v})^{\mathrm{H}}\boldsymbol{u} = \boldsymbol{v}^{\mathrm{H}}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{u}) = \boldsymbol{v}^{\mathrm{H}}\boldsymbol{0} = \boldsymbol{0}. \implies C(\boldsymbol{A}) \perp N(\boldsymbol{A}^{\mathrm{H}}) \Longleftrightarrow N(\boldsymbol{A}^{\mathrm{H}}) \perp C(\boldsymbol{A}).$$

- However, $N(\boldsymbol{A}^{\mathrm{H}})$ is *not always orthogonal* to $C(\boldsymbol{A})$ under the **old unconjugated inner product.**

  Example: If $\boldsymbol{A} = \begin{pmatrix} 1 & 1 \\ i & i \end{pmatrix}$, then $\boldsymbol{u} = \begin{pmatrix} 1 \\ i \end{pmatrix} \in C(\boldsymbol{A})$ and $\boldsymbol{v} = \begin{bmatrix} 1 \\ -i \end{bmatrix} \in N(\boldsymbol{A}^{\mathrm{H}})$.

  But $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v} = 2 \neq 0$.

(b)

- **Example:** Let $\boldsymbol{V} = \operatorname{span}\left\{\begin{pmatrix} 1 \\ i \end{pmatrix}\right\}$.

  Then since we have $\begin{pmatrix} 1 & i \end{pmatrix}\begin{pmatrix} 1 \\ i \end{pmatrix} = 0$, we see $\boldsymbol{V}^{\perp} = \boldsymbol{V}$. Thus $\boldsymbol{V} \cap \boldsymbol{V}^{\perp} = \boldsymbol{V}$!

- If we use $\boldsymbol{x}^{\mathrm{H}}\boldsymbol{v} = \boldsymbol{0}$ to define the orthogonal complement, then $\{\boldsymbol{0}\} \notin \boldsymbol{V} \cap \boldsymbol{V}^{\perp}$.

  Assume $\boldsymbol{V} \cap \boldsymbol{V}^{\perp}$ contains some nonzero vector $\boldsymbol{x}$, then $\boldsymbol{x}$ is orthogonal to itself:

  $$\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x} = 0.$$

  But $\boldsymbol{x}^{\mathrm{H}}\boldsymbol{x} = \|\boldsymbol{x}\|^2$, so $\boldsymbol{x} = \boldsymbol{0}$, which leads to a contradiction!

■

10.1.8. Solution to Assignment Eight

1. *Solution.* We factorize $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ into:

$$\boldsymbol{A} = \boldsymbol{U}\Sigma\boldsymbol{V}^{\mathrm{T}}$$

where $\boldsymbol{U}$ is a $n \times n$ **orthogonal** matrix, Σ is a $n \times n$ *diagonal* matrix, $\boldsymbol{V}$ is a $n \times n$ **orthogonal** matrix.

Thus we write $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ and $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ as:

$$\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{U}\Sigma\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}\Sigma^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}} = \boldsymbol{U}\Sigma\Sigma^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}} = \boldsymbol{U}\Sigma^2\boldsymbol{U}^{\mathrm{T}}. \qquad \text{Since } \boldsymbol{V}^{\mathrm{T}}\boldsymbol{V} = \boldsymbol{I} \text{ due to orthonormality.}$$
$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{V}\Sigma^{\mathrm{T}}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\Sigma\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{V}\Sigma^{\mathrm{T}}\Sigma\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{V}\Sigma^2\boldsymbol{V}^{\mathrm{T}}. \qquad \text{Since } \boldsymbol{U}^{\mathrm{T}}\boldsymbol{U} = \boldsymbol{I} \text{ due to orthonormality.}$$

If we set $\boldsymbol{S} = (\boldsymbol{V}^{\mathrm{T}})^{-1}\boldsymbol{U}^{\mathrm{T}} = \boldsymbol{V}\boldsymbol{U}^{\mathrm{T}}$, then the inverse is given by $\boldsymbol{S}^{-1} = (\boldsymbol{U}^{\mathrm{T}})^{-1}\boldsymbol{V}^{-1} = \boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}$.

Hnece there exists invertible $\boldsymbol{S} = \boldsymbol{V}\boldsymbol{U}^{\mathrm{T}}$ such that

$$\begin{aligned}\boldsymbol{S}^{-1}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})\boldsymbol{S} &= \boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})\boldsymbol{V}\boldsymbol{U}^{\mathrm{T}} \\ &= \boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}\Sigma^2\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{U}^{\mathrm{T}} \\ &= \boldsymbol{U}\Sigma^2\boldsymbol{U}^{\mathrm{T}} = \boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\end{aligned}$$

Hence $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is similiar to $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$, i.e. $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ and $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ are **similar**. ■

2. Let $\boldsymbol{A}$ be $m \times n$ $(m \geq n)$ matrix of rank n with singular value decomposition $\boldsymbol{U}\Sigma\boldsymbol{V}^{\mathrm{T}}$. Let Σ^{+} denote the $n \times m$ matrix

$$\begin{pmatrix} \frac{1}{\sigma_1} & & & 0 & \dots & 0 \\ & \ddots & & \vdots & \ddots & \vdots \\ & & \frac{1}{\sigma_n} & 0 & \dots & 0 \end{pmatrix}$$

And we define $\boldsymbol{A}^{+} = \boldsymbol{V}\Sigma^{+}\boldsymbol{U}^{\mathrm{T}}$

(a) Show that

$$\boldsymbol{A}\boldsymbol{A}^{+} = \begin{bmatrix} \boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{A}^{+}\boldsymbol{A} = \boldsymbol{I}_n.$$

(Note that $\boldsymbol{A}^{+}$ is called the **pseudo-inverse** of $\boldsymbol{A}$.)

(b) Show that $\hat{\boldsymbol{x}} = \boldsymbol{A}^{+}\boldsymbol{b}$ satisfies the normal equation $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$.

Solution. (a) We write Σ^{+} into block matrix:

$$\Sigma^{+} = \begin{bmatrix} \Sigma^{-1} & \boldsymbol{0}_{n\times(m-n)} \end{bmatrix}$$

where $\Sigma^{-1} := \operatorname{diag}(\frac{1}{\sigma_1}, \dots, \frac{1}{\sigma_n})$.

Hence $\Sigma\Sigma^{+} = \begin{bmatrix} \Sigma\Sigma^{-1} & \boldsymbol{0}_{m\times(m-n)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}_n & \boldsymbol{0}_{n\times(m-n)} \\ \boldsymbol{0}_{m-n} & \boldsymbol{0}_{(m-n)\times(m-n)} \end{bmatrix}$.

Thus we derive

$$\begin{aligned} \boldsymbol{A}\boldsymbol{A}^{+} &= \boldsymbol{U}\Sigma\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}\Sigma^{+}\boldsymbol{U}^{\mathrm{T}} \\ &= \boldsymbol{U}\Sigma\Sigma^{+}\boldsymbol{U}^{\mathrm{T}} = \boldsymbol{U}\begin{bmatrix} \boldsymbol{I}_n & \boldsymbol{0}_{n\times(m-n)} \\ \boldsymbol{0}_{m-n} & \boldsymbol{0}_{(m-n)\times(m-n)} \end{bmatrix}\boldsymbol{U}^{\mathrm{T}} \end{aligned}$$

We write $\boldsymbol{U}$ as block matrix:

$$\boldsymbol{U} = \begin{bmatrix} \boldsymbol{U}_1 & \boldsymbol{U}_2 \end{bmatrix}$$

where $\boldsymbol{U}_1$ is $m \times n$ matrix, $\boldsymbol{U}_2$ is $m \times (m-n)$ matrix.

Hence we derive

$$
\begin{aligned}
AA^+ &= U \begin{bmatrix} I_n & 0_{n\times(m-n)} \\ 0_{m-n} & 0_{(m-n)\times(m-n)} \end{bmatrix} U^{\mathrm{T}} \\
&= \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} I_n & 0_{n\times(m-n)} \\ 0_{m-n} & 0_{(m-n)\times(m-n)} \end{bmatrix} \begin{bmatrix} U_1^{\mathrm{T}} \\ U_2^{\mathrm{T}} \end{bmatrix} \\
&= \begin{bmatrix} U_1 I_n U_1^{\mathrm{T}} & 0_{n\times(m-n)} \\ 0_{m-n} & 0_{(m-n)\times(m-n)} \end{bmatrix} \\
&= \begin{bmatrix} I_n & 0_{n\times(m-n)} \\ 0_{m-n} & 0_{(m-n)\times(m-n)} \end{bmatrix} \qquad \text{due to the orthogonality of } U.
\end{aligned}
$$

Moreover, $A^+A = V\Sigma^+U^{\mathrm{T}}U\Sigma V^{\mathrm{T}} = V\Sigma^+\Sigma V^{\mathrm{T}}$.

You can verify by yourself that $\Sigma^+\Sigma = I$.

Hence $A^+A = VV^{\mathrm{T}} = I_n$.

(b) We only need to show $A^{\mathrm{T}}AA^+b = A^{\mathrm{T}}b$.

Since $\mathrm{rank}(A) = n$, the columns of A are ind. Hence $A^{\mathrm{T}}A$ is invertible.

- Firstly, we show $(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ is the right inverse of A:

$$
\begin{aligned}
A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}} &= U\Sigma V^{\mathrm{T}}(V\Sigma U^{\mathrm{T}}U\Sigma V^{\mathrm{T}})^{-1}V\Sigma U^{\mathrm{T}} \\
&= U\Sigma V^{\mathrm{T}}(V\Sigma^2 V^{\mathrm{T}})^{-1}V\Sigma U^{\mathrm{T}} = U\Sigma V^{\mathrm{T}}V\Sigma^{-2}V^{\mathrm{T}}V\Sigma U^{\mathrm{T}} \\
&= U\Sigma\Sigma^{-2}\Sigma U^{\mathrm{T}} \\
&= I
\end{aligned}
$$

- Since we also obtain $A^+A = I$, we derive

$$
A^+ = A^+I = A^+A\left[(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\right] = I\left[(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\right] = \left[(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\right]
$$

Thus we have $A^{\mathrm{T}}AA^+ = A^{\mathrm{T}} \implies A^{\mathrm{T}}AA^+b = A^{\mathrm{T}}b$.

■

3. *Proof.* (a)

$$\begin{aligned}
\|\boldsymbol{A}\|_F^2 &= \operatorname{trace}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) \\
&= \operatorname{trace}\left[\sum_{i=1}^{n}\sigma_i\boldsymbol{v}_i\boldsymbol{u}_i^{\mathrm{T}} \times \sum_{i=1}^{n}\sigma_i\boldsymbol{u}_i\boldsymbol{v}_i^{\mathrm{T}}\right] \\
&= \operatorname{trace}\left(\sum_{i=1}^{n}\sigma_i^2\boldsymbol{v}_i(\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{u}_i)\boldsymbol{v}_i^{\mathrm{T}} + \sum_{i\neq j}\sigma_i\sigma_j\boldsymbol{v}_i(\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{u}_j)\boldsymbol{v}_j^{\mathrm{T}}\right) \\
&= \operatorname{trace}\left(\sum_{i=1}^{n}\sigma_i^2\boldsymbol{v}_i\boldsymbol{v}_i^{\mathrm{T}} + \boldsymbol{0}\right) \qquad \text{due to orthogonality for } \boldsymbol{u}_i\text{'s and } \boldsymbol{v}_i\text{'s.} \\
&= \sum_{i=1}^{n}\sigma_i^2 \operatorname{trace}\left(\boldsymbol{v}_i\boldsymbol{v}_i^{\mathrm{T}}\right)
\end{aligned}$$

Suppose $\boldsymbol{v}_i = \begin{bmatrix} v_{1i} & v_{2i} & \dots & v_{ni}\end{bmatrix}^{\mathrm{T}}$, then due to the orthonormality of $\boldsymbol{v}_i$, we obtain

$$\operatorname{trace}\left(\boldsymbol{v}_i\boldsymbol{v}_i^{\mathrm{T}}\right) = \sum_{j=1}^{n} v_{ji}^2 = 1.$$

Hence $\|\boldsymbol{A}\|_F^2 = \sum_{i=1}^{n}\sigma_i^2 \operatorname{trace}\left(\boldsymbol{v}_i\boldsymbol{v}_i^{\mathrm{T}}\right) = \sum_{i=1}^{n}\sigma_i^2$.

(b)
- When $k < n$, it's obvious that

$$\boldsymbol{A}_k = \sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + \cdots + \sigma_k\boldsymbol{u}_k\boldsymbol{v}_k^{\mathrm{T}}.$$

Hence

$$\boldsymbol{A} - \boldsymbol{A}_k = \sigma_{k+1}\boldsymbol{u}_{k+1}\boldsymbol{v}_{k+1}^{\mathrm{T}} + \cdots + \sigma_n\boldsymbol{u}_n\boldsymbol{v}_n^{\mathrm{T}}.$$

And

$$\|\boldsymbol{A} - \boldsymbol{A}_k\|_F^2 = \operatorname{trace}\left(\sum_{i=k+1}^{n}\sigma_i\boldsymbol{v}_i\boldsymbol{u}_i^{\mathrm{T}} \times \sum_{i=k+1}^{n}\sigma_i\boldsymbol{u}_i\boldsymbol{v}_i^{\mathrm{T}}\right)$$

Similarly, we obtain

$$\|\boldsymbol{A} - \boldsymbol{A}_k\|_F^2 = \sum_{i=k+1}^{n}\sigma_i^2.$$

- Otherwise, $\boldsymbol{A}_k = \boldsymbol{A}$, thus $\|\boldsymbol{A} - \boldsymbol{A}_k\|_F^2 = 0$.

■

4. *Proof.* We only need to show that $\max_{\boldsymbol{x},\boldsymbol{y}}\|\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y}\|^2 = \sigma_1^2$:

- we find

$$
\begin{aligned}
\|\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y}\|^2 = \|\langle \boldsymbol{x}, \boldsymbol{A}\boldsymbol{y}\rangle\|^2 &\le \|\boldsymbol{x}\|^2 \cdot \|\boldsymbol{A}\boldsymbol{y}\|^2 \\
&= \|\boldsymbol{A}\boldsymbol{y}\|^2
\end{aligned}
$$

  The equality holds if and only if $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{y}$.

Thus

$$
\max_{\boldsymbol{x},\boldsymbol{y}} \|\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y}\|^2 = \max_{\boldsymbol{y}} \|\boldsymbol{A}\boldsymbol{y}\|^2 = \max_{\boldsymbol{y}} \boldsymbol{y}^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y}).
$$

We only need to show $\max_{\boldsymbol{y}} \boldsymbol{y}^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y}) = \sigma_1^2$:

- Since $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is real symmetric, there exists n orthogonal eigenvectors of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$. Moreover, we can divide these eigenvectors by their length to get n **orthonormal** eigenvectors $\boldsymbol{p}_1, \boldsymbol{p}_2, \dots, \boldsymbol{p}_n$ associated with eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ respectively.
  Without loss of generality, we set $\lambda_1 = \max_i \lambda_i$ for $i = 1, \dots, n$.
  Since they span $\mathbb{R}^n$, we can express arbitrary $\boldsymbol{y}$ as linear combination of $\boldsymbol{p}_1, \boldsymbol{p}_2, \dots, \boldsymbol{p}_n$:

$$
\boldsymbol{y} = \alpha_1\boldsymbol{p}_1 + \alpha_2\boldsymbol{p}_2 + \dots + \alpha_n\boldsymbol{p}_n.
$$

  Moreover, the product $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}$ is

$$
\begin{aligned}
\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y} &= \|\boldsymbol{y}\|^2 = 1 \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\alpha_j \boldsymbol{p}_i\boldsymbol{p}_j \\
&= \sum_{i=1}^{n} \alpha_i^2 = 1.
\end{aligned}
$$

- Moreover, the product $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y}$ is given by:

$$
\begin{aligned}
\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y} &= \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}(\alpha_1\boldsymbol{p}_1 + \alpha_2\boldsymbol{p}_2 + \dots + \alpha_n\boldsymbol{p}_n) \\
&= \alpha_1(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{p}_1) + \alpha_2(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{p}_2) + \dots + \alpha_n(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{p}_n) \\
&= \alpha_1\lambda_1\boldsymbol{p}_1 + \alpha_2\lambda_2\boldsymbol{p}_2 + \dots + \alpha_n\lambda_n\boldsymbol{p}_n
\end{aligned}
$$

Hence the product $\boldsymbol{y}^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y})$ is given by:

$$\begin{aligned}\boldsymbol{y}^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y}) &= \boldsymbol{y}^{\mathrm{T}}(\alpha_1\lambda_1\boldsymbol{p}_1 + \alpha_2\lambda_2\boldsymbol{p}_2 + \cdots + \alpha_n\lambda_n\boldsymbol{p}_n) \\ &= \left(\sum_{i=1}^{n}\alpha_i\boldsymbol{p}_i^{\mathrm{T}}\right)\left(\sum_{j=1}^{n}\alpha_j\lambda_j\boldsymbol{p}_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j\lambda_j\boldsymbol{p}_i^{\mathrm{T}}\boldsymbol{p}_j \\ &= \sum_{i=1}^{n}\alpha_i^2\lambda_i \\ &\le \lambda_1\sum_{i=1}^{n}\alpha_i^2 = \lambda_1.\end{aligned}$$

The equality is satisfied when $\boldsymbol{y} = \boldsymbol{p}_1$. Hence $\max_{\boldsymbol{y}}\boldsymbol{y}^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y}) = \lambda_1$. Since $\lambda_1 = \sigma_1^2$, we derive $\max_{\boldsymbol{y}}\boldsymbol{y}^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y}) = \sigma_1^2$.

■

5. *Proof.* • We do the eigendecomposition for $\boldsymbol{A}$:

$$\boldsymbol{A} = \boldsymbol{U}\Sigma\boldsymbol{U}^{\mathrm{T}}$$

where $\boldsymbol{U}$ is a $n \times n$ **orthogonal** matrix such that columns are eigenvectors of $\boldsymbol{A}^2$.

$\Sigma = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ is a $n \times n$ *diagonal* matrix, and $(\lambda_1, \dots, \lambda_n)$ are eigenvalues of $\boldsymbol{A}^2$.

And then we define $\sqrt{\Sigma} := \operatorname{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n})$. Obviously, we have $\Sigma = \sqrt{\Sigma}\sqrt{\Sigma}$.

Hence we could factorize $\boldsymbol{A}$ into

$$(\boldsymbol{U}\sqrt{\Sigma})(\boldsymbol{U}\sqrt{\Sigma})^{\mathrm{T}} = \boldsymbol{U}\sqrt{\Sigma}\sqrt{\Sigma}^{\mathrm{T}}\boldsymbol{U} = \boldsymbol{U}\Sigma\boldsymbol{U} = \boldsymbol{A}.$$

Thus we define $\boldsymbol{Q} := \boldsymbol{U}\sqrt{\Sigma}$, which means we can factorize $\boldsymbol{A}$ into $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}}$.

• Then we show the columns of $\boldsymbol{Q}$ are mutually orthogonal:

Suppose $\boldsymbol{U} = \begin{bmatrix} \boldsymbol{u}_1 & \dots & \boldsymbol{u}_n \end{bmatrix}$, and $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_n\}$ is orthonormal basis.

$$\boldsymbol{Q} = \boldsymbol{U}\sqrt{\Sigma} = \begin{bmatrix} \boldsymbol{u}_1 & \dots & \boldsymbol{u}_n \end{bmatrix} \begin{pmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{pmatrix} = \begin{bmatrix} \sqrt{\lambda_1}\boldsymbol{u}_1 & \sqrt{\lambda_2}\boldsymbol{u}_2 & \dots & \sqrt{\lambda_n}\boldsymbol{u}_n \end{bmatrix}$$

Since $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_n\}$ is orthonormal basis, we obtain:

$$\boldsymbol{u}_i \boldsymbol{u}_j = 0 \text{ for } i \neq j. \implies (\sqrt{\lambda_i}\boldsymbol{u}_i)(\sqrt{\lambda_j}\boldsymbol{u}_j) = 0 \text{ for } i \neq j.$$

which means columns of $\boldsymbol{Q}$ are mutually orthogonal.

■

10.2. Midterm Exam Solutions

10.2.1. Sample Exam Solution

1. (a)

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & c & 1 \\ 0 & -1 & 1 & 2 \\ 1 & 2 & 1 & -1 \end{bmatrix}$$

(b) The *augmented matrix* is given by

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

Then we compute its *row-reduced form:*

$$\left[\begin{array}{cccc|c} 1 & 1 & c & 1 & c \\ 0 & -1 & 1 & 2 & 0 \\ 1 & 2 & 1 & -1 & -c \end{array}\right] \xrightarrow[\text{Row 3=Row 3}-\text{Row 1}]{\text{Row 1=Row 1+Row 2}} \left[\begin{array}{cccc|c} 1 & 0 & c+1 & 3 & c \\ 0 & -1 & 1 & 2 & 0 \\ 0 & 1 & 1-c & -2 & -2c \end{array}\right]$$

$$\xrightarrow{\text{Row 2=Row 2}\times(-1)} \left[\begin{array}{cccc|c} 1 & 0 & c+1 & 3 & c \\ 0 & 1 & -1 & -2 & 0 \\ 0 & 1 & 1-c & -2 & -2c \end{array}\right]$$

$$\xrightarrow{\text{Row 3=Row 3}-\text{Row 2}} \left[\begin{array}{cccc|c} 1 & 0 & c+1 & 3 & c \\ 0 & 1 & -1 & -2 & 0 \\ 0 & 0 & 2-c & 0 & -2c \end{array}\right]$$

i. If $c = 2$, then we obtain:

$$\xrightarrow{\text{Row 3=Row 3}\times(-\frac{1}{4})} \left[\begin{array}{cccc|c} 1 & 0 & 3 & 3 & 2 \\ 0 & 1 & -1 & -2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right] \xrightarrow{\text{Row 1=Row 1}-2\times\text{Row 3}}$$

$$\left[\begin{array}{cccc|c} 1 & 0 & 3 & 3 & 0 \\ 0 & 1 & -1 & -2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right] \text{(rref)}$$

ii. Otherwise, we derive:

$$\xrightarrow{\text{Row 3=Row 3}\times(\frac{1}{2-c})} \left[\begin{array}{cccc|c} 1 & 0 & c+1 & 3 & c \\ 0 & 1 & -1 & -2 & 0 \\ 0 & 0 & 1 & 0 & \frac{2c}{c-2} \end{array}\right] \xrightarrow[\text{Row 2=Row 2+Row 3}]{\text{Row 1=Row 1}-\text{Row 3}\times(c+1)}$$

$$\left[\begin{array}{cccc|c} 1 & 0 & 0 & 3 & -\frac{c^2+4c}{c-2} \\ 0 & 1 & 0 & -2 & \frac{2c}{c-2} \\ 0 & 0 & 1 & 0 & \frac{2c}{c-2} \end{array}\right] \text{(rref)}$$

(c) i. If $c = 2$, there is no solution to this system.

ii. Otherwise, we convert this system into:

$$\begin{cases} x_1 + 3x_4 = -\dfrac{c^2+4c}{c-2} \\ x_2 - 2x_4 = \dfrac{2c}{c-2} \\ x_3 = \dfrac{2c}{c-2} \end{cases} \implies \begin{cases} x_1 = -\dfrac{c^2+4c}{c-2} - 3x_4 \\ x_2 = \dfrac{2c}{c-2} + 2x_4 \\ x_3 = \dfrac{2c}{c-2} \end{cases}$$

Hence the complete set of solutions is given by

$$\boldsymbol{x}_{\text{complete}} = \begin{pmatrix} -\frac{c^2+4c}{c-2} - 3x_4 \\ \frac{2c}{c-2} + 2x_4 \\ \frac{2c}{c-2} \\ x_4 \end{pmatrix} = \begin{pmatrix} -\frac{c^2+4c}{c-2} \\ \frac{2c}{c-2} \\ \frac{2c}{c-2} \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} -3 \\ 2 \\ 0 \\ 1 \end{pmatrix}.$$

(d) i. If $c = 2$, obviously, the rref of $\boldsymbol{A}$ is

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

Hence $\text{rank}(\boldsymbol{A}) = 2$.

ii. Otherwise, the rref of $\boldsymbol{A}$ is

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

Hence $\text{rank}(\boldsymbol{A}) = 3$.

In conclusion, $\text{rank}(\boldsymbol{A}) = \begin{cases} 3, & c \neq 2; \\ 2, & c = 2. \end{cases}$

(e) When $c = 0$, the complete solution is given by:

$$\boldsymbol{x}_{\text{complete}} = x_4 \begin{pmatrix} -3 \\ 2 \\ 0 \\ 1 \end{pmatrix}.$$

where x_4 is a scalar.

Hence a basis for the subspace of solutions is $\left\{ \begin{pmatrix} -3 \\ 2 \\ 0 \\ 1 \end{pmatrix} \right\}$.

2. • For *skew symmetric* matrix, once the lower triangular part is determined, the whole matrix is immediately determined. For example, if we know $a_{ij} = m(i > j)$, then the corresponding upper triangular entry is $a_{ji} = -m$. Thus our basis is given by:

$$\{\boldsymbol{A}_{ij}\} \text{ for } 1 \leq j \leq i \leq n.$$

where the entries $a_{st}(1 \le s,t \le n)$ for $\boldsymbol{A}_{ij}$ is given by

$$a_{st} = \begin{cases} 0, & (s,t) \neq (i,j) \text{ and } (s,t) \neq (j,i); \\ 1, & (s,t) = (i,j); \\ -1, & (s,t) = (j,i). \end{cases}$$

- Notice $ax^2 + bx + 2a + 3b = a(x^2+2) + b(x+3)$. And (x^2+2) and $(x+3)$ are obviously independent. Hence the basis is given by

$$\{(x^2+2), (x+3)\}.$$

- Firstly we show that $(x-1), (x+1), (2x^2-2)$ are independent:

$$\alpha_1(x-1) + \alpha_2(x+1) + \alpha_3(2x^2-2) = 0 \implies$$
$$2\alpha_3 x^2 + (\alpha_1+\alpha_2)x + (-\alpha_1+\alpha_2-2\alpha-3) = 0.$$

Hence we derive

$$\begin{cases} 2\alpha_3 = 0 \\ \alpha_1+\alpha_2 = 0 \\ -\alpha_1+\alpha_2-2\alpha_3 = 0 \end{cases} \implies \begin{cases} \alpha_1 = 0 \\ \alpha_2 = 0 \\ \alpha_3 = 0 \end{cases}$$

which means $(x-1), (x+1), (2x^2-2)$ are independent.

Hence one basis for this space is $\{(x-1), (x+1), (2x^2-2)\}$.

3. (a) Obviously, the entrie of $\boldsymbol{D}$ is

$$d_{ij} = \begin{cases} d_{ii}, & i = j; \\ 0, & i \neq j. \end{cases}$$

We set $\boldsymbol{E} = \boldsymbol{AD}, \boldsymbol{F} = \boldsymbol{DA}$. Hence the entries for $\boldsymbol{E}$ and $\boldsymbol{F}$ is given by:

$$e_{ij} = \sum_{t=1}^{n} a_{it}d_{tj} = a_{ij}d_{jj} \qquad f_{ij} = \sum_{t=1}^{n} d_{it}a_{tj} = d_{ii}a_{ij}$$

where $1 \leq i, j \leq n$.

In order to let $\boldsymbol{E} = \boldsymbol{F}$, we must let $e_{ij} = f_{ij}$ for $\forall 1 \leq i, j \leq n$.

$$\implies a_{ij}d_{jj} = d_{ii}a_{ij} \implies a_{ij}(d_{jj} - d_{ii}) = 0.$$

Since $d_{ii} \neq d_{jj}$ for $\forall i \neq j$, we derive $d_{jj} - d_{ii} \neq 0$. Hence $a_{ij} = 0$ for $\forall i \neq j$.

Considering the case $i = j$, then $d_{jj} - d_{ii} = d_{ii} - d_{ii} = 0$. Thus the value of a_{ij} is undetermined.

In conclusion, $\boldsymbol{A}$ could be any diagonal matrix.

(b) • We construct $\boldsymbol{B}^{ij}$ such that the (i, j)th entry of $\boldsymbol{B}^{ij}$ is 1, other entries are all zero.

• We set $\boldsymbol{A}\boldsymbol{B}^{ij} = \boldsymbol{E}^{ij}; \boldsymbol{B}^{ij}\boldsymbol{A} = \boldsymbol{F}^{ij}$. Hence the entries for $\boldsymbol{E}^{ij}$ and $\boldsymbol{F}^{ij}$ is given by:

$$e^{ij}_{pq} = \sum_{t=1}^{n} a_{pt}b_{tq} \qquad f^{ij}_{pq} = \sum_{t=1}^{n} b_{pt}a_{tq}$$

where $1 \leq p, q \leq n$.

Since $\boldsymbol{AB} = \boldsymbol{BA}$ is always true for any matrix $\boldsymbol{B}$, we have $\boldsymbol{A}\boldsymbol{B}^{ij} = \boldsymbol{B}^{ij}\boldsymbol{A}$.

Hence $e^{ij}_{pq} = f^{ij}_{pq}$.

• For $q \neq i$, we have $e^{ii}_{iq} = \sum_{t=1}^{n} a_{it}b_{tq} = 0$ since $b_{tq} = 0$ for $\forall t = 1, 2, \ldots, n$.

Also, $f^{ii}_{iq} = \sum_{t=1}^{n} b_{it}a_{tq} = a_{iq}$.

Hence $0 = a_{iq}$ for $\forall q \neq i$.

• For $i \neq j$, we have $e^{ij}_{ij} = \sum_{t=1}^{n} a_{it}b_{tj} = a_{ii}b_{ij} = a_{ii}$ and $f^{ij}_{ij} = \sum_{t=1}^{n} b_{it}a_{tj} = b_{ij}a_{jj} = a_{jj}$.

Hence $a_{ii} = a_{jj}$.

So, $\boldsymbol{A}$ is diagonal and all the diagonal entries of $\boldsymbol{A}$ are equal. Hence $\boldsymbol{A} = c\boldsymbol{I}$ for some scalar c.

4. (a)

$$\left[\begin{array}{cc|cc} 5 & 4 & 1 & 0 \\ 4 & 5 & 0 & 1 \end{array}\right] \xrightarrow{\text{Row 2}=5\times\text{Row 2}-4\times\text{Row 1}} \left[\begin{array}{cc|cc} 5 & 4 & 1 & 0 \\ 0 & 9 & -4 & 5 \end{array}\right]$$

$$\xrightarrow{\text{Row 1}=9\times\text{Row 1}-4\times\text{Row 2}} \left[\begin{array}{cc|cc} 45 & 0 & 25 & -20 \\ 0 & 9 & -4 & 5 \end{array}\right] \xrightarrow[\text{Row 2}=\frac{1}{9}\times\text{Row 2}]{\text{Row 1}=\frac{1}{45}\times\text{Row 1}}$$

$$\left[\begin{array}{cc|cc} 1 & 0 & \frac{5}{9} & -\frac{4}{9} \\ 0 & 1 & -\frac{4}{9} & \frac{5}{9} \end{array}\right]$$

Hence the inverse of the matrix $\begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}$ is $\begin{bmatrix} \frac{5}{9} & -\frac{4}{9} \\ -\frac{4}{9} & \frac{5}{9} \end{bmatrix}$.

(b)

$$\left[\begin{array}{cc|cc} a & b & 1 & 0 \\ c & d & 0 & 1 \end{array}\right] \xrightarrow{\text{Row 2}=a\times\text{Row 2}-c\times\text{Row 1}} \left[\begin{array}{cc|cc} a & b & 1 & 0 \\ 0 & ad-bc & -c & a \end{array}\right]$$

$$\xrightarrow{\text{Row 1}=(ad-bc)\times\text{Row 1}-b\times\text{Row 2}} \left[\begin{array}{cc|cc} a(ad-bc) & 0 & ad & -ab \\ 0 & ad-bc & -c & a \end{array}\right]$$

i. If $ad-bc=0$, then this process cannot continue, which means the inverse of $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ doesn't exist.

ii. If $ad-bc\neq 0$, without loss of generality, we assume $a\neq 0$.

(If $a=0$, then c must be nonzero. Then we only need to set the second row as pivot row to proceed similarly.)

Thus we obtain:

$$\xrightarrow[\text{Row 2}=\frac{1}{ad-bc}\times\text{Row 2}]{\text{Row 1}=\frac{1}{a(ad-bc)}\times\text{Row 1}} \left[\begin{array}{cc|cc} 1 & 0 & \frac{d}{ad-bc} & \frac{-b}{ad-bc} \\ 0 & 1 & \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{array}\right]$$

Hence the inverse of the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is $\begin{bmatrix} \frac{d}{ad-bc} & \frac{-b}{ad-bc} \\ \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{bmatrix}$.

5. (a) We set $\boldsymbol{A}=\boldsymbol{I}-\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}$.

- Firstly, we find that $\boldsymbol{u}\in N(\boldsymbol{A})$:

$$\boldsymbol{A}\boldsymbol{u}=(\boldsymbol{I}-\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}})\boldsymbol{u}=\boldsymbol{u}-\boldsymbol{u}(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u})=\boldsymbol{u}-\boldsymbol{u}=\boldsymbol{0}.$$

Moreover, $c\boldsymbol{u} \in N(\boldsymbol{A})$, where c is a scalar.

Hence any elements that parallel to $\boldsymbol{u}$ is in $N(\boldsymbol{A})$.

- Secondly, $\forall x \in N(\boldsymbol{A})$, we notice:

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0} \implies (\boldsymbol{I} - \boldsymbol{u}\boldsymbol{u}^{\mathrm{T}})\boldsymbol{x} = \boldsymbol{x} - \boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{0} \implies \boldsymbol{x} = \boldsymbol{u}(\boldsymbol{u}^{\mathrm{T}}\boldsymbol{x}).$$

Since $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{x}$ is a scalar, $\boldsymbol{x}$ is parallel to $\boldsymbol{u}$.

In other words, any elements in $N(\boldsymbol{A})$ is parallel to $\boldsymbol{u}$.

In conclusion, $N(\boldsymbol{A}) = \text{span}\{\boldsymbol{u}\}$. Hence $\dim(N(\boldsymbol{A})) = 1$.

Hence $\text{rank}(\boldsymbol{A}) = n - \dim(N(\boldsymbol{A})) = n - 1$.

(b) We find that

$$\begin{aligned}\boldsymbol{P}^2 &= \boldsymbol{P}\\ \boldsymbol{P}^5 &= \boldsymbol{P}.\end{aligned}$$

Hence $\text{rank}(\boldsymbol{P}^2) = \text{rank}(\boldsymbol{P}) = n - 1; \text{rank}(\boldsymbol{P}^5) = \text{rank}(\boldsymbol{P}) = n - 1$.

(c) i. If $\boldsymbol{I} - \boldsymbol{x}\boldsymbol{y}^{\mathrm{T}} = \boldsymbol{0}$, (for example, $\boldsymbol{x} = \begin{bmatrix}1\end{bmatrix}, \boldsymbol{y} = \begin{bmatrix}1\end{bmatrix}$.) then $\text{rank}(\boldsymbol{I} - \boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}) = 0$.

ii. Otherwise, we set $\boldsymbol{A} = \boldsymbol{I} - \boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}$.

- Firstly, for $\forall \boldsymbol{v} \in N(\boldsymbol{A})$, we notice:

$$\boldsymbol{A}\boldsymbol{v} = (\boldsymbol{I} - \boldsymbol{x}\boldsymbol{y}^{\mathrm{T}})\boldsymbol{v} = \boldsymbol{0} \implies \boldsymbol{v} = \boldsymbol{x}(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{v}).$$

Since $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{v}$ is a scalar, $\boldsymbol{v}$ is parallel to $\boldsymbol{x}$.

In other words, any elements in $N(\boldsymbol{A})$ is parallel to $\boldsymbol{x}$.

- Secondly, we discuss whether $\boldsymbol{x}$ is in $N(\boldsymbol{A})$:

$$\boldsymbol{x} \in N(\boldsymbol{A}) \iff \boldsymbol{A}\boldsymbol{x} = (\boldsymbol{I} - \boldsymbol{x}\boldsymbol{y}^{\mathrm{T}})\boldsymbol{x} = \boldsymbol{0} \iff \boldsymbol{x} = \boldsymbol{x}(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}). \tag{10.8}$$

A. If $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x} = 1$, then condition (10.8) is satisfied, then $\boldsymbol{x}$ is in $N(\boldsymbol{A})$.

Moreover, $c\boldsymbol{x} \in N(\boldsymbol{A})$, where c is a scalar.

Hence any elements that parallel to $\boldsymbol{x}$ is in $N(\boldsymbol{A})$.

In this case, we derive $N(\boldsymbol{A}) = \text{span}\{\boldsymbol{x}\}$. Hence $\dim(N(\boldsymbol{A})) = 1$.

$\text{rank}(\boldsymbol{A}) = n - \dim(N(\boldsymbol{A})) = n - 1$.

B. Otherwise, then condition (10.8) is **not** satisfied, thus $\boldsymbol{x}$ is not in $N(\boldsymbol{A})$.

Obviously, $c\boldsymbol{x} \notin N(\boldsymbol{A})$ for $\forall$ **nonzero** scalar c.

Hence any nonzero elements that parallel to $\boldsymbol{x}$ is not in $N(\boldsymbol{A})$.

In this case, we derive $N(\boldsymbol{A}) = \{\boldsymbol{0}\}$. Hence $\dim(N(\boldsymbol{A})) = 0$. $\text{rank}(\boldsymbol{A}) = n - \dim(N(\boldsymbol{A})) = n$.

In conclusion,

- When $\boldsymbol{I} - \boldsymbol{x}\boldsymbol{y}^{\mathrm{T}} = \boldsymbol{0}$, $\text{rank}(\boldsymbol{I} - \boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}) = 0$.

- Otherwise,

$$\text{rank}(\boldsymbol{I} - \boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}) = \begin{cases} n & \boldsymbol{y}^{\mathrm{T}}\boldsymbol{x} \neq 1; \\ n-1 & \boldsymbol{y}^{\mathrm{T}}\boldsymbol{x} = 1. \end{cases}$$

6. (a) No.

Reason: $(\boldsymbol{A}+\boldsymbol{B})(\boldsymbol{A}-\boldsymbol{B}) = \boldsymbol{A}^2 - \boldsymbol{B}^2 + (\boldsymbol{B}\boldsymbol{A} - \boldsymbol{A}\boldsymbol{B})$.

But $(\boldsymbol{B}\boldsymbol{A} - \boldsymbol{A}\boldsymbol{B})$ cannot always be zero. For example,

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \qquad \boldsymbol{B} = \begin{bmatrix} -2 & 0 \\ 2 & 0 \end{bmatrix}.$$

$$\text{But} \qquad \boldsymbol{A}\boldsymbol{B} = \begin{bmatrix} -2 & 0 \\ -2 & 0 \end{bmatrix}, \quad \boldsymbol{B}\boldsymbol{A} = \begin{bmatrix} -2 & 0 \\ 2 & 0 \end{bmatrix}.$$

(b) False.

Reason: For example,

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \boldsymbol{B} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Although $\boldsymbol{A}$ and $\boldsymbol{B}$ are invertible, $\boldsymbol{A}+\boldsymbol{B}$ is not invertible:

$$\boldsymbol{A}+\boldsymbol{B}=\begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix}.$$

(c) True.

Reason: If f_1 and f_2 is in this set, then the linear combination of f_1 and f_2 is also in this set. Why?

For function $\alpha_1 f_1+\alpha_2 f_2$, where α_1,α_2 are scalars, we obtain:

$$\begin{aligned}\alpha_1 f_1+\alpha_2 f_2(1) &= \alpha_1 f_1(1)+\alpha_2 f_2(1)\\ &= \alpha_1\times 0+\alpha_2\times 0\\ &= 0.\end{aligned}$$

Hence $\alpha_1 f_1+\alpha_2 f_2$ is also in this set. Hence this set is a vector space.

(d) True.

Reason: If $\boldsymbol{A}$ and $\boldsymbol{B}$ are invertible, then for the product $\boldsymbol{AB}$, we find

$$\boldsymbol{ABB}^{-1}\boldsymbol{A}^{-1}=\boldsymbol{A}(\boldsymbol{BB}^{-1})\boldsymbol{A}^{-1}=\boldsymbol{AIA}^{-1}=\boldsymbol{I}.$$

Hence $\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$ is the inverse of $\boldsymbol{AB}$. Hence the product $\boldsymbol{AB}$ is invertible.

(e) False.

Don't mix up this statement with the proposition: *Row transforamtion doesn't change the row space*.

Actually, in most case, the two matrices that have the same *reduced row echelon form* have **different** *column space.*

For example,

$$\boldsymbol{A}=\begin{bmatrix}1 & 3 & 3 & 4\\ 2 & 6 & 9 & 7\\ -1 & -3 & 3 & 4\end{bmatrix}\xrightarrow{\text{Row transform}}\boldsymbol{U}=\begin{bmatrix}1 & 3 & 0 & -1\\ 0 & 0 & 1 & 1\\ 0 & 0 & 0 & 0\end{bmatrix}$$

they have the same *reduced row echelon form*. However, the first column of $\boldsymbol{A}$ is $\begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} \notin \operatorname{col}(\boldsymbol{U})$. They have **different** *column space.*

(f) True.

Reason: Suppose $\boldsymbol{A}$ is $n \times n$ square matrix, if two columns of $\boldsymbol{A}$ are the same, then $\dim(\operatorname{col}(\boldsymbol{A})) = \operatorname{rank}(\boldsymbol{A}) < n$. Since $\boldsymbol{A}$ is not *full rank*, $\boldsymbol{A}$ *cannot* be invertible.

(g) False.

Don't mix up this statement with the equality:

$$\operatorname{rank}(\boldsymbol{A}) + \dim(N(\boldsymbol{A})) = n.$$

Actually, $\operatorname{rank}(\boldsymbol{A}) = \dim(\operatorname{row}(\boldsymbol{A})) = \dim(\operatorname{col}(\boldsymbol{A}))$.

10.2.2. Midterm Exam Solution

1. (a) We can write this system as:

$$\begin{cases} x - y + 3z = 1 \\ 2x + y = 5 \\ -x - 5y + 9z = -7 \end{cases}$$

We can convert it into matrix form:

$$\begin{bmatrix} 1 & -1 & 3 \\ 2 & 1 & 0 \\ -1 & -5 & 9 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \\ -7 \end{bmatrix}.$$

(b) The *augmented matrix* is given by:

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

And we perform *row transformation* on this matrix:

$$\left[\begin{array}{ccc|c} 1 & -1 & 3 & 1 \\ 2 & 1 & 0 & 5 \\ -1 & -5 & 9 & -7 \end{array}\right] \xrightarrow[\text{Row 3=Row 3+Row 1}]{\text{Row 2=Row 2}-2\times\text{Row 1}} \left[\begin{array}{ccc|c} 1 & -1 & 3 & 1 \\ 0 & 3 & -6 & 3 \\ 0 & -6 & 12 & -6 \end{array}\right] \xrightarrow{\text{Row 3=Row 3+2}\times\text{Row 2}}$$

$$\left[\begin{array}{ccc|c} 1 & -1 & 3 & 1 \\ 0 & 3 & -6 & 3 \\ 0 & 0 & 0 & 0 \end{array}\right] \xrightarrow{\text{Row 1=Row 1+}\frac{1}{3}\times\text{Row 2}} \left[\begin{array}{ccc|c} 1 & 0 & 1 & 2 \\ 0 & 3 & -6 & 3 \\ 0 & 0 & 0 & 0 \end{array}\right] \xrightarrow{\text{Row 2=Row 2}\times\frac{1}{3}}$$

$$\left[\begin{array}{ccc|c} 1 & 0 & 1 & 2 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right] \text{(rref)}$$

The reduced row echelon form of the augmented matrix for this system is

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

(c) We convert this system into:

$$\begin{cases} x+z=2 \\ y-2z=1 \end{cases} \implies \begin{cases} x=2-z \\ y=1+2z \end{cases}$$

Hence the complete set of solutions is given by

$$\boldsymbol{x}_{\text{complete}} = \begin{pmatrix} 2-z \\ 1+2z \\ z \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} + z \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}.$$

(d)

$$\boldsymbol{A} = \begin{bmatrix} 1 & -1 & 3 \\ 2 & 1 & 0 \\ -1 & -5 & 9 \end{bmatrix}$$

From part (b), we know that $\boldsymbol{A}$ is *singular*. Hence $\boldsymbol{A}^{-1}$ doesn't exist.

(e) From part (b), we know that $\boldsymbol{A}$ has 2 *pivot variables.* Hence $\text{rank}(\boldsymbol{A}) = 2$.

2. (a) The *coefficient matrix* for this equation is given by:

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

Hence x_1 is *pivot variable*, x_2, x_3, x_4 are *free variables.*

Moreover, $2x_1 - x_2 + 3x_3 = 0 \implies x_1 = \frac{x_2 - 3x_3}{2}$.

Hence the complete set of solutions is given by

$$\boldsymbol{x}_{\text{complete}} = \begin{pmatrix} \frac{x_2-3x_3}{2} \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = x_2 \begin{pmatrix} \frac{1}{2} \\ 1 \\ 0 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} -\frac{3}{2} \\ 0 \\ 1 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

(b) Obviously, the three vectors $\begin{pmatrix} \frac{1}{2} \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -\frac{3}{2} \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$ are ind.

Hence one basis for $\boldsymbol{V}$ is $\left\{ \begin{pmatrix} \frac{1}{2} \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -\frac{3}{2} \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right\}$.

Hence $\dim(\boldsymbol{V}) = 3$.

(c) The columns of $\boldsymbol{A}$ form a basis for $\boldsymbol{A}$.

Hence one matrix $\boldsymbol{A}$ is given by:

$$\boldsymbol{A} = \begin{bmatrix} \frac{1}{2} & -\frac{3}{2} & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

(d) We only need to find $\boldsymbol{B}$ such that

$$\boldsymbol{Bx} = \boldsymbol{0} \quad \text{where } \boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}.$$

Thus one possible matrix is $\boldsymbol{B} = \begin{bmatrix} 4 & -2 & 6 & 0 \end{bmatrix}$.

In this case, $\boldsymbol{Bx} = 2(2x_1 - x_2 + 3x_3) = 0$.

3. (a) $\boldsymbol{B} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$.

Verify: In this case, $\boldsymbol{B} = 2\boldsymbol{I}$.

Thus $\boldsymbol{BA} = 2\boldsymbol{IA} = 2\boldsymbol{A}$. for every $\boldsymbol{A}$.

(b) $\boldsymbol{B} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

Verify: In this case, $\boldsymbol{BA} = \boldsymbol{0A} = \boldsymbol{0}; 2\boldsymbol{B} = \boldsymbol{0}$.

Hence $\boldsymbol{BA} = 2\boldsymbol{B}$ for every $\boldsymbol{A}$.

(c) $\boldsymbol{B} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$.

Verify: In this case, $\boldsymbol{B}$ is an *elementary matrix*. It interchanges the first and the last rows of $\boldsymbol{A}$.

(d) Such $\boldsymbol{B}$ doesn't exist.

Reason: Suppose $\boldsymbol{A} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$, then $\boldsymbol{BA} = \begin{bmatrix} c & b & a \\ f & e & d \\ i & h & g \end{bmatrix}$.

However, if the first row of $\boldsymbol{B}$ is $\begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \end{bmatrix}$, then the $(1,1)$th entry of $\boldsymbol{BA}$ is

$$\alpha_1 a + \alpha_2 d + \alpha_3 g,$$

which makes it impossible to equal to c.

Hence such $\boldsymbol{B}$ doesn't exist.

4. (a) i. • *Sufficiency.* If there exists an $n \times m$ matrix $\boldsymbol{C}$ such that $\boldsymbol{AC} = \boldsymbol{I}_m$, then for $\forall \boldsymbol{b} \in \mathbb{R}^m$ we obtain:

$$\boldsymbol{ACb} = \boldsymbol{I}_m \boldsymbol{b} = \boldsymbol{b}.$$

If we set $\boldsymbol{x}_0 = \boldsymbol{Cb}$, then we derive $\boldsymbol{Ax}_0 = \boldsymbol{b}$. Hence $\boldsymbol{x}_0$ is one solution to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, which means $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ has at least one solution for $\forall \boldsymbol{b} \in \mathbb{R}^m$.

- *Necessity.* If $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ has at least one solution for $\forall \boldsymbol{b} \in \mathbb{R}^m$, then we construct $\boldsymbol{b} = \boldsymbol{e}_i$ for $i = 1, 2, \dots, m$.

  For $\forall i \in \{1, 2, \dots, m\}$, there exists $\boldsymbol{x}_i$ such that $\boldsymbol{A}\boldsymbol{x}_i = \boldsymbol{e}_i$.

  Thus we construct $\boldsymbol{C} = \begin{bmatrix} \boldsymbol{x}_1 & \Big| & \boldsymbol{x}_2 & \Big| & \cdots & \Big| & \boldsymbol{x}_m \end{bmatrix}$. $\boldsymbol{C}$ is an $n \times m$ matrix and

$$\begin{aligned}
\boldsymbol{A}\boldsymbol{C} &= \boldsymbol{A} \begin{bmatrix} \boldsymbol{x}_1 & \Big| & \boldsymbol{x}_2 & \Big| & \cdots & \Big| & \boldsymbol{x}_m \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{A}\boldsymbol{x}_1 & \Big| & \boldsymbol{A}\boldsymbol{x}_2 & \Big| & \cdots & \Big| & \boldsymbol{A}\boldsymbol{x}_m \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{e}_1 & \Big| & \boldsymbol{e}_2 & \Big| & \cdots & \Big| & \boldsymbol{e}_m \end{bmatrix} = \boldsymbol{I}.
\end{aligned}$$

  Thus $\boldsymbol{C}$ is *right inverse* of $\boldsymbol{A}$.

ii. The rank of $\boldsymbol{A}$ is the number of *nonzero* rows in the rref($\boldsymbol{A}$).

The linear system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ always has solution for $\forall \boldsymbol{b}$. We convert it into *augmented matrix form:*

$$\begin{bmatrix} \boldsymbol{A} & \Big| & \boldsymbol{b} \end{bmatrix} \xRightarrow{\text{Row transform}} \begin{bmatrix} \text{rref}(\boldsymbol{A}) & \Big| & \boldsymbol{b}^* \end{bmatrix}$$

Once the rref($\boldsymbol{A}$) has zero rows and the corresponding $\boldsymbol{b}^*$ has nonzero entries, this system has no solution. Hence rref($\boldsymbol{A}$) has **no** zero rows.

Since $\boldsymbol{A}$ is a $m \times n$ matrix, we have m nonzero rows for $\boldsymbol{A}$.

Thus $\text{rank}(\boldsymbol{A}) = m$.

(b) • For 1×3 matrix $\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 7\pi \end{pmatrix}$, $\text{rank}(\boldsymbol{A}) = 1$.

And there exists $\boldsymbol{x}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ such that $\boldsymbol{A}\boldsymbol{x}_1 = \boldsymbol{e}_1$.

Hence we construct $\boldsymbol{C} = \begin{bmatrix} \boldsymbol{x}_1 \end{bmatrix}$. We find that $\boldsymbol{A}\boldsymbol{C} = \begin{pmatrix} 1 & 2 & 7\pi \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} =$

$1 = \boldsymbol{I}$. Hence $\boldsymbol{C} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ is the *right inverse* of $\boldsymbol{A}$.

- For 3×1 matrix $\boldsymbol{B} = \begin{pmatrix} 1 \\ 2 \\ 7\pi \end{pmatrix}$, we find $\text{rank}(\boldsymbol{B}) = 1 \neq 3$.

  From part (a) we derive $\boldsymbol{B}$ has no *right inverse*.

5. (a) No, let's raise a counter-example:

$$\boldsymbol{A} = \begin{bmatrix} 3 & 1 \\ 5 & 3 \end{bmatrix} \implies \text{rank}(\boldsymbol{A}) = 2.$$

$$\boldsymbol{A}^{\mathrm{T}} = \begin{bmatrix} 3 & 5 \\ 1 & 3 \end{bmatrix} \implies \boldsymbol{A} + \boldsymbol{A}^{\mathrm{T}} = \begin{bmatrix} 6 & 6 \\ 6 & 6 \end{bmatrix}$$

Hence $\text{rank}(\boldsymbol{A} + \boldsymbol{A}^{\mathrm{T}}) = 1 \neq 2 = \text{rank}(\boldsymbol{A})$.

(b)
- Firstly, we show $N(\boldsymbol{A}) \subset N(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})$:

  For any $\boldsymbol{x}_0 \in N(\boldsymbol{A})$, we have $\boldsymbol{A}\boldsymbol{x}_0 = \boldsymbol{0}$. Thus by postmultiplying $\boldsymbol{A}^{\mathrm{T}}$ we have $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}_0 = \boldsymbol{0}$. Hence $\boldsymbol{x}_0 \in N(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})$.

- Then we show $N(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) \subset N(\boldsymbol{A})$:

  For any $\boldsymbol{x}_0 \in N(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})$, we have $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}_0 = \boldsymbol{0}$. Thus by postmultiplying $\boldsymbol{x}_0^{\mathrm{T}}$ we have $\boldsymbol{x}_0^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}_0 = \boldsymbol{0}$, which implies $\|\boldsymbol{A}\boldsymbol{x}_0\|^2 = \boldsymbol{x}_0^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}_0 = \boldsymbol{0}$. Hence $\boldsymbol{A}\boldsymbol{x}_0 = \boldsymbol{0}$. Hence $\boldsymbol{x}_0 \in N(\boldsymbol{A})$.

In conclusion, $N(\boldsymbol{A}) = N(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})$.

(c)
- Since $\boldsymbol{A}$ is $m \times n$ matrix, then $\text{rank}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) + \dim(N(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})) = n = \text{rank}(\boldsymbol{A}) + \dim(N(\boldsymbol{A}))$.

- Since $N(\boldsymbol{A}) = N(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})$, we derive $\dim(N(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})) = \dim(N(\boldsymbol{A}))$.

Thus $\text{rank}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = \text{rank}(\boldsymbol{A})$.

6. (a) Verify by yourself that the following matrices are *symmetric*:

$$(i) \quad A^2 - B^2$$
$$(iii) \quad ABA$$

(b) There are *infinitely* many solutions.

Reason:

- Since A is 5×8matrix, $\text{rank}(A) + \dim(N(A)) = 8 \implies \dim(N(A)) = 3$.
  Hence this system $Ax = b$ has **special solutions**.
- Moverover, since $\text{rank}(A) = 5$, we have 5 nonzero pivots, which means $\text{rref}(A)$ has no zero rows.
  Hence this system $Ax = b$ always has **particular solution.**

In conclusion, there are *infinitely* many solutions.

(c) False.

Reason: For example, if we have

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \qquad B = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

then $A + B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, which is obviously *nonsingular*.

(d) False.

Reason: For example, the set of 2×2 matrices with rank no more than $r = 1$ is **not** a vector space. Why?

$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, B = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ are both in this set since $\text{rank}(A) + \text{rank}(B) = 1$.

However, $A + B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ doesn't belong to this set since $\text{rank}(A + B) = 2$.

(e) False.

Reason: This set doesn't satisfy *vector addition rule* and *scalar multiplication rule.*

If f, g are both in this set, then $(f+g)(1) = f(1) + g(1) = 2 \neq 1$. Hence $f+g$ is not in this set.

Similarly, you can verify λf (λ is a scalar that not equal to 1) is not in this set.

Hence it cannot be a vector space.

10.3. Final Exam Solutions

10.3.1. Sample Exam Solution

1. (a) Since we have

$$\begin{aligned}
D(\sin x) &= 0\sin x + 1\cos x + 0\sin 2x + 0\cos 2x \\
D(\cos x) &= -1\sin x + 0\cos x + 0\sin 2x + 0\cos 2x \\
D(\sin 2x) &= 0\sin x + 0\cos x + 0\sin 2x + 2\cos 2x \\
D(\cos 2x) &= 0\sin x + 0\cos x + (-2)\sin 2x + 0\cos 2x.
\end{aligned}$$

the matrix representation for the basis $\{\sin x, \cos x, \sin 2x, \cos 2x\}$ is given by

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

(b) • Firstly, we show $\{\sin x, \cos x, \sin 2x, \cos 2x\}$ are four eigenvectors of $\boldsymbol{D}^2$:

$$\begin{aligned}
\boldsymbol{D}^2(\sin x) &= \frac{\mathrm{d}^2}{\mathrm{d}x^2}(\sin x) = (-1) \times \sin x \\
\boldsymbol{D}^2(\cos x) &= \frac{\mathrm{d}^2}{\mathrm{d}x^2}(\cos x) = (-1) \times \cos x \\
\boldsymbol{D}^2(\sin 2x) &= \frac{\mathrm{d}^2}{\mathrm{d}x^2}(\sin 2x) = (-4) \times \sin 2x \\
\boldsymbol{D}^2(\cos 2x) &= \frac{\mathrm{d}^2}{\mathrm{d}x^2}(\cos 2x) = (-4) \times \cos 2x
\end{aligned}$$

• Secondly, we show $\{\sin x, \cos x, \sin 2x, \cos 2x\}$ are independent:

Given

$$\alpha_1 \sin x + \alpha_2 \cos x + \alpha_3 \sin 2x + \alpha_4 \cos 2x = 0$$

where α_i's are scalars for $i = 1, 2, 3, 4$.

- If we set $x=0$, then we derive:

$$0\alpha_1+\alpha_2+0\alpha_3+\alpha_4=0.$$

- If we set $x=\pi$, then we derive:

$$0\alpha_1-\alpha_2+0\alpha_3+\alpha_4=0.$$

- If we set $x=\frac{\pi}{2}$, then we derive:

$$\alpha_1+0\alpha_2+0\alpha_3-\alpha_4=0.$$

- If we set $x=\frac{\pi}{4}$, then we derive:

$$\frac{\sqrt{2}}{2}\alpha_1+\frac{\sqrt{2}}{2}\alpha_2+\alpha_3+0\alpha_4=0.$$

Solving the linear system of equations $\begin{cases} 0\alpha_1+\alpha_2+0\alpha_3+\alpha_4=0 \\ 0\alpha_1-\alpha_2+0\alpha_3+\alpha_4=0 \\ \alpha_1+0\alpha_2+0\alpha_3-\alpha_4=0 \\ \frac{\sqrt{2}}{2}\alpha_1+\frac{\sqrt{2}}{2}\alpha_2+\alpha_3+0\alpha_4=0. \end{cases}$, we derive

$$\alpha_1=\alpha_2=\alpha_3=\alpha_4=0.$$

Hence $\{\sin x,\cos x,\sin 2x,\cos 2x\}$ are independent.

In conclusion, $\{\sin x,\cos x,\sin 2x,\cos 2x\}$ are four *linearly independent* eigenvectors of $\boldsymbol{D}^2$.

2. (a) We only need to find *least squares solution* $\boldsymbol{x}^*$ to $\boldsymbol{L}\boldsymbol{x}=\boldsymbol{b}$, where

$$\boldsymbol{L}=\begin{bmatrix}1&1\\1&2\\1&3\end{bmatrix} \qquad \boldsymbol{x}=\begin{bmatrix}C\\D\end{bmatrix} \qquad \boldsymbol{b}=\begin{bmatrix}2\\1\\3\end{bmatrix}.$$

Take on trust that we only need to solve $\boldsymbol{L}^{\mathrm{T}}\boldsymbol{L}\boldsymbol{x} = \boldsymbol{L}^{\mathrm{T}}\boldsymbol{b}$.

$$\text{Since } \boldsymbol{L}^{\mathrm{T}}\boldsymbol{L} = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}, \qquad \boldsymbol{L}^{\mathrm{T}}\boldsymbol{b} = \begin{bmatrix} 6 \\ 13 \end{bmatrix}$$

We derive

$$\boldsymbol{x} = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix}^{-1} \begin{bmatrix} 6 \\ 13 \end{bmatrix} = \frac{1}{3 \times 14 - 6 \times 6} \begin{bmatrix} 14 & -6 \\ -6 & 3 \end{bmatrix} \begin{bmatrix} 6 \\ 13 \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 6 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{1}{2} \end{bmatrix}.$$

Thus the fit line is $y = 1 + \frac{1}{2}x$.

(b) The eigenvalue for $\boldsymbol{P}$ is $\lambda = 1$. when $\boldsymbol{A}$ is $m \times n$ matrix with $m > n$; the eigenvalues for $\boldsymbol{P}$ are $\lambda = 0$ or $\lambda = 1$ when $\boldsymbol{A}$ is square matrix.

Reason: Suppose $\boldsymbol{A}$ is $m \times n (m \geq n)$ matrix with $\operatorname{rank}(\boldsymbol{A}) = n$.

- Firstly we notice that $\boldsymbol{P}$ is *idemponent*:

$$\begin{aligned} \boldsymbol{P}^2 &= \left[\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\right]\left[\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\right] \\ &= \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}} \\ &= \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{P}. \end{aligned}$$

- Secondly, we show that the possible eigenvalues for $\boldsymbol{P}$ could only be 0 or 1:

  If λ is the eigenvalue for $\boldsymbol{P}$, then there exists **nonzero** $\boldsymbol{x} \in \mathbb{R}^{m \times 1}$ s.t.

$$\boldsymbol{P}\boldsymbol{x} = \lambda\boldsymbol{x}$$

  By postmultiplying $\boldsymbol{P}$ we derive

$$\boldsymbol{P}^2\boldsymbol{x} = \lambda\boldsymbol{P}\boldsymbol{x} \implies \boldsymbol{P}\boldsymbol{x} = \lambda\boldsymbol{P}\boldsymbol{x} \implies (\lambda - 1)(\boldsymbol{P}\boldsymbol{x}) = \boldsymbol{0}.$$

  Hence we derive that $\lambda = 1$ or $\boldsymbol{P}\boldsymbol{x} = \boldsymbol{0}$.

– If $\boldsymbol{P}\boldsymbol{x} = \boldsymbol{0}$, where $\boldsymbol{x} \in \mathbb{R}^{m\times 1}$ is a nonzero vector,

then by postmultiplying $\boldsymbol{A}^{\mathrm{T}}$ we obtain:

$$\boldsymbol{A}^{\mathrm{T}}\left[\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\right]\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{0} = \boldsymbol{0} \implies \boldsymbol{A}^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{0}.$$

Since $\boldsymbol{A}$ has *independent columns*, we obtain $\dim(\text{col}(\boldsymbol{A})) = \text{rank}(\boldsymbol{A}) = n$.

Thus $\text{rank}(\boldsymbol{A}^{\mathrm{T}}) = n$.

Since $\text{rank}(\boldsymbol{A}^{\mathrm{T}}) + \dim(N(\boldsymbol{A}^{\mathrm{T}})) = m$, we derive $N(\boldsymbol{A}^{\mathrm{T}}) = m - n$.

i. If $m > n$, then $N(\boldsymbol{A}^{\mathrm{T}}) > 0$, 0 could be eigenvalue for $\boldsymbol{P}$.

 * We can construct an eigenvector for $\boldsymbol{P}$ associated with eigenvalue $\lambda = 0$:

 For any nonzero $\boldsymbol{x} \in N(\boldsymbol{A}^{\mathrm{T}})$, we have

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{0}.$$

 By postmultiplying $\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}$ we derive

$$\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{0} \implies \boldsymbol{P}\boldsymbol{x} = \boldsymbol{0}.$$

 which means $\boldsymbol{x}$ is the eigenvalue for $\boldsymbol{P}$ associated with eigenvalue $\lambda = 0$.

ii. If $m = n$, then $N(\boldsymbol{A}^{\mathrm{T}}) = 0$, 0 cannot be eigenvalue for $\boldsymbol{P}$.

- Finally we construct an eigenvector for $\boldsymbol{P}$ associated with eigenvalue $\lambda = 1$:

 For any $\boldsymbol{t} \in \mathbb{R}^{n\times 1}$, we construct $\hat{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{t}$. Then we notice

$$\boldsymbol{P}\hat{\boldsymbol{x}} = \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{t} = \boldsymbol{A}\boldsymbol{t} = \hat{\boldsymbol{x}}$$

Hence $\lambda = 1$ must be the eigenvalue for $\boldsymbol{P}$.

In conclusion, for $m \times n$ matrix $\boldsymbol{A}(m \geq n)$,

- When $m = n$, the only possible eigenvalue for $\boldsymbol{P}$ is $\lambda = 1$.
- When $m \geq n$, the possible eigenvalues for $\boldsymbol{P}$ are $\lambda = 0$ or $\lambda = 1$.

3. (a) True.

Reason: For symmetric $\boldsymbol{A} \succ 0$, $\boldsymbol{A}$ has all positive eigenvalues.

- Firstly we show $\boldsymbol{A}$ is invertible:

  We assume there exists $\boldsymbol{x}_0 \neq \boldsymbol{0}$ that is in $N(\boldsymbol{A})$. In other words, there exists $\boldsymbol{x}_0 \neq \boldsymbol{0}$ such that

$$\boldsymbol{A}\boldsymbol{x}_0 = \boldsymbol{0}$$

  which means 0 is the eigenvalue for $\boldsymbol{A}$. Since $\boldsymbol{A}$ has all positive eigenvalues, it makes a contradiction.

  Hence $N(\boldsymbol{A}) = \{\boldsymbol{0}\}$, $\boldsymbol{A}$ is invertible.

- Secondly, we show $\boldsymbol{A}^{-1} \succ 0$:

  Since $\boldsymbol{A} \succ 0$, $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} > 0$ for $\forall$ nonzero $\boldsymbol{x}$.

  We define $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, obviously, $\text{range}(\boldsymbol{A}) = \mathbb{R}^n - N(\boldsymbol{A}) = \mathbb{R}^n - \{\boldsymbol{0}\}$.

  Hence $\boldsymbol{y}$ also denotes arbitrary nonzero vector in $\mathbb{R}^n$. And

$$\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{y} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} > 0.$$

  Equivalently, $\boldsymbol{A}^{-1} \succ 0$.

(b) False.

Reason: Let me raise a counter example:

For $\boldsymbol{A} = \begin{bmatrix} 1 & i \\ 0 & 0 \end{bmatrix}$, $\boldsymbol{x} = \begin{bmatrix} -i \\ 1 \end{bmatrix}$ is in $N(\boldsymbol{A})$, $\boldsymbol{y} = \boldsymbol{A}^{\mathrm{T}} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ i \end{bmatrix}$ is in $C(\boldsymbol{A}^{\mathrm{T}})$.

But the inner product of $\boldsymbol{x}$ and $\boldsymbol{y}$ is not zero:

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{y}^{\mathrm{H}}\boldsymbol{x} = \begin{bmatrix} 1 & -i \end{bmatrix} \begin{bmatrix} -i \\ 1 \end{bmatrix} = -2i \neq 0.$$

Hence $\boldsymbol{x}$ and $\boldsymbol{y}$ are not perpendicular.

(c) True.

Reason: If $\text{rank}(\boldsymbol{A}) = 0$, then $\dim(\text{col}(\boldsymbol{A})) = 0$. However, any vector space with zero dimension could only be the space $\{\boldsymbol{0}\}$.

Hence the column space of $\boldsymbol{A}$ is $\{0\}$, which means all columns of $\boldsymbol{A}$ are $\boldsymbol{0}$. Hence all elements of $\boldsymbol{A}$ are 0. Thus $\boldsymbol{A} = \boldsymbol{0}$.

(d) True.

Reason: For $\forall \boldsymbol{x} \in N(\boldsymbol{A})$ and $\forall \boldsymbol{y} \in C(\boldsymbol{A}^{\mathrm{T}})$, there exists vector $\boldsymbol{u}$ such that $\boldsymbol{y} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}$.

Thus we derive

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = \langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{x}, \boldsymbol{A}^{\mathrm{T}}\boldsymbol{u} \rangle = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{0} = 0.$$

(e) True.

Reason: We do the eigendecomposition for $\boldsymbol{A}$ and $\boldsymbol{B}$:

$$\boldsymbol{A} = \boldsymbol{U}_1\Sigma_1\boldsymbol{U}_1^{\mathrm{T}} \qquad \boldsymbol{B} = \boldsymbol{U}_2\Sigma_2\boldsymbol{U}_2^{\mathrm{T}}$$

where $\boldsymbol{U}_1, \boldsymbol{U}_2$ are both orthogonal matrix.

Then we define $\boldsymbol{U} := \begin{bmatrix} \boldsymbol{U}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{U}_2 \end{bmatrix}$, we find that

$$\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U} = \begin{bmatrix} \boldsymbol{U}_1^{\mathrm{T}}\boldsymbol{U}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{U}_2^{\mathrm{T}}\boldsymbol{U}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} = \boldsymbol{I}.$$

Hence $\boldsymbol{U}$ is a matrix with orthonormal columns. Moreover, $\boldsymbol{U}$ is a square matrix. Hence it is a orthogonal matrix.

And we find that $\begin{bmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{B} \end{bmatrix}$ could be decomposed as

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{B} \end{bmatrix} = \begin{bmatrix} \boldsymbol{U}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{U}_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & \\ & \Sigma_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{U}_1^{\mathrm{T}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{U}_2^{\mathrm{T}} \end{bmatrix} = \boldsymbol{U} \begin{bmatrix} \Sigma_1 & \\ & \Sigma_2 \end{bmatrix} \boldsymbol{U}^{\mathrm{T}}$$

Hence $\begin{bmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{B} \end{bmatrix}$ is also diagonalizable.

4. (a) We set $\boldsymbol{u}_k = \begin{bmatrix} y_k \\ z_k \end{bmatrix}$. The rule

$$\begin{cases} y_{k+1} = 0.8y_k + 0.3z_k \\ z_{k+1} = 0.2y_k + 0.7z_k \end{cases}$$

can be written as $\boldsymbol{u}_{k+1} = \begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix} \boldsymbol{u}_k$. And $\boldsymbol{u}_0 = \begin{bmatrix} 0 \\ 5 \end{bmatrix}$.

We set $\boldsymbol{A} = \begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix}$ and $\boldsymbol{D} = \begin{bmatrix} 0.5 & 0 \\ 0 & 1 \end{bmatrix}$.

- In order to show $\boldsymbol{A}$ and $\boldsymbol{D}$ are similar, we construct our $\boldsymbol{S}$ such that

$$\boldsymbol{AS} = \boldsymbol{SD}$$

We set $\boldsymbol{S} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, then $\boldsymbol{AS} = \boldsymbol{SD}$ can be written as:

$$\begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{bmatrix} 0.5 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\implies \begin{bmatrix} 0.8a + 0.3c & 0.8b + 0.3d \\ 0.2a + 0.7c & 0.2b + 0.7d \end{bmatrix} = \begin{bmatrix} 0.5a & b \\ 0.5c & d \end{bmatrix}.$$

The linear system of equation could be converted as

$$\begin{cases} 0.8a + 0.3c = 0.5a \\ 0.8b + 0.3d = b \\ 0.2a + 0.7c = 0.5c \\ 0.2b + 0.7d = d \end{cases} \implies \begin{cases} a + c = 0 \\ 2b - 3d = 0 \end{cases}$$

If we set $a = 1, b = 3$, we get $c = -1, d = 2$.

Thus $\boldsymbol{S} = \begin{bmatrix} 1 & 3 \\ -1 & 2 \end{bmatrix}$ is one special solution.

Thus $\boldsymbol{AS} = \begin{bmatrix} 0.5 & 3 \\ -0.5 & 2 \end{bmatrix} = \boldsymbol{SD} \implies \boldsymbol{A} = \boldsymbol{SDS}^{-1}$. Hence $\boldsymbol{A}$ is similar to $\boldsymbol{D}$.

- And then we can compute $\boldsymbol{A}^k$:

$$\begin{aligned}
\boldsymbol{A}^k &= (\boldsymbol{SDS}^{-1})^k \\
&= \boldsymbol{SD}^k\boldsymbol{S}^{-1} \\
&= \begin{bmatrix} 1 & 3 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 0.5 & 0 \\ 0 & 1 \end{bmatrix}^k \begin{bmatrix} 1 & 3 \\ -1 & 2 \end{bmatrix}^{-1} \\
&= \begin{bmatrix} 1 & 3 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 0.5^k & 0 \\ 0 & 1 \end{bmatrix} \frac{1}{5} \begin{bmatrix} 2 & -3 \\ 1 & 1 \end{bmatrix} \\
&= \frac{1}{5} \begin{bmatrix} 2 \times (\frac{1}{2})^k + 3 & (-3) \times (\frac{1}{2})^k + 3 \\ (-2) \times (\frac{1}{2})^k + 2 & 3 \times (\frac{1}{2})^k + 2 \end{bmatrix}.
\end{aligned}$$

- Hence by induction, $\boldsymbol{u}_k = \boldsymbol{A}^k\boldsymbol{u}_0 = \frac{1}{5} \begin{bmatrix} 2 \times (\frac{1}{2})^k + 3 & (-3) \times (\frac{1}{2})^k + 3 \\ (-2) \times (\frac{1}{2})^k + 2 & 3 \times (\frac{1}{2})^k + 2 \end{bmatrix} \begin{bmatrix} 0 \\ 5 \end{bmatrix} = \begin{bmatrix} (-3) \times (\frac{1}{2})^k + 3 \\ 3 \times (\frac{1}{2})^k + 2 \end{bmatrix}$.

The general formula for y_k and z_k is $\begin{cases} y_k = (-3) \times (\dfrac{1}{2})^k + 3 \\ z_k = 3 \times (\dfrac{1}{2})^k + 2 \end{cases}$.

Thus $\begin{cases} \lim\limits_{k \to \infty} y_k = 3 \\ \lim\limits_{k \to \infty} z_k = 2 \end{cases}$.

(b) For real symmetric matrix $\boldsymbol{D} = \begin{bmatrix} 0.5 & 0 \\ 0 & 1 \end{bmatrix}$, SVD decomposition is just eigen-decomposition.

Obviously, the eigenvalues for $\boldsymbol{D}$ is $\lambda_1 = 0.5, \lambda_2 = 1$.

- When $\lambda = 0.5$, one eigenvector for $\boldsymbol{D}$ is $\boldsymbol{x}_1 = \begin{bmatrix} 1 & 0 \end{bmatrix}$.

- When $\lambda = 1$, one eigenvector for $\boldsymbol{D}$ is $\boldsymbol{x}_2 = \begin{bmatrix} 0 & 1 \end{bmatrix}$.

Hence we construct $\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

$\boldsymbol{D}$ has the factorization

$$\boldsymbol{D} = \boldsymbol{Q} \begin{pmatrix} 0.5 & \\ & 1 \end{pmatrix} \boldsymbol{Q}^{\mathrm{T}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0.5 & \\ & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

5. We do the eigendecomposition for $\boldsymbol{A}$:

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{D}\boldsymbol{Q}^{\mathrm{T}}.$$

where $\boldsymbol{Q}$ is orthogonal matrix, $\boldsymbol{D}$ is diagonal matrix.
Then if we set $\boldsymbol{y} := \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{x}$, we find that

$$R(\boldsymbol{x}, \boldsymbol{A}) = \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}} = \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{D}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}} = \frac{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{y}}{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}} = R(\boldsymbol{y}, \boldsymbol{D})$$

Given any $\boldsymbol{A}$, we can always convert it into diagonal matrix $\boldsymbol{D}$. Hence without loss of generality, we set $\boldsymbol{A}$ is a diagonal matrix such that

$$\boldsymbol{A} = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n).$$

For diagonal matrix $\boldsymbol{A}$, we derive

$$R(\boldsymbol{x}, \boldsymbol{A}) = \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}} = \frac{\sum_{i=1}^{n} \lambda_i x_i^2}{\sum_{i=1}^{n} x_i^2}$$

(a)

$$\sum_{i=1}^{n}\lambda_i x_i^2 \geq \sum_{i=1}^{n}\lambda_1 x_i^2 = \lambda_1\sum_{i=1}^{n}x_i^2 \implies R(\boldsymbol{x},\boldsymbol{A}) = \frac{\sum_{i=1}^{n}\lambda_i x_i^2}{\sum_{i=1}^{n}x_i^2} \geq \lambda_1, \qquad \forall \boldsymbol{x} \neq 0.$$

When $\boldsymbol{x} = (1,0,0,\dots,0)$, we can get the equality.

(b) Firstly we compute the eigenvector $\boldsymbol{x}_1$ for $\boldsymbol{A}$ associated with λ_1:

$$(\lambda_1\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x}_1 = \boldsymbol{0} \implies \begin{pmatrix} 0 & & & \\ & \lambda_2-\lambda_1 & & \\ & & \ddots & \\ & & & \lambda_n - \lambda_1 \end{pmatrix}\boldsymbol{x}_1 = \boldsymbol{0} \implies \boldsymbol{x}_1 = \begin{pmatrix}\alpha \\ 0 \\ \vdots \\ 0\end{pmatrix}.$$

where α is a scalar.

Hence $\boldsymbol{y} \perp \boldsymbol{x} \implies \boldsymbol{y} = (0, y_2, \dots, y_n)$. i.e. the first element of $\boldsymbol{y}$ is zero.

$$\sum_{i=1}^{n}\lambda_i y_i^2 = \sum_{i=2}^{n}\lambda_i y_i^2 \geq \sum_{i=2}^{n}\lambda_2 y_i^2 = \lambda_2\sum_{i=2}^{n}y_i^2 \implies R(\boldsymbol{y},\boldsymbol{A}) = \frac{\sum_{i=1}^{n}\lambda_i y_i^2}{\sum_{i=1}^{n}y_i^2} \geq \lambda_2,$$

for $\forall \boldsymbol{y} \in \boldsymbol{x}_1^{\perp} - \{\boldsymbol{0}\}$.

When $\boldsymbol{y} = (0,1,0,\dots,0)$, we get the equality.

(c) For $\forall \boldsymbol{v} = (b_1, b_2, \dots, b_n)$, there exists $(\beta_1, \beta_2) \neq \boldsymbol{0}$ such that $(\beta_1, \beta_2) \perp (b_1, b_2)$.

Hence we construct $\boldsymbol{y}_* = (\beta_1, \beta_2, 0, 0, \dots, 0)$. Then

$$\boldsymbol{y}_*^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y} = \lambda_1\beta_1^2 + \lambda_2\beta_2^2 \leq \lambda_2(\beta_1^2 + \beta_2^2) = \lambda_2\boldsymbol{y}_*^{\mathrm{T}}\boldsymbol{y}_* \implies R(\boldsymbol{y}_*,\boldsymbol{A}) \leq \lambda_2.$$

Moreover, $\boldsymbol{y}_*^{\mathrm{T}}\boldsymbol{v} = 0$. Thus we derive

$$\min_{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{v}=0} R(\boldsymbol{y},\boldsymbol{A}) \leq R(\boldsymbol{y}_*,\boldsymbol{A}) \leq \lambda_2.$$

6. (a) Suppose $\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^3$, then

$$\begin{aligned}
\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Z}\boldsymbol{x} &= 5x_1^2 + 5x_2^2 + 7x_3^2 + 2x_1x_2 + 8x_1x_3 + 6x_2x_3 \\
&= (x_1^2 + x_2^2 + 2x_1x_2)_{(}4x_1^2 + 4x_3^2 + 8x_1x_3) + (3x_2^2 + 3x_3^2 + 6x_2x_3) \\
&= (x_1 + x_2)^2 + 4(x_1 + x_3)^2 + 3(x_2 + x_3)^2 + x_2^2 \\
&\geq 0.
\end{aligned}$$

Hence $\boldsymbol{Z} \succeq 0$.

(b) Suppose $\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^n$, then

$$\begin{aligned}
\boldsymbol{x}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{x} &= \sum_{i,j=1}^{n} M_{ij}x_ix_j = \sum_{i=1}^{n} M_{ii}x_i^2 + \sum_{j\neq i} M_{ij}x_ix_j \\
&= 2\sum_{1\leq i<j\leq n} M_{ij}x_ix_j + \sum_{i=1}^{n} M_{ii}x_i^2 \\
&= \sum_{1\leq i<j\leq n} (2M_{ij}x_ix_j + |M_{ij}|x_i^2 + |M_{ij}|x_j^2) - \sum_{1\leq i<j\leq n} (|M_{ij}|x_i^2 + |M_{ij}|x_j^2) + \sum_{i=1}^{n} M_i \\
&= \sum_{1\leq i<j\leq n} (2M_{ij}x_ix_j + |M_{ij}|x_i^2 + |M_{ij}|x_j^2) + \sum_{i=1}^{n} (M_{ii}x_i^2 - \sum_{j\neq i} |M_{ij}|)x_i^2
\end{aligned}$$

Notice that $(M_{ii}x_i^2 - \sum_{j\neq i} |M_{ij}|) \geq 0$ since $\boldsymbol{M}$ is diagonal dominant.

And if we define $\sigma_{ij} = \begin{cases} 1, M_{ij} \geq 0 \\ 0, M_{ij} < 0 \end{cases}$, then we obtain:

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{x} = \sum_{1\leq i<j\leq n} |M_{ij}|(x_i + \sigma_{ij}x_j)^2 + \sum_{i=1}^{n} (M_{ii}x_i^2 - \sum_{j\neq i} |M_{ij}|)x_i^2 \geq 0.$$

Hence $\boldsymbol{M} \succeq 0$.

10.3.2. Final Exam Solution

1. (a) For $\forall f, g \in \{$**polynomials of degree** $\leq 4\}$, we obtain:

   - $$T(f+g) = (x-2)\frac{\mathrm{d}}{\mathrm{d}x}(f+g) = (x-2)\frac{\mathrm{d}}{\mathrm{d}f} + (x-2)\frac{\mathrm{d}}{\mathrm{d}g} = T(f) + T(g)$$

   - $$T(cf) = (x-2)\frac{\mathrm{d}}{\mathrm{d}x}(cf) = c(x-2)\frac{\mathrm{d}}{\mathrm{d}f} = cT(f).$$

     where c is a scalar.

   Since T satisfies the vector addition and scalar multiplication rule, it is a linear transformation.

   Moreover, we obtain:

$$\begin{aligned}
T(1) &= (x-2)\frac{\mathrm{d}1}{\mathrm{d}x} = 0 \\
T(x) &= (x-2)\frac{\mathrm{d}x}{\mathrm{d}x} = x-2 \\
T(x^2) &= (x-2)\frac{\mathrm{d}x^2}{\mathrm{d}x} = 2x(x-2) = 2x^2 - 4x \\
T(x^3) &= (x-2)\frac{\mathrm{d}x^3}{\mathrm{d}x} = 3x^2(x-2) = 3x^3 - 6x^2 \\
T(x^4) &= (x-2)\frac{\mathrm{d}x^4}{\mathrm{d}x} = 4x^3(x-2) = 4x^4 - 8x^3.
\end{aligned}$$

   Hence the matrix representation is given by:

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

   (b) - For $f = 1$, $T(f) = (x-2)\frac{\mathrm{d}f}{\mathrm{d}x} = 0 = 0f$.

   Hence $f = 1$ is an eigenvector of T associated with eigenvalue $\lambda = 0$.

- For $f = x - 2$, $T(f) = (x-2)\frac{\mathrm{d}f}{\mathrm{d}x} = x - 2 = f$.

  Hence $f = x - 2$ is an eigenvector of T associated with eigenvalue $\lambda = 1$.

Moreover, we have $\alpha_1 \times (1) + \alpha_2 \times (x-2) = 0$, where α_1, α_2 are scalars, then we derive

$$x(\alpha_1 + \alpha_2) - 2\alpha_2 = 0. \implies \alpha_1 = \alpha_2 = 0.$$

Hence $(x-2)$ and 1 are independent.

Hence two independent eigenvectors of T are 1 and $(x-2)$.

2. (a) Firstly, we set $\boldsymbol{x} = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}, \boldsymbol{y} = \begin{bmatrix} -1 \\ -1 \\ 4 \end{bmatrix}$. Obviously, they are independent.

Hence $\{\boldsymbol{x}, \boldsymbol{y}\}$ is the basis for column space of matrix $\begin{bmatrix} 1 & -1 \\ 1 & -1 \\ -2 & 4 \end{bmatrix}$.

Then we convert $\{\boldsymbol{x}, \boldsymbol{y}\}$ into orthogonal basis $\{\boldsymbol{q}_1, \boldsymbol{q}_2\}$:

-
$$\boldsymbol{q}_1 = \boldsymbol{x}$$

-
$$\boldsymbol{q}_2 = \boldsymbol{y} - \mathrm{Proj}_{\boldsymbol{y}}(\boldsymbol{q}_1) = \boldsymbol{y} - \frac{\langle \boldsymbol{y}, \boldsymbol{x} \rangle}{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}\boldsymbol{x} = \begin{bmatrix} \frac{2}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix}.$$

- The projection of $\boldsymbol{z}$ onto the vector $\boldsymbol{q}_1$ is

$$\mathrm{Proj}_{\boldsymbol{q}_1}(\boldsymbol{z}) = \frac{\langle \boldsymbol{x}, \boldsymbol{z} \rangle}{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}\boldsymbol{x} = \begin{bmatrix} -\frac{1}{6} \\ -\frac{1}{6} \\ \frac{1}{3} \end{bmatrix}.$$

- The projection of $\boldsymbol{z}$ onto the vector $\boldsymbol{q}_2$ is

$$\text{Proj}_{\boldsymbol{q}_2}(\boldsymbol{z}) = \frac{\langle \boldsymbol{q}_2, \boldsymbol{z} \rangle}{\langle \boldsymbol{q}_2, \boldsymbol{q}_2 \rangle} \boldsymbol{q}_2 = \begin{bmatrix} \frac{2}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix}.$$

Hence the projection of $\boldsymbol{z}$ onto span$\{\boldsymbol{x}, \boldsymbol{z}\}$ is given by:

$$\text{Proj}_{\text{span}\{\boldsymbol{q}_1, \boldsymbol{q}_2\}}(\boldsymbol{z}) = \text{Proj}_{\boldsymbol{q}_1}(\boldsymbol{z}) + \text{Proj}_{\boldsymbol{q}_2}(\boldsymbol{z}) = \begin{bmatrix} -\frac{1}{6} \\ -\frac{1}{6} \\ \frac{1}{3} \end{bmatrix} + \begin{bmatrix} \frac{2}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 1 \end{bmatrix}$$

Hence the projection onto the column space of $\begin{bmatrix} 1 & -1 \\ 1 & -1 \\ -2 & 4 \end{bmatrix}$ is $\begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 1 \end{bmatrix}$.

(b) We construct an isomorphism from $\mathbb{R}^{2\times 2}$ to $\mathbb{R}^{4\times 1}$:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \mapsto \begin{bmatrix} a & b & c & d \end{bmatrix}^{\mathrm{T}}.$$

The matrix representation $\boldsymbol{A}$ for the mapping

$$\begin{bmatrix} a \\ b \end{bmatrix} \mapsto \begin{bmatrix} a+b & a-b & -2a+4b & 0 \end{bmatrix}^{\mathrm{T}}$$

is given by:

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ -2 & 4 \\ 0 & 0 \end{bmatrix}.$$

We define $K = \{\boldsymbol{A}\boldsymbol{x} | \boldsymbol{x} \in \mathbb{R}^{2\times 1}\}$.

Hence we only need to find the best approximation of $\boldsymbol{b} := \begin{bmatrix} 1 \\ 2 \\ 7 \\ 1 \end{bmatrix}$ in the space K.

We define $\boldsymbol{x} := \begin{bmatrix} 1 \\ 1 \\ -2 \\ 0 \end{bmatrix}, \boldsymbol{y} := \begin{bmatrix} 1 \\ -1 \\ 4 \\ 0 \end{bmatrix}$. Then we convert $\{\boldsymbol{x}, \boldsymbol{y}\}$ into orthogonal vectors:

- We set $\boldsymbol{q}_1 = \boldsymbol{x}$.

- We set $\boldsymbol{q}_2 = \boldsymbol{y} - \mathrm{Proj}_{\boldsymbol{q}_1}(\boldsymbol{y})$. Hence

$$\begin{aligned} \boldsymbol{q}_2 &= \boldsymbol{y} - \mathrm{Proj}_{\boldsymbol{q}_1}(\boldsymbol{y}) \\ &= \boldsymbol{y} - \frac{\langle \boldsymbol{q}_1, \boldsymbol{y} \rangle}{\langle \boldsymbol{q}_1, \boldsymbol{q}_1 \rangle} \boldsymbol{q}_1 \\ &= \begin{bmatrix} \frac{7}{3} & \frac{1}{3} & \frac{4}{3} & 0 \end{bmatrix}^{\mathrm{T}}. \end{aligned}$$

Hence the projection of $\boldsymbol{b}$ onto the space K is:

$$\begin{aligned}
\text{Proj}_{\text{span}\{\boldsymbol{x},\boldsymbol{y}\}}(\boldsymbol{b}) &= \text{Proj}_{\text{span}\{\boldsymbol{q}_1,\boldsymbol{q}_2\}}(\boldsymbol{b}) \\
&= \text{Proj}_{\boldsymbol{q}_1}(\boldsymbol{b}) + \text{Proj}_{\boldsymbol{q}_2}(\boldsymbol{b}) \\
&= \frac{\langle \boldsymbol{q}_1, \boldsymbol{b}\rangle}{\langle \boldsymbol{q}_1, \boldsymbol{q}_1\rangle}\boldsymbol{q}_1 + \frac{\langle \boldsymbol{q}_2, \boldsymbol{b}\rangle}{\langle \boldsymbol{q}_2, \boldsymbol{q}_2\rangle}\boldsymbol{q}_2 \\
&= -\frac{6}{11}\begin{bmatrix} 1 \\ 1 \\ -2 \\ 0 \end{bmatrix} + \frac{37}{66}\begin{bmatrix} 7 \\ 1 \\ 4 \\ 0 \end{bmatrix} \\
&= \frac{1}{11}\begin{bmatrix} 23 \\ -14 \\ 65 \\ 0 \end{bmatrix}.
\end{aligned}$$

Hence the best approximation for $\boldsymbol{b} = \begin{bmatrix} 1 \\ 2 \\ 7 \\ 1 \end{bmatrix}$ is $\frac{1}{11}\begin{bmatrix} 23 \\ -14 \\ 65 \\ 0 \end{bmatrix}$.

Correspondingly, the best approximation for $\boldsymbol{B} = \begin{bmatrix} 1 & 2 \\ 7 & 1 \end{bmatrix}$ is $\frac{1}{11}\begin{bmatrix} 23 & -14 \\ 65 & 0 \end{bmatrix}$.

3. (a) False.

   Reason: For example, if $\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, $\boldsymbol{A}^{-1}$ doesn't exist.

   (b) True.

   Reason: For orthogonal matrix $\boldsymbol{Q}$, we obtain $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{I}$. Thus

$$\det(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}) = \det(\boldsymbol{I}) \implies \det(\boldsymbol{Q}^{\mathrm{T}})\det(\boldsymbol{Q}) = \det(\boldsymbol{I}) \implies [\det(\boldsymbol{Q})]^2 = 1$$

   Hence $\det(\boldsymbol{Q}) = \pm 1$.

   (c) True.

Reason: For real symmetric $\boldsymbol{A}$, $-\boldsymbol{A}$ is PSD. $-\boldsymbol{A}$ could be diagonalized by orthogona matrix $\boldsymbol{P}$:

$$\boldsymbol{P}^{\mathrm{T}}(-\boldsymbol{A})\boldsymbol{P} = \boldsymbol{D} \Longleftrightarrow \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{\mathrm{T}} = -\boldsymbol{A}$$

where $\boldsymbol{D} = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$, where λ_i's are eigenvalues for $-\boldsymbol{A}$.
Since $-\boldsymbol{A} = -\boldsymbol{A}^{\mathrm{T}}$, we obtain

$$(-\boldsymbol{A})(-\boldsymbol{A})^{\mathrm{T}} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{\mathrm{T}} = \boldsymbol{P}\boldsymbol{D}^2\boldsymbol{P}^{\mathrm{T}}.$$

Or equivalently, $\boldsymbol{D}^2 = \boldsymbol{P}^{\mathrm{T}}(-\boldsymbol{A})(-\boldsymbol{A})^{\mathrm{T}}\boldsymbol{P}$. where the eigenvalues for $(-\boldsymbol{A})(-\boldsymbol{A})^{\mathrm{T}}$ are on the diagonal of $\boldsymbol{D}^2$.
This shows that if λ is the eigenvalue for $-\boldsymbol{A}$, then λ^2 is the eigenvalue for $(-\boldsymbol{A})(-\boldsymbol{A})^{\mathrm{T}} = \boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$.
Since $-\boldsymbol{A}$ is PSD, all eigenvalues of $-\boldsymbol{A}$ are positive. Hence $\lambda = \sqrt{\lambda^2}$.
If λ is the eigenvalue for $-\boldsymbol{A}$, then $-\lambda$ is the eigenvalue for $\boldsymbol{A}$. Hence the absolute value of eigenvalues for $\boldsymbol{A}$ are the same as the singular values for $\boldsymbol{A}$.

(d) False.
Reason: For example, $\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, then $P_{\boldsymbol{A}}(t) = \begin{vmatrix} t & -1 \\ 0 & t \end{vmatrix} = t^2$.

(e) True.
Reason: $\operatorname{rank}(\boldsymbol{A})$ =the smallest number of rank 1 matrices with sum $\boldsymbol{A}$. Hence $\operatorname{rank}(\boldsymbol{A}) \le 5$.

4. (a)

$$|\lambda\boldsymbol{I} - \boldsymbol{A}| = 0 \implies \begin{vmatrix} \lambda & 1 \\ -4 & \lambda \end{vmatrix} = 0 \implies \lambda^2 + 4 = 0.$$

Hence the eigenvalues for $\boldsymbol{A}$ are $\lambda_1 = 2i, \lambda_2 = -2i$.

- When $\lambda = 2i$, $(\lambda\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0} \implies \boldsymbol{x} = \alpha \begin{pmatrix} 1 \\ -2i \end{pmatrix}$, where α is a scalar.

- When $\lambda = -2i$, $(\lambda \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0} \implies \boldsymbol{x} = \beta \begin{pmatrix} 1 \\ 2i \end{pmatrix}$, where β is a scalar.

Hence $\alpha \begin{pmatrix} 1 \\ -2i \end{pmatrix}$ are eigenvectors of $\boldsymbol{A}$ associated with eigenvalue $\lambda = 2i$;

$\beta \begin{pmatrix} 1 \\ 2i \end{pmatrix}$ are eigenvectors of $\boldsymbol{A}$ associated with eigenvalue $\lambda = -2i$.

Moreover, $\boldsymbol{u} = \begin{pmatrix} 1 \\ 2i \end{pmatrix} + \begin{pmatrix} 1 \\ -2i \end{pmatrix}$.

(b) - Firstly, $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \begin{bmatrix} 16 & 0 \\ 0 & 1 \end{bmatrix}$. And we have

$$|\lambda \boldsymbol{I} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}| = \begin{vmatrix} \lambda - 16 & 0 \\ 0 & \lambda - 1 \end{vmatrix} = (\lambda - 16)(\lambda - 1) = 0 \implies \lambda_1 = 16, \lambda_2 = 1.$$

  - When $\lambda = 16$, $(\lambda \boldsymbol{I} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})\boldsymbol{x} = \boldsymbol{0} \implies \boldsymbol{x} = \alpha \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, where α is a scalar.
  - When $\lambda = 1$, $(\lambda \boldsymbol{I} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})\boldsymbol{x} = \boldsymbol{0} \implies \boldsymbol{x} = \beta \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, where β is a scalar.

  Hence $\boldsymbol{x}_1 = \alpha \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ are eigenvectors of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ associated with $\lambda_1 = 16$;

  $\boldsymbol{x}_2 = \beta \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ are eigenvectors of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ associated with $\lambda_2 = 1$.

  Hence $\Sigma = \text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}) = \text{diag}(4, 1)$.

  If we set $\alpha = 1, \beta = 1$, then $\boldsymbol{V} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

- Secondly, Since we have known $\boldsymbol{A} = \boldsymbol{U}\Sigma\boldsymbol{V}^{\mathrm{T}}$, we derive

$$\boldsymbol{U} = \boldsymbol{A}\boldsymbol{V}\Sigma^{-1} = \begin{bmatrix} 0 & -1 \\ 4 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{4} & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

In conclusion, our SVD decomposition is given by:

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^{\mathrm{T}}$$

5. (a) Suppose $\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S} = \boldsymbol{D}_1$, $\boldsymbol{S}^{-1}\boldsymbol{B}\boldsymbol{S} = \boldsymbol{D}_2$, where $\boldsymbol{D}_1, \boldsymbol{D}_2$ are diagonal matrices. Then equivalently,

$$\boldsymbol{A} = \boldsymbol{S}\boldsymbol{D}_1\boldsymbol{S}^{-1} \quad \boldsymbol{B} = \boldsymbol{S}\boldsymbol{D}_2\boldsymbol{S}^{-1}$$

Hence the product $\boldsymbol{A}\boldsymbol{B}$ is given by:

$$\begin{aligned} \boldsymbol{A}\boldsymbol{B} &= (\boldsymbol{S}\boldsymbol{D}_1\boldsymbol{S}^{-1})(\boldsymbol{S}\boldsymbol{D}_2\boldsymbol{S}^{-1}) \\ &= \boldsymbol{S}\boldsymbol{D}_1\boldsymbol{D}_2\boldsymbol{S}^{-1} \\ &= \boldsymbol{S}\boldsymbol{D}_2\boldsymbol{D}_1\boldsymbol{S}^{-1} \\ &= \boldsymbol{S}\boldsymbol{D}_2\boldsymbol{S}^{-1})(\boldsymbol{S}\boldsymbol{D}_1\boldsymbol{S}^{-1}) \\ &= \boldsymbol{B}\boldsymbol{A}. \end{aligned}$$

(b) We let $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$ be linearly independent eigenvectors of $\boldsymbol{A}$ associated with n distinct eigenvalues $\lambda_1, \dots, \lambda_n$.

Thus $\boldsymbol{A}\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i$. By postmultiplying $\boldsymbol{B}$ we find that

$$\boldsymbol{B}\boldsymbol{A}\boldsymbol{v}_i = \lambda_i \boldsymbol{B}\boldsymbol{v}_i \text{ for } i = 1,2,\dots,n. \tag{10.9}$$

Notice that $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ spans the whole $\mathbb{R}^n$, thus any vector in $\mathbb{R}^n$ could be expressed as the linear combination of $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$. Hence for $\boldsymbol{B}\boldsymbol{v}_i \in \mathbb{R}^n$, we set

$$\boldsymbol{B}\boldsymbol{v}_i = \beta_1\boldsymbol{v}_1 + \beta_2\boldsymbol{v}_2 + \cdots + \beta_n\boldsymbol{v}_n. \tag{10.10}$$

By postmultiplying $\boldsymbol{A}$ for equation (10.10) we find that

$$\begin{aligned} \boldsymbol{A}\boldsymbol{B}\boldsymbol{v}_i &= \beta_1\boldsymbol{A}\boldsymbol{v}_1 + \beta_2\boldsymbol{A}\boldsymbol{v}_2 + \cdots + \beta_n\boldsymbol{A}\boldsymbol{v}_n \\ &= \beta_1\lambda_1\boldsymbol{v}_1 + \beta_2\lambda_2\boldsymbol{v}_2 + \cdots + \beta_n\lambda_n\boldsymbol{v}_n \end{aligned} \tag{10.11}$$

Also, by applying equation (10.10) into equation (10.9) we derive:

$$\begin{aligned} \boldsymbol{BAv}_i &= \lambda_i(\beta_1\boldsymbol{v}_1 + \beta_2\boldsymbol{v}_2 + \cdots + \beta_n\boldsymbol{v}_n) \\ &= \beta_1\lambda_i\boldsymbol{v}_1 + \beta_2\lambda_i\boldsymbol{v}_2 + \cdots + \beta_n\lambda_i\boldsymbol{v}_n \end{aligned} \tag{10.12}$$

Since $\boldsymbol{AB} = \boldsymbol{BA}$, we derive $\boldsymbol{ABv}_i = \boldsymbol{BAv}_i$. Combining equation (10.11) and (10.12) we obtain:

$$\mathbf{0} = \boldsymbol{ABv}_i - \boldsymbol{BAv}_i = \beta_1(\lambda_1 - \lambda_i)\boldsymbol{v}_1 + \beta_2(\lambda_2 - \lambda_i)\boldsymbol{v}_2 + \cdots + \beta_n(\lambda_n - \lambda_i)\boldsymbol{v}_n$$

Due to the independence of $\boldsymbol{v}_i$, we derive

$$\beta_1(\lambda_1 - \lambda_i) = \beta_2(\lambda_2 - \lambda_i) = \cdots = \beta_n(\lambda_n - \lambda_i) = 0.$$

Since eigenvalues of $\boldsymbol{A}$ are distinct, we get $\lambda_j - \lambda_i \neq 0$ for $j \neq i$. Hence $\beta_j = 0$ for $j \neq i$.

Considering equation (10.10), we derive $\boldsymbol{Bv}_i = \beta_i\boldsymbol{v}_i$, which means $\boldsymbol{v}_i$ is also the eigenvector of $\boldsymbol{B}$.

Hence $\boldsymbol{A}$ and $\boldsymbol{B}$ has the same eigenvectors $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$. Since $\boldsymbol{A}$ can be diagonalized by matrix $\boldsymbol{S} = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \dots & \boldsymbol{v}_n \end{bmatrix}$, $\boldsymbol{B}$ could be also diagonalized by matrix $\boldsymbol{S}$.

(c) We need to show that there exists $\boldsymbol{S}$ that can diagonalize $\boldsymbol{A}$ and $\boldsymbol{B}$:

Suppose $\lambda_1, \lambda_2, \dots, \lambda_h$ be the distinct eigenvalues of $\boldsymbol{A}$ with multiplicities $r_1, r_2, \dots, r_h$ respectively. Since $\boldsymbol{A}$ is diagonalizable, there exists $\boldsymbol{Q}$ satisfying

$$\boldsymbol{Q}^{-1}\boldsymbol{AQ} := \boldsymbol{D} = \operatorname{diag}(\lambda_1\boldsymbol{I}_{r_1}, \lambda_2\boldsymbol{I}_{r_2}, \dots, \lambda_h\boldsymbol{I}_{r_h}) = \begin{pmatrix} \lambda_1\boldsymbol{I}_{r_1} & & & \\ & \lambda_2\boldsymbol{I}_{r_2} & & \\ & & \ddots & \\ & & & \lambda_h\boldsymbol{I}_{r_h} \end{pmatrix}.$$

Also, we can obtain the product $\boldsymbol{Q}^{-1}\boldsymbol{BQ}$ and partition it into block matrix

(We partition it in the same way that $\boldsymbol{D}$ has been partitioned):

$$\boldsymbol{Q}^{-1}\boldsymbol{B}\boldsymbol{Q} := \boldsymbol{C} = \begin{bmatrix} \boldsymbol{C}_{11} & \boldsymbol{C}_{12} & \cdots & \boldsymbol{C}_{1h} \\ \boldsymbol{C}_{21} & \boldsymbol{C}_{22} & \cdots & \boldsymbol{C}_{2h} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{C}_{h1} & \boldsymbol{C}_{h2} & \cdots & \boldsymbol{C}_{hh} \end{bmatrix}$$

where $\boldsymbol{C}_{ij}$ is $r_i \times r_j$ matrix.

- Firstly, we show $\boldsymbol{C}$ is *block diagonal:*

  Note that $\boldsymbol{AB} = \boldsymbol{BA}$, thus we have

$$\begin{aligned} \boldsymbol{DC} &= (\boldsymbol{Q}^{-1}\boldsymbol{AQ})(\boldsymbol{Q}^{-1}\boldsymbol{BQ}) \\ &= \boldsymbol{Q}^{-1}\boldsymbol{ABQ} = \boldsymbol{Q}^{-1}\boldsymbol{BAQ} \\ &= (\boldsymbol{Q}^{-1}\boldsymbol{BQ})(\boldsymbol{Q}^{-1}\boldsymbol{AQ}) \\ &= \boldsymbol{CD}. \end{aligned}$$

  Notice that the (i,j)th submatrix of $\boldsymbol{DC}$ is equal to the (i,j)th submatrix of $\boldsymbol{CD}$, which yields $\lambda_i \boldsymbol{I}_{r_i}\boldsymbol{C}_{ij} = \boldsymbol{C}_{ij}\lambda_j \boldsymbol{I}_{r_j} \implies \lambda_i \boldsymbol{C}_{ij} = \lambda_j \boldsymbol{C}_{ij}$.
  Since $\lambda_i \neq \lambda_j$ for $i \neq j$, we derive $\boldsymbol{C}_{ij} = \boldsymbol{0}$ for $i \neq j$; thus

$$\boldsymbol{C} = \text{diag}(\boldsymbol{C}_{11}, \boldsymbol{C}_{22}, \dots, \boldsymbol{C}_{hh}) = \begin{pmatrix} \boldsymbol{C}_{11} & & & \\ & \boldsymbol{C}_{22} & & \\ & & \ddots & \\ & & & \boldsymbol{C}_{hh} \end{pmatrix}.$$

  is *block diagonal*.

- Then we show $\boldsymbol{C}$ is diagonalizable:

  Since $\boldsymbol{B}$ is diagonalizable, there exists $\boldsymbol{M}$ satisfying

$$\boldsymbol{M}^{-1}\boldsymbol{BM} = \boldsymbol{N} \implies \boldsymbol{B} = \boldsymbol{MNM}^{-1}$$

where $\boldsymbol{N}$ is diagonal. And since $\boldsymbol{Q}^{-1}\boldsymbol{B}\boldsymbol{Q} = \boldsymbol{C}$, we derive

$$\boldsymbol{Q}^{-1}\boldsymbol{M}\boldsymbol{N}\boldsymbol{M}^{-1}\boldsymbol{Q} = \boldsymbol{C} \implies (\boldsymbol{Q}^{-1}\boldsymbol{M})^{-1}\boldsymbol{C}(\boldsymbol{Q}^{-1}\boldsymbol{M}) = \boldsymbol{N}$$

If we define $\boldsymbol{T} := \boldsymbol{Q}^{-1}\boldsymbol{M}$, then $\boldsymbol{T}^{-1}\boldsymbol{C}\boldsymbol{T} = \boldsymbol{N}$. So $\boldsymbol{C}$ is also diagonalizable.

- Then we show each $\boldsymbol{C}_{ii}$ is diagonalizable:

  Moreover, if we partition $\boldsymbol{T}$ as:

$$\boldsymbol{T} = \begin{bmatrix} \boldsymbol{T}_{11} & \boldsymbol{T}_{12} & \cdots & \boldsymbol{T}_{1h} \\ \boldsymbol{T}_{21} & \boldsymbol{T}_{22} & \cdots & \boldsymbol{T}_{2h} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{T}_{h1} & \boldsymbol{T}_{h2} & \cdots & \boldsymbol{T}_{hh} \end{bmatrix},$$

  where $\boldsymbol{T}_{ij}$ is $r_i \times r_j$ matrix, then we find the product $\boldsymbol{C}\boldsymbol{T}$ is always block diagonal matrix.

  Similarly, the product $\boldsymbol{T}^{-1} \times (\boldsymbol{C}\boldsymbol{T})$ is also block diagonal matrix.

  Hence without loss of generailty, we can say there must exist block diagonal matrix $\boldsymbol{T}_* = \operatorname{diag}(\boldsymbol{T}_{11}, \boldsymbol{T}_{22}, \dots, \boldsymbol{T}_{hh})$ such that

$$\begin{aligned}\boldsymbol{T}_*^{-1}\boldsymbol{C}\boldsymbol{T}_* &= \begin{pmatrix} \boldsymbol{T}_{11}^{-1} & & & \\ & \boldsymbol{T}_{22}^{-1} & & \\ & & \ddots & \\ & & & \boldsymbol{T}_{hh}^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{C}_{11} & & & \\ & \boldsymbol{C}_{22} & & \\ & & \ddots & \\ & & & \boldsymbol{C}_{nn} \end{pmatrix} \begin{pmatrix} \boldsymbol{T}_{11} & & & \\ & \boldsymbol{T}_{22} & & \\ & & \ddots & \\ & & & \boldsymbol{T}_{hh} \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{T}_{11}^{-1}\boldsymbol{C}_{11}\boldsymbol{T}_{11} & & & \\ & \boldsymbol{T}_{22}^{-1}\boldsymbol{C}_{22}\boldsymbol{T}_{22} & & \\ & & \ddots & \\ & & & T_{hh}^{-1}\boldsymbol{C}_{hh}\boldsymbol{T}_{hh} \end{pmatrix} = \boldsymbol{N}.\end{aligned} \tag{10.13}$$

  Hence each $\boldsymbol{C}_{ii}$ is also diagonalizable.

- Finally, we set $\boldsymbol{P} = \boldsymbol{Q}\boldsymbol{T}_*$, we show that both $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ and $\boldsymbol{P}^{-1}\boldsymbol{B}\boldsymbol{P}$ are diagonal:

$$\begin{aligned}
\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} &= \boldsymbol{T}_*^{-1}\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{Q}\boldsymbol{T}_* = \boldsymbol{T}_*^{-1}\boldsymbol{D}\boldsymbol{T}_* \\
&= \operatorname{diag}(\boldsymbol{T}_{11}^{-1},\boldsymbol{T}_{22}^{-1},\dots,\boldsymbol{T}_{hh}^{-1})\operatorname{diag}(\lambda_1\boldsymbol{I}_{r_1},\lambda_2\boldsymbol{I}_{r_2},\dots,\lambda_h\boldsymbol{I}_{r_h})\operatorname{diag}(\boldsymbol{T}_{11},\boldsymbol{T}_{22},\dots,\boldsymbol{T}
\\
&= \operatorname{diag}(\lambda_1\boldsymbol{T}_{11}^{-1}\boldsymbol{T}_{11},\lambda_2\boldsymbol{T}_{22}^{-1}\boldsymbol{T}_{22},\dots,\lambda_h T_{hh}^{-1}T_{hh}) \\
&= \operatorname{diag}(\lambda_1\boldsymbol{I}_{r_1},\lambda_2\boldsymbol{I}_{r_2},\dots,\lambda_h\boldsymbol{I}_{r_h}) = \boldsymbol{D}
\end{aligned}$$

and

$$\begin{aligned}
\boldsymbol{P}^{-1}\boldsymbol{B}\boldsymbol{P} &= \boldsymbol{T}_*^{-1}\boldsymbol{Q}^{-1}\boldsymbol{B}\boldsymbol{Q}\boldsymbol{T}_* = \boldsymbol{T}_*^{-1}\boldsymbol{C}\boldsymbol{T}_* \\
&= \boldsymbol{N} \qquad \text{(You may check equation (10.13) to see why.)}
\end{aligned}$$

Hence both $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ and $\boldsymbol{P}^{-1}\boldsymbol{B}\boldsymbol{P}$ are diagonal. The proof is complete.

6. (a) Firstly, we extend the **Hadamard Product** into vectors:

For $\boldsymbol{u},\boldsymbol{v} \in \mathbb{R}^{n\times 1}$, we obtain:

$$\begin{bmatrix}\boldsymbol{u}\circ\boldsymbol{v}\end{bmatrix} = \begin{bmatrix}u_1v_1 & u_2v_2 & \dots & u_nv_n\end{bmatrix}^{\mathrm{T}}.$$

Secondly, it's easy for you to verify the two properties:

Proposition 10.1 For matrices $\boldsymbol{A},\boldsymbol{B},\boldsymbol{C} \in \mathbb{R}^{n\times n}$, we have

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

Proposition 10.2 For vectors $\boldsymbol{u}_1,\boldsymbol{v}_1,\boldsymbol{u}_2,\boldsymbol{v}_2 \in \mathbb{R}^{n\times 1}$, we have

$$(\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}})\circ(\boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}}) = (\boldsymbol{u}_1\circ\boldsymbol{u}_2)\times(\boldsymbol{v}_1\circ\boldsymbol{v}_2)^{\mathrm{T}}.$$

So we begin to show $\operatorname{rank}(\boldsymbol{A}\circ\boldsymbol{B}) \le \operatorname{rank}(\boldsymbol{A})\operatorname{rank}(\boldsymbol{B})$:

We let $r_1 = \operatorname{rank}(\boldsymbol{A}), r_2 = \operatorname{rank}(\boldsymbol{B})$. Due to the theorem (8.4), we can decompose $\boldsymbol{A}$ and $\boldsymbol{B}$ as:

$$\boldsymbol{A} = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^{\mathrm{T}} + \cdots + \sigma_{r_1} \boldsymbol{u}_{r_1} \boldsymbol{v}_{r_1}^{\mathrm{T}}$$
$$\boldsymbol{B} = \eta_1 \boldsymbol{w}_1 \boldsymbol{x}_1^{\mathrm{T}} + \eta_2 \boldsymbol{w}_2 \boldsymbol{x}_2^{\mathrm{T}} + \cdots + \eta_{r_2} \boldsymbol{w}_{r_2} \boldsymbol{x}_{r_2}^{\mathrm{T}}$$

where $\boldsymbol{u}_i, \boldsymbol{v}_i, \boldsymbol{w}_i, \boldsymbol{x}_i$'s are all $\mathbb{R}^{n\times 1}$ vectors.

Hence the Hadamard product $\boldsymbol{A} \circ \boldsymbol{B}$ is given by:

$$\begin{aligned}
\boldsymbol{A} \circ \boldsymbol{B} &= \left(\sum_{i=1}^{r_1} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^{\mathrm{T}}\right) \circ \left(\sum_{j=1}^{r_2} \eta_j \boldsymbol{w}_j \boldsymbol{x}_j^{\mathrm{T}}\right) \\
&= \sum_{i=1}^{r_1} \sum_{j=1}^{r_2} \sigma_i \eta_j (\boldsymbol{u}_i \boldsymbol{v}_i^{\mathrm{T}} \circ \boldsymbol{w}_j \boldsymbol{x}_j^{\mathrm{T}}) && \text{Due to the proposition (10.1)} \\
&= \sum_{i=1}^{r_1} \sum_{j=1}^{r_2} \sigma_i \eta_j (\boldsymbol{u}_i \circ \boldsymbol{w}_j)(\boldsymbol{v}_i \circ \boldsymbol{x}_j)^{\mathrm{T}} && \text{Due to the proposition (10.2)}
\end{aligned}$$

Notice that $(\boldsymbol{u}_i \circ \boldsymbol{w}_j)$ and $(\boldsymbol{v}_i \circ \boldsymbol{x}_j)$ are all $\mathbb{R}^{n\times 1}$ vectors, so $(\boldsymbol{u}_i \circ \boldsymbol{w}_j)(\boldsymbol{v}_i \circ \boldsymbol{x}_j)$ are rank 1 matrix.

Hence we express $\boldsymbol{A} \circ \boldsymbol{B}$ as the sum of $r_1 r_2$ matrices with rank 1.

Thus $\operatorname{rank}(\boldsymbol{A} \circ \boldsymbol{B}) \leq r_1 r_2 = \operatorname{rank}(\boldsymbol{A}) \operatorname{rank}(\boldsymbol{B})$.

(b) Since $\boldsymbol{A} \succeq$, we decompose $\boldsymbol{A}$ as:

$$\boldsymbol{A} = \boldsymbol{U}^{\mathrm{T}} \boldsymbol{U} \text{ where } \boldsymbol{U} \text{ is square.}$$

If we set $\boldsymbol{U} := \begin{bmatrix} \boldsymbol{u}_1^{\mathrm{T}} \\ \boldsymbol{u}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{u}_n^{\mathrm{T}} \end{bmatrix}$, we can write $\boldsymbol{A}$ as:

$$\boldsymbol{A} = \boldsymbol{U}^{\mathrm{T}} \boldsymbol{U} = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{u}_1^{\mathrm{T}} \\ \boldsymbol{u}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{u}_n^{\mathrm{T}} \end{bmatrix} = \boldsymbol{u}_1 \boldsymbol{u}_1^{\mathrm{T}} + \boldsymbol{u}_2 \boldsymbol{u}_2^{\mathrm{T}} + \cdots + \boldsymbol{u}_n \boldsymbol{u}_n^{\mathrm{T}}$$

Similarly, we can write $\boldsymbol{B}$ as:

$$\boldsymbol{B} = \boldsymbol{v}_1\boldsymbol{v}_1^{\mathrm{T}} + \boldsymbol{v}_2\boldsymbol{v}_2^{\mathrm{T}} + \cdots + \boldsymbol{v}_n\boldsymbol{v}_n^{\mathrm{T}}.$$

Hence $\boldsymbol{A} \circ \boldsymbol{B}$ can be written as

$$\begin{aligned}
\boldsymbol{A} \circ \boldsymbol{B} &= \left(\sum_{i=1}^{n} \boldsymbol{u}_i\boldsymbol{u}_i^{\mathrm{T}}\right) \circ \left(\sum_{j=1}^{n} \boldsymbol{v}_j\boldsymbol{v}_j^{\mathrm{T}}\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} (\boldsymbol{u}_i\boldsymbol{u}_i^{\mathrm{T}} \circ \boldsymbol{v}_j\boldsymbol{v}_j^{\mathrm{T}}) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} (\boldsymbol{u}_i \circ \boldsymbol{v}_j)(\boldsymbol{u}_i \circ \boldsymbol{v}_j)^{\mathrm{T}}
\end{aligned}$$

If we set $\boldsymbol{w}_{ij} = \boldsymbol{u}_i \circ \boldsymbol{v}_j$, then we obtain:

$$\boldsymbol{A} \circ \boldsymbol{B} = \sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{w}_{ij}\boldsymbol{w}_{ij}^{\mathrm{T}}$$

Hence for $\forall \boldsymbol{x} \in \mathbb{R}^n$, we derive

$$\begin{aligned}
\boldsymbol{x}^{\mathrm{T}}(\boldsymbol{A} \circ \boldsymbol{B})\boldsymbol{x} &= \sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{x}^{\mathrm{T}}\boldsymbol{w}_{ij}\boldsymbol{w}_{ij}^{\mathrm{T}}\boldsymbol{x} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \langle \boldsymbol{x}\boldsymbol{w}_{ij}, \boldsymbol{x}\boldsymbol{w}_{ij} \rangle \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} \|\boldsymbol{x}\boldsymbol{w}_{ij}\|^2 \geq 0.
\end{aligned}$$

By definition, $\boldsymbol{A} \circ \boldsymbol{B} \succeq 0$.