A FIRST COURSE

IN

LINEAR ALGEBRA

A FIRST COURSE

IN

LINEAR ALGEBRA

MAT2040 Notebook

Prof. Tom Luo

The Chinese University of Hongkong, Shenzhen

Prof. Ruoyu Sun

University of Illinois Urbana-Champaign

LOGO

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data:

Survey Methodology / Robert M. Groves . . . [et al.].
p. cm.—(Wiley series in survey methodology)
"Wiley-Interscience."
Includes bibliographical references and index.
ISBN 0-471-48348-6 (pbk.)
1. Surveys—Methodology. 2. Social sciences—Research—Statistical methods. I. Groves, Robert M. II. Series.

HA31.2.S873 2004
001.4'33—dc22 2004044064
Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

Contents

Contributors

Zhi-quan Luo, Shenzhen Research Institute of Big Data, Lecturer

Ruoyu Sun, Industrial and Enterprise Systems Engineering, Lecturer

Jie Wang, The Chinese University of Hongkong, Shenzhen, Typer

Foreword

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Preface

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

place
date

Acknowledgments

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

I. R. S.

Acronyms

ASTA	Arrivals See Time Averages
BHCA	Busy Hour Call Attempts
BR	Bandwidth Reservation
b.u.	bandwidth unit(s)
CAC	Call / Connection Admission Control
CBP	Call Blocking Probability(-ies)
CCS	Centum Call Seconds
CDTM	Connection Dependent Threshold Model
CS	Complete Sharing
DiffServ	Differentiated Services
EMLM	Erlang Multirate Loss Model
erl	The Erlang unit of traffic-load
FIFO	First in - First out
GB	Global balance
GoS	Grade of Service
ICT	Information and Communication Technology
IntServ	Integrated Services
IP	Internet Protocol
ITU-T	International Telecommunication Unit – Standardization sector
LB	Local balance
LHS	Left hand side

LIFO	Last in - First out
MMPP	Markov Modulated Poisson Process
MPLS	Multiple Protocol Labeling Switching
MRM	Multi-Retry Model
MTM	Multi-Threshold Model
PASTA	Poisson Arrivals See Time Averages
PDF	Probability Distribution Function
pdf	probability density function
PFS	Product Form Solution
QoS	Quality of Service
r.v.	random variable(s)
RED	random early detection
RHS	Right hand side
RLA	Reduced Load Approximation
SIRO	service in random order
SRM	Single-Retry Model
STM	Single-Threshold Model
TCP	Transport Control Protocol
TH	Threshold(s)
UDP	User Datagram Protocol

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.)

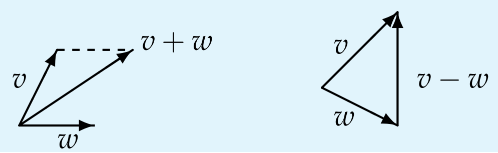
Visualize. Conncect between **Geometry** and **Algebra**.

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

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

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

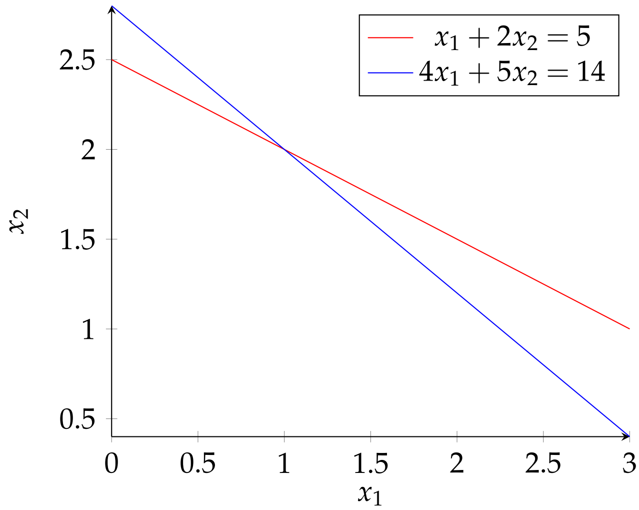
$$
\left[\begin{array}{cc|c}
1 & 2 & 5 \\
0 & 1 & 2
\end{array}\right].
$$

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.

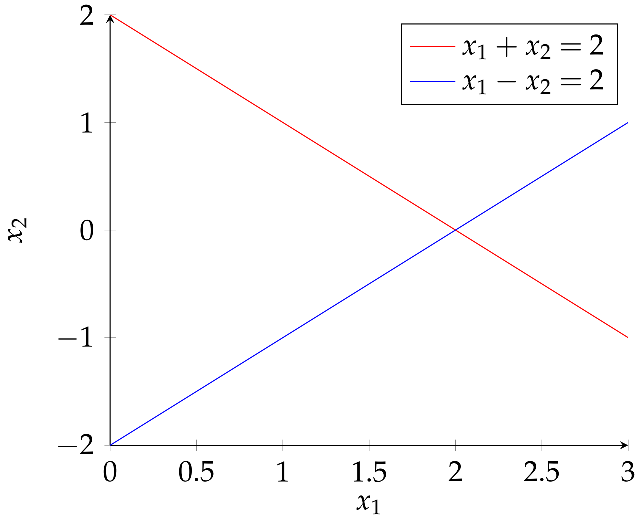


Column Picture. Focusing on the column of the system of equation, we can denote $\begin{bmatrix} 1 \\ 4 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 5 \end{bmatrix}$ as vectors in coordinate axis. Could the linear combinations of these two vectors form the vector $\begin{bmatrix} 5 \\ 14 \end{bmatrix}$?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

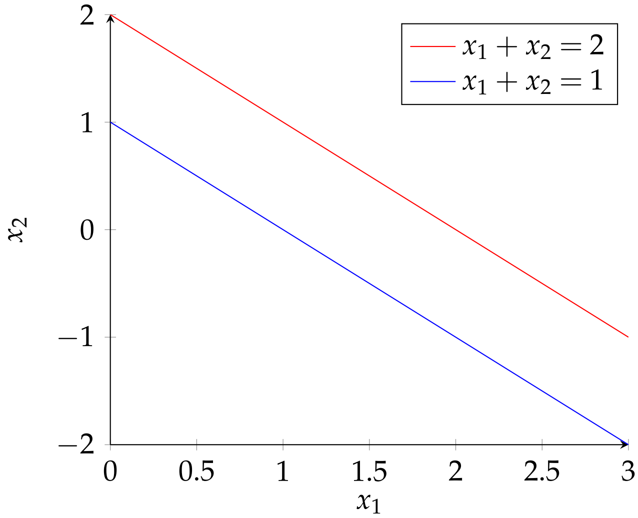
1.1.2.2. The solutions of the Linear System of Equations

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

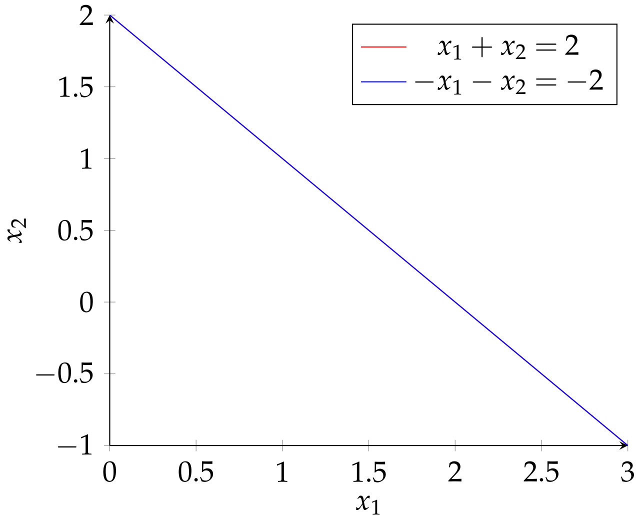
Case 1: unique solution. If two lines intersect at one point, then there is unique solution.



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**. $(k = 1, \ldots, m)$.

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:

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

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 $\boldsymbol{x}$:

$$\boldsymbol{Ax} = \begin{pmatrix} 2x_1 + x_2 + x_3 \\ 4x_1 - 6x_2 \\ -2x_1 + 7x_2 + 2x_3 \end{pmatrix} = \begin{pmatrix} \langle a_1, \boldsymbol{x} \rangle \\ \langle a_2, \boldsymbol{x} \rangle \\ \langle a_3, \boldsymbol{x} \rangle \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

Hence we finally write the system equation as:

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

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

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

■

1.2.2.2. Matrix Multiply Matrix

R Note that an $m \times n$ matrix $\boldsymbol{A}$ can be written as $\begin{bmatrix} a_{ij} \end{bmatrix}$, 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:

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

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

■ **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

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

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:

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

And the (i,j)th entry of the result should be the ith row of $\boldsymbol{A}$ multiply the jth column of $\boldsymbol{B}$. In this example, the result has 4×2 entries.

So 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,

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

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

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

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

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

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:

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

1.3.2. Elementary Matrix

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

(R) It seems that we only talk about elementary matrix of type III instead of other types. So 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* $\boldsymbol{E}_{31}$ 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 $\boldsymbol{E}_{21}\boldsymbol{E}_{31}$:

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

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

And 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 the result of $E_{21} \times (E_{31}b)$ is actually the same as $(E_{21}E_{31}) \times b$, which is one of the properties of matrix.

1.3.3. Properties of Matrix

Operations on matrix has the following properties:

1. $A(B + C) = AB + AC$.
2. $AB \neq BA$, this means AB doesn't *necessarily* equal to BA.

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

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

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

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

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

   ■ **Example 1.10** Given two matrix A and B, and we obtain $A \times B = C$. In order to compute the result of $\boldsymbol{C}$, we can partition $\boldsymbol{A}$ and $\boldsymbol{B}$ arbitrarily, for example,

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

The entries in $\boldsymbol{A}$ and $\boldsymbol{B}$ are picked arbitrarily. Thus we partition $\boldsymbol{C}$ into 4 blocks respectively:

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

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

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

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

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

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

And if we know $\boldsymbol{B}$ has k columns, we can partition $\boldsymbol{B}$ into k blocks to compute $\boldsymbol{A}\boldsymbol{B}$:

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

Also, if we know $\boldsymbol{A}$ has m rows, we can partition $\boldsymbol{A}$ into m blocks to compute

$\boldsymbol{AB}$:

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

1.3.4. Permutation Matrix

We notice that there is one type 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 $\boldsymbol{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, $\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, 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 $\boldsymbol{A}$. ■

But how to describe a matrix that exchanges row i and row j? We use notation $\boldsymbol{P}_{ij}$ to denote such matrix. A matrix that only interchange two rows is called *row exchange matrix*:

Definition 1.13 [Row Exchange Matrix] After a identity matrix's ith and jth row being exchanged, it is denoted by $\boldsymbol{P}_{ij}$. And $\boldsymbol{P}_{ij}$ is called **Row Exchange Matrix**. By postmultiplying $\boldsymbol{P}_{ij}$ for matrix $\boldsymbol{A}$ has the same effect of interchanging row i and row j of matrix $\boldsymbol{A}$. ■

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

■ **Example 1.11** P_{23} is used to exchange row 2 and row 3 of arbitarary matrix. And it is converted from 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:

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

(R) You may be confused about the concept between *permutation matrix* and *row exchange matrix*. Our P_{23} (row exchange matrix) is just one particular permutation matrix—it exchanges row 2 and row 3. Soon we will meet other permutation matrix, which can change the order of **several** rows. For example, rows 1,2,3 can be changed to 3,2,1.

Before we talk 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 $\boldsymbol{B}$ such that $\boldsymbol{AB} = \boldsymbol{BA} = \boldsymbol{I}$. And the matrix $\boldsymbol{B}$ is said to be the **inverse** of $\boldsymbol{A}$, and we can write $\boldsymbol{B} = \boldsymbol{A}^{-1}$.
3. After multiplying finite numbers of **elementary matrix**, $\boldsymbol{A}$ can be converted to identity matrix $\boldsymbol{I}$.
4. The system of equations $\boldsymbol{Ax} = \boldsymbol{b}$ has a unique solution.

If matrix $\boldsymbol{A}$ is **not** nonsingular, this matrix is called **singular**. ▪

And we are interested in the form of the inverse of permutation matrix.

Proposition 1.3 1. For a permutation matrix $\boldsymbol{P}$, it can always be decomposed into finite multiplication of row exchange matrix $\boldsymbol{P}_{ij}$:

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

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

$$\boldsymbol{P}_{ij} \boldsymbol{P}_{ij} = \boldsymbol{I} \Longleftrightarrow \boldsymbol{P}_{ij}^{-1} = \boldsymbol{P}_{ij}$$

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

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

4. For $n \times n$ permutation matrix and $n \times n$ matrix given by

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

we have:

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

Proof. [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 nothing!

- For proposition 3, it is because that we just need to do reverse order of our process in order to get the inverse matrix.

1.3.5. LU decomposition

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

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

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

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

Proposition 1.4

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

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

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

It has the same effect of adding $(-2)\times$ row 1 to row 2. How to get identity again? We just need to add $2\times$ row 1 to row 2. So we only need to postmultiply another elementary matrix:

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

Hence $\overline{\boldsymbol{E}_{21}}$ is the inverse matrix of $\boldsymbol{E}_{21}$.

The product of elementary matrix $\boldsymbol{E}_{ij}$ $(i < j)$ is lower triangular matrix. The product of elementary matrix $\boldsymbol{E}_{ij}$ $(i > j)$ is upper triangular matrix.

Let's look at an example to see what is LU decomposition:

■ **Example 1.12** 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 $E_{21} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $E_{31} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$, $E_{32} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$.
Finally we convert A into an upper triangular matrix U. Let's reverse the process to find some interesting conclusion:

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

where $L = E_{21}^{-1}E_{31}^{-1}E_{32}^{-1}$, which is lower triangular matrix.
And we successfully decompose matrix A into multiplication of a lower triangular matrix L and a upper triangular matrix U. ■

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 $Ax = b$. But computer wants to deal with A and b in separate equations. So LU decomposition can help us do the following things:

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

R *How does Solve work on?* First, we apply forward elimination to $\boldsymbol{b}$, which changes $\boldsymbol{b}$ to $\boldsymbol{y}$. In other words, we are actually solving $\boldsymbol{L}\boldsymbol{y} = \boldsymbol{b}$. After getting $\boldsymbol{y}$, we then do back substitution to solve $\boldsymbol{U}\boldsymbol{x} = \boldsymbol{y}$. In other words, in second step we are solving $\boldsymbol{U}\boldsymbol{x} = \boldsymbol{y}$. The original system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ is converted into two triangular systems:

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

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

Example 1.13 Forward elimination on $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ will result in equation $\boldsymbol{U}\boldsymbol{x} = \boldsymbol{y}$:

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \Longleftrightarrow \left\{ \begin{aligned} u + 2v &= 5 \\ 4u + 9v &= 21 \end{aligned} \right\} \qquad \text{becomes} \left\{ \begin{aligned} u + 2v &= 5 \\ v &= 1 \end{aligned} \right\} \Longleftrightarrow \boldsymbol{U}\boldsymbol{x} = \boldsymbol{y}.$$

How to use matrix to compute it more quickly?

- $\boldsymbol{L}\boldsymbol{y} = \boldsymbol{b}$: In this system of equation, we know $\boldsymbol{L} = \begin{bmatrix} 1 & 0 \\ 4 & 1 \end{bmatrix}$, in oder to solve $\boldsymbol{y}$, we only need to multiply the inverse of $\boldsymbol{L}$ both sides:

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

- $\boldsymbol{U}\boldsymbol{x} = \boldsymbol{y}$: In this system of equation, we know $\boldsymbol{U} = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$, 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

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

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

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

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

And we can translate LU decomposition into LDU decomposition by multiplying factor $\boldsymbol{D}\boldsymbol{D}^{-1}$:

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

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

Since we know exactly matrix D^{-1}, we can solve for matrix D:

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

In order to verify this solution you just need to multiply it by D^{-1} to ensure its result to be *identity matrix*. And we notice that the *diagonal* entries of D are all *pivots values* of U.

Similarly, we can proceed this step again to let *diagonal* entries of L to be 1. In conclusion, we decompose matrix A into the form:

$$A = LDU$$

where L is lower triangular matrix with unit entries in diagonal

D is diagonal matrix

U is upper triangular matrix with unit entries in diagonal

This decomposition is called **LDU decomposition**. Here is a property of LDU decomposition, we state it first without proof. We will proof it in the future.

Proposition 1.5 Let L be a lower triangular matrix, D diagonal, and U upper triangular. If $A = LDU$ and also $A = L_1D_1U_1$, then we have $L = L_1, D = D_1, U = U_1$. In other words, **LDU decomposition is unique to any matrix.**

1.3.7. LU Decomposition with row exchanges

How can we handle row exchange in our LU decomposition?

Assume we are going to do Gaussian Elimination with matrix A. We can postmultiply many elementary matrix $\boldsymbol{E}$ 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. But *how can we get all the row exchange matrix* $\boldsymbol{P}$ *out from among the* $\boldsymbol{L}$*?*

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

We skip the proof of this theorem, if you are interested in it, you may check the website:

`http://www.doc88.com/p-9387269167515.html`

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:

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

  And 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 nontrivally.* To explain it more specifically, let's use augmented matrix to represent $\boldsymbol{Ax} = \boldsymbol{b}$ (Let's assume it's 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 we focus on the columns, we may have the question: in which case does its columns could be linear 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*. And we will prove it in the end of this lecture. And if we focus on the rows, we may have the similar question. And its conclusion is similar.

- **matrix notation to describe Gaussian Elimination** Firstly let's consider we don't need to do row exchange case. For nonsingular matrix $\boldsymbol{A}$, 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:

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

And if we define $\boldsymbol{L} = \boldsymbol{E}_1^{-1} \dots \boldsymbol{E}_n^{-1}$, which is easy to verify that it is lower triangular matrix. So finally we decompose $\boldsymbol{A}$ into the product of two triangular matrix:

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

Further more, we can decompose $\boldsymbol{A}$ into product of three matrix to make the diagonal entries of $\boldsymbol{U}$ to be zero:

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

Note that the LDU decomposition is unique for any matrix, though we don't prove it.

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

$$\boldsymbol{E}\dots\boldsymbol{EP}\dots\boldsymbol{EP}\dots\boldsymbol{EA}=\boldsymbol{U}$$

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

$$\boldsymbol{E}_n\dots\boldsymbol{E}_1\boldsymbol{PA}=\boldsymbol{U}\implies\boldsymbol{PA}=\boldsymbol{LU}$$

And 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 called 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.

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

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

Note that we use notation $\begin{bmatrix}a_{:i}\end{bmatrix}$ to denote the ith column of $\boldsymbol{A}$. (MATLAB or Julia language.)

And given row vector $e_j^{\mathrm{T}} := \begin{bmatrix} 0 & 0 & \dots & 1 & \dots & 0 \end{bmatrix}$, the product of $e_j^{\mathrm{T}} \boldsymbol{A}$ is given by:

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

Note that we use notation $\begin{bmatrix} a_{j:} \end{bmatrix}$ to denote the jth row of $\boldsymbol{A}$.

- Secondly we want to compute the product $\mathbf{1}^{\mathrm{T}}\boldsymbol{A}\mathbf{1}$, where $\mathbf{1}$ denotes a column vector that all entres of $\mathbf{1}$ are 1.

  Let's first compute $\boldsymbol{A} \times \mathbf{1}$, where $\boldsymbol{A}$ is a $m \times n$ matrix and $\mathbf{1} \in \mathbb{R}^n$:

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

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

  where $\mathbf{1}^{\mathrm{T}}$ is a $1 \times m$ row vector.

- If vector $x \in \mathbb{R}^m$, $y \in \mathbb{R}^n$, we can compute $x^{\mathrm{T}}\boldsymbol{A}y$:

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

- If vector $x \in \mathbb{R}^n$, $y \in \mathbb{R}^n$, you should distinguish $x^{\mathrm{T}}y$ and xy^{T}:

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

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

- If vector $x \in \mathbb{R}^m$, $y \in \mathbb{R}^n$, we can compute $x^{\mathrm{T}}\boldsymbol{A}y$ by using block matrix:

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

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

where $\boldsymbol{A}_{11}$ is $m_1 \times n_1$ matrix, $\boldsymbol{A}_{22}$ is $m_2 \times n_2$ matrix. Then we partition vector x and y respectively:

$$x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \qquad y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

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 $x^{\mathrm{T}}\boldsymbol{A}y$:

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

-

Proposition 2.1 Postmultiplying $\boldsymbol{Q}$ for vector $v = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ rotates v in the plane anticlockwise by the angle θ:

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

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 $\boldsymbol{Q}$ and v:

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

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

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

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

If we aftermultiply this matrix into the matrix $\boldsymbol{A}$, we can flip $\boldsymbol{A}$ horizontally:

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

2.1.3. Inverse

Let's introduce the definition for inverse matrix:

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

And inverse matrix has some interesting properties:

Proposition 2.2 Matrix inverse is *unique*. In other words, if we have $\boldsymbol{AB}_1 = \boldsymbol{B}_1\boldsymbol{A} = \boldsymbol{I}$ and $\boldsymbol{AB}_2 = \boldsymbol{B}_2\boldsymbol{A} = \boldsymbol{I}$, then we obtain $\boldsymbol{B}_1 = \boldsymbol{B}_2$.

Proof.

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

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

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

Proof. On the one hand, we have

$$\boldsymbol{CAB} = \boldsymbol{C}(\boldsymbol{AB}) = \boldsymbol{CI} = \boldsymbol{C}$$

On the other hand, we obtain:

$$\boldsymbol{CAB} = (\boldsymbol{CA})\boldsymbol{B} = \boldsymbol{IB} = \boldsymbol{B}$$

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

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

By definition, we have $\boldsymbol{AX} = \boldsymbol{I}$, from the left side we derive

$$\boldsymbol{AX} = \boldsymbol{A}\begin{bmatrix} x_1 \mid x_2 \mid \dots \mid x_n \end{bmatrix}$$

from the right side we have

$$\boldsymbol{I} = \begin{bmatrix} e_1 \mid e_2 \mid \dots \mid e_n \end{bmatrix}$$

where $e_1, e_2, \dots, e_n$ are all unit vectors.
Hence we obtain $\boldsymbol{A}\begin{bmatrix} x_1 \mid x_2 \mid \dots \mid x_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}x_1 \mid \boldsymbol{A}x_2 \mid \dots \mid \boldsymbol{A}x_n \end{bmatrix} = \begin{bmatrix} e_1 \mid e_2 \mid \dots \mid e_n \end{bmatrix}$.
Thus we only need to compute n system of equations $\boldsymbol{A}x_i = e_i$, where $i = 1, 2, \dots, n$.
Hence we have to do n Gaussian Elimination to convert $\boldsymbol{A}$ into identity matrix $\boldsymbol{I}$. Once we have done that, we get the inverse of $\boldsymbol{A}$ 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 has three columns:

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

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

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

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

which is equivalent to $\boldsymbol{IX} = \begin{bmatrix} \frac{12}{16} & -\frac{5}{16} & -\frac{6}{16} \\ \frac{1}{2} & -\frac{3}{8} & -\frac{1}{4} \\ -1 & 1 & 1 \end{bmatrix}$.

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

Let's discuss in which case does the inverse exist:

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

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

$\Leftrightarrow$ none pivot values of $\boldsymbol{A}$ is zero. $\Leftrightarrow$ $\boldsymbol{Ax} = \boldsymbol{b}$ has a unique solution $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$. Finally let's discuss an interesting theorem that gives equivalent condition for columns combination and rows combination.

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

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

Proof. [Proofoutline.] Columns of $\boldsymbol{A}$ can be linearly combined to zero nontribally.

$\Leftrightarrow$ If $\boldsymbol{A} = \begin{bmatrix} a_1 & \vline & a_2 & \vline & \dots & \vline & a_n \end{bmatrix}$, then there exists $x_i \neq 0$ such that

$$a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = 0$$

$\Leftrightarrow$ $\boldsymbol{Ax} = \boldsymbol{0}$ has nonzero solution $\bar{\boldsymbol{x}}$.

$\Leftrightarrow$ $2\bar{\boldsymbol{x}}, 3\bar{\boldsymbol{x}}, \dots$ are also solutions to $\boldsymbol{Ax} = \boldsymbol{0}$.

$\Leftrightarrow$ $\boldsymbol{Ax} = \boldsymbol{0}$ has infinitely many solutions.

$\Leftrightarrow$ $\boldsymbol{A}^{-1}$ does not exist, otherwise we will only have unique solution $\boldsymbol{A}^{-1} \times \boldsymbol{0} = \boldsymbol{0}$.

$\Leftrightarrow$ Gaussian Elimination breaks down.

$\Leftrightarrow$ There exists zero row in the row echelon form.

$\Leftrightarrow$ Row vectors of $\boldsymbol{A}$ can be linearly combined to zero nontrivally.

2.2. Wednesday

2.2.1. Remarks on Gaussian Elimination

1. Gaussian Elimination to compute $\boldsymbol{A}^{-1} \leftrightarrow$ Solving $\boldsymbol{A}\boldsymbol{x}_i = e_i$ for $i = 1,2,\dots,n$.

   For each i solving $\boldsymbol{A}\boldsymbol{x}_{\boldsymbol{i}} = e_i$ takes $O(n^3)$.

   Hence solving nn such systems take $O(n^4)$ time.

   However, if we solve $\boldsymbol{A}\boldsymbol{x}_{\boldsymbol{i}} = e_i$ for $i = 1,2,\dots,n$ simultaneously(that means we write all b_i at the right side of the Augmented matrix.) ,by Gaussian Elimination takes $O(n^3)$ time.

2. 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.(We have $\sum_{j=0}^{\infty} N^j = \boldsymbol{A}^{-1}$.)

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

   $$\boldsymbol{E}\dots\boldsymbol{E}\boldsymbol{P}\boldsymbol{A} = \boldsymbol{D}$$

   where $\boldsymbol{D}$ is a diagonal matrix.(By doing row operation we can also convert $\boldsymbol{A}$ into $\boldsymbol{D}$.)

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

   $$\boldsymbol{D}^{-1}(\boldsymbol{E}\dots\boldsymbol{E}\boldsymbol{P}\boldsymbol{A}) = \boldsymbol{I}$$

   where the diagonals of $\boldsymbol{D}^{-1}$ are 1 divides by pivots.

   Hence we have

   $$\boldsymbol{D}^{-1}(\boldsymbol{E}\dots\boldsymbol{E}\boldsymbol{P}\boldsymbol{A}) = \boldsymbol{I} \implies (\boldsymbol{D}^{-1}\boldsymbol{E}\dots\boldsymbol{E}\boldsymbol{P})\boldsymbol{A} = \boldsymbol{I}$$

   $$\implies \boldsymbol{A} = (\boldsymbol{D}^{-1}\boldsymbol{E}\dots\boldsymbol{E}\boldsymbol{P})^{-1} = \boldsymbol{P}^{-1}\boldsymbol{E}^{-1}\dots\boldsymbol{E}^{-1}\boldsymbol{D}$$

2.2.2. Properties of matrix

1. If $\boldsymbol{A}$ is a diagonal matrix which is given by $\boldsymbol{A} = \begin{bmatrix} d_1 & & 0 \\ & \vdots & \\ 0 & & d_n \end{bmatrix}$,

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

2. If $\boldsymbol{D}_1, \boldsymbol{D}_2$ are diagonal and their product exists, then we have

$$\boldsymbol{D}_1 \boldsymbol{D}_2 = \boldsymbol{D}_2 \boldsymbol{D}_1$$

3. If $\boldsymbol{A}, \boldsymbol{B}$ are both invertible, then $\boldsymbol{AB}$ is also invertible. The inverse of product $\boldsymbol{AB}$ is

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

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

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

   Similarly, $\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$ times $\boldsymbol{AB}$ 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 $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ are nonsingular, then $(\boldsymbol{ABC})^{-1} = \boldsymbol{C}^{-1}\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$.

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

$$(\boldsymbol{I} - \boldsymbol{A})^{-1} = \sum_{i=1}^{\infty} \boldsymbol{A}^i \qquad \text{when } \boldsymbol{A} \text{ is ``small'', we will explain ``small'' later.}$$

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 $\boldsymbol{Ax} = \boldsymbol{I}$:

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

Note that this result is obtained by three row operations:

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

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

"Add $(-1)\times$ row 1 to row 2". ∎

Hence we know why a nonzero diagonal lower triangular matrix is invertible. Because only in this case can we 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. If $\boldsymbol{A}$ is $n \times n$ matrix and it is invertible. Then $\boldsymbol{A}$ could be decomposed into $\boldsymbol{LDU}$, such decomposition is unique. *Proof.* Assume $\boldsymbol{A}$ could be written as $\boldsymbol{A} = \boldsymbol{L}_1\boldsymbol{D}_1\boldsymbol{U}_1 = \boldsymbol{L}_2\boldsymbol{D}_2\boldsymbol{U}_2$.

   And for the latter equation, we aftermultiply $\boldsymbol{U}_1^{-1}$ to obtain:

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

And then by postmultiplying L_2^{-1} we obtain:

$$L_2^{-1}L_1D_1 = D_2U_2U_1^{-1}$$

Notice that the product of $L_2^{-1}L_1$ is lower triangular with unit diagonal. (This is because LDU decomposition requires L and U to have unit diagonal.) Hence the product $L_2^{-1}L_1D_1$ must be lower triangular matrix. Similarly, the product $D_2U_2U_1^{-1}$ must be upper triangular matrix. Hence they must be diagonal matrix. Also, notice that the diagonal of $L_2^{-1}L_1D_1$ is the same as the diagonal of D_1(This is because the diagonal of $L_2^{-1}L_1$ has unit diagonal.) Hence $L_2^{-1}L_1D_1 = D_1$. Similarly, $D_2U_2U_1^{-1} = D_2$. Hence we obtain $D_1 = D_2$.
And we also obtain $L_2^{-1}L_1D_1 = D_1 \implies L_2^{-1}L_1 = D_1D_1^{-1} = I \implies L_1 = L_2$. Similarly, we have $U_1 = U_2$.

2.2.3. matrix transpose

We introduce a new matrix, it is the **transpose** of A, which is denoted by A^{T}. *The columns of A^{T} are the rows of A.* 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 A is $m \times n$ matrix, the transpose is $n \times m$:

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

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

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

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 $(\boldsymbol{AB})^{\mathrm{T}} = (\boldsymbol{B})^{\mathrm{T}}(\boldsymbol{A})^{\mathrm{T}}$.

Proof. [Proofoutline.] To understand why the product rule holds, we start with $(\boldsymbol{A}x)^{\mathrm{T}} = x^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}$:

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

It is the same combination of the same vectors! So the transpose of the column $\boldsymbol{A}x$ is the row $x^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}$. 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 $\boldsymbol{B} = \left[\begin{array}{c|c|c|c} b_1 & b_2 & \dots & b_k \end{array}\right]$, 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 $\begin{bmatrix} b_1^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}} \\ b_2^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}} \\ \vdots \\ b_k^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}} \end{bmatrix}$, 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 2.3 [symmetric matrix] A matrix $\boldsymbol{A}$ is **symmetric matrix** if we have $\boldsymbol{A} = \boldsymbol{A}^{\mathrm{T}}$. This means that $\begin{bmatrix} a_{ij} \end{bmatrix} = \begin{bmatrix} a_{ji} \end{bmatrix}$.

Choose any matrix $\boldsymbol{A}$, probably rectangular. Postmultiply $\boldsymbol{A}^{\mathrm{T}}$ to $\boldsymbol{A}$. Then the product $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is automatically 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 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ is a different matrix from $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$.

We should make the transpose notation clear:

R For two vector x and y,

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

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

And we introduce a matrix that seems opposite to symmetric matrix:

Definition 2.4 [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**.

And moreover, any $n \times n$ matrix can be decomposed as the sum of a symmetric and skew-symmetric matrices. 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

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

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 2.5 [symmetric matrix] A $n \times n$ matrix $\boldsymbol{A}$ is **symmetric matrix** if we have $\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A}$, which means $\begin{bmatrix} a_{ij} \end{bmatrix} = \begin{bmatrix} a_{ji} \end{bmatrix}$.

For example, matrix $\boldsymbol{A}$ is symmetric matrix:

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

Definition 2.6 [skew-symmetric matrix] A $n \times n$ matrix $\boldsymbol{A}$ is **skew-symmetric matrix** or say, **anti-symmetric matrix** if we have $\boldsymbol{A} = -\boldsymbol{A}^{\mathrm{T}}$.

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

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

And there is an interesting theorem given by

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

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

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

2.4.2. Interaction of inverse and transpose

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

Proof.

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

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

Proof.

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

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

Corollary 2.2 Given matrix $M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$, matrix $M = M^{\mathrm{T}}$ if and only if $A = A^{\mathrm{T}}, D = D^{\mathrm{T}}, B^{\mathrm{T}} = C$.

Proposition 2.7 Suppose A is $n \times n$, symmetric, and nonsingular matrix. When we do LDU decomposition such that $A = LDU$, U is exactly L^{T}.

Proof. [Proofoutline.] Suppose $A = LDU$, then $A^{\mathrm{T}} = (LDU)^{\mathrm{T}} = U^{\mathrm{T}}D^{\mathrm{T}}L^{\mathrm{T}}$.

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

Hence $A^{\mathrm{T}} = U^{\mathrm{T}}DL^{\mathrm{T}} = A \implies U^{\mathrm{T}}DL^{\mathrm{T}} = LDU = A$.

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

Since LDU decomposition is unique, we obtain $U^{\mathrm{T}} = L, L^{\mathrm{T}} = U$.

Hence $A = LDU = LDL^{\mathrm{T}}$.

2.4.3. Vector Space

We move to a new chapter-vector spaces. We know matrix calculation(such as $\boldsymbol{A}x = \boldsymbol{b}$) involves many numbers. you may think they are linear combinations of n vectors. This chapter moves from numbers and vectors to a third level of understanding(the highest level).Instead of individual columns, we look at "spaces" of vectors. And this chapter ends with the "*Fundamental Theorem of Linear Algebra*".

We begin with the most important vector spaces. They are denoted as $\mathbb{R}^n$.

Definition 2.7 The space $\mathbb{R}^n$ contains all column vectors v such that v has n entries. ■

And we denote vectors as *a column between brackets*, or *along a line using commas and parentheses:*

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

Definition 2.8 [vector space] A **vector space** $\boldsymbol{V}$ is a set of vectors such that these vectors satisfy *vector addition* and *scalar multiplication*:

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

In other words, the set of vectors is **closed** under *addition* $v+w$ and *multiplication* cv. In short, **any linear combination is closed in vector space.**

Proposition 2.8 Every vector space must contain the zero vector.

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

■ **Example 2.3** $\boldsymbol{V} = \left\{ \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \\ \vdots \end{pmatrix} \mid \{a_n\} \text{ is infinite length sequences.} \right\}$ 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:

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

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

$\mathbb{R}.\}$ is also vector space. Here you may understand the notation "*span*", the span of v_1, v_2, v_3 contains all linear combinations of vectors v_1, v_2, v_3.Also, $\boldsymbol{V}$ is a vector space. How to check?

Given any two vectors u, w in $\boldsymbol{V}$, suppose $u = \alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3, v = \beta_1 v_1 + \beta_2 v_2 + \beta_3 v_3$, 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. The inner product of u and v is series:

$$< u, v >= \sum_{i \in \mathbb{N}} u_i v_i$$

■ **Example 2.4** $F = \{f(x) \mid f : [0,1] \mapsto \mathbb{R}\}$ is also a vector space. (verify it by yourself.) This vector space contains all real functions defined on $[0,1]$. And the vector space F is infinite dimensional.

Given two functions f and g in F, the inner product of f and g is given by

$$< f,g >= \int_0^1 f(x)g(x)\,\mathrm{d}x$$

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

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

This set F is also a vector space. ■

■ **Example 2.5** $V = \left\{ \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \mid a_{ij} \in \mathbb{R} \text{ for } i = 1,2; j = 1,2,3. \right\}$ is also a vector space. (easy to verify). Moreover, it is equivalent to the span of six basic vectors:

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

We say V is 6-dimension without introducing the definiton of dimension formally. ■

■ **Example 2.6** $V = \{ \begin{bmatrix} a_{ij} \end{bmatrix}_{3\times 3} \mid \text{any } 3 \times 3 \text{ matrices}\}$ is also a vector space.
Obviously, it is 9-dimension. We usually express it as $dim(V) = 9$.
$V_1 = \{ \begin{bmatrix} a_{ij} \end{bmatrix}_{3\times 3} \mid \text{any } 3 \times 3 \text{ symmetric matrices}\}$ is a special vector space.
Notice that $V_1 \subset V$, so we say V_1 is a *subspace* of V.
In the future we will know $dim(V_1) = 6 < 9$. ■

We use more examples to explain subspace:

■ **Example 2.7** Choose a plane through the origin $(0,0,0)$, note that this plane in three-dimensional space is not $\mathbb{R}^2$ (Even if it looks like $\mathbb{R}^2$). The vectors in the plane have three components and they belongs to $\mathbb{R}^3$. So this plane is a subspace of $\mathbb{R}^3$.

Notice that *Every subspace also contains the zero vector* since subspace is a special vector space. So Here is a list of all the possible subspaces of $\mathbb{R}^3$:

- $(\boldsymbol{L})$ Any line through $(0,0,0)$
- $(\boldsymbol{R})$ The whole space
- $(\boldsymbol{P})$ Any plane through $(0,0,0)$
- $(\boldsymbol{Z})$ The single vector $(0,0,0)$

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

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

Definition 2.9 [homogeneous equations] A system of linear equations is said to be **homogeneous** if the constants on the righthand side are all zero. In other words, $\boldsymbol{Ax} = \boldsymbol{0}$ is said to be **homogeneous**.

Definition 2.10 [column space] The column space consists of all linear combinations of the columns of matrix $\boldsymbol{A}$. In other words, if $m \times n$ matrix $\boldsymbol{A}$ is denoted by $\boldsymbol{A} = \begin{bmatrix} a_1 & \vline & a_2 & \vline & \dots & \vline & a_n \end{bmatrix}$, then the column space is denoted by $\boldsymbol{C}(\boldsymbol{A}) = span(a_1, a_2, \dots, a_n) \subset \mathbb{R}^m$.

Definition 2.11 [null space] The null space of $m \times n$ matrix $\boldsymbol{A}$ consists of all solutions to $\boldsymbol{Ax} = \boldsymbol{0}$. And null space can be denoted as $\boldsymbol{N}(\boldsymbol{A}) = \{\boldsymbol{x} \mid \boldsymbol{Ax} = \boldsymbol{0}\} \subset \mathbb{R}^n$.

Proposition 2.9 The null space $\boldsymbol{N}(\boldsymbol{A})$ is a vector space.

Proof. [Proofoutline.] For any two vectors $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{N}(\boldsymbol{A})$, we have $\boldsymbol{Ax} = \boldsymbol{0}, \boldsymbol{Ay} = \boldsymbol{0}$.

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

Hence the linear combination of $\boldsymbol{x}$ and $\boldsymbol{y}$ is also in $\boldsymbol{N}(\boldsymbol{A})$. Hence $\boldsymbol{N}(\boldsymbol{A})$ is a vector space.

■ **Example 2.8** 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:

$$\begin{cases} x_1 + 0x_2 = 0 \\ 5x_1 + 4x_2 = 0 \\ 2x_1 + 3x_2 = 0 \end{cases}$$

Then easily we 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 2.9** 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. And we find that $\boldsymbol{A}\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} = \boldsymbol{0}$.

Hence $\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$ is a special solution. Note that *the null space contains all linear combinations of special solutions.* Hence the null space is $\boldsymbol{N}(\boldsymbol{A}) = \left\{ c\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} \mid c \in \mathbb{R} \right\}$. ■

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

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

$$\boldsymbol{x}_{particular} \qquad \textbf{The particular solution solves } \boldsymbol{Ax_p} = \boldsymbol{b}$$

$$\boldsymbol{x}_{nullspace} \qquad \textbf{The special solutions solves } \boldsymbol{Ax_n} = \mathbf{0}$$

That's talk about a theorem to help us solve the complete solution to $\boldsymbol{Ax} = \boldsymbol{b}$.

Theorem 2.4 Solution set of $\boldsymbol{Ax} = \boldsymbol{b}$ can be represented as $\boldsymbol{x}_{complete} = \boldsymbol{x_p} + \boldsymbol{x_n}$.

Proof. *Proof.* [Sufficiency.]Given $\boldsymbol{x}_{complete} = \boldsymbol{x_p} + \boldsymbol{x_n}$, it suffices to show $\boldsymbol{x}_{complete}$ is the solution to $\boldsymbol{Ax} = \boldsymbol{b}$. And we notice 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 $\boldsymbol{Ax} = \boldsymbol{b}$. *Proof.* [Necessity.]Suppose $\boldsymbol{x}$ is another solution to $\boldsymbol{Ax} = \boldsymbol{b}$, it suffices to show $\boldsymbol{x} = \boldsymbol{x_p} + \boldsymbol{x_n}$.

Hence we only need to show $\boldsymbol{x} - \boldsymbol{x_p} \in \boldsymbol{N}(\boldsymbol{A})$.

Notice that $\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{x_p}) = \boldsymbol{Ax} - \boldsymbol{Ax_p} = \boldsymbol{b} - \boldsymbol{b} = \boldsymbol{0}$.

Hence $\boldsymbol{x} - \boldsymbol{x_p} \in \boldsymbol{N}(\boldsymbol{A})$. Thus $\boldsymbol{x} = \boldsymbol{x_p} + \boldsymbol{x_n}$.

■ **Example 2.10** There are $n = 2$ unknowns but only $m = 1$ equations:

$$x_1 + x_2 = 2.$$

It's easy to check that the particular solution can be $\boldsymbol{x_p} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, the special solutions could be $\boldsymbol{x_n} = c \begin{pmatrix} 1 \\ -1 \end{pmatrix}$, c can be taken arbitararily.

Hence the complete solution for the equations could be written as

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

So we summarize that if there are n unknowns and m equations such that $m < n$, then $\boldsymbol{Ax} = \boldsymbol{b}$ is **underdetermined** (It has many solutions).

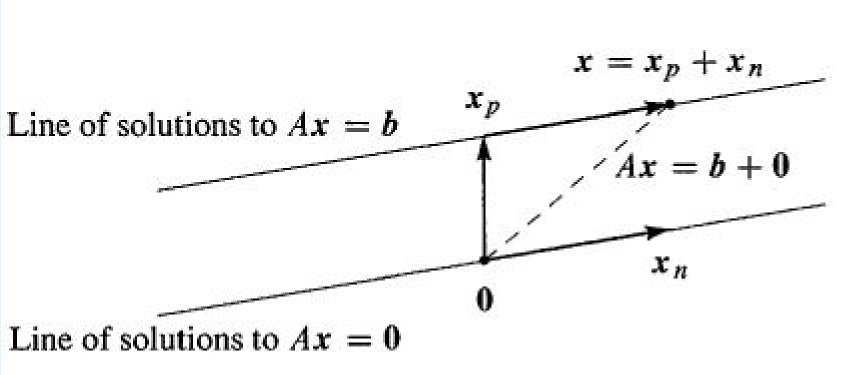


Figure 2.1: Complete solution = one particular solution + all nullspace solutions

2.4.3.3. Row-Echelon Matrices

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

$$\boldsymbol{PA} = \boldsymbol{LDU}$$

where $\boldsymbol{L}$ is $m \times m$ lower triangular matrix, $\boldsymbol{U}$ is $m \times n$ matrix that is of *row echelon form*. For example, here is a 4×7 row echelon matrix with the three pivots **1** highlighted in blue:

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

R

- 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 convert $\boldsymbol{U}$ into $\boldsymbol{R}$ that is of **reduced row echelon form**. For example, the $\boldsymbol{U}$ we listed above can be converted into:

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

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 'm' denotes **number of equations**, 'n' denotes **number of variables**. Assume 'r' denotes **number of pivots**, then we know 'r' is also **number of pivot variables**, '$n - r$' is **number of free variables**, and finally we have $m - r$ **redundant equations**.

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

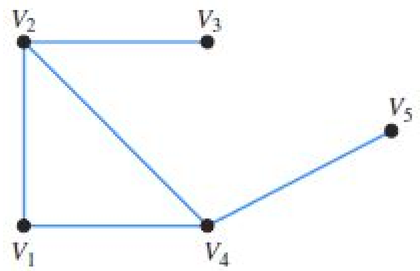
(b) $\boldsymbol{XA} + \boldsymbol{B} = \boldsymbol{C}$

(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. Why do we learn Linear Algebra?

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

- Baisis of AI/ML/SP/etc.

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

- Solving linear system of equations.

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

- Internal grace.

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

- Interview questions.

  LA is often used for interview questions for phd. Because the upper bound of difficulty for LA is **infinity**, interviewer often choose LA to question phd.

3.1.1.2. What is LA?

The main part of Mathematics is 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

Motivating question: How to solve linear system equations?

The basic method is **Gaussian Elimination** (To make equations *simpler*. The main idea is *induction*)

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

$$\implies \text{"if } a > 0\text{, no solution." or } x = \frac{b}{a}$$

By induction, *if you can sovle* $n \times n$ *systems, can you solve* $(n+1) \times (n+1)$ *systems?* In this 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:

For **singular** system, How to solve it?

- When will it has no solution, when it has infinitely many solutions? (Note that singular system don'y has unique solution.)
- If it has infinitely many solutions, how to find and express these solutions?

If we express system as matrix, we only to answer the question: **How to solve rectangular?**

3.1.3. Examples of solving equations

- For square case, we often convert the system into $\boldsymbol{R}\boldsymbol{x} = \boldsymbol{c}$, where $\boldsymbol{R}$ is of *row echelon form*.
- But for rectangular case, *row echelon form*(ref) is not enough, we must convert it into **reduced row echelon form**(rref):

$$\boldsymbol{U}(\text{ref}) = \begin{bmatrix} \boxed{1} & 0 & \times & \times & \times & 0 & \times \\ 0 & \boxed{1} & \times & \times & \times & 0 & \times \\ 0 & 0 & 0 & 0 & 0 & \boxed{1} & \times \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \implies \boldsymbol{R}(\text{rref}) = \begin{bmatrix} \boxed{1} & 0 & \times & \times & \times & 0 & \times \\ 0 & \boxed{1} & \times & \times & \times & 0 & \times \\ 0 & 0 & 0 & 0 & 0 & \boxed{1} & \times \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

■ **Example 3.1** We discuss how to solve square matrix of **rref**: If all rows have nonzero entry, we have:

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

We already sovled this system, but note that *the last row 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 result has 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 the $n \times n$ systems, does Gaussian Elimination work?

Answer: Almost, except "pivot=0"case.

- All pivots $\neq 0 \implies$ system has unique solution.
- Some pivots $= 0$ (The matrix is singular)
  1. No solution.
  2. Infinitely many solution.

■

3.1.3.1. What is G.E. doing? (Nonsingular case.)

Abstrastion: We use matrix to represent system of equations (Chinese mathematicians fail to do this.):

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

By postmultiplying $\boldsymbol{E}_{ij}$ and $\boldsymbol{P}_{ij}$ we do one step of elimination:

$$\boldsymbol{E}_{ij}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \qquad \boldsymbol{P}_{ij}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$

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

$$\hat{\boldsymbol{L}}\boldsymbol{P}\boldsymbol{A}\boldsymbol{x} = \hat{\boldsymbol{L}}\boldsymbol{P}\boldsymbol{b}$$

where $\hat{\boldsymbol{L}}\boldsymbol{P}\boldsymbol{A}$ represents a upper triangular matrix $\boldsymbol{U}$, $\hat{L}$ is the lower triangular matrix.

$$\implies \hat{\boldsymbol{L}}\boldsymbol{P}\boldsymbol{A} = \boldsymbol{U} \implies \boldsymbol{P}\boldsymbol{A} = \hat{\boldsymbol{L}}^{-1}\boldsymbol{U} \triangleq \boldsymbol{L}\boldsymbol{U}$$

So 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 an example for how to solve non-square system of rref, note that in last lecture we know the solution is given by:

$$\boldsymbol{x}_{\text{complete}} = \boldsymbol{x}_p + \boldsymbol{x}_{\text{special}}$$

Example 3.2 We try to solve the system $\begin{bmatrix} 1 & 3 & 0 & -1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}\boldsymbol{x} = \boldsymbol{c}$.

- **step 1:**Find null space. Thus we only need to solve $\boldsymbol{R}\boldsymbol{x} = \mathbf{0}$.

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

What should we do next? We want to express the **pivot variable** as the form of **free variable.**

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.

Hence the expression for x_1 and x_3 is given by:

$$\begin{cases} x_1 = -3x_2 + x_4 \\ x_2 = -x_4 \end{cases}$$

Hence all solutions to $\boldsymbol{R}\boldsymbol{x} = \mathbf{0}$ are

$$\boldsymbol{x}_{\text{special}} = \begin{bmatrix} -3x_2 + x_4 \\ x_2 \\ -x_4 \\ x_4 \end{bmatrix} = x_2 \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix}$$

where x_2 and x_4 can be taken arbitararily.

- **step 2:**find particular solution to $\boldsymbol{R}\boldsymbol{x} = \boldsymbol{c}$.

  The trick for this step is to set $x_2 = x_4 = 0$. (*set free variable to be zero and then derive the pivot variable.*)

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

$\implies$

- 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:** Assume $c_3 = 0$, then all solution to $\boldsymbol{Rx} = \boldsymbol{c}$ is given by:

$$\boldsymbol{x}_{complete} = \boldsymbol{x}_p + \boldsymbol{x}_{\text{special}} = \begin{bmatrix} c_1 \\ 0 \\ c_2 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix}$$

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:

- **step1:** Gaussian Elimination.

  With proper rows 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, clear row 2 and row 3 of this matrix:

$$\boldsymbol{A} \xRightarrow[E_{31}=\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix}]{E_{21}=\begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}} \begin{bmatrix} 1 & 1 & 2 & 3 \\ 0 & 0 & 4 & 4 \\ 0 & 0 & 4 & 4 \end{bmatrix} \xRightarrow[E_{32}=\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}]{E_{12}=\begin{bmatrix} 1 & -\frac{1}{2} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}} \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 0 & 4 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

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

  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 direction to derive these:

pivot $\Longrightarrow$ pivot columns $\Longrightarrow$ pivot columns $\Longrightarrow$ pivot variable

■ **Example 3.4** we want to identify **pivot variables** and **free variables** of $\boldsymbol{R}$:

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

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 $N(\boldsymbol{A})$, we only need to compute $N(\boldsymbol{R})$.
  - For each of $(n-r)$ free variables,

    set value of **it** to be 1.

    set other **free variables** to be 0.

    Then solve $\boldsymbol{R}\boldsymbol{x} = \boldsymbol{0}$ to get special solution y_j for $j = 1, 2, \ldots, n-r$.

■ **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 $\boldsymbol{R}\boldsymbol{x} = \boldsymbol{0}$, 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 $y_2 = \begin{bmatrix} -1 \\ 0 \\ -1 \\ 1 \end{bmatrix}$.

– When we get $(n-r)$ special solutions to $\boldsymbol{Rx} = \boldsymbol{0}$: $y_1, y_2, \dots, y_{n-r}$.

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

Example 3.6 We continue the example above, when we get all special solutions $y_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, y_2 = \begin{bmatrix} -1 \\ 0 \\ -1 \\ 1 \end{bmatrix}$, **the null space contains all linear combinations of the special solutions.**

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

where x_2, x_4 here could be arbitary.

- **step3:** Compute a particular solution $\boldsymbol{x}_p$.

  The easiest way is to "read" from $\boldsymbol{Rx} = \boldsymbol{c} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$:

  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. In order to have solution, we must have $c_{r+1} = \cdots = c_n = 0$. In other words, **For a solution to exist, zero rows in $\boldsymbol{R}$ must also be zero in $\boldsymbol{c}$.**

■ **Example 3.7** If $\boldsymbol{Rx} = \begin{bmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \boldsymbol{x} = \boldsymbol{c} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}$, then in order to have a solution, we must let $c_3 \neq 0$. ■

So we have to discuss the particular solution case by case:

- **case1:** one of $c_{r+1},\dots,c_n$ is nonzero, then the system has **no** solution.
- **case2:** $c_{r+1}=\dots=c_n$, then a particular solution exists:

$$\boldsymbol{x}_p=\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}$$

We set all **free variables** to be zero, and 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......

■ **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

$$\boldsymbol{x}_p=\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix}$$

Then we know x_2,x_4 are free variable, so $x_2=x_4=0$; x_1,x_3 are pivot variable, so we have $\begin{pmatrix}x_1\\x_3\end{pmatrix}=\begin{pmatrix}c_1\\c_2\end{pmatrix}$.

Hence the solution for $\boldsymbol{R}\boldsymbol{x}=\boldsymbol{c}$ is $\begin{bmatrix}c_1\\0\\c_2\\0\end{bmatrix}$. ■

- **Final step:** All solution of $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ are $\boldsymbol{x}_{\text{complete}}=\boldsymbol{x}_p+\boldsymbol{x}_{\text{special}}$, where $x_{\text{special}}\in$

$N(\boldsymbol{A})$. $\boldsymbol{x}_p$ 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 next lecture.

3.2. Thursday

3.2.1. Review

Last time you may be confused about how to compute $N(\boldsymbol{A})$ or $y_1, y_2, \ldots, y_{n-r}$ (step2). Now let's review the whole steps for solving rectangular bu using block matrix:

- Firstly, we convert our 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:

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

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

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

■

- Then our system equation is translated (We use 3×4 matrix to show the whole process.):

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

Because we have changed the columns, so here row 2 and row 3 is also switched respectively. And then x_1 and x_2 are pivot variables, x_3 and x_4 are free variables. Then we derive:

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

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

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

Hence all solutions to $\boldsymbol{Rx} = \boldsymbol{c}$ is obtained:

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} \boldsymbol{B}\begin{pmatrix} x_3 \\ x_4 \end{pmatrix} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ x_3 \\ x_4 \end{bmatrix}$$

Suppose $-\boldsymbol{B} = \begin{bmatrix} \hat{\boldsymbol{y}}_1 & \hat{\boldsymbol{y}}_2 \end{bmatrix}$, then pivot variables is equivalent to

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

- So the complete solution to the system is

$$
\begin{aligned}
\boldsymbol{x} &= \begin{pmatrix} c_1 \\ c_2 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} x_3\hat{\boldsymbol{y}}_1 + x_4\hat{\boldsymbol{y}}_2 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ x_3 \\ x_4 \end{pmatrix} && (3.1) \\
&= \begin{pmatrix} c_1 \\ c_2 \\ 0 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} \hat{\boldsymbol{y}}_1 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} \hat{\boldsymbol{y}}_2 \\ 0 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} && (3.2) \\
&= \underbrace{\begin{pmatrix} c_1 \\ c_2 \\ 0 \\ 0 \end{pmatrix}}_{\boldsymbol{x}_p} + \underbrace{x_3 \begin{pmatrix} \hat{\boldsymbol{y}}_1 \\ 1 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} \hat{\boldsymbol{y}}_2 \\ 0 \\ 1 \end{pmatrix}}_{\boldsymbol{x}_{\text{special}}} && (3.3)
\end{aligned}
$$

where x_3 and x_4 could be arbitarary.

- Notice that the block matrix is given by:

$$
\begin{pmatrix} \hat{\boldsymbol{y}}_1 & \hat{\boldsymbol{y}}_2 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -\boldsymbol{B} \\ \boldsymbol{I} \end{pmatrix} \implies \begin{pmatrix} \boldsymbol{I} & \boldsymbol{B} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \begin{bmatrix} -\boldsymbol{B} \\ \boldsymbol{I} \end{bmatrix} = \begin{bmatrix} -\boldsymbol{B} + \boldsymbol{B} \\ \boldsymbol{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}
$$

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

Answer: Also, we do G.E. to get rref, which will be of the form

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

3.2.2. Remarks on solving linear system equations

The two possibilities for linear equations depend on m and n:

Theorem 3.1 Let m denote number of equations, n denote number of variables. For 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.

Proof. [Proofoutline for $m < n$ case:] Recall we can convert $\boldsymbol{Ax} = \boldsymbol{b}$ into $\boldsymbol{Rx} = \boldsymbol{c}$:

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

Note that $x_1.x_2 \ldots, x_r$ is pivot variables (This is because of column switching). Hence we have $(n-r)$ free variables, thus $N(\boldsymbol{A})$ is spanned by $(n-r)$ special vectors $y_1, y_2, \ldots, y_{n-r}$.

Hence we only need to show $n > r$ *given the condition* $n > m$:

Obviously, $r \leq m$, and we have $n > m$, so we obtain $n > r$. So we get the proposition immediately:

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):

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

■

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

3.2.3. Linearly dependence

Definition 3.2 [linearly dependence] The vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ in linear space $\boldsymbol{V}$ are **linearly dependent** if there exists $c_1, c_2, \ldots, c_n \in \mathbb{R}$ 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 linear combination of others. When assuming $c_n \neq 0$, 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 [linearly independence] The vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n$ in linear space $\boldsymbol{V}$ are **linearly independent** if the two statements are equivalent:

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

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. And we often express *dependent* as *dep.*; we also sometimes express *linearly dependent* as *lin. dep.*; express *linearly independent* as *lin. ind.*

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

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

- 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

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 = \mathbf{0} \iff c_1 = c_2 = c_3 = 0.\text{(Why?)}$$

■

3.2.3.1. Relation between *lin.ind.* and *equations*

The following statements are equivalent:

- Vectors $a_1, a_2, \dots, a_n \in \mathbb{R}^m$ are dep.
- $\exists\, c_i$ not all zero s.t. $\sum_{i=1}^n c_i a_i = \mathbf{0}$.
- $\exists$ some $\boldsymbol{c} \neq \mathbf{0}$ s.t.
$$\boldsymbol{A}\boldsymbol{c} = \left[\begin{array}{c|c|c} a_1 & \dots & a_n \end{array}\right] \boldsymbol{c} = \mathbf{0}$$

So what if $m < n$, when checking corollary (3.1), we immediately 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 $\sum_{i=1}^n c_i a_i = \mathbf{0}$.

So we say the true equations are those linearly independent equations.

3.2.4. Basis and dimension

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

1. $v_1,\ldots,v_n$ are **linearly independent**.
2. $v_1,\ldots,v_n$ 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 don'y 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 feel that the number of vectors for 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; $w_1,w_2,\ldots,w_n$ is a basis for the same vector space $\boldsymbol{V}$, then $n=m$.

In order to proof it, let's try simple case first: *Proof.* [proofoutline.]

- Let's consider $\boldsymbol{V}=\mathbb{R}$ case first:

  For $\mathbb{R}$, 1 forms a basis.

Given any two vectors x and y, they are not a basis for $\mathbb{R}$. It is because 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.

  We want to show if $v_1, v_2, \dots, v_m$ is a basis, then $m = 3$.

  – Let's proof $m = 4$ is impossible (4 vectors in $\mathbb{R}^3$ cannot be a basis.):

    We only need to show for $\forall a_1, a_2, a_3, a_4 \in \mathbb{R}^3$ they must be dep.

    $\iff \boldsymbol{Ax} = \boldsymbol{0}$ has nonzero solutions, where $\boldsymbol{A} = \left[\begin{array}{c|c|c|c} a_1 & a_2 & \dots & a_4 \end{array}\right] \in \mathbb{R}^{3\times 4}$.

    By corollary (3.1), it is obviously true.

  – The same argument could show any basis for $\mathbb{R}^3$ satisfies $m \le 3$.

  – Then let's prove $m = 2$ is impossible (2 vectors in $\mathbb{R}^2$ cannot be a basis):

    We only need to show for $\forall a_1, a_2 \in \mathbb{R}^3$, they cannot span the whole space.

    If this is not true, then $\boldsymbol{Ax} = \boldsymbol{b}$ must have solution, where $\boldsymbol{A} = \left[\begin{array}{c|c} a_1 & a_2 \end{array}\right] \in \mathbb{R}^{3\times 2}$.

    However, this kind matrix may have no solution, which forms a contradiction.

  – The same arugment could show any basis for $\mathbb{R}^3$ satisfies $m \ge 3$.

- The same arugment could show any basis for $\mathbb{R}^n$ satisfies $m = n$.

- Next, let's consider general vector space:

  We assume $n < m$ (contradiction).

  We have known $v_1, \dots, v_n$ is a basis, our goal is to show $w_1, \dots, w_m$ cannot form a basis.

  $\Leftarrow \exists\, \boldsymbol{c} = \begin{bmatrix} c_1 & c_2 & \dots & c_m \end{bmatrix}^{\mathrm{T}} \neq \boldsymbol{0}$ s.t.

$$c_1 w_1 + c_2 w_2 + \cdots + c_m w_m = 0. \tag{3.4}$$

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

$$\begin{cases} w_1 = a_{11}v_1 + \cdots + a_{1n}v_n \\ \dots \\ w_m = a_{m1}v_1 + \cdots + a_{mn}v_n \end{cases} \tag{3.5}$$

By (3.5), we can write (3.4) 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, more specifically, we only need to let $\sum_{j=1}^{m} c_j a_{j1} = \sum_{j=1}^{m} c_j a_{j2} = \cdots = \sum_{j=1}^{m} c_j a_{jn} = 0$.
To write it into matrix form, we only need to let system $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{c} = \mathbf{0}$ has solution.
where $\boldsymbol{A} = \begin{bmatrix} a_{ij} \end{bmatrix}_{1 \le i \le m; 1 \le j \le n}$ and $\boldsymbol{c} = \begin{bmatrix} c_1 & c_2 & \dots & c_m \end{bmatrix}^{\mathrm{T}}$.
By corollary (3.1), since $\boldsymbol{A}^{\mathrm{T}}$ is $n \times m$ matrix where $n < m$, it has infinitely nonzero solution.

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 matrix form to express $\sum_{j=1}^{m} c_j a_{j1} = \sum_{j=1}^{m} c_j a_{j2} = \cdots = \sum_{j=1}^{m} c_j a_{jn} = 0$.

Next since all basis 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.

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 $\dim(\boldsymbol{P})$ is not 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?

1. Does plane has more point than a line?

   No, Cantor syas they have the same "number" of points by constructing a one-to-one mapping.

   Furthermore, $\mathbb{R}, \mathbb{R}^2, \ldots, \mathbb{R}^n$ has the same number of points.
2. However, the plane has bigger dimension 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.
5. For 3-body, they can perform dimension attack on you.

3.2.4.1. What is rank?

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

rank=dimension of row space of a matrix.

We will discuss it next lecture.

3.3. Friday

3.3.1. Review

Proposition 3.2 Undetermined system $Ax = b(m < n$ or number of equations < number of un
has no solution or infinitely many solutions.

We want to understand the meaning of rank: number of "true" equations.
Then we introduce definition of *linearly independence* and *linearly dependence.*
Dependence has relation with system:

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

Combining it with proposition (3.2) we derive the corollary:

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

Proposition 3.4 Undetermined system $Ax = b(m \geq n$ or number of equations $\geq$ number of ur
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 says **All basis of a given vector space have the same size.**
Hence 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 constraints:

$$\underbrace{\text{lin. ind.}}_{\text{not too many}} + \underbrace{\text{span the space}}_{\text{not too few}}$$

The **lin. 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:

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

Definition 3.7 [spanning set] $v_1, v_2, \dots, v_n$ is the spanning set of $\boldsymbol{V}$ iff. $\boldsymbol{V} = \text{span}\{v_1, v_2, \dots, v_n\}$. ■

■ **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 they still don't form a basis.

Then 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 I list the proof outline, you should follow the direction to prove it in detail.

Proof. [proofoutline.]

1. Suppose $v_1, v_2, \ldots, v_n$ are ind. and v is arbitarary vector in $\boldsymbol{V}$. Firstly, show that $v_1, v_2, \ldots, v_n, v$ is dep. Thus derive the equation $c_1 v_1 + c_2 v_2 + \cdots + c_n v_n + c_{n+1} v = \mathbf{0}$. Argue that scalar $c_{n+1} \neq 0$. Then express v in form of $v_1, v_2, \ldots, v_n$. It follows that $v_1, v_2, \ldots, v_n$ span $\boldsymbol{V}$.
2. Suppose $v_1, v_2, \ldots, v_n$ span $\boldsymbol{V}$. Assume $v_1, v_2, \ldots, v_n$ are dep. Then show that v_n could be written as form of other $(n-1)$ vectors, it follows that $v_1, v_2, \ldots, v_{n-1}$ still span $\boldsymbol{V}$. If $v_1, v_2, \ldots, v_{n-1}$ 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, $v_1, v_2, \ldots, v_n$ 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. $P_3 \triangleq \{\text{All polynomials with degree} \le 3\} = \text{span}\{1, x, x^2, x^3\} \implies \dim(P) = 4$.
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?

rank is *the number of nonzero pivots of rref of A.*

■ **Example 3.16**

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

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

However, the definition for rank is too complicated, can we define rank of A directly?

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

Answer: Dimension of row space.

Definition 3.8 [column space] The **column space** of a matrix is the subspace of $\mathbb{R}^n$ spanned by the columns.
In other words, suppose $A = \left[\begin{array}{c|c|c} a_1 & \dots & a_n \end{array}\right]$, the column space is given by

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

■

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

The **row space** of $\boldsymbol{A}$ is col$(\boldsymbol{A}^{\mathrm{T}})$, it is the column space of $\boldsymbol{A}^{\mathrm{T}}$.

In other words, suppose $\boldsymbol{A} = \begin{bmatrix} a_1 \\ \hline \dots \\ \hline a_n \end{bmatrix}$, the row space is given by

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

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 $\text{row}(newrows) \subset \text{row}(oldrows)$.

Assuming $\boldsymbol{A} \xRightarrow{\text{Row Transfom}} \boldsymbol{B}$, then we have $\text{row}(\boldsymbol{B}) \subset \text{row}(\boldsymbol{A})$.

Since row transformations are invertible, we also have $\boldsymbol{B} \xRightarrow{\text{Row Transfom}} \boldsymbol{A}$, hence we have $\text{row}(\boldsymbol{A}) \subset \text{row}(\boldsymbol{B})$.

Hence we obtain $\text{row}(\boldsymbol{B}) = \text{row}(\boldsymbol{A})$.

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

Hence we have a much simpler definition for rank:

Definition 3.10 [rank] The **dimension of the column space** is the **rank of a matrix**.

■

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

Theorem 3.4 The row space and column space both have dimension $\boldsymbol{r}$. Sometimes we call $\dim(\operatorname{col}(\boldsymbol{A}))$ as *column rank*, we call $\dim(\operatorname{row}(\boldsymbol{A}))$ as *row rank*.

Thus it says *column rank=row rank= rank*.

In other words, $\dim(\operatorname{row}(\boldsymbol{A})) = \dim(\operatorname{col}(\boldsymbol{A}))$.

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

■ **Example 3.17**

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

We notice that **column rank of** $\boldsymbol{A} = \mathbf{2}$ and **column rank of** $\boldsymbol{U} = \mathbf{2}$.

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

- *Wrong reason:* "$\boldsymbol{A}$ and $\boldsymbol{U}$ has the same column space". This is false. For example, the first column of $\boldsymbol{A}$ is $\begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} \notin \operatorname{col}(\boldsymbol{U})$. The column spaces of $\boldsymbol{A}$ and $\boldsymbol{U}$ are **different**, but the dimension of them are the **same**—-equal to 2.
- *Right reason:* The **same** combinations of the columns are zero (or nonzero) for $\boldsymbol{A}$ and $\boldsymbol{U}$. Say it in another way: $\boldsymbol{Ax} = \mathbf{0}$ iff. $\boldsymbol{Ux} = \mathbf{0}$. In other words, the r pivot columns(of both) are independent; the $(n - r)$ free columns(of both) are dependent.

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

For $\boldsymbol{A}$, column 1 and 3 are also ind.(pivot columns); column 2 and 4 are also dep.(free columns).

We will show **Row transformation doesn't change ind. relations of columns.**

Proposition 3.6 Suppose matrix $\boldsymbol{A}$ is converted into $\boldsymbol{B}$ by row transformation. If a set of columns of $\boldsymbol{A}$ are ind. then so are corresponding columns of $\boldsymbol{B}$.

Proof. Assume $\boldsymbol{A} = \left[\begin{array}{c|c|c} a_1 & \dots & a_n \end{array}\right], \boldsymbol{B} = \left[\begin{array}{c|c|c} b_1 & \dots & b_n \end{array}\right]$.

Without loss of generality (We often denote it as "WLOG"), we assume $a_1, a_2, \dots, a_k$ are ind.(We can achieve it by switching columns.)

We define $\hat{\boldsymbol{A}} = \left[\begin{array}{c|c|c} a_1 & \dots & a_k \end{array}\right], \hat{\boldsymbol{B}} = \left[\begin{array}{c|c|c} b_1 & \dots & b_k \end{array}\right]$.

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

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

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

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

Combining (1) and (2), $\hat{\boldsymbol{B}}\boldsymbol{x} = \boldsymbol{0}$ has only zero solution. Hence $b_1, b_2, \dots, b_k$ are ind.

Thus we can answer why $\boldsymbol{A}$ and $\boldsymbol{U}$ has the same column space dimension:

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

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

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

Since row transformations are invertible, we get $\boldsymbol{B} \xrightarrow{\text{row transform}} \boldsymbol{A}$.

Similarly, $\dim(\text{col}(\boldsymbol{A})) \geq \dim(\text{col}(\boldsymbol{B}))$.

Hence $\dim(\text{col}(\boldsymbol{A})) = \dim(\text{col}(\boldsymbol{B}))$. Thus using proposition (3.5) and (3.7) we can proof theorem (3.4): *Proof.* [Proof for theorem 3.4] Assume $\boldsymbol{A} \xrightarrow{\text{row transform}} \boldsymbol{U}(\text{rref})$.

- Proposition (3.5) $\implies \dim(\text{row}(\boldsymbol{A}) = \dim(\text{row}(\boldsymbol{U}))$.
- Proposition (3.7) $\implies \dim(\text{col}(\boldsymbol{A}) = \dim(\text{col}(\boldsymbol{U}))$.
- Notice that $\dim(\text{row}(\boldsymbol{U}))$ denotes number of pivots, $\dim(\text{col}(\boldsymbol{U}))$ denotes number of pivot columns. Obviously, $\dim(\text{row}(\boldsymbol{U})) = \dim(\text{col}(\boldsymbol{U}))$.

Hence $\dim(\text{row}(\boldsymbol{A})) = \dim(\text{col}(\boldsymbol{A}))$.

R $\dim(\text{row}(\boldsymbol{U}))$ denotes number of pivots, actually, it denotes the number of "true" equations. $\dim(\text{col}(\boldsymbol{U}))$ denotes the number of pivot columns, actually, it denotes the number of "true" variables.

Theorem 3.4 implies number of "true" equations should equal to number of "true" variables.

3.3.3.1. What is null space dimension?

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

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

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

where $\text{rank}(\boldsymbol{A}) = \dim(\text{col}(\boldsymbol{A}))$.

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

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

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

Proposition 3.8 If $\boldsymbol{Ax} = \boldsymbol{b}$ has at least one solution, then $\text{rank}(\boldsymbol{A}) = \text{rank}(\begin{bmatrix} \boldsymbol{A} & \boldsymbol{b} \end{bmatrix})$.

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

Proof. [Proofoutline.]

$$\boldsymbol{Ax} = \boldsymbol{b} \Longleftrightarrow \boldsymbol{b} \in \text{col}(\boldsymbol{A})$$

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

Proposition 3.9 If $\text{rank}(\boldsymbol{A}) \le n - 1$ for $m \times n$ matrix $\boldsymbol{A}$, then $\boldsymbol{Ax} = \boldsymbol{b}$ has no or infinitely many solutions.

Proof. [Proofoutline.]

$$\dim(\text{col}(\boldsymbol{A})) + \dim(N(\boldsymbol{A})) = n \implies \dim(N(\boldsymbol{A})) \ge 1$$

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

Definition 3.11 [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 $U = \begin{bmatrix} I_r & B \\ 0 & 0 \end{bmatrix}$(rref), where B is $r \times (n-r)$ matrix. We can represent the process in matrix notation:

$$PA = U(\text{rref})$$

P is the product of row transformation matrix, which is obviously invertible.

Since A is invertible, we let $A^{-1} = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix}$, where C_1 is an $r \times n$ matrix. Hence

$$P = PI_n = P(AA^{-1}) = (PA)A^{-1} = UA^{-1} = \begin{bmatrix} I_r & B \\ 0 & 0 \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} = \begin{bmatrix} C_1 + BC_2 \\ 0 \end{bmatrix}$$

Since P has $(n-r)$ rows as zero rows, so it is not invertible, which makes contradiction.

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

$$PA = I$$

where P is the product of row transformation matrix. Hence P is the left inverse of A, A is invertible.

3.3.3.2. Matrices of rank 1

■ **Example 3.19**

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

Here $\operatorname{rank}(A) = 1$. ■

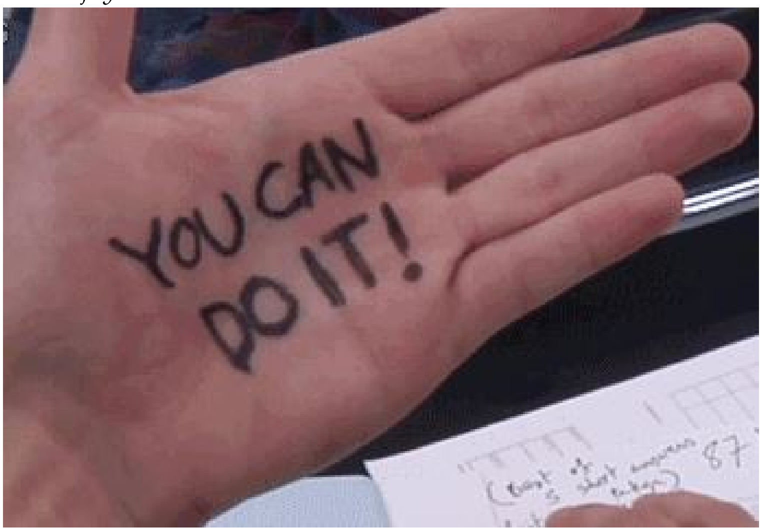
Proposition 3.10 Every rank 1 matrix A has the form $A = uv^{\mathrm{T}}$ = column vector $\times$ row vector.

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 $\boldsymbol{c}_1 = \begin{pmatrix} a_1b_1 & a_1b_2 & \dots & a_1b_n \end{pmatrix}$, where $a_1 \neq 0, b_i (i = 1, \dots, n)$ are not all zero.
Since $\text{rank}(\boldsymbol{A}) = 1$, we have $\dim(\text{row}(\boldsymbol{A})) = 1$. Hence $\boldsymbol{c}_i$ is dep. with $\boldsymbol{c}_1$. So we set $b_i = \frac{a_i}{a_1}$ 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?

Enjoy 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* $\boldsymbol{U}$ of $\boldsymbol{A}$.

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

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

   Note:Identify when there is *no solution*, and when the solution *exists*, write down all solutions in terms of b_1, b_2, b_3.

2. In each of the following, determine the *dimension* of the space:

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

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

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

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

   (e) $\operatorname{span}\{5, \cos 2x, \cos^2 x\}$ as a *subspace* of $C[-\pi, \pi]$.

   $C[-\pi, \pi]$ denotes the space of *continuous functions defined on the domain* $C[-\pi, \pi]$.

3. Let $\boldsymbol{A}$ be an $6 \times n$ matrix of rank r. For each pair of values of r and n below, how many solutions could one have for the linear system $\boldsymbol{Ax} = \boldsymbol{b}$? Explain your answers.

(a) $n = 7, r = 5$;

(b) $n = 7, r = 6$;

(c) $n = 5, r = 5$.

4. Prove the following proposition:

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

   (a) Any set of n *linearly independent* vectors in $\boldsymbol{V}$ form a basis.

   (b) Any set of n vectors that span $\boldsymbol{V}$ form a basis.

   Hint: refer to theorem(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 $\boldsymbol{U} \cap \boldsymbol{V} = \{\boldsymbol{0}\}$, prove that $\dim(\boldsymbol{U} + \boldsymbol{V}) = \dim(\boldsymbol{U}) + \dim(\boldsymbol{V})$.

6. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be $m \times n$ matrices. Prove that

$$\operatorname{rank}(\boldsymbol{A} + \boldsymbol{B}) \leq \operatorname{rank}(\boldsymbol{A}) + \operatorname{rank}(\boldsymbol{B}).$$

7. Let $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is an arbitrary matrix, $\boldsymbol{B} \in \mathbb{R}^{n \times n}$ is a square matrix. Prove that

   (a) $\operatorname{rank}(\boldsymbol{AB}) \leq \operatorname{rank}(\boldsymbol{A})$;

   (b) If $\operatorname{rank}(\boldsymbol{B}) = n$, then $\operatorname{rank}(\boldsymbol{AB}) = \operatorname{rank}(\boldsymbol{A})$.

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 $\boldsymbol{A}$, $\operatorname{rank}(\boldsymbol{A}) + \dim(\operatorname{row}(\boldsymbol{A})) = n$.

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

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \text{ for } \boldsymbol{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

   (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 $\operatorname{rank}(\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{A}) = \operatorname{rank}(\boldsymbol{A})$ 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 $\boldsymbol{A}$, $\operatorname{rank}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{A})$.

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

Start with a matrix $\boldsymbol{A}$. When multiplying $\boldsymbol{A}$ with a vector $\boldsymbol{v}$, it transform $\boldsymbol{v}$ 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 is **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 vector v_1, v_2 and scalar 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 linear transformation $L : \mathbb{R}^n \mapsto \mathbb{R}^m$, let's show that in order to study the output we only need 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 $i = 1, \dots, n$.

Notice that **The rule of linearity extends to combinations of three vectors or $\boldsymbol{n}$ vectors.**

Hence given any vector $\boldsymbol{x} = x_1e_1 + x_2e_2 + \cdots + x_ne_n \in \mathbb{R}^n$, we express its transformation in matrix multiplication form:

$$
\begin{aligned}
L(\boldsymbol{x}) &= L(x_1 e_1 + x_2 e_2 + \cdots + x_n e_n) \\
&= x_1 L(e_1) + x_2 L(e_2) + \cdots + x_n L(e_n) \\
&= x_1 a_1 + x_2 a_2 + \cdots + x_n a_n = \begin{bmatrix} a_1 & a_2 & \ldots & a_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{bmatrix} \\
&= \boldsymbol{A}\boldsymbol{x}
\end{aligned}
$$

where $\boldsymbol{A}$ is a $m \times n$ matrix with columns $a_1, \ldots, a_n$.

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 lienar operator.

R Transformations have a new "language". For example, for nonlinear transformation, if there is **no matrix**, we cannot talk about a **column space**. But this idea could be rescued. We know the *column space* consists of all outputs $\boldsymbol{A}\boldsymbol{v}$, the *nullspace* consists of all inputs for which $\boldsymbol{A}\boldsymbol{v} = \boldsymbol{0}$. We could translate 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 column space. If $L(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$, we have $\text{Range}(L) = \text{col}(\boldsymbol{A})$.

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 nullspace. 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:

$$N(\boldsymbol{A}) \mapsto \{\boldsymbol{0}\}$$
$$\boldsymbol{V} \mapsto \text{col}(\boldsymbol{A})$$

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 produces $L(\boldsymbol{v})$ for every other input vector $\boldsymbol{v}$.

Reason: Every $\boldsymbol{v}$ is a unique combination $c_1\boldsymbol{v}_1+\cdots+c_n\boldsymbol{v}_n$ of the basis vector $\boldsymbol{v}_i$. Suppose L is a linear transformation, $L(\boldsymbol{v})$ must be the **same combination** $c_1L(\boldsymbol{v}_1)+\cdots+c_nL(v_n)$ of the **known outputs** $L(\boldsymbol{v}_i)$.

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 $\boldsymbol{V}$ of *polynomials with degree*≤ 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 $\boldsymbol{v}=a+bx+cx^2+dx^3$. Thus 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 $1, x, x^2$. *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 is $\boldsymbol{A}$ the correct 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 and $2c$ and $3d$ appear when we multiply by matrix $\boldsymbol{A}$:

$$\textit{Take the derivative} \qquad \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} b \\ 2c \\ 3d \end{bmatrix}.$$

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 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 the Fundamental Theorem of Calculus. We see it now in linear Algebra. The integral transformation L^{-1} that *takes* *the integral from 0 to x* is 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:

$$\textbf{Take the integral} \qquad \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix} \begin{bmatrix} B \\ C \\ D \end{bmatrix} = \begin{bmatrix} 0 \\ B \\ \frac{1}{2}C \\ \frac{1}{3}D \end{bmatrix}.$$

If our input is $\boldsymbol{w} = B + Cx + Dx^2$, our integral is $0 + Bx + \frac{1}{2}Cx^2 + \frac{1}{3}Dx^3$.

Recall we have express derivative and integral as matrix:

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} \qquad \boldsymbol{A}^{-1} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix}$$

I want to call this matrix $\boldsymbol{A}^{-1}$, though rectangular matrices don't have inverses. We notice that $\boldsymbol{A}^{-1}$ is the **right inverse** of matrix $\boldsymbol{A}$! (Do you remember the definition that shown in mid-term?)

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \text{but} \qquad \boldsymbol{A}^{-1}\boldsymbol{A} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

This is reasonable. If you integrate a function and then differentiate, you get back to the start. Hence $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}$. But if you differentiate before integrating, the constant term is lost.

The integral of the derivative of 1 is zero:

$$L^{-1}L(1) = \text{integral of zero function} = 0.$$

Summary: In this example, we want to take the derivative. Then we let $\boldsymbol{V}$ be a vector space of polynomials with degree ≤ 3. Then its basis is given by $E = \{1, x, x^2, x^3\}$. Any $v \in \boldsymbol{V}$ there is a unique linear combination of the basis vectors that equals to v:

$$v = a + bx + cx^2 + dx^3$$

We write the coordinate vector of v relative to E:

$$[v]_E = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$$

Then we postmultiply $\boldsymbol{A}$ by $[v]_E$ to get the 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 coordinate vector:

Definition 5.4 [coordinate vector] Let $\boldsymbol{V}$ be a vector space of dimension n and let $B = \{v_1, v_2, \dots, v_n\}$ be an **ordered** basis for $\boldsymbol{V}$. Then for any $v \in \boldsymbol{V}$ there is a unique linear combination of the basis vectors such that

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n$$

where $\alpha_1, \dots, \alpha_n$ are scalars.

The **coordinate vector** of v relative to B is given by

$$[v]_B = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}$$

Hence vector v could be expressed as: $v = \begin{bmatrix} v_1 & v_2 & \dots & v_n \end{bmatrix} \times [v]_B$. ■

Also, there follows a theorem which is easy to verify:

Theorem 5.1 Let $E = \{v_1, \dots, v_n\}$ be a basis for $\boldsymbol{V}$; $F = \{w_1, \dots, w_m\}$ be a basis for $\boldsymbol{W}$. Given linear transformation $L : \boldsymbol{V} \mapsto \boldsymbol{W}$, for any vector $v \in \boldsymbol{V}$, there exists $m \times n$ matrix $\boldsymbol{A}$ such that

$$[L(v)]_F = \boldsymbol{A}[v]_E$$

And there is a corollary that is more commonly useful:

Corollary 5.1 Given linear transformation $L : \boldsymbol{V} \mapsto \boldsymbol{V}$. We set $E = \{\alpha_1, \dots, \alpha_n\}$ to be its basis. 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

Suppose $L : \boldsymbol{V} \mapsto \boldsymbol{V}$. $E = \{v_1, \dots, v_n\}$ is a basis for $\boldsymbol{V}$, $F = \{u_1, \dots, u_n\}$ is another basis for $\boldsymbol{V}$. Then vector $u_1, \dots, u_n$ could be expressed by vectors $v_1, \dots, v_n$. So we set

$$\begin{aligned}
u_1 &= s_{11}v_1 + s_{12}v_2 + \dots + s_{1n}v_n, \\
u_2 &= s_{21}v_1 + s_{22}v_2 + \dots + s_{2n}v_n, \\
&\dots \\
u_n &= s_{n1}v_1 + s_{n2}v_2 + \dots + s_{nn}v_n.
\end{aligned}$$

We could write this system into matrix form:

$$(u_1, \dots, u_n) = (v_1, \dots, v_n) \begin{pmatrix} s_{11} & s_{12} & \dots & s_{1n} \\ s_{21} & s_{22} & \dots & s_{2n} \\ \vdots & \vdots & \dots & \vdots \\ s_{n1} & s_{n2} & \dots & s_{nn} \end{pmatrix}.$$

We set $\boldsymbol{S} = (s_{ij})$. Hence we obtain:

$$(u_1, \ldots, u_n) = (v_1, \ldots, v_n)\boldsymbol{S}. \tag{5.1}$$

You should **prove it by yourself** that $\boldsymbol{S}$ is invertible. Hence we have:

$$(u_1, \ldots, u_n)\boldsymbol{S}^{-1} = (v_1, \ldots, v_n). \tag{5.2}$$

Given any vector $x \in \boldsymbol{V}$, we want to study its transformation relative to its coordinate vector. In other words, we want to study the relationship between $L(x)$ and $[x]_F$:

$$\begin{aligned}
L(x) &= \begin{bmatrix} v_1 & v_2 & \ldots & v_n \end{bmatrix} \times [L(x)]_E \\
&= \begin{bmatrix} v_1 & v_2 & \ldots & v_n \end{bmatrix} \times (\boldsymbol{A}[x]_E) \qquad \leftarrow \text{due to corollary (5.1)} \\
&= \begin{bmatrix} u_1 & u_2 & \ldots & u_n \end{bmatrix} \boldsymbol{S}^{-1} \times (\boldsymbol{A}[x]_E)
\end{aligned}$$

- And we claim that $[x]_E = \boldsymbol{S}[x]_F$:

  For any vector $x \in \boldsymbol{V}$, we obtain:

$$\begin{aligned}
x &= \begin{bmatrix} v_1 & v_2 & \ldots & v_n \end{bmatrix} \times [x]_E \\
&= \begin{bmatrix} u_1 & u_2 & \ldots & u_n \end{bmatrix} \times [x]_F \\
&= \begin{bmatrix} v_1 & v_2 & \ldots & v_n \end{bmatrix} \times \boldsymbol{S}[x]_F
\end{aligned}$$

  Hence $[x]_E = \boldsymbol{S}[x]_F$.

Hence $L(x)$ could be expressed as:

$$L(x) = \begin{bmatrix} u_1 & u_2 & \ldots & u_n \end{bmatrix} \boldsymbol{S}^{-1} \times (\boldsymbol{AS}[x]_F) = \begin{bmatrix} u_1 & u_2 & \ldots & u_n \end{bmatrix} \boldsymbol{S}^{-1}\boldsymbol{AS}[x]_F$$

What do the following process mean? We know that given basis $E = \{v_1, \ldots, v_n\}$, performing linear transformation on any vector x is just the same as matrix multiplication:

$$L(x) = \begin{bmatrix} v_1 & v_2 & \ldots & v_n \end{bmatrix} \times \boldsymbol{A}[x]_E$$

But what about changing another basis $F = \{u_1, \ldots, u_n\}$? Do we still mutliply the same matrix $\boldsymbol{A}$? The answer is no! We must change $\boldsymbol{A}$ into $\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}$:

$$L(x) = \begin{bmatrix} u_1 & u_2 & \ldots & u_n \end{bmatrix} \boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}[x]_F$$

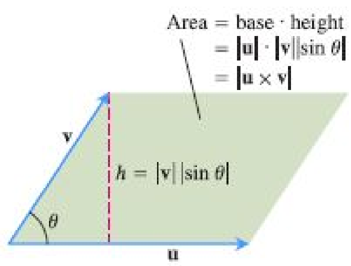
We give a definition for such phenomenon:

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 an 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:**



For example, suppose that $\boldsymbol{u} = u_1\boldsymbol{i} + u_2\boldsymbol{j} + u_3\boldsymbol{k}, \boldsymbol{v} = v_1\boldsymbol{i} + v_2\boldsymbol{j} + v_3\boldsymbol{k}$. In order to compute the area of the parallelogram determined by $\boldsymbol{u}$ and $\boldsymbol{v}$, we just need to compute the determinant $\begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix}$.

- **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:

$$\textbf{Product of pivots} \qquad a(d - \frac{c}{a}b) = ad - bc \qquad \textbf{which is} \quad \det \boldsymbol{A}$$

- **Compute determinants to find $\boldsymbol{A}^{-1}$ and $\boldsymbol{A}^{-1}\boldsymbol{b}$** (This formula is called **Cramer's Rule**.)

5.1.4.1. The properties of the Determinant

We don't intend to define the determinant 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, $\det \boldsymbol{A}$ and $|\boldsymbol{A}|$.

We will introduce three basic properties, then we will show how properties $1-3$ derive other properties.

1. **The determinant of the $\boldsymbol{n}$ by $\boldsymbol{n}$ 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 $1-3$. They completely determine the $\det \boldsymbol{A}$. We could stop here to find a formula for determinants. But 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. *Proof.* [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**.

$$\textbf{Triangular} \qquad \begin{vmatrix} a & b \\ 0 & d \end{vmatrix} = ad \qquad \text{and also} \qquad \begin{vmatrix} a & 0 \\ c & d \end{vmatrix} = ad$$

Suppose all diagonal entries of $\boldsymbol{A}$ are nonzero. We do Gaussian Elimination to convert $\boldsymbol{A}$ into diagonal matrix:

$$\det \begin{bmatrix} a_{11} & & & \text{\huge 0} \\ & a_{22} & & \\ & & \ddots & \\ \text{\huge 0} & & & a_{nn} \end{bmatrix} = a_{11}a_{22}\dots a_{nn}.$$

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 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ and $\boldsymbol{A}$ are the same.

   (b) Show the rank of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}, \boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}, \boldsymbol{A}^{\mathrm{T}}, \boldsymbol{A}$ 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).

$$\textbf{multiply row 1 by any number } \boldsymbol{t} \qquad \begin{vmatrix} ta & tb \\ c & d \end{vmatrix} = t \begin{vmatrix} a & b \\ c & d \end{vmatrix}$$

$$\textbf{Add row 1 of } \boldsymbol{A} \textbf{ to row 1 of } \boldsymbol{B}\textbf{:} \qquad \begin{vmatrix} a_1+a_2 & b_1+b_2 \\ c & d \end{vmatrix} = \begin{vmatrix} a_1 & b_1 \\ c & d \end{vmatrix} + \begin{vmatrix} a_2 & b_2 \\ c & d \end{vmatrix}$$

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 $ad - bc$.

We can break each row into two simpler rows:

$$\begin{vmatrix} a & b \end{vmatrix} = \begin{vmatrix} a & 0 \end{vmatrix} + \begin{vmatrix} 0 & b \end{vmatrix} \qquad \text{and} \qquad \begin{vmatrix} c & d \end{vmatrix} = \begin{vmatrix} c & 0 \end{vmatrix} + \begin{vmatrix} 0 & d \end{vmatrix}$$

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**:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & & \\ & a_{22} & \\ & & a_{33} \end{vmatrix} + \begin{vmatrix} & a_{12} & \\ & & a_{23} \\ a_{31} & & \end{vmatrix} + \begin{vmatrix} & & a_{13} \\ a_{21} & & \\ & a_{32} & \end{vmatrix}$$
$$+ \begin{vmatrix} a_{11} & & \\ & & a_{23} \\ & a_{32} & \end{vmatrix} + \begin{vmatrix} & a_{12} & \\ a_{21} & & \\ & & a_{33} \end{vmatrix} + \begin{vmatrix} & & a_{13} \\ & a_{22} & \\ a_{31} & & \end{vmatrix}$$

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:

$$
\begin{aligned}
\det \boldsymbol{A} = a_{11}a_{22}a_{33} \begin{vmatrix} 1 & & \\ & 1 & \\ & & 1 \end{vmatrix} + a_{12}a_{23}a_{31} \begin{vmatrix} & 1 & \\ & & 1 \\ 1 & & \end{vmatrix} + a_{13}a_{21}a_{32} \begin{vmatrix} & & 1 \\ 1 & & \\ & 1 & \end{vmatrix} \\
+ a_{11}a_{23}a_{32} \begin{vmatrix} 1 & & \\ & & 1 \\ & 1 & \end{vmatrix} + a_{12}a_{21}a_{33} \begin{vmatrix} & 1 & \\ 1 & & \\ & & 1 \end{vmatrix} + a_{13}a_{22}a_{31} \begin{vmatrix} & & 1 \\ & 1 & \\ 1 & & \end{vmatrix}
\end{aligned}
$$

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:

$$
\begin{aligned}
\det \boldsymbol{A} &= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31} \\
&= a_{11}(a_{22} - a_{33}) + a_{12}(a_{23}a_{31} - a_{21}a_{33}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})
\end{aligned}
$$

6.1.1.1. n by n formula

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 terms 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 determinant. $a_{1\alpha}a_{2\beta}\ldots a_{n\omega}$ 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

6.1.1.2. 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 quite more efficient.

6.1.1.3. Verify property

We can 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 1 and column 1. ■

Moreover, we can construct $\det \boldsymbol{A}$ from its properties. 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:

$$\textbf{Cofactors down column j:} \qquad \det \boldsymbol{A} = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}.$$

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 for a_{ij}.

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:

$$\boldsymbol{A}\boldsymbol{C}^{\mathrm{T}} = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix}\begin{bmatrix} C_{11} & \dots & C_{n1} \\ \vdots & \ddots & \vdots \\ C_{1n} & \dots & C_{nn} \end{bmatrix} = \begin{bmatrix} \det \boldsymbol{A} & & \\ & \det \boldsymbol{A} & \\ & & \det \boldsymbol{A} \end{bmatrix} \tag{6.1}$$

Explain:

- 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

$$\begin{array}{l} \textbf{Row 2 of } \boldsymbol{A} \\ \textbf{Row 1 of } \boldsymbol{C} \end{array} \qquad a_{21}C_{11} + a_{22}C_{12} + \cdots + a_{2n}C_{1n} = 0.$$

Answer: 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:

$$\begin{vmatrix} a_{21} & a_{22} & \dots & a_{2n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ a_{31} & a_{32} & \dots & a_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{21} & & & \\ & a_{22} & \dots & a_{2n} \\ & a_{32} & \dots & a_{3n} \\ & \vdots & \ddots & \vdots \\ & a_{n2} & \dots & a_{nn} \end{vmatrix} + \begin{vmatrix} & a_{22} & & \\ a_{21} & & \dots & a_{2n} \\ a_{31} & & \dots & a_{3n} \\ \vdots & & \ddots & \vdots \\ a_{n1} & & \dots & a_{nn} \end{vmatrix} + \cdots + \begin{vmatrix} & & & a_{2n} \\ a_{21} & a_{22} & a_{2(n-1)} & \\ a_{31} & a_{32} & a_{3(n-1)} & \\ \vdots & \vdots & \vdots & \\ a_{n1} & a_{n2} & a_{n(n-1)} & \end{vmatrix}$$

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 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 determinant of submatrix:

Definition 6.4 [Inverse] **The $(\boldsymbol{i}, \boldsymbol{j})$th entry of $\boldsymbol{A}^{-1}$ is the cofactor C_{ji} (not C_{ji}) divided by $\det \boldsymbol{A}$:**

$$\text{Formula for } \boldsymbol{A}^{-1} \qquad (\boldsymbol{A}^{-1})_{ij} = \frac{C_{ji}}{\det \boldsymbol{A}} \quad \text{and} \quad \boldsymbol{A}^{-1} = \frac{\boldsymbol{C}^{\mathrm{T}}}{\det \boldsymbol{A}}.$$

6.1.3.2. Cramer's Rule

Cramer's Rule solves $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$.

Assume $\boldsymbol{A}$ 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 $\boldsymbol{A}$ with a new matrix $\boldsymbol{C}_1$ to get $\boldsymbol{B}_1$:

$$\textbf{Key idea:} \qquad \boldsymbol{A}\boldsymbol{C}_1 = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 & 0 & 0 \\ x_2 & 1 & 0 \\ x_3 & 0 & 1 \end{bmatrix} = \begin{bmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{bmatrix} = \boldsymbol{B}_1$$

Taking determinants both sides, then we have

$$\det(\boldsymbol{A}\boldsymbol{C}_1) = \det(\boldsymbol{A})\det(\boldsymbol{C}_1) = \det(\boldsymbol{A})(x_1) = \det \boldsymbol{B}_1 \implies x_1 = \frac{\det \boldsymbol{B}_1}{\det \boldsymbol{A}_1}.$$

The matrix $\boldsymbol{C}_1$ is obtained by putting $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ into the *first* column of the *identity matrix*.

Similarly, we could get all x_j in this way. $(i = 1, \ldots, n)$.

Definition 6.5 [Cramer's Rule] If $\det \boldsymbol{A}$ is not zero, $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ could be solved by determinants:

$$x_1 = \frac{\det \boldsymbol{B}_1}{\det \boldsymbol{A}} \qquad x_2 = \frac{\det \boldsymbol{B}_2}{\det \boldsymbol{A}} \qquad \cdots\cdots \qquad x_n = \frac{\det \boldsymbol{B}_n}{\det \boldsymbol{A}}$$

The matrix $\boldsymbol{B}_j$ has the jth column of $\boldsymbol{A}$ replaced by the vector $\boldsymbol{b}$. In other words,

$$\boldsymbol{B}_j = \begin{bmatrix} a_{11} & \dots & b_1 & \dots & a_{1n} \\ a_{21} & \dots & b_2 & \dots & a_{2n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{n1} & \dots & b_n & \dots & a_{nn} \end{bmatrix} \qquad j = 1, \dots, n.$$

6.1.4. Orthogonality and Projection

Definition 6.6 [Orthogonal vectors]

Two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ are orthogonal when their inner product is zero:

$$< \boldsymbol{x}, \boldsymbol{y} > = \sum_{i=1}^{n} x_i y_i = 0.$$

(R) Note that the inner product of two vectors satisfies the *commutative rule*. In other words, $< \boldsymbol{x}, \boldsymbol{y} > = < \boldsymbol{y}, \boldsymbol{x} >$ for vectors $\boldsymbol{x}$ and $\boldsymbol{y}$. 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 $< \boldsymbol{x}, \boldsymbol{x} >$ gives the *length of $\boldsymbol{v}$ squared*.

Definition 6.7 [length/norm]

The **length(norm)** $\|\boldsymbol{x}\|$ of a vector $\boldsymbol{x} \in \mathbb{R}$ is the square root of $< \boldsymbol{x}, \boldsymbol{x} >$:

$$\textbf{length} = \|\boldsymbol{x}\| = \sqrt{< \boldsymbol{x}, \boldsymbol{x} >} = \sqrt{x_1^2 + \cdots + x_n^2}.$$

6.1.4.1. Function space

We can talk about inner product between functions under function space. For example, if we define $V = \{f(t) | \int_0^1 f^2(t)dt < \infty\}$, then we can define inner product and norm under V:

Definition 6.8 [Inner product; norm] The **inner product** of $f(x)$ and $g(x)$, and the **norm** are defined as:

$$< f, g > = \int_0^1 f(x)g(x)dx \quad \text{and} \quad \|f\|^2 = \sqrt{\int_0^1 f^2(x)dx}$$

Moreover, when $< f, g > = 0$, 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:

$$x_1 = \|\boldsymbol{x}\| \cos\theta \quad x_2 = \|\boldsymbol{x}\| \sin\theta \qquad y_1 = \|\boldsymbol{y}\| \cos\varphi \quad y_2 = \|\boldsymbol{y}\| \sin\varphi$$

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 great inequality:

Theorem 6.1 — Cauchy Schwarz Inequality.

$| < \boldsymbol{x}, \boldsymbol{y} > | \leq \|\boldsymbol{x}\|\|\boldsymbol{y}\|$ for two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$.

Proof. Firstly, we want to find t^* such that $\min \|\boldsymbol{x} - t\boldsymbol{y}\|^2 = \|\boldsymbol{x} - t^*\boldsymbol{y}\|^2$.

$$\begin{aligned}
\|\boldsymbol{x} - t\boldsymbol{y}\|^2 &= <\boldsymbol{x} - t\boldsymbol{y}, \boldsymbol{x} - t\boldsymbol{y}> = <\boldsymbol{x}, \boldsymbol{x}> + <-t\boldsymbol{y}, \boldsymbol{x}> + <\boldsymbol{x}, -t\boldsymbol{y}> + <-t\boldsymbol{y}, -t\boldsymbol{y}> \\
&= \|\boldsymbol{x}\|^2 - t <\boldsymbol{y}, \boldsymbol{x}> - t <\boldsymbol{x}, \boldsymbol{y}> + t^2\|\boldsymbol{y}\|^2 \\
&= \|\boldsymbol{x}\|^2 - 2t <\boldsymbol{x}, \boldsymbol{y}> + t^2\|\boldsymbol{y}\|^2
\end{aligned}$$

Hence the minimizer $t*$ must satisfy

$$\Delta = 0 \implies t^* = \frac{<\boldsymbol{x}, \boldsymbol{y}>}{\|\boldsymbol{y}\|^2}$$

Hence we have

$$\begin{aligned}
\|\boldsymbol{x} - t\boldsymbol{y}\|^2_{\min} = \|\boldsymbol{x} - t^*\boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 - \frac{<\boldsymbol{x}, \boldsymbol{y}>^2}{\|\boldsymbol{y}\|^2} \\
= \frac{\|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2 - <\boldsymbol{x}, \boldsymbol{y}>^2}{\|\boldsymbol{y}\|^2} \geq 0 \\
\implies \|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2 \geq <\boldsymbol{x}, \boldsymbol{y}>^2
\end{aligned}$$

Or equivalently,

$$| <\boldsymbol{x}, \boldsymbol{y}> | \leq \|\boldsymbol{x}\|\|\boldsymbol{y}\|.$$

If we consider functions f, g as vectors, then **Cauchy-Schwarz** inequality also holds:

$$\left[\int_0^1 f(t)g(t)dt\right] \leq \int_0^1 f^2 dt \int_0^1 g^2 dt$$

Since $| <\boldsymbol{x}, \boldsymbol{y}> | \leq \|\boldsymbol{x}\|\|\boldsymbol{y}\|$, we have

$$-1 \leq \frac{<\boldsymbol{x}, \boldsymbol{y}>}{\|\boldsymbol{x}\|\|\boldsymbol{y}\|} \leq 1$$

If we define $\frac{<\boldsymbol{x}, \boldsymbol{y}>}{\|\boldsymbol{x}\|\|\boldsymbol{y}\|} = \cos\theta$, then $<\boldsymbol{x}, \boldsymbol{y}> = \|\boldsymbol{x}\|\|\boldsymbol{y}\|\cos\theta$.

R This equality holds for **Hilbert space**, which will be discussed later.

6.1.4.3. Orthogonal for space

Also, we can discuss 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 $\boldsymbol{V}$:

$$\textbf{Orthogonal subspaces} \qquad \boldsymbol{u}^{\mathrm{T}}\boldsymbol{v} = 0 \text{ for all } \boldsymbol{u} \text{ in } \boldsymbol{U} \text{ and all } \boldsymbol{v} \text{ in } \boldsymbol{V}.$$

6.2. Thursday

6.2.1. Orthogonality and Projection

Two vectors are orthogonal if their inner product is zero:

$$\boldsymbol{u} \perp \boldsymbol{v} \iff \langle \boldsymbol{u}, \boldsymbol{v} \rangle = 0 \qquad (\text{if } \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n, \text{ then } \boldsymbol{u}^{\mathrm{T}}\boldsymbol{v} = 0.)$$

And orthogonality among vectors has an important property:

Proposition 6.1

If **nonzero** vectors $v_1, \ldots, v_k$ are mutually orthogonal (mutulally means $v_i \perp v_j$ for any $i \neq j$), then $\{v_1, \ldots, v_k\}$ must be ind.

Proof. We only need to show that

$$\text{if } \alpha_1 v_1 + \cdots + \alpha_k v_k = \mathbf{0}, \qquad \text{then } \alpha_i = 0 \text{ for any } i \in \{1, 2, \ldots, k\}.$$

- 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 $i = 1, \ldots, k$.

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 $\boldsymbol{V}$:

$$\textbf{Orthogonal subspaces} \qquad \boldsymbol{u} \perp \boldsymbol{v} \quad \forall \boldsymbol{u} \in \boldsymbol{U}, \boldsymbol{v} \in \boldsymbol{V}.$$

■ **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. Two planes (dimensions 2 and 2 in $\mathbb{R}^3$) cannot be orthogonal subspaces.

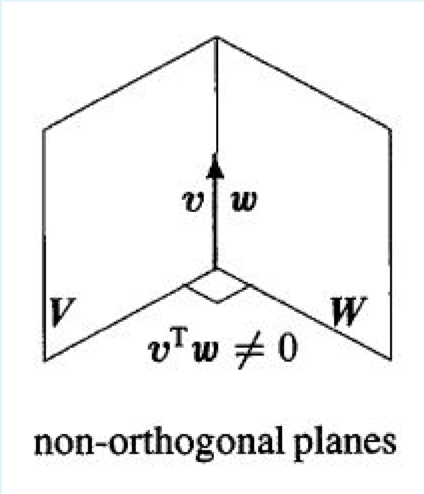


Figure 6.1: Orthogonality is impossible when $\dim \boldsymbol{U} + \dim \boldsymbol{V} > \dim(\boldsymbol{U} \cup \boldsymbol{V})$

■

(R) When a vector is in two orthogonal subspaces, it *must* be zero. It is **perpendicular** to itself.

The reason is clear: this vector $\boldsymbol{u} \in \boldsymbol{U}$ and $\boldsymbol{u} \in \boldsymbol{V}$, so $\langle \boldsymbol{u}, \boldsymbol{u} \rangle = 0$. It has to be zero vector.

If two subspaces are perpendicular, their basis must be ind.

Theorem 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

$$\alpha_1 u_1 + \dots + \alpha_k u_k + \beta_1 v_1 + \dots + \beta_l v_l = \mathbf{0}$$

then equibalently,

$$\alpha_1 u_1 + \cdots + \alpha_k u_k = -(\beta_1 v_1 + \cdots + \beta_l v_l)$$

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 If $\{u_1, u_2, \dots, u_k, v_1, v_2, \dots, v_l\} \in \boldsymbol{W}$, then $\dim(\boldsymbol{W}) \geq \dim(\boldsymbol{U}) + \dim(\boldsymbol{V})$. Note that $\boldsymbol{U} \cup \boldsymbol{V} \subset \boldsymbol{W}$.

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 $\boldsymbol{U}$. And 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 $\boldsymbol{V} = \text{span}\{\boldsymbol{e}_3\}$.

Moreover, $\boldsymbol{U}$ could also be expressed as $\text{span}\left\{\begin{pmatrix}1\\1\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}\right\}$. ■

Example:

Next let's show the nullspace is the orthogonal complement of the row space. (In $\mathbb{R}^n$).

Suppose $\boldsymbol{A}$ is a $m \times n$ matrix.

- Firstly, we show $\dim(N(\boldsymbol{A})) + \dim(C(\boldsymbol{A}^{\mathrm{T}})) = \dim(N(\boldsymbol{A}) \cup C(\boldsymbol{A}^{\mathrm{T}})) = \dim(\mathbb{R}^n) = n$:

We know $\dim(N(\boldsymbol{A})) = n - r$, where $r = \text{rank}(\boldsymbol{A})$. And $r = C(\boldsymbol{A}^{\mathrm{T}}))$.

Hence $\dim(N(\boldsymbol{A})) + \dim(C(\boldsymbol{A}^{\mathrm{T}})) = n$.

- Then we show $N(\boldsymbol{A}) \perp C(\boldsymbol{A}^{\mathrm{T}})$:

  For any $x \in N(\boldsymbol{A})$, if we set $\boldsymbol{A} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}$, then we obtain:

$$\boldsymbol{A}\boldsymbol{x} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

  Hence *every row has a zero product with* x. In other words, $\langle a_i, x \rangle = 0$ for $\forall i \in \{1, 2, \dots, m\}$.

  Hence for any $y = \sum_{i=1}^m \alpha_i a_i \in C(\boldsymbol{A}^{\mathrm{T}})$, we obtain:

$$\begin{aligned} \langle x, y \rangle = \langle y, x \rangle &= \langle \sum_{i=1}^m \alpha_i a_i, x \rangle \\ &= \sum_{i=1}^m \alpha_i \langle a_i, x \rangle = 0. \end{aligned}$$

  Hence $x \perp y$ for $\forall x \in N(\boldsymbol{A})$ and $y \in C(\boldsymbol{A}^{\mathrm{T}})$.

Hence $N(\boldsymbol{A})^{\perp} = C(\boldsymbol{A}^{\mathrm{T}})$.

If we applying this equation to $\boldsymbol{A}^{\mathrm{T}}$, then we have $N(\boldsymbol{A}^{\mathrm{T}})^{\perp} = C(\boldsymbol{A})$.

Theorem 6.3 — **Fundamental theorem for linear alegbra, part 2.**

$N(\boldsymbol{A})$ is the orthogonal complement of the row space $C(\boldsymbol{A}^{\mathrm{T}})$ (in $\mathbb{R}^n$).

$N(\boldsymbol{A}^{\mathrm{T}})$ is the orthogonal complement of the row space $C(\boldsymbol{A})$ (in $\mathbb{R}^m$).

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.

$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ is solvable. $\iff \boldsymbol{b} \in C(\boldsymbol{A})$. $\iff \boldsymbol{b} \in N(\boldsymbol{A}^{\mathrm{T}})^{\perp}$

$\iff \boldsymbol{y}^{\mathrm{T}}\boldsymbol{b} = 0$ for $\forall y \in N(\boldsymbol{A}^{\mathrm{T}})$ $\iff$ $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{0}$ implies $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{b}$=0. The Inverse Negative Propositions is more important:

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 $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{b} \neq 0$.

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 exists one linear combination of the row space to be zero.

Proof. [Necessity case.] Suppose $\exists \boldsymbol{y} \geq \boldsymbol{0}$ such that $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{0}$ and $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{b} \geq \boldsymbol{0}$. And we 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! The complete proof for this theorem is not required in this course.

■ **Example 6.5** Given the system

$$\begin{aligned} x_1 + x_2 &\geq 1 \\ -x_1 &\geq -1 \\ -x_2 &\geq 2 \end{aligned} \tag{6.2}$$

Eq(1)×1+Eq(2)×1+Eq(3)×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. Then if the right hand is larger than zero, then this system has no solution. ■

R

Corollary 6.4 If $A = A^{\mathrm{T}}$, then $N(A^{\mathrm{T}})^{\perp} = C(A) = C(A^{\mathrm{T}}) = N(A)$.

Corollary 6.5 The system $Ax = b$ may not have a solution, but $A^{\mathrm{T}}Ax = A^{\mathrm{T}}b$ always have at least one solution for $\forall b$.

Proof. Since $A^{\mathrm{T}}A$ is symmetric, we have $C(A^{\mathrm{T}}A) = C(AA^{\mathrm{T}})$. You can check by yourself that $C(AA^{\mathrm{T}}) = C(A^{\mathrm{T}})$. Hence $C(A^{\mathrm{T}}A) = C(A^{\mathrm{T}})$.

For any vector b we have $A^{\mathrm{T}}b \in C(A^{\mathrm{T}}) \implies A^{\mathrm{T}}b \in C(A^{\mathrm{T}}A)$, which means there exists a linear combination of the columns of $A^{\mathrm{T}}A$ that equals to b.

Equivalently, there exists a solution to $A^{\mathrm{T}}Ax = A^{\mathrm{T}}b$.

Corollary 6.6 $A^{\mathrm{T}}A$ is invertible if and only if columns of A are ind.

Proof. We have shown that $C(A^{\mathrm{T}}A) = C(A^{\mathrm{T}})$.

Hence $C(A^{\mathrm{T}}A)^{\perp} = C(A^{\mathrm{T}})^{\perp} \implies N(A^{\mathrm{T}}A) = N(A)$.

A has ind. columns $\iff N(A) = \{0\} \iff N(A^{\mathrm{T}}A) = \{0\} \iff A^{\mathrm{T}}A$ is invertible.

6.2.2. Least Squares Approximations

$Ax = b$ often has no solution, if so, what should we do?

We cannot always get the error $e = b - Ax$ down to zero, so we want to use least square method to minimize the error. In other words, our goal is to

$$\min_{x} e^2 = \min_{x} \|Ax - b\|^2 = \sum_{i=1}^{m} (a_i^{\mathrm{T}} x - b_i)^2$$

where $\boldsymbol{A} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}$ and $\boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$.
The minimizer $\boldsymbol{x}$ is called **linear least squares solution**.

6.2.2.1. Matrix Calculus

Firstly, you should know some basic calculus knowledge for matrix:

- $\frac{\partial (f^{\mathrm{T}} g)}{\partial x} = \frac{\partial f(x)}{\partial x} g(x) + \frac{\partial g(x)}{\partial x} f(x)$

Example:

- $\frac{\partial (a^{\mathrm{T}} \boldsymbol{x})}{\partial \boldsymbol{x}} = a$
- $\frac{\partial (a^{\mathrm{T}} \boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}} = \frac{\partial ((\boldsymbol{A}^{\mathrm{T}} a)^{\mathrm{T}} \boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{A}^{\mathrm{T}} a$
- $\frac{\partial (\boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{A}^{\mathrm{T}}$
- $\frac{\partial (\boldsymbol{x}^{\mathrm{T}} \boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{A}^{\mathrm{T}} \boldsymbol{x}$

Thus, in order to minimize $\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2 = (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})$, we only need to let its **partial derivative** with respect to $\boldsymbol{x}$ to be **zero.** (Since its second derivative is non-negative, we will talk about it 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. And this equation is called **normal equation**.

Theorem 6.5 The partial derivatives of $\|\boldsymbol{Ax}-\boldsymbol{b}\|^2$ are **zero** when $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$.

6.2.2.2. Fit a stright line

Given a collection of data $(\boldsymbol{x}_i, y_i)$ for $i=1,\dots,m$, we can fit the model parameters:

$$\begin{cases} y_1 = a_0 + a_1x_{1,1} + a_2x_{1,2} + \cdots + a_nx_{1,n} + \varepsilon_1 \\ y_2 = a_0 + a_1x_{2,1} + a_2x_{2,2} + \cdots + a_nx_{2,n} + \varepsilon_2 \\ \quad\vdots \\ y_m = a_0 + a_1x_{m,1} + a_2x_{m,2} + \cdots + a_nx_{m,n} + \varepsilon_m \end{cases}$$

Our fit line is

$$\hat{y} = a_0 + a_1x_1 + a_2x_2 + \cdots + a_nx_n$$

In *compact matrix form*, we have

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \dots & x_{1,n} \\ 1 & x_{2,1} & x_{2,2} & \dots & x_{2,n} \\ \vdots & \vdots & & & \\ 1 & x_{m,1} & x_{m,2} & \dots & x_{m,n} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix}$$

Or equivalently, we have

$$\boldsymbol{y} = \boldsymbol{Ax} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{A} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \dots & x_{1,n} \\ 1 & x_{2,1} & x_{2,2} & \dots & x_{2,n} \\ \vdots & \vdots & & & \\ 1 & x_{m,1} & x_{m,2} & \dots & x_{m,n} \end{bmatrix}_{m\times(n+1)}$, $\boldsymbol{x} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}_{(n+1)\times 1}$, $\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix}_{m\times 1}$.

Our goal is to minimize $\|\hat{\boldsymbol{y}}-\boldsymbol{y}\|^2 = \|\boldsymbol{Ax}-\boldsymbol{y}\|^2$. Then by theorem (6.5), we only need to sovle $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}$.

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 unique solution $\boldsymbol{x}^* = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$. Thus the error $\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^*$ is minimum. And $\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ **approximately** equals to $\boldsymbol{b}$.

- If $\boldsymbol{b}$ and $\boldsymbol{A}\boldsymbol{x}^*$ are exactly in the same space, then $\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{b}$.
- Otherwise, just as the Figure (6.2) shown, $\boldsymbol{A}\boldsymbol{x}^*$ is the projection of $\boldsymbol{b}$ to subspace $C(\boldsymbol{A})$.

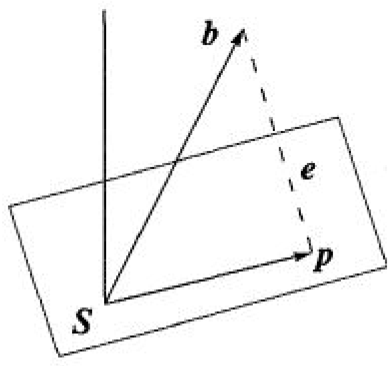


Figure 6.2: The projection of $\boldsymbol{b}$ onto a subspace $C(\boldsymbol{A})$.

Definition 6.12 [Projection] The projection of $\boldsymbol{b}$ onto the subspace $C(\boldsymbol{A})$ is denoted as $\text{Proj}_{C(\boldsymbol{A})}(\boldsymbol{b})$. ■

Definition 6.13 [Projection matrix] Given $\boldsymbol{A}\boldsymbol{x}^* = \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} = \text{Proj}_{C(\boldsymbol{A})}(\boldsymbol{b})$. Since $[\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}]\boldsymbol{b}$ is the projection of $\boldsymbol{b}$, we call $\boldsymbol{P} = \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}$ as **projection matrix**. ■

Definition 6.14 [Idempotent] Let $\boldsymbol{A}$ be a **square** matrix that satisfies $\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}$, then $\boldsymbol{A}$ is called a **idempotent** matrix. ■

Let's show 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

- If $\boldsymbol{b} \in C(\boldsymbol{A})$, then $\exists \boldsymbol{x}$ s.t. $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$. Moreover, 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 $\boldsymbol{a}$. (Shown in figure below.)

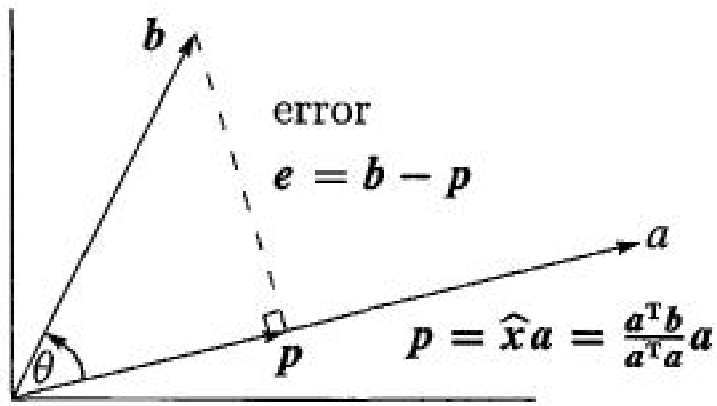


Figure 6.3: The projection of $\boldsymbol{b}$ onto a line $\boldsymbol{a}$.

More generally, we can write the projection of $\boldsymbol{b}$ as:

$$\text{Proj}_{\boldsymbol{a}}(\boldsymbol{b}) = \frac{\langle \boldsymbol{a}, \boldsymbol{b} \rangle}{\langle \boldsymbol{a}, \boldsymbol{a} \rangle} \boldsymbol{a}$$

Look at the figure above! The error is $\boldsymbol{b} - \text{Proj}_{\boldsymbol{a}}(\boldsymbol{b})$, which is obviously perpendicular to $\boldsymbol{a}$. And $\boldsymbol{b} - \text{Proj}_{\boldsymbol{a}}(\boldsymbol{b}) \in \text{span}\{\boldsymbol{a}, \boldsymbol{b}\}$.

If we define $\boldsymbol{b}' = \boldsymbol{b} - \text{Proj}_{\boldsymbol{a}}(\boldsymbol{b})$, then it's easy to check $\text{span}\{\boldsymbol{a}, \boldsymbol{b}'\} = \text{span}\{\boldsymbol{a}, \boldsymbol{b}\}$ and $\boldsymbol{a} \perp \boldsymbol{b}'$. Hence we convert a basis to another basis such that the elements are orthogonal to each other. We will discuss it in detail in next lecture.

6.3. Friday

This lecture has two goals. The first is to see **how orthogonality makes it easy to find projection matrix $\boldsymbol{P}$ and the projection** $\text{Proj}_{C(\boldsymbol{A})}\boldsymbol{b}$. *Orthogonality makes the product* $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ *a diagonal matrix*. The second goal is to **show how to construct orthogonal vectors**. For matrix $\boldsymbol{A} = \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix}$, the columns may not be orthogonal. Then we 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–just divide each vector by its length, then the vectors become **orthogonal unit vectors**. Their lengths are all 1. Then its basis is called **orthonormal**.

Definition 6.15 [orthonormal] The vectors $\boldsymbol{q}_1, \dots, \boldsymbol{q}_n$ are **orthonormal** if

$$\langle \boldsymbol{q}_i, \boldsymbol{q}_j \rangle = \begin{cases} 0 & \text{when } i \neq j \quad (\textbf{orthogonal} \text{ vectors}), \\ 1 & \text{when } i = j \quad (\textbf{unit} \text{ vectors: } \|\boldsymbol{q}_i\| = 1). \end{cases}$$

Moreover, if $\boldsymbol{q}_1, \dots, \boldsymbol{q}_n$ are **orthonormal**, then the basis $\{\boldsymbol{q}_1, \dots, q_n\}$ is called **orthonormal basis**.

Example 6.6 Unit vectors $\boldsymbol{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \boldsymbol{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \dots, \boldsymbol{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$ is an *orthonormal basis* for $\mathbb{R}^n$.

If we want to express vector $\boldsymbol{b}$ as a linear combination of arbitrary basis $\{\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_n\}$, what should you do?

Answer: To solve the system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, where $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \dots & \boldsymbol{q}_n \end{bmatrix}$.

What if $\{\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_n\}$ form an **orthogonal** basis? How to find solution $\boldsymbol{x}$ s.t.

$$\boldsymbol{b} = x_1\boldsymbol{q}_1 + x_2\boldsymbol{q}_2 + \dots + x_n\boldsymbol{q}_n?$$

Answer: We just do the inner product of each $\boldsymbol{q}_i$ with $\boldsymbol{b}$ to get the coefficient x_i:

$$\begin{aligned}\langle \boldsymbol{q}_i, \boldsymbol{b}\rangle &= x_1\langle \boldsymbol{q}_i, \boldsymbol{q}_1\rangle + x_2\langle \boldsymbol{q}_i, \boldsymbol{q}_2\rangle + \dots + x_n\langle \boldsymbol{q}_i, \boldsymbol{q}_n\rangle \\ &= x_i\langle \boldsymbol{q}_i, \boldsymbol{q}_i\rangle = x_i\end{aligned}$$

Since $x_i = \langle \boldsymbol{q}_i, \boldsymbol{b}\rangle$, 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, since $\{\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_n\}$ forms a basis, the columns of $\boldsymbol{A}$ must be ind. Hence $\boldsymbol{A}$ is invertible, then we get the solution to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$:

$$\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}. \tag{6.3}$$

Definition 6.16 [matrix with orthonormal columns]
Define $\boldsymbol{Q} = \begin{bmatrix} q_1 & q_2 & \dots & q_n \end{bmatrix}$. If vectors $\boldsymbol{q}_1, \dots, \boldsymbol{q}_n$ are **orthonormal**, then we say $\boldsymbol{Q}$ is a matrix with **orthonormal** columns.

R Note that a matrix with **orthonormal** columns is often denoted as $\boldsymbol{Q}$.

Such matrix **is easy to work with** because we have:

$$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{q}_1^{\mathrm{T}} \\ \boldsymbol{q}_2^{\mathrm{T}} \\ \dots \\ \boldsymbol{q}_n^{\mathrm{T}} \end{pmatrix} \begin{pmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \dots & \boldsymbol{q}_n \end{pmatrix} = \begin{pmatrix} \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{q}_1 & & \\ & \ddots & \\ & & \boldsymbol{q}_n^{\mathrm{T}}\boldsymbol{q}_n \end{pmatrix} = \boldsymbol{I}. \tag{6.4}$$

R Note that a matrix with orthonormal columns Q is *not required to be square*! Moreover, $\{q_1, \dots, q_n\}$ in Q is *not required to form a basis*.

Definition 6.17 [orthogonal matrix] An **square** that is a *matrix with orthonormal columns* is called **othogonal matrix**. ■

■ **Example 6.7**

If Q is a orthogonal matrix, while $\hat{Q}$ is a matrix with orthonormal columns that is **not square**. Do the products QQ^{T} and $\hat{Q}\hat{Q}^{\mathrm{T}}$ always be *identity matrix*?

Answer:

- QQ^{T} is always *identity matrix*. According to equation (6.4), we have $Q^{\mathrm{T}}Q = I$.

  Hence Q^{T} is the left inverse of square matrix Q.

  Hence $Q^{-1} = Q^{\mathrm{T}} \implies QQ^{\mathrm{T}} = QQ^{-1} = I$.

  Moreover, solving $Qx = b$ is equivalent to $x = Q^{-1}b = Q^{\mathrm{T}}b$, which is *exactly*

$$x = \begin{bmatrix} \langle q_1, b\rangle \\ \langle q_2, b\rangle \\ \vdots \\ \langle q_n, b\rangle \end{bmatrix}.$$

- But the product $\hat{Q}\hat{Q}^{\mathrm{T}}$ will never be identity matrix. Assume $\hat{Q}$ is a $m \times n$ matrix. ($m \neq n$.) Then it's easy to verify that $\mathrm{rank}(\hat{Q}\hat{Q}^{\mathrm{T}}) = \mathrm{rank}(\hat{Q})$.

  Since $\hat{Q}$ has orthonormal columns, the columns of $\hat{Q}$ are ind. Hence $\mathrm{rank}(\hat{Q}) = n$.

  But $\mathrm{rank}(\hat{Q}\hat{Q}^{\mathrm{T}}) = \mathrm{rank}(\hat{Q}) = n \neq m = \mathrm{rank}(I_m)$.

  Moreover, if $\hat{Q}$ has only one column $\hat{q}$, then $\hat{Q}\hat{Q}^{\mathrm{T}} = \hat{q}\hat{q}^{\mathrm{T}} = \mathrm{rank}(1) \neq I_m$.

■

Proposition 6.2

If Q has orthonormal columns, then it *leaves lengths unchanged*, in other words,

$$\textbf{Same length} \qquad \|Qx\| = \|x\| \quad \text{for every vector } x.$$

Also, $\boldsymbol{Q}$ preserves inner products for vectors:

$$\langle \boldsymbol{Q}\boldsymbol{x}, \boldsymbol{Q}\boldsymbol{y}\rangle = \langle \boldsymbol{x}, \boldsymbol{y}\rangle \quad \text{for every vectors } \boldsymbol{x} \text{ and } \boldsymbol{y}.$$

Proof. [Proofoutline.] $\|\boldsymbol{Q}\boldsymbol{x}\|^2 = \|\boldsymbol{x}\|^2$ because

$$\begin{aligned}\langle \boldsymbol{Q}\boldsymbol{x}, \boldsymbol{Q}\boldsymbol{x}\rangle &= \boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q})\boldsymbol{x} \\ &= \boldsymbol{x}^{\mathrm{T}}\boldsymbol{I}\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}\end{aligned}$$

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 numbers can never grow too large when lengths of vectors are fixed.
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 a unique solution $\hat{x} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}$. Why? Since $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{I}$, we derive $\hat{x} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}\hat{x} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}$.

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 $\mathrm{Proj}_{\mathrm{col}(\boldsymbol{Q})}(\boldsymbol{b}) = \boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b}$ doesn't equal to $\boldsymbol{b}$ in general.
For general $\boldsymbol{A}$, the projection matrix is $\boldsymbol{P} = \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}$.

6.3.2. Gram-Schmidt Process

"Orthogonal is good". So our goal for this section is: *Given ind. vectors, how to make them orthonormal?*
We start with three ind. vectors $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ in $\mathbb{R}^3$. In order to construct orthonormal vectors, firstly we construct three **orthogonal** vectors $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$. Then 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 $\boldsymbol{B} = \boldsymbol{b} - \text{Proj}_{\boldsymbol{A}}(\boldsymbol{b})$. Hence

$$\textbf{First Gram-Schmidt step} \qquad \boldsymbol{B} = \boldsymbol{b} - \frac{\langle \boldsymbol{A}, \boldsymbol{b} \rangle}{\langle \boldsymbol{A}, \boldsymbol{A} \rangle} \boldsymbol{A}.$$

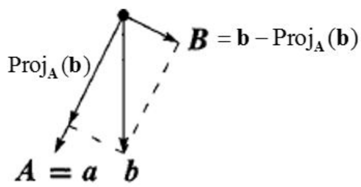
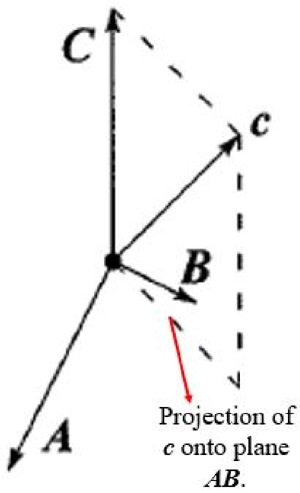


Figure 6.4: Subtract projection to get $\boldsymbol{B} = \boldsymbol{b} - \text{Proj}_{\boldsymbol{A}}\, \boldsymbol{b}$.

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 dep. We will show it later.)

Then we want to construct vector $\boldsymbol{C}$. $\boldsymbol{C}$ is not a linear combination of $\boldsymbol{A}$ and $\boldsymbol{B}$. (Because $\boldsymbol{c}$ is not a linear combination of $\boldsymbol{a}$ and $\boldsymbol{b}$.) But most likely $\boldsymbol{c}$ is **not** perpendicular to $\boldsymbol{A}$ and $\boldsymbol{B}$. Hence we *subtract $\boldsymbol{c}$ off its projections onto the space of $\boldsymbol{A}$ and $\boldsymbol{B}$.* to get $\boldsymbol{C}$:

$$\begin{aligned} & \boldsymbol{C} = \boldsymbol{c} - \text{Proj}_{\text{span}\{\boldsymbol{A},\boldsymbol{B}\}}(\boldsymbol{c}) \\ \textbf{Next Gram-Schmidt step} \qquad & \quad = \boldsymbol{c} - \text{Proj}_{\boldsymbol{A}}(\boldsymbol{c}) - \text{Proj}_{\boldsymbol{B}}(\boldsymbol{c}) \\ & \quad = \boldsymbol{c} - \frac{\langle \boldsymbol{A}, \boldsymbol{c} \rangle}{\langle \boldsymbol{A}, \boldsymbol{A} \rangle} \boldsymbol{A} - \frac{\langle \boldsymbol{B}, \boldsymbol{c} \rangle}{\langle \boldsymbol{B}, \boldsymbol{B} \rangle} \boldsymbol{B}. \end{aligned}$$



Finally we get $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$. Orthonormal vectors $\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3$ are obtained by dividing their lengths (shown in Figure (6.5)):

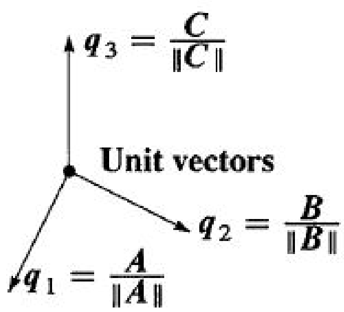


Figure 6.5: Final Gram-Schmidt step

Next we show an example of Gram-Schmidt step:

■ **Example 6.8** How to construct orthonormal vectors for $\boldsymbol{a} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{c} = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}$?

- 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 $\boldsymbol{Q}$:

$$Q = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{pmatrix}$$

But when will the Gram-Schmidt process "fail"? Let's describe this process in general case, then we answer this question.

6.3.2.1. Gram-Schmidt process in general case

Input: Ind. vectors $a_1, \dots, a_n$.

Firstly we want to construct orthogonal vectors $\boldsymbol{A}_1, \dots, \boldsymbol{A}_n$.

In step $j \in \{1, \dots, n\}$, we want to compute a_j minus its projection in the space spanned by $\{\boldsymbol{A}_1, \boldsymbol{A}_2, \dots, \boldsymbol{A}_{j-1}\}$:

$$\begin{aligned}
\boldsymbol{A}_j &= a_j - \text{Proj}_{\text{span}\{\boldsymbol{A}_1, \boldsymbol{A}_2, \dots, \boldsymbol{A}_{j-1}\}}(a_j) \\
&= a_j - \text{Proj}_{\boldsymbol{A}_1}(a_j) - \text{Proj}_{\boldsymbol{A}_2}(a_j) - \cdots - \text{Proj}_{\boldsymbol{A}_{j-1}}(a_j) \\
&= a_j - \frac{\langle \boldsymbol{A}_1, a_j \rangle}{\langle \boldsymbol{A}_1, \boldsymbol{A}_1 \rangle} \boldsymbol{A}_1 - \frac{\langle \boldsymbol{A}_2, a_j \rangle}{\langle \boldsymbol{A}_2, \boldsymbol{A}_2 \rangle} \boldsymbol{A}_2 - \cdots - \frac{\langle \boldsymbol{A}_{j-1}, a_j \rangle}{\langle \boldsymbol{A}_{j-1}, \boldsymbol{A}_{j-1} \rangle} \boldsymbol{A}_{j-1}
\end{aligned}$$

After we get $\boldsymbol{A}_1, \dots, \boldsymbol{A}_n$, 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 $a_1, a_2, \dots, a_n$ are ind.

Proof. [Proofoutline.] $\boldsymbol{A}_j = \boldsymbol{0} \iff a_j = \text{Proj}_{\text{span}\,\boldsymbol{A}_1, \dots, \boldsymbol{A}_{j-1}}(a_j)$

Hence we only need to prove $\exists j$ s.t. $\boldsymbol{A}_j = \boldsymbol{0}$ if and only if $a_1, a_2, \dots, a_n$ are dep.

Sufficiency. Given $\boldsymbol{A}_j = \boldsymbol{0}$, then $a_j = \text{Proj}_{\text{span}\,\boldsymbol{A}_1, \dots, \boldsymbol{A}_{j-1}}(a_j) \in \text{span}\{\boldsymbol{A}_1, \dots, \boldsymbol{A}_{j-1}\}$. It's easy to verify that $\text{span}\{\boldsymbol{A}_1, \dots, \boldsymbol{A}_{j-1}\} = \text{span}\{a_1, \dots, a_{j-1}\}$. Hence $a_j \in \text{span}\{a_1, \dots, a_{j-1}\}$. Hence $a_1, \dots, a_j$ are dep. Thus $a_1, \dots, a_n$ are dep.

Necessity. Given $a_1, a_2, \dots, a_n$ are dep. Then obviously, $a_n \in \text{span}\{a_1, \dots, a_{n-1}\}$. It's easy to verify that $a_n = \text{Proj}_{\text{span}\{a_1, \dots, a_{n-1}\}}(\boldsymbol{a}_n)$. Thus $a_n = \text{Proj}_{\text{span}\{\boldsymbol{A}_1, \dots, \boldsymbol{A}_{n-1}\}}(\boldsymbol{a}_n) \implies \boldsymbol{A}_n = \boldsymbol{0}$.

6.3.3. The Factorization $A = QR$

We know Gaussian Elimination leads to *LU decomposition*; in fact, Gram-Schmidt process leads to *QR factorization*. These two decomposition methods are quite important in LA, 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 matrix 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 $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$.

In general case, $\boldsymbol{a}_1, \dots, \boldsymbol{a}_k$ are combinations of $\boldsymbol{q}_1, \dots, \boldsymbol{q}_k$ at every step.

(In general suppose $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \dots & \boldsymbol{a}_n \end{bmatrix}, \boldsymbol{Q} = \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \dots & \boldsymbol{q}_n \end{bmatrix}$)

Let's discuss a specific example to show how to do factorization.

■ **Example 6.9** Given $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a} & \boldsymbol{b} & \boldsymbol{c} \end{bmatrix}$, whose columns are ind. We can write $\boldsymbol{A}$ as:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \boldsymbol{q}_3 \end{bmatrix} \begin{bmatrix} \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{a} & \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b} & \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{c} \\ 0 & \boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{b} & \boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{c} \\ 0 & 0 & \boldsymbol{q}_3^{\mathrm{T}}\boldsymbol{c} \end{bmatrix}$$

where $\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3$ are **orthonormal**.

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. ■

We have a theorem about QR faactorization (without proof):

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 a upper triangular matrix (always square).

We postmultiply $\boldsymbol{Q}^{\mathrm{T}}$ both sides for $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$ to obtain $\boldsymbol{R} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A}$. In fact, 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

$$\boldsymbol{R} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A} = \begin{bmatrix} \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{a}_1 & \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{a}_2 & \dots & \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{a}_n \\ 0 & \boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{a}_2 & \dots & \boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{a}_n \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \boldsymbol{q}_n^{\mathrm{T}}\boldsymbol{a}_n \end{bmatrix}$$

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 $\boldsymbol{A} = \boldsymbol{QR}$, then we have a simple way to solve $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$.

Proof. [Explain:] Since we have

$$\begin{aligned} \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} &= \boldsymbol{R}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{R}\boldsymbol{x} = \boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{x} \\ \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} &= \boldsymbol{R}^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{b} \end{aligned}$$

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 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 $\text{span}\{1,x,x^2\} \subset C[-1,1]$, firstly, how to define orthogonal for the basis $\{1,x,x^2\}$?

Pre-requisite: 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, we obtain:

$$\boldsymbol{A} = 1, \qquad \boldsymbol{B} = x, \qquad \boldsymbol{C} = x^2 - \frac{\langle 1,x^2\rangle}{\langle 1,1\rangle}1 - \frac{\langle x,x^2\rangle}{\langle x,x\rangle}x = x^2 - \frac{1}{3}$$

$\boldsymbol{A},\boldsymbol{B},\boldsymbol{C}$ are *orthogonal*. We can divide their length to obtain orthonormal basis:

$$\begin{aligned}
\boldsymbol{q}_1 &= \frac{\boldsymbol{A}}{\|\boldsymbol{A}\|} = \frac{1}{\sqrt{\int_{-1}^{1} 1^2\,\mathrm{d}x}} = \frac{1}{2} \\
\boldsymbol{q}_2 &= \frac{\boldsymbol{B}}{\|\boldsymbol{B}\|} = \frac{x}{\sqrt{\int_{-1}^{1} x^2\,\mathrm{d}x}} = \frac{x}{2/3} = \frac{3}{2}x \\
\boldsymbol{q}_3 &= \frac{\boldsymbol{C}}{\|\boldsymbol{C}\|} = \frac{x^2-\frac{1}{3}}{\sqrt{\int_{-1}^{1} (x^2-\frac{1}{3})^2\,\mathrm{d}x}} = \frac{x^2-\frac{1}{3}}{\frac{8}{45}} = \frac{45x^2-15}{8}
\end{aligned}$$

Hence $\{\boldsymbol{q}_1,\boldsymbol{q}_2,\boldsymbol{q}_3\}$ is the orthonormal basis for $\text{span}\{1,x,x^2\}$. ■

■ **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 $\int_0^{2\pi} fg \, dx = 0$. For example,

$$\langle \sin x, \sin 2x \rangle = \int_0^{2\pi} \sin x \sin 2x \, dx = \int_0^{2\pi} \frac{1}{2}(\cos x - \cos 3x) \, dx = 0.$$

And moreover, if $f = g$, we have $\int_0^{2\pi} f^2 \, dx = \pi$. For example,

$$\langle \sin 5x, \sin 5x \rangle = \int_0^{2\pi} \sin^2 5x \, dx = \int_0^{2\pi} \frac{1}{2}(1 + \cos 10x) \, dx = \pi.$$

In conclusion, the collection $\{1, \sin mx, \cos mx\}$ for $k = 1, 2, \dots$ are *orthogonal* in $C[0, 2\pi]$. Note that this set is **not orthonormal**! ■

This example motivates the fourier transformation:

6.3.5. Fourier Series

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 $f(x) \in C[0, 2\pi]$. We have an orthogonal basis! But what kind of function could be expressed in this way? There is a theorem for this condition (without proof):

Theorem 6.7 If a function f have the finite length in its function space $C[a, b]$, 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 6.6: Enjoy fourier series!

$$\langle f(x), \cos x \rangle = a_1 \langle \cos x, \cos x \rangle + 0 \implies a_1 = \frac{\langle f(x), \cos x \rangle}{\langle \cos x, \cos x \rangle} = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos x \, \mathrm{d}x$$

Similarly we derive

$$a_m = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos mx \, \mathrm{d}x \qquad b_m = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin mx \, \mathrm{d}x.$$

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 last two 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. Matrices calculation involve vectors, the columns $\boldsymbol{Ax}$ are linear combination of n vectors–columns of $\boldsymbol{A}$.

7.1.1.1. Determinants

And then we learn how to describle the **quantity of a matrix**–determinant. The determinant of a square matrix is a single number. That 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 $T(\boldsymbol{v}) = \boldsymbol{Av}$ gives a lienar transformation. If we consider a vector as a point in a vector space, then *the linear transformation allows movements in the space*. It "transforms" vector $\boldsymbol{v}$ to another vector $\boldsymbol{Av}$. In 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 $\boldsymbol{k}$ by $\boldsymbol{A}$ secondly by $\boldsymbol{B}$, actually, it has the same effect of transforming $\boldsymbol{k}$ by $\boldsymbol{BA}$.

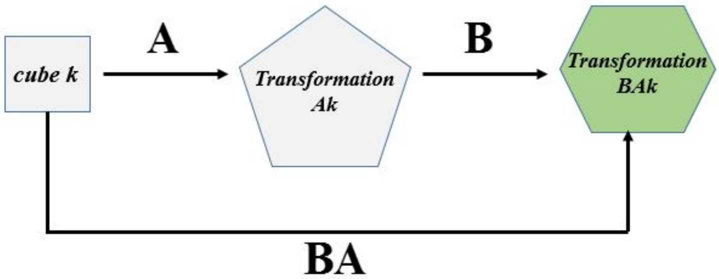


Figure 7.1: Transformation of a vector by $\boldsymbol{A}$, then by $\boldsymbol{B}$ has the same effect by $\boldsymbol{BA}$.

Hence if we denote the volumn on a graph, we find the volumn of $\boldsymbol{B}(\boldsymbol{Ak})$ is exactly the same as $(\boldsymbol{BA})\boldsymbol{k}$. Hence we have $\det(\boldsymbol{B})\det(\boldsymbol{A}) = \det(\boldsymbol{BA})$.
Morevoer, $\det(\boldsymbol{A}) = 0 \Longleftrightarrow \text{Volumn of } \boldsymbol{Ak} = 0 \Longleftrightarrow \dim(\boldsymbol{Ak}) = 0$.
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.)
Linear transformation has a matrix representation under certain basis. How to transform one basis into another basis? We have to use *similar matrices as matrix representation*.

7.1.1.3. Orthogonality

Why we learn orthogonality? It has two motivations:

1. Linear independence between vectors $\Longleftrightarrow$ Angle $\neq 0°$.

   Then we are interested in the special case: orthogonal$\Longleftrightarrow$ Angle $= 90°$.
2. Solving least squares (linear regression).

Input: x =age of propellant, Output: y =shear strength.

Our data contains $S = \{(x_1, y_1), \ldots, (x_n, y_n)\}$, $n = 20$ samples.

We want to find a best line that fit the data:

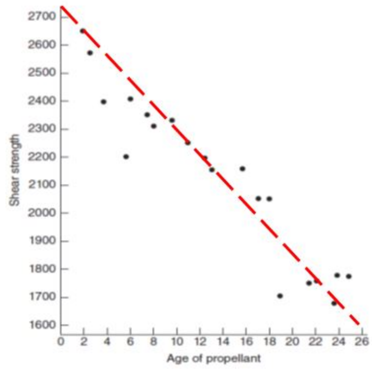


Figure 7.2: The relationship between x and y.

In other words, we want to find $\boldsymbol{x}$ s.t.

$$\Big(\underset{\text{age}}{\boldsymbol{A}} \; \underset{\text{coefficient}}{\boldsymbol{x}} \approx \underset{\text{strength}}{\boldsymbol{b}} \Big)$$

The general least square problem is given by:

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|^2$$

where $b \in \mathbb{R}^m$.

- If $m = n$, this optimization problem is converted into find the solution to equation $Ax = b$.
- Otherwise, the least square solution must satisfy $\frac{\partial}{\partial x}\|Ax - b\|^2 = \mathbf{0}$.

$$\implies A^{\mathrm{T}}Ax = A^{\mathrm{T}}b. \qquad \textbf{(normal equation.)}$$

This potimization problem also has geometric meaning:

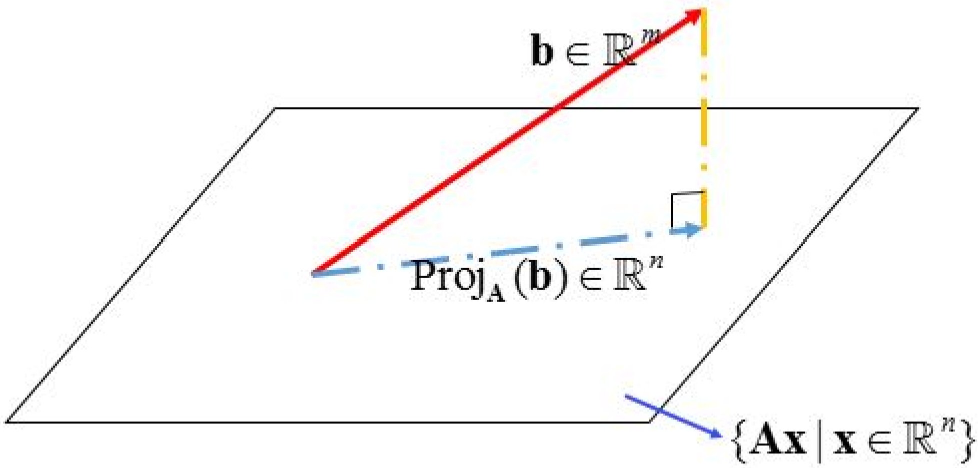


Figure 7.3: Least square problem: find x such that $Ax = \mathrm{Proj}_A(b)$.

So you need to memorize we only need to find x such that $Ax = \mathrm{Proj}_A(b)$. But how to find $\mathrm{Proj}_A(b)$? You can write it as inner product:

$$\mathrm{Proj}_A(b) = A\frac{1}{\langle A, A\rangle}\langle A, b\rangle = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}b.$$

R

- The projection of $\boldsymbol{b}$ onto a vector $\boldsymbol{a}$ is given by:

$$\operatorname{Proj}_{\boldsymbol{a}}(\boldsymbol{b}) = \frac{\langle \boldsymbol{a}, \boldsymbol{b} \rangle}{\langle \boldsymbol{a}, \boldsymbol{a} \rangle} \boldsymbol{a}$$

  Since the factor $\frac{\langle \boldsymbol{a}, \boldsymbol{b} \rangle}{\langle \boldsymbol{a}, \boldsymbol{a} \rangle}$ is a scalar, you can also write the projection as:

$$\operatorname{Proj}_{\boldsymbol{a}}(\boldsymbol{b}) = \boldsymbol{a} \frac{\langle \boldsymbol{a}, \boldsymbol{b} \rangle}{\langle \boldsymbol{a}, \boldsymbol{a} \rangle}$$

- However, the projection of $\boldsymbol{b}$ onto a subspace $\{\boldsymbol{A}\boldsymbol{x} | \boldsymbol{x} \in \mathbb{R}^n\}$ is given by

$$\operatorname{Proj}_{\boldsymbol{A}}(\boldsymbol{b}) = \boldsymbol{A} \frac{1}{\langle \boldsymbol{A}, \boldsymbol{A} \rangle} \langle \boldsymbol{A}, \boldsymbol{b} \rangle$$

  We **cannot** write this projection as $\frac{1}{\langle \boldsymbol{A}, \boldsymbol{A} \rangle} \langle \boldsymbol{A}, \boldsymbol{b} \rangle \boldsymbol{A}$, since the factor $\frac{1}{\langle \boldsymbol{A}, \boldsymbol{A} \rangle} \langle \boldsymbol{A}, \boldsymbol{b} \rangle$ is a vector instead of a scalar.

The least square solution is given by

$$\boldsymbol{x} = (\boldsymbol{A}^{\mathrm{T}} \boldsymbol{A})^{-1} \boldsymbol{A}^{\mathrm{T}} \boldsymbol{b}.$$

If $\boldsymbol{A} = \boldsymbol{Q}$, where $\boldsymbol{Q}$ is a *orthogonal matrix*, then the solution is converted as

$$\boldsymbol{x} = \boldsymbol{Q}^{\mathrm{T}} \boldsymbol{b}.$$

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 one linear transformation? This section gives us answer to this question.

When vectors are multiplied by A, almost all vectors change direction. If $\boldsymbol{x}$ has the same direction as $\boldsymbol{Ax}$, they are called **eigenvectors**.

The key equation is $\boldsymbol{Ax} = \lambda\boldsymbol{x}$, **The numebr λ is the eigenvalue of $\boldsymbol{A}$.**

Definition 7.1 [Eigenvectors and Eigenvalues] Let $\boldsymbol{A}$ be $n \times n$ matrix. A scalar λ is an **eigenvalue** of $\boldsymbol{A}$ iff $\exists$ a vector $\boldsymbol{x} \neq \boldsymbol{0}$ s.t. $\boldsymbol{Ax} = \lambda\boldsymbol{x}$. The vector $\boldsymbol{x}$ is called an **eigenvector** (corresponding to λ.)

■ **Example 7.1**

$$\boldsymbol{A} = \begin{bmatrix} 4 & -2 \\ 1 & 1 \end{bmatrix} \qquad \boldsymbol{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$\boldsymbol{Ax} = \begin{bmatrix} 6 \\ 3 \end{bmatrix} = 3\begin{bmatrix} 2 \\ 1 \end{bmatrix} = 3\boldsymbol{x}.$$

$\lambda = 3$ is the eigenvalue of $\boldsymbol{A}$.

$\boldsymbol{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ is the eigenvalue of $\boldsymbol{A}$ associated with $\lambda = 3$.

Proposition 7.1 If $\boldsymbol{x}$ is an eigenvector of $\boldsymbol{A}$, so is $\alpha\boldsymbol{x}$ for all *nonzero* scalar α. (These vectors have the same eigenvalue.)

7.1.2.2. Calculation

How to find λ and $\boldsymbol{x}$? In other words, how to solve the nonlinear equation $\boldsymbol{Ax} = \lambda\boldsymbol{x}$, where λ and $\boldsymbol{x}$ are unknowns? If we can know the eigenvalues λ, then we can solve the system $(\lambda\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$ to get the corresponding eigenvectors.

But how to find eigenvalues? $\boldsymbol{Ax} = \lambda\boldsymbol{x}$ 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 $\Longleftrightarrow \det(\lambda\boldsymbol{I} - \boldsymbol{A}) = 0$.

This is how to recognize an eigenvalue λ:

Proposition 7.2 The number λ is the eigenvalue of $\boldsymbol{A}$ if and only if $\lambda\boldsymbol{I}-\boldsymbol{A}$ is singular.

$$\textbf{Equation for the eigenvalues} \qquad \det(\lambda\boldsymbol{I}-\boldsymbol{A})=0. \tag{7.1}$$

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 $\boldsymbol{A}$.

And the equation $\det(\lambda\boldsymbol{I}-\boldsymbol{A})=0$ is called the **characteristic equation** for the matrix $\boldsymbol{A}$.

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:

$$P_{\boldsymbol{A}}(\lambda) = \lambda^n - a_1\lambda^{n-1} + a_2\lambda^{n-2} - \cdots + (-1)^n a_n$$

When n increases, it's hard to find its roots:

- When $n = 2$, solution to $ax^2 + bx + c = 0$ has *closed form*, which means we can express x in terms of a, b, c directly.
- When $n = 3$, solution to $ax^3 + bx^2 + cx + d = 0$ has *closed form*, which has been proved in 15th century.
- When $n = 4$, solution to $ax^4 + bx^3 + cx^2 + dx + e = 0$ also has *closed form*.
- However, when $n \geq 5$, the characteristic equation has *no closed form* solution, which has been proved by Galois and Abel.

Although we cannot find closed form solution for large n, does there exist such solution which is not closed form? 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, or say,** $\boldsymbol{A}$ **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. (such as Newton's method.)

7.1.3. Products and Sums of Eigenvalue

Suppose $P_{\boldsymbol{A}}(\lambda) = \det(\lambda \boldsymbol{I} - \boldsymbol{A})$ has n roots $\lambda_1, \dots, \lambda_n$, then we obtain:

$$P_{\boldsymbol{A}}(\lambda) = \det(\lambda \boldsymbol{I} - \boldsymbol{A}) = (\lambda - \lambda_1)\dots(\lambda - \lambda_n) \tag{7.2}$$

Why the coefficient for λ^n is 1 in equation (7.2)? If we expand $\det(\lambda \boldsymbol{I} - \boldsymbol{A})$, we find

$$\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \dots & -a_{nn} \\ -a_{21} & \lambda - a_{22} & \dots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & \dots & \dots & \lambda - a_{nn} \end{vmatrix} \tag{7.3}$$

So λ only appears in diagonal. If we expand the determinant, the coefficient is obviously 1.

Moreover, in (7.2), the coefficient of λ^{n-1} is

$$-(\lambda_1 + \lambda_2 + \cdots + \lambda_n)$$

In (7.3), λ^{n-1} only appears among $(\lambda - a_{11})(\lambda - a_{22})\dots(\lambda - a_{nn})$. Hence the coefficient of λ^{n-1} is

$$-(a_{11} + a_{22} + \cdots + a_{nn})$$

Consequently, we derive

$$\sum \lambda_i = \textbf{trace} = \sum a_{ii}$$

The sum of the entries on the main diagonal is called the **trace** of $\boldsymbol{A}$.

If we let $\lambda = 0$ in (7.2), then we obtain $\det(-\boldsymbol{A}) = (-1)^n \lambda_1 \lambda_2 \dots \lambda_n$.

And obviously, $\det(-\boldsymbol{A}) = (-1)^n \det(\boldsymbol{A})$.

Hence $(-1)^n \det(\boldsymbol{A}) = (-1)^n \lambda_1 \lambda_2 \dots \lambda_n \implies \det(\boldsymbol{A}) = \lambda_1 \lambda_2 \dots \lambda_n$. So we have two useful conclusions:

Theorem 7.2 ***The product of the n eigenvalues equals the determinant.***

The sum of the n eigenvalues equals the sum of the n diagonal entries.

7.1.4. Application: Page Rank and Web Search

If we do keyword search on google, every keyword will return 20k pages. But how to generate more useful pages for us? Our goal is to *compute a vector* $\boldsymbol{x} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}$ *in* $\mathbb{R}^n$*, each* x_i *represents the importance of the page.* Then we only need to generate the most important pages for us.

The information we use: links.

We use the number of links to judge whether a page is important. For example,

- A page *linked to by* 10^5 *pages* is more important than a page *linked by* 10 *pages*.
- If two pages *both linked to by* 100 *pages*, are they the same important?
  The answer is no. For example, in your own blogs, if you create 100 pages that all link to your home page, then obviously your home page is not so important. These 100 pages are created by yourself, which are called **fake pages**.

We do the following assumptions:

- Assume we have $100n$ people are visiting pages. We have n pages.
- We assume every page have 100 visitors. They follow the links on that page.
- And we assume *multiple of links are equally split.* (For example, if one page have 5 links, then there will be $100/5 = 20$ people follow each link.)

To start with, the distribution of people for n pages is given by:

$$\boldsymbol{x}^0 = \begin{pmatrix} x_1^0 \\ x_2^0 \\ \vdots \\ x_n^0 \end{pmatrix}$$

We assume the probability people will go from page j to page i is a_{ij}.
Question: If page j links to by 5 pages, what's a_{ij}?
Answer: due to assumption 3, roughly speaking, $a_{ij} = \frac{1}{5}$.
However, this answer is not absolutely right. Since assumption is not always true. If we consider *stochastic process*, then a_{ij} is given by:

$$a_{ij} = 0.85 \times \frac{1}{5} + 0.15 \times \frac{1}{n}.$$

If we write matrix $\boldsymbol{A} = \begin{bmatrix} a_{ij} \end{bmatrix}_{1 \le i,j \le n}$, then the (i,j) entry of $\boldsymbol{A}$ represents the probability that a random Web surfer will link from page j to page i.

Next step distribution:

Hence when each surfer follow one link of the pages, the distribution of people among pages is given by

$$\boldsymbol{x}^1 = \boldsymbol{A}\boldsymbol{x}^0 = \begin{pmatrix} x_1^1 \\ x_2^1 \\ \vdots \\ x_n^1 \end{pmatrix}$$

In general, we obtain $\boldsymbol{x}^{k+1} = \boldsymbol{A}\boldsymbol{x}^k$.

If the sequence $\{\boldsymbol{x}^k\}$ converges, then take the limit $k \to \infty$ suppose $\lim_{k\to\infty} \boldsymbol{x}^k = \boldsymbol{x}^*$. Then we obtain:

$$\boldsymbol{x}^{k+1} = \boldsymbol{A}\boldsymbol{x}^k \implies \boldsymbol{x}^* = \boldsymbol{A}\boldsymbol{x}^*$$

Hence we only need to get $\boldsymbol{x}^*$, which is the *eigenvector* of $\boldsymbol{A}$.

Once we get the distribution of people for pages, we get the importance of each page.

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* corresponding to λ.

- **How to compute eigenvalues and eigenvectors?** To solve the eigenvalue problem for an n by n matrix, you should follow these steps:
  - *Compute the determinant 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 an eigenvector* $\boldsymbol{x}$.

7.2.2. Similarity and eigenvalues

Which two matrices have the same eigenvalues? 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. Hence 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, all λ_i's are nonzero, and we obain:

$$\boldsymbol{x}_i = \lambda_i \boldsymbol{S}^{-1}\boldsymbol{x}_i \implies \frac{1}{\lambda_i}\boldsymbol{x}_i = \boldsymbol{S}^{-1}\boldsymbol{x}_i$$

Or equivalently, $\boldsymbol{S}^{-1}\boldsymbol{x}_i = \frac{1}{\lambda_i}\boldsymbol{x}_i$. $\frac{1}{\lambda_i}$'s are eigenvalues of $\boldsymbol{S}^{-1}$.

Since S^{-1} is $n \times n$ matrix, $\frac{1}{\lambda_i}$'s $(i = 1, \dots, n)$ are the only eigenvalues of $\boldsymbol{S}^{-1}$.

Hence the determinant of $\boldsymbol{S}^{-1}$ is the product of 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 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

$$\det(\boldsymbol{A}) = 0 \iff \exists \lambda_i = 0 \iff \exists \text{ nonzero } \boldsymbol{x} \text{ s.t. } \boldsymbol{A}\boldsymbol{x} = \lambda_i\boldsymbol{x} = 0\boldsymbol{x} = \boldsymbol{0}.$$

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.

- Q1: *What about the reverse direction?*
- What's the simplest form of matrix to have eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$?

  We can answer this question immediately. The matrix $\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$ has the simplest form. And we often write this matrix as $\operatorname{diag}(\lambda_1, \ldots, \lambda_n)$.

  Q2: What we want to ask is that *if* $\boldsymbol{A}$ *has eigenvalues* $\lambda_1, \ldots, \lambda_n$*, then* $\boldsymbol{A}$ *and* $\operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ *have the same eigenbalues. Are they similar?*

R Why the matrix $\operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ has eigenvalues $\lambda_1, \ldots, \lambda_n$?

Answer: Let's explain it with $n = 2$:

$$\begin{pmatrix} \lambda_1 & \\ & \lambda_2 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \lambda_1 \\ 0 \end{pmatrix} = \lambda_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \begin{pmatrix} \lambda_1 & \\ & \lambda_2 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ \lambda_2 \end{pmatrix} = \lambda_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

General n is also easy to verify.

The answer to question 1 and 2 are both No! Let's raise a counterexample to explain it:

■ **Example 7.3**

If $\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, then $P_{\boldsymbol{A}}(\lambda) = \det(\lambda \boldsymbol{I} - \boldsymbol{A}) = \begin{vmatrix} \lambda & -1 \\ 0 & \lambda \end{vmatrix}$. Hence its eigenvalues are $\lambda_1 = \lambda_2 = 0$.

And $\boldsymbol{A}$ and $\boldsymbol{D} = \operatorname{diag}(0,0)$ have the same eigenvalues. Are they similar?

We assume they are similar, which means there exists invertible matrix $\boldsymbol{S}$ such that

$$\boldsymbol{A} = \boldsymbol{S}^{-1}\boldsymbol{D}\boldsymbol{S} = \boldsymbol{S}^{-1}\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}\boldsymbol{S} = \boldsymbol{0}$$

which leads to a contradiction! So $\boldsymbol{A}$ and $\boldsymbol{D} = \text{diag}(\lambda_1, \lambda_2)$ are not similar. ■

Suppose $\boldsymbol{A}$ has eigenvalues $\lambda_1, \ldots, \lambda_n$, but $\boldsymbol{A}$ and $\text{diag}(\lambda_1, \ldots, \lambda_n)$ may not be similar! But which matrix is similar to its diagonal matrix $\text{diag}(\lambda_1, \ldots, \lambda_n)$?

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

$$\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S} = \boldsymbol{D}$$

We say $\boldsymbol{S}$ *diagonalize* $\boldsymbol{A}$. ■

You should remember the remarks below, they are very important. (We will show the proof for the remarks later.)

R

1. If $\boldsymbol{A}$ 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 $\boldsymbol{A}$.
2. The diagonalizing matrix $\boldsymbol{S}$ is not unique.
3. 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.
An $n \times n$ matrix $\boldsymbol{A}$ is *diagonalizable* iff $\boldsymbol{A}$ has n ind. eigenvectors.

Proof.
Necessity. Suppose $\boldsymbol{A}$ has n ind. eigenvectors $\boldsymbol{x}_i$ for $i = 1, \ldots, n$. And we assume $\exists \lambda_i$ such that

$$\boldsymbol{A}\boldsymbol{x}_i = \lambda_i \boldsymbol{x}_i \text{ for } i = 1, \ldots, n.$$

We 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 $\boldsymbol{A}\boldsymbol{x}_1$, that is $\lambda_1 \boldsymbol{x}_1$. Then we obtain:

$$\textbf{A times S} \qquad \boldsymbol{AS} = \boldsymbol{A} \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \dots & \boldsymbol{x}_n \end{bmatrix} = \begin{bmatrix} \lambda_1 \boldsymbol{x}_1 & \lambda_2 \boldsymbol{x}_2 & \dots & \lambda_n \boldsymbol{x}_n \end{bmatrix}.$$

The trick is to split this matrix $\boldsymbol{AS}$ into $\boldsymbol{S}$ times $\boldsymbol{D}$:

$$\textbf{S times D} \qquad \begin{bmatrix} \lambda_1 \boldsymbol{x}_1 & \lambda_2 \boldsymbol{x}_2 & \dots & \lambda_n \boldsymbol{x}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \dots & \boldsymbol{x}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} = \boldsymbol{SD}.$$

Hence we obtain $\boldsymbol{AS} = \boldsymbol{SD}$. Since $\boldsymbol{x}_i$'s are ind, there exists the inverse $\boldsymbol{S}^{-1}$.
So $\boldsymbol{D} = \boldsymbol{S}^{-1}\boldsymbol{AS}$.
Sufficiency. If $\boldsymbol{A}$ is diagonalizable, then there exists $\boldsymbol{S}$ and $\boldsymbol{D}$ such that

$$\boldsymbol{D} = \boldsymbol{S}^{-1}\boldsymbol{AS}$$

where $\boldsymbol{S}$ is nonsingular. And we assume $\boldsymbol{D} = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$.
Suppose $\boldsymbol{S} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \dots & \boldsymbol{x}_n \end{bmatrix}$, where $\boldsymbol{x}_i$'s are ind.
Then from equation $\boldsymbol{D} = \boldsymbol{S}^{-1}\boldsymbol{AS}$ we obtain $\boldsymbol{AS} = \boldsymbol{SD} \implies \boldsymbol{A}\boldsymbol{x}_i = \lambda_i \boldsymbol{x}_i$ for $i = 1, 2, \dots, n$.
Hence λ_i's are eigenvalues and $\boldsymbol{x}_i$'s are ind. eigenvectors of $\boldsymbol{A}$. For $n \times n$ matrix $\boldsymbol{A}$ which is *diagonalizable*, if its eigenvectors $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ form a basis, then for any $\boldsymbol{y} \in \mathbb{R}^n$, there exists $(c_1, c_2, \dots, c_n)$ such that

$$\boldsymbol{y} = c_1 \boldsymbol{x}_1 + c_2 \boldsymbol{x}_2 + \cdots + c_n \boldsymbol{x}_n$$

If we consider matrix $\boldsymbol{A}$ as representation of linear transformation, we obtain

$$\begin{aligned} \boldsymbol{Ay} &= c_1 \boldsymbol{A}\boldsymbol{x}_1 + \cdots + c_n \boldsymbol{A}\boldsymbol{x}_n \\ &= c_1 \lambda_1 \boldsymbol{x}_1 + \cdots + c_n \lambda_n \boldsymbol{x}_n \end{aligned}$$

So if we transform $\boldsymbol{y}$ into $\boldsymbol{A}\boldsymbol{y}$, it's equivalent to transform the coefficient $(c_1, \ldots, c_n)$ into $(c_1\lambda_1, \ldots, c_n\lambda_n)$.

$$\boldsymbol{y} \xRightarrow{\boldsymbol{A}} \boldsymbol{A}\boldsymbol{y}$$

$$(c_1, \ldots, c_n) \xRightarrow{\boldsymbol{D}=\mathrm{diag}(\lambda_1, \ldots, \lambda_n)} (c_1\lambda_1, \ldots, c_n\lambda_n) = (c_1, \ldots, c_n) \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

But is there an useful way to determine whether the eigenvectors of $\boldsymbol{A}$ is independent?

Theorem 7.4 If $\lambda_1,\dots,\lambda_k$ are *distinct* eigenvalues of a matrix $\boldsymbol{A}$ with 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 $k=2$. We assume $\lambda_1 \neq \lambda_2$ but $\boldsymbol{x}_1,\boldsymbol{x}_2$ are dep.

  That is to say, $\exists (c_1,c_2) \neq \boldsymbol{0}$ s.t.

  $$c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 = \boldsymbol{0}. \tag{7.4}$$

  If we multiply $\boldsymbol{A}$ both sides, we obtain

  $$\boldsymbol{A}(c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2) = \boldsymbol{0} \implies c_1\lambda_1\boldsymbol{x}_1 + c_2\lambda_2\boldsymbol{x}_2 = \boldsymbol{0}. \tag{7.5}$$

  Eq(7.4)$\times\lambda_2-$Eq(7.5):

  $$(c_1\lambda_2 - c_1\lambda_1)\boldsymbol{x} = \boldsymbol{0}. \implies c_1(\lambda_2 - \lambda_1)\boldsymbol{x} = \boldsymbol{0}.$$

  Since $\lambda_1 \neq \lambda_2, \boldsymbol{x} \neq \boldsymbol{0}$, we derive $c_2 = 0$.

  Similarly, if we let Eq(7.4)$\times\lambda_1-$Eq(7.5) to cancel c_2, then we get $c_1 = 0$.

  Hence $(c_1,c_2) = \boldsymbol{0}$ leads to contradiction!

- How to proof this statement for general k?

  Assume there exists $(c_1,\dots,c_k) \neq \boldsymbol{0}$ s.t.

  $$c_1\boldsymbol{x}_1 + \cdots + c_k\boldsymbol{x}_k = \boldsymbol{0} \tag{7.6}$$

  Then

  $$\boldsymbol{A}(c_1\boldsymbol{x}_1 + \cdots + c_k\boldsymbol{x}_k) = c_1\lambda_1 + c_2\lambda_2 + \cdots + c_k\lambda_k\boldsymbol{x}_k = \boldsymbol{0}. \tag{7.7}$$

  We can let $\text{Eq}(11.4) - \lambda_k\times\text{Eq}(11.3)$ to cancel $\boldsymbol{x}_k$:

  $$c_1(\lambda_1 - \lambda_k)\boldsymbol{x}_1 + \cdots + c_k(\lambda_{k-1} - \lambda_k)\boldsymbol{x}_{k-1} = \boldsymbol{0}. \tag{7.8}$$

We can continue this process to cancel $\boldsymbol{x}_{k-1}, \boldsymbol{x}_{k-2}, \ldots, \boldsymbol{x}_2$ to get:

$$c_1(\lambda_1 - \lambda_k)\ldots(\lambda_1 - \lambda_2)\boldsymbol{x}_1 = \mathbf{0} \qquad \text{which forces } c_1 = 0.$$

Similarly every $c_i = 0$ for $i = 1, \ldots, n$. 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 $\boldsymbol{A}$

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, $\boldsymbol{A}^k = (\boldsymbol{S}^{-1}\boldsymbol{D}\boldsymbol{S})\ldots(\boldsymbol{S}^{-1}\boldsymbol{D}\boldsymbol{S}) = \boldsymbol{S}^{-1}\boldsymbol{D}^k\boldsymbol{S}$.

We may ask if eigenvalues of $\boldsymbol{A}$ are $\lambda_1, \ldots, \lambda_n$, then what is the eigenvalues of $\boldsymbol{A}^k$? The answer is intuitive, the eigenvalues of $\boldsymbol{A}^k$ are $\lambda_1^k, \ldots, \lambda_n^k$. But you may use the wrong way to proof this statement:

Proposition 7.5 If eigenvalues of $n \times n$ matrix $\boldsymbol{A}$ are $\lambda_1, \ldots, \lambda_n$, then eigenvalues of $\boldsymbol{A}^k$ are $\lambda_1^k, \ldots, \lambda_n^k$.

Proof. [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 $\boldsymbol{D} = \text{diag}(\lambda_1, \ldots, \lambda_n)$, then $\boldsymbol{D}^k = \text{diag}(\lambda_1^k, \ldots, \lambda_n^k)$. Hence eigenvalues of $\boldsymbol{A}^k$ are $\lambda_1^k, \ldots, \lambda_n^k$.
This proof is wrong, because $\boldsymbol{A}$ may not be *diagonalizable*, which means $\boldsymbol{A}$ may not have the form $\boldsymbol{A} = \boldsymbol{S}^{-1}\boldsymbol{D}\boldsymbol{S}$. *Proof.* [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 $\boldsymbol{A}^k$. But it cannot derive this proposition.
Let's raise a counterexample: Let eigenvalues of $\boldsymbol{A}$ be $\lambda_1 = 1, \lambda_2 = 1, \lambda_3 = 2$, the eigenvalues of $\boldsymbol{A}^2$ be $1^2, 2^2, 2^2$. Then obviously, this $\boldsymbol{A}$ and $\boldsymbol{A}^2$ is a contradiction for this proof. Because 1,2 are the eigenvalues of $\boldsymbol{A}$, but this proof fails to determine its multiplicity!

7.2.5. Nondiagonalizable Matrices

Sometimes we face some matrices that have too few eigenvalues. (don't count with multiplicity)

For example, if $\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, it's easy to verify that its eigenvalue is $\lambda = 0$ and eigenvectors are of the form $\boldsymbol{x} = \begin{bmatrix} c \\ 0 \end{bmatrix}$.

However, this 2×2 matrix cannot be diagonalized. Why? Let's introduce a definition:

Definition 7.5 [Eigenspace] Suppose $\boldsymbol{A}$ has k distinct eigenvalues $\lambda_1, \dots, \lambda_k$. Then the eigenspace for $\boldsymbol{A}$ is the union of all eigenvectors. Or say, the eigenspace is the union of all null space $N(\lambda_i \boldsymbol{I} - \boldsymbol{A})$ for $i = 1, \dots, k$.

Moreover, the null space $N(\lambda_i \boldsymbol{I} - \boldsymbol{A})$ is called the **eigenspace corresponding to eigenvalue** λ_i.

Why this 2 by 2 matrix $\boldsymbol{A}$ cannot be diagonalizable? Because it has two repeated eigenvalues $\lambda_1 = \lambda_2 = 0$. And its eigenspace is of dimension $1 < 2$. In general, if a eigenspace for a $n \times n$ matrix has dimension $k < n$, then it cannot be diagonalizable.

7.3. Friday

7.3.1. Review

- **Diagonalization:** If a $n \times n$ matrix is diagonalizable, it's equivalent to say it has n ind. eigenvectors. So its eigenvectors form a basis for $\mathbb{R}^n$. $(*)$
- If *eigenvalues are distinct*, then $(*)$ holds.

7.3.2. Fibonacci Numbers

We show a famous example, where eigenvalues tell how to find the formula for Fibonacci Numbers.

Every new Fibonacci number come from two previous ones:

$$\textbf{Fibonacci Number: } 0,1,1,2,3,5,8,13,\dots$$

$$\textbf{Fibonacci Equation: } \boldsymbol{F}_{k+2} = \boldsymbol{F}_{k+1} + \boldsymbol{F}_k, \quad \boldsymbol{F}_0 = 0, \boldsymbol{F}_1 = 1.$$

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 $\boldsymbol{A}$:

$$\text{Let } \boldsymbol{u}_k = \begin{bmatrix} \boldsymbol{F}_{k+1} \\ \boldsymbol{F}_k \end{bmatrix}. \text{ The rule } \begin{cases} \boldsymbol{F}_{k+2} = \boldsymbol{F}_{k+1} + \boldsymbol{F}_k \\ \boldsymbol{F}_{k+1} = \boldsymbol{F}_{k+1} \end{cases} \text{ is } \boldsymbol{u}_{k+1} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \boldsymbol{u}_k. \qquad \boldsymbol{u}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Every step we mutliply $\boldsymbol{u}_0$ by $\boldsymbol{A}$. After 100 steps we obtain $\boldsymbol{u}_{100} = \boldsymbol{A}^{100}\boldsymbol{u}_0$:

$$\boldsymbol{u}_{100} = \begin{bmatrix} \boldsymbol{F}_{101} \\ \boldsymbol{F}_{100} \end{bmatrix} = \boldsymbol{A}^{100}\boldsymbol{u}_0 = \boldsymbol{A}^{100} \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

But how to compute $\boldsymbol{A}^{100}$? If possible, you can diagonalize $\boldsymbol{A}$.

It's eay to show that for matrix $\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$, we can decompose it into $\boldsymbol{A} = \boldsymbol{S}\boldsymbol{D}\boldsymbol{S}^{-1}$.

where $\boldsymbol{D} = \operatorname{diag}(\lambda_1, \lambda_2)$, $\boldsymbol{S} = \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix}$.

And $\begin{bmatrix} \lambda_1 \\ 1 \end{bmatrix}$ is the eigenvector corresponding to λ_1, $\begin{bmatrix} \lambda_2 \\ 1 \end{bmatrix}$ is the eigenvector corresponding to λ_2. You can verify $\lambda_1 = \frac{1+\sqrt{5}}{2}, \lambda_2 = \frac{1-\sqrt{5}}{2}$.

Thus we obtain $\boldsymbol{A}^{100} = \boldsymbol{S}\boldsymbol{D}^{100}\boldsymbol{S}^{-1}$. 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

$$\boldsymbol{F}_{100} = \frac{1}{\sqrt{5}} \left[\lambda_1^{100} - \lambda_2^{100} \right] = \frac{1}{\sqrt{5}} \left[\left(\frac{1+\sqrt{5}}{2} \right)^{100} - \left(\frac{1-\sqrt{5}}{2} \right)^{100} \right]$$

Another way to compute $\boldsymbol{F}_{100}$:

We wet $\boldsymbol{S} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 \end{bmatrix}$, where $\boldsymbol{x}_1 = \begin{bmatrix} \lambda_1 \\ 1 \end{bmatrix}, \boldsymbol{x}_2 = \begin{bmatrix} \lambda_2 \\ 1 \end{bmatrix}$. $\boldsymbol{x}_1, \boldsymbol{x}_2$ are eigenvectors of $\boldsymbol{A}$.

We let $\boldsymbol{u}_k = \begin{bmatrix} \boldsymbol{F}_{k+1} \\ \boldsymbol{F}_k \end{bmatrix}$.

Firstly, We want to find linear combination of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ to get $\boldsymbol{u}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$:

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{1}{\lambda_1 - \lambda_2} \left(\begin{bmatrix} \lambda_1 \\ 1 \end{bmatrix} - \begin{bmatrix} \lambda_2 \\ 1 \end{bmatrix} \right) \qquad \text{or} \qquad \boldsymbol{u}_0 = \frac{\boldsymbol{x}_1 - \boldsymbol{x}_2}{\lambda_1 - \lambda_2}$$

Then we multiply $\boldsymbol{u}_0$ by $\boldsymbol{A}^{100}$ to get $\boldsymbol{u}_{100}$:

$$\begin{aligned}
\boldsymbol{u}_{100} = \boldsymbol{A}^{100}\boldsymbol{u}_0 &= \frac{\boldsymbol{A}^{100}\boldsymbol{x}_1 - \boldsymbol{A}^{100}\boldsymbol{x}_2}{\lambda_1 - \lambda_2} \\
&= \frac{\boldsymbol{A}^{99}(\boldsymbol{A}\boldsymbol{x}_1) - \boldsymbol{A}^{99}(\boldsymbol{A}\boldsymbol{x}_2)}{\lambda_1 - \lambda_2} = \frac{\lambda_1\boldsymbol{A}^{99}\boldsymbol{x}_1 - \lambda_2\boldsymbol{A}^{99}\boldsymbol{x}_2}{\lambda_1 - \lambda_2} = \frac{\lambda_1^2\boldsymbol{A}^{98}\boldsymbol{x}_1 - \lambda_2^2\boldsymbol{A}^{98}\boldsymbol{x}_2}{\lambda_1 - \lambda_2} = \dots \\
&= \frac{\lambda_1^{100}\boldsymbol{x}_1 - \lambda_2^{100}\boldsymbol{x}_2}{\lambda_1 - \lambda_2}
\end{aligned}$$

Since $\lambda_1 - \lambda_2 = \sqrt{5}$, finally we obtain the same result.

7.3.3. Imaginary Eigenvalues

The eigenvalues might not be real numbers sometimes.

■ **Example 7.4**

Consider the rotation matrix given by $\boldsymbol{K} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. It rotates our vector by $90°$:

$$\boldsymbol{K}\begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}.$$

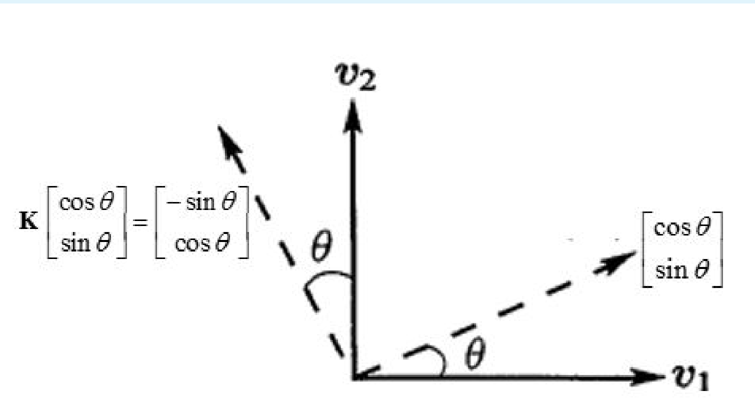


Figure 7.4: Rotate a vector by $90°$.

This rotation matrix exists eigenvector and eigenvalue, which means $\exists \boldsymbol{v} \neq \boldsymbol{0}$ and λ s.t.

$$\boldsymbol{K}\boldsymbol{v} = \lambda \boldsymbol{v}.$$

However, this equation means this rotaion matrix doesn't change the direction of $\boldsymbol{v}$. But in geometric meaning it rotates vector $\boldsymbol{v}$ by 90°. Why? 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 do similar transformation for $\boldsymbol{K}$:

$$\boldsymbol{D} = \boldsymbol{S}^{-1}\boldsymbol{K}\boldsymbol{S} = \begin{pmatrix} i & \\ & -i \end{pmatrix} \qquad \text{where } \boldsymbol{S} = \begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix}.$$

■

R For motion in vector space, eigenvalues are "speed" and eigenvectors are "directions" under basis $\boldsymbol{S} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \dots & \boldsymbol{x}_n \end{bmatrix}$.

$$\boldsymbol{v} = c_1\boldsymbol{x}_1 + \cdots + c_n\boldsymbol{x}_n \xrightarrow{\text{postmultiply } \boldsymbol{A}} \boldsymbol{A}\boldsymbol{v} = c_1\lambda_1\boldsymbol{x}_1 + \cdots + c_n\lambda_n\boldsymbol{x}_n.$$

$$\begin{pmatrix} c_1 & \dots & c_n \end{pmatrix} \xrightarrow{\text{rightmultiply } \boldsymbol{D} = \mathrm{diag}(\lambda_1, \dots, \lambda_n)} \begin{pmatrix} c_1\lambda_1 & \dots & c_n\lambda_n \end{pmatrix}.$$

7.3.4. Complex Numbers

Even when the matrix is real, its eigenvalues of this matrix may be complex numbers.

Example: A 2 by 2 rotation matrix has no real eigenvectors. It rotates a vector by 90°.

But it has complex eigenvalues i and $-i$.

Definition 7.6 [Complex Numbers] A cpmplex 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 $|\boldsymbol{x}| = \sqrt{a^2 + b^2} = \sqrt{\boldsymbol{x}\bar{\boldsymbol{x}}}$.

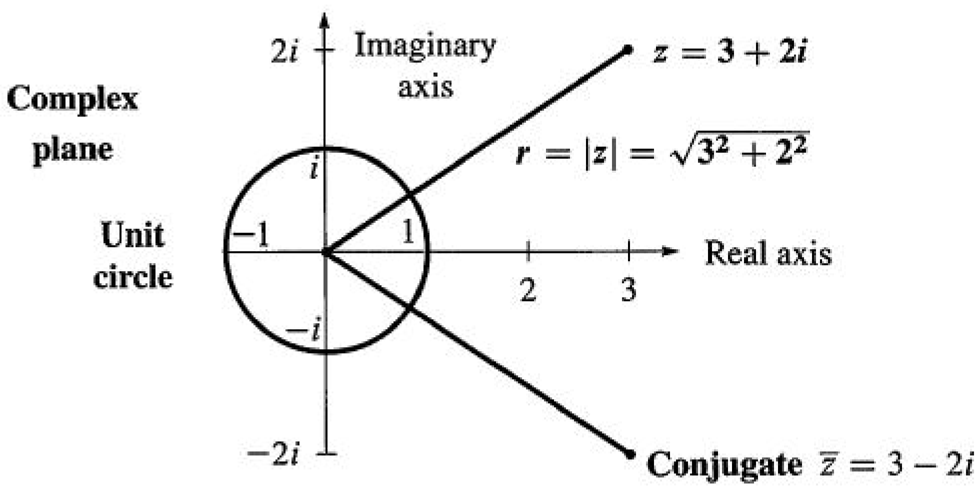


Figure 7.5: The number $z = a + bi$ corrsponds to the vector (a,b).

7.3.5. Complex Vectors

Definition 7.7 [Length (norm) for complex] For $z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \in \mathbb{C}^n$, 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 we introduce the definition of inner product for complex, let's introduce the *Hermitian* of a vector in $\mathbb{C}^n$:

Definition 7.8 [Hermitian] The hermitian of a vector in $\mathbb{C}^n$ is its *conjugate transpose*.

$$\boldsymbol{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} \qquad \boldsymbol{z}^{\mathrm{H}} = \bar{\boldsymbol{z}}^{\mathrm{T}} = \begin{bmatrix} \bar{z}_1 & \dots & \bar{z}_n \end{bmatrix}.$$

Definition 7.9 [Inner product] The inner product of real or complex vectors $\boldsymbol{z}$ and $\boldsymbol{w}$ is $\boldsymbol{w}^{\mathrm{H}}\boldsymbol{z}$, which is defined as

$$\langle \boldsymbol{z}, \boldsymbol{w} \rangle = \boldsymbol{w}^{\mathrm{H}}\boldsymbol{z} = \begin{bmatrix} \bar{w}_1 & \dots & \bar{w}_n \end{bmatrix} \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} = \bar{w}_1 z_1 + \cdots + \bar{w}_n z_n.$$

R Note that with complex 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, $\boldsymbol{z}^{\mathrm{H}}\boldsymbol{w} = \bar{z}_1 w_1 + \cdots + \bar{z}_n w_n$ is the complex conjugate of $\boldsymbol{w}^{\mathrm{H}}\boldsymbol{z}$.

Definition 7.10 [Orthogonal] The two vectors of real or complex 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 7.5

Given $\boldsymbol{z} = \begin{pmatrix} 1 \\ i \end{pmatrix}, \boldsymbol{w} = \begin{pmatrix} -i \\ 1 \end{pmatrix}$.

Although we have $\boldsymbol{z}^{\mathrm{T}}\boldsymbol{w} = 0$, the 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 7.6 The inner product of $\boldsymbol{u} = \begin{bmatrix} 1 \\ i \end{bmatrix}$ with $\boldsymbol{v} = \begin{bmatrix} i \\ 1 \end{bmatrix}$ is $\begin{bmatrix} -i & 1 \end{bmatrix} \begin{bmatrix} 1 \\ i \end{bmatrix} = 0.$

Although those vectors $(1, i)$ and $(i, 1)$ don't look perpendicular, actually they are! **A zero inner product still means vectors are orthogonal.**

Proposition 7.6 — **Conjugate symmetry.**

For two vectors $\boldsymbol{z}$ and $\boldsymbol{w} \in \mathbb{C}^n$, we have $\overline{\langle \boldsymbol{z}, \boldsymbol{w} \rangle} = \langle \boldsymbol{w}, \boldsymbol{z} \rangle$.

Proof. [Verify:]

$$\langle \boldsymbol{z}, \boldsymbol{w} \rangle = \boldsymbol{w}^{\mathrm{H}} \boldsymbol{z} = \bar{\boldsymbol{w}}^{\mathrm{T}} \boldsymbol{z} = \bar{w}_1 z_1 + \cdots + \bar{w}_n z_n$$
$$\langle \boldsymbol{w}, \boldsymbol{z} \rangle = \boldsymbol{z}^{\mathrm{H}} \boldsymbol{w} = \bar{\boldsymbol{z}}^{\mathrm{T}} \boldsymbol{w} = \bar{z}_1 w_1 + \cdots + \bar{z}_n w_n$$

And since we have $\overline{wv} = \bar{w}\bar{v}$ and $\overline{w+v} = \bar{w} + \bar{v}$, it's easy to find that

$$\overline{\bar{w}_1 z_1 + \cdots + \bar{w}_n z_n} = w_1 \bar{z}_1 + \cdots + w_n \bar{z}_n. = \bar{z}_1 w_1 + \cdots + \bar{z}_n w_n.$$

Hence $\overline{\langle \boldsymbol{z}, \boldsymbol{w} \rangle} = \langle \boldsymbol{w}, \boldsymbol{z} \rangle$.

Proposition 7.7 — **Sesquilinear.**

For two vectors $\boldsymbol{z}$ and $\boldsymbol{w} \in \mathbb{C}^n$, we have

$$\langle \alpha \boldsymbol{z}, \boldsymbol{w} \rangle = \alpha \langle \boldsymbol{z}, \boldsymbol{w} \rangle \tag{7.9}$$
$$\langle \boldsymbol{z}, \beta \boldsymbol{w} \rangle = \bar{\beta} \langle \boldsymbol{z}, \boldsymbol{w} \rangle \tag{7.10}$$

for scalars α and β.

Proof. [Verify:]

$$\begin{aligned} \langle \alpha \boldsymbol{z}, \boldsymbol{w} \rangle &= \boldsymbol{w}^{\mathrm{H}} (\alpha \boldsymbol{z}) \\ &= \alpha (\boldsymbol{w}^{\mathrm{H}} \boldsymbol{z}) \\ &= \alpha \langle \boldsymbol{z}, \boldsymbol{w} \rangle. \end{aligned}$$

For equation (7.10), 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

$$\langle \boldsymbol{z}, \beta \boldsymbol{w} \rangle = \overline{\beta \overline{\langle \boldsymbol{z}, \boldsymbol{w} \rangle}} = \bar{\beta} \langle \boldsymbol{z}, \boldsymbol{w} \rangle.$$

7.3.5.1. Hermitian of matrix

The Hermitian of a matrix $\boldsymbol{A}$ is given by

$$\boldsymbol{A}^{\mathrm{H}} := \bar{\boldsymbol{A}}^{\mathrm{T}}$$

And the rules for hermitian usually comes frim transpose. For example, the hermitian has the property

$$(\boldsymbol{A}\boldsymbol{B})^{\mathrm{H}} = \boldsymbol{B}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}.$$

$\mathbb{R}^n$	$\mathbb{C}^n$
$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$	$\langle \boldsymbol{z}, \boldsymbol{w} \rangle = \boldsymbol{w}^{\mathrm{H}}\boldsymbol{z}$
$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}$	$\boldsymbol{z}^{\mathrm{H}}\boldsymbol{w} = \overline{\boldsymbol{w}^{\mathrm{H}}\boldsymbol{z}}$
$\lVert\boldsymbol{x}\rVert^2 = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$	$\lVert\boldsymbol{z}\rVert^2 = \boldsymbol{z}^{\mathrm{H}}\boldsymbol{z}$
$\boldsymbol{x} \perp \boldsymbol{y} \iff \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = 0$	$\boldsymbol{z} \perp \boldsymbol{w} \iff \boldsymbol{w}^{\mathrm{H}}\boldsymbol{z} = 0$

R *What aspects of eigenvalues/eigenvectors are not nice?*

- Some matrix are *non-diagonalizable*. (or equivalently, eigenvectors don't form a basis.)
- Eigenvalues can be *complex*.

We may ask, what matrix has all real eigenvalues? Let's focus on *real* matrix first. For real symmetric matrix, its eigenvalues are all real!

You should remember the proposition below carefully, they are very important.

Proposition 7.8 For a *real symmetric* matrix $\boldsymbol{A}$,

- All eigenvalues are real numbers.
- Its eigenvectors corresponding to 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, suppose $\boldsymbol{x}$ is any eigenvectors of $\boldsymbol{A}$ corresponding to eigenvalue λ. Then 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 $\boldsymbol{A}^{\mathrm{H}} = \bar{\boldsymbol{A}}^{\mathrm{T}} = \boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A}$.
    Hence $\langle \boldsymbol{x}, \boldsymbol{A}^{\mathrm{H}}\boldsymbol{x} \rangle = \langle \boldsymbol{x}, \boldsymbol{A}\boldsymbol{x} \rangle$. Moreover, $\langle \boldsymbol{x}, \boldsymbol{A}\boldsymbol{x} \rangle = \langle \boldsymbol{x}, \lambda\boldsymbol{x} \rangle = \bar{\lambda}\langle \boldsymbol{x}, \boldsymbol{x} \rangle$.
    So we have $\langle \boldsymbol{x}, \boldsymbol{A}^{\mathrm{H}}\boldsymbol{x} \rangle = \bar{\lambda}\langle \boldsymbol{x}, \boldsymbol{x} \rangle$.

  Finally we have $\lambda\langle \boldsymbol{x}, \boldsymbol{x} \rangle = \langle \boldsymbol{A}\boldsymbol{x}, \boldsymbol{x} \rangle = \langle \boldsymbol{x}, \boldsymbol{A}^{\mathrm{H}}\boldsymbol{x} \rangle = \bar{\lambda}\langle \boldsymbol{x}, \boldsymbol{x} \rangle$.
  Since $\boldsymbol{x} \neq \boldsymbol{0}$, $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \neq 0$. Hence $\lambda = \bar{\lambda}$. So 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, $\langle \boldsymbol{x}_1, \boldsymbol{A}^{\mathrm{H}}\boldsymbol{x}_2 \rangle = \langle \boldsymbol{x}_1, \boldsymbol{A}\boldsymbol{x}_2 \rangle = \langle \boldsymbol{x}_1, \lambda_2\boldsymbol{x}_2 \rangle = \bar{\lambda}_2\langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle$. Since we have shown that all eigenvalues are real for $\boldsymbol{A}$, we obtain $\bar{\lambda} = \lambda$.
    Hence $\langle \boldsymbol{x}_1, \boldsymbol{A}^{\mathrm{H}}\boldsymbol{x}_2 \rangle = \lambda_2\langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle$.

  Hence $\lambda_1\langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle = \langle \boldsymbol{A}\boldsymbol{x}_1, \boldsymbol{x}_2 \rangle = \langle \boldsymbol{x}_1, \boldsymbol{A}^{\mathrm{H}}\boldsymbol{x}_2 \rangle = \lambda_2\langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle$.
  Since $\lambda_1 \neq \lambda_2$, we obtain $\langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle = 0$. That is to say, $\boldsymbol{x}_1 \perp \boldsymbol{x}_2$.

- The proof for the third part is not required.

7.3.6. Spectral Theorem

We state a theorem without proving it:

Theorem 7.5 — Spectral Theorem. Any real symmetric matrix $\boldsymbol{A}$ has the factorization

$$\boldsymbol{A} = \boldsymbol{Q}\Lambda\boldsymbol{Q}^{\mathrm{T}}, \tag{7.11}$$

where $\Lambda \in \mathbb{R}^{n\times n}$ is diagonal matrix, $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ is orthogonal matrix.

Proof. From proposition (7.8) we know that $\boldsymbol{A}$ is *diagonalizable*, which means there exists invertible matrix $\boldsymbol{Q}$ and diagonal matrix Λ such that

$$\boldsymbol{A} = \boldsymbol{Q}\Lambda\boldsymbol{Q}^{-1}$$

Then we know that all eigenvalues of $\boldsymbol{A}$ are real numbers, so Λ is a real matrix. Since all eigenvectors $\boldsymbol{x}_1,\dots,\boldsymbol{x}_n$ are orthogonal, matrix $\boldsymbol{Q} = \begin{bmatrix}\boldsymbol{x}_1 & \dots & \boldsymbol{x}_n\end{bmatrix}$ is orthogonal matrix.

R $\boldsymbol{A} = \boldsymbol{Q}\Lambda\boldsymbol{Q}^{\mathrm{T}} = \boldsymbol{Q}\Lambda\boldsymbol{Q}^{-1}$. So $\boldsymbol{A}$ could be diagonalized by an orthogonal matrix. If we let $\boldsymbol{Q} = \begin{bmatrix}q_1 & \dots & q_n\end{bmatrix}$, $\Lambda = \operatorname{diag}(\lambda_1,\dots,\lambda_n)$, then $\boldsymbol{A}$ could be written as:

$$\boldsymbol{A} = \boldsymbol{Q}\Lambda\boldsymbol{Q}^{\mathrm{T}} = \begin{bmatrix}q_1 & \dots & q_n\end{bmatrix}\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix}\begin{bmatrix}q_1^{\mathrm{T}} \\ \vdots \\ q_n^{\mathrm{T}}\end{bmatrix}$$

Or equivalently,

$$\boldsymbol{A} = \lambda_1 q_1 q_1^{\mathrm{T}} + \lambda_2 q_2 q_2^{\mathrm{T}} + \dots + \lambda_n q_n q_n^{\mathrm{T}} \tag{7.12}$$

Note that for each term $q_i q_i^{\mathrm{T}}$ is the **projection matrix** to q_i. Hence this theorm says that a real symmetric matrix is a combination of projection matrices.

■ **Example 7.7**

If we write $\boldsymbol{A}$ as combination of projection matrix, we can have a deep understanding for $\boldsymbol{Ax}$:

$$\boldsymbol{A} = \sum_{j=1}^{n} \lambda_j q_j q_j^{\mathrm{T}} \implies \boldsymbol{Ax} = \sum_{j=1}^{n} \lambda_j q_j q_j^{\mathrm{T}} \boldsymbol{x} = \sum_{j=1}^{n} \lambda_j (q_j q_j^{\mathrm{T}} \boldsymbol{x}).$$

If we set $n = 2$, it's clear to find that

$$\boldsymbol{x} = c_1 q_1 + c_2 q_2 \implies \boldsymbol{Ax} = \lambda_1 c_1 q_1 + \lambda_2 c_2 q_2$$

Showing in graph, we have

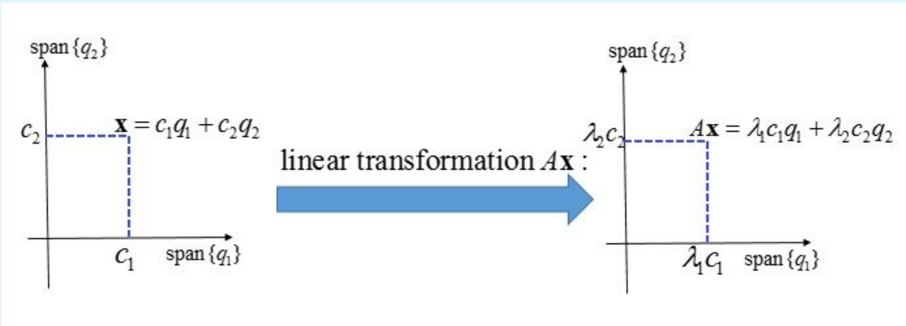


Figure 7.6: Linear transformation of $\boldsymbol{A}$.

The formula $\boldsymbol{A} = \sum_{j=1}^{n} \lambda_j q_j q_j^{\mathrm{T}}$ or $\boldsymbol{A} = \boldsymbol{Q}\Lambda\boldsymbol{Q}^{\mathrm{T}}$ are called **eigendecomposition** or **eigenvalue decomposition**.

And sometimes $\{\lambda_1, \dots, \lambda_n\}$ are called **spectum** of $\boldsymbol{A}$.

Also, we can extend our result from real symmetric matrix into complex:

7.3.7. Hermitian matrix

Definition 7.11 [Hermitian matrix] A matrix $\boldsymbol{M} \in \mathbb{C}^{n \times n}$ is said to be **Hermitian** if $\boldsymbol{M} = \boldsymbol{M}^{\mathrm{H}}$.

Example: $\boldsymbol{M} = \begin{bmatrix} 3 & 2-i \\ 2+i & 4 \end{bmatrix}$ is Hermitian matrix since $\boldsymbol{M} = \boldsymbol{M}^{\mathrm{H}}$.

If $\boldsymbol{M}$ is a real matrix, then $\boldsymbol{M} = \boldsymbol{M}^{\mathrm{H}} \iff \boldsymbol{M} = \boldsymbol{M}^{\mathrm{T}}$. So if the real matrix is Hermitian matrix, that is to say it is real symmetric matrix.

Hermitian matrix has many interesting properties:

Proposition 7.9 If $\boldsymbol{M} = \boldsymbol{M}^{\mathrm{H}}$, then $\boldsymbol{x}^{\mathrm{H}}\boldsymbol{M}\boldsymbol{x} \in \mathbb{R}$ for any complex vectors $\boldsymbol{x}$.

Proof. We set $\alpha := \boldsymbol{x}^{\mathrm{H}}\boldsymbol{M}\boldsymbol{x}$. Since α is a number (easy to check), we obtain $\alpha^{\mathrm{T}} = \alpha$. Thus $\bar{\alpha} = \alpha^{\mathrm{H}} = (\boldsymbol{x}^{\mathrm{H}}\boldsymbol{M}\boldsymbol{x})^{\mathrm{H}} = \boldsymbol{x}^{\mathrm{H}}\boldsymbol{M}\boldsymbol{x} = \alpha$.

Hence α is real.

Proposition 7.10 If $\boldsymbol{M} = \boldsymbol{M}^{\mathrm{H}}$, then $\langle \boldsymbol{x}, \boldsymbol{M}\boldsymbol{y} \rangle = \langle \boldsymbol{M}\boldsymbol{x}, \boldsymbol{y} \rangle$.

Proof. By definition,

$$\langle \boldsymbol{x}, \boldsymbol{M}\boldsymbol{y} \rangle = (\boldsymbol{M}\boldsymbol{y})^{\mathrm{H}}\boldsymbol{x} = \boldsymbol{y}^{\mathrm{H}}\boldsymbol{M}^{\mathrm{H}}\boldsymbol{x} = \boldsymbol{y}^{\mathrm{H}}\boldsymbol{M}\boldsymbol{x} = \langle \boldsymbol{M}\boldsymbol{x}, \boldsymbol{y} \rangle.$$

And we have general orthogonal matrices for complex matrices:

Definition 7.12 [Unitary] A unitary matrix is a complex matrix that has **orthonormal columns**. In other words, $\boldsymbol{U}$ is unitary if $\boldsymbol{U}^{\mathrm{H}}\boldsymbol{U} = \boldsymbol{I}$.

And the spectral theorm can also apply for Hermitian matrix:

Theorem 7.6 Any Hermitian matrix $\boldsymbol{M}$ can be factorized into

$$\boldsymbol{M} = \boldsymbol{U}\Lambda\boldsymbol{U}^{\mathrm{H}}$$

where Λ is a real diagonal matrix, $\boldsymbol{U}$ is a complex unitary matrix.

R What good points does Hermitian matrix have?

- It is diagonalizable.
- Its eigenvectors form 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
$$\lambda_n = \mathrm{trace}(\boldsymbol{A}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$$
   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 $\{\boldsymbol{0}\}$. 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:

$$x^2 + y^2 = 1 \implies \text{Circle.} \tag{8.1}$$

$$\frac{x^2}{2} + \frac{y^2}{5} = 1 \implies \text{Elipse.} \tag{8.2}$$

$$\frac{x^2}{2} - \frac{y^2}{5} = 1 \implies \text{Hyperbola.} \tag{8.3}$$

$$\left.\begin{aligned} x^2 &= \alpha y \\ y^2 &= \alpha x \end{aligned}\right\} \implies \text{Parabola.}$$

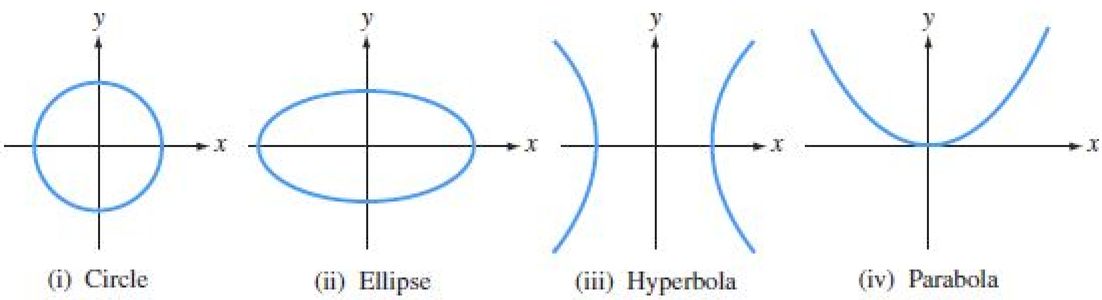


Figure 8.1: graph for quadratic form equations of two variables

The equations $(8.1) - (8.3)$ 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 $\boldsymbol{A}$ is $n \times n$ matrix, $\boldsymbol{x} \in \mathbb{R}^n$.

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

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = 1.$$

- Note that $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}$. Why?

  If we take the transpose of $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$, since it is a number, so we obtain

  $$(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x})^{\mathrm{T}} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}.$$

  Since $(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x})^{\mathrm{T}} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}$, finally we derive

  $$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}.$$

- 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 $\boldsymbol{A}$, if we want to study its quadratic form, we can always convert this matrix into symmetric matrix.

Hence without loss of generality, we assume $\boldsymbol{A} = \boldsymbol{A}^{\mathrm{T}}$ during the section of quadratic form.

■ **Example 8.1**

Given the equation $3x^2 + 2xy + 3y^2 = 1$, how we transform it into the conic section of quadratic form? And how can we determine its shape in view of matirx?

Actually, It can be written as

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 1. \qquad \textbf{conic section of quadratic form.} \tag{8.4}$$

And we define $\boldsymbol{A} = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$. If we do the eigendecomposition for $\boldsymbol{A}$, we obtain

$$\boldsymbol{A} = \boldsymbol{Q}\Lambda\boldsymbol{Q}^{\mathrm{T}}$$

where $\Lambda = \begin{pmatrix} \lambda_1 & \\ & \lambda_2 \end{pmatrix}$, $\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 \end{bmatrix}$. $\boldsymbol{x}_1$,$\boldsymbol{x}_2$ is the eigenvectors of $\boldsymbol{A}$ corresponding to eigenvalues λ_1,λ_2 respectively.

Thus we convert equation (8.4) into

$$\begin{pmatrix} x & y \end{pmatrix} \boldsymbol{Q}\Lambda\boldsymbol{Q}^{\mathrm{T}} \begin{pmatrix} x \\ y \end{pmatrix} = 1 \implies \tilde{\boldsymbol{x}}^{\mathrm{T}}\Lambda\tilde{\boldsymbol{x}} = 1.$$

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 to obtain:

$$\lambda_1\tilde{x}_1^2 + \lambda_2\tilde{x}_2^2 = 1.$$

After computation, we find $\lambda_1 = 4, \lambda_2 = 2$. Hence this equation is a **elipse.** ■

R

8.1.1.1. Matrix Calculus

Now we recall how to compute derivative for matrix:

- $\frac{\partial(f^{\mathrm{T}}g)}{\partial x} = \frac{\partial f(x)}{\partial x}g(x) + \frac{\partial g(x)}{\partial x}f(x)$

Example:

- $\frac{\partial(a^{\mathrm{T}}\boldsymbol{x})}{\partial \boldsymbol{x}} = a$
- $\frac{\partial(a^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}} = \frac{\partial((\boldsymbol{A}^{\mathrm{T}}a)^{\mathrm{T}}\boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{A}^{\mathrm{T}}a$
- $\frac{\partial(\boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{A}^{\mathrm{T}}$
- $\frac{\partial(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}$

■ **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:

$$\nabla f(\boldsymbol{x}^*) = \frac{\partial f(\boldsymbol{x}^*)}{\partial \boldsymbol{x}} = \boldsymbol{0} \implies \boldsymbol{A}\boldsymbol{x}^* + \boldsymbol{b} = \boldsymbol{0}.$$

There may follow these cases:

- If equation $\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}=\mathbf{0}$ has no solution, then $f(\boldsymbol{x})$ is unbounded.

  This statement is remained to be proved.

- If equation $\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}=\mathbf{0}$ 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

$$\boldsymbol{A}=\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}, \qquad \boldsymbol{b}=\mathbf{0}, \qquad \boldsymbol{x}=\begin{pmatrix}x_1\\ x_2\end{pmatrix},$$

  then $f(\boldsymbol{x})=\frac{1}{2}(x_1^2-x_2^2)$.

  One solution to $\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}=\mathbf{0}$ is $\boldsymbol{x}^*=\begin{pmatrix}0\\ 0\end{pmatrix}$. Obviously, $\boldsymbol{x}^*$ is not a optimal solution.

  Intutively, if we let $x_1=0, x_2\to\infty$, then $f(\boldsymbol{x})\to-\infty$!

8.1.1.2. Second optimality condition

If $\boldsymbol{x}^*$ is a optimal solution to $f(\boldsymbol{x})$, what else condition should $\boldsymbol{x}^*$ satisfy?

Let's take $f(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}$ as an example, we want to find $\boldsymbol{x}^*$ s.t. $\min f(\boldsymbol{x})=f(\boldsymbol{x}^*)$.

Firstly, we convert $f(\boldsymbol{x})$ into its *taylor expansion*:

$$f(\boldsymbol{x})=f(\boldsymbol{x}^*)+\langle\triangledown f(\boldsymbol{x}^*),\boldsymbol{x}-\boldsymbol{x}^*\rangle+\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}^*)^{\mathrm{T}}\triangledown^2 f(\boldsymbol{x}^*)(\boldsymbol{x}-\boldsymbol{x}^*).$$

Note that $\triangledown^2 f(\boldsymbol{x}^*)$ is the Hessian matrix of $f(\boldsymbol{x}^*)$, which is defined as

$$\triangledown^2 f(\boldsymbol{x}^*):=\left[\frac{\partial^2 f(\boldsymbol{x}^*)}{\partial x_i \partial x_j}\right]=\triangledown(\triangledown f(\boldsymbol{x}^*)).$$

Firstly we compute $\triangledown f(\boldsymbol{x})$ and $\triangledown^2 f(\boldsymbol{x})$:

$$\triangledown f(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}) + \boldsymbol{b}.$$
$$\triangledown^2 f(\boldsymbol{x}) = \triangledown \left[\frac{1}{2}(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}) + \boldsymbol{b}\right] = \frac{1}{2} \triangledown (\boldsymbol{A}\boldsymbol{x}) + \frac{1}{2} \triangledown (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{A} + \boldsymbol{A}^{\mathrm{T}}).$$

If we assume $\boldsymbol{A}$ is **symmetric**, then we have $\triangledown f(\boldsymbol{x}) = (\boldsymbol{A} + \boldsymbol{A}^{\mathrm{T}})\boldsymbol{x} + \boldsymbol{b}$ and $\triangledown^2 f(\boldsymbol{x}) = \boldsymbol{A}$.
Since the optimal solution $\boldsymbol{x}^*$ must satisfy $\triangledown f(\boldsymbol{x}^*) = \boldsymbol{0}$, we deive

$$\langle \triangledown f(\boldsymbol{x}^*), \boldsymbol{x} - \boldsymbol{x}^* \rangle = 0. \implies f(\boldsymbol{x}) = f(\boldsymbol{x}^*) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^*)^{\mathrm{T}} \triangledown^2 f(\boldsymbol{x}^*)(\boldsymbol{x} - \boldsymbol{x}^*).$$

Hence $f(\boldsymbol{x}) - f(\boldsymbol{x}^*) = \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{x}^*)$.
Since $\boldsymbol{x}^*$ is optimal that minimize $f(\boldsymbol{x})$, $LHS = f(\boldsymbol{x}) - f(\boldsymbol{x}^*) \ge 0$.

$$\implies \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{x}^*) \ge 0$$

Or equivalently,

$$(\boldsymbol{x} - \boldsymbol{x}^*)^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{x}^*) \ge 0 \text{ for } \forall \boldsymbol{x}. \iff \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} \ge 0 \text{ for } \forall \boldsymbol{x}.$$

Our conclusion is that if there exists a optimal solution for $f(\boldsymbol{x})$, then the matrix $\boldsymbol{A}$ should satisfy $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} \ge 0$ for $\forall \boldsymbol{x}$. We have a specific name for such $\boldsymbol{A}$.

8.1.2. Positive Definite Matrices

Definition 8.2 [Positive-definite]

- Matrix $\boldsymbol{A}$ is *positive-semidefinite* (PSD) if $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} \ge 0$ for $\forall \boldsymbol{x}$. And we denote it as $\boldsymbol{A} \succeq 0$.
- Matrix $\boldsymbol{A}$ is *positive-definite* (PD) if $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} > 0$ for $\forall \boldsymbol{x} \neq \boldsymbol{0}$. And we denote it as $\boldsymbol{A} \succ 0$.

- Matrix $\boldsymbol{A}$ is *indefinite* if there exist some $\boldsymbol{x}$ and $\boldsymbol{y}$ s.t.

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} < 0 < \boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y}.$$

R If a matrix is PSD or PD, it is usually assumed to be symmetric by default. Even in other textbooks, the definition for PSD and PD contains the *symmetric* condition.

Theorem 8.1 Let $\boldsymbol{A}$ be a $n \times n$ real symmetric matrix, the following 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 $\forall \boldsymbol{x} = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \end{pmatrix}^{\mathrm{T}} \neq \mathbf{0}$.

- For matrix $\boldsymbol{B}$, all *upper left square submatrices* is 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.*

- $(1) \implies (2)$: Suppose λ is any eigenvalue of $\boldsymbol{A}$. Then $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$ for some $\boldsymbol{x} \neq \boldsymbol{0}$. 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 $i = 1,2,\dots,n$.

  For $\forall \boldsymbol{x} \neq \boldsymbol{0}$, our goal is to show $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} > 0$:

  Since $\boldsymbol{A}$ is real symmetric matrix, we do eigendecomposition of $\boldsymbol{A}$:

$$\boldsymbol{A} = \boldsymbol{Q}\Lambda\boldsymbol{Q}^{\mathrm{T}} \qquad \boldsymbol{Q} \text{ is orthonormal matrix.}$$

  Hence

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\Lambda\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{x} = (\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{x})^{\mathrm{T}}\Lambda(\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{x}).$$

  If we set $\tilde{\boldsymbol{x}} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{x} = \begin{bmatrix} \tilde{x_1} & \dots & \tilde{x_n} \end{bmatrix}$, then $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}$ can be written 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 that $\sum_{i=1}^{n}\lambda_i\tilde{x_i}^2 \neq 0$. Actually we only need to show $\|\boldsymbol{x}\| \neq 0$:

  Since previously we have shown $\|\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{x}\| = \|\boldsymbol{x}\|$, we obtain:

$$\|\tilde{\boldsymbol{x}}\| = \|\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{x}\| = \|\boldsymbol{x}\| \neq 0.$$

- $(1) \implies (3)$: We only need to show $\det(\boldsymbol{A}_k) > 0$ for $k = 1,\dots,n$.

  Given any $\tilde{\boldsymbol{x}} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix} \in \mathbb{R}^k$, 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 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 $\boldsymbol{A} > 0$.

- $(3) \implies (4)$:
  - We want to show all pivots of $\boldsymbol{A}$ are positive first:

    We do row transform to convert $\boldsymbol{A}$ into upper triangular matrix $\tilde{\boldsymbol{A}}$:

    $$\begin{bmatrix}\times & \times & \times \\ \times & \times & \times \\ \times & \times & \times\end{bmatrix} \Longrightarrow \begin{bmatrix}\times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times\end{bmatrix}$$

    And during the row transformation, the determinant doesn't change. Moreover, the correponding *upper left submatrices* determinants don't change. In other words, we obtain

    $$\det(\tilde{\boldsymbol{A}}_i) = \det(\boldsymbol{A}_i) \text{ for } i = 1, \dots, n.$$

    And moreover, $\tilde{\boldsymbol{A}}_i$ always contains $\tilde{\boldsymbol{A}}_{i-1}$ on its upper left side:

    $$\tilde{\boldsymbol{A}}_i = \begin{bmatrix}\tilde{\boldsymbol{A}}_{i-1} & \boldsymbol{B} \\ \boldsymbol{0} & \tilde{a}_{ii}\end{bmatrix}$$

    And we notice $\tilde{\boldsymbol{A}}_i$'s are also upper triangular matrix. The determinant of a upper triangular matrix is the product of its diagonal entries. Hence we obtain

$$\det(\tilde{\boldsymbol{A}}_i) = \tilde{a_{ii}} \det(\tilde{\boldsymbol{A}}_{i-1}) \text{ for } i = 2, \dots, n.$$

Thus $\tilde{a_{ii}} = \frac{\det(\tilde{\boldsymbol{A}}_i)}{\det(\tilde{\boldsymbol{A}}_{i-1})} = \frac{\det(\boldsymbol{A}_i)}{\det(\boldsymbol{A}_{i-1})}$ for $i = 2, \dots, n$.

Due to (3) of this theorem, $\tilde{a_{ii}} > 0$ for $i = 2, \dots, n$. And $\tilde{a_{11}} = \det(\tilde{\boldsymbol{A}}_1) = \det(\boldsymbol{A}_1) > 0$.

In conclusion, all pivots $\tilde{a_{ii}} > 0$ for $i = 1, \dots, n$.

– Then we do the LDU composition for $\boldsymbol{A}$. Since $\boldsymbol{A}$ is symmetric, 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 $\boldsymbol{A}$. And $\boldsymbol{L}$ is lower triangular matrix with 1's on the diagonal entries.

Since all pivots of $\boldsymbol{A}$ are positive, we define $\sqrt{\boldsymbol{D}} := \operatorname{diag}(\sqrt{d_1}, \dots, \sqrt{d_n})$.

Hence $\boldsymbol{A}$ could be written as:

$$\boldsymbol{A} = \boldsymbol{L} \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix} \boldsymbol{L}^{\mathrm{T}} = \boldsymbol{L}\sqrt{\boldsymbol{D}}\sqrt{\boldsymbol{D}}\boldsymbol{L}^{\mathrm{T}} = (\sqrt{\boldsymbol{D}}\boldsymbol{L}^{\mathrm{T}})^{\mathrm{T}}(\sqrt{\boldsymbol{D}}\boldsymbol{L}^{\mathrm{T}}).$$

Fefine $\boldsymbol{R} = \sqrt{\boldsymbol{D}}\boldsymbol{L}^{\mathrm{T}}$. 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 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 we only need 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 $\|\boldsymbol{R}\boldsymbol{x}\| \neq 0$.

We may ask 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** 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 $\boldsymbol{A}$ 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 $k \in \{1, 2, \dots, n\}$.

Note that the leading principal submatrix is just the upper left submatrix we have mentioned before.

Corollary 8.1 If $\boldsymbol{A} \succ 0$, then all principal submatrices of $\boldsymbol{A} \succ 0$.

Proof. Our goal is to show $\boldsymbol{A}_{\alpha,\alpha} \succ 0$, where $\alpha \in \{1, 2, \dots, n\}$.

For any $\boldsymbol{x}_\alpha \in \mathbb{R}^{|\alpha|}$, we only need to show $\boldsymbol{x}_\alpha^{\mathrm{T}} \boldsymbol{A}_{\alpha,\alpha} \boldsymbol{x}_\alpha > 0$. Note that $|\alpha|$ denotes the number of elements in set α.

We construct $\boldsymbol{x} \in \mathbb{R}^n$ s.t. the ith entry of $\boldsymbol{x}$ is

$$\boldsymbol{x}_i = \begin{cases} (\boldsymbol{x}_\alpha)_i & i \in \alpha \\ 0 & i \notin \alpha \end{cases}$$

It's obvious that

$$\begin{aligned} \boldsymbol{x}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{x} &= \sum_{i,j=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_j \boldsymbol{A}_{ij} \\ &= \sum_{i,j \in \alpha} (\boldsymbol{x}_\alpha)_i (\boldsymbol{x}_\alpha)_j (\boldsymbol{A}_{\alpha,\alpha})_{ij} \\ &= \boldsymbol{x}_\alpha^{\mathrm{T}} \boldsymbol{A}_{\alpha,\alpha} \boldsymbol{x}_\alpha > 0. \end{aligned}$$

How to use this corollary to test the positive definiteness?

For example, given $\boldsymbol{A} = \begin{bmatrix} 2 & -1 & 1 \\ -1 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix}$, immediately we find one principal matrix is $\boldsymbol{A}_{2,2} = 0$. Hence it is not PD.

Also, there are many equivalent statements related to PSD.

Theorem 8.2 Let $\boldsymbol{A}$ be a $n \times n$ real symmetric matrix, the following are equivalent:

1. $\boldsymbol{A}$ is PSD.

2. All eigenvalues of $\boldsymbol{A}$ 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 $\boldsymbol{A}_{ij} \geq 0$? 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 real $n \times n$ matrix $\boldsymbol{A}$ and $\boldsymbol{B}$, the **Frobenius inner product** is given by

$$\langle \boldsymbol{A}, \boldsymbol{B} \rangle = \sum_{i,j=1}^{n} \boldsymbol{A}_{ij} \boldsymbol{B}_{ij}$$

Proposition 8.1 If two real $n \times n$ symmetric matrix $\boldsymbol{A} \succeq 0, \boldsymbol{B} \succeq 0$, 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 obtain $\langle \boldsymbol{A}, \boldsymbol{B} \rangle = \sum_{k=1}^{n} \boldsymbol{r}_k^{\mathrm{T}} \boldsymbol{B} \boldsymbol{r}_k \geq 0$.

8.2. Thursday

Three ways for matrix decomposition are significant in linear alegbra:

$$\begin{cases} \text{LU (from \textbf{elimination})} \\ \text{QR (from \textbf{orthogonalization})} \\ \text{SVD (from \textbf{eigenvectors})} \end{cases}$$

We have learnt the first two decomposition. And the third way is increasingly significant in the information age.

In the last lecture we talk about *eigendecomposition* for **real symmetric** matrices and *diagonalization*. However, can we get some **universal** decomposition? Is there any decomposition that can be applied to all matrices?

The anwer is yes. The key idea 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

Any matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ could be factorized into

$$\boldsymbol{A} = \boldsymbol{U}\Sigma\boldsymbol{V}^{\mathrm{T}}$$

where $\boldsymbol{U}$ is a $m \times m$ **orthogonal** matrix, Σ is a $m \times n$ "*diagonal*" (we will define it later) matrix, $\boldsymbol{V}$ is a $n \times n$ **orthogonal** matrix.

If $\boldsymbol{V}$=$\boldsymbol{U}$ (then consequently $m = n$), then this is exactly *eigendecomposition*.

Specifically speaking,

$\boldsymbol{U}$ is $m \times m$ matrix s.t. *columns are eigenvectors of* $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$.

$\boldsymbol{V}$ is $n \times n$ matrix s.t. *columns are eigenvectors of* $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$.

Σ is $m \times n$ matrix which has the form:

$$\Sigma = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix} \text{ if } m \geq n \quad \text{or } \Sigma = \begin{pmatrix} \sigma_1 & & & 0 & \dots & 0 \\ & \ddots & & \vdots & \ddots & \vdots \\ & & \sigma_m & 0 & \dots & 0 \end{pmatrix} \text{ if } m < n.$$

And $\sigma_i = \sqrt{\lambda_i}$ for $i = 1, 2, \dots, \min\{m, n\}$, where λ_i*'s are eigenvalues of* $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ *or* $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$. (if $m \geq n$), then λ_i's are eigenvalues of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$; otherwise λ_i's are eigenvalues of $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$.)

Theorem 8.3 SVD always exists for any **real** matrix.

Proof. For any $m \times n$ matrix $\boldsymbol{A}$, WLOG, we set $m \geq n$.

- Firstly, we consider the case that all $\lambda_j \neq 0$ for $j = 1, \dots, n$. (λ_j's are eigenvalues of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$.)

  Since $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is *real symmetric*, we do the eigendecomposition:

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^{\mathrm{T}}$$

  where $\boldsymbol{V}$ is *orthonormal* matrix and $\boldsymbol{D}$ is *diagonal* matrix.

  Also, the eigenvectors of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ are orthogonal (note that in proposition (16.3) we claim that the eigenvectors of diagonalizable matrix are orthogonal.):

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}_j = \lambda_j \boldsymbol{v}_j \text{ for } j = 1, \dots, n$$

  where $\boldsymbol{v}_j$'s are eigenvectors of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ s.t. they form orthonormal basis of $\mathbb{R}^n$.

Note that given any matrix $\boldsymbol{A}$, $\boldsymbol{V} = \begin{bmatrix} \boldsymbol{v}_1 & \dots & \boldsymbol{v}_n \end{bmatrix}$ is immediately defined. (This is because $\boldsymbol{v}_j$'s are eigenvectors of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$.)

If we want to show $\boldsymbol{A} = \boldsymbol{U}\Sigma\boldsymbol{V}^{\mathrm{T}}$, since $\boldsymbol{A}$ and $\boldsymbol{V}$ is defined, we only need to show there exists *special* $\boldsymbol{U}$ and Σ such that

$$\boldsymbol{U}\Sigma = \boldsymbol{A}(\boldsymbol{V}^{\mathrm{T}})^{-1} = \boldsymbol{A}\boldsymbol{V}.$$

– First step, we construct such $\boldsymbol{U}$ and Σ:

  Since λ_j's are eigenvalues of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ associated with eigenvectors $\boldsymbol{v}_j$, we obtain:

$$\|\boldsymbol{A}\boldsymbol{v}_j\|^2 = \boldsymbol{v}_j^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}_j) = \boldsymbol{v}_j^{\mathrm{T}}(\lambda_j\boldsymbol{v}_j) = \lambda_j(\boldsymbol{v}_j^{\mathrm{T}}\boldsymbol{v}_j) = \lambda_j\|\boldsymbol{v}_j\|^2.$$

  Hence $\lambda_j = \frac{\|\boldsymbol{A}\boldsymbol{v}_j\|^2}{\|\boldsymbol{v}_j\|^2} > 0$. (As we assume $\lambda_j \neq 0$, this is strictly inequality.)

  Hence we define $\boldsymbol{u}_j := \boldsymbol{A}\boldsymbol{v}_j\frac{1}{\sqrt{\lambda_j}} \in \mathbb{R}^{m\times 1}$. for $j = 1, \dots, n$.

  And then we construct $\boldsymbol{U}$ and Σ:

$$\begin{aligned} \boldsymbol{U} &:= \begin{bmatrix} \boldsymbol{u}_1 & \dots & \boldsymbol{u}_n \end{bmatrix} \in \mathbb{R}^{m\times n}. \\ \Sigma &:= \operatorname{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n}) \in \mathbb{R}^{n\times n}. \end{aligned}$$

  It's easy to verify that $\boldsymbol{U}\Sigma = \boldsymbol{A}\boldsymbol{V}$.

– Next step, we show that $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_n\}$ is orthonormal set:

  For any $\boldsymbol{u}_i, \boldsymbol{u}_j$, we have

$$\begin{aligned} \langle \boldsymbol{u}_i, \boldsymbol{u}_j \rangle &= \frac{1}{\sqrt{\lambda_i}\sqrt{\lambda_j}}\langle \boldsymbol{A}\boldsymbol{v}_i, \boldsymbol{A}\boldsymbol{v}_j \rangle \\ &= \frac{1}{\sqrt{\lambda_i}\sqrt{\lambda_j}}\langle \boldsymbol{v}_i, \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}_j \rangle \qquad \text{Due to the useful formula } \langle \boldsymbol{A}\boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{x}, \boldsymbol{A}^{\mathrm{H}}\boldsymbol{y} \rangle. \\ &= \frac{1}{\sqrt{\lambda_i}\sqrt{\lambda_j}}\langle \boldsymbol{v}_i, \lambda_j\boldsymbol{v}_j \rangle = \sqrt{\frac{\lambda_j}{\lambda_i}}\langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle \end{aligned}$$

Since $\{\boldsymbol{v}_1,\dots,\boldsymbol{v}_n\}$ are orthonormal, we obtain

$$\langle \boldsymbol{v}_i,\boldsymbol{v}_j\rangle = \begin{cases} 1, & i=j \\ 0, & i\neq j \end{cases} \implies \langle \boldsymbol{u}_i,\boldsymbol{u}_j\rangle = \begin{cases} 1, & i=j \\ 0, & i\neq j \end{cases}$$

Hence $\{\boldsymbol{u}_1,\dots,\boldsymbol{u}_n\}$ are orthonormal.

– Then we show that $\{\boldsymbol{u}_1,\dots,\boldsymbol{u}_n\}$ are eigenvectors of $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$:

  For $j=1,\dots,n$, we obtain:

$$\begin{aligned}
\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}_j &= \boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}_j\frac{1}{\sqrt{\lambda_j}} \qquad \text{by definition of } \boldsymbol{u}_j. \\
&= \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}_j)\frac{1}{\sqrt{\lambda_j}} = \boldsymbol{A}\lambda_j\boldsymbol{v}_j\frac{1}{\sqrt{\lambda_j}} \\
&= \sqrt{\lambda_j}\boldsymbol{A}\boldsymbol{v}_j = \lambda_j \times (\frac{1}{\sqrt{\lambda_j}}\boldsymbol{A}\boldsymbol{v}_j) \\
&= \lambda_j\boldsymbol{u}_j.
\end{aligned}$$

– We notice that in SVD $\boldsymbol{U}$ is m by m matrix, Σ is m by n matrix. Hence we need to *reconstruct* our $\boldsymbol{U}$ and Σ in step 1:

  Since $\{\boldsymbol{u}_1,\dots,\boldsymbol{u}_n\}$ are eigenvectors of $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$, and $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ has m orthogonal eigenvectors, so we pick $\boldsymbol{u}_{n+1},\dots,\boldsymbol{u}_m$ s.t. $\{\boldsymbol{u}_1,\dots,\boldsymbol{u}_n,\boldsymbol{u}_{n+1},\dots,\boldsymbol{u}_m\}$ are m orthonormal eigenvectors of $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$.

  Then we let

$$\boldsymbol{U} := \begin{bmatrix} \boldsymbol{u}_1 & \dots & \boldsymbol{u}_m \end{bmatrix} \in \mathbb{R}^{m\times m}.$$

$$\Sigma := \begin{pmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \\ 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix} \in \mathbb{R}^{m\times n}.$$

It's easy to verify that

$$U\Sigma = AV$$

Hence finally we obtain

$$U\Sigma V^{\mathrm{T}} = AVV^{\mathrm{T}} = A.$$

- For there exists some $\lambda_j = 0$ case, we discuss it in next section.

8.2.2. Remark on SVD decomposition

8.2.2.1. Remark 1

The eigenvalues for $A^{\mathrm{T}}A$ and AA^{T} are not always nonzero.

Proposition 8.2 For $m \times n$ matrix A, suppose $\text{rank}(A) = r$, and the eigenvalues for $A^{\mathrm{T}}A$ are

$$\text{eig}(A^{\mathrm{T}}A) = \{\lambda_1, \dots, \lambda_r, \overbrace{0, \dots\dots, 0}^{\text{Totally } n-r \text{ terms}}\},$$

then eigenvalues for AA^{T} are

$$\text{eig}(AA^{\mathrm{T}}) = \{\lambda_1, \dots, \lambda_r, \underbrace{0, \dots\dots, 0}_{\text{Totally } m-r \text{ terms}}\}.$$

Proof.

- Firstly let's prove a lemma:

$$\det(I_m - AB) = \det(I_n - BA).,$$

where $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times m}$.

We notice the two equality:

$$\begin{pmatrix} \boldsymbol{I}_n & \boldsymbol{B} \\ \boldsymbol{A} & \boldsymbol{I}_m \end{pmatrix} \begin{pmatrix} \boldsymbol{I}_n & -\boldsymbol{B} \\ \boldsymbol{0}_{m\times n} & \boldsymbol{I}_m \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}_n & \boldsymbol{0}_{n\times m} \\ \boldsymbol{A} & \boldsymbol{I}_m - \boldsymbol{A}\boldsymbol{B} \end{pmatrix} \tag{8.5}$$

$$\begin{pmatrix} \boldsymbol{I}_n & -\boldsymbol{B} \\ \boldsymbol{0}_{m\times n} & \boldsymbol{I}_m \end{pmatrix} \begin{pmatrix} \boldsymbol{I}_n & \boldsymbol{B} \\ \boldsymbol{A} & \boldsymbol{I}_m \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}_n - \boldsymbol{B}\boldsymbol{A} & \boldsymbol{0}_{n\times m} \\ \boldsymbol{A} & \boldsymbol{I}_m \end{pmatrix} \tag{8.6}$$

We take determinant both sides for the two equations above:

$$\begin{vmatrix} \boldsymbol{I}_n & \boldsymbol{B} \\ \boldsymbol{A} & \boldsymbol{I}_m \end{vmatrix} \begin{vmatrix} \boldsymbol{I}_n & -\boldsymbol{B} \\ \boldsymbol{0}_{m\times n} & \boldsymbol{I}_m \end{vmatrix} = \begin{vmatrix} \boldsymbol{I}_n & \boldsymbol{0}_{n\times m} \\ \boldsymbol{A} & \boldsymbol{I}_m - \boldsymbol{A}\boldsymbol{B} \end{vmatrix} \tag{8.7}$$

$$\begin{vmatrix} \boldsymbol{I}_n & -\boldsymbol{B} \\ \boldsymbol{0}_{m\times n} & \boldsymbol{I}_m \end{vmatrix} \begin{vmatrix} \boldsymbol{I}_n & \boldsymbol{B} \\ \boldsymbol{A} & \boldsymbol{I}_m \end{vmatrix} = \begin{vmatrix} \boldsymbol{I}_n - \boldsymbol{B}\boldsymbol{A} & \boldsymbol{0}_{n\times m} \\ \boldsymbol{A} & \boldsymbol{I}_m \end{vmatrix} \tag{8.8}$$

Since we have $\begin{vmatrix} \boldsymbol{I}_n & \boldsymbol{B} \\ \boldsymbol{A} & \boldsymbol{I}_m \end{vmatrix} \begin{vmatrix} \boldsymbol{I}_n & -\boldsymbol{B} \\ \boldsymbol{0}_{m\times n} & \boldsymbol{I}_m \end{vmatrix} = \begin{vmatrix} \boldsymbol{I}_n & -\boldsymbol{B} \\ \boldsymbol{0}_{m\times n} & \boldsymbol{I}_m \end{vmatrix} \begin{vmatrix} \boldsymbol{I}_n & \boldsymbol{B} \\ \boldsymbol{A} & \boldsymbol{I}_m \end{vmatrix}$, we derive

$$\begin{vmatrix} \boldsymbol{I}_n & \boldsymbol{0}_{n\times m} \\ \boldsymbol{A} & \boldsymbol{I}_m - \boldsymbol{A}\boldsymbol{B} \end{vmatrix} = \begin{vmatrix} \boldsymbol{I}_n - \boldsymbol{B}\boldsymbol{A} & \boldsymbol{0}_{n\times m} \\ \boldsymbol{A} & \boldsymbol{I}_m \end{vmatrix} \implies \det(\boldsymbol{I}_n)\det(\boldsymbol{I}_m - \boldsymbol{A}\boldsymbol{B}) = \det(\boldsymbol{I}_m)\det(\boldsymbol{I}_n - \boldsymbol{B}\boldsymbol{A})$$

Equivalently, $\det(\boldsymbol{I}_m - \boldsymbol{A}\boldsymbol{B}) = \det(\boldsymbol{I}_n - \boldsymbol{B}\boldsymbol{A})$.

- Secondly we prove that the nonzero eigenvalues of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ and $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ are exactly the same (counted with multiplicity):
  We only need to show $\frac{\det(\lambda \boldsymbol{I} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})}{\lambda^{n-r}} = \frac{\det(\lambda \boldsymbol{I} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})}{\lambda^{m-r}}$.
  And we find that

$$\begin{aligned} \det(\lambda \boldsymbol{I} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) &= \lambda^n \det(\boldsymbol{I} - \lambda^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) \\ &= \lambda^n \det(\boldsymbol{I} - \lambda^{-1}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}) && \text{Due to Sylvester's determinant identity} \\ & && \det(\boldsymbol{I}_m - \boldsymbol{A}\boldsymbol{B}) = \det(\boldsymbol{I}_n - \boldsymbol{B}\boldsymbol{A}). \\ & && \text{for } \boldsymbol{A} \in \mathbb{R}^{m\times n} \text{ and } \boldsymbol{B} \in \mathbb{R}^{n\times m}. \end{aligned}$$

Hence we obtain

$$\begin{aligned}
\frac{\det(\lambda \boldsymbol{I}-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})}{\lambda^{n-r}} &= \frac{\lambda^{n}\det(\boldsymbol{I}-\lambda^{-1}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})}{\lambda^{n-r}} \\
&= \frac{\lambda^{m}\det(\boldsymbol{I}-\lambda^{-1}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})}{\lambda^{m-r}} \\
&= \frac{\det(\lambda \boldsymbol{I}-\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})}{\lambda^{m-r}}
\end{aligned}$$

- Then we show the eigenvalues for $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ have exactly $(n-r)$ zeros; the eigenvalues for $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$ have exactly $(m-r)$ zeros.

  Assume there are n ind. eigenvectors $\{\boldsymbol{v}_1,\dots,\boldsymbol{v}_n\}$ for $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ corresponding to their eigenvalues $\{\lambda_1,\lambda_2,\dots,\lambda_n\}$. Hence we have

  $$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{v}_i = \lambda_i\boldsymbol{v}_i \text{ for } i=1,\dots,n.$$

  Since $\operatorname{rank}(\boldsymbol{A}) = r = \operatorname{rank}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})$, the dimension of the eigenspace for $\lambda=0$ is $n-r$.

  Hence among $\{\boldsymbol{v}_1,\dots,\boldsymbol{v}_n\}$ there are $n-r$ ind. eigenvectors belong to the eigenspace for $\lambda=0$.

  Thus there are exactly $(n-r)$ zeros for eigenvalues of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$.

- How to prove there are exactly $(m-r)$ zeros for eigenvalues of $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$? We just need to obtain $\operatorname{rank}(\boldsymbol{A}^{\mathrm{T}}) = r = \operatorname{rank}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})$ and proceed similarly.

R Maybe you think the second part proof ($A^{\mathrm{T}}A$ and AA^{T} have the same set of nonzero eigenvalues) is too thicky. Here is a proofoutline of another way to show this conclusion: *Proof.* [Proofoutline.]

Assume $\boldsymbol{v}_1,\dots,\boldsymbol{v}_n$'s are eigenvectors of $A^{\mathrm{T}}A$, then define $\boldsymbol{u}_j$ as in the proof of SVD theorem for nonzero λ_j's, where $j=1,\dots,r$.

Then by the equation $AA^{\mathrm{T}}\boldsymbol{u}_j=\lambda_j\boldsymbol{u}_j$, λ_j is also the eigenvalue of AA^{T} with eigenvector $\boldsymbol{u}_j$.

- If λ_j is eigenvalue of $A^{\mathrm{T}}A$ with mulplicity d, then by the eigendecomposition there are exactly d ind. eigenvectors $v_{j_1},\dots,v_{j_d}$ associated with λ_j; thus we get d ind. eigenvectors $\boldsymbol{u}_{j_1},\dots,\boldsymbol{u}_{j_d}$ associated with λ_j for AA^{T}. Thus the multiplicity of λ_j for $A^{\mathrm{T}}A$ and AA^{T} are the same.
- Thus $\{\lambda_1,\dots,\lambda_r\}$ (counted with multiplicity) is a subset of $\mathrm{eig}(AA^{\mathrm{T}})$.
- Similarly, the set of nonzero eigenvalues of AA^{T} is also a subset of $\mathrm{eig}(A^{\mathrm{T}}A)$. Thus these two matrices have the same set of nonzero eigenvalues.

Since all the rest eigenvalues must be zero, we get the desired result.

For 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:

$$\begin{cases} A\boldsymbol{v}_j=\sigma_j\boldsymbol{u}_j \\ A^{\mathrm{T}}\boldsymbol{u}_j=\sigma_j\boldsymbol{v}_j \end{cases}$$

And the columns of U ($\boldsymbol{u}_j$) are called **left singular vector** of A; the columns of V ($\boldsymbol{v}_j$) are called **right singular vector** of A; σ_j is called the **singular value**.

8.2.2.2. Remark 2: Four fundamental subspaces

The general SVD decomposition for $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is given by

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{u}_1 & \dots & \boldsymbol{u}_r & \dots & \boldsymbol{u}_m \end{bmatrix} \begin{pmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \\ 0 & \dots & \dots & \dots & \dots & 0 \\ 0 & \dots & \dots & \dots & \dots & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{v}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{v}_r^{\mathrm{T}} \\ \vdots \\ \boldsymbol{v}_n^{\mathrm{T}} \end{pmatrix} = \boldsymbol{U}\Sigma\boldsymbol{V}^{\mathrm{T}}$$

For such $\boldsymbol{A}$, the matrix $\boldsymbol{U}$ and $\boldsymbol{V}$ contain orthonormal basis for all four fundamental subspaces:

First r columns of $\boldsymbol{V}$: row space of $\boldsymbol{A}$.

last $n - r$ columns of $\boldsymbol{V}$: null space of $\boldsymbol{A}$.

First r columns of $\boldsymbol{U}$: column space of $\boldsymbol{A}$.

last $m - r$ columns of $\boldsymbol{U}$: null space of $\boldsymbol{A}^{\mathrm{T}}$.

Maybe it's easy to understand it in graph:

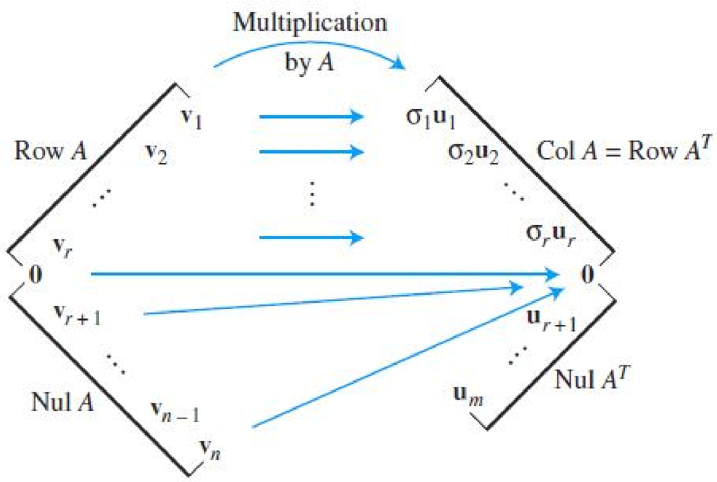


Figure 8.2: The fundamental spaces and the action of $\boldsymbol{A}$.

8.2.2.3. Remark 3: vector form

Recall we can write eigendecomposition in *vector form:*

$$\boldsymbol{A} = \lambda_1 \boldsymbol{v}_1 \boldsymbol{v}_1^{\mathrm{T}} + \cdots + \lambda_n \boldsymbol{v}_n \boldsymbol{v}_n^{\mathrm{T}}$$

Also, we could write the **general** SVD decomposition in remark 2 into *vector form*:

$$\boldsymbol{A} = \sigma \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \cdots + \sigma \boldsymbol{u}_r \boldsymbol{v}_r^{\mathrm{T}}$$

where $r = \operatorname{rank}(\boldsymbol{A}) =$ number of nonzero singular values. Here leads to the third meaning for the rank:

Proposition 8.3 The rank of $m \times n$ matrix $\boldsymbol{A}$ is the number of nonzero singular values.

Proof. Without loss of generality, we set $m \geq n$. And we assume there are exactly s zero singular values of $\boldsymbol{A}$, which means there are s zero eigenvalues of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ assocaiated with their s ind. eigenvectors. (Independence is due to the diagonalizable of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$.) In other words, the eigenspace of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ for $\lambda = 0$ has dimension s.

The eigenspace of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ for $\lambda = 0$ is given by

$$\{\boldsymbol{x} : \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}\}.$$

Hence its dimension is given by

$$\operatorname{rank}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{A}) := n - r.$$

Hence $s = n - r$. And obviously, the number of **nonzero** singular values is $n - s = r$.

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

8.2.2.4. Compact SVD

Hence any matrix with rank r can be factorized into

$$\begin{aligned}\boldsymbol{A} &= \boldsymbol{U}\Sigma\boldsymbol{V}^{\mathrm{T}} \\ &= \begin{bmatrix} \boldsymbol{u}_1 & \dots & \boldsymbol{u}_r \end{bmatrix} \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix} \begin{bmatrix} \boldsymbol{v}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{v}_r^{\mathrm{T}} \end{bmatrix}\end{aligned}$$

where $\boldsymbol{U} \in \mathbb{R}^{m\times r}$ and $\boldsymbol{V} \in \mathbb{R}^{r\times n}$ are both orthogonal matrix. And $\Sigma = \operatorname{diag}(\sigma_1, \dots, \sigma_r)$, where $\sigma_i > 0$ for $i = 1, 2, \dots, r$.

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 8.7 [perpendicular for matrix] For two real $n \times n$ matrix $\boldsymbol{A}$ and $\boldsymbol{B}$, they are said to be **perpendicular (orthogonal)** if the inner product between $\boldsymbol{A}$ and $\boldsymbol{B}$ is zero:

$$\langle \boldsymbol{A}, \boldsymbol{B} \rangle = \text{trace}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}) = \sum_{i,j=1}^{n} A_{ij}B_{ij} = 0.$$

Decompose $\boldsymbol{A} := \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^{\mathrm{T}}$. If we set $\boldsymbol{A}_i = \boldsymbol{u}_i \boldsymbol{v}_i^{\mathrm{T}} \sigma_i$, let's show $\boldsymbol{A}_i$'s are perpendicular:

$$\begin{aligned}
\langle \boldsymbol{A}_i, \boldsymbol{A}_j \rangle &= \text{trace}(\boldsymbol{A}_j^{\mathrm{T}} \boldsymbol{A}_i) \\
&= \text{trace}(\sigma_i \sigma_j \boldsymbol{v}_j \boldsymbol{u}_j^{\mathrm{T}} \boldsymbol{u}_i \boldsymbol{v}_i^{\mathrm{T}}) = \sigma_i \sigma_j \, \text{trace}(\boldsymbol{v}_j \boldsymbol{u}_j^{\mathrm{T}} \boldsymbol{u}_i \boldsymbol{v}_i^{\mathrm{T}}) \\
&= \sigma_i \sigma_j \, \text{trace}(\boldsymbol{v}_j (\boldsymbol{u}_j^{\mathrm{T}} \boldsymbol{u}_i) \boldsymbol{v}_i^{\mathrm{T}}) = \sigma_i \sigma_j \, \text{trace}(\boldsymbol{v}_j \boldsymbol{0} \boldsymbol{v}_i^{\mathrm{T}}) \\
&= 0.
\end{aligned}$$

So what is rank? How many rank 1 matrices do we need to pick to construct matrix $\boldsymbol{A}$? In fact, this number has no upper bound. For example, if we obtain

$$\boldsymbol{A} = \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \boldsymbol{u}_2 \boldsymbol{v}_2^{\mathrm{T}}$$

Then we can always decompose any rank 1 matrix into 2 rank 1 matrix:

$$\boldsymbol{A} = \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \frac{1}{2}\boldsymbol{u}_2 \boldsymbol{v}_2^{\mathrm{T}} + \frac{1}{2}\boldsymbol{u}_2 \boldsymbol{v}_2^{\mathrm{T}}.$$

But this number has a lower bound, that is rank. In other words, rank($\boldsymbol{A}$)=smallest number of rank 1 matrices with sum $\boldsymbol{A}$.

R Up till now, rank($\boldsymbol{A}$) has three meanings:

- $\text{rank}(\boldsymbol{A}) = \dim(\text{row}(\boldsymbol{A}))$
- $\text{rank}(\boldsymbol{A}) = \dim(\text{col}(\boldsymbol{A}))$
- $\text{rank}(\boldsymbol{A}) =$ smallest number of rank 1 matrices with sum $\boldsymbol{A}$.

8.2.3. Best Low-Rank Approximation

Given matrix $\boldsymbol{A}$. What is the *best rank k approximation*? In other words, given matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, what is the optimal solution for the model:

$$\begin{aligned} \min \quad & \|\boldsymbol{A} - \boldsymbol{Z}\|_F^2 \\ \text{s.t.} \quad & \operatorname{rank}(\boldsymbol{Z}) = k \\ & \boldsymbol{Z} \in \mathbb{R}^{m \times n} \end{aligned}$$

Firstly let's introduce the definition for Frobenius norm:

Definition 8.8 [Frobenius norm] The Frobenius norm for $m \times n$ matrix $\boldsymbol{A}$ is given by

$$\|\boldsymbol{A}\|_F = \sqrt{\langle \boldsymbol{A}, \boldsymbol{A} \rangle} = \sqrt{\operatorname{trace}(\boldsymbol{A}^{\mathrm{T}} \boldsymbol{A})}.$$

Theorem 8.4 Suppose the SVD if $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is given by

$$\boldsymbol{A} = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \cdots + \sigma_r \boldsymbol{u}_r \boldsymbol{v}_r^{\mathrm{T}}.$$

where $\boldsymbol{u}_i$'s and $\boldsymbol{v}_i$'s are $\mathbb{R}^n$ vectors. And we suppose $\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$. 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.

8.2.3.1. Analogy with least square problem

For least square 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 $\{\boldsymbol{A}\boldsymbol{x} | \boldsymbol{x} \in \mathbb{R}^n\}$:

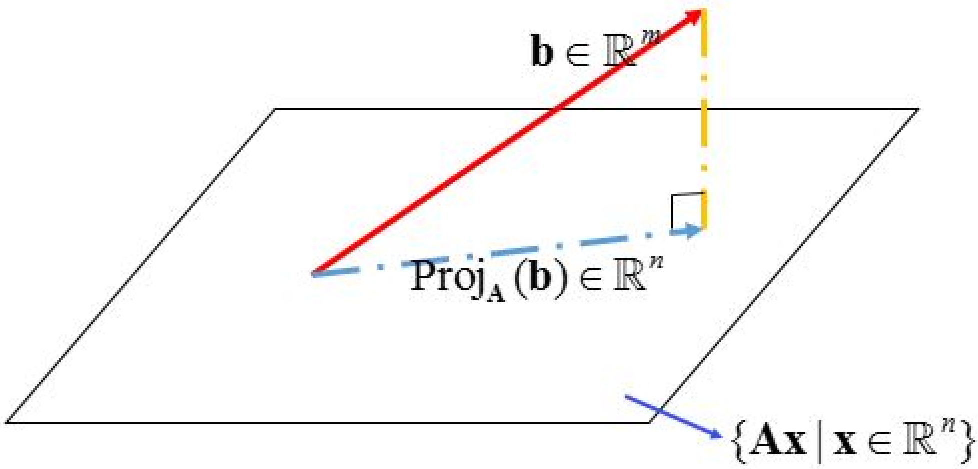


Figure 8.3: Least square problem: find $\boldsymbol{x}$ such that $\boldsymbol{Ax} = \mathrm{Proj}_{\boldsymbol{A}}(\boldsymbol{b})$.

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:

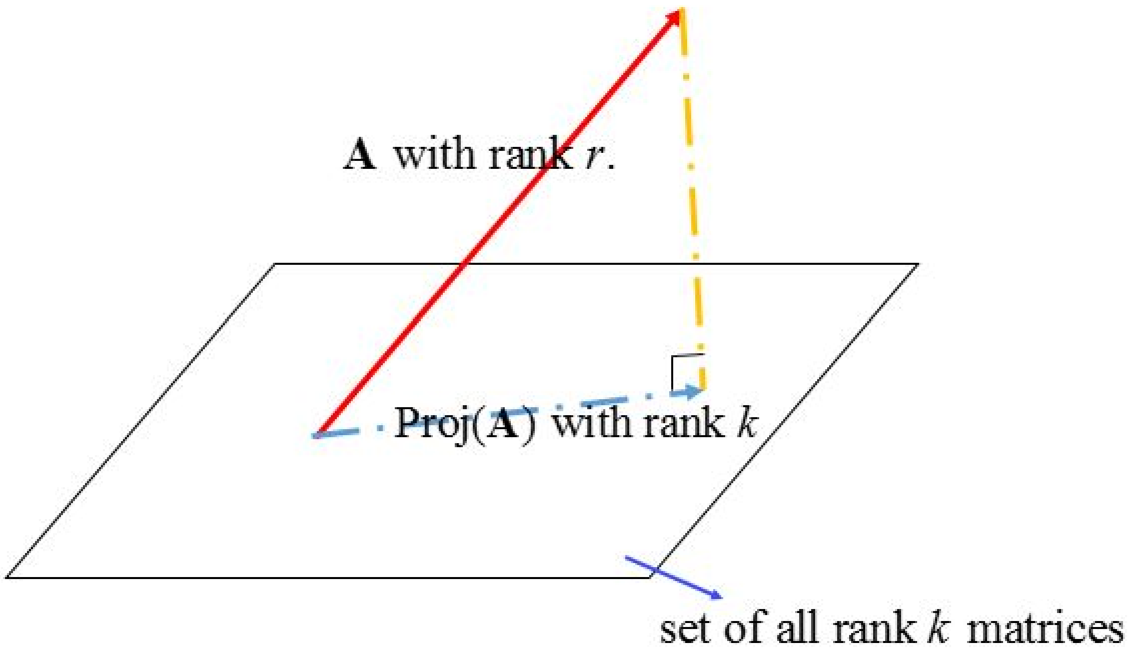


Figure 8.4: Best rank k approximation: find projection from rank r matrix to the plane that contains all rank k matrices

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 $\operatorname{span}\{\sin x, \cos x, \sin 2x, \cos 2x\}$.

   (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.

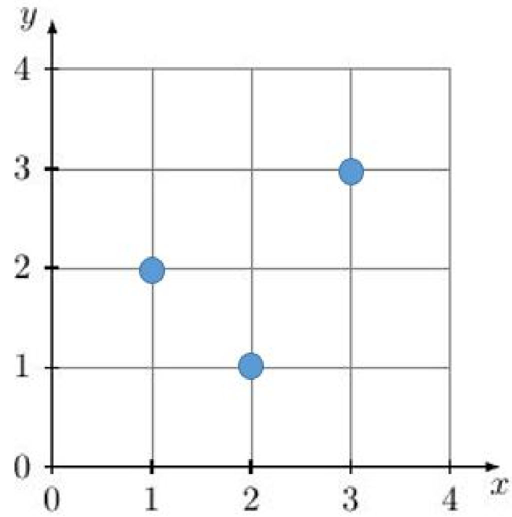
   Find 4 *linearly independent* eigenvectors of D^2. In other words, find f_k such that

   $$D^2(f_k) = \lambda_k f_k$$

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

$$R(\boldsymbol{x}) = \frac{\boldsymbol{x}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{x}}{\boldsymbol{x}^{\mathrm{T}} \boldsymbol{x}} \text{ for } \boldsymbol{x} \neq \mathbf{0}.$$

   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 $R(\boldsymbol{x})$.

$$\text{i.e. } \lambda_1 = \min_{\boldsymbol{x} \in \mathbb{R}^n - \{\mathbf{0}\}} R(\boldsymbol{x}).$$

   (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 $\forall i \in \{1, 2, \ldots, n\}$,

$$|\boldsymbol{M}_{ii}| \geq \sum_{j \neq i} |\boldsymbol{M}_{ij}|.$$

It is called **strictly diagonal dominant** if for $\forall i \in \{1, 2, \ldots, n\}$,

$$|\boldsymbol{M}_{ii}| > \sum_{j \neq i} |\boldsymbol{M}_{ij}|.$$

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 $\boldsymbol{A} = \mathbf{0}$.

(e) If $\boldsymbol{A}$ is the sum of 5 rank one matrices, then $\operatorname{rank}(\boldsymbol{A}) \leq 5$.

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 $\boldsymbol{AB} = \boldsymbol{BA}$.

Hint: Note that $\boldsymbol{A}$ *is said to be diagonalized by* $\boldsymbol{S}$ *if* $\boldsymbol{S}^{-1}\boldsymbol{AS}$ *is diagonal.*

(b) Suppose $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n\times n}$ satisfy $\boldsymbol{AB} = \boldsymbol{BA}$, 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* $\boldsymbol{Q}^{-1}\boldsymbol{AQ} = \boldsymbol{D}$, *where* $\boldsymbol{D}$ *is diagonal. Then you should express* $\boldsymbol{D}$. *Then you compute* $\boldsymbol{Q}^{-1}\boldsymbol{BQ} = \boldsymbol{C}$, *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) $\operatorname{rank}(\boldsymbol{A} \circ \boldsymbol{B}) \leq \operatorname{rank}(\boldsymbol{A}) \operatorname{rank}(\boldsymbol{B})$;

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. *Proof.* [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

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned} \tag{10.1}$$

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:

$$\begin{aligned} &a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \\ &= a_{i1}(m_1y_1 + m_2z_1) + a_{i2}(m_1y_2 + m_2z_2) + \cdots + a_{in}(m_1y_n + m_2z_n) \\ &= m_1(a_{i1}y_1 + a_{i2}y_2 + \cdots + a_{in}y_n) + m_2(a_{i1}z_1 + a_{i2}z_2 + \cdots + a_{in}z_n) \\ &= m_1b_i + m_2b_i = (m_1 + m_2)b_i = b_i. \end{aligned}$$

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. *Proof.* [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. *Proof.* [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)

$$\boldsymbol{C} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}; \boldsymbol{D} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \implies \boldsymbol{C}\boldsymbol{D} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} = -\boldsymbol{D}\boldsymbol{C}$$

(d)

$$E = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}; F = \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix} \implies EF = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \mathbf{0}$$

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 $i = 1,2,\ldots,m$ and $i = 1,2,\ldots,q$.

•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 $i = 1,2,\ldots,m$ and $i = 1,2,\ldots,q$.

Hence we have $h_{ij} = g_{ij}$, $i = 1,2,\ldots,m$ and $i = 1,2,\ldots,q$. Hence we have $\boldsymbol{H} = \boldsymbol{G} \implies (\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$.

6. *Proof.* [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. *Proof.* [Solution.]

$$A = \begin{bmatrix} a & a & a & a \\ a & b & b & b \\ a & b & c & c \\ a & b & c & d \end{bmatrix} \xrightarrow{E_{41}E_{31}E_{21}} \begin{bmatrix} a & a & a & a \\ 0 & b-a & b-a & b-a \\ 0 & b-a & c-a & c-a \\ 0 & b-a & c-a & d-a \end{bmatrix}$$

$$\xrightarrow{E_{42}E_{32}} \begin{bmatrix} a & a & a & a \\ 0 & b-a & b-a & b-a \\ 0 & 0 & c-b & c-b \\ 0 & 0 & c-b & d-b \end{bmatrix} \xrightarrow{E_{43}} \begin{bmatrix} a & a & a & a \\ 0 & b-a & b-a & b-a \\ 0 & 0 & c-b & c-b \\ 0 & 0 & 0 & d-c \end{bmatrix} = U$$

$$\implies E_{43}E_{42}E_{32}E_{41}E_{31}E_{21}A = U \implies A = E_{21}^{-1}E_{31}^{-1}E_{41}^{-1}E_{32}^{-1}E_{42}^{-1}E_{43}^{-1}U$$

$$\implies A = \begin{bmatrix} a & a & a & a \\ a & b & b & b \\ a & b & c & c \\ a & b & c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} a & a & a & a \\ 0 & b-a & b-a & b-a \\ 0 & 0 & c-b & c-b \\ 0 & 0 & 0 & d-c \end{bmatrix}$$

$$\implies L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}; \qquad U = \begin{bmatrix} a & a & a & a \\ 0 & b-a & b-a & b-a \\ 0 & 0 & c-b & c-b \\ 0 & 0 & 0 & d-c \end{bmatrix}$$

In order to get four pivots, we need to let the diagonal entries of 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.* [Proof.] *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. *Proof.* [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. *Proof.* [Solution.] The inverse are respectively given by $\begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ -\boldsymbol{C} & \boldsymbol{I} \end{bmatrix}$, $\begin{bmatrix} \boldsymbol{A}^{-1} & \boldsymbol{0} \\ -\boldsymbol{D}^{-1}\boldsymbol{CA}^{-1} & \boldsymbol{D}^{-1} \end{bmatrix}$,

   - $$\begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{C} & \boldsymbol{I} \end{bmatrix}\begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ -\boldsymbol{C} & \boldsymbol{I} \end{bmatrix} = \begin{bmatrix} \boldsymbol{II} + \boldsymbol{0}(-\boldsymbol{C}) & \boldsymbol{I0} + \boldsymbol{0I} \\ \boldsymbol{CI} + \boldsymbol{I}(-\boldsymbol{C}) & \boldsymbol{C0} + \boldsymbol{II} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix}$$

     Hence $\begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ -\boldsymbol{C} & \boldsymbol{I} \end{bmatrix}$ is the right inverse of $\begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{C} & \boldsymbol{I} \end{bmatrix}$, hence $\begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ -\boldsymbol{C} & \boldsymbol{I} \end{bmatrix}$ is the inverse of $\begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{C} & \boldsymbol{I} \end{bmatrix}$.

-
$$\begin{bmatrix} A & 0 \\ C & D \end{bmatrix} \begin{bmatrix} A^{-1} & 0 \\ -D^{-1}CA^{-1} & D^{-1} \end{bmatrix} = \begin{bmatrix} AA^{-1} + 0(-D^{-1}CA^{-1}) & A0 + 0D^{-1} \\ CA^{-1} + D(-D^{-1}CA^{-1}) & C0 + DD^{-1} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$$

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. *Proof.* [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. *Proof.* [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

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

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. *Proof.* [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. *Proof.* [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 $\boldsymbol{A} = \begin{bmatrix} 1 & 7 \\ 7 & 0 \end{bmatrix}$, $\boldsymbol{B} = \begin{bmatrix} 2 & 5 \\ 5 & 1 \end{bmatrix}$. Then $\boldsymbol{A} + \boldsymbol{B} = \begin{bmatrix} 3 & 12 \\ 12 & 1 \end{bmatrix}$, $\boldsymbol{A} - \boldsymbol{B} = \begin{bmatrix} -1 & 2 \\ 2 & -1 \end{bmatrix}$. The product $(\boldsymbol{A}+\boldsymbol{B})(\boldsymbol{A}-\boldsymbol{B})$ is given by:

$$(\boldsymbol{A}+\boldsymbol{B})(\boldsymbol{A}-\boldsymbol{B}) = \begin{bmatrix} 21 & -6 \\ -10 & 23 \end{bmatrix}$$

which is obviously *not symmetric.*

(c) $\boldsymbol{ABA}$ is symmetric. The reason is that

$$(\boldsymbol{ABA})^{\mathrm{T}} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{ABA}$$

(d) $\boldsymbol{ABAB}$ may not be symmetric, let me raise a counterexample to explain it:

Suppose $\boldsymbol{A} = \begin{bmatrix} 1 & 7 \\ 7 & 0 \end{bmatrix}$, $\boldsymbol{B} = \begin{bmatrix} 2 & 5 \\ 5 & 1 \end{bmatrix}$. Then the product $\boldsymbol{ABAB}$ is given by:

$$\boldsymbol{ABAB} = \begin{bmatrix} 1537 & 864 \\ 1008 & 1393 \end{bmatrix}$$

which is obviously *not symmetric.*

3. *Proof.* [Solution.] Starting from $\boldsymbol{A} = \boldsymbol{LDU}$, then $\boldsymbol{A} = \boldsymbol{L}(\boldsymbol{U}^{\mathrm{T}})^{-1} \times (\boldsymbol{U}^{\mathrm{T}}\boldsymbol{DU})$.

- $\boldsymbol{L}(\boldsymbol{U}^{\mathrm{T}})^{-1}$ is lower triangular with unit diagonals.

  Reason: $\boldsymbol{U}$ is upper triangular, hence $\boldsymbol{U}^{\mathrm{T}}$ is lower triangular, its inverse $(\boldsymbol{U}^{\mathrm{T}})^{-1}$ is also lower triangular. And $\boldsymbol{L}$ is also lower triangular. Hence the product $\boldsymbol{L}(\boldsymbol{U}^{\mathrm{T}})^{-1}$ remains lower triangular. Since $\boldsymbol{L}$ and $\boldsymbol{U}$ has unit diagonals, their transformation $\boldsymbol{L}(\boldsymbol{U}^{\mathrm{T}})^{-1}$ also has unit diagonals.

- $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{DU}$ is symmetric. The reason is that

$$(\boldsymbol{U}^{\mathrm{T}}\boldsymbol{DU})^{\mathrm{T}} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{D}^{\mathrm{T}}(\boldsymbol{U}^{\mathrm{T}})^{\mathrm{T}} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{DU}$$

In conclusion, here lists a new factorization of $\boldsymbol{A}$ into *triangular* times *symmetric*.

4. *Proof.* [Solution]

(a)

$$\boldsymbol{AX} + \boldsymbol{B} = \boldsymbol{C} \implies \boldsymbol{AX} = \boldsymbol{C} - \boldsymbol{B} \implies \boldsymbol{X} = \boldsymbol{A}^{-1}(\boldsymbol{C} - \boldsymbol{B}).$$

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)

$$\boldsymbol{XA} + \boldsymbol{B} = \boldsymbol{C} \implies \boldsymbol{XA} = \boldsymbol{C} - \boldsymbol{B} \implies \boldsymbol{X} = (\boldsymbol{C} - \boldsymbol{B})\boldsymbol{A}^{-1}.$$

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)

$$\boldsymbol{AX} + \boldsymbol{B} = \boldsymbol{X} \implies (\boldsymbol{A} - \boldsymbol{I})\boldsymbol{X} = -\boldsymbol{B} \implies \boldsymbol{X} = -(\boldsymbol{A} - \boldsymbol{I})^{-1}\boldsymbol{B}$$

Hence the soluion is given by

$$\boldsymbol{X} = -(\boldsymbol{A} - \boldsymbol{I})^{-1}\boldsymbol{B} = -\begin{bmatrix} 5-1 & 3 \\ 3 & 2-1 \end{bmatrix}^{-1} \begin{bmatrix} 6 & 2 \\ 2 & 4 \end{bmatrix} = -\frac{1}{4-9} \begin{bmatrix} 1 & -3 \\ -3 & 4 \end{bmatrix} \begin{bmatrix} 6 & 2 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 0 \\ - \end{bmatrix}$$

(d)

$$\boldsymbol{XA} + \boldsymbol{C} = \boldsymbol{X} \implies \boldsymbol{X}(\boldsymbol{A} - \boldsymbol{I}) = -\boldsymbol{C} \implies \boldsymbol{X} = -\boldsymbol{C}(\boldsymbol{A} - \boldsymbol{I})^{-1}$$

Hence the solution is given by

$$\boldsymbol{X} = -\boldsymbol{C}(\boldsymbol{A} - \boldsymbol{I})^{-1} = -\begin{bmatrix} 4 & -2 \\ -6 & 3 \end{bmatrix} \begin{bmatrix} -0.2 & 0.6 \\ 0.6 & -0.8 \end{bmatrix} = \begin{bmatrix} 2 & -4 \\ -3 & 6 \end{bmatrix}.$$

5. *Proof.* [Solution.] Firstly, we show $t_{jj} = u_{jj}r_{jj}$ for $j = 1, \dots, n$:

$$\begin{aligned}
t_{jj} &= \sum_{k=1}^{n} u_{jk}r_{kj} \\
&= \sum_{k=1, j<k} u_{jk}r_{kj} + u_{jj}r_{jj} + \sum_{k=1, j>k} u_{jk}r_{kj} \\
&= \sum_{k=1, j<k} u_{jk} \times 0 + u_{jj}r_{jj} + \sum_{k=1, j>k} 0 \times r_{kj} \\
&= u_{jj}r_{jj}
\end{aligned}$$

Secondly, we show that $t_{ij} = 0$ if $i > j$ for $i, j \in \{1, 2, \dots, n\}$:

$$\begin{aligned}
t_{ij} &= \sum_{k=1}^{n} u_{ik}r_{kj} \\
&= \sum_{k=1, k<i}^{n} u_{ik}r_{kj} + u_{ii}r_{ij} + \sum_{k=1, k>i}^{n} u_{ik}r_{kj} \\
&= \sum_{k=1, k<i}^{n} 0 \times r_{kj} + u_{ii} \times 0 + \sum_{k=1, k>i}^{n} u_{ik} \times 0 \\
&= 0
\end{aligned}$$

Hence $t_{ij} = 0$ for $i < j$. Hence $\boldsymbol{T}$ is upper triangular.

6. *Proof.* [Solution.]

(a)

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

(b)

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

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)

$$\boldsymbol{A}^3 = \begin{bmatrix} 2 & 4 & 1 & 4 & 1 \\ 4 & 2 & 3 & 5 & 1 \\ 1 & 3 & 0 & 1 & 1 \\ 4 & 5 & 1 & 2 & 3 \\ 1 & 1 & 1 & 3 & 0 \end{bmatrix}$$

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. *Proof.* [Solution.]

(a)

$$\begin{bmatrix} 1 & 2 & 3 & 1 & -3 \\ 2 & 5 & 5 & 4 & 9 \\ 3 & 7 & 8 & 5 & 6 \end{bmatrix} \xrightarrow[\text{Add } (-3)\times\text{Row 1 to Row 3}]{\text{Add } (-2)\times\text{Row 1 to Row 2}} \begin{bmatrix} 1 & 2 & 3 & 1 & -3 \\ 0 & 1 & -1 & 2 & 15 \\ 0 & 1 & -1 & 2 & 15 \end{bmatrix} \xrightarrow{\text{Add } (-1)\times\text{Row 2 to Row 3}}$$

$$\begin{bmatrix} 1 & 2 & 3 & 1 & -3 \\ 0 & 1 & -1 & 2 & 15 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow{\text{Add } (-2)\times\text{Row 2 to Row 1}} \begin{bmatrix} 1 & 0 & 5 & -3 & -33 \\ 0 & 1 & -1 & 2 & 15 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \text{(rref)}$$

(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 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ 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

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 4b_1 - b_2 - 5x_3 + 3x_4 + 33x_5 \\ -2b_1 + b_2 + x_3 - 2x_4 - 15x_5 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 4b_1 - b_2 \\ -2b_1 + b_2 \\ 0 \\ 0 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} -5 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} 3 \\ -2 \\ 0 \\ 1 \\ 0 \end{pmatrix} + x_5$$

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 $\boldsymbol{A}\boldsymbol{x} = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} x = \boldsymbol{0}$ has unique solution. Thus we only need to show $\boldsymbol{A} = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix}$ is invertible:

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & -3 \\ -2 & -2 & 3 \\ 2 & 4 & 6 \end{bmatrix} \xrightarrow[\text{Add } (-2)\times\text{Row 1 to Row 3}]{\text{Add } 2\times\text{Row 1 to Row 2}} \begin{bmatrix} 1 & 2 & -3 \\ 0 & 2 & -3 \\ 0 & 0 & 12 \end{bmatrix} \xrightarrow[\text{Row } 3\times\frac{1}{12}]{\text{Row } 2\times\frac{1}{2}} \begin{bmatrix} 1 & 2 & -3 \\ 0 & 1 & -\frac{3}{2} \\ 0 & 0 & 1 \end{bmatrix} \text{(rref)}$$

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 $\boldsymbol{U}\boldsymbol{x} = \boldsymbol{0}$.

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:

$$m_1(x-2)(x+2) + m_2x^2(x^4-2) + m_3(x^6-8) = (m_2+m_3)x^6 + (m_1-2m_2)x^2 + (-4m$$

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

  $$(m_2+m_3)x^6 + (m_1-2m_2)x^2 + (-4m_1-8m_3) \in \text{span}\{(x-2)(x+2), x^2(x^4-2),$$

  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

$$(m_2+m_3)x^6+(m_1-2m_2)x^2+(-4m_1-8m_3)\in \operatorname{span}\{x^4-4,x^6-8\}$$

$$\implies \operatorname{span}\{(x-2)(x+2),x^2(x^4-2),x^6-8\}\subset \operatorname{span}\{x^4-4,x^6-8\}$$

Conversely, by setting $m_1=2a_1+a_2, m_2=a_1, m_3=0$ we can show $\operatorname{span}\{x^4-4,x^6-8\}\subset \operatorname{span}\{(x-2)(x+2),x^2(x^4-2),x^6-8\}$.

Hence $\operatorname{span}\{x^4-4,x^6-8\}=\operatorname{span}\{(x-2)(x+2),x^2(x^4-2),x^6-8\}$

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 $\operatorname{span}\{(x-2)(x+2),x^2(x^4-2),x^6-8\}$.

Hence $\dim(\operatorname{span}\{(x-2)(x+2),x^2(x^4-2),x^6-8\})=2$.

(e) Firstly, it's easy to verify that 5 and $\cos^2 x$ are ind.

Next, let's show $\operatorname{span}\{5,\cos^2 x\}=\operatorname{span}\{5,\cos 2x,\cos^2 x\}$:

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 $\text{span}\{5, \cos 2x, \text{co}$
Hence $\dim(\text{span}\{5, \cos 2x, \cos^2 x\}) = 2$.

3. *Proof.* [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 $v_1, v_2, \ldots, v_n$, suppose v is the any vector in $\boldsymbol{V}$.

- Let's show $v_1, v_2, \ldots, v_n, v$ must be dep:

It suffices to show $c_1v_1 + \cdots + c_nv_n + c_{n+1}v = \mathbf{0}$ has nontrival solution for $c_1, \ldots, c_{n+1} \in \mathbb{R}$.

$$\iff \boldsymbol{A}\boldsymbol{x} = \mathbf{0} \text{ has nontrival solution, where } \boldsymbol{A} = \left[\begin{array}{c|c|c|c} v_1 & \ldots & v_n & v \end{array}\right]$$

which is obviously true since $\boldsymbol{A}$ is a $n \times n+1$ matrix $(n < n+1)$

- Hence there exists $(c_1, c_2, \ldots, c_{n+1}) \neq (0, 0, \ldots, 0)$ such that

$$c_1v_1 + \cdots + c_nv_n + c_{n+1}v = \mathbf{0}$$

If $c_{n+1} = 0$, then we have $(c_1, c_2, \ldots, c_n) \neq (0, 0, \ldots, 0)$ such that

$$c_1v_1 + \cdots + c_nv_n = \mathbf{0},$$

which contradicts that $v_1, \ldots, v_n$ are ind.

Hence $c_{n+1} \neq 0$. Then any $v \in \boldsymbol{V}$ could be expressed as:

$$v = -\frac{c_1}{c_{n+1}}v_1 - \frac{c_2}{c_{n+1}}v_2 - \cdots - \frac{c_n}{c_{n+1}}v_n$$

which means $v_1, v_2, \ldots, v_n$ spans $\boldsymbol{V}$. And they are ind.

So they form a basis for $\boldsymbol{V}$.

(b) Suppose $v_1 \ldots, v_n$ spans $\boldsymbol{V}$. We assume that they are dep. Hence there exists $(c_1, c_2, \ldots, c_n) \neq (0, 0, \ldots, 0)$ such that

$$c_1v_1 + c_2v_2 + \cdots + c_nv_n = \mathbf{0}$$

WLOG, we set $c_n \neq 0$. Hence we could express v_n as:

$$v_n = -\frac{c_1}{c_n}v_1 - \frac{c_2}{c_n}v_2 - \cdots - \frac{c_{n-1}}{c_n}v_{n-1}$$

- We claim that $v_1, v_2, \ldots, v_{n-1}$ still spans $\boldsymbol{V}$:

For any vector $v \in \boldsymbol{V}$, since $v_1, \ldots, v_n$ spans $\boldsymbol{V}$, v could be expressed in form of $v_1, \dots, v_n$:

$$v = m_1 v_1 + \cdots + m_n v_n$$

where $m_1, \dots, m_n \in \mathbb{R}$.

Hence it could also be expressed in form of $v_1, \dots, v_{n-1}$:

$$\begin{aligned} v &= m_1 v_1 + \cdots + m_n\left(-\frac{c_1}{c_n} v_1 - \frac{c_2}{c_n} v_2 - \cdots - \frac{c_{n-1}}{c_n} v_{n-1}\right) \\ &= \left(m_1 - \frac{m_n c_1}{c_n}\right) v_1 + \left(m_2 - \frac{m_n c_2}{c_n}\right) v_2 - \cdots - \left(m_{n-1} - \frac{m_n c_{n-1}}{c_n}\right) v_{n-1} \end{aligned}$$

Hence $v_1.v_2, \dots, v_{n-1}$ still spans $\boldsymbol{V}$.

- If $v-_1, v_2, \dots, v_n$ still dep, we continue eliminating vectors until we get ind. vectors, say, $v_1, v_2, \dots, v_k$. 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 $\boldsymbol{u}, \boldsymbol{v} = \mathbf{0}$ since $\boldsymbol{U} \cap \boldsymbol{V} = \{\mathbf{0}\}$. Thus we have

$$c_1 u_1 + c_2 u_2 + \cdots + c_m u_m = \mathbf{0}$$
$$c_{m+1} v_1 + c_{m+2} v_2 + \cdots + c_{m+n} v_n = \mathbf{0}$$

By the independence of $u_1, \ldots, u_m$ and the independence of $v_1, \ldots, v_n$ it follows that

$$c_1 = c_2 = \cdots = c_{m+n} = 0$$

- Thus $\{u_1, u_2, \ldots, u_m, v_1, v_2, \ldots, v_n\}$ 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 range$(\boldsymbol{A})$ is $\{a_1, \ldots, a_s\}$; $\boldsymbol{B} = \begin{bmatrix} B_1 & \Big| & \ldots & \Big| & B_n \end{bmatrix}$ one basis for range$(\boldsymbol{B})$ is $\{b_1, \ldots, b_t\}$. Thus we obtain:

$$\begin{aligned} \dim(\text{range}(\boldsymbol{A}) + \text{range}(\boldsymbol{B})) &= \dim(a_1, \ldots, a_s, b_1, \ldots, b_t) \\ &\leq s + t \\ &= \dim(\text{range}(\boldsymbol{A})) + \dim(\text{range}(\boldsymbol{B})) \\ &= \text{rank}(\boldsymbol{A}) + \text{rank}(\boldsymbol{B}) \end{aligned}$$

Hence we have

$$\begin{aligned}
\operatorname{rank}(\boldsymbol{A}+\boldsymbol{B}) &= \dim(\operatorname{range}(\boldsymbol{A}+\boldsymbol{B})) \\
&\le \dim(\operatorname{range}(\boldsymbol{A})+\operatorname{range}(\boldsymbol{B})) \\
&\le \operatorname{rank}(\boldsymbol{A})+\operatorname{rank}(\boldsymbol{B})
\end{aligned}$$

7. *Proof.*

   (a) We assume $\boldsymbol{A} = \left[\begin{array}{c|c|c} A_1 & \dots & A_n \end{array}\right]$, $\boldsymbol{B} = \left[\begin{array}{c|c|c} B_1 & \dots & B_n \end{array}\right]^{\mathrm{T}}$.
   Hence $\boldsymbol{AB}$ could be expressed as:

   $$\boldsymbol{AB} = A_1B_1 + \cdots + A_nB_n$$

   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 $\operatorname{rank}(\boldsymbol{AB}) = dim(\operatorname{col}(\boldsymbol{AB})) \le \dim(\operatorname{col}(\boldsymbol{A})) = \operatorname{rank}(\boldsymbol{A})$

   (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), $\operatorname{rank}(\boldsymbol{ABB}^{-1}) \le \operatorname{rank}(\boldsymbol{AB})$.
   In conclusion,

   $$\operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{ABB}^{-1}) \le \operatorname{rank}(\boldsymbol{AB}) \le \operatorname{rank}(\boldsymbol{A})$$

   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 $\boldsymbol{A} = \left[\begin{array}{c|c|c} v_1 & \dots & v_{n-1} \end{array}\right]$.
   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. *Proof.* [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. *Proof.* [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. *Proof.* [Solution.]

$$D(e^x) = 1 \cdot e^x + 0 \cdot xe^x + 0 \cdot x^2 e^x$$
$$D(xe^x) = 1 \cdot e^x + 1 \cdot xe^x + 0 \cdot x^2 e^x$$
$$D(x^2 e^x) = 0 \cdot e^x + 2 \cdot xe^x + 1 \cdot x^2 e^x$$

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. *Proof.* [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 $\text{col}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}) \subset \text{col}(\boldsymbol{A})$:

  For any $\boldsymbol{b} \in \text{col}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})$, there exists $\boldsymbol{x}_0$ such that $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_0 = \boldsymbol{b}$, which implies $\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_0) = \boldsymbol{b}$. Hence there exists vector $(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_0)$ such that

  $$\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_0) = \boldsymbol{b}$$

  Hence $\boldsymbol{b} \in \text{col}(\boldsymbol{A})$. Hence $\text{col}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}) \subset \text{col}(\boldsymbol{A})$.
- In part *b* we will show $\text{rank}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}) = \text{rank}(\boldsymbol{A})$. Hence $\dim(\text{col}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})) = \dim(\text{col}(\boldsymbol{A}))$.
- We assume $\dim(\text{col}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})) = \dim(\text{col}(\boldsymbol{A})) = n$, the basis for $\text{col}(\boldsymbol{A})$ is $\{v_1, v_2, \dots, v_n\}$. Thus since $\text{col}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}) \subset \text{col}(\boldsymbol{A})$, basis $\{v_1, v_2, \dots, v_n\}$ must span $\text{col}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})$. Since $\dim(\text{col}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})) = n$, $\{v_1, v_2, \dots, v_n\}$ must be the basis for $\text{col}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})$.
- Since $\text{col}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})$ and $\text{col}(\boldsymbol{A})$ have the same basis, we obtain $\text{col}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}) = \text{col}(\boldsymbol{A})$.

(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 $\boldsymbol{x}_0^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}_0 = \mathbf{0}$, 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 $\mathrm{rank}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) + \dim(N(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})) = n = \mathrm{rank}(\boldsymbol{A}) + \dim(N(\boldsymbol{A}))$.

- Since $\dim(N(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})) = \dim(N(\boldsymbol{A}))$, we obtain $\mathrm{rank}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = \mathrm{rank}(\boldsymbol{A})$.
- Similarly, we obtain $\mathrm{rank}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}) = \mathrm{rank}(\boldsymbol{A}^{\mathrm{T}})$ by changing $\boldsymbol{A}$ into $\boldsymbol{A}^{\mathrm{T}}$.
- Obviously, $\mathrm{rank}(\boldsymbol{A}^{\mathrm{T}}) = \dim(\mathrm{row}(\boldsymbol{A}^{\mathrm{T}})) = \dim(\mathrm{col}(\boldsymbol{A})) = \mathrm{rank}(\boldsymbol{A})$.

In conclusion, $\mathrm{rank}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}) = \mathrm{rank}(\boldsymbol{A}^{\mathrm{T}}) = \mathrm{rank}(\boldsymbol{A}) = \mathrm{rank}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})$.

10.1.6. Solution to Assignment Six

1. *Proof.* [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. *Proof.* [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 $Q^{\mathrm{T}}Q = I$:

$$Q^{\mathrm{T}}Q = \begin{bmatrix} q_1^{\mathrm{T}} \\ q_2^{\mathrm{T}} \\ \vdots \\ q_n^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} q_1 & q_2 & \dots & q_n \end{bmatrix} = \begin{bmatrix} q_1^{\mathrm{T}}q_1 & & & \\ & q_2^{\mathrm{T}}q_2 & & \\ & & \ddots & \\ & & & q_n^{\mathrm{T}}q_n \end{bmatrix} = I_n.$$

- Hence we derive

$$\begin{aligned} \|Qx\|^2 &= x^{\mathrm{T}}Q^{\mathrm{T}}Qx \\ &= x^{\mathrm{T}}(Q^{\mathrm{T}}Q)x = x^{\mathrm{T}}Ix \\ &= x^{\mathrm{T}}x \\ &= \|x\|^2 \end{aligned}$$

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. *Proof.* [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

  $$\begin{aligned} p_1(-x) &= \frac{p^*(-x)+p^*(x)}{2} = p_1(x) \\ p_2(-x) &= \frac{p^*(-x)-p^*(x)}{2} = -p_2(x), \end{aligned}$$

  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. *Proof.* [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. *Proof.* [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 $\boldsymbol{A}$:

$$\text{Define } \boldsymbol{A} := \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 \end{bmatrix} \qquad \langle \boldsymbol{a}_1, \boldsymbol{a}_2 \rangle = 0 \implies \text{Columns of } \boldsymbol{A} \text{ are orthogonal.}$$

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:

$$\boldsymbol{A}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}} \boldsymbol{b} \iff \boldsymbol{R}^{\mathrm{T}} \boldsymbol{Q}^{\mathrm{T}} \boldsymbol{Q} \boldsymbol{R} \boldsymbol{x} = \boldsymbol{R}^{\mathrm{T}} \boldsymbol{Q}^{\mathrm{T}} \boldsymbol{b} \iff \boldsymbol{R}^{\mathrm{T}} \boldsymbol{R} \boldsymbol{x} = \boldsymbol{R}^{\mathrm{T}} \boldsymbol{Q}^{\mathrm{T}} \boldsymbol{b}$$

$$\begin{aligned} \implies \boldsymbol{x} = \boldsymbol{R}^{-1} \boldsymbol{Q}^{\mathrm{T}} \boldsymbol{b} &= \frac{1}{5\sqrt{2}} \begin{bmatrix} \sqrt{10} & 0 \\ 0 & \sqrt{5} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ -\frac{2}{\sqrt{10}} & -\frac{1}{\sqrt{10}} & 0 & \frac{1}{\sqrt{10}} & \frac{2}{\sqrt{10}} \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ -1 \end{bmatrix} \end{aligned}$$

$$\text{Hence we have } \begin{cases} C = 1 \\ D = -1. \end{cases} \qquad \text{The best line is } \hat{y} = 1 - x.$$

10.1.7. Solution to Assignment Seven

1. *Proof.* [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 $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ are columns of $\boldsymbol{X}$, and $\boldsymbol{X}^{-1}\boldsymbol{X} = \boldsymbol{I}$, we obtain

$$\boldsymbol{X}^{-1}\boldsymbol{x}_i = \boldsymbol{e}_i \text{ for } i = 1, \ldots, k.$$

Hence

$$\begin{aligned}
\boldsymbol{B} =& \boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{X} \\
=& \begin{bmatrix} \lambda \boldsymbol{X}^{-1}\boldsymbol{x}_1 & \ldots & \lambda \boldsymbol{X}^{-1}\boldsymbol{x}_k & \boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{x}_{k+1} & \ldots & \boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{x}_n \end{bmatrix} \\
=& \begin{bmatrix} \lambda \boldsymbol{e}_1 & \ldots & \lambda \boldsymbol{X}^{-1}\boldsymbol{e}_k & \boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{x}_{k+1} & \ldots & \boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{x}_n \end{bmatrix} \\
=& \begin{bmatrix} \lambda & 0 & \ldots & 0 & b_{1(k+1)} & \ldots & b_{1n} \\ 0 & \lambda & \ldots & 0 & b_{2(k+1)} & \ldots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda & b_{k(k+1)} & \ldots & b_{kn} \\ 0 & 0 & \ldots & 0 & b_{(k+1)(k+1)} & \ldots & b_{(k+1)n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & b_{n(k+1)} & \ldots & b_{nn} \end{bmatrix}
\end{aligned}$$

If we write $\boldsymbol{B}$ in block matrix form, then we obtain:

$$\boldsymbol{B} = \begin{bmatrix} \lambda \boldsymbol{I} & \boldsymbol{B}_{12} \\ \boldsymbol{0} & \boldsymbol{B}_{22} \end{bmatrix}.$$

(b) For a fixed eigenvalue λ^*, $\boldsymbol{B}$ could be written as

$$\boldsymbol{B} = \begin{bmatrix} \lambda^* & 0 & \ldots & 0 & b_{1(k+1)} & \ldots & b_{1n} \\ 0 & \lambda^* & \ldots & 0 & b_{2(k+1)} & \ldots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda^* & b_{k(k+1)} & \ldots & b_{kn} \\ 0 & 0 & \ldots & 0 & b_{(k+1)(k+1)} & \ldots & b_{(k+1)n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & b_{n(k+1)} & \ldots & b_{nn} \end{bmatrix}$$

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 $\begin{vmatrix} \lambda \boldsymbol{I} - \boldsymbol{B} \end{vmatrix}$, we cross the first k columns to get

$$\begin{vmatrix} \lambda \boldsymbol{I} - \boldsymbol{B} \end{vmatrix} = (\lambda - \lambda^*)^k \begin{vmatrix} \lambda - b_{(k+1)(k+1)} & \dots & -b_{(k+1)n} \\ \vdots & \ddots & \vdots \\ -b_{n(k+1)} & \dots & \lambda - b_{nn} \end{vmatrix}$$

Hence the term $(\lambda - \lambda^*)$ appears at least k times in the characteristic polynomial of $\begin{vmatrix} \lambda \boldsymbol{I} - \boldsymbol{B} \end{vmatrix}$.

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. *Proof.* [Solution.]

(a) $\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x} \implies (\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{x} = \boldsymbol{0}$. Since $\lambda = 0$, we only need to investigate the dimension for $\boldsymbol{x}$, where $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$.

Since $\boldsymbol{A} = \boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}$, $\operatorname{rank}(\boldsymbol{A}) = 1$. 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),

$$\lambda_1 = \lambda_2 = \cdots = \lambda_{n-1} = 0.$$

The sum of the eigenvalues is the trace of $\boldsymbol{A}$ which equals to $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$. Thus

$$\sum_{i=1}^{n}\lambda_i = \lambda_n = \mathrm{trace}(\boldsymbol{A}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}.$$

Hence the remaining eigenvalue of $\boldsymbol{A}$ is $\lambda_n = \mathrm{trace}(\boldsymbol{A}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$.

(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 $\boldsymbol{D} = \mathrm{diag}(\lambda_1,\dots,\lambda_n)$, 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. *Proof.* [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 $\boldsymbol{D}^{k+1} = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} p & q \\ s & t \end{bmatrix} = \begin{bmatrix} 2p+s & 2q+ \\ 2s & 2t \end{bmatrix}$ $\begin{bmatrix} p(k+1) & q(k+1) \\ s(k+1) & t(k+1) \end{bmatrix}$.
- 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. *Proof.* [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 $\boldsymbol{I} + \boldsymbol{A} + \frac{1}{2!}\boldsymbol{A}^2 + \cdots := e^{\boldsymbol{A}}$, 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. *Proof.* [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 $\forall\boldsymbol{A}\boldsymbol{v} \in C(\boldsymbol{A})$,

  $$(\boldsymbol{A}\boldsymbol{v})^{\mathrm{T}}\boldsymbol{u} = \boldsymbol{v}^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{u}) = \boldsymbol{v}^{\mathrm{T}}\boldsymbol{0} = \boldsymbol{0}. \implies C(\boldsymbol{A}) \perp N(\boldsymbol{A}^{\mathrm{T}}) \Longleftrightarrow N(\boldsymbol{A}^{\mathrm{T}}) \perp C(\boldsymbol{A}).$$

- 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 $\forall \boldsymbol{A}\boldsymbol{v} \in C(\boldsymbol{A})$,

$$(\boldsymbol{A}\boldsymbol{v})^{\mathrm{H}}\boldsymbol{u} = \boldsymbol{v}^{\mathrm{H}}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{u}) = \boldsymbol{v}^{\mathrm{H}}\boldsymbol{0} = \boldsymbol{0}. \implies C(\boldsymbol{A}) \perp N(\boldsymbol{A}^{\mathrm{H}}) \iff N(\boldsymbol{A}^{\mathrm{H}}) \perp C(\boldsymbol{A}).$$

- 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. *Proof.* [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 $m \times n$ ($m \ge n$) 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 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$.

Proof. [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

$$
\begin{aligned}
\boldsymbol{A}\boldsymbol{A}^{+} &= \boldsymbol{U}\begin{bmatrix}\boldsymbol{I}_n & \mathbf{0}_{n\times(m-n)}\\ \mathbf{0}_{m-n} & \mathbf{0}_{(m-n)\times(m-n)}\end{bmatrix}\boldsymbol{U}^{\mathrm{T}}\\
&= \begin{bmatrix}\boldsymbol{U}_1 & \boldsymbol{U}_2\end{bmatrix}\begin{bmatrix}\boldsymbol{I}_n & \mathbf{0}_{n\times(m-n)}\\ \mathbf{0}_{m-n} & \mathbf{0}_{(m-n)\times(m-n)}\end{bmatrix}\begin{bmatrix}\boldsymbol{U}_1^{\mathrm{T}}\\ \boldsymbol{U}_2^{\mathrm{T}}\end{bmatrix}\\
&= \begin{bmatrix}\boldsymbol{U}_1\boldsymbol{I}_n\boldsymbol{U}_1^{\mathrm{T}} & \mathbf{0}_{n\times(m-n)}\\ \mathbf{0}_{m-n} & \mathbf{0}_{(m-n)\times(m-n)}\end{bmatrix}\\
&= \begin{bmatrix}\boldsymbol{I}_n & \mathbf{0}_{n\times(m-n)}\\ \mathbf{0}_{m-n} & \mathbf{0}_{(m-n)\times(m-n)}\end{bmatrix} \qquad \text{due to the orthogonality of } \boldsymbol{U}.
\end{aligned}
$$

Moreover, $\boldsymbol{A}^{+}\boldsymbol{A} = \boldsymbol{V}\Sigma^{+}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\Sigma\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{V}\Sigma^{+}\Sigma\boldsymbol{V}^{\mathrm{T}}$.

You can verify by yourself that $\Sigma^{+}\Sigma = \boldsymbol{I}$.

Hence $\boldsymbol{A}^{+}\boldsymbol{A} = \boldsymbol{V}\boldsymbol{V}^{\mathrm{T}} = \boldsymbol{I}_n$.

(b) We only need to show $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{A}^{+}\boldsymbol{b} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$.

Since $\operatorname{rank}(\boldsymbol{A}) = n$, the columns of $\boldsymbol{A}$ are ind. Hence $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is invertible.

- Firstly, we show $(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}$ is the right inverse of $\boldsymbol{A}$:

$$
\begin{aligned}
\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}} &= \boldsymbol{U}\Sigma\boldsymbol{V}^{\mathrm{T}}(\boldsymbol{V}\Sigma\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\Sigma\boldsymbol{V}^{\mathrm{T}})^{-1}\boldsymbol{V}\Sigma\boldsymbol{U}^{\mathrm{T}}\\
&= \boldsymbol{U}\Sigma\boldsymbol{V}^{\mathrm{T}}(\boldsymbol{V}\Sigma^{2}\boldsymbol{V}^{\mathrm{T}})^{-1}\boldsymbol{V}\Sigma\boldsymbol{U}^{\mathrm{T}} = \boldsymbol{U}\Sigma\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}\Sigma^{-2}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}\Sigma\boldsymbol{U}^{\mathrm{T}}\\
&= \boldsymbol{U}\Sigma\Sigma^{-2}\Sigma\boldsymbol{U}^{\mathrm{T}}\\
&= \boldsymbol{I}
\end{aligned}
$$

- Since we also obtain $\boldsymbol{A}^{+}\boldsymbol{A} = \boldsymbol{I}$, we derive

$$
\boldsymbol{A}^{+} = \boldsymbol{A}^{+}\boldsymbol{I} = \boldsymbol{A}^{+}\boldsymbol{A}\left[(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\right] = \boldsymbol{I}\left[(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\right] = \left[(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\right]
$$

Thus we have $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{A}^{+} = \boldsymbol{A}^{\mathrm{T}} \implies \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{A}^{+}\boldsymbol{b} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$.

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 $\boldsymbol{p}_1, \boldsymbol{p}_2, \dots, \boldsymbol{p}_n$:

$$\boldsymbol{y} = \alpha_1\boldsymbol{p}_1 + \alpha_2\boldsymbol{p}_2 + \cdots + \alpha_n\boldsymbol{p}_n.$$

  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:

$$\begin{aligned}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y} &= \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}(\alpha_1\boldsymbol{p}_1 + \alpha_2\boldsymbol{p}_2 + \cdots + \alpha_n\boldsymbol{p}_n) \\ &= \alpha_1(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{p}_1) + \alpha_2(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{p}_2) + \cdots + \alpha_n(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{p}_n) \\ &= \alpha_1\lambda_1\boldsymbol{p}_1 + \alpha_2\lambda_2\boldsymbol{p}_2 + \cdots + \alpha_n\lambda_n\boldsymbol{p}_n\end{aligned}$$

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 $\boldsymbol{A}^2$.
$\Sigma = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$ is a $n \times n$ *diagonal* matrix, and $(\lambda_1, \dots, \lambda_n)$ are eigenvalues of $\boldsymbol{A}^2$.
And then we define $\sqrt{\Sigma} := \mathrm{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n})$. 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)

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

(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 $\operatorname{rank}(\boldsymbol{A}) = 2$.

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 $\operatorname{rank}(\boldsymbol{A}) = 3$.

In conclusion, $\operatorname{rank}(\boldsymbol{A}) = \begin{cases} 3, & c \neq 2; \\ 2, & c = 2. \end{cases}$

(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 $a_{st}(1 \leq s,t \leq n)$ 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 $1 \le i,j \le n$.

In order to let $\boldsymbol{E} = \boldsymbol{F}$, we must let $e_{ij} = f_{ij}$ for $\forall 1 \le i,j \le n$.

$$\Longrightarrow a_{ij}d_{jj} = d_{ii}a_{ij} \Longrightarrow a_{ij}(d_{jj} - d_{ii}) = 0.$$

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:
$$e_{pq}^{ij} = \sum_{t=1}^{n} a_{pt}b_{tq} \qquad f_{pq}^{ij} = \sum_{t=1}^{n} b_{pt}a_{tq}$$
where $1 \le p,q \le n$.
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 $e_{pq}^{ij} = f_{pq}^{ij}$.

- For $q \neq i$, we have $e_{iq}^{ii} = \sum_{t=1}^{n} a_{it}b_{tq} = 0$ since $b_{tq} = 0$ for $\forall t = 1,2,\dots,n$.
Also, $f_{iq}^{ii} = \sum_{t=1}^{n} b_{it}a_{tq} = a_{iq}$.
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, $N(\boldsymbol{A}) = \operatorname{span}\{\boldsymbol{u}\}$. Hence $\dim(N(\boldsymbol{A})) = 1$.

Hence $\operatorname{rank}(\boldsymbol{A}) = n - \dim(N(\boldsymbol{A})) = n - 1$.

(b) We find that

$$\boldsymbol{P}^2 = \boldsymbol{P}$$
$$\boldsymbol{P}^5 = \boldsymbol{P}.$$

Hence $\operatorname{rank}(\boldsymbol{P}^2) = \operatorname{rank}(\boldsymbol{P}) = n - 1; \operatorname{rank}(\boldsymbol{P}^5) = \operatorname{rank}(\boldsymbol{P}) = n - 1$.

(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 $\operatorname{rank}(\boldsymbol{I} - \boldsymbol{x}\boldsymbol{y}^{\mathrm{T}}) = 0$.

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

$$\boldsymbol{B}\boldsymbol{x} = \boldsymbol{0} \quad \text{where } \boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}.$$

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 $\boldsymbol{Ax} = \boldsymbol{b}$, which means $\boldsymbol{Ax} = \boldsymbol{b}$ has at least one solution for $\forall \boldsymbol{b} \in \mathbb{R}^m$.

- *Necessity.* If $\boldsymbol{Ax} = \boldsymbol{b}$ 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 $\boldsymbol{Ax}_i = \boldsymbol{e}_i$.

  Thus we construct $\boldsymbol{C} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_m \end{bmatrix}$. $\boldsymbol{C}$ is an $n \times m$ matrix and

$$\begin{aligned} \boldsymbol{AC} &= \boldsymbol{A} \left[\boldsymbol{x}_1 \mid \boldsymbol{x}_2 \mid \cdots \mid \boldsymbol{x}_m \right] \\ &= \left[\boldsymbol{Ax}_1 \mid \boldsymbol{Ax}_2 \mid \cdots \mid \boldsymbol{Ax}_m \right] \\ &= \left[\boldsymbol{e}_1 \mid \boldsymbol{e}_2 \mid \cdots \mid \boldsymbol{e}_m \right] = \boldsymbol{I}. \end{aligned}$$

  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 $\boldsymbol{Ax} = \boldsymbol{b}$ always has solution for $\forall \boldsymbol{b}$. We convert it into *augmented matrix form:*

$$\left[\boldsymbol{A} \mid \boldsymbol{b} \right] \xRightarrow{\text{Row transform}} \left[\text{rref}(\boldsymbol{A}) \mid \boldsymbol{b}^* \right]$$

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 $\boldsymbol{Ax}_1 = \boldsymbol{e}_1$.

Hence we construct $\boldsymbol{C} = \begin{bmatrix} \boldsymbol{x}_1 \end{bmatrix}$. We find that $\boldsymbol{AC} = \begin{pmatrix} 1 & 2 & 7\pi \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} =$

$1 = \boldsymbol{I}$. 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*:

$$(i) \quad \boldsymbol{A}^2 - \boldsymbol{B}^2$$
$$(iii) \quad \boldsymbol{ABA}$$

(b) There are *infinitely* many solutions.

Reason:

- Since $\boldsymbol{A}$ is 5×8matrix, $\text{rank}(\boldsymbol{A}) + \dim(N(\boldsymbol{A})) = 8 \implies \dim(N(\boldsymbol{A})) = 3$.

  Hence this system $\boldsymbol{Ax} = \boldsymbol{b}$ has **special solutions**.

- Moverover, since $\text{rank}(\boldsymbol{A}) = 5$, we have 5 nonzero pivots, which means rref($\boldsymbol{A}$) has no zero rows.

  Hence this system $\boldsymbol{Ax} = \boldsymbol{b}$ always has **particular solution.**

In conclusion, there are *infinitely* many solutions.

(c) False.

Reason: For example, if we have

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

then $\boldsymbol{A} + \boldsymbol{B} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, 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?

$\boldsymbol{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \boldsymbol{B} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ are both in this set since $\text{rank}(\boldsymbol{A}) + \text{rank}(\boldsymbol{B}) = 1$.

However, $\boldsymbol{A} + \boldsymbol{B} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ doesn't belong to this set since $\text{rank}(\boldsymbol{A} + \boldsymbol{B}) = 2$.

(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 $\{\mathbf{0}\}$.

Hence the column space of $\boldsymbol{A}$ is $\{0\}$, which means all columns of $\boldsymbol{A}$ are $\mathbf{0}$. Hence all elements of $\boldsymbol{A}$ are 0. Thus $\boldsymbol{A} = \mathbf{0}$.

(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

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y} = \langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{x}, \boldsymbol{A}^{\mathrm{T}}\boldsymbol{u} \rangle = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{u}^{\mathrm{T}}\mathbf{0} = 0.$$

(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 $\boldsymbol{U} := \begin{bmatrix} \boldsymbol{U}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{U}_2 \end{bmatrix}$, we find that

$$\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U} = \begin{bmatrix} \boldsymbol{U}_1^{\mathrm{T}}\boldsymbol{U}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{U}_2^{\mathrm{T}}\boldsymbol{U}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I} \end{bmatrix} = \boldsymbol{I}.$$

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 $\begin{bmatrix} \boldsymbol{A} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{B} \end{bmatrix}$ could be decomposed as

$$\begin{bmatrix} \boldsymbol{A} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{B} \end{bmatrix} = \begin{bmatrix} \boldsymbol{U}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{U}_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & \\ & \Sigma_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{U}_1^{\mathrm{T}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{U}_2^{\mathrm{T}} \end{bmatrix} = \boldsymbol{U} \begin{bmatrix} \Sigma_1 & \\ & \Sigma_2 \end{bmatrix} \boldsymbol{U}^{\mathrm{T}}$$

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

$$\boldsymbol{A} = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_n).$$

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

$$\begin{aligned}
\boldsymbol{x}^{\mathrm{T}} \boldsymbol{M} \boldsymbol{x} &= \sum_{i,j=1}^{n} M_{ij} x_i x_j = \sum_{i=1}^{n} M_{ii} x_i^2 + \sum_{j \neq i} M_{ij} x_i x_j \\
&= 2 \sum_{1 \leq i < j \leq n} M_{ij} x_i x_j + \sum_{i=1}^{n} M_{ii} x_i^2 \\
&= \sum_{1 \leq i < j \leq n} (2M_{ij} x_i x_j + |M_{ij}| x_i^2 + |M_{ij}| x_j^2) - \sum_{1 \leq i < j \leq n} (|M_{ij}| x_i^2 + |M_{ij}| x_j^2) + \sum_{i=1}^{n} M_{ii} x_i^2 \\
&= \sum_{1 \leq i < j \leq n} (2M_{ij} x_i x_j + |M_{ij}| x_i^2 + |M_{ij}| x_j^2) + \sum_{i=1}^{n} (M_{ii} x_i^2 - \sum_{j \neq i} |M_{ij}|) x_i^2
\end{aligned}$$

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

$$\mathrm{Proj}_{\boldsymbol{q}_2}(\boldsymbol{z}) = \frac{\langle \boldsymbol{q}_2, \boldsymbol{z} \rangle}{\langle \boldsymbol{q}_2, \boldsymbol{q}_2 \rangle} \boldsymbol{q}_2 = \begin{bmatrix} \frac{2}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix}.$$

Hence the projection of $\boldsymbol{z}$ onto span$\{\boldsymbol{x}, \boldsymbol{z}\}$ is given by:

$$\mathrm{Proj}_{\mathrm{span}\{\boldsymbol{q}_1, \boldsymbol{q}_2\}}(\boldsymbol{z}) = \mathrm{Proj}_{\boldsymbol{q}_1}(\boldsymbol{z}) + \mathrm{Proj}_{\boldsymbol{q}_2}(\boldsymbol{z}) = \begin{bmatrix} -\frac{1}{6} \\ -\frac{1}{6} \\ \frac{1}{3} \end{bmatrix} + \begin{bmatrix} \frac{2}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 1 \end{bmatrix}$$

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:

$$\begin{aligned}
\operatorname{Proj}_{\operatorname{span}\{\boldsymbol{x},\boldsymbol{y}\}}(\boldsymbol{b}) &= \operatorname{Proj}_{\operatorname{span}\{\boldsymbol{q}_1,\boldsymbol{q}_2\}}(\boldsymbol{b}) \\
&= \operatorname{Proj}_{\boldsymbol{q}_1}(\boldsymbol{b}) + \operatorname{Proj}_{\boldsymbol{q}_2}(\boldsymbol{b}) \\
&= \frac{\langle \boldsymbol{q}_1, \boldsymbol{b}\rangle}{\langle \boldsymbol{q}_1, \boldsymbol{q}_1\rangle}\boldsymbol{q}_1 + \frac{\langle \boldsymbol{q}_2, \boldsymbol{b}\rangle}{\langle \boldsymbol{q}_2, \boldsymbol{q}_2\rangle}\boldsymbol{q}_2 \\
&= -\frac{6}{11}\begin{bmatrix} 1 \\ 1 \\ -2 \\ 0 \end{bmatrix} + \frac{37}{66}\begin{bmatrix} 7 \\ 1 \\ 4 \\ 0 \end{bmatrix} \\
&= \frac{1}{11}\begin{bmatrix} 23 \\ -14 \\ 65 \\ 0 \end{bmatrix}.
\end{aligned}$$

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 $\Sigma = \operatorname{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}) = \operatorname{diag}(4, 1)$.

  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:

$$\begin{aligned} \boldsymbol{B}\boldsymbol{A}\boldsymbol{v}_i &= \lambda_i(\beta_1\boldsymbol{v}_1 + \beta_2\boldsymbol{v}_2 + \cdots + \beta_n\boldsymbol{v}_n) \\ &= \beta_1\lambda_i\boldsymbol{v}_1 + \beta_2\lambda_i\boldsymbol{v}_2 + \cdots + \beta_n\lambda_i\boldsymbol{v}_n \end{aligned} \tag{10.12}$$

Since $\boldsymbol{A}\boldsymbol{B} = \boldsymbol{B}\boldsymbol{A}$, we derive $\boldsymbol{A}\boldsymbol{B}\boldsymbol{v}_i = \boldsymbol{B}\boldsymbol{A}\boldsymbol{v}_i$. Combining equation (10.11) and (10.12) we obtain:

$$\mathbf{0} = \boldsymbol{A}\boldsymbol{B}\boldsymbol{v}_i - \boldsymbol{B}\boldsymbol{A}\boldsymbol{v}_i = \beta_1(\lambda_1 - \lambda_i)\boldsymbol{v}_1 + \beta_2(\lambda_2 - \lambda_i)\boldsymbol{v}_2 + \cdots + \beta_n(\lambda_n - \lambda_i)\boldsymbol{v}_n$$

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 $\boldsymbol{B}\boldsymbol{v}_i = \beta_i\boldsymbol{v}_i$, 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

$$\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{Q} := \boldsymbol{D} = \operatorname{diag}(\lambda_1\boldsymbol{I}_{r_1}, \lambda_2\boldsymbol{I}_{r_2}, \dots, \lambda_h\boldsymbol{I}_{r_h}) = \begin{pmatrix} \lambda_1\boldsymbol{I}_{r_1} & & & \\ & \lambda_2\boldsymbol{I}_{r_2} & & \\ & & \ddots & \\ & & & \lambda_h\boldsymbol{I}_{r_h} \end{pmatrix}.$$

Also, we can obtain the product $\boldsymbol{Q}^{-1}\boldsymbol{B}\boldsymbol{Q}$ 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 $\boldsymbol{A}\boldsymbol{B} = \boldsymbol{B}\boldsymbol{A}$, thus we have

$$\begin{aligned} \boldsymbol{D}\boldsymbol{C} &= (\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{Q})(\boldsymbol{Q}^{-1}\boldsymbol{B}\boldsymbol{Q}) \\ &= \boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{B}\boldsymbol{Q} = \boldsymbol{Q}^{-1}\boldsymbol{B}\boldsymbol{A}\boldsymbol{Q} \\ &= (\boldsymbol{Q}^{-1}\boldsymbol{B}\boldsymbol{Q})(\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{Q}) \\ &= \boldsymbol{C}\boldsymbol{D}. \end{aligned}$$

  Notice that the (i,j)th submatrix of $\boldsymbol{D}\boldsymbol{C}$ is equal to the (i,j)th submatrix of $\boldsymbol{C}\boldsymbol{D}$, 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

$$\boldsymbol{C} = \operatorname{diag}(\boldsymbol{C}_{11}, \boldsymbol{C}_{22}, \dots, \boldsymbol{C}_{hh}) = \begin{pmatrix} \boldsymbol{C}_{11} & & & \\ & \boldsymbol{C}_{22} & & \\ & & \ddots & \\ & & & \boldsymbol{C}_{hh} \end{pmatrix}.$$

  is *block diagonal*.

- Then we show $\boldsymbol{C}$ is diagonalizable:

  Since $\boldsymbol{B}$ is diagonalizable, there exists $\boldsymbol{M}$ satisfying

$$\boldsymbol{M}^{-1}\boldsymbol{B}\boldsymbol{M} = \boldsymbol{N} \implies \boldsymbol{B} = \boldsymbol{M}\boldsymbol{N}\boldsymbol{M}^{-1}$$

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:

$$\begin{aligned}
\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{T}_*^{-1}\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{Q}\boldsymbol{T}_* = \boldsymbol{T}_*^{-1}\boldsymbol{D}\boldsymbol{T}_* \\
&= \operatorname{diag}(\boldsymbol{T}_{11}^{-1}, \boldsymbol{T}_{22}^{-1}, \dots, \boldsymbol{T}_{hh}^{-1}) \operatorname{diag}(\lambda_1 \boldsymbol{I}_{r_1}, \lambda_2 \boldsymbol{I}_{r_2}, \dots, \lambda_h \boldsymbol{I}_{r_h}) \operatorname{diag}(\boldsymbol{T}_{11}, \boldsymbol{T}_{22}, \dots, \boldsymbol{T}_{hh}) \\
&= \operatorname{diag}(\lambda_1 \boldsymbol{T}_{11}^{-1}\boldsymbol{T}_{11}, \lambda_2 \boldsymbol{T}_{22}^{-1}\boldsymbol{T}_{22}, \dots, \lambda_h T_{hh}^{-1} T_{hh}) \\
&= \operatorname{diag}(\lambda_1 \boldsymbol{I}_{r_1}, \lambda_2 \boldsymbol{I}_{r_2}, \dots, \lambda_h \boldsymbol{I}_{r_h}) = \boldsymbol{D}
\end{aligned}$$

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:

$$\begin{bmatrix} \boldsymbol{u} \circ \boldsymbol{v} \end{bmatrix} = \begin{bmatrix} u_1 v_1 & u_2 v_2 & \dots & u_n v_n \end{bmatrix}^{\mathrm{T}}.$$

   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, $\boldsymbol{A} \circ \boldsymbol{B} \succeq 0$.

Appendix A

This is Appendix Title

A.1. This is First Level Heading

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

A.1.1. This is Second Level Heading

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

A.1.1.1. This is Third Level Heading

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

This is Fourth Level Heading. Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.

This is Fifth Level Heading. Suspendisse vel felis. Ut lorem lorem, interdum eu, tincidunt sit amet, laoreet vitae, arcu. Aenean faucibus pede eu ante. Praesent enim elit, rutrum at, molestie non, nonummy vel, nisl. Ut lectus eros, malesuada sit amet, fermentum eu, sodales cursus, magna. Donec eu purus. Quisque vehicula, urna sed ultricies auctor, pede lorem egestas dui, et convallis elit erat sed nulla. Donec luctus. Curabitur et nunc. Aliquam dolor odio, commodo pretium, ultricies non, pharetra in, velit. Integer arcu est, nonummy in, fermentum faucibus, egestas vel, odio.

Theorem A.1 asfasf

Theorem A.2 asfasf