Concise Introduction to LINEAR ALGEBRA

Qingwen Hu

CPD CERTIFIED
The CPD Certification Service

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

Concise Introduction to

LINEAR ALGEBRA

Concise Introduction to
LINEAR ALGEBRA

Qingwen Hu

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed on acid-free paper
Version Date: 20170822

International Standard Book Number-13: 978-1-138-04449-4 (Hardback)

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

To my students.

Contents

Preface

This book provides sufficient materials for a one-semester linear algebra course at the sophomore level. It is based on the lecture notes for the linear algebra course that the author taught several years to undergraduate students in science and mathematics at the University of Texas at Dallas. The level and pace of the course can be adjusted by balancing the time for theoretical illustration and that for computational aspects of the subject. The author usually taught up to Chapter 7, spending one lecture per section on average, while the remaining two chapters can be left for students' reading homework or supervised individual study.

It seems that many undergraduate students have only one linear algebra course before graduation, and may have missed many important topics of linear algebra which may be remedied later by self-studying on demand. This book is written to accommodate the needs for classroom teaching in order to effectively deliver the essential topics of the subject, and for self-studying beyond a first linear algebra course.

The following is an introduction to each chapter of the book.

1. Chapter 1 deals with vectors, linear combinations and dot products in $\mathbb{R}^n$. In Section 1.3 we discuss matrix representations for linear systems and for elementary row operations.

2. Chapter 2 illustrates Guassian elimination and Gauss–Jordan elimination for solving linear systems, along with basic matrix theory, LU-decomposition and permutation matrices.

3. Chapter 3 starts with four subspaces of $\mathbb{R}^n$ associated with a real matrix. Then we discuss bases and dimensions of general vector spaces.

4. Chapter 4 deals with orthogonality between subspaces. Related topics include matrix representation of orthogonal projection, least squares solutions, Gram–Schmidt process and QR-decomposition.

5. Chapter 5 presents an axiomatic method of determinants which naturally leads to the permutation formula, co-factor expansion, product formula and Cramer's rule.

6. In Chapter 6 we introduce the notions of eigenvalues and eigenvectors which open the door for more applications of linear algebra, including the immediate application on diagonalizability, spectral decomposition of symmetric real matrices, quadratic forms, positive definite matrices and Rayleigh quotient.

7. **Chapter 7** continues to discuss the application of eigenvalues and eigenvectors and presents singular value decomposition of general matrices. Principal component analysis is also introduced as a real-world application of linear algebra.

8. **Chapter 8** discusses the matrix representation, range and null spaces for linear transformations on general vector spaces. Then we introduce invariant subspaces, decomposition of vector spaces and Jordan normal form and its computation, where the treatment of the Jordan normal form does not require a formal exposition of polynomial theory.

9. **Chapter 9** presents basic theory of linear programming along with the simplex method which is another concrete real-world application of linear algebra and which has been widely used in management and industry.

The book contains typical topics for linear algebra courses and can be used in many ways depending on the different mathematical background of the audiences. The book provides limited examples and exercises, while it is best used for readers who would like to have a broad coverage of the topics of linear algebra and who are motivated to customize questions for the materials of each section. Comments and suggestions from readers are highly appreciated and are welcome to be sent by e-mail to qingwen@utdallas.edu.

Qingwen Hu
January 2017

Chapter 1

Vectors and linear systems

A central goal of linear algebra is to solve systems of linear equations. We have seen the simplest linear equation $ax = b$, where $x \in \mathbb{R}$ (the symbol "$\in$" means "in") is the unknown variable and a, $b \in \mathbb{R}$ are constants. It is known that there are three scenarios for the solutions: 1) if $a \neq 0$, there is a unique solution $x = \frac{b}{a}$; 2) if $a = 0$, $b \neq 0$, there is no solution; 3) if $a = b = 0$, there are infinitely many solutions. We are then motivated to investigate systems of equations with multiple unknown variables. The following system

$$\begin{cases} x + 2y + 3z = 3 \\ 2x + 5y + 8z = 9 \\ 3x + 6y + 18z = 18 \end{cases} \tag{1.1}$$

is a system of linear equations with three equations and three unknowns. In this chapter, we learn how to use vectors to represent a linear system and learn the ideas of elimination which will be applied to solve systems of linear equations. The general form of linear systems is as follows:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m, \end{cases} \tag{1.2}$$

where $x = (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$ is the unknown vector in n-dimensional Euclidean space; $a_{i,j}$ and b_i with $i \in \{1, 2, \cdots, m\}$, $j \in \{1, 2, \cdots, n\}$ are constants.

1.1 Vectors and linear combinations

Before we discuss how to solve general linear systems, we use system (1.1) as a prototype to introduce the machinery of vectors. One may rewrite system (1.1) as

$$x \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + y \begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix} + z \begin{bmatrix} 3 \\ 8 \\ 18 \end{bmatrix} = \begin{bmatrix} 3 \\ 9 \\ 18 \end{bmatrix}. \tag{1.3}$$

System (1.3) makes sense only if we have defined addition and scalar multiplication of vectors in Euclidean spaces, where we have identified the vector (x_1, x_2, x_3) with the column of numbers

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix},$$

which is called a column **matrix.** In what follows we will always regard a vector in $\mathbb{R}^n$ as a column matrix.

Definition 1.1.1. Let $x = (x_1, x_2, \cdots, x_n)$, $y = (y_1, y_2, \cdots, y_n)$ be vectors in $\mathbb{R}^n$, α a scalar. We define addition $x + y$ and scalar multiplication αx by

$$\begin{aligned} x + y &= (x_1 + y_1, x_2 + y_2, \cdots, x_n + y_n), \\ \alpha x &= (\alpha x_1, \alpha x_2, \cdots, \alpha x_n). \end{aligned}$$

Definition 1.1.2. Let $x_1, x_2, \cdots, x_n \in \mathbb{R}^N$ be vectors, and $c_1, c_2 \cdots, c_n \in \mathbb{R}$ be scalars. We call

$$c_1 x_1 + c_2 x_2 + \cdots c_n x_n$$

a **linear combination** of $x_1, x_2, \cdots, x_n$.

System (1.3) now can be interpreted as finding a proper linear combination of the vectors (1, 2, 3), (2, 5, 6) and (3, 8, 18) to produce the given vector (3, 9, 18) on the right hand side. Certainly we can also interpret it as finding the common point (x, y, z) of three planes determined by each of the equations. If we visualize a linear system with this interpretation of a linear system, we obtain a row picture, while with the previous one, a column picture.

Example 1.1.3. 1. Let $v = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $w = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$. Then

$$3v + 5w = 3 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 5 \begin{bmatrix} 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 8 \\ 18 \end{bmatrix}$$

is a linear combination of v and w.

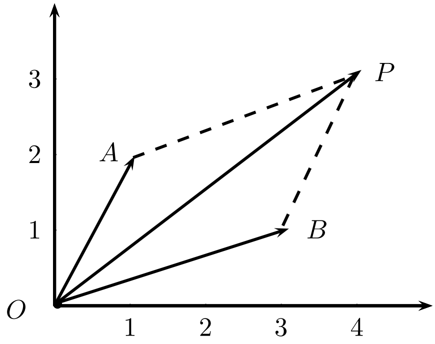

FIGURE 1.1: Slope of $\overrightarrow{OA} = \frac{2-0}{1-0} = 2$, $\overrightarrow{BP} = \frac{4-2}{2-1} = 2$. Slope of $\overrightarrow{OB} = \frac{3-0}{1-0} = 3$, $\overrightarrow{AP} = \frac{4-1}{3-2} = 3$.

2. Let $v = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. Then $\{cv : 0 \le c \le 2\}$ represents a line segment from (0, 0) to (2, 4) in $\mathbb{R}^2$.

3. Let $v = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $w = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Then $\{cv + dw : c \in \mathbb{R}, d \in \mathbb{R}\}$ represents the whole two dimensional plane $\mathbb{R}^2$.

4. Let $v = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ and $w = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. Then $S = \{cv + dw : c \in \mathbb{R}, d \in \mathbb{R}\}$ represents a two dimensional plane in $\mathbb{R}^3$, but not the whole space $\mathbb{R}^3$, because there exists the vector $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ which is not in S.

□

Example 1.1.4. (The parallelogram law for vector addition) A vector $x = (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$ can be visualized by the directed line segment $\overrightarrow{OA}$ from the origin $O = (0, 0, \cdots, 0)$ to the point $A = (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$. If we denote the end point of the vector $y = (y_1, y_2, \cdots, y_n)$ by B and that of $x + y$ by P, then we have a parallelogram $OAPB$, with OA parallel to BP and AP parallel to OB, since the opposite segments have the same slopes.

□

Exercise 1.1.5.

1. Let $u = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $v = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$. i) Sketch the directed line segments in $\mathbb{R}^2$ that represents u and v, respectively; ii) Use the parallelogram law to visualize the vector addition $u + v$; iii) Find $2u$, $2u + 5v$ and $2v - 5u$; iv) Solve the system of equations $xu + yv = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ for $(x, y) \in \mathbb{R}^2$ and draw the row picture and the column picture.

2. Let $u = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $v = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$. Is $w = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ a linear combination of u and v?

3. Is it true every vector $(x, y) \in \mathbb{R}^2$ can be represented as a linear combination of $v = (1, 0)$ and $w = (1, 1)$?

4. Find vectors $u, v, w \in \mathbb{R}^3$ such that the following system

$$\begin{cases} x + z = 1 \\ 2x + 5y + 8z = -1 \\ x + y = 1 \end{cases}$$

can be rewritten as $xu + yv + zw = b$, where $b = (1, -1, 1)$.

5. Show that $\mathbb{R}^2 = \left\{ x \begin{bmatrix} 1 \\ 0 \end{bmatrix} + y \begin{bmatrix} 0 \\ 1 \end{bmatrix} : x \in \mathbb{R}, y \in \mathbb{R} \right\}$.

1.2 Length, angle and dot products

In order to discuss geometry in Euclidean spaces, we introduce the notions of length and angle, which can be defined with dot products.

Definition 1.2.1. Let $x = (x_1, x_2, \cdots, x_n)$, $y = (y_1, y_2, \cdots, y_n)$ be vectors in $\mathbb{R}^n$; the dot product $x \cdot y$ is defined by

$$x \cdot y = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n = \sum_{i=1}^{n} x_i y_i.$$

Example 1.2.2. 1. Let $v = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $w = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$. Then

$$v \cdot w = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 3 \end{bmatrix} = 1 \cdot 2 + 1 \cdot 3 = 5.$$

2. Let $v = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ and $w = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. Then $v \cdot w = 1 \cdot 1 + 1 \cdot 1 + 1 \cdot 0 = 2$.

3. Let $v = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $w = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$. Then
$$v \cdot w = 1 \cdot 1 + 1 \cdot (-1) = 0.$$
We say v and w are orthogonal to each other and write $v \perp w$.

4. Consider the distance from $A = (1, 2)$ to the origin O. We have
$$\begin{aligned} \|\overrightarrow{OA}\| &= \sqrt{(1-0)^2 + (2-0)^2} \\ &= \sqrt{1 \cdot 1 + 2 \cdot 2}. \end{aligned}$$
If we denote by v the vector $\overrightarrow{OA}$, we have the length of v
$$\|v\| = \sqrt{v \cdot v}.$$

5. Consider unit vectors $u, v \in \mathbb{R}^2$. Then there exist $\alpha, \beta \in [0, 2\pi)$ such that
$$u = (\cos\alpha, \sin\alpha), \quad v = (\cos\beta, \sin\beta).$$
Then we have
$$u \cdot v = \cos\alpha\cos\beta + \sin\alpha\sin\beta = \cos(\alpha - \beta).$$

□

One can check directly that dot product satisfies the following

Lemma 1.2.3. Let $u, v, w \in \mathbb{R}^n$ be vectors. Then
$$\begin{aligned} u \cdot v &= v \cdot u, \\ u \cdot (v + w) &= u \cdot v + u \cdot w. \end{aligned}$$

Definition 1.2.4. Let $v = (v_1, v_2, \cdots, v_n)$ be a vector in $\mathbb{R}^n$. The length $\|v\|$ of v is defined by
$$\|v\| = \sqrt{v \cdot v} = \left(\sum_{i=1}^{n} v_i^2\right)^{\frac{1}{2}}.$$
A vector with unit length is called a unit vector.

Example 1.2.5.

Consider unit vectors $u, v \in \mathbb{R}^2$. Then there exist $\alpha, \beta \in [0, 2\pi)$ such that
$$u = (\cos\alpha, \sin\alpha), \quad v = (\cos\beta, \sin\beta).$$
Then we have
$$u \cdot v = \cos\alpha\cos\beta + \sin\alpha\sin\beta = \cos(\alpha - \beta).$$
There exists $\theta \in [0, \pi]$ such that $\cos\theta = \cos(\alpha - \beta)$. Then we call θ the angle between the vectors u and v.

Consider nonzero vectors $u,\, v \in \mathbb{R}^2$. Then $\frac{u}{\|u\|}$ and $\frac{v}{\|v\|}$ are unit vectors and there exist $\alpha,\, \beta \in [0,\, 2\pi)$ such that

$$\frac{u}{\|u\|} = (\cos\alpha,\, \sin\alpha), \qquad \frac{v}{\|v\|} = (\cos\beta,\, \sin\beta).$$

We have

$$\frac{u}{\|u\|} \cdot \frac{v}{\|v\|} = \cos(\alpha - \beta) = \cos\theta, \tag{1.4}$$

where $\theta \in [0,\, \pi]$ is the angle between $\frac{u}{\|u\|}$ and $\frac{v}{\|v\|}$. Notice that u and $\frac{u}{\|u\|}$ have the same direction, so do v and $\frac{v}{\|v\|}$. $\theta \in [0,\, \pi]$ is also the angle between u and v. By (1.4) we have

$$u \cdot v = \|u\|\|v\| \cos\theta,$$

where u and v can be zero. Then we have derived

Lemma 1.2.6. (**Cosine formula**) Let $u,\, v \in \mathbb{R}^2$. We have

$$u \cdot v = \|u\|\|v\| \cos\theta,$$

where $\theta \in [0,\, \pi]$ is the angle between u and v.

An immediate consequence of the cosine formula is that $|v \cdot w| = \|u\|\|v\|| \cos\theta| \leq \|u\|\|v\|$ which is the Schwarz inequality in $\mathbb{R}^2$. We show the general version of the Schwartz inequality in $\mathbb{R}^n$:

Lemma 1.2.7. (**Schwarz inequality**) Let $u,\, v \in \mathbb{R}^n$. We have

$$|u \cdot v| \leq \|u\|\|v\|.$$

Proof. The inequality is true if $v = 0$. We assume that $v \neq 0$ and let $w = u + tv$, $t \in \mathbb{R}$. Then $\|w\| \geq 0$ for every $t \in \mathbb{R}$. We have

$$\begin{aligned} 0 \leq \|w\| &= (u + tv) \cdot (u + tv) \\ &= u \cdot u + 2(u \cdot v)t + (v \cdot v)t^2, \end{aligned}$$

for every $t \in \mathbb{R}$. Therefore the discriminant of the quadratic polynomial $(u + tv,\, u + tv)$ of t satisfies

$$4(u \cdot v)^2 - 4(u \cdot u)(v \cdot v) \leq 0,$$

which is equivalent to $|u \cdot v| \leq \|u\|\|v\|$. $\square$

With the Schwarz inequality, we can then define angles between vectors in $\mathbb{R}^n$:

Definition 1.2.8. Let $u, v \in \mathbb{R}^n$. We define $\theta \in [0, \pi]$ such that

$$u \cdot v = \|u\|\|v\| \cos\theta,$$

the angle between u and v.

By properties of dot products and the Schwarz inequality, we have

Lemma 1.2.9. (**Triangle inequality**) Let $u, v \in \mathbb{R}^n$. We have

$$\|u + v\| \leq \|u\| + \|v\|.$$

Proof. We have

$$\begin{aligned}
\|u+v\|^2 =& (u+v) \cdot (u+v) \\
=& u \cdot u + 2u \cdot v + v \cdot v \\
\leq& u \cdot u + 2\|u\| \cdot \|v\| + v \cdot v \\
=& \|u\|^2 + 2\|u\| \cdot \|v\| + \|v\|^2 \\
=& (\|u\| + \|v\|)^2.
\end{aligned}$$

Therefore we have $\|u + v\| \leq \|u\| + \|v\|$. □

Exercise 1.2.10.

1. Let $u = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $v = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$. i) Find $u \cdot v$; ii) Find $\|u\|$ and $\|v\|$; iii) Find the angle θ between u and v; iv) Verify that $|u \cdot v| \leq \|u\|\|v\|$; v) Verify that $\|u + v\| \leq \|u\| + \|v\|$.

2. Find all possible real values of a such that the quadratic polynomial $x^2 + ax + 1$ has i) two positive roots; ii) two negative roots; iii) one negative and one positive root; iv) no real roots, respectively.

3. Let $u = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Find all possible vectors w such that $u \perp w$, i.e., $u \cdot w = 0$.

4. Let $u = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $v = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$ and $w = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$. i) Find $u \cdot v$ and $v \cdot w$. ii) Is it possible to find $(x, y) \neq (0, 0)$ such that $v = xu + yw$? Justify your answer.

5. Let $u, v \in \mathbb{R}^n$. Show that

$$|u \cdot v| = \|u\|\|v\|,$$

if and only if $v = 0$ or there exists a scalar $t \in \mathbb{R}$ such that $u = tv$.

1.3 Matrices

Recall that system (1.3) can be interpreted as finding a proper linear combination of the vectors $u = (1, 2, 3)$, $v = (2, 5, 8)$ and $w = (3, 6, 18)$ to produce the given vector $\mathbf{b} = (3, 9, 18)$ on the right hand side. That is, we are looking for scalars x, y, z such that

$$xu + yv + zw = \mathbf{b},$$

which looks to be a certain product between (u, v, w) and (x, y, z). To wit, we write

$$[u \quad v \quad w] \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{b},$$

which is a "row" multiplied by a "column." The reason why we put the letters for vectors horizontally becomes clear when we recover the values of u, v, w and $\mathbf{b}$:

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 8 \\ 3 & 6 & 18 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 3 \\ 9 \\ 18 \end{bmatrix}, \tag{1.5}$$

where we obtain a rectangular array of numbers called a **matrix**, and if u, v, w were placed vertically, we would not know how to place their values!

Let

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 8 \\ 3 & 6 & 18 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 3 \\ 9 \\ 18 \end{bmatrix}.$$

System (1.3) becomes the familiar form of

$$A\mathbf{x} = \mathbf{b}. \tag{1.6}$$

By comparing system (1.3) with system (1.5), we know that the so far undefined product $A\mathbf{x}$ between matrices A and $\mathbf{x}$ essentially consists of rows of A taking dot products with $\mathbf{x}$. That is,

$$\begin{bmatrix} (\text{Row } 1 \text{ of } A) \cdot \mathbf{x} \\ (\text{Row } 2 \text{ of } A) \cdot \mathbf{x} \\ (\text{Row } 3 \text{ of } A) \cdot \mathbf{x} \end{bmatrix} = \mathbf{b}.$$

Example 1.3.1.

$$\begin{bmatrix} 1 & 2 & 3 \\ -2 & -4 & 6 \end{bmatrix} \begin{bmatrix} 4 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} (1,2,3) \cdot (4,\ -1,\ 0) \\ (-2,\ -4,\ 6) \cdot (4,\ -1,\ 0) \end{bmatrix} = \begin{bmatrix} 2 \\ -4 \end{bmatrix}.$$

$$\begin{bmatrix} 3 & -2 \\ -1 & 0 \\ 2 & 5 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} (3,\ -2)\cdot(2,\ 1) \\ (-1,\ 0)\cdot(2,\ 1) \\ (2,\ 5)\cdot(2,\ 1) \end{bmatrix} = \begin{bmatrix} 4 \\ -2 \\ 9 \end{bmatrix}.$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

□

Remark 1.3.2.

For an $m \times n$ matrix A, we write $A = (a_{i\,j})$ when we emphasize the general form of its entries. We also write $(A)_{i\,j}$, $A(i,\ j)$ or simply $A_{i\,j}$ to denote the entry at the $(i,\ j)$-position.

If $A = (a_{i\,j})$ is an $n \times n$ square matrix, we call the entries a_{ii}, $i = 1,\ 2, \cdots, n$ the main diagonal entries. If every main diagonal entry of A is one, and every other entries are zero, that is,

$$a_{i\,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j, \end{cases}$$

we call A an **identity matrix** and denote it by I. Note that

> $Ix = x$ for every $x \in \mathbb{R}^n$.

Example 1.3.3. Let $u = (1,\ 0,\ 0)$, $v = (1,\ 1,\ 0)$ and $w = (1,\ 1,\ 1)$. $\mathbf{b} = (b_1,\ b_2,\ b_3)$. We solve system

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

where

$$A = [u \quad v \quad w] = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix},\ \mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

That is, we solve

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}.$$

We notice that A is a triangular matrix in the sense that the nonzero entries are above the main diagonal. Such type of matrix is convenient for solving the system by **back substitution**. Namely, we first solve for z, then y and x. We obtain

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} b_1 - b_2 - b_3 \\ b_2 - b_3 \\ b_3 \end{bmatrix}.$$

To have a solution resembling the solution $x = a^{-1}b$ of the single variable linear equation $ax = b$, $a \neq 0$, we wish to write (x, y, z) in terms of $\mathbf{b} = (b_1, b_2, b_3)$. We rewrite the solution as follows:

$$\begin{aligned}\begin{bmatrix} x \\ y \\ z \end{bmatrix} &= \begin{bmatrix} b_1 - b_2 - b_3 \\ b_2 - b_3 \\ b_3 \end{bmatrix} \\ &= \begin{bmatrix} b_1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -b_2 \\ b_2 \\ 0 \end{bmatrix} + \begin{bmatrix} -b_3 \\ -b_3 \\ b_3 \end{bmatrix} \\ &= b_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + b_2 \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + b_3 \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & -1 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}.\end{aligned}$$

Let

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

We have the solution $\mathbf{x} = B\mathbf{b}$. We write $B = A^{-1}$ and $\mathbf{x} = A^{-1}\mathbf{b}$. Note that we did not specify the values of $\mathbf{b}$. The system in question has a unique solution for every given $\mathbf{b} \in \mathbb{R}^3$. □

Example 1.3.4. Let $u = (1, 0, 0)$, $v = (1, 1, 0)$ and $w^* = (0, 1, 0)$. $\mathbf{b} = (b_1, b_2, b_3)$. We solve system

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

where

$$A = [u \quad v \quad w^*] = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

That is, we solve

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}.$$

We notice that A is also a triangular matrix but we cannot solve the system by back substitution. The third equation of the system is

$$0 = b_3,$$

which may or may not be true depending on the value of b_3.

If $b_3 \neq 0$, system $A\mathbf{x} = \mathbf{b}$ has no solution.

If $b_3 = 0$, system $A\mathbf{x} = \mathbf{b}$ becomes

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix},$$

which has a free variable z that can be parameterized by $z = t$, $t \in \mathbb{R}$. Then we have

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} b_1 - b_2 + t \\ b_2 - t \\ t \end{bmatrix} = \begin{bmatrix} b_1 - b_2 \\ b_2 \\ 0 \end{bmatrix} + t \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \; t \in \mathbb{R}, \tag{1.7}$$

which represents infinitely many solutions on a straight line in $\mathbb{R}^3$. □

Let us make some observations on the previous two examples. In Example 1.3.4, for arbitrary $\mathbf{b} \in \mathbb{R}^3$, we have a unique solution $x = A^{-1}\mathbf{b}$. That is, the vector equation

$$xu + yv + zw = \mathbf{b}$$

always has a unique solution for the linear combination coefficients (x, y, z). This implies that the set of vectors $\{u, v, w\}$ can span the whole space $\mathbb{R}^3$.

In Example 1.3.3, there exists $\mathbf{b} = (b_1, b_2, b_3) \in \mathbb{R}^3$ with $b_3 \neq 0$, which is not a linear combination of $\{u, v, w^*\}$. That is, the set of vectors $\{u, v, w^*\}$ cannot span the whole space $\mathbb{R}^3$. But why can $\{u, v, w\}$, while both sets have three different vectors? The answer is that $\{u, v, w^*\}$ has redundant vectors, but $\{u, v, w\}$ does not. Namely, the role of some vectors in $\{u, v, w^*\}$ can be replaced by other vectors. To identify the redundancy, we set up the following model:

$$x_1 u + x_2 v + x_3 w^* = 0,$$

solving for (x_1, x_2, x_3). By (1.7), we have at least one nonzero solution $(x_1, x_2, x_3) = (1, -1, 1)$. That is,

$$1u + (-1)v + 1\,w^* = 0 \Longleftrightarrow v = u + w^*.$$

That is, v can be replaced with $u + w^*$. Therefore, the spanning role of $\{u, v, w^*\}$ is the same as that of $\{u, w^*\}$, which cannot span a three dimensional space.

Next we verify that there is no redundancy in $\{u, v, w\}$ for spanning $\mathbb{R}^3$. We also set up the following model:

$$x_1 u + x_2 v + x_3 w = 0,$$

solving for (x_1, x_2, x_3). By the solution in Example 1.3.4, we have the only solution $(x_1, x_2, x_3) = (0, 0, 0)$. This means that none of the vectors in $\{u, v, w\}$ can be replaced by a linear combination of the other ones. They are **linearly independent**.

Definition 1.3.5. Let $\{u_1, u_2, \cdots, u_m\}$ be a set of vectors in $\mathbb{R}^n$. If the vector equation

$$x_1u_1 + x_2u_2 + \cdots + x_nu_n = 0$$

has only the trivial solution $x_1 = x_2 = \cdots = x_n = 0$, $\{u_1, u_2, \cdots, u_m\}$ is said to be linearly independent. Otherwise, $\{u_1, u_2, \cdots, u_m\}$ is said to be linearly dependent.

We finish this chapter with examples on matrix multiplication with elementary matrices.

Example 1.3.6. Elementary matrices

- Consider $E\mathbf{x} = \mathbf{b}$, where $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $E = \begin{bmatrix} 1 & 0 \\ l & 1 \end{bmatrix}$. Then $Ex = \begin{bmatrix} x_1 \\ x_2 + lx_1 \end{bmatrix}$. Note that the effect of multiplication by E from the left of $\mathbf{x}$ is "adding l-multiple of row 1 to row 2." The solution is

$$\begin{aligned} \mathbf{x} &= \begin{bmatrix} b_1 \\ b_2 - lb_1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ -l & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}. \end{aligned}$$

Denote by $E^{-1} = \begin{bmatrix} 1 & 0 \\ -l & 1 \end{bmatrix}$. We have the solution $\mathbf{x} = E^{-1}\mathbf{b}$. The effect of multiplication by E^{-1} from the left of $\mathbf{b}$ is "subtracting l-multiple of row 1 from row 2." Moreover, using $\mathbf{x} = E^{-1}\mathbf{b}$ and the original system $E\mathbf{x} = \mathbf{b}$, we have

$$E(E^{-1}\mathbf{b}) = \mathbf{b}, \quad E^{-1}(E\mathbf{x}) = \mathbf{x}.$$

That is, the multiplication actions from the left of a vector by E and E^{-1} are canceling each other. If we treat the action $A : \mathbf{x} \mapsto A\mathbf{x}$ as a function determined by the matrix A, then the effect from $E^{-1} \circ E$ and $E \circ E^{-1}$ is the same as the identity matrix I.

- Consider $E\mathbf{x} = \mathbf{b}$, where $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $E = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Then $Ex = \begin{bmatrix} x_2 \\ x_1 \end{bmatrix}$. Note that the effect of multiplication by E from the left of $\mathbf{x}$ is "exchanging positions of row 1 and row 2." The solution is

$$\begin{aligned} \mathbf{x} &= \begin{bmatrix} b_2 \\ b_1 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}. \end{aligned}$$

Denote by $E^{-1} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, which is identical to E itself. We have the solution $\mathbf{x} = E^{-1}\mathbf{b}$. The effect of multiplication by E^{-1} from the left of $\mathbf{b}$ is "exchanging positions of row 1 and row 2." Moreover, using $\mathbf{x} = E^{-1}\mathbf{b}$ and the original system $E\mathbf{x} = \mathbf{b}$, we have

$$E(E^{-1}\mathbf{b}) = \mathbf{b}, \quad E^{-1}(E\mathbf{x}) = \mathbf{x}.$$

That is, the multiplication actions from the left of a vector by E and E^{-1} are canceling each other. The multiplication effects from $E^{-1} \circ E$ and $E \circ E^{-1}$ are the same as the identity matrix I.

- Consider $E\mathbf{x} = \mathbf{b}$, where $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $E = \begin{bmatrix} 1 & 0 \\ 0 & c \end{bmatrix}$ with $c \neq 0$. Then

$Ex = \begin{bmatrix} x_1 \\ cx_2 \end{bmatrix}$. Note that the effect of multiplication by E from the left of $\mathbf{x}$ is "multiplying row 2 by c". The solution is

$$\begin{aligned} \mathbf{x} &= \begin{bmatrix} b_1 \\ \frac{1}{c} b_2 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{c} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}. \end{aligned}$$

Denote by $E^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{c} \end{bmatrix}$. We have the solution $\mathbf{x} = E^{-1}\mathbf{b}$. The effect of multiplication by E^{-1} from the left of $\mathbf{b}$ is "dividing row 2 by c". Moreover, using $\mathbf{x} = E^{-1}\mathbf{b}$ and the original system $E\mathbf{x} = \mathbf{b}$, we have

$$E(E^{-1}\mathbf{b}) = \mathbf{b}, \quad E^{-1}(E\mathbf{x}) = \mathbf{x}.$$

That is, the multiplication actions from the left of a vector by E and E^{-1} are canceling each other. If we treat the action $A : \mathbf{x} \mapsto A\mathbf{x}$ as a function determined by the matrix A, then the effect from $E^{-1} \circ E$ and $E \circ E^{-1}$ is the same as the identity matrix I.

The aforementioned three type of matrices are called **elementary matrices** which can be obtained by operating on the identity matrices with the elementary row operation in question. □

Exercise 1.3.7.

1. Let $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$. Compute i) $A + B$, $A + 2B$ and $A - 3B$; ii) AB and BA.

2. Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. i) Compute AB and BA; ii) Is $AB = BA$?