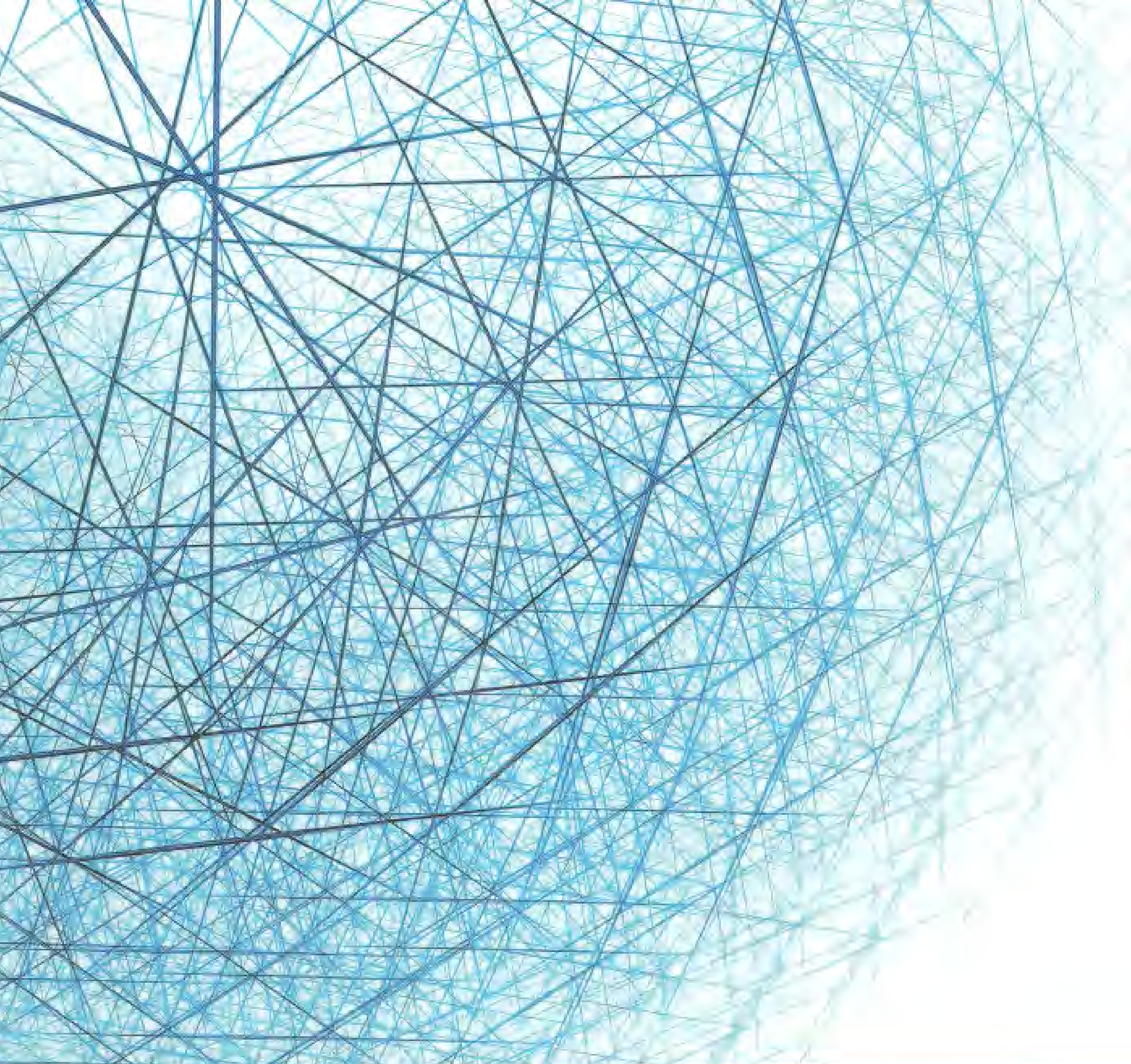


Linear Algebra

Concepts and Techniques on Euclidean Spaces

ACE ASIAN CUSTOMIZED EDITION

Ma Siu Lun • Ng Kah Loon • Victor Tan

LINEAR ALGEBRA

Concepts and Techniques on Euclidean Spaces

Ma Siu Lun

Ng Kah Loon

Victor Tan

McGraw Hill

Singapore • Boston • Burr Ridge, IL • Dubuque, IA • Madison, WI • New York
San Francisco • St. Louis • Bangkok • Kuala Lumpur • Lisbon • London • Madrid
Mexico City • Milan • Montreal • New Delhi • Seoul • Sydney • Taipei • Toronto

The McGraw-Hill Companies

Linear Algebra
Concepts and Techniques on Euclidean Spaces

Mc Graw Hill Higher Education

ACE ASIAN CUSTOMIZED EDITION The Asian Customized Edition is a speedy publishing service for lecturers and has not undergone the full McGraw-Hill Education (Asia) editorial process.

10 9 8 7 6 5 4 3 2 1
CTP MPM
20 15 14 13 12

When ordering this title, use **ISBN 978-1-259-01151-1** or **MHID 1-259-01151-8.**

Printed in Singapore

Preface

This publication is the successor of the third edition of the book "Linear Algebra I" published by the authors with their previous publisher. It is used as the course lecture notes for the undergraduate module MA1101R, Linear Algebra I, offered by the Department of Mathematics at the National University of Singapore. This module is the first course on linear algebra and it serves as an introduction to the basic concepts of linear algebra that are routinely applied in diverse fields such as science, engineering, statistics, economics and computing. Mindful that majority of the students taking this module are new to the subject, we have chosen to introduce the concepts of linear algebra in the context of Euclidean spaces rather than to jump straight into abstract vector spaces, which will be covered in the second course. The set up in Euclidean spaces also facilitates the connections between the algebraic and geometric viewpoints of linear algebra.

Formal proofs of most of the basic theorems in linear algebra have been included to enhance a proper understanding of the fundamental ideas and techniques. Several applications of linear algebra in some of the fields mentioned above are also highlighted. At the end of every chapter is a good collection of problems, all of which are culled from tutorial problems, test and examination questions from the same module taught by the authors in the past. These problems range from the straightforward computational ones to some highly challenging questions. In order to achieve a deeper understanding of the topic, students are advised to work through these problems.

Significant updates and revisions have been made in this new edition, including the discussion, examples and exercises. Many of the changes are done in response to feedback received from students, and teaching after-thoughts that the authors have accumulated over the past years. We believe that the revisions made to the presentation of materials will further enhance learners' understanding of the critical concepts in this subject.

Finally, the authors would like to thank their colleagues from the Mathematics Department in NUS who have contributed to the very first version of the lecture notes in 1998, especially Chan Onn, Tan Hwee Huat, Roger Tan Choon Ee and Tang Wai Shing. More recently, Toh Pee Choon (now with the National Institute of Education) has also given many useful comments on areas of improvement from the previous edition.

Contents

Chapter 1

Linear Systems and Gaussian Elimination

Section 1.1 Linear Systems and Their Solutions

Discussion 1.1.1 A line in the xy-plane can be represented algebraically by an equation of the form

$$ax + by = c$$

where a and b are not both zero. An equation of this kind is known as a linear equation in the variables of x and y. In general, we have the following definition.

Definition 1.1.2 A *linear equation* in n variables $x_1, x_2, \ldots, x_n$ has the form

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

where $a_1, a_2, \ldots, a_n$ and b are real constants. The variables in a linear equation are also called the *unknowns.*

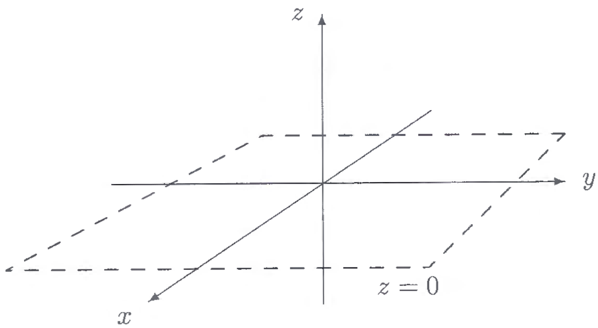
We do not need to assume that $a_1, a_2, \ldots, a_n$ are not all zero. If all $a_1, a_2, \ldots, a_n$ and b are zero, the equation is called a *zero equation.* A linear equation is called a *nonzero equation* if it is not a zero equation.

Example 1.1.3

1. The equations $x + 3y = 7$, $x_1 + 2x_2 + 2x_3 + x_4 = x_5$, $y = x - \frac{1}{2}z + 4.5$ and $x_1 + x_2 + \cdots + x_n = 1$ are linear.

2. The equations $xy = 2$, $\sin(\theta) + \cos(\phi) = 0.2$, $x_1^2 + x_2^2 + \cdots + x_n^2 = 1$ and $x = e^y$ are not linear.

3. The linear equation $ax + by + cz = d$, where a, b, c, d are constants and a, b, c are not all zero, represents a plane in the three dimensional space. For example, $z = 0$ (i.e. $0x + 0y + z = 0$) is the xy-plane contained inside the xyz-space.



Definition 1.1.4 Given n real numbers $s_1, s_2, \ldots, s_n$, we say that $x_1 = s_1, x_2 = s_2, \ldots, x_n = s_n$ is a *solution* to a linear equation $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$ if the equation is satisfied when we substitute the values into the equation accordingly. The set of all solutions to the equation is called the *solution set* of the equation and an expression that gives us all these solutions is called a *general solution* for the equation.

Example 1.1.5

1. Consider the linear equation $4x - 2y = 1$. It has a general solution

$$\begin{cases} x = t \\ y = 2t - \frac{1}{2} \end{cases} \quad \text{where } t \text{ is an arbitrary parameter.}$$

The equation also has another general solution

$$\begin{cases} x = \frac{1}{2}s + \frac{1}{4} \\ y = s \end{cases} \quad \text{where } s \text{ is an arbitrary parameter.}$$

Though the two general solutions above look different, they gives us the same set of solutions including

$$\begin{cases} x = 1 \\ y = 1.5, \end{cases} \quad \begin{cases} x = 1.5 \\ y = 2.5, \end{cases} \quad \begin{cases} x = -1 \\ y = -2.5, \end{cases}$$

and infinitely many other solutions.

2. Consider the equation $x_1 - 4x_2 + 7x_3 = 5$. It has a general solution

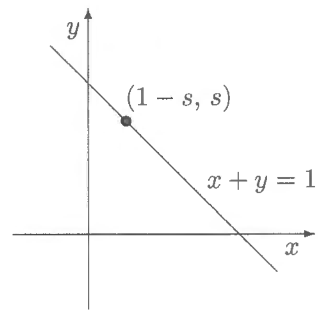
$$\begin{cases} x_1 = 5 + 4s - 7t \\ x_2 = s \\ x_3 = t \end{cases} \qquad \text{where } s,\ t \text{ are arbitrary parameters.}$$

3. **(Geometrical Interpretation)**

   (a) In the xy-plane, solutions to the equation $x + y = 1$ are points

   $$(x, y) = (1 - s,\ s)$$

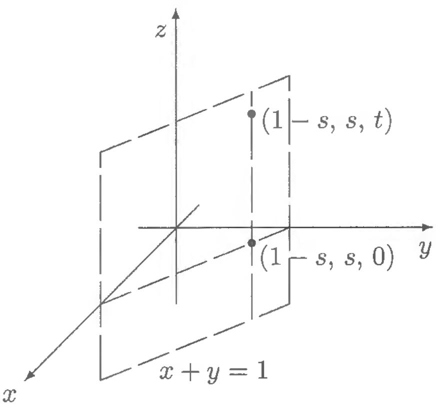
   where s is any real number. These points form a line as shown in the diagram on the right.



   (b) In the xyz-space, solutions to the equation $x + y = 1$ (i.e. $x + y + 0z = 1$) are points

   $$(x, y, z) = (1 - s,\ s,\ t)$$

   where s and t are any real numbers. These points form a plane as shown in the diagram on the right.



4. Consider the zero equation $0x_1 + 0x_2 + \cdots + 0x_n = 0$. Any values of $x_1,\ x_2,\ \ldots,\ x_n$ give us a solution. Thus the general solution is $x_1 = t_1,\ x_2 = t_2,\ \ldots,\ x_n = t_n$ where $t_1,\ t_2,\ \ldots,\ t_n$ are arbitrary parameters.

5. For an equation $0x_1 + 0x_2 + \cdots + 0x_n = b$, where b is nonzero, any values of $x_1,\ x_2,\ \ldots,\ x_n$ does not satisfy the equation and hence the equation has no solution.

Definition 1.1.6 A finite set of linear equations in the variables $x_1,\ x_2,\ \ldots,\ x_n$ is called a *system of linear equations* (or a *linear system*):

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \qquad \vdots \qquad\qquad\qquad\qquad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

where $a_{11}, a_{12}, \ldots, a_{mn}$ and $b_1, b_2, \ldots, b_m$ are real constants. If all $a_{11}, a_{12}, \ldots, a_{mn}$ and $b_1, b_2, \ldots, b_m$ are zero, the system is called a *zero system*. A linear system is called a *nonzero system* if it is not a zero system.

Given n real numbers $s_1, s_2, \ldots, s_n$, we say that $x_1 = s_1, x_2 = s_2, \ldots, x_n = s_n$ is a *solution* to the system if $x_1 = s_1, x_2 = s_2, \ldots, x_n = s_n$ is a solution to every equation in the system. The set of all solutions to the system is called the *solution set* of the system and an expression that gives us all these solutions is called a *general solution* for the system.

Example 1.1.7 Consider the system of linear equations

$$\begin{cases} 4x_1 - x_2 + 3x_3 = -1 \\ 3x_1 + x_2 + 9x_3 = -4. \end{cases}$$

$x_1 = 1$, $x_2 = 2$, $x_3 = -1$ is a solution to the system and $x_1 = 1$, $x_2 = 8$, $x_3 = 1$ is not a solution to the system.

Remark 1.1.8 Not all systems of linear equations have solutions. For example, the following system has no solution as it is impossible to have a solution that satisfies both equations simultaneously.

$$\begin{cases} x + y = 4 \\ 2x + 2y = 6. \end{cases}$$

Definition 1.1.9 A system of linear equations that has no solution is said to be *inconsistent*. A system that has at least one solution is called *consistent*.

Remark 1.1.10 Every system of linear equations has either *no solution, only one solution*, or *infinitely many solutions*. (See Question 2.22.)

Discussion 1.1.11

1. In the xy-plane, the two equations in the system

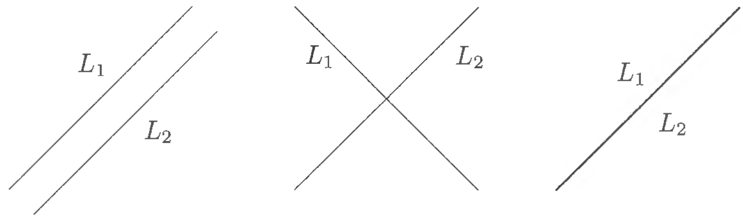
$$\begin{cases} a_1x + b_1y = c_1 & (L_1) \\ a_2x + b_2y = c_2, & (L_2) \end{cases}$$

   where a_1, b_1 are not both zero and a_2, b_2 are not both zero, represent two straight lines. A solution to the system is a point of intersection of the two lines.

   (a) The system has no solution if and only if L_1 and L_2 are different but parallel lines.

   (b) The system has only one solution if and only if L_1 and L_2 are not parallel lines.

   (c) The system has infinitely many solutions if and only if L_1 and L_2 are the same line.



No solution One solution Infinitely many solutions

Note that the two lines L_1 and L_2 are parallel if and only if $a_1 = ka_2$ and $b_1 = kb_2$ for a nonzero real number k.

2. In the xyz-space, the two equations in the system

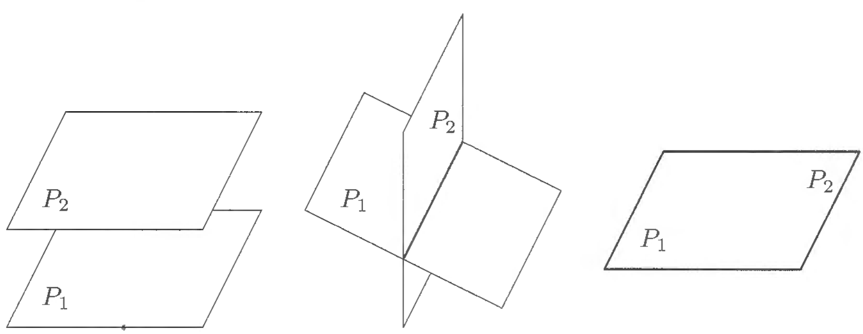
$$\begin{cases} a_1x + b_1y + c_1z = d_1 & (P_1) \\ a_2x + b_2y + c_2z = d_2, & (P_2) \end{cases}$$

where a_1, b_1, c_1 are not all zero and a_2, b_2, c_2 are not all zero, represents two planes. A solution to the system is a point of intersection of the two planes.

(a) The system has no solution if and only if P_1 and P_2 are different but parallel planes.

(b) The system cannot have only one solution.

(c) The system has infinitely many solutions if and only if either P_1 and P_2 intersect at a line or P_1 and P_2 are the same plane.



No solution Infinitely many solutions Infinitely many solutions

Note that the two planes P_1 and P_2 are parallel if and only if $a_1 = ka_2$, $b_1 = kb_2$ and $c_1 = kc_2$ for a nonzero real number k.

Section 1.2 Elementary Row Operations

Definition 1.2.1 A system of linear equations

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \quad\vdots \qquad\qquad\qquad\qquad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

can be represented by a rectangular array of numbers

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

This array is called the *augmented matrix* of the system.

Example 1.2.2 The array

$$\left(\begin{array}{ccc|c} 1 & 1 & 2 & 9 \\ 2 & 4 & -3 & 1 \\ 3 & 6 & -5 & 0 \end{array}\right)$$

is the augmented matrix of the system of linear equations

$$\begin{cases} x_1 + x_2 + 2x_3 = 9 \\ 2x_1 + 4x_2 - 3x_3 = 1 \\ 3x_1 + 6x_2 - 5x_3 = 0\,. \end{cases}$$

Discussion 1.2.3 The following are the basic techniques for solving a system of linear equations:

1. Multiply an equation by a nonzero constant.
2. Interchange two equations.
3. Add a multiple of one equation to another equation.

In terms of the augmented matrix, these correspond to:

1. Multiply a row by a nonzero constant.
2. Interchange two rows.
3. Add a multiple of one row to another row.

Definition 1.2.4 The three operations of the augmented matrix described above are known as *elementary row operations.*

Example 1.2.5 The following is an example of solving a system of linear equations by equation additions and subtractions. First, we start with the following system and its augmented matrix.

$$\begin{cases} x+\ y+3z=0 & (1)\\ 2x-2y+2z=4 & (2)\\ 3x+9y\qquad =3 & (3)\end{cases} \qquad \left(\begin{array}{ccc|c} 1 & 1 & 3 & 0\\ 2 & -2 & 2 & 4\\ 3 & 9 & 0 & 3\end{array}\right)$$

Add -2 times of Equation (1) to Equation (2) to obtain Equation (4). This is equivalent to adding -2 times of the first row of the matrix to the second row.

$$\begin{cases} x+\ y+3z=0 & (1)\\ \quad -4y-4z=4 & (4)\\ 3x+9y\qquad =3 & (3)\end{cases} \qquad \left(\begin{array}{ccc|c} 1 & 1 & 3 & 0\\ 0 & -4 & -4 & 4\\ 3 & 9 & 0 & 3\end{array}\right)$$

Add -3 times of Equation (1) to Equation (3) to obtain Equation (5). This is equivalent to adding -3 times of the first row of the matrix to the third row.

$$\begin{cases} x+\ y+3z=0 & (1)\\ \quad -4y-4z=4 & (4)\\ \quad\ \ 6y-9z=3 & (5)\end{cases} \qquad \left(\begin{array}{ccc|c} 1 & 1 & 3 & 0\\ 0 & -4 & -4 & 4\\ 0 & 6 & -9 & 3\end{array}\right)$$

Add $\frac{6}{4}$ times of Equation (4) to Equation (5) to obtain Equation (6). This is equivalent to adding $\frac{6}{4}$ times of the second row of the matrix to the third row.

$$\begin{cases} x+\quad y+\ 3z=0 & (1)\\ \quad -4y-\ 4z=4 & (4)\\ \qquad\quad -15z=9 & (6)\end{cases} \qquad \left(\begin{array}{ccc|c} 1 & 1 & 3 & 0\\ 0 & -4 & -4 & 4\\ 0 & 0 & -15 & 9\end{array}\right)$$

By Equation (6), $z=-\frac{3}{5}$. Substituting $z=-\frac{3}{5}$ into Equation (4),

$$-4y-4(-\tfrac{3}{5})=4 \quad\Leftrightarrow\quad y=-\tfrac{2}{5}.$$

Substituting $y=-\frac{2}{5}$ and $z=-\frac{3}{5}$ into Equation (1),

$$x+(-\tfrac{2}{5})+3(-\tfrac{3}{5})=0 \quad\Leftrightarrow\quad x=\tfrac{11}{5}.$$

Thus we find that $x=\frac{11}{5}$, $y=-\frac{2}{5}$, $z=-\frac{3}{5}$ is the only solution to the system.

Definition 1.2.6 Two augmented matrices are said to be *row equivalent* if one can be obtained from the other by a series of elementary row operations.

Theorem 1.2.7 If augmented matrices of two systems of linear equations are row equivalent, then the two systems have the same set of solutions.

Example 1.2.8 All augmented matrices in Example 1.2.5 are row equivalent. Thus all systems of linear equations in Example 1.2.5 have the same solution.

Remark 1.2.9 To see why Theorem 1.2.7 is true, we only need to check that every elementary row operation applied to an augmented matrix will not change the solution set of the corresponding linear system. Since it is easier to work with linear systems using the matrix representation discussed in Chapter 2, we shall postpone the proof of the theorem to Section 2.4, see Remark 2.4.6.

Section 1.3 Row-Echelon Forms

Definition 1.3.1 An augmented matrix is said to be in *row-echelon form* if it has the following properties:

1. If there are any rows that consist entirely of zeros, then they are grouped together at the bottom of the matrix.

2. In any two successive rows that do not consist entirely of zeros, the first nonzero number in the lower row occurs farther to the right than the first nonzero number in the higher row. (The first nonzero number in a row is called the *leading entry* of the row.)

In a row-echelon form, the leading entries of nonzero rows are also called *pivot points*. A column of a row-echelon form is called a *pivot column* if it contains a pivot point; otherwise, it is called a *non-pivot column*.

An augmented matrix is said to be in *reduced row-echelon form* if it is in row-echelon form and has the following additional properties:

3. The leading entry of every nonzero row is 1.

4. In each pivot column, except the pivot point, all other entries are zero .

Remark 1.3.2 In some textbooks, row-echelon forms are required to have the additional property 3 in Definition 1.3.1.

Example 1.3.3

1. The following augmented matrices are in reduced row-echelon form. The underlined numbers are pivot points (the leading entries of nonzero rows).

$$\left(\begin{array}{cc|c} \underline{1} & 2 & 3 \end{array}\right) \quad \left(\begin{array}{cc|c} 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right) \quad \left(\begin{array}{ccc|c} \underline{1} & 0 & 0 & 1 \\ 0 & \underline{1} & 0 & 2 \\ 0 & 0 & \underline{1} & 3 \end{array}\right) \quad \left(\begin{array}{cc|c} \underline{1} & 0 & 0 \\ 0 & \underline{1} & 0 \\ 0 & 0 & \underline{1} \end{array}\right) \quad \left(\begin{array}{cccc|c} 0 & \underline{1} & 2 & 0 & 1 \\ 0 & 0 & 0 & \underline{1} & 3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

2. The following augmented matrices are in row-echelon form but not in reduced row-echelon form. The underlined numbers are pivot points (the leading entries of nonzero rows).

$$\left(\begin{array}{cc|c} \underline{3} & 2 & 1 \end{array}\right) \quad \left(\begin{array}{cc|c} \underline{1} & -1 & 0 \\ 0 & \underline{1} & 0 \end{array}\right) \quad \left(\begin{array}{ccc|c} \underline{-1} & 2 & 3 & 4 \\ 0 & \underline{1} & 1 & 2 \\ 0 & 0 & \underline{2} & 3 \end{array}\right) \quad \left(\begin{array}{cc|c} \underline{2} & 1 & 0 \\ 0 & \underline{1} & 0 \\ 0 & 0 & \underline{1} \end{array}\right) \quad \left(\begin{array}{cccc|c} 0 & \underline{1} & 2 & 8 & 1 \\ 0 & 0 & 0 & \underline{4} & 3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

3. The following augmented matrices are neither in row-echelon form nor in reduced row-echelon form.

$$\left(\begin{array}{cc|c} 0 & 1 & 0 \\ 1 & 0 & 0 \end{array}\right) \quad \left(\begin{array}{ccc|c} 1 & 0 & 2 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 1 & 1 & 3 \end{array}\right) \quad \left(\begin{array}{cc|c} 0 & 0 & 1 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{array}\right) \quad \left(\begin{array}{cccc|c} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 & 1 \\ 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

Discussion 1.3.4 If the augmented matrix of a system of linear equations is in row-echelon form or reduced row-echelon form, we can get the solutions to the system easily.

Example 1.3.5

1. The array $\left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right)$ is the augmented matrix of a system of linear equations in 3 variables x_1, x_2, x_3:

$$\begin{cases} x_1 + 0x_2 + 0x_3 = 1 \\ 0x_1 + x_2 + 0x_3 = 2 \\ 0x_1 + 0x_2 + x_3 = 3 \end{cases} \quad \text{or} \quad \begin{cases} x_1 = 1 \\ x_2 = 2 \\ x_3 = 3 \,. \end{cases}$$

The system has only one solution: $x_1 = 1$, $x_2 = 2$, $x_3 = 3$.

2. The array $\left(\begin{array}{ccccc|c} 0 & 2 & 2 & 1 & -2 & 2 \\ 0 & 0 & 1 & 1 & 1 & 3 \\ 0 & 0 & 0 & 0 & 2 & 4 \end{array}\right)$ is the augmented matrix of a system of linear equations in 5 variables x_1, x_2, x_3, x_4, x_5:

$$\begin{cases} 0x_1 + 2x_2 + 2x_3 + x_4 - 2x_5 = 2 \\ 0x_1 + 0x_2 + x_3 + x_4 + x_5 = 3 \\ 0x_1 + 0x_2 + 0x_3 + 0x_4 + 2x_5 = 4 \end{cases} \quad \text{or} \quad \begin{cases} 2x_2 + 2x_3 + x_4 - 2x_5 = 2 \\ x_3 + x_4 + x_5 = 3 \\ 2x_5 = 4 \,. \end{cases}$$

The coefficients of x_1 are zero in all the three equations and this means that x_1 is arbitrary (see also Example 1.1.5.3(b)). By the third equation, $x_5 = 2$. Substituting $x_5 = 2$ into the second equation,

$$x_3 + x_4 + 2 = 3 \quad \Leftrightarrow \quad x_3 = 1 - x_4.$$

Substituting $x_3 = 1 - x_4$ and $x_5 = 2$ into the first equation,

$$2x_2 + 2(1 - x_4) + x_4 - 2 \cdot 2 = 2 \quad \Leftrightarrow \quad x_2 = 2 + \tfrac{1}{2}x_4.$$

Thus the linear system has a general solution

$$\begin{cases} x_1 = s \\ x_2 = 2 + \frac{1}{2}t \\ x_3 = 1 - t \\ x_4 = t \\ x_5 = 2, \end{cases} \quad \text{where } s,\ t \text{ are arbitrary parameters,}$$

and has infinitely many solutions. (The method we use here is called the *back-substitution.*)

3. The array $\left(\begin{array}{cccc|c} 1 & -1 & 0 & 3 & -2 \\ 0 & 0 & 1 & 2 & 5 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$ is the augmented matrix of a system of linear equations in 4 variables x_1, x_2, x_3, x_4:

$$\begin{cases} x_1 - x_2 + 0x_3 + 3x_4 = -2 \\ 0x_1 + 0x_2 + x_3 + 2x_4 = 5 \\ 0x_1 + 0x_2 + 0x_3 + 0x_4 = 0 \end{cases} \quad \text{or} \quad \begin{cases} x_1 - x_2 + 3x_4 = -2 \\ x_3 + 2x_4 = 5\,. \end{cases}$$

The linear system has a general solution

$$\begin{cases} x_1 = -2 + s - 3t \\ x_2 = s \\ x_3 = 5 - 2t \\ x_4 = t, \end{cases} \quad \text{where } s,\ t \text{ are arbitrary parameters,}$$

and has infinitely many solutions.

4. The array $\left(\begin{array}{ccc|c} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$ is the augmented matrix of a system of linear equations in 2 variables x_1, x_2, x_3:

$$\begin{cases} 0x_1 + 0x_2 + 0x_3 = 0 \\ 0x_1 + 0x_2 + 0x_3 = 0\,. \end{cases}$$

This is a zero system and the general solution is

$$\begin{cases} x_1 = r \\ x_2 = s \\ x_3 = t, \end{cases} \quad \text{where } r,\ s,\ t \text{ are arbitrary parameters,}$$

and has infinitely many solutions.

5. The array $\left(\begin{array}{cc|c} 3 & 1 & 4 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{array}\right)$ is the augmented matrix of a system of linear equations in 2 variables x_1, x_2:

$$\begin{cases} 3x_1 + x_2 = 4 \\ 0x_1 + 2x_2 = 1 \\ 0x_1 + 0x_2 = 1\,. \end{cases}$$

Since the last equation has no solution, the system is inconsistent.

Section 1.4 Gaussian Elimination

Definition 1.4.1 Let $\boldsymbol{A}$ and $\boldsymbol{R}$ be row-equivalent augmented matrices, i.e. $\boldsymbol{R}$ can be obtained from $\boldsymbol{A}$ by a series of elementary row operations. If $\boldsymbol{R}$ is in row-echelon form, $\boldsymbol{R}$ is called *a row-echelon form of* $\boldsymbol{A}$ and $\boldsymbol{A}$ is said to *have a row-echelon form* $\boldsymbol{R}$. Similarly, if $\boldsymbol{R}$ is in reduced row-echelon form, $\boldsymbol{R}$ is called *a reduced row-echelon form of* $\boldsymbol{A}$ and $\boldsymbol{A}$ is said to *have a reduced row-echelon form* $\boldsymbol{R}$.

Algorithm 1.4.2 (Gaussian Elimination) The following procedures reduce an augmented matrix to a row-echelon form by using elementary row operations.

Step 1: Locate the leftmost column that does not consist entirely of zeros.

Step 2: Interchange the top row with another row, if necessary, to bring a nonzero entry to the top of the column found in Step 1.

Step 3: For each row below the top row, add a suitable multiple of the top row to it so that the entry below the leading entry of the top row becomes zero.

Step 4: Now cover the top row in the matrix and begin again with Step 1 applied to the submatrix that remains. Continue in this way until the entire matrix is in row-echelon form.

Algorithm 1.4.3 (Gauss-Jordan Elimination) Given an augmented matrix, use Algorithm 1.4.2 to reduce it to a row-echelon form. Then follow the following procedures to reduce it to a reduced row-echelon form.

Step 5: Multiply a suitable constant to each row so that all the leading entries become 1.

Step 6: Beginning with the last nonzero row and working upward, add suitable multiples of each row to the rows above to introduce zeros above the leading entries.

Example 1.4.4

1. In the following we use Gaussian Elimination to reduce the augmented matrix

$$\left(\begin{array}{ccccc|c} 0 & 0 & 2 & 4 & 2 & 8 \\ 1 & 2 & 4 & 5 & 3 & -9 \\ -2 & -4 & -5 & -4 & 3 & 6 \end{array}\right)$$

to a row-echelon form.

Step 1: The first column is the leftmost nonzero column.

Step 2: Interchange the first and second rows.

$$\left(\begin{array}{ccccc|c} 1 & 2 & 4 & 5 & 3 & -9 \\ 0 & 0 & 2 & 4 & 2 & 8 \\ -2 & -4 & -5 & -4 & 3 & 6 \end{array}\right)$$

Step 3: Add 2 times of the first row to the third row.

$$\left(\begin{array}{ccccc|c} 1 & 2 & 4 & 5 & 3 & -9 \\ 0 & 0 & 2 & 4 & 2 & 8 \\ 0 & 0 & 3 & 6 & 9 & -12 \end{array}\right)$$

Step 4: Cover the first row and begin again with Step 1 applied to the submatrix that remains.

$$\left(\begin{array}{ccccc|c} 1 & 2 & 4 & 5 & 3 & -9 \\ \hline 0 & 0 & 2 & 4 & 2 & 8 \\ 0 & 0 & 3 & 6 & 9 & -12 \end{array}\right)$$

Step 1: The third column is the leftmost nonzero column.

Step 2: No action is needed.

Step 3: Add $-\frac{3}{2}$ times of the second row to the third row.

$$\left(\begin{array}{ccccc|c} 1 & 2 & 4 & 5 & 3 & -9 \\ \hline 0 & 0 & 2 & 4 & 2 & 8 \\ 0 & 0 & 0 & 0 & 6 & -24 \end{array}\right)$$

Step 4: The augmented matrix is already in row-echelon form.

2. Now we use the Gauss-Jordan Elimination to reduce the augmented matrix

$$\left(\begin{array}{ccccc|c} 0 & 0 & 2 & 4 & 2 & 8 \\ 1 & 2 & 4 & 5 & 3 & -9 \\ -2 & -4 & -5 & -4 & 3 & 6 \end{array}\right)$$

to a reduced row-echelon form. First, we follow the steps of the previous example to obtain a row-echelon form

$$\left(\begin{array}{ccccc|c} 1 & 2 & 4 & 5 & 3 & -9 \\ 0 & 0 & 2 & 4 & 2 & 8 \\ 0 & 0 & 0 & 0 & 6 & -24 \end{array}\right).$$

Step 5: Multiply $\frac{1}{2}$ and $\frac{1}{6}$ to the second and third rows respectively.

$$\left(\begin{array}{ccccc|c} 1 & 2 & 4 & 5 & 3 & -9 \\ 0 & 0 & 1 & 2 & 1 & 4 \\ 0 & 0 & 0 & 0 & 1 & -4 \end{array}\right)$$

Step 6: Add -1 times of the third row to the second row.

$$\left(\begin{array}{ccccc|c} 1 & 2 & 4 & 5 & 3 & -9 \\ 0 & 0 & 1 & 2 & 0 & 8 \\ 0 & 0 & 0 & 0 & 1 & -4 \end{array}\right)$$

Add -3 times of the third row to the first row.

$$\left(\begin{array}{ccccc|c} 1 & 2 & 4 & 5 & 0 & 3 \\ 0 & 0 & 1 & 2 & 0 & 8 \\ 0 & 0 & 0 & 0 & 1 & -4 \end{array}\right)$$

Add -4 times of the second row to the first row.

$$\left(\begin{array}{ccccc|c} 1 & 2 & 0 & -3 & 0 & -29 \\ 0 & 0 & 1 & 2 & 0 & 8 \\ 0 & 0 & 0 & 0 & 1 & -4 \end{array}\right)$$

The matrix is now in reduced row-echelon form.

Remark 1.4.5

1. Every matrix has a unique reduced row-echelon form but can have many different row-echelon forms.

2. In the actual implementation of the algorithms, the steps mentioned in Algorithm 1.4.2 and Algorithm 1.4.3 are usually modified to avoid the roundoff errors during the computation. (See Question 1.21 for effects of roundoff errors.)

Discussion 1.4.6 To solve a system of linear equations, what we need to do is to apply either Gaussian Elimination or Gauss-Jordan Elimination to reduce the augmented matrix to a row-echelon form or the reduced row-echelon form. As we have seen in Theorem 1.2.7, the row operations do not change the solutions to the system. Thus by solving the system corresponding to the augmented matrix in a row-echelon form or the reduced row-echelon form, we can find the solutions to the original system easily.

Example 1.4.7 Let us consider the system of linear equations

$$\begin{cases} \phantom{-2x_1 - 4x_2 -{}} 2x_3 + 4x_4 + 2x_5 = 8 \\ x_1 + 2x_2 + 4x_3 + 5x_4 + 3x_5 = -9 \\ -2x_1 - 4x_2 - 5x_3 - 4x_4 + 3x_5 = 6\,. \end{cases}$$

The augmented matrix is

$$\left(\begin{array}{ccccc|c} 0 & 0 & 2 & 4 & 2 & 8 \\ 1 & 2 & 4 & 5 & 3 & -9 \\ -2 & -4 & -5 & -4 & 3 & 6 \end{array}\right).$$

We solve the system in two different ways.

Method 1: Using Gaussian Elimination as in Example 1.4.4.1, we find a row-echelon form of the augmented matrix

$$\left(\begin{array}{ccccc|c} 1 & 2 & 4 & 5 & 3 & -9 \\ 0 & 0 & 2 & 4 & 2 & 8 \\ 0 & 0 & 0 & 0 & 6 & -24 \end{array}\right).$$

which corresponds to the system

$$\begin{cases} x_1 + 2x_2 + 4x_3 + 5x_4 + 3x_5 = -9 \\ \phantom{x_1 + 2x_2 +{}} 2x_3 + 4x_4 + 2x_5 = 8 \\ \phantom{x_1 + 2x_2 + 4x_3 + 5x_4 +{}} 6x_5 = -24\,. \end{cases}$$

By the third equation, $x_5 = -4$. Substituting $x_5 = -4$ into the second equation,

$$2x_3 + 4x_4 + 2(-4) = 8 \quad \Leftrightarrow \quad x_3 = 8 - 2x_4.$$

Substituting $x_3 = 8 - 2x_4$ and $x_5 = -4$ into the first equation,

$$x_1 + 2x_2 + 4(8 - 2x_4) + 5x_4 + 3(-4) = -9 \quad \Leftrightarrow \quad x_1 = -29 - 2x_2 + 3x_4.$$

So the given linear system has a general solution

$$\begin{cases} x_1 = -29 - 2s + 3t \\ x_2 = s \\ x_3 = 8 - 2t \\ x_4 = t \\ x_5 = -4 \end{cases} \qquad \text{where } s,\ t \text{ are arbitrary parameters.}$$

Method 2: Using Gauss-Jordan Elimination as in Example 1.4.4.2, we find the reduced row-echelon form of the augmented matrix

$$\left(\begin{array}{ccccc|c} 1 & 2 & 0 & -3 & 0 & -29 \\ 0 & 0 & 1 & 2 & 0 & 8 \\ 0 & 0 & 0 & 0 & 1 & -4 \end{array}\right)$$

which corresponds to the system

$$\begin{cases} x_1 + 2x_2 \qquad - 3x_4 & = -29 \\ \qquad\qquad x_3 + 2x_4 & = \quad 8 \\ \qquad\qquad\qquad\qquad x_5 & = \ -4\,. \end{cases}$$

So the given linear system has a general solution

$$\begin{cases} x_1 = -29 - 2s + 3t \\ x_2 = s \\ x_3 = 8 - 2t & \qquad\text{where } s,\ t \text{ are arbitrary parameters.} \\ x_4 = t \\ x_5 = -4 \end{cases}$$

For both methods, to get a general solution, the variables corresponding to non-pivot columns are first set to be arbitrary. Then we equate the other variables accordingly. In this example, the second and fourth columns are non-pivot columns. We set $x_2 = s$ and $x_4 = t$, where s, t are arbitrary parameters, and equate x_1, x_3, x_5 in terms of s and t.

Remark 1.4.8

1. A linear system is inconsistent, i.e. has no solution, if the last column of a row-echelon form of the augmented matrix is a pivot column, i.e. there is a row with nonzero last entry but zero elsewhere:

$$\left(\begin{array}{cccccc|c} & \otimes & * & & & & * \\ & & & \otimes & * & & * \\ & 0 & & & \ddots & & * \\ & & & & & \otimes \ * & * \\ \hline 0 & \cdots & & & \cdots & 0 & \otimes \\ \hline \end{array}\right) \longleftarrow \text{nonzero}$$

where each $\otimes$ represents a pivot point (the leading entry of a nonzero row).

For example, $\left(\begin{array}{ccc|c} 3 & 2 & 3 & 4 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 2 \end{array}\right)$ is a row-echelon form of the augmented matrix of an inconsistent system.

2. A consistent linear system has only one solution if except the last column, every column of a row-echelon form of the augmented matrix is a pivot column:

$$\left(\begin{array}{ccccc|c} \otimes & & & & & * \\ & \otimes & & & & * \\ 0 & & \ddots & & & * \\ & & & & \otimes & * \\ 0 & \cdots & & & 0 & 0 \end{array}\right)$$

where each $\otimes$ represents a pivot point (the leading entry of a nonzero row).

In other words, a consistent linear system has only one solution if the number of variables in the linear system is equal to the number of nonzero rows in a row-echelon form of the augmented matrix.

For example, $\left(\begin{array}{ccc|c} 1 & 2 & 3 & 4 \\ 0 & 2 & 0 & 1 \\ 0 & 0 & -1 & 2 \end{array}\right)$ and $\left(\begin{array}{cccc|c} 1 & 1 & 2 & 3 & 4 \\ 0 & 2 & 0 & 1 & -1 \\ 0 & 0 & 4 & -1 & 2 \\ 0 & 0 & 0 & -1 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$.

3. A consistent linear system has infinitely many solutions if apart from the last column, a row-echelon form of the augmented matrix has at least one more non-pivot column:

$$\left(\begin{array}{ccccccc|c} \otimes & * & & & & & & * \\ & & \otimes & * & & & & * \\ & 0 & & 0 & \ddots & & & * \\ & & & \vdots & & \otimes & * & * \\ 0 & & \cdots & & \cdots & & 0 & 0 \\ & & & 0 & & & & \end{array}\right)$$

$\uparrow$

non-pivot column
(no pivot point)

where each $\otimes$ represents a pivot point (the leading entry of a nonzero row).

In other words, a consistent linear system has infinitely many solutions if the number of variables in the linear system is greater than the number of nonzero rows in a row-echelon form of the augmented matrix.

For example, $\left(\begin{array}{cccc|c} 5 & 1 & 2 & 3 & 4 \\ 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 2 \end{array}\right)$ and $\left(\begin{array}{cccc|c} 0 & 1 & 2 & 3 & 4 \\ 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 2 \end{array}\right)$.

Notation 1.4.9 When doing elementary row operations, we adopt the following notation:

1. kR_i means "multiply the ith row by the constant k".
2. $R_i \leftrightarrow R_j$ means "interchange the ith and jth rows".
3. $R_j + kR_i$ means "add k times of the ith row to the jth row".

Example 1.4.10

1. What is the condition that must be satisfied by a, b, c so that the system of linear equations

$$\begin{cases} x + 2y - 3z = a \\ 2x + 6y - 11z = b \\ x - 2y + 7z = c \end{cases}$$

has at least one solution?

Solution In here, we just regard a, b, c as constants and apply the elementary row operations to the augmented matrix as before.

$$\left(\begin{array}{ccc|c} 1 & 2 & -3 & a \\ 2 & 6 & -11 & b \\ 1 & -2 & 7 & c \end{array}\right) \xrightarrow{R_2 - 2R_1} \left(\begin{array}{ccc|c} 1 & 2 & -3 & a \\ 0 & 2 & -5 & b-2a \\ 1 & -2 & 7 & c \end{array}\right) \xrightarrow{R_3 - R_1}$$
$$\left(\begin{array}{ccc|c} 1 & 2 & -3 & a \\ 0 & 2 & -5 & b-2a \\ 0 & -4 & 10 & c-a \end{array}\right) \xrightarrow{R_3 + 2R_2} \left(\begin{array}{ccc|c} 1 & 2 & -3 & a \\ 0 & 2 & -5 & b-2a \\ 0 & 0 & 0 & 2b+c-5a \end{array}\right)$$

The system has either no solution or infinitely many solutions. It has (infinitely many) solutions if and only if $2b + c - 5a = 0$.

2. Determine the values of b so that the system of linear equations

$$\begin{cases} x + 2y + z = 1 \\ 2x + by + 2z = 2 \\ 4x + 8y + b^2 z = 2b \end{cases}$$

has (a) no solution, (b) a unique solution, and (c) infinitely many solutions.

Solution

$$\left(\begin{array}{ccc|c} 1 & 2 & 1 & 1 \\ 2 & b & 2 & 2 \\ 4 & 8 & b^2 & 2b \end{array}\right) \xrightarrow[R_3 - 4R_1]{R_2 - 2R_1} \left(\begin{array}{ccc|c} 1 & 2 & 1 & 1 \\ 0 & b-4 & 0 & 0 \\ 0 & 0 & b^2-4 & 2b-4 \end{array}\right)$$

(a) The system has no solution if $b^2 - 4 = 0$ and $2b - 4 \neq 0$, i.e. $b = -2$.

(b) The system has a unique solution if $b - 4 \neq 0$ and $b^2 - 4 \neq 0$, i.e. $b \neq 4$, $b \neq 2$ and $b \neq -2$.

(c) The system has infinitely many solutions if either (i) $b - 4 = 0$ or (ii) $b^2 - 4 = 0$ and $2b - 4 = 0$, i.e. (i) $b = 4$ or (ii) $b = 2$.

3. Determine the values of a and b so that the system of linear equations

$$\begin{cases} ax + y = a \\ x + y + z = 1 \\ y + az = b \end{cases}$$

has (a) no solution, (b) a unique solution, and (c) infinitely many solutions.

Solution In doing elementary row operations, you cannot multiply a row by 0 or $\frac{1}{0}$ (when using the operation 1 in Discussion 1.2.3) and cannot add $\frac{1}{0}$ times of a row to another row (when using the operation 3 in Discussion 1.2.3). For this question, we need to be careful in handling the unknown constant a.

The augmented matrix of the system is

$$\left(\begin{array}{ccc|c} a & 1 & 0 & a \\ 1 & 1 & 1 & 1 \\ 0 & 1 & a & b \end{array}\right).$$

If we follow the Gaussian Elimination, the next step is to add $-\frac{1}{a}$ times of the first row to the second row. However, since the value of a is unknown, we need to consider two different situations.

Case 1: $a = 0$. After substituting $a = 0$ to the augmented matrix, we have

$$\left(\begin{array}{ccc|c} 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & b \end{array}\right) \xrightarrow{R_1 \leftrightarrow R_2} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & b \end{array}\right) \xrightarrow{R_3 - R_2} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & b \end{array}\right).$$

So under the assumption that $a = 0$, the system has no solution if $b \neq 0$; and the system has infinitely many solution if $b = 0$.

Case 2: $a \neq 0$. Then

$$\left(\begin{array}{ccc|c} a & 1 & 0 & a \\ 1 & 1 & 1 & 1 \\ 0 & 1 & a & b \end{array}\right) \xrightarrow{R_2 - \frac{1}{a}R_1} \left(\begin{array}{ccc|c} a & 1 & 0 & a \\ 0 & \frac{a-1}{a} & 1 & 0 \\ 0 & 1 & a & b \end{array}\right).$$

Again, if we follow the Gaussian Elimination, the next step is to add $-\frac{a}{a-1}$ times of the second row to the third row. We need to consider two cases.

Case 2a: $a = 1$. After substituting $a = 1$ into the last matrix, we have

$$\left(\begin{array}{ccc|c} 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & b \end{array}\right) \xrightarrow{R_2 \leftrightarrow R_3} \left(\begin{array}{ccc|c} 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & b \\ 0 & 0 & 1 & 0 \end{array}\right).$$

So under the assumption that $a = 1$, the system has only one solution.

Case 2b: $a \neq 1$. Then

$$\left(\begin{array}{ccc|c} a & 1 & 0 & a \\ 0 & \frac{a-1}{a} & 1 & 0 \\ 0 & 1 & a & b \end{array}\right) \xrightarrow{R_3 - \frac{a}{a-1}R_2} \left(\begin{array}{ccc|c} a & 1 & 0 & a \\ 0 & \frac{a-1}{a} & 1 & 0 \\ 0 & 0 & \frac{a^2-2a}{a-1} & b \end{array}\right).$$

So under the assumption that $a \neq 0$ and $a \neq 1$, the system has no solution if $\frac{a^2-2a}{a-1} = 0$ and $b \neq 0$, i.e. $a = 2$ and $b \neq 0$; the system has only one solution if $\frac{a^2-2a}{a-1} \neq 0$, i.e. $a \neq 2$; and the system has infinitely many solutions if $\frac{a^2-2a}{a-1} = 0$ and $b = 0$, i.e. $a = 2$ and $b = 0$.

Summarizing the results, we have

(a) the system has no solution if $b \neq 0$ and $a = 0$ or 2;

(b) the system has only one solution if $a \neq 0$ and $a \neq 2$; and

(c) the system has infinitely many solutions if $b = 0$ and $a = 0$ or 2.

If we rearrange the rows of the augmented matrix in the following way: the second row at the top, the third row in the middle and the first row at the bottom (try it), the problem will be much easier to be solved using Gaussian Elimination.

4. Given a cubic curve with equation

$$y = a + bx + cx^2 + dx^3,$$

where a, b, c, d are real constants, that passes through the points $(0, 10)$, $(1, 7)$, $(3, -11)$ and $(4, -14)$, find the values of a, b, c, d.

Solution By substituting $(x, y) = (0, 10)$, $(1, 7)$, $(3, -11)$ and $(4, -14)$ into the equation of the cubic curve, we obtain a system of linear equations

$$\begin{cases} a & = 10 \\ a + b + c + d & = 7 \\ a + 3b + 9c + 27d & = -11 \\ a + 4b + 16c + 64d & = -14 \end{cases}$$

where a, b, c, d are unknowns.

$$\left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 10 \\ 1 & 1 & 1 & 1 & 7 \\ 1 & 3 & 9 & 27 & -11 \\ 1 & 4 & 16 & 64 & -14 \end{array}\right) \begin{array}{c} R_2 - R_1 \\ R_3 - R_1 \\ \longrightarrow \\ R_4 - R_1 \end{array} \left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 10 \\ 0 & 1 & 1 & 1 & -3 \\ 0 & 3 & 9 & 27 & -21 \\ 0 & 4 & 16 & 64 & -24 \end{array}\right) \begin{array}{c} R_3 - 3R_2 \\ \longrightarrow \\ R_4 - 4R_2 \end{array}$$

$$\left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 10 \\ 0 & 1 & 1 & 1 & -3 \\ 0 & 0 & 6 & 24 & -12 \\ 0 & 0 & 12 & 60 & -12 \end{array}\right) \begin{array}{c} R_4 - 2R_3 \\ \longrightarrow \end{array} \left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 10 \\ 0 & 1 & 1 & 1 & -3 \\ 0 & 0 & 6 & 24 & -12 \\ 0 & 0 & 0 & 12 & 12 \end{array}\right) \begin{array}{c} \frac{1}{6}R_3 \\ \longrightarrow \\ \frac{1}{12}R_4 \end{array}$$

$$\left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 10 \\ 0 & 1 & 1 & 1 & -3 \\ 0 & 0 & 1 & 4 & -2 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right) \begin{array}{c} R_2 - R_4 \\ \longrightarrow \\ R_3 - 4R_4 \end{array} \left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 10 \\ 0 & 1 & 1 & 0 & -4 \\ 0 & 0 & 1 & 0 & -6 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right) \begin{array}{c} R_2 - R_3 \\ \longrightarrow \end{array} \left(\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 10 \\ 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & -6 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right)$$

So the solution is $a = 10$, $b = 2$, $c = -6$ and $d = 1$, i.e. the equation of the cubic curve is $y = 10 + 2x - 6x^2 + x^3$.

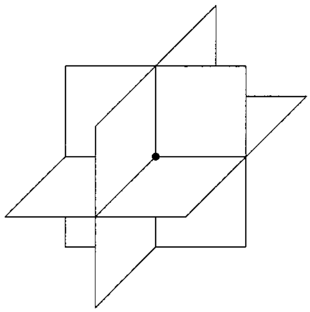
Discussion 1.4.11 Given a consistent system of linear equations in three variables x, y, z, each nonzero equation in the system represents a plane in the xyz-space. Solutions to the system are intersection points of these planes. (See Discussion 1.1.11.2 and Question 1.8.) Now, let us first reduce the augmented matrix of the linear system to a row-echelon form $\boldsymbol{R}$. Since the system is consistent, there are at most three nonzero rows in $\boldsymbol{R}$.

1. Suppose $\boldsymbol{R}$ has three nonzero rows. By Remark 1.4.8.2, the system has only one solution. On the other hand, the three rows of $\boldsymbol{R}$ represent three planes that intersect at a common point (see Figure A), which is the solution to the system.

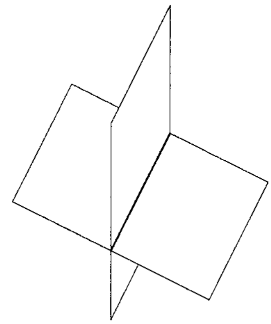
2. Suppose $\boldsymbol{R}$ has two nonzero rows. Apart from the last column, $\boldsymbol{R}$ has one more non-pivot column. So a general solution for the system needs only one arbitrary parameter. On the other hand, the two nonzero rows of $\boldsymbol{R}$ give us two non-parallel planes and they intersect at a common line (see Figure B). This means that the line of intersection is described by a general solution with one arbitrary parameter.

3. Suppose $\boldsymbol{R}$ has only one nonzero row. A general solution for the system needs two arbitrary parameters. On the other hand, the solutions to the system form a plane represented by the nonzero row in $\boldsymbol{R}$. This means that the plane is described by a general solution with two arbitrary parameters.

4. Suppose $\boldsymbol{R}$ has no nonzero row. The system is a zero system. A general solution needs three arbitrary parameters and it represents the whole xyz-space. (See Example 1.3.5.4.)



A: three planes intersect at a point



B: two planes intersect at a line

Example 1.4.12

1. Consider the system of linear equations

$$\begin{cases} x+y+2z=1 \\ x-y-\ \ z=0 \\ x+y-\ \ z=2. \end{cases}$$

Since

$$\left(\begin{array}{ccc|c} 1 & 1 & 2 & 1 \\ 1 & -1 & -1 & 0 \\ 1 & 1 & -1 & 2 \end{array}\right) \xrightarrow[\text{Elimination}]{\text{Gauss-Jordan}} \left(\begin{array}{ccc|c} 1 & 0 & 0 & \frac{2}{3} \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -\frac{1}{3} \end{array}\right),$$

the system has only one solution $x=\frac{2}{3}$, $y=1$ and $z=-\frac{1}{3}$, i.e. the three planes intersect at the point $(\frac{2}{3}, 1, -\frac{1}{3})$.

2. Consider the system of linear equations

$$\begin{cases} x+y+2z=1 \\ x-y-\ \ z=0 \\ 2x \quad\ \ +\ \ z=1 \\ 3x-y \qquad\ =1. \end{cases}$$

Since

$$\left(\begin{array}{ccc|c} 1 & 1 & 2 & 1 \\ 1 & -1 & -1 & 0 \\ 2 & 0 & 1 & 1 \\ 3 & -1 & 0 & 1 \end{array}\right) \xrightarrow[\text{Elimination}]{\text{Gauss-Jordan}} \left(\begin{array}{ccc|c} 1 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & \frac{3}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right),$$

the system has a general solution $x=\frac{1}{2}-\frac{1}{2}t$, $y=\frac{1}{2}-\frac{3}{2}t$ and $z=t$ where t is an arbitrary parameter, i.e. the four planes intersect at the line that consists of points $(\frac{1}{2}-\frac{1}{2}t,\ \frac{1}{2}-\frac{3}{2}t,\ t)$ for all real number t.

3. Consider the system of linear equations

$$\begin{cases} x + \ \ y + 2z = 1 \\ 3x + 3y + 6z = 3. \end{cases}$$

Since

$$\left(\begin{array}{ccc|c} 1 & 1 & 2 & 1 \\ 3 & 3 & 6 & 3 \end{array} \right) \xrightarrow[\text{Elimination}]{\text{Gauss-Jordan}} \left(\begin{array}{ccc|c} 1 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 \end{array} \right),$$

the system has a general solution $x = 1 - s - 2t$, $y = s$ and $z = t$ where s, t are arbitrary parameters, i.e. the two equations in the system represent the same planes that consists of points $(1 - s - 2t,\ s,\ t)$ for all real numbers s and t.

Section 1.5 Homogeneous Linear Systems

Definition 1.5.1 A system of linear equations is said to be *homogeneous* if it has the form

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = 0 \\ \qquad \vdots \qquad\qquad\qquad\qquad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = 0 \end{cases}$$

where $a_{11}, a_{12}, \ldots, a_{mn}$ are real constants. A linear system is called non-homogeneous if it is not homogeneous.

Note that $x_1 = 0$, $x_2 = 0$, $\ldots$, $x_n = 0$ is always a solution to the homogeneous system and it is called the *trivial solution.* Any solution other than the trivial solution is called a *non-trivial solution.*

Example 1.5.2

Given a quadric surface with equation

$$ax^2 + by^2 + cz^2 = d,$$

where a, b, c, d are real constants, that passes through the points $(1, 1, -1)$, $(1, 3, 3)$ and $(-2, 0, 2)$, find a formula for the quadric surface.

Solution By substituting $(x, y, z) = (1, 1, -1)$, $(1, 3, 3)$ and $(-2, 0, 2)$ into the equation, we obtain a homogeneous system of linear equations

$$\begin{cases} a + \ b + \ c - d = 0 \\ a + 9b + 9c - d = 0 \\ 4a \qquad + 4c - d = 0 \end{cases}$$

where a, b, c, d are unknowns. Solving the system, we obtain a general solution

$$\begin{cases} a = t \\ b = \frac{3}{4}t \\ c = -\frac{3}{4}t \\ d = t \end{cases} \quad \text{where } t \text{ is an arbitrary parameter.}$$

There are infinitely many solutions. However, any one of the nontrivial solutions gives us a formula for the same quadric surface. In particular, $4x^2 + 3y^2 - 3z^2 = 4$ is a formula for the quadric surface.

Discussion 1.5.3

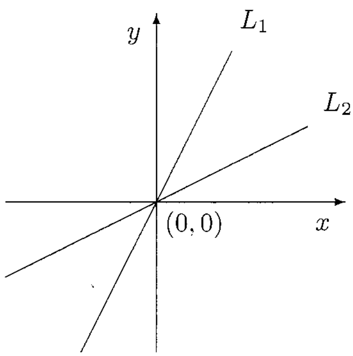
1. In xy-plane, the equations

$$\begin{cases} a_1x + b_1y = 0 & (L_1) \\ a_2x + b_2y = 0, & (L_2) \end{cases}$$

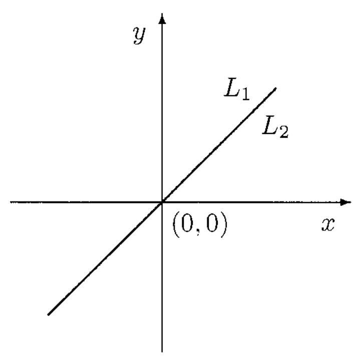
   where a_1, b_1 are not both zero and a_2, b_2 are not both zero, are straight lines through the origin (i.e. the point $(0, 0)$).

   (a) The system has only the trivial solution if and only if L_1 and L_2 are not the same line.

   (b) The system has non-trivial solutions if and only if L_1 and L_2 are the same line.



only the trivial solution

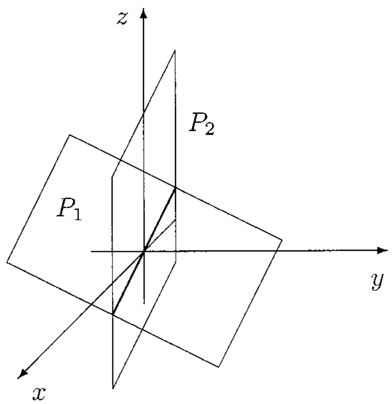


infinitely many solutions

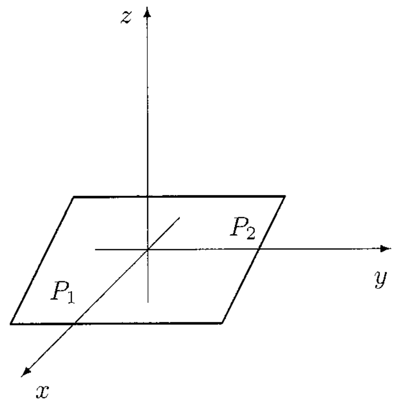
2. In the three dimensional space, the two equations in the system

$$\begin{cases} a_1x + b_1y + c_1z = 0 & (P_1) \\ a_2x + b_2y + c_2z = 0, & (P_2) \end{cases}$$

where a_1, b_1, c_1 are not all zero and a_2, b_2, c_2 are not all zero, represent two planes containing the origin (i.e. the point $(0,0,0)$). The system always has non-trivial solutions, see Remark 1.5.4.2.



infinitely many solutions



infinitely many solutions

Remark 1.5.4

1. A homogeneous system of linear equations has either only the trivial solution or infinitely many solutions in addition to the trivial solution.

2. A homogeneous system of linear equations with more unknowns than equations has infinitely many solutions.

Exercise 1

Question 1.1 to Question 1.15 are exercises for Sections 1.1 to 1.3.

1. Which of the following are linear equations in x_1, x_2 and x_3? In Parts (i)-(l), m is a constant.

(a) $\sqrt{3}x_1 - x_2 - x_3 = 0$, (b) $x_1x_3 + 2x_2 + x_1 = 4$,

(c) $x_1 = -7x_2 + 3x_3$, (d) $x_1 + 2x_2 + x_3^2 = 1$,

(e) $\sqrt{x_1} - 2x_2 + x_3 = 1$, (f) $x_1 + \frac{1}{2}x_2 + \frac{1}{3}x_3 = 1.333\pi$,

(g) $2^{x_1+x_2+x_3} = 5$, (h) $\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} = \frac{1}{10}$,

(i) $x_1 + x_2 + x_3 = \cos(m)$, (j) $\cos(x_1) + \cos(x_2) + \cos(x_3) = \cos(m)$,

(k) $3x_1 + x_2 - x_3^m = 0$, (l) $mx_1 - m^2x_2 = 9$.

2. Write down a general solution for each of the following linear equations.
   (a) $2x + 5y = 0$, (b) $8w - 2x - 5y + 6z = -1.5$,
   (c) $3x_1 - 8x_2 + 2x_3 + x_4 - 4x_5 = 1$.

3. (a) Find a linear equation in the variables x and y that has a general solution $x = 1 + 2t$ and $y = t$ where t is an arbitrary parameter.
   (b) Show that $x = t$ and $y = \frac{1}{2}t - \frac{1}{2}$, where t is an arbitrary parameter, is also a general solution for the equation in Part (a).

4. (a) Find a linear equation in the variables x, y and z that has a general solution

$$\begin{cases} x = 3 - 4s + t \\ y = s \\ z = t \end{cases} \quad \text{where } s, t \text{ are arbitrary parameters.}$$

   (b) Express a general solution for the equation in Part (a) in two other different ways.
   (c) Write down a linear system of two different nonzero linear equations such that the system has the same general solution as in Part (a).

5. (a) Give a geometrical interpretation for the linear equation $x + y + z = 1$.
   (b) Give geometrical interpretations for the linear equation $x - y = 0$ in (i) the xy-plane; and (ii) the xyz-space.
   (c) Give a geometrical interpretation for the solutions to the system of linear equations

$$\begin{cases} x + y + z = 1 \\ x - y \qquad = 0. \end{cases}$$

6. Consider the system of linear equations

$$\begin{cases} 3x + 4y - 5z = -8 \\ x - 2y + z = 2. \end{cases}$$

   (a) For any real number t, verify that $x = \frac{1}{5}(-4 + 3t)$, $y = \frac{1}{5}(-7 + 4t)$, $z = t$ is a solution to the linear system.
   (b) Write down two particular solutions to the system.

7. Each equation in the following linear system represents a line in the xy-plane:

$$\begin{cases} a_1x + b_1y = c_1 \\ a_2x + b_2y = c_2 \\ a_3x + b_3y = c_3, \end{cases}$$

where a_1, a_2, a_3, b_1, b_2, b_3, c_1, c_2, c_3 are constants and for each $i = 1, 2, 3$, a_i, b_i are not both zero. Discuss the relative positions of the three lines when the system

(a) has no solution;

(b) has only one solution;

(c) has infinitely many solutions.

8. Each equation in the following linear system represents a plane in the xyz-space:

$$\begin{cases} a_1x + b_1y + c_1z = d_1 \\ a_2x + b_2y + c_2z = d_2 \\ a_3x + b_3y + c_3z = d_3, \end{cases}$$

where a_1, a_2, a_3, b_1, b_2, b_3, c_1, c_2, c_3, d_1, d_2, d_3 are constants and for each $i = 1, 2, 3$, a_i, b_i, c_i are not all zero. Discuss the relative positions of the three planes when the linear system

(a) has no solution;

(b) has only one solution;

(c) has infinitely many solutions.

9. (a) Does an inconsistent linear system with more unknowns than equations exist?

(b) Does a linear system which has only one solution, but more equations than unknowns, exist?

(c) Does a linear system which has only one solution, but more unknowns than equations, exist?

(d) Does a linear system which has infinitely many solutions, but more equations than unknowns, exist?

Justify your answers.

10. Show that the following augmented matrices are row equivalent to each other.

$$\boldsymbol{A} = \left(\begin{array}{cc|c} 1 & 3 & 2 \\ 0 & 0 & 0 \end{array}\right), \quad \boldsymbol{B} = \left(\begin{array}{cc|c} 1 & 3 & 2 \\ 1 & 3 & 2 \end{array}\right),$$

$$\boldsymbol{C} = \left(\begin{array}{cc|c} 5 & 15 & 10 \\ 1 & 3 & 2 \end{array}\right), \quad \boldsymbol{D} = \left(\begin{array}{cc|c} 0 & 0 & 0 \\ 2 & 6 & 4 \end{array}\right).$$

11. Find a series of elementary row operations that bring the augmented matrix $\boldsymbol{A}$ to the augmented matrix $\boldsymbol{B}$ where

$$\boldsymbol{A} = \left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{array}\right) \quad \text{and} \quad \boldsymbol{B} = \left(\begin{array}{ccc|c} 3 & 0 & 0 & 0 \\ 0 & 6 & 0 & 0 \\ 0 & 0 & 9 & 9 \end{array}\right).$$

12. For each of the following augmented matrices, (i) determine whether the matrix is in row-echelon form, reduced row-echelon form, both, or neither; and (ii) find a system of linear equations corresponding to the augmented matrix and then solve the system (if possible). You may assume that the variables are x_1, x_2, x_3, etc.

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

(c) $\left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right)$; (d) $\left(\begin{array}{cccc|c} -2 & 0 & -1 & -7 & 8 \\ 0 & 3 & 0 & 3 & 2 \\ 0 & 0 & 0 & 1 & -1 \end{array}\right)$;

(e) $\left(\begin{array}{ccccc|c} 1 & 0 & 2 & -2 & 3 & -2 \\ 0 & 0 & 1 & 1 & 3 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 5 & 5 \end{array}\right)$; (f) $\left(\begin{array}{cccccc|c} 1 & 0 & -2 & 0 & 2 & 0 & -2 \\ 0 & 1 & 0 & 0 & 2 & 0 & 4 \\ 0 & 0 & 0 & 1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{array}\right)$.

13. Determine whether the following augmented matrices are row equivalent to each other.

$$\boldsymbol{A} = \left(\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 0 & 0 & 0 & 0 \\ 2 & 4 & 6 & 0 \end{array}\right), \quad \boldsymbol{B} = \left(\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 3 & 6 & 9 & 0 \\ 1 & 1 & 1 & 0 \end{array}\right), \quad \boldsymbol{C} = \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \end{array}\right).$$

14. The following is the reduced row-echelon form of the augmented matrix of a linear system:

$$\left(\begin{array}{ccc|c} a & 0 & 0 & d \\ 0 & b & 0 & e \\ 0 & 0 & c & f \end{array}\right),$$

where a, b, c, d, e, f are constants. Write down the necessary conditions on a, b, c, d, e, f so that the solution set of the linear system is a plane in the three dimensional space that does not contain the origin.

15. Consider the augmented matrix

$$\left(\begin{array}{ccc|c} a & b & c & d \\ e & f & g & h \end{array}\right),$$

where a, b, c, d, e, f, g, h are constants.

(a) Write down all possible reduced row-echelon forms for the augmented matrix.

(b) Which of the reduced row-echelon forms in Part (a) represent inconsistent systems?

(c) Which of the reduced row-echelon forms in Part (a) are row equivalent to each other?

Question 1.16 to Question 1.30 are exercises for Sections 1.4 and 1.5.

16. Solve each of the following systems by Gaussian Elimination or Gauss-Jordan Elimination.

(a) $\begin{cases} x_1 + x_2 + 2x_3 = 4 \\ -x_1 + 2x_2 - x_3 = 1 \\ 2x_1 \qquad + 3x_3 = -2 \end{cases}$

(b) $\begin{cases} w - x \qquad + z = -1 \\ 2w - x - y \qquad = 2 \\ -2w \qquad + 3y - z = 3 \end{cases}$

(c) $\begin{cases} x_1 + 3x_2 + 3x_3 + 2x_4 = 1 \\ 2x_1 + 6x_2 + 9x_3 + 5x_4 = 5 \\ -x_1 - 3x_2 + 3x_3 \qquad = 5 \end{cases}$

(d) $\begin{cases} x + y + 2z = 4 \\ x - y - z = -1 \\ 2x - 4y - 5z = 1 \end{cases}$

(e) $\begin{cases} u - v + 2w = 6 \\ 2u + 2v - 5w = 3 \\ 2u + 5v + w = 9 \end{cases}$

(f) $\begin{cases} w + 5x - 2y + 6z = 0 \\ 2w - 2x + y + 3z = 1 \\ w - 7x + 3y - 3z = 1 \\ 5w + x \qquad + 12z = 2 \end{cases}$

(g) $\begin{cases} 2x_2 + x_3 + 2x_4 - x_5 = 4 \\ x_2 \qquad + x_4 - x_5 = 3 \\ 4x_1 + 6x_2 + x_3 + 4x_4 - 3x_5 = 8 \\ 2x_1 + 2x_2 \qquad + x_4 - x_5 = 2 \end{cases}$

17. Solve the following system of non-linear equations:

$$\begin{cases} x^2 - y^2 + 2z^2 = 6 \\ 2x^2 + 2y^2 - 5z^2 = 3 \\ 2x^2 + 5y^2 + z^2 = 9. \end{cases}$$

18. Solve the following system of non-linear equations if $0 \le \theta < 2\pi$ and $0 \le \phi < 2\pi$:

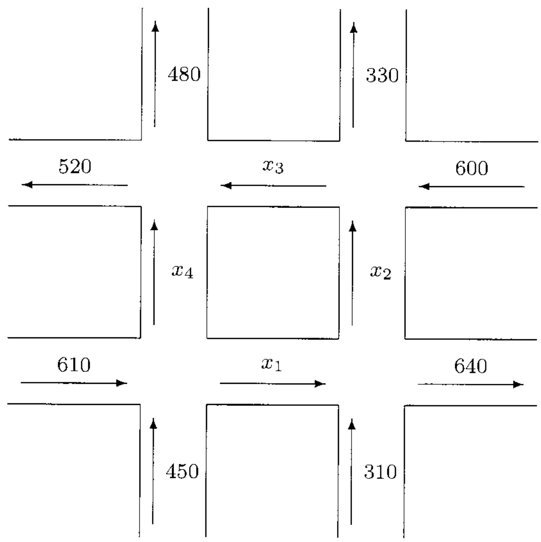
$$\begin{cases} \cos(\theta) - \sin(\phi) - \tan(\phi) = 0 \\ 3\cos(\theta) - \sin(\phi) - 2\tan(\phi) = 0. \end{cases}$$

19. In a three-commodity market, the supply and demand for each commodity depend on the prices of the commodities. Let D_1, D_2, D_3 be the respective demands for products 1, 2 and 3, S_1, S_2, S_3 the respective supplies and P_1, P_2, P_3 the respective prices for the commodities. Suppose the market can be described by the linear equations

$$\begin{aligned} D_1 &= -2P_1 + P_2 + P_3 + 4, & S_1 &= -P_1 + 2P_2 + 2P_3 - 1, \\ D_2 &= P_1 + 2P_2 + P_3 - 1, & S_2 &= 2P_1 + P_2 + 2P_3 - 2, \\ D_3 &= 2P_1 + P_2 + P_3 + 4, & S_3 &= P_1 + 2P_2 + 3P_3 - 1. \end{aligned}$$

Use Gaussian Elimination or Gauss-Jordan Elimination to find the *equilibrium solution* (where supplies equal to demands).

20. In the downtown section of a certain city, two sets of one-way streets intersect as shown in the following:



The average hourly volume of traffic entering and leaving this section during rush hour is given in the diagram.

(a) Do we have enough information to find the traffic volumes x_1, x_2, x_3, x_4? Explain your answer.

(b) Given that $x_4 = 500$, find x_1, x_2, x_3.

(The average hourly volume of traffic entering an intersection must be equal to the volume that leaving.)

21. Consider the two linear systems

$$\text{(i)} \begin{cases} x + \qquad y = 0 \\ x + (1+\delta)y = 1 \end{cases} \qquad \text{and} \qquad \text{(ii)} \begin{cases} x + \qquad y = 0 \\ x - (1+\delta)y = 1 \end{cases}$$

where δ is a real number close to zero.

(a) Solve the two linear systems for $\delta = 0.0001$ and 0.0002.

(b) Compare the effect on the solution to the two systems when we change δ from 0.0001 to 0.0002.

(c) Explain your answer in Part (b) geometrically.
(Hint: Plot the lines to see the difference between the two systems.)

(This question shows that solutions of some linear systems can be largely affected by small changes in the coefficients, say, due to roundoff errors.)

22. For each of the following linear systems, determine the values of a such that the system has (i) no solution; (ii) only one solution; and (iii) infinitely many solutions.

(a) $\begin{cases} x + y + az = 0 \\ x + ay + z = 0 \\ ax + y + z = 0 \end{cases}$ (b) $\begin{cases} x + y + z = 1 \\ 2x + ay + 2z = 2 \\ 4x + 4y + a^2 z = 2a \end{cases}$

(c) $\begin{cases} ax + ay = 1 \\ ax - |a|y = 1 \end{cases}$

23. Determine the values of a and b so that the linear system

$$\begin{cases} ax \qquad + bz = 2 \\ ax + ay + 4z = 4 \\ \qquad ay + 2z = b \end{cases}$$

(a) has no solution;
(b) has only one solution;
(c) has infinitely many solutions and a general solution has one arbitrary parameter;
(d) has infinitely many solutions and a general solution has two arbitrary parameters.

24. Determine the values of a, b and c so that the linear system

$$\begin{cases} ax + ay + az = c \\ \qquad by + bz = a \\ \qquad\qquad cz = b \end{cases}$$

(a) has no solution;
(b) has only one solution;
(c) has infinitely many solutions.

25. Without using pencil and paper, determine which of the following homogeneous systems have non-trivial solutions.

(a) $\begin{cases} x - y = 0 \\ x + y = 0 \end{cases}$ (b) $\begin{cases} x - 4y + 6z = 0 \\ 2y - 5z = 0 \\ 3z = 0 \end{cases}$

(c) $\begin{cases} 3x_1 - x_2 + 11x_3 = 0 \\ x_1 + x_2 - 7x_3 + x_4 = 0 \\ 5x_1 + x_2 - 4x_3 + 2x_4 = 0 \end{cases}$ (d) $\begin{cases} ax + by + cz = 0 \\ dx + ey + fz = 0 \\ (a+d)x + (b+e)y + (c+f)z = 0 \end{cases}$

For Part (d), a, b, c, d, e, f are constants.

26. When propane gas burns, the propane combines with oxygen to form carbon dioxide and water:

$$w\,C_3H_8 + x\,O_2 \;\rightarrow\; y\,CO_2 + z\,H_2O$$

where w and x are the numbers of propane and oxygen molecules, respectively, required for the combustion; and y and z are the numbers of carbon dioxide and water molecules, respectively, produced.

    (a) By equating the numbers of carbon, hydrogen and oxygen atoms, respectively, on both sides of the chemical equation, write down a homogeneous system of three equations in terms of w, x, y, z.

    (b) Find a general solution for the homogeneous system obtained in Part (a).

    (c) Find the (non-trivial) solution of w, x, y, z with smallest values. (Note that w, x, y, z are positive integers.)

27. Consider the homogeneous system of equations

$$\begin{cases} ax + by + cz = 0 \\ dx + ey + fz = 0 \end{cases}$$

where a, b, c, d, e, f are constants.

    (a) Let $x = x_0$, $y = y_0$, $z = z_0$ be a solution to the system and k is a constant. Show that $x = kx_0$, $y = ky_0$, $z = kz_0$ is also a solution to the system.

    (b) Let $x = x_0$, $y = y_0$, $z = z_0$ and $x = x_1$, $y = y_1$, $z = z_1$ be two solutions to the system. Show that $x = x_0 + x_1$, $y = y_0 + y_1$, $z = z_0 + z_1$ is also a solution to the system.

28. Consider the homogeneous linear system

$$\begin{cases} a_1x + b_1y + c_1z = 0 \\ a_2x + b_2y + c_2z = 0 \\ a_3x + b_3y + c_3z = 0 \end{cases}$$

where a_1, a_2, a_3, b_1, b_2, b_3, c_1, c_2, c_3 are constants. Determine all possible reduced row-echelon forms of the augmented matrix of the system and describe the geometrical meaning of the solutions obtained from various reduced row-echelon forms.

29. The following is the reduced row-echelon form of the augmented matrix of a linear system:

$$\left(\begin{array}{ccc|c} a & b & c & d \\ 0 & e & f & g \\ 0 & 0 & h & k \end{array}\right)$$

where a, b, c, d, e, f, g, h, k are constants. Suppose the solution set of this system is represented by a line that passes through the origin and the point $(1, 1, 1)$. Find the values of a, b, c, d, e, f, g, h, k. Justify your answers.

30. Determine which of the following statements are true. Justify your answer.

    (a) A homogeneous system can have a non-trivial solution.

    (b) A non-homogeneous system can have a trivial solution.

    (c) If a homogeneous system has the trivial solution, then it cannot have a non-trivial solution.

    (d) If a homogeneous system has a non-trivial solution, then it cannot have a trivial solution.

    (e) If a homogeneous system has a unique solution, then the solution has to be trivial.

    (f) If a homogeneous system has the trivial solution, then the solution has to be unique.

    (g) If a homogeneous system has a non-trivial solution, then there are infinitely many solutions to the system.

Chapter 2

Matrices

Section 2.1 Introduction to Matrices

Definition 2.1.1 A *matrix* (plural *matrices*) is a rectangular array of numbers. The numbers in the array are called *entries* in the matrix. The *size* of a matrix is given by $m \times n$ where m is the number of rows and n is the number of columns. The (i, j)-*entry* of a matrix is the number which is in the ith row and jth column of the matrix.

Example 2.1.2

1. $\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 0 & -1 \end{pmatrix}$ is a 3×2 matrix.

   The (1,2)-entry of the matrix is 2 and the (3,1)-entry is 0.

2. $\begin{pmatrix} 2 & 1 & 0 \end{pmatrix}$ is a 1×3 matrix.

3. $\begin{pmatrix} \sqrt{2} & 3.1 & -2 \\ 3 & \frac{1}{2} & 0 \\ 0 & \pi & 0 \end{pmatrix}$ is a 3×3 matrix.

4. $\begin{pmatrix} 4 \end{pmatrix}$ is a 1×1 matrix.

   (A 1×1 matrix will usually be treated as a real number in computation.)

5. $\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$ is a 3×1 matrix.

Definition 2.1.3 A *column matrix* (or a *column vector*) is a matrix with only one column. A *row matrix* (or a *row vector*) is a matrix with only one row.

Example 2.1.4 The matrix in Example 2.1.2.5 is a column matrix and the matrix in Example 2.1.2.2 is a row matrix. The matrix in Example 2.1.2.4 is both a column and row matrix.

Notation 2.1.5 In general, an $m \times n$ matrix can be written as

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

or simply $\boldsymbol{A} = (a_{ij})_{m \times n}$, where a_{ij} is the (i, j)-entry of $\boldsymbol{A}$. Sometimes, if the size of the matrix is already known, we may just write $\boldsymbol{A} = (a_{ij})$.

Example 2.1.6 Write down the following matrices explicitly:

1. $\boldsymbol{A} = (a_{ij})_{2\times 3}$ where $a_{ij} = i + j$.
2. $\boldsymbol{B} = (b_{ij})_{3\times 2}$ where $b_{ij} = \begin{cases} 1 & \text{if } i+j \text{ is even} \\ -1 & \text{if } i+j \text{ is odd.} \end{cases}$

Answers

1. $\boldsymbol{A} = \begin{pmatrix} 2 & 3 & 4 \\ 3 & 4 & 5 \end{pmatrix}$; 2. $\boldsymbol{B} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \\ 1 & -1 \end{pmatrix}$.

Definition 2.1.7 The following are some special types of matrices:

1. A matrix is called a *square matrix* if it has the same number of rows and columns. In particular, an $n \times n$ square matrix is called a square matrix of *order* n.
2. Given a square matrix $\boldsymbol{A} = (a_{ij})$ of order n, the *diagonal* of $\boldsymbol{A}$ is the sequence of entries $a_{11}, a_{22}, \ldots, a_{nn}$. The entries a_{ii} are called the *diagonal entries* while a_{ij}, $i \neq j$, are called *non-diagonal entries.* A square matrix is called a *diagonal matrix* if all its non-diagonal entries are zero, i.e.

$$\boldsymbol{A} = (a_{ij})_{n\times n} \text{ is diagonal} \quad \Leftrightarrow \quad a_{ij} = 0 \text{ whenever } i \neq j.$$

3. A diagonal matrix is called a *scalar matrix* if all its diagonal entries are the same, i.e.

$$\boldsymbol{A} = (a_{ij})_{n\times n} \text{ is scalar} \quad \Leftrightarrow \quad a_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ c & \text{if } i = j \end{cases} \quad \text{for a constant } c.$$

4. A diagonal matrix is called an *identity matrix* if all its diagonal entries are 1. We use $\boldsymbol{I}_n$ to denote the identity matrix of order n. Sometimes we write $\boldsymbol{I}$ instead of $\boldsymbol{I}_n$ when there is no danger of confusion.

5. A matrix with all entries equal to zero is called a *zero matrix.* We denote the $m \times n$ zero matrix by $\mathbf{0}_{m\times n}$, or simply $\mathbf{0}$.

6. A square matrix (a_{ij}) is called *symmetric* if $a_{ij} = a_{ji}$ for all i, j.

7. A square matrix (a_{ij}) is called *upper triangular* if $a_{ij} = 0$ whenever $i > j$; and a square matrix (a_{ij}) is called *lower triangular* if $a_{ij} = 0$ whenever $i < j$. Both upper and lower triangular matrices are called *triangular matrices.*

Example 2.1.8

1. The following are some examples of square matrices.

$$(4)\,, \quad \begin{pmatrix} 0 & 4 \\ 1 & 2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ -1 & 3 & 2 \\ 0 & 0 & 2 \end{pmatrix}, \quad \begin{pmatrix} 2 & 1 & 6 & 2 \\ 0 & 3 & 9 & -1 \\ 1 & 0 & 0 & 0 \\ -2 & 3 & 0 & 1 \end{pmatrix}.$$

2. The following are some examples of diagonal matrices.

$$(4)\,, \quad \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \quad \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

3. The following are some examples of scalar matrices.

$$(4)\,, \quad \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

4. The following are some examples of identity matrices.

$$\boldsymbol{I}_1 = (1)\,, \quad \boldsymbol{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \boldsymbol{I}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{I}_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

5. The following are some examples of zero matrices.

$$\mathbf{0}_{1\times 1} = (0)\,, \quad \mathbf{0}_{2\times 4} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{0}_{4\times 3} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

6. The following are some examples of symmetric matrices.

$$(4), \quad \begin{pmatrix} 0 & 4 \\ 4 & 2 \end{pmatrix}, \quad \begin{pmatrix} 1 & -1 & 0 \\ -1 & 3 & 2 \\ 0 & 2 & 2 \end{pmatrix}, \quad \begin{pmatrix} 2 & 1 & 6 & -2 \\ 1 & 3 & 0 & -1 \\ 6 & 0 & 0 & 0 \\ -2 & -1 & 0 & 1 \end{pmatrix}.$$

7. The following are some examples of triangular matrices. The first three matrices are upper triangular while the first and the last matrices are lower triangular.

$$(4), \quad \begin{pmatrix} 0 & 4 \\ 0 & 2 \end{pmatrix}, \quad \begin{pmatrix} 1 & -1 & 0 \\ 0 & 3 & 2 \\ 0 & 0 & 2 \end{pmatrix}, \quad \begin{pmatrix} 2 & 0 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 6 & 0 & 0 & 0 \\ -2 & -1 & 0 & 1 \end{pmatrix}.$$

Section 2.2 Matrix Operations

Definition 2.2.1 Two matrices are said to be *equal* if they have the same size and their corresponding entries are equal. That is, given $\boldsymbol{A} = (a_{ij})_{m\times n}$ and $\boldsymbol{B} = (b_{ij})_{p\times q}$, $\boldsymbol{A}$ is equal to $\boldsymbol{B}$ if $m = p$, $n = q$ and $a_{ij} = b_{ij}$ for all i, j.

Example 2.2.2 Let $\boldsymbol{A} = \begin{pmatrix} 1 & x \\ 2 & 4 \end{pmatrix}$, $\boldsymbol{B} = \begin{pmatrix} 1 & -1 \\ 2 & 4 \end{pmatrix}$ and $\boldsymbol{C} = \begin{pmatrix} 1 & -1 & 0 \\ 2 & 4 & 0 \end{pmatrix}$ where x is a constant. Then

1. $\boldsymbol{A} = \boldsymbol{B}$ if and only if $x = -1$.
2. $\boldsymbol{A} \neq \boldsymbol{C}$ for all values of x.

Definition 2.2.3 Let $\boldsymbol{A} = (a_{ij})_{m\times n}$, $\boldsymbol{B} = (b_{ij})_{m\times n}$ and c a real constant. We define the matrices $\boldsymbol{A} + \boldsymbol{B}$, $\boldsymbol{A} - \boldsymbol{B}$ and $c\boldsymbol{A}$ as follows:

1. **(Matrix Addition)** $\boldsymbol{A} + \boldsymbol{B} = (a_{ij} + b_{ij})_{m\times n}$.
2. **(Matrix Subtraction)** $\boldsymbol{A} - \boldsymbol{B} = (a_{ij} - b_{ij})_{m\times n}$.
3. **(Scalar Multiplication)** $c\boldsymbol{A} = (ca_{ij})_{m\times n}$, where c is usually called a *scalar*.

Example 2.2.4 Let $\boldsymbol{A} = \begin{pmatrix} 2 & 3 & 4 \\ 4 & 5 & 6 \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} 1 & 2 & 3 \\ -1 & -1 & -1 \end{pmatrix}$.

1. $\boldsymbol{A}+\boldsymbol{B} = \begin{pmatrix} 2+1 & 3+2 & 4+3 \\ 4+(-1) & 5+(-1) & 6+(-1) \end{pmatrix} = \begin{pmatrix} 3 & 5 & 7 \\ 3 & 4 & 5 \end{pmatrix}$.

2. $\boldsymbol{A}-\boldsymbol{B} = \begin{pmatrix} 2-1 & 3-2 & 4-3 \\ 4-(-1) & 5-(-1) & 6-(-1) \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 5 & 6 & 7 \end{pmatrix}$.

3. $4\boldsymbol{A} = \begin{pmatrix} 4\cdot 2 & 4\cdot 3 & 4\cdot 4 \\ 4\cdot 4 & 4\cdot 5 & 4\cdot 6 \end{pmatrix} = \begin{pmatrix} 8 & 12 & 16 \\ 16 & 20 & 24 \end{pmatrix}$.

Remark 2.2.5

1. Given a matrix $\boldsymbol{A}$, we normally use $-\boldsymbol{A}$ to denote the matrix $(-1)\boldsymbol{A}$.
2. The matrix subtraction can be defined using the matrix addition: Given two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ of the same size, $\boldsymbol{A}-\boldsymbol{B}$ is defined to be the matrix $\boldsymbol{A}+(-\boldsymbol{B})$.

Theorem 2.2.6 Let $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ be matrices of the same size and c, d scalars. Then

1. **(Commutative Law for Matrix Addition)** $\boldsymbol{A}+\boldsymbol{B} = \boldsymbol{B}+\boldsymbol{A}$,
2. **(Associative Law for Matrix Addition)** $\boldsymbol{A}+(\boldsymbol{B}+\boldsymbol{C}) = (\boldsymbol{A}+\boldsymbol{B})+\boldsymbol{C}$,
3. $c(\boldsymbol{A}+\boldsymbol{B}) = c\boldsymbol{A}+c\boldsymbol{B}$,
4. $(c+d)\boldsymbol{A} = c\boldsymbol{A}+d\boldsymbol{A}$,
5. $c(d\boldsymbol{A}) = (cd)\boldsymbol{A} = d(c\boldsymbol{A})$,
6. $\boldsymbol{A}+\boldsymbol{0} = \boldsymbol{0}+\boldsymbol{A} = \boldsymbol{A}$,
7. $\boldsymbol{A}-\boldsymbol{A} = \boldsymbol{0}$ and
8. $0\boldsymbol{A} = \boldsymbol{0}$ (in here, 0 on the LHS is the number zero and $\boldsymbol{0}$ on the RHS is a zero matrix).

Proof Since the sizes of all the matrices are the same, to verify the rules above, we only need to show that the (i,j)-entries of the resulting matrix on LHS are equal to the (i,j)-entries of the resulting matrix on RHS. In the following, we illustrate the proof of $\boldsymbol{A}+(\boldsymbol{B}+\boldsymbol{C}) = (\boldsymbol{A}+\boldsymbol{B})+\boldsymbol{C}$:

Let $\boldsymbol{A} = (a_{ij})$, $\boldsymbol{B} = (b_{ij})$ and $\boldsymbol{C} = (c_{ij})$. Then for any i, j,

$$\begin{aligned}
&\text{the } (i,j)\text{-entry of } \boldsymbol{A}+(\boldsymbol{B}+\boldsymbol{C}) \\
&= a_{ij} + [\text{the } (i,j)\text{-entry of } \boldsymbol{B}+\boldsymbol{C}] \\
&= a_{ij} + [b_{ij}+c_{ij}] \\
&= [a_{ij}+b_{ij}] + c_{ij} \qquad \text{(by the Associative Law for real number addition)} \\
&= [\text{the } (i,j)\text{-entry of } \boldsymbol{A}+\boldsymbol{B}] + c_{ij} \\
&= \text{the } (i,j)\text{-entry of } (\boldsymbol{A}+\boldsymbol{B})+\boldsymbol{C}.
\end{aligned}$$

(Proofs of the other parts of the theorem are left as exercises. See Question 2.18.)

Remark 2.2.7 Let $\boldsymbol{A_1}$, $\boldsymbol{A_2}$, ..., $\boldsymbol{A_k}$ be matrices of the same size. By the Associative Law for Matrix Addition, we can write $\boldsymbol{A_1} + \boldsymbol{A_2} + \cdots + \boldsymbol{A_k}$ to represent the sum of the matrices without using any parentheses to indicate the order of the matrix additions.

Definition 2.2.8 (Matrix Multiplication) Let $\boldsymbol{A} = (a_{ij})_{m\times p}$ and $\boldsymbol{B} = (b_{ij})_{p\times n}$ be two matrices. The *product* $\boldsymbol{AB}$ is defined to be an $m \times n$ matrix whose (i, j)-entry is

$$a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ip}b_{pj} = \sum_{k=1}^{p} a_{ik}b_{kj}$$

for $i = 1, 2, \dots, m$ and $j = 1, 2, \dots, n$. (See Question 2.3 for the meaning of the notation $\sum$.)

Example 2.2.9

1. $$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & 3 \\ -1 & -2 \end{pmatrix} = \begin{pmatrix} 1\cdot 1 + 2\cdot 2 + 3\cdot(-1) & 1\cdot 1 + 2\cdot 3 + 3\cdot(-2) \\ 4\cdot 1 + 5\cdot 2 + 6\cdot(-1) & 4\cdot 1 + 5\cdot 3 + 6\cdot(-2) \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 8 & 7 \end{pmatrix}.$$

2. $$\begin{pmatrix} 1 & 1 \\ 2 & 3 \\ -1 & -2 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} 1\cdot 1 + 1\cdot 4 & 1\cdot 2 + 1\cdot 5 & 1\cdot 3 + 1\cdot 6 \\ 2\cdot 1 + 3\cdot 4 & 2\cdot 2 + 3\cdot 5 & 2\cdot 3 + 3\cdot 6 \\ (-1)\cdot 1 + (-2)\cdot 4 & (-1)\cdot 2 + (-2)\cdot 5 & (-1)\cdot 3 + (-2)\cdot 6 \end{pmatrix} = \begin{pmatrix} 5 & 7 & 9 \\ 14 & 19 & 24 \\ -9 & -12 & -15 \end{pmatrix}.$$

Remark 2.2.10

1. We can only multiply two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ (in the manner $\boldsymbol{AB}$) when the number of columns of $\boldsymbol{A}$ is equal to the number of rows of $\boldsymbol{B}$.

2. The matrix multiplication is not commutative, i.e. in general, $\boldsymbol{AB}$ and $\boldsymbol{BA}$ are two different matrices even if the products exist.

   For example, let $\boldsymbol{A} = \begin{pmatrix} -1 & 0 \\ 2 & 3 \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} 1 & 2 \\ 3 & 0 \end{pmatrix}$. Then

   $$\boldsymbol{AB} = \begin{pmatrix} -1 & -2 \\ 11 & 4 \end{pmatrix} \quad \text{and} \quad \boldsymbol{BA} = \begin{pmatrix} 3 & 6 \\ -3 & 0 \end{pmatrix}$$

   and hence $\boldsymbol{AB} \neq \boldsymbol{BA}$.

3. In view of Part 2 above, it would be ambiguous to say "the multiplication of a matrix $\boldsymbol{A}$ to another matrix $\boldsymbol{B}$" since it could mean $\boldsymbol{AB}$ or $\boldsymbol{BA}$. To distinguish the two, we refer to $\boldsymbol{AB}$ as the *pre-multiplication* of $\boldsymbol{A}$ to $\boldsymbol{B}$ and $\boldsymbol{BA}$ as the *post-multiplication* of $\boldsymbol{A}$ to $\boldsymbol{B}$.

4. $\boldsymbol{AB} = \boldsymbol{0}$ does not imply $\boldsymbol{A} = \boldsymbol{0}$ or $\boldsymbol{B} = \boldsymbol{0}$.

   For example, let $\boldsymbol{A} = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$. We have $\boldsymbol{A} \neq \boldsymbol{0}$ and $\boldsymbol{B} \neq \boldsymbol{0}$ while

$$\boldsymbol{AB} = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = \boldsymbol{0}.$$

Theorem 2.2.11

1. **(Associative Law for Matrix Multiplication)**
   If $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ are $m \times p$, $p \times q$ and $q \times n$ matrices respectively, then

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

2. **(Distributive Laws for Matrix Addition and Multiplication)**
   If $\boldsymbol{A}$, $\boldsymbol{B_1}$ and $\boldsymbol{B_2}$ are $m \times p$, $p \times n$ and $p \times n$ matrices respectively, then

$$\boldsymbol{A}(\boldsymbol{B_1} + \boldsymbol{B_2}) = \boldsymbol{AB_1} + \boldsymbol{AB_2}.$$

   If $\boldsymbol{A}$, $\boldsymbol{C_1}$ and $\boldsymbol{C_2}$ are $p \times n$, $m \times p$ and $m \times p$ matrices respectively, then

$$(\boldsymbol{C_1} + \boldsymbol{C_2})\boldsymbol{A} = \boldsymbol{C_1 A} + \boldsymbol{C_2 A}.$$

3. If $\boldsymbol{A}$, $\boldsymbol{B}$ are $m \times p$, $p \times n$ matrices, respectively, and c is a scalar, then

$$c(\boldsymbol{AB}) = (c\boldsymbol{A})\boldsymbol{B} = \boldsymbol{A}(c\boldsymbol{B}).$$

4. If $\boldsymbol{A}$ is an $m \times n$ matrix, then

$$\boldsymbol{A0}_{n\times q} = \boldsymbol{0}_{m\times q}, \quad \boldsymbol{0}_{p\times m}\boldsymbol{A} = \boldsymbol{0}_{p\times n} \quad \text{and} \quad \boldsymbol{AI}_n = \boldsymbol{I}_m\boldsymbol{A} = \boldsymbol{A}.$$

Proof To prove the matrix identities above, we need to check that

(i) the size of the resulting matrix on the LHS is equal to that on the RHS, and

(ii) the (i, j)-entries of the resulting matrix on the LHS are equal to those on the RHS.

In the following, we illustrate the proof of $\boldsymbol{A}(\boldsymbol{B_1} + \boldsymbol{B_2}) = \boldsymbol{AB_1} + \boldsymbol{AB_2}$:

(i) Since the size of $\boldsymbol{A}$ is $m \times p$ and the size of $\boldsymbol{B_1} + \boldsymbol{B_2}$ is $p \times n$, the size of $\boldsymbol{A}(\boldsymbol{B_1} + \boldsymbol{B_2})$ is $m \times n$. On the other hand, both the sizes of $\boldsymbol{AB_1}$ and $\boldsymbol{AB_2}$ are $m \times n$ and hence the size of $\boldsymbol{AB_1} + \boldsymbol{AB_2}$ is $m \times n$.

Thus the sizes of the resulting matrices on both sides of the identity are the same.

(ii) Let $\boldsymbol{A} = (a_{ij})$, $\boldsymbol{B_1} = (b_{ij})$ and $\boldsymbol{B_2} = (b'_{ij})$. For any i, j,

$$\begin{aligned}
&\text{the } (i,j)\text{-entry of } \boldsymbol{A}(\boldsymbol{B_1}+\boldsymbol{B_2}) \\
&= a_{i1}[\text{the } (1,j)\text{-entry of } \boldsymbol{B_1}+\boldsymbol{B_2}] + a_{i2}[\text{the } (2,j)\text{-entry of } \boldsymbol{B_1}+\boldsymbol{B_2}] \\
&\qquad + \cdots + a_{ip}[\text{the } (p,j)\text{-entry of } \boldsymbol{B_1}+\boldsymbol{B_2}] \\
&= a_{i1}[b_{1j}+b'_{1j}] + a_{i2}[b_{2j}+b'_{2j}] + \cdots + a_{ip}[b_{pj}+b'_{pj}] \\
&= [a_{i1}b_{1j}+a_{i1}b'_{1j}] + [a_{i2}b_{2j}+a_{i2}b'_{2j}] + \cdots + [a_{ip}b_{pj}+a_{ip}b'_{pj}] \\
&= [a_{i1}b_{1j}+a_{i2}b_{2j}+\cdots+a_{ip}b_{pj}] + [a_{i1}b'_{1j}+a_{i2}b'_{2j}+\cdots+a_{ip}b'_{pj}] \\
&= [\text{the } (i,j)\text{-entry of } \boldsymbol{AB_1}] + [\text{the } (i,j)\text{-entry of } \boldsymbol{AB_2}] \\
&= \text{the } (i,j)\text{-entry of } \boldsymbol{AB_1}+\boldsymbol{AB_2}.
\end{aligned}$$

By (i) and (ii), we have proved $\boldsymbol{A}(\boldsymbol{B_1}+\boldsymbol{B_2}) = \boldsymbol{AB_1}+\boldsymbol{AB_2}$.

(Proofs of the other parts of the theorem are left as exercises. See Question 2.19.)

Definition 2.2.12 (Powers of Square Matrices) Let $\boldsymbol{A}$ be a square matrix and n a nonnegative integer. We define $\boldsymbol{A}^n$ as follows:

$$\boldsymbol{A}^n = \begin{cases} \boldsymbol{I} & \text{if } n = 0 \\ \underbrace{\boldsymbol{AA}\cdots\boldsymbol{A}}_{n \text{ times}} & \text{if } n \geq 1. \end{cases}$$

Example 2.2.13 Let $\boldsymbol{A} = \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix}$. Then

$$\boldsymbol{A}^3 = \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 11 & 30 \\ 15 & 41 \end{pmatrix}.$$

Remark 2.2.14

1. Let $\boldsymbol{A}$ be a square matrix and m, n nonnegative integers. Then $\boldsymbol{A}^m\boldsymbol{A}^n = \boldsymbol{A}^{m+n}$.

2. Since the matrix multiplication is not commutative, in general, for two square matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ of the same order, $(\boldsymbol{AB})^n$ may not be equal to $\boldsymbol{A}^n\boldsymbol{B}^n$.

   For example, let $\boldsymbol{A} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Then $(\boldsymbol{AB})^2 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ and $\boldsymbol{A}^2\boldsymbol{B}^2 = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}$ and hence $(\boldsymbol{AB})^2 \neq \boldsymbol{A}^2\boldsymbol{B}^2$.

Notation 2.2.15 Let

$$\boldsymbol{A} = (a_{ij})_{m\times p} = \begin{pmatrix} \boldsymbol{a_1} \\ \boldsymbol{a_2} \\ \vdots \\ \boldsymbol{a_m} \end{pmatrix}, \quad \text{where } \boldsymbol{a_i} = \begin{pmatrix} a_{i1} & a_{i2} & \cdots & a_{ip} \end{pmatrix} \text{ is the } i\text{th row of } \boldsymbol{A},$$

and let

$$\boldsymbol{B} = (b_{ij})_{p\times n} = \begin{pmatrix} \boldsymbol{b_1} & \boldsymbol{b_2} & \cdots & \boldsymbol{b_n} \end{pmatrix}, \quad \text{where } \boldsymbol{b_j} = \begin{pmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{pj} \end{pmatrix} \text{ is the } j\text{th column of } \boldsymbol{B}.$$

Then

$$\boldsymbol{AB} = \begin{pmatrix} \boldsymbol{a_1b_1} & \boldsymbol{a_1b_2} & \cdots & \boldsymbol{a_1b_n} \\ \boldsymbol{a_2b_1} & \boldsymbol{a_2b_2} & \cdots & \boldsymbol{a_2b_n} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{a_mb_1} & \boldsymbol{a_mb_2} & \cdots & \boldsymbol{a_mb_n} \end{pmatrix} \tag{2.1}$$

where $\boldsymbol{a_ib_j} = \begin{pmatrix} a_{i1} & a_{i2} & \cdots & a_{ip} \end{pmatrix} \begin{pmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{pj} \end{pmatrix} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ip}b_{pj}$ is the (i, j)-entry of $\boldsymbol{AB}$. Also we can write

$$\boldsymbol{AB} = \begin{pmatrix} \boldsymbol{Ab_1} & \boldsymbol{Ab_2} & \cdots & \boldsymbol{Ab_n} \end{pmatrix}, \tag{2.2}$$

where $\boldsymbol{Ab_j}$ is the jth column of $\boldsymbol{AB}$, or

$$\boldsymbol{AB} = \begin{pmatrix} \boldsymbol{a_1B} \\ \boldsymbol{a_2B} \\ \vdots \\ \boldsymbol{a_mB} \end{pmatrix}, \tag{2.3}$$

where $\boldsymbol{a_iB}$ is the ith row of $\boldsymbol{AB}$. (See Question 2.23 for other ways of multiplying matrices in blocks.)

Example 2.2.16 Let

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}, \quad \boldsymbol{a_1} = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix}, \quad \boldsymbol{a_2} = \begin{pmatrix} 4 & 5 & 6 \end{pmatrix},$$

$$\boldsymbol{B} = \begin{pmatrix} 1 & 1 \\ 2 & 3 \\ -1 & -2 \end{pmatrix}, \quad \boldsymbol{b_1} = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}, \quad \boldsymbol{b_2} = \begin{pmatrix} 1 \\ 3 \\ -2 \end{pmatrix}.$$

Note that $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 \end{pmatrix}$. We work out the RHS of Equations (2.1), (2.2) and (2.3) in Notation 2.2.15:

$$\begin{pmatrix} \boldsymbol{a}_1\boldsymbol{b}_1 & \boldsymbol{a}_1\boldsymbol{b}_2 \\ \boldsymbol{a}_2\boldsymbol{b}_1 & \boldsymbol{a}_2\boldsymbol{b}_2 \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} 1 & 2 & 3 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} & \begin{pmatrix} 1 & 2 & 3 \end{pmatrix}\begin{pmatrix} 1 \\ 3 \\ -2 \end{pmatrix} \\ \begin{pmatrix} 4 & 5 & 6 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} & \begin{pmatrix} 4 & 5 & 6 \end{pmatrix}\begin{pmatrix} 1 \\ 3 \\ -2 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 8 & 7 \end{pmatrix},$$

$$\begin{pmatrix} \boldsymbol{A}\boldsymbol{b}_1 & \boldsymbol{A}\boldsymbol{b}_2 \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} & \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}\begin{pmatrix} 1 \\ 3 \\ -2 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 8 & 7 \end{pmatrix},$$

$$\begin{pmatrix} \boldsymbol{a}_1\boldsymbol{B} \\ \boldsymbol{a}_2\boldsymbol{B} \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} 1 & 2 & 3 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 2 & 3 \\ -1 & -2 \end{pmatrix} \\ \begin{pmatrix} 4 & 5 & 6 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 2 & 3 \\ -1 & -2 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 8 & 7 \end{pmatrix}.$$

All three forms give us the matrix $\boldsymbol{AB} = \begin{pmatrix} 2 & 1 \\ 8 & 7 \end{pmatrix}$.

Remark 2.2.17 Let $\boldsymbol{A} = (a_{ij})_{m\times n}$, $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ and $\boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$.

The system of linear equations

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \quad\vdots \qquad\qquad\qquad\qquad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

can be written as

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

or

$$\boldsymbol{Ax} = \boldsymbol{b}.$$

The matrix $\boldsymbol{A}$ is called the *coefficient matrix*, $\boldsymbol{x}$ the *variable matrix* and $\boldsymbol{b}$ the *constant matrix* of the linear system. Furthermore, a solution $x_1 = u_1$, $x_2 = u_2$, $\ldots$, $x_n = u_n$ to the linear system can be represented by an $n \times 1$ matrix

$$\boldsymbol{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix},$$

i.e. $\boldsymbol{u}$ is said to be a *solution* to the linear system $\boldsymbol{Ax} = \boldsymbol{b}$ if $\boldsymbol{Au} = \boldsymbol{b}$.

Let $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{c_1} & \boldsymbol{c_2} & \cdots & \boldsymbol{c_n} \end{pmatrix}$ where $\boldsymbol{c_j}$ is the jth column of $\boldsymbol{A}$. The linear system can also be written as

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

i.e. $x_1\boldsymbol{c_1} + x_2\boldsymbol{c_2} + \cdots + x_n\boldsymbol{c_n} = \boldsymbol{b}$ or $\displaystyle\sum_{j=1}^{n} x_j\boldsymbol{c_j} = \boldsymbol{b}$.

Example 2.2.18 The system of linear equations

$$\begin{cases} 4x + 5y + 6z = 5 \\ x - y = 2 \\ y - z = 3 \end{cases}$$

can be written as

$$\begin{pmatrix} 4 & 5 & 6 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \\ 3 \end{pmatrix}$$

or

$$x \begin{pmatrix} 4 \\ 1 \\ 0 \end{pmatrix} + y \begin{pmatrix} 5 \\ -1 \\ 1 \end{pmatrix} + z \begin{pmatrix} 6 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \\ 3 \end{pmatrix}.$$

Also, $\begin{pmatrix} 3 \\ 1 \\ -2 \end{pmatrix}$ is the only solution to the system. (Check it.)

Definition 2.2.19 Let $\boldsymbol{A} = (a_{ij})$ be an $m \times n$ matrix. The *transpose* of $\boldsymbol{A}$, denoted by $\boldsymbol{A}^{\mathrm{T}}$ (or $\boldsymbol{A}^{\mathrm{t}}$), is the $n \times m$ matrix whose (i, j)-entry is a_{ji}.

Example 2.2.20 Let $\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{pmatrix}$. Then $\boldsymbol{A}^{\mathrm{T}} = \begin{pmatrix} 1 & 5 \\ 2 & 6 \\ 3 & 7 \\ 4 & 8 \end{pmatrix}$.

Remark 2.2.21

1. For a matrix $\boldsymbol{A}$, the rows of $\boldsymbol{A}$ is the columns of $\boldsymbol{A}^{\mathrm{T}}$ and vice versa.
2. In Definition 2.1.7.6, a square matrix $\boldsymbol{A} = (a_{ij})$ is called symmetric if $a_{ij} = a_{ji}$ for all i, j. Thus a square matrix $\boldsymbol{A}$ is symmetric if and only if $\boldsymbol{A} = \boldsymbol{A}^{\mathrm{T}}$.

Theorem 2.2.22 Let $\boldsymbol{A}$ be an $m \times n$ matrix.

1. $(\boldsymbol{A}^{\mathrm{T}})^{\mathrm{T}} = \boldsymbol{A}$.
2. If $\boldsymbol{B}$ is an $m \times n$ matrix, then $(\boldsymbol{A} + \boldsymbol{B})^{\mathrm{T}} = \boldsymbol{A}^{\mathrm{T}} + \boldsymbol{B}^{\mathrm{T}}$.
3. If c is a scalar, then $(c\boldsymbol{A})^{\mathrm{T}} = c\boldsymbol{A}^{\mathrm{T}}$.
4. If $\boldsymbol{B}$ is an $n \times p$ matrix, then $(\boldsymbol{AB})^{\mathrm{T}} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}$.

Proof In the following, we illustrate the proof of $(\boldsymbol{AB})^{\mathrm{T}} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}$:

(i) Since the size of $\boldsymbol{AB}$ is $m \times p$, the size of $(\boldsymbol{AB})^{\mathrm{T}}$ is $p \times m$. On the other hand, since the size of $\boldsymbol{B}^{\mathrm{T}}$ is $p \times n$ and $\boldsymbol{A}^{\mathrm{T}}$ is $n \times m$, the size of $\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}$ is $p \times m$. Thus the sizes of the resulting matrices on both sides of the identity are the same.

(ii) Let $\boldsymbol{A} = (a_{ij})$ and $\boldsymbol{B} = (b_{ij})$. The (i, j)-entry of $\boldsymbol{AB}$ is $a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}$. So by Definition 2.2.19,

$$\begin{aligned} \text{the } (i,j)\text{-entry of } (\boldsymbol{AB})^{\mathrm{T}} &= \text{the } (j,i)\text{-entry of } \boldsymbol{AB} \\ &= a_{j1}b_{1i} + a_{j2}b_{2i} + \cdots + a_{jn}b_{ni}. \end{aligned}$$

On the other hand, let $\boldsymbol{A}^{\mathrm{T}} = (a'_{ij})$ and $\boldsymbol{B}^{\mathrm{T}} = (b'_{ij})$. By Definition 2.2.19, $a'_{ij} = a_{ji}$ and $b'_{ij} = b_{ji}$. We have

$$\begin{aligned} \text{the } (i,j)\text{-entry of } \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}} &= b'_{i1}a'_{1j} + b'_{i2}a'_{2j} + \cdots + b'_{in}a'_{nj} \\ &= b_{1i}a_{j1} + b_{2i}a_{j2} + \cdots + b_{ni}a_{jn} \\ &= a_{j1}b_{1i} + a_{j2}b_{2i} + \cdots + a_{jn}b_{ni}. \end{aligned}$$

Hence the (i, j)-entry of $(\boldsymbol{AB})^{\mathrm{T}} =$ the (i, j)-entry of $\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}$.

By (i) and (ii), we have proved $(\boldsymbol{AB})^{\mathrm{T}} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}$.

(Proofs of the other parts of the theorem are left as exercises. See Question 2.20.)

Section 2.3 Inverses of Square Matrices

Discussion 2.3.1 Let a, b be two real numbers such that $a \neq 0$. Then the solution to the equation $ax = b$ is $x = \frac{b}{a} = a^{-1}b$. Now, let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two matrices. It is much harder to solve the matrix equation $\boldsymbol{AX} = \boldsymbol{B}$ because we do not have "division" for matrices. However, for some square matrices, we can find their "inverses" which have the similar property as a^{-1} in the computation of the solution to $ax = b$ above.

Definition 2.3.2 Let $\boldsymbol{A}$ be a square matrix of order n. Then $\boldsymbol{A}$ is said to be *invertible* if there exists a square matrix $\boldsymbol{B}$ of order n such that $\boldsymbol{AB} = \boldsymbol{I}$ and $\boldsymbol{BA} = \boldsymbol{I}$. Such a matrix $\boldsymbol{B}$ is called an *inverse* of $\boldsymbol{A}$. A square matrix is called *singular* if it has no inverse.

Example 2.3.3

1. Let $\boldsymbol{A} = \begin{pmatrix} 2 & -5 \\ -1 & 3 \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} 3 & 5 \\ 1 & 2 \end{pmatrix}$. Then

$$\boldsymbol{AB} = \begin{pmatrix} 2 & -5 \\ -1 & 3 \end{pmatrix}\begin{pmatrix} 3 & 5 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \boldsymbol{I}$$

and

$$\boldsymbol{BA} = \begin{pmatrix} 3 & 5 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 2 & -5 \\ -1 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \boldsymbol{I}.$$

So $\boldsymbol{A}$ is invertible and $\boldsymbol{B}$ is an inverse of $\boldsymbol{A}$.

2. Consider the matrix equation

$$\begin{pmatrix} 2 & -5 \\ -1 & 3 \end{pmatrix}\boldsymbol{X} = \begin{pmatrix} 4 \\ 0 \end{pmatrix}.$$

By pre-multiplying the matrix $\begin{pmatrix} 3 & 5 \\ 1 & 2 \end{pmatrix}$, which is an inverse of $\begin{pmatrix} 2 & -5 \\ -1 & 3 \end{pmatrix}$, to both sides of the equation, we have

$$\begin{pmatrix} 3 & 5 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 2 & -5 \\ -1 & 3 \end{pmatrix}\boldsymbol{X} = \begin{pmatrix} 3 & 5 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 4 \\ 0 \end{pmatrix} \quad \Rightarrow \quad \boldsymbol{IX} = \begin{pmatrix} 12 \\ 4 \end{pmatrix} \quad \Rightarrow \quad \boldsymbol{X} = \begin{pmatrix} 12 \\ 4 \end{pmatrix}.$$

3. Show that $\boldsymbol{A} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$ is singular.

Solution (Proof by Contradiction) Suppose on the contrary $\boldsymbol{A}$ has an inverse $\boldsymbol{B} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. By the definition of inverses, we have

$$\boldsymbol{BA} = \boldsymbol{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

On the other hand,

$$\boldsymbol{BA} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} a+b & 0 \\ c+d & 0 \end{pmatrix}.$$

Thus $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a+b & 0 \\ c+d & 0 \end{pmatrix}$ which is impossible.

(In Section 2.4, we shall see a systematic method to check whether a square matrix is invertible. See Remark 2.4.10.)

Remark 2.3.4

1. **(Cancellation Law for Matrices)** Let $\boldsymbol{A}$ be an invertible $m \times m$ matrix.
   (a) If $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ are $m \times n$ matrices with $\boldsymbol{AB}_1 = \boldsymbol{AB}_2$, then $\boldsymbol{B}_1 = \boldsymbol{B}_2$.
   (b) If $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ are $n \times m$ matrices with $\boldsymbol{C}_1\boldsymbol{A} = \boldsymbol{C}_2\boldsymbol{A}$, then $\boldsymbol{C}_1 = \boldsymbol{C}_2$.
2. If the matrix $\boldsymbol{A}$ in Part 1 is singular, then the Cancellation Law may not hold. For example, take $\boldsymbol{A} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$, $\boldsymbol{B}_1 = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$ and $\boldsymbol{B}_2 = \begin{pmatrix} 2 & 0 \\ 1 & 3 \end{pmatrix}$. Then $\boldsymbol{AB}_1 = \boldsymbol{AB}_2 = \begin{pmatrix} 2 & 0 \\ 2 & 0 \end{pmatrix}$ but $\boldsymbol{B}_1 \neq \boldsymbol{B}_2$. (As an exercise, find two 2×2 matrices $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ such that $\boldsymbol{C}_1\boldsymbol{A} = \boldsymbol{C}_2\boldsymbol{A}$ but $\boldsymbol{C}_1 \neq \boldsymbol{C}_2$.)

Theorem 2.3.5 (Uniqueness of Inverses) If $\boldsymbol{B}$ and $\boldsymbol{C}$ are inverses of a square matrix $\boldsymbol{A}$, then $\boldsymbol{B} = \boldsymbol{C}$.

Proof Since $\boldsymbol{B}$ is an inverse of $\boldsymbol{A}$, we have

$$\boldsymbol{BA} = \boldsymbol{I} \quad \text{and} \quad \boldsymbol{AB} = \boldsymbol{I}.$$

Also, since $\boldsymbol{C}$ is an inverse of $\boldsymbol{A}$, we have

$$\boldsymbol{CA} = \boldsymbol{I} \quad \text{and} \quad \boldsymbol{AC} = \boldsymbol{I}.$$

So $\boldsymbol{AB} = \boldsymbol{I} \Rightarrow \boldsymbol{CAB} = \boldsymbol{CI} \Rightarrow \boldsymbol{IB} = \boldsymbol{C} \Rightarrow \boldsymbol{B} = \boldsymbol{C}$.

Notation 2.3.6 Let $\boldsymbol{A}$ be an invertible matrix. By Theorem 2.3.5, we know that there is only one inverse of $\boldsymbol{A}$ and we use the symbol $\boldsymbol{A}^{-1}$ to denote this unique inverse of $\boldsymbol{A}$.

Remark 2.3.7 If we are asked to show that $\boldsymbol{A}^{-1} = \boldsymbol{B}$ where $\boldsymbol{A}$ and $\boldsymbol{B}$ are square matrices of the same size, what we need to do is to check that $\boldsymbol{AB} = \boldsymbol{I}$ and $\boldsymbol{BA} = \boldsymbol{I}$. Actually, in the next section, we shall learn that we only need to check any one of the two conditions. (See Theorem 2.4.12.)

Example 2.3.8 Let $\boldsymbol{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. If $ad - bc \neq 0$, show that $\boldsymbol{A}$ is invertible and

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

Solution Let $\boldsymbol{B} = \begin{pmatrix} \frac{d}{ad-bc} & \frac{-b}{ad-bc} \\ \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{pmatrix}$. Note that $\boldsymbol{B}$ is well-defined if $ad - bc \neq 0$. It suffices to show that $\boldsymbol{B}$ is the inverse of $\boldsymbol{A}$. By Remark 2.3.7, we need to check that $\boldsymbol{AB} = \boldsymbol{I}$ and $\boldsymbol{BA} = \boldsymbol{I}$:

$$\boldsymbol{AB} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} \frac{d}{ad-bc} & \frac{-b}{ad-bc} \\ \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{pmatrix} = \begin{pmatrix} \frac{ad-bc}{ad-bc} & \frac{-ab+ba}{ad-bc} \\ \frac{cd-dc}{ad-bc} & \frac{-cb+da}{ad-bc} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \boldsymbol{I},$$

$$\boldsymbol{BA} = \begin{pmatrix} \frac{d}{ad-bc} & \frac{-b}{ad-bc} \\ \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \frac{ad-bc}{da-bc} & \frac{-db+bd}{ad-bc} \\ \frac{-ca+ac}{ad-bc} & \frac{-cb+ad}{ad-bc} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \boldsymbol{I}.$$

Thus $\boldsymbol{A}$ is invertible and $\boldsymbol{A}^{-1} = \boldsymbol{B}$.

Theorem 2.3.9 Let $\boldsymbol{A}, \boldsymbol{B}$ be two invertible matrices of the same size and c a nonzero scalar. Then

1. $c\boldsymbol{A}$ is invertible and $(c\boldsymbol{A})^{-1} = \frac{1}{c}\boldsymbol{A}^{-1}$.
2. $\boldsymbol{A}^{\mathrm{T}}$ is invertible and $(\boldsymbol{A}^{\mathrm{T}})^{-1} = (\boldsymbol{A}^{-1})^{\mathrm{T}}$.
3. $\boldsymbol{A}^{-1}$ is invertible and $(\boldsymbol{A}^{-1})^{-1} = \boldsymbol{A}$.
4. $\boldsymbol{AB}$ is invertible and $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$.

Proof In the following, we only illustrate the proof of Part 2 of the theorem:

To show that $\boldsymbol{A}^{\mathrm{T}}$ is invertible, we only need to verify that $(\boldsymbol{A}^{-1})^{\mathrm{T}}$ is the inverse of $\boldsymbol{A}^{\mathrm{T}}$. Note that

$$\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}^{-1})^{\mathrm{T}} = (\boldsymbol{A}^{-1}\boldsymbol{A})^{\mathrm{T}} = \boldsymbol{I}^{\mathrm{T}} = \boldsymbol{I} \quad \text{and} \quad (\boldsymbol{A}^{-1})^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}} = (\boldsymbol{A}\boldsymbol{A}^{-1})^{\mathrm{T}} = \boldsymbol{I}^{\mathrm{T}} = \boldsymbol{I}.$$

By Remark 2.3.7, we have proved that $\boldsymbol{A}^{\mathrm{T}}$ is invertible and $(\boldsymbol{A}^{\mathrm{T}})^{-1} = (\boldsymbol{A}^{-1})^{\mathrm{T}}$.

(Proofs of the other parts of the theorem are left as exercises. See Question 2.30.)

Remark 2.3.10 By Theorem 2.3.9.4, if $\boldsymbol{A_1}, \boldsymbol{A_2}, \ldots, \boldsymbol{A_k}$ are invertible matrices of the same size, then $\boldsymbol{A_1}\boldsymbol{A_2}\cdots\boldsymbol{A_k}$ is invertible and $(\boldsymbol{A_1}\boldsymbol{A_2}\cdots\boldsymbol{A_k})^{-1} = \boldsymbol{A_k}^{-1}\cdots\boldsymbol{A_2}^{-1}\boldsymbol{A_1}^{-1}$.

Definition 2.3.11 (Negative Powers of Square Matrices) Let $\boldsymbol{A}$ be an invertible matrix and n a positive integer. We define $\boldsymbol{A}^{-n}$ as follows:

$$\boldsymbol{A}^{-n} = (\boldsymbol{A}^{-1})^n = \underbrace{\boldsymbol{A}^{-1}\boldsymbol{A}^{-1}\cdots\boldsymbol{A}^{-1}}_{n \text{ times}}.$$

Example 2.3.12 Let $\boldsymbol{A} = \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix}$. By Example 2.3.8, $\boldsymbol{A}^{-1} = \begin{pmatrix} 3 & -2 \\ -1 & 1 \end{pmatrix}$. Then

$$\boldsymbol{A}^{-3} = (\boldsymbol{A}^{-1})^3 = \begin{pmatrix} 3 & -2 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 3 & -2 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 3 & -2 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 41 & -30 \\ -15 & 11 \end{pmatrix}.$$

Remark 2.3.13 Let $\boldsymbol{A}$ be an invertible matrix.

1. $\boldsymbol{A}^r\boldsymbol{A}^s = \boldsymbol{A}^{r+s}$ for any integers r and s.
2. $\boldsymbol{A}^n$ is invertible and $(\boldsymbol{A}^n)^{-1} = \boldsymbol{A}^{-n}$ for any integer n.

Section 2.4 Elementary Matrices

Definition 2.4.1 In Definition 1.2.4, Definition 1.2.6, Definition 1.3.1, Definition 1.4.1, Algorithm 1.4.2 and Algorithm 1.4.3, the concepts of *elementary row operations*, *row equivalent matrices*, *row-echelon forms*, *reduced row-echelon forms*, *Gaussian Elimination*, *Gauss-Jordan Elimination* are defined for augmented matrices. Form now on, these terms will also be used for matrices.

Discussion 2.4.2 Consider the three types of elementary row operations on matrices.

1. **Multiply a row by a constant:**

   Let $\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 2 & 3 \\ 2 & -1 & 3 & 6 \\ 1 & 4 & 4 & 0 \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} 1 & 0 & 2 & 3 \\ 4 & -2 & 6 & 12 \\ 1 & 4 & 4 & 0 \end{pmatrix}$. Note that $\boldsymbol{B}$ is obtained from $\boldsymbol{A}$ by multiplying the second row of $\boldsymbol{A}$ by 2.

   Let $\boldsymbol{E}_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. Observe that

$$\boldsymbol{E}_1\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 2 & 3 \\ 2 & -1 & 3 & 6 \\ 1 & 4 & 4 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 2 & 3 \\ 4 & -2 & 6 & 12 \\ 1 & 4 & 4 & 0 \end{pmatrix} = \boldsymbol{B}.$$

In general, let $\boldsymbol{A}$ be an $m \times n$ matrix and

$$\boldsymbol{E} = \left(\begin{array}{ccc|c|ccc} 1 &&&&&& \\ & \ddots &&&& 0 & \\ && 1 &&&& \\ \hline &&& k &&& \\ \hline &&&& 1 && \\ & 0 &&&& \ddots & \\ &&&&&& 1 \end{array}\right) \begin{array}{l} \\ \\ \\ \leftarrow i\text{th row} \\ \\ \\ \\ \end{array}$$

$$\uparrow$$
$$i\text{th column}$$

be a square matrix of order m where all entries not specified are zero. Then $\boldsymbol{EA}$ is the matrix which can be obtained from $\boldsymbol{A}$ by multiplying the ith row of $\boldsymbol{A}$ by k.

If $k \neq 0$, the matrix $\boldsymbol{E}$ is invertible and

$$\boldsymbol{E}^{-1} = \left(\begin{array}{ccc|c|ccc} 1 &&&&&& \\ & \ddots &&&& 0 & \\ && 1 &&&& \\ \hline &&& \frac{1}{k} &&& \\ \hline &&&& 1 && \\ & 0 &&&& \ddots & \\ &&&&&& 1 \end{array}\right) \begin{array}{l} \\ \\ \\ \leftarrow i\text{th row} \\ \\ \\ \\ \end{array}$$

$$\uparrow$$
$$i\text{th column}$$

corresponds to the row operation of multiplying the ith row of a matrix by $\frac{1}{k}$.

2. **Interchange two rows:**

Let $\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 2 & 3 \\ 2 & -1 & 3 & 6 \\ 1 & 4 & 4 & 0 \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} 1 & 0 & 2 & 3 \\ 1 & 4 & 4 & 0 \\ 2 & -1 & 3 & 6 \end{pmatrix}$. Note that $\boldsymbol{B}$ is obtained from $\boldsymbol{A}$ by interchanging the second and third rows of $\boldsymbol{A}$.

Let $\boldsymbol{E_2} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$. Observe that

$$\boldsymbol{E_2A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 2 & 3 \\ 2 & -1 & 3 & 6 \\ 1 & 4 & 4 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 2 & 3 \\ 1 & 4 & 4 & 0 \\ 2 & -1 & 3 & 6 \end{pmatrix} = \boldsymbol{B}.$$

In general, let $\boldsymbol{A}$ be an $m \times n$ matrix and

$$\boldsymbol{E} = \begin{array}{cl}
\left(\begin{array}{ccc|c|ccc|c|ccc}
1 &&&&&&&&&& \\
& \ddots &&&&&&&& \mathbf{0} & \\
&& 1 &&&&&&&& \\ \hline
&&& 0 &&&& 1 &&& \\ \hline
&&&& 1 &&&&&& \\
&&&&& \ddots &&&&& \\
&&&&&& 1 &&&& \\ \hline
&&& 1 &&&& 0 &&& \\ \hline
&&&&&&&& 1 && \\
& \mathbf{0} &&&&&&&& \ddots & \\
&&&&&&&&&& 1
\end{array}\right)
&
\begin{array}{l}
\\ \\ \\ \leftarrow i\text{th row} \\ \\ \\ \\ \leftarrow j\text{th row} \\ \\ \\ \\
\end{array}
\\
\begin{array}{cc}
\uparrow & \uparrow \\
i\text{th column} & j\text{th column}
\end{array}
&
\end{array}$$

be a square matrix of order m where all entries not specified are zero. Then $\boldsymbol{EA}$ is the matrix which can be obtained from $\boldsymbol{A}$ by interchanging the ith and jth rows of $\boldsymbol{A}$.

The matrix $\boldsymbol{E}$ is invertible and $\boldsymbol{E}^{-1} = \boldsymbol{E}$.

3. **Add a multiple of a row to another row:**

Let $\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 2 & 3 \\ 2 & -1 & 3 & 6 \\ 1 & 4 & 4 & 0 \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} 1 & 0 & 2 & 3 \\ 2 & -1 & 3 & 6 \\ 3 & 4 & 8 & 6 \end{pmatrix}$. Note that $\boldsymbol{B}$ is obtained from $\boldsymbol{A}$ by adding 2 times of the first row of $\boldsymbol{A}$ to the third row.

Let $\boldsymbol{E_3} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix}$. Observe that

$$\boldsymbol{E_3 A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 2 & 3 \\ 2 & -1 & 3 & 6 \\ 1 & 4 & 4 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 2 & 3 \\ 2 & -1 & 3 & 6 \\ 3 & 4 & 8 & 6 \end{pmatrix} = \boldsymbol{B}.$$

In general, let $\boldsymbol{A}$ be an $m \times n$ matrix and

$$\boldsymbol{E} = \begin{array}{cl}
\left(\begin{array}{ccc|c|cc|c|ccc}
1 &&&&&&&&& \\
& \ddots &&&&&&& \mathbf{0} & \\
&& 1 &&&&&&& \\
&&& 1 &&&&&& \\
&&&& \ddots &&&&& \\ \hline
&&& k &&& 1 &&& \\ \hline
&&&&&&& 1 && \\
& \mathbf{0} &&&&&&& \ddots & \\
&&&&&&&&& 1
\end{array}\right)
&
\begin{array}{l}
\\ \\ \\ \\ \\ \leftarrow j\text{th row} \\ \\ \\ \\
\end{array}
\\
\begin{array}{cc}
\uparrow & \uparrow \\
i\text{th column} & j\text{th column}
\end{array}
&
\end{array}
\qquad \text{if } i < j$$

$$
\text{or}\quad
\begin{array}{cc}
\begin{pmatrix}
1 & & & & & & & & \\
& \ddots & & & & & & \mathbf{0} & \\
& & 1 & & & & & & \\
& & & 1 & & k & & & \\
& & & & \ddots & & & & \\
& & & & & 1 & & & \\
& & & & & & 1 & & \\
& \mathbf{0} & & & & & & \ddots & \\
& & & & & & & & 1
\end{pmatrix}
&
\begin{array}{l}
\\ \\ \\ \leftarrow j\text{th row} \\ \\ \\ \\ \\ \\
\end{array}
\\
\uparrow \quad\quad \uparrow & \\
j\text{th column} \quad i\text{th column} &
\end{array}
\qquad \text{if } i > j
$$

be a square matrix of order m where all entries not specified are zero. Then $\boldsymbol{EA}$ is the matrix which can be obtained from $\boldsymbol{A}$ by adding k times of the ith row of $\boldsymbol{A}$ to the jth row.

The matrix $\boldsymbol{E}$ is invertible and

$$
\boldsymbol{E}^{-1} =
\begin{array}{cc}
\begin{pmatrix}
1 & & & & & & & & \\
& \ddots & & & & & & \mathbf{0} & \\
& & 1 & & & & & & \\
& & & 1 & & & & & \\
& & & & \ddots & & & & \\
& & & -k & & 1 & & & \\
& & & & & & 1 & & \\
& \mathbf{0} & & & & & & \ddots & \\
& & & & & & & & 1
\end{pmatrix}
&
\begin{array}{l}
\\ \\ \\ \\ \\ \leftarrow j\text{th row} \\ \\ \\ \\
\end{array}
\\
\uparrow \quad\quad \uparrow & \\
i\text{th column} \quad j\text{th column} &
\end{array}
\qquad \text{if } i < j
$$

$$
\text{or}\quad
\begin{array}{cc}
\begin{pmatrix}
1 & & & & & & & & \\
& \ddots & & & & & & \mathbf{0} & \\
& & 1 & & & & & & \\
& & & 1 & & -k & & & \\
& & & & \ddots & & & & \\
& & & & & 1 & & & \\
& & & & & & 1 & & \\
& \mathbf{0} & & & & & & \ddots & \\
& & & & & & & & 1
\end{pmatrix}
&
\begin{array}{l}
\\ \\ \\ \leftarrow j\text{th row} \\ \\ \\ \\ \\ \\
\end{array}
\\
\uparrow \quad\quad \uparrow & \\
j\text{th column} \quad i\text{th column} &
\end{array}
\qquad \text{if } i > j
$$

corresponds to the row operation of adding $-k$ times of the ith row of a matrix to the jth row.

Definition 2.4.3 A square matrix is called an *elementary matrix* if it can be obtained from an identity matrix by performing a single elementary row operation.

Remark 2.4.4

1. The matrices $\boldsymbol{E}$ described in Discussion 2.4.2 are elementary matrices and every elementary matrix is of one of these three types.

2. All elementary matrices are invertible and their inverses are also elementary matrices.

Example 2.4.5 Let

$$\boldsymbol{A} = \begin{pmatrix} 0 & 4 & 2 \\ -2 & 1 & -3 \\ 1 & 0 & 2 \end{pmatrix} \quad \text{and} \quad \boldsymbol{B} = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

where $\boldsymbol{A}$ and $\boldsymbol{B}$ are row equivalent. In the following, we show how $\boldsymbol{B}$ is obtained from $\boldsymbol{A}$ by elementary row operations and for each of these row operations, we write down the corresponding elementary matrices:

$$\boldsymbol{A} = \begin{pmatrix} 0 & 4 & 2 \\ -2 & 1 & -3 \\ 1 & 0 & 2 \end{pmatrix} \xrightarrow[\boldsymbol{E_1}]{R_1 \leftrightarrow R_3} \begin{pmatrix} 1 & 0 & 2 \\ -2 & 1 & -3 \\ 0 & 4 & 2 \end{pmatrix} \xrightarrow[\boldsymbol{E_2}]{R_2 + 2R_1} \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 4 & 2 \end{pmatrix}$$

$$\xrightarrow[\boldsymbol{E_3}]{R_3 - 4R_2} \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & -2 \end{pmatrix} \xrightarrow[\boldsymbol{E_4}]{-\frac{1}{2}R_3} \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} = \boldsymbol{B}.$$

where

$$\boldsymbol{E_1} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \boldsymbol{E_2} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{E_3} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -4 & 1 \end{pmatrix}, \quad \boldsymbol{E_4} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -\frac{1}{2} \end{pmatrix}.$$

By Discussion 2.4.2, we know that

$$\boldsymbol{E_4 E_3 E_2 E_1 A} = \boldsymbol{B}$$

and hence

$$\boldsymbol{A} = \boldsymbol{E_1}^{-1} \boldsymbol{E_2}^{-1} \boldsymbol{E_3}^{-1} \boldsymbol{E_4}^{-1} \boldsymbol{B}$$

where

$$\boldsymbol{E_1}^{-1} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \boldsymbol{E_2}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{E_3}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 4 & 1 \end{pmatrix}, \quad \boldsymbol{E_4}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$

(Recall that the inverses of $\boldsymbol{E_1}$, $\boldsymbol{E_2}$, $\boldsymbol{E_3}$, $\boldsymbol{E_4}$ have been given in Discussion 2.4.2.)

Remark 2.4.6 (Proof of Theorem 1.2.7) Theorem 1.2.7 states that if augmented matrices of two systems of linear equations are row equivalent, then the two systems have the same set of solutions. Following Remark 1.2.9, we need to show the following: Let $(\boldsymbol{A} \mid \boldsymbol{b})$ and $(\boldsymbol{C} \mid \boldsymbol{d})$ be two augmented matrices such that $(\boldsymbol{C} \mid \boldsymbol{d})$ can be obtained from $(\boldsymbol{A} \mid \boldsymbol{b})$ by an elementary row operation. Then the linear systems $\boldsymbol{Ax} = \boldsymbol{b}$ and $\boldsymbol{Cx} = \boldsymbol{d}$ have the same set of solutions.

Proof By Discussion 2.4.2, there exists an elementary matrix $\boldsymbol{E}$ such that

$$(\boldsymbol{C} \mid \boldsymbol{d}) = \boldsymbol{E}(\boldsymbol{A} \mid \boldsymbol{b}) = (\boldsymbol{EA} \mid \boldsymbol{Eb}),$$

i.e. $\boldsymbol{C} = \boldsymbol{EA}$ and $\boldsymbol{d} = \boldsymbol{Eb}$.

If $\boldsymbol{u}$ is a solution to $\boldsymbol{Ax} = \boldsymbol{b}$, then

$$\boldsymbol{Au} = \boldsymbol{b} \quad \Rightarrow \quad \boldsymbol{EAu} = \boldsymbol{Eb} \quad \Rightarrow \quad \boldsymbol{Cu} = \boldsymbol{d}$$

and hence $\boldsymbol{u}$ is a solution to $\boldsymbol{Cx} = \boldsymbol{d}$.

If $\boldsymbol{v}$ is a solution to $\boldsymbol{Cx} = \boldsymbol{d}$, then

$$\boldsymbol{Cv} = \boldsymbol{d} \quad \Rightarrow \quad \boldsymbol{EAv} = \boldsymbol{Eb} \quad \Rightarrow \quad \boldsymbol{E}^{-1}\boldsymbol{EAv} = \boldsymbol{E}^{-1}\boldsymbol{Eb} \quad \Rightarrow \quad \boldsymbol{IAv} = \boldsymbol{Ib} \quad \Rightarrow \quad \boldsymbol{Av} = \boldsymbol{b}$$

and hence $\boldsymbol{v}$ is a solution to $\boldsymbol{Ax} = \boldsymbol{b}$.

So we have shown that the linear systems $\boldsymbol{Ax} = \boldsymbol{b}$ and $\boldsymbol{Cx} = \boldsymbol{d}$ have the same set of solutions.

Theorem 2.4.7 (This theorem is part of our main theorem on invertible matrices, see Theorem 6.1.8.) If $\boldsymbol{A}$ is a square matrix, then the following statements are equivalent:

1. $\boldsymbol{A}$ is invertible.

2. The linear system $\boldsymbol{Ax} = \boldsymbol{0}$ has only the trivial solution.

3. The reduced row-echelon form of $\boldsymbol{A}$ is an identity matrix.

4. $\boldsymbol{A}$ can be expressed as a product of elementary matrices.

Proof

$1 \Rightarrow 2$: If $\boldsymbol{A}$ is invertible, then $\boldsymbol{Ax} = \boldsymbol{0}$ implies

$$\boldsymbol{x} = \boldsymbol{Ix} = \boldsymbol{A}^{-1}\boldsymbol{Ax} = \boldsymbol{A}^{-1}\boldsymbol{0} = \boldsymbol{0}$$

and hence the system has only the trivial solution $\boldsymbol{x} = \boldsymbol{0}$.

2 $\Rightarrow$ 3 : Suppose $\boldsymbol{Ax} = \boldsymbol{0}$ has only the trivial solution. Since the number of columns in $\boldsymbol{A}$ is equal to the number of rows in $\boldsymbol{A}$, the reduced row-echelon form of the augmented matrix $\left(\boldsymbol{A}\,\middle|\,\begin{smallmatrix}0\\ \vdots\\ 0\end{smallmatrix}\right)$ of the system $\boldsymbol{Ax} = \boldsymbol{0}$ cannot have any zero rows (see Remark 1.4.8.2), i.e. the reduced row-echelon form of $\left(\boldsymbol{A}\,\middle|\,\begin{smallmatrix}0\\ \vdots\\ 0\end{smallmatrix}\right)$ is $\left(\boldsymbol{I}\,\middle|\,\begin{smallmatrix}0\\ \vdots\\ 0\end{smallmatrix}\right)$. Hence the reduced row-echelon form of $\boldsymbol{A}$ is $\boldsymbol{I}$.

3 $\Rightarrow$ 4 : Since the reduced row-echelon form of $\boldsymbol{A}$ is $\boldsymbol{I}$, there exist elementary matrices $\boldsymbol{E}_1, \boldsymbol{E}_2, \ldots, \boldsymbol{E}_k$ such that

$$\boldsymbol{E}_k \cdots \boldsymbol{E}_2\boldsymbol{E}_1\boldsymbol{A} = \boldsymbol{I}$$

and hence

$$\boldsymbol{A} = (\boldsymbol{E}_k \cdots \boldsymbol{E}_2\boldsymbol{E}_1)^{-1}\boldsymbol{I} = (\boldsymbol{E}_k \cdots \boldsymbol{E}_2\boldsymbol{E}_1)^{-1} = {\boldsymbol{E}_1}^{-1}{\boldsymbol{E}_2}^{-1} \cdots {\boldsymbol{E}_k}^{-1}$$

where ${\boldsymbol{E}_1}^{-1}, {\boldsymbol{E}_2}^{-1}, \ldots, {\boldsymbol{E}_k}^{-1}$ are also elementary matrices.

4 $\Rightarrow$ 1 : Suppose $\boldsymbol{A}$ is a product of elementary matrices. Since all elementary matrices are invertible, $\boldsymbol{A}$ is invertible by Remark 2.3.10.

Discussion 2.4.8 Let $\boldsymbol{A}$ be an invertible matrix of order n and let $\boldsymbol{E}_1, \boldsymbol{E}_2, \ldots, \boldsymbol{E}_k$ be elementary matrices as in the proof of Theorem 2.4.7, i.e.

$$\boldsymbol{E}_k \cdots \boldsymbol{E}_2\boldsymbol{E}_1\boldsymbol{A} = \boldsymbol{I}.$$

Post-multiply $\boldsymbol{A}^{-1}$ to both sides of the equation:

$$\boldsymbol{E}_k \cdots \boldsymbol{E}_2\boldsymbol{E}_1\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}\boldsymbol{A}^{-1} \quad \Rightarrow \quad \boldsymbol{E}_k \cdots \boldsymbol{E}_2\boldsymbol{E}_1 = \boldsymbol{A}^{-1}.$$

Consider the $n \times 2n$ matrix $(\boldsymbol{A} \mid \boldsymbol{I})$. We have

$$\begin{aligned}\boldsymbol{E}_k \cdots \boldsymbol{E}_2\boldsymbol{E}_1(\boldsymbol{A} \mid \boldsymbol{I}) &= \boldsymbol{E}_k \cdots \boldsymbol{E}_2(\boldsymbol{E}_1\boldsymbol{A} \mid \boldsymbol{E}_1\boldsymbol{I}) \\ &\ \ \vdots \\ &= (\boldsymbol{E}_k \cdots \boldsymbol{E}_2\boldsymbol{E}_1\boldsymbol{A} \mid \boldsymbol{E}_k \cdots \boldsymbol{E}_2\boldsymbol{E}_1\boldsymbol{I}) \\ &= (\boldsymbol{I} \mid \boldsymbol{A}^{-1}).\end{aligned}$$

Since pre-multiplications of a matrix by elementary matrices correspond to performing elementary row operations on the matrix, we can use elementary row operations to transform the matrix $(\boldsymbol{A} \mid \boldsymbol{I})$ to $(\boldsymbol{I} \mid \boldsymbol{A}^{-1})$. This provides us with a method to find the inverse of $\boldsymbol{A}$.

Example 2.4.9 Find the inverse of $\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 3 \\ 1 & 0 & 8 \end{pmatrix}$ if it exists.

Solution

$$\left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 2 & 5 & 3 & 0 & 1 & 0 \\ 1 & 0 & 8 & 0 & 0 & 1 \end{array}\right) \xrightarrow[R_3 - R_1]{R_2 - 2R_1} \left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & -3 & -2 & 1 & 0 \\ 0 & -2 & 5 & -1 & 0 & 1 \end{array}\right) \xrightarrow{R_3 + 2R_2}$$

$$\left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & -3 & -2 & 1 & 0 \\ 0 & 0 & -1 & -5 & 2 & 1 \end{array}\right) \xrightarrow{-R_3} \left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & -3 & -2 & 1 & 0 \\ 0 & 0 & 1 & 5 & -2 & -1 \end{array}\right) \xrightarrow[R_2 + 3R_3]{R_1 - 3R_3}$$

$$\left(\begin{array}{ccc|ccc} 1 & 2 & 0 & -14 & 6 & 3 \\ 0 & 1 & 0 & 13 & -5 & -3 \\ 0 & 0 & 1 & 5 & -2 & -1 \end{array}\right) \xrightarrow{R_1 - 2R_2} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -40 & 16 & 9 \\ 0 & 1 & 0 & 13 & -5 & -3 \\ 0 & 0 & 1 & 5 & -2 & -1 \end{array}\right)$$

So $\boldsymbol{A}^{-1} = \begin{pmatrix} -40 & 16 & 9 \\ 13 & -5 & -3 \\ 5 & -2 & -1 \end{pmatrix}$.

Remark 2.4.10 Theorem 2.4.7 tells us that a square matrix is invertible if and only if its reduced row-echelon form is an identity matrix. This can be used to check whether a square matrix is invertible. Actually, to check whether a square matrix is invertible, we only need to reduce the matrix to a row-echelon form. If the row-echelon form of a square matrix has no zero row, the matrix is invertible; otherwise, the matrix is singular. (Why?)

Example 2.4.11

1. Let $\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 4 & 4 \\ 3 & 6 & 3 \end{pmatrix}$. Since

$$\begin{pmatrix} 1 & 2 & 0 \\ 2 & 4 & 4 \\ 3 & 6 & 3 \end{pmatrix} \xrightarrow[R_3 - 3R_1]{R_2 - 2R_1} \begin{pmatrix} 1 & 2 & 0 \\ 0 & 0 & 4 \\ 0 & 0 & 3 \end{pmatrix} \xrightarrow{R_3 - \frac{3}{4}R_2} \begin{pmatrix} 1 & 2 & 0 \\ 0 & 0 & 4 \\ 0 & 0 & 0 \end{pmatrix},$$

   $\boldsymbol{A}$ is singular.

2. Let $\boldsymbol{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Show that $\boldsymbol{A}$ is invertible if and only if $ad - bc \neq 0$.

   Solution By Example 2.3.8, we already know that $\boldsymbol{A}$ is invertible if $ad - bc \neq 0$. So we only need to show that if $\boldsymbol{A}$ is invertible, then $ad - bc \neq 0$:

   If both $a = 0$ and $c = 0$, then the reduced row-echelon form of $\boldsymbol{A} = \begin{pmatrix} 0 & b \\ 0 & d \end{pmatrix}$ can never be the identity matrix (why?) and hence $\boldsymbol{A}$ is not invertible. Thus we have $a \neq 0$ or $c \neq 0$. First, assume $a \neq 0$. Then

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \xrightarrow{R_2 - \frac{c}{a}R_1} \begin{pmatrix} a & b \\ 0 & \frac{ad-bc}{a} \end{pmatrix}.$$

So if $\boldsymbol{A}$ is invertible, $ad - bc$ cannot be zero. Next, assume $a = 0$ and hence $c \neq 0$. Then

$$\begin{pmatrix} 0 & b \\ c & d \end{pmatrix} \xrightarrow{R_1 \leftrightarrow R_2} \begin{pmatrix} c & d \\ 0 & b \end{pmatrix}.$$

So if $\boldsymbol{A}$ is invertible, then b cannot be zero and again we have $ad - bc = -bc \neq 0$.

(Note that when doing row operations, we must be very careful if we want to multiply or divide a row by an unknown constant, see Example 1.4.10.3.)

Theorem 2.4.12 Let $\boldsymbol{A}$, $\boldsymbol{B}$ be square matrices of the same size. If $\boldsymbol{AB} = \boldsymbol{I}$, then $\boldsymbol{A}$, $\boldsymbol{B}$ are both invertible,

$$\boldsymbol{A}^{-1} = \boldsymbol{B}, \quad \boldsymbol{B}^{-1} = \boldsymbol{A} \quad \text{and} \quad \boldsymbol{BA} = \boldsymbol{I}.$$

Proof Consider the homogeneous system of linear equations $\boldsymbol{Bx} = \boldsymbol{0}$. Suppose $\boldsymbol{u}$ is a solution to the system, i.e. $\boldsymbol{Bu} = \boldsymbol{0}$. Then

$$\boldsymbol{ABu} = \boldsymbol{Iu} \quad \Rightarrow \quad \boldsymbol{A0} = \boldsymbol{u} \quad \Rightarrow \quad \boldsymbol{0} = \boldsymbol{u}.$$

The system $\boldsymbol{Bx} = \boldsymbol{0}$ has only the trivial solution. By Theorem 2.4.7, $\boldsymbol{B}$ is invertible.

Since $\boldsymbol{B}$ is invertible, we post-multiply $\boldsymbol{B}^{-1}$ to both sides of $\boldsymbol{AB} = \boldsymbol{I}$:

$$\boldsymbol{ABB}^{-1} = \boldsymbol{IB}^{-1} \quad \Rightarrow \quad \boldsymbol{AI} = \boldsymbol{B}^{-1} \quad \Rightarrow \quad \boldsymbol{A} = \boldsymbol{B}^{-1}.$$

Hence by Theorem 2.3.9.3, $\boldsymbol{A}$ is invertible,

$$\boldsymbol{A}^{-1} = (\boldsymbol{B}^{-1})^{-1} = \boldsymbol{B} \quad \text{and} \quad \boldsymbol{BA} = \boldsymbol{BB}^{-1} = \boldsymbol{I}.$$

Example 2.4.13 Let $\boldsymbol{A}$ be a square matrix such that

$$\boldsymbol{A}^2 - 3\boldsymbol{A} - 6\boldsymbol{I} = \boldsymbol{0}.$$

Show that $\boldsymbol{A}$ is invertible.

Solution Note that

$$\boldsymbol{A}(\boldsymbol{A} - 3\boldsymbol{I}) = \boldsymbol{A}^2 - 3\boldsymbol{AI} = \boldsymbol{A}^2 - 3\boldsymbol{A} = 6\boldsymbol{I}.$$

So $\boldsymbol{A}\left[\frac{1}{6}(\boldsymbol{A} - 3\boldsymbol{I})\right] = \boldsymbol{I}$ and by Theorem 2.4.12, $\boldsymbol{A}$ is invertible.

Theorem 2.4.14 Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two square matrices of the same order. If $\boldsymbol{A}$ is singular, then $\boldsymbol{AB}$ and $\boldsymbol{BA}$ are singular.

(The proof of the theorem is left as an exercise. See Question 2.42.)

Discussion 2.4.15 From Discussion 2.4.2, we learn that to pre-multiply an elementary matrix to a matrix $\boldsymbol{A}$ is equivalent to doing an elementary row operation on $\boldsymbol{A}$. What will happen if we post-multiply an elementary matrix to a matrix?

Let $\boldsymbol{A}$ be a $p \times m$ matrix.

1. If $\boldsymbol{E}$ is the elementary matrix defined in Discussion 2.4.2.1, then $\boldsymbol{AE}$ is the matrix which can be obtained from $\boldsymbol{A}$ by multiplying the ith column of $\boldsymbol{A}$ by k.
2. If $\boldsymbol{E}$ is the elementary matrix defined in Discussion 2.4.2.2, then $\boldsymbol{AE}$ is the matrix which can be obtained from $\boldsymbol{A}$ by interchanging the ith and jth columns of $\boldsymbol{A}$.
3. If $\boldsymbol{E}$ is the elementary matrix defined in Discussion 2.4.2.3, then $\boldsymbol{AE}$ is the matrix which can be obtained from $\boldsymbol{A}$ by adding k times of the jth column of $\boldsymbol{A}$ to the ith column.

The three operations on matrices described above are called *elementary column operations.*

Example 2.4.16 Let $\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 2 & 3 \\ 2 & -1 & 3 & 6 \\ 1 & 4 & 4 & 0 \end{pmatrix}$.

1. Let $\boldsymbol{E}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$. Then

$$\boldsymbol{AE}_1 = \begin{pmatrix} 1 & 0 & 2 & 3 \\ 2 & -1 & 3 & 6 \\ 1 & 4 & 4 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 4 & 3 \\ 2 & -1 & 6 & 6 \\ 1 & 4 & 8 & 0 \end{pmatrix}$$

   which can be obtained from $\boldsymbol{A}$ by multiplying the third column of $\boldsymbol{A}$ by 2.

2. Let $\boldsymbol{E}_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$. Then

$$\boldsymbol{AE}_2 = \begin{pmatrix} 1 & 0 & 2 & 3 \\ 2 & -1 & 3 & 6 \\ 1 & 4 & 4 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 3 & 2 & 0 \\ 2 & 6 & 3 & -1 \\ 1 & 0 & 4 & 4 \end{pmatrix}$$

   which can be obtained from $\boldsymbol{A}$ by interchanging the second and fourth columns of $\boldsymbol{A}$.

3. Let $E_3 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 2 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$. Then

$$AE_3 = \begin{pmatrix} 1 & 0 & 2 & 3 \\ 2 & -1 & 3 & 6 \\ 1 & 4 & 4 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 2 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 0 & 2 & 3 \\ 8 & -1 & 3 & 6 \\ 7 & 4 & 4 & 0 \end{pmatrix}$$

which can be obtained from $\boldsymbol{A}$ by adding 2 times of the third column of $\boldsymbol{A}$ to the first column.

Section 2.5 Determinants

Discussion 2.5.1 By Example 2.4.11.2, we know that a 2×2 matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible if and only if $ad - bc \neq 0$. Actually, we have a similar formula to determine whether a square matrix of higher order is invertible. The formula involves a quantity called "determinant". Unfortunately, there is no easy way to define it. In the following, we use an inductive approach to define this quantity.

Definition 2.5.2 Let $\boldsymbol{A} = (a_{ij})$ be an $n \times n$ matrix. Let $\boldsymbol{M_{ij}}$ be an $(n-1) \times (n-1)$ matrix obtained from $\boldsymbol{A}$ by deleting the ith row and the jth column. Then the *determinant* of $\boldsymbol{A}$ is defined as

$$\det(\boldsymbol{A}) = \begin{cases} a_{11} & \text{if } n = 1 \\ a_{11}A_{11} + a_{12}A_{12} + \cdots + a_{1n}A_{1n} & \text{if } n > 1 \end{cases}$$

where

$$A_{ij} = (-1)^{i+j} \det(\boldsymbol{M_{ij}}).$$

The number A_{ij} is called the (i, j)-*cofactor* of $\boldsymbol{A}$.

The way we defined "determinant" above is known as the *cofactor expansion* (see also Theorem 2.5.6).

Notation 2.5.3 For an $n \times n$ matrix $\boldsymbol{A} = (a_{ij})$, $\det(\boldsymbol{A})$ is usually written as

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}.$$

Example 2.5.4

1. Let $\boldsymbol{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Then $\boldsymbol{M_{11}} = (d)$ and $\boldsymbol{M_{12}} = (c)$. Hence

$$A_{11} = (-1)^{1+1}\det(\boldsymbol{M_{11}}) = d \quad \text{and} \quad A_{12} = (-1)^{1+2}\det(\boldsymbol{M_{12}}) = -c.$$

By Definition 2.5.2, $\det(\boldsymbol{A}) = aA_{11} + bA_{12} = ad - bc$.

2. Let $\boldsymbol{B} = \begin{pmatrix} -3 & -2 & 4 \\ 4 & 3 & 1 \\ 0 & 2 & 4 \end{pmatrix}$. By Definition 2.5.2,

$$\begin{aligned} \det(\boldsymbol{B}) &= (-3)\begin{vmatrix} 3 & 1 \\ 2 & 4 \end{vmatrix} - (-2)\begin{vmatrix} 4 & 1 \\ 0 & 4 \end{vmatrix} + 4\begin{vmatrix} 4 & 3 \\ 0 & 2 \end{vmatrix} \\ &= -3(3\cdot 4 - 1\cdot 2) + 2(4\cdot 4 - 1\cdot 0) + 4(4\cdot 2 - 3\cdot 0) = 34. \end{aligned}$$

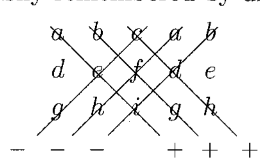
3. Let $\boldsymbol{C} = \begin{pmatrix} 0 & -1 & 2 & 0 \\ 2 & -3 & 3 & -2 \\ 0 & 2 & 4 & 0 \\ 0 & 0 & 2 & -1 \end{pmatrix}$. By Definition 2.5.2,

$$\begin{aligned} \det(\boldsymbol{C}) &= 0\begin{vmatrix} -3 & 3 & -2 \\ 2 & 4 & 0 \\ 0 & 2 & -1 \end{vmatrix} - (-1)\begin{vmatrix} 2 & 3 & -2 \\ 0 & 4 & 0 \\ 0 & 2 & -1 \end{vmatrix} + 2\begin{vmatrix} 2 & -3 & -2 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{vmatrix} - 0\begin{vmatrix} 2 & -3 & 3 \\ 0 & 2 & 4 \\ 0 & 0 & 2 \end{vmatrix} \\ &= 0 + \left[2\begin{vmatrix} 4 & 0 \\ 2 & -1 \end{vmatrix} - 3\begin{vmatrix} 0 & 0 \\ 0 & -1 \end{vmatrix} + (-2)\begin{vmatrix} 0 & 4 \\ 0 & 2 \end{vmatrix}\right] \\ &\quad + 2\left[2\begin{vmatrix} 2 & 0 \\ 0 & -1 \end{vmatrix} - (-3)\begin{vmatrix} 0 & 0 \\ 0 & -1 \end{vmatrix} + (-2)\begin{vmatrix} 0 & 2 \\ 0 & 0 \end{vmatrix}\right] + 0 \\ &= [2\cdot(-4) - 3\cdot 0 - 2\cdot 0] + 2\,[2\cdot(-2) + 3\cdot 0 - 2\cdot 0] = -16. \end{aligned}$$

Remark 2.5.5 Let $\boldsymbol{A} = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$. Then

$$\det(\boldsymbol{A}) = aei + bfg + cdh - ceg - afh - bdi.$$

(Check it.) The formula can be easily remembered by using the following diagram:



However, the method shown here cannot be generalized to higher order. For example, if we apply the method to the 4×4 matrix $\boldsymbol{C}$ in Example 2.5.4.3, we cannot obtain the value of $\det(\boldsymbol{C})$. (Try it.)

Theorem 2.5.6 (Cofactor Expansions) For an $n \times n$ matrix $\boldsymbol{A} = (a_{ij})$, $\det(\boldsymbol{A})$ can be expressed as a cofactor expansion using any row or column of $\boldsymbol{A}$:

$$\begin{aligned}\det(\boldsymbol{A}) &= a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in} && \text{(cofactor expansion along the } i\text{th row)}\\ &= a_{1j}A_{1j} + a_{2j}A_{2j} + \cdots + a_{nj}A_{nj} && \text{(cofactor expansion along the } j\text{th column)}\end{aligned}$$

for any $i = 1, 2, \ldots n$ and $j = 1, 2, \ldots n$.

(Since the proof involves some deeper knowledge of determinants, we use this result without proving it.)

Example 2.5.7 We use Example 2.5.4.2 to illustrate Theorem 2.5.6. If we expand $\boldsymbol{B}$ along the second row,

$$\det(\boldsymbol{B}) = -4\begin{vmatrix} -2 & 4 \\ 2 & 4 \end{vmatrix} + 3\begin{vmatrix} -3 & 4 \\ 0 & 4 \end{vmatrix} - \begin{vmatrix} -3 & -2 \\ 0 & 2 \end{vmatrix} = 34.$$

If we expand B along the third column,

$$\det(\boldsymbol{B}) = 4\begin{vmatrix} 4 & 3 \\ 0 & 2 \end{vmatrix} - \begin{vmatrix} -3 & -2 \\ 0 & 2 \end{vmatrix} + 4\begin{vmatrix} -3 & -2 \\ 4 & 3 \end{vmatrix} = 34.$$

Theorem 2.5.8 If $\boldsymbol{A}$ is a triangular matrix, then the determinant of $\boldsymbol{A}$ is equal to the product of the diagonal entries of $\boldsymbol{A}$.

Proof We shall prove the following assertion by using mathematical induction on n.

$$\text{For any } \boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a_{nn} \end{pmatrix}, \quad \det(\boldsymbol{A}) = a_{11}a_{22}\cdots a_{nn}. \tag{P}$$

(The proof for lower triangular matrices is similar.)

The procedure of mathematical induction consists of two steps:

Initial Step: For $n = 1$, $\boldsymbol{A} = (a_{11})$ and hence $\det(\boldsymbol{A}) = a_{11}$. So (P) is true for $n = 1$.

Induction Step: Assume (P) is true for $n = k$ (this is called the *inductive assumption*),

$$\text{i.e.} \quad \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ 0 & a_{22} & \cdots & a_{2k} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a_{kk} \end{vmatrix} = a_{11}a_{22}\cdots a_{kk}.$$

Let $\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} & a_{1,k+1} \\ 0 & a_{22} & \cdots & a_{2k} & a_{2,k+1} \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & a_{kk} & a_{k,k+1} \\ 0 & \cdots & \cdots & 0 & a_{k+1,k+1} \end{pmatrix}$. By expanding along the last row of $\boldsymbol{A}$,

$$\det(\boldsymbol{A}) = (-1)^{(k+1)+(k+1)} a_{k+1,k+1} \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ 0 & a_{22} & \cdots & a_{2k} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a_{kk} \end{vmatrix} = a_{11}a_{22}\cdots a_{kk}a_{k+1,k+1}.$$

Hence we have shown that if (P) is true for $n = k$, then (P) is true for $n = k + 1$.

By the initial step, (P) is true for $n = 1$. Applying the induction step with $k = 1$, (P) is true for $n = 1 + 1 = 2$. Again applying the induction step with $k = 2$, (P) is true for $n = 2 + 1 = 3$. By repeating the induction step, (P) is true for $n = 3 + 1 = 4$, $n = 4 + 1 = 5$, and so on. Eventually, (P) is true for all $n \geq 1$.

Example 2.5.9

1. $\det(\boldsymbol{I}) = 1$.
2. $\begin{vmatrix} -1 & 3.5 & 14 \\ 0 & 5 & -26 \\ 0 & 0 & 2 \end{vmatrix} = (-1) \cdot 5 \cdot 2 = -10.$
3. $\begin{vmatrix} -2 & 0 & 0 \\ 99 & 0 & 0 \\ 10 & -4.5 & 10 \end{vmatrix} = (-2) \cdot 0 \cdot 10 = 0.$

Theorem 2.5.10 If $\boldsymbol{A}$ is a square matrix, then $\det(\boldsymbol{A}) = \det(\boldsymbol{A}^{\mathrm{T}})$.

Proof We prove the theorem by using mathematical induction on the order of $\boldsymbol{A}$:

If $\boldsymbol{A}$ is a 1×1 matrix, then $\boldsymbol{A} = \boldsymbol{A}^{\mathrm{T}}$ and hence $\det(\boldsymbol{A}) = \det(\boldsymbol{A}^{\mathrm{T}})$.

Assume $\det(\boldsymbol{A}) = \det(\boldsymbol{A}^{\mathrm{T}})$ for any $k \times k$ matrix $\boldsymbol{A}$.

Now, let $\boldsymbol{A}$ be a $(k+1) \times (k+1)$ matrix. We need to show $\det(\boldsymbol{A}) = \det(\boldsymbol{A}^{\mathrm{T}})$ by using the inductive assumption. If we expand along the first row of $\boldsymbol{A}$, we have

$$\det(\boldsymbol{A}) = a_{11}\det(\boldsymbol{M_{11}}) - a_{12}\det(\boldsymbol{M_{12}}) + \cdots + (-1)^{1+(k+1)} a_{1,k+1}\det(\boldsymbol{M_{1,k+1}})$$

where $\boldsymbol{M_{ij}}$ is the $k \times k$ matrix obtained from $\boldsymbol{A}$ by deleting the ith row and the jth column. On the other hand, if we expand $\boldsymbol{A}^{\mathrm{T}}$ along the first column, we get

$$\det(\boldsymbol{A}^{\mathrm{T}}) = a_{11}\det(\boldsymbol{M_{11}}^{\mathrm{T}}) - a_{12}\det(\boldsymbol{M_{12}}^{\mathrm{T}}) + \cdots + (-1)^{1+(k+1)} a_{1,k+1}\det(\boldsymbol{M_{1,k+1}}^{\mathrm{T}}).$$

By the inductive assumption, $\det(\boldsymbol{M}_{ij}) = \det(\boldsymbol{M}_{ij}{}^{\mathrm{T}})$ for all i, j. So $\det(\boldsymbol{A}) = \det(\boldsymbol{A}^{\mathrm{T}})$. Hence by mathematical induction, $\det(\boldsymbol{A}) = \det(\boldsymbol{A}^{\mathrm{T}})$ for any square matrix $\boldsymbol{A}$.

Example 2.5.11 Let

$$\boldsymbol{C} = \begin{pmatrix} 0 & -1 & 2 & 0 \\ 2 & -3 & 3 & -2 \\ 0 & 2 & 4 & 0 \\ 0 & 0 & 2 & -1 \end{pmatrix}.$$

By Example 2.5.4.3, $\det(\boldsymbol{C}) = -16$. Thus by Theorem 2.5.10, $\det(\boldsymbol{C}^{\mathrm{T}}) = \det(\boldsymbol{C}) = -16$. (As an exercise, verify the result by computing $\det(\boldsymbol{C}^{\mathrm{T}})$ directly.)

Theorem 2.5.12

1. The determinant of a square matrix with two identical rows is zero.
2. The determinant of a square matrix with two identical columns is zero.

(The proof of the theorem is left as an exercise. See Question 2.58.)

Example 2.5.13 The following matrices have zero determinants:

$$\begin{pmatrix} 4 & -2 \\ 4 & -2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 4 \\ -1 & 10 & 4 \\ 1 & 2 & 4 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 & 1 \\ -1 & -3 & -3 & 9 \\ 2 & 4 & 4 & 0 \\ 0 & -2 & -2 & -1 \end{pmatrix}.$$

Discussion 2.5.14 Consider the three types of elementary row operations on matrices.

1. **Multiply a row by a constant:**

   Let $\boldsymbol{A} = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} a & b & c \\ kd & ke & kf \\ g & h & i \end{pmatrix}$. Note that $\boldsymbol{B}$ is obtained from $\boldsymbol{A}$ by multiplying the second row of $\boldsymbol{A}$ by a constant k. Then

   $$\begin{aligned} \det(\boldsymbol{B}) &= a(ke)i + b(kf)g + c(kd)h - c(ke)g - a(kf)h - b(kd)i \\ &= k(aei + bfg + cdh - ceg - afh - bdi) = k\det(\boldsymbol{A}). \end{aligned}$$

   Let $\boldsymbol{E} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & k & 0 \\ 0 & 0 & 1 \end{pmatrix}$. Then $\boldsymbol{B} = \boldsymbol{E}\boldsymbol{A}$. Since $\det(\boldsymbol{E}) = k$, we have

   $$\det(\boldsymbol{E}\boldsymbol{A}) = \det(\boldsymbol{E})\det(\boldsymbol{A}).$$

2. **Interchange two rows:**

Let $\boldsymbol{A} = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} a & b & c \\ g & h & i \\ d & e & f \end{pmatrix}$. Note that $\boldsymbol{B}$ is obtained from $\boldsymbol{A}$ by interchanging the second and third rows of $\boldsymbol{A}$. Then

$$\begin{aligned}\det(\boldsymbol{B}) &= ahf + bid + cge - chd - aie - bgf \\ &= -(aei + bfg + cdh - ceg - afh - bdi) = -\det(\boldsymbol{A}).\end{aligned}$$

Let $\boldsymbol{E} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$. Then $\boldsymbol{B} = \boldsymbol{E}\boldsymbol{A}$. Since $\det(\boldsymbol{E}) = -1$, we have

$$\det(\boldsymbol{E}\boldsymbol{A}) = \det(\boldsymbol{E})\det(\boldsymbol{A}).$$

3. **Add a multiple of a row to another row:**

Let $\boldsymbol{A} = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} a & b & c \\ d & e & f \\ g+ka & h+kb & i+kc \end{pmatrix}$. Note that $\boldsymbol{B}$ is obtained from $\boldsymbol{A}$ by adding k times of the first row of $\boldsymbol{A}$ to the third row. Then

$$\begin{aligned}\det(\boldsymbol{B}) &= ae(i+kc) + bf(g+ka) + cd(h+kb) - ce(g+ka) - af(h+kb) - bd(i+kc) \\ &= aei + bfg + cdh - ceg - afh - bdi = \det(\boldsymbol{A}).\end{aligned}$$

Let $\boldsymbol{E} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ k & 0 & 1 \end{pmatrix}$. Then $\boldsymbol{B} = \boldsymbol{E}\boldsymbol{A}$. Since $\det(\boldsymbol{E}) = 1$, we have

$$\det(\boldsymbol{E}\boldsymbol{A}) = \det(\boldsymbol{E})\det(\boldsymbol{A}).$$

Theorem 2.5.15 Let $\boldsymbol{A}$ be a square matrix.

1. If $\boldsymbol{B}$ is a square matrix obtained from $\boldsymbol{A}$ by multiplying one row of $\boldsymbol{A}$ by a constant k, then $\det(\boldsymbol{B}) = k\det(\boldsymbol{A})$.
2. If $\boldsymbol{B}$ is a square matrix obtained from $\boldsymbol{A}$ by interchanging two rows of $\boldsymbol{A}$, then $\det(\boldsymbol{B}) = -\det(\boldsymbol{A})$.
3. If $\boldsymbol{B}$ is a square matrix obtained from $\boldsymbol{A}$ by adding a multiple of one row of $\boldsymbol{A}$ to another row, then $\det(\boldsymbol{B}) = \det(\boldsymbol{A})$.
4. Let $\boldsymbol{E}$ be an elementary matrix of the same size as $\boldsymbol{A}$. Then $\det(\boldsymbol{E}\boldsymbol{A}) = \det(\boldsymbol{E})\det(\boldsymbol{A})$.

Proof In the following, we only illustrate the proof of Part 3: Let $\boldsymbol{A} = (a_{ij})$ be a square matrix of order n and $\boldsymbol{B}$ a square matrix obtained from $\boldsymbol{A}$ by adding k times of the ith row of $\boldsymbol{A}$ to the jth row. Then by expanding along the jth row of $\boldsymbol{B}$,

$$\begin{aligned}\det(\boldsymbol{B}) &= (a_{j1} + ka_{i1})A_{j1} + (a_{j2} + ka_{i2})A_{j2} + \cdots + (a_{jn} + ka_{in})A_{jn} \\ &= \det(\boldsymbol{A}) + k\left(a_{i1}A_{j1} + a_{i2}A_{j2} + \cdots + a_{in}A_{jn}\right).\end{aligned}$$

Note that

$$\begin{aligned} a_{i1}A_{j1} + a_{i2}A_{j2} + \cdots + a_{in}A_{jn} &= \begin{array}{|cccc|l} a_{11} & a_{12} & \cdots & a_{1n} & \\ \vdots & \vdots & & \vdots & \\ a_{i1} & a_{i2} & \cdots & a_{in} & \leftarrow \text{ the } i\text{th row} \\ \vdots & \vdots & & \vdots & \\ a_{j-1,1} & a_{j-1,2} & \cdots & a_{j-1,n} & \\ a_{i1} & a_{i2} & \cdots & a_{in} & \leftarrow \text{ the } j\text{th row} \\ a_{j+1,1} & a_{j+1,2} & \cdots & a_{j+1,n} & \\ \vdots & \vdots & & \vdots & \\ a_{n1} & a_{n2} & \cdots & a_{nn} & \end{array} \\ &= 0 \end{aligned} \tag{2.4}$$

by Theorem 2.5.12.1. So $\det(\boldsymbol{B}) = \det(\boldsymbol{A})$.

(Proofs of the other parts of the theorem are left as exercises. See Question 2.59.)

Remark 2.5.16 By Theorem 2.5.15, we can use elementary row operations to transform a square matrix to a triangular matrix and then by Theorem 2.5.8, compute the determinant accordingly.

Example 2.5.17

1. Use elementary row operations to find the determinant of the matrix $\begin{pmatrix} 3 & -1 & 1 & 1 \\ 3 & -1 & 2 & 1 \\ 0 & 2 & 4 & 0 \\ 0 & 0 & 2 & -1 \end{pmatrix}$.

   Solution

$$\begin{vmatrix} 3 & -1 & 1 & 1 \\ 3 & -1 & 2 & 1 \\ 0 & 2 & 4 & 0 \\ 0 & 0 & 2 & -1 \end{vmatrix} \overset{R_2 - R_1}{=} \begin{vmatrix} 3 & -1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 2 & 4 & 0 \\ 0 & 0 & 2 & -1 \end{vmatrix} \overset{R_2 \leftrightarrow R_3}{=} -\begin{vmatrix} 3 & -1 & 1 & 1 \\ 0 & 2 & 4 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & -1 \end{vmatrix}$$

$$\overset{R_4 - 2R_3}{=} -\begin{vmatrix} 3 & -1 & 1 & 1 \\ 0 & 2 & 4 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{vmatrix} = -3 \cdot 2 \cdot 1 \cdot (-1) = 6.$$

2. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be square matrices such that $\boldsymbol{B}$ can be obtained from $\boldsymbol{A}$ by elementary row operations:

$$\boldsymbol{A} \xrightarrow{R_1 + \frac{2}{9}R_2} \xrightarrow{R_2 \leftrightarrow R_3} \xrightarrow{4R_2} \boldsymbol{B}.$$

If $\boldsymbol{B} = \begin{pmatrix} 5 & 0 & 8 & -1 \\ 0 & -2 & 1 & 0 \\ 0 & 0 & 1 & 9 \\ 0 & 0 & 0 & \frac{1}{3} \end{pmatrix}$, find $\det(\boldsymbol{A})$.

Solution Since $\det(\boldsymbol{B}) = 5 \cdot (-2) \cdot 1 \cdot \frac{1}{3} = -\frac{10}{3}$, $\det(\boldsymbol{A}) = -\frac{1}{4}\det(\boldsymbol{B}) = \frac{5}{6}$.

Remark 2.5.18 Since $\det(\boldsymbol{A}) = \det(\boldsymbol{A}^{\mathrm{T}})$ for any square matrix $\boldsymbol{A}$, Theorem 2.5.15 still holds if we change "rows" to "columns": Let $\boldsymbol{A}$ be a square matrix.

1. If $\boldsymbol{B}$ is a square matrix obtained from $\boldsymbol{A}$ by multiplying one column of $\boldsymbol{A}$ by a constant k, then $\det(\boldsymbol{B}) = k\det(\boldsymbol{A})$.

2. If $\boldsymbol{B}$ is a square matrix obtained from $\boldsymbol{A}$ by interchanging two columns of $\boldsymbol{A}$, then $\det(\boldsymbol{B}) = -\det(\boldsymbol{A})$.

3. If $\boldsymbol{B}$ is a square matrix obtained from $\boldsymbol{A}$ by adding a multiple of one column of $\boldsymbol{A}$ to another column, then $\det(\boldsymbol{B}) = \det(\boldsymbol{A})$.

4. Let $\boldsymbol{E}$ be an elementary matrix of the same size as $\boldsymbol{A}$. Then $\det(\boldsymbol{AE}) = \det(\boldsymbol{A})\det(\boldsymbol{E})$.

Theorem 2.5.19 A square matrix $\boldsymbol{A}$ is invertible if and only if $\det(\boldsymbol{A}) \neq 0$.

Proof Let $\boldsymbol{A}$ be a square matrix and $\boldsymbol{E_1}, \boldsymbol{E_2}, \ldots, \boldsymbol{E_k}$ elementary matrices such that $\boldsymbol{B} = \boldsymbol{E_k} \cdots \boldsymbol{E_2}\boldsymbol{E_1}\boldsymbol{A}$ is the reduced row-echelon form of $\boldsymbol{A}$. By Theorem 2.5.15.4, $\det(\boldsymbol{B}) = \det(\boldsymbol{E_k}) \cdots \det(\boldsymbol{E_2})\det(\boldsymbol{E_1})\det(\boldsymbol{A})$.

If $\boldsymbol{A}$ is invertible, then by Theorem 2.4.7, $\boldsymbol{B}$ is an identity matrix and hence $\det(\boldsymbol{A})$ cannot be zero.

Suppose $\boldsymbol{A}$ is singular. Then $\boldsymbol{B}$ has a row consisting entirely of zeros. So $\det(\boldsymbol{B}) = 0$ (why?). Since $\det(\boldsymbol{E_i}) \neq 0$ for all i, $\det(\boldsymbol{A})$ must be zero.

Example 2.5.20

1. Let $\boldsymbol{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. By Example 2.5.4.1, $\det(\boldsymbol{A}) = ad - bc$. Then by Theorem 2.5.19, $\boldsymbol{A}$ is invertible if and only if $ad - bc \neq 0$. The result is the same as in Example 2.4.11.2.

2. The matrices $\boldsymbol{B}$ and $\boldsymbol{C}$ in Example 2.5.4.2 and Example 2.5.4.3, respectively, are invertible.

Remark 2.5.21 By Theorem 2.5.19, we can use the determinant to check whether a given square matrix is invertible. However, for matrices of higher order, to compute determinants may not be easier than the method discussed in Remark 2.4.10.

Theorem 2.5.22 Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two square matrices of order n and c a scalar. Then

1. $\det(c\boldsymbol{A}) = c^n \det(\boldsymbol{A})$;
2. $\det(\boldsymbol{AB}) = \det(\boldsymbol{A})\det(\boldsymbol{B})$; and
3. if $\boldsymbol{A}$ is invertible, $\det(\boldsymbol{A}^{-1}) = \dfrac{1}{\det(\boldsymbol{A})}$.

Proof

1. Since $c\boldsymbol{A}$ is obtained from $\boldsymbol{A}$ by multiplying c to every row of $\boldsymbol{A}$, by Theorem 2.5.15.1, $\det(c\boldsymbol{A}) = c^n \det(\boldsymbol{A})$.
2. If $\boldsymbol{A}$ is singular, by Theorem 2.4.14, $\boldsymbol{AB}$ is singular and hence
$$\det(\boldsymbol{AB}) = 0 = \det(\boldsymbol{A})\det(\boldsymbol{B}).$$
Suppose $\boldsymbol{A}$ is invertible. By Theorem 2.4.7, $\boldsymbol{A} = \boldsymbol{E}_1\boldsymbol{E}_2\cdots\boldsymbol{E}_k$ where $\boldsymbol{E}_1, \boldsymbol{E}_2, \ldots, \boldsymbol{E}_k$ are elementary matrices. Then by using Theorem 2.5.15.4 repeatedly,
$$\begin{aligned}\det(\boldsymbol{AB}) &= \det(\boldsymbol{E}_1\boldsymbol{E}_2\cdots\boldsymbol{E}_k\boldsymbol{B})\\ &= \det(\boldsymbol{E}_1)\det(\boldsymbol{E}_2)\cdots\det(\boldsymbol{E}_k)\det(\boldsymbol{B})\\ &= \det(\boldsymbol{E}_1\boldsymbol{E}_2\cdots\boldsymbol{E}_k)\det(\boldsymbol{B})\\ &= \det(\boldsymbol{A})\det(\boldsymbol{B}).\end{aligned}$$
3. Suppose $\boldsymbol{A}$ is invertible. Since $\boldsymbol{AA}^{-1} = \boldsymbol{I}$,
$$\det(\boldsymbol{A})\det(\boldsymbol{A}^{-1}) = \det(\boldsymbol{AA}^{-1}) = \det(\boldsymbol{I}) = 1.$$
So $\det(\boldsymbol{A}^{-1}) = \dfrac{1}{\det(\boldsymbol{A})}$.

Example 2.5.23 Let $\boldsymbol{A} = \begin{pmatrix} -3 & -2 & 4 \\ 4 & 3 & 1 \\ 0 & 2 & 4 \end{pmatrix}$. Note that $\det(\boldsymbol{A}) = 34$.

1. By Theorem 2.5.22.1, $\det(2\boldsymbol{A}) = 2^3\det(\boldsymbol{A}) = 2^3 \cdot 34 = 272$.

2. Let $\boldsymbol{B} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 2 \end{pmatrix}$. Note that $\det(\boldsymbol{B}) = -1$.

   By Theorem 2.5.22.2, $\det(\boldsymbol{AB}) = \det(\boldsymbol{A})\det(\boldsymbol{B}) = -34$.

3. By Theorem 2.5.22.3, $\det(\boldsymbol{A}^{-1}) = \dfrac{1}{\det(\boldsymbol{A})} = \dfrac{1}{34}$.

(As an exercise, verify the results above by computing $\det(2\boldsymbol{A})$, $\det(\boldsymbol{AB})$ and $\det(\boldsymbol{A}^{-1})$ directly.)

Definition 2.5.24 Let $\boldsymbol{A}$ be a square matrix of order n. Then the (*classical*) *adjoint* of $\boldsymbol{A}$ is the $n \times n$ matrix

$$\mathbf{adj}(\boldsymbol{A}) = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{pmatrix}^{\mathrm{T}} = \begin{pmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \vdots & \vdots & & \vdots \\ A_{1n} & A_{2n} & \cdots & A_{nn} \end{pmatrix}$$

where A_{ij} is the (i, j)-cofactor of $\boldsymbol{A}$.

Theorem 2.5.25 Let $\boldsymbol{A}$ be a square matrix. If $\boldsymbol{A}$ is invertible, then

$$\boldsymbol{A}^{-1} = \frac{1}{\det(\boldsymbol{A})}\mathbf{adj}(\boldsymbol{A}).$$

Proof Let $\boldsymbol{A} = (a_{ij})$ be an $n \times n$ matrix and let $\boldsymbol{A}\,[\mathbf{adj}(\boldsymbol{A})] = (b_{ij})$. Then

$$b_{ij} = a_{i1}A_{j1} + a_{i2}A_{j2} + \cdots + a_{in}A_{jn}.$$

By Theorem 2.5.6, $b_{ii} = \det(\boldsymbol{A})$. By Equation (2.4) in the proof of Theorem 2.5.15.3, $b_{ij} = 0$ if $i \neq j$. So

$$\boldsymbol{A}\,[\mathbf{adj}(\boldsymbol{A})] = \det(\boldsymbol{A})\boldsymbol{I} \;\Rightarrow\; \boldsymbol{A}\left[\frac{1}{\det(\boldsymbol{A})}\mathbf{adj}(\boldsymbol{A})\right] = \boldsymbol{I}.$$

By Theorem 2.4.12, $\boldsymbol{A}^{-1} = \dfrac{1}{\det(\boldsymbol{A})}\mathbf{adj}(\boldsymbol{A})$.

Example 2.5.26

1. Let $\boldsymbol{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ where $ad - bc \neq 0$. Then

$$\mathbf{adj}(\boldsymbol{A}) = \begin{pmatrix} d & -c \\ -b & a \end{pmatrix}^{\mathrm{T}} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

   Since $\det(\boldsymbol{A}) = ad - bc$, $\boldsymbol{A}^{-1} = \dfrac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$.

2. Let $\boldsymbol{B} = \begin{pmatrix} 1 & -1 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 3 \end{pmatrix}$. Then

$$\mathbf{adj}(\boldsymbol{B}) = \begin{pmatrix} \begin{vmatrix} -1 & 0 \\ 0 & 3 \end{vmatrix} & -\begin{vmatrix} 0 & 0 \\ 1 & 3 \end{vmatrix} & \begin{vmatrix} 0 & -1 \\ 1 & 0 \end{vmatrix} \\ -\begin{vmatrix} -1 & 1 \\ 0 & 3 \end{vmatrix} & \begin{vmatrix} 1 & 1 \\ 1 & 3 \end{vmatrix} & -\begin{vmatrix} 1 & -1 \\ 1 & 0 \end{vmatrix} \\ \begin{vmatrix} -1 & 1 \\ -1 & 0 \end{vmatrix} & -\begin{vmatrix} 1 & 1 \\ 0 & 0 \end{vmatrix} & \begin{vmatrix} 1 & -1 \\ 0 & -1 \end{vmatrix} \end{pmatrix}^{\mathrm{T}} = \begin{pmatrix} -3 & 0 & 1 \\ 3 & 2 & -1 \\ 1 & 0 & -1 \end{pmatrix}^{\mathrm{T}} = \begin{pmatrix} -3 & 3 & 1 \\ 0 & 2 & 0 \\ 1 & -1 & -1 \end{pmatrix}.$$

Since $\det(\boldsymbol{B}) = -2$, $\boldsymbol{B}^{-1} = -\dfrac{1}{2}\begin{pmatrix} -3 & 3 & 1 \\ 0 & 2 & 0 \\ 1 & -1 & -1 \end{pmatrix}$.

Theorem 2.5.27 (Cramer's Rule) Suppose $\boldsymbol{Ax} = \boldsymbol{b}$ is a linear system where $\boldsymbol{A}$ is an $n \times n$ matrix. Let $\boldsymbol{A_i}$ be the matrix obtained from $\boldsymbol{A}$ by replacing the ith column of $\boldsymbol{A}$ by $\boldsymbol{b}$. If A is invertible, then the system has only one solution

$$\boldsymbol{x} = \frac{1}{\det(\boldsymbol{A})}\begin{pmatrix} \det(\boldsymbol{A_1}) \\ \det(\boldsymbol{A_2}) \\ \vdots \\ \det(\boldsymbol{A_n}) \end{pmatrix}.$$

Proof Let $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ and $\boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$. Since

$$\boldsymbol{Ax} = \boldsymbol{b} \quad \Leftrightarrow \quad \boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b} = \frac{1}{\det(\boldsymbol{A})}\,[\mathbf{adj}(\boldsymbol{A})]\,\boldsymbol{b},$$

we have

$$x_i = \frac{b_1 A_{1i} + b_2 A_{2i} + \cdots + b_n A_{ni}}{\det(\boldsymbol{A})} = \frac{\det(\boldsymbol{A_i})}{\det(\boldsymbol{A})}$$

for $i = 1, 2, \ldots, n$.

Example 2.5.28 Use Cramer's rule to solve the system of linear equations

$$\begin{cases} x + \ \ y + 3z = 0 \\ 2x - 2y + 2z = 4 \\ 3x + 9y \qquad = 3. \end{cases}$$

Solution Rewrite the linear system as $\begin{pmatrix} 1 & 1 & 3 \\ 2 & -2 & 2 \\ 3 & 9 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 4 \\ 3 \end{pmatrix}$. Then by Cramer's rule,

$$x = \frac{\begin{vmatrix} 0 & 1 & 3 \\ 4 & -2 & 2 \\ 3 & 9 & 0 \end{vmatrix}}{\begin{vmatrix} 1 & 1 & 3 \\ 2 & -2 & 2 \\ 3 & 9 & 0 \end{vmatrix}} = 2.2, \qquad y = \frac{\begin{vmatrix} 1 & 0 & 3 \\ 2 & 4 & 2 \\ 3 & 3 & 0 \end{vmatrix}}{\begin{vmatrix} 1 & 1 & 3 \\ 2 & -2 & 2 \\ 3 & 9 & 0 \end{vmatrix}} = -0.4, \qquad z = \frac{\begin{vmatrix} 1 & 1 & 0 \\ 2 & -2 & 4 \\ 3 & 9 & 3 \end{vmatrix}}{\begin{vmatrix} 1 & 1 & 3 \\ 2 & -2 & 2 \\ 3 & 9 & 0 \end{vmatrix}} = -0.6.$$

Exercise 2

Question 2.1 to Question 2.24 are exercises for Sections 2.1 and 2.2.

1. Let $\boldsymbol{A} = (a_{ij})_{3\times 4}$, where $a_{ij} = 2i - 3j$, $\boldsymbol{B} = \boldsymbol{I}_4$, $\boldsymbol{C} = \boldsymbol{0}_{3\times 3}$,

$$\boldsymbol{D} = (d_{ij})_{4\times 3} \text{ where } d_{ij} = \begin{cases} -1 & \text{if } i+j \text{ is even} \\ 1 & \text{if } i+j \text{ is odd.} \end{cases}$$

$$\boldsymbol{E} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{pmatrix}, \quad \boldsymbol{F} = \begin{pmatrix} 5 & -1 \\ 9 & 1 \\ 2 & 0 \end{pmatrix}, \quad \text{and} \quad \boldsymbol{G} = \begin{pmatrix} 1 \\ -1 \\ 3 \\ 2 \end{pmatrix}.$$

Evaluate the following, wherever possible.

(a) $\boldsymbol{AD}$, (b) $\boldsymbol{DA} - 3\boldsymbol{B}$, (c) $\boldsymbol{D}^2$, (d) $\boldsymbol{E}^2 + \boldsymbol{C}^3$,
(e) $\boldsymbol{DE} + 2\boldsymbol{D}$, (f) $\boldsymbol{EA}$, (g) $\boldsymbol{DB}$, (h) $\boldsymbol{CF}$,
(i) $\boldsymbol{AG}$, (j) $\boldsymbol{FE}$, (k) $\boldsymbol{EF}$, (l) $\boldsymbol{CA}$,
(m) $\boldsymbol{E} - \boldsymbol{E}^{\mathrm{T}}$, (n) $\boldsymbol{F} - \boldsymbol{F}^{\mathrm{T}}$, (o) $\boldsymbol{GG}^{\mathrm{T}}$, (p) $\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}$.

2. Solve the following matrix equation for a, b, c and d :

$$\begin{pmatrix} a-b & a+c \\ -a+c & a+b-d \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}.$$

3. The symbol $\sum$ is used to denote the sum of a sequence of numbers. For example,

$$\sum_{i=1}^{n} a_i = a_1 + a_2 + \cdots + a_n,$$

$$\sum_{x=0}^{m} f(x) = f(0) + f(1) + \cdots + f(m),$$

$$\sum_{k=1}^{r} c_{ik}d_{kj} = c_{i1}d_{1j} + c_{i2}d_{2j} + \cdots + c_{ir}d_{rj}.$$

Let $\boldsymbol{A} = (a_{ij})$ be an $m \times n$ matrix and $\boldsymbol{B} = (b_{ij})$ an $n \times m$ matrix, with $m, n \geq 5$.

(a) Each of the following sums represents an entry of either $\boldsymbol{AB}$ or $\boldsymbol{BA}$. Determine which matrix product is involved and which entry of that product is represented in each case:

(i) $\displaystyle\sum_{k=1}^{n} a_{3k}b_{k4}$, (ii) $\displaystyle\sum_{p=1}^{n} a_{4p}b_{p1}$,

(iii) $\displaystyle\sum_{q=1}^{m} a_{q2}b_{3q}$, (vi) $\displaystyle\sum_{x=1}^{m} b_{2x}a_{x5}$.

(b) Use the symbol $\sum$ to express the following entries symbolically.

(i) In $\boldsymbol{AB}$, the entry in the third row and second column.

(ii) In $\boldsymbol{BA}$, the entry in the fourth row and first column.

4. Given $\boldsymbol{A} = (a_{ij})_{n\times p}$, $\boldsymbol{B} = (b_{ij})_{p\times q}$ and $\boldsymbol{C} = (c_{ij})_{p\times p}$, write down the (i, j) entries of

(a) $\boldsymbol{AB}$, (b) $\boldsymbol{C}^2$, (c) $\boldsymbol{AC}^{\mathrm{T}}$.

5. Find an example of a nonzero 3×3 matrix $\boldsymbol{A}$ such that $\boldsymbol{A}^{\mathrm{T}} = -\boldsymbol{A}$. What is the general form of the matrix $\boldsymbol{A}$?

6. Find examples of nonzero 3×3 matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ for each of the following cases:

(a) $\boldsymbol{AB} = \boldsymbol{0}$, (b) $\boldsymbol{AB} \neq \boldsymbol{BA}$, (c) $\boldsymbol{BA} = \boldsymbol{CA}$ but $\boldsymbol{B} \neq \boldsymbol{C}$.

7. Give an example of a 2×3 matrix $\boldsymbol{A}$ such that the solution set of the linear system $\boldsymbol{Ax} = \boldsymbol{0}$ is the plane $2x + 3y - z = 0$.

8. Let S be the set of points (x, y, z) of the form

$$\begin{cases} x = t + 1 \\ y = t \\ z = 3 \end{cases} \qquad \text{where } t \text{ is an arbitrary parameter.}$$

(a) Describe S geometrically.

(b) Find an example of a 2×3 matrix $\boldsymbol{A}$ and a 2×1 matrix $\boldsymbol{b}$ such that the solution set of the linear system $\boldsymbol{Ax} = \boldsymbol{b}$ is S. Give a geometric interpretation for the linear system.

9. Suppose the homogeneous system $\boldsymbol{Ax} = \boldsymbol{0}$ has non-trivial solution. Show that the linear system $\boldsymbol{Ax} = \boldsymbol{b}$ has either no solution or infinitely many solutions.

10. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be $m \times n$ and $n \times p$ matrices respectively.

(a) Suppose the homogeneous linear system $\boldsymbol{Bx} = \boldsymbol{0}$ has infinitely many solutions. How many solutions does the system $\boldsymbol{ABx} = \boldsymbol{0}$ have?

(b) Suppose $\boldsymbol{Bx} = \boldsymbol{0}$ has only the trivial solution. Can we tell how many solutions are there for $\boldsymbol{ABx} = \boldsymbol{0}$?

11. Let $\boldsymbol{A} = (a_{ij})_{n \times n}$ be an square matrix. The *trace* of $\boldsymbol{A}$, denoted by $\operatorname{tr}(\boldsymbol{A})$, is defined to be the sum of the entries on the diagonal of $\boldsymbol{A}$, i.e.

$$\operatorname{tr}(\boldsymbol{A}) = a_{11} + a_{22} + \cdots + a_{nn} = \sum_{i=1}^{n} a_{ii}.$$

(a) Find the trace of each of the following square matrices.

$$\text{(i)} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \quad \text{(ii)} \begin{pmatrix} -1 & 3 & -4 \\ 2 & 4 & 1 \\ -4 & 2 & -9 \end{pmatrix}, \quad \text{(iii)} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 1 & 3 & 5 & 0 \\ 1 & 3 & 5 & 7 \end{pmatrix}.$$

(b) Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be any square matrices of the same size. Show that

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

(c) Let $\boldsymbol{A}$ be any square matrix and c a scalar. Show that

$$\operatorname{tr}(c\boldsymbol{A}) = c \operatorname{tr}(\boldsymbol{A}).$$

(d) Let $\boldsymbol{C}$ and $\boldsymbol{D}$ be $m \times n$ and $n \times m$ matrices respectively. Show that

$$\operatorname{tr}(\boldsymbol{CD}) = \operatorname{tr}(\boldsymbol{DC}).$$

(e) Show that there are no square matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ such that $\boldsymbol{AB} - \boldsymbol{BA} = \boldsymbol{I}$.

12. A square matrix $\boldsymbol{A}$ is called *orthogonal* if

$$\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{I} \qquad \text{and} \qquad \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{I}.$$

   (a) Determine which of the following matrices are orthogonal:

$$\text{(i)} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \text{(ii)} \begin{pmatrix} \frac{3}{5} & \frac{4}{5} \\ \frac{4}{5} & -\frac{3}{5} \end{pmatrix}.$$

   (b) Suppose $\boldsymbol{A}$ and $\boldsymbol{B}$ are orthogonal matrices of the same size. Show that $\boldsymbol{AB}$ is orthogonal.

13. A square matrix $\boldsymbol{A}$ is called *nilpotent* if $\boldsymbol{A}^k = \boldsymbol{0}$ for some positive integer k.

   (a) Show that $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and $\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ are nilpotent.

   (b) Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be square matrices of the same size such that $\boldsymbol{AB} = \boldsymbol{BA}$ and $\boldsymbol{A}$ is nilpotent. Show that $\boldsymbol{AB}$ is nilpotent.

   (c) If $\boldsymbol{AB} \neq \boldsymbol{BA}$ in Part (b), must $\boldsymbol{AB}$ be nilpotent?

14. Let $\boldsymbol{A} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Consider the matrix equation

$$\boldsymbol{AX} = \boldsymbol{XA} \tag{2.5}$$

   where $\boldsymbol{X}$ is a 2×2 unknown matrix.

   (a) Determine which of the following matrices satisfy Equation (2.5):

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

   (b) Prove that if $\boldsymbol{P}$ and $\boldsymbol{Q}$ satisfy Equation (2.5), then $\boldsymbol{P} + \boldsymbol{Q}$ and $\boldsymbol{PQ}$ also satisfy Equation (2.5).

   (c) Find conditions on p, q, r, s which determine precisely when $\begin{pmatrix} p & q \\ r & s \end{pmatrix}$ satisfy Equation (2.5).

15. (a) Let $\boldsymbol{D} = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix}$ where a, b, c are real numbers. Show that, for all positive integer k, $\boldsymbol{D}^k = \begin{pmatrix} a^k & 0 & 0 \\ 0 & b^k & 0 \\ 0 & 0 & c^k \end{pmatrix}$.

(b) Find a diagonal matrix $\boldsymbol{A}$ such that $\boldsymbol{A}^5 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$.

(c) Find all diagonal matrices $\boldsymbol{B}$ such that $\boldsymbol{B}^2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 9 \end{pmatrix}$.

16. Let $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ be three nonzero $n \times n$ matrices. Suppose $\boldsymbol{AB} = \boldsymbol{BA}$ and $\boldsymbol{AC} = \boldsymbol{CA}$.

    (a) Is $\boldsymbol{CB} = \boldsymbol{BC}$?

    (b) Among the matrices $\boldsymbol{ABC}, \boldsymbol{ACB}, \boldsymbol{BAC}, \boldsymbol{BCA}, \boldsymbol{CAB}, \boldsymbol{CBA}$, which of them are equal to one another?

    Justify your answers.

17. Consider the population of certain endangered species of wild animals: On the average, each adult will give birth to one baby each year; 50% of the new born babies will survive the first year; 60% of the one-year-old cubs will survive the second year and become adults; and 70% of the adults will survive each year.

    Define $\boldsymbol{A} = \begin{pmatrix} 0 & 0 & 1 \\ 0.5 & 0 & 0 \\ 0 & 0.6 & 0.7 \end{pmatrix}$. Let x_0, y_0 and z_0 be the numbers of babies, one-year-old cubs and adults, respectively, at the end of a particular year.

    (a) Let $\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} = \boldsymbol{A} \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix}$. What information do the numbers x_1, y_1 and z_1 give us?

    (b) Let $\begin{pmatrix} x_n \\ y_n \\ z_n \end{pmatrix} = \boldsymbol{A}^n \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix}$ where n is a positive integer. Interpret the numbers x_n, y_n and z_n.

    (c) Suppose initially, $x_0 = 0$, $y_0 = 0$ and $z_0 = 100$. What is the total population three years later?

18. Complete the proof of Theorem 2.2.6:

    Let $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ be matrices of the same size and c, d scalars. Show that

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

    (b) $c(\boldsymbol{A} + \boldsymbol{B}) = c\boldsymbol{A} + c\boldsymbol{B}$,

    (c) $(c + d)\boldsymbol{A} = c\boldsymbol{A} + d\boldsymbol{A}$,

    (d) $c(d\boldsymbol{A}) = (cd)\boldsymbol{A} = d(c\boldsymbol{A})$,

    (e) $\boldsymbol{A} + \boldsymbol{0} = \boldsymbol{0} + \boldsymbol{A} = \boldsymbol{A}$,

    (f) $\boldsymbol{A} - \boldsymbol{A} = \boldsymbol{0}$,

    (g) $0\boldsymbol{A} = \boldsymbol{0}$.

19. Complete the proof of Theorem 2.2.11:

 (a) If $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ are $m \times p$, $p \times q$ and $q \times n$ matrices respectively, show that
$$\boldsymbol{A}(\boldsymbol{BC}) = (\boldsymbol{AB})\boldsymbol{C}.$$

 (b) If $\boldsymbol{A}$, $\boldsymbol{C_1}$ and $\boldsymbol{C_2}$ are $p \times n$, $m \times p$ and $m \times p$ matrices respectively, show that
$$(\boldsymbol{C_1} + \boldsymbol{C_2})\boldsymbol{A} = \boldsymbol{C_1 A} + \boldsymbol{C_2 A}.$$

 (c) If $\boldsymbol{A}$, $\boldsymbol{B}$ are $m \times p$, $p \times n$ matrices, respectively, and c is a scalar, show that
$$c(\boldsymbol{AB}) = (c\boldsymbol{A})\boldsymbol{B} = \boldsymbol{A}(c\boldsymbol{B}).$$

 (d) If $\boldsymbol{A}$ is an $m \times n$ matrix, show that
$$\boldsymbol{A0}_{n\times q} = \boldsymbol{0}_{m\times q}, \quad \boldsymbol{0}_{p\times m}\boldsymbol{A} = \boldsymbol{0}_{p\times n} \quad \text{and} \quad \boldsymbol{AI}_n = \boldsymbol{I}_m\boldsymbol{A} = \boldsymbol{A}.$$

20. Complete the proof of Theorem 2.2.22:

 Let $\boldsymbol{A}$ be an $m \times n$ matrix.

 (a) Show that $(\boldsymbol{A}^{\mathrm{T}})^{\mathrm{T}} = \boldsymbol{A}$.
 (b) If $\boldsymbol{B}$ is an $m \times n$ matrix, show that $(\boldsymbol{A} + \boldsymbol{B})^{\mathrm{T}} = \boldsymbol{A}^{\mathrm{T}} + \boldsymbol{B}^{\mathrm{T}}$.
 (c) If c is a scalar, show that $(c\boldsymbol{A})^{\mathrm{T}} = c\boldsymbol{A}^{\mathrm{T}}$.

21. Given that $\boldsymbol{A}$ is a 3×3 matrix such that
$$\boldsymbol{A}\begin{pmatrix}1\\0\\0\end{pmatrix} = \begin{pmatrix}1\\1\\1\end{pmatrix} \quad \text{and} \quad \boldsymbol{A}\begin{pmatrix}0\\0\\1\end{pmatrix} = \begin{pmatrix}0\\0\\1\end{pmatrix}.$$

 Find a matrix $\boldsymbol{X}$ such that
$$\boldsymbol{AX} = \begin{pmatrix}1 & 0 & 4\\1 & 0 & 4\\1 & 0 & 7\end{pmatrix}.$$

 (Hint: Write $\boldsymbol{X} = \begin{pmatrix}\boldsymbol{x_1} & \boldsymbol{x_2} & \boldsymbol{x_3}\end{pmatrix}$ where $\boldsymbol{x_i}$ is the ith column of $\boldsymbol{X}$.)

22. Prove Remark 1.1.10:

 Show that a linear system $\boldsymbol{Ax} = \boldsymbol{b}$ has either no solution, only one solution or infinitely many solutions.

 (Hint: Suppose $\boldsymbol{Ax} = \boldsymbol{b}$ has two different solutions $\boldsymbol{u}$ and $\boldsymbol{v}$. Use $\boldsymbol{u}$ and $\boldsymbol{v}$ to construct infinitely many other solutions.)

23. Let $\boldsymbol{A}$ be an $m \times n$ matrix.

 (a) Let $\boldsymbol{B_1}$ and $\boldsymbol{B_2}$ be $n \times p$ and $n \times q$ matrices respectively. Show that
$$\boldsymbol{A}\begin{pmatrix}\boldsymbol{B_1} & \boldsymbol{B_2}\end{pmatrix} = \begin{pmatrix}\boldsymbol{AB_1} & \boldsymbol{AB_2}\end{pmatrix}.$$

(In here, $\begin{pmatrix} B_1 & B_2 \end{pmatrix}$ is an $n \times (p+q)$ matrix such that its jth column is equal to the jth column of B_1 if $j \leq p$ and equal to the $(j-p)$th column of B_2 if $j > p$.)

(b) Let C_1 and C_2 be $r \times m$ matrices. Is it true that $\begin{pmatrix} C_1 & C_2 \end{pmatrix} A = \begin{pmatrix} C_1 A & C_2 A \end{pmatrix}$?

(c) Let D_1 and D_2 be $s \times m$ and $t \times m$ matrices respectively. Show that

$$\begin{pmatrix} D_1 \\ D_2 \end{pmatrix} A = \begin{pmatrix} D_1 A \\ D_2 A \end{pmatrix}.$$

(In here, $\begin{pmatrix} D_1 \\ D_2 \end{pmatrix}$ is an $(s+t) \times m$ matrix such that its ith row is equal to the ith row of D_1 if $i \leq s$ and equal to the $(i-s)$th row of D_2 if $i > s$.)

24. Determine which of the following statements are true. Justify your answer.

(a) If A and B are two diagonal matrices of the same size, then $AB = BA$.

(b) If A is a square matrix, then $\frac{1}{2}(A + A^{\mathrm{T}})$ is symmetric.

(c) For all matrices A and B, $(A + B)^2 = A^2 + B^2 + 2AB$.

(d) If A and B are symmetric matrices of the same size, then $A - B$ is symmetric.

(e) If A and B are symmetric matrices of the same size, then AB is symmetric.

(f) If A is a square matrix such that $A^2 = 0$, then $A = 0$.

(g) If A is a matrix such that $AA^{\mathrm{T}} = 0$, then $A = 0$.

Question 2.25 to Question 2.46 are exercises for Sections 2.3 and 2.4.

25. Let $A = \begin{pmatrix} 2 & -1 & -1 \\ 0 & 3 & 1 \\ 0 & 1 & 3 \end{pmatrix}$.

(a) Verify that $A^2 - 6A + 8I = 0$.

(b) Show that $A^{-1} = \frac{1}{8}(6I - A)$ without computing the inverse of A explicitly.

26. Let A be a square matrix.

(a) Show that if $A^2 = 0$, then $I - A$ is invertible and $(I - A)^{-1} = I + A$.

(b) Show that if $A^3 = 0$, then $I - A$ is invertible and $(I - A)^{-1} = I + A + A^2$.

(c) If $A^n = 0$ for $n \geq 4$, is $I - A$ invertible?

27. (a) Give three examples of 2×2 matrices A such that $A^2 = A$.

(b) Let A be a square matrix such that $A^2 = A$. Show that $I + A$ is invertible and $(I + A)^{-1} = \frac{1}{2}(2I - A)$.

28. Determine which of the following statements are true. Justify your answer.

 (a) If $\boldsymbol{A}$ and $\boldsymbol{B}$ are invertible matrices of the same size, then $\boldsymbol{A}+\boldsymbol{B}$ is also invertible.

 (b) If $\boldsymbol{A}$ and $\boldsymbol{B}$ are singular matrices of the same size, then $\boldsymbol{A}+\boldsymbol{B}$ is also singular.

29. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be invertible matrices of the same size. Suppose $\boldsymbol{A}+\boldsymbol{B}$ is invertible. Show that $\boldsymbol{A}^{-1}+\boldsymbol{B}^{-1}$ is invertible and $(\boldsymbol{A}+\boldsymbol{B})^{-1}=\boldsymbol{A}^{-1}(\boldsymbol{A}^{-1}+\boldsymbol{B}^{-1})^{-1}\boldsymbol{B}^{-1}$.

30. Complete the proof of Theorem 2.3.9:

 Let $\boldsymbol{A}$, $\boldsymbol{B}$ be two invertible matrices of the same size and c a nonzero scalar. Show that

 (a) $c\boldsymbol{A}$ is invertible and $(c\boldsymbol{A})^{-1}=\frac{1}{c}\boldsymbol{A}^{-1}$.

 (b) $\boldsymbol{A}^{-1}$ is invertible and $(\boldsymbol{A}^{-1})^{-1}=\boldsymbol{A}$.

 (c) $\boldsymbol{AB}$ is invertible and $(\boldsymbol{AB})^{-1}=\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$.

31. (a) Let $\boldsymbol{A}$, $\boldsymbol{P}$ and $\boldsymbol{D}$ be square matrices of the same size such that $\boldsymbol{A}=\boldsymbol{PDP}^{-1}$. Show that $\boldsymbol{A}^k=\boldsymbol{PD}^k\boldsymbol{P}^{-1}$ for all positive integer k.

 (b) Let $\boldsymbol{A}=\begin{pmatrix}-7 & 5\\ -10 & 8\end{pmatrix}$. Verify that

$$\boldsymbol{A}=\begin{pmatrix}1 & 1\\ 1 & 2\end{pmatrix}\begin{pmatrix}-2 & 0\\ 0 & 3\end{pmatrix}\begin{pmatrix}1 & 1\\ 1 & 2\end{pmatrix}^{-1}.$$

 Hence or otherwise, find $\boldsymbol{A}^{10}$.

32. Express the matrix $\boldsymbol{A}=\begin{pmatrix}5 & -2 & 6 & 0\\ -2 & 1 & 3 & 1\end{pmatrix}$ in the form $\boldsymbol{E_1E_2\cdots E_nR}$ where $\boldsymbol{E_1}$, $\boldsymbol{E_2}$, $\dots$, $\boldsymbol{E_n}$ are elementary matrices and $\boldsymbol{R}$ is the reduced row-echelon form of $\boldsymbol{A}$.

33. Let $\boldsymbol{A}$ be the 4×4 matrix obtained from $\boldsymbol{I}$ by the following sequence of elementary row operations:

$$\boldsymbol{I}\xrightarrow{\frac{1}{2}R_2}\xrightarrow{R_1-R_2}\xrightarrow{R_2\leftrightarrow R_4}\xrightarrow{R_3+3R_1}\boldsymbol{A}$$

 (a) Write $\boldsymbol{A}$ as a product of four elementary matrices.

 (b) Write $\boldsymbol{A}^{-1}$ as a product of four elementary matrices.

34. Let $\boldsymbol{A}$, $\boldsymbol{B}$ be 3×3 matrices such that $\boldsymbol{A} = \boldsymbol{E_1 E_2 E_3 E_4 B}$ where

$$\boldsymbol{E_1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \quad \boldsymbol{E_2} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{pmatrix}, \quad \boldsymbol{E_3} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad \boldsymbol{E_4} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

(a) Describe how $\boldsymbol{A}$ is obtained from $\boldsymbol{B}$ by elementary row operations.

(b) If $\boldsymbol{A}$ is invertible, is $\boldsymbol{B}$ invertible? Justify your answer.

35. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be 4×4 matrices such that $\boldsymbol{E_1 E_2 A} = \boldsymbol{E_3 E_4 B}$ where

$$\boldsymbol{E_1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad \boldsymbol{E_2} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$\boldsymbol{E_3} = \begin{pmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad \boldsymbol{E_4} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

Describe how $\boldsymbol{B}$ can be obtained from $\boldsymbol{A}$ by elementary row operations.

36. (a) Let $\boldsymbol{A} = \begin{pmatrix} a & b \\ 0 & c \end{pmatrix}$ where $ac \neq 0$. Express $\boldsymbol{A}$ as a product of three elementary matrices.

(b) Let $\boldsymbol{B} = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 3 \\ -1 & -1 & 4 \end{pmatrix}$. Express $\boldsymbol{B}$ as a product of four elementary matrices.

37. For each of the following matrices, determine if the matrix is invertible. Also, for each of the invertible matrices, find the inverse of the matrix.

(a) $\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$, (b) $\begin{pmatrix} -1 & 3 & -4 \\ 2 & 4 & 1 \\ -4 & 2 & -9 \end{pmatrix}$, (c) $\begin{pmatrix} 2 & 0 & 1 \\ 0 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}$;

(d) $\begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 2 & 0 & 0 \\ 1 & 2 & 3 & 0 \\ 1 & 2 & 3 & 4 \end{pmatrix}$, (e) $\begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & 2 & 6 & 3 \\ 1 & -2 & -6 & -4 \\ 1 & 1 & 1 & 0 \end{pmatrix}$, (f) $\begin{pmatrix} 1 & 3 & 1 & 1 \\ 2 & 5 & 2 & 2 \\ 1 & 3 & 8 & 9 \\ 1 & 3 & 2 & 2 \end{pmatrix}$.

38. Solve the matrix equation $\begin{pmatrix} 2 & 1 & 1 \\ 0 & 1 & 2 \\ 1 & 3 & 2 \end{pmatrix} \boldsymbol{X} = \begin{pmatrix} 2 & 3 & 4 & 1 \\ 1 & 0 & 3 & 7 \\ 2 & 1 & 1 & 2 \end{pmatrix}$.

39. A manufacturer makes three types of chairs A, B, C. The company has available 260 units of wood, 60 units of upholstery and 240 units of labor. The manufacturer wants a production schedule that uses all of these resources. The various products require the following amounts of resources.

	A	B	C
Wood	4	4	3
Upholstery	0	1	2
Labor	2	4	5

(a) Find the inverse of the data matrix above and hence determine how many pieces of each product should be manufactured.

(b) If the amount of wood is increased by 10 units, how will this change the number of type C chairs produced?

40. Determine the value(s) of a so that the matrix $\begin{pmatrix} 1 & 0 & a \\ 0 & a & 1 \\ a & 1 & 0 \end{pmatrix}$ is invertible. What is the inverse of the matrix if it exists?

41. (a) Determine the values of a, b and c so that the homogeneous system

$$\begin{cases} x + y + z = 0 \\ ax + by + cz = 0 \\ a^2x + b^2y + c^2z = 0 \end{cases}$$

has non-trivial solution.

(b) Write down the conditions so that the matrix $\begin{pmatrix} 1 & 1 & 1 \\ a & b & c \\ a^2 & b^2 & c^2 \end{pmatrix}$ is invertible.

42. Prove Theorem 2.4.14:

Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two square matrices of the same order. Prove that if $\boldsymbol{A}$ is singular, then $\boldsymbol{AB}$ and $\boldsymbol{BA}$ are singular. (Since we use Theorem 2.4.14 to prove Theorem 2.5.22.2, we cannot use determinants to do this question. Work out the proof using the definition of inverses together with Theorem 2.4.12.)

43. Let $\boldsymbol{A}$ be an $m \times n$ matrix which is row equivalent to the following matrix:

$$\begin{pmatrix} \boldsymbol{R} \\ 0 \cdots 0 \end{pmatrix}$$

where the last row is a zero row and $\boldsymbol{R}$ is an $(m-1) \times n$ matrix. Show that there exists an $n \times 1$ matrix $\boldsymbol{b}$ such that the linear system $\boldsymbol{Ax} = \boldsymbol{b}$ is inconsistent.

(Hint: If $\boldsymbol{A}$ is row equivalent to a matrix $\boldsymbol{C}$, then $\boldsymbol{A} = \boldsymbol{E_k} \cdots \boldsymbol{E}_1\boldsymbol{C}$ for some elementary matrices $\boldsymbol{E}_1, \ldots, \boldsymbol{E_k}$.)

44. Let $\boldsymbol{A}$ be an $m \times n$ matrix and $\boldsymbol{B}$ an $n \times m$ matrix.

    (a) Suppose $\boldsymbol{A}$ is the matrix described in Question 2.43. Show that $\boldsymbol{AB}$ is singular.

    (b) If $m > n$, can $\boldsymbol{AB}$ be invertible? Justify your answer. (Hint: How will a row-echelon form of $\boldsymbol{A}$ look like if $m > n$?)

    (c) When $m = 2$ and $n = 3$, give an example of $\boldsymbol{A}$ and $\boldsymbol{B}$ such that $\boldsymbol{AB}$ is invertible.

45. Let $\boldsymbol{A}$ be a square matrix and let $\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_n$ be elementary row operations such that

$$\boldsymbol{A} \xrightarrow{\mathcal{R}_1} \xrightarrow{\mathcal{R}_2} \overset{\cdots}{\cdots} \xrightarrow{\mathcal{R}_n} \boldsymbol{I}.$$

Suppose $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_n$ are the elementary column operations (see Discussion 2.4.15) such that $\mathcal{R}_i$ and $\mathcal{C}_i$ correspond to the same elementary matrix for each i.

Show that

$$\boldsymbol{A} \xrightarrow{\mathcal{C}_n} \xrightarrow{\mathcal{C}_{n-1}} \overset{\cdots}{\cdots} \xrightarrow{\mathcal{C}_1} \boldsymbol{I}$$

46. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two invertible matrices of the same size. In each of the following cases, describe how $\boldsymbol{B}^{-1}$ is related to $\boldsymbol{A}^{-1}$.

    (a) $\boldsymbol{B}$ can be obtained from $\boldsymbol{A}$ by multiplying a constant to a row.

    (b) $\boldsymbol{B}$ can be obtained from $\boldsymbol{A}$ by interchanging two rows.

    (c) $\boldsymbol{B}$ can be obtained from $\boldsymbol{A}$ by adding a multiple of a row to another row.

Question 2.47 to Question 2.61 are exercises for Section 2.5.

47. For each of the following matrices,

$$\text{(a)} \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \quad \text{(b)} \begin{pmatrix} -1 & 3 & -4 \\ 2 & 4 & 1 \\ -4 & 2 & -9 \end{pmatrix}, \quad \text{(c)} \begin{pmatrix} 2 & 0 & 1 \\ 0 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}, \quad \text{(d)} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 2 & 0 & 0 \\ 1 & 2 & 3 & 0 \\ 1 & 2 & 3 & 4 \end{pmatrix}.$$

(i) compute the determinant using cofactor expansion;

(ii) compute the determinant using the method discussed in Remark 2.5.16; and

(iii) if the matrix is invertible, compute the inverse using Theorem 2.5.25.

48. Solve the following linear systems using Cramer's Rule.

(a) $\begin{cases} 9x + y = 8 \\ x - 9y = 10 \end{cases}$ (b) $\begin{cases} x - y = 0 \\ y - z = -1 \\ x + z = 2 \end{cases}$

(c) $\begin{cases} x + y + z = -1 \\ 2x - y - z = 4 \\ x + 2y - 3z = 7 \end{cases}$ (d) $\begin{cases} w - x = 0 \\ 2w + x - y = 0 \\ 3w + 2x + y - z = 1 \\ 4w + 3x + 2y + z = -1 \end{cases}$

49. Let $\boldsymbol{A} = \begin{pmatrix} a & b & c \\ 0 & b & c \\ 0 & 0 & c \end{pmatrix}$.

(a) Find $\det(\boldsymbol{A})$.

(b) Determine the values of a, b, c for which $\boldsymbol{A}$ is invertible and find $\boldsymbol{A}^{-1}$.

50. Let

$$\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 3 & -1 \\ -2 & 1 & 0 & -2 \\ 0 & 0 & 2 & 1 \end{pmatrix}, \quad \boldsymbol{C} = \begin{pmatrix} -1 & 3 & 4 & -2 \\ 0 & 10 & 1 & 1 \\ 0 & 0 & 0 & 5 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} 2 \\ 4 \\ 6 \\ 8 \end{pmatrix}, \quad \boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}.$$

(a) Find $\det(\boldsymbol{C})$.

(b) Without computing the matrix $\boldsymbol{AC}$, explain why the homogeneous linear system $\boldsymbol{ACx} = \boldsymbol{0}$ has infinitely many solutions.

51. Find all values of λ for which $\det(\boldsymbol{A}) = 0$.

(a) $\boldsymbol{A} = \begin{pmatrix} \lambda - 2 & 1 \\ -5 & \lambda + 4 \end{pmatrix}$, (b) $\boldsymbol{A} = \begin{pmatrix} \lambda - 4 & 0 & 0 \\ 0 & \lambda & 2 \\ 0 & 3 & \lambda - 1 \end{pmatrix}$;

(c) $\boldsymbol{A} = \begin{pmatrix} 1 & \lambda & \lambda & \lambda \\ 2 & \lambda & \lambda & \lambda \\ \lambda + 1 & 1 & 2 & 0 \\ 4 & 0 & 1 & 2\lambda \end{pmatrix}$, (d) $\boldsymbol{A} = \begin{pmatrix} 1 & 1 & 2 & 3 \\ 1 & 2 - \lambda^2 & 2 & 3 \\ 2 & 3 & 1 & 5 \\ 2 & 3 & 1 & 9 - \lambda^2 \end{pmatrix}$.

52. Show that $\begin{vmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{vmatrix} = (b - a)(c - a)(c - b)$.

53. Let $\boldsymbol{A}$ be a 4×4 matrix such that $\det(\boldsymbol{A}) = 9$. Find

(a) $\det(3\boldsymbol{A})$, (b) $\det(\boldsymbol{A}^{-1})$, (c) $\det(3\boldsymbol{A}^{-1})$, (d) $\det((3\boldsymbol{A})^{-1})$.

54. Let $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ be matrices such that both $\boldsymbol{A}$ and $\boldsymbol{B}$ can be obtained from $\boldsymbol{C}$ by elementary row operations:

$$\boldsymbol{C} \xrightarrow{3R_2} \xrightarrow{R_3+2R_1} \boldsymbol{A}, \qquad \boldsymbol{C} \xrightarrow{R_1+R_2} \xrightarrow{R_2 \leftrightarrow R_3} \xrightarrow{R_4-R_2} \boldsymbol{B}.$$

   (a) Describe how $\boldsymbol{A}$ can be obtained from $\boldsymbol{B}$ by elementary row operations.

   (b) Let $\boldsymbol{A} = \begin{pmatrix} 1 & -1 & 7 & \frac{1}{11} \\ 0 & 2 & 1 & 1 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & -1 \end{pmatrix}$. Find $\det(\boldsymbol{B})$.

55. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be 3×3 matrices such that

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 3 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \boldsymbol{A} = \begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \boldsymbol{B}.$$

   (a) Describe how $\boldsymbol{A}$ can be obtained from $\boldsymbol{B}$ by elementary row operations.

   (b) If $\det(\boldsymbol{A}) = 4$, find $\det(\boldsymbol{B})$.

56. Let $\boldsymbol{A} = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$ where a, b, c, d, e, f, g, h, i are either 0 or 1. Find the largest possible value and the smallest possible value of $\det(\boldsymbol{A})$.

57. Let $\boldsymbol{A}$ be a 2×2 orthogonal matrix (see Question 2.12).

   (a) Prove that $\det(\boldsymbol{A}) = \pm 1$.

   (b) If $\det(\boldsymbol{A}) = 1$, show that $\boldsymbol{A} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$ for some real number θ.

   (c) If $\det(\boldsymbol{A}) = -1$, show that $\boldsymbol{A} = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{pmatrix}$ for some real number θ.

   (Hint: If a and b are real numbers such that $a^2+b^2 = 1$, then $a = \cos(\theta)$ and $b = \sin(\theta)$ for some real number θ.)

58. Prove Theorem 2.5.12:

   (a) Prove that the determinant of a square matrix with two identical rows is zero.

   (Hint: First prove that the statement is true for 2×2 matrices. Then assume the statement is true for $k \times k$ matrices where $k \geq 2$. Let $\boldsymbol{A}$ be a $(k+1) \times (k+1)$ matrix such that the ith and jth rows of $\boldsymbol{A}$ are the same. Take any $m = 1, 2, \ldots, k+1$ and $m \neq i, j$. Compute $\det(\boldsymbol{A})$ by expanding along the mth row of $\boldsymbol{A}$.)

(b) Prove that the determinant of a square matrix with two identical columns is zero.

59. Complete the proof of Theorem 2.5.15:

Let $\boldsymbol{A}$ be a square matrix.

(a) If $\boldsymbol{B}$ is a square matrix obtained from $\boldsymbol{A}$ by multiplying one row of $\boldsymbol{A}$ by a constant k, show that $\det(\boldsymbol{B}) = k\det(\boldsymbol{A})$.

(b) If $\boldsymbol{B}$ is a square matrix obtained from $\boldsymbol{A}$ by interchanging two rows of $\boldsymbol{A}$, show that $\det(\boldsymbol{B}) = -\det(\boldsymbol{A})$.

(c) Let $\boldsymbol{E}$ be an elementary matrix of the same size as $\boldsymbol{A}$. Show that $\det(\boldsymbol{E}\boldsymbol{A}) = \det(\boldsymbol{E})\det(\boldsymbol{A})$.

60. Let $\boldsymbol{A}$ be an $n \times n$ invertible matrix.

(a) Show that $\mathbf{adj}(\boldsymbol{A})$ is invertible.

(b) Find $\det(\mathbf{adj}(\boldsymbol{A}))$ and $\mathbf{adj}(\boldsymbol{A})^{-1}$.

(c) If $\det(\boldsymbol{A}) = 1$, show that $\mathbf{adj}(\mathbf{adj}(\boldsymbol{A})) = \boldsymbol{A}$.

61. Determine which of the following statements are true. Justify your answer.

(a) If $\boldsymbol{A}$ and $\boldsymbol{B}$ are square matrices of the same size, then $\det(\boldsymbol{A}+\boldsymbol{B}) = \det(\boldsymbol{A}) + \det(\boldsymbol{B})$.

(b) If $\boldsymbol{A}$ is a square matrix, then $\det(\boldsymbol{A}+\boldsymbol{I}) = \det(\boldsymbol{A}^T + \boldsymbol{I})$.

(c) If $\boldsymbol{A}$ and $\boldsymbol{B}$ are square matrices of the same size such that $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{B}\boldsymbol{P}^{-1}$ for some invertible matrix $\boldsymbol{P}$, then $\det(\boldsymbol{A}) = \det(\boldsymbol{B})$.

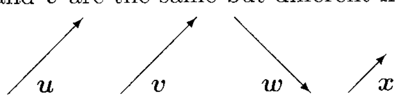
(d) If $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ are square matrices of the same size such that $\det(\boldsymbol{A}) = \det(\boldsymbol{B})$, then $\det(\boldsymbol{A}+\boldsymbol{C}) = \det(\boldsymbol{B}+\boldsymbol{C})$.

Chapter 3

Vector Spaces

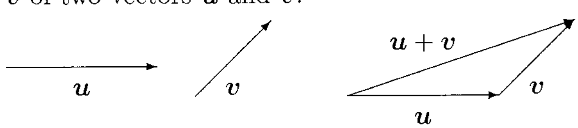
Section 3.1 Euclidean n-Spaces

Discussion 3.1.1 (Geometric Vectors) A (nonzero) *vector* is represented geometrically by a directed line segment or an arrow; the *direction* of the arrow specifies the direction of the vector and the *length* of the arrow describes its magnitude. The *zero vector* is represented by a point or a degenerated arrow with zero length and no direction. Two (nonzero) vectors are regarded as *equal* if they have the same length and direction. For example, in the following diagram, the two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are the same but different from the vectors $\boldsymbol{w}$ and $\boldsymbol{x}$.



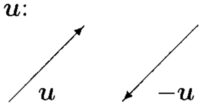
The addition, negative, difference and scalar multiple of vectors can be defined geometrically as follows.

(a) The *addition* $\boldsymbol{u}+\boldsymbol{v}$ of two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$:



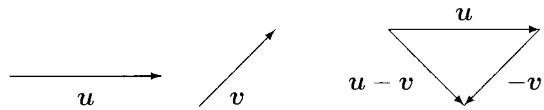
Note that $\boldsymbol{u}+\boldsymbol{v}$ is the same as $\boldsymbol{v}+\boldsymbol{u}$.

(b) The *negative* $-\boldsymbol{u}$ of a vector $\boldsymbol{u}$:



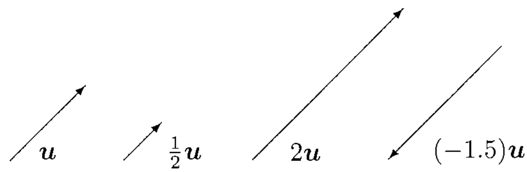
The vector $-\boldsymbol{u}$ has the same length as $\boldsymbol{u}$ but in the reverse direction.

(c) The *difference* $\boldsymbol{u} - \boldsymbol{v}$ of two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$:



Note that $\boldsymbol{u} - \boldsymbol{v}$ is the same as $\boldsymbol{u} + (-\boldsymbol{v})$.

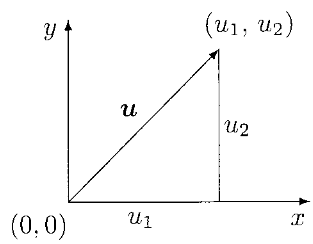
(d) The *scalar multiple* $c\boldsymbol{u}$ of a vector $\boldsymbol{u}$ where c is a real number:



If c is positive, the vector $c\boldsymbol{u}$ has the same direction as $\boldsymbol{u}$ and its length is c times of the length of $\boldsymbol{u}$. If c is negative, the vector $c\boldsymbol{u}$ is in the reverse direction of $\boldsymbol{u}$ and its length is $|c|$ times of the length of $\boldsymbol{u}$. Note that $0\boldsymbol{u}$ is the zero vector and $(-1)\boldsymbol{u}$ is the negative of $\boldsymbol{u}$.

Discussion 3.1.2 (Coordinate Systems)

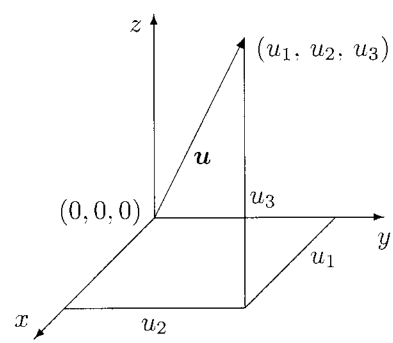
1. **Vectors in $\boldsymbol{xy}$-plane:** Suppose we position a vector $\boldsymbol{u}$ in the xy-plane such that its initial point is at the origin $(0, 0)$. The coordinates (u_1, u_2) of the end point of $\boldsymbol{u}$ are called the components of $\boldsymbol{u}$ and we write $\boldsymbol{u} = (u_1, u_2)$.



   (a) The addition of two vectors: Let $\boldsymbol{u} = (u_1, u_2)$ and $\boldsymbol{v} = (v_1, v_2)$. Then $\boldsymbol{u} + \boldsymbol{v} = (u_1 + v_1, u_2 + v_2)$.

   (b) The scalar multiple of a vector: Let $\boldsymbol{u} = (u_1, u_2)$. Then $c\boldsymbol{u} = (cu_1, cu_2)$ for any real number c.

2. **Vectors in $\boldsymbol{xyz}$-space:** Similarly, by placing a vector $\boldsymbol{u}$ with its initial point at the origin $(0, 0, 0)$, the coordinates (u_1, u_2, u_3) of the end point of $\boldsymbol{u}$ are the components of $\boldsymbol{u}$ and we write $\boldsymbol{u} = (u_1, u_2, u_3)$.



(a) The addition of two vectors: Let $\boldsymbol{u} = (u_1, u_2, u_3)$ and $\boldsymbol{v} = (v_1, v_2, v_3)$. Then $\boldsymbol{u} + \boldsymbol{v} = (u_1 + v_1, u_2 + v_2, u_3 + v_3)$.

(b) The scalar multiple of a vector: Let $\boldsymbol{u} = (u_1, u_2, u_3)$. Then $c\boldsymbol{u} = (cu_1, cu_2, cu_3)$ for any real number c.

Definition 3.1.3 An *n-vector* or *ordered n-tuple* of real numbers has the form

$$(u_1, u_2, \ldots, u_i, \ldots, u_n)$$

where $u_1, u_2, \ldots, u_n$ are real numbers. The number u_i in the ith position of an n-vector is called the ith *component* or the ith *coordinate* of the n-vector.

Let $\boldsymbol{u} = (u_1, u_2, \ldots, u_n)$ and $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)$ be two n-vectors.

1. We say that $\boldsymbol{u}$ and $\boldsymbol{v}$ are *equal* if and only if $u_i = v_i$ for all $i = 1, 2, \cdots, n$.

2. The *addition* $\boldsymbol{u} + \boldsymbol{v}$ of $\boldsymbol{u}$ and $\boldsymbol{v}$ is defined by

$$\boldsymbol{u} + \boldsymbol{v} = (u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n).$$

3. Let c be a real number. The *scalar multiple* $c\boldsymbol{u}$ of $\boldsymbol{u}$ is defined by

$$c\boldsymbol{u} = (cu_1, cu_2, \ldots, cu_n).$$

4. The n-vector $(0, 0, \ldots, 0)$ is called the *zero vector* and it is denoted by $\mathbf{0}$.

5. We define the *negative* of $\boldsymbol{u}$ to be $(-1)\boldsymbol{u}$ and denote it by $-\boldsymbol{u}$, i.e.

$$-\boldsymbol{u} = (-u_1, -u_2, \ldots, -u_n).$$

6. The *subtraction* $\boldsymbol{u} - \boldsymbol{v}$ of $\boldsymbol{v}$ from $\boldsymbol{u}$ is defined by $\boldsymbol{u} + (-\boldsymbol{v})$, i.e.

$$\boldsymbol{u} - \boldsymbol{v} = (u_1 - v_1, u_2 - v_2, \ldots, u_n - v_n).$$

Example 3.1.4 Let $\boldsymbol{u} = (1, 2, 3, 4)$ and $\boldsymbol{v} = (-1, -2, -3, 0)$. Then

$$\begin{aligned}
\boldsymbol{u} + \boldsymbol{v} &= (1 + (-1), 2 + (-2), 3 + (-3), 4 + 0) = (0, 0, 0, 4), \\
\boldsymbol{u} - \boldsymbol{v} &= (1 - (-1), 2 - (-2), 3 - (-3), 4 - 0) = (2, 4, 6, 4), \\
3\boldsymbol{u} &= (3 \cdot 1, 3 \cdot 2, 3 \cdot 3, 3 \cdot 4) = (3, 6, 9, 12), \\
3\boldsymbol{u} + 4\boldsymbol{v} &= (3 + (-4), 6 + (-8), 9 + (-12), 12 + 0) = (-1, -2, -3, 12).
\end{aligned}$$

Notation 3.1.5 The features in the definition of n-vectors are similar to those of matrices. In certain context, we can identify an n-vector $(u_1, u_2, \ldots, u_n)$ with a $1 \times n$ matrix $\begin{pmatrix} u_1 & u_2 & \cdots & u_n \end{pmatrix}$ (row vector) or an $n \times 1$ matrix $\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$ (column vector).

Theorem 3.1.6 Let $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ be n-vectors and c, d real numbers. Then

1. $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$;
2. $\boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w}) = (\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w}$;
3. $\boldsymbol{u} + \boldsymbol{0} = \boldsymbol{u} = \boldsymbol{0} + \boldsymbol{u}$;
4. $\boldsymbol{u} + (-\boldsymbol{u}) = \boldsymbol{0}$;
5. $c(d\boldsymbol{u}) = (cd)\boldsymbol{u}$;
6. $c(\boldsymbol{u} + \boldsymbol{v}) = c\boldsymbol{u} + c\boldsymbol{v}$;
7. $(c + d)\boldsymbol{u} = c\boldsymbol{u} + d\boldsymbol{u}$;
8. $1\boldsymbol{u} = \boldsymbol{u}$.

Proof All these properties of n-vectors can be easily verified by expressing the n-vectors in their component forms.

Definition 3.1.7 The set of all n-vectors of real numbers is called the *Euclidean n-space* or simply *n-space*. We use $\mathbb{R}$ to denote the set of all real numbers and $\mathbb{R}^n$ to denote the Euclidean n-space. If we say that $\boldsymbol{u}$ is a *vector* in $\mathbb{R}^n$, we mean that $\boldsymbol{u}$ is an n-vector, i.e. $\boldsymbol{u} \in \mathbb{R}^n$ if and only if $\boldsymbol{u} = (u_1, u_2, \ldots, u_n)$ for some $u_1, u_2, \ldots, u_n \in \mathbb{R}$.

Example 3.1.8 The following are some subsets of $\mathbb{R}^n$.

1. Let $B = \{\, (u_1, u_2, u_3, u_4) \mid u_1 = 0 \text{ and } u_2 = u_4 \,\}$, i.e. B is a subset of $\mathbb{R}^4$ such that a vector (u_1, u_2, u_3, u_4) is an element of B if and only if $u_1 = 0$ and $u_2 = u_4$. For example, $(0, 0, 0, 0)$, $(0, 0, 10, 0)$, $(0, 1, 3, 1)$ and $(0, \pi, \pi, \pi)$ are some elements of B. In general, the elements of B are of the form $(0, a, b, a)$ for arbitrary real numbers a and b. So we can also write $B = \{\, (0, a, b, a) \mid a, b \in \mathbb{R} \,\}$.

2. **(The solution set of a system of linear equations)** If a system of linear equations has n variables, then its solution set is a subset (may be empty) of $\mathbb{R}^n$. For example, the general solution of the system
$$\begin{cases} x + y + z = 0 \\ x - y + 2z = 1 \end{cases}$$
can be expressed in vector form: $(x, y, z) = (\frac{1}{2} - \frac{3}{2}t,\ -\frac{1}{2} + \frac{1}{2}t,\ t)$ where $t \in \mathbb{R}$.

We can therefore describe the solution set of this system in two ways:

$$\{\,(x,y,z) \mid x+y+z=0 \text{ and } x-y+2z=1\,\} \text{ (implicit) or}$$
$$\{\,(\tfrac{1}{2}-\tfrac{3}{2}t,\ -\tfrac{1}{2}+\tfrac{1}{2}t,\ t) \mid t \in \mathbb{R}\,\} \text{ (explicit).}$$

Geometrically, this solution set represents a line in $\mathbb{R}^3$. (See also Discussion 1.4.11.)

3. **(Lines in $\mathbb{R}^2, \mathbb{R}^3$ and planes in $\mathbb{R}^3$)**

(a) A line in $\mathbb{R}^2$ can be expressed implicitly in set notation by

$$\{\,(x,y) \mid ax+by=c\,\}$$

where a, b, c are real constants and a, b are not both zero. Explicitly, the line can also be written as

$$\left\{ \left(\frac{c-bt}{a},\ t \right) \middle|\ t \in \mathbb{R} \right\} \text{ if } a \neq 0, \text{ or}$$
$$\left\{ \left(t,\ \frac{c-at}{b} \right) \middle|\ t \in \mathbb{R} \right\} \text{ if } b \neq 0.$$

(b) A plane in $\mathbb{R}^3$ can be expressed implicitly in set notation by

$$\{\,(x,y,z) \mid ax+by+cz=d\,\}$$

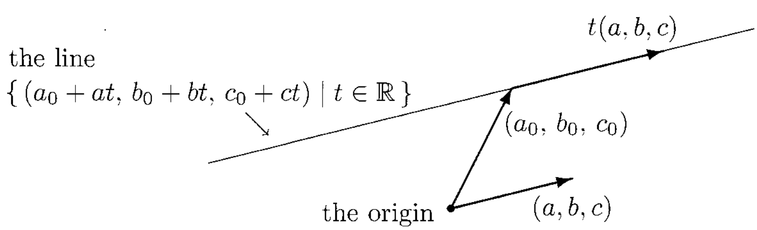
where a, b, c, d are real constants and a, b, c are not all zero. Explicitly, the plane can also be written as

$$\left\{ \left(\frac{d-bs-ct}{a},\ s,\ t \right) \middle|\ s,\ t \in \mathbb{R} \right\} \text{ if } a \neq 0, \text{ or}$$
$$\left\{ \left(s,\ \frac{d-as-ct}{b},\ t \right) \middle|\ s,\ t \in \mathbb{R} \right\} \text{ if } b \neq 0, \text{ or}$$
$$\left\{ \left(s,\ t,\ \frac{d-as-bt}{c} \right) \middle|\ s,\ t \in \mathbb{R} \right\} \text{ if } c \neq 0.$$

(c) Usually, a line in $\mathbb{R}^3$ is represented explicitly in set notation by

$$\{\,(a_0+at,\ b_0+bt,\ c_0+ct) \mid t \in \mathbb{R}\,\} = \{\,(a_0,\ b_0,\ c_0)+t(a,b,c) \mid t \in \mathbb{R}\,\}$$

where a_0, b_0, c_0, a, b, c are real constants and a, b, c are not all zero.



In here, $(a_0,\ b_0,\ c_0)$ is a point on the line and (a,b,c) is the *direction* of the line.

Note that in $\mathbb{R}^3$, a line cannot be represented by a single equation as in the case of $\mathbb{R}^2$. (In Example 3.1.8.2, we have an example of how a line is represented by a system of two equations.)

Notation 3.1.9 Let S be a finite set. We use $|S|$ to denote the number of elements contained in S.

Example 3.1.10 Let $S_1 = \{1, 2, 3, 4\}$, $S_2 = \{(1,2,3,4)\}$, and $S_3 = \{(1,2,3), (2,3,4)\}$. Then $|S_1| = 4$, $|S_2| = 1$ and $|S_3| = 2$.

Section 3.2 Linear Combinations and Linear Spans

Definition 3.2.1 Let $\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}$ be vectors in $\mathbb{R}^n$. For any real numbers $c_1, c_2, \ldots, c_k$, the vector

$$c_1\boldsymbol{u_1} + c_2\boldsymbol{u_2} + \cdots + c_k\boldsymbol{u_k}$$

is called a *linear combination* of $\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}$.

Example 3.2.2

1. Let $\boldsymbol{u_1} = (2,1,3)$, $\boldsymbol{u_2} = (1,-1,2)$ and $\boldsymbol{u_3} = (3,0,5)$. For each of the following vectors, determine whether it is a linear combination of $\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}$.

   (a) $\boldsymbol{v} = (3,3,4)$; (b) $\boldsymbol{w} = (1,2,4)$.

   Solution

   (a) Write $\boldsymbol{v} = a\boldsymbol{u_1} + b\boldsymbol{u_2} + c\boldsymbol{u_3}$, i.e.

   $$(3,3,4) = a(2,1,3) + b(1,-1,2) + c(3,0,5) = (2a+b+3c,\ a-b,\ 3a+2b+5c).$$

   So we obtain a system of linear equations

   $$\begin{cases} 2a + \ \ b + 3c = 3 \\ \ \ a - \ \ b \qquad = 3 \\ 3a + 2b + 5c = 4. \end{cases}$$

   Since

   $$\left(\begin{array}{ccc|c} 2 & 1 & 3 & 3 \\ 1 & -1 & 0 & 3 \\ 3 & 2 & 5 & 4 \end{array}\right) \xrightarrow[\text{Elimination}]{\text{Gaussian}} \left(\begin{array}{ccc|c} 2 & 1 & 3 & 3 \\ 0 & -\frac{3}{2} & -\frac{3}{2} & \frac{3}{2} \\ 0 & 0 & 0 & 0 \end{array}\right),$$

   the system is consistent. So $\boldsymbol{v}$ is a linear combination of $\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}$.

By using back-substitution, we obtain a general solution

$$(a, b, c) = (2-t, -1-t, t) \quad \text{where } t \text{ is an arbitrary parameter.}$$

For example, we have particular solutions $(2, -1, 0)$, $(1, -2, 1)$, etc. So we can write $\boldsymbol{v}$ as linear combinations

$$\boldsymbol{v} = 2\boldsymbol{u_1} - \boldsymbol{u_2} + 0\boldsymbol{u_3}, \quad \boldsymbol{v} = \boldsymbol{u_1} - 2\boldsymbol{u_2} + \boldsymbol{u_3}, \quad \text{etc.}$$

(b) We do the same for $\boldsymbol{w}$. After the Gaussian Elimination, we get

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

Since the system is inconsistent, $\boldsymbol{w}$ is not a linear combination of $\boldsymbol{u_1}$, $\boldsymbol{u_2}$, $\boldsymbol{u_3}$.

2. Let $\boldsymbol{e_1} = (1, 0, 0)$, $\boldsymbol{e_2} = (0, 1, 0)$ and $\boldsymbol{e_3} = (0, 0, 1)$. Every vector $\boldsymbol{u} = (x, y, z)$ in $\mathbb{R}^3$ is a linear combination of $\boldsymbol{e_1}$, $\boldsymbol{e_2}$ and $\boldsymbol{e_3}$. Note that $\boldsymbol{u} = x\boldsymbol{e_1} + y\boldsymbol{e_2} + z\boldsymbol{e_3}$. (Geometrically, $\boldsymbol{e_1}$, $\boldsymbol{e_2}$ and $\boldsymbol{e_3}$ are called the *directional vectors* of the x-axis, y-axis and z-axis, respectively, of $\mathbb{R}^3$.)

Definition 3.2.3 Let $S = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}\}$ be a set of vectors in $\mathbb{R}^n$. Then the set of all linear combinations of $\boldsymbol{u_1}$, $\boldsymbol{u_2}$, $\ldots$, $\boldsymbol{u_k}$,

$$\{\, c_1\boldsymbol{u_1} + c_2\boldsymbol{u_2} + \cdots + c_k\boldsymbol{u_k} \mid c_1, c_2, \ldots, c_k \in \mathbb{R} \,\},$$

is called the *linear span* of S (or the *linear span* of $\boldsymbol{u_1}$, $\boldsymbol{u_2}$, $\ldots$, $\boldsymbol{u_k}$) and is denoted by $\text{span}(S)$ (or $\text{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}\}$).

Example 3.2.4

1. In Example 3.2.2.1, $\boldsymbol{v} \in \text{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\}$ and $\boldsymbol{w} \notin \text{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\}$.

2. Let $S = \{(1, 0, 0, -1), (0, 1, 1, 0)\} \subseteq \mathbb{R}^4$. Every element in $\text{span}(S)$ is of the form

$$a(1, 0, 0, -1) + b(0, 1, 1, 0) = (a, b, b, -a)$$

where a, b are any real numbers. So $\text{span}(S) = \{\, (a, b, b, -a) \mid a,\ b \in \mathbb{R} \,\}$.

3. Let $V = \{\, (2a + b,\ a,\ 3b - a) \mid a, b \in \mathbb{R} \,\} \subseteq \mathbb{R}^3$. For any $a, b \in \mathbb{R}$,

$$(2a + b,\ a,\ 3b - a) = a(2, 1, -1) + b(1, 0, 3).$$

So $V = \text{span}\{(2, 1, -1), (1, 0, 3)\}$.

4. Show that span$\{(1,0,1), (1,1,0), (0,1,1)\} = \mathbb{R}^3$.

Solution To show span$\{(1,0,1), (1,1,0), (0,1,1)\} = \mathbb{R}^3$, we need to verify that for any vector (x,y,z) in $\mathbb{R}^3$, (x,y,z) is a linear combination of $(1,0,1)$, $(1,1,0)$ and $(0,1,1)$, i.e. there exist real numbers a, b, c such that

$$a(1,0,1) + b(1,1,0) + c(0,1,1) = (x,y,z).$$

This is equivalent to show that the linear system

$$\begin{cases} a+b \phantom{{}+c} = x \\ \phantom{a+{}} b+c = y \\ a \phantom{{}+b} +c = z, \end{cases}$$

where a, b and c are variables, is consistent for all $x, y, z \in \mathbb{R}$. Since

$$\left(\begin{array}{ccc|c} 1 & 1 & 0 & x \\ 0 & 1 & 1 & y \\ 1 & 0 & 1 & z \end{array}\right) \xrightarrow[\text{Elimination}]{\text{Gaussian}} \left(\begin{array}{ccc|c} 1 & 1 & 0 & x \\ 0 & 1 & 1 & y \\ 0 & 0 & 2 & z-x+y \end{array}\right),$$

the system is consistent regardless of the values of x, y, z. So we have shown that span$\{(1,0,1), (1,1,0), (0,1,1)\} = \mathbb{R}^3$.

(Explicitly, if we solve the system, we get $a = \frac{1}{2}(x-y+z)$, $b = \frac{1}{2}(x+y-z)$ and $c = \frac{1}{2}(-x+y+z)$.)

5. Show that span$\{(1,1,1), (1,2,0), (2,1,3), (2,3,1)\} \neq \mathbb{R}^3$.

Solution We follow the method used in Part 1. For any vector (x,y,z) in $\mathbb{R}^3$, we solve the vector equation

$$a(1,1,1) + b(1,2,0) + c(2,1,3) + d(2,3,1) = (x,y,z) \tag{3.1}$$

where a, b, c, d are variables. The linear system is

$$\begin{cases} a + b + 2c + 2d = x \\ a + 2b + c + 3d = y \\ a \phantom{{}+2b} + 3c + d = z. \end{cases}$$

Since

$$\left(\begin{array}{cccc|c} 1 & 1 & 2 & 2 & x \\ 1 & 2 & 1 & 3 & y \\ 1 & 0 & 3 & 1 & z \end{array}\right) \xrightarrow[\text{Elimination}]{\text{Gaussian}} \left(\begin{array}{cccc|c} 1 & 1 & 2 & 2 & x \\ 0 & 1 & -1 & 1 & y-x \\ 0 & 0 & 0 & 0 & y+z-2x \end{array}\right),$$

the system is inconsistent if $y+z-2x \neq 0$.

For example, if $(x,y,z) = (1,0,0)$ where $y+z-2x = -2 \neq 0$, then Equation (3.1) does not have a solution and hence $(1,0,0) \notin$ span$\{(1,1,1), (1,2,0), (2,1,3), (2,3,1)\}$.

So span$\{(1,1,1), (1,2,0), (2,1,3), (2,3,1)\} \neq \mathbb{R}^3$.

Discussion 3.2.5 In the following, we present a method to determine whether a set of vectors spans the whole $\mathbb{R}^n$:

Let $S = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}\} \subseteq \mathbb{R}^n$ where $\boldsymbol{u_1} = (a_{11}, a_{12}, \ldots, a_{1n})$, $\boldsymbol{u_2} = (a_{21}, a_{22}, \ldots, a_{2n})$, $\ldots$, $\boldsymbol{u_k} = (a_{k1}, a_{k2}, \ldots, a_{kn})$. For any $\boldsymbol{v} = (v_1, v_2, \ldots, v_n) \in \mathbb{R}^n$, $\boldsymbol{v}$ is contained in span(S) if and only if the vector equation

$$c_1\boldsymbol{u_1} + c_2\boldsymbol{u_2} + \cdots + c_k\boldsymbol{u_k} = \boldsymbol{v}$$

has a solution for $c_1, c_2, \ldots, c_k$, i.e. the linear system

$$\begin{cases} a_{11}c_1 + a_{21}c_2 + \cdots + a_{k1}c_k = v_1 \\ a_{12}c_1 + a_{22}c_2 + \cdots + a_{k2}c_k = v_2 \\ \qquad \vdots \qquad\qquad\qquad\qquad \vdots \\ a_{1n}c_1 + a_{2n}c_2 + \cdots + a_{kn}c_k = v_n \end{cases} \tag{3.2}$$

is consistent.

Let $\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{k1} \\ a_{12} & a_{22} & \cdots & a_{k2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{kn} \end{pmatrix}$.

1. If a row-echelon form of $\boldsymbol{A}$ does not have any zero row, then the linear system (3.2) is always consistent regardless of the values of $v_1, v_2, \ldots, v_n$ and hence span$(S) = \mathbb{R}^n$.

2. If a row-echelon form of $\boldsymbol{A}$ has at least one zero row, then the linear system (3.2) is not always consistent and hence span$(S) \neq \mathbb{R}^n$. (See Question 2.43.)

Example 3.2.6

1. In Example 3.2.4.4, $\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$ has a row-echelon form $\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix}$ which does not have any zero row. So span$\{(1, 0, 1), (1, 1, 0), (0, 1, 1)\} = \mathbb{R}^3$.

2. In Example 3.2.4.5, $\begin{pmatrix} 1 & 1 & 2 & 2 \\ 1 & 2 & 1 & 3 \\ 1 & 0 & 3 & 1 \end{pmatrix}$ has a row-echelon form $\begin{pmatrix} 1 & 1 & 2 & 2 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ which has a zero row. So span$\{(1, 1, 1), (1, 2, 0), (2, 1, 3), (2, 3, 1)\} \neq \mathbb{R}^3$.

Theorem 3.2.7 Let $S = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}\}$ be a set of vectors in $\mathbb{R}^n$. If $k < n$, then S cannot span $\mathbb{R}^n$.

Proof Follow the notation in Discussion 3.2.5. Since $k < n$, a row-echelon form of the matrix $\boldsymbol{A}$ must has at least one zero row. Thus span$(S) \neq \mathbb{R}^n$.

Example 3.2.8

1. One vector cannot span $\mathbb{R}^2$.
2. One vector or two vectors cannot span $\mathbb{R}^3$.

Theorem 3.2.9 Let $S = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}\} \in \mathbb{R}^n$.

1. $\mathbf{0} \in \text{span}(S)$.
2. For any $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_r} \in \text{span}(S)$ and $c_1, c_2, \ldots, c_r \in \mathbb{R}$,

$$c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_r\boldsymbol{v_r} \in \text{span}(S).$$

Proof

1. Since $\mathbf{0} = 0\boldsymbol{u_1} + 0\boldsymbol{u_2} + \cdots + 0\boldsymbol{u_k}$, $\mathbf{0} \in \text{span}(S)$.

2. Since $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_r} \in \text{span}(S)$, each $\boldsymbol{v_i}$ is a linear combination of $\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}$, i.e.

$$\begin{aligned}
\boldsymbol{v_1} &= a_{11}\boldsymbol{u_1} + a_{12}\boldsymbol{u_2} + \cdots + a_{1k}\boldsymbol{u_k}, \\
\boldsymbol{v_2} &= a_{21}\boldsymbol{u_1} + a_{22}\boldsymbol{u_2} + \cdots + a_{2k}\boldsymbol{u_k}, \\
&\vdots \\
\boldsymbol{v_r} &= a_{r1}\boldsymbol{u_1} + a_{r2}\boldsymbol{u_2} + \cdots + a_{rk}\boldsymbol{u_k},
\end{aligned}$$

where $a_{11}, a_{12}, \ldots, a_{rk}$ are real numbers. Thus

$$\begin{aligned}
&c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_r\boldsymbol{v_r} \\
&= c_1(a_{11}\boldsymbol{u_1} + a_{12}\boldsymbol{u_2} + \cdots + a_{1k}\boldsymbol{u_k}) + c_2(a_{21}\boldsymbol{u_1} + a_{22}\boldsymbol{u_2} + \cdots + a_{2k}\boldsymbol{u_k}) \\
&\quad + \cdots + c_r(a_{r1}\boldsymbol{u_1} + a_{r2}\boldsymbol{u_2} + \cdots + a_{rk}\boldsymbol{u_k}) \\
&= (c_1a_{11} + c_2a_{21} + \cdots + c_ra_{r1})\boldsymbol{u_1} + (c_1a_{12} + c_2a_{22} + \cdots + c_ra_{r2})\boldsymbol{u_2} \\
&\quad + \cdots + (c_1a_{1k} + c_2a_{2k} + \cdots + c_ra_{rk})\boldsymbol{u_k}
\end{aligned}$$

which is a linear combination of $\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}$. Hence $c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_r\boldsymbol{v_r} \in \text{span}(S)$.

Theorem 3.2.10 Let $S_1 = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}\}$ and $S_2 = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_m}\}$ be subsets of $\mathbb{R}^n$. Then $\text{span}(S_1) \subseteq \text{span}(S_2)$ if and only if each $\boldsymbol{u_i}$ is a linear combination of $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_m}$.

Proof

($\Rightarrow$) Suppose $\text{span}(S_1) \subseteq \text{span}(S_2)$. Since

$$S_1 \subseteq \text{span}(S_1) \subseteq \text{span}(S_2),$$

each $\boldsymbol{u_i}$ in S_1 is a linear combination of $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_m}$.

($\Leftarrow$) Suppose each $\boldsymbol{u_i}$ is a linear combination of $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_m}$. It means $\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k} \in \text{span}(S_2)$. Let $\boldsymbol{w}$ be any vector in $\text{span}(S_1)$, i.e. $\boldsymbol{w} = c_1\boldsymbol{u_1} + c_2\boldsymbol{u_2} + \cdots + c_k\boldsymbol{u_k}$ for some real numbers $c_1, c_2, \ldots, c_k$. Then by Theorem 3.2.9.2, $\boldsymbol{w}$ is also a vector in $\text{span}(S_2)$.

So we have shown that $\text{span}(S_1) \subseteq \text{span}(S_2)$.

Example 3.2.11

1. Let $\boldsymbol{u_1} = (1,0,1)$, $\boldsymbol{u_2} = (1,1,2)$, $\boldsymbol{u_3} = (-1,2,1)$ and $\boldsymbol{v_1} = (1,2,3)$, $\boldsymbol{v_2} = (2,-1,1)$. Show that

$$\text{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\} = \text{span}\{\boldsymbol{v_1}, \boldsymbol{v_2}\}.$$

Solution For two sets A and B, if we want to prove $A = B$, we need to show that $A \subseteq B$ and $B \subseteq A$.

For this question, we first show that $\text{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\} \subseteq \text{span}\{\boldsymbol{v_1}, \boldsymbol{v_2}\}$. By Theorem 3.2.10, we need to show that each $\boldsymbol{u_i}$ can be written as $a\boldsymbol{v_1} + b\boldsymbol{v_2}$ for some real number a, b, i.e. to solve the following three systems of linear equations:

$$\text{(a)} \begin{cases} a + 2b = 1 \\ 2a - b = 0 \\ 3a + b = 1 \end{cases} \quad \text{(for } \boldsymbol{u_1}\text{)}, \qquad \text{(b)} \begin{cases} a + 2b = 1 \\ 2a - b = 1 \\ 3a + b = 2 \end{cases} \quad \text{(for } \boldsymbol{u_2}\text{)},$$

$$\text{(c)} \begin{cases} a + 2b = -1 \\ 2a - b = 2 \\ 3a + b = 1 \end{cases} \quad \text{(for } \boldsymbol{u_3}\text{)}.$$

Since the row operations required to solve the three systems are all the same, we can work them out together:

$$\left(\begin{array}{cc|c|c|c} 1 & 2 & 1 & 1 & -1 \\ 2 & -1 & 0 & 1 & 2 \\ 3 & 1 & 1 & 2 & 1 \end{array}\right) \xrightarrow[\text{Elimination}]{\text{Gauss-Jordan}} \left(\begin{array}{cc|c|c|c} 1 & 0 & \frac{1}{5} & \frac{3}{5} & \frac{3}{5} \\ 0 & 1 & \frac{2}{5} & \frac{1}{5} & -\frac{4}{5} \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

We obtain

$$\boldsymbol{u_1} = \tfrac{1}{5}\boldsymbol{v_1} + \tfrac{2}{5}\boldsymbol{v_2}, \quad \boldsymbol{u_2} = \tfrac{3}{5}\boldsymbol{v_1} + \tfrac{1}{5}\boldsymbol{v_2} \quad \text{and} \quad \boldsymbol{u_3} = \tfrac{3}{5}\boldsymbol{v_1} - \tfrac{4}{5}\boldsymbol{v_2}.$$

So $\text{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\} \subseteq \text{span}\{\boldsymbol{v_1}, \boldsymbol{v_2}\}$.

Similarly, we find that

$$\boldsymbol{v_1} = -\boldsymbol{u_1} + 2\boldsymbol{u_2} + 0\boldsymbol{u_3} \quad \text{and} \quad \boldsymbol{v_2} = 3\boldsymbol{u_1} - \boldsymbol{u_2} + 0\boldsymbol{u_3}.$$

(The expressions of $\boldsymbol{v_1}$ and $\boldsymbol{v_2}$ as linear combinations of $\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}$ are not unique.) So $\text{span}\{\boldsymbol{v_1}, \boldsymbol{v_2}\} \subseteq \text{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\}$.

Thus we have shown that $\text{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\} = \text{span}\{\boldsymbol{v_1}, \boldsymbol{v_2}\}$.

(Actually, in showing $\operatorname{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\} \subseteq \operatorname{span}\{\boldsymbol{v_1}, \boldsymbol{v_2}\}$ and vice versa, we do not need to find the solutions of the linear systems explicitly. The proper procedures will be shown in the next example.)

2. Let $\boldsymbol{u_1} = (1,0,0,1)$, $\boldsymbol{u_2} = (0,1,-1,2)$, $\boldsymbol{u_3} = (2,1,-1,4)$ and $\boldsymbol{v_1} = (1,1,1,1)$, $\boldsymbol{v_2} = (-1,1,-1,1)$, $\boldsymbol{v_3} = (-1,1,1,-1)$. Show that

$$\operatorname{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\} \subseteq \operatorname{span}\{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}\}$$

but

$$\operatorname{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\} \neq \operatorname{span}\{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}\}.$$

Solution To show that each $\boldsymbol{u_i}$, $i = 1, 2, 3$, is a linear combination of $\boldsymbol{v_1}$, $\boldsymbol{v_2}$, $\boldsymbol{v_3}$, we follow the same procedure as the previous example and solve the three linear systems together (what are the three systems?):

$$\left(\begin{array}{ccc|c|c|c} 1 & -1 & -1 & 1 & 0 & 2 \\ 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & -1 & 1 & 0 & -1 & -1 \\ 1 & 1 & -1 & 1 & 2 & 4 \end{array}\right) \xrightarrow[\text{Elimination}]{\text{Gaussian}} \left(\begin{array}{ccc|c|c|c} 1 & -1 & -1 & 1 & 0 & 2 \\ 0 & 2 & 2 & -1 & 1 & -1 \\ 0 & 0 & 2 & -1 & -1 & -3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

Since all three systems are consistent, all $\boldsymbol{u_i}$ are linear combinations of $\boldsymbol{v_1}$, $\boldsymbol{v_2}$, $\boldsymbol{v_3}$. So

$$\operatorname{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\} \subseteq \operatorname{span}\{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}\}.$$

On the other hand,

$$\left(\begin{array}{ccc|c|c|c} 1 & 0 & 2 & 1 & -1 & -1 \\ 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & -1 & -1 & 1 & -1 & 1 \\ 1 & 2 & 4 & 1 & 1 & -1 \end{array}\right) \xrightarrow[\text{Elimination}]{\text{Gaussian}} \left(\begin{array}{ccc|c|c|c} 1 & 0 & 2 & 1 & -1 & -1 \\ 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 2 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

Since not all three systems are consistent, some $\boldsymbol{v_i}$'s are not linear combinations of $\boldsymbol{u_1}$, $\boldsymbol{u_2}$, $\boldsymbol{u_3}$. So $\operatorname{span}\{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}\} \not\subseteq \operatorname{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\}$ and hence

$$\operatorname{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\} \neq \operatorname{span}\{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}\}.$$

(Can you tell which $\boldsymbol{v_i}$'s are not linear combinations of $\boldsymbol{u_1}$, $\boldsymbol{u_2}$, $\boldsymbol{u_3}$?)

Theorem 3.2.12 Let $\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}$ be vectors in $\mathbb{R}^n$. If $\boldsymbol{u_k}$ is a linear combination of $\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_{k-1}}$, then

$$\operatorname{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_{k-1}}\} = \operatorname{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_{k-1}}, \boldsymbol{u_k}\}.$$

Proof It is obvious that $\operatorname{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_{k-1}}\} \subseteq \operatorname{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_{k-1}}, \boldsymbol{u_k}\}$.

On the other hand, since $\boldsymbol{u_k}$ is a linear combination of $\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_{k-1}}$, by Theorem 3.2.10, $\text{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_{k-1}}, \boldsymbol{u_k}\} \subseteq \text{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_{k-1}}\}$.

So we have $\text{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_{k-1}}\} = \text{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_{k-1}}, \boldsymbol{u_k}\}$.

Example 3.2.13 Let $\boldsymbol{u_1} = (1, 1, 0, 2)$, $\boldsymbol{u_2} = (1, 0, 0, 1)$ and $\boldsymbol{u_3} = (0, 1, 0, 1)$. It is easy to check that $\boldsymbol{u_3} = \boldsymbol{u_1} - \boldsymbol{u_2}$. Thus by Theorem 3.2.12, $\text{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}\} = \text{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\}$.

Section 3.3 Subspaces

Definition 3.3.1 Let V be a subset of $\mathbb{R}^n$. Then V is called a *subspace* of $\mathbb{R}^n$ if $V = \text{span}(S)$ where $S = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}\}$ for some vectors $\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k} \in \mathbb{R}^n$. More precisely, V is called the *subspace spanned* by S (or the *subspace spanned* by $\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}$). We also say that S *spans* (or $\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}$ *span*) the subspace V.

Remark 3.3.2

1. Let $\mathbf{0}$ be the zero vector in $\mathbb{R}^n$. Then the set $\{\mathbf{0}\} = \text{span}\{\mathbf{0}\}$ is a subspace of $\mathbb{R}^n$ and is known as the *zero space*.

2. Let $\boldsymbol{e_1} = (1, 0, \ldots, 0)$, $\boldsymbol{e_2} = (0, 1, 0, \ldots, 0)$, $\ldots$, $\boldsymbol{e_n} = (0, \ldots, 0, 1)$ be vectors in $\mathbb{R}^n$ where for each i, the ith coordinate of $\boldsymbol{e_i}$ is 1 while all other coordinates of $\boldsymbol{e_i}$ are 0. Any vector $\boldsymbol{u} = (u_1, u_2, \ldots, u_n) \in \mathbb{R}^n$ can be written as

$$\boldsymbol{u} = u_1\boldsymbol{e_1} + u_2\boldsymbol{e_2} + \cdots + u_n\boldsymbol{e_n}.$$

   Thus $\mathbb{R}^n = \text{span}\{\boldsymbol{e_1}, \boldsymbol{e_2}, \ldots, \boldsymbol{e_n}\}$ is a subspace of $\mathbb{R}^n$. (See also Example 3.5.9.3.)

3. In more advanced textbooks on linear algebra, "subspace" is defined differently, see Remark 3.3.8.

Example 3.3.3

1. Let $V_1 = \{\, (a+4b,\ a) \mid a, b \in \mathbb{R}\,\} \subseteq \mathbb{R}^2$. For any $a, b \in \mathbb{R}$, $(a+4b,\ a) = a(1, 1) + b(4, 0)$. So $V_1 = \text{span}\{(1, 1),\ (4, 0)\}$ is a subspace of $\mathbb{R}^2$. (Actually, $V_1 = \mathbb{R}^2$.)

2. Let $V_2 = \{\, (x, y, z) \mid x + y - z = 0\,\} \subseteq \mathbb{R}^3$. The equation $x + y - z = 0$ has a general solution

$$(x, y, z) = (-s + t,\ s,\ t) = s(-1, 1, 0) + t(1, 0, 1) \quad \text{where } s, t \text{ are arbitrary parameters.}$$

   So $V_2 = \text{span}\{(-1, 1, 0),\ (1, 0, 1)\}$ is a subspace of $\mathbb{R}^3$. (V_2 is a plane in $\mathbb{R}^3$ containing the origin.)

3. Let $V_3 = \{ (1, a) \mid a \in \mathbb{R} \} \subseteq \mathbb{R}^2$. V_3 is not a subspace of $\mathbb{R}^2$.

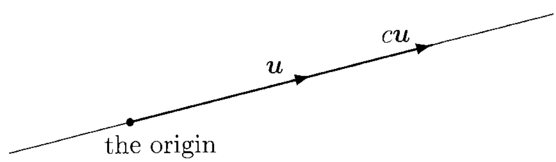
   Proof Assume that V_3 is a subspace, i.e. $V_3 = \text{span}(S)$ where S is a finite set of vectors in $\mathbb{R}^2$. By Theorem 3.2.9.1, $(0, 0) \in \text{span}(S) = V_3$ but this is impossible as $(0, 0) \neq (1, a)$ for any $a \in \mathbb{R}$. So V_3 is not a subspace of $\mathbb{R}^2$.

4. Let $V_4 = \{ (x, y, z) \mid x^2 \leq y^2 \leq z^2 \} \subseteq \mathbb{R}^3$. V_4 is not a subspace of $\mathbb{R}^3$.

   Proof Assume that V_4 is a subspace, i.e. $V_4 = \text{span}(T)$ where T is a finite set of vectors in $\mathbb{R}^3$. Observe that $(1, 1, 2), (1, 1, -2) \in V_4$. By Theorem 3.2.9.2, $(1, 1, 2) + (1, 1, -2) \in \text{span}(T) = V_4$ but this is impossible as $(1, 1, 2) + (1, 1, -2) = (2, 2, 0) \notin V_4$. So V_4 is not a subspace of $\mathbb{R}^3$.

Discussion 3.3.4

1. Let $\boldsymbol{u}$ be a nonzero vector in $\mathbb{R}^2$ or $\mathbb{R}^3$. Then $\text{span}\{\boldsymbol{u}\}$ consists of all vectors of the form $c\boldsymbol{u}$ for any real number c. It is a line through the origin.

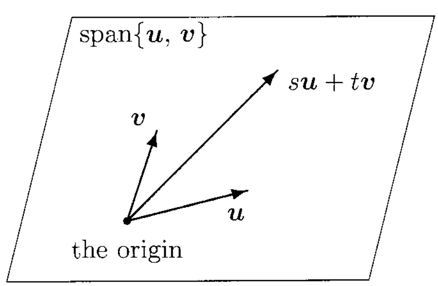


   In $\mathbb{R}^2$, if $\boldsymbol{u} = (u_1, u_2)$, then $\text{span}\{\boldsymbol{u}\} = \{ c\boldsymbol{u} \mid c \in \mathbb{R} \} = \{ (x, y) \mid u_2 x - u_1 y = 0 \}$.

   In $\mathbb{R}^3$, if $\boldsymbol{u} = (u_1, u_2, u_3)$, then $\text{span}\{\boldsymbol{u}\} = \{ c\boldsymbol{u} \mid c \in \mathbb{R} \} = \{ (cu_1, cu_2, cu_3) \mid c \in \mathbb{R} \}$.

   Let us recall that in $\mathbb{R}^3$, a line cannot be represented by a single equation as in the case of $\mathbb{R}^2$.

2. Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be two nonzero vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$. Then $\text{span}\{\boldsymbol{u}, \boldsymbol{v}\}$ consists of all vectors of the form $s\boldsymbol{u} + t\boldsymbol{v}$ for any real numbers s, t. If $\boldsymbol{u}$ and $\boldsymbol{v}$ are not parallel, then $\text{span}\{\boldsymbol{u}, \boldsymbol{v}\}$ is a plane containing the origin.



   In $\mathbb{R}^2$, $\text{span}\{\boldsymbol{u}, \boldsymbol{v}\} = \mathbb{R}^2$.

In $\mathbb{R}^3$, if $\boldsymbol{u} = (u_1, u_2, u_3)$ and $\boldsymbol{v} = (v_1, v_2, v_3)$, then

$$\text{span}\{\boldsymbol{u},\ \boldsymbol{v}\} = \{\, s\boldsymbol{u} + t\boldsymbol{v} \mid s, t \in \mathbb{R}\,\} = \{\,(x, y, z) \mid ax + by + cz = 0\,\}$$

where (a, b, c) is a non-trivial solution to the system of the two linear equations $u_1a + u_2b + u_3c = 0$ and $v_1a + v_2b + v_3c = 0$.

Remark 3.3.5

1. The following are all the subspaces of $\mathbb{R}^2$:
   (a) $\{\mathbf{0}\}$ (given by span$\{\mathbf{0}\}$);
   (b) lines through the origin (given by span$\{\boldsymbol{u}\}$ for nonzero $\boldsymbol{u} \in \mathbb{R}^2$);
   (c) $\mathbb{R}^2$ (given by span$\{(1, 0),\ (0, 1)\}$).
2. The following are all the subspaces of $\mathbb{R}^3$:
   (a) $\{\mathbf{0}\}$ (given by span$\{\mathbf{0}\}$);
   (b) lines through the origin (given by span$\{\boldsymbol{u}\}$ for nonzero $\boldsymbol{u} \in \mathbb{R}^3$);
   (c) planes containing the origin (given by span$\{\boldsymbol{u},\ \boldsymbol{v}\}$ for nonzero $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^3$ which are not parallel to each other);
   (d) $\mathbb{R}^3$ (given by span$\{(1, 0, 0),\ (0, 1, 0),\ (0, 0, 1)\}$).

Theorem 3.3.6 The solution set of a homogeneous system of linear equations in n variables is a subspace of $\mathbb{R}^n$. (This subspace is called the *solution space* of the system, see also Section 4.3.)

Proof If the homogeneous system has only the trivial solution, then the solution set is $\{\mathbf{0}\}$ which is the zero space (see Remark 3.3.2.1).

Suppose the homogeneous system has infinitely many solution. Let $x_1, x_2, \ldots, x_n$ be the variables of the system. By solving the system, say, using Gauss-Jordan Elimination, a general solution can be expressed in the form

$$\begin{cases} x_1 = r_{11}t_1 + r_{12}t_2 + \cdots + r_{1k}t_k \\ x_2 = r_{21}t_1 + r_{22}t_2 + \cdots + r_{2k}t_k \\ \quad\vdots \\ x_n = r_{n1}t_1 + r_{n2}t_2 + \cdots + r_{nk}t_k \end{cases}$$

for some arbitrary parameters $t_1, t_2, \ldots, t_k$, where $r_{11}, r_{12}, \ldots, r_{nk}$ are real numbers. We can rewrite this general solution as

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = t_1 \begin{pmatrix} r_{11} \\ r_{21} \\ \vdots \\ r_{n1} \end{pmatrix} + t_2 \begin{pmatrix} r_{12} \\ r_{22} \\ \vdots \\ r_{n2} \end{pmatrix} + \cdots + t_k \begin{pmatrix} r_{1k} \\ r_{2k} \\ \vdots \\ r_{nk} \end{pmatrix}.$$

So the solution set is

$$\operatorname{span}\{(r_{11}, r_{21}, \ldots, r_{n1}),\ (r_{12}, r_{22}, \ldots, r_{n2}),\ \ldots,\ (r_{1k}, r_{2k}, \ldots, r_{nk})\}$$

and hence is a subspace of $\mathbb{R}^n$.

Example 3.3.7

1. The linear system

$$\begin{cases} x - 2y + 3z = 0 \\ 2x - 4y + 6z = 0 \\ 3x - 6y + 9z = 0 \end{cases}$$

has a general solution

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2s - 3t \\ s \\ t \end{pmatrix} = s\begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} + t\begin{pmatrix} -3 \\ 0 \\ 1 \end{pmatrix} \quad \text{where } s,\ t \text{ are arbitrary parameters.}$$

The solution space is $\{\,(2s - 3t, s, t) \mid s,\ t \in \mathbb{R}\,\} = \operatorname{span}\{(2, 1, 0), (-3, 0, 1)\}$.

In fact, it is a plane in $\mathbb{R}^3$ containing the origin.

2. The linear system

$$\begin{cases} x - 2y + 3z = 0 \\ -3x + 7y - 8z = 0 \\ -2x + 4y - 6z = 0 \end{cases}$$

has a general solution

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -5t \\ -t \\ t \end{pmatrix} = t\begin{pmatrix} -5 \\ -1 \\ 1 \end{pmatrix} \quad \text{where } t \text{ is an arbitrary parameter.}$$

The solution space is $\{\,(-5t, -t, t) \mid t \in \mathbb{R}\,\} = \operatorname{span}\{(-5, -1, 1)\}$.

In fact, it is a line in $\mathbb{R}^3$ through the origin.

3. The linear system

$$\begin{cases} x - 2y + 3z = 0 \\ -3x + 7y - 8z = 0 \\ 4x + y + 2z = 0 \end{cases}$$

has a general solution

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

The solution space is $\{(0, 0, 0)\}$, the zero space.

Remark 3.3.8 Let V be a non-empty subset of $\mathbb{R}^n$. Then V is a subspace of $\mathbb{R}^n$ if and only if

$$\text{for all } \boldsymbol{u}, \boldsymbol{v} \in V \text{ and } c, d \in \mathbb{R}, \quad c\boldsymbol{u} + d\boldsymbol{v} \in V.$$

This is actually the definition of subspaces in abstract linear algebra. The proof of the result requires some knowledge from the next section. So we do not prove it here. See Question 3.31.

Section 3.4 Linear Independence

Discussion 3.4.1 Let $\boldsymbol{u_1} = (1, 1, 0, 2)$, $\boldsymbol{u_2} = (1, 0, 0, 1)$ and $\boldsymbol{u_3} = (0, 1, 0, 1)$. From Example 3.2.13, we learn that $\text{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}\} = \text{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\}$. So we can say that the vector $\boldsymbol{u_3}$ is "redundant" in $\text{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\}$. In order to make this notion of "redundancy" explicit, we need to introduce another important concept in the study of vectors.

Definition 3.4.2 Let $S = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \dots, \boldsymbol{u_k}\}$ be a set of vectors in $\mathbb{R}^n$. Consider the equation

$$c_1\boldsymbol{u_1} + c_2\boldsymbol{u_2} + \cdots + c_k\boldsymbol{u_k} = \boldsymbol{0}, \tag{3.3}$$

where $c_1, c_2, \dots, c_k$ are variables. Note that $c_1 = 0$, $c_2 = 0$, $\dots$, $c_k = 0$ satisfies Equation (3.3) and hence is a solution to Equation (3.3). This solution is called the *trivial solution.* (See also Definition 1.5.1.)

1. S is called a *linearly independent set* and $\boldsymbol{u_1}, \boldsymbol{u_2}, \dots, \boldsymbol{u_k}$ are said to be *linearly independent* if Equation (3.3) has only the trivial solution.

2. S is called a *linearly dependent set* and $\boldsymbol{u_1}, \boldsymbol{u_2}, \dots, \boldsymbol{u_k}$ are said to be *linearly dependent* if Equation (3.3) has non-trivial solutions, i.e. there exist real numbers $a_1, a_2, \dots, a_k$, not all of them are zero, such that $a_1\boldsymbol{u_1} + a_2\boldsymbol{u_2} + \cdots + a_k\boldsymbol{u_k} = \boldsymbol{0}$.

Example 3.4.3

1. Determine whether the vectors $(1, -2, 3)$, $(5, 6, -1)$, $(3, 2, 1)$ are linearly independent.

   Solution The equation

   $$c_1(1, -2, 3) + c_2(5, 6, -1) + c_3(3, 2, 1) = (0, 0, 0)$$

   gives us a linear system

   $$\begin{cases} c_1 + 5c_2 + 3c_3 = 0 \\ -2c_1 + 6c_2 + 2c_3 = 0 \\ 3c_1 - c_2 + c_3 = 0. \end{cases}$$

By Gaussian Elimination, we find that there are infinitely many solutions, i.e. there exist non-trivial solutions. So the vectors are linearly dependent.

2. Determine whether the vectors $(1, 0, 0, 1)$, $(0, 2, 1, 0)$, $(1, -1, 1, 1)$ are linearly independent.

   Solution The equation

$$c_1(1, 0, 0, 1) + c_2(0, 2, 1, 0) + c_3(1, -1, 1, 1) = (0, 0, 0, 0)$$

   gives us a linear system

$$\begin{cases} c_1 \quad\quad + c_3 = 0 \\ \quad 2c_2 - c_3 = 0 \\ \quad\quad c_2 + c_3 = 0 \\ c_1 \quad\quad + c_3 = 0. \end{cases}$$

   By Gaussian Elimination, we find that there is only one solution $c_1 = 0$, $c_2 = 0$, $c_3 = 0$. So the vectors are linearly independent.

3. Let $S = \{\boldsymbol{u}\}$ be a subset of $\mathbb{R}^n$. Then S is linearly dependent if and only if $\boldsymbol{u} = \boldsymbol{0}$.

4. Let $S = \{\boldsymbol{u}, \boldsymbol{v}\}$ be a subset of $\mathbb{R}^n$. Then S is linearly dependent if and only if $\boldsymbol{u} = a\boldsymbol{v}$ for some real number a or $\boldsymbol{v} = b\boldsymbol{u}$ for some real number b.

5. Let S be a finite subset of $\mathbb{R}^n$. If $\boldsymbol{0} \in S$, then S is linearly dependent.

Theorem 3.4.4 Let $S = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}\}$ be a set of vectors in $\mathbb{R}^n$ where $k \geq 2$. Then

1. S is linearly dependent if and only if at least one vector $\boldsymbol{u_i}$ in S can be written as a linear combination of other vectors in S, i.e.

$$\boldsymbol{u_i} = a_1\boldsymbol{u_1} + a_2\boldsymbol{u_2} + \cdots + a_{i-1}\boldsymbol{u_{i-1}} + a_{i+1}\boldsymbol{u_{i+1}} + \cdots + a_k\boldsymbol{u_k}$$

   for some real numbers $a_1, a_2, \ldots, a_{i-1}, a_{i+1}, \ldots, a_k$; and

2. S is linearly independent if and only if no vector in S can be written as a linear combination of other vectors in S.

Proof The two statements 1 and 2 are logically equivalent. We only need to prove one of them. In the following, the first statement is proved.

($\Rightarrow$) Suppose S is linearly dependent, i.e. there exist real numbers $a_1, a_2, \ldots, a_k$, not all of them are zero, such that $a_1\boldsymbol{u_1} + a_2\boldsymbol{u_2} + \cdots + a_k\boldsymbol{u_k} = \boldsymbol{0}$. Let a_i be one of the nonzero coefficients. Then

$$\boldsymbol{u_i} = -\frac{a_1}{a_i}\boldsymbol{u_1} - \frac{a_2}{a_i}\boldsymbol{u_2} - \cdots - \frac{a_{i-1}}{a_i}\boldsymbol{u_{i-1}} - \frac{a_{i+1}}{a_i}\boldsymbol{u_{i+1}} - \cdots - \frac{a_k}{a_i}\boldsymbol{u_k}.$$

Here $\boldsymbol{u_i}$ is a linear combination of $\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_{i-1}}, \boldsymbol{u_{i+1}}, \ldots, \boldsymbol{u_k}$.

($\Leftarrow$) Suppose there exists $\boldsymbol{u_i}$ such that

$$\boldsymbol{u_i} = a_1\boldsymbol{u_1} + a_2\boldsymbol{u_2} + \cdots + a_{i-1}\boldsymbol{u_{i-1}} + a_{i+1}\boldsymbol{u_{i+1}} + \cdots + a_k\boldsymbol{u_k}$$

for some real numbers $a_1, a_2, \ldots, a_{i-1}, a_{i+1}, \ldots, a_k$. Put $a_i = -1$. Then

$$\begin{aligned} & a_1\boldsymbol{u_1} + a_2\boldsymbol{u_2} + \cdots + a_k\boldsymbol{u_k} \\ &= a_1\boldsymbol{u_1} + a_2\boldsymbol{u_2} + \cdots + a_{i-1}\boldsymbol{u_{i-1}} + a_i\boldsymbol{u_i} + a_{i+1}\boldsymbol{u_{i+1}} + \cdots + a_k\boldsymbol{u_k} \\ &= a_1\boldsymbol{u_1} + a_2\boldsymbol{u_2} + \cdots + a_{i-1}\boldsymbol{u_{i-1}} + a_{i+1}\boldsymbol{u_{i+1}} + \cdots + a_k\boldsymbol{u_k} - \boldsymbol{u_i} \\ &= \boldsymbol{u_i} - \boldsymbol{u_i} = \boldsymbol{0}. \end{aligned}$$

Since we have $a_1\boldsymbol{u_1} + a_2\boldsymbol{u_2} + \cdots + a_k\boldsymbol{u_k} = \boldsymbol{0}$ where not all the coefficients are zero ($a_i = -1 \neq 0$), S is linearly dependent.

Remark 3.4.5 By Theorem 3.4.4, "linearly dependent" gives us an explicit description of the concept of "redundancy" discussed in Discussion 3.4.1.

1. If a set of vectors is linearly dependent, then there exists at least one "redundant" vector in the set.

2. If a set is linearly independent, then there is no "redundant" vector in the set.

Example 3.4.6

1. Let $S_1 = \{(1,0), (0,4), (2,4)\} \subseteq \mathbb{R}^2$. Note that S_1 is linearly dependent (check it). We see that
$$(2,4) = 2(1,0) + (0,4)$$
i.e. $(2,4)$ can be expressed as a linear combination of $(1,0)$ and $(0,4)$.

2. Let $S_2 = \{(-1,0,0), (0,3,0), (0,0,7)\} \subseteq \mathbb{R}^3$. Note that S_2 is linearly independent (check it). By comparing the first components of the three vectors, we see that $(-1,0,0)$ cannot be expressed as a linear combination of $(0,3,0)$ and $(0,0,7)$. Similarly each of $(0,3,0)$ and $(0,0,7)$ cannot be expressed as a linear combination of the other two vectors.

Theorem 3.4.7 Let $S = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}\}$ be a set of vectors in $\mathbb{R}^n$. If $k > n$, then S is linearly dependent.

Proof Let $\boldsymbol{u_1} = (a_{11}, a_{12}, \ldots, a_{1n})$, $\boldsymbol{u_2} = (a_{21}, a_{22}, \ldots, a_{2n})$, $\ldots$, $\boldsymbol{u_k} = (a_{k1}, a_{k2}, \ldots, a_{kn})$. Consider the equation
$$c_1\boldsymbol{u_1} + c_2\boldsymbol{u_2} + \cdots + c_k\boldsymbol{u_k} = \boldsymbol{0}.$$

We obtain a homogeneous system of n linear equations in variables $c_1, c_2, \ldots, c_k$:

$$\begin{cases} a_{11}c_1 + a_{21}c_2 + \cdots + a_{k1}c_k = 0 \\ a_{12}c_1 + a_{22}c_2 + \cdots + a_{k2}c_k = 0 \\ \qquad \vdots \qquad\qquad\qquad \vdots \\ a_{1n}c_1 + a_{2n}c_2 + \cdots + a_{kn}c_k = 0. \end{cases}$$

The system has k unknowns and n equations. Since $k > n$, by Remark 1.5.4.2, the system has non-trivial solutions. Therefore, S is linearly dependent.

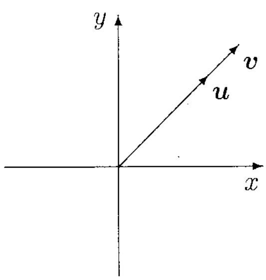
Example 3.4.8

1. In $\mathbb{R}^2$, a set of three or more vectors must be linearly dependent.

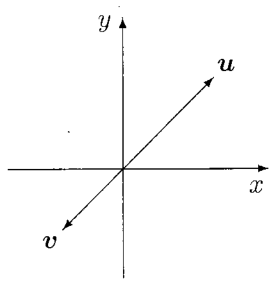
2. In $\mathbb{R}^3$, a set of four or more vectors must be linearly dependent.

Discussion 3.4.9

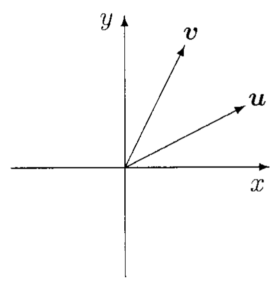
1. In $\mathbb{R}^2$ (or $\mathbb{R}^3$), two vectors $\boldsymbol{u}, \boldsymbol{v}$ are linearly dependent if and only if they lie on the same line (when they are placed with their initial points at the origin).



$\boldsymbol{u}, \boldsymbol{v}$ are linearly dependent

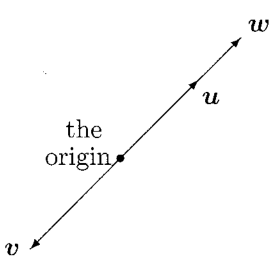


$\boldsymbol{u}, \boldsymbol{v}$ are linearly dependent

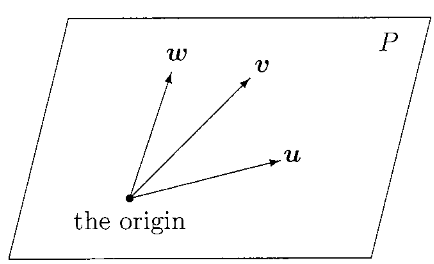


$\boldsymbol{u}, \boldsymbol{v}$ are linearly independent

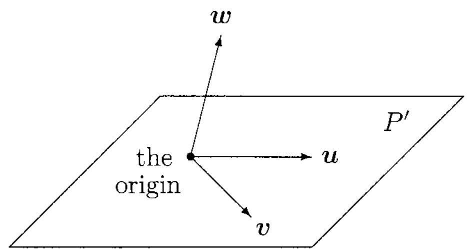
2. In $\mathbb{R}^3$, three vectors $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ are linearly dependent if and only if they lie on the same line or the same plane (when they are placed with their initial points at the origin).



$\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ are linearly dependent



$\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ are linearly dependent
($P = \text{span}\{\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}\}$)



$\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ are linearly independent
$(P' = \operatorname{span}\{\boldsymbol{u}, \boldsymbol{v}\})$

Theorem 3.4.10 Let $\boldsymbol{u_1}$, $\boldsymbol{u_2}$, ..., $\boldsymbol{u_k}$ be linearly independent vectors in $\mathbb{R}^n$. If $\boldsymbol{u_{k+1}}$ is a vector in $\mathbb{R}^n$ and it is not a linear combination of $\boldsymbol{u_1}$, $\boldsymbol{u_2}$, ..., $\boldsymbol{u_k}$, then $\boldsymbol{u_1}$, $\boldsymbol{u_2}$, ..., $\boldsymbol{u_k}$, $\boldsymbol{u_{k+1}}$ are linearly independent.

Proof We need to solve the vector equation

$$c_1\boldsymbol{u_1} + c_2\boldsymbol{u_2} + \cdots + c_k\boldsymbol{u_k} + c_{k+1}\boldsymbol{u_{k+1}} = \boldsymbol{0}. \tag{3.4}$$

First, we show that $c_{k+1} = 0$. Assume not, i.e. $c_{k+1} \neq 0$. Then

$$\boldsymbol{u_{k+1}} = -\frac{c_1}{c_{k+1}}\boldsymbol{u_1} - \frac{c_2}{c_{k+1}}\boldsymbol{u_2} - \cdots - \frac{c_k}{c_{k+1}}\boldsymbol{u_k}$$

which contradicts that $\boldsymbol{u_{k+1}}$ is not a linear combination of $\boldsymbol{u_1}$, $\boldsymbol{u_2}$, ..., $\boldsymbol{u_k}$. So $c_{k+1} = 0$.

Now, substituting $c_{k+1} = 0$ into Equation (3.4), we get

$$c_1\boldsymbol{u_1} + c_2\boldsymbol{u_2} + \cdots + c_k\boldsymbol{u_k} = \boldsymbol{0}.$$

Since $\boldsymbol{u_1}$, $\boldsymbol{u_2}$, ..., $\boldsymbol{u_k}$ are linearly independent, this vector equation can have only the trivial solution, i.e. $c_1 = 0$, $c_2 = 0$, ..., $c_k = 0$.

Summarizing the works above, the only solution to Equation (3.4) is $c_1 = 0$, $c_2 = 0$, ..., $c_k = 0$, $c_{k+1} = 0$. It means $\boldsymbol{u_1}$, $\boldsymbol{u_2}$, ..., $\boldsymbol{u_k}$, $\boldsymbol{u_{k+1}}$ are linearly independent.

Section 3.5 Bases

Discussion 3.5.1 From now on, we shall adopt the following conventions in using the terms "vector space" and "subspace".

1. A set V is called a *vector space* if either $V = \mathbb{R}^n$ or V is a subspace of $\mathbb{R}^n$ for some positive integer n. (In more advanced textbooks on linear algebra, "vector space" is defined abstractly.)

2. Let W be a vector space. A set V is called a *subspace* of W if V is a vector space contained in W.

Example 3.5.2 Let

$$U = \text{span}\{(1,1,1)\}, \quad V = \text{span}\{(1,1,-1)\} \text{ and } W = \text{span}\{(1,0,0),\ (0,1,1)\}.$$

Since U, V and W are subspaces of $\mathbb{R}^3$, they are vector spaces.

Note that $(1,1,1) = (1,0,0) + (0,1,1)$. By Theorem 3.2.10, $U \subseteq W$. So U is a subspace of W. On the other hand, $(1,1,-1) \notin W$ (check it) and hence $V \not\subseteq W$. So V is not a subspace of W.

Discussion 3.5.3 Given a vector space V, we want to find a set S, as small as possible, so that every vector in V is a linear combination of the elements in S. Such a set can then be used to build a "coordinate system" for V.

Definition 3.5.4 Let $S = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}\}$ be a subset of a vector space V. Then S is called a *basis* (plural *bases*) for V if

1. S is linearly independent and
2. S spans V.

Example 3.5.5

1. Show that $S = \{(1,2,1), (2,9,0), (3,3,4)\}$ is a basis for $\mathbb{R}^3$.

   Solution

   (a) The equation
   $$c_1(1,2,1) + c_2(2,9,0) + c_4(3,3,4) = (0,0,0)$$
   gives us a system of linear equations
   $$\begin{cases} c_1 + 2c_2 + 3c_3 = 0 \\ 2c_1 + 9c_2 + 3c_3 = 0 \\ c_1 \qquad\quad + 4c_3 = 0 \end{cases}$$
   which has only the trivial solution (check it). So S is linearly independent.

   (b) Since $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 9 & 3 \\ 1 & 0 & 4 \end{pmatrix}$ has a row-echelon form $\begin{pmatrix} 1 & 2 & 3 \\ 0 & 5 & -3 \\ 0 & 0 & -\frac{1}{5} \end{pmatrix}$ which does not have any zero row, by Discussion 3.2.5, $\text{span}(S) = \mathbb{R}^3$.

   By (a) and (b), S is a basis for $\mathbb{R}^3$.

2. Let $V = \text{span}\{(1, 1, 1, 1),\ (1, -1, -1, 1),\ (1, 0, 0, 1)\}$ and $S = \{(1, 1, 1, 1),\ (1, -1, -1, 1)\}$. Show that S is a basis for V.

   Solution

   (a) The equation
   $$c_1(1, 1, 1, 1) + c_2(1, -1, -1, 1) = (0, 0, 0, 0)$$
   gives us a system of linear equations
   $$\begin{cases} c_1 + c_2 = 0 \\ c_1 - c_2 = 0 \\ c_1 - c_2 = 0 \\ c_1 + c_2 = 0 \end{cases}$$
   which has only the trivial solution (check it). So S is linearly independent.

   (b) Since $(1, 0, 0, 1) = \frac{1}{2}(1, 1, 1, 1) + \frac{1}{2}(1, -1, -1, 1)$, by Theorem 3.2.12, $\text{span}(S) = V$.

   By (a) and (b), S is a basis for V.

3. Is $S = \{(1, 1, 1, 1),\ (0, 0, 1, 2),\ (-1, 0, 0, 1)\}$ a basis for $\mathbb{R}^4$?

   Solution By Theorem 3.2.7, $\text{span}(S) \neq \mathbb{R}^4$. So S is not a basis for $\mathbb{R}^4$.

4. Let $V = \text{span}(S)$ where $S = \{(1, 1, 1),\ (0, 0, 1),\ (1, 1, 0)\}$. Is S a basis for V?

   Solution Note that S is linearly dependent (check it). So S is not a basis for V.

Remark 3.5.6

1. A basis for a vector space V contains the smallest possible number of vectors that can span V.
2. For convenience, we say that the empty set, $\emptyset$, is the basis for the zero space.
3. Except the zero space, any vector space has infinitely many different bases.

Theorem 3.5.7 If $S = \{\boldsymbol{u_1},\ \boldsymbol{u_2},\ \ldots,\ \boldsymbol{u_k}\}$ is a basis for a vector space V, then every vector $\boldsymbol{v}$ in V can be expressed in the form
$$\boldsymbol{v} = c_1\boldsymbol{u_1} + c_2\boldsymbol{u_2} + \cdots + c_k\boldsymbol{u_k}$$
in exactly one way, where $c_1,\ c_2,\ \ldots,\ c_k$ are real numbers.

Proof Since S spans V, every vector $\boldsymbol{v}$ can be expressed in the form $\boldsymbol{v} = c_1\boldsymbol{u_1} + c_2\boldsymbol{u_2} + \cdots + c_k\boldsymbol{u_k}$. Suppose that a vector $\boldsymbol{v}$ can be expressed in two ways
$$\boldsymbol{v} = c_1\boldsymbol{u_1} + c_2\boldsymbol{u_2} + \cdots + c_k\boldsymbol{u_k} \quad \text{and} \quad \boldsymbol{v} = d_1\boldsymbol{u_1} + d_2\boldsymbol{u_2} + \cdots + d_k\boldsymbol{u_k}.$$

where $c_1, c_2, \ldots, c_k, d_1, d_2, \ldots, d_k$ are real numbers. Subtracting the second equation to the first equation, we obtain

$$(c_1 - d_1)\boldsymbol{u_1} + (c_2 - d_2)\boldsymbol{u_2} + \cdots + (c_k - d_k)\boldsymbol{u_k} = \mathbf{0}.$$

Since S is linearly independent, the only possible solution is

$$c_1 - d_1 = 0,\ c_2 - d_2 = 0,\ \ldots,\ c_k - d_k = 0,$$

i.e. $c_1 = d_1$, $c_2 = d_2$, $\ldots$, $c_k = d_k$. So the expression is unique.

Definition 3.5.8 Let $S = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}\}$ be a basis for a vector space V and $\boldsymbol{v}$ a vector in V. By Theorem 3.5.7, $\boldsymbol{v}$ is expressed uniquely as a linear combination

$$\boldsymbol{v} = c_1\boldsymbol{u_1} + c_2\boldsymbol{u_2} + \cdots + c_k\boldsymbol{u_k}.$$

The coefficients $c_1, c_2, \ldots, c_k$ are called the *coordinates* of $\boldsymbol{v}$ relative to the basis S. The vector

$$(\boldsymbol{v})_S = (c_1, c_2, \ldots, c_k) \in \mathbb{R}^k$$

is called the *coordinate vector* of $\boldsymbol{v}$ relative to the basis S. (Here we assume the vectors of S are in a fixed order so that $\boldsymbol{u_1}$ is the first vector, $\boldsymbol{u_2}$ is the second vector, etc.)

Example 3.5.9

1. Let $S = \{(1,2,1), (2,9,0), (3,3,4)\}$. It is easy to check that S is a basis for $\mathbb{R}^3$.

   (a) Find the coordinate vector of $\boldsymbol{v} = (5,-1,9)$ relative to S.

   (b) Find a vector $\boldsymbol{w}$ in $\mathbb{R}^3$ such that $(\boldsymbol{w})_S = (-1,3,2)$.

   Solution

   (a) Solving the equation

   $$a(1,2,1) + b(2,9,0) + c(3,3,4) = (5,-1,9),$$

   we obtain only one solution $a = 1$, $b = -1$ and $c = 2$, i.e.

   $$\boldsymbol{v} = (1,2,1) - (2,9,0) + 2(3,3,4)$$

   and $(\boldsymbol{v})_S = (1,-1,2)$.

   (b) $\boldsymbol{w} = -(1,2,1) + 3(2,9,0) + 2(3,3,4) = (11,31,7)$.

2. Let $\boldsymbol{v} = (2,3) \in \mathbb{R}^2$. Find the coordinate vector of $\boldsymbol{v}$ relative to each of the following bases for $\mathbb{R}^2$.

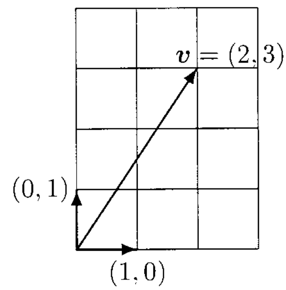
(a) $S_1 = \{(1,0), (0,1)\}$;

(b) $S_2 = \{(1,-1), (1,1)\}$;

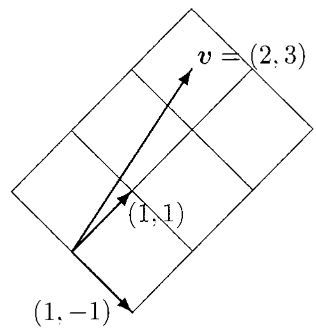
(c) $S_3 = \{(1,0), (1,1)\}$.

Solution

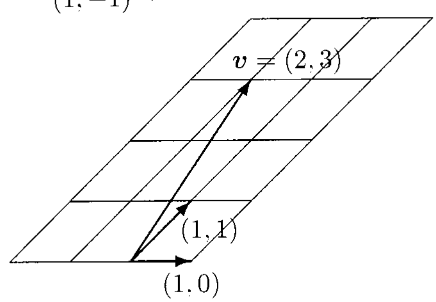
(a) Since $\boldsymbol{v} = (2,3) = 2(1,0) + 3(0,1)$, $(\boldsymbol{v})_{S_1} = (2,3)$.



(b) Since $\boldsymbol{v} = (2,3) = -\frac{1}{2}(1,-1) + \frac{5}{2}(1,1)$, $(\boldsymbol{v})_{S_2} = (-\frac{1}{2}, \frac{5}{2})$.



(c) Since $\boldsymbol{v} = (2,3) = -(1,0) + 3(1,1)$, $(\boldsymbol{v})_{S_3} = (-1,3)$.



3. **(Standard Basis for $\mathbb{R}^n$)** Let $E = \{\boldsymbol{e_1}, \boldsymbol{e_2}, \ldots, \boldsymbol{e_n}\}$ where $\boldsymbol{e_1} = (1,0,\ldots,0)$, $\boldsymbol{e_2} = (0,1,0,\ldots,0)$, $\ldots$, $\boldsymbol{e_n} = (0,\ldots,0,1)$ are vectors in $\mathbb{R}^n$. By Remark 3.3.2.2, E spans $\mathbb{R}^n$. Also, it is easy to show that E is linearly independent. Thus E is a basis for $\mathbb{R}^n$ which is called the *standard basis* for $\mathbb{R}^n$. For any $\boldsymbol{u} = (u_1, u_2, \ldots, u_n) \in \mathbb{R}^n$,

$$(\boldsymbol{u})_E = (u_1, u_2, \ldots, u_n) = \boldsymbol{u}.$$

Remark 3.5.10 The following are some useful rules in working with coordinate vectors:

Let S be a basis for a vector space V.

1. For any $\boldsymbol{u}, \boldsymbol{v} \in V$, $\boldsymbol{u} = \boldsymbol{v}$ if and only if $(\boldsymbol{u})_S = (\boldsymbol{v})_S$.

2. For any $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_r} \in V$ and $c_1, c_2, \ldots, c_r \in \mathbb{R}$,
$$(c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_r\boldsymbol{v_r})_S = c_1(\boldsymbol{v_1})_S + c_2(\boldsymbol{v_2})_S + \cdots + c_r(\boldsymbol{v_r})_S.$$

Theorem 3.5.11 Let S be a basis for a vector space V where $|S| = k$. Let $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_r}$ be vectors in V. Then

1. $\boldsymbol{v_1}$, $\boldsymbol{v_2}$, ..., $\boldsymbol{v_r}$ are linearly dependent (respectively, independent) vectors in V if and only if $(\boldsymbol{v_1})_S$, $(\boldsymbol{v_2})_S$, ..., $(\boldsymbol{v_r})_S$ are linearly dependent (respectively, independent) vectors in $\mathbb{R}^k$; and

2. $\text{span}\{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_r}\} = V$ if and only if $\text{span}\{(\boldsymbol{v_1})_S, (\boldsymbol{v_2})_S, \ldots, (\boldsymbol{v_r})_S\} = \mathbb{R}^k$.

Proof

1. By Remark 3.5.10,
$$\begin{aligned} & c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_r\boldsymbol{v_r} = \boldsymbol{0} \\ \Leftrightarrow\quad & (c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_r\boldsymbol{v_r})_S = (\boldsymbol{0})_S \\ \Leftrightarrow\quad & c_1(\boldsymbol{v_1})_S + c_2(\boldsymbol{v_2})_S + \cdots + c_r(\boldsymbol{v_r})_S = (\boldsymbol{0})_S, \end{aligned}$$
where $(\boldsymbol{0})_S = (0, \ldots, 0)$ is the zero vector in $\mathbb{R}^k$. The equation $c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_r\boldsymbol{v_r} = \boldsymbol{0}$ has non-trivial solution if and only if the equation $c_1(\boldsymbol{v_1})_S + c_2(\boldsymbol{v_2})_S + \cdots + c_r(\boldsymbol{v_r})_S = (\boldsymbol{0})_S$ has non-trivial solution. Hence $\boldsymbol{v_1}$, $\boldsymbol{v_2}$, ..., $\boldsymbol{v_r}$ are linearly dependent (respectively, independent) vectors in V if and only if $(\boldsymbol{v_1})_S$, $(\boldsymbol{v_2})_S$, ..., $(\boldsymbol{v_r})_S$ are linearly dependent (respectively, independent) vectors in $\mathbb{R}^k$.

2. Let $S = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}\}$.

   ($\Rightarrow$) Suppose $\text{span}\{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_r}\} = V$. Take any vector $(a_1, a_2, \ldots, a_k) \in \mathbb{R}^k$. Let $\boldsymbol{w} = a_1\boldsymbol{u_1} + a_2\boldsymbol{u_2} + \cdot + a_k\boldsymbol{u_k} \in V$. Since $\boldsymbol{v_1}$, $\boldsymbol{v_2}$, ..., $\boldsymbol{v_r}$ span V, there exist real numbers $c_1, c_2, \ldots, c_r$ such that $\boldsymbol{w} = c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_r\boldsymbol{v_r}$. By Remark 3.5.10,
$$\begin{aligned} c_1(\boldsymbol{v_1})_S + c_2(\boldsymbol{v_2})_S + \cdots + c_r(\boldsymbol{v_r})_S &= (c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_r\boldsymbol{v_r})_S \\ &= (\boldsymbol{w})_S = (a_1, a_2, \ldots, a_k). \end{aligned}$$
   Since every vector in $\mathbb{R}^k$ is a linear combination of $(\boldsymbol{v_1})_S$, $(\boldsymbol{v_2})_S$, ..., $(\boldsymbol{v_r})_S$, we have $\text{span}\{(\boldsymbol{v_1})_S, (\boldsymbol{v_2})_S, \ldots, (\boldsymbol{v_r})_S\} = \mathbb{R}^k$.

   ($\Leftarrow$) Suppose $\text{span}\{(\boldsymbol{v_1})_S, (\boldsymbol{v_2})_S, \ldots, (\boldsymbol{v_r})_S\} = \mathbb{R}^k$. Take any vector $\boldsymbol{w} \in V$. Since $(\boldsymbol{w})_S$ is a vector in $\mathbb{R}^k$ and $(\boldsymbol{v_1})_S$, $(\boldsymbol{v_2})_S$, ..., $(\boldsymbol{v_r})_S$ span $\mathbb{R}^k$, there exist real numbers $c_1, c_2, \ldots, c_r$ such that
$$\begin{aligned} (\boldsymbol{w})_S &= c_1(\boldsymbol{v_1})_S + c_2(\boldsymbol{v_2})_S + \cdots + c_r(\boldsymbol{v_r})_S \\ &= (c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_r\boldsymbol{v_r})_S \end{aligned}$$

and hence $\boldsymbol{w} = c_1\boldsymbol{v_1} + c_2\boldsymbol{v_2} + \cdots + c_r\boldsymbol{v_r}$. Since every vector in V is a linear combination of $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_r}$, we have $\mathrm{span}\{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_r}\} = V$.

Section 3.6 Dimensions

Theorem 3.6.1 Let V be a vector space which has a basis with k vectors. Then

1. any subset of V with more than k vectors is always linearly dependent;

2. any subset of V with less than k vectors cannot span V.

Proof Let S be a basis for V and $|S| = k$.

1. Let $T = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_r}\}$ be a subset of V with $r > k$. Since $(\boldsymbol{v_1})_S, (\boldsymbol{v_2})_S, \ldots, (\boldsymbol{v_r})_S$ are vectors in $\mathbb{R}^k$, by Theorem 3.4.7, $(\boldsymbol{v_1})_S, (\boldsymbol{v_2})_S, \ldots, (\boldsymbol{v_r})_S$ are linearly dependent. Thus by Theorem 3.5.11.1, T is linearly dependent.

2. Let $T = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_r}\}$ be a subset of V with $r < k$. By Theorem 3.2.7, $(\boldsymbol{v_1})_S$, $(\boldsymbol{v_2})_S, \ldots, (\boldsymbol{v_r})_S$ cannot span $\mathbb{R}^k$. Thus by Theorem 3.5.11.2, T cannot span V.

Remark 3.6.2 By Theorem 3.6.1, all bases for a vector space have the same number of vectors. This number gives us a way to measure the "size" of a vector space.

Definition 3.6.3 The *dimension* of a vector space V, denoted by $\dim(V)$, is defined to be the number of vectors in a basis for V. In addition, we define the dimension of the zero space to be zero.

Example 3.6.4

1. $\dim(\mathbb{R}^n) = n$.

2. Except $\{\boldsymbol{0}\}$ and $\mathbb{R}^2$, all subspaces of $\mathbb{R}^2$ are lines through the origin and they are of dimension 1.

3. Except $\{\boldsymbol{0}\}$ and $\mathbb{R}^3$, all subspaces of $\mathbb{R}^3$ are either lines through the origin, which are of dimension 1, or planes containing the origin, which are of dimension 2.

4. Find a basis for and determine the dimension of the subspace $W = \{ (x, y, z) \mid y = z \}$ of $\mathbb{R}^3$.

   Solution Every vector in W is of the form
$$(x, y, y) = x(1, 0, 0) + y(0, 1, 1).$$
   So $W = \text{span}\{(1, 0, 0), (0, 1, 1)\}$. It is easy to check that $(1, 0, 0)$, $(0, 1, 1)$ are linearly independent. Hence $\{(1, 0, 0), (0, 1, 1)\}$ is a basis for W and $\dim(W) = 2$.

Discussion 3.6.5 Given a homogeneous linear system $\boldsymbol{Ax} = \mathbf{0}$, we want to find a basis for and determine the dimension of the solution space.

First, use Gauss-Jordan Elimination to convert the augmented matrix to the reduced row-echelon form. Then set the variables corresponding to non-pivot columns as arbitrary and equate the other variables accordingly. Finally, write a general solution for the system in the form of $t_1\boldsymbol{u_1} + t_2\boldsymbol{u_2} + \cdots + t_k\boldsymbol{u_k}$ where $t_1, t_2, \ldots, t_k$ are arbitrary parameters and $\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}$ are fixed vectors. In this way, the vectors $\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}$ are always linearly independent and hence $\{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}\}$ is a basis for the solution space.

Let $\boldsymbol{R}$ be a row-echelon form of $\boldsymbol{A}$. Then the number of arbitrary parameters needed in the general solution is equal to the number of non-pivot columns in $\boldsymbol{R}$. So the dimension of the solution space of the system is equal to the number of non-pivot columns in $\boldsymbol{R}$. (See also Theorem 4.3.4.)

Example 3.6.6 Find a basis for and determine the dimension of the solution space of the homogeneous system
$$\begin{cases} 2v + 2w - x + z = 0 \\ -v - w + 2x - 3y + z = 0 \\ x + y + z = 0 \\ v + w - 2x - z = 0. \end{cases}$$

Solution Since
$$\left(\begin{array}{ccccc|c} 2 & 2 & -1 & 0 & 1 & 0 \\ -1 & -1 & 2 & -3 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & -2 & 0 & -1 & 0 \end{array}\right) \xrightarrow[\text{Elimination}]{\text{Gauss-Jordan}} \left(\begin{array}{ccccc|c} 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right),$$
the linear system has a general solution
$$\begin{pmatrix} v \\ w \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -s-t \\ s \\ -t \\ 0 \\ t \end{pmatrix} = s\begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + t\begin{pmatrix} -1 \\ 0 \\ -1 \\ 0 \\ 1 \end{pmatrix} \quad \text{where } s, t \text{ are arbitrary parameters.}$$

Then $\{(-1,1,0,0,0), (-1,0,-1,0,1)\}$ is a basis for the solution space and the dimension of the solution space is 2.

Theorem 3.6.7 Let V be a vector space of dimension k and S a subset of V. The following are equivalent:

1. S is a basis for V.
2. S is linearly independent and $|S| = k$.
3. S spans V and $|S| = k$.

Proof "$1 \Rightarrow 2$" and "$1 \Rightarrow 3$" follows from Remark 3.6.2 directly.

$2 \Rightarrow 1$: Suppose S is linearly independent and $|S| = k$. Assume that S is not a basis for V, i.e. $\text{span}(S) \neq V$. Take a vector $\boldsymbol{u}$ in V but not in $\text{span}(S)$. By Theorem 3.4.10, $S' = S \cup \{\boldsymbol{u}\}$ is a set of $k+1$ linearly independent vectors. But this contradicts Theorem 3.6.1.

$3 \Rightarrow 1$: Suppose S spans V and $|S| = k$. Assume that S is not a basis for V, i.e. S is linearly dependent. Take a vector $\boldsymbol{v}$ in S which is a linear combination of other vectors in S. By Theorem 3.2.12, $S'' = S - \{\boldsymbol{v}\}$ is a set of $k-1$ vectors that spans V. But this also contradicts Theorem 3.6.1.

Example 3.6.8 Show that $\boldsymbol{u_1} = (2,0,-1)$, $\boldsymbol{u_2} = (4,0,7)$, $\boldsymbol{u_3} = (-1,1,4)$ form a basis for $\mathbb{R}^3$.

Solution Consider the vector equation

$$c_1\boldsymbol{u_1} + c_2\boldsymbol{u_2} + c_3\boldsymbol{u_3} = \boldsymbol{0},$$

i.e. $c_1(2,0,-1) + c_2(4,0,7) + c_3(-1,1,4) = (0,0,0)$. This implies

$$\begin{cases} 2c_1 + 4c_2 - c_3 = 0 \\ c_3 = 0 \\ -c_1 + 7c_3 + 4c_3 = 0 \end{cases}$$

On solving the system, we get $c_1 = 0$, $c_2 = 0$ and $c_3 = 0$. So $\boldsymbol{u_1}$, $\boldsymbol{u_2}$, $\boldsymbol{u_3}$ are linearly independent. As $\dim(\mathbb{R}^3) = 3$, by Theorem 3.6.7, $\{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\}$ is a basis for $\mathbb{R}^3$.

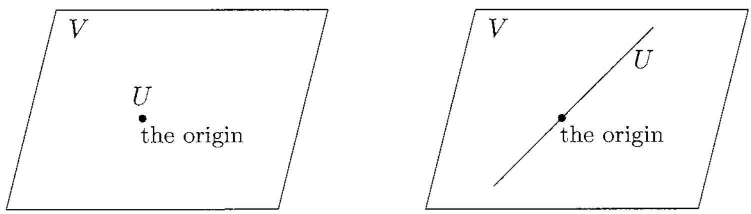
Theorem 3.6.9 Let U be a subspace of a vector space V. Then $\dim(U) \leq \dim(V)$. Furthermore, if $U \neq V$, then $\dim(U) < \dim(V)$.

Proof Let S be a basis for U. Since $U \subseteq V$, S is a linearly independent subset of V. By Theorem 3.6.1.1, $\dim(U) = |S| \leq \dim(V)$.

Assume $\dim(U) = \dim(V)$. As S is linearly independent and $|S| = \dim(U) = \dim(V)$, by Theorem 3.6.7, S is a basis for V. But then $U = \text{span}(S) = V$. Hence if $U \neq V$, then $\dim(U) < \dim(V)$.

Example 3.6.10 Let V be a plane in $\mathbb{R}^3$ containing the origin. Note that V is a vector space of dimension 2, see Example 3.6.4.3.

Suppose U is a subspace of V such that $U \neq V$. By Theorem 3.6.9, $\dim(U) < 2$. Hence U is either $\{(0, 0, 0)\}$ (which is of dimension 0) or a line through the origin (which is of dimension 1).



Theorem 3.6.11 (This theorem is a continuation of Theorem 2.4.7 and forms part of our main theorem on invertible matrices, see Theorem 6.1.8.) Let $\boldsymbol{A}$ be an $n \times n$ matrices. The following statements are equivalent.

1. $\boldsymbol{A}$ is invertible.

2. The linear system $\boldsymbol{Ax} = \boldsymbol{0}$ has only the trivial solution.

3. The reduced row-echelon form of $\boldsymbol{A}$ is an identity matrix.

4. $\boldsymbol{A}$ can be expressed as a product of elementary matrices.

5. $\det(\boldsymbol{A}) \neq 0$.

6. The rows of $\boldsymbol{A}$ form a basis for $\mathbb{R}^n$.

7. The columns of $\boldsymbol{A}$ form a basis for $\mathbb{R}^n$.

Proof By Theorem 2.4.7, statements 1 to 4 are equivalent. By Theorem 2.5.19, we have "$1 \Leftrightarrow 5$".

The rows of $\boldsymbol{A}$ are the columns of $\boldsymbol{A}^{\mathrm{T}}$. Since $\boldsymbol{A}$ is invertible if and only if $\boldsymbol{A}^{\mathrm{T}}$ is invertible (see Theorem 2.3.9), we only need to show either "$1 \Leftrightarrow 6$" or "$1 \Leftrightarrow 7$".

In the following, we prove "$1 \Leftrightarrow 7$":

Let $\boldsymbol{A} = (\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n)$ where $\boldsymbol{a}_i$ is the ith column of $\boldsymbol{A}$.

$$
\begin{aligned}
& \{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n\} \text{ is a basis for } \mathbb{R}^n & \\
\Leftrightarrow\ & \operatorname{span}\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n\} = \mathbb{R}^n & \text{(by Theorem 3.6.7)} \\
\Leftrightarrow\ & \text{a row echelon form of } \boldsymbol{A} \text{ has no zero row} & \text{(by Discussion 3.2.5)} \\
\Leftrightarrow\ & \boldsymbol{A} \text{ is invertible.} & \text{(by Remark 2.4.10)}
\end{aligned}
$$

Example 3.6.12

1. Let $\boldsymbol{u_1} = (1, 1, 1)$, $\boldsymbol{u_2} = (-1, 1, 2)$ and $\boldsymbol{u_3} = (1, 0, 1)$. Since

$$\begin{vmatrix} 1 & -1 & 1 \\ 1 & 1 & 0 \\ 1 & 2 & 1 \end{vmatrix} = 3,$$

   $\{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\}$ is a basis for $\mathbb{R}^3$.

2. Let $\boldsymbol{u_1} = (1, 1, 1, 1)$, $\boldsymbol{u_2} = (1, -1, 1, -1)$, $\boldsymbol{u_3} = (0, 1, -1, 0)$ and $\boldsymbol{u_4} = (2, 1, 1, 0)$. Since

$$\begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 0 & 1 & -1 & 0 \\ 2 & 1 & 1 & 0 \end{vmatrix} = 0,$$

   $\{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}, \boldsymbol{u_4}\}$ is not a basis for $\mathbb{R}^4$.

Section 3.7 Transition Matrices

Notation 3.7.1 Let $S = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}\}$ be a basis for a vector space V and let $\boldsymbol{v}$ be a vector in V. Recall that if $\boldsymbol{v} = c_1\boldsymbol{u_1} + c_2\boldsymbol{u_2} + \cdots + c_k\boldsymbol{u_k}$, then the vector

$$(\boldsymbol{v})_S = (c_1, c_2, \ldots, c_k)$$

is called the coordinate vector of $\boldsymbol{v}$ relative to S. Sometimes, it is more convenient to write the coordinate vector in the form of a column vector. Thus we define

$$[\boldsymbol{v}]_S = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix}$$

and also call it the *coordinate vector* of $\boldsymbol{v}$ relative to S.

Discussion 3.7.2 Let $S = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \dots, \boldsymbol{u_k}\}$ and $T = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_k}\}$ be two bases for a vector space V. For any vector $\boldsymbol{w}$ in V, we want to study the relation between $[\boldsymbol{w}]_S$ and $[\boldsymbol{w}]_T$.

Suppose $\boldsymbol{w} = c_1\boldsymbol{u_1} + c_2\boldsymbol{u_2} + \cdots + c_k\boldsymbol{u_k}$, i.e.

$$[\boldsymbol{w}]_S = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix}.$$

Since T is a basis for V, we can write each $\boldsymbol{u_i}$ as a linear combination of $\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_k}$, say

$$\begin{aligned} \boldsymbol{u_1} &= a_{11}\boldsymbol{v_1} + a_{21}\boldsymbol{v_2} + \cdots + a_{k1}\boldsymbol{v_k}, \\ \boldsymbol{u_2} &= a_{12}\boldsymbol{v_1} + a_{22}\boldsymbol{v_2} + \cdots + a_{k2}\boldsymbol{v_k}, \\ &\vdots \\ \boldsymbol{u_k} &= a_{1k}\boldsymbol{v_1} + a_{2k}\boldsymbol{v_2} + \cdots + a_{kk}\boldsymbol{v_k}, \end{aligned}$$

where $a_{11}, a_{12}, \dots, a_{kk}$ are real constants, i.e.

$$[\boldsymbol{u_1}]_T = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{k1} \end{pmatrix}, \quad [\boldsymbol{u_2}]_T = \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{k2} \end{pmatrix}, \quad \dots, \quad [\boldsymbol{u_k}]_T = \begin{pmatrix} a_{1k} \\ a_{2k} \\ \vdots \\ a_{kk} \end{pmatrix}.$$

Then

$$\begin{aligned} \boldsymbol{w} &= c_1\boldsymbol{u_1} + c_2\boldsymbol{u_2} + \cdots + c_k\boldsymbol{u_k} \\ &= c_1(a_{11}\boldsymbol{v_1} + a_{21}\boldsymbol{v_2} + \cdots + a_{k1}\boldsymbol{v_k}) + c_2(a_{12}\boldsymbol{v_1} + a_{22}\boldsymbol{v_2} + \cdots + a_{k2}\boldsymbol{v_k}) \\ &\qquad + \cdots + c_k(a_{1k}\boldsymbol{v_1} + a_{2k}\boldsymbol{v_2} + \cdots + a_{kk}\boldsymbol{v_k}) \\ &= (c_1a_{11} + c_2a_{12} + \cdots + c_ka_{1k})\boldsymbol{v_1} + (c_1a_{21} + c_2a_{22} + \cdots + c_ka_{2k})\boldsymbol{v_2} \\ &\qquad + \cdots + (c_1a_{k1} + c_2a_{k2} + \cdots + c_ka_{kk})\boldsymbol{v_k} \end{aligned}$$

i.e.

$$\begin{aligned} [\boldsymbol{w}]_T = \begin{pmatrix} c_1a_{11} + c_2a_{12} + \cdots + c_ka_{1k} \\ c_1a_{21} + c_2a_{22} + \cdots + c_ka_{2k} \\ \vdots \\ c_1a_{k1} + c_2a_{k2} + \cdots + c_ka_{kk} \end{pmatrix} &= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix} \\ &= \begin{pmatrix} [\boldsymbol{u_1}]_T & [\boldsymbol{u_2}]_T & \cdots & [\boldsymbol{u_k}]_T \end{pmatrix} [\boldsymbol{w}]_S. \end{aligned}$$

Let $\boldsymbol{P} = \begin{pmatrix} [\boldsymbol{u_1}]_T & [\boldsymbol{u_2}]_T & \cdots & [\boldsymbol{u_k}]_T \end{pmatrix}$. Then

$$[\boldsymbol{w}]_T = \boldsymbol{P}[\boldsymbol{w}]_S \quad \text{for all } \boldsymbol{w} \in V.$$

Definition 3.7.3 Let $S = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}\}$ and T be two bases for a vector space. The square matrix $\boldsymbol{P} = \begin{pmatrix}[\boldsymbol{u_1}]_T & [\boldsymbol{u_2}]_T & \cdots & [\boldsymbol{u_k}]_T\end{pmatrix}$ in Discussion 3.7.2 is called the *transition matrix* from S to T.

Example 3.7.4

1. Consider the following two bases S and T for $\mathbb{R}^3$:
   $S = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\}$, where $\boldsymbol{u_1} = (1, 0, -1)$, $\boldsymbol{u_2} = (0, -1, 0)$ and $\boldsymbol{u_3} = (1, 0, 2)$, and
   $T = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}\}$, where $\boldsymbol{v_1} = (1, 1, 1)$, $\boldsymbol{v_2} = (1, 1, 0)$ and $\boldsymbol{v_3} = (-1, 0, 0)$.

   (a) Find the transition matrix from S to T.

   (b) Let $\boldsymbol{w}$ be a vector in $\mathbb{R}^3$ such that $(\boldsymbol{w})_S = (2, -1, 2)$. Find $(\boldsymbol{w})_T$.

   Solution

   (a) First, we need to find $a_{11}, a_{21}, \ldots, a_{33}$ such that

   $$\begin{aligned}\boldsymbol{u_1} &= a_{11}\boldsymbol{v_1} + a_{21}\boldsymbol{v_2} + a_{31}\boldsymbol{v_3},\\ \boldsymbol{u_2} &= a_{12}\boldsymbol{v_1} + a_{22}\boldsymbol{v_2} + a_{32}\boldsymbol{v_3},\\ \boldsymbol{u_3} &= a_{13}\boldsymbol{v_1} + a_{23}\boldsymbol{v_2} + a_{33}\boldsymbol{v_3}.\end{aligned}$$

   By

   $$\left(\begin{array}{ccc|c|c|c} 1 & 1 & -1 & 1 & 0 & 1\\ 1 & 1 & 0 & 0 & -1 & 0\\ 1 & 0 & 0 & -1 & 0 & 2\end{array}\right) \xrightarrow[\text{Elimination}]{\text{Gauss-Jordan}} \left(\begin{array}{ccc|c|c|c} 1 & 0 & 0 & -1 & 0 & 2\\ 0 & 1 & 0 & 1 & -1 & -2\\ 0 & 0 & 1 & -1 & -1 & -1\end{array}\right),$$

   we have

   $$\begin{aligned}\boldsymbol{u_1} &= -\boldsymbol{v_1} + \boldsymbol{v_2} - \boldsymbol{v_3},\\ \boldsymbol{u_2} &= \quad\quad - \boldsymbol{v_2} - \boldsymbol{v_3},\\ \boldsymbol{u_3} &= 2\boldsymbol{v_1} - 2\boldsymbol{v_2} - \boldsymbol{v_3}.\end{aligned}$$

   (See Example 3.2.11 for the method of finding the linear combinations.)

   So $\boldsymbol{P} = \begin{pmatrix} -1 & 0 & 2\\ 1 & -1 & -2\\ -1 & -1 & -1\end{pmatrix}$ is the transition matrix from S to T.

   (b) Since $[\boldsymbol{w}]_T = \boldsymbol{P}[\boldsymbol{w}]_S = \begin{pmatrix} -1 & 0 & 2\\ 1 & -1 & -2\\ -1 & -1 & -1\end{pmatrix}\begin{pmatrix}2\\-1\\2\end{pmatrix} = \begin{pmatrix}2\\-1\\-3\end{pmatrix}$, $(\boldsymbol{w})_T = (2, -1, -3)$.

2. Let $S = \{\boldsymbol{u_1}, \boldsymbol{u_2}\}$, where $\boldsymbol{u_1} = (1,1)$ and $\boldsymbol{u_2} = (1,-1)$, and let $T = \{\boldsymbol{v_1}, \boldsymbol{v_2}\}$, where $\boldsymbol{v_1} = (1,0)$ and $\boldsymbol{v_2} = (1,1)$. Note that S and T are two bases for $\mathbb{R}^2$.

   It is easy to check that
$$\begin{aligned}\boldsymbol{u_1} &= \phantom{2\boldsymbol{v_1} -{}} \boldsymbol{v_2}, \\ \boldsymbol{u_2} &= 2\boldsymbol{v_1} - \boldsymbol{v_2}.\end{aligned}$$

   Thus $\boldsymbol{P} = \begin{pmatrix} 0 & 2 \\ 1 & -1 \end{pmatrix}$ is the transition matrix from S to T.

   Also, it is easy to get
$$\begin{aligned}\boldsymbol{v_1} &= \tfrac{1}{2}\boldsymbol{u_1} + \tfrac{1}{2}\boldsymbol{u_2}, \\ \boldsymbol{v_2} &= \phantom{\tfrac{1}{2}}\boldsymbol{u_1} \qquad\quad .\end{aligned}$$

   Thus $\boldsymbol{Q} = \begin{pmatrix} \frac{1}{2} & 1 \\ \frac{1}{2} & 0 \end{pmatrix}$ is the transition matrix from T to S. Note that $\boldsymbol{Q} = \boldsymbol{P}^{-1}$.

Theorem 3.7.5 Let S and T be two bases of a vector space and let $\boldsymbol{P}$ be the transition matrix from S to T. Then

1. $\boldsymbol{P}$ is invertible; and
2. $\boldsymbol{P}^{-1}$ is the transition matrix from T to S.

Proof Let $\boldsymbol{Q}$ be the transition matrix from T to S. By Theorem 2.4.12, it suffices to show that $\boldsymbol{QP} = \boldsymbol{I}$.

Suppose $S = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}\}$. Note that
$$[\boldsymbol{u_1}]_S = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad [\boldsymbol{u_2}]_S = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots, \quad [\boldsymbol{u_k}]_S = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

Thus for $i = 1, 2, \ldots, k$,
$$\text{the } i\text{th column of } \boldsymbol{QP} = \boldsymbol{QP}[\boldsymbol{u_i}]_S = \boldsymbol{Q}[\boldsymbol{u_i}]_T = [\boldsymbol{u_i}]_S.$$

So $\boldsymbol{QP} = \begin{pmatrix}[\boldsymbol{u_1}]_S & [\boldsymbol{u_2}]_S & \cdots & [\boldsymbol{u_k}]_S\end{pmatrix} = \boldsymbol{I}$.

Example 3.7.6 In Example 3.7.4.1, the transition matrix from S to T is
$$\boldsymbol{P} = \begin{pmatrix} -1 & 0 & 2 \\ 1 & -1 & -2 \\ -1 & -1 & -1 \end{pmatrix}.$$

So the transition matrix from T to S is

$$\boldsymbol{P}^{-1} = \begin{pmatrix} -1 & 0 & 2 \\ 1 & -1 & -2 \\ -1 & -1 & -1 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & -\frac{2}{3} \\ -1 & -1 & 0 \\ \frac{2}{3} & \frac{1}{3} & -\frac{1}{3} \end{pmatrix}.$$

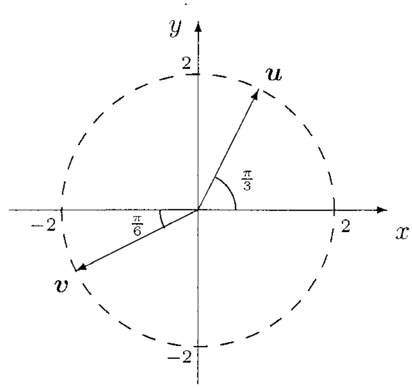
Note that

$$\boldsymbol{P}^{-1}[\boldsymbol{w}]_T = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & -\frac{2}{3} \\ -1 & -1 & 0 \\ \frac{2}{3} & \frac{1}{3} & -\frac{1}{3} \end{pmatrix} \begin{pmatrix} 2 \\ -1 \\ -3 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \\ 2 \end{pmatrix} = [\boldsymbol{w}]_S.$$

Exercise 3

Question 3.1 to Question 3.14 are exercises for Sections 3.1 and 3.2.

1. Find the vectors $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{u}+\boldsymbol{v}$ and $3\boldsymbol{u}-2\boldsymbol{v}$ in $\mathbb{R}^2$ where $\boldsymbol{u}$ and $\boldsymbol{v}$ are shown in the figure below.



2. Express each of the following by the set notation in both implicit and explicit form:

   (a) the line in $\mathbb{R}^2$ passing through the points $(1, 2)$ and $(2, -1)$.

   (b) the plan in $\mathbb{R}^3$ containing the points $(0, 1, -1)$, $(1, -1, 0)$ and $(0, 2, 0)$.

   (c) the line in $\mathbb{R}^3$ passing through the points $(0, 1, -1)$ and $(1, -1, 0)$.

3. Consider the following subsets of $\mathbb{R}^3$:

$$\begin{aligned}
A &= \text{a line passes through the origin and } (9,9,9)\\
B &= \{\,(k,k,k) \mid k \in \mathbb{R}\,\},\\
C &= \{\,(x_1,\, x_2,\, x_3) \mid x_1 = x_2 = x_3\,\},\\
D &= \{\,(x,y,z) \mid 2x - y - z = 0\,\},\\
E &= \{\,(a,b,c) \mid 2a - b - c = 0 \text{ and } a + b + c = 0\,\},\\
F &= \{\,(u,v,w) \mid 2u - v - w = 0 \text{ and } 3u - 2v - w = 0\,\}.
\end{aligned}$$

Which of these subsets are the same?

4. Let U, V, W be three planes in $\mathbb{R}^3$ where

$$\begin{aligned}
U &= \{\,(x,y,z) \mid 2x - y + 3z = 0\,\},\ V = \{\,(x,y,z) \mid 3x + 2y - z = 0\,\},\\
W &= \{\,(x,y,z) \mid x - 3y - 2z = 1\,\}.
\end{aligned}$$

(a) Determine which of U, V, W contain the origin.

(b) Write down the sets $U \cap V$ and $V \cap W$ explicitly.

5. Let $A = \{\,(1+t, 1+2t, 1+3t) \mid t \in \mathbb{R}\,\}$ be a subset of $\mathbb{R}^3$.

(a) Describe A geometrically.

(b) Show that $A = \{\,(x,y,z) \mid x + y - z = 1 \text{ and } x - 2y + z = 0\,\}$.

(c) Write down a matrix equation $\boldsymbol{Mx} = \boldsymbol{b}$ where $\boldsymbol{M}$ is a 3×3 matrix and $\boldsymbol{b}$ is a 3×1 matrix such that its solution set is A.

6. Determine whether the following subsets of $\mathbb{R}^4$ are equal to each other.

$$\begin{aligned}
S &= \{\,(p,q,p,q) \mid p, q \in \mathbb{R}\,\},\\
T &= \{\,(x,y,z,w) \mid x + y - z - w = 0\,\},\\
V &= \left\{ (a,b,c,d) \;\middle|\; \begin{vmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ a & b & c & d \end{vmatrix} = 0 \right\}.
\end{aligned}$$

Briefly explain why one subset is equal (or not equal) to another subset.

7. Let P represent a plane in $\mathbb{R}^3$ with equation $x - y + z = 1$ and A, B, C represent three different lines given by the following set notation:

$$A = \{\,(a,a,1) \mid a \in \mathbb{R}\,\}, \quad B = \{\,(b,0,0) \mid b \in \mathbb{R}\,\}, \quad C = \{\,(c,0,-c) \mid c \in \mathbb{R}\,\}.$$

(a) Express the plane P in explicit set notation.

(b) Does any of the three lines above lie completely on the plane P? Briefly explain your answer.

(c) Find all the points of intersection of the line B with the plane P.

(d) Find the equation of another plane that is parallel to (but not overlapping) the plane P, and contains exactly one of the three lines above.

(e) Can you find a nonzero linear system whose solution set contains all the three lines? Justify your answer.

8. Let $\boldsymbol{u_1} = (2, 1, 0, 3)$, $\boldsymbol{u_2} = (3, -1, 5, 2)$, and $\boldsymbol{u_3} = (-1, 0, 2, 1)$. Which of the following vectors are linear combinations of $\boldsymbol{u_1}$, $\boldsymbol{u_2}$, $\boldsymbol{u_3}$?

   (a) $(2, 3, -7, 3)$, (b) $(0, 0, 0, 0)$, (c) $(1, 1, 1, 1)$, (d) $(-4, 6, -13, 4)$.

9. For each of the following sets, determine whether the set spans $\mathbb{R}^3$.

   (a) $S_1 = \{(1, 1, -1), (-2, 2, 1)\}$.

   (b) $S_2 = \{(1, 1, -1), (-2, -2, 2)\}$.

   (c) $S_3 = \{(1, 1, -1), (-2, 2, 1), (1, 5, -2)\}$.

   (d) $S_4 = \{(1, 1, -1), (-2, 2, 1), (4, 0, 3)\}$.

   (e) $S_5 = \{(1, 1, -1), (-2, 2, 1), (1, 5, -2)\ (0, 8, -2)\}$.

   (f) $S_6 = \{(1, 1, -1), (-2, 2, 1), (4, 0, 3), (2, 6, -3)\}$.

10. Let $V = \{\, (x, y, z) \mid x - y - z = 0 \,\}$ be a subset of $\mathbb{R}^3$.

   (a) Let $S = \{(1, 1, 0), (5, 2, 3)\}$. Show that $\mathrm{span}(S) = V$.

   (b) Let $S' = \{(1, 1, 0), (5, 2, 3), (0, 0, 1)\}$. Show that $\mathrm{span}(S') = \mathbb{R}^3$.

11. Determine whether $\mathrm{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\} = \mathrm{span}\{\boldsymbol{v_1}, \boldsymbol{v_2}\}$ if

   (a) $\boldsymbol{u_1} = (2, -2, 0)$, $\boldsymbol{u_2} = (-1, 1, -1)$, $\boldsymbol{u_3} = (0, 0, 9)$, $\boldsymbol{v_1} = (1, -1, -5)$, $\boldsymbol{v_2} = (0, 1, 1)$.

   (b) $\boldsymbol{u_1} = (1, 6, 4)$, $\boldsymbol{u_2} = (2, 4, -1)$, $\boldsymbol{u_3} = (-1, 2, 5)$, $\boldsymbol{v_1} = (1, -2, -5)$, $\boldsymbol{v_2} = (0, 8, 9)$.

12. Let $\boldsymbol{u_1} = (2, 0, 2, -4)$, $\boldsymbol{u_2} = (1, 0, 2, 5)$, $\boldsymbol{u_3} = (0, 3, 6, 9)$, $\boldsymbol{u_4} = (1, 1, 2, -1)$, $\boldsymbol{v_1} = (-1, 2, 1, 0)$, $\boldsymbol{v_2} = (3, 1, 4, 0)$, $\boldsymbol{v_3} = (0, 1, 1, 3)$, $\boldsymbol{v_4} = (-4, 3, -1, 6)$. Determine if the following are true.

   (a) $\mathrm{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}, \boldsymbol{u_4}\} \subseteq \mathrm{span}\{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}, \boldsymbol{v_4}\}$.

(b) $\text{span}\{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}, \boldsymbol{v_4}\} \subseteq \text{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}, \boldsymbol{u_4}\}$.

(c) $\text{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}, \boldsymbol{u_4}\} = \mathbb{R}^4$.

(d) $\text{span}\{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}, \boldsymbol{v_4}\} = \mathbb{R}^4$.

13. Let $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ be vectors in $\mathbb{R}^n$ and let

$$S_1 = \{\boldsymbol{u}, \boldsymbol{v}\}, \quad S_2 = \{\boldsymbol{u}-\boldsymbol{v}, \boldsymbol{v}-\boldsymbol{w}, \boldsymbol{w}-\boldsymbol{u}\}, \quad S_3 = \{\boldsymbol{u}-\boldsymbol{v}, \boldsymbol{v}-\boldsymbol{w}, \boldsymbol{u}+\boldsymbol{w}\},$$
$$S_4 = \{\boldsymbol{u}, \boldsymbol{u}+\boldsymbol{v}, \boldsymbol{u}+\boldsymbol{v}+\boldsymbol{w}\}, \quad S_5 = \{\boldsymbol{u}+\boldsymbol{v}, \boldsymbol{v}+\boldsymbol{w}, \boldsymbol{u}+\boldsymbol{w}, \boldsymbol{u}+\boldsymbol{v}+\boldsymbol{w}\}.$$

Suppose $n = 3$ and $\text{span}\{\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}\} = \mathbb{R}^3$. Determine which of the sets above span $\mathbb{R}^3$.

14. Determine which of the following statements are true. Justify your answer.

(a) If $\boldsymbol{u}$ is a nonzero vector in $\mathbb{R}^1$, then $\text{span}\{\boldsymbol{u}\} = \mathbb{R}^1$.

(b) If $\boldsymbol{u}$, $\boldsymbol{v}$ are nonzero vectors in $\mathbb{R}^2$ such that $\boldsymbol{u} \neq \boldsymbol{v}$, then $\text{span}\{\boldsymbol{u}, \boldsymbol{v}\} = \mathbb{R}^2$.

(c) If S_1 and S_2 are two subsets of $\mathbb{R}^n$, then $\text{span}(S_1 \cap S_2) = \text{span}(S_1) \cap \text{span}(S_2)$.

(d) If S_1 and S_2 are two subsets of $\mathbb{R}^n$, then $\text{span}(S_1 \cup S_2) = \text{span}(S_1) \cup \text{span}(S_2)$.

Question 3.15 to Question 3.31 are exercises for Sections 3.3 and 3.4.

15. Determine which of the following are subspaces of $\mathbb{R}^3$. Justify your answer.

(a) $\{(0,0,0)\}$.

(b) $\{(1,1,1)\}$.

(c) $\{(0,0,0), (1,1,1)\}$.

(d) $\{\,(0,0,c) \mid c \text{ is an integer}\,\}$.

(e) $\{\,(0,0,c) \mid c \text{ is a real number}\,\}$.

(f) $\{\,(1,1,c) \mid c \text{ is a real number}\,\}$.

(g) $\{\,(a,b,c) \mid a,\ b,\ c \text{ are real numbers and } abc = 0\,\}$.

(h) $\{\,(a,b,c) \mid a,\ b,\ c \text{ are real numbers and } a \geq b \geq c\,\}$.

(i) $\{\,(a,b,c) \mid a,\ b \text{ are real numbers and } 4a = 3b\,\}$.

(j) $\{\,(a,b,b) \mid a,\ b \text{ are real numbers}\,\}$.

(k) $\{\,(a,b,ab) \mid a,\ b \text{ are real numbers}\,\}$.

16. Determine which of the following are subspaces of $\mathbb{R}^4$. Justify your answers.

(a) $\{\,(w,x,y,z) \mid w + x = y + z\,\}$.

(b) $\{\,(w,x,y,z) \mid wx = yz\,\}$.

(c) $\{\,(w,x,y,z) \mid w + x + y = z^2\,\}$.

(d) $\{\,(w,x,y,z) \mid w = 0 \ \text{ and } \ y = 0\,\}$.

(e) $\{(w,x,y,z) \mid w=0 \text{ or } y=0\}$.

(f) $\{(w,x,y,z) \mid w=1 \text{ and } y=0\}$.

(g) $\{(w,x,y,z) \mid w+z=0 \text{ and } x+y-4z=0 \text{ and } 4w+y-z=0\}$.

(h) $\{(w,x,y,z) \mid w+z=0 \text{ or } x+y-4z=0 \text{ or } 4w+y-z=0\}$.

17. Give an example of a 2×3 matrix $\boldsymbol{A}$, if possible, such that the solution space of the linear system $\boldsymbol{Ax} = \boldsymbol{0}$ is

(a) $\mathbb{R}^3$.

(b) the plane $\{(x,y,z) \mid 2x+3y-z=0\}$.

(c) the line $\{(t,2t,3t) \mid t \in \mathbb{R}\}$.

(d) the zero subspace.

18. Let W be a subspace of $\mathbb{R}^n$ and let $\boldsymbol{v} \in \mathbb{R}^n$. The set

$$W + \boldsymbol{v} = \{\boldsymbol{u} + \boldsymbol{v} \mid \boldsymbol{u} \in W\}$$

is called a *coset* of W containing $\boldsymbol{v}$. For each of the following, give a geometric interpretation for the coset $W + \boldsymbol{v}$.

(a) $W = \{(x,y) \mid x+y=0\}$ and $\boldsymbol{v} = (1,1)$.

(b) $W = \{c(1,1,1) \mid c \in \mathbb{R}\}$ and $\boldsymbol{v} = (0,0,1)$.

(c) $W = \{(x,y,z) \mid x+y+z=0\}$ and $\boldsymbol{v} = (2,0,-1)$.

19. Let U, V and W be the three planes defined in Question 3.4. Is $U \cap V$ a subspace of $\mathbb{R}^3$? Is $V \cap W$ a subspace of $\mathbb{R}^3$? Justify your answers.

20. Let V and W be subspaces of $\mathbb{R}^n$. Define $V + W = \{\boldsymbol{v} + \boldsymbol{w} \mid \boldsymbol{v} \in V \text{ and } \boldsymbol{w} \in W\}$.

(a) Show that $V + W$ is a subspace of $\mathbb{R}^n$.

(b) Write down the subspace $V + W$ explicitly if

(i) $V = \{(t,0) \mid t \in \mathbb{R}\}$ and $W = \{(0,t) \mid t \in \mathbb{R}\}$.

(ii) $V = \{(t,t,t) \mid t \in \mathbb{R}\}$ and $W = \{(t,-t,0) \mid t \in \mathbb{R}\}$.

21. (All vectors in this question are written as column vectors.) Let $\boldsymbol{A}$ be an $m \times n$ matrix. Define $V_{\boldsymbol{A}}$ to be the subset $\{\boldsymbol{Au} \mid \boldsymbol{u} \in \mathbb{R}^n\}$ of $\mathbb{R}^m$.

(a) Show that $V_{\boldsymbol{A}}$ is a subspace of $\mathbb{R}^m$.

(b) Write down the subspace $V_{\boldsymbol{A}}$ explicitly if

(i) $\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \end{pmatrix}$. (ii) $\boldsymbol{A} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \\ 3 & 1 \end{pmatrix}$.

22. (All vectors in this question are written as column vectors.) Let $\boldsymbol{A}$ be an $n \times n$ matrix. Define $W_{\boldsymbol{A}}$ to be the subset $\{\, \boldsymbol{u} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{u} = \boldsymbol{u} \,\}$ of $\mathbb{R}^n$.

    (a) Show that $W_{\boldsymbol{A}}$ is a subspace of $\mathbb{R}^n$.

    (b) Let $\boldsymbol{A} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$. Write down the subspace $W_{\boldsymbol{A}}$ explicitly.

23. Determine which of the following statements are true. Justify your answer.

    (a) $\mathbb{R}^2$ is a subspace of $\mathbb{R}^3$.

    (b) The solution set of $x + 2y - z = 0$ is a subspace of $\mathbb{R}^3$.

    (c) The solution set of $x + 2y - z = 1$ is a subspace of $\mathbb{R}^3$.

    (d) If S_1 and S_2 are two subsets of $\mathbb{R}^n$, then $\operatorname{span}(S_1 \cup S_2) = \operatorname{span}(S_1) + \operatorname{span}(S_2)$. (See Question 3.20.)

24. Let V and W be subspaces of $\mathbb{R}^n$.

    (a) Show that $V \cap W$ is a subspace of $\mathbb{R}^n$. (Hint: Use Remark 3.3.8.)

    (b) Give an example of V and W in $\mathbb{R}^2$ such that $V \cup W$ is not a subspace.

    (c) Show that $V \cup W$ is a subspace of $\mathbb{R}^n$ if and only if $V \subseteq W$ or $W \subseteq V$.

25. For each of the sets S_1 to S_6 in Question 3.9, determine whether the set is linearly independent

26. (a) Let $\boldsymbol{R} = \begin{pmatrix} 1 & 1 & 1 & 2 & 2 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$.

    Show that the three nonzero rows of $\boldsymbol{R}$ are linearly independent vectors.

    (b) For a nonzero matrix in row-echelon form, is it true that the nonzero rows are always linearly independent?

27. In Question 3.13, suppose $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ are linearly independent vectors in $\mathbb{R}^n$. Determine which of the sets S_1 to S_5 are linearly independent.

28. Let $\boldsymbol{u_1} = (a, 1, -1)$, $\boldsymbol{u_2} = (-1, a, 1)$, $\boldsymbol{u_3} = (1, -1, a)$. For what values of a are $\boldsymbol{u_1}$, $\boldsymbol{u_2}$, $\boldsymbol{u_3}$ linearly independent?

29. Let $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ be vectors in $\mathbb{R}^3$ such that $V = \operatorname{span}\{\boldsymbol{u}, \boldsymbol{v}\}$ and $W = \operatorname{span}\{\boldsymbol{u}, \boldsymbol{w}\}$ are planes in $\mathbb{R}^3$. Find $V \cap W$ if

    (a) $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ are linearly independent.

    (b) $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ are not linearly independent.

30. (All vectors in this question are written as column vectors.) Let $\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}$ be vectors in $\mathbb{R}^n$ and $\boldsymbol{P}$ a square matrix or order n.

    (a) Show that if $\boldsymbol{Pu_1}, \boldsymbol{Pu_2}, \ldots, \boldsymbol{Pu_k}$ are linearly independent, then $\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}$ are linearly independent.

    (b) Suppose $\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}$ are linearly independent.

    (i) Show that if $\boldsymbol{P}$ is invertible, then $\boldsymbol{Pu_1}, \boldsymbol{Pu_2}, \ldots, \boldsymbol{Pu_k}$ are linearly independent.

    (ii) If $\boldsymbol{A}$ is not invertible, are $\boldsymbol{Pu_1}, \boldsymbol{Pu_2}, \ldots, \boldsymbol{Pu_k}$ linearly independent?

31. Prove Remark 3.3.8:

    Let V be a non-empty subset of $\mathbb{R}^n$. Show that V is a subspace of $\mathbb{R}^n$ if and only if for all $\boldsymbol{u}, \boldsymbol{v} \in V$ and $c, d \in \mathbb{R}$, $c\boldsymbol{u} + d\boldsymbol{v} \in V$. (Hint: For the "if" part, you need to find a finite set S of vectors that spans V. By Theorem 3.4.7, there are at most n linearly independent vectors in V. When $V \neq \{\boldsymbol{0}\}$, let S be a largest set of linearly independent vectors in V. Then show that $\operatorname{span}(S) = V$.)

Question 3.32 to Question 3.49 are exercises for Sections 3.5 to 3.7.

32. Determine which of the following sets are bases for $\mathbb{R}^3$.

    (a) $S_1 = \{(1, 0, -1), (-1, 2, 3)\}$.

    (b) $S_2 = \{(1, 0, -1), (-1, 2, 3), (0, 3, 0)\}$.

    (c) $S_3 = \{(1, 0, -1), (-1, 2, 3), (0, 3, 3)\}$.

    (d) $S_4 = \{(1, 0, -1), (-1, 2, 3), (0, 3, 0), (1, -1, 1)\}$.

33. Find a basis for the solution space of each of the following homogeneous systems.

    (a) $x_1 + 3x_2 - x_3 + 2x_4 = 0$.

    (b) $\begin{cases} x_1 + 3x_2 - x_3 + 2x_4 = 0 \\ \quad\;\; -3x_2 + x_3 \qquad\quad = 0. \end{cases}$

    (c) $\begin{cases} x_1 + 3x_2 - x_3 + 2x_4 = 0 \\ \quad\;\; -3x_2 + x_3 \qquad\quad = 0 \\ x_1 \qquad\qquad\quad - x_4 = 0. \end{cases}$

34. For each of the following cases, find the coordinate vector of $\boldsymbol{v}$ relative to the basis $S = \{\boldsymbol{u_1},\ \boldsymbol{u_2},\ \boldsymbol{u_3}\}$.

    (a) $\boldsymbol{v} = (1, -2, 6),\ \boldsymbol{u_1} = (1, 1, 1),\ \boldsymbol{u_2} = (0, 2, 2),\ \boldsymbol{u_3} = (0, 0, 3)$.

    (b) $\boldsymbol{v} = (0, 0, 1),\ \boldsymbol{u_1} = (1, 1, 2),\ \boldsymbol{u_2} = (-1, 1, -2),\ \boldsymbol{u_3} = (1, 3, 3)$.

35. Let $V = \{\, (a+b, a+c, c+d, b+d) \mid a, b, c, d \in \mathbb{R}\,\}$ and $S = \{(1, 1, 0, 0), (1, 0, -1, 0), (0, -1, 0, 1)\}$.

    (a) Show that V is a subspace of $\mathbb{R}^4$ and S is a basis for V.

    (b) Find the coordinate vector of $\boldsymbol{u} = (1, 2, 3, 2)$ relative to S.

    (c) Find a vector $\boldsymbol{v}$ such that $(\boldsymbol{v})_S = (1, 3, -1)$.

36. Find a basis for and determine the dimension of each of the following subspaces of $\mathbb{R}^3$:

    (a) the plane $x - y + z = 0$,

    (b) the plane $x = y$,

    (c) the line $x = t$, $y = -t$ and $z = 2t$ for $t \in \mathbb{R}$.

37. Find a basis for and determine the dimension of each of the following subspaces of $\mathbb{R}^4$:

    (a) the subspace containing all vectors of the form $(w, 0, y, 0)$.

    (b) the subspace containing all vectors of the form (w, x, x, w).

    (c) the subspace containing all vectors of the form (w, x, y, z) with $w = 2x = 3y$.

    (d) the solution space of

$$\begin{cases} 2w + 3x + \ \ y + \ \ z = 0 \\ -3w + \ \ x + 4y - 7z = 0 \\ w + 2x + \ \ y \qquad\ \ = 0. \end{cases}$$

    (e) the subspace $\{\, (w, x, y, z) \mid y = w + x \ \text{ and } \ z = w - x \,\}$.

38. Let $\{\boldsymbol{u_1},\ \boldsymbol{u_2},\ \boldsymbol{u_3}\}$ be a basis for a vector space V. Determine whether $\{\boldsymbol{v_1},\ \boldsymbol{v_2},\ \boldsymbol{v_3}\}$ is a basis for V if

    (a) $\boldsymbol{v_1} = \boldsymbol{u_1},\ \boldsymbol{v_2} = \boldsymbol{u_1} + \boldsymbol{u_2},\ \boldsymbol{v_3} = \boldsymbol{u_1} + \boldsymbol{u_2} + \boldsymbol{u_3}$.

    (b) $\boldsymbol{v_1} = \boldsymbol{u_1} - \boldsymbol{u_2},\ \boldsymbol{v_2} = \boldsymbol{u_2} - \boldsymbol{u_3},\ \boldsymbol{v_3} = \boldsymbol{u_3} - \boldsymbol{u_1}$.

39. Give an example of a family of subspaces V_1, V_2, $\dots$, V_n of $\mathbb{R}^n$ such that $\dim(V_i) = i$ for $i = 1, 2, \dots n$ and
$$V_1 \subseteq V_2 \subseteq \cdots \subseteq V_n.$$
Justify your answer.

40. Let $\boldsymbol{u_1} = (1, 0, 1, 1)$, $\boldsymbol{u_2} = (-3, 3, 7, 1)$, $\boldsymbol{u_3} = (-1, 3, 9, 3)$, $\boldsymbol{u_4} = (-5, 3, 5, -1)$ and let $S = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}, \boldsymbol{u_4}\}$ and $V = \text{span}(S)$.

    (a) Find a non-trivial solution to the equation
    $$a\boldsymbol{u_1} + b\boldsymbol{u_2} + c\boldsymbol{u_3} + d\boldsymbol{u_4} = \boldsymbol{0}.$$

    (b) Express $\boldsymbol{u_3}$ and $\boldsymbol{u_4}$ (separately) as linear combinations of $\boldsymbol{u_1}$ and $\boldsymbol{u_2}$.

    (c) Find a basis for and determine the dimension of V.

    (d) Find a subspace W of $\mathbb{R}^4$ such that $\dim(W) = 3$ and $\dim(W \cap V) = 2$. Justify your answer.

41. Let V be a vector space.

    (a) Suppose S is a finite subset of V such that $\text{span}(S) = V$. Show that there exists a subset S' of S such that S' is a basis for V.

    (b) Suppose T is a finite subset of V such that T is linearly independent. Show that there exists a basis T^* for V such that $T \subseteq T^*$.

42. Let V be a vector space of dimension n. Show that there exists $n + 1$ vectors $\boldsymbol{u_1}$, $\boldsymbol{u_2}$, $\cdots$, $\boldsymbol{u_{n+1}}$ such that every vector in V can be expressed as a linear combination of $\boldsymbol{u_1}$, $\boldsymbol{u_2}$, $\cdots$, $\boldsymbol{u_{n+1}}$ with non-negative coefficients.

43. Let V and W be subspaces of $\mathbb{R}^n$. Show that
$$\dim(V + W) = \dim(V) + \dim(W) - \dim(V \cap W).$$
(See Question 3.20 and Question 3.24.)

44. Let $U = \text{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$ and $V = \text{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ be subspaces of $\mathbb{R}^5$ such that $\dim(U \cap V) = 2$. Suppose W is the smallest subspace of $\mathbb{R}^5$ that contains both U and V. Determine all possible dimensions of W. Justify your answers.

45. Determine which of the following statements are true. Justify your answer.

    (a) If S_1 and S_2 are basis for V and W respectively, where V and W are subspaces of a vector space, then $S_1 \cap S_2$ is a basis for $V \cap W$. (See Question 3.24.)

(b) If S_1 and S_2 are basis for V and W respectively, where V and W are subspaces of a vector space, then $S_1 \cup S_2$ is a basis for $V + W$. (See Question 3.20.)

(c) If V and W are subspaces of a vector space, then there exists a basis S_1 for V and a basis S_2 for W such that $S_1 \cap S_2$ is a basis for $V \cap W$.

(d) If V and W are subspaces of a vector space, then there exists a basis S_1 for V and a basis S_2 for W such that $S_1 \cup S_2$ is a basis for $V + W$.

46. (a) Let $\boldsymbol{u_1} = (1, 2, -1)$, $\boldsymbol{u_2} = (0, 2, 1)$, $\boldsymbol{u_3} = (0, -1, 3)$. Show that $S = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\}$ forms a basis for $\mathbb{R}^3$.

(b) Suppose $\boldsymbol{w} = (1, 1, 1)$. Find the coordinate vector of $\boldsymbol{w}$ relative to S.

(c) Let $T = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}\}$ be another basis for $\mathbb{R}^3$ where $\boldsymbol{v_1} = (1, 5, 4)$, $\boldsymbol{v_2} = (-1, 3, 7)$, $\boldsymbol{v_3} = (2, 2, 4)$. Find the transition matrix from T to S.

(d) Find the transition matrix from S to T.

(e) Use the vector $\boldsymbol{w}$ in Part (b). Find the coordinate vector of $\boldsymbol{w}$ relative to T.

47. Let $S = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\}$ where $\boldsymbol{u_1} = (3, -2, 5)$, $\boldsymbol{u_2} = (1, -4, 4)$, $\boldsymbol{u_3} = (0, 3, -2)$.

(a) Show that S is a basis for $\mathbb{R}^3$.

(b) Show that $T = \{\boldsymbol{u_1} - \boldsymbol{u_2}, \boldsymbol{u_1} + 2\boldsymbol{u_2} - \boldsymbol{u_3}, \boldsymbol{u_2} + 2\boldsymbol{u_3}\}$ is also a basis for $\mathbb{R}^3$.

(c) Find the coordinate vector of $\boldsymbol{v} = (1, 0, 1)$ relative to S.

(d) Find a vector $\boldsymbol{w}$ in $\mathbb{R}^3$ such that $(\boldsymbol{w})_T = (1, 0, 1)$.

(e) Find the transition matrix from T to S and the transition matrix from S to T.

(f) Let $\boldsymbol{x}$ be a vector in $\mathbb{R}^3$ such that $(\boldsymbol{x})_T = (1, 1, 2)$. Find $(\boldsymbol{x})_S$.

48. Consider the vector space $V = \{\, (x, y, z) \mid 2x - y + z = 0 \,\}$. Let $S = \{(0, 1, 1), (1, 2, 0)\}$ and $T = \{(1, 1, -1), (1, 0, -2)\}$.

(a) Show that both S and T are bases for V.

(b) Find the transition matrix from T to S and the transition matrix from S to T.

(c) Find $(\boldsymbol{w})_S$ and $(\boldsymbol{w})_T$ where $\boldsymbol{w} = (1, -1, -3)$.

49. Let $S = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\}$ be a basis for $\mathbb{R}^3$ and $T = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}\}$ where

$$\boldsymbol{v_1} = \boldsymbol{u_1} + \boldsymbol{u_2} + \boldsymbol{u_3}, \quad \boldsymbol{v_2} = \boldsymbol{u_2} + \boldsymbol{u_3} \quad \text{and} \quad \boldsymbol{v_3} = \boldsymbol{u_2} - \boldsymbol{u_3}.$$

(a) Show that T is a basis for $\mathbb{R}^3$.

(b) Find the transition matrix from S to T.

Chapter 4

Vector Spaces Associated with Matrices

Section 4.1 Row Spaces and Column Spaces

Discussion 4.1.1 Each $m \times n$ matrix is naturally associated with three vector spaces, namely the row space, the column space and the nullspace. These three vector spaces provide us with insights into the relationships between solutions of linear systems and the coefficient matrix. In the following, we shall first discuss the row and column spaces of a matrix.

Definition 4.1.2 Let $\boldsymbol{A} = (a_{ij})$ be an $m \times n$ matrix, i.e.

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

The *row space* of $\boldsymbol{A}$ is the subspace of $\mathbb{R}^n$ spanned by the rows of $\boldsymbol{A}$. The *column space* of $\boldsymbol{A}$ is the subspace of $\mathbb{R}^m$ spanned by the columns of $\boldsymbol{A}$. Let $\boldsymbol{r_1}, \boldsymbol{r_2}, \ldots, \boldsymbol{r_m}$ be the m rows of $\boldsymbol{A}$, i.e.

$$\begin{aligned} \boldsymbol{r_1} &= \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \end{pmatrix}, \\ \boldsymbol{r_2} &= \begin{pmatrix} a_{21} & a_{22} & \ldots & a_{2n} \end{pmatrix}, \\ &\vdots \\ \boldsymbol{r_m} &= \begin{pmatrix} a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}, \end{aligned}$$

and let $\boldsymbol{c_1}, \boldsymbol{c_2}, \ldots, \boldsymbol{c_n}$ be the n columns, i.e.

$$\boldsymbol{c_1} = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \quad \boldsymbol{c_2} = \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix}, \quad \ldots, \quad \boldsymbol{c_n} = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}.$$

Then

$$\text{the row space of } \boldsymbol{A} = \operatorname{span}\{\boldsymbol{r_1}, \boldsymbol{r_2}, \ldots, \boldsymbol{r_m}\} \subseteq \mathbb{R}^n$$

and

$$\text{the column space of } \boldsymbol{A} = \operatorname{span}\{\boldsymbol{c_1}, \boldsymbol{c_2}, \ldots, \boldsymbol{c_n}\} \subseteq \mathbb{R}^m.$$

Remark 4.1.3 The row space of $\boldsymbol{A}$ is the same as the column space of $\boldsymbol{A}^{\mathrm{T}}$ while the column space of $\boldsymbol{A}$ is the same as the row space of $\boldsymbol{A}^{\mathrm{T}}$.

Example 4.1.4

1. Consider the matrix

$$\boldsymbol{A} = \begin{pmatrix} 2 & -1 & 0 \\ 1 & -1 & 3 \\ -5 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

The rows of $\boldsymbol{A}$ are

$$\begin{aligned} \boldsymbol{r_1} &= \begin{pmatrix} 2 & -1 & 0 \end{pmatrix}, \\ \boldsymbol{r_2} &= \begin{pmatrix} 1 & -1 & 3 \end{pmatrix}, \\ \boldsymbol{r_3} &= \begin{pmatrix} -5 & 1 & 0 \end{pmatrix}, \\ \boldsymbol{r_4} &= \begin{pmatrix} 1 & 0 & 1 \end{pmatrix} \end{aligned}$$

and

$$\begin{aligned} &\text{the row space of } \boldsymbol{A} \\ &= \operatorname{span}\{\boldsymbol{r_1}, \boldsymbol{r_2}, \boldsymbol{r_3}, \boldsymbol{r_4}\} \\ &= \left\{ a \begin{pmatrix} 2 & -1 & 0 \end{pmatrix} + b \begin{pmatrix} 1 & -1 & 3 \end{pmatrix} + c \begin{pmatrix} -5 & 1 & 0 \end{pmatrix} + d \begin{pmatrix} 1 & 0 & 1 \end{pmatrix} \,\middle|\, a, b, c, d \in \mathbb{R} \right\} \\ &= \left\{ \begin{pmatrix} 2a + b - 5c + d & -a - b + c & 3b + d \end{pmatrix} \,\middle|\, a, b, c, d \in \mathbb{R} \right\} \end{aligned}$$

which is a subspace of $\mathbb{R}^3$.

The columns of $\boldsymbol{A}$ are

$$\boldsymbol{c_1} = \begin{pmatrix} 2 \\ 1 \\ -5 \\ 1 \end{pmatrix}, \quad \boldsymbol{c_2} = \begin{pmatrix} -1 \\ -1 \\ 1 \\ 0 \end{pmatrix}, \quad \boldsymbol{c_3} = \begin{pmatrix} 0 \\ 3 \\ 0 \\ 1 \end{pmatrix}$$

and

$$\begin{aligned}
\text{the column space of } \boldsymbol{A} &= \operatorname{span}\{\boldsymbol{c_1}, \boldsymbol{c_2}, \boldsymbol{c_3}\} \\
&= \left\{ a\begin{pmatrix} 2 \\ 1 \\ -5 \\ 1 \end{pmatrix} + b\begin{pmatrix} -1 \\ -1 \\ 1 \\ 0 \end{pmatrix} + c\begin{pmatrix} 0 \\ 3 \\ 0 \\ 1 \end{pmatrix} \,\middle|\, a, b, c \in \mathbb{R} \right\} \\
&= \left\{ \begin{pmatrix} 2a - b \\ a - b + 3c \\ -5a + b \\ a + c \end{pmatrix} \,\middle|\, a, b, c \in \mathbb{R} \right\}
\end{aligned}$$

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

2. Find a basis for the row space and a basis for the column space of $\boldsymbol{A}$ given in Part 1. Hence state the dimension of the row space and the dimension of the column space of $\boldsymbol{A}$.

   Solution (In Remark 4.1.9 and Remark 4.1.13, we have methods for finding bases for row and column spaces of matrices. For this example, we just do it by brute force.)

   We first consider the row space of $\boldsymbol{A}$. The row space of $\boldsymbol{A}$ is a subspace of $\mathbb{R}^3$ and hence any basis of the row space of $\boldsymbol{A}$ contains at most three vectors. Since the rows $\boldsymbol{r_1}$, $\boldsymbol{r_2}$ and $\boldsymbol{r_3}$ of $\boldsymbol{A}$ are linearly independent (check it), we see that they form a basis for the row space of $\boldsymbol{A}$.

   All three columns $\boldsymbol{c_1}$, $\boldsymbol{c_2}$ and $\boldsymbol{c_3}$ of $\boldsymbol{A}$ are linearly independent (check it) and so form a basis for the column space of $\boldsymbol{A}$.

   In this example, both the row space and the column space of $\boldsymbol{A}$ have the same dimension, which is 3. (In general, both the row space and the column space of a matrix always have the same dimension. See Theorem 4.2.1.)

Notation 4.1.5 As mentioned in Notation 3.1.5, a vector in $\mathbb{R}^n$ can be identified as a $1 \times n$ matrix or an $n \times 1$ matrix. From now on, to make the context clear, when a vector in $\mathbb{R}^n$ is written as $(u_1, u_2, \ldots, u_n)$, it is a row vector and is identified with the $1 \times n$ matrix $\begin{pmatrix} u_1 & u_2 & \cdots & u_n \end{pmatrix}$; and if it is written as $(u_1, u_2, \ldots, u_n)^{\mathrm{T}}$, it is a column vector and is identified with the $n \times 1$ matrix $\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$.

For example, in Example 4.1.4.1, we may write the rows of $\boldsymbol{A}$ as $\boldsymbol{r_1} = (2, -1, 0)$, $\boldsymbol{r_2} = (1, -1, 3)$, $\boldsymbol{r_3} = (-5, 1, 0)$, $\boldsymbol{r_4} = (1, 0, 1)$ and hence

$$\text{the row space of } \boldsymbol{A} = \{\, (2a + b - 5c + d,\ -a - b + c,\ 3b + d) \mid a, b, c, d \in \mathbb{R} \,\}.$$

Also, we may write the columns of $\boldsymbol{A}$ as $\boldsymbol{c_1} = (2, 1, -5, 1)^{\mathrm{T}}$, $\boldsymbol{c_2} = (-1, -1, 1, 0)^{\mathrm{T}}$, $\boldsymbol{c_3} = (0, 3, 0, 1)^{\mathrm{T}}$ and hence

$$\text{the column space of } \boldsymbol{A} = \{\, (2a - b,\ a - b + 3c,\ -5a + b,\ a + c)^{\mathrm{T}} \mid a, b, c \in \mathbb{R} \,\}.$$

Discussion 4.1.6 In Definition 1.2.4 (and Definition 2.4.1), we say that two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are *row equivalent* if one can be obtained from the other by a series of elementary row operations. Using the concept of row equivalent matrices, we shall develop methods to find bases for row spaces and column spaces. First, let us note some observations:

1. The notion of row equivalence is an equivalence relation on matrices of the same size (check it).

2. A matrix is row equivalent to its row-echelon forms. In particular, if two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ have a same (reduced) row-echelon form, then $\boldsymbol{A}$ and $\boldsymbol{B}$ are row equivalent.

Theorem 4.1.7 Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be row equivalent matrices. Then the row space of $\boldsymbol{A}$ and the row space of $\boldsymbol{B}$ are identical, i.e. elementary row operations preserve the row space of a matrix.

Proof Let $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{r_1} \\ \boldsymbol{r_2} \\ \vdots \\ \boldsymbol{r_m} \end{pmatrix}$ where $\boldsymbol{r_i}$ is the ith row of $\boldsymbol{A}$. We can make use of Theorem 3.2.10 to show that the three types of elementary row operations preserve the row space of $\boldsymbol{A}$. In the following, we show the proof for the elementary row operation of the first type.

Suppose $\boldsymbol{B} = \begin{pmatrix} \boldsymbol{r_1} \\ \vdots \\ \boldsymbol{r_{i-1}} \\ k\boldsymbol{r_i} \\ \boldsymbol{r_{i+1}} \\ \vdots \\ \boldsymbol{r_m} \end{pmatrix}$ which is obtained from $\boldsymbol{A}$ by multiplying k to the ith row.

Since $k\boldsymbol{r_i} \in \operatorname{span}\{\boldsymbol{r_1}, \boldsymbol{r_2}, \ldots, \boldsymbol{r_m}\}$, by Theorem 3.2.10,

$$\operatorname{span}\{\boldsymbol{r_1}, \ldots, \boldsymbol{r_{i-1}}, k\boldsymbol{r_i}, \boldsymbol{r_{i+1}}, \ldots, \boldsymbol{r_m}\} \subseteq \operatorname{span}\{\boldsymbol{r_1}, \boldsymbol{r_2}, \ldots, \boldsymbol{r_m}\}.$$

On the other hand, $r_i = \frac{1}{k}(kr_i) \in \text{span}\{r_1, \ldots, r_{i-1}, kr_i, r_{i+1}, \ldots, r_m\}$ and hence by Theorem 3.2.10,

$$\text{span}\{r_1, r_2, \ldots, r_m\} \subseteq \text{span}\{r_1, \ldots, r_{i-1}, kr_i, r_{i+1}, \ldots, r_m\}.$$

Thus the row space of $\boldsymbol{A}$ and the row space of $\boldsymbol{B}$ are the same.

Proofs for the other two elementary row operations are similar.

Example 4.1.8

1. Let $\boldsymbol{A} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 2 & 4 \\ \frac{1}{2} & 1 & 2 \end{pmatrix}$, $\boldsymbol{B} = \begin{pmatrix} \frac{1}{2} & 1 & 2 \\ 0 & 2 & 4 \\ 0 & 0 & 1 \end{pmatrix}$, $\boldsymbol{C} = \begin{pmatrix} 1 & 2 & 4 \\ 0 & 2 & 4 \\ 0 & 0 & 1 \end{pmatrix}$, $\boldsymbol{D} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 4 \\ 0 & 0 & 1 \end{pmatrix}$. Note that

$$\boldsymbol{A} \xrightarrow{R_1 \leftrightarrow R_3} \boldsymbol{B} \xrightarrow{2R_1} \boldsymbol{C} \xrightarrow{R_1 - R_2} \boldsymbol{D}.$$

The four matrices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$ are row equivalent to one another and hence they all have the same row space. In particular,

$$\text{span}\{(0,0,1),\ (0,2,4),\ (\tfrac{1}{2},1,2)\} = \text{span}\{(1,0,0),\ (0,2,4),\ (0,0,1)\}.$$

2. Let

$$\boldsymbol{A} = \begin{pmatrix} 2 & 2 & -1 & 0 & 1 \\ -1 & -1 & 2 & -3 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & -2 & 0 & -1 \end{pmatrix}.$$

The following is a row-echelon form of $\boldsymbol{A}$ (check it):

$$\boldsymbol{R} = \begin{pmatrix} 2 & 2 & -1 & 0 & 1 \\ 0 & 0 & \frac{3}{2} & -3 & \frac{3}{2} \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

So the row space of $\boldsymbol{A}$ can be given by

$$\text{span}\{(2,2,-1,0,1),\ (0,0,\tfrac{3}{2},-3,\tfrac{3}{2}),\ (0,0,0,3,0)\}.$$

It is easy to check that the three vectors are linearly independent and hence

$$\{(2,2,-1,0,1),\ (0,0,\tfrac{3}{2},-3,\tfrac{3}{2}),\ (0,0,0,3,0)\}$$

is a basis for the row space of $\boldsymbol{A}$ (see Remark 4.1.9).

Remark 4.1.9 Let $\boldsymbol{A}$ be a matrix and $\boldsymbol{R}$ a row-echelon form of $\boldsymbol{A}$. Then the set of nonzero rows in $\boldsymbol{R}$ is a basis for the row space of $\boldsymbol{A}$. (See Question 3.26)

Discussion 4.1.10 Since the column space of a matrix $\boldsymbol{A}$ is the row space of $\boldsymbol{A}^{\mathrm{T}}$, a basis for the column space of $\boldsymbol{A}$ can be obtained from a row-echelon form of $\boldsymbol{A}^{\mathrm{T}}$ as discussed in Remark 4.1.9. However, we shall study another method for finding a basis for the column space of $\boldsymbol{A}$ based on the information obtained from a row-echelon form of $\boldsymbol{A}$. Before we discuss the method, we must be aware that elementary row operations may not preserve the column space of a matrix, i.e. the column spaces of row equivalent matrices may not be the same. For example the following matrices

$$\boldsymbol{A} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \boldsymbol{B} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

are row equivalent but their column spaces are different.

Theorem 4.1.11 Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be row equivalent matrices. Then the following statements hold:

1. A given set of columns of $\boldsymbol{A}$ is linearly independent if and only if the set of corresponding columns of $\boldsymbol{B}$ is linearly independent.
2. A given set of columns of $\boldsymbol{A}$ forms a basis for the column space of $\boldsymbol{A}$ if and only if the set of corresponding columns of $\boldsymbol{B}$ forms a basis for the column space of $\boldsymbol{B}$.

(The proof of the theorem is left as an exercise. See Question 4.18.)

Example 4.1.12 (Example 4.1.8.2 revisited) Let

$$\boldsymbol{A} = \begin{pmatrix} 2 & 2 & -1 & 0 & 1 \\ -1 & -1 & 2 & -3 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & -2 & 0 & -1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{R} = \begin{pmatrix} 2 & 2 & -1 & 0 & 1 \\ 0 & 0 & \frac{3}{2} & -3 & \frac{3}{2} \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

where $\boldsymbol{R}$ is a row-echelon form of $\boldsymbol{A}$.

1. The first, third and fifth columns of $\boldsymbol{R}$ are linearly dependent (check it). Thus by Theorem 4.1.11.1, the first, third and fifth columns of $\boldsymbol{A}$ are linearly dependent.
2. The first, third and fourth columns of $\boldsymbol{R}$ are linearly independent. Thus by Theorem 4.1.11.1, the first, third and fourth columns of $\boldsymbol{A}$ are linearly independent. Furthermore, the first, third and fourth columns of $\boldsymbol{R}$ form a basis for the column space of $\boldsymbol{R}$. So by Theorem 4.1.11.2, the first, third and fourth columns of $\boldsymbol{A}$ form a basis for the column space of $\boldsymbol{A}$, i.e.

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

is a basis for the column space of $\boldsymbol{A}$. (See Remark 4.1.13.)

Remark 4.1.13 Let $\boldsymbol{A}$ be a matrix and $\boldsymbol{R}$ a row-echelon form of $\boldsymbol{A}$. A basis for the column space of $\boldsymbol{A}$ can be obtained by taking the columns of $\boldsymbol{A}$ that correspond to the pivot columns in $\boldsymbol{R}$.

Example 4.1.14 We use the notion of row spaces and column spaces of matrices to obtain bases for subspaces of $\mathbb{R}^n$.

1. Consider vectors
$$\begin{aligned} \boldsymbol{u_1} &= (1,2,2,1), & \boldsymbol{u_2} &= (3,6,6,3), \\ \boldsymbol{u_3} &= (4,9,9,5), & \boldsymbol{u_4} &= (-2,-1,-1,1), \\ \boldsymbol{u_5} &= (5,8,9,4), & \boldsymbol{u_6} &= (4,2,7,3). \end{aligned}$$

   Find a basis for the subspace of $\mathbb{R}^4$ spanned by $\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}, \boldsymbol{u_4}, \boldsymbol{u_5}, \boldsymbol{u_6}$.

   Solution

   Method 1: We place the vectors $\boldsymbol{u_1}$, $\boldsymbol{u_2}$, $\boldsymbol{u_3}$, $\boldsymbol{u_4}$, $\boldsymbol{u_5}$, $\boldsymbol{u_6}$ in the form of rows in a 6×4 matrix as shown.
$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 2 & 1 \\ 3 & 6 & 6 & 3 \\ 4 & 9 & 9 & 5 \\ -2 & -1 & -1 & 1 \\ 5 & 8 & 9 & 4 \\ 4 & 2 & 7 & 3 \end{pmatrix}.$$

   We perform row operations to obtain a row-echelon form of $\boldsymbol{A}$:
$$\begin{pmatrix} 1 & 2 & 2 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

   Thus by Remark 4.1.9, $\{(1,2,2,1), (0,1,1,1), (0,0,1,1)\}$ is a basis for the subspace spanned by $\boldsymbol{u_1}$, $\boldsymbol{u_2}$, $\boldsymbol{u_3}$, $\boldsymbol{u_4}$, $\boldsymbol{u_5}$, $\boldsymbol{u_6}$.

   Method 2: We place the vectors $\boldsymbol{u_1}$, $\boldsymbol{u_2}$, $\boldsymbol{u_3}$, $\boldsymbol{u_4}$, $\boldsymbol{u_5}$, $\boldsymbol{u_6}$ in the form of columns in a 4×6 matrix as shown.
$$\boldsymbol{B} = \begin{pmatrix} 1 & 3 & 4 & -2 & 5 & 4 \\ 2 & 6 & 9 & -1 & 8 & 2 \\ 2 & 6 & 9 & -1 & 9 & 7 \\ 1 & 3 & 5 & 1 & 4 & 3 \end{pmatrix}.$$

   We perform row operations to obtain a row-echelon form of $\boldsymbol{B}$:

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

We see that the first, third and fifth columns are pivot columns. Therefore by Remark 4.1.13, $\{\boldsymbol{u}_1, \boldsymbol{u}_3, \boldsymbol{u}_5\}$ is a basis for the subspace spanned by $\boldsymbol{u}_1$, $\boldsymbol{u}_2$, $\boldsymbol{u}_3$, $\boldsymbol{u}_4$, $\boldsymbol{u}_5$, $\boldsymbol{u}_6$.

2. Let $S = \{(1, 4, -2, 5, 1), (2, 9, -1, 8, 2), (2, 9, -1, 9, 3)\}$. Note that S is linearly independent (check it). Extend S to a basis for $\mathbb{R}^5$.

   Solution The following is a procedure for solving problems of this type.

   Step 1: Form a matrix $\boldsymbol{A}$ using the vectors in S ($\subseteq \mathbb{R}^n$) as rows:

$$\boldsymbol{A} = \begin{pmatrix} 1 & 4 & -2 & 5 & 1 \\ 2 & 9 & -1 & 8 & 2 \\ 2 & 9 & -1 & 9 & 3 \end{pmatrix}.$$

   Step 2: Reduce $\boldsymbol{A}$ to a row-echelon form $\boldsymbol{R}$:

$$\boldsymbol{R} = \begin{pmatrix} 1 & 4 & -2 & 5 & 1 \\ 0 & 1 & 3 & -2 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}.$$

   Step 3: Identify the non-pivot columns in $\boldsymbol{R}$:
   In this example, the third and fifth columns are non-pivot columns.

   Step 4: For each non-pivot column identified in Step 3, get a vector such that the leading entry of the vector is at that column:
   In this example, we get one vector of the form $(0, 0, x, *, *)$ and one of the form $(0, 0, 0, 0, y)$ where x, y are any nonzero real numbers. In particular, we choose $(0, 0, 1, 0, 0)$ and $(0, 0, 0, 0, 1)$.

   Step 5 Now, $S \cup$ (the set of vectors obtained in Step 4) is a basis for $\mathbb{R}^n$:
   In this example,

$$\{(1, 4, -2, 5, 1), (2, 9, -1, 8, 2), (2, 9, -1, 9, 3), (0, 0, 1, 0, 0), (0, 0, 0, 0, 1)\}$$

   is a basis for $\mathbb{R}^5$.

Discussion 4.1.15 (System of linear equations revisited) Consider the system of equations

$$\begin{cases} 2x - y \phantom{{}+3z} = -1 \\ x - y + 3z = 4 \\ -5x + y \phantom{{}+3z} = -2 \\ x \phantom{{}-y} + z = 3. \end{cases}$$

We examine the system by writing it in the form $\boldsymbol{Ax} = \boldsymbol{b}$ as follows

$$\begin{pmatrix} 2 & -1 & 0 \\ 1 & -1 & 3 \\ -5 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -1 \\ 4 \\ -2 \\ 3 \end{pmatrix}.$$

Following Remark 2.2.17, we rewrite the linear system as

$$x \begin{pmatrix} 2 \\ 1 \\ -5 \\ 1 \end{pmatrix} + y \begin{pmatrix} -1 \\ -1 \\ 1 \\ 0 \end{pmatrix} + z \begin{pmatrix} 0 \\ 3 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 4 \\ -2 \\ 3 \end{pmatrix}.$$

We see that a solution to the system is a way of writing the vector $\boldsymbol{b}$ as a linear combination of the columns of $\boldsymbol{A}$. For example, $x = 1$, $y = 3$ and $z = 2$ is a solution to the system. Hence $(-1, 4, -2, 3)^{\mathrm{T}}$ can be written as the linear combination

$$1 \begin{pmatrix} 2 \\ 1 \\ -5 \\ 1 \end{pmatrix} + 3 \begin{pmatrix} -1 \\ -1 \\ 1 \\ 0 \end{pmatrix} + 2 \begin{pmatrix} 0 \\ 3 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 4 \\ -2 \\ 3 \end{pmatrix}.$$

The discussion here leads to the following theorem.

Theorem 4.1.16 Let A be an $m \times n$ matrix. Then

$$\text{the column space of } \boldsymbol{A} = \{\, \boldsymbol{Au} \mid \boldsymbol{u} \in \mathbb{R}^n \,\}.$$

Hence a system of linear equations $\boldsymbol{Ax} = \boldsymbol{b}$ is consistent if and only if $\boldsymbol{b}$ lies in the column space of $\boldsymbol{A}$.

Proof Write $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{c_1} & \boldsymbol{c_2} & \cdots & \boldsymbol{c_n} \end{pmatrix}$ where $\boldsymbol{c_j}$ is the jth column of $\boldsymbol{A}$.

For any $\boldsymbol{u} = (u_1, u_2, \ldots, u_n)^{\mathrm{T}} \in \mathbb{R}^n$,

$$\begin{aligned} \boldsymbol{Au} &= \begin{pmatrix} \boldsymbol{c_1} & \boldsymbol{c_2} & \cdots & \boldsymbol{c_n} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \\ &= u_1\boldsymbol{c_1} + u_2\boldsymbol{c_2} + \cdots + u_n\boldsymbol{c_n} \\ &\in \operatorname{span}\{\boldsymbol{c_1}, \boldsymbol{c_2}, \ldots, \boldsymbol{c_n}\} = \text{the column space of } \boldsymbol{A}. \end{aligned}$$

Thus $\{\, \boldsymbol{Au} \mid \boldsymbol{u} \in \mathbb{R}^n \,\} \subseteq$ the column space of $\boldsymbol{A}$.

Conversely, suppose $\boldsymbol{b}$ is in the column space of $\boldsymbol{A}$, i.e. $\boldsymbol{b} \in \text{span}\{\boldsymbol{c_1}, \boldsymbol{c_2}, \ldots, \boldsymbol{c_n}\}$. Hence there exists $u_1, u_2, \ldots, u_n \in \mathbb{R}$ such that

$$\boldsymbol{b} = u_1\boldsymbol{c_1} + u_2\boldsymbol{c_2} + \cdots + u_n\boldsymbol{c_n} = \boldsymbol{A}\boldsymbol{u},$$

where $\boldsymbol{u} = (u_1,\, u_2,\, \ldots,\, u_n)^{\mathrm{T}}$. Thus the column space of $\boldsymbol{A} \subseteq \{\, \boldsymbol{A}\boldsymbol{u} \mid \boldsymbol{u} \in \mathbb{R}^n \,\}$.

So we have shown that the column space of $\boldsymbol{A} = \{\, \boldsymbol{A}\boldsymbol{u} \mid \boldsymbol{u} \in \mathbb{R}^n \,\}$.

Finally, a system of linear equations $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ is consistent if and only if there exists $\boldsymbol{u} \in \mathbb{R}^n$ such that $\boldsymbol{A}\boldsymbol{u} = \boldsymbol{b}$, i.e. $\boldsymbol{b} \in \{\, \boldsymbol{A}\boldsymbol{u} \mid \boldsymbol{u} \in \mathbb{R}^n \,\} =$ the column space of $\boldsymbol{A}$.

Section 4.2 Ranks

Theorem 4.2.1 The row space and column space of a matrix have the same dimension.

Proof Let $\boldsymbol{A}$ be a matrix and $\boldsymbol{R}$ a row-echelon form of $\boldsymbol{A}$. Since the row space of $\boldsymbol{A}$ coincides with that of $\boldsymbol{R}$ by Remark 4.1.9, we see that the dimension of the row space of $\boldsymbol{A}$ is the number of nonzero rows in $\boldsymbol{R}$ which is equal to the number of pivot columns in $\boldsymbol{R}$. On the other hand, by Remark 4.1.13, the columns of $\boldsymbol{A}$ that correspond to the pivot columns in $\boldsymbol{R}$ form a basis for the column space of $\boldsymbol{A}$. It follows that the dimension of the column space of $\boldsymbol{A}$ is also equal to the number of pivot columns in $\boldsymbol{R}$. This completes the proof.

Example 4.2.2 Let

$$\boldsymbol{C} = \begin{pmatrix} 2 & 0 & 3 & -1 & 8 \\ 2 & 1 & 1 & -2 & 5 \\ -4 & -3 & 0 & 5 & -7 \end{pmatrix}.$$

By using Gaussian Elimination, we get a row-echelon form of $\boldsymbol{C}$:

$$\begin{pmatrix} 2 & 0 & 3 & -1 & 8 \\ 0 & 1 & -2 & -1 & -3 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

By Remark 4.1.9, $\{(2, 0, 3, -1, 8), (0, 1, -2, -1, -3)\}$ is a basis for the row space of $\boldsymbol{C}$ and hence the dimension of the row space of $\boldsymbol{C}$ is 2.

By Remark 4.1.13, $\{(2, 2, -4)^{\mathrm{T}}, (0, 1, -3)^{\mathrm{T}}\}$ is a basis for the column space of $\boldsymbol{C}$ and hence the dimension of the column space of $\boldsymbol{C}$ is also 2.

Definition 4.2.3 The *rank* of a matrix is the dimension of its row space (or column space). We denote the rank of a matrix $\boldsymbol{A}$ by $\text{rank}(\boldsymbol{A})$. Note that $\text{rank}(\boldsymbol{A})$ is equal to the number of nonzero rows as well as the number of pivot columns in a row-echelon form of $\boldsymbol{A}$.

Example 4.2.4

1. $\text{rank}(\mathbf{0}) = 0$ and $\text{rank}(\boldsymbol{I_n}) = n$.
2. In Example 4.2.2, $\text{rank}(\boldsymbol{C}) = 2$.
3. In Example 4.1.14.1, $\text{rank}(\boldsymbol{A}) = \text{rank}(\boldsymbol{B}) = 3$. Note that $\boldsymbol{B}^{\mathrm{T}} = \boldsymbol{A}$.
4. What is the largest possible rank of a 5×3 matrix?

   Solution What we need to do is to find out the largest possible number of pivot columns in a row-echelon form of a 5×3 matrix. The answer is 3.

Remark 4.2.5

1. For an $m \times n$ matrix $\boldsymbol{A}$, $\text{rank}(\boldsymbol{A}) \le \min\{m, n\}$. If $\text{rank}(\boldsymbol{A}) = \min\{m, n\}$, $\boldsymbol{A}$ is said to have *full rank.*
2. A square matrix $\boldsymbol{A}$ is of full rank if and only if $\det(\boldsymbol{A}) \neq 0$.
3. $\text{rank}(\boldsymbol{A}) = \text{rank}(\boldsymbol{A}^{\mathrm{T}})$ for any matrix $\boldsymbol{A}$ because the row space of $\boldsymbol{A}$ is the column space of $\boldsymbol{A}^{\mathrm{T}}$.

Remark 4.2.6 (System of linear equations revisited) A linear system $\boldsymbol{Ax} = \boldsymbol{b}$ is consistent if and only if $\boldsymbol{A}$ and the augmented matrix $(\boldsymbol{A} \mid \boldsymbol{b})$ have the same rank.

Example 4.2.7 Consider the system of linear equations

$$\begin{cases} 2x - y \phantom{{}+3z} = 1 \\ x - y + 3z = 0 \\ -5x + y \phantom{{}+3z} = 0 \\ x \phantom{{}-y} + z = 0. \end{cases}$$

Writing it in a matrix form we have

$$\begin{pmatrix} 2 & -1 & 0 \\ 1 & -1 & 3 \\ -5 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

The augmented matrix $(\boldsymbol{A} \mid \boldsymbol{b})$ has a row-echelon form

$$\left(\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & 5 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right).$$

We see that the system is inconsistent. Note from the row-echelon form that the rank of $\boldsymbol{A}$ is 3 while the rank of $(\boldsymbol{A} \mid \boldsymbol{b})$ is 4.

Theorem 4.2.8 Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be $m \times n$ and $n \times p$ matrices respectively. Then

$$\operatorname{rank}(\boldsymbol{AB}) \leq \min\{\operatorname{rank}(\boldsymbol{A}),\ \operatorname{rank}(\boldsymbol{B})\}.$$

Proof Let $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 & \cdots & \boldsymbol{b}_n \end{pmatrix}$ where $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$ are ith columns of $\boldsymbol{A}$ and $\boldsymbol{B}$ respectively. Following Notation 2.2.15,

$$\boldsymbol{AB} = \begin{pmatrix} \boldsymbol{Ab}_1 & \boldsymbol{Ab}_2 & \cdots & \boldsymbol{Ab}_n \end{pmatrix}$$

where $\boldsymbol{Ab}_i$ is the ith column of $\boldsymbol{AB}$. By Theorem 4.1.16,

$$\boldsymbol{Ab}_i \in \text{the column space of } \boldsymbol{A} = \operatorname{span}\{\boldsymbol{a}_1,\ \boldsymbol{a}_2,\ \ldots,\ \boldsymbol{a}_n\}$$

and hence by Theorem 3.2.10,

$$\begin{aligned} \text{the column space of } \boldsymbol{AB} &= \operatorname{span}\{\boldsymbol{Ab}_1,\ \boldsymbol{Ab}_2,\ \ldots,\ \boldsymbol{Ab}_p\} \\ &\subseteq \operatorname{span}\{\boldsymbol{a}_1,\ \boldsymbol{a}_2,\ \ldots,\ \boldsymbol{a}_n\} = \text{the column space of } \boldsymbol{A}. \end{aligned}$$

So

$$\begin{aligned} \operatorname{rank}(\boldsymbol{AB}) &= \dim(\text{the column space of } \boldsymbol{AB}) \\ &\leq \dim(\text{the column space of } \boldsymbol{A}) = \operatorname{rank}(\boldsymbol{A}). \end{aligned} \tag{4.1}$$

On the other hand, by applying Equation 4.1 to $\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}$, we have $\operatorname{rank}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}) \leq \operatorname{rank}(\boldsymbol{B}^{\mathrm{T}})$ and hence

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

Thus we have shown $\operatorname{rank}(\boldsymbol{AB}) \leq \min\{\operatorname{rank}(\boldsymbol{A}),\ \operatorname{rank}(\boldsymbol{B})\}$.

Section 4.3 Nullspaces and Nullities

Definition 4.3.1 Let $\boldsymbol{A}$ be an $m \times n$ matrix. The solution space of the homogeneous system of linear equations $\boldsymbol{Ax} = \boldsymbol{0}$ is known as the *nullspace* of $\boldsymbol{A}$.

The dimension of the nullspace of a matrix $\boldsymbol{A}$ is called the *nullity* of $\boldsymbol{A}$ and is denoted by $\operatorname{nullity}(\boldsymbol{A})$. If $\boldsymbol{A}$ is an $m \times n$ matrix, it is clear that $\operatorname{nullity}(\boldsymbol{A}) \leq n$ since the nullspace is a subspace of $\mathbb{R}^n$.

Notation 4.3.2 From now on, vectors in nullspaces, as well as solution sets of linear systems, will always be written as column vectors, see Notation 4.1.5.

Example 4.3.3

1. (Example 4.1.8.2 revisited) Find a basis for the nullspace of the matrix

$$\boldsymbol{A} = \begin{pmatrix} 2 & 2 & -1 & 0 & 1 \\ -1 & -1 & 2 & -3 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & -2 & 0 & -1 \end{pmatrix}.$$

Solution (See Discussion 3.6.5.) The reduced row-echelon form of the augmented matrix $(\boldsymbol{A} \mid \boldsymbol{0})$ of the homogeneous system of linear equations $\boldsymbol{Ax} = \boldsymbol{0}$ is given by

$$\left(\begin{array}{ccccc|c} 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

The homogeneous system $\boldsymbol{Ax} = \boldsymbol{0}$ has a general solution

$$\boldsymbol{x} = s\begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + t\begin{pmatrix} -1 \\ 0 \\ -1 \\ 0 \\ 1 \end{pmatrix} \quad \text{where } s, t \text{ are arbitrary parameters.}$$

Thus a basis for the nullspace of $\boldsymbol{A}$ is given by $\{(-1, 1, 0, 0, 0)^{\mathrm{T}}, (-1, 0, -1, 0, 1)^{\mathrm{T}}\}$. Here $\text{nullity}(\boldsymbol{A}) = 2$.

2. Let $\boldsymbol{B} = \begin{pmatrix} 2 & 1 & -5 & 1 \\ -1 & -1 & 1 & 0 \\ 0 & 3 & 0 & 1 \end{pmatrix}$. Determine the nullity of $\boldsymbol{B}$.

Solution We consider the homogeneous system $\boldsymbol{Bx} = \boldsymbol{0}$. The reduced row-echelon form of the augmented matrix $(\boldsymbol{B} \mid \boldsymbol{0})$ is

$$\left(\begin{array}{cccc|c} 1 & 0 & 0 & -\frac{7}{9} & 0 \\ 0 & 1 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 1 & -\frac{4}{9} & 0 \end{array}\right).$$

The system has a general solution

$$\boldsymbol{x} = t\begin{pmatrix} \frac{7}{9} \\ -\frac{1}{3} \\ \frac{4}{9} \\ 1 \end{pmatrix} \quad \text{where } t \text{ is an arbitrary parameter.}$$

Thus the nullspace of $\boldsymbol{B}$ is spanned by the vector $(7, -3, 4, 9)^{\mathrm{T}}$ and $\text{nullity}(\boldsymbol{B}) = 1$.

Theorem 4.3.4 (Dimension Theorem for Matrices) Let $\boldsymbol{A}$ be a matrix with n columns. Then

$$\operatorname{rank}(\boldsymbol{A}) + \operatorname{nullity}(\boldsymbol{A}) = n.$$

Proof Let $\boldsymbol{R}$ be a row-echelon form of $\boldsymbol{A}$. The columns of $\boldsymbol{R}$ can be classified into two types: pivot columns and non-pivot columns. Since the rank of $\boldsymbol{A}$ is the number of pivot columns in $\boldsymbol{R}$ and the nullity of $\boldsymbol{A}$ is the number of non-pivot columns in $\boldsymbol{R}$ (see Discussion 3.6.5), the theorem follows.

Example 4.3.5

1. In Example 4.3.3.2, nullity($\boldsymbol{B}$) = 1. By Theorem 4.3.4,

$$\operatorname{rank}(\boldsymbol{B}) = \text{the number of columns in } \boldsymbol{B} - \operatorname{nullity}(\boldsymbol{B}) = 4 - 1 = 3.$$

2. In each of the following cases, use Theorem 4.3.4 to find rank($\boldsymbol{A}$), nullity($\boldsymbol{A}$) and nullity($\boldsymbol{A}^{\mathrm{T}}$).

   (a) $\boldsymbol{A}$ is a 3×4 matrix and $\operatorname{rank}(\boldsymbol{A}) = 3$.

   (b) $\boldsymbol{A}$ is a 7×5 matrix and $\operatorname{nullity}(\boldsymbol{A}) = 3$.

   (c) $\boldsymbol{A}$ is a 3×2 matrix and $\operatorname{nullity}(\boldsymbol{A}^{\mathrm{T}}) = 3$.

   Answers The values of rank($\boldsymbol{A}$), nullity($\boldsymbol{A}$) and nullity($\boldsymbol{A}^{\mathrm{T}}$) are respectively

   (a) 3, 1, 0; (b) 2, 3, 5; (c) 0, 2, 3 (a matrix of rank 0 is a zero matrix).

Theorem 4.3.6 (System of linear equations revisited) Suppose the system of linear equations $\boldsymbol{Ax} = \boldsymbol{b}$ has a solution $\boldsymbol{v}$. Then the solution set of the system is given by

$$M = \{\, \boldsymbol{u} + \boldsymbol{v} \mid \boldsymbol{u} \text{ is an element of the nullspace of } \boldsymbol{A} \,\}.$$

That is, $\boldsymbol{Ax} = \boldsymbol{b}$ has a general solution

$$\boldsymbol{x} = (\text{a general solution for } \boldsymbol{Ax} = \boldsymbol{0}) + (\text{one particular solution to } \boldsymbol{Ax} = \boldsymbol{b}).$$

(Using the terminology defined in Question 3.18, the solution set M is a coset of the nullspace of $\boldsymbol{A}$.)

Proof Since $\boldsymbol{v}$ is a solution to the system, we have $\boldsymbol{Av} = \boldsymbol{b}$.

Let $\boldsymbol{w}$ be any solution to the system, i.e. $\boldsymbol{Aw} = \boldsymbol{b}$. Let $\boldsymbol{u} = \boldsymbol{w} - \boldsymbol{v}$. Then

$$\boldsymbol{Au} = \boldsymbol{A}(\boldsymbol{w} - \boldsymbol{v}) = \boldsymbol{Aw} - \boldsymbol{Av} = \boldsymbol{b} - \boldsymbol{b} = \boldsymbol{0}.$$

Hence $\boldsymbol{u}$ is contained in the nullspace of $\boldsymbol{A}$. Since $\boldsymbol{w} = \boldsymbol{u} + \boldsymbol{v}$, $\boldsymbol{w} \in M$. So we have shown that the solution set of the system is a subset of M.

On the other hand, take any element $\boldsymbol{w}$ of M, i.e. $\boldsymbol{w} = \boldsymbol{u} + \boldsymbol{v}$ where $\boldsymbol{u}$ is an element of the nullspace of $\boldsymbol{A}$. Then

$$\boldsymbol{A}\boldsymbol{w} = \boldsymbol{A}(\boldsymbol{u} + \boldsymbol{v}) = \boldsymbol{A}\boldsymbol{u} + \boldsymbol{A}\boldsymbol{v} = \boldsymbol{0} + \boldsymbol{b} = \boldsymbol{b}.$$

Thus $\boldsymbol{w}$ is a solution to the system. So we have shown that M is a subset of the solution set of the system. This completes the proof.

Remark 4.3.7 By Theorem 4.3.6, a linear system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ has only one solution if and only if the nullspace of $\boldsymbol{A}$ is equal to $\{\boldsymbol{0}\}$.

Example 4.3.8 Consider the linear system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ where

$$\boldsymbol{A} = \begin{pmatrix} 2 & 2 & -1 & 0 & 1 \\ -1 & -1 & 2 & -3 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & -2 & 0 & -1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{pmatrix} 0 \\ 0 \\ 3 \\ -3 \end{pmatrix}.$$

By Example 4.3.3.1, we learn that

$$\text{the nullspace of } \boldsymbol{A} = \left\{ s\begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + t\begin{pmatrix} -1 \\ 0 \\ -1 \\ 0 \\ 1 \end{pmatrix} \;\middle|\; s, t \in \mathbb{R} \right\}.$$

It is easy to check that $(1, -1, 1, 1, 1)^{\mathrm{T}}$ is a solution to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ (by substituting the vector into the equation). So the system has a general solution

$$\boldsymbol{x} = s\begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + t\begin{pmatrix} -1 \\ 0 \\ -1 \\ 0 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ -1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \quad \text{where } s, t \text{ are arbitrary parameters.}$$

Remark 4.3.9 (For students who have learnt ordinary differential equations) Consider the ordinary differential equation

$$a\frac{d^2y}{dx^2} + b\frac{dy}{dx} + cy = f(x)$$

where a, b, c are real constants. There is a general solution of the form

$$y = \left(\text{a general solution for } a\frac{d^2y}{dx^2} + b\frac{dy}{dx} + cy = 0\right) + \left(\text{one particular solution to } a\frac{d^2y}{dx^2} + b\frac{dy}{dx} + cy = f(x)\right).$$

The form looks very similar to the solutions to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ discussed in Theorem 4.3.6. This is not a coincidence. To explain the relation between these two different types of equations, we need the concept of "abstract" vector spaces that will be covered in a more advanced linear algebra course.

Exercise 4

1. For each of the following $m \times n$ matrices,

   (i) find a basis for the row space and a basis for the column space;

   (ii) extend the basis for the row space in (i) to a basis for $\mathbb{R}^m$;

   (iii) extend the basis for the column space in (i) to a basis for $\mathbb{R}^n$;

   (iv) find a basis for the nullspace;

   (v) find the rank and nullity of the matrix and hence verify the Dimension Theorem for Matrices; and

   (vi) determine if the matrix has full rank.

   (a) $\boldsymbol{A} = \begin{pmatrix} 1 & 4 & 0 & 5 & 2 \\ 2 & 1 & 0 & 3 & 0 \\ -1 & 3 & 0 & 2 & 2 \\ 1 & -1 & 1 & -1 & 1 \end{pmatrix}$, (b) $\boldsymbol{B} = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ -1 & 3 & 6 \\ 2 & 1 & 0 \\ 3 & 1 & -1 \end{pmatrix}$,

   (c) $\boldsymbol{C} = \begin{pmatrix} 2 & 1 & 4 & 1 & 2 \\ 4 & 2 & 2 & 3 & 2 \\ 2 & 1 & -2 & 2 & 0 \\ 6 & 3 & 6 & 4 & 4 \end{pmatrix}$, (d) $\boldsymbol{D} = \begin{pmatrix} 1 & 4 & 5 & 8 \\ -1 & 4 & 3 & 0 \\ 2 & 0 & 2 & 1 \end{pmatrix}$.

2. Let W be the subspace of $\mathbb{R}^5$ spanned by the following vectors: $\boldsymbol{u_1} = (1, -2, 0, 0, 3)$, $\boldsymbol{u_2} = (2, -5, -3, -2, 6)$, $\boldsymbol{u_3} = (0, 5, 15, 10, 0)$, $\boldsymbol{u_4} = (2, 1, 15, 8, 6)$.

   (a) Find a basis for W.

   (b) What is $\dim(W)$?

   (c) Extend the basis for W found in Part (a) to a basis for $\mathbb{R}^5$.

3. For each of the following vector spaces $V = \operatorname{span}(S)$, find a subset S' of S such that S' forms a basis for V.

   (a) $S = \{(1, 0, 1, 3), (2, -1, 0, 1), (-1, 3, 5, 12), (0, 1, 2, 5), (3, -1, 1, 4)\}$

   (b) $S = \{(1, 0, 1, 3, 4), (2, 1, -2, 1, 0), (-3, -2, 5, 1, 4), (0, 5, 2, 1, 1), (0, 4, 6, 6, 9)\}$

4. Find a basis for the following subspace of $\mathbb{R}^5$:

   $V = \{\, (a+b+3c+3d,\ b+2c+d,\ a+c+2d,\ -a-b-3c-3d,\ a+c+2d) \mid a, b, c, d \in \mathbb{R}\,\}$.

5. Let $\boldsymbol{A} = \begin{pmatrix} 1 & 0 & -1 & 1 & 1 \\ 1 & 1 & 2 & 0 & 3 \\ -1 & 0 & 2 & -1 & 0 \end{pmatrix}$ and $\boldsymbol{R} = \begin{pmatrix} 1 & 0 & 0 & 1 & 2 \\ 0 & 1 & 0 & -1 & -1 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix}$.

(a) Show that $\boldsymbol{R}$ is the reduced row-echelon form of $\boldsymbol{A}$.

(b) Let $S = \{(1, 0, -1, 1, 1),\ (1, 1, 2, 0, 3),\ (-1, 0, 2, -1, 0)\}$ and $T = \{(1, 0, 0, 1, 2),\ (0, 1, 0, -1, -1),\ (0, 0, 1, 0, 1)\}$.

(i) Explain why S is a basis for span(T).

(ii) Find the transition matrix from S to T.

6. Let $V = \text{span}\{\boldsymbol{u_1},\ \boldsymbol{u_2},\ \boldsymbol{u_3},\ \boldsymbol{u_4}\}$ where $\boldsymbol{u_1} = (1,\ 1,\ 1,\ 1,\ 1)$, $\boldsymbol{u_2} = (1,\ a,\ a,\ a,\ a)$, $\boldsymbol{u_3} = (1,\ a,\ a^2,\ a,\ a^2)$, $\boldsymbol{u_4} = (1,\ a^3,\ a,\ 2a - a^3,\ a)$ for some constant a. Find a basis for V and determine the dimension of V.

7. Let $V = \text{span}\{\,(1, 1, 0, 0),\ (-1, 0, 1, 0)\,\}$ and $W = \text{span}\{\,(-1, 2, 3, 0),\ (2, -1, 2, -1)\,\}$. Find a basis for $V + W$. (See Question 3.20.)

8. Let $V = \{\, a(1, 2, 0, 0) + b(0, -1, 1, 0) + c(0, 0, 0, 1) \mid a, b, c \in \mathbb{R}\,\}$.

(a) Find a 4×4 matrix $\boldsymbol{A}$ such that the row space of $\boldsymbol{A}$ is V.

(b) Find a 4×4 matrix $\boldsymbol{B}$ such that the column space of $\boldsymbol{B}$ is V.

(c) Find a 4×4 matrix $\boldsymbol{C}$ such that the nullspace of $\boldsymbol{C}$ is V.

9. Let $\boldsymbol{A}$ be a 3×4 matrix. Suppose that $x_1 = 1$, $x_2 = 0$, $x_3 = -1$, $x_4 = 0$ is a solution to a non-homogeneous linear system $\boldsymbol{Ax} = \boldsymbol{b}$ and that the homogeneous system $\boldsymbol{Ax} = \boldsymbol{0}$ has a general solution $x_1 = t - 2s$, $x_2 = s + t$, $x_3 = s$, $x_4 = t$ where s, t are arbitrary parameters.

(a) Find a basis for the nullspace of $\boldsymbol{A}$ and determine the nullity of $\boldsymbol{A}$.

(b) Find a general solution for the system $\boldsymbol{Ax} = \boldsymbol{b}$.

(c) Write down the reduced row-echelon form of $\boldsymbol{A}$.

(d) Find a basis for the row space of $\boldsymbol{A}$ and determine the rank of $\boldsymbol{A}$.

(e) Do we have enough information for us to find the column space of $\boldsymbol{A}$?

10. Let $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{a_1} & \boldsymbol{a_2} & \boldsymbol{a_3} & \boldsymbol{a_4} & \boldsymbol{a_5} \end{pmatrix}$ be a 4×5 matrix such that the columns $\boldsymbol{a_1}$, $\boldsymbol{a_2}$, $\boldsymbol{a_3}$ are linearly independent while $\boldsymbol{a_4} = \boldsymbol{a_1} - 2\boldsymbol{a_2} + \boldsymbol{a_3}$ and $\boldsymbol{a_5} = \boldsymbol{a_2} + \boldsymbol{a_3}$.

(a) Determine the reduced row-echelon form of $\boldsymbol{A}$. (Hint: The linear relations between columns will not be changed by row operations. In this question, the fifth column of $\boldsymbol{A}$ is the sum of the second and the third columns of $\boldsymbol{A}$. Then the fifth column of the reduced row-echelon form $\boldsymbol{R}$ is still the sum of the second and the third columns of $\boldsymbol{R}$.)

(b) Find a basis for the row space of $\boldsymbol{A}$ and a basis for the column space of $\boldsymbol{A}$.

11. For each of the following $\boldsymbol{A}$ and $\boldsymbol{b}$, solve the linear system $\boldsymbol{Ax} = \boldsymbol{b}$. Show that $\boldsymbol{b}$ belongs to the column space of $\boldsymbol{A}$ by expressing $\boldsymbol{b}$ as a linear combination of the columns of $\boldsymbol{A}$.

(a) $\boldsymbol{A} = \begin{pmatrix} 1 & 1 & 5 \\ 2 & 3 & 4 \\ 0 & 1 & -1 \\ 1 & 1 & 2 \end{pmatrix}$ and $\boldsymbol{b} = \begin{pmatrix} 16 \\ 13 \\ -4 \\ 7 \end{pmatrix}$,

(b) $\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 2 & -1 & 1 \\ 1 & 1 & -1 & 2 & 0 \\ 3 & 1 & 0 & 0 & -1 \end{pmatrix}$ and $\boldsymbol{b} = \begin{pmatrix} -1 \\ 9 \\ 4 \end{pmatrix}$.

12. For each of the following cases, write down a matrix with the required property or explain why no such matrix exists.

(a) The column space contains vectors $(1, 0, 0)^{\mathrm{T}}$, $(0, 0, 1)^{\mathrm{T}}$ and the row space contains vectors $(1, 1)$, $(1, 2)$.

(b) The column space $= \mathbb{R}^4$ and the row space $= \mathbb{R}^3$.

(c) The column space $=$ the row space $= \operatorname{span}\{(1, 2, 3)\}$.

(d) A 2×2 matrix with the column space $=$ the nullspace.

13. For each of the following, find the largest possible value for the rank of the matrix $\boldsymbol{A}$ and the smallest possible value for the nullity of $\boldsymbol{A}$.

(a) $\boldsymbol{A}$ is 5×5. (b) $\boldsymbol{A}$ is 4×6. (c) $\boldsymbol{A}$ is 8×3.

14. Consider a 2×2 matrix $\boldsymbol{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. What can you say about the values of a, b, c, d for each of the following cases?

(a) $\operatorname{rank}(\boldsymbol{A}) = 0$. (b) $\operatorname{rank}(\boldsymbol{A}) = 2$. (c) $\operatorname{rank}(\boldsymbol{A}) = 1$.

15. Determine the possible rank and nullity of each of the following matrices:

(a) $\boldsymbol{A} = \begin{pmatrix} 1 & 1 & a \\ 1 & a & 1 \\ a & 1 & 1 \end{pmatrix}$, (b) $\boldsymbol{B} = \begin{pmatrix} 0 & 0 & b \\ 0 & 0 & c \\ d & e & f \end{pmatrix}$,

where a, b, c, d, e, f are real numbers.

16. Let $\boldsymbol{X_n} = (x_{ij})$ be a $(2n+1) \times (2n+1)$ matrix such that

$$x_{ij} = \begin{cases} 1 & \text{if } i = j \text{ or } i = 2n + 2 - j \\ 0 & \text{otherwise.} \end{cases}$$

(a) Write down $\boldsymbol{X_1}$, $\boldsymbol{X_2}$ and find their rank and nullity.

(b) In general, what are $\text{rank}(\boldsymbol{X_n})$ and $\text{nullity}(\boldsymbol{X_n})$?

17. Let $\boldsymbol{A}$ be a 3×3 matrix. Describe geometrically the solution set of the linear system $\boldsymbol{Ax} = \boldsymbol{0}$ for each of the cases when $\text{rank}(\boldsymbol{A}) = 0, 1, 2, 3$.

18. Prove Theorem 4.1.11:

    Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be row equivalent matrices. Prove the following statements.

    (a) A given set of columns of $\boldsymbol{A}$ is linearly independent if and only if the set of corresponding columns of $\boldsymbol{B}$ is linearly independent. (Hint: Read Question 3.30.)

    (b) A given set of columns of $\boldsymbol{A}$ forms a basis for the column space of $\boldsymbol{A}$ if and only if the set of corresponding columns of $\boldsymbol{B}$ forms a basis for the column space of $\boldsymbol{B}$.

19. Let $\boldsymbol{B}$ be an $m \times n$ matrix. If there exists an $n \times m$ matrix $\boldsymbol{C}$ such that $\boldsymbol{BC} = \boldsymbol{I}$, then $\boldsymbol{C}$ is called a *right inverse* of $\boldsymbol{B}$.

    (a) Let $\boldsymbol{B} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix}$. Solve each of the following linear systems:

    $$\text{(i) } \boldsymbol{B}\begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \qquad \text{(ii) } \boldsymbol{B}\begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \qquad \text{(iii) } \boldsymbol{B}\begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

    Hence, or otherwise, find a right inverse of $\boldsymbol{B}$. Is the right inverse unique?

    (b) Give an example of a nonzero matrix that has no right inverse.

    (c) Show that an $m \times n$ matrix $\boldsymbol{B}$ has a right inverse if and only if $\text{rank}(\boldsymbol{B}) = m$.

20. Suppose $\boldsymbol{A}$ and $\boldsymbol{B}$ are two matrices such that $\boldsymbol{AB} = \boldsymbol{0}$. Show that the column space of $\boldsymbol{B}$ is contained in the nullspace of $\boldsymbol{A}$.

21. Show that there is no matrix whose row space and nullspace both contain the same nonzero vector.

22. Let $\boldsymbol{A}$ be an $m \times n$ matrix and $\boldsymbol{P}$ an $m \times m$ matrix.

    (a) If $\boldsymbol{P}$ is invertible, show that $\text{rank}(\boldsymbol{PA}) = \text{rank}(\boldsymbol{A})$.

    (b) Give an example such that $\text{rank}(\boldsymbol{PA}) < \text{rank}(\boldsymbol{A})$.

(c) Suppose $\operatorname{rank}(\boldsymbol{PA}) = \operatorname{rank}(\boldsymbol{A})$. Is it true that $\boldsymbol{P}$ must be invertible? Justify your answer.

23. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two matrices of the same size. Show that

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

24. Let $\boldsymbol{A}$ be an $m \times n$ matrix. Suppose the linear system $\boldsymbol{Ax} = \boldsymbol{b}$ is consistent for all $\boldsymbol{b} \in \mathbb{R}^m$. Show that the linear system $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{0}$ has only the trivial solution.

25. Let $\boldsymbol{A}$ be an $m \times n$ matrix.

(a) Show that the nullspace of $\boldsymbol{A}$ is equal to the nullspace of $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$.

(b) Show that $\operatorname{nullity}(\boldsymbol{A}) = \operatorname{nullity}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})$ and $\operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})$.

(c) Is it true that $\operatorname{nullity}(\boldsymbol{A}) = \operatorname{nullity}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})$? Justify your answer.

(d) Is it true that $\operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})$? Justify your answer.

26. Determine which of the following statements are true. Justify your answer.

(a) If $\boldsymbol{A}$ and $\boldsymbol{B}$ are two row equivalent matrices, then the row space of $\boldsymbol{A}^{\mathrm{T}}$ and the row space of $\boldsymbol{B}^{\mathrm{T}}$ are the same.

(b) If $\boldsymbol{A}$ and $\boldsymbol{B}$ are two row equivalent matrices, then the column space of $\boldsymbol{A}^{\mathrm{T}}$ and the column space of $\boldsymbol{B}^{\mathrm{T}}$ are the same.

(c) If $\boldsymbol{A}$ and $\boldsymbol{B}$ are two row equivalent matrices, then the nullspace of $\boldsymbol{A}^{\mathrm{T}}$ and the nullspace of $\boldsymbol{B}^{\mathrm{T}}$ are the same.

(d) If $\boldsymbol{A}$ and $\boldsymbol{B}$ are two matrices of the same size, then $\operatorname{rank}(\boldsymbol{A}+\boldsymbol{B}) = \operatorname{rank}(\boldsymbol{A}) + \operatorname{rank}(\boldsymbol{B})$.

(e) If $\boldsymbol{A}$ and $\boldsymbol{B}$ are two matrices of the same size, then $\operatorname{nullity}(\boldsymbol{A}+\boldsymbol{B}) = \operatorname{nullity}(\boldsymbol{A}) + \operatorname{nullity}(\boldsymbol{B})$.

(f) If $\boldsymbol{A}$ is an $n \times m$ matrix and $\boldsymbol{B}$ is an $m \times n$ matrix, then $\operatorname{rank}(\boldsymbol{AB}) = \operatorname{rank}(\boldsymbol{BA})$.

(g) If $\boldsymbol{A}$ is an $n \times m$ matrix and $\boldsymbol{B}$ is an $m \times n$ matrix, then $\operatorname{nullity}(\boldsymbol{AB}) = \operatorname{nullity}(\boldsymbol{BA})$.

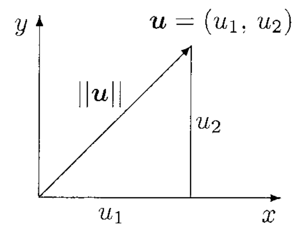
Chapter 5

Orthogonality

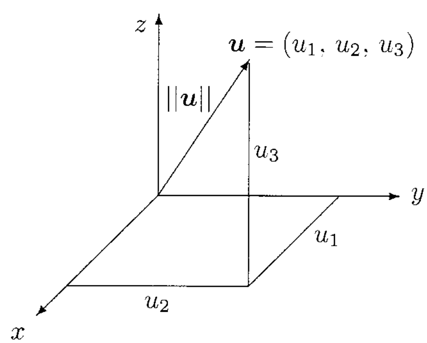
Section 5.1 The Dot Product

Discussion 5.1.1

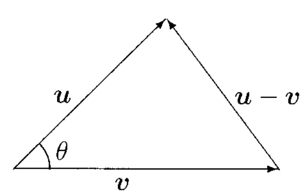
1. Let $\boldsymbol{u} = (u_1, u_2)$ be a vector in $\mathbb{R}^2$. Then the length of $\boldsymbol{u}$ is given by
$$||\boldsymbol{u}|| = \sqrt{u_1^2 + u_2^2}.$$



2. Similarly, the length of a vector $\boldsymbol{u} = (u_1, u_2, u_3)$ in $\mathbb{R}^3$ is given by
$$||\boldsymbol{u}|| = \sqrt{u_1^2 + u_2^2 + u_3^2}.$$



3. Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be two vectors from $\mathbb{R}^2$ or $\mathbb{R}^3$. The distance between $\boldsymbol{u}$ and $\boldsymbol{v}$ is equal to the length of the vector $\boldsymbol{u} - \boldsymbol{v}$, i.e. $d(\boldsymbol{u}, \boldsymbol{v}) = ||\boldsymbol{u} - \boldsymbol{v}||$. Let θ be the angle between the two vectors.



By the *cosine rule* of trigonometry, we have

$$||\boldsymbol{u}-\boldsymbol{v}||^2 = ||\boldsymbol{u}||^2 + ||\boldsymbol{v}||^2 - 2||\boldsymbol{u}||\,||\boldsymbol{v}||\cos(\theta)$$

and hence

$$\theta = \cos^{-1}\left(\frac{||\boldsymbol{u}||^2 + ||\boldsymbol{v}||^2 - ||\boldsymbol{u}-\boldsymbol{v}||^2}{2||\boldsymbol{u}||\,||\boldsymbol{v}||}\right).$$

If $\boldsymbol{u} = (u_1,\, u_2)$ and $\boldsymbol{v} = (v_1,\, v_2)$ are vectors in $\mathbb{R}^2$, then

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

and

$$\begin{aligned}\theta &= \cos^{-1}\left(\frac{\left[u_1^2+u_2^2\right] + \left[v_1^2+v_2^2\right] - \left[(u_1-v_1)^2+(u_2-v_2)^2\right]}{2||\boldsymbol{u}||\,||\boldsymbol{v}||}\right)\\ &= \cos^{-1}\left(\frac{u_1v_1+u_2v_2}{||\boldsymbol{u}||\,||\boldsymbol{v}||}\right).\end{aligned}$$

Similarly, if $\boldsymbol{u} = (u_1,\, u_2,\, u_3)$ and $\boldsymbol{v} = (v_1,\, v_2,\, v_3)$ are vectors in $\mathbb{R}^3$, then

$$d(\boldsymbol{u},\boldsymbol{v}) = ||\boldsymbol{u}-\boldsymbol{v}|| = \sqrt{(u_1-v_1)^2 + (u_2-v_2)^2 + (u_3-v_3)^2}.$$

and

$$\theta = \cos^{-1}\left(\frac{u_1v_1+u_2v_2+u_3v_3}{||\boldsymbol{u}||\,||\boldsymbol{v}||}\right).$$

Definition 5.1.2 Let $\boldsymbol{u} = (u_1,\, u_2,\, \ldots,\, u_n)$ and $\boldsymbol{v} = (v_1,\, v_2,\, \ldots,\, v_n)$ be two vectors in $\mathbb{R}^n$.

1. The *dot product* (or *inner product*) of $\boldsymbol{u}$ and $\boldsymbol{v}$ is defined to be the value

$$\boldsymbol{u}\cdot\boldsymbol{v} = u_1v_1 + u_2v_2 + \cdots + u_nv_n.$$

2. The *norm* (or *length*) of $\boldsymbol{u}$ is defined to be

$$||\boldsymbol{u}|| = \sqrt{\boldsymbol{u}\cdot\boldsymbol{u}} = \sqrt{u_1^2+u_2^2+\cdots+u_n^2}.$$

In particular, vectors of norm 1 are called *unit vectors.*

3. The *distance* between $\boldsymbol{u}$ and $\boldsymbol{v}$ is

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

4. The *angle* between $\boldsymbol{u}$ and $\boldsymbol{v}$ is

$$\cos^{-1}\left(\frac{\boldsymbol{u}\cdot\boldsymbol{v}}{||\boldsymbol{u}||\,||\boldsymbol{v}||}\right).$$

(By Question 5.4(a), $-1 \le \dfrac{\boldsymbol{u}\cdot\boldsymbol{v}}{||\boldsymbol{u}||\,||\boldsymbol{v}||} \le 1$ and hence the angle is well-defined.)

Remark 5.1.3 Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be vectors in $\mathbb{R}^n$. Suppose $\boldsymbol{u}$ and $\boldsymbol{v}$ are written as row vectors, i.e. $\boldsymbol{u} = \begin{pmatrix} u_1 & u_2 & \cdots & u_n \end{pmatrix}$ and $\boldsymbol{v} = \begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix}$. Then

$$\boldsymbol{u} \cdot \boldsymbol{v} = \begin{pmatrix} u_1 & u_2 & \cdots & u_n \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}.$$

Suppose $\boldsymbol{u}$ and $\boldsymbol{v}$ are written as column vectors, i.e. $\boldsymbol{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$ and $\boldsymbol{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$. Then

$$\boldsymbol{u} \cdot \boldsymbol{v} = \begin{pmatrix} u_1 & u_2 & \cdots & u_n \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{v}.$$

Example 5.1.4 Let $\boldsymbol{u} = (1, -2, 2, -1)$ and $\boldsymbol{v} = (1, 0, 2, 0)$. Then

$$\begin{aligned} \boldsymbol{u} \cdot \boldsymbol{v} &= 1 \cdot 1 + (-2) \cdot 0 + 2 \cdot 2 + (-1) \cdot 0 = 5 \\ ||\boldsymbol{u}|| &= \sqrt{1^2 + (-2)^2 + 2^2 + (-1)^2} = \sqrt{10}, \\ ||\boldsymbol{v}|| &= \sqrt{1^2 + 0^2 + 2^2 + 0^2} = \sqrt{5}, \\ d(\boldsymbol{u}, \boldsymbol{v}) &= \sqrt{(1-1)^2 + (-2-0)^2 + (2-2)^2 + (-1-0)^2} = \sqrt{5} \end{aligned}$$

and the angle between $\boldsymbol{u}$ and $\boldsymbol{v}$ is $\cos^{-1}\left(\dfrac{5}{\sqrt{10}\sqrt{5}}\right) = \cos^{-1}\left(\dfrac{1}{\sqrt{2}}\right) = \dfrac{\pi}{4}$.

Theorem 5.1.5 Let $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ be vectors in $\mathbb{R}^n$ and c a scalar. Then

1. $\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{v} \cdot \boldsymbol{u}$;
2. $(\boldsymbol{u} + \boldsymbol{v}) \cdot \boldsymbol{w} = \boldsymbol{u} \cdot \boldsymbol{w} + \boldsymbol{v} \cdot \boldsymbol{w}$ and $\boldsymbol{w} \cdot (\boldsymbol{u} + \boldsymbol{v}) = \boldsymbol{w} \cdot \boldsymbol{u} + \boldsymbol{w} \cdot \boldsymbol{v}$;
3. $(c\boldsymbol{u}) \cdot \boldsymbol{v} = \boldsymbol{u} \cdot (c\boldsymbol{v}) = c(\boldsymbol{u} \cdot \boldsymbol{v})$;
4. $||c\boldsymbol{u}|| = |c| \, ||\boldsymbol{u}||$.
5. $\boldsymbol{u} \cdot \boldsymbol{u} \geq 0$; and $\boldsymbol{u} \cdot \boldsymbol{u} = 0$ if and only if $\boldsymbol{u} = \boldsymbol{0}$.

Proof We only prove the last part of the theorem:

Let $\boldsymbol{u} = (u_1, u_2, \ldots, u_n)$. Then

$$\boldsymbol{u} \cdot \boldsymbol{u} = u_1^2 + u_2^2 + \cdots + u_n^2 \geq 0.$$

If $\boldsymbol{u} = \boldsymbol{0}$, it is obvious that $\boldsymbol{u} \cdot \boldsymbol{u} = 0$.

On the other hand, suppose $\boldsymbol{u} \cdot \boldsymbol{u} = 0$. Then $u_1^2 + u_2^2 + \cdots + u_n^2 = 0$ and this implies $u_1 = 0$, $u_2 = 0$, ..., $u_n = 0$, i.e. $\boldsymbol{u} = \boldsymbol{0}$.

(Proofs of the other parts of the theorem are left as exercises. See Question 5.3.)

Discussion 5.1.6 (Information Retrieval) The growth of digital libraries on the Internet has led to dramatic improvements in the storage and retrieval of information. In a typical situation, a database consists of a collection of documents and we want to search the collection and find documents that are best matched by some keywords. Let us consider a simple example:

Suppose there are 6 documents, D_1 to D_6, and 5 words, k_1 to k_5. We associate each document D_j with a 5×1 column vector, $\boldsymbol{d_j}$, such that the ith component of $\boldsymbol{d_j}$ is 1 if the word k_i appears in the document D_j and 0 otherwise.

	$\boldsymbol{d_1}$	$\boldsymbol{d_2}$	$\boldsymbol{d_3}$	$\boldsymbol{d_4}$	$\boldsymbol{d_5}$	$\boldsymbol{d_6}$
k_1	1	1	0	0	1	0
k_2	1	1	1	1	0	0
k_3	0	0	1	0	0	1
k_4	1	0	0	0	1	1
k_5	1	0	0	1	1	1

For example, we want to search for documents that consist of words k_1, k_4 and k_5. Let $\boldsymbol{x} = (1, 0, 0, 1, 1)^{\mathrm{T}}$. Note that $\boldsymbol{d_j} \cdot \boldsymbol{x}$ tells us how many words among k_1, k_4 and k_5 appear in the document D_j. To do the search systematically, we define a 5×6 matrix $\boldsymbol{Q}$ such that $\boldsymbol{d_j}$ is the j the column of the matrix, i.e.

$$\boldsymbol{Q} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}.$$

Then the jth component of $\boldsymbol{y} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{x}$ gives us the value of $\boldsymbol{d_j} \cdot \boldsymbol{x}$ and hence $\boldsymbol{y}$ tells us results of the search. For this example,

$$\boldsymbol{y} = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 0 \\ 1 \\ 3 \\ 2 \end{pmatrix}.$$

Hence documents D_1 and D_5 match the search.

Suppose, in addition, instead of telling us whether a word appears in the document D_j, we redefine $\boldsymbol{d_j}$ such that its ith component gives us the number of times that the word k_i appears in the document.

	$\boldsymbol{d_1}$	$\boldsymbol{d_2}$	$\boldsymbol{d_3}$	$\boldsymbol{d_4}$	$\boldsymbol{d_5}$	$\boldsymbol{d_6}$
k_1	2	1	0	0	3	0
k_2	1	3	1	2	0	0
k_3	0	0	5	0	0	6
k_4	2	0	0	0	5	1
k_5	1	0	0	1	1	4

There are a lot of ways to determine how close a document vector $\boldsymbol{d_j}$ and the search vector $\boldsymbol{x}$ are. One commonly used method is to compare the angle θ_j between $\boldsymbol{d_j}$ and $\boldsymbol{x}$ by computing

$$\cos(\theta_j) = \frac{\boldsymbol{d_j} \cdot \boldsymbol{x}}{||\boldsymbol{d_j}||\,||\boldsymbol{x}||}.$$

Let $\boldsymbol{x'} = \dfrac{1}{||\boldsymbol{x}||}\boldsymbol{x}$ and let $\boldsymbol{Q'}$ be the 5×6 matrix such that the jth column of $\boldsymbol{Q'}$ is $\dfrac{1}{||\boldsymbol{d_j}||}\boldsymbol{d_j}$, i.e.

$$\boldsymbol{x'} = \begin{pmatrix} 0.5774 \\ 0 \\ 0 \\ 0.5774 \\ 0.5774 \end{pmatrix}$$

and

$$\boldsymbol{Q'} = \begin{pmatrix} 0.6325 & 0.3162 & 0 & 0 & 0.5071 & 0 \\ 0.3162 & 0.9487 & 0.1961 & 0.8944 & 0 & 0 \\ 0 & 0 & 0.9801 & 0 & 0 & 0.8242 \\ 0.6325 & 0 & 0 & 0 & 0.8452 & 0.1374 \\ 0.3162 & 0 & 0 & 0.4472 & 0.1690 & 0.5494 \end{pmatrix}.$$

Then the jth component of $\boldsymbol{y'} = \boldsymbol{Q'}^{\mathrm{T}}\boldsymbol{x'}$ is equal to $\cos(\theta_j)$. For our example,

$$\boldsymbol{y'} = \begin{pmatrix} 0.6325 & 0.3162 & 0 & 0.6325 & 0.3162 \\ 0.3162 & 0.9487 & 0 & 0 & 0 \\ 0 & 0.1961 & 0.9801 & 0 & 0 \\ 0 & 0.8944 & 0 & 0 & 0.4472 \\ 0.5071 & 0 & 0 & 0.8452 & 0.1690 \\ 0 & 0 & 0.8242 & 0.1374 & 0.5494 \end{pmatrix} \begin{pmatrix} 0.5774 \\ 0 \\ 0 \\ 0.5774 \\ 0.5774 \end{pmatrix} = \begin{pmatrix} 0.9130 \\ 0.1826 \\ 0 \\ 0.2582 \\ 0.8784 \\ 0.6868 \end{pmatrix}.$$

Since $\cos(\theta_1) = 0.9130$ and $\cos(\theta_5) = 0.8784$, the document D_1 is closer to the search vector. Note that appearances of the words k_1, k_4 and k_5 in D_1 are more evenly distributed than that in D_5.

The example above illustrates some of the basic ideas of database searches. We can improve the search technique by applying some advanced results in linear algebra.

Section 5.2 Orthogonal and Orthonormal Bases

Definition 5.2.1

1. Two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{R}^n$ are called *orthogonal* if $\boldsymbol{u} \cdot \boldsymbol{v} = 0$.
2. A set S of vectors in $\mathbb{R}^n$ is called *orthogonal* if every pair of distinct vectors in S are orthogonal.
3. A set S of vectors in $\mathbb{R}^n$ is called *orthonormal* if S is orthogonal and every vector in S is a unit vector.

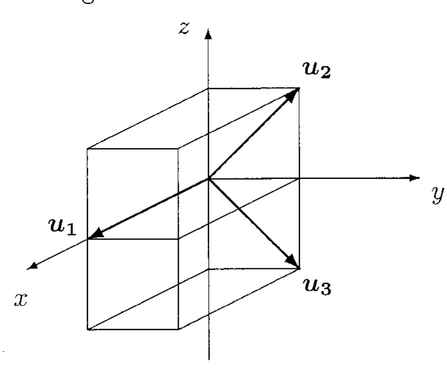
Remark 5.2.2 Given two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{R}^n$, if they are orthogonal, then the angle between them is equal to

$$\cos^{-1}\left(\frac{\boldsymbol{u} \cdot \boldsymbol{v}}{||\boldsymbol{u}||\,||\boldsymbol{v}||}\right) = \cos^{-1}(0) = \frac{\pi}{2}.$$

Thus the concept of "orthogonal" in $\mathbb{R}^n$ is the same as the concept of "perpendicular" in $\mathbb{R}^2$ and $\mathbb{R}^3$.

Example 5.2.3

1. Since $(1, 2, 2, -1) \cdot (1, 1, -1, 1) = 1 \cdot 1 + 2 \cdot 1 + 2 \cdot (-1) + (-1) \cdot 1 = 0$, $(1, 2, 2, -1)$ and $(1, 1, -1, 1)$ are orthogonal.
2. Let $\boldsymbol{u_1} = (2, 0, 0)$, $\boldsymbol{u_2} = (0, 1, 1)$ and $\boldsymbol{u_3} = (0, 1, -1)$. Since
$$\boldsymbol{u_1} \cdot \boldsymbol{u_2} = 0, \quad \boldsymbol{u_1} \cdot \boldsymbol{u_3} = 0 \quad \text{and} \quad \boldsymbol{u_2} \cdot \boldsymbol{u_3} = 0,$$
the set $\{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\}$ is orthogonal.



Let

$$\begin{aligned}
\boldsymbol{v_1} &= \frac{1}{||\boldsymbol{u_1}||}\boldsymbol{u_1} = \tfrac{1}{2}(2,0,0) = (1,0,0),\\
\boldsymbol{v_2} &= \frac{1}{||\boldsymbol{u_2}||}\boldsymbol{u_2} = \tfrac{1}{\sqrt{2}}(0,1,1) = (0,\tfrac{1}{\sqrt{2}},\tfrac{1}{\sqrt{2}}) \quad \text{and}\\
\boldsymbol{v_3} &= \frac{1}{||\boldsymbol{u_3}||}\boldsymbol{u_3} = \tfrac{1}{\sqrt{2}}(0,1,-1) = (0,\tfrac{1}{\sqrt{2}},-\tfrac{1}{\sqrt{2}}).
\end{aligned}$$

Then by Theorem 5.1.5,

$$||\boldsymbol{v_i}|| = \left|\left|\frac{1}{||\boldsymbol{u_i}||}\boldsymbol{u_i}\right|\right| = \frac{1}{||\boldsymbol{u_i}||}||\boldsymbol{u_i}|| = 1 \quad \text{for all } i$$

and

$$\boldsymbol{v_i}\cdot\boldsymbol{v_j} = \left(\frac{1}{||\boldsymbol{u_i}||}\boldsymbol{u_i}\right)\cdot\left(\frac{1}{||\boldsymbol{u_j}||}\boldsymbol{u_j}\right) = \frac{1}{||\boldsymbol{u_i}||\,||\boldsymbol{u_j}||}(\boldsymbol{u_i}\cdot\boldsymbol{u_j}) = 0 \quad \text{if } i \neq j.$$

So the set $\{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}\}$ is orthonormal.

The process of converting an orthogonal set to an orthonormal set by multiplying each vector $\boldsymbol{u}$ by $\dfrac{1}{||\boldsymbol{u}||}$ is called *normalizing.*

3. Consider the standard basis $E = \{\boldsymbol{e_1}, \boldsymbol{e_2}, \ldots, \boldsymbol{e_n}\}$ for $\mathbb{R}^n$. It is easy to check that $||\boldsymbol{e_i}|| = 1$ for all i and $\boldsymbol{e_i}\cdot\boldsymbol{e_j} = 0$ for $i \neq j$. So E is an orthonormal set.

Theorem 5.2.4 Let S be an orthogonal set of nonzero vectors in a vector space. Then S is linearly independent.

Proof Let $S = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}\}$. Consider the vector equation

$$c_1\boldsymbol{u_1} + c_2\boldsymbol{u_2} + \cdots + c_k\boldsymbol{u_k} = \boldsymbol{0}. \tag{5.1}$$

Since S is orthogonal, $\boldsymbol{u_i}\cdot\boldsymbol{u_j} = 0$ for $i \neq j$. For $i = 1, 2, \ldots, k$, by Theorem 5.1.5,

$$\begin{aligned}
(c_1\boldsymbol{u_1} + c_2\boldsymbol{u_2} + \cdots + c_k\boldsymbol{u_k})\cdot\boldsymbol{u_i} &= (c_1\boldsymbol{u_1})\cdot\boldsymbol{u_i} + (c_2\boldsymbol{u_2})\cdot\boldsymbol{u_i} + \cdots + (c_k\boldsymbol{u_k})\cdot\boldsymbol{u_i}\\
&= c_1(\boldsymbol{u_1}\cdot\boldsymbol{u_i}) + c_2(\boldsymbol{u_2}\cdot\boldsymbol{u_i}) + \cdots + c_{i-1}(\boldsymbol{u_{i-1}}\cdot\boldsymbol{u_i})\\
&\qquad + c_i(\boldsymbol{u_i}\cdot\boldsymbol{u_i}) + c_{i+1}(\boldsymbol{u_{i+1}}\cdot\boldsymbol{u_i}) + \cdots + c_k(\boldsymbol{u_k}\cdot\boldsymbol{u_i})\\
&= c_i(\boldsymbol{u_i}\cdot\boldsymbol{u_i}).
\end{aligned} \tag{5.2}$$

Taking dot product on both sides of Equation (5.1) with $\boldsymbol{u_i}$, we have

$$c_i(\boldsymbol{u_i}\cdot\boldsymbol{u_i}) = (c_1\boldsymbol{u_1} + c_2\boldsymbol{u_2} + \cdots + c_k\boldsymbol{u_k})\cdot\boldsymbol{u_i} = \boldsymbol{0}\cdot\boldsymbol{u_i} = 0.$$

Given that $\boldsymbol{u_i} \neq \boldsymbol{0}$, by Theorem 5.1.5.5, $\boldsymbol{u_i}\cdot\boldsymbol{u_i} \neq 0$. This means $c_i = 0$.
Since Equation (5.1) has only the trivial solution, S is linearly independent.

Definition 5.2.5

1. A basis S for a vector space is called an *orthogonal basis* if S is orthogonal.
2. A basis S for a vector space is called an *orthonormal basis* if S is orthonormal.

Remark 5.2.6 By Theorem 5.2.4 and Theorem 3.6.7, to determine whether a set S of nonzero vectors in a vector space of dimension k is an orthogonal (respectively, orthonormal) basis, we only need to check (i) S is orthogonal (respectively, orthonormal) and (ii) $|S| = k$.

Example 5.2.7

1. The standard basis for $\mathbb{R}^n$ is an orthogonal basis as well as an orthonormal basis.
2. By Remark 5.2.6, the sets $\{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\}$ and $\{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}\}$ in Example 5.2.3.2 are orthogonal bases for $\mathbb{R}^3$ while the set $\{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}\}$ is an orthonormal basis for $\mathbb{R}^3$.

Theorem 5.2.8

1. If $S = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}\}$ is an orthogonal basis for a vector space V, then for any vector $\boldsymbol{w}$ in V,
$$\boldsymbol{w} = \frac{\boldsymbol{w}\cdot\boldsymbol{u_1}}{\boldsymbol{u_1}\cdot\boldsymbol{u_1}}\boldsymbol{u_1} + \frac{\boldsymbol{w}\cdot\boldsymbol{u_2}}{\boldsymbol{u_2}\cdot\boldsymbol{u_2}}\boldsymbol{u_2} + \cdots + \frac{\boldsymbol{w}\cdot\boldsymbol{u_k}}{\boldsymbol{u_k}\cdot\boldsymbol{u_k}}\boldsymbol{u_k},$$
i.e. $(\boldsymbol{w})_S = \left(\dfrac{\boldsymbol{w}\cdot\boldsymbol{u_1}}{\boldsymbol{u_1}\cdot\boldsymbol{u_1}}, \dfrac{\boldsymbol{w}\cdot\boldsymbol{u_2}}{\boldsymbol{u_2}\cdot\boldsymbol{u_2}}, \ldots, \dfrac{\boldsymbol{w}\cdot\boldsymbol{u_k}}{\boldsymbol{u_k}\cdot\boldsymbol{u_k}}\right)$.
2. If $T = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_k}\}$ is an orthonormal basis for a vector space V, then for any vector $\boldsymbol{w}$ in V,
$$\boldsymbol{w} = (\boldsymbol{w}\cdot\boldsymbol{v_1})\boldsymbol{v_1} + (\boldsymbol{w}\cdot\boldsymbol{v_2})\boldsymbol{v_2} + \cdots + (\boldsymbol{w}\cdot\boldsymbol{v_k})\boldsymbol{v_k},$$
i.e. $(\boldsymbol{w})_T = (\boldsymbol{w}\cdot\boldsymbol{v_1}, \boldsymbol{w}\cdot\boldsymbol{v_2}, \ldots, \boldsymbol{w}\cdot\boldsymbol{v_k})$.

Proof

1. Let $(\boldsymbol{w})_S = (c_1, c_2, \ldots, c_k)$, i.e. $\boldsymbol{w} = c_1\boldsymbol{u_1} + c_2\boldsymbol{u_2} + \cdots + c_k\boldsymbol{u_k}$. For $i = 1, 2, \ldots, k$,
$$\boldsymbol{w}\cdot\boldsymbol{u_i} = (c_1\boldsymbol{u_1} + c_2\boldsymbol{u_2} + \cdots + c_k\boldsymbol{u_k})\cdot\boldsymbol{u_i} = c_i(\boldsymbol{u_i}\cdot\boldsymbol{u_i})$$
(see Equation (5.2) in the proof of Theorem 5.2.4). Thus $c_i = \dfrac{\boldsymbol{w}\cdot\boldsymbol{u_i}}{\boldsymbol{u_1}\cdot\boldsymbol{u_1}}$.
2. The result follows from Part 1 because $\boldsymbol{v_i}\cdot\boldsymbol{v_i} = ||\boldsymbol{v_i}||^2 = 1$ for all i.

Example 5.2.9

1. Consider the orthonormal basis $S = \{\boldsymbol{v_1}, \boldsymbol{v_2}\}$ in $\mathbb{R}^2$ where $\boldsymbol{v_1} = (\frac{3}{5}, \frac{4}{5})$ and $\boldsymbol{v_2} = (\frac{4}{5}, -\frac{3}{5})$. For any vector $\boldsymbol{w} = (x, y)$,
$$\boldsymbol{w}\cdot\boldsymbol{v_1} = \frac{3x+4y}{5} \quad\text{and}\quad \boldsymbol{w}\cdot\boldsymbol{v_2} = \frac{4x-3y}{5}.$$

So
$$\boldsymbol{w} = \frac{3x+4y}{5}\boldsymbol{v_1} + \frac{4x-3y}{5}\boldsymbol{v_2}$$
and
$$(\boldsymbol{w})_S = \left(\frac{3x+4y}{5}, \frac{4x-3y}{5}\right).$$

2. Consider the orthogonal basis $S = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\}$ in $\mathbb{R}^3$ where $\boldsymbol{u_1} = (1, 1, 1)$, $\boldsymbol{u_2} = (1, 0, -1)$ and $\boldsymbol{u_3} = (1, -2, 1)$. Let $\boldsymbol{w} = (1, -1, 0)$. Then
$$(\boldsymbol{w})_S = \left(\frac{\boldsymbol{w}\cdot\boldsymbol{u_1}}{\boldsymbol{u_1}\cdot\boldsymbol{u_1}}, \frac{\boldsymbol{w}\cdot\boldsymbol{u_2}}{\boldsymbol{u_2}\cdot\boldsymbol{u_2}}, \frac{\boldsymbol{w}\cdot\boldsymbol{u_3}}{\boldsymbol{u_3}\cdot\boldsymbol{u_3}}\right) = \left(0, \frac{1}{2}, \frac{1}{2}\right).$$

Definition 5.2.10 Let V be a subspace of $\mathbb{R}^n$. A vector $\boldsymbol{u} \in \mathbb{R}^n$ is said to be *orthogonal* (or *perpendicular*) to V if $\boldsymbol{u}$ is orthogonal to all vectors in V.

Example 5.2.11

1. Let V be a plane in $\mathbb{R}^3$ defined by the equation $ax + by + cz = 0$. Let $\boldsymbol{n} = (a, b, c)$. For any vector $\boldsymbol{u} = (x, y, z)$ in V,
$$\boldsymbol{n}\cdot\boldsymbol{u} = ax + by + cz = 0.$$
Thus $\boldsymbol{n}$ is orthogonal to V. In fact,
$$V = \{\,(x, y, z) \in \mathbb{R}^3 \mid ax + by + cz = 0\} = \{\boldsymbol{u} \in \mathbb{R}^3 \mid \boldsymbol{n}\cdot\boldsymbol{u} = 0\,\}.$$
The vector $\boldsymbol{n}$ is called a *normal vector* of V.

2. Let $V = \operatorname{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}\}$ be a subspace of $\mathbb{R}^4$ where $\boldsymbol{u_1} = (1, 1, 1, 0)$ and $\boldsymbol{u_2} = (0, -1, -1, 1)$. Find all vectors that are orthogonal to V.

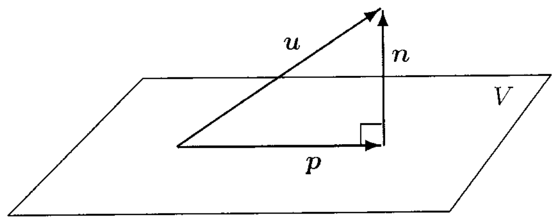
Solution Let $\boldsymbol{v} = (w, x, y, z)$ be a vector in $\mathbb{R}^4$. Then
$$\begin{aligned} & \boldsymbol{v}\cdot(a\boldsymbol{u_1} + b\boldsymbol{u_2}) = 0 \quad \text{for all } a, b \in \mathbb{R} \\ \Leftrightarrow\quad & \boldsymbol{v}\cdot\boldsymbol{u_1} = 0 \quad\text{and}\quad \boldsymbol{v}\cdot\boldsymbol{u_2} = 0 \\ \Leftrightarrow\quad & \left\{\begin{array}{l} w + x + y \qquad = 0 \\ \quad\ - x - y + z = 0 \end{array}\right. \\ \Leftrightarrow\quad & (w, x, y, z) = (-t,\ -s + t,\ s,\ t) \quad \text{for some } s, t \in \mathbb{R}. \end{aligned}$$
So a vector $\boldsymbol{v}$ is orthogonal to V if and only if
$$\boldsymbol{v} = (-t,\ -s + t,\ s,\ t) = s(0, -1, 1, 0) + t(-1, 1, 0, 1)$$
for some $s, t \in \mathbb{R}$, i.e. $\boldsymbol{v} \in \operatorname{span}\{\,(0, -1, 1, 0), (-1, 1, 0, 1)\,\}$.

Remark 5.2.12 In general, if $V = \operatorname{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}\}$ is a subspace of $\mathbb{R}^n$, then a vector $\boldsymbol{v} \in \mathbb{R}^n$ is orthogonal to V if and only if $\boldsymbol{v}\cdot\boldsymbol{u_i} = 0$ for $i = 1, 2, \ldots, k$.

Definition 5.2.13 Let V be a subspace of $\mathbb{R}^n$. Every vector $\boldsymbol{u} \in \mathbb{R}^n$ can be written uniquely as

$$\boldsymbol{u} = \boldsymbol{n} + \boldsymbol{p}$$

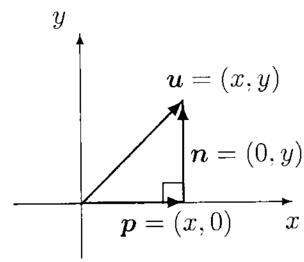
such that $\boldsymbol{n}$ is a vector orthogonal to V and $\boldsymbol{p}$ is a vector in V (see Question 5.18). The vector $\boldsymbol{p}$ is called the (*orthogonal*) *projection* of $\boldsymbol{u}$ onto V.



Example 5.2.14 We follow the notation of Definition 5.2.13 in the two examples below.

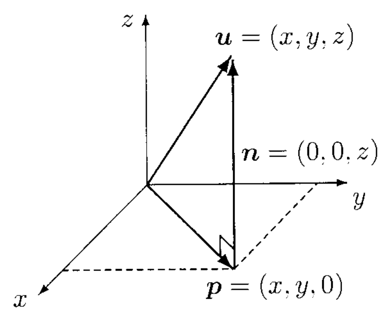
1. The projection of $\boldsymbol{u} = (x, y)$ onto the x-axis is $\boldsymbol{p} = (x, 0)$.

   In here, $\boldsymbol{n} = (0, y)$.



2. The projection of $\boldsymbol{u} = (x, y, z)$ onto the xy-plane is $\boldsymbol{p} = (x, y, 0)$.

   In here, $\boldsymbol{n} = (0, 0, z)$.



Theorem 5.2.15 Let V be a subspace of $\mathbb{R}^n$ and $\boldsymbol{w}$ a vector in $\mathbb{R}^n$.

1. If $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k\}$ is an orthogonal basis for V, then

$$\frac{\boldsymbol{w} \cdot \boldsymbol{u}_1}{\boldsymbol{u}_1 \cdot \boldsymbol{u}_1}\boldsymbol{u}_1 + \frac{\boldsymbol{w} \cdot \boldsymbol{u}_2}{\boldsymbol{u}_2 \cdot \boldsymbol{u}_2}\boldsymbol{u}_2 + \cdots + \frac{\boldsymbol{w} \cdot \boldsymbol{u}_k}{\boldsymbol{u}_k \cdot \boldsymbol{u}_k}\boldsymbol{u}_k;$$

   is the projection of $\boldsymbol{w}$ onto V.

2. If $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$ is an orthonormal basis for V, then

$$(\boldsymbol{w} \cdot \boldsymbol{v}_1)\boldsymbol{v}_1 + (\boldsymbol{w} \cdot \boldsymbol{v}_2)\boldsymbol{v}_2 + \cdots + (\boldsymbol{w} \cdot \boldsymbol{v}_k)\boldsymbol{v}_k$$

   is the projection of $\boldsymbol{w}$ onto V.

Proof We only prove the first part of the theorem. The second part is a simple consequence of the first part:

Let
$$p = \frac{w \cdot u_1}{u_1 \cdot u_1}u_1 + \frac{w \cdot u_2}{u_2 \cdot u_2}u_2 + \cdots + \frac{w \cdot u_k}{u_k \cdot u_k}u_k \quad \text{and} \quad n = w - p. \tag{5.3}$$
Then for $i = 1, 2, \dots, k$,

$$\begin{aligned} n \cdot u_i &= w \cdot u_i - p \cdot u_i \\ &= w \cdot u_i - \frac{w \cdot u_1}{u_1 \cdot u_1}(u_1 \cdot u_i) - \frac{w \cdot u_2}{u_2 \cdot u_2}(u_2 \cdot u_i) - \cdots - \frac{w \cdot u_k}{u_k \cdot u_k}(u_k \cdot u_i) \\ &= w \cdot u_i - \frac{w \cdot u_i}{u_i \cdot u_i}(u_i \cdot u_i) = 0. \end{aligned}$$

By Remark 5.2.12, n is orthogonal to V. Since $w = n + p$ where n is orthogonal to V and p is a vector in V, p is the projection of w onto V.

Example 5.2.16 Let V be a subspace of $\mathbb{R}^3$ spanned by the orthogonal vectors $u_1 = (1, 0, 1)$ and $u_2 = (1, 0, -1)$. Then the projection of $w = (1, 1, 0)$ onto V is equal to

$$\frac{w \cdot u_1}{u_1 \cdot u_1}u_1 + \frac{w \cdot u_2}{u_2 \cdot u_2}u_2 = \frac{1}{2}(1, 0, 1) + \frac{1}{2}(1, 0, -1) = (1, 0, 0).$$

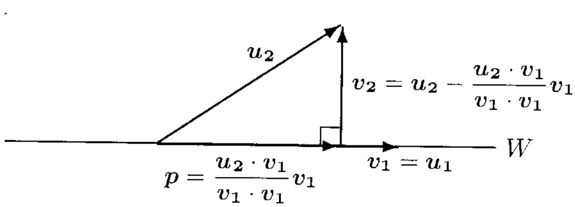
Remark 5.2.17 Theorem 5.2.8 can be regarded as a particular case of Theorem 5.2.15 when w is contained in V, i.e. $w = p$ and $n = 0$ in formulae (5.3).

Discussion 5.2.18

1. Let $\{u_1, u_2\}$ be a basis for a vector space V where V is either $\mathbb{R}^2$ or a plane in $\mathbb{R}^3$ containing the origin. Let W be the subspace of V spanned by u_1. (W is a line through the origin.) By Theorem 5.2.15, the projection of u_2 onto W is

$$p = \frac{u_2 \cdot u_1}{u_1 \cdot u_1}u_1.$$

   Define $v_1 = u_1$ and $v_2 = u_2 - p = u_2 - \dfrac{u_2 \cdot v_1}{v_1 \cdot v_1}v_1$.



   Then $\{v_1, v_2\}$ is an orthogonal basis for V.

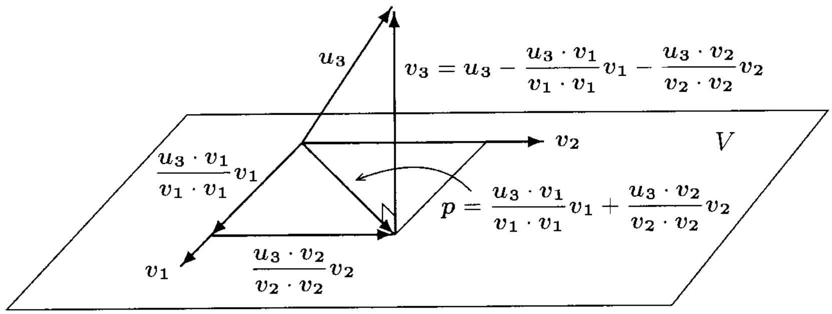
2. Let $\{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\}$ be a basis for $\mathbb{R}^3$ and let V be the subspace of $\mathbb{R}^3$ spanned by $\boldsymbol{u_1}$, $\boldsymbol{u_2}$. (V is a plane containing the origin.) Define

$$\boldsymbol{v_1} = \boldsymbol{u_1} \quad \text{and} \quad \boldsymbol{v_2} = \boldsymbol{u_2} - \frac{\boldsymbol{u_2} \cdot \boldsymbol{v_1}}{\boldsymbol{v_1} \cdot \boldsymbol{v_1}} \boldsymbol{v_1}.$$

By Part 1, $\{\boldsymbol{v_1}, \boldsymbol{v_2}\}$ is an orthogonal basis for V. By Theorem 5.2.15, the projection of $\boldsymbol{u_3}$ onto V is

$$\boldsymbol{p} = \frac{\boldsymbol{u_3} \cdot \boldsymbol{v_1}}{\boldsymbol{v_1} \cdot \boldsymbol{v_1}} \boldsymbol{v_1} + \frac{\boldsymbol{u_3} \cdot \boldsymbol{v_2}}{\boldsymbol{v_2} \cdot \boldsymbol{v_2}} \boldsymbol{v_2}.$$

Define $\boldsymbol{v_3} = \boldsymbol{u_3} - \boldsymbol{p} = \boldsymbol{u_3} - \frac{\boldsymbol{u_3} \cdot \boldsymbol{v_1}}{\boldsymbol{v_1} \cdot \boldsymbol{v_1}} \boldsymbol{v_1} - \frac{\boldsymbol{u_3} \cdot \boldsymbol{v_2}}{\boldsymbol{v_2} \cdot \boldsymbol{v_2}} \boldsymbol{v_2}.$



Then $\{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}\}$ is an orthogonal basis for $\mathbb{R}^3$.

Theorem 5.2.19 (Gram-Schmidt Process) Let $\{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}\}$ be a basis for a vector space V. Let

$$\begin{aligned}
\boldsymbol{v_1} &= \boldsymbol{u_1}, \\
\boldsymbol{v_2} &= \boldsymbol{u_2} - \frac{\boldsymbol{u_2} \cdot \boldsymbol{v_1}}{\boldsymbol{v_1} \cdot \boldsymbol{v_1}} \boldsymbol{v_1}, \\
\boldsymbol{v_3} &= \boldsymbol{u_3} - \frac{\boldsymbol{u_3} \cdot \boldsymbol{v_1}}{\boldsymbol{v_1} \cdot \boldsymbol{v_1}} \boldsymbol{v_1} - \frac{\boldsymbol{u_3} \cdot \boldsymbol{v_2}}{\boldsymbol{v_2} \cdot \boldsymbol{v_2}} \boldsymbol{v_2}, \\
&\vdots \\
\boldsymbol{v_k} &= \boldsymbol{u_k} - \frac{\boldsymbol{u_k} \cdot \boldsymbol{v_1}}{\boldsymbol{v_1} \cdot \boldsymbol{v_1}} \boldsymbol{v_1} - \frac{\boldsymbol{u_k} \cdot \boldsymbol{v_2}}{\boldsymbol{v_2} \cdot \boldsymbol{v_2}} \boldsymbol{v_2} - \cdots - \frac{\boldsymbol{u_k} \cdot \boldsymbol{v_{k-1}}}{\boldsymbol{v_k} \cdot \boldsymbol{v_k}} \boldsymbol{v_{k-1}}.
\end{aligned}$$

Then $\{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_k}\}$ is an orthogonal basis for V. Furthermore, let

$$\boldsymbol{w_1} = \frac{1}{||\boldsymbol{v_1}||} \boldsymbol{v_1}, \quad \boldsymbol{w_2} = \frac{1}{||\boldsymbol{v_2}||} \boldsymbol{v_2}, \quad \ldots, \quad \boldsymbol{w_k} = \frac{1}{||\boldsymbol{v_k}||} \boldsymbol{v_k}.$$

Then $\{\boldsymbol{w_1}, \boldsymbol{w_2}, \ldots, \boldsymbol{w_k}\}$ is an orthonormal basis for V.

Example 5.2.20 Apply the Gram-Schmidt Process to transform the basis $\{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\}$, where $\boldsymbol{u_1} = (1,-1,2)$, $\boldsymbol{u_2} = (2,1,0)$ and $\boldsymbol{u_3} = (0,0,1)$, for $\mathbb{R}^3$ into an orthonormal basis.

Solution Let

$$\begin{aligned}
\boldsymbol{v_1} &= \boldsymbol{u_1} = (1,-1,2),\\
\boldsymbol{v_2} &= \boldsymbol{u_2} - \frac{\boldsymbol{u_2}\cdot\boldsymbol{v_1}}{\boldsymbol{v_1}\cdot\boldsymbol{v_1}}\boldsymbol{v_1}\\
&= (2,1,0) - \frac{1}{6}(1,-1,2) = \left(\frac{11}{6}, \frac{7}{6}, -\frac{1}{3}\right),\\
\boldsymbol{v_3} &= \boldsymbol{u_3} - \frac{\boldsymbol{u_3}\cdot\boldsymbol{v_1}}{\boldsymbol{v_1}\cdot\boldsymbol{v_1}}\boldsymbol{v_1} - \frac{\boldsymbol{u_3}\cdot\boldsymbol{v_2}}{\boldsymbol{v_2}\cdot\boldsymbol{v_2}}\boldsymbol{v_2}\\
&= (0,0,1) - \frac{2}{6}(1,-1,2) - \frac{-1/3}{29/6}\left(\frac{11}{6}, \frac{7}{6}, -\frac{1}{3}\right) = \left(-\frac{6}{29}, \frac{12}{29}, \frac{9}{29}\right).
\end{aligned}$$

Then

$$\begin{aligned}
&\left\{\frac{1}{||\boldsymbol{v_1}||}\boldsymbol{v_1}, \frac{1}{||\boldsymbol{v_2}||}\boldsymbol{v_2}, \frac{1}{||\boldsymbol{v_3}||}\boldsymbol{v_3}\right\}\\
&= \left\{\frac{1}{\sqrt{6}}(1,-1,2), \frac{1}{\sqrt{29/6}}\left(\frac{11}{6}, \frac{7}{6}, -\frac{1}{3}\right), \frac{1}{\sqrt{9/29}}\left(-\frac{6}{29}, \frac{12}{29}, \frac{9}{29}\right)\right\}\\
&= \left\{\left(\frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}\right), \left(\frac{11}{\sqrt{174}}, \frac{7}{\sqrt{174}}, -\frac{2}{\sqrt{174}}\right), \left(-\frac{2}{\sqrt{29}}, \frac{4}{\sqrt{29}}, \frac{3}{\sqrt{29}}\right)\right\}
\end{aligned}$$

is an orthonormal basis for $\mathbb{R}^3$.

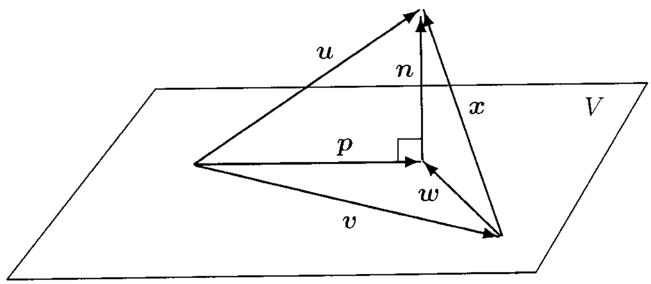
Section 5.3 Best Approximations

Discussion 5.3.1 One of the most important applications of the concept of orthogonality is in the study of approximations. For example, in the more advance theory of linear algebra, functions are regarded as vectors and orthogonal functions are used to compute the approximate values of other functions. The following theorem is the foundation for finding the best approximation. Although the theorem is stated in Euclidean n-spaces, it can be generalized to other "abstract" vector spaces.

Theorem 5.3.2 Let V be a subspace in $\mathbb{R}^n$. If $\boldsymbol{u}$ is a vector in $\mathbb{R}^n$ and $\boldsymbol{p}$ is the projection of $\boldsymbol{u}$ onto V, then

$$d(\boldsymbol{u}, \boldsymbol{p}) \le d(\boldsymbol{u}, \boldsymbol{v}) \quad \text{for all } \boldsymbol{v} \in V,$$

i.e. $\boldsymbol{p}$ is the *best approximation* of $\boldsymbol{u}$ in V.



Proof Take any vector $\boldsymbol{v}$ in V. Define

$$\boldsymbol{n} = \boldsymbol{u} - \boldsymbol{p}, \quad \boldsymbol{w} = \boldsymbol{p} - \boldsymbol{v} \quad \text{and} \quad \boldsymbol{x} = \boldsymbol{u} - \boldsymbol{v}.$$

Observe that

(a) $\boldsymbol{x} = \boldsymbol{n} + \boldsymbol{w}$; and

(b) since $\boldsymbol{n}$ is orthogonal to V and $\boldsymbol{w}$ is a vector in V, the vectors $\boldsymbol{n}$ and $\boldsymbol{w}$ are orthogonal, i.e. $\boldsymbol{n} \cdot \boldsymbol{w} = 0$.

Thus

$$\begin{aligned} ||\boldsymbol{x}||^2 = \boldsymbol{x} \cdot \boldsymbol{x} = (\boldsymbol{n} + \boldsymbol{w}) \cdot (\boldsymbol{n} + \boldsymbol{w}) &= \boldsymbol{n} \cdot \boldsymbol{n} + \boldsymbol{n} \cdot \boldsymbol{w} + \boldsymbol{w} \cdot \boldsymbol{n} + \boldsymbol{w} \cdot \boldsymbol{w} \\ &= \boldsymbol{n} \cdot \boldsymbol{n} + \boldsymbol{w} \cdot \boldsymbol{w} = ||\boldsymbol{n}||^2 + ||\boldsymbol{w}||^2 \\ &\geq ||\boldsymbol{n}||^2. \end{aligned}$$

Since $d(\boldsymbol{u}, \boldsymbol{p}) = ||\boldsymbol{u} - \boldsymbol{p}|| = ||\boldsymbol{n}||$ and $d(\boldsymbol{u}, \boldsymbol{v}) = ||\boldsymbol{u} - \boldsymbol{v}|| = ||\boldsymbol{x}||$, we have $d(\boldsymbol{u}, \boldsymbol{p}) \leq d(\boldsymbol{u}, \boldsymbol{v})$.

Example 5.3.3 Let $V = \text{span}\{(1, 0, 1), (1, 1, 1)\}$ which is a plane in $\mathbb{R}^3$ containing the origin. Find the (shortest) distance from $\boldsymbol{u} = (1, 2, 3)$ to V.

Solution By Theorem 5.3.2, the shortest distance is $d(\boldsymbol{u}, \boldsymbol{p})$ where $\boldsymbol{p}$ is the projection of $\boldsymbol{u}$ onto V. Applying the Gram-Schmidt Process, the vectors

$$(1, 0, 1) \quad \text{and} \quad (1, 1, 1) - \frac{(1, 1, 1) \cdot (1, 0, 1)}{(1, 0, 1) \cdot (1, 0, 1)}(1, 0, 1) = (0, 1, 0)$$

forms an orthogonal basis for V. Thus by Theorem 5.2.15,

$$\boldsymbol{p} = \frac{(1, 2, 3) \cdot (1, 0, 1)}{(1, 0, 1) \cdot (1, 0, 1)}(1, 0, 1) + \frac{(1, 2, 3) \cdot (0, 1, 0)}{(0, 1, 0) \cdot (0, 1, 0)}(0, 1, 0) = (2, 2, 2)$$

and hence the distance from $\boldsymbol{u}$ to V is $d(\boldsymbol{u}, \boldsymbol{p}) = ||\boldsymbol{u} - \boldsymbol{p}|| = ||(-1, 0, 1)|| = \sqrt{2}$.

Discussion 5.3.4 Using Theorem 5.3.2, we shall derive a useful mathematical tool for experimental scientists called the "least squares method". In analyzing experimental results, scientists always face a problem of fitting experimental data to an equation. We illustrate the situation using the following example.

Example 5.3.5 Suppose r, s and t are physical quantities that satisfy the rule

$$t = cr^2 + ds + e$$

for some constants c, d and e. An experiment was conducted in order to find the constants c, d and e. In the experiment, six measurements of t were taken with various settings for values of r and s.

i	1	2	3	4	5	6	
r_i	0	0	1	1	2	2	
s_i	0	1	2	0	1	2	
t_i	0.5	1.6	2.8	0.8	5.1	5.9	$\leftarrow$ Experimental values

If there are no experimental errors, we have

$$\begin{cases} cr_1^2 + ds_1 + e = t_1 \\ cr_2^2 + ds_2 + e = t_2 \\ \qquad \vdots \\ cr_6^2 + ds_6 + e = t_6 \end{cases} \quad \Leftrightarrow \quad \begin{pmatrix} r_1^2 & s_1 & 1 \\ r_2^2 & s_2 & 1 \\ \vdots & \vdots & \vdots \\ r_6^2 & s_6 & 1 \end{pmatrix} \begin{pmatrix} c \\ d \\ e \end{pmatrix} = \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_6 \end{pmatrix}.$$

Let $\boldsymbol{A} = \begin{pmatrix} r_1^2 & s_1 & 1 \\ r_2^2 & s_2 & 1 \\ \vdots & \vdots & \vdots \\ r_6^2 & s_6 & 1 \end{pmatrix}$, $\boldsymbol{x} = \begin{pmatrix} c \\ d \\ e \end{pmatrix}$ and $\boldsymbol{b} = \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_6 \end{pmatrix}$. By solving the linear system $\boldsymbol{Ax} = \boldsymbol{b}$, we can obtain the values of c, d and e. However, due to experimental errors, we do not expect to get the exact values of t_i's and the system $\boldsymbol{Ax} = \boldsymbol{b}$ is usually inconsistent. Hence we cannot obtain the values of c, d and e directly.

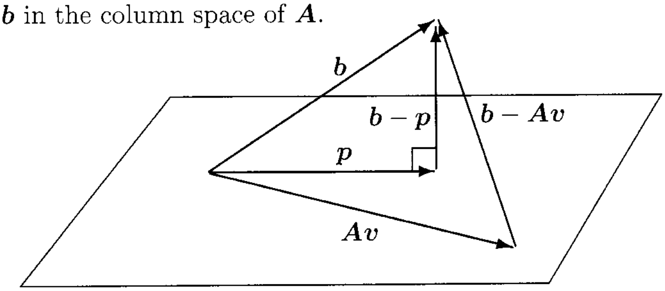
The usual scheme is to get the approximate values of c, d and e that minimize the *sum of squares of errors* (SSE):

$$\begin{aligned} &\sum_{i=1}^{6} \left[t_i - (cr_i^2 + ds_i + e)\right]^2 \\ &= \left[t_1 - (cr_1^2 + ds_1 + e)\right]^2 + \left[t_2 - (cr_2^2 + ds_2 + e)\right]^2 + \cdots + \left[t_6 - (cr_6^2 + ds_6 + e)\right]^2 \\ &= ||\boldsymbol{b} - \boldsymbol{Ax}||^2. \end{aligned}$$

That is, we need to find $\boldsymbol{x}$ that minimizes the value of $||\boldsymbol{b} - \boldsymbol{Ax}||$. We shall find this value of $\boldsymbol{x}$ in Example 5.3.11.2 after introducing the concept of least squares solutions.

Definition 5.3.6 Let $\boldsymbol{Ax} = \boldsymbol{b}$ be a linear system where $\boldsymbol{A}$ is an $m \times n$ matrix. A vector $\boldsymbol{u} \in \mathbb{R}^n$ is called the *least squares solution* to the linear system if $||\boldsymbol{b} - \boldsymbol{Au}|| \leq ||\boldsymbol{b} - \boldsymbol{Av}||$ for all $\boldsymbol{v} \in \mathbb{R}^n$.

Discussion 5.3.7 By Theorem 4.1.16, $\{ \boldsymbol{Av} \mid \boldsymbol{v} \in \mathbb{R}^n \}$ is equal to the column space of $\boldsymbol{A}$. Thus to find the least squares solution to $\boldsymbol{Ax} = \boldsymbol{b}$, we first need to find the best approximation of $\boldsymbol{b}$ in the column space of $\boldsymbol{A}$.



the column space of $\boldsymbol{A}$

Theorem 5.3.8 Let $\boldsymbol{Ax} = \boldsymbol{b}$ be a linear system, where $\boldsymbol{A}$ is an $m \times n$ matrix, and let $\boldsymbol{p}$ be the projection of $\boldsymbol{b}$ onto the column space of $\boldsymbol{A}$. Then

$$||\boldsymbol{b}-\boldsymbol{p}|| \leq ||\boldsymbol{b}-\boldsymbol{Av}|| \quad \text{for all } \boldsymbol{v} \in \mathbb{R}^n,$$

i.e. $\boldsymbol{u}$ is a least squares solution to $\boldsymbol{Ax} = \boldsymbol{b}$ if and only if $\boldsymbol{Au} = \boldsymbol{p}$.

Proof Let V be the column space of $\boldsymbol{A}$. By Theorem 5.3.2,

$$||\boldsymbol{b}-\boldsymbol{p}|| = d(\boldsymbol{b},\boldsymbol{p}) \leq d(\boldsymbol{b},\boldsymbol{w}) = ||\boldsymbol{b}-\boldsymbol{w}|| \quad \text{for all } \boldsymbol{w} \in V.$$

Since $V = \{ \boldsymbol{Av} \mid \boldsymbol{v} \in \mathbb{R}^n \}$, $||\boldsymbol{b}-\boldsymbol{p}|| \leq ||\boldsymbol{b}-\boldsymbol{Av}||$ for all $\boldsymbol{v} \in \mathbb{R}^n$.

Example 5.3.9 Let $\boldsymbol{A} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}$, $\boldsymbol{b} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ and

$$V = \text{the column space of } \boldsymbol{A} = \text{span}\left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}.$$

By Example 5.3.3, the projection of $\boldsymbol{b}$ onto V is $\boldsymbol{p} = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}$. By Theorem 5.3.8, $\begin{pmatrix} x \\ y \end{pmatrix}$ is a least squares solution to $\boldsymbol{Ax} = \boldsymbol{b}$ if and only if

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix} \quad \Leftrightarrow \quad \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}.$$

Theorem 5.3.10 (A method to find the least squares solution) Let $\boldsymbol{Ax} = \boldsymbol{b}$ be a linear system. Then $\boldsymbol{u}$ is a least squares solution to $\boldsymbol{Ax} = \boldsymbol{b}$ if and only if $\boldsymbol{u}$ is a solution to $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$.

Proof Let $\boldsymbol{A} = \begin{pmatrix}\boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n\end{pmatrix}$, where $\boldsymbol{a}_i$ is the ith column of $\boldsymbol{A}$, and let V be the column space of $\boldsymbol{A}$, i.e. $V = \text{span}\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n\} = \{\, \boldsymbol{A}\boldsymbol{v} \mid \boldsymbol{v} \in \mathbb{R}^n \,\}$. Then

$$
\begin{aligned}
& \boldsymbol{u} \text{ is the least squares solution to } \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\
\Leftrightarrow\ & \boldsymbol{A}\boldsymbol{u} \text{ is the projection of } \boldsymbol{b} \text{ onto } V && \text{(by Theorem 5.3.8)} \\
\Leftrightarrow\ & \boldsymbol{b} - \boldsymbol{A}\boldsymbol{u} \text{ is orthogonal to } V && \text{(by Definition 5.2.13)} \\
\Leftrightarrow\ & \boldsymbol{b} - \boldsymbol{A}\boldsymbol{u} \text{ is orthogonal to } \boldsymbol{a}_1,\ \boldsymbol{a}_2,\ \ldots,\ \boldsymbol{a}_n && \text{(by Remark 5.2.12)} \\
\Leftrightarrow\ & \boldsymbol{a}_1 \cdot (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{u}) = 0,\quad \boldsymbol{a}_2 \cdot (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{u}) = 0,\quad \ldots,\quad \boldsymbol{a}_n \cdot (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{u}) = 0 \\
\Leftrightarrow\ & \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{u}) = \boldsymbol{0} && \left(\text{becasue } \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{u}) = \begin{pmatrix} \boldsymbol{a}_1 \cdot (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{u}) \\ \boldsymbol{a}_2 \cdot (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{u}) \\ \vdots \\ \boldsymbol{a}_n \cdot (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{u}) \end{pmatrix}\right) \\
\Leftrightarrow\ & \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{u} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}.
\end{aligned}
$$

Example 5.3.11

1. Use $\boldsymbol{A}$ and $\boldsymbol{b}$ in Example 5.3.9. By Theorem 5.3.10, to find the least square solution to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, we do not need to compute the projection $\boldsymbol{p}$. For this case, the equation $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ is
$$\begin{pmatrix} 2 & 2 \\ 2 & 3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 4 \\ 6 \end{pmatrix}.$$
Solving this linear system, we obtain the least squares solution $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$.

2. For Example 5.3.5, the linear system is
$$\left\{\begin{aligned} e &= 0.5 \\ d + e &= 1.6 \\ c + 2d + e &= 2.8 \\ c \qquad + e &= 0.8 \\ 4c + \ d + e &= 5.1 \\ 4c + 2d + e &= 5.9 \end{aligned}\right. \quad\Leftrightarrow\quad \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 0 & 1 \\ 4 & 1 & 1 \\ 4 & 2 & 1 \end{pmatrix}\begin{pmatrix} c \\ d \\ e \end{pmatrix} = \begin{pmatrix} 0.5 \\ 1.6 \\ 2.8 \\ 0.8 \\ 5.1 \\ 5.9 \end{pmatrix} \quad\Leftrightarrow\quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$
where
$$\boldsymbol{A} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 0 & 1 \\ 4 & 1 & 1 \\ 4 & 2 & 1 \end{pmatrix}, \quad \boldsymbol{x} = \begin{pmatrix} c \\ d \\ e \end{pmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{pmatrix} 0.5 \\ 1.6 \\ 2.8 \\ 0.8 \\ 5.1 \\ 5.9 \end{pmatrix}.$$

Then the equation $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ is

$$\begin{pmatrix} 34 & 14 & 10 \\ 14 & 10 & 6 \\ 10 & 6 & 6 \end{pmatrix} \begin{pmatrix} c \\ d \\ e \end{pmatrix} = \begin{pmatrix} 47.6 \\ 24.1 \\ 16.7 \end{pmatrix}.$$

Solving this linear system, we obtain the least squares solution $c = 0.9275$, $d = 0.9225$ and $e = 0.3150$.

3. In this example, we demonstrate how to find the projection using a least squares solution:

   Let $V = \mathrm{span}\{(1,-1,1,-1), (1,2,0,1), (2,1,1,0)\}$. Find the projection of $(1,1,1,1)$ onto V.

   Solution Let $\boldsymbol{A} = \begin{pmatrix} 1 & 1 & 2 \\ -1 & 2 & 1 \\ 1 & 0 & 1 \\ -1 & 1 & 0 \end{pmatrix}$ and $\boldsymbol{b} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$. We first obtain a least squares solution of $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$. The equation $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ is

$$\begin{pmatrix} 4 & -2 & 2 \\ -2 & 6 & 4 \\ 2 & 4 & 6 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 4 \\ 4 \end{pmatrix}.$$

   Solving this linear system, we have $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -t + \frac{2}{5} \\ -t + \frac{4}{5} \\ t \end{pmatrix}$ where t is an arbitrary parameter. Take any one of the least squares solutions, say $\boldsymbol{u} = \begin{pmatrix} \frac{2}{5} \\ \frac{4}{5} \\ 0 \end{pmatrix}$, and compute $\boldsymbol{A}\boldsymbol{u}$:

$$\begin{pmatrix} 1 & 1 & 2 \\ -1 & 2 & 1 \\ 1 & 0 & 1 \\ -1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \frac{2}{5} \\ \frac{4}{5} \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{6}{5} \\ \frac{6}{5} \\ \frac{2}{5} \\ \frac{2}{5} \end{pmatrix}.$$

   By Theorem 5.3.8, $(\frac{6}{5}, \frac{6}{5}, \frac{2}{5}, \frac{2}{5})$ is the projection of $(1,1,1,1)$ onto V.

Section 5.4 Orthogonal Matrices

Discussion 5.4.1 Since coordinate systems built upon orthonormal bases have a lot of advantages over coordinate systems using other bases, orthonormal bases are frequently used both in theoretic studies and in applications. On the other hand, in the studies of vector spaces, it is quite often that we need to shift between bases. Thus we would like to know more about transition matrices concerning orthonormal bases.

Let $S = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_k}\}$ and $T = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_k}\}$ be two bases for a vector space V. Be reminded that the transition matrix from S to T is the matrix

$$\boldsymbol{P} = \begin{pmatrix} [\boldsymbol{u_1}]_T & [\boldsymbol{u_2}]_T & \cdots & [\boldsymbol{u_k}]_T \end{pmatrix}.$$

Note that $[\boldsymbol{w}]_T = \boldsymbol{P}[\boldsymbol{w}]_S$ for all $\boldsymbol{w} \in V$ (see Discussion 3.7.2).

Example 5.4.2 Let $E = \{\boldsymbol{e_1}, \boldsymbol{e_2}, \boldsymbol{e_3}\}$ be the standard basis for $\mathbb{R}^3$, i.e.

$$\boldsymbol{e_1} = (1,0,0), \quad \boldsymbol{e_2} = (0,1,0) \quad \text{and} \quad \boldsymbol{e_3} = (0,0,1),$$

and let $S = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\}$ where

$$\boldsymbol{u_1} = \frac{1}{\sqrt{3}}(1,1,1), \quad \boldsymbol{u_2} = \frac{1}{\sqrt{2}}(1,0,-1) \quad \text{and} \quad \boldsymbol{u_3} = \frac{1}{\sqrt{6}}(1,-2,1).$$

Both E and S are orthonormal basis for $\mathbb{R}^3$.

Obviously,

$$\begin{aligned} \boldsymbol{u_1} &= \tfrac{1}{\sqrt{3}}\boldsymbol{e_1} + \tfrac{1}{\sqrt{3}}\boldsymbol{e_2} + \tfrac{1}{\sqrt{3}}\boldsymbol{e_3}, \\ \boldsymbol{u_2} &= \tfrac{1}{\sqrt{2}}\boldsymbol{e_1} \qquad\quad - \tfrac{1}{\sqrt{2}}\boldsymbol{e_3}, \\ \boldsymbol{u_3} &= \tfrac{1}{\sqrt{6}}\boldsymbol{e_1} - \tfrac{2}{\sqrt{6}}\boldsymbol{e_2} + \tfrac{1}{\sqrt{6}}\boldsymbol{e_3}. \end{aligned}$$

Thus the transition matrix from S to E is $\boldsymbol{P} = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{pmatrix}$.

As S is an orthonormal basis for $\mathbb{R}^3$, by Theorem 5.2.8,

$$\begin{aligned} \boldsymbol{e_1} &= (\boldsymbol{e_1} \cdot \boldsymbol{u_1})\boldsymbol{u_1} + (\boldsymbol{e_1} \cdot \boldsymbol{u_2})\boldsymbol{u_2} + (\boldsymbol{e_1} \cdot \boldsymbol{u_3})\boldsymbol{u_3}, \\ \boldsymbol{e_2} &= (\boldsymbol{e_2} \cdot \boldsymbol{u_1})\boldsymbol{u_1} + (\boldsymbol{e_2} \cdot \boldsymbol{u_2})\boldsymbol{u_2} + (\boldsymbol{e_2} \cdot \boldsymbol{u_3})\boldsymbol{u_3}, \\ \boldsymbol{e_3} &= (\boldsymbol{e_3} \cdot \boldsymbol{u_1})\boldsymbol{u_1} + (\boldsymbol{e_3} \cdot \boldsymbol{u_2})\boldsymbol{u_2} + (\boldsymbol{e_3} \cdot \boldsymbol{u_3})\boldsymbol{u_3}. \end{aligned}$$

We have

$$\begin{aligned} \boldsymbol{e_1} &= \tfrac{1}{\sqrt{3}}\boldsymbol{u_1} + \tfrac{1}{\sqrt{2}}\boldsymbol{u_2} + \tfrac{1}{\sqrt{6}}\boldsymbol{u_3}, \\ \boldsymbol{e_2} &= \tfrac{1}{\sqrt{3}}\boldsymbol{u_1} \qquad\quad - \tfrac{2}{\sqrt{6}}\boldsymbol{u_3}, \\ \boldsymbol{e_3} &= \tfrac{1}{\sqrt{3}}\boldsymbol{u_1} - \tfrac{1}{\sqrt{2}}\boldsymbol{u_2} + \tfrac{1}{\sqrt{6}}\boldsymbol{u_3}. \end{aligned}$$

Thus the transition matrix from E to S is $\boldsymbol{Q} = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{pmatrix}$.

Note that $\boldsymbol{Q} = \boldsymbol{P}^{\mathrm{T}}$. Furthermore, by Theorem 3.7.5, $\boldsymbol{Q} = \boldsymbol{P}^{-1}$. So $\boldsymbol{P}^{-1} = \boldsymbol{P}^{\mathrm{T}}$

Definition 5.4.3 A square matrix $\boldsymbol{A}$ is called *orthogonal* if $\boldsymbol{A}^{-1} = \boldsymbol{A}^{\mathrm{T}}$. (See also Question 2.12.)

Remark 5.4.4 By Theorem 2.4.12, a square matrix $\boldsymbol{A}$ is orthogonal if and only if $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{I}$ (or $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{I}$).

Example 5.4.5 The following are examples of orthogonal matrices.

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{6}} & 0 \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \end{pmatrix}, \qquad \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}.$$

Theorem 5.4.6 Let $\boldsymbol{A}$ be a square matrix of order n. The following statements are equivalent:

1. $\boldsymbol{A}$ is orthogonal.
2. The rows of $\boldsymbol{A}$ form an orthonormal basis for $\mathbb{R}^n$.
3. The columns of $\boldsymbol{A}$ form an orthonormal basis for $\mathbb{R}^n$.

Proof Let $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_n \end{pmatrix}$ where $\boldsymbol{a}_i$ is the ith row of $\boldsymbol{A}$. By Remark 5.2.6, to show "1 $\Leftrightarrow$ 2", it suffices to show that $\boldsymbol{A}$ is orthogonal if and only if $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$ are orthonormal.

Observe that

$$\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \begin{pmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_n \end{pmatrix} \begin{pmatrix} \boldsymbol{a}_1{}^{\mathrm{T}} & \boldsymbol{a}_2{}^{\mathrm{T}} & \cdots & \boldsymbol{a}_n{}^{\mathrm{T}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{a}_1\boldsymbol{a}_1{}^{\mathrm{T}} & \boldsymbol{a}_1\boldsymbol{a}_2{}^{\mathrm{T}} & \cdots & \boldsymbol{a}_1\boldsymbol{a}_n{}^{\mathrm{T}} \\ \boldsymbol{a}_2\boldsymbol{a}_1{}^{\mathrm{T}} & \boldsymbol{a}_2\boldsymbol{a}_2{}^{\mathrm{T}} & \cdots & \boldsymbol{a}_2\boldsymbol{a}_n{}^{\mathrm{T}} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{a}_n\boldsymbol{a}_1{}^{\mathrm{T}} & \boldsymbol{a}_n\boldsymbol{a}_2{}^{\mathrm{T}} & \cdots & \boldsymbol{a}_n\boldsymbol{a}_n{}^{\mathrm{T}} \end{pmatrix}$$

$$= \begin{pmatrix} \boldsymbol{a}_1 \cdot \boldsymbol{a}_1 & \boldsymbol{a}_1 \cdot \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_1 \cdot \boldsymbol{a}_n \\ \boldsymbol{a}_2 \cdot \boldsymbol{a}_1 & \boldsymbol{a}_2 \cdot \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_2 \cdot \boldsymbol{a}_n \\ \vdots & \vdots & & \vdots \\ \boldsymbol{a}_n \cdot \boldsymbol{a}_1 & \boldsymbol{a}_n \cdot \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \cdot \boldsymbol{a}_n \end{pmatrix}.$$

Thus by Remark 5.4.4,

$$\begin{aligned}
\boldsymbol{A} \text{ is orthogonal} \quad &\Leftrightarrow \quad \boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{I} \\
&\Leftrightarrow \quad \text{for all } i, j, \quad \boldsymbol{a_i} \cdot \boldsymbol{a_j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \\
&\Leftrightarrow \quad \boldsymbol{a_1}, \boldsymbol{a_2}, \dots, \boldsymbol{a_n} \text{ are orthonormal.}
\end{aligned}$$

So we have shown "1 $\Leftrightarrow$ 2". The proof of "1 $\Leftrightarrow$ 3" is similar except we compute $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ instead.

Theorem 5.4.7 Let S and T be two orthonormal bases for a vector space and let $\boldsymbol{P}$ be the transition matrix from S to T. Then $\boldsymbol{P}$ is orthogonal and $\boldsymbol{P}^{\mathrm{T}}$ is the transition matrix from T to S.

Proof Let $S = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \dots, \boldsymbol{u_k}\}$ and $T = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_k}\}$. Since T is orthonormal, by Theorem 5.2.8,

$$\begin{aligned}
\boldsymbol{u_1} &= (\boldsymbol{u_1} \cdot \boldsymbol{v_1})\boldsymbol{v_1} + (\boldsymbol{u_1} \cdot \boldsymbol{v_2})\boldsymbol{v_2} + \cdots + (\boldsymbol{u_1} \cdot \boldsymbol{v_k})\boldsymbol{v_k}, \\
\boldsymbol{u_2} &= (\boldsymbol{u_2} \cdot \boldsymbol{v_1})\boldsymbol{v_1} + (\boldsymbol{u_2} \cdot \boldsymbol{v_2})\boldsymbol{v_2} + \cdots + (\boldsymbol{u_2} \cdot \boldsymbol{v_k})\boldsymbol{v_k}, \\
&\vdots \\
\boldsymbol{u_k} &= (\boldsymbol{u_k} \cdot \boldsymbol{v_1})\boldsymbol{v_1} + (\boldsymbol{u_k} \cdot \boldsymbol{v_2})\boldsymbol{v_2} + \cdots + (\boldsymbol{u_k} \cdot \boldsymbol{v_k})\boldsymbol{v_k}.
\end{aligned}$$

Thus the transition matrix from S to T is $\boldsymbol{P} = \begin{pmatrix} \boldsymbol{u_1} \cdot \boldsymbol{v_1} & \boldsymbol{u_2} \cdot \boldsymbol{v_1} & \cdots & \boldsymbol{u_k} \cdot \boldsymbol{v_1} \\ \boldsymbol{u_1} \cdot \boldsymbol{v_2} & \boldsymbol{u_2} \cdot \boldsymbol{v_2} & \cdots & \boldsymbol{u_k} \cdot \boldsymbol{v_2} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{u_1} \cdot \boldsymbol{v_k} & \boldsymbol{u_2} \cdot \boldsymbol{v_k} & \cdots & \boldsymbol{u_k} \cdot \boldsymbol{v_k} \end{pmatrix}$.

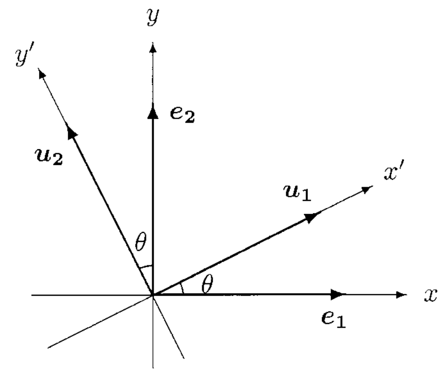
Similarly, the transition matrix from T to S is $\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{v_1} \cdot \boldsymbol{u_1} & \boldsymbol{v_2} \cdot \boldsymbol{u_1} & \cdots & \boldsymbol{v_k} \cdot \boldsymbol{u_1} \\ \boldsymbol{v_1} \cdot \boldsymbol{u_2} & \boldsymbol{v_2} \cdot \boldsymbol{u_2} & \cdots & \boldsymbol{v_k} \cdot \boldsymbol{u_2} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{v_1} \cdot \boldsymbol{u_k} & \boldsymbol{v_2} \cdot \boldsymbol{u_k} & \cdots & \boldsymbol{v_k} \cdot \boldsymbol{u_k} \end{pmatrix}$.

Since $\boldsymbol{u_i} \cdot \boldsymbol{v_j} = \boldsymbol{v_j} \cdot \boldsymbol{u_i}$ for all i, j, we have $\boldsymbol{Q} = \boldsymbol{P}^{\mathrm{T}}$. By Theorem 3.7.5, $\boldsymbol{Q} = \boldsymbol{P}^{-1}$. Thus $\boldsymbol{P}^{-1} = \boldsymbol{P}^{\mathrm{T}}$ and $\boldsymbol{P}$ is orthogonal.

Example 5.4.8

1. **(Rotation of xy-coordinates)** Let $E = \{\boldsymbol{e_1}, \boldsymbol{e_2}\}$ be the standard basis for $\mathbb{R}^2$, i.e. $\boldsymbol{e_1} = (1, 0)$ and $\boldsymbol{e_2} = (0, 1)$. Note that $\boldsymbol{e_1}$ is in the same direction as the x-axis and $\boldsymbol{e_2}$ in the same direction as the y-axis.

   Consider a new $x'y'$-coordinate system obtained by rotating the original xy-coordinates anti-clockwise about the origin through an angle θ.



Let $\boldsymbol{u_1}$ and $\boldsymbol{u_2}$ be unit vectors such that $\boldsymbol{u_1}$ is in the direction of the x'-axis and $\boldsymbol{u_2}$ in the direction of the y'-axis. It is obvious that $S = \{\boldsymbol{u_1}, \boldsymbol{u_2}\}$ is an orthonormal basis for $\mathbb{R}^2$. Note that

$$\begin{aligned}\boldsymbol{u_1} = (\cos(\theta), \sin(\theta)) &= \cos(\theta)\boldsymbol{e_1} + \sin(\theta)\boldsymbol{e_2},\\ \boldsymbol{u_2} = (-\sin(\theta), \cos(\theta)) &= -\sin(\theta)\boldsymbol{e_1} + \cos(\theta)\boldsymbol{e_2}.\end{aligned}$$

Thus the transition matrix from S to E is $\boldsymbol{P} = \begin{pmatrix}\cos(\theta) & -\sin(\theta)\\ \sin(\theta) & \cos(\theta)\end{pmatrix}$.

By Theorem 5.4.7, the transition matrix from E to S is $\boldsymbol{P}^{\mathrm{T}} = \begin{pmatrix}\cos(\theta) & \sin(\theta)\\ -\sin(\theta) & \cos(\theta)\end{pmatrix}$.

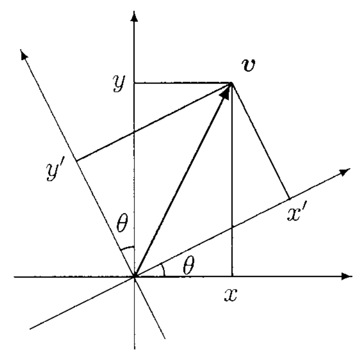
Let $\boldsymbol{v} = (x, y)$ be a vector in $\mathbb{R}^2$ and let $(\boldsymbol{v})_T = (x', y')$. In here, (x', y') can be regarded as the coordinates of $\boldsymbol{v}$ using the new $x'y'$-coordinate system.

Since $\begin{pmatrix}x\\ y\end{pmatrix} = [\boldsymbol{v}]_S = \boldsymbol{P}[\boldsymbol{v}]_T = \boldsymbol{P}\begin{pmatrix}x'\\ y'\end{pmatrix}$, we have

$$\begin{pmatrix}x'\\ y'\end{pmatrix} = \boldsymbol{P}^{\mathrm{T}}\begin{pmatrix}x\\ y\end{pmatrix} = \begin{pmatrix}\cos(\theta) & \sin(\theta)\\ -\sin(\theta) & \cos(\theta)\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix},$$

i.e.

$$\begin{cases} x' = x\cos(\theta) + y\sin(\theta)\\ y' = -x\sin(\theta) + y\cos(\theta).\end{cases}$$



2. Let $S = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\}$ and $T = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}\}$ where

$$\boldsymbol{u_1} = \frac{1}{\sqrt{3}}(1, 1, 1), \quad \boldsymbol{u_2} = \frac{1}{\sqrt{2}}(1, 0, -1), \quad \boldsymbol{u_3} = \frac{1}{\sqrt{6}}(1, -2, 1)$$

and

$$\boldsymbol{v_1} = (0, 0, 1), \quad \boldsymbol{v_2} = \frac{1}{\sqrt{2}}(1, -1, 0), \quad \boldsymbol{v_3} = \frac{1}{\sqrt{2}}(1, 1, 0).$$

Both S and T are orthonormal bases for $\mathbb{R}^3$.

For $i = 1, 2, 3$,

$$u_i = (u_i \cdot v_1)v_1 + (u_i \cdot v_2)v_2 + (u_i \cdot v_3)v_3$$

and hence

$$\begin{aligned} u_1 &= \tfrac{1}{\sqrt{3}}v_1 && + \tfrac{2}{\sqrt{6}}v_3, \\ u_2 &= -\tfrac{1}{\sqrt{2}}v_1 + \tfrac{1}{2}v_2 + && \tfrac{1}{2}v_3, \\ u_3 &= \tfrac{1}{\sqrt{6}}v_1 + \tfrac{3}{\sqrt{12}}v_2 - && \tfrac{1}{\sqrt{12}}v_3. \end{aligned}$$

Thus the transition matrix from S to T is $P = \begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{1}{2} & \frac{3}{\sqrt{12}} \\ \frac{2}{\sqrt{6}} & \frac{1}{2} & -\frac{1}{\sqrt{12}} \end{pmatrix}$.

By Theorem 5.4.7, the transition matrix from T to S is $P^{\mathrm{T}} = \begin{pmatrix} \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{\sqrt{6}} & \frac{3}{\sqrt{12}} & -\frac{1}{\sqrt{12}} \end{pmatrix}$.

Exercise 5

Question 5.1 to Question 5.20 are exercises for Sections 5.1 and 5.2.

1. For each of the following, find $||u||$, $||v||$, $d(u, v)$, $u \cdot v$ and the angle between u and v.

   (a) $u = (2, 3)$ and $v = (1, 1)$.

   (b) $u = (1, -1)$ and $v = (-1, 3)$.

   (c) $u = (1, 2, 3)$ and $v = (0, -3, 2)$.

   (d) $u = (1, -1, 1, -1)$ and $v = (2, 1, 1, 2)$.

2. Consider a triangle in $\mathbb{R}^4$ with vertices $A = (1, 1, 0, 0)$, $B = (1, -1, 0, 0)$ and $C = (2, 0, 0, 1)$.

   (a) Find the lengths of the sides of the triangle.

   (b) Find the angle between AB and AC.

   (c) Verify the cosine rule: $2|AB|\,|AC|\cos(\theta) = |AB|^2 + |AC|^2 - |BC|^2$, where θ is the angle between AB and AC.

3. Complete the proof of Theorem 5.1.5:

   Let $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ be vectors in $\mathbb{R}^n$ and c a scalar. Show that

   (a) $\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{v} \cdot \boldsymbol{u}$;

   (b) $(\boldsymbol{u} + \boldsymbol{v}) \cdot \boldsymbol{w} = \boldsymbol{u} \cdot \boldsymbol{w} + \boldsymbol{v} \cdot \boldsymbol{w}$ and $\boldsymbol{w} \cdot (\boldsymbol{u} + \boldsymbol{v}) = \boldsymbol{w} \cdot \boldsymbol{u} + \boldsymbol{w} \cdot \boldsymbol{v}$;

   (c) $(c\boldsymbol{u}) \cdot \boldsymbol{v} = \boldsymbol{u} \cdot (c\boldsymbol{v}) = c(\boldsymbol{u} \cdot \boldsymbol{v})$;

   (d) $||c\boldsymbol{u}|| = |c|\,||\boldsymbol{u}||$.

4. Let $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{w}$ be any three vectors in $\mathbb{R}^n$. Prove the following inequalities.

   (a) $|\boldsymbol{u} \cdot \boldsymbol{v}| \leq ||\boldsymbol{u}||\,||\boldsymbol{v}||$ (the *Cauchy-Schwarz Inequality*).

   (b) $||\boldsymbol{u} + \boldsymbol{v}|| \leq ||\boldsymbol{u}|| + ||\boldsymbol{v}||$ (the *Triangle Inequality*).

   (c) $d(\boldsymbol{u}, \boldsymbol{w}) \leq d(\boldsymbol{u}, \boldsymbol{v}) + d(\boldsymbol{v}, \boldsymbol{w})$.

   Interpret the result in Part (b) geometrically in $\mathbb{R}^2$.

5. Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be any two vectors in $\mathbb{R}^n$. Prove the following equalities.

   (a) $||\boldsymbol{u} + \boldsymbol{v}||^2 + ||\boldsymbol{u} - \boldsymbol{v}||^2 = 2||\boldsymbol{u}||^2 + 2||\boldsymbol{v}||^2$.

   (b) $\boldsymbol{u} \cdot \boldsymbol{v} = \frac{1}{4}||\boldsymbol{u} + \boldsymbol{v}||^2 - \frac{1}{4}||\boldsymbol{u} - \boldsymbol{v}||^2$.

   Interpret the result in Part (a) geometrically in $\mathbb{R}^2$.

6. For each of the following vectors, find all vectors that are orthogonal to it.

   (a) $(1, 1)$, (b) $(1, 0, 3)$, (c) $(1, -1, 1, -1)$.

   Interpret the results in Part (a) and (b) geometrically.

7. Let W be a subspace of $\mathbb{R}^n$. Define $W^\perp = \{\, \boldsymbol{u} \in \mathbb{R}^n \mid \boldsymbol{u}$ is orthogonal to $W \,\}$.

   (a) Let $W = \text{span}\{(1, 0, 1, 1), (1, -1, 0, 2), (1, 2, 3, -1)\}$. Find $W^\perp$.

   (b) Show that $W^\perp$ is a subspace of $\mathbb{R}^n$. (Hint: Show that $W^\perp$ is a solution set of a homogeneous system of linear equations.)

8. Let $S = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\}$, where $\boldsymbol{u_1}$, $\boldsymbol{u_2}$, $\boldsymbol{u_3}$ are vectors in $\mathbb{R}^3$, and let $T = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}\}$ where $\boldsymbol{v_1} = \frac{3}{5}\boldsymbol{u_2} + \frac{4}{5}\boldsymbol{u_3}$, $\boldsymbol{v_2} = \frac{4}{5}\boldsymbol{u_2} - \frac{3}{5}\boldsymbol{u_3}$ and $\boldsymbol{v_3} = \boldsymbol{u_1}$.

   (a) Show that $\text{span}(S) = \text{span}(T)$.

   (b) If S is orthonormal, show that T is also orthonormal.

9. Let $\{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_n}\}$ be an orthogonal set of vectors in a vector space. Show that

$$||\boldsymbol{u_1} + \boldsymbol{u_2} + \cdots + \boldsymbol{u_n}||^2 = ||\boldsymbol{u_1}||^2 + ||\boldsymbol{u_2}||^2 + \cdots + ||\boldsymbol{u_n}||^2.$$

For $n = 2$, interpret the result geometrically in $\mathbb{R}^2$.

10. Let $\boldsymbol{u_1} = (1, 2, 2, -1)$, $\boldsymbol{u_2} = (1, 1, -1, 1)$, $\boldsymbol{u_3} = (-1, 1, -1, -1)$, $\boldsymbol{u_4} = (-2, 1, 1, 2)$.

    (a) Show that $S = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}, \boldsymbol{u_4}\}$ is an orthogonal set.

    (b) Obtain an orthonormal set S' by normalizing $\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}, \boldsymbol{u_4}$.

    (c) Is S' an orthonormal basis for $\mathbb{R}^4$?

    (d) If $\boldsymbol{w} = (0, 1, 2, 3)$, find $(\boldsymbol{w})_S$ and $(\boldsymbol{w})_{S'}$.

    (e) Let $V = \text{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\}$. Find all vectors that are orthogonal to V.

    (f) Find the projection of $\boldsymbol{w}$ onto V.

11. Let $\boldsymbol{u_1} = (-2, -4, 1)$, $\boldsymbol{u_2} = (3, -1, 2)$ and $\boldsymbol{u_3} = (1, -1, -2)$.

    (a) Show that $\{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\}$ is an orthogonal basis for $\mathbb{R}^3$.

    (b) Let $V = \text{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}\}$ and $W = \text{span}\{\boldsymbol{u_3}\}$. Write each of the following vectors as a sum of two vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ such that $\boldsymbol{v} \in V$ and $\boldsymbol{w} \in W$:

        (i) $(0, 0, 1)$, (ii) $(1, 1, 0)$.

12. Use Gram-Schmidt Process to transform each of the following bases for $\mathbb{R}^3$ to an orthonormal basis.

    (a) $\{(1, 0, 1), (0, 1, 2), (2, 1, 0)\}$.

    (b) $\{(1, 1, 1), (1, -1, 1), (1, 1, -1)\}$.

13. Use Gram-Schmidt Process to transform the following basis for $\mathbb{R}^4$ to an orthonormal basis: $\{(2, 1, 0, 0), (-1, 0, 0, 1), (2, 0, -1, 1), (0, 0, 1, 1)\}$.

14. (a) Find an orthonormal basis for the solution space of the equation $x + y - z = 0$.

    (b) Find the projection of $(1, 0, -1)$ onto the plane $x + y - z = 0$.

    (c) Extend the set obtained in Part (a) to an orthonormal basis for $\mathbb{R}^3$.

15. Let $W = \text{span}\{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}, \boldsymbol{u_4}, \boldsymbol{u_5}\}$ be a subspace of $\mathbb{R}^4$ where $\boldsymbol{u_1} = (1, 1, 0, 0)$, $\boldsymbol{u_2} = (1, 0, 0, 1)$, $\boldsymbol{u_3} = (1, 0, 1, 0)$, $\boldsymbol{u_4} = (3, 1, 1, 1)$, $\boldsymbol{u_5} = (-1, -1, 1, -1)$.

    (a) Show that $\{\boldsymbol{u_1}, \boldsymbol{u_3}, \boldsymbol{u_4}\}$ is a basis for W.

(b) Apply the Gram-Schmidt Process to transform $\{\boldsymbol{u_1}, \boldsymbol{u_3}, \boldsymbol{u_4}\}$ into an orthonormal basis for W.

(c) Extend the set obtained in Part (b) to an orthonormal basis for $\mathbb{R}^4$.

16. Let $\boldsymbol{V} = \text{span}\{(1,1,1), (1,a,a)\}$ where a is a real number.

(a) Find an orthonormal basis for $\boldsymbol{V}$.

(b) Compute the projection of $(5,3,1)$ onto V.

17. Let $\boldsymbol{A} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}$, $\boldsymbol{u_1} = (1,1,1,0)^{\mathrm{T}}$, $\boldsymbol{u_2} = (1,1,1,1)^{\mathrm{T}}$ and $\boldsymbol{u_3} = (0,0,1,1)^{\mathrm{T}}$.

(a) Use the Gram-Schmidt Process to transform $\{\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}\}$ into an orthonormal basis $\{\boldsymbol{w_1}, \boldsymbol{w_2}, \boldsymbol{w_3}\}$ for the column space of $\boldsymbol{A}$. (Do not change the order of $\boldsymbol{u_1}$, $\boldsymbol{u_2}$, $\boldsymbol{u_3}$ when applying the Gram-Schmidt Process.)

(b) Write each of $\boldsymbol{u_1}$, $\boldsymbol{u_2}$, $\boldsymbol{u_3}$ as a linear combination of $\boldsymbol{w_1}$, $\boldsymbol{w_2}$, $\boldsymbol{w_3}$.

(c) Hence, or otherwise, write $\boldsymbol{A} = \boldsymbol{QR}$ where $\boldsymbol{Q}$ is a 4×3 matrix with orthonormal columns and $\boldsymbol{R}$ is a 3×3 upper triangular matrix with positive entries along its diagonal.

(The process of writing a matrix in the form described in Part (c) is called the *QR factorization.* It is widely used in computer algorithms for various computations concerning matrices.)

18. Prove the uniqueness of (orthogonal) projection:

Let V be a subspace of $\mathbb{R}^n$ and $\boldsymbol{u}$ a vector in $\mathbb{R}^n$. Show that $\boldsymbol{u}$ can be written uniquely as $\boldsymbol{u} = \boldsymbol{n} + \boldsymbol{p}$ such that $\boldsymbol{n}$ is a vector orthogonal to V and $\boldsymbol{p}$ is a vector in V.

(Hint: We need to prove that if $\boldsymbol{u} = \boldsymbol{n_1} + \boldsymbol{p_1} = \boldsymbol{n_2} + \boldsymbol{p_2}$ where $\boldsymbol{n_1}, \boldsymbol{n_2}$ are orthogonal to V and $\boldsymbol{p_1}, \boldsymbol{p_2} \in V$, then $\boldsymbol{n_1} = \boldsymbol{n_2}$ and $\boldsymbol{p_1} = \boldsymbol{p_2}$.)

19. (All vectors in this question are written as column vectors.) Let $\boldsymbol{A}$ be a square matrix of order n such that $\boldsymbol{A}^2 = \boldsymbol{A}$ and $\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A}$.

(a) For any two vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$, show that $(\boldsymbol{Au}) \cdot \boldsymbol{v} = \boldsymbol{u} \cdot (\boldsymbol{Av})$.

(b) For any vector $\boldsymbol{w} \in \mathbb{R}^n$, show that $\boldsymbol{Aw}$ is the projection of $\boldsymbol{w}$ onto the subspace $V = \{\boldsymbol{u} \in \mathbb{R}^n \mid \boldsymbol{Au} = \boldsymbol{u}\}$ of $\mathbb{R}^n$.

20. Determine which of the following statements are true. Justify your answer.

(a) If $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$ are vectors in $\mathbb{R}^n$ such that $||\boldsymbol{u}|| = ||\boldsymbol{v}||$, then $||\boldsymbol{u} + \boldsymbol{w}|| = ||\boldsymbol{v} + \boldsymbol{w}||$.

(b) If $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$ are vectors in $\mathbb{R}^n$ such that $||\boldsymbol{u}|| = ||\boldsymbol{v}||$ and $\boldsymbol{w}$ is orthogonal to both $\boldsymbol{u}$ and $\boldsymbol{v}$, then $||\boldsymbol{u}+\boldsymbol{w}|| = ||\boldsymbol{v}+\boldsymbol{w}||$.

(c) If $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$ are vectors in $\mathbb{R}^n$ such that $\boldsymbol{u}$ is orthogonal to both $\boldsymbol{v}$ and $\boldsymbol{w}$, then $\boldsymbol{u}$ and $\boldsymbol{v}+\boldsymbol{w}$ are orthogonal.

(d) If $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$ are vectors in $\mathbb{R}^n$ such that $\boldsymbol{u}$, $\boldsymbol{v}$ are orthogonal and $\boldsymbol{v}$, $\boldsymbol{w}$ are orthogonal, then $\boldsymbol{u}$ and $\boldsymbol{w}$ are orthogonal.

Question 5.21 to Question 5.34 are exercises for Sections 5.3 and 5.4.

21. (a) In $\mathbb{R}^2$, find the distance from the point $(1,5)$ to the line $x-y=0$.

    (b) In $\mathbb{R}^3$, find the distance from the point $(1,0,-2)$ to the plane $2x+y-2z=0$.

    (c) In $\mathbb{R}^3$, find the distance from the point $(1,0,-2)$ to the line $x=t$, $y=2t$ and $z=2t$ for $t\in\mathbb{R}$.

22. There are two costs involved if we want to publish a book. C is a fixed cost due to typesetting and editing and D is the printing and binding cost for each additional book we want to produce.

    Suppose we expect b, the total cost of producing t books to be a linear function of t. We shall apply the least squares method (see Example 5.3.5, Theorem 5.3.10 and Example 5.3.11.2) to find a straight line $b=C+Dt$ that "best fits" the following set of data:

    $$b_1 = 3 \text{ when } t_1 = 1, \quad b_2 = 5 \text{ when } t_2 = 2 \quad \text{and} \quad b_3 = 6 \text{ when } t_3 = 3.$$

    (a) Write down a linear system with three equations and two variables using the data set.

    (b) Obtain the least squares solution for C and D.

23. A father wishes to distribute an amount of money among his three sons Jack, Jim and John.

    (a) Show that it is not possible to have a distribution such that the following conditions are all satisfied.

        (i) The amount Jack receives plus twice the amount Jim receives is \$300.

        (ii) The amount Jim receives plus the amount John receives is \$300.

        (iii) Jack receives \$300 more than twice of what John receives.

    (b) Since there is no solution to the distribution problem above, find a least squares solution.

        (Make sure that your least squares solution is feasible. For example, one cannot give a negative amount of money to anybody.)

(d) Let $\boldsymbol{R} = \boldsymbol{QP}$. Is $\boldsymbol{R}$ the transition matrix from E to V'?

29. Suppose an $x'y'$-coordinate system is obtained from the xy-coordinate system by an anti-clockwise rotation through an angle $\theta = \pi/3$.

    (a) Let P be the point such that its xy-coordinates are $(2, 1)$. Find the $x'y'$-coordinates of P.

    (b) Let Q be the point such that its $x'y'$-coordinates are $(2, 1)$. Find the xy-coordinates of Q.

    (c) Let L be the line $x + y = 1$. Write down the equation of L using the $x'y'$-coordinates.

30. Suppose an $x'y'z'$-coordinate system is obtained from the xyz-coordinate system by an anti-clockwise rotation about the z-axis through an angle θ. Let $\boldsymbol{u} = (x, y, z)^{\mathrm{T}}$ and $\boldsymbol{u'} = (x', y', z')^{\mathrm{T}}$ be the xyz-coordinates and $x'y'z'$-coordinates, respectively, of the same point. Find a 3×3 matrix $\boldsymbol{A}$ such that $\boldsymbol{Au} = \boldsymbol{u'}$.

    (Hint: The z-axis is fixed under the rotation.)

31. (a) Let $S_1 = \{(1, 0), (0, 1)\}$, $S_2 = \{(1, -1), (2, 1)\}$ and $S_3 = \{(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}), (-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})\}$. Clearly, S_1, S_2 and S_3 are three bases for $\mathbb{R}^2$.

    Let $\boldsymbol{u} = (1, 4)$ and $\boldsymbol{v} = (-1, 1)$. Compute $(\boldsymbol{u})_{S_i}$, $(\boldsymbol{v})_{S_i}$ and $(\boldsymbol{u})_{S_i} \cdot (\boldsymbol{v})_{S_i}$ for $i = 1, 2, 3$. What do you observe?

    (b) Prove that if S and T are two orthonormal bases for a vector space V, then for any vectors $\boldsymbol{u}, \boldsymbol{v} \in V$, $(\boldsymbol{u})_S \cdot (\boldsymbol{v})_S = (\boldsymbol{u})_T \cdot (\boldsymbol{v})_T$.

32. (All vectors in this question are written as column vectors.) Let $\boldsymbol{A}$ be an orthogonal matrix of order n and let $\boldsymbol{u}$, $\boldsymbol{v}$ be any two vectors in $\mathbb{R}^n$. Show that

    (a) $||\boldsymbol{u}|| = ||\boldsymbol{Au}||$;

    (b) $d(\boldsymbol{u}, \boldsymbol{v}) = d(\boldsymbol{Au}, \boldsymbol{Av})$; and

    (c) the angle between $\boldsymbol{u}$ and $\boldsymbol{v}$ is equal to the angle between $\boldsymbol{Au}$ and $\boldsymbol{Av}$.

33. (All vectors in this question are written as column vectors.) Let $\boldsymbol{A}$ be an orthogonal matrix of order n and let $S = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_n}\}$ be a basis for $\mathbb{R}^n$.

    (a) Show that $T = \{\boldsymbol{Au_1}, \boldsymbol{Au_2}, \ldots, \boldsymbol{Au_n}\}$ is a basis for $\mathbb{R}^n$.

    (b) If S is orthogonal, show that T is orthogonal.

    (c) If S is orthonormal, is T orthonormal?

34. Determine which of the following statements are true. Justify your answer.

(a) If $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{c_1} & \boldsymbol{c_2} & \cdots & \boldsymbol{c_k} \end{pmatrix}$ is an $n \times k$ matrix such that $\boldsymbol{c_1}, \boldsymbol{c_2}, \ldots, \boldsymbol{c_k}$ are orthonormal, then $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{I}_k$.

(b) If $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{c_1} & \boldsymbol{c_2} & \cdots & \boldsymbol{c_k} \end{pmatrix}$ is an $n \times k$ matrix such that $\boldsymbol{c_1}, \boldsymbol{c_2}, \ldots, \boldsymbol{c_k}$ are orthonormal, then $\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{I}_n$.

(c) If $\boldsymbol{A}$ and $\boldsymbol{B}$ are orthogonal matrices, then $\boldsymbol{A} + \boldsymbol{B}$ is an orthogonal matrix.

(d) If $\boldsymbol{A}$ and $\boldsymbol{B}$ are orthogonal matrices, then $\boldsymbol{AB}$ is an orthogonal matrix.

Chapter 6

Diagonalization

In this chapter, all vectors are written as column vectors.

Section 6.1 Eigenvalues and Eigenvectors

Example 6.1.1 Each year 4% of the rural population moves to the urban district while 1% of the urban population moves to the rural district. Suppose we want to study the long term effect if things keep going like this.

Let a_n and b_n be the rural population and urban population, respectively, after n years. Observe that for $n = 1, 2, 3, \dots$,

$$\begin{cases} a_n = 0.96\, a_{n-1} + 0.01\, b_{n-1} \\ b_n = 0.04\, a_{n-1} + 0.99\, b_{n-1} \end{cases} \quad \Leftrightarrow \quad \begin{pmatrix} a_n \\ b_n \end{pmatrix} = \begin{pmatrix} 0.96 & 0.01 \\ 0.04 & 0.99 \end{pmatrix} \begin{pmatrix} a_{n-1} \\ b_{n-1} \end{pmatrix}.$$

Let $\boldsymbol{x_n} = \begin{pmatrix} a_n \\ b_n \end{pmatrix}$ and $\boldsymbol{A} = \begin{pmatrix} 0.96 & 0.01 \\ 0.04 & 0.99 \end{pmatrix}$. Then

$$\boldsymbol{x_n} = \boldsymbol{A}\boldsymbol{x_{n-1}} = \boldsymbol{A}^2\boldsymbol{x_{n-2}} = \cdots = \boldsymbol{A}^n\boldsymbol{x_0}.$$

To study the long term effect, we need to compute $\boldsymbol{A}^n$ for large n. If possible, we want to find $\lim\limits_{n\to\infty} \boldsymbol{x_n} = \lim\limits_{n\to\infty} \boldsymbol{A}^n\boldsymbol{x_0}$. It happens that we can write

$$\boldsymbol{A} = \begin{pmatrix} 0.96 & 0.01 \\ 0.04 & 0.99 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 4 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0.95 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 4 & -1 \end{pmatrix}^{-1} \quad \text{(check it)}.$$

Let $\boldsymbol{P} = \begin{pmatrix} 1 & 1 \\ 4 & -1 \end{pmatrix}$ and $\boldsymbol{D} = \begin{pmatrix} 1 & 0 \\ 0 & 0.95 \end{pmatrix}$. Then

$$\begin{aligned}
\boldsymbol{A}^n &= (\boldsymbol{PDP}^{-1})^n \\
&= \underbrace{(\boldsymbol{PDP}^{-1})(\boldsymbol{PDP}^{-1})\cdots(\boldsymbol{PDP}^{-1})}_{n \text{ times}} \\
&= \boldsymbol{PD}(\boldsymbol{P}^{-1}\boldsymbol{P})\boldsymbol{D}(\boldsymbol{P}^{-1}\boldsymbol{P})\cdots(\boldsymbol{P}^{-1}\boldsymbol{P})\boldsymbol{DP}^{-1} \\
&= \boldsymbol{PDIDI}\cdots\boldsymbol{IDP}^{-1} \\
&= \boldsymbol{PDD}\cdots\boldsymbol{DP}^{-1} = \boldsymbol{PD}^n\boldsymbol{P}^{-1}.
\end{aligned}$$

Since

$$\boldsymbol{D}^n = \begin{pmatrix} 1 & 0 \\ 0 & 0.95 \end{pmatrix}^n = \begin{pmatrix} 1^n & 0 \\ 0 & 0.95^n \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0.95^n \end{pmatrix},$$

we have $\lim_{n\to\infty} \boldsymbol{D}^n = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$. Hence

$$\begin{aligned}
\lim_{n\to\infty} \boldsymbol{A}^n &= \lim_{n\to\infty} \boldsymbol{PD}^n\boldsymbol{P}^{-1} \\
&= \begin{pmatrix} 1 & 1 \\ 4 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 4 & -1 \end{pmatrix}^{-1} = \begin{pmatrix} 0.2 & 0.2 \\ 0.8 & 0.8 \end{pmatrix}
\end{aligned}$$

and

$$\lim_{n\to\infty} \begin{pmatrix} a_n \\ b_n \end{pmatrix} = \lim_{n\to\infty} \boldsymbol{x_n} = \lim_{n\to\infty} \boldsymbol{A}^n\boldsymbol{x_0} = \begin{pmatrix} 0.2 & 0.2 \\ 0.8 & 0.8 \end{pmatrix}\begin{pmatrix} a_0 \\ b_0 \end{pmatrix},$$

i.e. $\lim_{n\to\infty} a_n = 0.2(a_0 + b_0)$ and $\lim_{n\to\infty} b_n = 0.8(a_0 + b_0)$. So in the long run, 20% of the total population will stay in the rural district and 80% of the population will stay in the urban district.

Remark 6.1.2 In the example above, the crucial step of the calculation is to express $\boldsymbol{A}$ in the form $\boldsymbol{PDP}^{-1}$ where $\boldsymbol{D}$ is a diagonal matrix. This leads us to the problem of "diagonalizing" square matrices. First, we need to study the concept of eigenvalues and eigenvectors.

Definition 6.1.3 Let $\boldsymbol{A}$ be a square matrix of order n. A nonzero column vector $\boldsymbol{u}$ in $\mathbb{R}^n$ is called an *eigenvector* of $\boldsymbol{A}$ if

$$\boldsymbol{Au} = \lambda\boldsymbol{u}$$

for some scalar λ. The scalar λ is called an *eigenvalue* of $\boldsymbol{A}$ and $\boldsymbol{u}$ is said to be an eigenvector of $\boldsymbol{A}$ *associated* with the eigenvalue λ.

Example 6.1.4

1. Let $\boldsymbol{A}$ be the 2×2 matrix in Example 6.1.1. Let $\boldsymbol{u} = \begin{pmatrix} 1 \\ 4 \end{pmatrix}$ and $\boldsymbol{v} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Then

$$\boldsymbol{A}\boldsymbol{u} = \begin{pmatrix} 0.96 & 0.01 \\ 0.04 & 0.99 \end{pmatrix} \begin{pmatrix} 1 \\ 4 \end{pmatrix} = \begin{pmatrix} 1 \\ 4 \end{pmatrix} = \boldsymbol{u}$$

and

$$\boldsymbol{A}\boldsymbol{v} = \begin{pmatrix} 0.96 & 0.01 \\ 0.04 & 0.99 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0.95 \\ -0.95 \end{pmatrix} = 0.95 \begin{pmatrix} 1 \\ -1 \end{pmatrix} = 0.95\boldsymbol{v}.$$

So 1 and 0.95 are eigenvalues of $\boldsymbol{A}$, $\boldsymbol{u}$ is an eigenvector of $\boldsymbol{A}$ associated with the eigenvalue 1 and $\boldsymbol{v}$ is an eigenvector of $\boldsymbol{A}$ associated with the eigenvalue 0.95.

2. Let $\boldsymbol{B} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ and let $\boldsymbol{u} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$, $\boldsymbol{v} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$ and $\boldsymbol{w} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$. Then

$$\boldsymbol{B}\boldsymbol{u} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \\ 3 \end{pmatrix} = 3 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 3\boldsymbol{u},$$

$$\boldsymbol{B}\boldsymbol{v} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = 0 \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = 0\boldsymbol{v}$$

and

$$\boldsymbol{B}\boldsymbol{w} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = 0 \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} = 0\boldsymbol{w}.$$

So 3 and 0 are eigenvalues of $\boldsymbol{B}$, $\boldsymbol{u}$ is an eigenvector of $\boldsymbol{B}$ associated with the eigenvalue 3 and $\boldsymbol{v}, \boldsymbol{w}$ are eigenvectors of $\boldsymbol{B}$ associated with the eigenvalue 0. Note that

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & -2 \\ 1 & -1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & -2 \\ 1 & -1 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

(check it).

Remark 6.1.5 Let $\boldsymbol{A}$ be a square matrix of order n. Then

$$\begin{aligned} & \lambda \text{ is an eigenvalue of } \boldsymbol{A} \\ \Leftrightarrow\quad & \boldsymbol{A}\boldsymbol{u} = \lambda\boldsymbol{u} \text{ for some nonzero column vector } \boldsymbol{u} \\ \Leftrightarrow\quad & \lambda\boldsymbol{u} - \boldsymbol{A}\boldsymbol{u} = \boldsymbol{0} \text{ for some nonzero column vector } \boldsymbol{u} \\ \Leftrightarrow\quad & (\lambda\boldsymbol{I} - \boldsymbol{A})\boldsymbol{u} = \boldsymbol{0} \text{ for some nonzero column vector } \boldsymbol{u} \\ \Leftrightarrow\quad & \text{the linear system } (\lambda\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0} \text{ has non-trival solution} \\ \Leftrightarrow\quad & \det(\lambda\boldsymbol{I} - \boldsymbol{A}) = 0 \quad \text{(by Theorem 3.6.11)}. \end{aligned}$$

If expanded, $\det(\lambda\boldsymbol{I} - \boldsymbol{A})$ is a polynomial in λ of degree n.

Definition 6.1.6 Let $\boldsymbol{A}$ be a square matrix of order n. The equation

$$\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = 0$$

is called the *characteristic equation* of $\boldsymbol{A}$ and the polynomial

$$\det(\lambda \boldsymbol{I} - \boldsymbol{A})$$

is called the *characteristic polynomial* of $\boldsymbol{A}$.

Example 6.1.7

1. Let $\boldsymbol{A}$ be the 2×2 matrix in Example 6.1.1. The characteristic polynomial of $\boldsymbol{A}$ is

$$\begin{aligned}\det(\lambda \boldsymbol{I} - \boldsymbol{A}) &= \det\left(\lambda\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} - \begin{pmatrix}0.96 & 0.01\\ 0.04 & 0.99\end{pmatrix}\right)\\ &= \begin{vmatrix}\lambda - 0.96 & -0.01\\ -0.04 & \lambda - 0.99\end{vmatrix}\\ &= (\lambda - 0.96)(\lambda - 0.99) - (-0.01)(-0.04)\\ &= \lambda^2 - 1.95\lambda + 0.95\\ &= (\lambda - 1)(\lambda - 0.95).\end{aligned}$$

Hence $\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = 0$ if and only if $\lambda = 1$ or 0.95. The eigenvalues of $\boldsymbol{A}$ are 1 and 0.95.

2. Let $\boldsymbol{B}$ be the 3×3 matrix in Example 6.1.4.2. The characteristic polynomial of $\boldsymbol{B}$ is

$$\begin{aligned}\det(\lambda \boldsymbol{I} - \boldsymbol{B}) &= \begin{vmatrix}\lambda - 1 & -1 & -1\\ -1 & \lambda - 1 & -1\\ -1 & -1 & \lambda - 1\end{vmatrix}\\ &= \lambda^3 - 3\lambda^2 = (\lambda - 3)(\lambda - 0)^2.\end{aligned}$$

Hence $\det(\lambda \boldsymbol{I} - \boldsymbol{B}) = 0$ if and only if $\lambda = 3$ or 0. The eigenvalues of $\boldsymbol{B}$ are 3 and 0.

3. Let $\boldsymbol{C} = \begin{pmatrix}0 & -1 & 0\\ 0 & 0 & 2\\ 1 & 1 & 1\end{pmatrix}$. The characteristic polynomial of $\boldsymbol{C}$ is

$$\begin{aligned}\det(\lambda \boldsymbol{I} - \boldsymbol{C}) &= \begin{vmatrix}\lambda & 1 & 0\\ 0 & \lambda & -2\\ -1 & -1 & \lambda - 1\end{vmatrix}\\ &= \lambda^3 - \lambda^2 - 2\lambda + 2 = (\lambda - 1)(\lambda - \sqrt{2})(\lambda + \sqrt{2}).\end{aligned}$$

Hence $\det(\lambda \boldsymbol{I} - \boldsymbol{C}) = 0$ if and only if $\lambda = 1$, $\sqrt{2}$ or $-\sqrt{2}$. The eigenvalues of $\boldsymbol{C}$ are 1, $\sqrt{2}$ and $-\sqrt{2}$.

Theorem 6.1.8 (The Main Theorem on Invertible Matrices) Let $\boldsymbol{A}$ be an $n \times n$ matrices. The following statements are equivalent.

1. $\boldsymbol{A}$ is invertible.
2. The linear system $\boldsymbol{Ax} = \boldsymbol{0}$ has only the trivial solution.
3. The reduced row-echelon form of $\boldsymbol{A}$ is an identity matrix.
4. $\boldsymbol{A}$ can be expressed as a product of elementary matrices.
5. $\det(\boldsymbol{A}) \neq 0$.
6. The rows of $\boldsymbol{A}$ form a basis for $\mathbb{R}^n$.
7. The columns of $\boldsymbol{A}$ form a basis for $\mathbb{R}^n$.
8. $\operatorname{rank}(\boldsymbol{A}) = n$.
9. 0 is not an eigenvalue of $\boldsymbol{A}$.

Proof By Theorem 3.6.11, statements 1 to 7 are equivalent. By Remark 4.2.5.2, "5 $\Leftrightarrow$ 8". Since

$$\det(0\boldsymbol{I} - \boldsymbol{A}) = \det(-\boldsymbol{A}) = (-1)^n \det(\boldsymbol{A}),$$

by Remark 6.1.5, $\det(\boldsymbol{A}) \neq 0$ if and only if 0 is not an eigenvalue of $\boldsymbol{A}$, i.e. "5 $\Leftrightarrow$ 9".

Theorem 6.1.9 If $\boldsymbol{A}$ is a triangular matrix (either upper triangular matrix or lower triangular matrix), the eigenvalues of $\boldsymbol{A}$ are the diagonal entries of $\boldsymbol{A}$.

Proof Suppose $\boldsymbol{A} = (a_{ij})_{n\times n}$ is a triangular matrix. Then $\lambda\boldsymbol{I} - \boldsymbol{A}$ is a triangular matrix with diagonal entries $\lambda - a_{11}$, $\lambda - a_{22}$, $\ldots$, $\lambda - a_{nn}$. By Theorem 2.5.8,

$$\det(\lambda\boldsymbol{I} - \boldsymbol{A}) = (\lambda - a_{11})(\lambda - a_{22})\cdots(\lambda - a_{nn}).$$

Hence the diagonal entries, $a_{11}, a_{22}, \ldots, a_{nn}$, of $\boldsymbol{A}$ are the eigenvalues of $\boldsymbol{A}$.

Example 6.1.10

1. The eigenvalues of $\begin{pmatrix} -1 & 3.5 & 14 \\ 0 & 5 & -26 \\ 0 & 0 & 2 \end{pmatrix}$ are -1, 5 and 2.
2. The eigenvalues of $\begin{pmatrix} -2 & 0 & 0 \\ 99 & 0 & 0 \\ 10 & -4.5 & 10 \end{pmatrix}$ are -2, 0 and 10.

Definition 6.1.11 Let $\boldsymbol{A}$ be a square matrix of order n and λ an eigenvalue of $\boldsymbol{A}$. Then the solution space of the linear system $(\lambda\boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}=\boldsymbol{0}$ is called the *eigenspace* of $\boldsymbol{A}$ *associated* with the eigenvalue λ and is denoted by E_λ.

Note that if $\boldsymbol{u}$ is a nonzero vector in E_λ, then $\boldsymbol{u}$ is an eigenvector of $\boldsymbol{A}$ associated with the eigenvalue λ.

Example 6.1.12

1. Let $\boldsymbol{A}$ be the 2×2 matrix in Example 6.1.1. The eigenvalues of $\boldsymbol{A}$ are 1 and 0.95.

   For $\lambda=1$, the linear system $(\lambda\boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}=\boldsymbol{0}$ is

   $$\begin{pmatrix}1-0.96 & -0.01\\ -0.04 & 1-0.99\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix}=\begin{pmatrix}0\\ 0\end{pmatrix} \quad\Leftrightarrow\quad \begin{pmatrix}0.04 & -0.01\\ -0.04 & 0.01\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix}=\begin{pmatrix}0\\ 0\end{pmatrix}.$$

   Solving the system, we have $\begin{pmatrix}x\\ y\end{pmatrix}=t\begin{pmatrix}0.25\\ 1\end{pmatrix}$ where t is an arbitrary parameter. Hence the eigenspace of $\boldsymbol{A}$ associated with the eigenvalue 1 is

   $$E_1=\operatorname{span}\left\{\begin{pmatrix}0.25\\ 1\end{pmatrix}\right\}.$$

   For $\lambda=0.95$, the linear system $(\lambda\boldsymbol{I}-\boldsymbol{A})\boldsymbol{x}=\boldsymbol{0}$ is

   $$\begin{pmatrix}0.95-0.96 & -0.01\\ -0.04 & 0.95-0.99\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix}=\begin{pmatrix}0\\ 0\end{pmatrix} \quad\Leftrightarrow\quad \begin{pmatrix}-0.01 & -0.01\\ -0.04 & -0.4\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix}=\begin{pmatrix}0\\ 0\end{pmatrix}.$$

   Solving the system, we have $\begin{pmatrix}x\\ y\end{pmatrix}=t\begin{pmatrix}-1\\ 1\end{pmatrix}$ where t is an arbitrary parameter. Hence the eigenspace of $\boldsymbol{A}$ associated with the eigenvalue 0.95 is

   $$E_{0.95}=\operatorname{span}\left\{\begin{pmatrix}-1\\ 1\end{pmatrix}\right\}.$$

2. Let $\boldsymbol{B}$ be the 3×3 matrix in Example 6.1.4.2. The eigenvalues of $\boldsymbol{B}$ are 3 and 0.

   For $\lambda=3$, the linear system $(\lambda\boldsymbol{I}-\boldsymbol{B})\boldsymbol{x}=\boldsymbol{0}$ is

   $$\begin{pmatrix}3-1 & -1 & -1\\ -1 & 3-1 & -1\\ -1 & -1 & 3-1\end{pmatrix}\begin{pmatrix}x\\ y\\ z\end{pmatrix}=\begin{pmatrix}0\\ 0\\ 0\end{pmatrix} \quad\Leftrightarrow\quad \begin{pmatrix}2 & -1 & -1\\ -1 & 2 & -1\\ -1 & -1 & 2\end{pmatrix}\begin{pmatrix}x\\ y\\ z\end{pmatrix}=\begin{pmatrix}0\\ 0\\ 0\end{pmatrix}.$$

   Solving the system, we have $\begin{pmatrix}x\\ y\\ z\end{pmatrix}=t\begin{pmatrix}1\\ 1\\ 1\end{pmatrix}$ where t is an arbitrary parameter. Hence the eigenspace of $\boldsymbol{B}$ associated with the eigenvalue 3 is

$$E_3 = \text{span}\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}.$$

For $\lambda = 0$, the linear system $(\lambda \boldsymbol{I} - \boldsymbol{B})\boldsymbol{x} = \boldsymbol{0}$ is

$$\begin{pmatrix} 0-1 & -1 & -1 \\ -1 & 0-1 & -1 \\ -1 & -1 & 0-1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad \Leftrightarrow \quad \begin{pmatrix} -1 & -1 & -1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Solving the system, we have $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = s \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$ where s, t are arbitrary parameters. Hence the eigenspace of $\boldsymbol{B}$ associated with the eigenvalue 0 is

$$E_0 = \text{span}\left\{ \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

3. Let $\boldsymbol{C}$ be the 3×3 matrix in Example 6.1.7.3. The eigenvalues of $\boldsymbol{C}$ are 1, $\sqrt{2}$ and $-\sqrt{2}$.

For $\lambda = 1$, the linear system $(\lambda \boldsymbol{I} - \boldsymbol{C})\boldsymbol{x} = \boldsymbol{0}$ is

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & -2 \\ -1 & -1 & 1-1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Solving the system, we have $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = t \begin{pmatrix} -2 \\ 2 \\ 1 \end{pmatrix}$ where t is an arbitrary parameter. Hence the eigenspace of $\boldsymbol{C}$ associated with the eigenvalue 1 is

$$E_1 = \text{span}\left\{ \begin{pmatrix} -2 \\ 2 \\ 1 \end{pmatrix} \right\}.$$

Similarly, the eigenspaces of $\boldsymbol{C}$ associated with the eigenvalues $\sqrt{2}$ and $-\sqrt{2}$ are

$$E_{\sqrt{2}} = \text{span}\left\{ \begin{pmatrix} -1 \\ \sqrt{2} \\ 1 \end{pmatrix} \right\} \quad \text{and} \quad E_{-\sqrt{2}} = \text{span}\left\{ \begin{pmatrix} -1 \\ -\sqrt{2} \\ 1 \end{pmatrix} \right\}.$$

4. Let $\boldsymbol{M} = \begin{pmatrix} 2 & 0 \\ 1 & 2 \end{pmatrix}$. By Theorem 6.1.9, M has only one eigenvalue 2. It is easy to check that the eigenspace of $\boldsymbol{M}$ associated with the eigenvalue 2 is

$$E_2 = \text{span}\left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}.$$

Section 6.2 Diagonalization

Definition 6.2.1 A square matrix $\boldsymbol{A}$ is called *diagonalizable* if there exists an invertible matrix $\boldsymbol{P}$ such that $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ is a diagonal matrix. Here the matrix $\boldsymbol{P}$ is said to *diagonalize* $\boldsymbol{A}$.

Example 6.2.2

1. The matrix $\boldsymbol{A} = \begin{pmatrix} 0.96 & 0.01 \\ 0.04 & 0.99 \end{pmatrix}$ in Example 6.1.1 is diagonalizable because

$$\begin{pmatrix} 1 & 1 \\ 4 & -1 \end{pmatrix}^{-1} \begin{pmatrix} 0.96 & 0.01 \\ 0.04 & 0.99 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 4 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0.95 \end{pmatrix}.$$

2. The matrix $\boldsymbol{B} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ in Example 6.1.4.2 is diagonalizable because

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & -2 \\ 1 & -1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & -2 \\ 1 & -1 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

3. The matrix $\boldsymbol{M} = \begin{pmatrix} 2 & 0 \\ 1 & 2 \end{pmatrix}$ in Example 6.1.12.4 is not diagonalizable, i.e. there is no invertible matrix that can diagonalize $\boldsymbol{M}$.

Proof Assume the contrary, i.e. there exists an invertible matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ such that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} \begin{pmatrix} 2 & 0 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}.$$

Then $\begin{pmatrix} 2 & 0 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}$ which implies

$$\begin{cases} 2a & & & = \lambda a \\ & 2b & & = \mu b \\ a & & +2c & = \lambda c \\ & b & & +2d = \mu d. \end{cases}$$

Solving the equations, we obtain $a = 0$ and $b = 0$. However, $\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ c & d \end{pmatrix}$ is not invertible, a contradiction.

(We shall introduce a systematic approach to test whether a square matrix is diagonalizable in Algorithm 6.2.4.)

Theorem 6.2.3 Let $\boldsymbol{A}$ be a square matrix of order n. Then $\boldsymbol{A}$ is diagonalizable if and only if $\boldsymbol{A}$ has n linearly independent eigenvectors.

Proof

($\Rightarrow$) Suppose $\boldsymbol{A}$ is diagonalizable. Let $\boldsymbol{P}$ be an invertible matrix such that

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{D}$$

where $\boldsymbol{D} = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}$.

Let $\boldsymbol{P} = \begin{pmatrix} \boldsymbol{u_1} & \boldsymbol{u_2} & \cdots & \boldsymbol{u_n} \end{pmatrix}$ where $\boldsymbol{u_i}$ is the ith column of $\boldsymbol{P}$. Since $\boldsymbol{AP} = \boldsymbol{PD}$, we have

$$\boldsymbol{A}\begin{pmatrix} \boldsymbol{u_1} & \boldsymbol{u_2} & \cdots & \boldsymbol{u_n} \end{pmatrix} = \begin{pmatrix} \boldsymbol{u_1} & \boldsymbol{u_2} & \cdots & \boldsymbol{u_n} \end{pmatrix}\begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}$$

$$\Rightarrow \quad \begin{pmatrix} \boldsymbol{Au_1} & \boldsymbol{Au_2} & \cdots & \boldsymbol{Au_n} \end{pmatrix} = \begin{pmatrix} \lambda_1\boldsymbol{u_1} & \lambda_2\boldsymbol{u_2} & \cdots & \lambda_n\boldsymbol{u_n} \end{pmatrix}.$$

So $\boldsymbol{Au_i} = \lambda_i\boldsymbol{u_i}$ for all i, i.e. $\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_n}$ are eigenvectors of $\boldsymbol{A}$. Since $\boldsymbol{P}$ is invertible, by Theorem 3.6.11, $\{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_n}\}$ is a basis for $\mathbb{R}^n$ and hence must be linearly independent.

($\Leftarrow$) Suppose $\boldsymbol{A}$ has n linearly independent eigenvectors, say, $\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_n}$ are linearly independent eigenvectors of $\boldsymbol{A}$ associated with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ respectively. By Theorem 3.6.7, $\{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_n}\}$ is a basis for $\mathbb{R}^n$.

Define $\boldsymbol{P} = \begin{pmatrix} \boldsymbol{u_1} & \boldsymbol{u_2} & \cdots & \boldsymbol{u_n} \end{pmatrix}$. Then

$$\boldsymbol{AP} = \begin{pmatrix} \boldsymbol{Au_1} & \boldsymbol{Au_2} & \cdots & \boldsymbol{Au_n} \end{pmatrix} = \begin{pmatrix} \lambda_1 \boldsymbol{u_1} & \lambda_2 \boldsymbol{u_2} & \cdots & \lambda_n \boldsymbol{u_n} \end{pmatrix} = \boldsymbol{PD}$$

where $\boldsymbol{D} = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}$.

By Theorem 3.6.11, $\boldsymbol{P}$ is invertible. So $\boldsymbol{P}^{-1}\boldsymbol{AP} = \boldsymbol{D}$, i.e. $\boldsymbol{A}$ is diagonalizable.

Algorithm 6.2.4 Given a square matrix $\boldsymbol{A}$ of order n, we want to determine whether $\boldsymbol{A}$ is diagonalizable. Also, if $\boldsymbol{A}$ is diagonalizable, find an invertible matrix $\boldsymbol{P}$ such that $\boldsymbol{P}^{-1}\boldsymbol{AP}$ is a diagonal matrix.

Step 1: Find all distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$. (By Remark 6.1.5, eigenvalues can be obtained by solving the characteristic equation $\det(\lambda\boldsymbol{I} - \boldsymbol{A}) = 0$.)

Step 2: For each eigenvalue λ_i, find a basis S_{λ_i} for the eigenspace E_{λ_i}.

Step 3: Let $S = S_{\lambda_1} \cup S_{\lambda_2} \cup \cdots \cup S_{\lambda_k}$.

(a) If $|S| < n$, then $\boldsymbol{A}$ is not diagonalizable.

(b) If $|S| = n$, say $S = \{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_n}\}$, then $\boldsymbol{P} = \begin{pmatrix} \boldsymbol{u_1} & \boldsymbol{u_2} & \cdots & \boldsymbol{u_n} \end{pmatrix}$ is an invertible matrix that diagonalizes $\boldsymbol{A}$.

Remark 6.2.5 The following are some remarks on Algorithm 6.2.4:

1. In Step 1, sometimes, the matrix $\boldsymbol{A}$ may have eigenvalues that are not real numbers but complex numbers, i.e. the characteristic equation $\det(\lambda\boldsymbol{I} - \boldsymbol{A}) = 0$ has complex solutions. We can still use the algorithm to diagonalize the matrix. (See Question 6.10.) However, to discuss the theory, we need the concept of vector spaces over complex numbers.

2. Suppose the characteristic polynomial of the matrix $\boldsymbol{A}$ can be factorized as

$$\det(\lambda\boldsymbol{I} - \boldsymbol{A}) = (\lambda - \lambda_1)^{r_1}(\lambda - \lambda_2)^{r_2} \cdots (\lambda - \lambda_k)^{r_k}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are distinct eigenvalues of $\boldsymbol{A}$. Then for each eigenvalue λ_i,

$$\dim(E_{\lambda_i}) \le r_i.$$

Furthermore, $\boldsymbol{A}$ is diagonalizable if and only if in Step 2, for each eigenvalue λ_i, $\dim(E_{\lambda_i}) = r_i$, i.e. $|S_{\lambda_i}| = r_i$.

(The proof of this result requires some advanced knowledge of linear algebra which is beyond the scope of this book.)

3. In Step 3, the set S is always linearly independent. (See Question 6.22.)

Example 6.2.6

1. Let $\boldsymbol{B}$ be the 3×3 matrix in Example 6.1.4.2.

   Step 1: By Example 6.1.4.2, the eigenvalues are 3 and 0.

   Step 2: By Example 6.1.12.2,

   $$\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\} \text{ is a basis for } E_3,$$

   $$\left\{ \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right\} \text{ is a basis for } E_0.$$

   Step 3: Let $\boldsymbol{P} = \begin{pmatrix} 1 & -1 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$. Then $\boldsymbol{P}^{-1}\boldsymbol{B}\boldsymbol{P} = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

   In Step 3, if we let $\boldsymbol{Q} = \begin{pmatrix} -1 & 1 & -1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$, then $\boldsymbol{Q}^{-1}\boldsymbol{B}\boldsymbol{Q} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

2. Let $\boldsymbol{C}$ be the 3×3 matrix in Example 6.1.7.3.

   Step 1: By Example 6.1.7.3, the eigenvalues are 1, $\sqrt{2}$ and $-\sqrt{2}$.

   Step 2: By Example 6.1.12.3,

   $$\left\{ \begin{pmatrix} -2 \\ 2 \\ 1 \end{pmatrix} \right\} \text{ is a basis for } E_1,$$

   $$\left\{ \begin{pmatrix} -1 \\ \sqrt{2} \\ 1 \end{pmatrix} \right\} \text{ is a basis for } E_{\sqrt{2}} \text{ and}$$

   $$\left\{ \begin{pmatrix} -1 \\ -\sqrt{2} \\ 1 \end{pmatrix} \right\} \text{ is a basis for } E_{-\sqrt{2}}.$$

   Step 3: Let $\boldsymbol{P} = \begin{pmatrix} -2 & -1 & -1 \\ 2 & \sqrt{2} & -\sqrt{2} \\ 1 & 1 & 1 \end{pmatrix}$. Then $\boldsymbol{P}^{-1}\boldsymbol{C}\boldsymbol{P} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & -\sqrt{2} \end{pmatrix}$.

3. Let $\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ -3 & 5 & 2 \end{pmatrix}$.

Step 1: The eigenvalues are 1 and 2. (We can get the eigenvalues of $\boldsymbol{A}$ without solving the equation $\det(\lambda \boldsymbol{I} - \boldsymbol{A}) = 0$. Why?)

Step 2: For $\lambda = 1$, the linear system $(\lambda \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$ is

$$\begin{pmatrix} 0 & 0 & 0 \\ -1 & -1 & 0 \\ 3 & -5 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Solving the system, we have $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = t \begin{pmatrix} 1 \\ -1 \\ 8 \end{pmatrix}$ where t is an arbitrary parameter.

So $\left\{ \begin{pmatrix} 1 \\ -1 \\ 8 \end{pmatrix} \right\}$ is a basis for E_1.

For $\lambda = 2$, the linear system $(\lambda \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$ is

$$\begin{pmatrix} 1 & 0 & 0 \\ -1 & 0 & 0 \\ 3 & -5 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Solving the system, we have $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = t \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ where t is an arbitrary parameter.

So $\left\{ \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$ is a basis for E_2.

Step 3: Since we only have two linearly independent eigenvectors, $\boldsymbol{A}$ is not diagonalizable.

Theorem 6.2.7 Let $\boldsymbol{A}$ be a square matrix of order n. If $\boldsymbol{A}$ has n distinct eigenvalues, then $\boldsymbol{A}$ is diagonalizable.

Proof Suppose $\boldsymbol{A}$ has n distinct eigenvalues. In Step 2 of Algorithm 6.2.4, we can find one eigenvector for each eigenvalue and hence we have n eigenvectors. By Remark 6.2.5.3, these eigenvectors are linearly independent. So $\boldsymbol{A}$ is diagonalizable.

Example 6.2.8 Let $\boldsymbol{A} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 2 & 2 \\ 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 4 \end{pmatrix}$.

By Theorem 6.1.9, $\boldsymbol{A}$ has eigenvalues 1, 2, 3 and 4. Thus by Theorem 6.2.7, $\boldsymbol{A}$ is diagonalizable.

Remark 6.2.9 The converse of Theorem 6.2.7 is not true, i.e. a diagonalizable matrix of order n may not need to have n distinct eigenvalues. For example, the matrix $\boldsymbol{B}$ in Example 6.2.6.1 is a 3×3 diagonalizable matrix but has only two eigenvalues 3 and 0. See also Remark 6.2.5.2.

Discussion 6.2.10 As we have seen in Example 6.1.1, one of the applications of diagonalization is to compute powers of square matrices: Let $\boldsymbol{A}$ be a square matrix of order n and $\boldsymbol{P}$ an invertible matrix such that

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}.$$

1. For all positive integer m, $\quad \boldsymbol{A}^m = \boldsymbol{P}\begin{pmatrix} \lambda_1^m & & & 0 \\ & \lambda_2^m & & \\ & & \ddots & \\ 0 & & & \lambda_n^m \end{pmatrix}\boldsymbol{P}^{-1}.$

2. Suppose $\boldsymbol{A}$ is invertible. By Theorem 6.1.8, we know that $\lambda_i \neq 0$ for all i. Then

$$\boldsymbol{A}^{-1} = \boldsymbol{P}\begin{pmatrix} \lambda_1^{-1} & & & 0 \\ & \lambda_2^{-1} & & \\ & & \ddots & \\ 0 & & & \lambda_n^{-1} \end{pmatrix}\boldsymbol{P}^{-1}$$

and hence for all positive integer m, $\quad \boldsymbol{A}^{-m} = \boldsymbol{P}\begin{pmatrix} \lambda_1^{-m} & & & 0 \\ & \lambda_2^{-m} & & \\ & & \ddots & \\ 0 & & & \lambda_n^{-m} \end{pmatrix}\boldsymbol{P}^{-1}.$

Example 6.2.11

1. Let $\boldsymbol{A} = \begin{pmatrix} -4 & 0 & -6 \\ 2 & 1 & 2 \\ 3 & 0 & 5 \end{pmatrix}$. Following Algorithm 6.2.4, we find an invertible matrix

$$\boldsymbol{P} = \begin{pmatrix} -2 & 0 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \text{ such that } \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

So

$$A^m = P\begin{pmatrix} (-1)^m & 0 & 0 \\ 0 & 1^m & 0 \\ 0 & 0 & 2^m \end{pmatrix} P^{-1}.$$

In particular,

$$\begin{pmatrix} -4 & 0 & -6 \\ 2 & 1 & 2 \\ 3 & 0 & 5 \end{pmatrix}^{10} = \begin{pmatrix} -2 & 0 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1024 \end{pmatrix}\begin{pmatrix} -2 & 0 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} -1022 & 0 & -2046 \\ 0 & 1 & 0 \\ 1023 & 0 & 2047 \end{pmatrix}.$$

2. Let $(a_0, a_1, a_2, \dots)$ be a sequence of numbers such that $a_0 = 0$, $a_1 = 1$ and $a_n = a_{n-1} + a_{n-2}$ for $n \geq 2$. These numbers are known as the *Fibonacci numbers*. In the following, we demonstrate a method to find the value of a_n by using the eigenvalue technique.

   In order to formulate the problem in terms of a matrix equation, for $n = 1, 2, 3, \dots$, we write

$$\begin{cases} a_n &= \quad a_n \\ a_{n+1} &= a_{n-1} + a_n \end{cases} \quad \Leftrightarrow \quad \begin{pmatrix} a_n \\ a_{n+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} a_{n-1} \\ a_n \end{pmatrix}.$$

   Let $\boldsymbol{x_n} = \begin{pmatrix} a_n \\ a_{n+1} \end{pmatrix}$ and $\boldsymbol{A} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$. Then

$$\boldsymbol{x_n} = \boldsymbol{A}\boldsymbol{x_{n-1}} = \boldsymbol{A}^2\boldsymbol{x_{n-2}} = \cdots = \boldsymbol{A}^n\boldsymbol{x_0}.$$

   Following Algorithm 6.2.4, we find an invertible matrix $\boldsymbol{P} = \begin{pmatrix} 1 & 1 \\ \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} \end{pmatrix}$ such that $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{pmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{pmatrix}$. Then

$$\begin{pmatrix} a_n \\ a_{n+1} \end{pmatrix} = \boldsymbol{x_n} = \boldsymbol{A}^n\boldsymbol{x_0}$$

$$= \begin{pmatrix} 1 & 1 \\ \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} \end{pmatrix}\begin{pmatrix} \left(\frac{1+\sqrt{5}}{2}\right)^n & 0 \\ 0 & \left(\frac{1-\sqrt{5}}{2}\right)^n \end{pmatrix}\begin{pmatrix} 1 & 1 \\ \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} \end{pmatrix}^{-1}\begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^n \\ \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \frac{1}{\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^{n+1} \end{pmatrix}.$$

   Thus $a_n = \dfrac{1}{\sqrt{5}}\left(\dfrac{1+\sqrt{5}}{2}\right)^n - \dfrac{1}{\sqrt{5}}\left(\dfrac{1-\sqrt{5}}{2}\right)^n.$

Section 6.3 Orthogonal Diagonalization

Discussion 6.3.1 Consider the 3×3 matrix $\boldsymbol{B} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$. In Example 6.2.6.1, we use bases $\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}$ and $\left\{ \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right\}$ for E_3 and E_0, respectively, to obtain the matrix $\boldsymbol{P} = \begin{pmatrix} 1 & -1 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$ so that $\boldsymbol{P}^{-1}\boldsymbol{B}\boldsymbol{P} = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

Let us use orthonormal bases $\left\{ \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix} \right\}$ and $\left\{ \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}, \begin{pmatrix} -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{pmatrix} \right\}$ for E_3 and E_0 instead.

Let $\boldsymbol{R} = \begin{pmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \end{pmatrix}$. Note that $\boldsymbol{R}$ is an orthogonal matrix, i.e. $\boldsymbol{R}^{-1} = \boldsymbol{R}^{\mathrm{T}}$ (check it). Thus $\boldsymbol{R}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{R} = \boldsymbol{R}^{-1}\boldsymbol{B}\boldsymbol{R} = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

Definition 6.3.2 A square matrix $\boldsymbol{A}$ is called *orthogonally diagonalizable* if there exits an orthogonal matrix $\boldsymbol{P}$ such that $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}$ is a diagonal matrix. Here the matrix $\boldsymbol{P}$ is said to *orthogonally diagonalize* $\boldsymbol{A}$.

Remark 6.3.3 The orthogonal diagonalization gives us a perfect tool to study a widely used mathematical object called "quadratic forms" which we shall discuss in Section 6.4.

Theorem 6.3.4 A square matrix is orthogonally diagonalizable if and only if it is symmetric.

Proof We only prove that if a square matrix is orthogonally diagonalizable, then it is symmetric. The proof of the converse requires some advanced knowledge of linear algebra which is beyond the scope of this book.

Suppose $\boldsymbol{A}$ is orthogonally diagonalizable, i.e. there exists an orthogonal matrix such that $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{D}$ where $\boldsymbol{D}$ is a diagonal matrix. Since $\boldsymbol{P}^{\mathrm{T}} = \boldsymbol{P}^{-1}$, we have

$$\boldsymbol{A} = (\boldsymbol{P}^{\mathrm{T}})^{-1}\boldsymbol{D}\boldsymbol{P}^{-1} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{\mathrm{T}}.$$

Observe that $\boldsymbol{D}^{\mathrm{T}} = \boldsymbol{D}$. So

$$\boldsymbol{A}^{\mathrm{T}} = (\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{\mathrm{T}})^{\mathrm{T}} = (\boldsymbol{P}^{\mathrm{T}})^{\mathrm{T}}\boldsymbol{D}^{\mathrm{T}}\boldsymbol{P}^{\mathrm{T}} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{\mathrm{T}} = \boldsymbol{A}$$

and hence $\boldsymbol{A}$ is symmetric.

Algorithm 6.3.5 Given a symmetric matrix $\boldsymbol{A}$ of order n, we want to find an orthogonal matrix $\boldsymbol{P}$ such that $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}$ is a diagonal matrix.

Step 1: Find all distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$.

Step 2: For each eigenvalue λ_i,

Step 2a: find a basis S_{λ_i} for the eigenspace E_{λ_i} and then

Step 2b: use the Gram-Schmidt Process (Theorem 5.2.19) to transform S_{λ_i} to an orthonormal basis T_{λ_i}.

Step 3: Let $T = T_{\lambda_1} \cup T_{\lambda_2} \cup \cdots \cup T_{\lambda_k}$, say $T = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_n}\}$.
Then $\boldsymbol{P} = \begin{pmatrix} \boldsymbol{v_1} & \boldsymbol{v_2} & \cdots & \boldsymbol{v_n} \end{pmatrix}$ is an orthogonal matrix that diagonalizes $\boldsymbol{A}$.

Remark 6.3.6 The following are some remarks on Algorithm 6.3.5:

1. In Step 1, the eigenvalues of a symmetric matrix are always real numbers.
2. Since $\boldsymbol{A}$ is diagonalizable, by Remark 6.2.5.2, we have the following result:

   Suppose the characteristic polynomial of the matrix $\boldsymbol{A}$ can be factorized as

   $$\det(\lambda\boldsymbol{I} - \boldsymbol{A}) = (\lambda - \lambda_1)^{r_1}(\lambda - \lambda_2)^{r_2}\cdots(\lambda - \lambda_k)^{r_k}$$

   where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are distinct eigenvalues of $\boldsymbol{A}$. Then in Step 2, for each eigenvalue λ_i, $\dim(E_{\lambda_i}) = r_i$, i.e. $|S_{\lambda_i}| = |T_{\lambda_i}| = r_i$.
3. In Step 3, the set T is always orthonormal. (See Question 6.26.)
4. Since T is always orthonormal, by Theorem 5.4.6, the square matrix $\boldsymbol{P}$ in Step 3 is always orthogonal.

Example 6.3.7

1. Let $\boldsymbol{A} = \begin{pmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{pmatrix}$.

   Step 1: Since

   $$\det(\lambda\boldsymbol{I} - \boldsymbol{A}) = \lambda^2 - 2\lambda + \tfrac{3}{4} = \left(\lambda - \tfrac{1}{2}\right)\left(\lambda - \tfrac{3}{2}\right),$$

   the eigenvalues are $\frac{1}{2}$ and $\frac{3}{2}$.

Step 2:

For $\lambda = \frac{1}{2}$, the linear system $(\lambda \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$ is

$$\begin{pmatrix} -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Solving the system, we have $\begin{pmatrix} x \\ y \end{pmatrix} = t \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ where t is an arbitrary parameter. So $\left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$ is a basis for $E_{\frac{1}{2}}$. By normalizing $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$, we obtain the unit vector $\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$. Hence $\left\{ \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \right\}$ is an orthonormal basis for $E_{\frac{1}{2}}$.

For $\lambda = \frac{3}{2}$, the linear system $(\lambda \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$ is

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Solving the system, we have $\begin{pmatrix} x \\ y \end{pmatrix} = t \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ where t is an arbitrary parameter. So $\left\{ \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right\}$ is a basis for $E_{\frac{3}{2}}$. By normalizing $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$, we obtain the unit vector $\begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$. Hence $\left\{ \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \right\}$ is an orthonormal basis for $E_{\frac{3}{2}}$.

Step 3: Let $\boldsymbol{P} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$. Then $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{3}{2} \end{pmatrix}$.

In Step 3, if we let $\boldsymbol{Q} = \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$, then $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{Q} = \begin{pmatrix} \frac{3}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$.

2. Let $\boldsymbol{B} = \begin{pmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 3 & 1 \\ -1 & 1 & 1 & 3 \end{pmatrix}$.

Step 1: Since

$$\det(\lambda \boldsymbol{I} - \boldsymbol{B}) = \lambda^4 - 8\lambda^3 + 16\lambda^2 = \lambda^2(\lambda - 4)^2,$$

the eigenvalues are 0 and 4.

Step 2:

For $\lambda = 0$, the linear system $(\lambda \boldsymbol{I} - \boldsymbol{B})\boldsymbol{x} = \boldsymbol{0}$ is

$$\begin{pmatrix} -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & -3 & -1 \\ 1 & -1 & -1 & -3 \end{pmatrix} \begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Solving the system, we have $\begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix} = s \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} + t \begin{pmatrix} 2 \\ 0 \\ -1 \\ 1 \end{pmatrix}$ where s, t are arbitrary parameters. So $\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \\ -1 \\ 1 \end{pmatrix} \right\}$ is a basis for E_0. By the Gram-Schmidt Process, we obtain an orthonormal basis $\left\{ \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \right\}$ for E_0.

For $\lambda = 4$, the linear system $(\lambda \boldsymbol{I} - \boldsymbol{B})\boldsymbol{x} = \boldsymbol{0}$ is

$$\begin{pmatrix} 3 & 1 & -1 & 1 \\ 1 & 3 & 1 & -1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Solving the system, we have $\begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix} = s \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ 0 \\ 1 \end{pmatrix}$ where s, t are arbitrary parameters. So $\left\{ \begin{pmatrix} 1 \\ -1 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \\ 2 \end{pmatrix} \right\}$ is a basis for E_4. By the Gram-Schmidt Process, we obtain an orthonormal basis $\left\{ \begin{pmatrix} \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \\ 0 \end{pmatrix}, \begin{pmatrix} -\frac{1}{\sqrt{12}} \\ \frac{1}{\sqrt{12}} \\ \frac{1}{\sqrt{12}} \\ \frac{3}{\sqrt{12}} \end{pmatrix} \right\}$ for E_4.

Step 3: Let $\boldsymbol{P} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{2} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{12}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{2} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{12}} \\ 0 & -\frac{1}{2} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{12}} \\ 0 & \frac{1}{2} & 0 & \frac{3}{\sqrt{12}} \end{pmatrix}$. Then $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{P} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}$.

Section 6.4 Quadratic Forms and Conic Sections

Definition 6.4.1 The expression

$$\begin{aligned} Q(x_1, x_2, \ldots, x_n) &= \sum_{i=1}^{n} \sum_{j=i}^{n} q_{ij}\, x_i x_j \\ &= q_{11}\, x_1^2 + q_{12}\, x_1 x_2 + \cdots + q_{1n}\, x_1 x_n \\ &\qquad\qquad q_{22}\, x_2^2 + \cdots + q_{2n}\, x_2 x_n \\ &\qquad\qquad\qquad + \cdots \quad \cdots \\ &\qquad\qquad\qquad\qquad + q_{nn}\, x_n^2 , \end{aligned}$$

where q_{ij}'s are real numbers, is called a *quadratic form* in n variables $x_1, x_2, \ldots, x_n$. Define an $n \times n$ symmetric matrix $\boldsymbol{A} = (a_{ij})$ such that

$$a_{ij} = \begin{cases} q_{ii} & \text{if } i = j \\ \frac{1}{2}\, q_{ij} & \text{if } i < j \\ \frac{1}{2}\, q_{ji} & \text{if } i > j \end{cases}$$

and let $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$. Then

$$Q(x_1, x_2, \ldots, x_n) = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix} \begin{pmatrix} q_{11} & \frac{1}{2}\, q_{12} & \cdots & \frac{1}{2}\, q_{1n} \\ \frac{1}{2}\, q_{12} & q_{22} & \cdots & \frac{1}{2}\, q_{2n} \\ \vdots & \vdots & & \vdots \\ \frac{1}{2}\, q_{1n} & \frac{1}{2}\, q_{2n} & \cdots & q_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}.$$

Thus the quadratic form can also be regarded as a mapping $Q : \mathbb{R}^n \to \mathbb{R}$ defined by

$$Q(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} \quad \text{for } \boldsymbol{x} \in \mathbb{R}^n.$$

Example 6.4.2

1. Let $Q_1(x, y) = x^2 - xy + y^2$ which is a quadratic form in x, y. Then

$$Q_1(x, y) = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

2. Let $Q_2(x, y, z) = x^2 + 2y^2 + z^2 + 2xz$ which is a quadratic form in x, y, z. Then

$$Q_2(x, y, z) = \begin{pmatrix} x & y & z \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Remark 6.4.3 Quadratic forms appear quite often in various areas of study. For example, in probability and statistics, the density function of a multivariate normal distribution of n random variables is given by

$$f(x_1, x_2, \ldots, x_n) = \frac{1}{\sqrt{(2\pi)^n \det(\boldsymbol{A})}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{A} (\boldsymbol{x}-\boldsymbol{\mu})} \quad \text{for } \boldsymbol{x} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}} \in \mathbb{R}^n,$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)^{\mathrm{T}} \in \mathbb{R}^n$ and $\boldsymbol{A}$ is an $n \times n$ symmetric matrix.

Discussion 6.4.4 Let $Q(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{x}$ be a quadratic form in n variables $x_1, x_2, \ldots, x_n$ where $\boldsymbol{A}$ is an $n \times n$ symmetric matrix and $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$. The following is a standard technique to simplify the quadratic form:

First, use Algorithm 6.3.5 to find an orthogonal matrix $\boldsymbol{P}$ that diagonalizes the symmetric matrix $\boldsymbol{A}$, i.e.

$$\boldsymbol{P}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{P} = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}.$$

Define new variables $y_1, y_2, \cdots, y_n$ such that $\boldsymbol{y} = \boldsymbol{P}^{\mathrm{T}} \boldsymbol{x} = \boldsymbol{P}^{-1} \boldsymbol{x}$ where $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)^{\mathrm{T}}$. Note that $\boldsymbol{x} = \boldsymbol{P} \boldsymbol{y}$. Then the quadratic form becomes

$$\begin{aligned} Q(\boldsymbol{x}) = Q(\boldsymbol{P}\boldsymbol{y}) &= (\boldsymbol{P}\boldsymbol{y})^{\mathrm{T}} \boldsymbol{A} (\boldsymbol{P}\boldsymbol{y}) \\ &= \boldsymbol{y}^{\mathrm{T}} \boldsymbol{P}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{P} \boldsymbol{y} \\ &= \begin{pmatrix} y_1 & y_2 & \cdots & y_n \end{pmatrix} \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \\ &= \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2. \end{aligned}$$

Example 6.4.5

1. Consider the quadratic form $Q_1(x, y) = x^2 - xy + y^2$ in Example 6.4.2.1. By Algorithm 6.3.5, we find an orthogonal matrix $\begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$ such that

$$\begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{3}{2} \end{pmatrix}.$$

Define new variables x', y' such that

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}}(x+y) \\ \frac{1}{\sqrt{2}}(-x+y) \end{pmatrix}.$$

Then $Q_1(x,y) = \begin{pmatrix} x' & y' \end{pmatrix} \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{3}{2} \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} = \frac{1}{2}x'^2 + \frac{3}{2}y'^2 = \frac{1}{4}(x+y)^2 + \frac{3}{4}(-x+y)^2.$

2. Consider the quadratic form $Q_2(x,y,z) = x^2 + 2y^2 + z^2 + 2xz$ in Example 6.4.2.2. By Algorithm 6.3.5, we find an orthogonal matrix $\begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{pmatrix}$ such that

$$\begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Define new variables x', y', z' such that

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}}(x+z) \\ y \\ \frac{1}{\sqrt{2}}(-x+z) \end{pmatrix}.$$

Then $Q_2(x,y,z) = \begin{pmatrix} x' & y' & z' \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = 2x'^2 + 2y'^2 = (x+z)^2 + 2y^2.$

Definition 6.4.6 A *quadratic equation* in two variables x and y is an equation of the form

$$ax^2 + bxy + cy^2 + dx + ey = f$$

where a, b, c, d, e, f are real numbers. We can rewrite the equation in the form

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a & \frac{1}{2}b \\ \frac{1}{2}b & c \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} d & e \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = f.$$

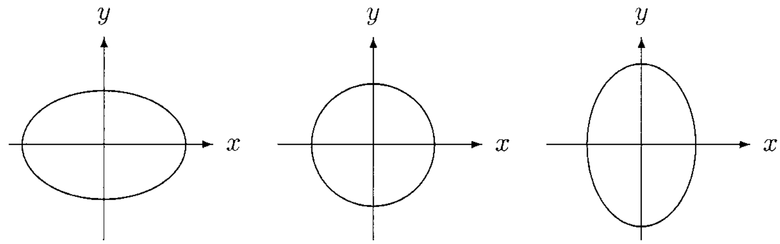
Let $\boldsymbol{x} = \begin{pmatrix} x \\ y \end{pmatrix}$, $\boldsymbol{A} = \begin{pmatrix} a & \frac{1}{2}b \\ \frac{1}{2}b & c \end{pmatrix}$ and $\boldsymbol{b} = \begin{pmatrix} d \\ e \end{pmatrix}$. Then the equation becomes

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x} = f.$$

The term $ax^2 + bxy + cy^2$ $(= \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x})$ is called the quadratic form *associated* with the quadratic equation.

The graph of a quadratic equation is known as a *conic section.* A conic section is called *degenerate* if it is the empty set, a point, a line or a pair of lines; and it is called *non-degenerate* if it is a circle, an ellipse, a hyperbola or a parabola. The equation of a non-degenerate conic section is said to be in *standard form* if it belongs to one of the following cases:

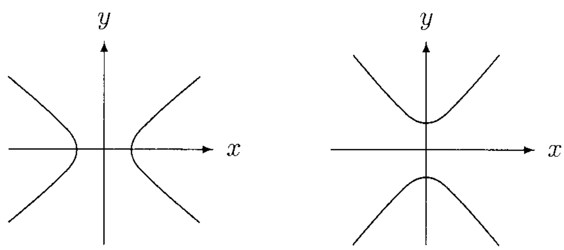
1. **(circle or ellipse)** $\dfrac{x^2}{\alpha^2} + \dfrac{y^2}{\beta^2} = 1$, i.e. $\begin{pmatrix} x & y \end{pmatrix}\begin{pmatrix} \frac{1}{\alpha^2} & 0 \\ 0 & \frac{1}{\beta^2} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = 1.$



$\alpha > \beta$ $\qquad$ $\alpha = \beta$ $\qquad$ $\alpha < \beta$

2. **(hyperbola)** $\dfrac{x^2}{\alpha^2} - \dfrac{y^2}{\beta^2} = 1$, i.e. $\begin{pmatrix} x & y \end{pmatrix}\begin{pmatrix} \frac{1}{\alpha^2} & 0 \\ 0 & -\frac{1}{\beta^2} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = 1,$

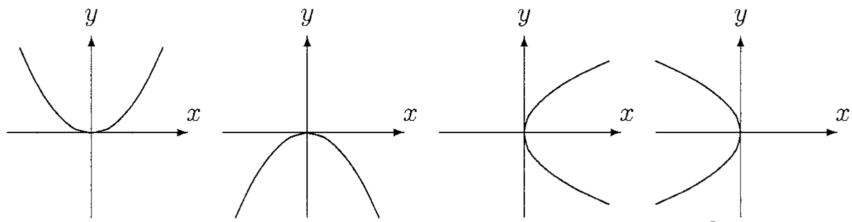
   or $-\dfrac{x^2}{\alpha^2} + \dfrac{y^2}{\beta^2} = 1$, i.e. $\begin{pmatrix} x & y \end{pmatrix}\begin{pmatrix} -\frac{1}{\alpha^2} & 0 \\ 0 & \frac{1}{\beta^2} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = 1.$



$\frac{x^2}{\alpha^2} - \frac{y^2}{\beta^2} = 1$ $\qquad$ $-\frac{x^2}{\alpha^2} + \frac{y^2}{\beta^2} = 1$

3. **(parabola)** $x^2 = \alpha y$, i.e. $\begin{pmatrix} x & y \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 0 & -\alpha \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = 0,$

   or $y^2 = \alpha x$, i.e. $\begin{pmatrix} x & y \end{pmatrix}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} -\alpha & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = 0.$



$x^2 = \alpha y$, $\alpha > 0$ $\qquad$ $x^2 = \alpha y$, $\alpha < 0$ $\qquad$ $y^2 = \alpha x$, $\alpha > 0$ $\qquad$ $y^2 = \alpha x$, $\alpha < 0$

Example 6.4.7

1. Consider the quadratic equation $x^2 - xy + y^2 - x - y = 1$, i.e.

$$\begin{pmatrix} x & y \end{pmatrix}\begin{pmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} -1 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = 1.$$

Note that the quadratic form $x^2 - xy + y^2$ is the same as $Q_1(x, y)$ in Example 6.4.5.1. Following the discussion in Example 6.4.5.1, using the new variables x', y' defined by

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}}(x+y) \\ \frac{1}{\sqrt{2}}(-x+y) \end{pmatrix},$$

we have

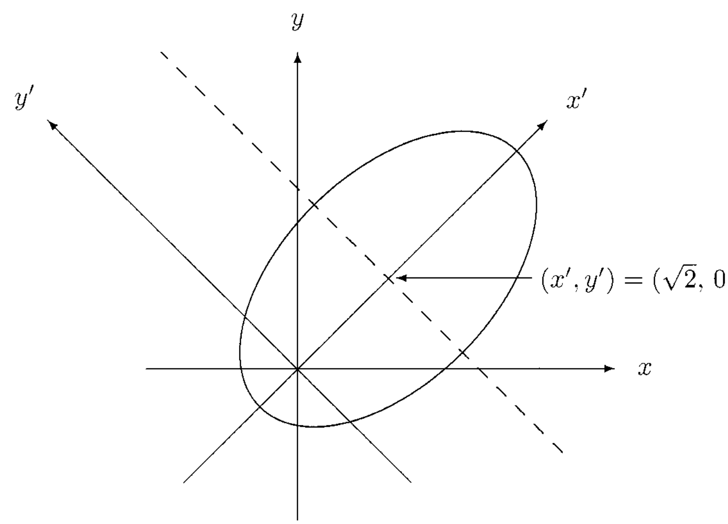
$$x^2 - xy + y^2 = \begin{pmatrix} x & y \end{pmatrix}\begin{pmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x' & y' \end{pmatrix}\begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{3}{2} \end{pmatrix}\begin{pmatrix} x' \\ y' \end{pmatrix} = \tfrac{1}{2}x'^2 + \tfrac{3}{2}y'^2.$$

(We can regard (x', y') as the coordinates of the point (x, y) using a new coordinate system with x'-axis in the direction of $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ and y'-axis in the direction of $(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$.)

Substituting x' and y' into the quadratic equation, we have

$$\begin{aligned}
& \begin{pmatrix} x & y \end{pmatrix}\begin{pmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} -1 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = 1 \\
\Leftrightarrow\quad & \begin{pmatrix} x' & y' \end{pmatrix}\begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{3}{2} \end{pmatrix}\begin{pmatrix} x' \\ y' \end{pmatrix} + \begin{pmatrix} -1 & -1 \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}\begin{pmatrix} x' \\ y' \end{pmatrix} = 1 \\
\Leftrightarrow\quad & \tfrac{1}{2}x'^2 + \tfrac{3}{2}y'^2 - \sqrt{2}x' = 1 \\
\Leftrightarrow\quad & \tfrac{1}{2}(x' - \sqrt{2})^2 + \tfrac{3}{2}y'^2 = 2 \\
\Leftrightarrow\quad & \frac{(x' - \sqrt{2})^2}{4} + \frac{y'^2}{4/3} = 1
\end{aligned}$$

which resembles the standard form of an ellipse.



2. Consider the quadratic equation $2x^2 + 24xy + 9y^2 + 20x - 6y = 5$, i.e.

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 2 & 12 \\ 12 & 9 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 20 & -6 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 5.$$

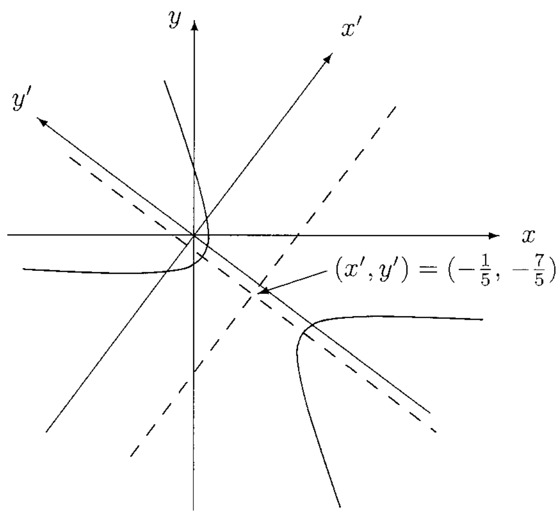
Following Algorithm 6.3.5, we obtain an orthogonal matrix $\begin{pmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{pmatrix}$ such that

$$\begin{pmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} 2 & 12 \\ 12 & 9 \end{pmatrix} \begin{pmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{pmatrix} = \begin{pmatrix} 18 & 0 \\ 0 & -7 \end{pmatrix}.$$

Let $\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} x \\ y \end{pmatrix}$. Then the quadratic equation becomes

$$\begin{aligned} & \begin{pmatrix} x' & y' \end{pmatrix} \begin{pmatrix} 18 & 0 \\ 0 & -7 \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} + \begin{pmatrix} 20 & -6 \end{pmatrix} \begin{pmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} = 5 \\ \Leftrightarrow \quad & 18x'^2 - 7y'^2 + \tfrac{36}{5}x' - \tfrac{98}{5}y' = 5 \\ \Leftrightarrow \quad & 18\left(x' + \tfrac{1}{5}\right)^2 - 7\left(y' + \tfrac{7}{5}\right)^2 = -8 \\ \Leftrightarrow \quad & -\frac{\left(x' + \frac{1}{5}\right)^2}{4/9} + \frac{\left(y' + \frac{7}{5}\right)^2}{8/7} = 1 \end{aligned}$$

which resembles the standard form of a hyperbola.



Exercise 6

Question 6.1 to Question 6.23 are exercises for Sections 6.1 and 6.2.

1. For each of the following, (i) find the characteristic equation of $\boldsymbol{A}$; (ii) find all the eigenvalues of $\boldsymbol{A}$; and (iii) find a basis for the eigenspace associated with each eigenvalue of $\boldsymbol{A}$.

   (a) $\boldsymbol{A} = \begin{pmatrix} 3 & 0 \\ 8 & -1 \end{pmatrix}$, (b) $\boldsymbol{A} = \begin{pmatrix} 1 & 1 \\ -1 & 3 \end{pmatrix}$,

   (c) $\boldsymbol{A} = \begin{pmatrix} 0 & 4 \\ 1 & 0 \end{pmatrix}$, (d) $\boldsymbol{A} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$,

   (e) $\boldsymbol{A} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix}$, (f) $\boldsymbol{A} = \begin{pmatrix} 0 & 1 & 0 \\ 9 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix}$,

   (g) $\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$, (h) $\boldsymbol{A} = \begin{pmatrix} -1 & 1 & 1 \\ -2 & 2 & 1 \\ 2 & -1 & 0 \end{pmatrix}$,

   (i) $\boldsymbol{A} = \begin{pmatrix} 4 & 0 & 0 & 0 \\ 3 & 3 & 0 & 0 \\ 2 & 2 & 2 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}$, (j) $\boldsymbol{A} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$.

2. Let $\boldsymbol{A}$ be a 2×2 matrix. Suppose $\lambda^2 + m\lambda + n$ is the characteristic polynomial of $\boldsymbol{A}$.

   (a) Show that $m = -\mathrm{tr}(\boldsymbol{A})$ (see Question 2.11) and $n = \det(\boldsymbol{A})$.

   (b) Show that $\boldsymbol{A}^2 + m\boldsymbol{A} + n\boldsymbol{I} = \boldsymbol{0}$.

3. Let λ be an eigenvalue of a square matrix $\boldsymbol{A}$.

   (a) Show that λ^n is an eigenvalue of $\boldsymbol{A}^n$ where n is a positive integer.

   (b) If $\boldsymbol{A}$ is invertible, show that $\frac{1}{\lambda}$ is an eigenvalue of $\boldsymbol{A}^{-1}$ (by Theorem 6.1.8, $\lambda \neq 0$).

   (c) Show that λ is an eigenvalue of $\boldsymbol{A}^{\mathrm{T}}$.

4. Let $\boldsymbol{A}$ be a square matrix such that $\boldsymbol{A}^2 = \boldsymbol{A}$.

   (a) Show that if λ is an eigenvalue of $\boldsymbol{A}$, then $\lambda = 0$ or 1.

   (b) Find all 2×2 matrices $\boldsymbol{A}$ such that $\boldsymbol{A}^2 = \boldsymbol{A}$ and $\boldsymbol{A}$ has eigenvalues 0 and 1.

5. Let $\boldsymbol{A}$ be a nonzero $n \times n$ matrix such that $\boldsymbol{A}^2 = \boldsymbol{0}$.

   (a) Show that if λ is an eigenvalue of $\boldsymbol{A}$, then $\lambda = 0$.

   (b) Can $\boldsymbol{A}$ be diagonalizable? Justify your answer.

   (c) Let $\boldsymbol{u}$ be a vector in $\mathbb{R}^n$ such that $\boldsymbol{A}\boldsymbol{u} \neq \boldsymbol{0}$. Prove that $\boldsymbol{u}$ and $\boldsymbol{A}\boldsymbol{u}$ are linearly independent.

   (d) For $n = 2$, show that there exists an invertible 2×2 matrix $\boldsymbol{P}$ such that
   $$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

6. Let $\boldsymbol{A} = \begin{pmatrix} 0 & -1 & 0 \\ 2 & -3 & 0 \\ 0 & 0 & -1 \end{pmatrix}$.

   (a) Show that -1 is an eigenvalue of $\boldsymbol{A}$.

   (b) Show that $\dim(E_{-1}) = 2$.

   (c) Find a 3×3 matrix $\boldsymbol{B}$ such that -3 is an eigenvalue of $\boldsymbol{B}\boldsymbol{A}$.

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

   (a) Show that 2 is an eigenvalue of $\boldsymbol{A}$.

   (b) Find a basis for the eigenspace associated with 2.

(c) If $\boldsymbol{B}$ is another 3×3 matrix with an eigenvalue λ such that the dimension of the eigenspace associated with λ is 2, prove that $2 + \lambda$ is an eigenvalue of the matrix $\boldsymbol{A} + \boldsymbol{B}$.

8. Let $\{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_n}\}$ be a basis for $\mathbb{R}^n$ and let $\boldsymbol{A}$ be an $n \times n$ matrix such that $\boldsymbol{Au_i} = \boldsymbol{u_{i+1}}$ for $i = 1, 2, \ldots, n-1$ and $\boldsymbol{Au_n} = \boldsymbol{0}$. Show that the only eigenvalue of $\boldsymbol{A}$ is 0 and find all the eigenvectors of $\boldsymbol{A}$.

9. For each matrix $\boldsymbol{A}$ in Question 6.1, (i) determine whether $\boldsymbol{A}$ is diagonalizable; and (ii) if $\boldsymbol{A}$ is diagonalizable, find a matrix $\boldsymbol{P}$ that diagonalizes $\boldsymbol{A}$ and determine $\boldsymbol{P}^{-1}\boldsymbol{AP}$.

10. (This question is only for students who are familiar with computations using complex numbers.) Each matrix $\boldsymbol{A}$ below has complex eigenvalues. Following the discussion in Remark 6.2.5.1, find a matrix $\boldsymbol{P}$ that diagonalizes $\boldsymbol{A}$ and determine $\boldsymbol{P}^{-1}\boldsymbol{AP}$.

(a) $\boldsymbol{A} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, (b) $\boldsymbol{A} = \begin{pmatrix} 1 & 1 \\ -2 & 3 \end{pmatrix}$, (c) $\boldsymbol{A} = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 2 & 1 \\ 0 & -1 & 2 \end{pmatrix}$.

11. Let $\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 3 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{pmatrix}$.

(a) Find a matrix $\boldsymbol{P}$ that diagonalizes $\boldsymbol{A}$.

(b) Compute $\boldsymbol{A}^{10}$.

(c) Find a matrix $\boldsymbol{B}$ such that $\boldsymbol{B}^2 = \boldsymbol{A}$.

12. Find a 3×3 matrix which has eigenvalues 1, 0 and -1 with corresponding eigenvectors $(0, 1, 1)^{\mathrm{T}}$, $(1, -1, 1)^{\mathrm{T}}$ and $(1, 0, 0)^{\mathrm{T}}$ respectively.

13. Determine the values of a and b so that the matrix $\begin{pmatrix} a & 1 \\ 0 & b \end{pmatrix}$ is diagonalizable.

14. Let $\boldsymbol{B}$ be a 4×4 matrix and $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3, \boldsymbol{u}_4\}$ a basis for $\mathbb{R}^4$. Suppose

$$\boldsymbol{Bu}_1 = 2\boldsymbol{u}_1, \quad \boldsymbol{Bu}_2 = \boldsymbol{0}, \quad \boldsymbol{Bu}_3 = \boldsymbol{u}_4, \quad \boldsymbol{Bu}_4 = \boldsymbol{u}_3.$$

(a) Write down all the eigenvalues of $\boldsymbol{B}$.

(b) For each eigenvalue of $\boldsymbol{B}$, write down one eigenvector associated with it.

(c) Is $\boldsymbol{B}$ a diagonalizable matrix? Justify your answer.

15. Two square matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are said to be *similar* if there exists an invertible matrix $\boldsymbol{P}$ such that $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{B}$.

    (a) Suppose $\boldsymbol{A}$ and $\boldsymbol{B}$ are similar matrices.

    (i) Show that $\boldsymbol{A}^n$ is similar to $\boldsymbol{B}^n$ for all positive integer n.

    (ii) If $\boldsymbol{A}$ is invertible, show that $\boldsymbol{B}$ is invertible and $\boldsymbol{A}^{-1}$ is similar to $\boldsymbol{B}^{-1}$.

    (iii) If $\boldsymbol{A}$ is diagonalizable, show that $\boldsymbol{B}$ is diagonalizable.

    (b) Show that the following two matrices are similar:

$$\boldsymbol{A} = \begin{pmatrix} 0 & 1 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & -1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{B} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

16. A square matrix $(a_{ij})_{n\times n}$ is called a *stochastic matrix* if all the entries are non-negative and the sum of entries of each column is 1, i.e. $a_{1i}+a_{2i}+\cdots+a_{ni} = 1$ for $i = 1, 2, \ldots, n$.

    (a) Let $\boldsymbol{A}$ be a stochastic matrix.

    (i) Show that 1 is an eigenvalue of $\boldsymbol{A}$.

    (ii) If λ is an eigenvalue of $\boldsymbol{A}$, then $|\lambda| \le 1$.

    (b) Let $\boldsymbol{B} = \begin{pmatrix} 0.95 & 0 & 0 \\ 0.05 & 0.95 & 0.05 \\ 0 & 0.05 & 0.95 \end{pmatrix}$.

    (i) Is $\boldsymbol{B}$ a stochastic matrix?

    (ii) Find a 3×3 invertible matrix $\boldsymbol{P}$ that diagonalizes $\boldsymbol{B}$.

17. A utility company finds that, in general, if a customer pays a bill late one month, there is about $\frac{1}{2}$ of the time that person will pay before the due date next month; and if a customer pays early one month, there are about $\frac{8}{10}$ of the time that person pays early again the following month. In January, the company finds that all 10,000 customers pay their bills on time. Estimate the number of customers that will pay on time in April. Will the number of customers that pay on time stabilize in the long run? If so, estimate the number of customers that pay on time each month in the long run.

18. In a large city, the soft-drink market was 100% dominated by brand A. Four months ago, two new brands B and C were introduced to the market. According to the market research, for each month, about 1% and 2% of the customers of brand A switch to brands B and C respectively; about 1% and 2% of the customers of brand B switch to brands A and C respectively; and about 2% and 2% of the customers of brand C switch to brands A and B respectively. Compute the present market shares of the three brands of soft drink. Will the market shares stabilize in the long run if the trend continues? If so, estimate the market shares in the long run.

19. Let $\boldsymbol{A}$ be a square matrix. The *exponential* of $\boldsymbol{A}$ is defined to be the matrix

$$e^{\boldsymbol{A}} = \boldsymbol{I} + \boldsymbol{A} + \frac{1}{2!}\boldsymbol{A}^2 + \frac{1}{3!}\boldsymbol{A}^3 + \cdots = \sum_{n=0}^{\infty} \frac{1}{n!}\boldsymbol{A}^n.$$

For each of the following, compute $e^{\boldsymbol{A}}$.

(a) $\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$, (b) $\boldsymbol{A} = \begin{pmatrix} 3 & 0 \\ 8 & -1 \end{pmatrix}$, (c) $\boldsymbol{A} = \begin{pmatrix} -1 & 1 & 1 \\ -2 & 2 & 1 \\ 2 & -1 & 0 \end{pmatrix}$.

20. Following the procedure discussed in Example 6.2.11.2, solve the following recurrence relations.

(a) $a_n = 3a_{n-1} - 2a_{n-2}$ with $a_0 = 0$ and $a_1 = 1$.

(b) $a_n = a_{n-1} + 2a_{n-2}$ with $a_0 = 1$ and $a_1 = 0$.

21. Let d_n be the determinant of the following $n \times n$ matrix:

$$\begin{pmatrix} 3 & 1 & & & & \\ 1 & 3 & 1 & & \text{\Large 0} & \\ & 1 & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & \\ & \text{\Large 0} & & \ddots & 3 & 1 \\ & & & & 1 & 3 \end{pmatrix}.$$

Show that $d_n = 3d_{n-1} - d_{n-2}$. Hence, or otherwise, find d_n.

22. (This is an induction step for proving Remark 6.2.5.3.) Let $\boldsymbol{A}$ be a square matrix of order n. By Theorem 6.2.3, to diagonalize $\boldsymbol{A}$, we need to find n linearly independent eigenvectors.

Suppose we already have m $(< n)$ linearly independent eigenvectors $\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_m}$, say, $\boldsymbol{Au_i} = \lambda\boldsymbol{u_i}$ for $i = 1, 2, \ldots, m$ where $\lambda_1, \lambda_2, \ldots, \lambda_m$ are not necessarily distinct.

For a new eigenvalue μ $(\mu \neq \lambda_i$ for $i = 1, 2, \ldots, m)$ of $\boldsymbol{A}$, let $\{\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_p}\}$ be a basis for the eigenspace E_μ. Prove that $\{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_m}, \boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_p}\}$ is linearly independent.

(Hint: Consider the vector equation

$$a_1\boldsymbol{u_1} + a_2\boldsymbol{u_2} + \cdots + a_m\boldsymbol{u_m} + b_1\boldsymbol{v_1} + b_2\boldsymbol{v_2} + \cdots + b_p\boldsymbol{v_p} = \boldsymbol{0}.$$

By using the property of eigenvectors, show that

$$a_1(\lambda_1 - \mu)\boldsymbol{u_1} + a_2(\lambda_2 - \mu)\boldsymbol{u_2} + \cdots + a_m(\lambda_m - \mu)\boldsymbol{u_m} = \boldsymbol{0}.$$

Then make use of the linearly independent assumption on $\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_m}$, as well as $\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_p}$, to finish the proof.)

23. Determine which of the following statements are true. Justify your answer.

    (a) If $\boldsymbol{A}$ is a diagonalizable matrix, then $\boldsymbol{A}^{\mathrm{T}}$ is diagonalizable.

    (b) If $\boldsymbol{A}$ and $\boldsymbol{B}$ are diagonalizable matrices of the same size, then $\boldsymbol{A}+\boldsymbol{B}$ is diagonalizable.

    (c) If $\boldsymbol{A}$ and $\boldsymbol{B}$ are diagonalizable matrices of the same size, then $\boldsymbol{AB}$ is diagonalizable.

Question 6.24 to Question 6.34 are exercises for Sections 6.3 and 6.4.

24. For each of the following, find a matrix $\boldsymbol{P}$ that orthogonally diagonalizes $\boldsymbol{A}$ and determine $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{AP}$.

    (a) $\boldsymbol{A} = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$; (b) $\boldsymbol{A} = \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix}$,

    (c) $\boldsymbol{A} = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}$, (d) $\boldsymbol{A} = \begin{pmatrix} 2 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 2 \end{pmatrix}$,

    (e) $\boldsymbol{A} = \begin{pmatrix} 0 & -2 & 1 \\ -2 & 3 & -2 \\ 1 & -2 & 0 \end{pmatrix}$, (f) $\boldsymbol{A} = \begin{pmatrix} 0 & -1 & 1 \\ -1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$,

    (g) $\boldsymbol{A} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$, (h) $\boldsymbol{A} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$.

25. Let $\boldsymbol{u}$ be a column matrix.

    (a) Show that $\boldsymbol{I} - \boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}$ is orthogonally diagonalizable.

    (b) Find a matrix $\boldsymbol{P}$ that orthogonally diagonalizes $\boldsymbol{I} - \boldsymbol{u}\boldsymbol{u}^{\mathrm{T}}$ if $\boldsymbol{u} = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$.

26. Let $\boldsymbol{A}$ be a symmetric matrix. If $\boldsymbol{u}$ and $\boldsymbol{v}$ are two eigenvectors of $\boldsymbol{A}$ associated with eigenvalues λ and μ, respectively, where $\lambda \neq \mu$, show that $\boldsymbol{u} \cdot \boldsymbol{v} = 0$.

    (Hint: Compute $\boldsymbol{v}^{\mathrm{T}}\boldsymbol{Au}$ in two different ways.)

27. Let $\boldsymbol{A}$ be a 3×3 symmetric matrix with two eigenvalues 1 and -1. Suppose the eigenspace associated with the eigenvalue 1 represents the plane $x + y - z = 0$. Determine the matrix $\boldsymbol{A}$.

28. Let $\boldsymbol{A}$ be a 4×4 matrix with eigenspaces given by $\text{span}\{(1, 0, 1, 0)^{\mathrm{T}}, (1, 1, 1, 1)^{\mathrm{T}}\}$ and $\text{span}\{(1, 1, -1, -1)^{\mathrm{T}}, (1, -1, -1, 1)^{\mathrm{T}}\}$. Show that $\boldsymbol{A}$ is symmetric.

29. Let $\boldsymbol{A} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$ and $\boldsymbol{u} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$.

    (a) Show that $\boldsymbol{u}$ is an eigenvector of $\boldsymbol{A}$.

    (b) Let $\boldsymbol{v} = (a, b, c, d)^{\mathrm{T}}$ be a nonzero vector. Show that if $\boldsymbol{v} \cdot \boldsymbol{u} = 0$, then $\boldsymbol{v}$ is an eigenvector of $\boldsymbol{A}$.

    (c) Suppose $\boldsymbol{P} = \begin{pmatrix} \frac{1}{2} & a_1 & a_2 & a_3 \\ \frac{1}{2} & b_1 & b_2 & b_3 \\ \frac{1}{2} & c_1 & c_2 & c_3 \\ \frac{1}{2} & d_1 & d_2 & d_3 \end{pmatrix}$ is an orthogonal matrix. Find $\boldsymbol{P}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{P}$.

30. Determine which of the following statements are true. Justify your answer.

    (a) If $\boldsymbol{A}$ and $\boldsymbol{B}$ are orthogonally diagonalizable matrices of the same size, then $\boldsymbol{A}+\boldsymbol{B}$ is orthogonally diagonalizable.

    (b) If $\boldsymbol{A}$ and $\boldsymbol{B}$ are orthogonally diagonalizable matrices of the same size, then $\boldsymbol{AB}$ is orthogonally diagonalizable.

31. For each of the quadratic forms below, (i) rewrite the form using matrix notation; and (ii) simplify the form by following the procedure in Discussion 6.4.4 and Example 6.4.5.

    (a) $Q_1(x, y) = 5x^2 + 5y^2 - 4xy$,

    (b) $Q_2(x, y, z) = 7x^2 + 6y^2 + 5z^2 - 4xy + 4yz$. (Hint: For Part (ii), 3 is one of the eigenvalues of the corresponding matrix.)

32. (a) Let λ_1, λ_2 and λ_3 be real numbers such that $\lambda_1 \leq \lambda_2 \leq \lambda_3$.

    (i) Show that λ_1 is the minimum value of $\lambda_1 x_1^2 + \lambda_2 x_2^2 + \lambda_3 x_3^2$ for all real numbers x_1, x_2, x_3 satisfying $x_1^2 + x_2^2 + x_3^2 = 1$.

(ii) Show that λ_3 is the maximum value of $\lambda_1 x_1^2+\lambda_2 x_2^2+\lambda_3 x_3^2$ for all real numbers x_1, x_2, x_3 satisfying $x_1^2+x_2^2+x_3^2=1$.

(b) For each of the following 3×3 matrices $\boldsymbol{Q}$, find the minimum and maximum values of $\boldsymbol{u}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{u}$ for all unit vectors $\boldsymbol{u}$ (i.e. $\|\boldsymbol{u}\|=1$) in $\mathbb{R}^3$

(i) $\boldsymbol{Q}=\begin{pmatrix}1&0&0\\0&2&0\\0&0&3\end{pmatrix}$, (ii) $\boldsymbol{Q}=\begin{pmatrix}2&-1&0\\-1&2&-1\\0&-1&2\end{pmatrix}$.

33. For each of the following quadratic equations, find the quadratic form associated with it, name the conic section and give its equation in the standard form.

(a) $x^2+2y^2-2x+8y+8=0$,

(b) $x^2-4x+4y+4=0$,

(c) $2x^2-4xy-y^2+8=0$,

(d) $x^2+xy+y^2=6$,

(e) $11x^2+24xy+4y^2-15=0$,

(f) $9x^2-4xy+6y^2-10x-20y-5=0$,

(g) $9x^2+6xy+y^2-10\sqrt{10}x+10\sqrt{10}y+90=0$.

34. Let $\boldsymbol{A}$ be a 2×2 symmetric matrix such that the eigenvalues of $\boldsymbol{A}$ are 1 and 4. The equation

$$\begin{pmatrix}x & y\end{pmatrix}\boldsymbol{A}\begin{pmatrix}x\\y\end{pmatrix}=8$$

represents a non-degenerate conic section in $\mathbb{R}^2$. Name the conic section and write its equation in the standard form. Justify your answer.

Chapter 7

Linear Transformations

In this chapter, all vectors are written as column vectors.

Section 7.1 Linear Transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$

Definition 7.1.1 A *linear transformation* is a mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ of the form

$$T\left(\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}\right) = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix} \quad \text{for } \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$$

where $a_{11}, a_{12}, \ldots, a_{mn}$ are real numbers. In particular, if $n = m$, T is also called a *linear operator* on $\mathbb{R}^n$. We can rewrite the formula of T as

$$T\left(\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}\right) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

The matrix $(a_{ij})_{m\times n}$ above is called the *standard matrix* for T.

Example 7.1.2

1. Let $I : \mathbb{R}^n \to \mathbb{R}^n$ be the *identity transformation* defined by

$$I(\boldsymbol{u}) = \boldsymbol{u} \quad \text{for } \boldsymbol{u} \in \mathbb{R}^n.$$

   It is a linear operator on $\mathbb{R}^n$ and the standard matrix is the identity matrix $\boldsymbol{I_n}$.

2. Let $O : \mathbb{R}^n \to \mathbb{R}^m$ be the *zero transformation* defined by

$$O(\boldsymbol{u}) = \boldsymbol{0} \quad \text{for } \boldsymbol{u} \in \mathbb{R}^n.$$

It is a linear transformation and the standard matrix is the zero matrix $\boldsymbol{0}_{m \times n}$.

3. Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be defined by

$$T\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} x+y \\ 2x \\ -3y \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & 0 \\ 0 & -3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad \text{for } \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2.$$

It is a linear transformation and the standard matrix is $\begin{pmatrix} 1 & 1 \\ 2 & 0 \\ 0 & -3 \end{pmatrix}$.

Remark 7.1.3 For abstract vector spaces, linear transformations have a different definition: Let V and W be vector spaces. A mapping $T : V \to W$ is called a *linear transformation* if $T(c\boldsymbol{u} + d\boldsymbol{v}) = cT(\boldsymbol{u}) + dT(\boldsymbol{v})$ for all $\boldsymbol{u}, \boldsymbol{v} \in V$ and $c, d \in \mathbb{R}$.

It can be shown that the two definitions are the same if $V = \mathbb{R}^n$ and $W = \mathbb{R}^m$. (See Question 7.4.)

Theorem 7.1.4 Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation.

1. $T(\boldsymbol{0}) = \boldsymbol{0}$.
2. If $\boldsymbol{u_1}, \boldsymbol{u_2}, \dots, \boldsymbol{u_k} \in \mathbb{R}^n$ and $c_1, c_2, \dots, c_k \in \mathbb{R}$, then

$$T(c_1\boldsymbol{u_1} + c_2\boldsymbol{u_2} + \cdots + c_k\boldsymbol{u_k}) = c_1T(\boldsymbol{u_1}) + c_2T(\boldsymbol{u_2}) + \cdots + c_kT(\boldsymbol{u_k}).$$

Proof Let $\boldsymbol{A}$ be the standard matrix for T, i.e. $T(\boldsymbol{u}) = \boldsymbol{A}\boldsymbol{u}$ for all $\boldsymbol{u} \in \mathbb{R}^n$.

1. $T(\boldsymbol{0}) = \boldsymbol{A}\boldsymbol{0} = \boldsymbol{0}$.
2. $$\begin{aligned} T(c_1\boldsymbol{u_1} + c_2\boldsymbol{u_2} + \cdots + c_k\boldsymbol{u_k}) &= \boldsymbol{A}(c_1\boldsymbol{u_1} + c_2\boldsymbol{u_2} + \cdots + c_k\boldsymbol{u_k}) \\ &= c_1\boldsymbol{A}\boldsymbol{u_1} + c_2\boldsymbol{A}\boldsymbol{u_2} + \cdots + c_k\boldsymbol{A}\boldsymbol{u_k} \\ &= c_1T(\boldsymbol{u_1}) + c_2T(\boldsymbol{u_2}) + \cdots + c_kT(\boldsymbol{u_k}). \end{aligned}$$

Example 7.1.5

1. Let $T_1 : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$T_1\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} x+1 \\ y+3 \end{pmatrix} \quad \text{for } \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2.$$

It is a not a linear transformation.

Note that $T_1\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$ which violates Theorem 7.1.4.1.

2. Let $T_2 : \mathbb{R}^3 \to \mathbb{R}^2$ be defined by

$$T_2\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = \begin{pmatrix} x^2 \\ yz \end{pmatrix} \quad \text{for} \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3.$$

It is a not a linear transformation.

Note that T_5 violates Theorem 7.1.4.2. For example,

$$T_2\left(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}\right) = T_2\left(\begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} 4 \\ 0 \end{pmatrix}$$

and

$$T_2\left(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}\right) + T_2\left(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}.$$

Discussion 7.1.6 Let $\{\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_n}\}$ be a basis for $\mathbb{R}^n$. Any vector $\boldsymbol{v} \in \mathbb{R}^n$ can be written as

$$\boldsymbol{v} = c_1\boldsymbol{u_1} + c_2\boldsymbol{u_2} + \cdots + c_n\boldsymbol{u_n}$$

for some $c_1, c_2, \ldots, c_n \in \mathbb{R}$. Suppose $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation. By Theorem 7.1.4.2,

$$T(\boldsymbol{v}) = c_1T(\boldsymbol{u_1}) + c_2T(\boldsymbol{u_2}) + \cdots + c_nT(\boldsymbol{u_n}).$$

It follows that the image $T(\boldsymbol{v})$ of $\boldsymbol{v}$ is completely determined by the images $T(\boldsymbol{u_1})$, $T(\boldsymbol{u_2})$, $\ldots$, $T(\boldsymbol{u_n})$ of the basis vectors $\boldsymbol{u_1}$, $\boldsymbol{u_2}$, $\ldots$, $\boldsymbol{u_n}$.

Example 7.1.7 Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be the linear transformation such that

$$T\left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\right) = \begin{pmatrix} 1 \\ 3 \end{pmatrix}, \quad T\left(\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}\right) = \begin{pmatrix} -1 \\ 2 \end{pmatrix}, \quad T\left(\begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix}\right) = \begin{pmatrix} 4 \\ -1 \end{pmatrix}.$$

1. Find the image of the vector $(-1, 4, 6)^{\mathrm{T}}$ under T.

2. Find the formula of T.

Solution

1. We first observe that the vectors

$$\left\{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix}\right\}$$

form a basis for $\mathbb{R}^3$ (check it). Thus the given information determines T completely.

To find the image of $(-1,4,6)^{\mathrm{T}}$, we first write it as a linear combination of vectors from our basis. That is, we find constants c_1, c_2 and c_3 such that

$$\begin{pmatrix}-1\\4\\6\end{pmatrix} = c_1\begin{pmatrix}1\\1\\1\end{pmatrix} + c_2\begin{pmatrix}0\\1\\1\end{pmatrix} + c_3\begin{pmatrix}2\\0\\-1\end{pmatrix}.$$

Solving the equation, we get $c_1 = 3$, $c_2 = 1$ and $c_3 = -2$. Thus the required image is

$$\begin{aligned}
T\left(\begin{pmatrix}-1\\4\\6\end{pmatrix}\right) &= T\left(3\begin{pmatrix}1\\1\\1\end{pmatrix} + \begin{pmatrix}0\\1\\1\end{pmatrix} - 2\begin{pmatrix}2\\0\\-1\end{pmatrix}\right)\\
&= 3\,T\left(\begin{pmatrix}1\\1\\1\end{pmatrix}\right) + T\left(\begin{pmatrix}0\\1\\1\end{pmatrix}\right) - 2\,T\left(\begin{pmatrix}2\\0\\-1\end{pmatrix}\right)\\
&= 3\begin{pmatrix}1\\3\end{pmatrix} + \begin{pmatrix}-1\\2\end{pmatrix} - 2\begin{pmatrix}4\\-1\end{pmatrix} = \begin{pmatrix}-6\\13\end{pmatrix}.
\end{aligned}$$

2. Let $(x,y,z)^{\mathrm{T}}$ be any vector in $\mathbb{R}^3$. Following the procedure of Part (a), by solving the equation

$$\begin{pmatrix}x\\y\\z\end{pmatrix} = c_1\begin{pmatrix}1\\1\\1\end{pmatrix} + c_2\begin{pmatrix}0\\1\\1\end{pmatrix} + c_3\begin{pmatrix}2\\0\\-1\end{pmatrix},$$

where x, y, z are regarded as constants, we obtain $c_1 = x-2y+2z$, $c_2 = -x+3y-2z$ and $c_3 = y-z$, i.e.

$$\begin{pmatrix}x\\y\\z\end{pmatrix} = (x-2y+2z)\begin{pmatrix}1\\1\\1\end{pmatrix} + (-x+3y-2z)\begin{pmatrix}0\\1\\1\end{pmatrix} + (y-z)\begin{pmatrix}2\\0\\-1\end{pmatrix}.$$

So the general formula of T is

$$\begin{aligned}
T\left(\begin{pmatrix}x\\y\\z\end{pmatrix}\right) &= (x-2y+2z)\begin{pmatrix}1\\3\end{pmatrix} + (-x+3y-2z)\begin{pmatrix}-1\\2\end{pmatrix} + (y-z)\begin{pmatrix}4\\-1\end{pmatrix}\\
&= \begin{pmatrix}2x-y\\x-y+3z\end{pmatrix}.
\end{aligned}$$

Discussion 7.1.8 Let $T: \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation with the standard matrix $\boldsymbol{A}$. Let $\{\boldsymbol{e_1}, \boldsymbol{e_2}, \ldots, \boldsymbol{e_n}\}$ be the standard basis for $\mathbb{R}^n$. By Discussion 7.1.6, we know that the images $T(\boldsymbol{e_1}), T(\boldsymbol{e_2}), \ldots, T(\boldsymbol{e_n})$ completely define T. Furthermore, since for each $\boldsymbol{e_i}$,

$$T(\boldsymbol{e_i}) = \boldsymbol{A}\boldsymbol{e_i} = \text{the } i\text{th column of } \boldsymbol{A},$$

we have $\boldsymbol{A} = \begin{pmatrix}T(\boldsymbol{e_1}) & T(\boldsymbol{e_2}) & \cdots & T(\boldsymbol{e_n})\end{pmatrix}$.

Example 7.1.9 We use the linear transformation defined in Example 7.1.7. Instead of finding the formula of T directly as in Example 7.1.7.2, we compute the standard matrix following Discussion 7.1.8.

By

$$\left(\begin{array}{ccc|c|c|c} 1 & 0 & 2 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & -1 & 0 & 0 & 1 \end{array}\right) \xrightarrow[\text{Elimination}]{\text{Gauss-Jordan}} \left(\begin{array}{ccc|c|c|c} 1 & 0 & 0 & 1 & -2 & 2 \\ 0 & 1 & 0 & -1 & 3 & -2 \\ 0 & 0 & 1 & 0 & 1 & -1 \end{array}\right),$$

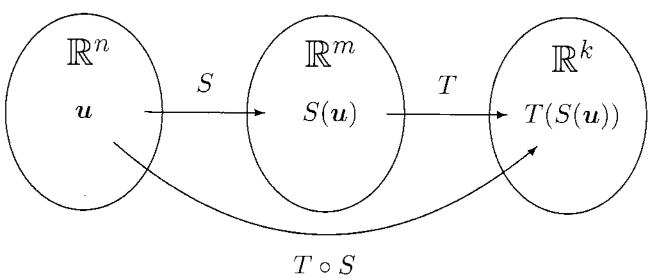
$$\begin{pmatrix}1\\0\\0\end{pmatrix} = \begin{pmatrix}1\\1\\1\end{pmatrix} - \begin{pmatrix}0\\1\\1\end{pmatrix}, \quad \begin{pmatrix}0\\1\\0\end{pmatrix} = -2\begin{pmatrix}1\\1\\1\end{pmatrix} + 3\begin{pmatrix}0\\1\\1\end{pmatrix} + \begin{pmatrix}2\\0\\-1\end{pmatrix}, \quad \begin{pmatrix}0\\0\\1\end{pmatrix} = 2\begin{pmatrix}1\\1\\1\end{pmatrix} - 2\begin{pmatrix}0\\1\\1\end{pmatrix} - \begin{pmatrix}2\\0\\-1\end{pmatrix}.$$

Then we have $T\left(\begin{pmatrix}1\\0\\0\end{pmatrix}\right) = T\left(\begin{pmatrix}1\\1\\1\end{pmatrix}\right) - T\left(\begin{pmatrix}0\\1\\1\end{pmatrix}\right) = \begin{pmatrix}1\\3\end{pmatrix} - \begin{pmatrix}-1\\2\end{pmatrix} = \begin{pmatrix}2\\1\end{pmatrix}$ and similarly, $T\left(\begin{pmatrix}0\\1\\0\end{pmatrix}\right) = \begin{pmatrix}-1\\-1\end{pmatrix}$, $T\left(\begin{pmatrix}0\\0\\1\end{pmatrix}\right) = \begin{pmatrix}0\\3\end{pmatrix}$. So the standard matrix for T is

$$\left(T\left(\begin{pmatrix}1\\0\\0\end{pmatrix}\right) \quad T\left(\begin{pmatrix}0\\1\\0\end{pmatrix}\right) \quad T\left(\begin{pmatrix}0\\0\\1\end{pmatrix}\right)\right) = \begin{pmatrix}2 & -1 & 0\\1 & -1 & 3\end{pmatrix}.$$

Definition 7.1.10 Let $S : \mathbb{R}^n \to \mathbb{R}^m$ and $T : \mathbb{R}^m \to \mathbb{R}^k$ be linear transformations. The *composition* of T with S, denoted by $T \circ S$, is a mapping from $\mathbb{R}^n$ to $\mathbb{R}^k$ such that

$$(T \circ S)(\boldsymbol{u}) = T(S(\boldsymbol{u})) \quad \text{for } \boldsymbol{u} \in \mathbb{R}^n.$$



Theorem 7.1.11 If $S : \mathbb{R}^n \to \mathbb{R}^m$ and $T : \mathbb{R}^m \to \mathbb{R}^k$ are linear transformations, then $T \circ S : \mathbb{R}^n \to \mathbb{R}^k$ is again a linear transformation. Furthermore, if $\boldsymbol{A}$ and $\boldsymbol{B}$ are the standard matrices for the linear transformations S and T respectively, then the standard matrix for the composition $T \circ S$ is given by $\boldsymbol{BA}$.

Proof For all $\boldsymbol{u} \in \mathbb{R}^n$,

$$(T \circ S)(\boldsymbol{u}) = T(S(\boldsymbol{u})) = T(\boldsymbol{A}\boldsymbol{u}) = \boldsymbol{B}\boldsymbol{A}\boldsymbol{u}.$$

So $T \circ S$ is a linear transformation and its standard matrix is $\boldsymbol{BA}$.

Example 7.1.12 Let $S : \mathbb{R}^3 \to \mathbb{R}^2$ be the linear transformation defined by

$$S\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = \begin{pmatrix} x+y \\ z \end{pmatrix} \quad \text{for} \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3$$

and $T : \mathbb{R}^2 \to \mathbb{R}^3$ the linear transformation defined by

$$T\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} y \\ y \\ x \end{pmatrix} \quad \text{for} \quad \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2.$$

Then $T \circ S : \mathbb{R}^3 \to \mathbb{R}^3$ is given by

$$(T \circ S)\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = T\left(S\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right)\right) = T\left(\begin{pmatrix} x+y \\ z \end{pmatrix}\right) = \begin{pmatrix} z \\ z \\ x+y \end{pmatrix} \quad \text{for} \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3.$$

The standard matrices for S, T and $T \circ S$ are

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

respectively. Note that

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Section 7.2 Ranges and Kernels

Definition 7.2.1 Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. The *range* of T, denoted by $\mathrm{R}(T)$, is the set of images of T, i.e.

$$\mathrm{R}(T) = \{\, T(\boldsymbol{u}) \mid \boldsymbol{u} \in \mathbb{R}^n \,\} \subseteq \mathbb{R}^m.$$

Example 7.2.2 Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be a linear transformation defined by

$$T\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} x+y \\ y \\ x \end{pmatrix} \quad \text{for} \quad \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2.$$

Then

$$\mathrm{R}(T) = \left\{ \begin{pmatrix} x+y \\ y \\ x \end{pmatrix} \,\middle|\, x, y \in \mathbb{R} \right\}.$$

Since $(x+y,\, y,\, x)^{\mathrm{T}} = x(1,0,1)^{\mathrm{T}} + y(1,1,0)^{\mathrm{T}}$,

$$\mathrm{R}(T) = \mathrm{span}\{(1,0,1)^{\mathrm{T}},\ (1,1,0)^{\mathrm{T}}\}.$$

In fact, $\mathrm{R}(T)$ is the plane $x - y - z = 0$ in $\mathbb{R}^3$.

Discussion 7.2.3 Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and let $\{\boldsymbol{u_1}, \boldsymbol{u_2}, \dots, \boldsymbol{u_n}\}$ be a basis for $\mathbb{R}^n$. In Discussion 7.1.6, the image of every vector $\boldsymbol{v} \in \mathbb{R}^n$ under T is a linear combination of $T(\boldsymbol{u_1}), T(\boldsymbol{u_2}), \dots, T(\boldsymbol{u_n})$, i.e. $T(\boldsymbol{v}) \in \mathrm{span}\{T(\boldsymbol{u_1}), T(\boldsymbol{u_2}), \dots, T(\boldsymbol{u_n})\}$ for all $\boldsymbol{v} \in \mathbb{R}^n$. Hence $\mathrm{R}(T) \subseteq \mathrm{span}\{T(\boldsymbol{u_1}), T(\boldsymbol{u_2}), \dots, T(\boldsymbol{u_n})\}$.

Conversely, every linear combination of $T(\boldsymbol{u_1}), T(\boldsymbol{u_2}), \dots, T(\boldsymbol{u_n})$ is an element of $\mathrm{R}(T)$ (why?).

So $\mathrm{R}(T) = \mathrm{span}\{T(\boldsymbol{u_1}), T(\boldsymbol{u_2}), \dots, T(\boldsymbol{u_n})\}$.

Theorem 7.2.4 Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and $\boldsymbol{A}$ the standard matrix for T. Then

$$\mathrm{R}(T) = \text{the column space of } \boldsymbol{A}$$

which is a subspace of $\mathbb{R}^m$.

Proof Take the standard basis $\{\boldsymbol{e_1}, \boldsymbol{e_2}, \dots, \boldsymbol{e_n}\}$ for $\mathbb{R}^n$. By Discussion 7.1.8, $T(\boldsymbol{e_1}), T(\boldsymbol{e_2}), \dots, T(\boldsymbol{e_n})$ are the n columns of $\boldsymbol{A}$. Following Discussion 7.2.3, we have

$$\mathrm{R}(T) = \mathrm{span}\{T(\boldsymbol{e_1}), T(\boldsymbol{e_2}), \dots, T(\boldsymbol{e_n})\} = \text{the column space of } \boldsymbol{A}.$$

Definition 7.2.5 Let T be a linear transformation. The dimension of $\mathrm{R}(T)$ is called the *rank* of T and is denoted by $\mathrm{rank}(T)$.

By Theorem 7.2.4, if $\boldsymbol{A}$ is the standard matrix for T, then $\mathrm{rank}(T) = \mathrm{rank}(\boldsymbol{A})$.

Example 7.2.6 Let $T : \mathbb{R}^4 \to \mathbb{R}^4$ be a linear transformation defined by

$$T\left(\begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix}\right) = \begin{pmatrix} x+2y+z \\ x+3y \\ x+4y-z \\ y-z \end{pmatrix} \quad \text{for} \quad \begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix} \in \mathbb{R}^4.$$

Find a basis for the range of T and determine the rank of T.

Solution The range of T is equal to the column space of the matrix

$$\boldsymbol{A} = \begin{pmatrix} 0 & 1 & 2 & 1 \\ 0 & 1 & 3 & 0 \\ 0 & 1 & 4 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix}.$$

By Gaussian Elimination, we reduce $\boldsymbol{A}$ to

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

So $\{(1,1,1,0)^{\mathrm{T}},\ (2,3,4,1)^{\mathrm{T}}\}$ is a basis for $\mathrm{R}(T)$ and

$$\mathrm{rank}(T) = \dim(\mathrm{R}(T)) = 2.$$

Definition 7.2.7 Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. The *kernel* of T, denoted by $\mathrm{Ker}(T)$, is the set of vectors in $\mathbb{R}^n$ whose image is the zero vector in $\mathbb{R}^m$, i.e.

$$\mathrm{Ker}(T) = \{\, \boldsymbol{u} \mid T(\boldsymbol{u}) = \boldsymbol{0} \,\} \subseteq \mathbb{R}^n.$$

Example 7.2.8

1. Let $T_1 : \mathbb{R}^3 \to \mathbb{R}^4$ be a linear transformation defined by

$$T_1\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = \begin{pmatrix} 2x - y \\ x - y + 3z \\ -5x + y \\ x - z \end{pmatrix} \quad \text{for} \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3.$$

The kernel of T_1 is the set of vectors $\boldsymbol{u} \in \mathbb{R}^3$ such that $T_1(\boldsymbol{u}) = \boldsymbol{0}$. That is, we need to solve

$$T_1\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

This is equivalent to

$$\begin{pmatrix} 2 & -1 & 0 \\ 1 & -1 & 3 \\ -5 & 1 & 0 \\ 1 & 0 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

By solving the linear system, we get only the trivial solution $x = 0$, $y = 0$, $z = 0$. Thus

$$\mathrm{Ker}(T_1) = \{(0,0,0)^{\mathrm{T}}\}.$$

2. Let $T_2 : \mathbb{R}^3 \to \mathbb{R}^3$ be a linear transformation defined by

$$T_2\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = \begin{pmatrix} z - y \\ 0 \\ x \end{pmatrix} \quad \text{for } \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3.$$

By equating $\begin{pmatrix} z - y \\ 0 \\ x \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$, we get $x = 0$ and $y = z$. So

$$\mathrm{Ker}(T_2) = \left\{ \begin{pmatrix} 0 \\ y \\ y \end{pmatrix} \middle| \; y \in \mathbb{R} \right\} = \mathrm{span}\{(0, 1, 1)^{\mathrm{T}}\}.$$

Theorem 7.2.9 Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and $\boldsymbol{A}$ the standard matrix for T. Then

$$\mathrm{Ker}(T) = \text{the nullspace of } \boldsymbol{A}$$

which is a subspace of $\mathbb{R}^n$.

Proof $\mathrm{Ker}(T) = \{\, \boldsymbol{u} \mid T(\boldsymbol{u}) = \boldsymbol{0} \,\} = \{\, \boldsymbol{u} \mid \boldsymbol{A}\boldsymbol{u} = \boldsymbol{0} \,\}$ which is the nullspace of $\boldsymbol{A}$.

Definition 7.2.10 Let T be a linear transformation. The dimension of $\mathrm{Ker}(T)$ is called the *nullity* of T and is denoted by $\mathrm{nullity}(T)$.

By Theorem 7.2.9, if $\boldsymbol{A}$ is the standard matrix for T, then $\mathrm{nullity}(T) = \mathrm{nullity}(\boldsymbol{A})$.

Example 7.2.11

1. The nullity of the linear transformation in Example 7.2.8.1 is 0 while the nullity of the linear transformation in Example 7.2.8.2 is 1.

2. Let $T : \mathbb{R}^4 \to \mathbb{R}^4$ be a linear transformation defined by

$$T\left(\begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix}\right) = \begin{pmatrix} x + 2y + z \\ x + 3y \\ x + 4y - z \\ y - z \end{pmatrix} \quad \text{for } \begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix} \in \mathbb{R}^4.$$

Find a basis for the kernel of T and determine the nullity of T.

Solution The kernel is found by solving

$$T\left(\begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix}\right) = \begin{pmatrix} x + 2y + z \\ x + 3y \\ x + 4y - z \\ y - z \end{pmatrix} = \begin{pmatrix} 0 & 1 & 2 & 1 \\ 0 & 1 & 3 & 0 \\ 0 & 1 & 4 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} w \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

The linear system has a general solution $w = s$, $x = -3t$, $y = t$, $z = t$ where s, t are arbitrary parameters. Thus the kernel is the subspace

$$\mathrm{Ker}(T) = \left\{ s\begin{pmatrix}1\\0\\0\\0\end{pmatrix} + t\begin{pmatrix}0\\-3\\1\\1\end{pmatrix} \;\middle|\; s, t \in \mathbb{R} \right\},$$

$\{(1,0,0,0)^{\mathrm{T}}, (0,-3,1,1)^{\mathrm{T}}\}$ is a basis for the kernel and

$$\mathrm{nullity}(T) = \dim(\mathrm{Ker}(T)) = 2.$$

Theorem 7.2.12 (Dimension Theorem for Linear Transformations) If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, then

$$\mathrm{rank}(T) + \mathrm{nullity}(T) = n.$$

Proof The theorem is a restatement of Theorem 4.3.4, the Dimension Theorem for Matrices: If $\boldsymbol{A}$ is an $m \times n$ matrix, then $\mathrm{rank}(\boldsymbol{A}) + \mathrm{nullity}(\boldsymbol{A}) = n$.

If we define a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ by $T(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$ for $\boldsymbol{x} \in \mathbb{R}^n$. Since $\mathrm{rank}(T) = \mathrm{rank}(\boldsymbol{A})$ and $\mathrm{nullity}(T) = \mathrm{nullity}(\boldsymbol{A})$, we have

$$\mathrm{rank}(T) + \mathrm{nullity}(T) = \mathrm{rank}(\boldsymbol{A}) + \mathrm{nullity}(\boldsymbol{A}) = n.$$

Section 7.3 Geometric Linear Transformations

Discussion 7.3.1 Several well-known geometric transformations on $\mathbb{R}^2$ and $\mathbb{R}^3$ such as scalings, reflections about lines and planes through the origin, and rotations about the origin are in fact linear transformations. In the following, we shall examine some examples of these transformations. Recall that every linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is completely described by the images of the vectors in a basis for $\mathbb{R}^n$. In particular, we can use the image of the standard basis $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n\}$ for $\mathbb{R}^n$ to find the standard matrix for a linear transformation (see Discussion 7.1.8).

Example 7.3.2 (Scalings in $\mathbb{R}^2$)

Suppose $S : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear transformation such that

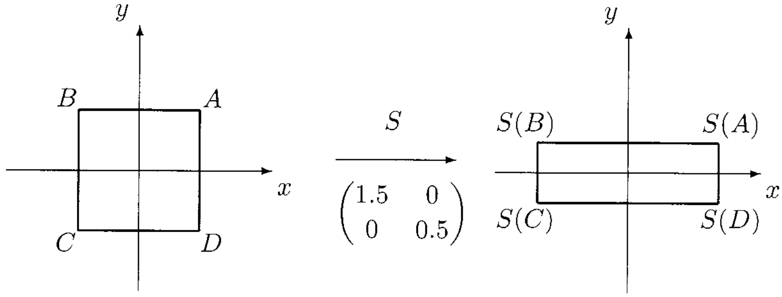
$$S\left(\begin{pmatrix}1\\0\end{pmatrix}\right) = \begin{pmatrix}\lambda_1\\0\end{pmatrix} \quad \text{and} \quad S\left(\begin{pmatrix}0\\1\end{pmatrix}\right) = \begin{pmatrix}0\\\lambda_2\end{pmatrix}$$

for some positive real numbers λ_1 and λ_2. The standard matrix for S is $\begin{pmatrix}\lambda_1 & 0\\0 & \lambda_2\end{pmatrix}$. So

$$S\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \lambda_1 x \\ \lambda_2 y \end{pmatrix}.$$

The effect of this linear transformation is to scale by a factor of λ_1 along the x-axis and by a factor of λ_2 along the y-axis.

For example, in the following, we show the graph of the image of the square with vertices $A = (1,1)^{\mathrm{T}}$, $B = (-1,1)^{\mathrm{T}}$, $C = (-1,-1)^{\mathrm{T}}$, $D = (1,-1)^{\mathrm{T}}$ under S with $\lambda_1 = 1.5$ and $\lambda_2 = 0.5$:



In general, S is called a scaling along the x and y-axes by factors of λ_1 and λ_2 respectively. For the special case when $\lambda_1 = \lambda_2 = \lambda$, the scaling is known as a *dilation* if $\lambda > 1$ and a *contraction* if $0 < \lambda < 1$.

Remark 7.3.3 Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation such that $T(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$ for $x \in \mathbb{R}^2$ where $\boldsymbol{A}$ is a 2×2 matrix.

Suppose there exists a 2×2 invertible matrix $\boldsymbol{P} = \begin{pmatrix} \boldsymbol{u_1} & \boldsymbol{u_2} \end{pmatrix}$ such that

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

for some positive real numbers λ_1 and λ_2. Then

$$T(\boldsymbol{u_1}) = \boldsymbol{A}\boldsymbol{u_1} = \lambda_1 \boldsymbol{u_1} \quad \text{and} \quad T(\boldsymbol{u_2}) = \boldsymbol{A}\boldsymbol{u_2} = \lambda_2 \boldsymbol{u_2}.$$

Thus T can be regarded as a scaling that scales along axes in the directions of $\boldsymbol{u_1}$ and $\boldsymbol{u_2}$ by factors of λ_1 and λ_2 respectively. (In here, the new axes may not be perpendicular to each other.)

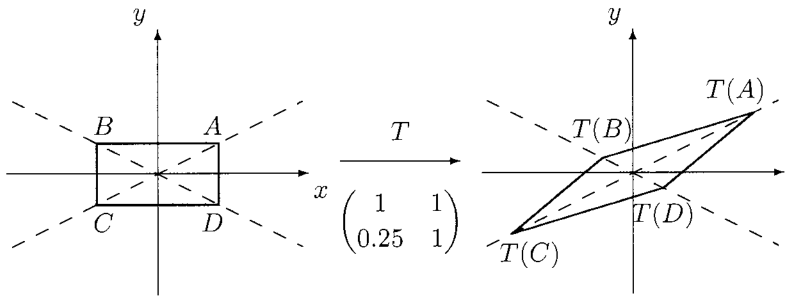
The observation above gives us a geometric interpretation for diagonalizable matrices.

Example 7.3.4 Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation such that the standard matrix for T is $\begin{pmatrix} 1 & 1 \\ 0.25 & 1 \end{pmatrix}$. It is easy to check that

$$\begin{pmatrix} 2 & -2 \\ 1 & 1 \end{pmatrix}^{-1}\begin{pmatrix} 1 & 1 \\ 0.25 & 1 \end{pmatrix}\begin{pmatrix} 2 & -2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1.5 & 0 \\ 0 & 0.5 \end{pmatrix}.$$

Thus T is a scaling that scales along axes in the directions of $(2,1)^{\mathrm{T}}$ and $(-2,1)^{\mathrm{T}}$ by factors of 1.5 and 0.5 respectively.

The following graph shows the image of the rectangle with vertices $A = (2,1)^{\mathrm{T}}$, $B = (-2,1)^{\mathrm{T}}$, $C = (-2,-1)^{\mathrm{T}}$, $D = (2,-1)^{\mathrm{T}}$ under the transformation T:



Example 7.3.5 (Scalings in $\mathbb{R}^3$)

Similar to Example 7.3.2, the standard matrix for the scaling along the x, y and z-axes in $\mathbb{R}^3$ by factors of λ_1, λ_2 and λ_3, respectively, is $\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}$.

For the special case when $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$, the scaling is known as a *dilation* if $\lambda > 1$ and a *contraction* if $0 < \lambda < 1$.

Example 7.3.6 (Reflections in $\mathbb{R}^2$)

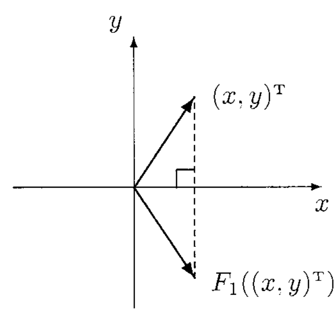
1. Reflections about the usual coordinate axes of $\mathbb{R}^2$:

   Let $F_1 : \mathbb{R}^2 \to \mathbb{R}^2$ be the reflection about the x-axis, i.e.

   $$F_1\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad F_1\left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}.$$

   So the standard matrix for F_1 is given by $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ and

   $$F_1\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ -y \end{pmatrix}.$$

   Similarly, the reflection $F_2 : \mathbb{R}^2 \to \mathbb{R}^2$ about the y-axis has the standard matrix $\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$ and the formula $F_2\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} -x \\ y \end{pmatrix}$.

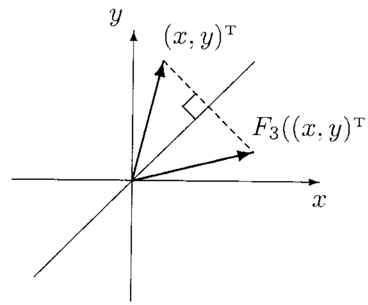
2. The reflection about the line $y = x$:

   Let $F_3 : \mathbb{R}^2 \to \mathbb{R}^2$ be the reflection about the line $y = x$. We have

   $$F_3\left(\begin{pmatrix}1\\0\end{pmatrix}\right) = \begin{pmatrix}0\\1\end{pmatrix} \quad \text{and} \quad F_3\left(\begin{pmatrix}0\\1\end{pmatrix}\right) = \begin{pmatrix}1\\0\end{pmatrix}.$$

   So the standard matrix for F_3 is given by $\begin{pmatrix}0 & 1\\1 & 0\end{pmatrix}$ and

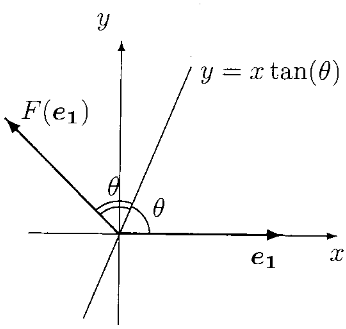
   $$F_3\left(\begin{pmatrix}x\\y\end{pmatrix}\right) = \begin{pmatrix}0 & 1\\1 & 0\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}y\\x\end{pmatrix}.$$

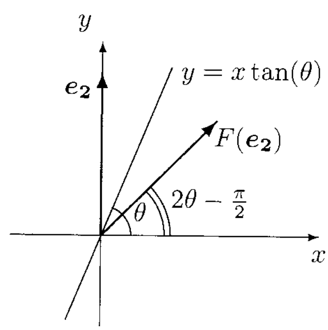


3. Reflections about any line in $\mathbb{R}^2$ that passes through the origin:

   First of all, recall that any line in $\mathbb{R}^2$ (except $x = 0$) that passes through the origin has an equation of the form $y = mx$ where m is the gradient of the line. If θ is the angle between the x-axis and the line $y = mx$, then $m = \tan(\theta)$.

   Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be the reflection about such line. To obtain the standard matrix (and thus the formula) for F, we consider the images of the standard basis vectors $\boldsymbol{e_1} = (1, 0)^{\mathrm{T}}$ and $\boldsymbol{e_2} = (0, 1)^{\mathrm{T}}$ for $\mathbb{R}^2$.





   We have

   $$F\left(\begin{pmatrix}1\\0\end{pmatrix}\right) = F(\boldsymbol{e_1}) = \begin{pmatrix}||F(\boldsymbol{e_1})||\cos(2\theta)\\||F(\boldsymbol{e_1})||\sin(2\theta)\end{pmatrix} = \begin{pmatrix}\cos(2\theta)\\\sin(2\theta)\end{pmatrix}$$

   and

   $$F\left(\begin{pmatrix}0\\1\end{pmatrix}\right) = F(\boldsymbol{e_2}) = \begin{pmatrix}||F(\boldsymbol{e_2})||\cos(2\theta - \frac{\pi}{2})\\||F(\boldsymbol{e_2})||\sin(2\theta - \frac{\pi}{2})\end{pmatrix} = \begin{pmatrix}\cos(\frac{\pi}{2} - 2\theta)\\-\sin(\frac{\pi}{2} - 2\theta)\end{pmatrix}$$
   $$= \begin{pmatrix}\sin(2\theta)\\-\cos(2\theta)\end{pmatrix}.$$

So the standard matrix for F is given by $\begin{pmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{pmatrix}$.

We check that, when $\theta = 0$, $\frac{\pi}{2}$ and $\frac{\pi}{4}$, the standard matrix gives us reflections about the x-axis, y-axis and the line $y = x$ respectively.

Remark 7.3.7 The formula of the reflection F in Example 7.3.6.3 can also be written as

$$F(\boldsymbol{u}) = \boldsymbol{u} - 2(\boldsymbol{u} \cdot \boldsymbol{n})\boldsymbol{n} \quad \text{for } \boldsymbol{u} \in \mathbb{R}^2$$

where $\boldsymbol{n} = (\sin(\theta), -\cos(\theta))^{\mathrm{T}}$. (Check it.) Can you explain the formula geometrically?

Example 7.3.8 (Reflections in $\mathbb{R}^3$)

Similar to Example 7.3.6.1, the standard matrices for reflections about xy-plane, xz-plane and yz-plane in $\mathbb{R}^3$ are

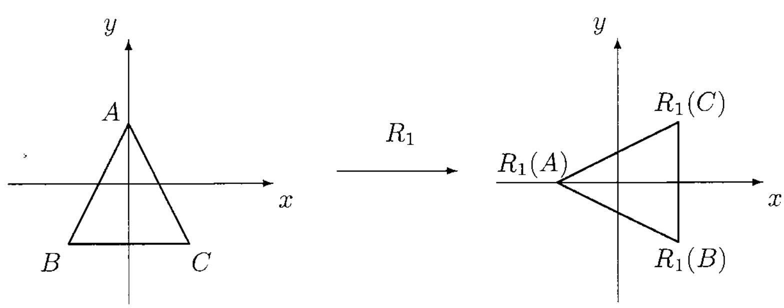
$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

respectively. (See also Question 7.26.)

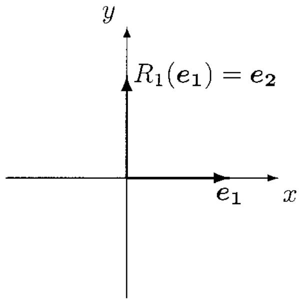
Example 7.3.9 (Rotations in $\mathbb{R}^2$)

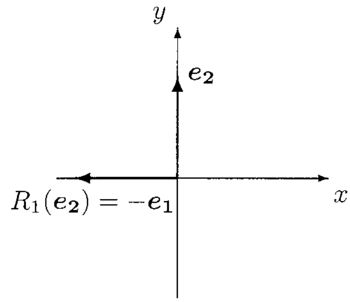
1. Anti-clockwise rotations about the origin through angles $\frac{\pi}{2}$, π and $\frac{3\pi}{2}$:

   Let $R_1 : \mathbb{R}^2 \to \mathbb{R}^2$ be the anti-clockwise rotation about the origin through an angle $\frac{\pi}{2}$. The following graph shows the image of the triangle with vertices $A = (0, 1)^{\mathrm{T}}$, $B = (-1, -1)^{\mathrm{T}}$, $C = (1, -1)^{\mathrm{T}}$ under the rotation R_1:



   To obtain the standard matrix for R_1, we consider the images of the standard basis vectors $\boldsymbol{e_1} = (1, 0)^{\mathrm{T}}$ and $\boldsymbol{e_2} = (0, 1)^{\mathrm{T}}$ for $\mathbb{R}^2$.





We have

$$R_1\left(\begin{pmatrix}1\\0\end{pmatrix}\right) = \begin{pmatrix}0\\1\end{pmatrix} \quad \text{and} \quad R_1\left(\begin{pmatrix}0\\1\end{pmatrix}\right) = \begin{pmatrix}-1\\0\end{pmatrix}.$$

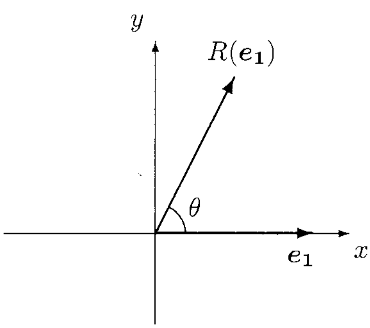
So the standard matrix for R_1 is $\begin{pmatrix}0 & -1\\1 & 0\end{pmatrix}$.

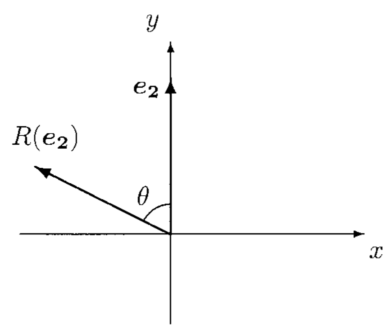
The image of A is $\begin{pmatrix}0 & -1\\1 & 0\end{pmatrix}\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}-1\\0\end{pmatrix}$; the image of B is $\begin{pmatrix}0 & -1\\1 & 0\end{pmatrix}\begin{pmatrix}-1\\-1\end{pmatrix} = \begin{pmatrix}1\\-1\end{pmatrix}$; and the image of C is $\begin{pmatrix}0 & -1\\1 & 0\end{pmatrix}\begin{pmatrix}1\\-1\end{pmatrix} = \begin{pmatrix}1\\1\end{pmatrix}$.

Similarly, the standard matrices for anti-clockwise rotations through angles π and $\frac{3\pi}{2}$ are $\begin{pmatrix}-1 & 0\\0 & -1\end{pmatrix}$ and $\begin{pmatrix}0 & 1\\-1 & 0\end{pmatrix}$ respectively.

2. The anti-clockwise rotation through an angle θ:

   To obtain the standard matrix for R, we consider the images of the standard basis vectors $\boldsymbol{e_1} = (1,0)^{\mathrm{T}}$ and $\boldsymbol{e_2} = (0,1)^{\mathrm{T}}$ for $\mathbb{R}^2$.





We have

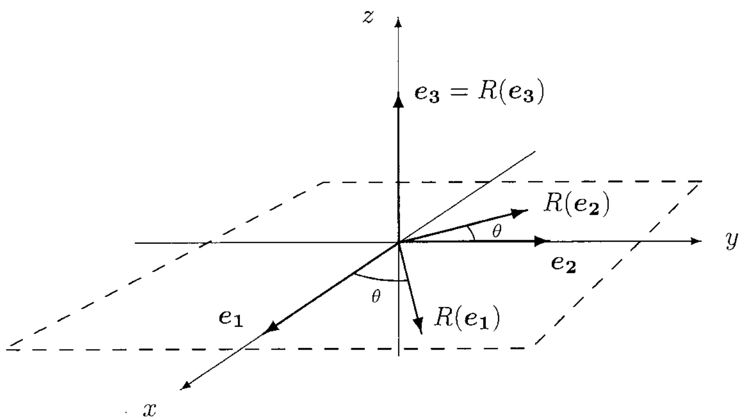
$$R\left(\begin{pmatrix}1\\0\end{pmatrix}\right) = R(\boldsymbol{e_1}) = \begin{pmatrix}||R(\boldsymbol{e_1})||\cos(\theta)\\||R(\boldsymbol{e_1})||\sin(\theta)\end{pmatrix} = \begin{pmatrix}\cos(\theta)\\\sin(\theta)\end{pmatrix}$$

and
$$R\left(\begin{pmatrix}0\\1\end{pmatrix}\right) = R(\boldsymbol{e_2}) = \begin{pmatrix}||R(\boldsymbol{e_2})||\cos(\frac{\pi}{2}+\theta)\\ ||R(\boldsymbol{e_2})||\sin(\frac{\pi}{2}+\theta)\end{pmatrix} = \begin{pmatrix}-\sin(\theta)\\ \cos(\theta)\end{pmatrix}.$$

So the standard matrix for R is given by $\begin{pmatrix}\cos(\theta) & -\sin(\theta)\\ \sin(\theta) & \cos(\theta)\end{pmatrix}$.

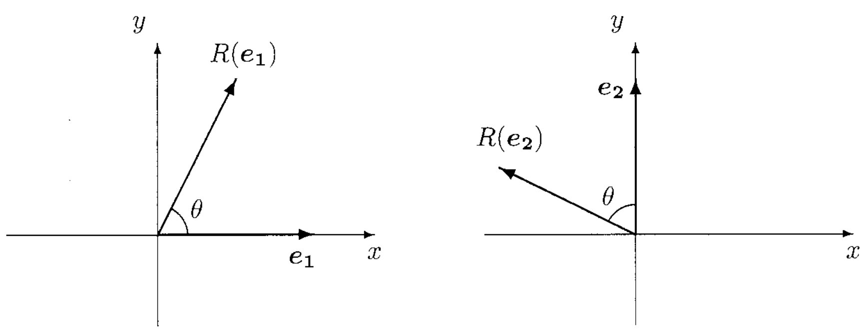
Example 7.3.10 (Rotations in $\mathbb{R}^3$)

Let $R : \mathbb{R}^3 \to \mathbb{R}^3$ be the anti-clockwise rotation about the z-axis through an angle θ. We consider the image of the standard basis vectors $\boldsymbol{e_1} = (1,0,0)^{\mathrm{T}}$, $\boldsymbol{e_2} = (0,1,0)^{\mathrm{T}}$ and $\boldsymbol{e_3} = (0,0,1)^{\mathrm{T}}$ for $\mathbb{R}^3$.



We have the following observations:

1. The z-axis is fixed under the rotation, i.e. $R(\boldsymbol{e_3}) = \boldsymbol{e_3}$.
2. The xy-plane is rotated in the same manner as the rotation in $\mathbb{R}^2$ discussed in Example 7.3.9.2. It is easier to check the images of $\boldsymbol{e_1}$ and $\boldsymbol{e_2}$ by drawing the xy-plane only. (You are reminded that $\boldsymbol{e_1} = (1,0,0)^{\mathrm{T}}$ and $\boldsymbol{e_2} = (0,1,0)^{\mathrm{T}}$ in the following graphs.)



Thus $R(\boldsymbol{e_1}) = [\cos(\theta)]\,\boldsymbol{e_1} + [\sin(\theta)]\,\boldsymbol{e_2}$ and $R(\boldsymbol{e_2}) = -[\sin(\theta)]\,\boldsymbol{e_1} + [\cos(\theta)]\,\boldsymbol{e_2}$.

Summarizing the results above, we have

$$R\left(\begin{pmatrix}1\\0\\0\end{pmatrix}\right)=\begin{pmatrix}\cos(\theta)\\ \sin(\theta)\\ 0\end{pmatrix},\quad R\left(\begin{pmatrix}0\\1\\0\end{pmatrix}\right)=\begin{pmatrix}-\sin(\theta)\\ \cos(\theta)\\ 0\end{pmatrix},\quad R\left(\begin{pmatrix}0\\0\\1\end{pmatrix}\right)=\begin{pmatrix}0\\0\\1\end{pmatrix}.$$

The standard matrix for R is $\begin{pmatrix}\cos(\theta) & -\sin(\theta) & 0\\ \sin(\theta) & \cos(\theta) & 0\\ 0 & 0 & 1\end{pmatrix}$.

Similarly, the standard matrices for rotations in $\mathbb{R}^3$ about x-axis and y-axis are

$$\begin{pmatrix}1 & 0 & 0\\ 0 & \cos(\theta) & -\sin(\theta)\\ 0 & \sin(\theta) & \cos(\theta)\end{pmatrix} \quad\text{and}\quad \begin{pmatrix}\cos(\theta) & 0 & \sin(\theta)\\ 0 & 1 & 0\\ -\sin(\theta) & 0 & \cos(\theta)\end{pmatrix}$$

respectively.

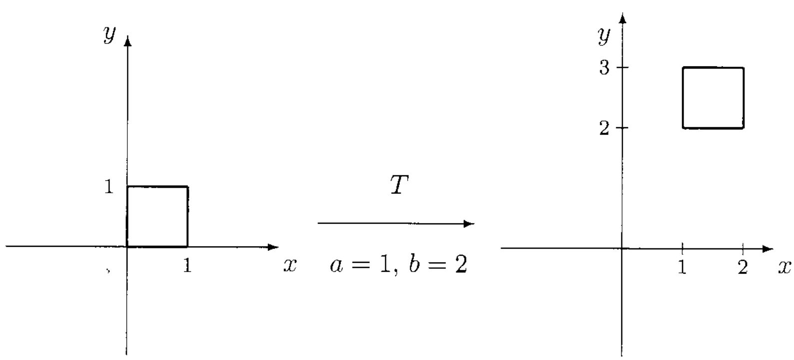
Example 7.3.11 (Translations)

1. Let $T:\mathbb{R}^2\to\mathbb{R}^2$ be a translation in $\mathbb{R}^2$ such that

$$T\left(\begin{pmatrix}x\\y\end{pmatrix}\right)=\begin{pmatrix}x+a\\y+b\end{pmatrix}\quad\text{for } (x,y)^{\mathrm{T}}\in\mathbb{R}^2$$

where a and b are real constants.

For example, the following graph shows the image of the square with vertices $(1,1)^{\mathrm{T}}$, $(0,1)^{\mathrm{T}}$, $(0,0)^{\mathrm{T}}$, $(1,0)^{\mathrm{T}}$ under the translation with $a=1$ and $b=2$:



T is not a linear transformation except when $a=b=0$.

If a,b are not both zero, $T\left(\begin{pmatrix}0\\0\end{pmatrix}\right)=\begin{pmatrix}a\\b\end{pmatrix}\neq\begin{pmatrix}0\\0\end{pmatrix}$ which violates Theorem 7.1.4.1.

2. Let $T' : \mathbb{R}^3 \to \mathbb{R}^3$ be a translation in $\mathbb{R}^3$ such that

$$T'\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = \begin{pmatrix} x+a \\ y+b \\ z+c \end{pmatrix} \quad \text{for all } \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3$$

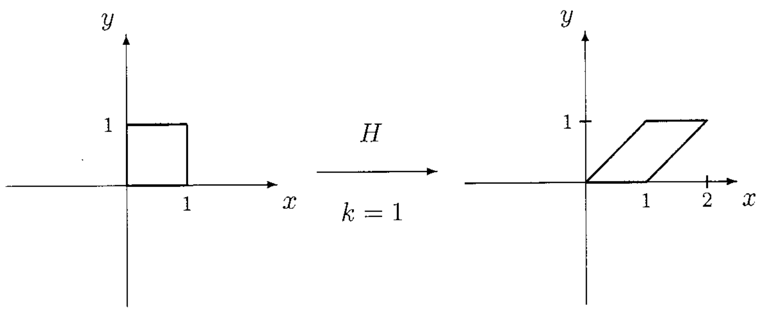
where a, b and c are real constants. Same as translations in $\mathbb{R}^2$, T' is not a linear transformation except when $a = b = c = 0$.

Example 7.3.12 (Shears)

1. A mapping $H : \mathbb{R}^2 \to \mathbb{R}^2$ is called a *shear* in the x-direction by a factor of k if

$$H\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} x+ky \\ y \end{pmatrix} \quad \text{for all } (x, y)^{\mathrm{T}} \in \mathbb{R}^2.$$

For example, the following graph shows the image of the square with vertices $(1, 1)^{\mathrm{T}}$, $(0, 1)^{\mathrm{T}}$, $(0, 0)^{\mathrm{T}}$, $(1, 0)^{\mathrm{T}}$ under the shear in the x-direction with $k = 1$:



Note that any point on the line $y = 1$ is moved to the right by a distance of $k = 1$ under the action of H.

We observe that for all $(x, y)^{\mathrm{T}}$ in $\mathbb{R}^2$,

$$H\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} x+ky \\ y \end{pmatrix} = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.$$

Thus H is a linear transformation and the standard matrix for H is $\begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$.

Similarly, the standard matrix for the shear in the y-direction by a factor of k is $\begin{pmatrix} 1 & 0 \\ k & 1 \end{pmatrix}$.

2. A mapping $H' : \mathbb{R}^3 \to \mathbb{R}^3$ is called a *shear* in the x-direction by a factor of k_1 and in the y-direction by a factor of k_2 if

$$H'\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = \begin{pmatrix} x + k_1 z \\ y + k_2 z \\ z \end{pmatrix} \quad \text{for all } (x, y, z)^{\mathrm{T}} \in \mathbb{R}^3.$$

Same as Part 1, H' is a linear transformation with the standard matrix $\begin{pmatrix} 1 & 0 & k_1 \\ 0 & 1 & k_2 \\ 0 & 0 & 1 \end{pmatrix}$.

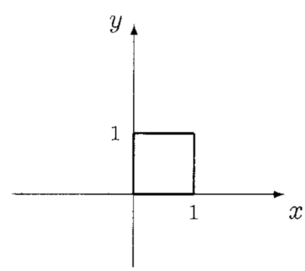
Note that any point on the plane $z = 1$ is translated by k_1 in the x-direction and by k_2 in the y-direction under the action of H'.

Discussion 7.3.13 (2D Computer Graphic)

In 2D (dimension two) computer graphic, a figure is drawn by connecting a set of points by lines. If a figure is drawn by connecting n points, we can store it by a $2 \times n$ matrix. For example, the matrix

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

gives us the square with vertices $(1,1)^{\mathrm{T}}$, $(0,1)^{\mathrm{T}}$, $(0,0)^{\mathrm{T}}$, $(1,0)^{\mathrm{T}}$.

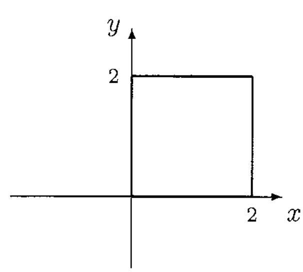


We can transform a figure by changing the positions of the vertices and then redrawing the figure. If the transformation is linear, it can be carried out by pre-multiplying the standard matrix for the transformation to the matrix representing the figure.

For example, if we want to double both the width and the height of the square above, we only need to pre-multiply $\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ to $\boldsymbol{A}$, i.e.

$$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 & 2 & 2 \\ 2 & 2 & 0 & 0 & 2 \end{pmatrix},$$

which gives us the square with vertices $(2,2)^{\mathrm{T}}$, $(0,2)^{\mathrm{T}}$, $(0,0)^{\mathrm{T}}$, $(2,0)^{\mathrm{T}}$.



There are four primary geometric transformations that are used in 2D computer graphics: scalings, reflections, rotations and translations. We know that scalings, reflections and rotations are linear transformations but translations are not. One method to handle translations is to use a new system of coordinates called homogeneous coordinates:

The *homogeneous coordinate system* is formed by equaling each vector in $\mathbb{R}^2$ with a vector in $\mathbb{R}^3$ having the same first two coordinates and having 1 as its third coordinate. For example, the matrix $\boldsymbol{A}$ representing the square with vertices $(1,1)^{\mathrm{T}}$, $(0,1)^{\mathrm{T}}$, $(0,0)^{\mathrm{T}}$, $(1,0)^{\mathrm{T}}$ becomes

$$\boldsymbol{A}' = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

If we want to draw the figure, we simply ignore the third coordinate.

Suppose $\boldsymbol{P}$ is the standard matrix for a geometric linear transformation on $\mathbb{R}^2$ such as a scaling, a reflection or a rotation. Then the matrix

$$\boldsymbol{P}' = \begin{pmatrix} \boldsymbol{P} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 \end{matrix} & 1 \end{pmatrix}$$

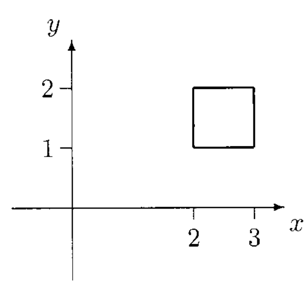
will transform $\boldsymbol{A}'$ accordingly.

For example, the matrix $\begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ can double both the width and the height of the square:

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 & 2 & 2 \\ 2 & 2 & 0 & 0 & 2 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

To do a translation, we need to use a shear defined in $\mathbb{R}^3$. For example, if we want to translate the square by a distance of 2 in the x-direction and by a distance of 1 in the y-direction, the shear with the standard matrix $\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$ will do the job:

$$\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 2 & 2 & 3 & 3 \\ 2 & 2 & 1 & 1 & 2 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$



Exercise 7

Question 7.1 to Question 7.17 are exercises for Sections 7.1 and 7.2.

1. Determine whether the following are linear transformations. Write down the standard matrix for each of the linear transformations.

   (a) $T_1 : \mathbb{R}^2 \to \mathbb{R}^2$ such that $T_1\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} x+y \\ y-x \end{pmatrix}$ for $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$.

   (b) $T_2 : \mathbb{R}^2 \to \mathbb{R}^2$ such that $T_2\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} 2^x \\ 0 \end{pmatrix}$ for $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$.

   (c) $T_3 : \mathbb{R}^2 \to \mathbb{R}^3$ such that $T_3\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} x+y \\ 0 \\ 0 \end{pmatrix}$ for $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$.

   (d) $T_4 : \mathbb{R}^3 \to \mathbb{R}^3$ such that $T_4\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = \begin{pmatrix} 1 \\ y-x \\ y-z \end{pmatrix}$ for $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3$.

   (e) $T_5 : \mathbb{R}^n \to \mathbb{R}$ such that $T_5(\boldsymbol{x}) = \boldsymbol{x} \cdot \boldsymbol{y}$ for $\boldsymbol{x} \in \mathbb{R}^n$ where $\boldsymbol{y} = (y_1, y_2, \dots, y_n)^{\mathrm{T}}$ is a fixed vector in $\mathbb{R}^n$.

   (f) $T_6 : \mathbb{R}^n \to \mathbb{R}$ such that $T_6(\boldsymbol{x}) = \boldsymbol{x} \cdot \boldsymbol{x}$ for $\boldsymbol{x} \in \mathbb{R}^n$.

   (In Parts (e) and (f), $\mathbb{R}$ is regarded as $\mathbb{R}^1$.)

2. For each of the following linear transformations, (i) determine whether there is enough information for us to find the formula of T; and (ii) find the formula and the standard matrix for T if possible.

   (a) $T : \mathbb{R}^3 \to \mathbb{R}^4$ such that

   $$T\left(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} 1 \\ 3 \\ 0 \\ 1 \end{pmatrix}, \quad T\left(\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} 2 \\ 2 \\ -1 \\ 4 \end{pmatrix} \quad \text{and} \quad T\left(\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\right) = \begin{pmatrix} 0 \\ 4 \\ 1 \\ 6 \end{pmatrix}.$$

   (b) $T : \mathbb{R}^2 \to \mathbb{R}^3$ such that

   $$T\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right) = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad T\left(\begin{pmatrix} 2 \\ 2 \end{pmatrix}\right) = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

   (c) $T : \mathbb{R}^2 \to \mathbb{R}^2$ such that

   $$T\left(\begin{pmatrix} 1 \\ -1 \end{pmatrix}\right) = \begin{pmatrix} 2 \\ 0 \end{pmatrix}, \quad T\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right) = \begin{pmatrix} 0 \\ 2 \end{pmatrix} \quad \text{and} \quad T\left(\begin{pmatrix} 2 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} 2 \\ 2 \end{pmatrix}.$$

(d) $T : \mathbb{R}^3 \to \mathbb{R}^2$ such that

$$T\left(\begin{pmatrix}1\\1\\1\end{pmatrix}\right) = \begin{pmatrix}2\\3\end{pmatrix}, \quad T\left(\begin{pmatrix}2\\1\\3\end{pmatrix}\right) = \begin{pmatrix}-1\\0\end{pmatrix} \quad \text{and} \quad T\left(\begin{pmatrix}-1\\1\\2\end{pmatrix}\right) = \begin{pmatrix}0\\1\end{pmatrix}.$$

(e) $T : \mathbb{R}^3 \to \mathbb{R}^3$ such that

$$T\left(\begin{pmatrix}1\\0\\1\end{pmatrix}\right) = \begin{pmatrix}1\\1\\2\end{pmatrix}, \quad T\left(\begin{pmatrix}1\\1\\0\end{pmatrix}\right) = \begin{pmatrix}1\\0\\1\end{pmatrix} \quad \text{and} \quad T\left(\begin{pmatrix}1\\0\\0\end{pmatrix}\right) = \begin{pmatrix}0\\0\\1\end{pmatrix}.$$

(f) $T : \mathbb{R}^3 \to \mathbb{R}$ such that

$$T\left(\begin{pmatrix}1\\-1\\0\end{pmatrix}\right) = -1, \quad T\left(\begin{pmatrix}0\\1\\-1\end{pmatrix}\right) = 1 \quad \text{and} \quad T\left(\begin{pmatrix}-1\\0\\1\end{pmatrix}\right) = 0.$$

3. Let S and T be linear transformations as defined below. Determine the formulae of the compositions $S \circ T$ and $T \circ S$ whenever they are defined.

(a) $S : \mathbb{R}^2 \to \mathbb{R}^3$ such that $S\left(\begin{pmatrix}x\\y\end{pmatrix}\right) = \begin{pmatrix}x+y\\x-y\\x\end{pmatrix}$;

$T : \mathbb{R}^2 \to \mathbb{R}^2$ such that $T\left(\begin{pmatrix}x\\y\end{pmatrix}\right) = \begin{pmatrix}x+y\\x-y\end{pmatrix}$.

(b) $S : \mathbb{R}^2 \to \mathbb{R}^3$ such that $S\left(\begin{pmatrix}x\\y\end{pmatrix}\right) = \begin{pmatrix}y\\x\\x+y\end{pmatrix}$;

$T : \mathbb{R}^3 \to \mathbb{R}^2$ such that $T\left(\begin{pmatrix}x\\y\\z\end{pmatrix}\right) = \begin{pmatrix}-x-y+3z\\-2x-y+3z\end{pmatrix}$.

4. Prove Remark 7.1.3:

Show that a mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation if and only if

$$T(c\boldsymbol{u} + d\boldsymbol{v}) = cT(\boldsymbol{u}) + dT(\boldsymbol{v}) \quad \text{for all } \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n \text{ and } c, d \in \mathbb{R}.$$

5. (a) Let $T_1 : \mathbb{R}^n \to \mathbb{R}^m$ and $T_2 : \mathbb{R}^n \to \mathbb{R}^m$ be linear transformations with standard matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ respectively. Define a mapping $T_1 + T_2 : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$(T_1 + T_2)(\boldsymbol{u}) = T_1(\boldsymbol{u}) + T_2(\boldsymbol{u}) \quad \text{for } \boldsymbol{u} \in \mathbb{R}^n.$$

Show that $T_1 + T_2$ is a linear transformation and find the standard matrix.

(b) Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation with the standard matrix $\boldsymbol{A}$ and let λ be a scalar. Define a mapping $\lambda T : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$(\lambda T)(\boldsymbol{u}) = \lambda T(\boldsymbol{u}) \quad \text{for } \boldsymbol{u} \in \mathbb{R}^n.$$

Show that λT is a linear transformation and find the standard matrix.

6. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear operator. If there exists a linear operator $S : \mathbb{R}^n \to \mathbb{R}^n$ such that $S \circ T$ is the identity transformation, i.e.

$$(S \circ T)(\boldsymbol{u}) = \boldsymbol{u} \text{ for all } \boldsymbol{u} \in \mathbb{R}^n,$$

then T is said to be *invertible* and S is called the *inverse* of T.

(a) For each of the following, determine whether T is invertible and find the inverse of T if possible.

(i) $T : \mathbb{R}^2 \to \mathbb{R}^2$ such that $T\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} y \\ x \end{pmatrix}$ for $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$.

(ii) $T : \mathbb{R}^2 \to \mathbb{R}^2$ such that $T\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} x+y \\ 0 \end{pmatrix}$ for $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$.

(b) Suppose T is invertible and $\boldsymbol{A}$ is the standard matrix for T. Find the standard matrix for the inverse of T.

7. Let $\boldsymbol{n}$ be a unit vector in $\mathbb{R}^n$. Define $P : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$P(\boldsymbol{x}) = \boldsymbol{x} - (\boldsymbol{n} \cdot \boldsymbol{x})\boldsymbol{n} \quad \text{for } \boldsymbol{x} \in \mathbb{R}^n.$$

(a) Show that P is a linear transformation and find the standard matrix for P.

(b) Prove that $P \circ P = P$.

8. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation such that $T \circ T = T$.

(a) If T is not the zero transformation, show that there exists a nonzero vector $\boldsymbol{u} \in \mathbb{R}^n$ such that $T(\boldsymbol{u}) = \boldsymbol{u}$.

(b) If T is not the identity transformation, show that there exists a nonzero vector $\boldsymbol{v} \in \mathbb{R}^n$ such that $T(\boldsymbol{v}) = \boldsymbol{0}$.

(c) Find all linear transformations $T : \mathbb{R}^2 \to \mathbb{R}^2$ such that $T \circ T = T$. (Hint: See Question 6.4.)

9. Let $\boldsymbol{n}$ be a unit vector in $\mathbb{R}^n$. Define $F : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$F(\boldsymbol{x}) = \boldsymbol{x} - 2(\boldsymbol{n} \cdot \boldsymbol{x})\boldsymbol{n} \quad \text{for } \boldsymbol{x} \in \mathbb{R}^n.$$

(See also Remark 7.3.7 and Question 7.26.)

(a) Show that F is a linear transformation and find the standard matrix for F.

(b) Prove that $F \circ F$ is the identity transformation.

(c) Show that the standard matrix for F is an orthogonal matrix.

10. A linear operator T on $\mathbb{R}^n$ is called an *isometry* if $||T(\boldsymbol{u})|| = ||\boldsymbol{u}||$ for all $\boldsymbol{u} \in \mathbb{R}^n$.

(a) If T is an isometry on $\mathbb{R}^n$, show that $T(\boldsymbol{u}) \cdot T(\boldsymbol{v}) = \boldsymbol{u} \cdot \boldsymbol{v}$ for all $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$. (Hint: Compute $T(\boldsymbol{u}+\boldsymbol{v}) \cdot T(\boldsymbol{u}+\boldsymbol{v})$ in two different ways.)

(b) Let $\boldsymbol{A}$ be the standard matrix for a linear operator T. Show that T is an isometry if and only if $\boldsymbol{A}$ is an orthogonal matrix. (See also Question 5.32.)

(c) Find all isometries on $\mathbb{R}^2$. (Hint: See Question 2.57.)

11. Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be a linear transformation defined by

$$T\left(\begin{pmatrix} x \\ y \\ z \end{pmatrix}\right) = \begin{pmatrix} 2x + y \\ x - y + z \end{pmatrix} \quad \text{for} \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3.$$

(a) Find a basis for the range of T.

(b) Find a basis for the kernel of T.

(c) Use this example to verify the Dimension Theorem for Linear Transformation.

(d) Extend the basis found in Part (b) to a basis for $\mathbb{R}^3$.

12. Let $T : \mathbb{R}^4 \to \mathbb{R}^3$ be a linear transformation with the standard matrix

$$\begin{pmatrix} 3 & -1 & 2 & 7 \\ 1 & 2 & 3 & 0 \\ 0 & 1 & 1 & -1 \end{pmatrix}.$$

(a) Find a basis for the range of T.

(b) Find a basis for the kernel of T.

(c) Use this example to verify the Dimension Theorem for Linear Transformation.

13. In each of the following parts, use the given information to find the nullity of the linearly transformation T.

(a) $T : \mathbb{R}^4 \to \mathbb{R}^6$ has rank 2.

(b) The range of $T : \mathbb{R}^6 \to \mathbb{R}^4$ is $\mathbb{R}^4$.

(c) The reduced row-echelon form of the standard matrix for $T : \mathbb{R}^6 \to \mathbb{R}^6$ has 4 nonzero rows.

14. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear operator defined by $T(\boldsymbol{v}) = 2\boldsymbol{v}$.

    (a) What is the kernel of T?

    (b) What is the range of T?

15. Let V be a subspace of $\mathbb{R}^n$. Define a mapping $P : \mathbb{R}^n \to \mathbb{R}^n$ such that for $\boldsymbol{u} \in \mathbb{R}^n$, $P(\boldsymbol{u})$ is the projection of $\boldsymbol{u}$ onto V.

    (a) Show that P is a linear transformation.

    (b) Suppose $n = 3$ and V is the plane $ax + by + cz = 0$ where a, b, c are not all zero. Find $\mathrm{Ker}(P)$ and $\mathrm{R}(P)$.

16. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Show that $\mathrm{Ker}(T) = \{\boldsymbol{0}\}$ if and only if T is one-to-one, i.e. for any two vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$, if $\boldsymbol{u} \neq \boldsymbol{v}$, then $T(\boldsymbol{u}) \neq T(\boldsymbol{v})$.

17. Let $S : \mathbb{R}^n \to \mathbb{R}^m$ and $T : \mathbb{R}^m \to \mathbb{R}^k$ be linear transformations.

    (a) Show that $\mathrm{Ker}(S) \subseteq \mathrm{Ker}(T \circ S)$.

    (b) Show that $\mathrm{R}(T \circ S) \subseteq \mathrm{R}(T)$.

Question 7.18 to Question 7.28 are exercises for Section 7.3.

18. Describe geometrically the effect of each of the following linear operators $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$.

    (a) $\boldsymbol{A} = \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix}$, (b) $\boldsymbol{A} = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}$,

    (c) $\boldsymbol{A} = \begin{pmatrix} -\frac{\sqrt{3}}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}$, (d) $\boldsymbol{A} = \begin{pmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}$,

    (e) $\boldsymbol{A} = \begin{pmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}$, (f) $\boldsymbol{A} = \begin{pmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{\sqrt{3}}{2} \end{pmatrix}$,

    (g) $\boldsymbol{A} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}^{-1}$.

19. Let $F_1 : \mathbb{R}^2 \to \mathbb{R}^2$ and $F_2 : \mathbb{R}^2 \to \mathbb{R}^2$ be two reflections about lines $y = x\tan(\theta)$ and $y = x\tan(\phi)$ respectively. Show that $F_2 \circ F_1$ is an anti-clockwise rotation about the origin through an angle $2(\phi - \theta)$. (Hint: Compute the standard matrix for $F_2 \circ F_1$.)

20. For each of the following linear operators on $\mathbb{R}^2$, (i) find the standard matrices for $T_1 \circ T_2$ and $T_2 \circ T_1$; and (ii) determine whether $T_1 \circ T_2 = T_2 \circ T_1$.

    (a) T_1 is the reflection about the line $x - y = 0$ and T_2 is the dilation by a factor of 2.

    (b) T_1 is the reflection about the line $x - y = 0$ and T_2 is the scaling along the x and y-axes by factors of 1 and 2 respectively.

    (c) T_1 is the reflection about the line $x - y = 0$ and T_2 is the anti-clockwise rotation about the origin through an angle $\frac{\pi}{2}$.

21. Determine which of the following statements are true. Justify your answer.

    (a) If R_1 and R_2 are two rotations about the origin in $\mathbb{R}^2$, then $R_2 \circ R_1$ is a rotation about the origin in $\mathbb{R}^2$.

    (b) If R_1 and R_2 are two rotations about the origin in $\mathbb{R}^2$, then $R_1 \circ R_2 = R_2 \circ R_1$.

    (c) If R is a rotation about the origin and F is a reflection about a line through the origin in $\mathbb{R}^2$, then $F \circ R$ is a reflection about a line through the origin in $\mathbb{R}^2$.

    (d) If R is a rotation about the origin and F is a reflection about a line through the origin in $\mathbb{R}^2$, then $R \circ F = F \circ R$.

    (e) If F_1 and F_2 are two reflections about lines through the origin in $\mathbb{R}^2$, then $F_2 \circ F_1$ is a reflection about a line through the origin in $\mathbb{R}^2$.

    (f) If F_1 and F_2 are two reflections about lines through the origin in $\mathbb{R}^2$, then $F_1 \circ F_2 = F_2 \circ F_1$.

22. Describe geometrically the effect of each of the following linear operators $T : \mathbb{R}^3 \to \mathbb{R}^3$ defined by $T(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$.

    (a) $\boldsymbol{A} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$, (b) $\boldsymbol{A} = \begin{pmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix}$,

    (c) $\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{4}{5} & -\frac{3}{5} \\ 0 & \frac{3}{5} & \frac{4}{5} \end{pmatrix}$, (d) $\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{3}{5} & \frac{4}{5} \\ 0 & \frac{4}{5} & -\frac{3}{5} \end{pmatrix}$,

    (e) $\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{4}{5} & -\frac{3}{5} \\ 0 & \frac{3}{5} & \frac{4}{5} \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0.5 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{4}{5} & -\frac{3}{5} \\ 0 & \frac{3}{5} & \frac{4}{5} \end{pmatrix}^{\mathrm{T}}$.

23. For each of the following linear operators on $\mathbb{R}^3$, (i) find the standard matrices for $T_1 \circ T_2$ and $T_2 \circ T_1$; and (ii) determine whether $T_1 \circ T_2 = T_2 \circ T_1$.

    (a) T_1 is the anti-clockwise rotation about the z-axis through an angle $\frac{\pi}{2}$ and T_2 is the contraction by a factor of 0.5.

    (b) T_1 is the anti-clockwise rotation about the z-axis through an angle $\frac{\pi}{2}$ and T_2 is the reflection about the xy-plane.

    (c) T_1 is the anti-clockwise rotation about the z-axis through an angle $\frac{\pi}{2}$ and T_2 is the reflection about the yz-plane.

24. Let $R : \mathbb{R}^3 \to \mathbb{R}^3$ be the (anti-clockwise) rotation about the axis in the direction $\boldsymbol{n} = (0, 1, 1)^{\mathrm{T}}$ through an angle π. Find the formula and the standard matrix for R. (Hint: Check the images of $\boldsymbol{e_1}$, $\boldsymbol{e_2}$ and $\boldsymbol{e_3}$ under R geometrically.)

25. Let $R : \mathbb{R}^3 \to \mathbb{R}^3$ be the anti-clockwise rotation about the axis in the direction of $\boldsymbol{n} = (1, 1, -1)^{\mathrm{T}}$ through an angle θ. Let $\{\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}\}$ be the basis for $\mathbb{R}^3$ obtained in Question 5.14(c). Find $R(\boldsymbol{v_1})$, $R(\boldsymbol{v_2})$ and $R(\boldsymbol{v_3})$. (Hint: Check the effect of the rotation on the vectors in the plane $x + y - z = 0$.)

26. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be a mapping such that

$$T(\boldsymbol{u}) = \boldsymbol{u} - 2\left(\frac{\boldsymbol{u} \cdot \boldsymbol{n}}{\boldsymbol{n} \cdot \boldsymbol{n}}\right)\boldsymbol{n} \quad \text{for } \boldsymbol{u} \in \mathbb{R}^3$$

    where $\boldsymbol{n} = (a, b, c)^{\mathrm{T}}$ is a nonzero vector. Show that T is the reflection about the plane $ax + by + cz = 0$ in $\mathbb{R}^3$ and write down the standard matrix for T.

27. Let $\boldsymbol{B} = \begin{pmatrix} 0 & 0 & 2 & 2 \\ 0 & 2 & 0 & 2 \\ 1 & 1 & 1 & 1 \end{pmatrix}$. The column vectors of $\boldsymbol{B}$ represent the homogeneous coordinates of points in the xy-plane (see Discussion 7.3.13).

    (a) Sketch the figure represented by $\boldsymbol{B}$.

    (b) For each of the following, sketch the figure represented by $\boldsymbol{PB}$ and describe geometrically the effect of the corresponding transformation.

    (i) $\boldsymbol{P} = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}$, (ii) $\boldsymbol{P} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$,

    (iii) $\boldsymbol{P} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, (iv) $\boldsymbol{P} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$.

28. Consider the homogeneous coordinates discussed in Discussion 7.3.13.

    (a) Write down the matrix that corresponds to the translation that moves $(x, y)^{\mathrm{T}}$ to $(x-1,\ y+2)^{\mathrm{T}}$.

    (b) Write down the inverse of the matrix obtain in Part (a) and describe geometrically the effect of the corresponding transformation.

    (c) Write down the matrix that corresponds to the anti-clockwise rotation about the origin through an angle $\frac{\pi}{4}$.

    (d) Hence, or otherwise, find the matrix that corresponds to the anti-clockwise rotation about the point $(-1, 2)^{\mathrm{T}}$ through an angle $\frac{\pi}{4}$.

The McGraw-Hill Companies

ISBN 978-1-259-01151-1
MHID 1-259-01151-8

9 781259 011511

Visit McGraw-Hill Education (Asia) at
www.mheducation.asia