

# Linear Algebra for 21st C Application

A. J. Roberts \*

University of Adelaide, South Australia, 5005

October 31, 2017

\* <http://orcid.org/0000-0001-8930-1552>

---

# Contents

---

<b>1</b>	<b>Vectors</b>	<b>4</b>
1.1	Vectors have magnitude and direction . . . . .	8
1.2	Adding and stretching vectors . . . . .	23
1.3	The dot product determines angles and lengths . . .	44
1.4	The cross product . . . . .	75
1.5	Use MATLAB/Octave for vector computation . . . .	97
<b>2</b>	<b>Systems of linear equations</b>	<b>104</b>
2.1	Introduction to systems of linear equations . . . . .	110

2.2	Directly solve linear systems . . . . .	120
2.3	Linear combinations span sets . . . . .	151
<b>3</b>	<b>Matrices encode system interactions</b>	<b>167</b>
3.1	Matrix operations and algebra . . . . .	171
3.2	The inverse of a matrix . . . . .	228
3.3	Factorise to the singular value decomposition . . . .	276
3.4	Subspaces, basis and dimension . . . . .	324
3.5	Project to solve inconsistent equations . . . . .	373
3.6	Introducing linear transformations . . . . .	445
<b>4</b>	<b>Eigenvalues and eigenvectors of symmetric matrices</b>	<b>479</b>
4.1	Introduction to eigenvalues and eigenvectors . . . .	482
4.2	Beautiful properties for symmetric matrices . . . .	512

<b>5</b>	<b>Approximate matrices</b>	<b>544</b>
5.1	Measure changes to matrices . . . . .	546
5.2	Regularise linear equations . . . . .	600
<b>6</b>	<b>Determinants distinguish matrices</b>	<b>617</b>
6.1	Geometry underlies determinants . . . . .	620
6.2	Laplace expansion theorem for determinants . . . . .	640
<b>7</b>	<b>Eigenvalues and eigenvectors in general</b>	<b>664</b>
7.1	Find eigenvalues and eigenvectors of matrices . . . . .	673
7.2	Linear independent vectors may form a basis . . . . .	725
7.3	Diagonalisation identifies the transformation . . . . .	761

---

# 1 Vectors

---

## Chapter Contents

1.1	Vectors have magnitude and direction . . . . .	8
1.2	Adding and stretching vectors . . . . .	23
1.2.1	Basic operations . . . . .	24
1.2.2	Parametric equation of a line . . . . .	36
1.2.3	Manipulation requires algebraic properties . .	40
1.3	The dot product determines angles and lengths . . .	44
1.3.1	Work done involves the dot product . . . . .	54
1.3.2	Algebraic properties of the dot product . . .	57

1.3.3	Orthogonal vectors are at right-angles . . . .	62
1.3.4	Normal vectors and equations of a plane . . .	66
1.4	The cross product . . . . .	75
1.5	Use MATLAB/Octave for vector computation . . . .	97

This chapter is a relatively concise introduction to vectors, their properties, and a little computation with MATLAB/Octave. Skim or study as needed.

Mathematics started with counting. The natural numbers  $1, 2, 3, \dots$  quantify how many objects have been counted. Historically, there were many existential arguments over many centuries about whether negative numbers and zero are meaningful. Nonetheless, eventually negative numbers and the zero were included to form the **integers**  $\dots, -2, -1, 0, 1, 2, \dots$ . In the mean time people needed to quantify fractions such as two and half a bags, or a third of a cup which led to the rational numbers such as  $\frac{1}{3}$  or  $2\frac{1}{2} = \frac{5}{2}$ . Now **rational numbers** are defined as all numbers writeable in the form  $\frac{p}{q}$  for integers  $p$  and  $q$  ( $q$  non-zero). Roughly two thousand years ago, Pythagoras was forced to recognise that for many triangles the length of a side could not be rational, and hence there must be more numbers in the world about us than rationals could provide.

To cope with non-rational numbers such as  $\sqrt{2} = 1.41421 \dots$  and  $\pi = 3.14159 \dots$ , mathematicians define the **real numbers** to be all numbers which in principle can be written as a decimal expansion such as  $\sqrt{2}$ ,  $\pi$ ,

$$\frac{9}{7} = 1.285714285714 \dots \quad \text{or} \quad e = 2.718281828459 \dots .$$

Such decimal expansions may terminate or repeat or may need to continue on indefinitely (as denoted by the three dots, called an ellipsis). The frequently invoked symbol  $\mathbb{R}$  denotes the set of all possible real numbers.

In the sixteenth century Gerolamo Cardano developed a procedure to solve cubic polynomial equations. But the procedure involved manipulating  $\sqrt{-1}$  which seemed a crazy figment of imagination. Nonetheless the procedure worked. Subsequently, many practical uses were found for  $\sqrt{-1}$ , now denoted by  $i$  (or  $j$  in some disciplines). Consequently, many areas of modern science and engineering use **complex numbers** which are those of the form  $a + bi$  for real numbers  $a$  and  $b$ . The symbol  $\mathbb{C}$  denotes the set of all possible complex numbers. This book mostly uses integers and real numbers, but eventually we need the marvellous complex numbers.

This book uses the term **scalar** to denote a number that could be integer, real or complex. The term ‘scalar’ arises because such numbers are often used to scale the length of a ‘vector’.



## 1.1 Vectors have magnitude and direction

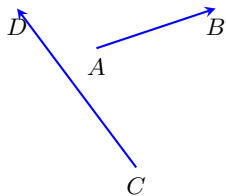
### Section Contents

There are more things in heaven and earth, Horatio,  
than are dreamt of in your philosophy.

(*Hamlet I.5:159–167*)

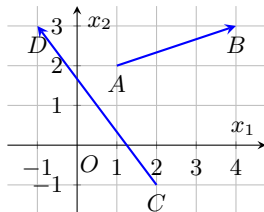
In the eighteenth century, astronomers needed to describe both the position and velocity of the planets. Such a description required quantities which have both a magnitude and a direction. Step outside, a wind blowing at 8 m/s from the south-west also has both a magnitude and direction. Quantities that have the properties of both a magnitude and a direction are called **vectors** (from the Latin for *carrier*).

**Example 1.1.1** (displacement vector). An important class of vectors are the so-called **displacement vectors**. Given two points in space, say  $A$  and  $B$ , the displacement vector  $\overrightarrow{AB}$  is the directed line

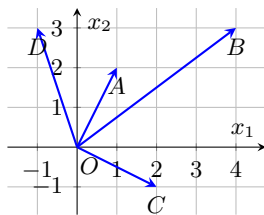
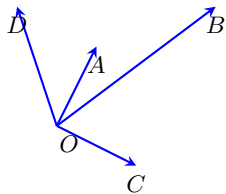


segment from the point  $A$  to the point  $B$ —as illustrated by the two displacement vectors  $\overrightarrow{AB}$  and  $\overrightarrow{CD}$  in the margin. For example, if your home is at position  $A$  and your school at position  $B$ , then travelling from home to school is to move by the amount of the displacement vector  $\overrightarrow{AB}$ .

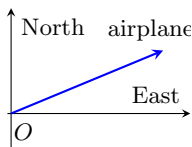
To be able to manipulate vectors we describe them with numbers. For such numbers to have meaning they must be set in the context of a coordinate system. So choose an origin for the coordinate system, usually denoted  $O$ , and draw coordinate axes in the plane (or space), as illustrated for the above two displacement vectors. Here the displacement vector  $\overrightarrow{AB}$  goes three units to the right and one unit up, so we denote it by the ordered pair of numbers  $\overrightarrow{AB} = (3, 1)$ . Whereas the displacement vector  $\overrightarrow{CD}$  goes three units to the left and four units up, so we denote it by the ordered pair of numbers  $\overrightarrow{CD} = (-3, 4)$ . ■



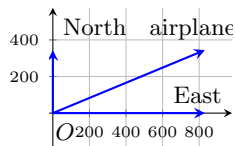
**Example 1.1.2** (position vector). The next important class of vectors are the **position vectors**. Given some chosen fixed origin in space, then  $\overrightarrow{OA}$  is the position vector of the point  $A$ . The marginal picture illustrates the position vectors of four points in the plane, given a chosen origin  $O$ .



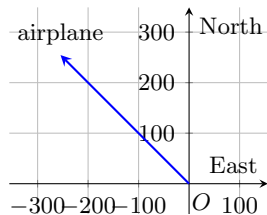
Again, to be able to manipulate such vectors we describe them with numbers, and such numbers have meaning via a coordinate system. So draw coordinate axes in the plane (or space), as illustrated for the above four position vectors. Here the position vector  $\overrightarrow{OA}$  goes one unit to the right and two units up so we denote it by  $\overrightarrow{OA} = (1, 2)$ . Similarly, the position vectors  $\overrightarrow{OB} = (4, 3)$ ,  $\overrightarrow{OC} = (2, -1)$ , and  $\overrightarrow{OD} = (-1, 3)$ . Recognise that the ordered pairs of numbers in the position vectors are exactly the coordinates of each of the specified end-points. ■



**Example 1.1.3** (velocity vector). Consider an airplane in level flight at 900 km/hr to the east-north-east. Choosing coordinate axes oriented to the East and the North, the direction of the airplane



is at an angle  $22.5^\circ$  from the East, as illustrated in the margin. Trigonometry then tells us that the Eastward part of the speed of the airplane is  $900 \cos(22.5^\circ) = 831.5 \text{ km/hr}$ , whereas the Northward part of the speed is  $900 \sin(22.5^\circ) = 344.4 \text{ km/hr}$  (as indicated in the margin). Further, the airplane is in level flight, not going up or down, so in the third direction of space (vertically) its speed component is zero. Putting these together forms the velocity vector  $(831.5, 344.4, 0)$  in  $\text{km/hr}$  in space.

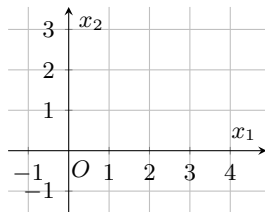


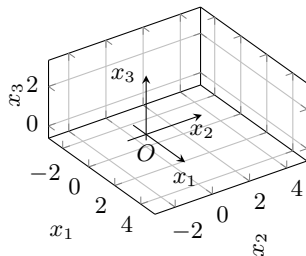
Another airplane takes off from an airport at  $360 \text{ km/hr}$  to the northwest and climbs at  $2 \text{ m/s}$ . The direction northwest is  $45^\circ$  to the East-West lines and  $45^\circ$  to the North-South lines. Trigonometry then tells us that the Westward speed of the airplane is  $360 \cos(45^\circ) = 360 \cos(\frac{\pi}{4}) = 254.6 \text{ km/hr}$ , whereas the Northward speed is  $360 \sin(45^\circ) = 360 \sin(\frac{\pi}{4}) = 254.6 \text{ km/hr}$  as illustrated in the margin. But West is the opposite direction to East, so if the coordinate system treats East as positive, then West must be negative. Consequently, together with the climb in the vertical, the velocity vector is  $(-254.6 \text{ km/hr}, 254.6 \text{ km/hr}, 2 \text{ m/s})$ . But it is best to avoid mixing units within a vector, so here convert all speeds

to m/s: here 360 km/hr upon dividing by 3600 secs/hr and multiplying by 1000 m/km gives 360 km/hr = 100 m/s. Then the North and West speeds are both  $100 \cos(\frac{\pi}{4}) = 70.71$  m/s. Consequently, the velocity vector of the climbing airplane should be described as  $(-70.71, 70.71, 2)$  in m/s. ■

In applications, as these examples illustrate, the ‘physical’ vector exists before the coordinate system. It is only when we choose a specific coordinate system that a ‘physical’ vector gets expressed by numbers. Throughout, unless otherwise specified, this book assumes that vectors are expressed in what is called a **standard coordinate system**.

- In the two dimensions of the plane the standard coordinate system has two coordinate axes, one horizontal and one vertical at right-angles to each other, often labelled  $x_1$  and  $x_2$  respectively (as illustrated in the margin), although labels  $x$  and  $y$  are also common.
- In the three dimensions of space the standard coordinate





system has three coordinate axes, two horizontal and one vertical all at right-angles to each other, often labelled  $x_1$ ,  $x_2$  and  $x_3$  respectively (as illustrated in the margin), although labels  $x$ ,  $y$  and  $z$  are also common.

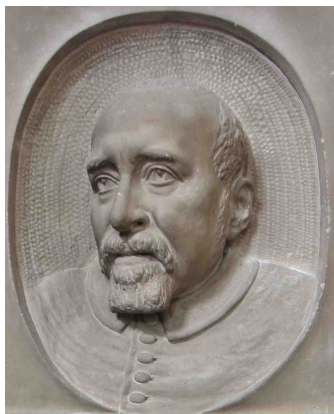
- Correspondingly, in so-called ‘ $n$  dimensions’ the standard coordinate system has  $n$  coordinate axes, all at right-angles to each other, and often labelled  $x_1, x_2, \dots, x_n$ , respectively.

**Definition 1.1.4.** *Given a standard coordinate system with  $n$  coordinate axes, all at right-angles to each other, a **vector** is an ordered  $n$ -tuple of real numbers  $x_1, x_2, \dots, x_n$  equivalently written either as a row in parentheses or as a column in brackets,*

$$(x_1, x_2, \dots, x_n) = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

*(they mean the same, it is just more convenient to usually use a row in parentheses in text, and a column in brackets in displayed*

mathematics). The real numbers  $x_1, x_2, \dots, x_n$  are called the **components** of the vector, and the number of components is termed its **size** (here  $n$ ). The components are determined such that letting  $X$  be the point with coordinates  $(x_1, x_2, \dots, x_n)$  then the position vector  $\overrightarrow{OX}$  has the same magnitude and direction as the vector denoted  $(x_1, x_2, \dots, x_n)$ . Two vectors of the same size are **equal**,  $=$ , if all their corresponding components are equal (vectors with different sizes are never equal).



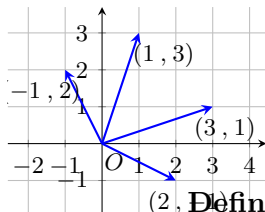
Robert Recorde invented the equal sign circa 1557 “because noe 2 thynges can be moare equalle”.

Examples 1.1.1 and 1.1.2 introduced some vectors and wrote them as a row in parentheses, such as  $\overrightarrow{AB} = (3, 1)$ . In this book exactly the same thing is meant by the columns in brackets: for example,

$$\overrightarrow{AB} = (3, 1) = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \quad \overrightarrow{CD} = (-3, 4) = \begin{bmatrix} -3 \\ 4 \end{bmatrix},$$

$$\overrightarrow{OC} = (2, -1) = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \quad (-70.71, 70.71, 2) = \begin{bmatrix} -70.71 \\ 70.71 \\ 2 \end{bmatrix}.$$

However, as defined subsequently, a row of numbers within brackets is quite different:  $(3, 1) \neq [3 \ 1]$ , and  $(831, 344, 0) \neq [831 \ 344 \ 0]$ .



The *ordering* of the components is very important. For example, as illustrated in the margin, the vector  $(3, 1)$  is very different from the vector  $(1, 3)$ ; similarly, the vector  $(2, -1)$  is very different from the vector  $(-1, 2)$ .

**Definition 1.1.5.** *The set of all vectors with  $n$  components is denoted  $\mathbb{R}^n$ . The vector with all components zero,  $(0, 0, \dots, 0)$ , is called the **zero vector** and denoted by  $\mathbf{0}$ .*

**Example 1.1.6.**

- All the vectors we can draw and imagine in the two dimensional plane form  $\mathbb{R}^2$ . Sometimes we write that  $\mathbb{R}^2$  is the plane because of this very close connection.
- All the vectors we can draw and imagine in three dimensional space form  $\mathbb{R}^3$ . Again, sometimes we write that  $\mathbb{R}^3$  is three dimensional space because of the close connection.
- The set  $\mathbb{R}^1$  is the set of all vectors with one component, and that one component is measured along one axis. Hence  $\mathbb{R}^1$  is effectively the same as the set of real numbers labelling that axis.





As just introduced for the zero vector  $\mathbf{0}$ , this book generally denotes vectors by a bold letter (except for displacement vectors). The other common notation you may see elsewhere is to denote vectors by a small over-arrow such as in the “zero vector  $\vec{0}$ ”. Less commonly, some books and articles use an over- or under-tilde ( $\sim$ ) to denote vectors. Be aware of this different notation in reading other books.

Question: why do we need vectors with  $n$  components, in  $\mathbb{R}^n$ , when the world around us is only three dimensional? Answer: because vectors can encode much more than spatial structure as in the next example.

**Example 1.1.7** (linguistic vectors). Consider the following four sentences.

- (a) The dog sat on the mat.
- (b) The cat scratched the dog.
- (c) The cat and dog sat on the mat.

(d) The dog scratched.

These four sentences involve up to three objects, cat, dog and mat, and two actions, sat and scratched. Some characteristic of the sentences is captured simply by counting the number of times each of these three objects and two actions appear in each sentence, and then forming a vector from the counts. Let's use vectors  $\mathbf{w} = (N_{\text{cat}}, N_{\text{dog}}, N_{\text{mat}}, N_{\text{sat}}, N_{\text{scratched}})$  where the various  $N$  are the counts of each word ( $\mathbf{w}$  for words). The previous statement implicitly specifies that we use five coordinate axes, perhaps labelled “cat”, “dog”, “mat”, “sat” and “scratched”, and that distance along each axis represents the number of times the corresponding word is used. These word vectors are in  $\mathbb{R}^5$ . Then

- (a) “The dog sat on the mat” is summarised by the vector  $\mathbf{w} = (0, 1, 1, 1, 0)$ .
- (b) “The cat scratched the dog” is summarised by the vector  $\mathbf{w} = (1, 1, 0, 0, 1)$ .
- (c) “The cat and dog sat on the mat” is summarised by the vector  $\mathbf{w} = (1, 1, 1, 1, 0)$ .

- (d) “The dog scratched” is summarised by the vector  $\mathbf{w} = (0, 1, 0, 0, 1)$ .
- (e) An empty sentence is the zero vector  $\mathbf{w} = (0, 0, 0, 0, 0)$ .
- (f) Together, the two sentences “The dog sat on the mat. The cat scratched the dog.” are summarised by the vector  $\mathbf{w} = (1, 2, 1, 1, 1)$ .

Using such crude summary representations of some text, even of entire documents, empowers us to use powerful mathematical techniques to relate documents together, compare and contrast, express similarities, look for type clusters, and so on. In application we would not just count words for objects (nouns) and actions (verbs), but also qualifications (adjectives and adverbs).

People generally know and use thousands of words. Consequently, in practice, such word vectors typically have thousands of components corresponding to coordinate axes of thousands of distinct words. To cope with such vectors of many components, modern linear algebra has been developed to powerfully handle problems involving vectors with thousands, millions or even an ‘infinite num-

ber' of components. ■

**Activity 1.1.8.** Given word vectors  $\mathbf{w} = (N_{\text{cat}}, N_{\text{dog}}, N_{\text{mat}}, N_{\text{sat}}, N_{\text{scratched}})$  as in [Example 1.1.7](#), which of the following has word vector  $\mathbf{w} = (2, 2, 0, 2, 1)$ ?

- (a) “The dog scratched the cat on the mat.”
  - (b) “A dog sat. A cat scratched the dog. The cat sat.”
  - (c) “Which cat sat by the dog on the mat, and then scratched the dog.”
  - (d) “A dog and cat both sat on the mat which the dog had scratched.”
-

---

**King – man + woman = queen**

Computational linguistics has dramatically changed the way researchers study and understand language. The ability to number-crunch huge amounts of words for the first time has led to entirely new ways of thinking about words and their relationship to one another.

This number-crunching shows exactly how often a word appears close to other words, an important factor in how they are used. So the word Olympics might appear close to words like running, jumping, and throwing but less often next to words like electron or stegosaurus. This set of relationships can be thought of as a multidimensional vector that describes how the word Olympics is used within a language, which itself can be thought of as a vector space.

And therein lies this massive change. This new approach allows languages to be treated like vector spaces with precise mathematical properties. Now the study of language is becoming a problem of vector space mathematics. *Technology Review, 2015*

---

**Definition 1.1.9** (Pythagoras). *For every vector  $\mathbf{v} = (v_1, v_2, \dots, v_n)$  in  $\mathbb{R}^n$ , define the **length** (or **magnitude**) of vector  $\mathbf{v}$  to be the real number  $(\geq 0)$*

$$|\mathbf{v}| := \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}.$$

*A vector of length one is called a **unit vector**. (Many people and books denote the length of a vector with a pair of double lines, as in  $\|\mathbf{v}\|$ . Either notation is good.)*

**Example 1.1.10.** Find the lengths of the following vectors.

(a)  $\mathbf{a} = (-3, 4)$

(b)  $\mathbf{b} = (3, 3)$

(c)  $\mathbf{c} = (1, -2, 3)$

(d)  $\mathbf{d} = (1, -1, -1, 1)$



**Example 1.1.11.** Write down three different vectors, all three with the same number of components, that are (a) of length 5, (b) of length 3, and (c) of length  $-2$ . ■

**Activity 1.1.12.** What is the length of the vector  $(2, -3, 6)$ ?

- (a) 5                      (b) 11                      (c)  $\sqrt{11}$                       (d) 7

**Theorem 1.1.13.** *The zero vector is the only vector of length zero:  $|\mathbf{v}| = 0$  if and only if  $\mathbf{v} = \mathbf{0}$ .*

## 1.2 Adding and stretching vectors

### Section Contents

1.2.1	Basic operations . . . . .	24
	Distance . . . . .	33
1.2.2	Parametric equation of a line . . . . .	36
1.2.3	Manipulation requires algebraic properties . .	40

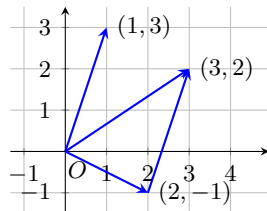
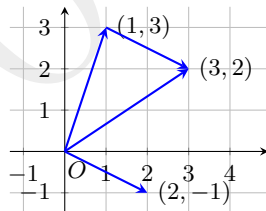
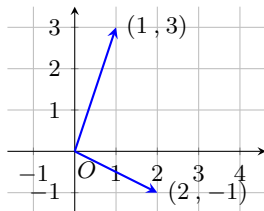
We want to be able to make sense of statements such as “king – man + women = queen”. To do so we need to define operations on vectors. Useful operations on vectors are those which are physically meaningful. Then our algebraic manipulations will derive powerful results in applications. The first two vector operations are addition and scalar multiplication.



### 1.2.1 Basic operations

**Example 1.2.1.** Vectors of the same size are added component-wise. Equivalently, obtain the same result by geometrically joining the two vectors ‘head-to-tail’ and drawing the vector from the start to the finish.

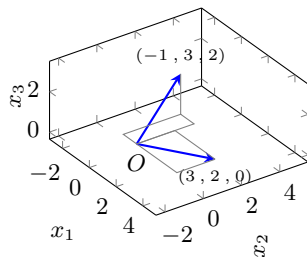
- (a)  $(1, 3) + (2, -1) = (1 + 2, 3 + (-1)) = (3, 2)$  as illustrated below where (given the two vectors plotted in the margin) the vector  $(2, -1)$  is drawn from the end of  $(1, 3)$ , and the end point of the result determines the vector addition  $(3, 2)$ , as shown below-left.



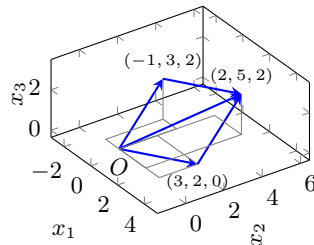
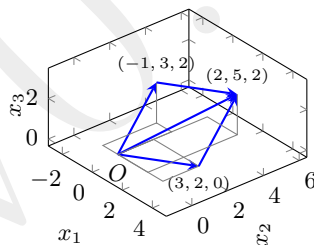
This result  $(3, 2)$  is the same if the vector  $(1, 3)$  is drawn from the end of  $(2, -1)$  as shown above-right. That is,

$(2, -1) + (1, 3) = (1, 3) + (2, -1)$ . That the order of addition is immaterial is the commutative law of vector addition that is established in general by [Theorem 1.2.19a](#).

- (b)  $(3, 2, 0) + (-1, 3, 2) = (3 + (-1), 2 + 3, 0 + 2) = (2, 5, 2)$  as illustrated below where (given the two vectors as plotted in the margin) the vector  $(-1, 3, 2)$  is drawn from the end of  $(3, 2, 0)$ , and the end point of the result determines the vector addition  $(2, 5, 2)$ . As below, find the same result by drawing the vector  $(3, 2, 0)$  from the end of  $(-1, 3, 2)$ .



I implement such cross-eyed stereo so that these stereo images are useful when projected on a large screen.



As drawn above, many of the three-D plots in this book are **stereo pairs** drawing the plot from two slightly different

viewpoints: cross your eyes to merge two of the images, and then focus on the pair of plots to see the three-D effect. With practice viewing such three-D stereo pairs becomes less difficult.

- (c) The addition  $(1, 3) + (3, 2, 0)$  is not defined and cannot be done as the two vectors have a different number of components, different sizes.



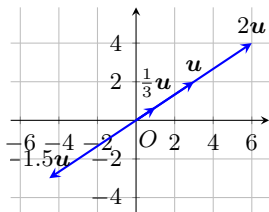
**Example 1.2.2.** To multiply a vector by a scalar, a number, multiply each component by the scalar. Equivalently, visualise the result through stretching the vector by a factor of the scalar.

- (a) Let the vector  $\mathbf{u} = (3, 2)$  then, as illustrated in the margin,

$$2\mathbf{u} = 2(3, 2) = (2 \cdot 3, 2 \cdot 2) = (6, 4),$$

$$\frac{1}{3}\mathbf{u} = \frac{1}{3}(3, 2) = \left(\frac{1}{3} \cdot 3, \frac{1}{3} \cdot 2\right) = \left(1, \frac{2}{3}\right),$$

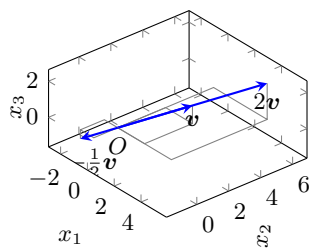
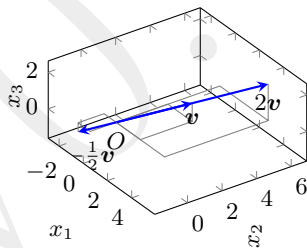
$$(-1.5)\mathbf{u} = (-1.5 \cdot 3, -1.5 \cdot 2) = (-4.5, -3).$$



(b) Let the vector  $\mathbf{v} = (2, 3, 1)$  then, as illustrated below in stereo,

$$2\mathbf{v} = 2 \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \cdot 2 \\ 2 \cdot 3 \\ 2 \cdot 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \\ 2 \end{bmatrix},$$

$$\left(-\frac{1}{2}\right)\mathbf{v} = -\frac{1}{2} \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \cdot 2 \\ -\frac{1}{2} \cdot 3 \\ -\frac{1}{2} \cdot 1 \end{bmatrix} = \begin{bmatrix} -1 \\ -\frac{3}{2} \\ -\frac{1}{2} \end{bmatrix}.$$



**Activity 1.2.3.** Combining multiplication and addition, what is  $\mathbf{u} + 2\mathbf{v}$  for vectors  $\mathbf{u} = (4, 1)$  and  $\mathbf{v} = (-1, -3)$ ?

- (a)  $(1, -8)$       (b)  $(2, -5)$       (c)  $(5, -8)$       (d)  $(3, -2)$



**Definition 1.2.4.** Let two vectors in  $\mathbb{R}^n$  be  $\mathbf{u} = (u_1, u_2, \dots, u_n)$  and  $\mathbf{v} = (v_1, v_2, \dots, v_n)$ , and let  $c$  be a scalar. Then the **sum** or **addition** of  $\mathbf{u}$  and  $\mathbf{v}$ , denoted  $\mathbf{u} + \mathbf{v}$ , is the vector obtained by joining  $\mathbf{v}$  to  $\mathbf{u}$  ‘head-to-tail’, and is computed as

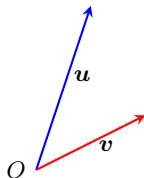
$$\mathbf{u} + \mathbf{v} := (u_1 + v_1, u_2 + v_2, \dots, u_n + v_n).$$

The **scalar multiplication** of  $\mathbf{u}$  by  $c$ , denoted  $c\mathbf{u}$ , is the vector of length  $|c||\mathbf{u}|$  in the direction of  $\mathbf{u}$  when  $c > 0$  but in the opposite direction when  $c < 0$ , and is computed as

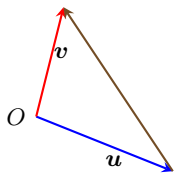
$$c\mathbf{u} := (cu_1, cu_2, \dots, cu_n).$$

The **negative** of  $\mathbf{u}$  denoted  $-\mathbf{u}$ , is defined as the scalar multiple  $-\mathbf{u} = (-1)\mathbf{u}$ , and is a vector of the same length as  $\mathbf{u}$  but in exactly

the opposite direction. The **difference**  $\mathbf{u} - \mathbf{v}$  is defined as  $\mathbf{u} + (-\mathbf{v})$  and is equivalently the vector drawn from the end of  $\mathbf{v}$  to the end of  $\mathbf{u}$ .

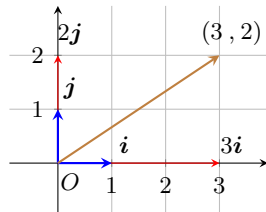


**Example 1.2.5.** For the vectors  $\mathbf{u}$  and  $\mathbf{v}$  shown in the margin, draw the vectors  $\mathbf{u} + \mathbf{v}$ ,  $\mathbf{v} + \mathbf{u}$ ,  $\mathbf{u} - \mathbf{v}$ ,  $\mathbf{v} - \mathbf{u}$ ,  $\frac{1}{2}\mathbf{u}$  and  $-\mathbf{v}$ . ■



**Activity 1.2.6.** For the vectors  $\mathbf{u}$  and  $\mathbf{v}$  shown in the margin, what is the result vector that is also shown? ■

- (a)  $\mathbf{v} - \mathbf{u}$       (b)  $\mathbf{u} - \mathbf{v}$       (c)  $\mathbf{u} + \mathbf{v}$       (d)  $\mathbf{v} + \mathbf{u}$



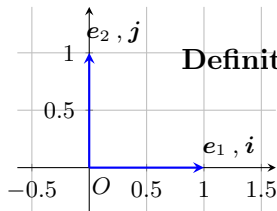
Using vector addition and scalar multiplication we often write vectors in terms of so-called standard unit vectors. In the plane and drawn in the margin are the two unit vectors  $\mathbf{i}$  and  $\mathbf{j}$  (length one) in the direction of the two coordinate axes. Then, for example,

$$\begin{aligned}
 (3, 2) &= (3, 0) + (0, 2) \quad (\text{by addition}) \\
 &= 3(1, 0) + 2(0, 1) \quad (\text{by scalar mult}) \\
 &= 3\mathbf{i} + 2\mathbf{j} \quad (\text{by definition of } \mathbf{i} \text{ and } \mathbf{j}).
 \end{aligned}$$

Similarly, in three dimensional space we often write vectors in terms of the three vectors  $\mathbf{i}$ ,  $\mathbf{j}$  and  $\mathbf{k}$ , each of length one, aligned along the three coordinate axes. For example,

$$\begin{aligned}
 (2, 3, -1) &= (2, 0, 0) + (0, 3, 0) + (0, 0, -1) \quad (\text{by addition}) \\
 &= 2(1, 0, 0) + 3(0, 1, 0) - (0, 0, 1) \quad (\text{by scalar mult}) \\
 &= 2\mathbf{i} + 3\mathbf{j} - \mathbf{k} \quad (\text{by definition of } \mathbf{i}, \mathbf{j} \text{ and } \mathbf{k}).
 \end{aligned}$$

The next definition generalises these standard unit vectors to vectors in  $\mathbb{R}^n$ .

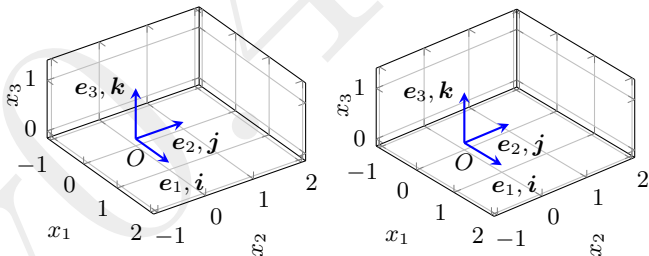


**Definition 1.2.7.**

Given a standard coordinate system with  $n$  coordinate axes, all at right-angles to each other, the **standard unit vectors**  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$  are the vectors of length one in the direction of the corresponding coordinate axis (as illustrated in the margin for  $\mathbb{R}^2$  and below for  $\mathbb{R}^3$ ). That is,

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad \dots \quad \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

In  $\mathbb{R}^2$  and  $\mathbb{R}^3$  the symbols  $\mathbf{i}$ ,  $\mathbf{j}$  and  $\mathbf{k}$  are often used as synonyms for  $\mathbf{e}_1$ ,  $\mathbf{e}_2$  and  $\mathbf{e}_3$ , respectively (as also illustrated).



That is, for three examples, the following are equivalent ways of writing the same vector:

$$(3, 2) = \begin{bmatrix} 3 \\ 2 \end{bmatrix} = 3\mathbf{i} + 2\mathbf{j} = 3\mathbf{e}_1 + 2\mathbf{e}_2;$$



$$(2, 3, -1) = \begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix} = 2\mathbf{i} + 3\mathbf{j} - \mathbf{k} = 2\mathbf{e}_1 + 3\mathbf{e}_2 - \mathbf{e}_3;$$

$$(0, -3.7, 0, 0.1, -3.9) = \begin{bmatrix} 0 \\ -3.7 \\ 0 \\ 0.1 \\ -3.9 \end{bmatrix} = -3.7\mathbf{e}_2 + 0.1\mathbf{e}_4 - 3.9\mathbf{e}_5.$$

**Activity 1.2.8.** Which of the following is the same as the vector  $3\mathbf{e}_2 + \mathbf{e}_5$ ?

(a)  $(0, 3, 0, 0, 1)$

(b)  $(5, 0, 2)$

(c)  $(3, 1)$

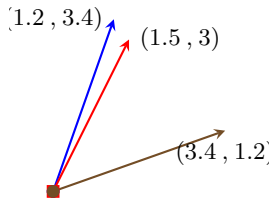
(d)  $(0, 3, 0, 1)$



## Distance

Defining a ‘distance’ between vectors empowers us to compare vectors concisely.

**Example 1.2.9.** We would like to say that  $(1.2, 3.4) \approx (1.5, 3)$  to an error 0.5 (as illustrated in the margin). Why 0.5? Because the difference between the vectors  $(1.5, 3) - (1.2, 3) = (0.3, -0.4)$  has length  $\sqrt{0.3^2 + (-0.4)^2} = 0.5$ .

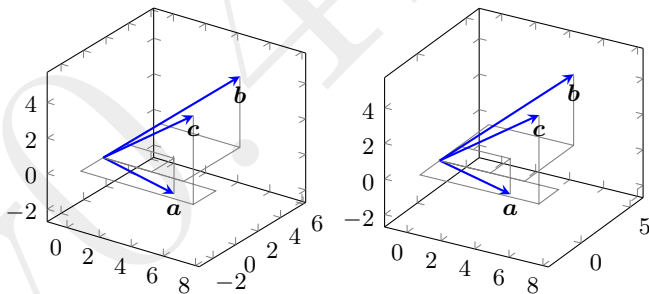


Conversely, we would like to recognise that vectors  $(1.2, 3.4)$  and  $(3.4, 1.2)$  are very different (as also illustrated in the margin)—there is a large ‘distance’ between them. Why is there a large ‘distance’? Because the difference between the vectors  $(1.2, 3.4) - (3.4, 1.2) = (-2.2, 2.2)$  has length  $\sqrt{(-2.2)^2 + 2.2^2} = 2.2\sqrt{2} = 3.1113$  which is relatively large. ■

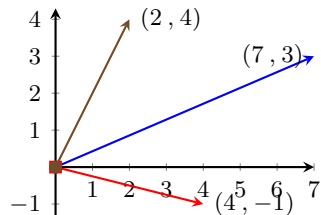
This concept of distance between two vectors  $\mathbf{u}$  and  $\mathbf{v}$ , directly corresponding to the distance between two points, is the length  $|\mathbf{u} - \mathbf{v}|$ .

**Definition 1.2.10.** The *distance* between vectors  $\mathbf{u}$  and  $\mathbf{v}$  in  $\mathbb{R}^n$  is the length of their difference,  $|\mathbf{u} - \mathbf{v}|$ .

**Example 1.2.11.** Given three vectors  $\mathbf{a} = 3\mathbf{i} + 2\mathbf{j} - 2\mathbf{k}$ ,  $\mathbf{b} = 5\mathbf{i} + 5\mathbf{j} + 4\mathbf{k}$  and  $\mathbf{c} = 7\mathbf{i} - 2\mathbf{j} + 5\mathbf{k}$  (shown below in stereo): which pair are the closest to each other? and which pair are furthest from each other?



**Activity 1.2.12.** Which pair of the following vectors are closest—have the smallest distance between them?  $\mathbf{a} = (7, 3)$ ,  $\mathbf{b} = (4, -1)$ ,  $\mathbf{c} = (2, 4)$



(a) two of the pairs

(b)  $\mathbf{a}$ ,  $\mathbf{b}$

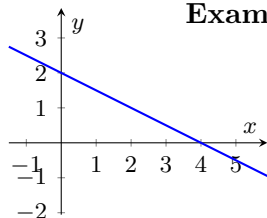
(c)  $\mathbf{b}$ ,  $\mathbf{c}$

(d)  $\mathbf{a}$ ,  $\mathbf{c}$



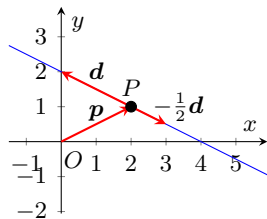
### 1.2.2 Parametric equation of a line

We are familiar with lines in the plane, and equations that describe them. Let's now consider such equations from a vector view. The insights empower us to generalise the descriptions to lines in space, and then in any number of dimensions.

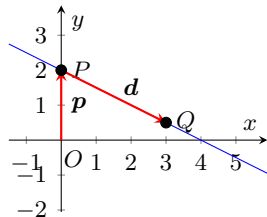


**Example 1.2.13.**

Consider the line drawn in the margin in some chosen coordinate system. Recall one way to find an equation of the line is to find the intercepts with the axes, here at  $x = 4$  and at  $y = 2$ , then write down  $\frac{x}{4} + \frac{y}{2} = 1$  as an equation of the line. Algebraic rearrangement gives various other forms, such as  $x + 2y = 4$  or  $y = 2 - x/2$ .



The alternative is to describe the line with vectors. Choose any point  $P$  on the line, such as  $(2, 1)$  as drawn in the margin. Then view every other point on the line as having position vector that is the vector sum of  $\overrightarrow{OP}$  and a vector aligned along the line. Denote  $\overrightarrow{OP}$  by  $\mathbf{p}$  as drawn. Then, for example, the point  $(0, 2)$  on the line has position vector  $\mathbf{p} + \mathbf{d}$  for vector  $\mathbf{d} = (-2, 1)$  because  $\mathbf{p} + \mathbf{d} = (2, 1) + (-2, 1) = (0, 2)$ . Other points on the line are also given using

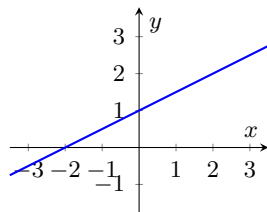


the same vectors,  $\mathbf{p}$  and  $\mathbf{d}$ : for example, the point  $(3, \frac{1}{2})$  has position vector  $\mathbf{p} - \frac{1}{2}\mathbf{d}$  (as drawn) because  $\mathbf{p} - \frac{1}{2}\mathbf{d} = (2, 1) - \frac{1}{2}(-2, 1) = (3, \frac{1}{2})$ ; and the point  $(-2, 3)$  has position vector  $\mathbf{p} + 2\mathbf{d} = (2, 1) + 2(-2, 1)$ . In general, every point on the line may be expressed as  $\mathbf{p} + t\mathbf{d}$  for some scalar  $t$ .

For any given line, there are many possible choices of  $\mathbf{p}$  and  $\mathbf{d}$  in such a vector representation. A different looking, but equally valid form is obtained from any pair of points on the line. For example, one could choose point  $P$  to be  $(0, 2)$  and point  $Q$  to be  $(3, \frac{1}{2})$ , as drawn in the margin. Let position vector  $\mathbf{p} = \overrightarrow{OP} = (0, 2)$  and the vector  $\mathbf{d} = \overrightarrow{PQ} = (3, -\frac{3}{2})$ , then every point on the line has position vector  $\mathbf{p} + t\mathbf{d}$  for some scalar  $t$ :

$$\begin{aligned}(2, 1) &= (0, 2) + (2, -1) = (0, 2) + \frac{2}{3}(3, -\frac{3}{2}) = \mathbf{p} + \frac{2}{3}\mathbf{d}; \\(6, -1) &= (0, 2) + (6, -3) = (0, 2) + 2(3, -\frac{3}{2}) = \mathbf{p} + 2\mathbf{d}; \\(-1, \frac{5}{2}) &= (0, 2) + (-1, \frac{1}{2}) = (0, 2) - \frac{1}{3}(3, -\frac{3}{2}) = \mathbf{p} - \frac{1}{3}\mathbf{d}.\end{aligned}$$

Other choices of points  $P$  and  $Q$  give other valid vector equations for a given line. ■



**Activity 1.2.14.** Which one of the following is *not* a valid vector equation for the line plotted in the margin?

(a)  $(-2, 0) + (-4, -2)t$

(b)  $(0, 1) + (2, 1)t$

(c)  $(-1, 1/2) + (2, -1)t$

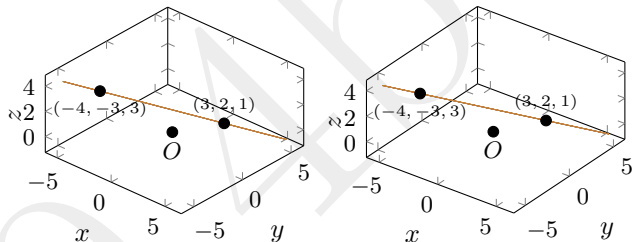
(d)  $(2, 2) + (1, 1/2)t$



**Definition 1.2.15.** A *parametric equation* of a line is  $\mathbf{x} = \mathbf{p} + t\mathbf{d}$  where  $\mathbf{p}$  is the position vector of some point on the line, the so-called *direction vector*  $\mathbf{d}$  is parallel to the line ( $\mathbf{d} \neq \mathbf{0}$ ), and the scalar *parameter*  $t$  varies over all real values to give all position vectors  $\mathbf{x}$  on the line.

Beautifully, this definition applies for lines in any number of dimensions by using vectors with the corresponding number of components.

**Example 1.2.16.** Given that the line drawn below in space goes through points  $(-4, -3, 3)$  and  $(3, 2, 1)$ , find a parametric equation of the line.



**Example 1.2.17.** Given the parametric equation of a line in space (in stereo) is  $\mathbf{x} = (-4 + 2t, 3 - t, -1 - 4t)$ , find the value of the parameter  $t$  that gives each of the following points on the line:  $(-1.6, 1.8, -5.8)$ ,  $(-3, 2.5, -3)$ , and  $(-6, 4, 4)$ .



### 1.2.3 Manipulation requires algebraic properties

It seems to be nothing other than that art which they call by the barbarous name of ‘algebra’, if only it could be disentangled from the multiple numbers and inexplicable figures that overwhelm it ... *Descartes*

To unleash the power of algebra on vectors, we need to know the properties of vector operations. Many of the following properties are familiar as they directly correspond to familiar properties of arithmetic operations on scalars. Moreover, the proofs show the vector properties follow directly from the familiar properties of arithmetic operations on scalars.

**Example 1.2.18.** Let vectors  $\mathbf{u} = (1, 2)$ ,  $\mathbf{v} = (3, 1)$ , and  $\mathbf{w} = (-2, 3)$ , and let scalars  $a = -\frac{1}{2}$  and  $b = \frac{5}{2}$ . Verify the following properties hold:

- (a)  $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$  (commutative law);
- (b)  $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$  (associative law);
- (c)  $\mathbf{u} + \mathbf{0} = \mathbf{u}$ ;

- (d)  $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$ ;
- (e)  $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$  (a distributive law);
- (f)  $(a + b)\mathbf{u} = a\mathbf{u} + b\mathbf{u}$  (a distributive law);
- (g)  $(ab)\mathbf{u} = a(b\mathbf{u})$ ;
- (h)  $1\mathbf{u} = \mathbf{u}$ ;
- (i)  $0\mathbf{u} = \mathbf{0}$ ;
- (j)  $|a\mathbf{u}| = |a| \cdot |\mathbf{u}|$ .



Now let's state and prove these properties in general.

**Theorem 1.2.19.** *For all vectors  $\mathbf{u}$ ,  $\mathbf{v}$  and  $\mathbf{w}$  with  $n$  components (that is, in  $\mathbb{R}^n$ ), and for all scalars  $a$  and  $b$ , the following properties hold:*

- (a)  $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$  (commutative law);
- (b)  $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$  (associative law);

$$(c) \mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u};$$

$$(d) \mathbf{u} + (-\mathbf{u}) = (-\mathbf{u}) + \mathbf{u} = \mathbf{0};$$

$$(e) a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v} \quad (a \text{ distributive law});$$

$$(f) (a + b)\mathbf{u} = a\mathbf{u} + b\mathbf{u} \quad (a \text{ distributive law});$$

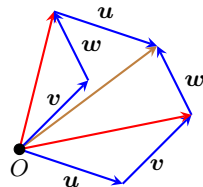
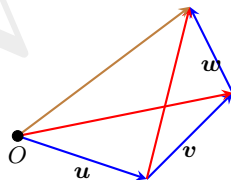
$$(g) (ab)\mathbf{u} = a(b\mathbf{u});$$

$$(h) 1\mathbf{u} = \mathbf{u};$$

$$(i) 0\mathbf{u} = \mathbf{0};$$

$$(j) |a\mathbf{u}| = |a| \cdot |\mathbf{u}|.$$

**Example 1.2.20.** Which of the following two diagrams best illustrates the associative law 1.2.19b? Give reasons.





We frequently use the algebraic properties of [Theorem 1.2.19](#) in rearranging and solving vector equations.

**Example 1.2.21.** Find the vector  $\mathbf{x}$  such that  $3\mathbf{x} - 2\mathbf{u} = 6\mathbf{v}$ .



**Example 1.2.22.** Rearrange  $3\mathbf{x} - \mathbf{a} = 2(\mathbf{a} + \mathbf{x})$  to write vector  $\mathbf{x}$  in terms of  $\mathbf{a}$ : give excruciating detail of the justification using [Theorem 1.2.19](#).



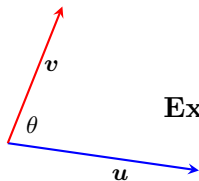
## 1.3 The dot product determines angles and lengths

### Section Contents

1.3.1	Work done involves the dot product . . . . .	54
1.3.2	Algebraic properties of the dot product . . .	57
1.3.3	Orthogonal vectors are at right-angles . . . .	62
1.3.4	Normal vectors and equations of a plane . . .	66

The previous [Section 1.2](#) discussed how to add, subtract and stretch vectors. Question: can we multiply two vectors? The answer is that ‘vector multiplication’ has major differences to the multiplication of scalar numbers. This section introduces the so-called dot product of two vectors that, among other attributes, gives a valuable way to determine the angle between the two vectors.

Often the angle between vectors is denoted by the Greek letter theta,  $\theta$ .



**Example 1.3.1.** Consider the two vectors  $u = (7, -1)$  and  $v = (2, 5)$  plotted in the margin. What is the angle  $\theta$  between the two vectors? ■

The interest in this [Example 1.3.1](#) is the number nine on the right-hand side of  $|\mathbf{u}||\mathbf{v}|\cos\theta = 9$ . The reason is that 9 just happens to be  $14 - 5$ , which in turn just happens to be  $7 \cdot 2 + (-1) \cdot 5$ , and it is no coincidence that this expression is the same as  $u_1v_1 + u_2v_2$  in terms of vector components  $\mathbf{u} = (u_1, u_2) = (7, -1)$  and  $\mathbf{v} = (v_1, v_2) = (2, 5)$ . Repeat this example for many pairs of vectors  $\mathbf{u}$  and  $\mathbf{v}$  to find that always  $|\mathbf{u}||\mathbf{v}|\cos\theta = u_1v_1 + u_2v_2$  (??). This equality suggests that the sum of products of corresponding components of  $\mathbf{u}$  and  $\mathbf{v}$  is closely connected to the angle between the vectors.

**Definition 1.3.2.** *For every two vectors in  $\mathbb{R}^n$ ,  $\mathbf{u} = (u_1, u_2, \dots, u_n)$  and  $\mathbf{v} = (v_1, v_2, \dots, v_n)$ , define the **dot product** (or **inner product**), denoted by a dot between the two vectors, as the scalar*

$$\mathbf{u} \cdot \mathbf{v} := u_1v_1 + u_2v_2 + \cdots + u_nv_n.$$

The dot product of two vectors gives a scalar result, a number, not a vector result.

When writing the vector dot product, the dot between the two vectors is essential. We sometimes also denote the scalar product

by such a dot (to clarify a product) and sometimes omit the dot between the scalars, for example  $a \cdot b = ab$  for scalars. But for the vector dot product the dot must not be omitted: ' $\mathbf{uv}$ ' is meaningless.

**Example 1.3.3.** Compute the dot product between the following pairs of vectors.

(a)  $\mathbf{u} = (-2, 5, -2)$ ,  $\mathbf{v} = (3, 3, -2)$

(b)  $\mathbf{u} = (1, -3, 0)$ ,  $\mathbf{v} = (1, 2)$

(c)  $\mathbf{a} = (-7, 3, 0, 2, 2)$ ,  $\mathbf{b} = (-3, 4, -4, 2, 0)$

(d)  $\mathbf{p} = (-0.1, -2.5, -3.3, 0.2)$ ,  $\mathbf{q} = (-1.6, 1.1, -3.4, 2.2)$



**Activity 1.3.4.** What is the dot product of the two vectors  $\mathbf{u} = 2\mathbf{i} - \mathbf{j}$  and  $\mathbf{v} = 3\mathbf{i} + 4\mathbf{j}$ ?

(a) 2

(b) 5

(c) 8

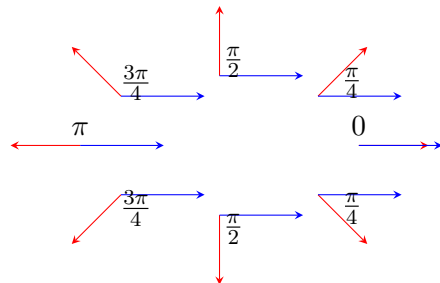
(d) 10



**Theorem 1.3.5.** For every two non-zero vectors  $\mathbf{u}$  and  $\mathbf{v}$  in  $\mathbb{R}^n$ , the *angle*  $\theta$  between the vectors is determined by

$$\cos \theta = \frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{u}||\mathbf{v}|}, \quad 0 \leq \theta \leq \pi \quad (0 \leq \theta \leq 180^\circ).$$

This picture illustrates the range of angles between two vectors: when they point in the same direction the angle is zero; when they are at right-angles to each other the angle is  $\pi/2$ , or equivalently  $90^\circ$ ;





when they point in opposite directions the angle is  $\pi$ , or equivalently  $180^\circ$ .

**Example 1.3.6.** Determine the angle between the following pairs of vectors.

- (a)  $(4, 3)$  and  $(5, 12)$
- (b)  $(3, 1)$  and  $(-2, 1)$
- (c)  $(4, -2)$  and  $(-1, -2)$



**Activity 1.3.7.** What is the angle between the two vectors  $(1, \sqrt{3})$  and  $(\sqrt{3}, 1)$ ?

- (a)  $60^\circ$
- (b)  $77.50^\circ$
- (c)  $30^\circ$
- (d)  $64.34^\circ$

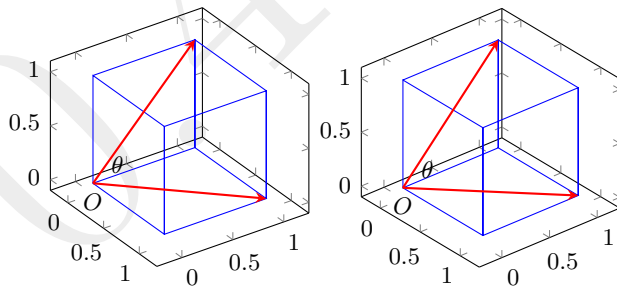


Table 1.1: when a cosine is one of these tabulated special values, then we know the corresponding angle exactly. In other cases we usually use a calculator (`arccos` or  $\cos^{-1}$ ) or computer (`acos()`) to compute the angle numerically.

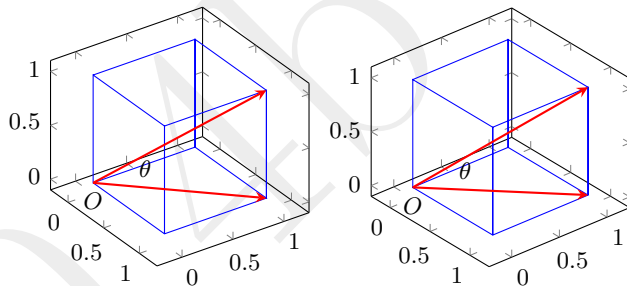
$\theta$	$\theta$	$\cos \theta$	$\cos \theta$
0	0°	1	1.
$\pi/6$	30°	$\sqrt{3}/2$	0.8660
$\pi/4$	45°	$1/\sqrt{2}$	0.7071
$\pi/3$	60°	1/2	0.5
$\pi/2$	90°	0	0.
$2\pi/3$	120°	-1/2	-0.5
$3\pi/4$	135°	$-1/\sqrt{2}$	-0.7071
$5\pi/6$	150°	$-\sqrt{3}/2$	-0.8660
$\pi$	180°	-1	-1.

**Example 1.3.8.** In chemistry one computes the angles between bonds in molecules and crystals. In engineering one needs the angles between beams and struts in complex structures. The dot product determines such angles.

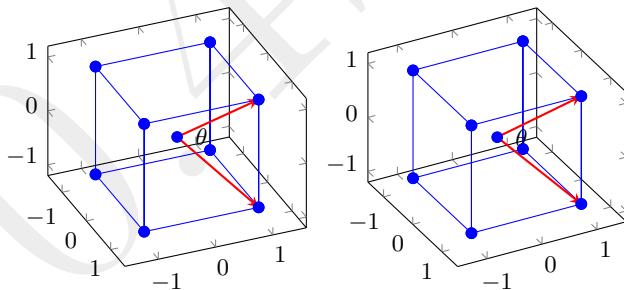
- (a) Consider the cube drawn in stereo below, and compute the angle between the diagonals on two adjacent faces.



- (b) Consider the cube drawn in stereo below: what is the angle between a diagonal on a face and a diagonal of the cube?



- (c) A body-centered cubic lattice (such as that formed by caesium chloride crystals) has one lattice point in the center of the unit cell as well as the eight corner points. Consider the body-centered cube of atoms drawn in stereo below with the center of the cube at the origin: what is the angle between the center atom and any two adjacent corner atoms?



- Example 1.3.9** (semantic similarity). Recall that [Example 1.1.7](#) introduced the encoding of sentences and documents as word count vectors. In the example, a word vector has five components,  $(N_{\text{cat}}, N_{\text{dog}}, N_{\text{mat}}, N_{\text{sat}}, N_{\text{scratched}})$  where the various  $N$  are the counts of each word in any sentence or document. For example,
- (a) “The dog sat on the mat” has word vector  $\mathbf{a} = (0, 1, 1, 1, 0)$ .
  - (b) “The cat scratched the dog” has word vector  $\mathbf{b} = (1, 1, 0, 0, 1)$ .
  - (c) “The cat and dog sat on the mat” has word vector  $\mathbf{c} = (1, 1, 1, 1, 0)$ .

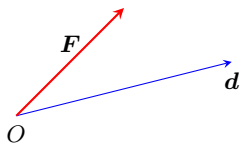
Use the angle between these three word vectors to characterise the similarity of the sentences: a small angle means the sentences are somehow close; a large angle means the sentences are disparate. ■

### 1.3.1 Work done involves the dot product

In physics and engineering, “work” has a precise meaning related to energy: when a force of magnitude  $F$  acts on a body and that body moves a distance  $d$ , then the work done by the force is  $W = Fd$ . This formula applies only for one dimensional force and displacement, the case when the force and the displacement are all in the same direction. For example, if a 5 kg barbell drops downwards 2 m under the force of gravity (9.8 newtons/kg), then the work done by gravity on the barbell during the drop is the product

$$W = F \times d = (5 \times 9.8) \times 2 = 98 \text{ joules.}$$

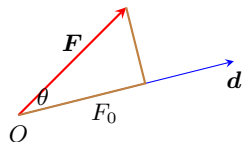
This work done goes to the kinetic energy of the falling barbell. The kinetic energy dissipates when the barbell hits the floor.



In general, the applied force and the displacement are not in the same direction (as illustrated in the margin). Consider the general case when a vector force  $\mathbf{F}$  acts on a body which moves a displacement vector  $\mathbf{d}$ . Then the work done by the force on the body is the length of the displacement times the component of the force in

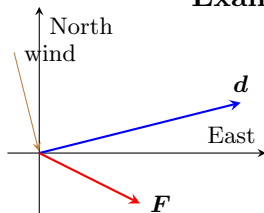
the direction of the displacement—the component of the force at right-angles to the displacement does no work.

As illustrated in the margin, draw a right-angled triangle to decompose the force  $\mathbf{F}$  into the component  $F_0$  in the direction of the displacement, and an unnamed component at right-angles. Then by the scalar formula, the work done is  $W = F_0|\mathbf{d}|$ . As drawn, the force  $\mathbf{F}$  makes an angle  $\theta$  to the displacement  $\mathbf{d}$ : the dot product determines this angle via  $\cos \theta = (\mathbf{F} \cdot \mathbf{d}) / (|\mathbf{F}||\mathbf{d}|)$  (Theorem 1.3.5). By basic trigonometry, the adjacent side of the force triangle has length  $F_0 = |\mathbf{F}| \cos \theta = |\mathbf{F}| \frac{\mathbf{F} \cdot \mathbf{d}}{|\mathbf{F}||\mathbf{d}|} = \frac{\mathbf{F} \cdot \mathbf{d}}{|\mathbf{d}|}$ . Finally, the work done  $W = F_0|\mathbf{d}| = \frac{\mathbf{F} \cdot \mathbf{d}}{|\mathbf{d}|}|\mathbf{d}| = \mathbf{F} \cdot \mathbf{d}$ , the dot product of the vector force and vector displacement.



### Example 1.3.10.

A sailing boat travels a distance of 40 m East and 10 m North, as drawn in the margin. The wind from abeam of strength and direction  $(1, -4)$  m/s generates a force  $\mathbf{F} = (20, -10)$  (newtons) on the sail, as drawn. What is the work done by the wind? ■





**Activity 1.3.11.** Recall the force of gravity on an object is the mass of the object times the acceleration of gravity,  $9.8 \text{ m/s}^2$ . A 3 kg ball is thrown horizontally from a height of 2 m and lands 10 m away on the ground: what is the total work done by gravity on the ball?

- (a) 58.8 joules                      (b) 98 joules  
(c) 19.6 joules                      (d) 29.4 joules



Finding components of vectors in various directions is called projection. Such projection is surprisingly common in applications and is developed much further by [Subsection 3.5.3](#).

### 1.3.2 Algebraic properties of the dot product

To manipulate the dot product in algebraic expressions, we need to know its basic algebraic rules. The following rules of [Theorem 1.3.13](#) are analogous to well known rules for scalar multiplication.

**Example 1.3.12.** Given vectors  $\mathbf{u} = (-2, 5, -2)$ ,  $\mathbf{v} = (3, 3, -2)$  and  $\mathbf{w} = (2, 0, -5)$ , and scalar  $a = 2$ , verify that (properties [1.3.13c](#) and [1.3.13d](#))

- $a(\mathbf{u} \cdot \mathbf{v}) = (a\mathbf{u}) \cdot \mathbf{v} = \mathbf{u} \cdot (a\mathbf{v})$  (a form of associativity);
- $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$  (distributivity).



**Theorem 1.3.13** (dot properties). *For all vectors  $\mathbf{u}$ ,  $\mathbf{v}$  and  $\mathbf{w}$  in  $\mathbb{R}^n$ , and for all scalars  $a$ , the following properties hold:*

- (a)  $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$  (commutative law);
- (b)  $\mathbf{u} \cdot \mathbf{0} = \mathbf{0} \cdot \mathbf{u} = 0$ ;

$$(c) \ a(\mathbf{u} \cdot \mathbf{v}) = (a\mathbf{u}) \cdot \mathbf{v} = \mathbf{u} \cdot (a\mathbf{v});$$

$$(d) \ (\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w} \quad (\text{distributive law});$$

$$(e) \ \mathbf{u} \cdot \mathbf{u} \geq 0, \text{ and moreover, } \mathbf{u} \cdot \mathbf{u} = 0 \text{ if and only if } \mathbf{u} = \mathbf{0}.$$

**Activity 1.3.14.** For vectors  $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ , which of the following statements is not generally true?

$$(a) \ \mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$$

$$(b) \ (2\mathbf{u}) \cdot (2\mathbf{v}) = 2(\mathbf{u} \cdot \mathbf{v})$$

$$(c) \ (\mathbf{u} - \mathbf{v}) \cdot (\mathbf{u} + \mathbf{v}) = \mathbf{u} \cdot \mathbf{u} - \mathbf{v} \cdot \mathbf{v}$$

$$(d) \ \mathbf{u} \cdot \mathbf{v} - \mathbf{v} \cdot \mathbf{u} = 0$$



The above proof of [Theorem 1.3.13e](#), that  $\mathbf{u} \cdot \mathbf{u} = 0$  if and only if  $\mathbf{u} = \mathbf{0}$ , may look uncannily familiar. The reason is that this last part is essentially the same as the proof of [Theorem 1.1.13](#) that the

zero vector is the only vector of length zero. The upcoming [Theorem 1.3.17](#) establishes that this connection between dot products and lengths is no coincidence.

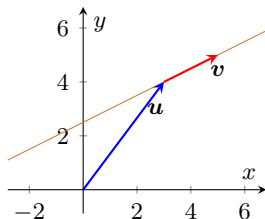
**Example 1.3.15.** For the two vectors  $\mathbf{u} = (3, 4)$  and  $\mathbf{v} = (2, 1)$  verify the following three properties:

- (a)  $\sqrt{\mathbf{u} \cdot \mathbf{u}} = |\mathbf{u}|$ , the length of  $\mathbf{u}$ ;
- (b)  $|\mathbf{u} \cdot \mathbf{v}| \leq |\mathbf{u}||\mathbf{v}|$  (Cauchy–Schwarz inequality);
- (c)  $|\mathbf{u} + \mathbf{v}| \leq |\mathbf{u}| + |\mathbf{v}|$  (triangle inequality).



The Cauchy–Schwarz inequality is one point of distinction between this ‘vector multiplication’ and scalar multiplication: for scalars  $|ab| = |a||b|$ , but the dot product of vectors is typically less,  $|\mathbf{u} \cdot \mathbf{v}| \leq |\mathbf{u}||\mathbf{v}|$ .

**Example 1.3.16.** The general proof of the Cauchy–Schwarz inequality involves a trick, so let’s introduce the trick using the vectors of [Example 1.3.15](#). Let vectors  $\mathbf{u} = (3, 4)$  and  $\mathbf{v} = (2, 1)$  and consider the line given parametrically ([Definition 1.2.15](#)) as the position vectors  $\mathbf{x} = \mathbf{u} + t\mathbf{v} = (3 + 2t, 4 + t)$  for scalar parameter  $t$ —illustrated in the margin. The position vector  $\mathbf{x}$  of any point on the line has length  $\ell$  ([Definition 1.1.9](#)) where



$$\begin{aligned}\ell^2 &= (3 + 2t)^2 + (4 + t)^2 \\ &= 9 + 12t + 4t^2 + 16 + 8t + t^2 \\ &= \underbrace{25}_c + \underbrace{20}_b t + \underbrace{5}_a t^2,\end{aligned}$$

a quadratic polynomial in  $t$ . We know that the length  $\ell > 0$  (the line does not pass through the origin so no  $\mathbf{x}$  is zero). Hence the quadratic in  $t$  cannot have any zeros. By the known properties of quadratic equations it follows that the discriminant  $b^2 - 4ac < 0$ . Indeed it is: here  $b^2 - 4ac = 20^2 - 4 \cdot 5 \cdot 25 = 400 - 500 = -100 < 0$ . Usefully, here  $a = 5 = |\mathbf{v}|^2$ ,  $c = 25 = |\mathbf{u}|^2$  and  $b = 20 = 2 \cdot 10 = 2(\mathbf{u} \cdot \mathbf{v})$ . So  $b^2 - 4ac < 0$ , written as  $\frac{1}{4}b^2 < ac$ , becomes the statement that  $\frac{1}{4}[2(\mathbf{u} \cdot \mathbf{v})]^2 = (\mathbf{u} \cdot \mathbf{v})^2 < |\mathbf{v}|^2|\mathbf{u}|^2$ . Taking the square-root of both

sides verifies the Cauchy–Schwarz inequality. The proof of the next theorem establishes it in general. ■

**Theorem 1.3.17.** *For all vectors  $\mathbf{u}$  and  $\mathbf{v}$  in  $\mathbb{R}^n$  the following properties hold:*

- (a)  $\sqrt{\mathbf{u} \cdot \mathbf{u}} = |\mathbf{u}|$ , the length of  $\mathbf{u}$ ;
- (b)  $|\mathbf{u} \cdot \mathbf{v}| \leq |\mathbf{u}||\mathbf{v}|$  (*Cauchy–Schwarz inequality*);
- (c)  $|\mathbf{u} \pm \mathbf{v}| \leq |\mathbf{u}| + |\mathbf{v}|$  (*triangle inequality*).

**Example 1.3.18.** Verify the Cauchy–Schwarz inequality (+ case) and the triangle inequality for the vectors  $\mathbf{a} = (-1, -2, 1, 3, -2)$  and  $\mathbf{b} = (-3, -2, 10, 2, 2)$ . ■

### 1.3.3 Orthogonal vectors are at right-angles

Of all the angles that vectors can make with each other, the two most important angles are, firstly, when the vectors are aligned with each other, and secondly, when the vectors are at right-angles to each other. Recall [Theorem 1.3.5](#) gives the angle  $\theta$  between two vectors via  $\cos \theta = \frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{u}| |\mathbf{v}|}$ . For vectors at right-angles  $\theta = 90^\circ$  and so  $\cos \theta = 0$  and hence non-zero vectors are at right-angles only when the dot product  $\mathbf{u} \cdot \mathbf{v} = 0$ . We give a special name to vectors at right-angles.

**Definition 1.3.19.** *Two vectors  $\mathbf{u}$  and  $\mathbf{v}$  in  $\mathbb{R}^n$  are termed **orthogonal** (or **perpendicular**) if and only if their dot product  $\mathbf{u} \cdot \mathbf{v} = 0$ .*

*The term ‘orthogonal’ derives from the Greek for ‘right-angled’.*

By convention the zero vector  $\mathbf{0}$  is orthogonal to all other vectors. However, in practice, we almost always use the notion of orthogonality only in connection with *non-zero* vectors. Often the requirement that the orthogonal vectors are non-zero is explicitly made, but beware that sometimes the requirement may be implicit in the problem.

**Example 1.3.20.** The standard unit vectors ([Definition 1.2.7](#)) are orthogonal to each other. For example, consider the standard unit vectors  $\mathbf{i}$ ,  $\mathbf{j}$  and  $\mathbf{k}$  in  $\mathbb{R}^3$ :

- $\mathbf{i} \cdot \mathbf{j} = (1, 0, 0) \cdot (0, 1, 0) = 0 + 0 + 0 = 0$ ;
- $\mathbf{j} \cdot \mathbf{k} = (0, 1, 0) \cdot (0, 0, 1) = 0 + 0 + 0 = 0$ ;
- $\mathbf{k} \cdot \mathbf{i} = (0, 0, 1) \cdot (1, 0, 0) = 0 + 0 + 0 = 0$ .

By [Definition 1.3.19](#) these are orthogonal to each other. ■

**Example 1.3.21.** Which pairs of the following vectors, if any, are perpendicular to each other?  $\mathbf{u} = (-1, 1, -3, 0)$ ,  $\mathbf{v} = (2, 4, 2, -6)$  and  $\mathbf{w} = (-1, 6, -2, 3)$ . ■



**Activity 1.3.22.** Which pair of the following three vectors are orthogonal to each other?  $\mathbf{x} = \mathbf{i} - 2\mathbf{k}$ ,  $\mathbf{y} = -3\mathbf{i} - 4\mathbf{j}$ ,  $\mathbf{z} = -\mathbf{i} - 2\mathbf{j} + 2\mathbf{k}$

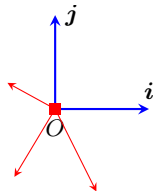
- (a) no pair      (b)  $\mathbf{x}, \mathbf{z}$       (c)  $\mathbf{y}, \mathbf{z}$       (d)  $\mathbf{x}, \mathbf{y}$



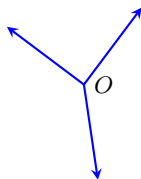
**Example 1.3.23.** Find the number  $b$  such that vectors  $\mathbf{a} = \mathbf{i} + 4\mathbf{j} + 2\mathbf{k}$  and  $\mathbf{b} = \mathbf{i} + b\mathbf{j} - 3\mathbf{k}$  are at right-angles. ■

**Key properties** The next couple of innocuous looking theorems are vital keys to important results in subsequent chapters.

To introduce the first theorem, consider the 2D plane and try to draw a non-zero vector at right-angles to both the two standard unit vectors  $\mathbf{i}$  and  $\mathbf{j}$ . The red vectors in the margin illustrate three failed attempts to draw a vector at right-angles to both  $\mathbf{i}$  and  $\mathbf{j}$ . It cannot be done. No vector in the plane can be at right angles to both the standard unit vectors in the plane.



**Theorem 1.3.24.** *There is no non-zero vector orthogonal to all  $n$  standard unit vectors in  $\mathbb{R}^n$ .*



To introduce the second theorem, imagine trying to draw three unit vectors in any orientation in the 2D plane such that all three are at right-angles to each other. The margin illustrates one attempt. It cannot be done. There are at most two vectors in 2D that are all at right-angles to each other.

**Theorem 1.3.25** (orthogonal completeness). *In a set of orthogonal unit vectors in  $\mathbb{R}^n$ , there can be no more than  $n$  vectors in the set.*

### 1.3.4 Normal vectors and equations of a plane

This section uses the dot product to find equations of a plane in 3D. The key is to write points in the plane as all those at right-angles to a certain direction. This direction is perpendicular to the required plane, and is called a normal. Let's start with an example of the idea in 2D.

**Example 1.3.26.** First find the equation of the line that is perpendicular to the vector  $(2, 3)$  and that passes through the origin. Second, find the equation of the line that passes through the point  $(4, 1)$  (instead of the origin). ■

**Activity 1.3.27.** What is an equation of the line through the point  $(4, 2)$  and that is a right-angles to the vector  $(1, 3)$ ?

(a)  $4x + y = 11$

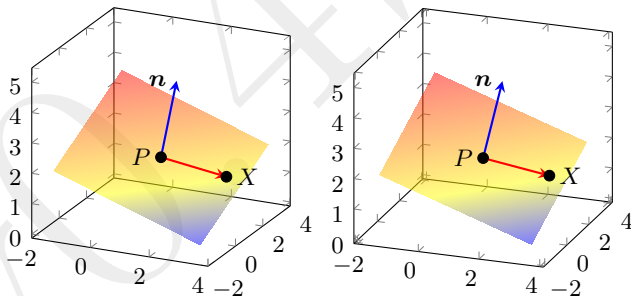
(b)  $4x + 2y = 10$

(c)  $2x + 3y = 11$

(d)  $x + 3y = 10$



Now use the same approach to finding an equation of a plane in 3D. The problem is to find the equation of the plane that goes through a given point  $P$  and is perpendicular to a given vector  $\mathbf{n}$ , called a **normal vector**. As illustrated in stereo below, that means to find all points  $X$  such that  $\overrightarrow{PX}$  is orthogonal to  $\mathbf{n}$ .



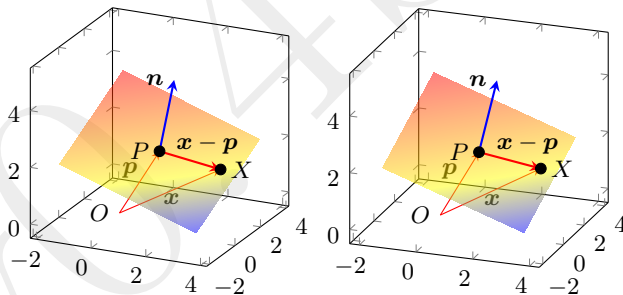
Denote the position vector of  $P$  by  $\mathbf{p} = (x_0, y_0, z_0)$ , the position vector of  $X$  by  $\mathbf{x} = (x, y, z)$ , and let the normal vector be  $\mathbf{n} = (a, b, c)$ . Then, as drawn below, the displacement vector  $\overrightarrow{PX} = \mathbf{x} - \mathbf{p} = (x - x_0, y - y_0, z - z_0)$  and so for  $\overrightarrow{PX}$  to be orthogonal

to  $\mathbf{n}$  requires  $\mathbf{n} \cdot (\mathbf{x} - \mathbf{p}) = 0$ ; that is, an **equation of the plane** is

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0,$$

equivalently, an equation of the plane is

$$ax + by + cz = d \quad \text{for constant } d = ax_0 + by_0 + cz_0.$$



**Example 1.3.28.** Find an equation of the plane through point  $P = (1, 1, 2)$  that has normal vector  $\mathbf{n} = (1, -1, 3)$ . (This is the case in the above illustrations.) Hence write down three distinct points on the plane. ■

**Example 1.3.29.** Write down a normal vector to each of the following planes:

(a)  $3x - 6y + z = 4$ ;

(b)  $z = 0.2x - 3.3y - 1.9$ .

**Activity 1.3.30.** Which of the following is a normal vector to the plane  $x_2 + 2x_3 + 4 = x_1$ ?

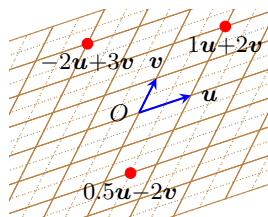
(a)  $(-1, 1, 2)$

(b) none of these

(c)  $(1, 2, 1)$

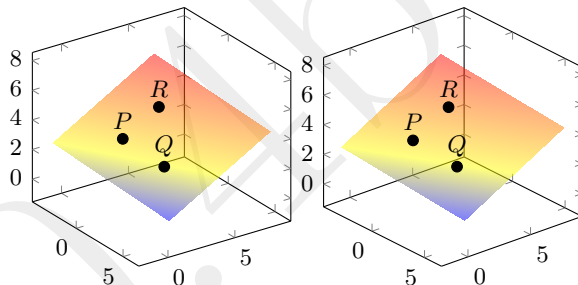
(d)  $(1, 2, 4)$

**Parametric equation of a plane** An alternative way of describing a plane is via a parametric equation analogous to the parametric equation of a line ([Subsection 1.2.2](#)). Such a parametric representation generalises to every dimension ([Section 2.3](#)).



The basic idea, as illustrated in the margin, is that given any plane (through the origin for the moment), then choosing almost any two vectors in the plane allows us to write all points in the plane as a sum of multiples of the two vectors. With the given vectors  $\mathbf{u}$  and  $\mathbf{v}$  shown in the margin, illustrated are the points  $\mathbf{u} + 2\mathbf{v}$ ,  $\frac{1}{2}\mathbf{u} - 2\mathbf{v}$  and  $-2\mathbf{u} + 3\mathbf{v}$ . Similarly, all points in the plane have a position vector in the form  $s\mathbf{u} + t\mathbf{v}$  for some scalar parameters  $s$  and  $t$ . The grid shown in the margin illustrates the sum of integral and half-integral multiples. The formula  $\mathbf{x} = s\mathbf{u} + t\mathbf{v}$  for parameters  $s$  and  $t$  is called a parametric equation of the plane.

**Example 1.3.31.** Find a parametric equation of the plane that passes through the three points  $P = (-1, 2, 3)$ ,  $Q = (2, 3, 2)$  and  $R = (0, 4, 5)$ , drawn below in stereo.



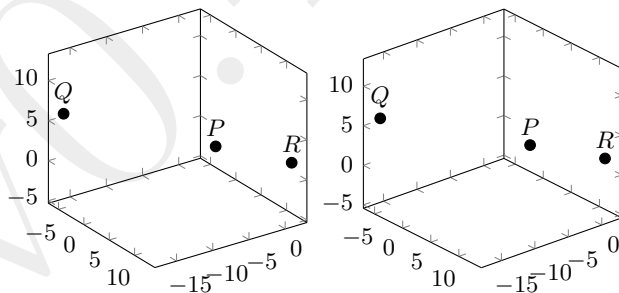
**Definition 1.3.32.** A *parametric equation* of a plane is  $\mathbf{x} = \mathbf{p} + s\mathbf{u} + t\mathbf{v}$  where  $\mathbf{p}$  is the position vector of some point in the plane, the two vectors  $\mathbf{u}$  and  $\mathbf{v}$  are parallel to the plane ( $\mathbf{u}, \mathbf{v} \neq \mathbf{0}$  and are at a non-zero/non- $\pi$  angle to each other), and the scalar **parameters**  $s$  and  $t$  vary over all real values to give position vectors of all points



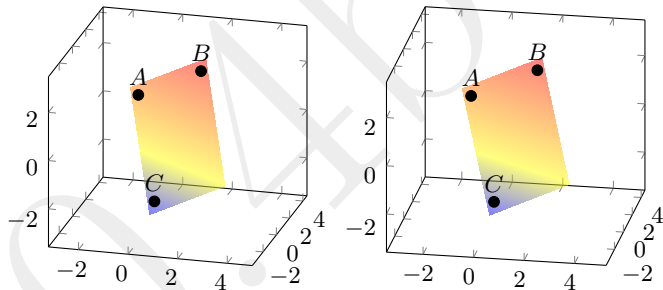
*in the plane.*

The beauty of this definition is that it applies for planes in any number of dimensions. To do so the parametric equations just uses two vectors with the corresponding number of components.

**Example 1.3.33.** Find a parametric equation of the plane that passes through the three points  $P = (6, -4, 3)$ ,  $Q = (-4, -18, 7)$  and  $R = (11, 3, 1)$ , drawn below in stereo.



**Example 1.3.34.** Find a parametric equation of the plane that passes through the three points  $A = (-1.2, 2.4, 0.8)$ ,  $B = (1.6, 1.4, 2.4)$  and  $C = (0.2, -0.4, -2.5)$ , drawn below in stereo.



**Activity 1.3.35.** Which of the following is *not* a parametric equation of a plane?

- (a)  $(4, 1, 4) + (3, 6, 3)s + (2, 4, 2)t$
- (b)  $(3s + 2t, 4 + 2s + t, 4 + 3t)$

(c)  $(-1, 1, -1)s + (4, 2, -1)t$

(d)  $\mathbf{i} + s\mathbf{j} + t\mathbf{k}$



## 1.4 The cross product

### Section Contents

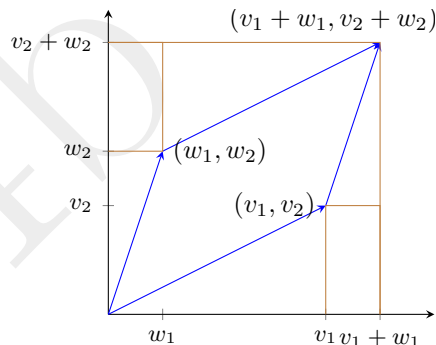
Area of a parallelogram . . . . .	76
Normal vector to a plane . . . . .	78
Definition of a cross product . . . . .	80
Geometry of a cross product . . . . .	83
Algebraic properties of a cross product . . . .	89
Volume of a parallelepiped . . . . .	92

This section is optional for us, but is vital in many topics of science and engineering.

The dot product is not the only way to multiply vectors. In the three dimensions of the world we live in there is another way to multiply vectors, called the cross product. But for more than three dimensions, qualitatively different techniques are developed in subsequent chapters.

## Area of a parallelogram

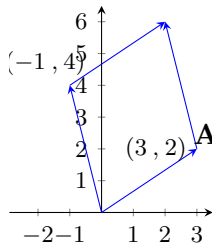
Consider the parallelogram drawn in blue. It has sides given by vectors  $\mathbf{v} = (v_1, v_2)$  and  $\mathbf{w} = (w_1, w_2)$  as shown. Let's determine the area of the parallelogram. Its area is the containing rectangle less the two small rectangles and the four small triangles. The two small rectangles have the same area, namely  $w_1v_2$ . The two small triangles on the left and the right also have the same area, namely  $\frac{1}{2}w_1w_2$ . The two small triangles on the top and the bottom similarly have the same area, namely  $\frac{1}{2}v_1v_2$ . Thus, the parallelogram has



$$\begin{aligned}
 \text{area} &= (v_1 + w_1)(v_2 + w_2) - 2w_1v_2 - 2 \cdot \frac{1}{2}w_1w_2 - 2 \cdot \frac{1}{2}v_1v_2 \\
 &= v_1v_2 + v_1w_2 + w_1v_2 + w_1w_2 - 2w_1v_2 - w_1w_2 - v_1v_2 \\
 &= v_1w_2 - v_2w_1.
 \end{aligned}$$

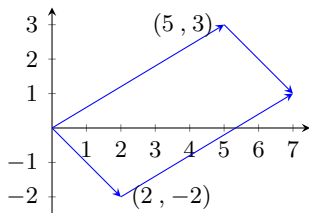
In application, sometimes this right-hand side expression is negative because vectors  $\mathbf{v}$  and  $\mathbf{w}$  are the ‘wrong way’ around. Thus in general the parallelogram area =  $|v_1w_2 - v_2w_1|$ .

**Example 1.4.1.** What is the area of the parallelogram (illustrated in the margin) whose edges are formed by the vectors  $(3, 2)$  and  $(-1, 4)$ ? ■



**Activity 1.4.2.** What is the area of the parallelogram (illustrated in the margin) whose edges are formed by the vectors  $(5, 3)$  and  $(2, -2)$ ? ■

- (a) 4                      (b) 16                      (c) 19                      (d) 11



Interestingly, we meet this expression for area,  $v_1w_2 - v_2w_1$ , in another context: that of equations for a plane and its normal vector.

## Normal vector to a plane

Recall [Subsection 1.3.4](#) introduced that we describe planes either via an equation such as  $x - y + 3z = 6$  or via a parametric description such as  $\mathbf{x} = (1, 1, 2) + (1, 1, 0)s + (0, 3, 1)t$ . These determine the same plane, just different algebraic descriptions. One converts between these two descriptions using the cross product.

**Example 1.4.3.** Derive that the plane described parametrically by  $\mathbf{x} = (1, 1, 2) + (1, 1, 0)s + (0, 3, 1)t$  has normal equation  $x - y + 3z = 6$ . ■

**Activity 1.4.4.** Use the procedure of [Example 1.4.3](#) to derive a normal vector to the plane described in parametric form as  $\mathbf{x} = (4, -1, -2) + (1, -2, 1)s + (2, -3, -2)t$ . Which of the following is your computed normal vector?

(a)  $(2, -2, 5)$

(b)  $(-4, 4, -10)$

(c)  $(5, 6, 7)$

(d)  $(7, 4, 1)$



VO.410



## Definition of a cross product

**General formula** The procedure used in [Example 1.4.3](#) to derive a normal vector leads to an algebraic formula. Let's apply the same procedure to two general vectors  $\mathbf{v} = (v_1, v_2, v_3)$  and  $\mathbf{w} = (w_1, w_2, w_3)$ . The procedure computes

$$\mathbf{n} = \begin{vmatrix} \mathbf{i} & v_1 & w_1 \\ \mathbf{j} & v_2 & w_2 \\ \mathbf{k} & v_3 & w_3 \end{vmatrix}$$

(cross out 1st column and each row, multiplying each by common entry, with alternating sign)

$$\begin{aligned} &= i \begin{vmatrix} \cancel{\mathbf{i}} & \cancel{v_1} & \cancel{w_1} \\ \mathbf{j} & v_2 & w_2 \\ \mathbf{k} & v_3 & w_3 \end{vmatrix} - j \begin{vmatrix} \cancel{\mathbf{i}} & v_1 & w_1 \\ \cancel{\mathbf{j}} & \cancel{v_2} & \cancel{w_2} \\ \mathbf{k} & v_3 & w_3 \end{vmatrix} + k \begin{vmatrix} \cancel{\mathbf{i}} & v_1 & w_1 \\ \mathbf{j} & v_2 & w_2 \\ \cancel{\mathbf{k}} & \cancel{v_3} & \cancel{w_3} \end{vmatrix} \\ &= i \begin{vmatrix} v_2 & w_2 \\ v_3 & w_3 \end{vmatrix} - j \begin{vmatrix} v_1 & w_1 \\ v_3 & w_3 \end{vmatrix} + k \begin{vmatrix} v_1 & w_1 \\ v_2 & w_2 \end{vmatrix} \end{aligned}$$

(draw diagonals, then subtract product of red diagonal from product of the blue)

$$\begin{aligned}
&= i \begin{vmatrix} v_2 & w_2 \\ v_3 & w_3 \end{vmatrix} - j \begin{vmatrix} v_1 & w_1 \\ v_3 & w_3 \end{vmatrix} + k \begin{vmatrix} v_1 & w_1 \\ v_2 & w_2 \end{vmatrix} \\
&= i(v_2w_3 - v_3w_2) - j(v_1w_3 - v_3w_1) + k(v_1w_2 - v_2w_1).
\end{aligned}$$

We use this formula to define the cross product algebraically, and then see what it means geometrically.

**Definition 1.4.5.** Let  $\mathbf{v} = (v_1, v_2, v_3)$  and  $\mathbf{w} = (w_1, w_2, w_3)$  be two vectors in  $\mathbb{R}^3$ . The **cross product** (or **vector product**)  $\mathbf{v} \times \mathbf{w}$  is defined algebraically as

$$\mathbf{v} \times \mathbf{w} := i(v_2w_3 - v_3w_2) + j(v_3w_1 - v_1w_3) + k(v_1w_2 - v_2w_1).$$

**Example 1.4.6.** Among the standard unit vectors, derive that

- |   |  |
|---|--|
| (a) $\mathbf{i} \times \mathbf{j} = \mathbf{k}$ ,   | (b) $\mathbf{j} \times \mathbf{i} = -\mathbf{k}$ , |
| (c) $\mathbf{j} \times \mathbf{k} = \mathbf{i}$ ,   | (d) $\mathbf{k} \times \mathbf{j} = -\mathbf{i}$ , |
| (e) $\mathbf{k} \times \mathbf{i} = \mathbf{j}$ ,   | (f) $\mathbf{i} \times \mathbf{k} = -\mathbf{j}$ , |
| (g) $\mathbf{i} \times \mathbf{i} = \mathbf{j} \times \mathbf{j} = \mathbf{k} \times \mathbf{k} = \mathbf{0}$ . |  |

The cross products of this [Example 1.4.6](#) most clearly demonstrate the orthogonality of a cross product to its two argument vectors ([Theorem 1.4.10a](#)), and that the direction is in the so-called right-hand sense ([Theorem 1.4.10b](#)). ■

**Activity 1.4.7.** Use [Definition 1.4.5](#) to find the cross product of  $(-4, 1, -1)$  and  $(-2, 2, 1)$  is which one of the following:

(a)  $(-3, -6, 6)$

(b)  $(3, 6, -6)$

(c)  $(3, -6, -6)$

(d)  $(-3, -6, 6)$



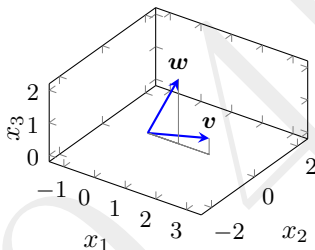
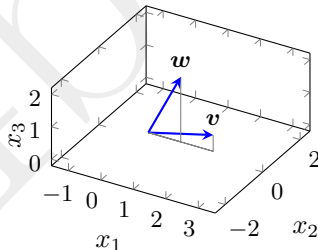
## Geometry of a cross product

**Example 1.4.8** (parallelogram area). Let's revisit the introduction to this section. Consider the parallelogram in the  $x_1x_2$ -plane with edges formed by the  $\mathbb{R}^3$  vectors  $\mathbf{v} = (v_1, v_2, 0)$  and  $\mathbf{w} = (w_1, w_2, 0)$ . At the start of this [Section 1.4](#) we derived that the parallelogram formed by these vectors has area  $= |v_1w_2 - v_2w_1|$ . Compare this area with the cross product

$$\begin{aligned}\mathbf{v} \times \mathbf{w} &= \mathbf{i}(v_2 \cdot 0 - 0 \cdot w_2) + \mathbf{j}(0 \cdot w_1 - v_1 \cdot 0) + \mathbf{k}(v_1w_2 - v_2w_1) \\ &= \mathbf{i}0 + \mathbf{j}0 + \mathbf{k}(v_1w_2 - v_2w_1) \\ &= \mathbf{k}(v_1w_2 - v_2w_1).\end{aligned}$$

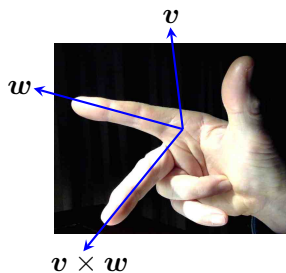
Consequently, the length of this cross product equals the area of the parallelogram formed by  $\mathbf{v}$  and  $\mathbf{w}$  ([Theorem 1.4.10d](#)). (Also the direction of the cross product,  $\pm\mathbf{k}$ , is orthogonal to the  $x_1x_2$ -plane containing the two vectors—[Theorem 1.4.10a](#)). ■

**Activity 1.4.9.** Using property 1.4.10b of the next theorem, in which direction is the cross product  $\mathbf{v} \times \mathbf{w}$  for the two vectors illustrated in stereo below?

(a)  $+\mathbf{j}$ (b)  $-\mathbf{i}$ (c)  $+\mathbf{i}$ (d)  $-\mathbf{j}$ 

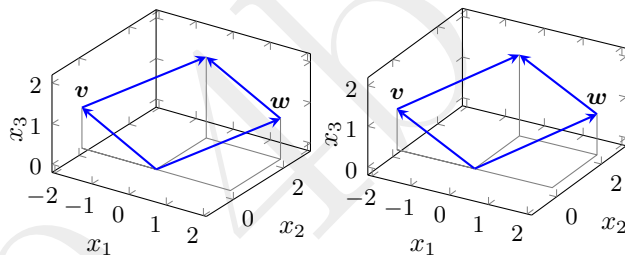
**Theorem 1.4.10** (cross product geometry). *Let  $\mathbf{v}$  and  $\mathbf{w}$  be two vectors in  $\mathbb{R}^3$ :*

(a) *the vector  $\mathbf{v} \times \mathbf{w}$  is orthogonal to both  $\mathbf{v}$  and  $\mathbf{w}$ ;*

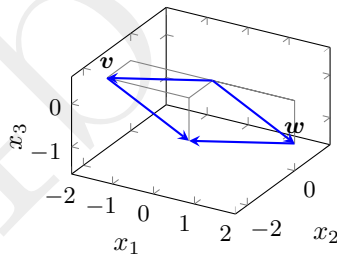
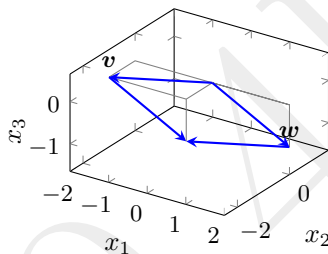


- (b) the direction of  $\mathbf{v} \times \mathbf{w}$  is in the **right-hand sense** in that if  $\mathbf{v}$  is in the direction of your thumb, and  $\mathbf{w}$  is in the direction of your straight index finger, then  $\mathbf{v} \times \mathbf{w}$  is in the direction of your bent second/longest finger—all on your right-hand as illustrated in the margin;
- (c)  $|\mathbf{v} \times \mathbf{w}| = |\mathbf{v}| |\mathbf{w}| \sin \theta$  where  $\theta$  is the angle between vectors  $\mathbf{v}$  and  $\mathbf{w}$  ( $0 \leq \theta \leq \pi$ , equivalently  $0^\circ \leq \theta \leq 180^\circ$ ); and
- (d) the length  $|\mathbf{v} \times \mathbf{w}|$  is the area of the parallelogram with edges  $\mathbf{v}$  and  $\mathbf{w}$ .

**Example 1.4.11.** Find the area of the parallelogram with edges formed by vectors  $\mathbf{v} = (-2, 0, 1)$  and  $\mathbf{w} = (2, 2, 1)$ —as in stereo below.



**Activity 1.4.12.** What is the area of the parallelogram (in stereo below) with edges formed by vectors  $\mathbf{v} = (-2, 1, 0)$  and  $\mathbf{w} = (2, 0, -1)$ ?



(a) 5

(b)  $\sqrt{5}$ 

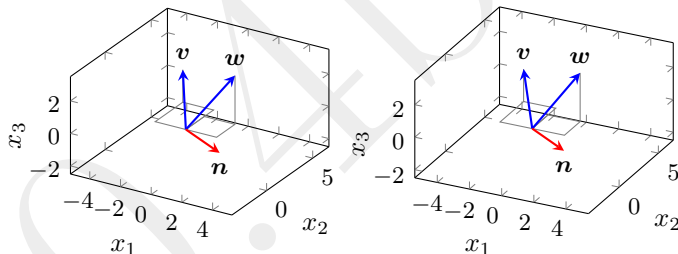
(c) 1

(d) 3





**Example 1.4.13.** Find a normal vector to the plane containing the two vectors  $\mathbf{v} = -2\mathbf{i} + 3\mathbf{j} + 2\mathbf{k}$  and  $\mathbf{w} = 2\mathbf{i} + 2\mathbf{j} + 3\mathbf{k}$  —illustrated below. Hence find an equation of the plane given parametrically as  $\mathbf{x} = -2\mathbf{i} - \mathbf{j} + 3\mathbf{k} + (-2\mathbf{i} + 3\mathbf{j} + 2\mathbf{k})s + (2\mathbf{i} + 2\mathbf{j} + 3\mathbf{k})t$ .



## Algebraic properties of a cross product

Exercises ??–?? establish three of the following four useful algebraic properties of the cross product.

**Theorem 1.4.14** (cross product properties). *Let  $\mathbf{u}$ ,  $\mathbf{v}$  and  $\mathbf{w}$  be vectors in  $\mathbb{R}^3$ , and  $c$  be a scalar:*

- (a)  $\mathbf{v} \times \mathbf{v} = \mathbf{0}$ ;
- (b)  $\mathbf{w} \times \mathbf{v} = -(\mathbf{v} \times \mathbf{w})$  (not commutative);
- (c)  $(c\mathbf{v}) \times \mathbf{w} = c(\mathbf{v} \times \mathbf{w}) = \mathbf{v} \times (c\mathbf{w})$ ;
- (d)  $\mathbf{u} \times (\mathbf{v} + \mathbf{w}) = \mathbf{u} \times \mathbf{v} + \mathbf{u} \times \mathbf{w}$  (distributive law).

**Example 1.4.15.** As an example of [Theorem 1.4.14b](#), [Example 1.4.6](#) shows that  $\mathbf{i} \times \mathbf{j} = \mathbf{k}$ , whereas reversing the order of the cross product gives the negative  $\mathbf{j} \times \mathbf{i} = -\mathbf{k}$ . Given [Example 1.4.13](#) derived  $\mathbf{v} \times \mathbf{w} = 5\mathbf{i} + 10\mathbf{j} - 10\mathbf{k}$  in the case when  $\mathbf{v} = -2\mathbf{i} + 3\mathbf{j} + 2\mathbf{k}$  and  $\mathbf{w} = 2\mathbf{i} + 2\mathbf{j} + 3\mathbf{k}$ , what is  $\mathbf{w} \times \mathbf{v}$ ? ■

**Example 1.4.16.** Given  $(\mathbf{i} + \mathbf{j} + \mathbf{k}) \times (-2\mathbf{i} - \mathbf{j}) = \mathbf{i} - 2\mathbf{j} + \mathbf{k}$ , what is  $(3\mathbf{i} + 3\mathbf{j} + 3\mathbf{k}) \times (-2\mathbf{i} - \mathbf{j})$ ? ■

**Activity 1.4.17.** For vectors  $\mathbf{u} = -\mathbf{i} + 3\mathbf{k}$ ,  $\mathbf{v} = \mathbf{i} + 3\mathbf{j} + 5\mathbf{k}$ , and  $\mathbf{w} = -2\mathbf{i} + \mathbf{j} - \mathbf{k}$  you are given that

$$\mathbf{u} \times \mathbf{v} = -9\mathbf{i} + 8\mathbf{j} - 3\mathbf{k},$$

$$\mathbf{u} \times \mathbf{w} = -3\mathbf{i} - 7\mathbf{j} - \mathbf{k},$$

$$\mathbf{v} \times \mathbf{w} = -8\mathbf{i} - 9\mathbf{j} + 7\mathbf{k}.$$

Which is the cross product  $(-\mathbf{i} + 3\mathbf{k}) \times (-\mathbf{i} + 4\mathbf{j} + 4\mathbf{k})$ ?

(a)  $-17\mathbf{i} - \mathbf{j} + 4\mathbf{k}$

(b)  $\mathbf{i} - 17\mathbf{j} + 10\mathbf{k}$

(c)  $-12\mathbf{i} + \mathbf{j} - 4\mathbf{k}$

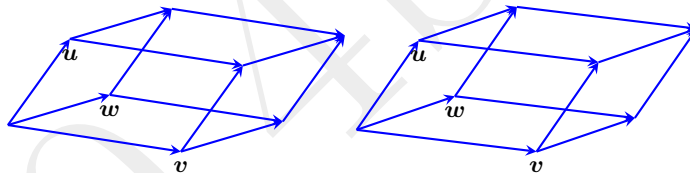
(d)  $-11\mathbf{i} - 16\mathbf{j} + 6\mathbf{k}$

Also, which is  $(\mathbf{i} + 3\mathbf{j} + 5\mathbf{k}) \times (-3\mathbf{i} + \mathbf{j} + 2\mathbf{k})$ ? ■

**Example 1.4.18.** The properties of [Theorem 1.4.14](#) empower algebraic manipulation. Use such algebraic manipulation, and the identities among standard unit vectors of [Example 1.4.6](#), compute the cross product  $(\mathbf{i} - \mathbf{j}) \times (4\mathbf{i} + 2\mathbf{k})$ . ■

## Volume of a parallelepiped

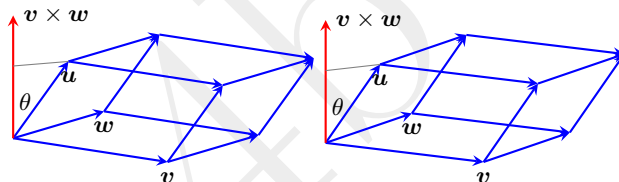
Consider the parallelepiped with edges formed by three vectors  $\mathbf{u}$ ,  $\mathbf{v}$  and  $\mathbf{w}$  in  $\mathbb{R}^3$ , as illustrated in stereo below. Our challenge is to derive that the volume of the parallelepiped is  $|\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})|$ .



Let's use that we know the volume of the parallelepiped is the area of its base times its height.

- The base of the parallelepiped is the parallelogram formed with edges  $\mathbf{v}$  and  $\mathbf{w}$ . Hence the base has area  $|\mathbf{v} \times \mathbf{w}|$  ([Theorem 1.4.10d](#)).
- The height of the parallelepiped is then that part of  $\mathbf{u}$  in the direction of a normal vector to  $\mathbf{v}$  and  $\mathbf{w}$ . We know that  $\mathbf{v} \times \mathbf{w}$  is orthogonal to both  $\mathbf{v}$  and  $\mathbf{w}$  ([Theorem 1.4.10a](#)),

so by trigonometry the height must be  $|\mathbf{u}| \cos \theta$  for angle  $\theta$  between  $\mathbf{u}$  and  $\mathbf{v} \times \mathbf{w}$ , as illustrated below.



To cater for cases where  $\mathbf{v} \times \mathbf{w}$  points in the opposite direction to that shown, the height is  $|\mathbf{u}| |\cos \theta|$ . The dot product determines this cosine ([Theorem 1.3.5](#)):

$$\cos \theta = \frac{\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})}{|\mathbf{u}| |\mathbf{v} \times \mathbf{w}|}.$$

The height of the parallelepiped is then

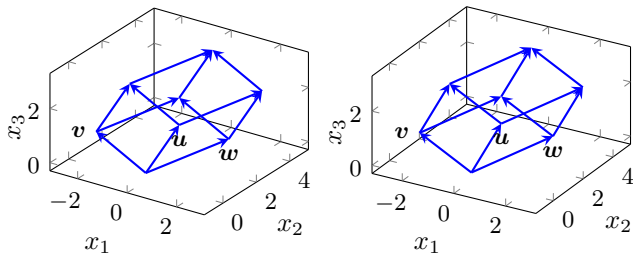
$$|\mathbf{u}| |\cos \theta| = |\mathbf{u}| \frac{|\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})|}{|\mathbf{u}| |\mathbf{v} \times \mathbf{w}|} = \frac{|\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})|}{|\mathbf{v} \times \mathbf{w}|}.$$

Consequently, the volume of the parallelepiped equals

$$\text{base} \cdot \text{height} = |\mathbf{v} \times \mathbf{w}| \frac{|\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})|}{|\mathbf{v} \times \mathbf{w}|} = |\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})|.$$

**Definition 1.4.19.** For every three vectors  $\mathbf{u}$ ,  $\mathbf{v}$  and  $\mathbf{w}$  in  $\mathbb{R}^3$ , the **scalar triple product** is  $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})$ .

**Example 1.4.20.** Use the scalar triple product to find the area of the parallelepiped formed by vectors  $\mathbf{u} = (0, 2, 1)$ ,  $\mathbf{v} = (-2, 0, 1)$  and  $\mathbf{w} = (2, 2, 1)$ —as illustrated in stereo below.





Using the procedure of [Example 1.4.3](#) to find a scalar triple product establishes a strong connection to the determinants of [Chapter 6](#). In the second solution to the previous [Example 1.4.20](#), in finding  $\mathbf{u} \times \mathbf{w}$ , the unit vectors  $\mathbf{i}$ ,  $\mathbf{j}$  and  $\mathbf{k}$  just acted as place holding symbols to eventually ensure a multiplication by the correct component of  $\mathbf{v}$  in the dot product. We could seamlessly combine the two products by replacing the symbols  $\mathbf{i}$ ,  $\mathbf{j}$  and  $\mathbf{k}$  directly with the corresponding component of  $\mathbf{v}$ :

$$\begin{aligned} \mathbf{v} \cdot (\mathbf{u} \times \mathbf{w}) &= \begin{vmatrix} -2 & 0 & 2 \\ 0 & 2 & 2 \\ 1 & 1 & 1 \end{vmatrix} \\ &= -2 \begin{vmatrix} -2 & 0 & 2 \\ 0 & 2 & 2 \\ 1 & 1 & 1 \end{vmatrix} - 0 \begin{vmatrix} -2 & 0 & 2 \\ 0 & 2 & 2 \\ 1 & 1 & 1 \end{vmatrix} + 1 \begin{vmatrix} -2 & 0 & 2 \\ 0 & 2 & 2 \\ 1 & 1 & 1 \end{vmatrix} \end{aligned}$$



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

Hence the parallelepiped formed by  $\mathbf{u}$ ,  $\mathbf{v}$  and  $\mathbf{w}$  has volume  $|-4|$ , as before. Here the volume follows from the above manipulations of the matrix of numbers formed with columns of the matrix being the vectors  $\mathbf{u}$ ,  $\mathbf{v}$  and  $\mathbf{w}$ . Chapter 6 shows that this computation of volume generalises to determining, via analogous matrices of vectors, the ‘volume’ of objects formed by vectors with any number of components.

## 1.5 Use Matlab/Octave for vector computation

### Section Contents

It is the science of *calculation*,—which becomes continually more necessary at each step of our progress, and which must ultimately govern the whole of the applications of science to the arts of life.

*Charles Babbage, 1832*

Subsequent chapters invoke the computer packages MATLAB/Octave to perform calculations that would be tedious and error prone when done by hand. This section introduces MATLAB/Octave so that you can start to become familiar with it on small problems. You should directly compare the computed answer with your calculation by hand. The aim is to develop some basic confidence with MATLAB/Octave before later using it to save considerable time in longer tasks.

- MATLAB is commercial software available from Mathworks.

It is also useable over the internet as MATLAB-Online or MATLAB-Mobile.

- Octave is free software, that for our purposes is almost identical to MATLAB, and downloadable over the internet. Octave is also freely useable over the internet.
- Alternatively, your home institution may provide MATLAB/Octave via a web service that is useable via smart phones, tablets and computers.

**Example 1.5.1.** Use the MATLAB/Octave command `norm()` to compute the length/magnitude of the following vectors ([Definition 1.1.9](#)).

(a)  $(2, -1)$

(b)  $(-1, 1, -5, 4)$

(c)  $(-0.3, 4.3, -2.5, -2.8, 7, -1.9)$



Table 1.2: Use MATLAB/Octave to help compute vector results with the following basics. This and subsequent tables throughout the book summarise MATLAB/Octave for our use.

- Real numbers are limited to being zero or of magnitude from  $10^{-323}$  to  $10^{+308}$ , both positive and negative (called the **floating point** numbers). Real numbers are computed and stored to a maximum precision of nearly sixteen significant digits.
- MATLAB/Octave potentially uses complex numbers ( $\mathbb{C}$ ), but mostly we stay within real numbers ( $\mathbb{R}$ ).
- Each MATLAB/Octave command is usually typed on one line by itself.
- `[ . ; . ; . ]` where each dot denotes a number, forms vectors in  $\mathbb{R}^3$  (or use newlines instead of the semi-colons). Use  $n$  numbers separated by semi-colons for vectors in  $\mathbb{R}^n$ .
- `=` assigns the result of the expression to the right of the `=` to the variable name on the left.

If the result of an expression is not explicitly assigned to a variable, then by default it is assigned to the variable `ans`.

- Variable names are alphanumeric starting with a letter.
- `size(v)` returns the number of components of the vector (Definition 1.1.4): if the vector  $\mathbf{v}$  is in  $\mathbb{R}^m$ , then `size(v)`

**Example 1.5.2.** Use MATLAB/Octave operators  $+$ ,  $-$ ,  $*$  to compute the value of the expressions  $\mathbf{u} + \mathbf{v}$ ,  $\mathbf{u} - \mathbf{v}$ ,  $3\mathbf{u}$  for vectors  $\mathbf{u} = (-4.1, 1.7, 4.1)$  and  $\mathbf{v} = (2.9, 0.9, -2.4)$  (Definition 1.2.4). ■

**Example 1.5.3.** Use MATLAB/Octave to confirm that  $2(2\mathbf{p} - 3\mathbf{q}) + 6(\mathbf{q} - \mathbf{p}) = -2\mathbf{p}$  for vectors  $\mathbf{p} = (1, 0, 2, -6)$  and  $\mathbf{q} = (2, 4, 3, 5)$ . ■

**Example 1.5.4.** Use MATLAB/Octave to confirm the commutative law (Theorem 1.2.19a)  $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$  for vectors  $\mathbf{u} = (8, -6, -4, -2)$  and  $\mathbf{v} = (4, 3, -1)$ . ■

**Activity 1.5.5.** You enter the two vectors into MATLAB by typing  $u=[1.1;3.7;-4.5]$  and  $v=[1.7;0.6;-2.6]$ .

- Which of the following is the result of typing the command  $u-v$ ?

(a) 2.8000  
Error using \*  
Inner matrix  
dimensions must  
agree.

(b) 4.3000  
-7.1000

(c) -0.6000  
3.1000  
-1.9000

(d) 2.2000  
7.4000  
-9.0000

- Which is the result of typing the command  $2*u$ ?
- Which is the result of typing the command  $u*v$ ?



**Example 1.5.6.** Use MATLAB/Octave to compute the angles between the pair of vectors  $(4, 3)$  and  $(5, 12)$  ([Theorem 1.3.5](#)). ■

**Example 1.5.7.** Verify the distributive law for the dot product  $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$  ([Theorem 1.3.13d](#)) for vectors  $\mathbf{u} = (-0.1, -3.1, -2.9, -1.3)$ ,  $\mathbf{v} = (-3, 0.5, 6.4, -0.9)$  and  $\mathbf{w} = (-1.5, -0.2, 0.4, -3.1)$ . ■

**Activity 1.5.8.** Given two vectors  $\mathbf{u}$  and  $\mathbf{v}$  that have already been typed into MATLAB/Octave, which of the following expressions could check the identity that  $(\mathbf{u} - 2\mathbf{v}) \cdot (\mathbf{u} + \mathbf{v}) = \mathbf{u} \cdot \mathbf{u} - \mathbf{u} \cdot \mathbf{v} - 2\mathbf{v} \cdot \mathbf{v}$ ?

- (a) `dot(u-2*v,u+v)-dot(u,u)+dot(u,v)+2*dot(v,v)`
- (b) None of the others
- (c) `dot(u-2*v,u+v)-dot(u,u)+dot(u,v)+2*dot(v,v)`
- (d) `(u-2*v)*(u+v)-u*u+u*v+2*v*v`



Many other books (Quarteroni & Saleri 2006, §§1.1–3, e.g.) give more details about the basics than the essentials that are introduced here.

On two occasions I have been asked [by members of Parliament!], “Pray, Mr. Babbage, if you put into the machine wrong figures, will the right answers come out?”

I am not able rightly to apprehend the kind of confusion of ideas that could provoke such a question.

*Charles Babbage*



---

## 2 Systems of linear equations

---

### Chapter Contents

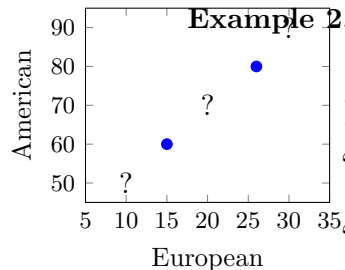
2.1	Introduction to systems of linear equations . . . . .	110
2.2	Directly solve linear systems . . . . .	120
2.2.1	Compute a system's solution . . . . .	121
2.2.2	Algebraic manipulation solves systems . . . . .	134
2.2.3	Three possible numbers of solutions . . . . .	144
2.3	Linear combinations span sets . . . . .	151

Linear relationships are commonly identified in science and engineering, and are commonly expressed as linear equations. One

of the reasons is that scientists and engineers can do amazingly powerful algebraic transformations with linear equations. Such transformations and their practical implications are the subject of this book.

One vital use in science and engineering is in the scientific task of taking scattered experimental data and inferring a general algebraic relation between the quantities measured. In computing science this task is often called ‘data mining’, ‘knowledge discovery’ or even ‘artificial intelligence’—although the algebraic relation is then typically discussed as a computational procedure. But appearing within these tasks is always linear equations to be solved.

I am sure you can guess where we are going with this example, but let’s pretend we do not know.



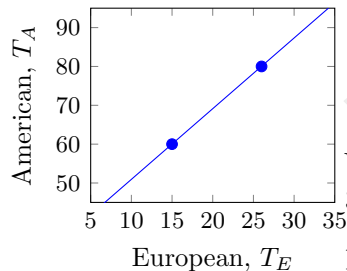
**Example 2.0.1 (scientific inference).** Two colleagues, an American and a European, discuss the weather; in particular, they discuss the temperature. The American says “yesterday the temperature was  $80^{\circ}$  but today is much cooler at  $60^{\circ}$ ”. The European says, “that’s not what I heard, I heard the temperature was  $26^{\circ}$  and today is  $15^{\circ}$ ”. (The marginal figure plots these two data points.) “Hmmm, we must be using a different temperature scale”, they say. Being scientists they start to use linear algebra to *infer*,

from the two days of temperature data, a general relation between their temperature scales—a relationship valid over a wide range of temperatures (denoted by the question marks in the marginal figure). Let's assume that, in terms of the European temperature  $T_E$ , the American temperature  $T_A = cT_E + d$  for some constants  $c$  and  $d$  they and we aim to find. The two days of data then give that

$$80 = c26 + d \quad \text{and} \quad 60 = c15 + d.$$

To find the constants  $c$  and  $d$ :

- subtract the second equation from the first to deduce  $80 - 60 = 26c + d - 15c - d$  which simplifies to  $20 = 11c$ , that is,  $c = 20/11 = 1.82$  to two decimal places (2 d.p.);
- use this value of  $c$  in either equation, say the second, gives  $60 = \frac{20}{11}15 + d$  which rearranges to  $d = 360/11 = 32.73$  to two decimal places (2 d.p.).



We deduce that the temperature relationship is  $T_A = 1.82T_E + 32.73$  (as plotted in the marginal figure). The two colleagues now *predict* that they will be able to use this formula to translate their temperature into that of the other, and vice versa.

You may quite rightly object that the two colleagues *assumed* a linear relation, they do *not know* it is linear. You may also object that the predicted relation is erroneous as it should be  $T_A = \frac{9}{5}T_E + 32$  (the relation between Celsius and Fahrenheit). Absolutely, you should object. Scientifically, the deduced relation  $T_A = 1.82T_E + 32.73$  is only a conjecture that fits the known data. More data and more linear algebra together empower us to both confirm the linearity (or not as the case may be), and also to improve the accuracy of the coefficients. Such progressive refinement is fundamental scientific methodology—and central to it is the algebra of linear equations. ■

Linear algebra and equations are also crucial for nonlinear relationships. [Figure 2.1](#) shows four plots of the same nonlinear curve, but on successively smaller scales. Zooming in on the point  $(0, 1)$  we see the curve looks straighter and straighter until on the microscale (bottom-right) it is effectively a straight line. The same is true for everywhere on every smooth curve: we discover that every smooth curve looks like a straight line on the microscale. Thus we may view

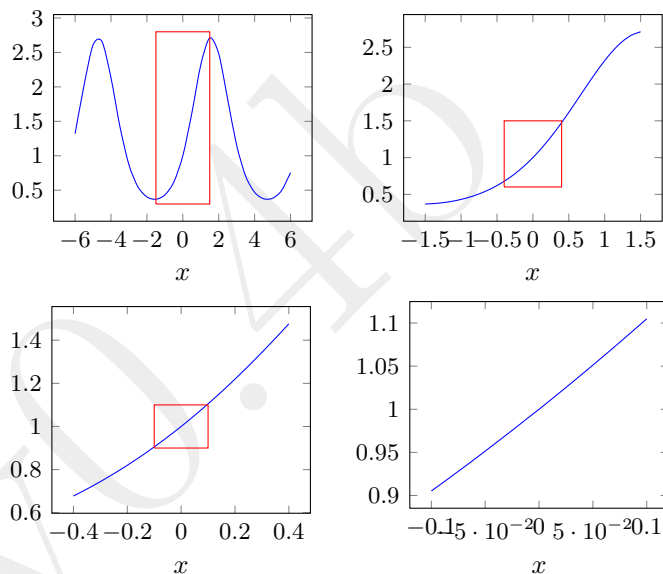


Figure 2.1: zoom in anywhere on any smooth nonlinear curve, such as the plotted  $f(x)$ , and we discover that the curve looks like a straight line on the microscale. The (red) rectangles show the region plotted in the next graph in the sequence.

any smooth curve as roughly being made up of lots of microscale straight line segments. Linear equations and their algebra on this microscale empower our understanding of nonlinear relations—for example, microscale linearity underwrites all of calculus.

## 2.1 Introduction to systems of linear equations

### Section Contents

The great aspect of linear equations is that we straightforwardly manipulate them algebraically to deduce results: some results are not only immensely useful in applications but also in further theory.

**Example 2.1.1** (simple algebraic manipulation). Following [Example 2.0.1](#), recall that the temperature in Fahrenheit  $T_F = \frac{9}{5}T_C + 32$  in terms of the temperature in Celsius,  $T_C$ . Straightforward algebra answers the following questions.

- What is a formula for the Celsius temperature as a function of the temperature in Fahrenheit? Answer by rearranging the equation: subtract 32 from both sides,  $T_F - 32 = \frac{9}{5}T_C$ ; multiply both sides by  $\frac{5}{9}$ , then  $\frac{5}{9}(T_F - 32) = T_C$ ; that is,  $T_C = \frac{5}{9}T_F - \frac{160}{9}$ .
- What temperature has the same *numerical value* in the two scales? That is, when is  $T_F = T_C$ ? Answer by algebra:

we want  $T_C = T_F = \frac{9}{5}T_C + 32$ ; subtract  $\frac{9}{5}T_C$  from both sides to give  $-\frac{4}{5}T_C = 32$ ; multiply both sides by  $-\frac{5}{4}$ , then  $T_C = -\frac{5}{4} \times 32 = -40$ . Algebra discovers that  $-40^\circ\text{C}$  is the same temperature as  $-40^\circ\text{F}$ . ■

Linear equations are characterised by each unknown never being multiplied or divided by another unknown, or itself, nor inside ‘curvaceous’ functions. Table 2.1 lists examples of both. The power of linear algebra is especially important for large numbers of unknown variables: thousands or millions of variables are common in modern applications. Generally we say there are  $n$  unknown variables. The value of  $n$  maybe two or three as in many examples, or may be thousands or millions in many modern applications.

**Definition 2.1.2.** A *linear equation* in the  $n$  variables  $x_1, x_2, \dots, x_n$  is an equation that can be written in the form

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



Table 2.1: examples of linear equations, and equations that are not linear (called nonlinear equations).

linear	nonlinear
$-3x + 2 = 0$	$x^2 - 3x + 2 = 0$
$2x - 3y = -1$	$2xy = 3$
$-1.2x_1 + 3.4x_2 - x_3 = 5.6$	$x_1^2 + 2x_2^2 = 4$
$r - 5s = 2 - 3s + 2t$	$r/s = 2 + t$
$\sqrt{3}t_1 + \frac{\pi}{2}t_2 - t_3 = 0$	$3\sqrt{t_1} + t_2^3/t_3 = 0$
$(\cos \frac{\pi}{6})x + e^2y = 1.23$	$x + e^{2y} = 1.23$

where the **coefficients**  $a_1, a_2, \dots, a_n$  and the **constant term**  $b$  are given scalar constants. An equation that cannot be written in this form is called a **nonlinear equation**. A **system** of linear equations is a set of one or more linear equations in one or more variables (usually more than one).

**Example 2.1.3** (two equations in two variables). Graphically and algebraically solve each of the following systems.

(a) 
$$\begin{aligned}x + y &= 3 \\ 2x - 4y &= 0\end{aligned}$$

(b) 
$$\begin{aligned}2x - 3y &= 2 \\ -4x + 6y &= 3\end{aligned}$$

(c) 
$$\begin{aligned}x + 2y &= 4 \\ 2x + 4y &= 8\end{aligned}$$



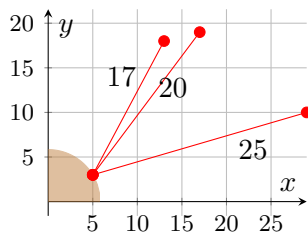
**Activity 2.1.4.** Solve the system  $x + 5y = 9$  and  $x + 2y = 3$  to find the solution is which of the following?

- (a)  $(1, 1)$       (b)  $(1, 2)$       (c)  $(-1, 1)$       (d)  $(-1, 2)$



**Example 2.1.5** (Global Positioning System). The Global Positioning System (GPS) is a network of 24 satellites orbiting the Earth. Each satellite knows very accurately its position at all times, and broadcasts this position by radio. Receivers, such as smart-phones, pick up these signals and from the time taken for the signals to arrive know the distance to those satellites within ‘sight’. Each receiver solves a system of equations and informs you of its precise position.

Let’s solve a definite example problem, but in two dimensions for simplicity. Suppose you and your smart-phone are at some unknown location  $(x, y)$  in the 2D-plane, on the Earth’s surface where the Earth has radius about 6 Mm (here all distances are measured in units of Megametres, Mm, thousands of km). But your smart-phone picks up the broadcast from three GPS satellites, and then determines their distance from you. From the broadcast and the timing, suppose you then know that a satellite at  $(29, 10)$  is 25 away (all in Mm), one at  $(17, 19)$  is 20 away, and one at  $(13, 18)$  is 17 away (as drawn in the margin). Find your location  $(x, y)$ .



If the  $x$ -axis is a line through the equator, and the  $y$ -axis goes through the North pole, then trigonometry gives that your location

would be at latitude  $\tan^{-1} \frac{3}{5} = 0.5404 = 30.96^\circ\text{N}$ . ■

**Example 2.1.6** (three equations in three variables).  
and algebraically solve the system

Graph the surfaces

$$\begin{aligned}x_1 + x_2 - x_3 &= -2, \\x_1 + 3x_2 + 5x_3 &= 8, \\x_1 + 2x_2 + x_3 &= 1.\end{aligned}$$

The sequence of marginal graphs in the previous [Example 2.1.6](#) illustrate the equations at each main step in the algebraic manipulations. Apart from keeping the solution point fixed, the sequence of graphs looks rather chaotic. Indeed there is no particular geometric pattern or interpretation of the steps in this algebra. One feature of [Section 3.3](#) is that we discover how the so-called ‘singular value decomposition’ solves linear equations via a great method with a

strong geometric interpretation. This geometric interpretation then empowers further methods useful in applications.

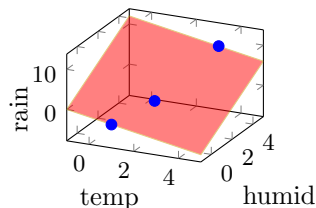
**Transform into abstract setting** Linear algebra has an important aspect crucial in applications. A crucial skill in applying linear algebra is that it takes an application problem and transforms it into an abstract setting. [Example 2.0.1](#) transformed the problem of inferring a line through two data points into solving two linear equations. The next [Example 2.1.7](#) similarly transforms the problem of inferring a plane through three data points into solving three linear equations. The original application is often not easily recognisable in the abstract version. Nonetheless, it is the abstraction by linear algebra that empowers immense results.

**Example 2.1.7** (infer a surface through three points). This example illustrates the previous paragraph. Given a geometric problem of inferring what plane passes through three given points, we transform this problem into the linear algebra task of finding the intersection point of three specific planes. This task we do.

Suppose we observe that at some given temperature and humid-

Table 2.2: in some artificial units, this table lists measured temperature, humidity, and rainfall.

'temp'	'humid'	'rain'
1	-1	-2
3	5	8
2	1	1



ity we get some rainfall: let's find a formula that predicts the rainfall from temperature and humidity measurements. In some *completely artificial units*, Table 2.2 lists measured temperature ('temp'), humidity ('humid'), and rainfall ('rain').

■

The solution of three linear equations in three variables leads to finding the intersection point of three planes. Figure 2.2 illustrates the three general possibilities: a unique solution (as in Exam-

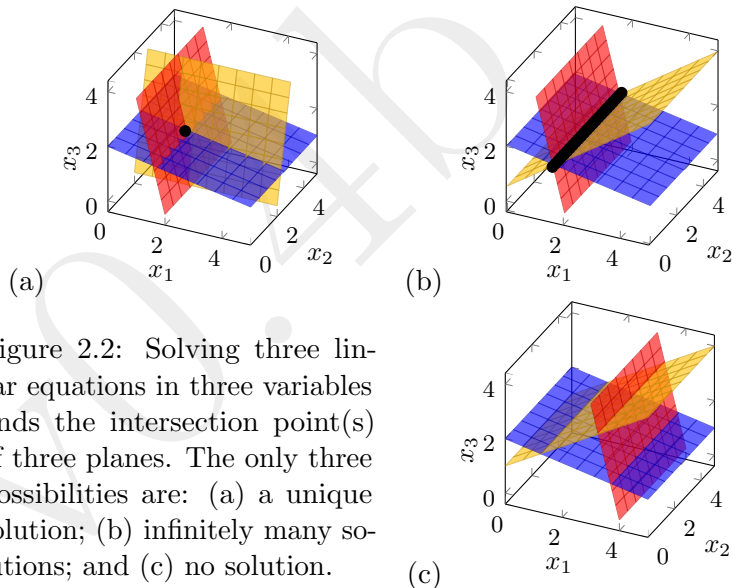


Figure 2.2: Solving three linear equations in three variables finds the intersection point(s) of three planes. The only three possibilities are: (a) a unique solution; (b) infinitely many solutions; and (c) no solution.

ple 2.1.6), or infinitely many solutions, or no solution. The solution of two linear equations in two variables also has the same three possibilities—as deduced and illustrated in Example 2.1.3. The next section establishes the general key property of a system of any number of linear equations in any number of variables: the system has either

- a unique solution (a consistent system), or
- infinitely many solutions (a consistent system), or
- no solutions (an inconsistent system).



## 2.2 Directly solve linear systems

### Section Contents

2.2.1	Compute a system's solution . . . . .	121
2.2.2	Algebraic manipulation solves systems . . . .	134
2.2.3	Three possible numbers of solutions . . . . .	144

The previous [Section 2.1](#) solved some example systems of linear equations by hand algebraic manipulation. We continue to do so for small systems. However, such by-hand solutions are tedious for systems bigger than say four equations in four unknowns. For bigger systems with tens to millions of equations—which are typical in applications—we use computers to find solutions because computers are ideal for tedious repetitive calculations.

### 2.2.1 Compute a system's solution

It is unworthy of excellent persons to lose hours like slaves in the labour of calculation.

*Gottfried Wilhelm von Leibniz*

Computers primarily deal with numbers, not algebraic equations, so we have to abstract the coefficients of a system into a numerical data structure. We use matrices and vectors.

**Example 2.2.1.** The first system of [Example 2.1.3a](#)

$$\begin{array}{l} x + y = 3 \\ 2x - 4y = 0 \end{array} \quad \text{is written} \quad \underbrace{\begin{bmatrix} 1 & 1 \\ 2 & -4 \end{bmatrix}}_A \underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_x = \underbrace{\begin{bmatrix} 3 \\ 0 \end{bmatrix}}_b.$$

That is, the system  $\begin{cases} x + y = 3 \\ 2x - 4y = 0 \end{cases}$  is equivalent to  $Ax = b$  for

- the so-called coefficient matrix  $A = \begin{bmatrix} 1 & 1 \\ 2 & -4 \end{bmatrix}$ ,

- right-hand side vector  $\mathbf{b} = (3, 0)$ , and
- vector of variables  $\mathbf{x} = (x, y)$ .



The beauty of the form  $A\mathbf{x} = \mathbf{b}$  is that the numbers involved in the system are abstracted into the matrix  $A$  and vector  $\mathbf{b}$ : MATLAB/Octave handles such numerical matrices and vectors. For some of you, writing a system in this matrix-vector form  $A\mathbf{x} = \mathbf{b}$  (Definition 2.2.2 below) will appear to be just some mystic rearrangement of symbols—such an interpretation is sufficient for this chapter. However, those of you who have met matrix multiplication will recognise that  $A\mathbf{x} = \mathbf{b}$  is an expression involving natural operations for matrices and vectors: Section 3.1 defines and explores such useful operations.

In this chapter, the two character symbol ‘ $A\mathbf{x}$ ’ is just a shorthand for all the left-hand sides of the linear equations in a system. Section 3.1 then defines a crucial multiplicative meaning to this composite symbol.

**Definition 2.2.2** (matrix-vector form). *For every given system of  $m$  linear equations in  $n$  variables*

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1,$$

$$a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 ,$$

$$\vdots$$

$$a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m ,$$

its **matrix-vector form** is  $A\mathbf{x} = \mathbf{b}$  for the  $m \times n$  **matrix** of coefficients

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

and vectors  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  and  $\mathbf{b} = (b_1, b_2, \dots, b_m)$ . If  $m = n$  (the number of equations is the same as the number of variables), then  $A$  is called a **square matrix** (the number of rows is the same as the number of columns).

**Example 2.2.3** (matrix-vector form).  
matrix-vector form.

Write the following systems in

$$\begin{array}{ll} (a) & \begin{array}{l} x_1 + x_2 - x_3 = -2, \\ x_1 + 3x_2 + 5x_3 = 8, \\ x_1 + 2x_2 + x_3 = 1. \end{array} \\ (b) & \begin{array}{l} -2r + 3s = 6, \\ s - 4t = -\pi. \end{array} \end{array}$$



**Activity 2.2.4.** Which of the following systems corresponds to the matrix-vector equation

$$\begin{bmatrix} -1 & 3 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} u \\ w \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} ?$$

$$\begin{array}{ll} (a) & \begin{array}{l} -u + 3w = 1 \\ u + 2w = 0 \end{array} \\ (b) & \begin{array}{l} -x + y = 1 \\ 3x + 2y = 0 \end{array} \\ (c) & \begin{array}{l} -u + w = 1 \\ 3u + 2w = 0 \end{array} \\ (d) & \begin{array}{l} -x + 3y = 1 \\ x + 2y = 0 \end{array} \end{array}$$



**Procedure 2.2.5** (unique solution). *In MATLAB/Octave, to solve the matrix-vector system  $A\mathbf{x} = \mathbf{b}$  for a square matrix  $A$ , use commands listed in Table 1.2 and 2.3 to:*

1. *form matrix  $A$  and column vector  $\mathbf{b}$ ;*
2. *check `rcond(A)` exists and is not too small,  $1 \geq \text{good} > 10^{-2} > \text{poor} > 10^{-4} > \text{bad} > 10^{-8} > \text{terrible}$ , (`rcond(A)` is always between zero and one inclusive);*
3. *if `rcond(A)` both exists and is acceptable, then execute  $\mathbf{x} = A \backslash \mathbf{b}$  to compute the solution vector  $\mathbf{x}$ .*

Checking `rcond(A)` avoids gross mistakes. Subsection 3.3.2 discovers what `rcond()` is, and why `rcond()` avoids mistakes. In practice, decisions about acceptability are rarely black and white, and so the qualitative ranges of `rcond()` reflects practical reality.

In theory, there is no difference between theory and

practice. But, in practice, there is.

*Jan L. A. van de Snepscheut*

**Example 2.2.6.** Use MATLAB/Octave to solve the system (from [Example 2.1.6](#))

$$x_1 + x_2 - x_3 = -2,$$

$$x_1 + 3x_2 + 5x_3 = 8,$$

$$x_1 + 2x_2 + x_3 = 1.$$

■

**Activity 2.2.7.** Use MATLAB/Octave to solve the system  $7x + 8y = 42$  and  $32x + 38y = 57$ , to find the answer for  $(x, y)$  is

(a)  $\begin{bmatrix} 114 \\ -94.5 \end{bmatrix}$       (b)  $\begin{bmatrix} -94.5 \\ 114 \end{bmatrix}$       (c)  $\begin{bmatrix} 73.5 \\ 342 \end{bmatrix}$       (d)  $\begin{bmatrix} 342 \\ 73.5 \end{bmatrix}$

■

Table 2.3: To realise [Procedure 2.2.5](#), and other procedures, we need these basics of MATLAB/Octave as well as that of [Table 1.2](#).

- The floating point numbers are extended by `Inf`, denoting ‘infinity’, and `NaN`, denoting ‘not a number’ such as the indeterminate  $0/0$ .
- `[ ... ; ... ; ... ]` forms both matrices and vectors, or use newlines instead of the semi-colons.
- `rcond(A)` of a square matrix  $A$  *estimates* the reciprocal of the so-called condition number (defined precisely by [Definition 3.3.16](#)).
- `x=A\b` computes an ‘answer’ to  $A\mathbf{x} = \mathbf{b}$ —but it may not be a solution unless `rcond(A)` exists and is not small;
- Change an element of an array or vector by assigning a new value with assignments `A(i,j)=...` or `b(i)=...` where  $i$  and  $j$  denote some indices.
- For a vector (or matrix)  $\mathbf{t}$  and an exponent  $p$ , the operation `t.^p` computes the  $p$ th power of each element in the vector; for example, if `t=[1;2;3;4;5]` then `t.^2` results in `[1;4;9;16;25]`.
- The function `ones(m,1)` gives a (column) vector of  $m$  ones,  $(1, 1, \dots, 1)$ .
- Lastly, always remember that ‘the answer’ by a computer is



**Example 2.2.8.** Following the previous [Example 2.2.6](#), solve each of the two systems:

$$\begin{array}{ll} x_1 + x_2 - x_3 = -2, & x_1 + x_2 - x_3 = -2, \\ \text{(a) } x_1 + 3x_2 + 5x_3 = 5, & \text{(b) } x_1 + 3x_2 - 2x_3 = 5, \\ x_1 - 3x_2 + x_3 = 1; & x_1 - 3x_2 + x_3 = 1. \end{array}$$

**Example 2.2.9.** Use MATLAB/Octave to solve the system

$$\begin{aligned} x_1 - 2x_2 + 3x_3 + x_4 + 2x_5 &= 7, \\ -2x_1 - 6x_2 - 3x_3 - 2x_4 + 2x_5 &= -1, \\ 2x_1 + 3x_2 - 2x_5 &= -9, \\ -2x_1 + x_2 &= -3, \\ -2x_1 - 2x_2 + x_3 + x_4 - 2x_5 &= 5. \end{aligned}$$

**Example 2.2.10.** What system of linear equations are represented by the following matrix-vector expression? and what is the result of using [Procedure 2.2.5](#) for this system?

$$\begin{bmatrix} -7 & 3 \\ 7 & -5 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} 3 \\ -2 \\ 1 \end{bmatrix}.$$

**Example 2.2.11** (partial fraction decomposition). Recall that mathematical methods sometimes need to separate a rational function into a sum of simpler ‘partial’ fractions. For example, for some purposes the fraction  $\frac{3}{(x-1)(x+2)}$  needs to be written as  $\frac{1}{x-1} - \frac{1}{x+2}$ . Solving linear equations helps:

- here pose that  $\frac{3}{(x-1)(x+2)} = \frac{A}{x-1} + \frac{B}{x+2}$  for some unknown  $A$  and  $B$ ;

- then write the right-hand side over the common denominator,

$$\begin{aligned}\frac{A}{x-1} + \frac{B}{x+2} &= \frac{A(x+2) + B(x-1)}{(x-1)(x+2)} \\ &= \frac{(A+B)x + (2A-B)}{(x-1)(x+2)}\end{aligned}$$

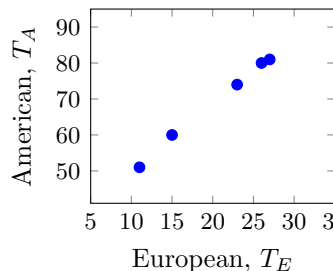
and this equals  $\frac{3}{(x-1)(x+2)}$  only if both  $A+B = 0$  and  $2A-B = 3$ ;

- solving these two linear equations gives the required  $A = 1$  and  $B = -1$  to determine the decomposition  $\frac{3}{(x-1)(x+2)} = \frac{1}{x-1} - \frac{1}{x+2}$ .

Now find the partial fraction decomposition of  $\frac{-4x^3+8x^2-5x+2}{x^2(x-1)^2}$ . ■

**Example 2.2.12** (`rcond` avoids disaster). In [Example 2.0.1](#) an American and European compared temperatures and using two days temperatures discovered the approximation that the American temperature  $T_A = 1.82 T_E + 32.73$  where  $T_E$  denotes the European temperature.

Continuing the story, three days later they again meet and compare the temperatures they experienced: the American reporting that “for the last three days it has been  $51^\circ$ ,  $74^\circ$  and  $81^\circ$ ”, whereas the European reports “why, I recorded it as  $11^\circ$ ,  $23^\circ$  and  $27^\circ$ ”. The marginal figure plots this data with the original two data points, apparently confirming a reasonable linear relationship between the two temperature scales.



Let's fit a polynomial to this temperature data. ■

The previous [Example 2.2.12](#) also illustrates one of the ‘rules of thumb’ in science and engineering: *for data fitting, avoid using polynomials of degree higher than cubic.*

**Example 2.2.13** (Global Positioning System in space-time). Recall the [Example 2.1.5](#). Consider the GPS receiver in your smart-phone. The phone's clock is generally in error, it may only be by a second but the GPS needs micro-second precision. Because of such a timing unknown, five satellites determine our precise position in space *and* time.

Suppose at some time (according to our smart-phone) the phone receives from a GPS satellite that it is at 3D location  $(6, 12, 23)$  Mm (Megametres) and that the signal was sent at a true time  $0.04$  s (seconds) before the phone's time. But the phone's time is different to the true time by some unknown amount, say  $t$ . Consequently, the travel time of the signal from the satellite to the phone is actually  $t + 0.04$ . Given the speed of light is  $c = 300$  Mm/s, this is a distance of  $300(t + 0.04) = 300t + 12$  —linear in the discrepancy of the phone's clock to the GPS clocks. Let  $(x, y, z)$  be you and your phone's position in 3D space, then the distance to the satellite is also  $\sqrt{(x - 6)^2 + (y - 12)^2 + (z - 23)^2}$ . Equating the squares of these two gives one equation

$$(x - 6)^2 + (y - 12)^2 + (z - 23)^2 = (300t + 12)^2.$$

Similarly other satellites give other equations that help determine our position. But writing “ $300t$ ” all the time is a bit tedious, so replace it with the new unknown  $w = 300t$ .

Given your phone also detects that four other satellites broadcast the following position and time information:  $(13, 20, 12)$  time

shift 0.04 s before; (17, 14, 10) time shift  $0.033\cdots$  s before; (8, 21, 10) time shift  $0.033\cdots$  s before; and (22, 9, 8) time shift 0.04 s before. Adapting the approach of [Example 2.1.5](#), use linear algebra to determine your phone's location in space.



### 2.2.2 Algebraic manipulation solves systems

A variant of GE [Gaussian Elimination] was used by the Chinese around the first century AD; the *Jiu Zhang Suanshu* (Nine Chapters of the Mathematical Art) contains a worked example for a system of five equations in five unknowns *Higham (1996) [p.195]*

This and the next subsection are not essential, but many further courses currently assume knowledge of the content. Theorems 2.2.27 and 2.2.31 are convenient to establish in the next subsection, but could alternatively be established using Procedure 3.3.15.

To solve linear equations with non-square matrices, or with poorly conditioned matrices we need to know much more details about linear algebra.

This subsection systematises the algebraic working of Examples 2.1.3 and 2.1.6. The systematic approach empowers by-hand solution of systems of linear equations, together with two general properties on the number of solutions possible. The algebraic methodology invoked here also reinforces algebraic skills that will help in further courses.

In hand calculations we often want to minimise writing, so the discussion here uses two forms side-by-side for the linear equations:

one form with all symbols recorded for best clarity; and beside it, one form where only coefficients are recorded for quickest writing. Translating from one to the other is crucial even in a computing era as the computer also primarily deals with arrays of numbers, and we must interpret what those arrays of numbers mean in terms of linear equations.

**Example 2.2.14.** Recall the system of linear equations of [Example 2.1.6](#):

$$\begin{aligned}x_1 + x_2 - x_3 &= -2, \\x_1 + 3x_2 + 5x_3 &= 8, \\x_1 + 2x_2 + x_3 &= 1.\end{aligned}$$

The first crucial level of abstraction is to write this in the matrix-vector form, [Example 2.2.3](#),

$$\underbrace{\begin{bmatrix} 1 & 1 & -1 \\ 1 & 3 & 5 \\ 1 & 2 & 1 \end{bmatrix}}_A \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} -2 \\ 8 \\ 1 \end{bmatrix}}_{\mathbf{b}}$$

A second step of abstraction omits the symbols “ $\mathbf{x} =$ ”—often we draw a vertical (dotted) line to show where the symbols “ $\mathbf{x} =$ ”



were, but this line is not essential and the theoretical statements ignore such a drawn line. Here this second step of abstraction represents this linear system by the so-called augmented matrix

$$\left[ \begin{array}{ccc|c} 1 & 1 & -1 & -2 \\ 1 & 3 & 5 & 8 \\ 1 & 2 & 1 & 1 \end{array} \right]$$



**Definition 2.2.15.** The *augmented matrix* of the system of linear equations  $A\mathbf{x} = \mathbf{b}$  is the matrix  $[A : \mathbf{b}]$ .

**Example 2.2.16.** Write down augmented matrices for the two following systems:

(a) 
$$\begin{aligned} -2r + 3s &= 6, \\ s - 4t &= -\pi, \end{aligned}$$

(b) 
$$\begin{aligned} -7y + 3z &= 3, \\ 7y - 5z &= -2, \\ y - 2z &= 1. \end{aligned}$$



**Activity 2.2.17.** Which of the following *cannot* be an augmented matrix for the system  $p + 4q = 3$  and  $-p + 2q = -2$ ?

(a)  $\left[ \begin{array}{cc|c} 2 & -1 & 3 \\ 4 & 1 & -2 \end{array} \right]$

(b)  $\left[ \begin{array}{cc|c} 4 & 1 & 3 \\ 2 & -1 & -2 \end{array} \right]$

(c)  $\left[ \begin{array}{cc|c} 1 & 4 & 3 \\ -1 & 2 & -2 \end{array} \right]$

(d)  $\left[ \begin{array}{cc|c} -1 & 2 & -2 \\ 1 & 4 & 3 \end{array} \right]$



Recall that Examples 2.1.3 and 2.1.6 manipulate the linear equations to deduce solution(s) to systems of linear equations. The following theorem validates such manipulations in general, and gives the basic operations a collective name.

**Theorem 2.2.18.** *The following **elementary row operations** can be performed on either a system of linear equations or on its corresponding augmented matrix without changing the solutions:*

(a) *interchange two equations/rows; or*

- (b) multiply an equation/row by a nonzero constant; or
- (c) add a multiple of an equation/row to another.

**Example 2.2.19.** Use elementary row operations to find the only solution of the following system of linear equations:

$$\begin{aligned}x + 2y + z &= 1, \\2x - 3y &= 2, \\-3y - z &= 2.\end{aligned}$$

Confirm with MATLAB/Octave. ■

**Definition 2.2.20.** A system of linear equations or (augmented) matrix is in **reduced row echelon form** (RREF) if:

- (a) any equations with all zero coefficients, or rows of the matrix consisting entirely of zeros, are at the bottom;
- (b) in each nonzero equation/row, the first nonzero coefficient/entry is a one (called the **leading one**), and is in a variable/column to the left of any leading ones below it; and

(c) each variable/column containing a leading one has zero coefficients/entries in every other equation/row.

A **free variable** is any variable which is not multiplied by a leading one when the row reduced echelon form is translated to its corresponding algebraic equations.

**Example 2.2.21** (reduced row echelon form). Which of the following are in reduced row echelon form (RREF)? For those that are, identify the leading ones, and treating other variables as free variables write down the most general solution of the system of linear equations.

$$(a) \begin{cases} x_1 + x_2 + 0x_3 - 2x_4 = -2 \\ 0x_1 + 0x_2 + x_3 + 4x_4 = 5 \end{cases}$$

$$(b) \begin{bmatrix} 1 & 0 & -1 & : & 1 \\ 0 & 1 & -1 & : & -2 \\ 0 & 0 & 0 & : & 4 \end{bmatrix}$$

$$(c) \begin{bmatrix} 1 & 0 & -1 & : & 1 \\ 0 & 1 & -1 & : & -2 \\ 0 & 0 & 0 & : & 0 \end{bmatrix}$$

$$(d) \begin{cases} x + 2y = 3 \\ 0x + y = -2 \end{cases}$$

$$(e) \left[ \begin{array}{cccc|c} -1 & 4 & 1 & 6 & -1 \\ 3 & 0 & 1 & -2 & -2 \end{array} \right]$$



**Activity 2.2.22.** Which one of the following augmented matrices is *not* in reduced row echelon form?

$$(a) \left[ \begin{array}{ccc|c} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 2 \end{array} \right]$$

$$(b) \left[ \begin{array}{ccc|c} 1 & 0 & 1 & 2 \\ 0 & 1 & 0 & 1 \end{array} \right]$$

$$(c) \left[ \begin{array}{ccc|c} 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{array} \right]$$

$$(d) \left[ \begin{array}{ccc|c} 1 & 1 & 0 & 0 \\ 0 & -1 & 1 & -1 \end{array} \right]$$



**Activity 2.2.23.** Which one of the following is a general solution to the system with augmented matrix in reduced row echelon form of

$$\left[ \begin{array}{ccc|c} 1 & 0 & -0.2 & 0.4 \\ 0 & 1 & -1.2 & -0.6 \end{array} \right] ?$$

- (a)  $(0.2t + 0.4, 1.2t - 0.6, t)$       (b)  $(0.4, -0.6, 0)$   
(c) solution does not exist      (d)  $(0.2 + 0.4t, 1.2 + 0.6t, t)$



The previous [Example 2.2.21](#) showed that given a system of linear equations in reduced row echelon form we can either immediately write down all solutions, or immediately determine if none exists. Generalising [Example 2.2.19](#), the following Gauss–Jordan procedure uses elementary row operations ([Theorem 2.2.18](#)) to find an equivalent system of equations in reduced row echelon form. From such a form we then write down a general solution.

Computers and graphics calculators perform Gauss–Jordan elimination for you; for example, `A\` in MATLAB/Octave. However, when `rcond` indicates `A\` is inappropriate, then the singular value decomposition of [Section 3.3](#) is a far better choice than such Gauss–Jordan elimination.

**Procedure 2.2.24** (Gauss–Jordan elimination).

1. Write down either the full symbolic form of the system of linear equations, or the augmented matrix of the system of linear equations.
2. Use elementary row operations to reduce the system/augmented matrix to reduced row echelon form.
3. If the resulting system is consistent, then solve for the leading variables in terms of any remaining free variables.

**Example 2.2.25.** Use Gauss–Jordan elimination, [Procedure 2.2.24](#), to find all possible solutions to the system

$$\begin{aligned}-x - y &= -3, \\ x + 4y &= -1, \\ 2x + 4y &= c,\end{aligned}$$

depending upon the parameter  $c$ .



**Example 2.2.26.** Use Gauss–Jordan elimination, [Procedure 2.2.24](#), to find all possible solutions to the system

$$\begin{cases} -2v + 3w = -1, \\ 2u + v + w = -1. \end{cases}$$





### 2.2.3 Three possible numbers of solutions

The number of possible solutions to a system of equations is fundamental. We need to know all the possibilities. As seen in previous examples, the following theorem says there are only three possibilities for linear equations.

**Theorem 2.2.27.** *For every system of linear equations  $A\mathbf{x} = \mathbf{b}$ , exactly one of the following is true:*

- *there is no solution;*
- *there is a unique solution;*
- *there are infinitely many solutions.*

An important class of linear equations always has at least one solution, never none. For example, modify [Example 2.2.25](#) to

$$\begin{aligned}-x - y &= 0, \\ x + 4y &= 0, \\ 2x + 4y &= 0,\end{aligned}$$

and then  $x = y = 0$  is immediately a solution. The reason is that the right-hand side is all zeros and so  $x = y = 0$  makes the left-hand sides also zero.

**Definition 2.2.28.** *A system of linear equations is called **homogeneous** if the (right-hand side) constant term in each equation is zero; that is, when the system may be written  $A\mathbf{x} = \mathbf{0}$ . Otherwise the system is termed **non-homogeneous**.*

**Example 2.2.29.**

(a)  $\begin{cases} 3x_1 - 3x_2 = 0 \\ -x_1 - 7x_2 = 0 \end{cases}$  is homogeneous. Solving, the first equation gives  $x_1 = x_2$  and substituting in the second then gives  $-x_2 - 7x_2 = 0$  so that  $x_1 = x_2 = 0$  is the only solution. It must have  $\mathbf{x} = \mathbf{0}$  as a solution as the system is homogeneous.

(b)  $\begin{cases} 2r + s - t = 0 \\ r + s + 2t = 0 \\ -2r + s = 3 \\ 2r + 4s - t = 0 \end{cases}$  is not homogeneous because there is a

non-zero constant on the right-hand side.

- (c)  $\begin{cases} -2 + y + 3z = 0 \\ 2x + y + 2z = 0 \end{cases}$  is not homogeneous because there is a non-zero constant in the first equation, the  $(-2)$ , even though it is here sneakily written on the left-hand side.

- (d)  $\begin{cases} x_1 + 2x_2 + 4x_3 - 3x_4 = 0 \\ x_1 + 2x_2 - 3x_3 + 6x_4 = 0 \end{cases}$  is homogeneous. Use Gauss–Jordan elimination, [Procedure 2.2.24](#), to solve:

$$\begin{cases} x_1 + 2x_2 + 4x_3 - 3x_4 = 0 \\ x_1 + 2x_2 - 3x_3 + 6x_4 = 0 \end{cases} \iff \begin{bmatrix} 1 & 2 & 4 & -3 & : & 0 \\ 1 & 2 & -3 & 6 & : & 0 \end{bmatrix}$$

Subtract the first row from the second.

$$\begin{cases} x_1 + 2x_2 + 4x_3 - 3x_4 = 0 \\ 0x_1 + 0x_2 - 7x_3 + 9x_4 = 0 \end{cases} \iff \begin{bmatrix} 1 & 2 & 4 & -3 & : & 0 \\ 0 & 0 & -7 & 9 & : & 0 \end{bmatrix}$$

Divide the second row by  $(-7)$ .

$$\begin{cases} x_1 + 2x_2 + 4x_3 - 3x_4 = 0 \\ 0x_1 + 0x_2 + x_3 - \frac{9}{7}x_4 = 0 \end{cases} \iff \begin{bmatrix} 1 & 2 & 4 & -3 & : & 0 \\ 0 & 0 & 1 & -\frac{9}{7} & : & 0 \end{bmatrix}$$

Subtract four times the second row from the first.

$$\begin{cases} x_1 + 2x_2 + 0x_3 + \frac{15}{7}x_4 = 0 \\ 0x_1 + 0x_2 + x_3 - \frac{9}{7}x_4 = 0 \end{cases} \iff \begin{bmatrix} 1 & 2 & 0 & \frac{15}{7} & : & 0 \\ 0 & 0 & 1 & -\frac{9}{7} & : & 0 \end{bmatrix}$$

The system is now in reduced row echelon form. The second and fourth columns are those of free variables so set the second and fourth component  $x_2 = s$  and  $x_4 = t$  for arbitrary  $s$  and  $t$ . Then the first row gives  $x_1 = -2s - \frac{15}{7}t$ , and the second row gives  $x_3 = \frac{9}{7}t$ . That is, the solutions are  $\mathbf{x} = (-2s - \frac{15}{7}t, s, \frac{9}{7}t, t) = (-2, 1, 0, 0)s + (-\frac{15}{7}, 0, \frac{9}{7}, 1)t$  for arbitrary  $s$  and  $t$ . These solutions include  $\mathbf{x} = \mathbf{0}$  via the choice  $s = t = 0$ .



**Activity 2.2.30.** Which one of the following systems of equations for  $x$  and  $y$  is homogeneous?

(a)  $-3x - y = 0$  and  
 $7x + 5y = 3$

(b)  $3x + 1 = 0$  and  $-x - y = 0$

(c)  $5y = 3x$  and  $4x = 2y$

(d)  $-2x + y - 3 = 0$  and  
 $x + 4 = 2y$



As [Example 2.2.29d](#) illustrates, a further subclass of homogeneous systems is immediately known to have an infinite number of solutions. Namely, if the number of equations is less than the number of unknowns (two is less than four in the last example), then a homogeneous system always has an infinite number of solutions.

**Theorem 2.2.31.** *If  $A\mathbf{x} = \mathbf{0}$  is a homogeneous system of  $m$  linear equations with  $n$  variables where  $m < n$ , then the system has infinitely many solutions.*

Remember that this theorem says nothing about the cases where there are at least as many equations as variables ( $m \geq n$ ), when there may or may not be an infinite number of solutions.

### **Prefer a matrix/vector level**

Working at the element level in this way leads to a profusion of symbols, superscripts, and subscripts that tend to obscure the mathematical structure and hinder insights being drawn into the underlying process. One of the key developments in the last century was the recognition that it is much more profitable to work at the matrix level. *(Higham 2015, §2)*

A large part of this and preceding sections is devoted to arithmetic and algebraic manipulations on the individual coefficients and variables in the system. This is working at the ‘element level’ commented on by Higham. But as Higham also comments, we need to work more at a whole matrix level. This means we need to discuss and manipulate matrices as a whole, not get enmeshed in

the intricacies of the element operations. This has close intellectual parallels in computing where abstract data structures empower us to encode complex tasks: here the analogous abstract data structures are matrices and vectors, and working with matrices and vectors as objects in their own right empowers linear algebra. The next chapter proceeds to develop linear algebra at the matrix level. But first, the next [Section 2.3](#) establishes some necessary fundamental aspects at the vector level.

## 2.3 Linear combinations span sets

### Section Contents

A common feature in the solution to linear equations is the appearance of combinations of several vectors. For example, the general solution to [Example 2.2.29d](#) is

$$\begin{aligned}\mathbf{x} &= \left(-2s - \frac{15}{7}t, s, \frac{9}{7}t, t\right) \\ &= \underbrace{s(-2, 1, 0, 0) + t\left(-\frac{15}{7}, 0, \frac{9}{7}, 1\right)}_{\text{linear combination}}.\end{aligned}$$

The general solution to [Example 2.2.21a](#) is

$$\begin{aligned}\mathbf{x} &= (-2 - s + 2t, s, 5 - 4t, t) \\ &= \underbrace{1 \cdot (-2, 0, 5, 0) + s(-1, 1, 0, 0) + t(2, 0, -4, 1)}_{\text{linear combination}}.\end{aligned}$$

Such so-called linear combinations occur in many other contexts. Recall the standard unit vectors in  $\mathbb{R}^3$  are  $\mathbf{e}_1 = (1, 0, 0)$ ,  $\mathbf{e}_2 =$



$(0, 1, 0)$  and  $\mathbf{e}_3 = (0, 0, 1)$  (Definition 1.2.7): so any other vector in  $\mathbb{R}^3$  may be written as

$$\begin{aligned}\mathbf{x} &= (x_1, x_2, x_3) \\ &= x_1(1, 0, 0) + x_2(0, 1, 0) + x_3(0, 0, 1) \\ &= \underbrace{x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3}_{\text{linear combination}}.\end{aligned}$$

The wide-spread appearance of such combinations calls for the following definition.

**Definition 2.3.1.** *A vector  $\mathbf{v}$  is a **linear combination** of vectors  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$  if there are scalars  $c_1, c_2, \dots, c_k$  (called the **coefficients**) such that  $\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_k\mathbf{v}_k$ .*

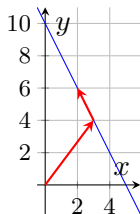
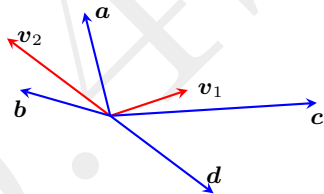
**Example 2.3.2.** Estimate roughly each of the blue vectors as a linear combination of the given red vectors in the following graphs (estimate coefficients to say roughly 10% error).



**Activity 2.3.3.** Choose any one of these linear combinations:

$$2\mathbf{v}_1 - 0.5\mathbf{v}_2; \quad 0\mathbf{v}_1 - \mathbf{v}_2; \quad -0.5\mathbf{v}_1 + 0.5\mathbf{v}_2; \quad \mathbf{v}_1 + \mathbf{v}_2.$$

Then in the plot below, which vector,  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$  or  $\mathbf{d}$ , corresponds to the chosen linear combination?

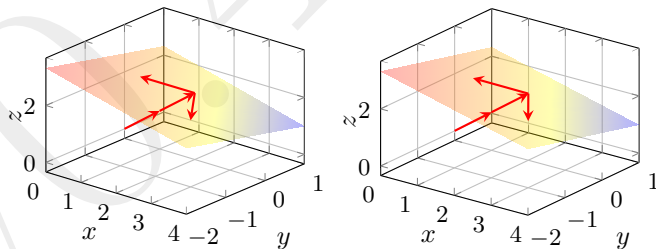


**Example 2.3.4.** Parametric descriptions of lines and planes involve linear combinations (Sections 1.2–1.3).

- (a) For each value of  $t$ , the expression  $(3, 4) + t(-1, 2)$  is a linear combination of the two vectors  $(3, 4)$  and  $(-1, 2)$ . Over all

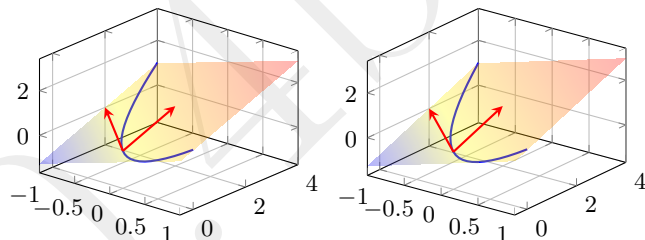
values of parameter  $t$  it describes the line illustrated in the margin. (The line is alternatively described as  $2x + y = 10$ .)

- (b) For each value of  $s$  and  $t$ , the expression  $2(1, 0, 1) + s(-1, -\frac{1}{2}, \frac{1}{2}) + t(1, -1, 0)$  is a linear combination of the three vectors  $(1, 0, 1)$ ,  $(-1, -\frac{1}{2}, \frac{1}{2})$  and  $(1, -1, 0)$ . Over all values of the parameters  $s$  and  $t$  it describes the plane illustrated below. (Alternatively the plane could be described as  $x + y + 3z = 8$ ).



- (c) The expression  $t(-1, 2, 0) + t^2(0, 2, 1)$  is a linear combination of the two vectors  $(-1, 2, 0)$  and  $(0, 2, 1)$  as the vectors are multiplied by scalars and then added. That a coefficient is a nonlinear function of some parameter is irrelevant to

the property of linear combination. This expression is a parametric description of a parabola in  $\mathbb{R}^3$ , as illustrated below, and very soon we will be able to say it is a parabola in the plane spanned by  $(-1, 2, 0)$  and  $(0, 2, 1)$ .



The matrix-vector form  $A\mathbf{x} = \mathbf{b}$  of a system of linear equations involves a linear combination on the left-hand side.

**Example 2.3.5.** Recall from [Definition 2.2.2](#) that  $\begin{bmatrix} -5 & 4 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$  is our abstract abbreviation for the system of two equations

$$\begin{aligned} -5x + 4y &= 1, \\ 3x + 2y &= -2. \end{aligned}$$

Form both sides into a vector so that

$$\begin{bmatrix} -5x + 4y \\ 3x + 2y \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}.$$

Write the left-hand side as the sum of two vectors:

$$\begin{bmatrix} -5x \\ 3x \end{bmatrix} + \begin{bmatrix} 4y \\ 2y \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}.$$

By scalar multiplication the system becomes

$$\begin{bmatrix} -5 \\ 3 \end{bmatrix} x + \begin{bmatrix} 4 \\ 2 \end{bmatrix} y = \begin{bmatrix} 1 \\ -2 \end{bmatrix}.$$

That is, the left-hand side is a linear combination of  $(-5, 3)$  and  $(4, 2)$ , the two columns of the matrix. ■

**Example 2.3.6.** Let's repeat the previous example in general. Recall from [Definition 2.2.2](#) that  $A\mathbf{x} = \mathbf{b}$  is our abstract abbreviation for the system of  $m$  equations

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

Form both sides into a vector so that

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

Then use addition and scalar multiplication of vectors ([Defini-](#)

tion 1.2.4) to rewrite the left-hand side vector as

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

This left-hand side is a linear combination of the columns of matrix  $A$ : define from the columns of  $A$  the  $n$  vectors,  $\mathbf{a}_1 = (a_{11}, a_{21}, \dots, a_{m1})$ ,  $\mathbf{a}_2 = (a_{12}, a_{22}, \dots, a_{m2})$ ,  $\dots$ ,  $\mathbf{a}_n = (a_{1n}, a_{2n}, \dots, a_{mn})$ , then the left-hand side is a linear combination of these vectors, with the coefficients of the linear combination being  $x_1, x_2, \dots, x_n$ . That is, the system  $A\mathbf{x} = \mathbf{b}$  is identical to the linear combination  $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b}$ . ■

Be aware of a subtle twist going on here: for the general [Example 2.3.6](#) this theorem turns a question about the existence of an  $n$  variable solution  $\mathbf{x}$ , into a question about vectors with  $m$  components; and vice-versa.

### Theorem 2.3.7.

A system of linear equations  $A\mathbf{x} = \mathbf{b}$  is consistent ([Procedure 2.2.24](#)) if and only if the right-hand side vector  $\mathbf{b}$  is a linear combination of the columns of  $A$ .



**Example 2.3.8.** This first example considers the simplest cases when the matrix has only one column, and so any linear combination is only a scalar multiple of that column. Compare the consistency of the equations with the right-hand side being a linear combination of the column of the matrix.

(a)  $\begin{bmatrix} -1 \\ 2 \end{bmatrix} x = \begin{bmatrix} -2 \\ 4 \end{bmatrix}.$

(b)  $\begin{bmatrix} -1 \\ 2 \end{bmatrix} x = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$

(c)  $\begin{bmatrix} 1 \\ a \end{bmatrix} x = \begin{bmatrix} 3 \\ -6 \end{bmatrix}$  depending upon parameter  $a$ .



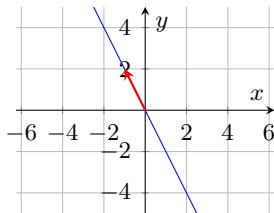
**Activity 2.3.9.** For what value of  $a$  is the system  $\begin{bmatrix} 3-a \\ -2a \end{bmatrix} x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$  consistent?

- (a)  $a = -3$       (b)  $a = -\frac{1}{2}$       (c)  $a = 2$       (d)  $a = 1$



In the Examples 2.3.4 and 2.3.6 of linear combination, the coefficients mostly are a variable parameter or unknown. Consequently, mostly we are interested in the range of possibilities encompassed by a given set of vectors.

**Definition 2.3.10.** Let a set of  $k$  vectors in  $\mathbb{R}^n$  be  $S = \{v_1, v_2, \dots, v_k\}$ , then the set of all linear combinations of  $v_1, v_2, \dots, v_k$  is called the **span** of  $v_1, v_2, \dots, v_k$ , and is denoted by  $\text{span}\{v_1, v_2, \dots, v_k\}$  or  $\text{span } S$ .

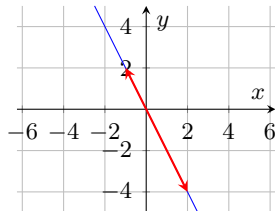


- Example 2.3.11.** (a) Let the set  $S = \{(-1, 2)\}$  with just one vector. Then  $\text{span } S = \text{span}\{(-1, 2)\}$  is the set of all vectors encompassed by the form  $t(-1, 2)$ . From the parametric equation of a line ([Definition 1.2.15](#)),  $\text{span } S$  is all vectors in the line  $y = -2x$  as shown in the margin.
- (b) With two vectors in the set,  $\text{span}\{(-1, 2), (3, 4)\} = \mathbb{R}^2$  is the entire 2D plane. To see this, recall that any point in the span must be of the form  $s(-1, 2) + t(3, 4)$ . Given any vector  $(x_1, x_2)$  in  $\mathbb{R}^2$  we choose  $s = (-4x_1 + 3x_2)/10$  and  $t = (2x_1 + x_2)/10$  and then the linear combination

$$\begin{aligned} s \begin{bmatrix} -1 \\ 2 \end{bmatrix} + t \begin{bmatrix} 3 \\ 4 \end{bmatrix} &= \frac{-4x_1 + 3x_2}{10} \begin{bmatrix} -1 \\ 2 \end{bmatrix} + \frac{2x_1 + x_2}{10} \begin{bmatrix} 3 \\ 4 \end{bmatrix} \\ &= x_1 \left( \frac{-4}{10} \begin{bmatrix} -1 \\ 2 \end{bmatrix} + \frac{2}{10} \begin{bmatrix} 3 \\ 4 \end{bmatrix} \right) \\ &\quad + x_2 \left( \frac{3}{10} \begin{bmatrix} -1 \\ 2 \end{bmatrix} + \frac{1}{10} \begin{bmatrix} 3 \\ 4 \end{bmatrix} \right) \\ &= x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} \end{aligned}$$

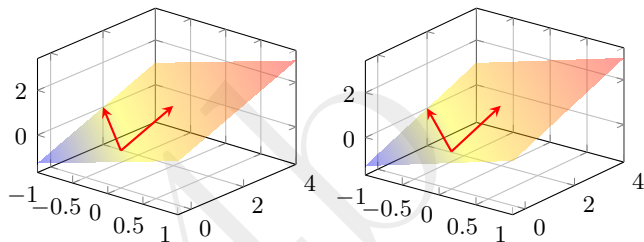
$$= \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Since every vector in  $\mathbb{R}^2$  can be expressed as  $s(-1, 2) + t(3, 4)$ , then  $\mathbb{R}^2 = \text{span}\{(-1, 2), (3, 4)\}$



- (c) But if two vectors are proportional to each other then their span is a line. For example,  $\text{span}\{(-1, 2), (2, -4)\}$  is the set of all vectors of the form  $r(-1, 2) + s(2, -4) = r(-1, 2) + (-2s)(-1, 2) = (r - 2s)(-1, 2) = t(-1, 2)$  for  $t = r - 2s$ . That is,  $\text{span}\{(-1, 2), (2, -4)\} = \text{span}\{(-1, 2)\}$  as illustrated in the margin.

- (d) In 3D,  $\text{span}\{(-1, 2, 0), (0, 2, 1)\}$  is the set of all linear combinations  $s(-1, 2, 0) + t(0, 2, 1)$  which here is a parametric form of the plane illustrated below ([Definition 1.3.32](#)). The plane passes through the origin  $\mathbf{0}$ , obtained when  $s = t = 0$ .



One could also check that the vector  $(2, 1, -2)$  is orthogonal to these two vectors, hence is a normal to the plane, and so the plane may be also expressed as  $2x + y - 2z = 0$ .

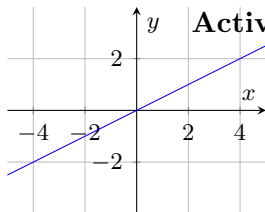
- (e) For the complete set of  $n$  standard unit vectors in  $\mathbb{R}^n$  (Definition 1.2.7),  $\text{span}\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\} = \mathbb{R}^n$ . This is because every vector  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  in  $\mathbb{R}^n$  may be written as the linear combination  $\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \dots + x_n\mathbf{e}_n$ , and hence every vector is in  $\text{span}\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ .
- (f) The homogeneous system (Definition 2.2.28) of linear equations from Example 2.2.29d has solutions  $\mathbf{x} = (-2s - \frac{15}{7}t, s, \frac{9}{7}t, t) = (-2, 1, 0, 0)s + (-\frac{15}{7}, 0, \frac{9}{7}, 1)t$  for arbitrary  $s$  and  $t$ . That is, the set of solutions is  $\text{span}\{(-2, 1, 0, 0), (-\frac{15}{7}, 0, \frac{9}{7}, 1)\}$ , a subset

of  $\mathbb{R}^4$ .

Generally, the set of solutions to a homogeneous system is the span of some set.

- (g) However, the set of solutions to a non-homogeneous system is generally not the span of some set. For example, the solutions to [Example 2.2.26](#) are all of the form  $(u, v, w) = (-\frac{3}{4} - \frac{1}{4}t, \frac{1}{2} + \frac{3}{2}t, t) = (-\frac{3}{4}, \frac{1}{2}, 0) + t(-\frac{1}{4}, \frac{3}{2}, 1)$  for arbitrary  $t$ . True, each of these solutions is a linear combination of vectors  $(-\frac{3}{4}, \frac{1}{2}, 0)$  and  $(-\frac{1}{4}, \frac{3}{2}, 1)$ . But the multiple of  $(-\frac{3}{4}, \frac{1}{2}, 0)$  is always fixed, whereas the span invokes *all* multiples. Consequently, all the possible solutions cannot be the same as the span of a set of vectors.



**Activity 2.3.12.**

In the margin is drawn a line: for which one of the following vectors  $\mathbf{u}$  is  $\text{span}\{\mathbf{u}\}$  *not* the drawn line?

(a)  $(-1, -0.5)$

(b)  $(4, 2)$

(c)  $(-1, -2)$

(d)  $(2, 1)$

**Example 2.3.13.** Describe in other words  $\text{span}\{\mathbf{i}, \mathbf{k}\}$  in  $\mathbb{R}^3$ .

**Example 2.3.14.** Find a set  $S$  such that  $\text{span } S = \{(3b, a + b, -2a - 4b) : a, b \text{ scalars}\}$ . Similarly, find a set  $T$  such that  $\text{span } T = \{(-a - 2b - 2, -b + 1, -3b - 1) : a, b \text{ scalars}\}$ .

Geometrically, the span of a set of vectors is always all vectors lying in either a line, a plane, or a higher dimensional hyper-plane, that passes *through the origin* (discussed further by [Section 3.4](#)).

---

## 3 Matrices encode system interactions

---

### Chapter Contents

3.1	Matrix operations and algebra . . . . .	171
3.1.1	Basic matrix terminology . . . . .	173
3.1.2	Addition, subtraction and multiplication with matrices . . . . .	180
3.1.3	Familiar algebraic properties of matrix oper- ations . . . . .	223
3.2	The inverse of a matrix . . . . .	228
3.2.1	Introducing the unique inverse . . . . .	230
3.2.2	Diagonal matrices stretch and shrink . . . . .	243



3.2.3	Orthogonal matrices rotate . . . . .	261
3.3	Factorise to the singular value decomposition . . . .	276
3.3.1	Introductory examples . . . . .	278
3.3.2	The SVD solves general systems . . . . .	285
3.3.3	Prove the SVD Theorem 3.3.6 . . . . .	313
3.4	Subspaces, basis and dimension . . . . .	324
3.4.1	Subspaces are lines, planes, and so on . . . .	326
3.4.2	Orthonormal bases form a foundation . . . .	345
3.4.3	Is it a line? a plane? The dimension answers	361
3.5	Project to solve inconsistent equations . . . . .	373
3.5.1	Make a minimal change to the problem . . .	375
3.5.2	Compute the smallest appropriate solution	395
3.5.3	Orthogonal projection resolves vector components . . . . .	409

3.6	Introducing linear transformations . . . . .	445
3.6.1	Matrices correspond to linear transformations	456
3.6.2	The pseudo-inverse of a matrix . . . . .	463
3.6.3	Function composition connects to matrix inverse . . . . .	472

Section 2.2 introduced matrices in the matrix-vector form  $A\mathbf{x} = \mathbf{b}$  of a system of linear equations. This chapter starts with Sections 3.1 and 3.2 developing the basic operations on matrices that make them so useful in applications and theory—including making sense of the ‘product’  $A\mathbf{x}$ . Section 3.3 then explores how the so-called “singular value decomposition (SVD)” of a matrix empowers us to understand how to solve general linear systems of equations, and a graphical meaning of a matrix in terms of rotations and stretching. The structures discovered by an SVD lead to further conceptual development (Section 3.4) that underlies the at first paradoxical solution of inconsistent equations (Section 3.5). Finally, Section 3.6 unifies the geometric views invoked.

the language of mathematics reveals itself unreasonably effective in the natural sciences ... a wonderful gift which we neither understand nor deserve. We should be grateful for it and hope that it will remain valid in future research and that it will extend, for better or for worse, to our pleasure even though perhaps also to our bafflement, to wide branches of learning

*Wigner, 1960 (Mandelbrot 1982, p.3)*

## 3.1 Matrix operations and algebra

### Section Contents

3.1.1	Basic matrix terminology . . . . .	173
3.1.2	Addition, subtraction and multiplication with matrices . . . . .	180
	Matrix addition and subtraction . . . . .	181
	Scalar multiplication of matrices . . . . .	184
	Matrix-vector multiplication transforms . . .	186
	Matrix-matrix multiplication . . . . .	195
	The transpose of a matrix . . . . .	202
	Compute in MATLAB/Octave . . . . .	208
3.1.3	Familiar algebraic properties of matrix oper- ations . . . . .	223

This section introduces basic matrix concepts, operations and algebra. Many of you will have met some of it in previous study.

### 3.1.1 Basic matrix terminology

Let's start with some basic definitions of terminology.

- As already introduced by [Section 2.2](#), a **matrix** is a rectangular array of real numbers, written inside **brackets**  $[\cdots]$ , such as these six examples:

$$\begin{bmatrix} -2 & -5 & 4 \\ 1 & -3 & 0 \\ 2 & 4 & 0 \end{bmatrix}, \quad \begin{bmatrix} -2.33 & 3.66 \\ -4.17 & -0.36 \end{bmatrix}, \quad \begin{bmatrix} 0.56 \\ 3.99 \\ -5.22 \end{bmatrix}, \\ \begin{bmatrix} 1 & -\sqrt{3} & \pi \\ -5/3 & \sqrt{5} & -1 \end{bmatrix}, \quad \begin{bmatrix} 1 & \frac{10}{3} & \frac{\pi^2}{4} \end{bmatrix}, \quad [0.35]. \quad (3.1)$$

- The **size** of a matrix is its number of rows and columns—written  $m \times n$  where  $m$  is the number of rows and  $n$  is the number of columns. The six example matrices of (3.1) are of size, respectively,  $3 \times 3$ ,  $2 \times 2$ ,  $3 \times 1$ ,  $2 \times 3$ ,  $1 \times 3$ , and  $1 \times 1$ .

Recall from [Definition 2.2.2](#) that if the number of rows equals the number of columns,  $m = n$ , then it is called a square

matrix. For example, the first, second and last matrices in (3.1) are square; the others are not.

- To correspond with vectors, we often invoke the term **column vector** which means a matrix with only one column; that is, a matrix of size  $m \times 1$  for some  $m$ . For convenience and compatibility with vectors, we often write a column vector horizontally within **parentheses**  $(\cdots)$ . The third matrix of (3.1) is an example, and may also be written as  $(0.56, 3.99, -5.22)$ .

Occasionally we refer to a **row vector** to mean a matrix with one row; that is, a  $1 \times n$  matrix for some  $n$ , such as the fifth matrix of (3.1). Remember the distinction: a row of numbers written within brackets,  $[\cdots]$ , is a row vector, whereas a row of numbers written within parentheses,  $(\cdots)$ , is a column vector.

- The numbers appearing in a matrix are called the **entries**, **elements** or **components** of the matrix. For example, the first matrix in (3.1) has entries/elements/components of the

numbers  $-5$ ,  $-3$ ,  $-2$ ,  $0$ ,  $1$ ,  $2$  and  $4$ .

- But it is important to identify where the numbers appear in a matrix: the **double subscript** notation identifies the location of an entry. For a matrix  $A$ , the entry in row  $i$  and column  $j$  is denoted by  $a_{ij}$ : by convention we use capital (uppercase) letters for a matrix, and the corresponding lowercase letter subscripted for its entries. For example, let matrix

$$A = \begin{bmatrix} -2 & -5 & 4 \\ 1 & -3 & 0 \\ 2 & 4 & 0 \end{bmatrix},$$

then entries  $a_{12} = -5$ ,  $a_{22} = -3$  and  $a_{31} = 2$ .

- The first of two special matrices is a **zero matrix** of all zeros and of any size: the symbol  $O_{m \times n}$  denotes the  $m \times n$  zero matrix, such as

$$O_{2 \times 4} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The symbol  $O_n$  denotes the square zero matrix of size  $n \times n$ ,



whereas the plain symbol  $O$  denotes a zero matrix whose size is apparent from the context.

- Arising from the nature of matrix multiplication ([Subsection 3.1.2](#)), the second special matrix is the **identity matrix**: the symbol  $I_n$  denotes a  $n \times n$  square matrix which has zero entries except for the diagonal from the top-left to the bottom-right which are all ones. Occasionally we invoke non-square ‘identity’ matrices denoted by  $I_{m \times n}$ . For examples,

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad I_{2 \times 3} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad I_{4 \times 2} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

The plain symbol  $I$  denotes an identity matrix whose size is apparent from the context.

- Using the double subscript notation, and as already used in

**Definition 2.2.2**, a general  $m \times n$  matrix

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

Often, as already seen in [Example 2.3.6](#), it is useful to write a matrix  $A$  in terms of its  $n$  column vectors  $\mathbf{a}_j$ ,  $A = [\mathbf{a}_1 \ \mathbf{a}_2 \ \cdots \ \mathbf{a}_n]$ . For example, matrix

$$B = \begin{bmatrix} 1 & -\sqrt{3} & \pi \\ -5/3 & \sqrt{5} & -1 \end{bmatrix} = [\mathbf{b}_1 \ \mathbf{b}_2 \ \mathbf{b}_3]$$

for the three column vectors

$$\mathbf{b}_1 = \begin{bmatrix} 1 \\ -5/3 \end{bmatrix}, \quad \mathbf{b}_2 = \begin{bmatrix} -\sqrt{3} \\ \sqrt{5} \end{bmatrix}, \quad \mathbf{b}_3 = \begin{bmatrix} \pi \\ -1 \end{bmatrix}.$$

Alternatively these column vectors are written as  $\mathbf{b}_1 = (1, -5/3)$ ,  $\mathbf{b}_2 = (-\sqrt{3}, \sqrt{5})$  and  $\mathbf{b}_3 = (\pi, -1)$ .

- Lastly, two matrices are **equal** ( $=$ ) if they both have the same size *and* their corresponding entries are equal. Otherwise the matrices are not equal. For example, consider matrices

$$A = \begin{bmatrix} 2 & \pi \\ 3 & 9 \end{bmatrix}, \quad B = \begin{bmatrix} \sqrt{4} & \pi \\ 2+1 & 3^2 \end{bmatrix},$$
$$C = \begin{bmatrix} 2 & \pi \end{bmatrix}, \quad D = \begin{bmatrix} 2 \\ \pi \end{bmatrix} = (2, \pi).$$

The matrices  $A = B$  because they are the same size and their corresponding entries are equal, such as  $a_{11} = 2 = \sqrt{4} = b_{11}$ . Matrix  $A$  cannot be equal to  $C$  because their sizes are different. Matrices  $C$  and  $D$  are not equal, despite having the same elements in the same order, because they have different sizes:  $1 \times 2$  and  $2 \times 1$  respectively.

**Activity 3.1.1.** Which of the following matrices equals  $\begin{bmatrix} 3 & -1 & 4 \\ -2 & 0 & 1 \end{bmatrix}$ ?

(a)  $\begin{bmatrix} 3 & -1 \\ 4 & -2 \\ 0 & 1 \end{bmatrix}$

(b)  $\begin{bmatrix} \sqrt{9} & -1 & 2^2 \\ -2 & 0 & \cos 0 \end{bmatrix}$

(c)  $\begin{bmatrix} 3 & -2 \\ -1 & 0 \\ 4 & 1 \end{bmatrix}$

(d)  $\begin{bmatrix} 3 & 1-2 & \sqrt{16} \\ 3-2 & 0 & e^0 \end{bmatrix}$



### 3.1.2 Addition, subtraction and multiplication with matrices

A matrix is not just an array of numbers: associated with a matrix is a suite of operations that empower a matrix to be useful in applications. We start with addition and multiplication: ‘division’ is addressed by [Section 3.2](#) and others.

An analogue in computing science is the concept of object orientated programming. In object oriented programming one defines not just data structures, but also the functions that operate on those structures. Analogously, an array is just a group of numbers, but a matrix is an array together with many operations explicitly available. The power and beauty of matrices results from the ramifications of its associated operations.

## Matrix addition and subtraction

Corresponding to vector addition and subtraction ([Definition 1.2.4](#)), matrix addition and subtraction is done component wise, but only between matrices of the same size.

**Example 3.1.2.** Let matrices

$$A = \begin{bmatrix} 4 & 0 \\ -5 & -4 \\ 0 & -3 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 & 2 \\ -3 & 0 & 3 \end{bmatrix}, \quad C = \begin{bmatrix} -4 & -1 \\ -4 & -1 \\ 1 & 4 \end{bmatrix},$$
$$D = \begin{bmatrix} -2 & -1 & -3 \\ 1 & 3 & 0 \end{bmatrix}, \quad E = \begin{bmatrix} 5 & -2 & -2 \\ 0 & -3 & 2 \\ -4 & 7 & -1 \end{bmatrix}.$$

Then the addition and subtraction

$$\begin{aligned} A + C &= \begin{bmatrix} 4 & 0 \\ -5 & -4 \\ 0 & -3 \end{bmatrix} + \begin{bmatrix} -4 & -1 \\ -4 & -1 \\ 1 & 4 \end{bmatrix} \\ &= \begin{bmatrix} 4 + (-4) & 0 + (-1) \\ -5 + (-4) & -4 + (-1) \\ 0 + 1 & -3 + 4 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ -9 & -5 \\ 1 & 1 \end{bmatrix}, \end{aligned}$$

$$\begin{aligned}
 B - D &= \begin{bmatrix} 1 & 0 & 2 \\ -3 & 0 & 3 \end{bmatrix} - \begin{bmatrix} -2 & -1 & -3 \\ 1 & 3 & 0 \end{bmatrix} \\
 &= \begin{bmatrix} 1 - (-2) & 0 - (-1) & 2 - (-3) \\ -3 - 1 & 0 - 3 & 3 - 0 \end{bmatrix} = \begin{bmatrix} 3 & 1 & 5 \\ -4 & -3 & 3 \end{bmatrix}.
 \end{aligned}$$

But because the matrices are of different sizes, the following are not defined and must not be attempted:  $A + B$ ,  $A - D$ ,  $E - A$ ,  $B + C$ ,  $E - C$ , for example. ■

In general, when  $A$  and  $B$  are both  $m \times n$  matrices, with entries  $a_{ij}$  and  $b_{ij}$  respectively, then we define their **sum** or **addition**,  $A + B$ , as the  $m \times n$  matrix whose  $(i, j)$ th entry is  $a_{ij} + b_{ij}$ . Similarly, define the **difference** or **subtraction**  $A - B$  as the  $m \times n$  matrix whose  $(i, j)$ th entry is  $a_{ij} - b_{ij}$ . That is,

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

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

Consequently, letting  $O$  denote the zero matrix of the appropriate size,

$$A \pm O = A, \quad O + A = A, \quad \text{and} \quad A - A = O.$$

**Activity 3.1.3.** Given the two matrices  $A = \begin{bmatrix} 3 & -2 \\ 1 & -1 \end{bmatrix}$  and  $B = \begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix}$ , which of the following is the matrix  $\begin{bmatrix} 5 & -1 \\ -2 & -3 \end{bmatrix}$ ?

- (a)  $A - B$       (b)  $A + B$       (c) none of the others      (d)  $B - A$





## Scalar multiplication of matrices

Corresponding to multiplication of a vector by a scalar ([Definition 1.2.4](#)), multiplication of a matrix by a scalar means that every entry of the matrix is multiplied by the scalar.

**Example 3.1.4.** Let the three matrices

$$A = \begin{bmatrix} 5 & 2 \\ -2 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 0 \\ -6 \end{bmatrix}, \quad C = \begin{bmatrix} 5 & -6 & 4 \\ -1 & -3 & -3 \end{bmatrix}.$$

Then the scalar multiplications

$$3A = \begin{bmatrix} 3 \cdot 5 & 3 \cdot 2 \\ 3 \cdot (-2) & 3 \cdot 3 \end{bmatrix} = \begin{bmatrix} 15 & 6 \\ -6 & 9 \end{bmatrix},$$

$$-B = (-1)B = \begin{bmatrix} (-1) \cdot 1 \\ (-1) \cdot 0 \\ (-1) \cdot (-6) \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 6 \end{bmatrix},$$

$$-\pi C = (-\pi)C = \begin{bmatrix} 5\pi & -6\pi & 4\pi \\ -\pi & -3\pi & -3\pi \end{bmatrix}.$$



In general, when  $A$  is an  $m \times n$  matrix, with entries  $a_{ij}$ , then we define the **scalar product** by  $c$ , either  $cA$  or  $Ac$ , as the  $m \times n$  matrix whose  $(i, j)$ th entry is  $ca_{ij}$ . That is,

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

## Matrix-vector multiplication transforms

Recall that the matrix-vector form of a system of linear equations, [Definition 2.2.2](#), wrote  $A\mathbf{x} = \mathbf{b}$ . In this form,  $A\mathbf{x}$  denotes a matrix-vector product. As implied by [Definition 2.2.2](#), we define the general **matrix-vector product**

$$A\mathbf{x} := \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix}$$

for  $m \times n$  matrix  $A$  and vector  $\mathbf{x}$  in  $\mathbb{R}^n$  with entries/components

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

This product is only defined when the number of columns of matrix  $A$  are the same as the number of components of vector  $\mathbf{x}$ . If not, then the product cannot be used.

**Example 3.1.5.** Let matrices

$$A = \begin{bmatrix} 3 & 2 \\ -2 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 5 & -6 & 4 \\ -1 & -3 & -3 \end{bmatrix},$$

and vectors  $\mathbf{x} = (2, -3)$  and  $\mathbf{b} = (1, 0, 4)$ . Then the matrix-vector products

$$A\mathbf{x} = \begin{bmatrix} 3 & 2 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ -3 \end{bmatrix} = \begin{bmatrix} 3 \cdot 2 + 2 \cdot (-3) \\ (-2) \cdot 2 + 1 \cdot (-3) \end{bmatrix} = \begin{bmatrix} 0 \\ -7 \end{bmatrix},$$

$$\begin{aligned} B\mathbf{b} &= \begin{bmatrix} 5 & -6 & 4 \\ -1 & -3 & -3 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 4 \end{bmatrix} \\ &= \begin{bmatrix} 5 \cdot 1 + (-6) \cdot 0 + 4 \cdot 4 \\ (-1) \cdot 1 + (-3) \cdot 0 + (-3) \cdot 4 \end{bmatrix} = \begin{bmatrix} 9 \\ -13 \end{bmatrix}. \end{aligned}$$

The combinations  $A\mathbf{b}$  and  $B\mathbf{x}$  are not defined as the number of columns of the matrices are not equal to the number of components in the vectors.

Further, we do not here define vector-matrix products such as  $\mathbf{x}A$  or  $\mathbf{b}B$ : the order of multiplication matters with matrices and so

these are not in the scope of the definition. ■

**Activity 3.1.6.** Which of the following is the result of the matrix-vector product  $\begin{bmatrix} 4 & 1 \\ 3 & -2 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix}$ ?

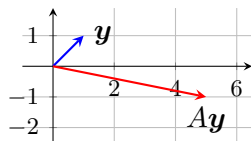
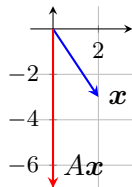
(a)  $\begin{bmatrix} 15 \\ 2 \end{bmatrix}$

(b)  $\begin{bmatrix} 21 \\ -2 \end{bmatrix}$

(c)  $\begin{bmatrix} 18 \\ -1 \end{bmatrix}$

(d)  $\begin{bmatrix} 14 \\ 5 \end{bmatrix}$

■



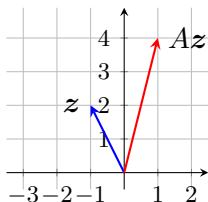
**Geometric interpretation** Multiplication of a vector by a *square matrix* transforms the vector into another in the same space. The margin shows the example of  $Ax$  from [Example 3.1.5](#). For another vector  $y = (1, 1)$  and the same matrix  $A$  the product

$$Ay = \begin{bmatrix} 3 & 2 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \cdot 1 + 2 \cdot 1 \\ (-2) \cdot 1 + 1 \cdot 1 \end{bmatrix} = \begin{bmatrix} 5 \\ -1 \end{bmatrix},$$

as illustrated in the second marginal picture. Similarly, for the

vector  $\mathbf{z} = (-1, 2)$  and the same matrix  $A$  the product

$$A\mathbf{z} = \begin{bmatrix} 3 & 2 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \cdot (-1) + 2 \cdot 2 \\ (-2) \cdot (-1) + 1 \cdot 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \end{bmatrix},$$



as illustrated in the third marginal picture. Such a geometric interpretation underlies the use of matrix multiplication in video and picture processing, for example. Such video/picture processing employs stretching and shrinking ([Subsection 3.2.2](#)), rotations ([Subsection 3.2.3](#)), among more general transformations ([Section 3.6](#)).

**Example 3.1.7.** Recall  $I_n$  is the  $n \times n$  identity matrix. Then the products

$$I_2 \mathbf{x} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ -3 \end{bmatrix} = \begin{bmatrix} 1 \cdot 2 + 0 \cdot (-3) \\ 0 \cdot 2 + 1 \cdot (-3) \end{bmatrix} = \begin{bmatrix} 2 \\ -3 \end{bmatrix},$$

$$I_3 \mathbf{b} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 \cdot 1 + 0 \cdot 0 + 0 \cdot 4 \\ 0 \cdot 1 + 1 \cdot 0 + 0 \cdot 4 \\ 0 \cdot 1 + 0 \cdot 0 + 1 \cdot 4 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 4 \end{bmatrix}.$$

That is, and justifying its name of “identity”, the products with an identity matrix give the result that is the vector itself:  $I_2 \mathbf{x} = \mathbf{x}$  and

$I_3 \mathbf{b} = \mathbf{b}$ . Multiplication by the identity matrix leaves the vector unchanged (Theorem 3.1.25e). ■



**Example 3.1.8** (rabbits multiply). In 1202 Fibonacci famously considered the breeding of rabbits—such as the following question. One pair of rabbits can give birth to another pair of rabbits (called kittens) every month, say. Each kitten becomes fertile after it has aged a month, when it becomes adult and is called a buck (male) or doe (female). The new bucks and does then also start breeding. How many rabbits are there after six months?

Fibonacci's real name is Leonardo Bonacci. He lived circa 1175 to 1250, travelled extensively from Pisa, and is considered to be one of the most talented Western mathematician of the Middle Ages.

Let's just count the females, the does, and the female kittens. At the start of any month let there be  $x_1$  kittens (female) and  $x_2$  does. Then at the end of the month:

- because all the female kittens grow up to be does, the number of does is now  $x'_2 = x_2 + x_1$ ;
- and because all the does at the start month have bred another pair of kittens, of which we expect one to be female, the new number of female kittens just born is  $x'_1 = x_2$ , on average.

Then  $x'_1$  and  $x'_2$  is the number of kittens and does at the start of the next month. Write this as a matrix vector system. Let the female population be  $\mathbf{x} = (x_1, x_2)$  and the population one month later be  $\mathbf{x}' = (x'_1, x'_2)$ . Then our model is that

$$\mathbf{x}' = \begin{bmatrix} x'_1 \\ x'_2 \end{bmatrix} = \begin{bmatrix} x_2 \\ x_1 + x_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = L\mathbf{x} \quad \text{for } L = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix},$$

called a Leslie matrix.

- At the start there is one adult pair, one doe, so the initial population is  $\mathbf{x} = (0, 1)$ .
- After one month, the does  $\mathbf{x}' = L\mathbf{x} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ .
- After two months, the does  $\mathbf{x}'' = L\mathbf{x}' = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ .
- After three months, the does  $\mathbf{x}''' = L\mathbf{x}'' = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ .
- After four months, the does  $\mathbf{x}^{iv} = L\mathbf{x}''' = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$ .



- After five months, the does  $\mathbf{x}^v = L\mathbf{x}^{iv} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \end{bmatrix} = \begin{bmatrix} 5 \\ 8 \end{bmatrix}$ .
- After six months, the does  $\mathbf{x}^{vi} = L\mathbf{x}^v = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 5 \\ 8 \end{bmatrix} = \begin{bmatrix} 8 \\ 13 \end{bmatrix}$ .

Fibonacci's model predicts the rabbit population grows rapidly according to the famous Fibonacci numbers 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, .... ■

**Example 3.1.9** (age structured population). An ecologist studies an isolated population of a species of animal. The growth of the population depends primarily upon the females so it is only these that are counted. The females are grouped into three ages: female pups (in their first year), juvenile females (one year old), and mature females (two years or older). During the study, the ecologist observes the following happens over the period of a year:

- half of the female pups survive and become juvenile females;
- one-third of the juvenile females survive and become mature

females;

- each mature female breeds and produces four female pups;
- one-third of the mature females survive to breed in the following year;
- female pups and juvenile females do not breed.

- (a) Let  $x_1$ ,  $x_2$  and  $x_3$  be the number of females at the start of a year, of ages zero, one and two+ respectively, and let  $x'_1$ ,  $x'_2$  and  $x'_3$  be their number at the start of the next year. Use the observations to write  $x'_1$ ,  $x'_2$  and  $x'_3$  as a function of  $x_1$ ,  $x_2$  and  $x_3$  (this is called a Markov chain).
- (b) Letting vectors  $\mathbf{x} = (x_1, x_2, x_3)$  and  $\mathbf{x}' = (x'_1, x'_2, x'_3)$  write down your function as the matrix-vector product  $\mathbf{x}' = L\mathbf{x}$  for some matrix  $L$  (called a Leslie matrix).
- (c) Suppose the ecologist observes the numbers of females at the start of a given year is  $\mathbf{x} = (60, 70, 20)$ , use your matrix to predict the numbers  $\mathbf{x}'$  at the start of the next year. Continue similarly to predict the numbers after two years ( $\mathbf{x}''$ )? and

three years ( $x'''$ )?



## Matrix-matrix multiplication

Matrix-vector multiplication explicitly uses the vector in its equivalent form as an  $n \times 1$  matrix—a matrix with one column. Such multiplication immediately generalises to the case of a right-hand matrix with multiple columns.

**Example 3.1.10.** Let two matrices

$$A = \begin{bmatrix} 3 & 2 \\ -2 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 5 & -6 & 4 \\ -1 & -3 & -3 \end{bmatrix},$$

then the matrix multiplication  $AB$  may be done as the matrix  $A$  multiplying each of the three columns in  $B$ . That is, in detail write

$$\begin{aligned} AB &= A \begin{bmatrix} 5 & -6 & 4 \\ -1 & -3 & -3 \end{bmatrix} \\ &= A \begin{bmatrix} 5 & \vdots & -6 & \vdots & 4 \\ -1 & \vdots & -3 & \vdots & -3 \end{bmatrix} \\ &= \left[ A \begin{bmatrix} 5 \\ -1 \end{bmatrix} : A \begin{bmatrix} -6 \\ -3 \end{bmatrix} : A \begin{bmatrix} 4 \\ -3 \end{bmatrix} \right] \end{aligned}$$

$$\begin{aligned} &= \left[ \begin{bmatrix} 13 \\ -11 \end{bmatrix} : \begin{bmatrix} -24 \\ 9 \end{bmatrix} : \begin{bmatrix} 6 \\ -11 \end{bmatrix} \right] \\ &= \begin{bmatrix} 13 & -24 & 6 \\ -11 & 9 & -11 \end{bmatrix}. \end{aligned}$$

Conversely, the product  $BA$  cannot be done because if we try the same procedure then

$$\begin{aligned} BA &= B \begin{bmatrix} 3 & 2 \\ -2 & 1 \end{bmatrix} \\ &= B \left[ \begin{bmatrix} 3 \\ -2 \end{bmatrix} : \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right] \\ &= \left[ B \begin{bmatrix} 3 \\ -2 \end{bmatrix} : B \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right], \end{aligned}$$

and neither of these matrix-vector products can be done as, for example,

$$B \begin{bmatrix} 3 \\ -2 \end{bmatrix} = \begin{bmatrix} 5 & -6 & 4 \\ -1 & -3 & -3 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \end{bmatrix}$$

the number of columns of the left matrix is not equal to the number of elements of the vector on the right. Hence the product  $BA$  is

not defined. ■

**Example 3.1.11.** Let matrices

$$C = \begin{bmatrix} -4 & -1 \\ -4 & -1 \\ 1 & 4 \end{bmatrix}, \quad D = \begin{bmatrix} -2 & -1 & -3 \\ 1 & 3 & 0 \end{bmatrix}.$$

Compute, if possible,  $CD$  and  $DC$ ; compare these products. ■

**Definition 3.1.12** (matrix product). *Let matrix  $A$  be  $m \times n$ , and matrix  $B$  be  $n \times p$ , then the **matrix product**  $C = AB$  is the  $m \times p$  matrix whose  $(i, j)$ th entry is*

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}.$$

This formula looks like a dot product ([Definition 1.3.2](#)) of two vectors : indeed we do use that the expression for the  $(i, j)$ th entry

is the dot product of the  $i$ th row of  $A$  and the  $j$ th column of  $B$  as illustrated by

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} b_{11} & \cdots & b_{1j} & \cdots & b_{1p} \\ b_{21} & \cdots & b_{2j} & \cdots & b_{2p} \\ \vdots & & \vdots & & \vdots \\ b_{n1} & \cdots & b_{nj} & \cdots & b_{np} \end{bmatrix}.$$

As seen in the examples, although the two matrices  $A$  and  $B$  may be of different sizes, the number of columns of  $A$  must equal the number of rows of  $B$  in order for the product  $AB$  to be defined.

**Activity 3.1.13.** Which one of the following matrix products is not defined?

(a)  $\begin{bmatrix} 3 & -1 \end{bmatrix} \begin{bmatrix} -3 & 1 \\ 7 & -3 \end{bmatrix}$

(b)  $\begin{bmatrix} -1 & 2 \\ 1 & -3 \end{bmatrix} \begin{bmatrix} -3 & -1 & -1 \\ 0 & -4 & -1 \end{bmatrix}$

(c)  $\begin{bmatrix} 8 & 9 & 3 \\ 2 & 5 & 1 \end{bmatrix} \begin{bmatrix} -2 & 8 \\ 3 & -2 \end{bmatrix}$

(d)  $\begin{bmatrix} 2 & 5 & -3 \end{bmatrix} \begin{bmatrix} -3 & 1 \\ -5 & -1 \\ 2 & -2 \end{bmatrix}$



**Example 3.1.14.** Matrix multiplication leads to powers of a square matrix.  
Let matrix

$$A = \begin{bmatrix} 3 & 2 \\ -2 & 1 \end{bmatrix},$$

then by  $A^2$  we mean the product

$$AA = \begin{bmatrix} 3 & 2 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 3 & 2 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 8 \\ -8 & -3 \end{bmatrix},$$

and by  $A^3$  we mean the product

$$AAA = AA^2 = \begin{bmatrix} 3 & 2 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 5 & 8 \\ -8 & -3 \end{bmatrix} = \begin{bmatrix} -1 & 18 \\ -18 & -19 \end{bmatrix},$$

and so on.





In general, for an  $n \times n$  square matrix  $A$  and a positive integer exponent  $p$  we define the **matrix power**

$$A^p = \underbrace{AA \cdots A}_{p \text{ factors}}.$$

The matrix powers  $A^p$  are also  $n \times n$  square matrices.

**Example 3.1.15** (age structured population). Matrix powers occur naturally in modelling populations by ecologists such as the animals of [Example 3.1.9](#). Recall that given the numbers of female pups, juveniles and mature aged formed into a vector  $\mathbf{x} = (x_1, x_2, x_3)$ , the number in each age one year later (indicated here by a dash) is  $\mathbf{x}' = L\mathbf{x}$  for Leslie matrix

$$L = \begin{bmatrix} 0 & 0 & 4 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} \end{bmatrix}.$$

Hence the number in each age category two years later (indicated here by two dashes) is

$$\mathbf{x}'' = L\mathbf{x}' = L(L\mathbf{x}) = (LL)\mathbf{x} = L^2\mathbf{x},$$

provided that matrix multiplication is associative (established by [Theorem 3.1.25c](#)) to enable us to write  $L(L\mathbf{x}) = (LL)\mathbf{x}$ . Then the matrix square

$$L^2 = \begin{bmatrix} 0 & 0 & 4 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 0 & 0 & 4 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} \end{bmatrix} = \begin{bmatrix} 0 & \frac{4}{3} & \frac{4}{3} \\ 0 & 0 & 2 \\ \frac{1}{6} & \frac{1}{9} & \frac{1}{9} \end{bmatrix}.$$

Continuing to use such associativity, the number in each age category three years later (indicated here by three dashes) is

$$\mathbf{x}''' = L\mathbf{x}'' = L(L^2\mathbf{x}) = (LL^2)\mathbf{x} = L^3\mathbf{x},$$

where the matrix cube

$$L^3 = LL^2 = \begin{bmatrix} 0 & 0 & 4 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 0 & \frac{4}{3} & \frac{4}{3} \\ 0 & 0 & 2 \\ \frac{1}{6} & \frac{1}{9} & \frac{1}{9} \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & \frac{4}{9} & \frac{4}{9} \\ 0 & \frac{2}{3} & \frac{2}{3} \\ \frac{1}{18} & \frac{1}{27} & \frac{19}{27} \end{bmatrix}.$$

That is, the powers of the Leslie matrix help predict what happens two, three, or more years into the future. ■

## The transpose of a matrix

The operations so far defined for matrices correspond directly to analogous operations for real numbers. The transpose has no corresponding analogue. At first mysterious, the transpose occurs frequently—often due to it linking the dot product of vectors with matrix multiplication. The transpose also reflects symmetry in applications ([Chapter 4](#)), such as Newton's law that every action has an equal and opposite reaction.

**Example 3.1.16.** Let matrices

$$A = \begin{bmatrix} -4 & 2 \\ -3 & 4 \\ -1 & -7 \end{bmatrix}, \quad B = [2 \quad 0 \quad -1], \quad C = \begin{bmatrix} 1 & 1 & 1 \\ -1 & -3 & 0 \\ 2 & 3 & 2 \end{bmatrix}.$$

Then obtain the transpose of each of these three matrices by writing each of their rows as columns, in order:

$$A^T = \begin{bmatrix} -4 & -3 & -1 \\ 2 & 4 & -7 \end{bmatrix}, \quad B^T = \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix}, \quad C^T = \begin{bmatrix} 1 & -1 & 2 \\ 1 & -3 & 3 \\ 1 & 0 & 2 \end{bmatrix}.$$



These examples illustrate the following definition.

**Definition 3.1.17** (transpose). *The **transpose** of an  $m \times n$  matrix  $A$  is the  $n \times m$  matrix, denoted  $A^T$ , obtained by writing the  $i$ th row of  $A$  as the  $i$ th column of  $A^T$ , or equivalently by writing the  $j$ th column of  $A$  to be the  $j$ th row of  $A^T$ . That is, if  $B = A^T$ , then  $b_{ij} = a_{ji}$ .*

**Activity 3.1.18.** Which of the following matrices is the transpose of the matrix

$$\begin{bmatrix} 1 & -0.5 & 2.9 \\ -1.4 & -1.4 & -0.2 \\ 0.9 & -2.3 & 1.6 \end{bmatrix} ?$$

(a)  $\begin{bmatrix} 0.9 & -2.3 & 1.6 \\ -1.4 & -1.4 & -0.2 \\ 1 & -0.5 & 2.9 \end{bmatrix}$

(b)  $\begin{bmatrix} 2.9 & -0.5 & 1 \\ -0.2 & -1.4 & -1.4 \\ 1.6 & -2.3 & 0.9 \end{bmatrix}$

$$(c) \begin{bmatrix} 1.6 & -2.3 & 0.9 \\ -0.2 & -1.4 & -1.4 \\ 2.9 & -0.5 & 1 \end{bmatrix} \quad (d) \begin{bmatrix} 1 & -1.4 & 0.9 \\ -0.5 & -1.4 & -2.3 \\ 2.9 & -0.2 & 1.6 \end{bmatrix}$$



**Example 3.1.19** (transpose and dot product). Consider two vectors in  $\mathbb{R}^n$ , say  $\mathbf{u} = (u_1, u_2, \dots, u_n)$  and  $\mathbf{v} = (v_1, v_2, \dots, v_n)$ ; that is,

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}.$$

Then the dot product between the two vectors

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n \quad (\text{Defn. 1.3.2 of dot})$$

$$\begin{aligned}
&= \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \quad (\text{Defn. 3.1.12 of mult.}) \\
&= \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}^T \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \quad (\text{transpose Defn. 3.1.17}) \\
&= \mathbf{u}^T \mathbf{v}.
\end{aligned}$$

Subsequent sections and chapters often use this identity, that the dot product  $\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v}$ . ■

**Definition 3.1.20** (symmetry). *A (real) matrix  $A$  is a **symmetric matrix** if  $A^T = A$ ; that is, if the matrix is equal to its transpose.*

A symmetric matrix must be a square matrix—as otherwise the sizes of  $A$  and  $A^T$  would be different and so the matrices could not be equal.

**Example 3.1.21.** None of the three matrices in [Example 3.1.16](#) are symmetric: the first two matrices are not square so cannot be symmetric, and the third matrix  $C \neq C^T$ . The following matrix is symmetric:

$$D = \begin{bmatrix} 2 & 0 & 1 \\ 0 & -6 & 3 \\ 1 & 3 & 4 \end{bmatrix} = D^T.$$

When is the following general  $2 \times 2$  matrix symmetric?

$$E = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$



Symmetric matrices of note are the  $n \times n$  identity matrix and  $n \times n$  zero matrix,  $I_n$  and  $O_n$ .

**Activity 3.1.22.** Which one of the following matrices is a symmetric matrix?

- (a)  $\begin{bmatrix} 2.3 & -1.3 & -2 \\ -3.2 & -1 & -1.3 \\ -3 & -3.2 & 2.3 \end{bmatrix}$
- (b)  $\begin{bmatrix} -2.6 & 0.3 & -1.3 \\ 0.3 & -0.2 & 0 \\ -1.3 & 0 & -2 \end{bmatrix}$
- (c)  $\begin{bmatrix} 2.2 & -0.9 & -1.2 \\ -0.9 & -1.2 & -3.1 \end{bmatrix}$
- (d)  $\begin{bmatrix} 0 & -3.2 & -0.8 \\ 3.2 & 0 & 3.2 \\ 0.8 & -3.2 & 0 \end{bmatrix}$





## Compute in Matlab/Octave

MATLAB/Octave empowers us to compute all these operations quickly, especially for the large problems found in applications: after all, MATLAB is an abbreviation of *Matrix Laboratory*. [Table 3.1](#) summarises the MATLAB/Octave version of the operations introduced so far, and used in the rest of this book.

**Matrix size and elements** Let the matrix

$$A = \begin{bmatrix} 0 & 0 & -2 & -11 & 5 \\ 0 & 1 & -1 & 11 & -8 \\ -4 & 2 & 10 & 2 & -3 \end{bmatrix}.$$

We readily see this is a  $3 \times 5$  matrix, but to check that MATLAB/Octave agrees, execute the following in MATLAB/Octave:

```
A=[0 0 -2 -11 5
    0 1 -1 11 -8
    -4 2 10 2 -3]
size(A)
```



Table 3.1: As well as the basics of MATLAB/Octave listed in [Table 1.2](#) and [2.3](#), we need these matrix operations.

- `size(A)` returns the number of rows and columns of matrix  $A$ : if  $A$  is  $m \times n$ , then `size(A)` returns  $[m \ n]$ .
- `A(i,j)` is the  $(i, j)$ th entry of a matrix  $A$ , `A(:,j)` is the  $j$ th column, `A(i,:)` is the  $i$ th row; either to use the value(s) or to assign value(s).
- `+, -, *` is matrix/vector/scalar addition, subtraction, and multiplication, but only provided the sizes of the two operands are compatible.
- `A^p` for scalar  $p$  computes the  $p$ th power of square matrix  $A$  (in contrast to `A.^p` which computes the  $p$ th power of *each element* of  $A$ , [Table 2.3](#)).
- The character single quote, `A'`, transposes the matrix  $A$ . But when using complex numbers be wary: `A'` is the complex conjugate transpose (which is what we usually want); whereas `A.'` is the transpose without complex conjugation ([Subsection 7.1.5](#)).
- Predefined matrices include:
  - `zeros(m,n)` is the zero matrix  $O_{m \times n}$ ;
  - `eye(m,n)` is  $m \times n$  ‘identity matrix’  $I_{m \times n}$ ;
  - `ones(m,n)` is the  $m \times n$  matrix where all entries are one;

The answer, “3 5”, confirms  $A$  is  $3 \times 5$ . MATLAB/Octave accesses individual elements, rows and columns. For example, execute each of the following:

- $A(2,4)$  gives  $a_{24}$  which here results in 11;
- $A(:,5)$  is the the fifth column vector, here  $\begin{bmatrix} 5 \\ -8 \\ -3 \end{bmatrix}$ ;
- $A(1,:)$  is the first row, here  $[0 \ 0 \ -2 \ -11 \ 5]$ .

One may also use these constructs to change the elements in matrix  $A$ : for example, executing  $A(2,4)=9$  changes matrix  $A$  to

$$A = \begin{bmatrix} 0 & 0 & -2 & -11 & 5 \\ 0 & 1 & -1 & 9 & -8 \\ -4 & 2 & 10 & 2 & -3 \end{bmatrix}$$

then  $A(:,5)=[2;-3;1]$  changes matrix  $A$  to

$$A =$$

0	0	-2	-11	2
0	1	-1	9	-3
-4	2	10	2	1

whereas  $A(1,:) = [1 \ 2 \ 3 \ 4 \ 5]$  changes matrix  $A$  to

$A =$

1	2	3	4	5
0	1	-1	9	-3
-4	2	10	2	1

**Matrix addition and subtraction** To illustrate further operations let's use some random matrices generated by MATLAB/Octave: you will generate different matrices to the following, but the operations will work the same. [Table 3.1](#) mentions that `randn(m)` and `randn(m,n)` generate random matrices so execute say

`A=randn(4)`

`B=randn(4)`

`C=randn(4,2)`

and obtain matrices such as (2 d.p.)



A =

-1.31	2.07	0.08	2.05
1.25	-1.35	-1.00	1.94
1.08	1.79	-0.99	0.93
1.34	-0.99	-0.23	-0.22

B =

1.21	-0.46	0.09	0.58
1.67	-1.96	1.26	1.93
0.24	-0.46	2.77	-0.59
0.03	-0.28	-0.76	0.13

C =

1.14	0.85
-0.48	0.17
0.37	-0.64
0.62	-1.17

Then A+B gives here the sum

ans =

-0.10	1.62	0.17	2.63
2.92	-3.31	0.26	3.87

```
1.31    1.33    1.78    0.34
1.37   -1.27   -0.99   -0.09
```

and A-B the difference

```
ans =
-2.52    2.53   -0.01    1.46
-0.41    0.62   -2.25    0.01
 0.84    2.26   -3.76    1.52
 1.31   -0.71    0.53   -0.35
```

You could check that  $B+A$  gives the same matrix as  $A+B$  ([Theorem 3.1.23a](#)) by seeing that their difference is the  $3 \times 5$  zero matrix: execute  $(A+B)-(B+A)$  (the parentheses control the order of evaluation). However, expressions such as  $B+C$  and  $A-C$  give an error, because the matrices are of incompatible sizes, reported by MATLAB as

```
Error using +
Matrix dimensions must agree.
```

or reported by Octave as

error: operator +: nonconformant arguments

**Scalar multiplication of matrices** In MATLAB/Octave the asterisk indicates multiplication. Scalar multiplication can be done either way around. For example, generate a random  $4 \times 3$  matrix  $A$  and compute  $2A$  and  $A\frac{1}{10}$ . These commands

```
A=randn(4,3)
```

```
2*A
```

```
A*0.1
```

might give the following (2 d.p.)

```
A =
```

0.82	2.54	-0.98
2.30	0.05	2.63
-1.45	2.15	0.89
-2.58	-0.09	-0.55

```
>> 2*A
```

```
ans =
```



```
1.64    5.07   -1.97
4.61    0.10    5.25
-2.90    4.30    1.77
-5.16   -0.18   -1.11
```

```
>> A*0.1
ans =
0.08    0.25   -0.10
0.23    0.00    0.26
-0.15    0.21    0.09
-0.26   -0.01   -0.06
```

Division by a scalar is also defined in MATLAB/Octave and means multiplication by the reciprocal; for example, the product `A*0.1` could equally well be computed as `A/10`.

In mathematical algebra we would not normally accept statements such as  $A + 3$  or  $2A - 5$  because addition and subtraction with matrices has only been defined between matrices of the same size. However, MATLAB/Octave usefully extends addition and subtraction so that `A+3` and `2*A-5` mean add three to *every* element of  $A$



and subtract five from *every* element of  $2A$ . For example, with the above random  $4 \times 3$  matrix  $A$ ,

```
>> A+3
ans =
    3.82    5.54    2.02
    5.30    3.05    5.63
    1.55    5.15    3.89
    0.42    2.91    2.45

>> 2*A-5
ans =
   -3.36    0.07   -6.97
   -0.39   -4.90    0.25
   -7.90   -0.70   -3.23
  -10.16   -5.18   -6.11
```

This last computation illustrates that in any expression the operations of multiplication and division are performed before additions and subtractions—as normal in mathematics.

**Matrix multiplication** In MATLAB/Octave the asterisk also invokes matrix-matrix and matrix-vector multiplication. For example, generate and multiply two random matrices say of size  $3 \times 4$  and  $4 \times 2$  with

```
A=randn(3,4)
```

```
B=randn(4,2)
```

```
C=A*B
```

might give the following result (2 d.p.)

```
A =
```

```
-0.02    1.31   -0.74   -0.49  
-0.36   -1.30   -0.23    0.41  
-0.88   -0.34    0.28   -0.99
```

```
B =
```

```
-1.32   -0.79  
 0.71    1.48  
-0.48    2.79  
 1.40   -0.41
```



```
>> C=A*B
C =
    0.62    0.10
    0.24   -2.44
   -0.60    1.38
```

Without going into excruciating arithmetic detail this product is hard to check. However, we can check several things such as  $c_{11}$  comes from the first row of  $A$  times the first column of  $B$  by computing  $A(1,:)*B(:,1)$  and seeing it does give 0.62 as required. Also check that the two columns of  $C$  may be viewed as the two matrix-vector products  $A\mathbf{b}_1$  and  $A\mathbf{b}_2$  by comparing  $C$  with  $[A*B(:,1) \ A*B(:,2)]$  and seeing they are the same.

Recall that in a matrix product the number of columns of the left matrix have to be the same as the number of rows of the right matrix. MATLAB/Octave gives an error message if this is not the case, such as occurs upon asking it to compute  $B*A$  when MATLAB reports

```
Error using *  
Inner matrix dimensions must agree.
```

and Octave reports

```
error: operator *: nonconformant arguments
```

The caret symbol,  $\wedge$ , computes matrix powers in MATLAB/Octave, such as the cube  $A^3$ . But such matrix powers only makes sense and works for square matrices  $A$ . For example, if matrix  $A$  was  $3 \times 4$ , then  $A^2 = AA$  would involve multiplying a  $3 \times 4$  matrix by a  $3 \times 4$  matrix: since the number of columns of the left  $A$  is not the same as the number of rows of the right  $A$  such a multiplication is not allowed.

**The transpose and symmetry** In MATLAB/Octave the single apostrophe denotes matrix transpose. For example, see it transpose a couple of random matrices with

```
A=randn(3,4)  
B=randn(4,2)  
A'
```



B'

giving here for example (2 d.p.)

A =

0.80	0.30	-0.12	-0.57
0.07	-0.51	-0.81	1.95
0.29	-0.10	0.17	0.70

B =

-0.71	-0.34
-0.33	-0.73
1.11	-0.21
0.41	0.33

>> A'

ans =

0.80	0.07	0.29
0.30	-0.51	-0.10
-0.12	-0.81	0.17
-0.57	1.95	0.70

```
>> B'
ans =
    -0.71    -0.33     1.11     0.41
    -0.34    -0.73    -0.21     0.33
```

One can do further operations after the transposition, such as checking the multiplication rule that  $(AB)^T = B^T A^T$  ([Theorem 3.1.28d](#)) by verifying the result of  $(A*B)' - B'*A'$  is the zero matrix, here  $O_{2 \times 3}$ .

You can generate a symmetric matrix by adding a square matrix to its transpose ([Theorem 3.1.28f](#)): for example, generate a random square matrix by first  $C=\text{randn}(3)$  then  $C=C+C'$  makes a random symmetric matrix such as the following (2 d.p.)

```
>> C=randn(3)
C =
    -0.33     0.65    -0.62
    -0.43    -2.18    -0.28
     1.86    -1.00    -0.52

>> C=C+C'
```

```
C =  
-0.65    0.22    1.24  
 0.22   -4.36   -1.28  
 1.24   -1.28   -1.04
```

```
>> C-C'  
ans =  
 0.00    0.00    0.00  
 0.00    0.00    0.00  
 0.00    0.00    0.00
```

That the resulting matrix  $C$  is symmetric is checked by this last step which computes the difference between  $C$  and  $C^T$  and confirming the difference is zero. Hence  $C$  and  $C^T$  must be equal.

### 3.1.3 Familiar algebraic properties of matrix operations

Almost all of the familiar algebraic properties of scalar addition, subtraction and multiplication—namely commutativity, associativity and distributivity—hold for matrix addition, subtraction and multiplication.

The one outstanding exception is that matrix multiplication is *not* commutative: for matrices  $A$  and  $B$  the products  $AB$  and  $BA$  are usually not equal. We are used to such non-commutativity in life. For example, when you go home, to enter your house you first open the door, second walk in, and third close the door. You cannot swap the order and try to walk in before opening the door—these operations do not commute. Similarly, for another example, I often teach classes on the third floor of a building next to my office: after finishing classes, first I walk downstairs to ground level, and second I cross the road to my office. If I try to cross the road before going downstairs, then the force of gravity has something very painful to say about the outcome—the operations do not commute. Similar to these analogues, the result of a matrix multiplication depends upon the order of the matrices in the multiplication.



**Theorem 3.1.23** (Properties of addition and scalar multiplication). *Let matrices  $A$ ,  $B$  and  $C$  be of the same size, and let  $c$  and  $d$  be scalars. Then:*

- (a)  $A + B = B + A$  (commutativity of addition);
- (b)  $(A + B) + C = A + (B + C)$  (associativity of addition);
- (c)  $A \pm O = A = O + A$ ;
- (d)  $c(A \pm B) = cA \pm cB$  (distributivity over matrix addition);
- (e)  $(c \pm d)A = cA \pm dA$  (distributivity over scalar addition);
- (f)  $c(dA) = (cd)A$  (associativity of scalar multiplication);
- (g)  $1A = A$ ; and
- (h)  $0A = O$ .

**Example 3.1.24** (geometry of associativity). Many properties of matrix multiplication have a useful geometric interpretation such as that discussed for matrix-vector products. Recall the earlier [Example 3.1.15](#) invoked the associativity [Theorem 3.1.25c](#). For another

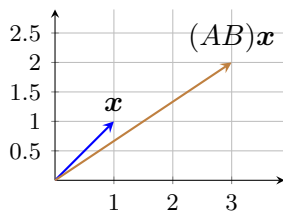
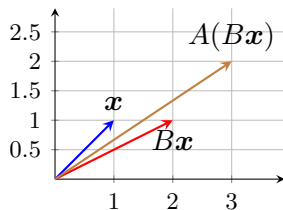
example, consider the two matrices and vector

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 0 \\ 2 & -1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Now the transform  $\mathbf{x}' = B\mathbf{x} = (2, 1)$ , and then transforming with  $A$  gives  $\mathbf{x}'' = A\mathbf{x}' = A(B\mathbf{x}) = (3, 2)$ , as illustrated in the margin. This is the same results as forming the product

$$AB = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 2 & -1 \end{bmatrix} = \begin{bmatrix} 4 & -1 \\ 2 & 0 \end{bmatrix}$$

and then computing  $(AB)\mathbf{x} = (3, 2)$  as also illustrated in the margin. Such associativity asserts that  $A(B\mathbf{x}) = (AB)\mathbf{x}$ : that is, the geometric transform of  $\mathbf{x}$  by matrix  $B$  followed by the transform of matrix  $A$  is the same result as just transforming by the matrix formed from the product  $AB$ —as assured by [Theorem 3.1.25c](#). ■



**Theorem 3.1.25** (properties of matrix multiplication). *Let matrices  $A$ ,  $B$  and  $C$  be of sizes such that the following expressions are defined, and let  $c$  be a scalar, then:*

- (a)  $A(B \pm C) = AB \pm AC$  (distributivity of matrix multiplication);
- (b)  $(A \pm B)C = AC \pm BC$  (distributivity of matrix multiplication);
- (c)  $A(BC) = (AB)C$  (associativity of matrix multiplication);
- (d)  $c(AB) = (cA)B = A(cB)$ ;
- (e)  $I_m A = A = A I_n$  for  $m \times n$  matrix  $A$  (multiplicative identity);
- (f)  $O_m A = O_{m \times n} = A O_n$  for  $m \times n$  matrix  $A$ ;
- (g)  $A^p A^q = A^{p+q}$ ,  $(A^p)^q = A^{pq}$  and  $(cA)^p = c^p A^p$  for square  $A$  and for positive integers  $p$  and  $q$ .

**Example 3.1.26.** Show that  $(A + B)^2 \neq A^2 + 2AB + B^2$  in general. ■

**Example 3.1.27.** Show that the matrix  $J = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$  is not a multiplicative identity (despite having ones down a diagonal, this diagonal is the wrong one for an identity). ■

**Theorem 3.1.28** (properties of transpose). *Let matrices  $A$  and  $B$  be of sizes such that the following expressions are defined, then:*

- (a)  $(A^T)^T = A$ ;
- (b)  $(A \pm B)^T = A^T \pm B^T$ ;
- (c)  $(cA)^T = c(A^T)$  for any scalar  $c$ ;
- (d)  $(AB)^T = B^T A^T$ ;
- (e)  $(A^p)^T = (A^T)^p$  for all positive integer exponents  $p$ ;
- (f)  $A + A^T$ ,  $A^T A$  and  $AA^T$  are symmetric matrices.

*Remember the reversed order in the identity  $(AB)^T = B^T A^T$ .*

## 3.2 The inverse of a matrix

### Section Contents

3.2.1	Introducing the unique inverse . . . . .	230
3.2.2	Diagonal matrices stretch and shrink . . . . .	243
	Solve systems whose matrix is diagonal . . . . .	247
	But do not divide by zero . . . . .	252
	Stretch or squash the unit square . . . . .	253
	Sketch convenient coordinates . . . . .	258
3.2.3	Orthogonal matrices rotate . . . . .	261
	Orthogonal set of vectors . . . . .	263
	Orthogonal matrices . . . . .	266

The previous [Section 3.1](#) introduced addition, subtraction, multiplication, and other operations of matrices. Conspicuously missing

from the list is ‘division’ by a matrix. This section develops ‘division’ by a matrix as multiplication by the inverse of a matrix. The analogue in ordinary arithmetic is that division by ten is the same as multiplying by its reciprocal, one-tenth. But the inverse of a matrix looks nothing like a reciprocal.

### 3.2.1 Introducing the unique inverse

Let's start with an example that illustrates an analogy with the reciprocal/inverse of a scalar number.

**Example 3.2.1.** Recall that a crucial property is that a number multiplied by its reciprocal/inverse is one: for example,  $2 \times 0.5 = 1$  so 0.5 is the reciprocal/inverse of 2. Similarly, show that matrix

$$B = \begin{bmatrix} -3 & 1 \\ -4 & 1 \end{bmatrix} \text{ is an inverse of } A = \begin{bmatrix} 1 & -1 \\ 4 & -3 \end{bmatrix}$$

by showing their product is the  $2 \times 2$  identity matrix  $I_2$ . ■

The previous [Example 3.2.1](#) shows at least one case when we can do some sort of matrix ‘division’: that is, multiplying by  $B$  is equivalent to ‘dividing’ by  $A$ . One restriction is that a clearly defined ‘division’ only works for square matrices. Part of the reason is because we need to be able to compute both  $AB$  and  $BA$ .

**Definition 3.2.2** (inverse). For every  $n \times n$  square matrix  $A$ , an **inverse** of  $A$  is an  $n \times n$  matrix  $B$  such that both  $AB = I_n$  and  $BA = I_n$ . If such a matrix  $B$  exists, then matrix  $A$  is called **invertible**.

(By saying “an inverse” this definition allows for many inverses, but [Theorem 3.2.6](#) establishes that the inverse is unique.)

**Example 3.2.3.** Show that matrix

$$B = \begin{bmatrix} 0 & -\frac{1}{4} & -\frac{1}{8} \\ \frac{3}{2} & 1 & \frac{7}{8} \\ \frac{1}{2} & \frac{1}{4} & \frac{3}{8} \end{bmatrix} \text{ is an inverse of } A = \begin{bmatrix} 1 & -1 & 5 \\ -5 & -1 & 3 \\ 2 & 2 & -6 \end{bmatrix}.$$





- Activity 3.2.4.** What value of  $b$  makes the matrix  $\begin{bmatrix} -1 & b \\ 1 & 2 \end{bmatrix}$  to be the inverse of  $\begin{bmatrix} 2 & 3 \\ -1 & -1 \end{bmatrix}$ ?
- (a) 1                      (b) 3                      (c) -2                      (d) -3



But even among square matrices, there are many non-zero matrices which do not have an inverse! A matrix which is not invertible is sometimes called a **singular matrix**. The next [Section 3.3](#) further explores why some matrices do not have an inverse: the reason is associated with both `rcond` being zero ([Procedure 2.2.5](#)) and/or the so-called determinant being zero ([Chapter 6](#)).

**Example 3.2.5** (no inverse). Prove that the matrix

$$A = \begin{bmatrix} 1 & -2 \\ -3 & 6 \end{bmatrix}$$

does not have an inverse. ■

**Theorem 3.2.6** (unique inverse). *If  $A$  is an invertible matrix, then its inverse is unique (and denoted by  $A^{-1}$ ).*

In the elementary case of  $1 \times 1$  matrices, that is  $A = [a_{11}]$ , the inverse is simply the reciprocal of the entry, that is  $A^{-1} = [1/a_{11}]$  provided  $a_{11}$  is non-zero. The reason is that  $AA^{-1} = [a_{11} \cdot \frac{1}{a_{11}}] = [1] = I_1$  and  $A^{-1}A = [\frac{1}{a_{11}} \cdot a_{11}] = [1] = I_1$ .

In the case of  $2 \times 2$  matrices the inverse is a little more complicated, but should be remembered. (For larger sized matrices, any direct general formulas for an inverse are too complicated to remember.)

**Theorem 3.2.7** ( $2 \times 2$  inverse). *Let  $2 \times 2$  matrix  $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ . Then  $A$  is invertible if the **determinant**  $ad - bc \neq 0$ , in which case*

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}. \quad (3.2)$$

*If the determinant  $ad - bc = 0$ , then  $A$  is not invertible.*

**Example 3.2.8.** (a) Recall that [Example 3.2.1](#) verified that

$$B = \begin{bmatrix} -3 & 1 \\ -4 & 1 \end{bmatrix} \text{ is an inverse of } A = \begin{bmatrix} 1 & -1 \\ 4 & -3 \end{bmatrix}.$$

Formula [\(3.2\)](#) gives this inverse from the matrix  $A$ : its elements are  $a = 1$ ,  $b = -1$ ,  $c = 4$  and  $d = -3$  so the determinant  $ad - bc = 1 \cdot (-3) - (-1) \cdot 4 = 1$  and hence formula [\(3.2\)](#) derives the inverse

$$A^{-1} = \frac{1}{1} \begin{bmatrix} -3 & -(-1) \\ -4 & 1 \end{bmatrix} = \begin{bmatrix} -3 & 1 \\ -4 & 1 \end{bmatrix} = B.$$

(b) Further, recall [Example 3.2.5](#) proved there is no inverse for matrix

$$A = \begin{bmatrix} 1 & -2 \\ -3 & 6 \end{bmatrix}.$$

[Theorem 3.2.7](#) also establishes this matrix is not invertible because the matrix determinant  $ad - bc = 1 \cdot 6 - (-2) \cdot (-3) = 6 - 6 = 0$ .

**Activity 3.2.9.** Which of the following matrices is invertible?

(a)  $\begin{bmatrix} -2 & 1 \\ 4 & -2 \end{bmatrix}$

(b)  $\begin{bmatrix} -2 & -1 & 4 \\ 3 & 1 & 1 \end{bmatrix}$

(c)  $\begin{bmatrix} 0 & -3 \\ 4 & -2 \end{bmatrix}$

(d)  $\begin{bmatrix} -4 & -2 \\ 2 & 2 \\ -3 & 1 \end{bmatrix}$

Almost anything you can do with  $A^{-1}$  can be done without it.

*G. E. Forsythe and C. B. Moler, 1967 ([Higham 1996](#), p.261)*

**Computer considerations** Except for easy cases such as  $2 \times 2$  matrices, we rarely explicitly compute the inverse of a matrix. Computationally there are (almost) always better ways such as the MATLAB/Octave operation  $A \backslash \mathbf{b}$  of [Procedure 2.2.5](#). The inverse is a crucial theoretical device, rarely a practical computational tool.

The following [Theorem 3.2.10](#) is an example: for a system of linear equations the theorem connects the existence of a unique solution to the invertibility of the matrix of coefficients. Further, [Subsection 3.3.2](#) connects solutions to the `rcond` invoked by [Procedure 2.2.5](#). Although in theoretical statements we write expressions like  $\mathbf{x} = A^{-1}\mathbf{b}$ , practically, once we know a solution exists (`rcond` is acceptable), we generally compute a solution without ever constructing  $A^{-1}$ .

**Theorem 3.2.10.** *If  $A$  is an invertible  $n \times n$  matrix, then the system of linear equations  $A\mathbf{x} = \mathbf{b}$  has the unique solution  $\mathbf{x} = A^{-1}\mathbf{b}$  for every  $\mathbf{b}$  in  $\mathbb{R}^n$ .*

One consequence is the following: if a system of linear equations has no solution or an infinite number of solutions ([Theorem 2.2.27](#)),

then this theorem establishes that the matrix of the system is not invertible.

**Example 3.2.11.** Use the matrices of Examples 3.2.1, 3.2.3 and 3.2.5 to decide whether each of the following systems have a unique solution, or not.

$$\begin{array}{ll} \text{(a)} \quad \begin{cases} x - y = 4, \\ 4x - 3y = 3. \end{cases} & \text{(b)} \quad \begin{cases} u - v + 5w = 2, \\ -5u - v + 3w = 5, \\ 2u + 2v - 6w = 1. \end{cases} \\ \text{(c)} \quad \begin{cases} r - 2s = -1, \\ -3r + 6s = 3. \end{cases} & \end{array}$$

**Example 3.2.12.** Given the following information about solutions of systems of linear equations, write down if the matrix associated with each system is invertible, or not, or there is not enough given information to decide. Give reasons.

- (a) A general solution is  $(1, -5, 0, 3)$ .  
(b) A general solution is  $(3, -5 + 3t, 3 - t, -1)$ .  
(c) A solution of a system is  $(-3/2, -2, -\pi, 2, -4)$ .  
(d) A solution of a homogeneous system is  $(1, 2, -8)$ .



Recall from [Section 3.1](#) the properties of scalar multiplication, matrix powers, transpose, and their computation ([Table 3.1](#)). The next theorem incorporates the inverse into this suite of properties.

**Theorem 3.2.13** (properties of the inverse). *Let  $A$  and  $B$  be invertible matrices of the same size, then:*

- (a) *matrix  $A^{-1}$  is invertible and  $(A^{-1})^{-1} = A$ ;*  
(b) *if scalar  $c \neq 0$ , then matrix  $cA$  is invertible and  $(cA)^{-1} = \frac{1}{c}A^{-1}$ ;*  
(c) *matrix  $AB$  is invertible and  $(AB)^{-1} = B^{-1}A^{-1}$ ;*

*Remember the reversed order in the identity  $(AB)^{-1} = B^{-1}A^{-1}$ .*

- (d) matrix  $A^T$  is invertible and  $(A^T)^{-1} = (A^{-1})^T$ ;
- (e) matrices  $A^p$  are invertible for all  $p = 1, 2, 3, \dots$  and  $(A^p)^{-1} = (A^{-1})^p$ .

**Activity 3.2.14.** The matrix  $\begin{bmatrix} 3 & -5 \\ 4 & -7 \end{bmatrix}$  has inverse  $\begin{bmatrix} 7 & -5 \\ 4 & -3 \end{bmatrix}$ .

- What is the inverse of the matrix  $\begin{bmatrix} 6 & -10 \\ 8 & -14 \end{bmatrix}$ ?

(a)  $\begin{bmatrix} 14 & -10 \\ 8 & -3 \end{bmatrix}$

(b)  $\begin{bmatrix} 3.5 & 2 \\ -2.5 & -1.5 \end{bmatrix}$

(c)  $\begin{bmatrix} 7 & 4 \\ -5 & -3 \end{bmatrix}$

(d)  $\begin{bmatrix} 3.5 & -2.5 \\ 2 & -1.5 \end{bmatrix}$

- Further, which of the above is the inverse of  $\begin{bmatrix} 3 & 4 \\ -5 & -7 \end{bmatrix}$ ?





**Definition 3.2.15** (non-positive powers). *For every invertible matrix  $A$ , define  $A^0 := I$  and for every positive integer  $p$  define  $A^{-p} := (A^{-1})^p$  (or by [Theorem 3.2.13e](#) equivalently as  $(A^p)^{-1}$ ).*

**Example 3.2.16.** Recall from [Example 3.2.1](#) that matrix

$$A = \begin{bmatrix} 1 & -1 \\ 4 & -3 \end{bmatrix} \text{ has inverse } A^{-1} = \begin{bmatrix} -3 & 1 \\ -4 & 1 \end{bmatrix}.$$

Compute  $A^{-2}$  and  $A^{-4}$ . ■

**Activity 3.2.17.** [Example 3.2.16](#) gives the inverse of a matrix  $A$  and determines  $A^{-2}$ : what is  $A^{-3}$ ?

(a)  $\begin{bmatrix} -7 & 3 \\ -12 & 5 \end{bmatrix}$

(b)  $\begin{bmatrix} 3 & -7 \\ 5 & -12 \end{bmatrix}$

(c)  $\begin{bmatrix} -7 & -12 \\ 3 & 5 \end{bmatrix}$

(d)  $\begin{bmatrix} 3 & 5 \\ -7 & -12 \end{bmatrix}$  ■

**Example 3.2.18** (predict the past). Recall [Example 3.1.9](#) introduced how to use a Leslie matrix to predict the future population of an animal. If  $\mathbf{x} = (60, 70, 20)$  is the current number of pups, juveniles, and mature females respectively, then by the modelling the predicted population numbers after a year is  $\mathbf{x}' = L\mathbf{x}$ , after two years is  $\mathbf{x}'' = L\mathbf{x}' = L^2\mathbf{x}$ , and so on. In these formulas, and for this example, the Leslie matrix

$$L = \begin{bmatrix} 0 & 0 & 4 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} \end{bmatrix}, \quad \text{which has inverse } L^{-1} = \begin{bmatrix} 0 & 2 & 0 \\ -\frac{1}{4} & 0 & 3 \\ \frac{1}{4} & 0 & 0 \end{bmatrix}.$$

Assume the same rule applies for earlier years.

- Letting the population numbers a year ago be denoted by  $\mathbf{x}^-$  then by the modelling the current population  $\mathbf{x} = L\mathbf{x}^-$ . Multiply by the inverse of  $L$ :  $L^{-1}\mathbf{x} = L^{-1}L\mathbf{x}^- = \mathbf{x}^-$ ; that is, the population a year before the current is  $\mathbf{x}^- = L^{-1}\mathbf{x}$ .
- Similarly, letting the population numbers two years ago be denoted by  $\mathbf{x}^=$  then by the modelling  $\mathbf{x}^- = L\mathbf{x}^=$  and multiplication by  $L^{-1}$  gives  $\mathbf{x}^= = L^{-1}\mathbf{x}^- = L^{-1}L^{-1}\mathbf{x} = L^{-2}\mathbf{x}$ .

- One more year earlier, letting the population numbers two years ago be denoted by  $\mathbf{x}^{\equiv}$  then by the modelling  $\mathbf{x}^{\equiv} = L\mathbf{x}^{\equiv}$  and multiplication by  $L^{-1}$  gives  $\mathbf{x}^{\equiv} = L^{-1}\mathbf{x}^{\equiv} = L^{-1}L^{-2}\mathbf{x} = L^{-3}\mathbf{x}$ .

Hence use the inverse powers of  $L$  to predict the earlier history of the population of female animals in the given example: but first verify the given inverse is correct. ■

**Example 3.2.19.** As an alternative to the hand calculations of [Example 3.2.18](#), predict earlier populations by computing in MATLAB/Octave without ever explicitly finding the inverse or powers of the inverse. The procedure is to solve the linear system  $L\mathbf{x}^{-} = \mathbf{x}$  for the population  $\mathbf{x}^{-}$  a year ago, and then similarly solve  $L\mathbf{x}^{\equiv} = \mathbf{x}^{-}$ ,  $L\mathbf{x}^{\equiv} = \mathbf{x}^{\equiv}$ , and so on. ■

### 3.2.2 Diagonal matrices stretch and shrink

Recall that the identity matrices are zero except for a diagonal of ones from the top-left to the bottom-right of the matrix. Because of the nature of matrix multiplication it is this diagonal that is special. Because of the special nature of this diagonal, this section explores matrices which are zero except for the numbers (not generally ones) in the top-left to bottom-right diagonal.

**Example 3.2.20.** That is, this section explores the nature of so-called diagonal matrices such as

$$\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}, \quad \begin{bmatrix} 0.58 & 0 & 0 \\ 0 & -1.61 & 0 \\ 0 & 0 & 2.17 \end{bmatrix}, \quad \begin{bmatrix} \pi & 0 & 0 \\ 0 & \sqrt{3} & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

We use the term diagonal matrix to also include non-square matrices such as

$$\begin{bmatrix} -\sqrt{2} & 0 \\ 0 & \frac{1}{2} \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \pi & 0 & 0 & 0 \\ 0 & 0 & e & 0 & 0 \end{bmatrix}.$$

The term diagonal matrix does *not* describe matrices such as

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 2 & 0 \\ -\frac{1}{2} & 0 & 0 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} -0.17 & 0 & 0 & 0 \\ 0 & -4.22 & 0 & 0 \\ 0 & 0 & 0 & 3.05 \end{bmatrix}.$$



Amazingly, the singular value decomposition of [Section 3.3](#) proves that diagonal matrices lie at the very heart of the action of *every* matrix.

**Definition 3.2.21** (diagonal matrix). *For every  $m \times n$  matrix  $A$ , the **diagonal entries** of  $A$  are  $a_{11}, a_{22}, \dots, a_{pp}$  where  $p = \min(m, n)$ . A matrix whose non-diagonal entries are all zero is called a **diagonal matrix**.*

*For brevity we may write  $\text{diag}(v_1, v_2, \dots, v_n)$  to denote the  $n \times n$  square matrix with diagonal entries  $v_1, v_2, \dots, v_n$ , or  $\text{diag}_{m \times n}(v_1, v_2, \dots, v_p)$  for an  $m \times n$  matrix with diagonal entries  $v_1, v_2, \dots, v_p$ .*

**Example 3.2.22.** The five diagonal matrices of [Example 3.2.20](#) could equivalently be written as  $\text{diag}(3, 2)$ ,  $\text{diag}(0.58, -1.61, 2.17)$ ,  $\text{diag}(\pi, \sqrt{3}, 0)$ ,  $\text{diag}_{3 \times 2}(-\sqrt{2}, \frac{1}{2})$  and  $\text{diag}_{3 \times 5}(1, \pi, e)$ , respectively. ■

Diagonal matrices may also have zeros on the diagonal, as well as the required zeros for the non-diagonal entries.

**Activity 3.2.23.** Which of the following matrices are not diagonal?

(a) 
$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

(b) 
$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

(c)  $O_n$

(d)  $I_n$

■

Table 3.2: As well as the basics of MATLAB/Octave listed in Tables 1.2, 2.3 and 3.1, we need these matrix operations.

- **diag(v)** where **v** is a row/column vector of length  $p$  generates the  $p \times p$  matrix

$$\text{diag}(v_1, v_2, \dots, v_p) = \begin{bmatrix} v_1 & 0 & \cdots & 0 \\ 0 & v_2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & v_p \end{bmatrix}.$$

- In MATLAB/Octave (but not usually in algebra), **diag** also does the opposite: for an  $m \times n$  matrix  $A$  such that both  $m, n \geq 2$ , **diag(A)** returns the (column) vector  $(a_{11}, a_{22}, \dots, a_{pp})$  of diagonal entries where the result vector length  $p = \min(m, n)$ .
- The dot operators **./** and **.\*** do element-by-element division and multiplication of two matrices/vectors of the same size. For example,  
 $[5 \ 14 \ 33] ./ [5 \ 7 \ 3] = [1 \ 2 \ 11]$
- [Section 3.5](#) also needs to compute the logarithm of data: **log10(v)** finds the logarithm to base 10 of each component of **v** and returns the results in a vector of the same size;

## Solve systems whose matrix is diagonal

Solving a system of linear equations ([Definition 2.1.2](#)) is particularly straightforward when the matrix of the system is diagonal. Indeed much mathematics in both theory and applications is devoted to transforming a given problem so that the matrix appearing in the system is diagonal (e.g., sections [2.2.2](#) and [3.3.2](#), and Chapters [4](#) and [7](#)).

**Example 3.2.24.** Solve

$$\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ -5 \end{bmatrix}$$



**Example 3.2.25.** Solve

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{2}{3} & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$



**Activity 3.2.26.** What is the solution to the system  $\begin{bmatrix} 0.4 & 0 \\ 0 & 0.1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 0.1 \\ -0.2 \end{bmatrix}$ ?

(a)  $(\frac{1}{4}, -2)$       (b)  $(4, -\frac{1}{2})$       (c)  $(4, -2)$       (d)  $(\frac{1}{4}, -\frac{1}{2})$

**Theorem 3.2.27** (inverse of diagonal matrix). *For every  $n \times n$  diagonal matrix  $D = \text{diag}(d_1, d_2, \dots, d_n)$ , if all the diagonal entries are nonzero,  $d_i \neq 0$  for  $i = 1, 2, \dots, n$ , then  $D$  is invertible and the inverse  $D^{-1} = \text{diag}(1/d_1, 1/d_2, \dots, 1/d_n)$ .*

**Example 3.2.28.** The previous [Example 3.2.25](#) gave the inverse of a  $3 \times 3$  matrix. For the  $2 \times 2$  matrix  $D = \text{diag}(3, 2) = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$  the inverse is  $D^{-1} = \text{diag}(\frac{1}{3}, \frac{1}{2}) = \begin{bmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$ . Then the solution to

$$\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 2 \\ -5 \end{bmatrix} \quad \text{is } \mathbf{x} = \begin{bmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 2 \\ -5 \end{bmatrix} = \begin{bmatrix} 2/3 \\ -5/2 \end{bmatrix}.$$

■

**Compute in Matlab/Octave.** To solve the matrix-vector equation  $D\mathbf{x} = \mathbf{b}$  recognise that this equation means

$$\begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} d_1 x_1 \\ d_2 x_2 \\ \vdots \\ d_n x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

$$\Leftrightarrow \begin{cases} d_1 x_1 = b_1 \\ d_2 x_2 = b_2 \\ \vdots \\ d_n x_n = b_n \end{cases} \Leftrightarrow \begin{cases} x_1 = b_1/d_1 \\ x_2 = b_2/d_2 \\ \vdots \\ x_n = b_n/d_n \end{cases} \quad (3.3)$$

- Suppose you have a column vector **d** of the diagonal entries of *D* and a column vector **b** of the RHS; then compute a solution by, for example,

```
d=[2;2/3;-1]
b=[1;2;3]
x=b./d
```

to find the answer `[0.5;3;-3]`. Here the MATLAB/Octave operation `./` does element-by-element division ([Table 3.2](#)).

- When you have the diagonal matrix in full: extract the diagonal elements into a column vector with `diag()` ([Table 3.2](#)); then execute the element-by-element division; for example,

```
D=[2 0 0;0 2/3 0;0 0 -1]
b=[1;2;3]
```



```
x=b./diag(D)
```

**But do not divide by zero**

Dividing by zero is almost always nonsense. Instead use reasoning. Consider solving  $D\mathbf{x} = \mathbf{b}$  for diagonal  $D = \text{diag}(d_1, d_2, \dots, d_n)$  where  $d_n = 0$  (and similarly for any others that are zero). From (3.3) we need to solve  $d_n x_n = b_n$ , which here is  $0 \cdot x_n = b_n$ , that is,  $0 = b_n$ . There are two cases:

- if  $b_n \neq 0$ , then there is no solution; conversely
- if  $b_n = 0$ , then there is an infinite number of solutions as any  $x_n$  satisfies  $0 \cdot x_n = 0$ .

**Example 3.2.29.** Solve the two systems (the only difference is the last component on the RHS)

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{2}{3} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{2}{3} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$$

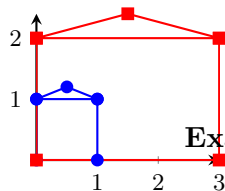


## Stretch or squash the unit square

Equations are just the boring part of mathematics. I attempt to see things in terms of geometry.

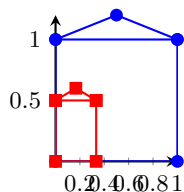
*Stephen Hawking, 2005*

Multiplication by matrices transforms shapes: multiplication by diagonal matrices just stretches or squashes and/or reflects in the direction of the coordinate axes. The next [Subsection 3.2.3](#) introduces matrices that rotate.



**Example 3.2.30.**

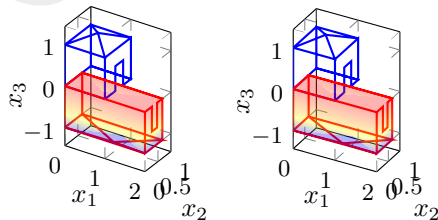
Consider  $A = \text{diag}(3, 2) = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$ . The marginal pictures



shows this matrix stretches the (blue) unit square (drawn with a ‘roof’) by a factor of three horizontally and two vertically (to the red). Recall that  $(x_1, x_2)$  denotes the corresponding column vector. As seen in the corner points of the graphic in the margin,  $A \times (1, 0) = (3, 0)$ ,  $A \times (0, 1) = (0, 2)$ ,  $A \times (0, 0) = (0, 0)$ , and  $A \times (1, 1) = (3, 2)$ . The ‘roof’ just helps us to track which corner goes where.

The inverse  $A^{-1} = \text{diag}(\frac{1}{3}, \frac{1}{2}) = \begin{bmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$  undoes the stretching of the matrix  $A$  by squashing in both the horizontal and vertical directions (from blue to red). ■

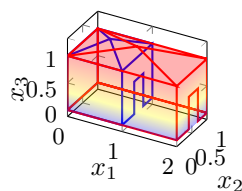
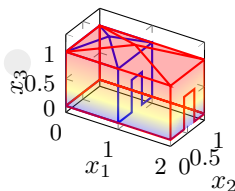
**Example 3.2.31.** Consider  $\text{diag}(2, \frac{2}{3}, -1) = \begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{2}{3} & 0 \\ 0 & 0 & -1 \end{bmatrix}$ : the stereo pair below illustrates how this diagonal matrix stretches in one direction, squashes in another, and reflects in the vertical. By multiplying the matrix by corner vectors  $(1, 0, 0)$ ,  $(0, 1, 0)$ ,  $(0, 0, 1)$ , and so on, we see that the blue unit cube (with ‘roof’ and ‘door’) maps to the red.



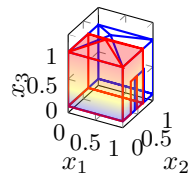
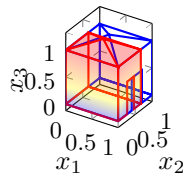


One great aspect of a diagonal matrix is that it is easy to separate its effects into each coordinate direction. For example, the above  $3 \times 3$  matrix is the same as the combined effects of the following three.

$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ . Stretch by a factor of two in the  $x_1$  direction.

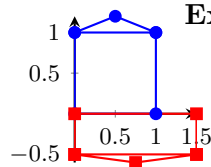
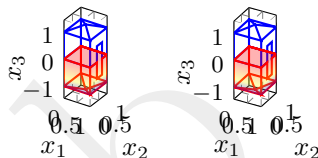


$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{2}{3} & 0 \\ 0 & 0 & 1 \end{bmatrix}$ . Squash by a factor of  $2/3$  in the  $x_2$  direction.



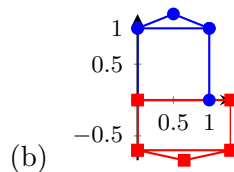
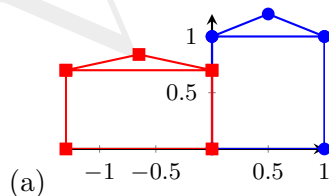


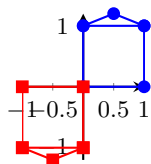
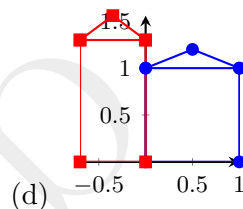
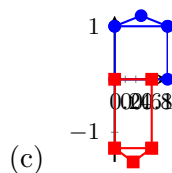
$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$
 Reflect in  
 the vertical  $x_3$   
 direction.



**Example 3.2.32.** What diagonal matrix transforms the blue unit square to the red in the illustration in the margin? ■

**Activity 3.2.33.** Which of the following diagrams represents the transformation from the (blue) unit square to the (red) rectangle by the matrix  $\text{diag}(-1.3, 0.7)$ ?





**Some diagonal matrices rotate** Now consider the transformation of multiplying by matrix  $\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$ : the two reflections of this diagonal matrix, the two  $-1$ s, have the same effect as one rotation, here by  $180^\circ$ , as shown to the left. Matrices that rotate are incredibly useful and is the topic of the next [Subsection 3.2.3](#).

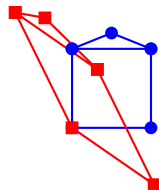
## Sketch convenient coordinates

This optional subsubsection is a preliminary to diagonalisation.

One of the fundamental principles of applying mathematics in science and engineering is that the real world—nature—does its thing irrespective of our mathematical description. Hence we often simplify our mathematical description of real world applications by choosing a coordinate system to suit its nature. That is, although this book (almost) always draws the  $x$  or  $x_1$  axis horizontally, and the  $y$  or  $x_2$  axis vertically, in applications it is often better to draw the axes in some other directions—directions which are convenient for the application. This example illustrates the principle.

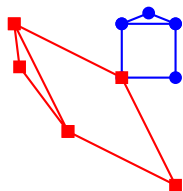
### Example 3.2.34.

Consider the transformation shown in the margin (it might arise from the deformation of some material and we need to know the internal stretching and shrinking to predict failure). The drawing has no coordinate axes shown because it is supposed to be some transformation in nature. Now we impose on nature our mathematical description. Draw approximate coordinate axes, with origin at the common point at the lower-left corner, so the transformation becomes that of the diagonal matrix  $\text{diag}(\frac{1}{2}, 2) =$



$$\begin{bmatrix} 0.5 & 0 \\ 0 & 2 \end{bmatrix}.$$

■

**Example 3.2.35.**

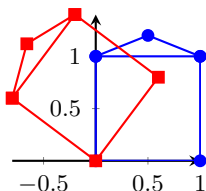
Consider the transformation shown in the margin. It has no coordinate axes shown because it supposed to be some transformation in nature. Now impose on nature our mathematical description. Draw approximate coordinate axes, with origin at the common corner point, so the transformation becomes that of the diagonal matrix  $\text{diag}(3, -1) = \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix}$ .

■

Finding such coordinate systems in which a given real world transformation is diagonal is important in science, engineering, and computer science. Systematic methods for such diagonalisation are developed in [Section 3.3](#), and Chapters [4](#) and [7](#). These rely

on understanding the algebra and geometry of not only diagonal matrices, but also rotations, which is our next topic.

### 3.2.3 Orthogonal matrices rotate



Whereas diagonal matrices stretch and squash, the so-called ‘orthogonal matrices’ represent just rotations (and/or reflection). For example, this section shows that multiplying by the ‘orthogonal matrix’  $\begin{bmatrix} 3/5 & -4/5 \\ 4/5 & 3/5 \end{bmatrix}$  rotates by  $53.13^\circ$  as shown in the marginal picture. Orthogonal matrices are the best to compute with, such as to solve linear equations, since they all have `rcond` = 1. To see these and related marvellous properties, we must invoke the geometry of lengths and angles.

**Recall the dot product determines lengths and angles** [Section 1.3](#) introduced the dot product between two vectors ([Definition 1.3.2](#)). For any two vectors in  $\mathbb{R}^n$ ,  $\mathbf{u} = (u_1, \dots, u_n)$  and  $\mathbf{v} = (v_1, \dots, v_n)$ , define the **dot product**

$$\begin{aligned} \mathbf{u} \cdot \mathbf{v} &= (u_1, \dots, u_n) \cdot (v_1, \dots, v_n) \\ &= u_1 v_1 + u_2 v_2 + \dots + u_n v_n. \end{aligned}$$

Considering the two vectors as column matrices, the dot product is the same as the matrix product ([Example 3.1.19](#))

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v} = \mathbf{v}^T \mathbf{u} = \mathbf{v} \cdot \mathbf{u}.$$

Also ([Theorem 1.3.17a](#)), the length of a vector  $\mathbf{v} = (v_1, v_2, \dots, v_n)$  in  $\mathbb{R}^n$  is the real number

$$|\mathbf{v}| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{v_1^2 + v_2^2 + \dots + v_n^2},$$

and that unit vectors are those of length one. For two non-zero vectors  $\mathbf{u}, \mathbf{v}$  in  $\mathbb{R}^n$ , [Theorem 1.3.5](#) defines the angle  $\theta$  between the vectors via

$$\cos \theta = \frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{u}||\mathbf{v}|}, \quad 0 \leq \theta \leq \pi.$$

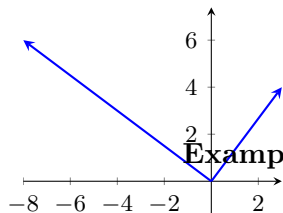
If the two vectors are at right-angles, then the dot product is zero and the two vectors are termed orthogonal ([Definition 1.3.19](#)).

## Orthogonal set of vectors

We need sets of orthogonal vectors (non-zero vectors which are all at right-angles to each other). One example is the set of standard unit vectors  $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$  aligned with the coordinate axes in  $\mathbb{R}^n$ .

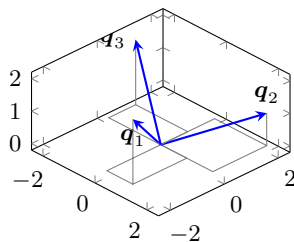
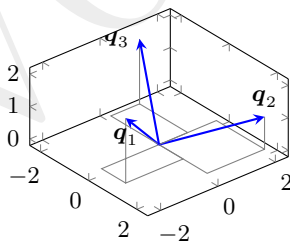
### Example 3.2.36.

The set of two vectors  $\{(3, 4), (-8, 6)\}$  shown in the margin is an orthogonal set as the two vectors have dot product  $= 3 \cdot (-8) + 4 \cdot 6 = -24 + 24 = 0$ . ■



### Example 3.2.37.

Let vectors  $\mathbf{q}_1 = (1, -2, 2)$ ,  $\mathbf{q}_2 = (2, 2, 1)$  and  $\mathbf{q}_3 = (-2, 1, 2)$ , illustrated in stereo below. Is  $\{\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3\}$  an orthogonal set?





**Definition 3.2.38.** A set of non-zero vectors  $\{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_k\}$  in  $\mathbb{R}^n$  is called an **orthogonal set** if all pairs of distinct vectors in the set are orthogonal: that is,  $\mathbf{q}_i \cdot \mathbf{q}_j = 0$  whenever  $i \neq j$  for  $i, j = 1, 2, \dots, k$ . A set of vectors in  $\mathbb{R}^n$  is called an **orthonormal set** if it is an orthogonal set of unit vectors. ■

A single non-zero vector always forms an orthogonal set. A single unit vector always forms an orthonormal set.

**Example 3.2.39.** Any set, or subset, of standard unit vectors in  $\mathbb{R}^n$  (Definition 1.2.7) are an orthonormal set as they are all at right-angles (orthogonal), and all of length one. ■

**Example 3.2.40.** Let vectors  $\mathbf{q}_1 = (\frac{1}{3}, -\frac{2}{3}, \frac{2}{3})$ ,  $\mathbf{q}_2 = (\frac{2}{3}, \frac{2}{3}, \frac{1}{3})$ ,  $\mathbf{q}_3 = (-\frac{2}{3}, \frac{1}{3}, \frac{2}{3})$ . Show the set  $\{\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3\}$  is an orthonormal set. ■

**Activity 3.2.41.** Which one of the following sets of vectors is *not* an orthogonal set?

(a)  $\{(-5, 4)\}$

(b)  $\{(-2, 3), (6, 4)\}$

(c)  $\{(2, 3), (4, -1)\}$

(d)  $\{\mathbf{i}, \mathbf{k}\}$



## Orthogonal matrices

**Example 3.2.42.** Example 3.2.36 showed  $\{(3, 4), (-8, 6)\}$  is an orthogonal set. The vectors have lengths five and ten, respectively, so dividing each by their length means  $\{(\frac{3}{5}, \frac{4}{5}), (-\frac{4}{5}, \frac{3}{5})\}$  is an orthonormal set. Form the matrix  $Q$  with these two vectors as its columns:

$$Q = \begin{bmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{bmatrix}.$$

Then consider

$$Q^T Q = \begin{bmatrix} \frac{3}{5} & \frac{4}{5} \\ -\frac{4}{5} & \frac{3}{5} \end{bmatrix} \begin{bmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{bmatrix} = \begin{bmatrix} \frac{9+16}{25} & \frac{-12+12}{25} \\ \frac{-12+12}{25} & \frac{16+9}{25} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Similarly  $QQ^T = I_2$ . Consequently, the transpose  $Q^T$  is here the inverse of  $Q$  (Definition 3.2.2). The transpose being the inverse is no accident.

Also no accident is that multiplication by this  $Q$  gives the rotation illustrated at the start of this section, (§3.2.3). ■

**Definition 3.2.43** (orthogonal matrices). A square  $n \times n$  matrix  $Q$  is called an **orthogonal matrix** if  $Q^T Q = I_n$ . Because of its special properties ([Theorem 3.2.48](#)), multiplication by an orthogonal matrix is called a **rotation and/or reflection**; for brevity and depending upon the circumstances it may be called just a **rotation** or just a **reflection**.

**Activity 3.2.44.** For which of the following values of  $p$  is the matrix

$$Q = \begin{bmatrix} \frac{1}{2} & p \\ -p & \frac{1}{2} \end{bmatrix} \text{ orthogonal?}$$

(a) some other value

(b)  $p = 3/4$

(c)  $p = -1/2$

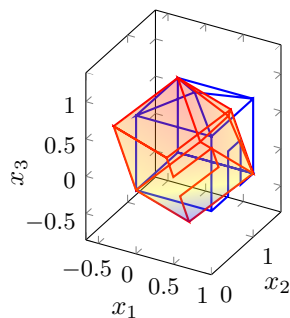
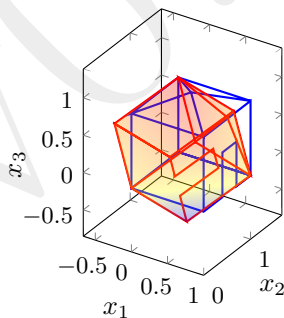
(d)  $p = \sqrt{3}/2$



**Example 3.2.45.** In the following equation, check that the matrix is orthogonal, and hence solve the equation  $Q\mathbf{x} = \mathbf{b}$ :

$$\begin{bmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\ -\frac{2}{3} & \frac{1}{3} & \frac{2}{3} \end{bmatrix} \mathbf{x} = \begin{bmatrix} 0 \\ 2 \\ -1 \end{bmatrix}.$$

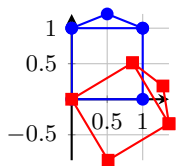
The stereo pair below illustrates the rotation of the unit cube under multiplication by the matrix  $Q$ : every point  $\mathbf{x}$  in the (blue) unit cube, is mapped to the point  $Q\mathbf{x}$  to form the (red) result.



**Example 3.2.46.** Given the matrix is orthogonal, solve the linear equation

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1 \\ -1 \\ 1 \\ 3 \end{bmatrix}.$$

**Example 3.2.47.** The marginal graph shows a rotation of the unit square. From the graph estimate roughly the matrix  $Q$  such that multiplication by  $Q$  performs the rotation. Confirm that your estimated matrix is orthogonal (approximately).



Because orthogonal matrices represent rotations, they arise frequently in engineering and scientific mechanics of bodies. Also,

the ease in solving equations with orthogonal matrices puts orthogonal matrices at the heart of coding and decoding photographs (jpeg), videos (mpeg), signals (Fourier transforms), and so on. Furthermore, an extension of orthogonal matrices to complex valued matrices, the so-called unitary matrices, is at the core of quantum physics and quantum computing. Moreover, the next [Section 3.3](#) establishes that orthogonal matrices express the orientation of the action of *every* matrix and hence are a vital component of solving linear equations in general. But to utilise orthogonal matrices across the wide range of applications we need to establish the following properties.

**Theorem 3.2.48.** *For every square matrix  $Q$ , the following statements are equivalent:*

- (a)  *$Q$  is an orthogonal matrix;*
- (b) *the column vectors of  $Q$  form an orthonormal set;*
- (c)  *$Q$  is invertible and  $Q^{-1} = Q^T$ ;*
- (d)  *$Q^T$  is an orthogonal matrix;*

- (e) *the row vectors of  $Q$  form an orthonormal set;*
- (f) *multiplication by  $Q$  preserves all lengths and angles (and hence corresponds to our intuition of a rotation and/or reflection).*

Another important property, proved by ??, is that the product of orthogonal matrices is also an orthogonal matrix.

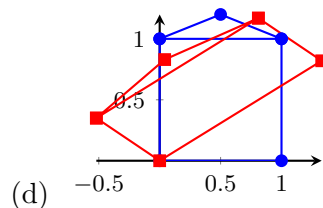
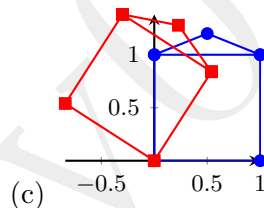
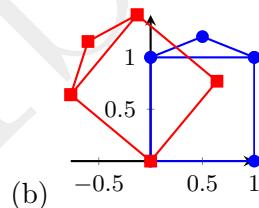
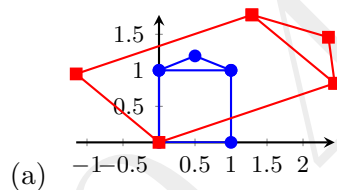
**Example 3.2.49.** Show that these matrices are orthogonal and hence write down their inverses:

$$\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}.$$



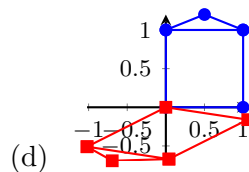
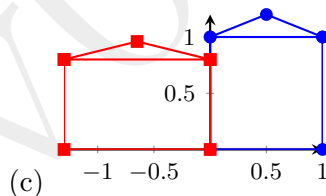
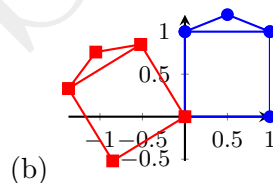
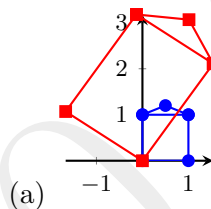


**Example 3.2.50.** The following graphs illustrate the transformation of the unit square through multiplying by some different matrices. Using [Theorem 3.2.48f](#), which transformations appear to be that of multiplying by an orthogonal matrix?



**Activity 3.2.51.** The following graphs illustrate the transformation of the unit square through multiplying by some different matrices.

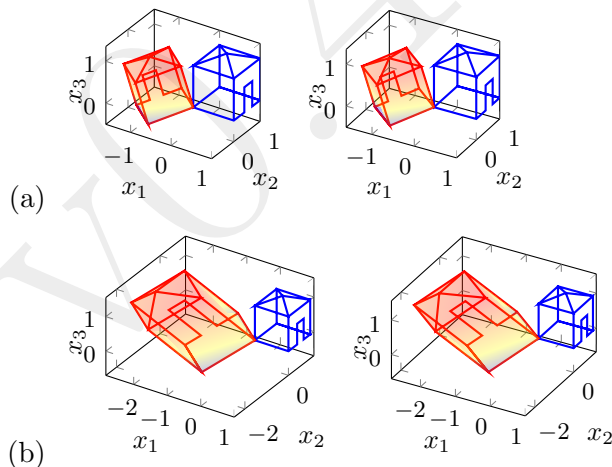
- Which transformation appears to be that of multiplying by an orthogonal matrix?

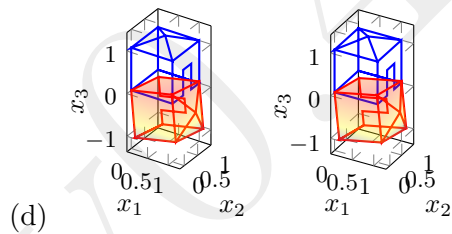
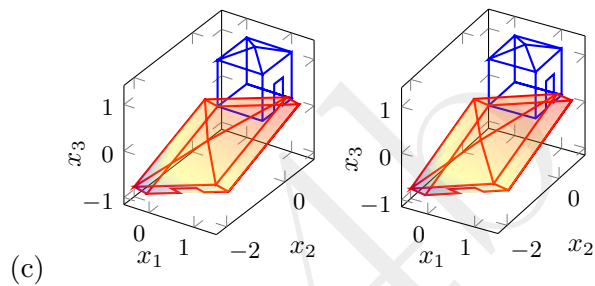


- Further, which of the above transformations appear to be that of multiplying by a diagonal matrix?



**Example 3.2.52.** The following stereo pairs illustrate the transformation of the unit cube through multiplying by some different matrices: using [Theorem 3.2.48f](#), which transformations appear to be that of multiplying by an orthogonal matrix?





## 3.3 Factorise to the singular value decomposition

### Section Contents

3.3.1	Introductory examples . . . . .	278
3.3.2	The SVD solves general systems . . . . .	285
	Computers empower use of the SVD . . . . .	290
	Condition number and rank determine the possibilities . . . . .	295
3.3.3	Prove the SVD Theorem 3.3.6 . . . . .	313
	Prelude to the proof . . . . .	314
	Detailed proof of the SVD Theorem 3.3.6 . .	317

Beltrami first derived the SVD in 1873. The first reliable method for computing an SVD was developed by Golub and Kahan in 1965, and only thereafter did applications proliferate.

The singular value decomposition (SVD) is sometimes called the jewel in the crown of linear algebra. Its importance is certified by the many names by which it is invoked in scientific and engineering applications: principal component analysis, singular spectrum analysis, principal orthogonal decomposition, latent semantic indexing,

Schmidt decomposition, correspondence analysis, Lanczos methods, dimension reduction, and so on. Let's start seeing what it can do for us.

### 3.3.1 Introductory examples

Let's introduce an analogous problem so the SVD procedure follows more easily.

You are a contestant in a quiz show. The final million dollar question is:

in your head, without a calculator, solve  $42x = 1554$   
within twenty seconds,

your time starts now .....

**Activity 3.3.1.** Given  $154 = 2 \cdot 7 \cdot 11$ , solve in your head  $154x = 8008$  or 9856 or 12628 or 13090 or 14322 (teacher to choose): first to answer wins. ■

Such examples show factorisation can turn a hard problem into several easy problems. We adopt an analogous matrix factorisation to solve and understand general linear equations.

To illustrate the procedure to come, let's write the above solution steps in detail: we solve  $42x = 1554$ .

1. Factorise the coefficient  $42 = 2 \cdot 3 \cdot 7$  so the equation becomes

$$2 \cdot \underbrace{3 \cdot \overbrace{7 \cdot x}^{=y}}_{=z} = 1554,$$

and introduce two intermediate unknowns  $y$  and  $z$  as indicated above.

2. Solve  $2z = 1554$  to get  $z = 777$ .
3. Solve  $3y = z = 777$  to get  $y = 259$ .
4. Solve  $7x = y = 259$  to get  $x = 37$  —the answer.

Now let's proceed to small matrix examples. These introduce the general matrix procedure empowered by the factorisation of a matrix to its singular value decomposition (SVD).

**Example 3.3.2.** Solve the  $2 \times 2$  system

$$\begin{bmatrix} 10 & 2 \\ 5 & 11 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 18 \\ -1 \end{bmatrix}$$



given the matrix factorisation

$$\begin{bmatrix} 10 & 2 \\ 5 & 11 \end{bmatrix} = \begin{bmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{bmatrix} \begin{bmatrix} 10\sqrt{2} & 0 \\ 0 & 5\sqrt{2} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}^T$$

(note the transpose on the last matrix). ■

**Activity 3.3.3.** Let's solve the system  $\begin{bmatrix} 12 & -41 \\ 34 & -12 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 94 \\ 58 \end{bmatrix}$  using the factorisation

$$\begin{bmatrix} 12 & -41 \\ 34 & -12 \end{bmatrix} = \begin{bmatrix} \frac{4}{5} & -\frac{3}{5} \\ \frac{3}{5} & \frac{4}{5} \end{bmatrix} \begin{bmatrix} 50 & 0 \\ 0 & 25 \end{bmatrix} \begin{bmatrix} \frac{3}{5} & \frac{4}{5} \\ -\frac{4}{5} & \frac{3}{5} \end{bmatrix}^T$$

in which the first and third matrices on the right-hand side are orthogonal. After solving  $\begin{bmatrix} \frac{4}{5} & -\frac{3}{5} \\ \frac{3}{5} & \frac{4}{5} \end{bmatrix} \mathbf{z} = \begin{bmatrix} 94 \\ 58 \end{bmatrix}$ , the next step is to solve which of the following?

$$(a) \begin{bmatrix} 50 & 0 \\ 0 & 25 \end{bmatrix} \mathbf{y} = \begin{bmatrix} 110 \\ -10 \end{bmatrix}$$

$$(b) \begin{bmatrix} 50 & 0 \\ 0 & 25 \end{bmatrix} \mathbf{y} = \begin{bmatrix} \frac{202}{5} \\ \frac{514}{5} \end{bmatrix}$$

$$(c) \begin{bmatrix} 50 & 0 \\ 0 & 25 \end{bmatrix} \mathbf{y} = \begin{bmatrix} 10 \\ 110 \end{bmatrix}$$

$$(d) \begin{bmatrix} 50 & 0 \\ 0 & 25 \end{bmatrix} \mathbf{y} = \begin{bmatrix} \frac{514}{5} \\ -\frac{202}{5} \end{bmatrix}$$

**Example 3.3.4.**Solve the  $3 \times 3$  system

$$A\mathbf{x} = \begin{bmatrix} 10 \\ 2 \\ -2 \end{bmatrix} \quad \text{for matrix } A = \begin{bmatrix} -4 & -2 & 4 \\ -8 & -1 & -4 \\ 6 & 6 & 0 \end{bmatrix}$$

using the following given matrix factorisation (note the last is transposed)

$$A = \begin{bmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\ -\frac{2}{3} & \frac{1}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} 12 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} -\frac{8}{9} & -\frac{1}{9} & -\frac{4}{9} \\ -\frac{4}{9} & \frac{4}{9} & \frac{7}{9} \\ -\frac{1}{9} & -\frac{8}{9} & \frac{4}{9} \end{bmatrix}^T.$$



**Warning: do *not* solve in reverse order**

**Example 3.3.5.** Reconsider [Example 3.3.2](#) wrongly.

(a) After writing the system using the SVD as

$$\underbrace{\begin{bmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{bmatrix}}_{=z} \underbrace{\begin{bmatrix} 10\sqrt{2} & 0 \\ 0 & 5\sqrt{2} \end{bmatrix}}_{=y} \underbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}}_{=y}^T \mathbf{x} = \begin{bmatrix} 18 \\ -1 \end{bmatrix},$$

one might be inadvertently tempted to ‘solve’ the system by using the matrices in reverse order as in the following: *do not do this*.

(b) First solve  $\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}^T \mathbf{x} = \begin{bmatrix} 18 \\ -1 \end{bmatrix}$ : this matrix is orthogonal,

so multiplying by itself (the transpose of the transpose) gives

$$\mathbf{x} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 18 \\ -1 \end{bmatrix} = \begin{bmatrix} 19/\sqrt{2} \\ 17/\sqrt{2} \end{bmatrix}.$$

(c) Inappropriately ‘solve’  $\begin{bmatrix} 10\sqrt{2} & 0 \\ 0 & 5\sqrt{2} \end{bmatrix} \mathbf{y} = \begin{bmatrix} 19/\sqrt{2} \\ 17/\sqrt{2} \end{bmatrix}$ : this matrix is diagonal, so dividing by the diagonal elements gives

$$\mathbf{y} = \begin{bmatrix} 10\sqrt{2} & 0 \\ 0 & 5\sqrt{2} \end{bmatrix}^{-1} \begin{bmatrix} 19/\sqrt{2} \\ 17/\sqrt{2} \end{bmatrix} = \begin{bmatrix} \frac{19}{20} \\ \frac{17}{10} \end{bmatrix}.$$

(d) Inappropriately ‘solve’  $\begin{bmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{bmatrix} \mathbf{z} = \begin{bmatrix} \frac{19}{20} \\ \frac{17}{10} \end{bmatrix}$ : this matrix is orthogonal, so multiplying by the transpose gives

$$\mathbf{z} = \begin{bmatrix} \frac{3}{5} & \frac{4}{5} \\ -\frac{4}{5} & \frac{3}{5} \end{bmatrix} \begin{bmatrix} \frac{19}{20} \\ \frac{17}{10} \end{bmatrix} = \begin{bmatrix} 1.93 \\ 0.26 \end{bmatrix}.$$

And then, since the solution is to be called  $\mathbf{x}$ , we might inappropriately call what we just calculated as the solution  $\mathbf{x} = (1.93, 0.26)$ .



Avoid this reverse process as it is wrong. Matrix multiplicative is *not* commutative ([Subsection 3.1.3](#)). We must use an SVD factorisation in the correct order: to solve linear equations use the matrices in an SVD from left to right.

### 3.3.2 The SVD solves general systems

The previous examples depended upon a matrix being factored into a product of three matrices: two orthogonal and one diagonal. Amazingly, such factorisation is always possible.

<http://www.youtube.com/watch?v=JEYLfIVvR9I> is an entertaining prelude

**Theorem 3.3.6** (SVD factorisation). *Every  $m \times n$  real matrix  $A$  can be factored into a product of three matrices*

$$A = USV^T, \quad (3.4)$$

*called a **singular value decomposition** (SVD), where*

- $m \times m$  matrix  $U = [\mathbf{u}_1 \ \mathbf{u}_2 \ \cdots \ \mathbf{u}_m]$  is orthogonal,
- $n \times n$  matrix  $V = [\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_n]$  is orthogonal, and
- $m \times n$  diagonal matrix  $S$  is zero except for non-negative diagonal elements called **singular values**  $\sigma_1, \sigma_2, \dots, \sigma_{\min(m,n)}$ , which are unique when ordered from largest to smallest so that  $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min(m,n)} \geq 0$ .

The symbol  $\sigma$  is the Greek letter sigma, and denotes singular values.

The orthonormal vectors  $\mathbf{u}_j$  and  $\mathbf{v}_j$  are called **singular vectors**.

Importantly, the singular values are unique (when ordered), although the orthogonal matrices  $U$  and  $V$  are not unique (e.g., one may change the sign of any column in  $U$  together with its corresponding column in  $V$ ). Nonetheless, although there are many SVDs of a matrix, all SVDs are equivalent in application.

Some may be disturbed by the non-uniqueness of an SVD. But the non-uniqueness is analogous to the non-uniqueness of row reduction upon re-ordering of equations, and/or re-ordering the variables in the equations.

**Example 3.3.7.**     [Example 3.3.2](#) invoked the SVD

$$\begin{bmatrix} 10 & 2 \\ 5 & 11 \end{bmatrix} = \begin{bmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{bmatrix} \begin{bmatrix} 10\sqrt{2} & 0 \\ 0 & 5\sqrt{2} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}^T,$$

where the two outer matrices are orthogonal (check), so the singular values of this matrix are  $\sigma_1 = 10\sqrt{2}$  and  $\sigma_2 = 5\sqrt{2}$ .

**Example 3.3.4** invoked the SVD

$$\begin{bmatrix} -4 & -2 & 4 \\ -8 & -1 & -4 \\ 6 & 6 & 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\ -\frac{2}{3} & \frac{1}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} 12 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} -\frac{8}{9} & -\frac{1}{9} & -\frac{4}{9} \\ -\frac{4}{9} & \frac{4}{9} & \frac{7}{9} \\ -\frac{1}{9} & -\frac{8}{9} & \frac{4}{9} \end{bmatrix}^T,$$

where the two outer matrices are orthogonal (check), so the singular values of this matrix are  $\sigma_1 = 12$ ,  $\sigma_2 = 6$  and  $\sigma_3 = 3$ . ■

**Example 3.3.8.** Any orthogonal matrix  $Q$ , say  $n \times n$ , has an SVD  $Q = QI_nI_n^T$ ; that is,  $U = Q$ ,  $S = V = I_n$ . Hence every  $n \times n$  orthogonal matrix has singular values  $\sigma_1 = \sigma_2 = \cdots = \sigma_n = 1$ . ■

**Example 3.3.9** (some non-uniqueness).

has an SVD  $I_n = I_nI_nI_n^T$ .

- An identity matrix, say  $I_n$ ,
- Additionally, for *every*  $n \times n$  orthogonal matrix  $Q$ , the identity  $I_n$  also has the SVD  $I_n = QI_nQ^T$ —as this right-hand side  $QI_nQ^T = QQ^T = I_n$ .



- Further, any constant multiple of an identity, say  $sI_n = \text{diag}(s, s, \dots, s)$ , has the same non-uniqueness: an SVD is  $sI_n = USV^T$  for matrices  $U = Q$ ,  $S = sI_n$  and  $V = Q$  for every  $n \times n$  orthogonal  $Q$  (provided  $s \geq 0$ ).

The matrices in this example are characterised by all their singular values having an identical value. In general, analogous non-uniqueness in  $U$  and  $V$  occurs whenever two or more singular values are identical in value. ■

**Activity 3.3.10.** [Example 3.3.8](#) commented that  $QI_nI_n^T$  is an SVD of an orthogonal matrix  $Q$ . Which of the following is also an SVD of an  $n \times n$  orthogonal matrix  $Q$ ?

(a)  $I_nI_n(Q^T)^T$

(b)  $Q(-I_n)(-I_n)^T$

(c)  $I_nQI_n^T$

(d)  $I_nI_nQ^T$



**Example 3.3.11** (positive ordering). Find an SVD of the diagonal matrix

$$D = \begin{bmatrix} 2.7 & 0 & 0 \\ 0 & -3.9 & 0 \\ 0 & 0 & -0.9 \end{bmatrix}.$$



## Computers empower use of the SVD

Except for simple cases such as  $2 \times 2$  matrices (Example 3.3.32), constructing an SVD is usually far too laborious by hand. Typically, this book either gives an SVD (as in the earlier two examples) or asks you to compute an SVD in MATLAB/Octave with `[U,S,V]=svd(A)` (Table 3.3).

The SVD theorem asserts that every matrix is the product of two orthogonal matrices and a diagonal matrix. Because, in a matrix's SVD factorisation, the rotations (and/or reflection) by the two orthogonal matrices are so 'nice', any 'badness' or 'trickiness' in the matrix is represented in the diagonal matrix  $S$  of the singular values.

The following examples illustrate the cases of either no or infinite solutions, to complement the case of unique solutions of the first two examples.

**Example 3.3.12** (rate sport teams/players). Consider three table tennis players, Anne, Bob and Chris: Anne beat Bob 3 games to 2 games; Anne beat Chris 3-1; Bob beat Chris 3-2. How good are they? What is their rating?



Table 3.3: As well as the MATLAB/Octave commands and operations listed in Tables 1.2, 2.3, 3.1 and 3.2, we need these matrix operations.

- $[U, S, V] = \text{svd}(A)$  computes the three matrices  $U$ ,  $S$  and  $V$  in a singular value decomposition (SVD) of the  $m \times n$  matrix:  $A = USV^T$  for  $m \times m$  orthogonal matrix  $U$ ,  $n \times n$  orthogonal matrix  $V$ , and  $m \times n$  non-negative diagonal matrix  $S$  (Theorem 3.3.6).  
 $\text{svd}(A)$  just reports the singular values in a vector.
- Complementing information of Table 3.1, to extract and compute with a subset of rows/columns of a matrix, specify the vector of indices. For examples:
  - $V(:, 1:r)$  selects the first  $r$  columns of  $V$ ;
  - $A([2 \ 3 \ 5], :)$  selects the second, third and fifth row of matrix  $A$ ;
  - $B(4:6, 1:3)$  selects the  $3 \times 3$  submatrix of the first three columns of the fourth, fifth and sixth rows.

**Compute in Matlab/Octave.** As seen in the previous example, often we need to compute with a subset of the components of matrices (Table 3.3):

- $\mathbf{b}(1:r)$  selects the first  $r$  entries of vector  $\mathbf{b}$
- $\mathbf{S}(1:r, 1:r)$  selects the top-left  $r \times r$  submatrix of  $\mathbf{S}$ ;
- $\mathbf{V}(:, 1:r)$  selects the first  $r$  columns of matrix  $\mathbf{V}$ .

**Example 3.3.13.** But what if Bob beat Chris 3-1? ■

Section 3.5 further explores systems with no solution and uses the SVD to determine a good approximate solution (Example 3.5.3).

**Example 3.3.14.** Find the value(s) of the parameter  $c$  such that the following system has a solution, and find a general solution for that (those) parameter value(s):

$$\begin{bmatrix} -9 & -15 & -9 & -15 \\ -10 & 2 & -10 & 2 \\ 8 & 4 & 8 & 4 \end{bmatrix} \mathbf{x} = \begin{bmatrix} c \\ 8 \\ -5 \end{bmatrix}.$$



**Procedure 3.3.15** (general solution). *Obtain a general solution of the system  $A\mathbf{x} = \mathbf{b}$  using an SVD and via intermediate unknowns.*

1. *Obtain an SVD factorisation  $A = USV^T$ .*
2. *Solve  $U\mathbf{z} = \mathbf{b}$  by  $\mathbf{z} = U^T\mathbf{b}$  (unique given  $U$ ).*
3. *When possible, solve  $S\mathbf{y} = \mathbf{z}$  as follows. Identify the non-zero and the zero singular values: suppose  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$  and  $\sigma_{r+1} = \dots = \sigma_{\min(m,n)} = 0$ :*
  - *if  $z_i \neq 0$  for any  $i = r + 1, \dots, m$ , then there is no solution (the equations are **inconsistent**);*
  - *otherwise (when  $z_i = 0$  for all  $i = r + 1, \dots, m$ ) determine the  $i$ th component of  $\mathbf{y}$  by  $y_i = z_i/\sigma_i$  for  $i = 1, \dots, r$  (for which  $\sigma_i > 0$ ), and let  $y_i$  be a free variable for  $i = r + 1, \dots, n$ .*

4. Solve  $V^T \mathbf{x} = \mathbf{y}$  (unique given  $V$  and for each  $\mathbf{y}$ ) to derive a general solution is  $\mathbf{x} = V\mathbf{y}$ .

This [Procedure 3.3.15](#) determines for us that there is either none, one or an infinite number of solutions, as [Theorem 2.2.27](#) requires.

However, MATLAB/Octave's "`A\`" gives one 'answer' for all of these cases, even when there is no solution or an infinite number of solutions. The function `rcond(A)` indicates whether the 'answer' is a good unique solution of  $A\mathbf{x} = \mathbf{b}$  ([Procedure 2.2.5](#)). [Section 3.5](#) addresses what the 'answer' by MATLAB/Octave means in the other cases of no or infinite solutions.

## Condition number and rank determine the possibilities

The expression ‘ill-conditioned’ is sometimes used merely as a term of abuse . . . It is characteristic of ill-conditioned sets of equations that small percentage errors in the coefficients given may lead to large percentage errors in the solution.

*Alan Turing, 1934 (Higham 1996, p.131)*

The MATLAB/Octave function `rcond()` roughly estimates the reciprocal of what is called the condition number (estimates it to within a factor of two or three).

**Definition 3.3.16.** *For every  $m \times n$  matrix  $A$ , the **condition number** of  $A$  is the ratio of the largest to smallest of its singular values:  $\text{cond } A := \sigma_1 / \sigma_{\min(m,n)}$ . By convention: if  $\sigma_{\min(m,n)} = 0$ , then  $\text{cond } A := \infty$  (infinity); also, for zero matrices  $\text{cond } O_{m \times n} := \infty$ .*



**Example 3.3.17.** [Example 3.3.7](#) gives the singular values of two matrices: for the  $2 \times 2$  matrix the condition number  $\sigma_1/\sigma_2 = (10\sqrt{2})/(5\sqrt{2}) = 2$  (for which `rcond` = 0.5); for the  $3 \times 3$  matrix the condition number  $\sigma_1/\sigma_3 = 12/3 = 4$  (for which `rcond` = 0.25). [Example 3.3.8](#) comments that every  $n \times n$  orthogonal matrix has singular values  $\sigma_1 = \cdots = \sigma_n = 1$ ; hence an orthogonal matrix has condition number one (`rcond` = 1). Such small condition numbers (non-small `rcond`) indicate all orthogonal matrices are “good” matrices (as classified by [Procedure 2.2.5](#)).

However, the matrix in the sports ranking [Example 3.3.12](#) has singular values  $\sigma_1 = \sigma_2 = \sqrt{3}$  and  $\sigma_3 = 0$  so its condition number  $\sigma_1/\sigma_3 = \sqrt{3}/0 = \infty$  (correspondingly, `rcond` = 0) which indicates that the equations are likely to be unsolvable. (In MATLAB/Octave, see that  $\sigma_3 = 2 \cdot 10^{-17}$  so a numerical calculation would give condition number  $1.7321/\sigma_3 = 7 \cdot 10^{16}$  which is effectively infinite.) ■

**Activity 3.3.18.** What is the condition number of the matrix of [Example 3.3.14](#),

$$\begin{bmatrix} -9 & -15 & -9 & -15 \\ -10 & 2 & -10 & 2 \\ 8 & 4 & 8 & 4 \end{bmatrix},$$

given it has an SVD (2 d.p.)

$$\begin{bmatrix} -0.86 & 0.43 & 0.29 \\ -0.29 & -0.86 & 0.43 \\ 0.43 & 0.29 & 0.86 \end{bmatrix} \begin{bmatrix} 28 & 0 & 0 & 0 \\ 0 & 14 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.50 & 0.50 & -0.19 & -0.68 \\ 0.50 & -0.50 & 0.68 & -0.19 \\ 0.50 & 0.50 & 0.19 & 0.68 \\ 0.50 & -0.50 & -0.68 & 0.19 \end{bmatrix}^T$$

(a) 0.5

(b)  $\infty$

(c) 2

(d) 0



In practice, a condition number  $> 10^8$  is effectively infinite (equivalently  $\mathbf{rcond} < 10^{-8}$  is effectively zero, and hence called “terrible” by [Procedure 2.2.5](#)). The closely related important property of a matrix is the *number* of singular values that are nonzero. When

applying the following definition in practical computation (e.g., MATLAB/Octave), any singular values  $< 10^{-8}\sigma_1$  are effectively zero.

**Definition 3.3.19.** *The **rank** of a matrix  $A$  is the number of nonzero singular values in an SVD,  $A = USV^T$ : letting  $r = \text{rank } A$ ,*

$$S = \begin{bmatrix} \sigma_1 & \cdots & 0 & & \\ \vdots & \ddots & \vdots & & \\ 0 & \cdots & \sigma_r & & \\ & & & O_{r \times (n-r)} & \\ & & & & O_{(m-r) \times (n-r)} \end{bmatrix},$$

*equivalently  $S = \text{diag}_{m \times n}(\sigma_1, \sigma_2, \dots, \sigma_r, 0, \dots, 0)$ .*

**Example 3.3.20.** In the four matrices of [Example 3.3.17](#), the respective ranks are 2, 3,  $n$  and 2. ■

[Theorem 3.3.6](#) asserts the singular values are unique for a given matrix, so the rank of a matrix is independent of its different SVDs.

**Activity 3.3.21.** What is the rank of the matrix of [Example 3.3.14](#),

$$\begin{bmatrix} -9 & -15 & -9 & -15 \\ -10 & 2 & -10 & 2 \\ 8 & 4 & 8 & 4 \end{bmatrix},$$

given it has an SVD (2 d.p.)

$$\begin{bmatrix} -0.86 & 0.43 & 0.29 \\ -0.29 & -0.86 & 0.43 \\ 0.43 & 0.29 & 0.86 \end{bmatrix} \begin{bmatrix} 28 & 0 & 0 & 0 \\ 0 & 14 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.50 & 0.50 & -0.19 & -0.68 \\ 0.50 & -0.50 & 0.68 & -0.19 \\ 0.50 & 0.50 & 0.19 & 0.68 \\ 0.50 & -0.50 & -0.68 & 0.19 \end{bmatrix}^T$$

(a) 1

(b) 3

(c) 2

(d) 4



**Example 3.3.22.** Use MATLAB/Octave to find the ranks of the two matrices

(a) 
$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & -1 \\ 1 & 0 & -1 \\ 2 & 0 & -2 \end{bmatrix}$$

(b) 
$$\begin{bmatrix} 1 & -2 & -1 & 2 & 1 \\ -2 & -2 & -0 & 2 & -0 \\ -2 & -3 & 1 & -1 & 1 \\ -3 & 0 & 1 & -0 & -1 \\ 2 & 1 & 1 & 2 & -1 \end{bmatrix}$$

■

**Theorem 3.3.23.** *For every matrix  $A$ , let an SVD of  $A$  be  $USV^T$ , then the transpose  $A^T$  has an SVD of  $V(S^T)U^T$ . Further,  $\text{rank}(A^T) = \text{rank } A$ .*

**Example 3.3.24.** From earlier examples, write down an SVD of the matrices

$$\begin{bmatrix} 10 & 5 \\ 2 & 11 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -4 & -8 & 6 \\ -2 & -1 & 6 \\ 4 & -4 & 0 \end{bmatrix}.$$



**Activity 3.3.25.** Recall that

$$\begin{bmatrix} -0.86 & 0.43 & 0.29 \\ -0.29 & -0.86 & 0.43 \\ 0.43 & 0.29 & 0.86 \end{bmatrix} \begin{bmatrix} 28 & 0 & 0 & 0 \\ 0 & 14 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.50 & 0.50 & -0.19 & -0.68 \\ 0.50 & -0.50 & 0.68 & -0.19 \\ 0.50 & 0.50 & 0.19 & 0.68 \\ 0.50 & -0.50 & -0.68 & 0.19 \end{bmatrix}^T$$

is an SVD (2 d.p.) of the matrix of [Example 3.3.14](#),

$$\begin{bmatrix} -9 & -15 & -9 & -15 \\ -10 & 2 & -10 & 2 \\ 8 & 4 & 8 & 4 \end{bmatrix}.$$

Which of the following is an SVD of the transpose of this matrix?

$$\begin{aligned}
\text{(a)} \quad & \begin{bmatrix} 0.50 & 0.50 & -0.19 & -0.68 \\ 0.50 & -0.50 & 0.68 & -0.19 \\ 0.50 & 0.50 & 0.19 & 0.68 \\ 0.50 & -0.50 & -0.68 & 0.19 \end{bmatrix} \begin{bmatrix} 28 & 0 & 0 \\ 0 & 14 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -0.86 & 0.43 & 0.29 \\ -0.29 & -0.86 & 0.43 \\ 0.43 & 0.29 & 0.86 \end{bmatrix}^T \\
\text{(b)} \quad & \begin{bmatrix} -0.86 & 0.43 & 0.29 \\ -0.29 & -0.86 & 0.43 \\ 0.43 & 0.29 & 0.86 \end{bmatrix} \begin{bmatrix} 28 & 0 & 0 \\ 0 & 14 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.50 & 0.50 & -0.19 & -0.68 \\ 0.50 & -0.50 & 0.68 & -0.19 \\ 0.50 & 0.50 & 0.19 & 0.68 \\ 0.50 & -0.50 & -0.68 & 0.19 \end{bmatrix}^T \\
\text{(c)} \quad & \begin{bmatrix} 0.50 & 0.50 & 0.50 & 0.50 \\ 0.50 & -0.50 & 0.50 & -0.50 \\ -0.19 & 0.68 & 0.19 & -0.68 \\ -0.68 & -0.19 & 0.68 & 0.19 \end{bmatrix} \begin{bmatrix} 28 & 0 & 0 \\ 0 & 14 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -0.86 & -0.29 & 0.43 \\ 0.43 & -0.86 & 0.29 \\ 0.29 & 0.43 & 0.86 \end{bmatrix}^T \\
\text{(d)} \quad & \begin{bmatrix} -0.86 & -0.29 & 0.43 \\ 0.43 & -0.86 & 0.29 \\ 0.29 & 0.43 & 0.86 \end{bmatrix} \begin{bmatrix} 28 & 0 & 0 \\ 0 & 14 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.50 & 0.50 & 0.50 & 0.50 \\ 0.50 & -0.50 & 0.50 & -0.50 \\ -0.19 & 0.68 & 0.19 & -0.68 \\ -0.68 & -0.19 & 0.68 & 0.19 \end{bmatrix}^T
\end{aligned}$$



Let's now return to the topic of linear equations and connect new

concepts to the task of solving linear equations. In particular, the following theorem addresses when a unique solution exists to a system of linear equations. Concepts developed in subsequent sections extend this theorem further (Theorems 3.4.43 and 7.2.41).

**Theorem 3.3.26** (Unique Solutions: version 1). *For every  $n \times n$  square matrix  $A$ , the following statements are equivalent:*

- (a)  $A$  is invertible;
- (b)  $A\mathbf{x} = \mathbf{b}$  has a unique solution for every  $\mathbf{b}$  in  $\mathbb{R}^n$ ;
- (c)  $A\mathbf{x} = \mathbf{0}$  has only the zero solution;
- (d) all  $n$  singular values of  $A$  are nonzero;
- (e) the condition number of  $A$  is finite ( $\text{rcond} > 0$ );
- (f)  $\text{rank } A = n$ .

Optional: this discussion and theorem reinforces why we must check condition numbers in computation.



**Practical shades of grey** The preceding Unique Solution [Theorem 3.3.26](#) is ‘black-and-white’: either a solution exists, or it does not. This is a great theory. But in applications, problems arise in ‘all shades of grey’. Practical issues in applications are better phrased in terms of reliability, uncertainty, and error estimates. For example, suppose in an experiment you measure quantities  $\mathbf{b}$  to three significant digits, then solve the linear equations  $A\mathbf{x} = \mathbf{b}$  to estimate quantities of interest  $\mathbf{x}$ : how accurate are your estimates of the interesting quantities  $\mathbf{x}$ ? or are your estimates complete nonsense?

**Example 3.3.27.** Consider the following innocuous looking system of linear equations

$$\begin{cases} -2q + r = 3 \\ p - 5q + r = 8 \\ -3p + 2q + 3r = -5 \end{cases}$$

Solve by hand ([Procedure 2.2.24](#)) to find the unique solution is  $(p, q, r) = (2, -1, 1)$ .

But, and it is a big but in practical applications, what happens if

the right-hand side comes from experimental measurements with a relative error of 1%? Let's explore by writing the system in matrix-vector form and using MATLAB/Octave to solve with various example errors.

- (a) First solve the system as stated. Denoting the unknowns by vector  $\mathbf{x} = (p, q, r)$ , write the system as  $A\mathbf{x} = \mathbf{b}$  for matrix

$$A = \begin{bmatrix} 0 & -2 & 1 \\ 1 & -5 & 1 \\ -3 & 2 & 3 \end{bmatrix}, \quad \text{and right-hand side } \mathbf{b} = \begin{bmatrix} 3 \\ 8 \\ -5 \end{bmatrix}.$$

Use [Procedure 2.2.5](#) to solve the system in MATLAB/Octave:

- i. enter the matrix and vector with

```
A=[0 -2 1; 1 -5 1; -3 2 3]
b=[3;8;-5]
```

- ii. find `rcond(A)` is 0.0031 which is poor, but we proceed anyway;

- iii. then `x=A\b` gives the solution  $\mathbf{x} = (2, -1, 1)$  as before.



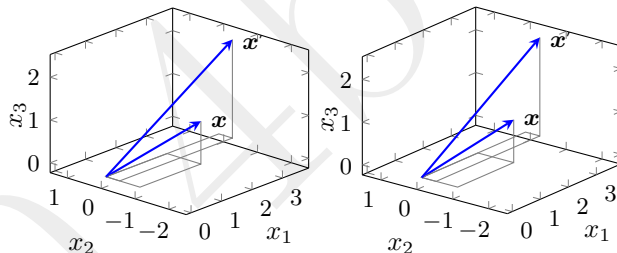
- (b) Now let's recognise that the right-hand side comes from experimental measurements with a 1% error. In MATLAB/Octave, `norm(b)` computes the length  $|\mathbf{b}| = 9.90$  (2 d.p.). Thus a 1% error corresponds to changing  $\mathbf{b}$  by  $0.01 \times 9.90 \approx 0.1$ . Let's say the first component of  $\mathbf{b}$  is in error by this amount and see what the new solution would be:

- i. executing `x1=A\b+[0.1;0;0]` adds the 1% error  $(0.1, 0, 0)$  to  $\mathbf{b}$  and then solves the new system to find  $\mathbf{x}' = (3.7, -0.4, 2.3)$ . This solution is very different to the original solution  $\mathbf{x} = (2, -1, 1)!$
- ii. `relerr1=norm(x-x1)/norm(x)` computes its relative error  $|\mathbf{x} - \mathbf{x}'|/|\mathbf{x}|$  to be 0.91, that is, 91%—rather large.

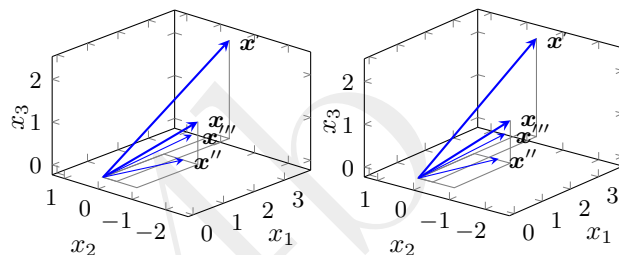
As illustrated below, the large difference between  $\mathbf{x}$  and  $\mathbf{x}'$  indicates 'the solution'  $\mathbf{x}$  is almost complete nonsense. How can a 1% error in  $\mathbf{b}$  turn into the astonishingly large 91% error in solution  $\mathbf{x}$ ? [Theorem 3.3.29](#) below shows it is no accident that the magnification of the error by a factor of 91 is of the same order of magnitude as the condition number = 152.27



computed via  $\mathbf{s}=\text{svd}(\mathbf{A})$  and then  $\text{condA}=\mathbf{s}(1)/\mathbf{s}(3)$ .



- (c) To explore further, let's say the second component of  $\mathbf{b}$  is in error by 1% of  $\mathbf{b}$ , that is, by 0.1. As in the previous case, add  $(0, 0.1, 0)$  to the right-hand side and solve to find now  $\mathbf{x}'' = (1.2, -1.3, 0.4)$  which is quite different to both  $\mathbf{x}$  and  $\mathbf{x}'$ , as illustrated below. Compute its relative error  $|\mathbf{x} - \mathbf{x}''|/|\mathbf{x}| = 0.43$ . At 43%, the relative error in solution  $\mathbf{x}''$  is also much larger than the 1% error in  $\mathbf{b}$ .



- (d) Lastly, let's say the third component of  $\mathbf{b}$  is in error by 1% of  $\mathbf{b}$ , that is, by 0.1. As in the previous cases, add  $(0, 0, 0.1)$  to the right-hand side and solve to find now  $\mathbf{x}''' = (1.7, -1.1, 0.8)$  which, as illustrated above, is at least is roughly  $\mathbf{x}$ . Compute its relative error  $|\mathbf{x} - \mathbf{x}'''|/|\mathbf{x}| = 0.15$ . At 15%, the relative error in solution  $\mathbf{x}'''$  is significantly larger than the 1% error in  $\mathbf{b}$ .

This example shows that the apparently innocuous matrix  $A$  variously multiplies measurement errors in  $\mathbf{b}$  by factors of 91, 41 or 15 when finding 'the solution'  $\mathbf{x}$  to  $A\mathbf{x} = \mathbf{b}$ . The matrix  $A$  must, after all, be a bad matrix. [Theorem 3.3.29](#) shows this badness is quantified by its condition number 152.27, and its poor reciprocal

$$\text{rcond}(\mathbf{A}) = 0.0031 \text{ (estimated).}$$

**Example 3.3.28.** Consider solving the system of linear equations

$$\begin{bmatrix} 0.4 & 0.4 & -0.2 & 0.8 \\ -0.2 & 0.8 & -0.4 & -0.4 \\ 0.4 & -0.4 & -0.8 & -0.2 \\ -0.8 & -0.2 & -0.4 & 0.4 \end{bmatrix} \mathbf{x} = \begin{bmatrix} -3 \\ 3 \\ -9 \\ -1 \end{bmatrix}.$$

Use MATLAB/Octave to explore the effect on the solution  $\mathbf{x}$  of 1% errors in the right-hand side vector.

The condition number determines the reliability of the solution of a system of linear equations. This is why we should always precede the computation of a solution with an estimate of the condition number such as that provided by the reciprocal `rcond()` ([Procedure 2.2.5](#)). The next theorem establishes that the condition number characterises the amplification of errors that occurs in solving a linear system. Hence solving a system of linear equations

with a large condition number (small `rcond`) means that errors are amplified by a large factor as happens in [Example 3.3.27](#).

**Theorem 3.3.29** (error magnification).

*The symbol  $\epsilon$  is the Greek letter epsilon, and often denotes errors.*

*Consider solving  $A\mathbf{x} = \mathbf{b}$  for  $n \times n$  matrix  $A$  with full rank  $A = n$ . Suppose the right-hand side  $\mathbf{b}$  has relative error of size  $\epsilon$ , then the solution  $\mathbf{x}$  has relative error  $\leq \epsilon \text{cond } A$ , with equality in the worst case.*

**Example 3.3.30.**

Each of the following cases involves solving a linear system  $A\mathbf{x} = \mathbf{b}$  to determine quantities of interest  $\mathbf{x}$  from some measured quantities  $\mathbf{b}$ . From the given information estimate the maximum relative error in  $\mathbf{x}$ , if possible, otherwise say so.

- (a) Quantities  $\mathbf{b}$  are measured to a relative error 0.001, and matrix  $A$  has condition number of ten.
- (b) Quantities  $\mathbf{b}$  are measured to three significant digits and `rcond(A)` = 0.025.
- (c) Measurements are accurate to two decimal places, and matrix  $A$  has condition number of twenty.

- (d) Measurements are correct to two significant digits and  $\text{rcond}(\mathbf{A}) = 0.002$ .



**Activity 3.3.31.** In some experiment the components of  $\mathbf{b}$ ,  $|\mathbf{b}| = 5$ , are measured to two decimal places. We compute a vector  $\mathbf{x}$  by solving  $\mathbf{A}\mathbf{x} = \mathbf{b}$  with a matrix  $\mathbf{A}$  for which we compute  $\text{rcond}(\mathbf{A}) = 0.02$ . What is our estimate of the relative error in  $\mathbf{x}$ ?

- (a) 20%      (b) 5%      (c) 2%      (d) 0.1%



This issue of the amplification of errors occurs in other contexts. The eminent mathematician Henri Poincaré (1854–1912) was the first to detect possible chaos in the orbits of the planets.

If we knew exactly the laws of nature and the situation of the universe at the initial moment, we could predict



exactly the situation of that same universe at a succeeding moment. But even if it were the case that the natural laws had no longer any secret for us, we could still only know the initial situation approximately. If that enabled us to predict the succeeding situation with the same approximation, that is all we require, and we should say that the phenomenon had been predicted, that it is governed by laws. But it is not always so; it may happen that small differences in the initial conditions produce very great ones in the final phenomena. A small error in the former will produce an enormous error in the latter. Prediction becomes impossible, and we have the fortuitous phenomenon. *Poincaré, 1903*

The analogue for us in solving linear equations such as  $A\mathbf{x} = \mathbf{b}$  is the following: it may happen that a small error in the elements of  $\mathbf{b}$  will produce an enormous error in the final  $\mathbf{x}$ . The condition number warns when this happens by characterising the amplification.

### 3.3.3 Prove the SVD **Theorem 3.3.6**

When doing maths there's this great feeling. You start with a problem that just mystifies you. You can't understand it, it's so complicated, you just can't make head nor tail of it. But then when you finally resolve it, you have this incredible feeling of how beautiful it is, how it all fits together so elegantly. *Andrew Wiles, C1993*

This proof may be delayed until the last week of a semester. It may be given together with the closely related classic proof of **Theorem 4.2.16** on the eigenvectors of symmetric matrices.

Two preliminary examples introduce the structure of the general proof that an SVD exists. As in this example prelude, the proof of a general singular value decomposition is similarly constructive.

## Prelude to the proof

These first two examples are optional: their purpose is to introduce key parts of the general proof in a definite setting.

**Example 3.3.32** (a  $2 \times 2$  case). Recall [Example 3.3.2](#) factorised the matrix

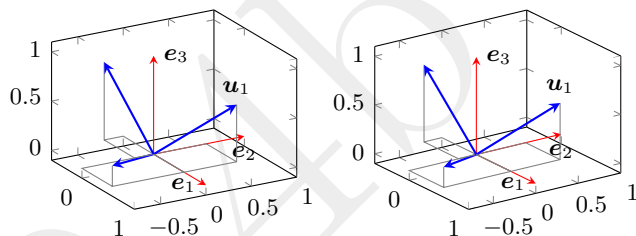
$$A = \begin{bmatrix} 10 & 2 \\ 5 & 11 \end{bmatrix} = \begin{bmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{bmatrix} \begin{bmatrix} 10\sqrt{2} & 0 \\ 0 & 5\sqrt{2} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}^T.$$

Find this factorisation,  $A = USV^T$ , by maximising  $|A\mathbf{v}|$  over all unit vectors  $\mathbf{v}$ . ■

**Example 3.3.33** (a  $3 \times 1$  case). Find the following SVD for the  $3 \times 1$  matrix

$$A = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{3}} & \cdot & \cdot \\ \frac{1}{\sqrt{3}} & \cdot & \cdot \\ \frac{1}{\sqrt{3}} & \cdot & \cdot \end{bmatrix} \begin{bmatrix} \sqrt{3} \\ 0 \\ 0 \end{bmatrix} [1]^T = USV^T,$$

where we do not worry about the elements denoted by dots as they are multiplied by the zeros in  $S = (\sqrt{3}, 0, 0)$ .



■

**Outline of the general proof** We use induction on the size  $m \times n$  of the matrix.

- First zero matrices have trivial SVD, and  $m \times 1$  and  $1 \times n$  matrices have straightforward SVD (as in [Example 3.3.33](#)).
- Choose  $\mathbf{v}_1$  to maximise  $|\mathbf{A}\mathbf{v}|^2$  for unit vectors  $\mathbf{v}$  in  $\mathbb{R}^n$ .

- Crucially, we then establish that for every vector  $\mathbf{v}$  orthogonal to  $\mathbf{v}_1$ , the vector  $A\mathbf{v}$  is orthogonal to  $A\mathbf{v}_1$ .
- Then rotate the standard unit vectors to align one with  $\mathbf{v}_1$ . Similarly for  $A\mathbf{v}_1$ .
- This rotation transforms the matrix  $A$  to strip off the leading singular value, and effectively leave an  $(m - 1) \times (n - 1)$  matrix.
- By induction on the size, an SVD exists for all sizes.

This proof corresponds closely to the proof of the spectral [Theorem 4.2.16](#) for symmetric matrices of [Section 4.2](#).

Detailed proof of the SVD **Theorem 3.3.6**

Use induction on the size  $m \times n$  of the matrix  $A$ : we assume an SVD exists for all  $(m-1) \times (n-1)$  matrices, and prove that consequently an SVD must exist for all  $m \times n$  matrices. There are three base cases to establish: one for  $m \leq n$ , one for  $m \geq n$ , and one for matrix  $A = O$ ; then the induction extends to all sized matrices.

**Case  $A = O_{m \times n}$ :** When  $m \times n$  matrix  $A = O_{m \times n}$  then choose  $U = I_m$  (orthogonal),  $S = O_{m \times n}$  (diagonal), and  $V = I_n$  (orthogonal) so then  $USV^T = I_m O_{m \times n} I_n^T = O_{m \times n} = A$ .

Consequently, the rest of the proof only considers the non-trivial cases when the matrix  $A$  is not all zero.

**Case  $m \times 1$  ( $n = 1$ ):** Here the  $m \times 1$  nonzero matrix  $A = [\mathbf{a}_1]$  for  $\mathbf{a}_1 = (a_{11}, a_{21}, \dots, a_{m1})$ . Set the singular value  $\sigma_1 = |\mathbf{a}_1| = \sqrt{a_{11}^2 + a_{21}^2 + \dots + a_{m1}^2}$  and unit vector  $\mathbf{u}_1 = \mathbf{a}_1/\sigma_1$ . Set  $1 \times 1$  orthogonal matrix  $V = [1]$ ;  $m \times 1$  diagonal matrix  $S = (\sigma_1, 0, \dots, 0)$ ; and  $m \times m$  orthogonal matrix  $U = [\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_m]$ . Matrix  $U$

exists because we can take the orthonormal set of standard unit vectors in  $\mathbb{R}^m$  and rotate them all together so that the first lines up with  $\mathbf{u}_1$ : the other  $(m-1)$  unit vectors then become the other  $\mathbf{u}_j$ . Then an SVD for the  $m \times 1$  matrix  $A$  is

$$\begin{aligned} USV^T &= [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_m] \begin{bmatrix} \sigma_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \mathbf{1}^T \\ &= \sigma_1 \mathbf{u}_1 = [\mathbf{a}_1] = A. \end{aligned}$$

**Case  $1 \times n$  ( $m=1$ ):** use an exactly complementary argument to the preceding  $m \times 1$  case.

**Induction** Assume an SVD exists for all  $(m-1) \times (n-1)$  matrices: we proceed to prove that consequently an SVD must exist for all  $m \times n$  matrices. Consider any  $m \times n$  nonzero matrix  $A$  for  $m, n \geq 2$ . Set vector  $\mathbf{v}_1$  in  $\mathbb{R}^n$  to be a *unit vector* that maximises  $|A\mathbf{v}|^2$  for

unit vectors  $\mathbf{v}$  in  $\mathbb{R}^n$ ; that is, vector  $\mathbf{v}_1$  achieves the maximum in  $\max_{|\mathbf{v}|=1} |A\mathbf{v}|^2$ .

1. *Such a maximum exists by the Extreme Value Theorem in Calculus.* This theorem is proved in higher level analysis.

As matrix  $A$  is nonzero, there exists  $\mathbf{v}$  such that  $|A\mathbf{v}| > 0$ . Since  $\mathbf{v}_1$  maximises  $|A\mathbf{v}|$  it follows that  $|A\mathbf{v}_1| > 0$ .

*The vector  $\mathbf{v}_1$  is not unique:* for example, the negative  $-\mathbf{v}_1$  is another unit vector that achieves the maximum value. Sometimes there are other unit vectors that achieve the maximum value. Choose any one of them.

Nonetheless, the maximum value of  $|A\mathbf{v}|^2$  is unique, and so the following singular value  $\sigma_1$  is unique.

2. *Set the singular value  $\sigma_1 := |A\mathbf{v}_1| > 0$  and unit vector  $\mathbf{u}_1 := (A\mathbf{v}_1)/\sigma_1$  in  $\mathbb{R}^m$ . For every unit vector  $\mathbf{v}$  orthogonal to  $\mathbf{v}_1$  we now prove that the vector  $A\mathbf{v}$  is orthogonal to  $\mathbf{u}_1$ . Let  $\mathbf{u} := A\mathbf{v}$  in  $\mathbb{R}^m$  and consider  $f(t) := |A(\mathbf{v}_1 \cos t + \mathbf{v} \sin t)|^2$ . Since  $\mathbf{v}_1$  achieves the maximum, and  $\mathbf{v}_1 \cos t + \mathbf{v} \sin t$  is a*



unit vector for all  $t$  (??), then  $f(t)$  must have a maximum at  $t = 0$  (maybe at other  $t$  as well), and so  $f'(0) = 0$  (from the Calculus of a maximum). On the other hand,

$$\begin{aligned} f(t) &= |A\mathbf{v}_1 \cos t + A\mathbf{v} \sin t|^2 \\ &= |\sigma_1 \mathbf{u}_1 \cos t + \mathbf{u} \sin t|^2 \\ &= (\sigma_1 \mathbf{u}_1 \cos t + \mathbf{u} \sin t) \cdot (\sigma_1 \mathbf{u}_1 \cos t + \mathbf{u} \sin t) \\ &= \sigma_1^2 \cos^2 t + \sigma_1 \mathbf{u} \cdot \mathbf{u}_1 2 \sin t \cos t + |\mathbf{u}|^2 \sin^2 t; \end{aligned}$$

differentiating  $f(t)$  and evaluating at zero gives  $f'(0) = \sigma_1 \mathbf{u} \cdot \mathbf{u}_1$ . But from the maximum this derivative is zero, so  $\sigma_1 \mathbf{u} \cdot \mathbf{u}_1 = 0$ . Since the singular value  $\sigma_1 > 0$ , we must have  $\mathbf{u} \cdot \mathbf{u}_1 = 0$  and so  $\mathbf{u}_1$  and  $\mathbf{u}$  are orthogonal ([Definition 1.3.19](#)).

3. Consider the orthonormal set of standard unit vectors in  $\mathbb{R}^n$ : rotate them so that the first unit vector lines up with  $\mathbf{v}_1$ , and let the other  $(n - 1)$  rotated unit vectors become the columns of the  $n \times (n - 1)$  matrix  $\bar{V}$ . Then set the  $n \times n$  matrix  $V_1 := [\mathbf{v}_1 \ \bar{V}]$  which is orthogonal as its columns are orthonormal ([Theorem 3.2.48b](#)). Similarly set an  $m \times m$

orthogonal matrix  $U_1 := [\mathbf{u}_1 \quad \bar{U}]$ . Compute the  $m \times n$  matrix

$$\begin{aligned} A_1 &:= U_1^T A V_1 = \begin{bmatrix} \mathbf{u}_1^T \\ \bar{U}^T \end{bmatrix} A \begin{bmatrix} \mathbf{v}_1 & \bar{V} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{u}_1^T A \mathbf{v}_1 & \mathbf{u}_1^T A \bar{V} \\ \bar{U}^T A \mathbf{v}_1 & \bar{U}^T A \bar{V} \end{bmatrix} \end{aligned}$$

where

- the top-left entry  $\mathbf{u}_1^T A \mathbf{v}_1 = \mathbf{u}_1^T \sigma_1 \mathbf{u}_1 = \sigma_1 |\mathbf{u}_1|^2 = \sigma_1$ ,
- the bottom-left column  $\bar{U}^T A \mathbf{v}_1 = \bar{U}^T \sigma_1 \mathbf{u}_1 = O_{m-1 \times 1}$  as the columns of  $\bar{U}$  are orthogonal to  $\mathbf{u}_1$ ,
- the top-right row  $\mathbf{u}_1^T A \bar{V} = O_{1 \times n-1}$  as each column of  $\bar{V}$  is orthogonal to  $\mathbf{v}_1$  and hence each column of  $A \bar{V}$  is orthogonal to  $\mathbf{u}_1$ ,
- and set the bottom-right block  $B := \bar{U}^T A \bar{V}$  which is an  $(m-1) \times (n-1)$  matrix as  $\bar{U}^T$  is  $(m-1) \times m$  and  $\bar{V}$  is  $n \times (n-1)$ .

Consequently,

$$A_1 = \begin{bmatrix} \sigma_1 & O_{1 \times n-1} \\ O_{m-1 \times 1} & B \end{bmatrix}.$$

Note: rearranging  $A_1 := U_1^T A V_1$  gives  $A V_1 = U_1 A_1$ .

4. *By induction assumption,  $(m-1) \times (n-1)$  matrix  $B$  has an SVD, and so we now construct an SVD for  $m \times n$  matrix  $A$ . Let  $B = \hat{U} \hat{S} \hat{V}^T$  be an SVD for  $B$ . Then construct*

$$U := U_1 \begin{bmatrix} 1 & 0 \\ 0 & \hat{U} \end{bmatrix}, \quad V := V_1 \begin{bmatrix} 1 & 0 \\ 0 & \hat{V} \end{bmatrix}, \quad S := \begin{bmatrix} \sigma_1 & 0 \\ 0 & \hat{S} \end{bmatrix}.$$

Matrices  $U$  and  $V$  are orthogonal as each are the product of two orthogonal matrices (??). Also, matrix  $S$  is diagonal. These form an SVD for matrix  $A$  since

$$\begin{aligned} AV &= A V_1 \begin{bmatrix} 1 & 0 \\ 0 & \hat{V} \end{bmatrix} = U_1 A_1 \begin{bmatrix} 1 & 0 \\ 0 & \hat{V} \end{bmatrix} \\ &= U_1 \begin{bmatrix} \sigma_1 & 0 \\ 0 & B \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \hat{V} \end{bmatrix} = U_1 \begin{bmatrix} \sigma_1 & 0 \\ 0 & B \hat{V} \end{bmatrix} \end{aligned}$$

$$\begin{aligned} &= U_1 \begin{bmatrix} \sigma_1 & 0 \\ 0 & \hat{U}\hat{S} \end{bmatrix} = U_1 \begin{bmatrix} 1 & 0 \\ 0 & \hat{U} \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \hat{S} \end{bmatrix} \\ &= US. \end{aligned}$$

Hence  $A = USV^T$ . By induction, an SVD exists for all  $m \times n$  matrices.

This argument establishes the SVD [Theorem 3.3.6](#).

## 3.4 Subspaces, basis and dimension

### Section Contents

3.4.1	Subspaces are lines, planes, and so on . . . .	326
3.4.2	Orthonormal bases form a foundation . . . .	345
3.4.3	Is it a line? a plane? The dimension answers	361

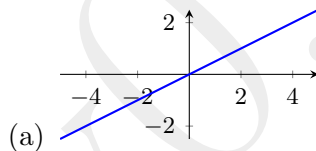
[Nature] is written in that great book which ever lies before our eyes—I mean the universe—but we cannot understand it if we do not first learn the language and grasp the symbols in which it is written. The book is written in the mathematical language, and the symbols are triangles, circles, and other geometric figures, without whose help it is impossible to comprehend a single word of it; without which one wanders in vain through a dark labyrinth.

*Galileo Galilei, 1610*

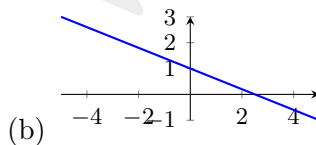
Some of the most fundamental geometric structures in mathematics, especially linear algebra, are the lines or planes through the origin, and higher dimensional analogues. For example, a general solution of linear equations often involve linear combinations such as  $(-2, 1, 0, 0)s + (-\frac{15}{7}, 0, \frac{9}{7}, 1)t$  (Example 2.2.29d) and  $y_3\mathbf{v}_3 + y_4\mathbf{v}_4$  (Example 3.3.14): such combinations for all values of the free variables forms a plane through the origin (Subsection 1.3.4). The aim of this section is to connect geometric structures, such as lines and planes, to the information in a singular value decomposition. The structures are called subspaces.

### 3.4.1 Subspaces are lines, planes, and so on

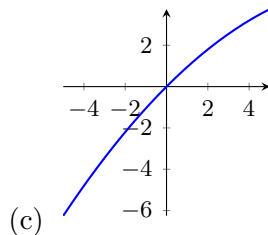
**Example 3.4.1.** The following graphs illustrate the concept of subspaces through examples (imagine the graphs extend to infinitely as appropriate).



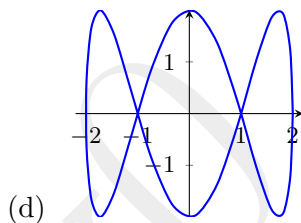
is a subspace as it is a straight line through the origin.



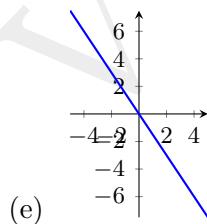
is *not* a subspace as it does not include the origin.



is *not* a subspace as it  
curves.

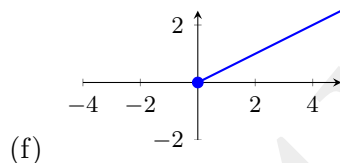


is *not* a subspace as it  
not only curves, but does  
not include the origin.

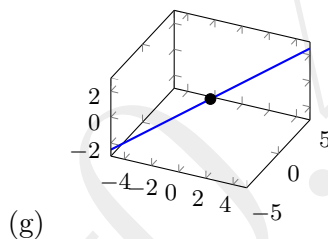


is a subspace.

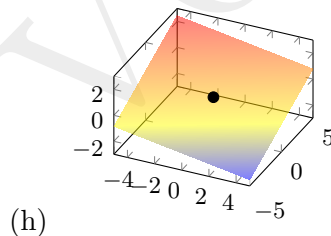




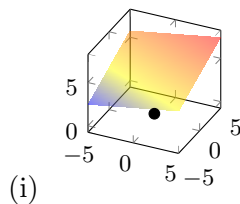
where the disc indicates an end to the line, is *not* a subspace as it does not extend infinitely in both directions.



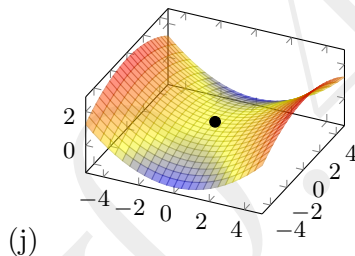
is a subspace as it is a line through the origin (marked in these 3D plots).



is a subspace as it is a plane through the origin.



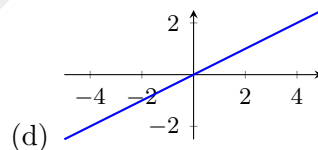
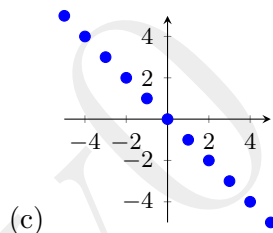
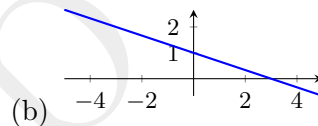
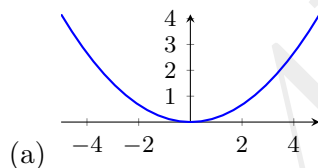
is *not* a subspace as it does not go through the origin.



is *not* a subspace as it curves.



**Activity 3.4.2.** Given the examples and comments of [Example 3.4.1](#), which of the following is a subspace?



The following definition expresses precisely in algebra the concept of a subspace. This book uses the ‘blackboard bold’ font, such as

$\mathbb{W}$  and  $\mathbb{R}$ , for names of spaces and subspaces.

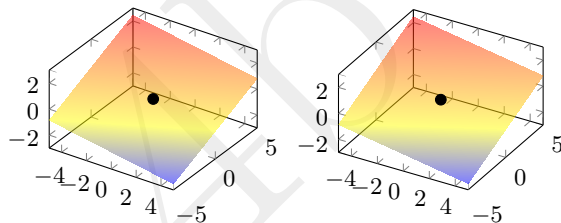
Recall that the mathematical symbol “ $\in$ ” means “in” or “in the set” or “is an element of the set”. For two examples: “ $c \in \mathbb{R}$ ” means “ $c$  is a real number”; whereas “ $\mathbf{v} \in \mathbb{R}^3$ ” means “ $\mathbf{v}$  is a vector with three components. Hereafter, this book uses “ $\in$ ” extensively.

**Definition 3.4.3.** *A **subspace**  $\mathbb{W}$  of  $\mathbb{R}^n$ , is a set of vectors with  $\mathbf{0} \in \mathbb{W}$  and such that  $\mathbb{W}$  is closed under addition and scalar multiplication: that is, for all  $c \in \mathbb{R}$  and for all  $\mathbf{u}, \mathbf{v} \in \mathbb{W}$ , then both  $\mathbf{u} + \mathbf{v} \in \mathbb{W}$  and  $c\mathbf{u} \in \mathbb{W}$ .*

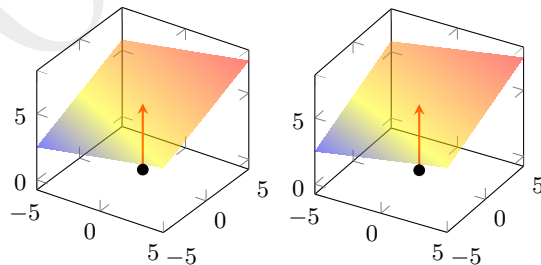
**Example 3.4.4.** Use [Definition 3.4.3](#) to show why each of the following are subspaces, or not.

- (a) All vectors in the line  $y = x/2$  ([Example 3.4.1a](#)).
- (b) All vectors  $(x, y)$  such that  $y = x - x^2/20$  ([Example 3.4.1c](#)).
- (c) All vectors  $(x, y)$  in the line  $y = x/2$  for  $x, y \geq 0$  ([Example 3.4.1f](#)).

- (d) All vectors  $(x, y, z)$  in the plane  $z = -x/6 + y/3$  (Example 3.4.1h).



- (e) All vectors  $(x, y, z)$  in the plane  $z = 5 + x/6 + y/3$  (Example 3.4.1i).

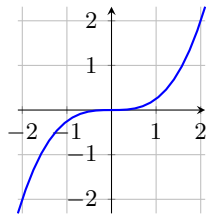


(f)  $\{\mathbf{0}\}$ .

(g)  $\mathbb{R}^n$ .

■

**Activity 3.4.5.** The following pairs of vectors are all in the set shown in the margin (in the sense that their end-points lie on the plotted curve). The sum of which pair proves that the curve plotted in the margin is not a subspace?



(a)  $(2, 2), (-2, -2)$

(b)  $(1, \frac{1}{4}), (0, 0)$

(c)  $(-1, -\frac{1}{4}), (2, 2)$

(d)  $(0, 0), (2, 2)$

■

In summary:

- in two dimensions ( $\mathbb{R}^2$ ), subspaces are the origin  $\mathbf{0}$ , a line through  $\mathbf{0}$ , or the entire plane  $\mathbb{R}^2$ ;

- in three dimensions ( $\mathbb{R}^3$ ), subspaces are the origin  $\mathbf{0}$ , a line through  $\mathbf{0}$ , a plane through  $\mathbf{0}$ , or the entire space  $\mathbb{R}^3$ ;
- and analogously for higher dimensions ( $\mathbb{R}^n$ ).

Recall that the set of all linear combinations of a set of vectors, such as  $(-2, 1, 0, 0)s + (-\frac{15}{7}, 0, \frac{9}{7}, 1)t$  (Example 2.2.29d), is called the span of that set (Definition 2.3.10).

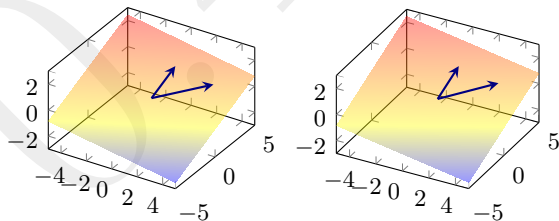
**Theorem 3.4.6.** *Let  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$  be  $k$  vectors in  $\mathbb{R}^n$ , then  $\text{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$  is a subspace of  $\mathbb{R}^n$ .*

**Example 3.4.7.**  $\text{span}\{(1, \frac{1}{2})\}$  is the subspace  $y = x/2$ . The reason is that a vector  $\mathbf{u} \in \text{span}\{(1, \frac{1}{2})\}$  only if there is some constant  $a_1$  such that  $\mathbf{u} = a_1(1, \frac{1}{2}) = (a_1, a_1/2)$ . That is, the  $y$ -component is half the  $x$ -component and hence it lies on the line  $y = x/2$ .

$\text{span}\{(1, \frac{1}{2}), (-2, -1)\}$  is also the subspace  $y = x/2$  since every linear combination  $a_1(1, \frac{1}{2}) + a_2(-2, -1) = (a_1 - 2a_2, a_1/2 - a_2)$  satisfies that the  $y$ -component is half the  $x$ -component and hence the linear combination lies on the line  $y = x/2$ . ■

**Example 3.4.8.**

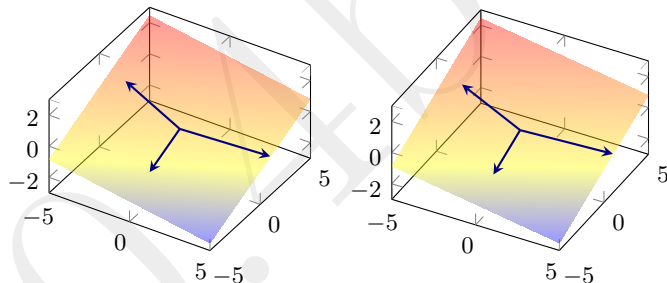
The plane  $z = -x/6 + y/3$  may be written as  $\text{span}\{(3, 3, 1/2), (0, 3, 1)\}$ , as illustrated in stereo below, since every linear combination of these two vectors fills out the plane:  $a_1(3, 3, 1/2) + a_2(0, 3, 1) = (3a_1, 3a_1 + 3a_2, a_1/2 + a_2)$  and so lies in the plane as  $-x/6 + y/3 - z = -\frac{1}{6}3a_1 + \frac{1}{3}(3a_1 + 3a_2) - (a_1/2 + a_2) = -\frac{1}{2}a_1 + a_1 + a_2 - \frac{1}{2}a_1 - a_2 = 0$  for all  $a_1$  and  $a_2$  (although such arguments do not establish that the linear combinations cover the whole plane—we need [Theorem 3.4.14](#)).



Also,  $\text{span}\{(5, 1, -1/2), (0, -3, -1), (-4, 1, 1)\}$  is the plane  $z = -x/6 + y/3$ , as illustrated below. The reason is that every linear combination of these three vectors fills out the plane:  $a_1(5, 1, -1/2) + a_2(0, -3, -1) + a_3(-4, 1, 1) = (5a_1 - 4a_3, a_1 - 3a_2 + a_3, -a_1/2 -$



$a_2 + a_3$ ) and so lies in the plane as  $-x/6 + y/3 - z = -\frac{1}{6}(5a_1 - 4a_3) + \frac{1}{3}(a_1 - 3a_2 + a_3) - (-a_1/2 - a_2 + a_3) = -\frac{5}{6}a_1 + \frac{2}{3}a_3 + \frac{1}{3}a_1 - a_2 + \frac{1}{3}a_3 + \frac{1}{2}a_1 + a_2 - a_3 = 0$  for all  $a_1, a_2$  and  $a_3$ .



**Example 3.4.9.** Find a set of two vectors that spans the plane  $x - 2y + 3z = 0$ .

Such subspaces connect with matrices. The connection is via a

matrix whose columns are the vectors appearing within the span. Although sometimes we also use the rows of the matrix to be the vectors in the span.

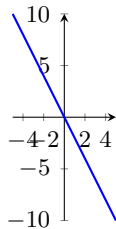
**Definition 3.4.10.** (a) The **column space** of any  $m \times n$  matrix  $A$  is the subspace of  $\mathbb{R}^m$  spanned by the  $n$  column vectors of  $A$ .

(b) The **row space** of any  $m \times n$  matrix  $A$  is the subspace of  $\mathbb{R}^n$  spanned by the  $m$  row vectors of  $A$ .

**Example 3.4.11.** Examples 3.4.7–3.4.9 provide some cases.

- From Example 3.4.7, the column space of  $A = \begin{bmatrix} 1 & -2 \\ \frac{1}{2} & -1 \end{bmatrix}$  is the line  $y = x/2$ .

The row space of this matrix  $A$  is  $\text{span}\{(1, -2), (\frac{1}{2}, -1)\}$ . This row space is the set of all vectors of the form  $(1, -2)s + (\frac{1}{2}, -1)t = (s + t/2, -2s - t) = (1, -2)(s + t/2) = (1, -2)t'$  is the line  $y = -2x$  as illustrated in the margin. That the row space and the column space are both lines, albeit different lines, is not a coincidence (Theorem 3.4.32).



- Example 3.4.8 shows that the column space of matrix

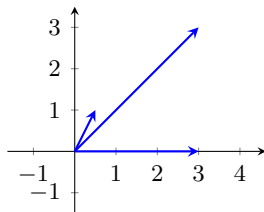
$$B = \begin{bmatrix} 3 & 0 \\ 3 & 3 \\ \frac{1}{2} & 1 \end{bmatrix}$$

is the plane  $z = -x/6 + y/3$  in  $\mathbb{R}^3$ .

The row space of matrix  $B$  is  $\text{span}\{(3, 0), (3, 3), (\frac{1}{2}, 1)\}$  which is a subspace of  $\mathbb{R}^2$ , whereas the column space is a subspace of  $\mathbb{R}^3$ . Here the span is all of  $\mathbb{R}^2$  as for each  $(x, y) \in \mathbb{R}^2$  choose the linear combination  $\frac{x-y}{3}(3, 0) + \frac{y}{3}(3, 3) + 0(\frac{1}{2}, 1) = (x - y + y + 0, 0 + y + 0) = (x, y)$  so each  $(x, y)$  is in the span, and hence all of the  $\mathbb{R}^2$  plane is the span. That the column space and the row space are both planes is no coincidence (Theorem 3.4.32).

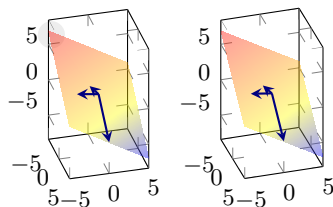
- Example 3.4.8 also shows that the column space of matrix

$$C = \begin{bmatrix} 5 & 0 & -4 \\ 1 & -3 & 1 \\ -\frac{1}{2} & -1 & 1 \end{bmatrix}$$



is also the plane  $z = -x/6 + y/3$  in  $\mathbb{R}^3$ .

Now,  $\text{span}\{(5, 0, -4), (1, -3, 1), (-\frac{1}{2}, -1, 1)\}$  is the row space of matrix  $C$ . It is not readily apparent but we can check that this space is the plane  $4x + 3y + 5z = 0$  as illustrated below in stereo. To see this, consider all linear combinations  $a_1(5, 0, -4) + a_2(1, -3, 1) + a_3(-\frac{1}{2}, -1, 1) = (5a_1 + a_2 - a_3/2, -3a_2 - a_3, -4a_1 + a_2 + a_3)$  satisfy  $4x + 3y + 5z = 4(5a_1 + a_2 - a_3/2) + 3(-3a_2 - a_3) + 5(-4a_1 + a_2 + a_3) = 20a_1 + 4a_2 - 2a_3 - 9a_2 - 3a_3 - 20a_1 + 5a_2 + 5a_3 = 0$ .



Again, it is no coincidence that the row and column spaces of  $C$  are both planes ([Theorem 3.4.32](#)).



**Activity 3.4.12.** Which one of the following vectors is in the column space of the matrix

$$\begin{bmatrix} 6 & 2 \\ -3 & 5 \\ -2 & -1 \end{bmatrix}?$$

(a)  $\begin{bmatrix} 2 \\ 2 \\ -3 \end{bmatrix}$

(b)  $\begin{bmatrix} 2 \\ -3 \\ -3 \end{bmatrix}$

(c)  $\begin{bmatrix} 8 \\ 2 \\ -3 \end{bmatrix}$

(d)  $\begin{bmatrix} 8 \\ 5 \\ -2 \end{bmatrix}$



**Example 3.4.13.** Is vector  $\mathbf{b} = (-0.6, 0, -2.1, 1.9, 1.2)$  in the column space of matrix

$$A = \begin{bmatrix} 2.8 & -3.1 & 3.4 \\ 4.0 & 1.7 & 0.8 \\ -0.4 & -0.1 & 4.4 \\ 1.0 & -0.4 & -4.7 \\ -0.3 & 1.9 & 0.7 \end{bmatrix} ?$$

What about vector  $\mathbf{c} = (15.2, 5.4, 3.8, -1.9, -3.7)$ ?



Another subspace associated with matrices is the set of possible solutions to a homogeneous system of linear equations.

**Theorem 3.4.14.** *For any  $m \times n$  matrix  $A$ , define the set  $\text{null}(A)$  to be all the solutions  $\mathbf{x}$  of the homogeneous system  $A\mathbf{x} = \mathbf{0}$ . The set  $\text{null}(A)$  is a subspace of  $\mathbb{R}^n$  called the **nullspace** of  $A$ .*

**Example 3.4.15.**

- [Example 2.2.29a](#) showed that the only solution of the homogeneous system  $\begin{cases} 3x_1 - 3x_2 = 0 \\ -x_1 - 7x_2 = 0 \end{cases}$  is  $\mathbf{x} = \mathbf{0}$ . Thus its set of solutions is  $\{\mathbf{0}\}$  which is a subspace ([Example 3.4.4f](#)). Thus  $\{\mathbf{0}\}$  is the nullspace of matrix  $\begin{bmatrix} 3 & -3 \\ -1 & -7 \end{bmatrix}$ .
- Recall the homogeneous system of linear equations from [Example 2.2.29d](#) has solutions  $\mathbf{x} = (-2s - \frac{15}{7}t, s, \frac{9}{7}t, t) = (-2, 1, 0, 0)s + (-\frac{15}{7}, 0, \frac{9}{7}, 1)t$  for arbitrary  $s$  and  $t$ . That is, the set of solutions is  $\text{span}\{(-2, 1, 0, 0), (-\frac{15}{7}, 0, \frac{9}{7}, 1)\}$ . Since the set is a span ([Theorem 3.4.6](#)), the set of solutions is a subspace of  $\mathbb{R}^4$ . Thus this set of solutions is the nullspace of the matrix  $\begin{bmatrix} 1 & 2 & 4 & -3 \\ 1 & 2 & -3 & 6 \end{bmatrix}$ .
- In contrast, [Example 2.2.26](#) shows that the set of solutions of the *non*-homogeneous system  $\begin{cases} -2v + 3w = -1, \\ 2u + v + w = -1. \end{cases}$  is  $(u, v, w) = (-\frac{3}{4} - \frac{1}{4}t, \frac{1}{2} + \frac{3}{2}t, t) = (-\frac{3}{4}, \frac{1}{2}, 0) + (-\frac{1}{4}, \frac{3}{2}, 1)t$  over all values of parameter  $t$ . But there is no value of parameter  $t$  giving  $\mathbf{0}$  as a solution: for the last component to be zero requires  $t = 0$ , but

when  $t = 0$  neither of the other components are zero, so they cannot all be zero. Since the origin  $\mathbf{0}$  is not in the set of solutions, the set does not form a subspace. A *non*-homogeneous system does not form a subspace of solutions. ■

**Example 3.4.16.** Given the matrix

$$A = \begin{bmatrix} 3 & 1 & 0 \\ -5 & -1 & -4 \end{bmatrix},$$

is vector  $\mathbf{v} = (-2, 6, 1)$  in the null space of  $A$ ? What about vector  $\mathbf{w} = (1, -3, 2)$ ? ■



**Activity 3.4.17.** Which vector is in the nullspace of the matrix

$$\begin{bmatrix} 4 & 5 & 1 \\ 4 & 3 & -1 \\ 4 & 2 & -2 \end{bmatrix}?$$

(a)  $\begin{bmatrix} -1 \\ 0 \\ 4 \end{bmatrix}$

(b)  $\begin{bmatrix} 2 \\ -2 \\ 2 \end{bmatrix}$

(c)  $\begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}$

(d)  $\begin{bmatrix} 3 \\ -4 \\ 0 \end{bmatrix}$



**Summary** Three common ways that subspaces arise from a matrix are as the column space, row space, and nullspace.

### 3.4.2 Orthonormal bases form a foundation

The importance of orthogonal basis functions in interpolation and approximation cannot be overstated.

(*Cuyt 2015*, §5.3)

Given that subspaces arise frequently in linear algebra, and that there are many ways of representing the same subspace (as seen in some previous examples), is there a ‘best’ way of representing subspaces? The next definition and theorems largely answer this question.

We prefer to use an orthonormal set of vectors to span a subspace. The virtue is that orthonormal sets have many practically useful properties. For example, orthonormal sets underpin JPEG images, our understanding of vibrations, and reliable weather forecasting. Recall that an orthonormal set is composed of vectors that are both at right-angles to each other (their dot products are zero) and all of unit length ([Definition 3.2.38](#)).

**Definition 3.4.18.** An *orthonormal basis* for a subspace  $\mathbb{W}$  of  $\mathbb{R}^n$  is an orthonormal set of vectors that span  $\mathbb{W}$ .

**Example 3.4.19.** Recall that  $\mathbb{R}^n$  is itself a subspace of  $\mathbb{R}^n$  (Example 3.4.4g).

- (a) The  $n$  standard unit vectors  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$  in  $\mathbb{R}^n$  form a set of  $n$  orthonormal vectors. They span the subspace  $\mathbb{R}^n$  as every vector in  $\mathbb{R}^n$  can be written as a linear combination  $\mathbf{x} = (x_1, x_2, \dots, x_n) = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \dots + x_n\mathbf{e}_n$ . Hence the set of standard unit vectors in  $\mathbb{R}^n$  are an orthonormal basis for the subspace  $\mathbb{R}^n$ .
- (b) The  $n$  columns  $\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n$  of an  $n \times n$  orthogonal matrix  $Q$  also form an orthonormal basis for the subspace  $\mathbb{R}^n$ . The reasons are: first, Theorem 3.2.48b establishes the column vectors of  $Q$  are orthonormal; and second they span the subspace  $\mathbb{R}^n$  as for every vector  $\mathbf{x} \in \mathbb{R}^n$  there exists a linear combination  $\mathbf{x} = c_1\mathbf{q}_1 + c_2\mathbf{q}_2 + \dots + c_n\mathbf{q}_n$  obtained by solving  $Q\mathbf{c} = \mathbf{x}$  through calculating  $\mathbf{c} = Q^T\mathbf{x}$  since  $Q^T$  is the inverse of an orthogonal matrix  $Q$  (Theorem 3.2.48c).

This example also illustrates that generally there are many different

orthonormal bases for a given subspace. ■

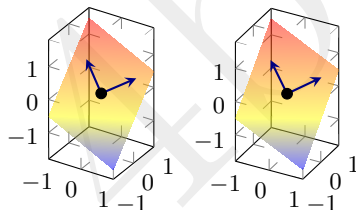
**Activity 3.4.20.** Which of the following sets is an orthonormal basis for  $\mathbb{R}^2$ ?

- (a)  $\{\frac{1}{2}(1, \sqrt{3}), \frac{1}{2}(-\sqrt{3}, 1)\}$       (b)  $\{(1, 1), (1, -1)\}$   
(c)  $\{\mathbf{0}, \mathbf{i}, \mathbf{j}\}$       (d)  $\{\frac{1}{5}(3, -4), \frac{1}{13}(12, 5)\}$
- 

**Example 3.4.21.** Find an orthonormal basis for the line  $x = y = z$  in  $\mathbb{R}^3$ . ■

For subspaces that are planes in  $\mathbb{R}^n$ , orthonormal bases have more details to confirm as in the next example. The SVD then empowers us to find such bases as in the next [Procedure 3.4.23](#).

**Example 3.4.22.** Confirm that the plane  $-x + 2y - 2z = 0$  has an orthonormal basis  $\{\mathbf{u}_1, \mathbf{u}_2\}$  where  $\mathbf{u}_1 = (-\frac{2}{3}, \frac{1}{3}, \frac{2}{3})$ , and  $\mathbf{u}_2 = (\frac{2}{3}, \frac{2}{3}, \frac{1}{3})$  as illustrated in stereo below.



■

**Procedure 3.4.23** (orthonormal basis for a span). *Let  $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}$  be a set of  $n$  vectors in  $\mathbb{R}^m$ , then the following procedure finds an orthonormal basis for the subspace  $\text{span}\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}$ .*

1. Form matrix  $A := [\mathbf{a}_1 \ \mathbf{a}_2 \ \cdots \ \mathbf{a}_n]$ .
2. Factorise  $A$  into an SVD,  $A = USV^T$ , let  $\mathbf{u}_j$  denote the columns of  $U$  (singular vectors), and let  $r = \text{rank } A$  be the

number of nonzero singular values (*Definition 3.3.19*).

3. Then  $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r\}$  is an orthonormal basis for the subspace  $\text{span}\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}$ .

**Example 3.4.24.** Compute an orthonormal basis for  $\text{span}\{(1, \frac{1}{2}), (-2, -1)\}$ . ■

**Example 3.4.25.** Recall that *Example 3.4.8* found the plane  $z = -x/6 + y/3$  could be written as  $\text{span}\{(3, 3, 1/2), (0, 3, 1)\}$  or as  $\text{span}\{(5, 1, -1/2), (0, -3, -1), (-4, 1, 1)\}$ . Use each of these spans to find two different orthonormal bases for the plane. ■

**Activity 3.4.26.** The matrix

$$A = \begin{bmatrix} 4 & 5 & 1 \\ 4 & 3 & -1 \\ 4 & 2 & -2 \end{bmatrix}$$

has the following SVD computed via MATLAB/Octave command  $[U,S,V]=\text{svd}(A)$ : what is an orthonormal basis for the column space of the matrix  $A$  (2 d.p.)?

$$U = \begin{pmatrix} -0.67 & 0.69 & 0.27 \\ -0.55 & -0.23 & -0.80 \\ -0.49 & -0.69 & 0.53 \end{pmatrix}$$

$$S = \begin{pmatrix} 9.17 & 0 & 0 \\ 0 & 2.83 & 0 \\ 0 & 0 & 0.00 \end{pmatrix}$$

$$V = \begin{pmatrix} -0.75 & -0.32 & -0.58 \\ -0.66 & 0.49 & 0.58 \\ 0.09 & 0.81 & -0.58 \end{pmatrix}$$

- (a)  $\{(-0.75, -0.32, -0.58), (-0.66, 0.49, 0.58)\}$
- (b)  $\{(-0.67, -0.55, -0.49), (0.69, -0.23, -0.69)\}$
- (c)  $\{(-0.67, 0.69, 0.27), (-0.55, -0.23, -0.80)\}$

$$(d) \{(-0.75, -0.66, 0.09), (-0.32, 0.49, 0.81)\}$$

Extension: recalling [Theorem 3.3.23](#), which of the above is an orthonormal basis for the row space of  $A$ ? ■

**Example 3.4.27** (data reduction). Every four or five years the phenomenon of El Nino makes a large impact on the world's weather: from drought in Australia to floods in South America. We would like to predict El Nino in advance to save lives and economies. El Nino is correlated significantly with the difference in atmospheric pressure between Darwin and Tahiti—the so-called Southern Oscillation Index (SOI). This example seeks patterns in the SOI in order to be able to predict the SOI and hence predict El Nino.

[Figure 3.1](#) plots the yearly average SOI each year for fifty years up to 1993. A strong regular structure is apparent, but there are significant variations and complexities in the year-to-year signal. The challenge of this example is to explore the full details of this signal.



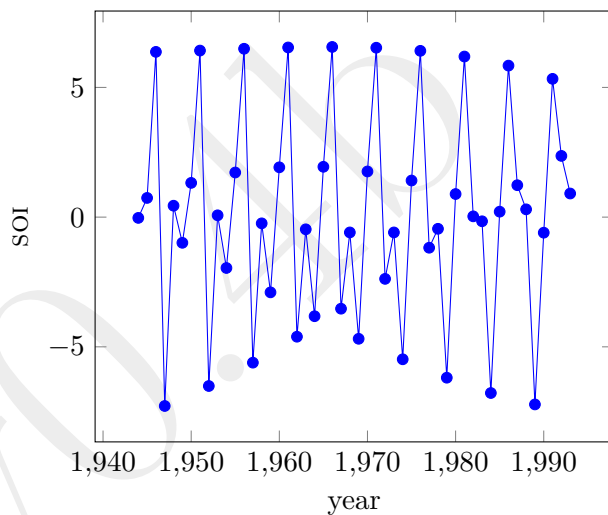


Figure 3.1: yearly average SOI over fifty years ('smoothed' somewhat for the purposes of the example). The nearly regular behaviour suggests it should be predictable.

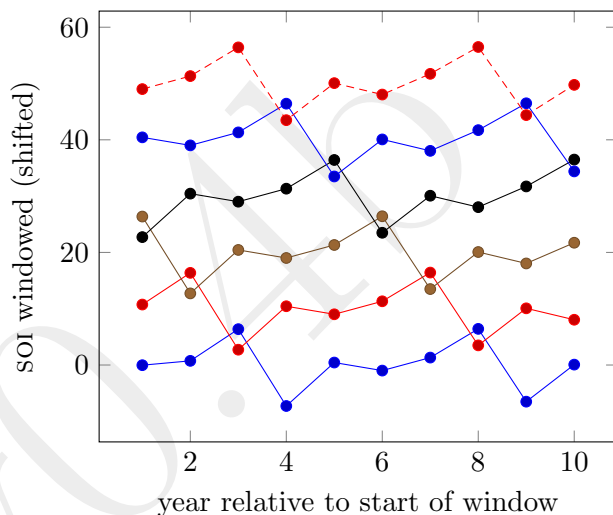


Figure 3.2: the first six windows of the SOI data of Figure 3.1—displaced vertically for clarity. Each window is of length ten years: lowest, the first window is data 1944–1953; second lowest, the second is 1945–1954; third lowest, covers 1946–1955; and so on to the 41st window is data 1984–1993, not shown.

Let's use a general technique called a Singular Spectrum Analysis. Consider a window of ten years of the SOI, and let the window 'slide' across the data to give us many 'local' pictures of the evolution in time. For example, [Figure 3.2](#) plots six windows (each displaced vertically for clarity) each of length ten years. As the 'window' slides across the fifty year data of [Figure 3.1](#) there are 41 such local views of the data of length ten years. Let's invoke the concept of subspaces to detect regularity in the data via these windows.

The fundamental property is that if the data has regularities, then it should lie in some subspace. We detect such subspaces using the SVD of a matrix.

- First form the 41 data windows of length ten into a matrix of size  $10 \times 41$ . The numerical values of the SOI data of [Figure 3.1](#) are the following:

```
year=(1944:1993)'  
soi=[-0.03; 0.74; 6.37; -7.28; 0.44; -0.99; 1.32  
6.42; -6.51; 0.07; -1.96; 1.72; 6.49; -5.61  
-0.24; -2.90; 1.92; 6.54; -4.61; -0.47; -3.82
```

1.94; 6.56; -3.53; -0.59; -4.69; 1.76; 6.53  
 -2.38; -0.59; -5.48; 1.41; 6.41; -1.18; -0.45  
 -6.19; 0.89; 6.19; 0.03; -0.16; -6.78; 0.21; 5.84  
 1.23; 0.30; -7.22; -0.60; 5.33; 2.36; 0.91 ]

- Second form the  $10 \times 41$  matrix of the windows of the data: the first seven columns being

A =  
 Columns 1 through 7

-0.03	0.74	6.37	-7.28	0.44	-0.99	1.32
0.74	6.37	-7.28	0.44	-0.99	1.32	6.42
6.37	-7.28	0.44	-0.99	1.32	6.42	-6.51
-7.28	0.44	-0.99	1.32	6.42	-6.51	0.07
0.44	-0.99	1.32	6.42	-6.51	0.07	-1.96
-0.99	1.32	6.42	-6.51	0.07	-1.96	1.72
1.32	6.42	-6.51	0.07	-1.96	1.72	6.49
6.42	-6.51	0.07	-1.96	1.72	6.49	-5.61
-6.51	0.07	-1.96	1.72	6.49	-5.61	-0.24
0.07	-1.96	1.72	6.49	-5.61	-0.24	-2.90

Figure 3.2 plots the first six of these columns. The simplest way to form this matrix in MATLAB/Octave—useful for all such shifting windows of data—is to invoke the `hankel()` function:

```
A=hankel(soi(1:10),soi(10:50))
```

In MATLAB/Octave the command `hankel(s(1:w),s(w:n))` forms the  $w \times (n - w + 1)$  so-called Hankel matrix

$$\begin{bmatrix} s_1 & s_2 & s_3 & \cdots & s_{n-w} & s_{n-w+1} \\ s_2 & s_3 & \vdots & & s_{n-w+1} & \vdots \\ s_3 & \vdots & s_w & & \vdots & \vdots \\ \vdots & s_w & s_{w+1} & & \vdots & s_{n-1} \\ s_w & s_{w+1} & s_{w+2} & \cdots & s_{n-1} & s_n \end{bmatrix}$$

- Lastly, compute the SVD of the matrix of these windows:

```
[U,S,V]=svd(A);  
singValues=diag(S)  
plot(U(:,1:4))
```



The computed singular values are 44.63, 43.01, 39.37, 36.69, 0.03, 0.03, 0.02, 0.02, 0.02, 0.01. In practice, treat the six small singular values as zero. Since there are four ‘non-zero’ singular values, the windows of data lie in a subspace spanned by the first four columns of  $U$ .

That is, all the structure seen in the fifty year SOI data of [Figure 3.1](#) can be expressed in terms of the orthonormal basis of the four ten-year vectors plotted in [Figure 3.3](#). This analysis implies the SOI data is composed of two cycles of two different frequencies. ■

[Example 3.4.25](#) obtained two different orthonormal bases for the one plane. Although the bases are different, they both had the same number of vectors. The next theorem establishes that this same number always occurs.

**Theorem 3.4.28.**     *For every given subspace, any two orthonormal bases have the same number of vectors.*

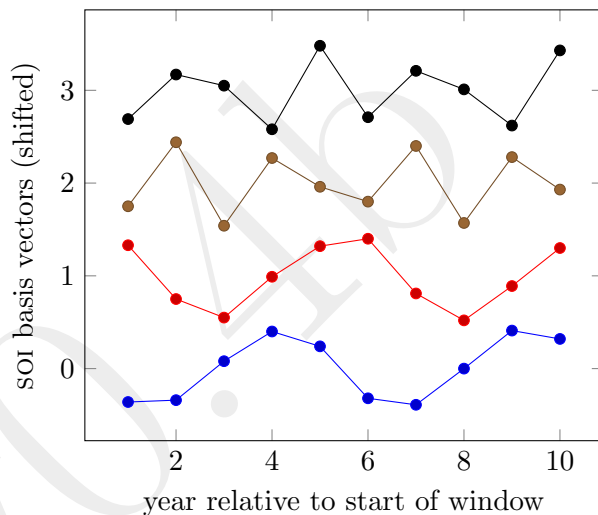


Figure 3.3: first four singular vectors of the SOI data—displaced vertically for clarity. The bottom two form a pair to show a five year cycle. The top two are a pair that show a two–three year cycle. The combination of these two cycles leads to the structure of the SOI in [Figure 3.1](#).

The following optional theorem and proof settles an existential issue.

**An existential issue** How do we know that every subspace has an orthonormal basis? We know many subspaces, such as row and column spaces, have an orthonormal basis because they are the span of rows and columns of a matrix, and then [Procedure 3.4.23](#) assures us they have an orthonormal basis. But do all subspaces have an orthonormal basis? The following theorem certifies that they do.

**Theorem 3.4.29** (existence of basis). *Let  $\mathbb{W}$  be a subspace of  $\mathbb{R}^n$ , then there exists an orthonormal basis for  $\mathbb{W}$ .*

**Ensemble simulation makes better weather forecasts** Near the end of the twentieth century weather forecasts were becoming amazingly good at predicting the chaotic weather days in advance. However, there were notable failures: occasionally the weather forecast would give no hint of storms that developed (such as the severe 1999 storm in Sydney) Why?

Occasionally the weather is both near a ‘tipping point’ where small changes may cause a storm, and where the errors in measuring the



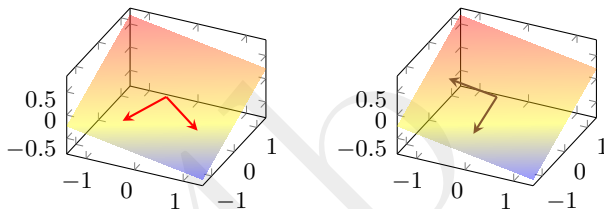
current weather are of the size of the necessary changes. Then the storm would be within the possibilities, but it would not be forecast if the measurements were, by chance error, the ‘other side’ of the tipping point. Meteorologists now mostly overcome this problem by executing on their computers an ensemble of simulations, perhaps an ensemble of a hundred different forecast simulations (Roulstone & Norbury 2013, pp.274–80, e.g.). Such a set of 100 simulations essentially lie in a subspace spanned by 100 vectors in the vastly larger space, say  $\mathbb{R}^{1000,000,000}$ , of the maybe billion variables in the weather model. But what happens in the computational simulations is that the ensemble of simulations degenerate in time: so the meteorologists continuously ‘renormalise’ the ensemble of simulations by rewriting the ensemble in terms of an *orthonormal basis* of 100 vectors. Such an orthonormal basis for the ensemble reasonably ensures unusual storms are retained in the range of possibilities explored by the ensemble forecast.

### 3.4.3 Is it a line? a plane? The dimension answers

*physical dimension.* It is an intuitive notion that appears to go back to an archaic state before Greek geometry, yet deserves to be taken up again.

*Mandelbrot (1982)*

One of the beauties of an orthonormal basis is that, being orthonormal, they look just like a rotated version of the standard unit vectors. That is, the two orthonormal basis of a plane could form the two ‘standard unit vectors’ of a coordinate system in that plane. [Example 3.4.25](#) found the plane  $z = -x/6 + y/3$  could have the following two orthonormal bases: either of these orthonormal bases, or indeed any other pair of orthonormal vectors, could act as a pair of ‘standard unit vectors’ of the given planar subspace.



Similarly in other dimensions for other subspaces. Just as  $\mathbb{R}^n$  is called  $n$ -dimensional and has  $n$  standard unit vectors, so we analogously define the dimension of any subspace.

**Definition 3.4.30.** *Let  $\mathbb{W}$  be a subspace of  $\mathbb{R}^n$ . The number of vectors in an orthonormal basis for  $\mathbb{W}$  is called the **dimension** of  $\mathbb{W}$ , denoted  $\dim \mathbb{W}$ . By convention,  $\dim\{\mathbf{0}\} = 0$ .*

**Example 3.4.31.**

- [Example 3.4.21](#) finds that the linear subspace  $x = y = z$  is spanned by the orthonormal basis  $\{(1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})\}$ . With one vector in the basis, the line is one dimensional.
- [Example 3.4.22](#) finds that the planar subspace  $-x + 2y - 2z = 0$  is spanned by the orthonormal basis  $\{\mathbf{u}_1, \mathbf{u}_2\}$  where  $\mathbf{u}_1 =$

$(-\frac{2}{3}, \frac{1}{3}, \frac{2}{3})$ , and  $\mathbf{u}_2 = (\frac{2}{3}, \frac{2}{3}, \frac{1}{3})$ . With two vectors in the basis, the plane is two dimensional.

- Subspace  $\mathbb{W} = \text{span}\{(5, 1, -1/2), (0, -3, -1), (-4, 1, 1)\}$  of [Example 3.4.25](#) is found to have an orthonormal basis of the vectors  $(-0.99, -0.01, 0.16)$  and  $(-0.04, -0.95, -0.31)$ . With two vectors in the basis, the subspace is two dimensional; that is,  $\dim \mathbb{W} = 2$ .
- Since the subspace  $\mathbb{R}^n$  ([Example 3.4.4g](#)) has an orthonormal basis of the  $n$  standard unit vectors,  $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ , then  $\dim \mathbb{R}^n = n$ .
- The El Nino windowed data of [Example 3.4.27](#) is effectively spanned by four orthonormal vectors. Despite the apparent complexity of the signal, the data effectively lies in a subspace of dimension four (that of two oscillators).



**Theorem 3.4.32.** *The row space and column space of a matrix  $A$  have the same dimension. Further, given an SVD of the matrix, say  $A = USV^T$ , an orthonormal basis for the column space is the first  $\text{rank } A$  columns of  $U$ , and that for the row space is the first  $\text{rank } A$  columns of  $V$ .*

**Example 3.4.33.** Find an SVD of the matrix  $A = \begin{bmatrix} 1 & -4 \\ 1/2 & -2 \end{bmatrix}$  and compare the column space and the row space of the matrix. ■

**Activity 3.4.34.** Using the SVD of [Example 3.4.33](#), what is the dimension of the nullspace of the matrix  $\begin{bmatrix} 1 & -4 \\ 1/2 & -2 \end{bmatrix}$ ?

(a) 1

(b) 0

(c) 2

(d) 3



**Example 3.4.35.** Use the SVD of the matrix  $B$  in [Example 3.4.25](#) to compare the column space and the row space of matrix  $B$ . ■

**Definition 3.4.36.** The *nullity* of a matrix  $A$  is the dimension of its nullspace (defined in [Theorem 3.4.14](#)), and is denoted by  $\text{nullity}(A)$ .

**Example 3.4.37.** [Example 3.4.15](#) finds the nullspace of the two matrices

$$\begin{bmatrix} 3 & -3 \\ -1 & -7 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 2 & 4 & -3 \\ 1 & 2 & -3 & 6 \end{bmatrix}.$$

- The first matrix has nullspace  $\{\mathbf{0}\}$  which has dimension zero and hence the nullity of the matrix is zero.
- The second matrix,  $2 \times 4$ , has nullspace written as  $\text{span}\{(-2, 1, 0, 0), (-\frac{15}{7}, 0, \frac{9}{7}, 1)\}$ . Being spanned by two vectors not proportional to each other, we expect the dimension of the nullspace, the nullity, to be two. To check, compute the singular values of the matrix whose columns are these vectors: calling the matrix  $N$  for nullspace,



```
N=[-2 1 0 0; -15/7 0 9/7 1]'
```

```
svd(N)
```

which computes the singular values

```
3.2485
```

```
1.3008
```

Since there are two non-zero singular values, there are two orthonormal vectors spanning the subspace, the nullspace, hence its dimension, the nullity, is two.



**Example 3.4.38.** For the matrix

$$C = \begin{bmatrix} -1 & -2 & 2 & 1 \\ -3 & 3 & 1 & 0 \\ 2 & -5 & 1 & 1 \end{bmatrix},$$

find an orthonormal basis for its nullspace and hence determine its nullity.



This [Example 3.4.38](#) indicates that the nullity is determined by the number of zero columns in the diagonal matrix  $S$  of an SVD. Conversely, the rank of a matrix is determined by the number of non-zero columns in the diagonal matrix  $S$  of an SVD. Put these two facts together in general and we get the following theorem that helps characterise solutions of linear equations.

**Theorem 3.4.39 (rank theorem).** *For every  $m \times n$  matrix  $A$ ,  $\text{rank } A + \text{nullity } A = n$ , the number of columns of  $A$ .*

**Example 3.4.40.** Compute SVDs to determine the rank and nullity of each of the given matrices.

(a) 
$$\begin{bmatrix} 1 & -1 & 2 \\ 2 & -2 & 4 \end{bmatrix}$$

(b) 
$$\begin{bmatrix} 1 & -1 & -1 \\ 1 & 0 & -1 \\ -1 & 3 & 1 \end{bmatrix}$$



(c) 
$$\begin{bmatrix} 0 & 0 & -1 & -3 & 2 \\ -2 & -2 & 1 & 0 & 1 \\ 1 & -1 & 2 & 8 & -2 \\ -1 & 1 & 0 & -2 & -2 \\ -3 & -1 & 0 & -5 & 1 \end{bmatrix}$$

**Activity 3.4.41.** The matrix

$$\begin{bmatrix} -2 & 1 & 4 & 0 & -4 \\ -1 & 1 & 0 & -2 & 0 \\ -3 & 1 & 3 & 2 & -3 \\ 0 & 0 & 1 & 0 & -1 \end{bmatrix} \text{ has singular values } \begin{matrix} 8.1975 \\ 2.6561 \\ 1.6572 \\ 0.0000 \end{matrix}$$

computed with `svd()`. What is its nullity?

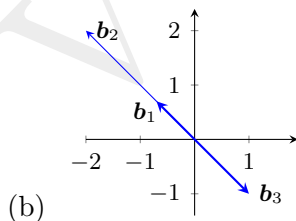
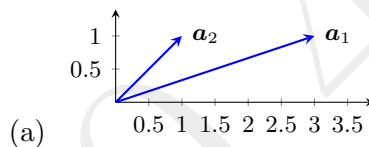
(a) 2

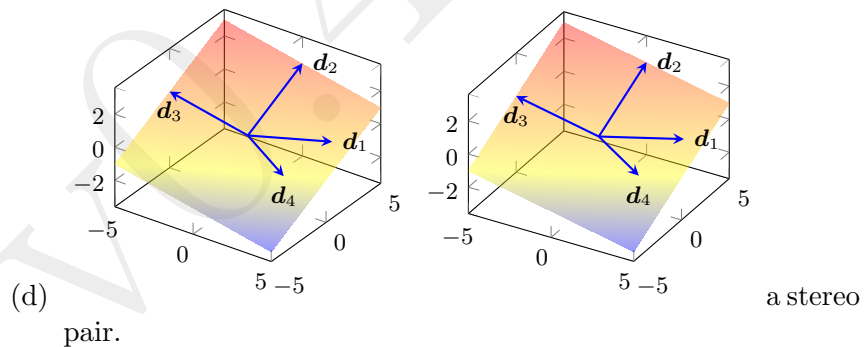
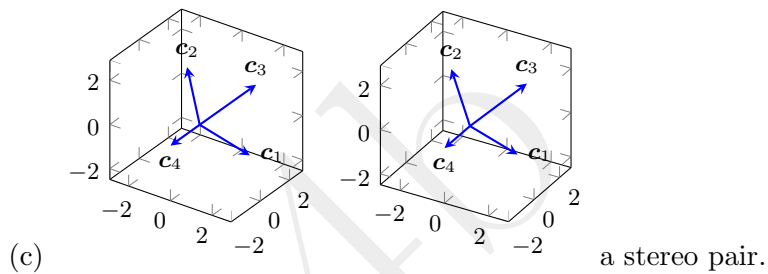
(b) 0

(c) 1

(d) 3

**Example 3.4.42.** Each of the following graphs plot all the column vectors of a matrix. What is the nullity of each of the matrices? Give reasons.





The recognition of these new concepts associated with matrices and linear equations, then empowers us to extend the list of exact properties that ensure a system of linear equations has a unique solution.

**Theorem 3.4.43** (Unique Solutions: version 2). *For every  $n \times n$  square matrix  $A$ , and extending [Theorem 3.3.26](#), the following statements are equivalent:*

- (a)  $A$  is invertible;
- (b)  $A\mathbf{x} = \mathbf{b}$  has a unique solution for every  $\mathbf{b} \in \mathbb{R}^n$ ;
- (c)  $A\mathbf{x} = \mathbf{0}$  has only the zero solution;
- (d) all  $n$  singular values of  $A$  are nonzero;
- (e) the condition number of  $A$  is finite ( $\mathbf{rcond} > 0$ );
- (f)  $\text{rank } A = n$ ;
- (g)  $\text{nullity } A = 0$ ;
- (h) the column vectors of  $A$  span  $\mathbb{R}^n$ ;

(i) the row vectors of  $A$  span  $\mathbb{R}^n$ .

## 3.5 Project to solve inconsistent equations

### Section Contents

3.5.1	Make a minimal change to the problem . . .	375
3.5.2	Compute the smallest appropriate solution .	395
3.5.3	Orthogonal projection resolves vector components . . . . .	409
	Project onto a direction . . . . .	410
	Project onto a subspace . . . . .	416
	Orthogonal decomposition separates . . . . .	428

As well as being fundamental to engineering, scientific and computational inference, approximately solving inconsistent equations also introduces the linear transformation of projection.

Agreement with experiment is the sole criterion of truth  
for a physical theory. *Pierre Duhem, 1906*

The scientific method is to infer general laws from data and then validate the laws. This section addresses some aspects of the

inference of general laws from data. A big challenge is that data is typically corrupted by noise and errors. So this section shows how the singular value decomposition (SVD) leads to understanding ‘least square methods’ for handling noisy errors.

### 3.5.1 Make a minimal change to the problem

**Example 3.5.1** (rationalise contradictions). I weighed myself the other day. I weighed myself four times, each time separated by a few minutes: the scales reported my weight in kg as 84.8, 84.1, 84.7 and 84.4. The measurements give four different weights! What sense can we make of this apparently contradictory data? Traditionally we just average and say my weight is  $x \approx (84.8 + 84.1 + 84.7 + 84.4)/4 = 84.5$  kg. Let's see this same answer from a new linear algebra view.

In the linear algebra view my weight  $x$  is an unknown and the four experimental measurements give four equations for this one unknown:

$$x = 84.8, \quad x = 84.1, \quad x = 84.7, \quad x = 84.4.$$

Despite being manifestly impossible to satisfy all four equations, let's see what linear algebra can do for us. Linear algebra writes



these four equations as the matrix-vector system

$$Ax = \mathbf{b}, \quad \text{namely} \quad \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} x = \begin{bmatrix} 84.8 \\ 84.1 \\ 84.7 \\ 84.4 \end{bmatrix}.$$

The linear algebra [Procedure 3.3.15](#) is to ‘solve’ this system, despite its contradictions, via an SVD and some intermediaries:

$$Ax = U \underbrace{S \overbrace{V^T x}^{=y}}_{=z} = \mathbf{b}.$$

- (a) You are given that this particular matrix  $A$  of a column of ones has an SVD of

$$A = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \end{bmatrix} [1]^T = USV^T$$

(perhaps check the columns of  $U$  are orthonormal).

(b) Solve  $U\mathbf{z} = \mathbf{b}$  by computing

$$\mathbf{z} = U^T \mathbf{b} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 84.8 \\ 84.1 \\ 84.7 \\ 84.4 \end{bmatrix} = \begin{bmatrix} 169 \\ -0.1 \\ 0.2 \\ 0.5 \end{bmatrix}.$$

(c) Now try to solve  $Sy = \mathbf{z}$ , that is,

$$\begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \end{bmatrix} y = \begin{bmatrix} 169 \\ -0.1 \\ 0.2 \\ 0.5 \end{bmatrix}.$$

But we cannot because the last three components in the equation are impossible: we cannot satisfy any of

$$0y = -0.1, \quad 0y = 0.2, \quad 0y = 0.5.$$

Instead of seeking an *exact* solution, ask what is the *smallest change* we can make to  $\mathbf{z} = (169, -0.1, 0.2, 0.5)$  so that we

can report a solution to a slightly different problem? Answer: we *have to* adjust the last three components to zero. Further, any adjustment to the first component is unnecessary, would make the change to  $z$  bigger than necessary, and so we do not adjust the first component. Hence we solve a slightly different problem, that of

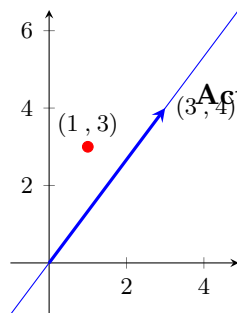
$$\begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \end{bmatrix} y = \begin{bmatrix} 169 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

with solution  $y = 84.5$ . We treat this exact solution to a slightly different problem as an *approximate* solution to the original problem.

- (d) Lastly, solve  $V^T x = y$  by computing  $x = Vy = 1y = y = 84.5$  kg (upon including the physical units). That is, this linear algebra procedure gives my weight as  $x = 84.5$  kg (approximately).

This linear algebra procedure recovers the traditional answer of averaging measurements. ■

The answer to the previous [Example 3.5.1](#) illustrates how traditional averaging emerges from trying to make sense of apparently inconsistent information. Importantly, the principle of making the smallest possible change to the intermediary  $\mathbf{z}$  is equivalent to making the smallest possible change to the original data vector  $\mathbf{b}$ . The reason is that  $\mathbf{b} = U\mathbf{z}$  for an orthogonal matrix  $U$ : since  $U$  is an orthogonal matrix, multiplication by  $U$  preserves distances and angles ([Theorem 3.2.48](#)) and so the smallest possible change to  $\mathbf{b}$  is the same as the smallest possible change to  $\mathbf{z}$ . Scientists and engineers implicitly use this same ‘smallest change’ approach to approximately solve many sorts of inconsistent linear equations.



### Activity 3.5.2.

Consider the inconsistent equations  $3x = 1$  and  $4x = 3$  formed as the system

$$\begin{bmatrix} 3 \\ 4 \end{bmatrix} x = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \quad \text{and where} \quad \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} \frac{3}{5} & \frac{4}{5} \\ \frac{4}{5} & -\frac{3}{5} \end{bmatrix} \begin{bmatrix} 5 \\ 0 \end{bmatrix} [1]^T$$

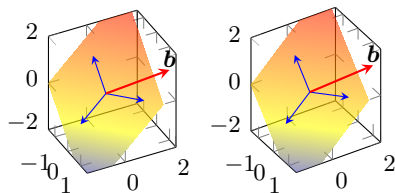
is an SVD factorisation of the  $2 \times 1$  matrix. Following the procedure

of the previous example, what is the ‘best’ approximate solution to these inconsistent equations?

- (a)  $x = 4/7$       (b)  $x = 1/3$       (c)  $x = 3/5$       (d)  $x = 3/4$



**Example 3.5.3.** Recall the table tennis player rating [Example 3.3.13](#). There we found that we could not solve the equations to find some ratings because the equations were inconsistent. In our new terminology of the previous [Section 3.4](#), the right-hand side vector  $\mathbf{b}$  is not in the column space of the matrix  $A$  ([Definition 3.4.10](#)): the stereo picture below illustrates the 2D column space spanned by the three columns of  $A$  and that the vector  $\mathbf{b}$  lies outside the column space.



Now reconsider Step 3 in [Example 3.3.13](#).

- (a) We need to interpret and ‘solve’  $S\mathbf{y} = \mathbf{z}$  which here is

$$\begin{bmatrix} 1.7321 & 0 & 0 \\ 0 & 1.7321 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{y} = \begin{bmatrix} -2.0412 \\ -2.1213 \\ 0.5774 \end{bmatrix}.$$

The third line of this system says  $0y_3 = 0.5774$  which is impossible for any  $y_3$ : we cannot have zero on the left-hand side equalling 0.5774 on the right-hand side. Instead of seeking an *exact* solution, ask what is the *smallest change* we can make to  $\mathbf{z} = (-2.0412, -2.1213, 0.5774)$  so that we can report a solution, albeit to a slightly different problem? Answer: we *must* change the last component of  $\mathbf{z}$  to zero. But any change to the first two components is unnecessary, would make the change bigger than necessary, and so we do not change the first two components. Hence find an approximate solution to the player ratings via solving

$$\begin{bmatrix} 1.7321 & 0 & 0 \\ 0 & 1.7321 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{y} = \begin{bmatrix} -2.0412 \\ -2.1213 \\ 0 \end{bmatrix}.$$

Here a general solution is  $\mathbf{y} = (-1.1785, -1.2247, y_3)$  from  $\mathbf{y} = \mathbf{z}(1:2) ./ \text{diag}(\mathbf{S}(1:2, 1:2))$ . Varying the free variable  $y_3$  gives equally good approximate solutions.

- (b) Lastly, solve  $V^T \mathbf{x} = \mathbf{y}$ , via computing  $\mathbf{x} = V(:, 1:2) * \mathbf{y}$ , to determine

$$\begin{aligned}\mathbf{x} = V\mathbf{y} &= \begin{bmatrix} 0.0000 & -0.8165 & 0.5774 \\ -0.7071 & 0.4082 & 0.5774 \\ 0.7071 & 0.4082 & 0.5774 \end{bmatrix} \begin{bmatrix} -1.1785 \\ -1.2247 \\ y_3 \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ \frac{1}{3} \\ -\frac{4}{3} \end{bmatrix} + \frac{y_3}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.\end{aligned}$$

As before, it is only the relative ratings that are important so we choose any particular (approximate) solution by setting  $y_3$  to anything we like, such as zero. The predicted ratings are then  $\mathbf{x} = (1, \frac{1}{3}, -\frac{4}{3})$  for Anne, Bob and Chris, respectively. ■

The reliability and likely error of such approximate solutions are

the province of Statistics courses. We focus on the geometry and linear algebra of obtaining the ‘best’ approximate solution.

**Procedure 3.5.4 (approximate solution).** *Obtain the so-called ‘least square’ approximate solution(s) of inconsistent equations  $A\mathbf{x} = \mathbf{b}$  using an SVD and via intermediate unknowns:*

1. *factorise  $A = USV^T$  and set  $r = \text{rank } A$  (remembering that relatively small singular values are effectively zero);*
2. *solve  $U\mathbf{z} = \mathbf{b}$  by  $\mathbf{z} = U^T\mathbf{b}$ ;*
3. *disregard the equations for  $i = r + 1, \dots, m$  as errors, set  $y_i = z_i/\sigma_i$  for  $i = 1, \dots, r$  (as these  $\sigma_i > 0$ ), and otherwise  $y_i$  is free for  $i = r + 1, \dots, n$ ;*
4. *solve  $V^T\mathbf{x} = \mathbf{y}$  to obtain a general approximate solution as  $\mathbf{x} = V\mathbf{y}$ .*

**Example 3.5.5.** You are given the choice of two different types of concrete mix. One type contains 40% cement, 40% gravel, and 20% sand; whereas the other type contains 20% cement, 10% gravel,



Table 3.4: the results of six games played in a round robin: the scores are games/goals/points scored by each when playing the others. For example, Dee beat Anne 3 to 1.

	Anne	Bob	Chris	Dee
Anne	-	3	3	1
Bob	2	-	2	4
Chris	0	1	-	2
Dee	3	0	3	-

and 70% sand. How many kilograms of each type should you mix together to obtain a concrete mix as close as possible to 3 kg of cement, 2 kg of gravel, and 4 kg of sand. ■

**Example 3.5.6** (round robin tournament). Consider four players (or teams) that play in a round robin sporting event: Anne, Bob, Chris and Dee. Table 3.4 summarises the results of the six games played. From these results estimate the relative player ratings of the four players. As in many real-life situations, the information appears

contradictory such as Anne beats Bob, who beats Dee, who in turn beats Anne. Assume that the rating  $x_i$  of player  $i$  is to reflect, as best we can, the difference in scores upon playing player  $j$ : that is, pose the difference in ratings,  $x_i - x_j$ , should equal the difference in the scores when they play. ■

When rating players or teams based upon results, be clear the purpose. For example, is the purpose to summarise past performance? or to predict future contests? If the latter, then my limited experience suggests that one should fit the win-loss record instead of the scores. Explore the alternatives for your favourite sport.

**Activity 3.5.7.** Listed below are four approximate solutions to the system  $A\mathbf{x} = \mathbf{b}$ ,

$$\begin{bmatrix} 5 & 3 \\ 3 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 9 \\ 2 \\ 10 \end{bmatrix}.$$

Setting vector  $\tilde{\mathbf{b}} = A\mathbf{x}$  for each, which one minimises the distance between the original right-hand side  $\mathbf{b} = (9, 2, 10)$  and the approxi-

---

Be aware of Kenneth Arrow's Impossibility Theorem (one of the great theorems of the 20th century): *all 1D ranking systems are flawed!* Wikipedia [2014] described the theorem this way (in the context of voting systems): that among

three or more distinct alternatives (options), no rank order voting system can convert the ranked preferences of individuals into a community-wide (complete and transitive) ranking while also meeting [four sensible] criteria . . . called unrestricted domain, non-dictatorship, Pareto efficiency, and independence of irrelevant alternatives.

In rating sport players/teams:

- the “distinct alternatives” are the players/teams;
- the “ranked preferences of individuals” are the individual results of each game played; and
- the “community-wide ranking” is the assumption that we can rate each player/team by a one-dimensional numerical rating.

Arrow's theorem assures us that every such scheme must violate at least one of four sensible criteria. Every ranking scheme is thus open to criticism. But every alternative scheme will also be open to criticism by also violating one of the criteria.

---

mate  $\tilde{\mathbf{b}}$ ?

$$(a) \mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad (b) \mathbf{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \quad (c) \mathbf{x} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \quad (d) \mathbf{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

■

**Theorem 3.5.8 (smallest change).** *All approximate solutions obtained by [Procedure 3.5.4](#) solve the linear system  $A\mathbf{x} = \tilde{\mathbf{b}}$  for the unique*

*The over-tilde on  $\tilde{\mathbf{b}}$  is to suggest a consistent right-hand side vector  $\tilde{\mathbf{b}}$  that minimises the distance  $|\tilde{\mathbf{b}} - \mathbf{b}|$ .*

**Example 3.5.9 (life expectancy).** [Table 3.5](#) lists life expectancies of people born in a given year; [Figure 3.4](#) plots the data points. Over the decades the life expectancies have increased. Let's quantify the overall trend to be able to draw, as in [Figure 3.4](#), the best straight line to the female life expectancy. Solve the approximation problem with an SVD and confirm it gives the same solution as  $A \backslash \mathbf{b}$  in MATLAB/Octave.

■

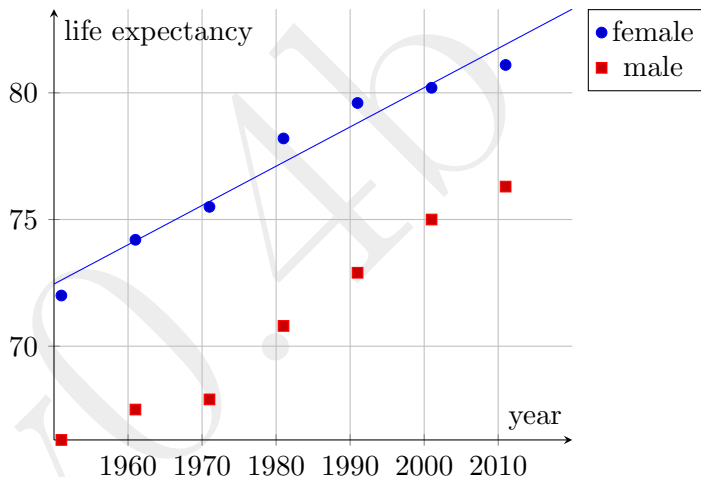


Figure 3.4: the life expectancies in years of females and males born in the given years (Table 3.5). Also plotted is the best straight line fit to the female data obtained by Example 3.5.9.

Table 3.5: life expectancy in years of (white) females and males born in the given years [<http://www.infoplease.com/ipa/A0005140.html>, 2014]. Used by Example 3.5.9.

year	1951	1961	1971	1981	1991	2001	2011
female	72.0	74.2	75.5	78.2	79.6	80.2	81.1
male	66.3	67.5	67.9	70.8	72.9	75.0	76.3

**Activity 3.5.10.** In calibrating a vortex flowmeter the following flow rates were obtained for various applied voltages.

voltage (V)	1.18	1.85	2.43	2.81
flow rate (litre/s)	0.18	0.57	0.93	1.27

Letting  $v_i$  be the voltages and  $f_i$  the flow rates, which of the following is a reasonable model to seek? (for coefficients  $x_1, x_2, x_3$ )

(a)  $v_i = x_1 + x_2 f_i$

(b)  $f_i = x_1$

(c)  $v_i = x_1 + x_2 f_i + x_3 f_i^2$

(d)  $f_i = x_1 + x_2 v_i$

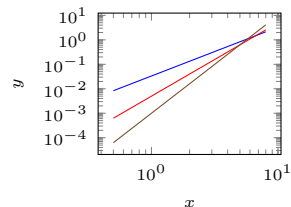
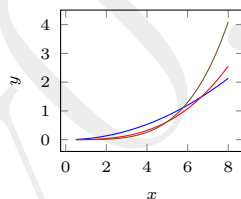
**Example 3.5.11** (planetary orbital periods). Table 3.6 lists each orbital period of the planets of the solar system; Figure 3.5 plots the data points as a function of the distance of the planets from the sun. Let's infer Kepler's law that the period grows as the distance to the power  $3/2$ : shown by the straight line fit in Figure 3.5. Use the data for Mercury to Uranus to infer the law with an SVD, confirm it gives the same solution as  $A \backslash \mathbf{b}$  in MATLAB/Octave, and use the fit to predict Neptune's period from its distance. ■

**Compute in Matlab/Octave.**

There are two separate important computational issues.

- Many books approximate solutions of  $A\mathbf{x} = \mathbf{b}$  by solving the associated normal equation  $(A^T A)\mathbf{x} = (A^T \mathbf{b})$ . For theoretical purposes this normal equation is very useful. However, in practical computation avoid the normal equation because forming  $A^T A$ , and then manipulating it, is both expensive

**Power laws and the log-log plot** Hundreds of power-laws have been identified in engineering, physics, biology and the social sciences. These laws were typically detected via log-log plots. A log-log plot is a two-dimensional graph of the numerical data that uses a logarithmic scale on both the horizontal and vertical axes, as in Figure 3.5. Then curvaceous relationships of the form  $y = cx^a$  between the vertical variable,  $y$ , and the horizontal variable,  $x$ , appear as straight lines on a log-log plot. For example, below-left is plotted the three curves  $y \propto x^2$ ,  $y \propto x^3$ , and  $y \propto x^4$ . It is hard to tell which is which.



However, plot the same curves on the above-right log-log plot and it distinguishes the curves as different straight lines: the steepest line is the curve with the largest exponent,  $y \propto x^4$ , whereas the least-steep line is the curve with the smallest exponent,  $y \propto x^2$ .

For example, suppose you make three measurements that at  $x = 1.8, 3.3, 6.7$  the value of  $y = 0.9, 4.6, 29.1$ , respectively. The graph



Table 3.6: orbital periods for the eight planets of the solar system: the periods are in (Earth) days; the distance is the length of the semi-major axis of the orbits [Wikipedia, 2014]. Used by [Example 3.5.11](#)

planet	distance (Gigametres)	period (days)
Mercury	57.91	87.97
Venus	108.21	224.70
Earth	149.60	365.26
Mars	227.94	686.97
Jupiter	778.55	4332.59
Saturn	1433.45	10759.22
Uranus	2870.67	30687.15
Neptune	4498.54	60190.03

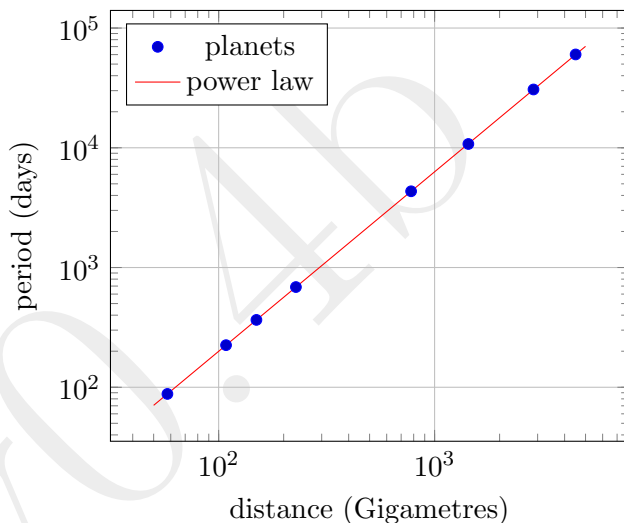


Figure 3.5: the planetary periods as a function of the distance from the data of [Table 3.6](#): the graph is a log-log plot to show the excellent power law. Also plotted is the power law fit computed by [Example 3.5.11](#).

and error enhancing (especially in large problems). For example,  $\text{cond}(A^T A) = (\text{cond } A)^2$  (??) so matrix  $A^T A$  typically has a much worse condition number than matrix  $A$  ([Procedure 2.2.5](#)).

- The last two examples observe that  $A \backslash b$  gives an answer that was identical to what the SVD procedure gives. Thus  $A \backslash b$  can serve as a very useful short-cut to finding a best approximate solution. For non-square matrices with more rows than columns (more equations than variables),  $A \backslash b$  generally does this (without comment as MATLAB/Octave assume you know what you are doing).

### 3.5.2 Compute the smallest appropriate solution

I'm thinking of two numbers. Their average is three.  
What are the numbers? *Cleve Moler, The world's  
simplest impossible problem  
(1990)*

**The Matlab/Octave operation  $A \setminus b$**  Recall that Examples 3.5.5, 3.5.9 and 3.5.11 observed that  $A \setminus b$  gives an answer identical to the best approximate solution given by the SVD Procedure 3.5.4. But there are just as many circumstances when  $A \setminus b$  is not 'the approximate answer' that you want. Beware.

**Example 3.5.12.** Use  $x = A \setminus b$  to 'solve' the problems of Examples 3.5.1, 3.5.3 and 3.5.6.

- With Octave, observe the answer returned is the *particular* solution determined by the SVD Procedure 3.5.4 (whether approximate or exact): respectively 84.5 kg; ratings  $(1, \frac{1}{3}, -\frac{4}{3})$ ; and ratings  $(\frac{1}{2}, 1, -\frac{5}{4}, -\frac{1}{4})$ .

- With MATLAB, the computed answers are often different: respectively 84.5 kg (the same); ratings (NaN, Inf, Inf) with a warning; and ratings ( $\frac{3}{4}, \frac{5}{4}, -1, 0$ ) with a warning.

How do we make sense of such differences in computed answers? ■

Recall that systems of linear equations may not have unique solutions (as in the rating examples): what does  $A \backslash b$  compute when there are an infinite number of solutions?

- For systems of equations with the number of equations not equal to the number of variables,  $m \neq n$ , the Octave operation  $A \backslash b$  computes for you the *smallest solution* of all valid solutions ([Theorem 3.5.13](#)): often ‘exact’ when  $m < n$ , or approximate when  $m > n$  ([Theorem 3.5.8](#)). Using  $A \backslash b$  is the most efficient computationally, but using the SVD helps us understand what it does.
- MATLAB (R2013b) does something different with  $A \backslash b$  in the case of fewer equations than variables,  $m < n$ . MATLAB’s

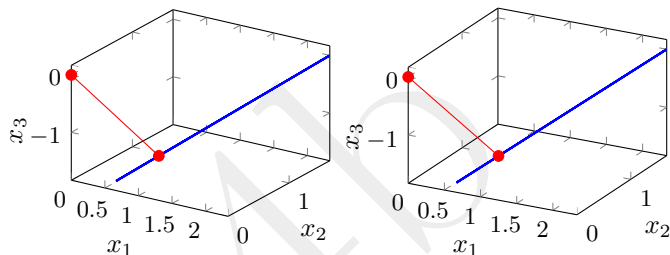
different ‘answer’ does reinforce that a choice of one solution among many is a subjective decision. But Octave’s choice of the smallest valid solution is often more appealing.

**Theorem 3.5.13 (smallest solution).** *Obtain the smallest solution, whether exact or as an approximation, to a system of linear equations by invoking Procedures [3.3.15](#) or [3.5.4](#), respectively, and then setting to zero the free variables,  $y_{r+1} = \cdots = y_n = 0$ .*

**Example 3.5.14.** In the table tennis ratings of [Example 3.5.3](#) the procedure found the ratings were any of

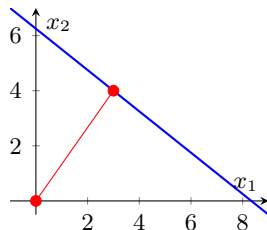
$$\mathbf{x} = \begin{bmatrix} 1 \\ \frac{1}{3} \\ -\frac{4}{3} \end{bmatrix} + \frac{y_3}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},$$

as illustrated in stereo below (blue). Verify  $|\mathbf{x}|$  is a minimum only when the free variable  $y_3 = 0$  (a disc in the plot).



■

**Example 3.5.15** (closest point to the origin). What is the point on the line  $3x_1 + 4x_2 = 25$  that is closest to the origin? I am sure you could think of several methods, perhaps inspired by the marginal graph, but here use an SVD and [Theorem 3.5.13](#). Confirm the Octave computation `A\b` gives this same closest point, but MATLAB gives a different ‘answer’ (one that is not relevant here). ■



**Activity 3.5.16.** What is the closest point to the origin of the plane  $2x + 3y + 6z = 98$ ? Use the SVD

$$\begin{bmatrix} 2 & 3 & 6 \end{bmatrix} = \begin{bmatrix} 1 \end{bmatrix} \begin{bmatrix} 7 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{2}{7} & -\frac{3}{7} & -\frac{6}{7} \\ \frac{3}{7} & \frac{6}{7} & -\frac{2}{7} \\ \frac{6}{7} & -\frac{2}{7} & \frac{3}{7} \end{bmatrix}^T.$$

- (a)  $(-12, -4, 6)$  (b)  $(-3, 6, -2)$   
(c)  $(2, 3, 6)$  (d)  $(4, 6, 12)$

**Example 3.5.17** (computed tomography).

A CT-scan, also called X-ray computed tomography (X-ray CT) or computerized axial tomography scan (CAT scan), makes use of computer-processed combinations of many X-ray images taken from different angles to produce cross-sectional (tomographic) images (virtual



Table 3.7: As well as the MATLAB/Octave commands and operations listed in Tables 1.2, 2.3, 3.1, 3.2, and 3.3 we may invoke these functions for drawing images—functions which are otherwise not needed.

- **reshape(A,p,q)** for a  $m \times n$  matrix/vector  $A$ , provided  $mn = pq$ , generates a  $p \times q$  matrix with entries taken column-wise from  $A$ . Either  $p$  or  $q$  can be `[]` in which case MATLAB/Octave uses  $p = mn/q$  or  $q = mn/p$  respectively.
- **colormap(gray)** MATLAB/Octave usually draws graphs with colour, but for many images we need grayscale; this command changes the current figure to 64 shades of gray.  
(**colormap(jet)** is the default, **colormap(hot)** is good for both colour and grayscale reproductions, **colormap('list')** lists the available colormaps you can try.)
- **imagesc(A)** where  $A$  is a  $m \times n$  matrix of values draws an  $m \times n$  image in the current figure window using the values of  $A$  (scaled to fit) to determine the colour from the current colormap (e.g., grayscale).
- **log(x)** where  $x$  is a matrix, vector or scalar computes the natural logarithm to the base  $e$  of each element, and returns the result(s) as a correspondingly sized matrix, vector or scalar.

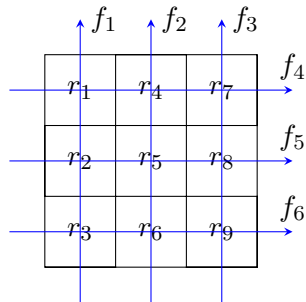
'slices') of specific areas of a scanned object, allowing the user to see inside the object without cutting.

*Wikipedia, 2015*

Importantly for medical diagnosis and industrial purposes, the computed tomography answer must not have artificial features. Artificial features must not be generated because of deficiencies in the measurements. If there is any ambiguity about the answer, then the answer computed should be the 'greyest'—the 'greyest' corresponds to the mathematical smallest solution.

$r_1$	$r_4$	$r_7$
$r_2$	$r_5$	$r_8$
$r_3$	$r_6$	$r_9$

Let's analyse a toy example. Suppose we divide a cross-section of a body into nine squares (large pixels) in a  $3 \times 3$  grid. Inside each square the body's material has some unknown density represented by transmission factors,  $r_1, r_2, \dots, r_9$  as shown in the margin. The CT-scan is to find these transmission factors. The factor  $r_j$  is the fraction of the incident X-ray that emerges after passing through the  $j$ th square: typically, smaller  $r_i$  corresponds to higher density in the body.



Computers almost always use “log” to denote the natural logarithm, so we do too. Herein, unsubscripted “log” means the same as “ln”.

As indicated next in the margin, six X-ray measurements are made through the body where  $f_1, f_2, \dots, f_6$  denote the fraction of energy in the measurements relative to the incident power of the X-ray beam. Thus we need to solve six equations for the nine unknown transmission factors:

$$\begin{aligned} r_1 r_2 r_3 &= f_1, & r_4 r_5 r_6 &= f_2, & r_7 r_8 r_9 &= f_3, \\ r_1 r_4 r_7 &= f_4, & r_2 r_5 r_8 &= f_5, & r_3 r_6 r_9 &= f_6. \end{aligned}$$

Turn such nonlinear equations into linear equations that we can handle by taking the logarithm (to any base, but here say the natural logarithm to base  $e$ ) of both sides of all equations:

$$r_i r_j r_k = f_l \iff (\log r_i) + (\log r_j) + (\log r_k) = (\log f_l).$$

That is, letting new unknowns  $x_i = \log r_i$  and new right-hand sides  $b_i = \log f_i$ , we solve six linear equations for nine unknowns:

$$\begin{aligned} x_1 + x_2 + x_3 &= b_1, & x_4 + x_5 + x_6 &= b_2, & x_7 + x_8 + x_9 &= b_3, \\ x_1 + x_4 + x_7 &= b_4, & x_2 + x_5 + x_8 &= b_5, & x_3 + x_6 + x_9 &= b_6. \end{aligned}$$

This forms the matrix-vector system  $A\mathbf{x} = \mathbf{b}$  for  $6 \times 9$  matrix

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

For example, let's find an answer for the factors when the measurements give vector  $\mathbf{b} = (-0.91, -1.04, -1.54, -1.52, -1.43, -0.53)$  (all negative as they are the logarithms of fractions  $f_i$  less than one)



```
A=[1 1 1 0 0 0 0 0 0
   0 0 0 1 1 1 0 0 0
   0 0 0 0 0 0 1 1 1
   1 0 0 1 0 0 1 0 0
   0 1 0 0 1 0 0 1 0
   0 0 1 0 0 1 0 0 1 ]
b=[-0.91 -1.04 -1.54 -1.52 -1.43 -0.53] '
x=A\b
```

```
r=reshape(exp(x),3,3)
colormap(gray),imagesc(r)
```

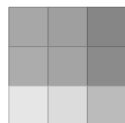
- The answer from Octave is (2 d.p.)

$$\mathbf{x} = (-.42, -.39, -.09, -.47, -.44, -.14, -.63, -.60, -.30).$$

These are logarithms so to get the corresponding physical transmission factors compute the exponential of each component, denoted as  $\exp(\mathbf{x})$ ,

$$\mathbf{r} = \exp(\mathbf{x}) = (.66, .68, .91, .63, .65, .87, .53, .55, .74),$$

although it is perhaps more appealing to put these factors into the shape of the  $3 \times 3$  array of pixels as in (and as illustrated in the margin)



$$\begin{bmatrix} r_1 & r_4 & r_7 \\ r_2 & r_5 & r_8 \\ r_3 & r_6 & r_9 \end{bmatrix} = \begin{bmatrix} 0.66 & 0.63 & 0.53 \\ 0.68 & 0.65 & 0.55 \\ 0.91 & 0.87 & 0.74 \end{bmatrix}.$$

Octave's answer predicts that there is less transmitting, more absorbing, denser, material to the top-right; and more transmitting, less absorbing, less dense, material to the bottom-left.

- However, the answer from MATLAB's  $\mathbf{A} \backslash \mathbf{b}$  is (2 d.p.)

$$\mathbf{x} = (-0.91, 0, 0, -0.61, -1.43, 1.01, 0, 0, -1.54),$$

as illustrated below—the leftmost picture—which is quite different!



Furthermore, MATLAB could give other ‘answers’ as illustrated in the other pictures above. Reordering the rows in the matrix  $\mathbf{A}$  and right-hand side  $\mathbf{b}$  does not change the system of equations. But after such reordering the answer from MATLAB's  $\mathbf{x} = \mathbf{A} \backslash \mathbf{b}$  variously predicts each of the above four pictures.

The reason for such multiplicity of mathematically valid answers is that the problem is underdetermined. There are nine unknowns but only six equations, so in linear algebra there are typically an infinity of valid answers (as in [Theorem 2.2.31](#)): just five of

these are illustrated above. *In this application to CT-scans* we add the additional information that we desire the answer that is the ‘greyest’, the most ‘washed out’, the answer with fewest features. Finding the answer  $\mathbf{x}$  that minimises  $|\mathbf{x}|$  is a reasonable way to quantify this desire.

The SVD procedure guarantees that we find such a smallest answer. [Procedure 3.5.4](#) in MATLAB/Octave gives the following process to satisfy the experimental measurements expressed in  $A\mathbf{x} = \mathbf{b}$ .

- (a) First, find an SVD,  $A = USV^T$ , via `[U,S,V]=svd(A)` and get (2 d.p.)

```
U =
-0.41 -0.00  0.82 -0.00  0.00  0.41
-0.41 -0.00 -0.41 -0.57 -0.42  0.41
-0.41 -0.00 -0.41  0.57  0.42  0.41
-0.41  0.81 -0.00  0.07 -0.09 -0.41
-0.41 -0.31 -0.00 -0.45  0.61 -0.41
-0.41 -0.50  0.00  0.38 -0.52 -0.41
```

```
S =
2.45      0      0      0      0      0      0      0      0
      0  1.73      0      0      0      0      0      0      0
```



$$\begin{array}{cccccccc}
 0 & 0 & 1.73 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 1.73 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 1.73 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0.00 & 0 & 0 \\
 \mathbf{V} = & & & & & & & \\
 -0.33 & 0.47 & 0.47 & 0.04 & -0.05 & 0.03 & -0.58 & -0.21 & -0.25 \\
 -0.33 & -0.18 & 0.47 & -0.26 & 0.35 & -0.36 & 0.49 & -0.27 & -0.07 \\
 -0.33 & -0.29 & 0.47 & 0.22 & -0.30 & 0.33 & 0.09 & 0.47 & 0.33 \\
 -0.33 & 0.47 & -0.24 & -0.29 & -0.29 & -0.48 & 0.11 & 0.37 & 0.26 \\
 -0.33 & -0.18 & -0.24 & -0.59 & 0.11 & 0.41 & -0.24 & -0.27 & 0.38 \\
 -0.33 & -0.29 & -0.24 & -0.11 & -0.54 & 0.07 & 0.13 & -0.10 & -0.64 \\
 -0.33 & 0.47 & -0.24 & 0.37 & 0.19 & 0.45 & 0.47 & -0.16 & -0.00 \\
 -0.33 & -0.18 & -0.24 & 0.07 & 0.59 & -0.05 & -0.25 & 0.53 & -0.31 \\
 -0.33 & -0.29 & -0.24 & 0.55 & -0.06 & -0.40 & -0.22 & -0.37 & 0.32
 \end{array}$$

(b) Solve  $Uz = \mathbf{b}$  by  $\mathbf{z} = U' * \mathbf{b}$  to find

$$\mathbf{z} = (2.85, -0.52, 0.31, 0.05, -0.67, -0.00).$$

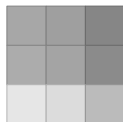
(c) Because the sixth singular value is zero, ignore the sixth equation: because  $z_6 = 0.00$  this is only a small inconsistency error. Now set  $y_i = z_i/\sigma_i$  for  $i = 1, \dots, 5$  and *for the smallest*



magnitude answer set the free variables  $y_6 = y_7 = y_8 = y_9 = 0$  (Theorem 3.5.13). Obtain the non-zero values via `y=z(1:5)./diag(S(1:5,1:5))` to find

$$\mathbf{y} = (1.16, -0.30, 0.18, 0.03, -0.39, 0, 0, 0, 0)$$

- (d) Then solve  $V^T \mathbf{x} = \mathbf{y}$  to determine the smallest solution via `x=V(:,1:5)*y` is  $\mathbf{x} = (-0.42, -0.39, -0.09, -0.47, -0.44, -0.14, -0.63, -0.60, -0.30)$ . This is the same answer as computed by Octave's `A\b` to give the pixel image shown that has minimal artifices.



In practice, *each* slice of a real CT-scan would involve finding the absorption of tens of millions of pixels. That is, a CT-scan needs to best solve many systems of tens of millions of equations in tens of millions of unknowns! ■

### 3.5.3 Orthogonal projection resolves vector components

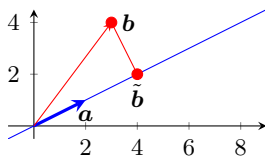
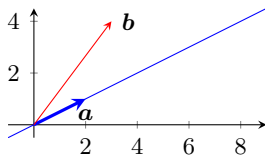
This optional section does usefully support least square approximation, and provides examples of transformations for the next [Section 3.6](#). Such orthogonal projections are extensively used in applications.

Reconsider the task of making a minimal change to the right-hand side of a system of linear equations, and let's connect it to the so-called orthogonal projection. This important connection occurs because of the geometry that the closest point on a line or plane to another given point is the one which forms a right-angle; that is, is forms an orthogonal vector.

## Project onto a direction

**Example 3.5.18.** Consider ‘solving’ the inconsistent system  $\mathbf{a}x = \mathbf{b}$  where  $\mathbf{a} = (2, 1)$  and  $\mathbf{b} = (3, 4)$ ; that is, solve

$$\begin{bmatrix} 2 \\ 1 \end{bmatrix} x = \begin{bmatrix} 3 \\ 4 \end{bmatrix}.$$



As illustrated in the margin, the impossible task is to find some multiple of the vector  $\mathbf{a} = (2, 1)$  (all multiples plotted) that equals  $\mathbf{b} = (3, 4)$ . It cannot be done. Question: how may we change the right-hand side vector  $\mathbf{b}$  so that the task is possible? A partial answer is to replace  $\mathbf{b}$  by some vector  $\tilde{\mathbf{b}}$  which is in the column space of matrix  $A = [\mathbf{a}]$ . But we could choose any  $\tilde{\mathbf{b}}$  in the column space, so any answer is possible! Surely any answer is not acceptable. Instead, the preferred answer is, out of all vectors in the column space of matrix  $A = [\mathbf{a}]$ , find the vector  $\tilde{\mathbf{b}}$  which is closest to  $\mathbf{b}$ , as illustrated in the margin here.

The SVD approach of [Procedure 3.5.4](#) to find  $\tilde{\mathbf{b}}$  and  $x$  is the following.

- (a) Use  $[\mathbf{U}, \mathbf{S}, \mathbf{V}] = \text{svd}([2; 1])$  to find here the SVD factorisation

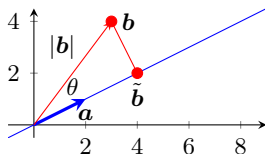
$$A = USV^T = \begin{bmatrix} 0.89 & -0.45 \\ 0.45 & 0.89 \end{bmatrix} \begin{bmatrix} 2.24 \\ 0 \end{bmatrix} [1]^T \text{ (2 d.p.)}.$$

(b) Then  $\mathbf{z} = U^T \mathbf{b} = (4.47, 2.24)$ .

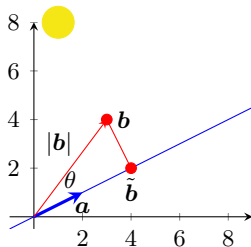
(c) Treat the second component of  $Sy = \mathbf{z}$  as an error—it is the magnitude  $|\mathbf{b} - \tilde{\mathbf{b}}|$ —to deduce  $y = 4.47/2.24 = 2.00$  (2 d.p.) from the first component.

(d) Then  $x = Vy = 1y = 2$  solves the changed problem.

From this solution, the vector  $\tilde{\mathbf{b}} = \mathbf{a}x = (2, 1)2 = (4, 2)$ , as is recognisable in the graphs. ■



Now let's derive the same result but with two differences: firstly, use more elementary arguments, not the SVD; and secondly, derive the result for general vectors  $\mathbf{a}$  and  $\mathbf{b}$  (although continuing to use the same illustration). Start with the crucial observation that the closest point/vector  $\tilde{\mathbf{b}}$  in the column space of  $A = [\mathbf{a}]$  is such that  $\mathbf{b} - \tilde{\mathbf{b}}$  is at right-angles, orthogonal, to  $\mathbf{a}$ . (If  $\mathbf{b} - \tilde{\mathbf{b}}$  were not orthogonal, then we would be able to slide  $\tilde{\mathbf{b}}$  along the line  $\text{span}\{\mathbf{a}\}$  to reduce the length of  $\mathbf{b} - \tilde{\mathbf{b}}$ .) Thus we form a right-angle triangle



with hypotenuse of length  $|b|$  and angle  $\theta$  as shown in the margin. Trigonometry then gives the adjacent length  $|\tilde{b}| = |b| \cos \theta$ . But the angle  $\theta$  is that between the given vectors  $a$  and  $b$ , so the dot product gives the cosine as  $\cos \theta = a \cdot b / (|a||b|)$  (Theorem 1.3.5). Hence the adjacent length  $|\tilde{b}| = |b|a \cdot b / (|a||b|) = a \cdot b / |a|$ . To approximately solve  $ax = b$ , replace the inconsistent  $ax = b$  by the consistent  $ax = \tilde{b}$ . Then as  $x$  is a scalar we solve this consistent equation via the ratio of lengths,  $x = |\tilde{b}|/|a| = a \cdot b / |a|^2$ . For Example 3.5.18, this gives ‘solution’  $x = (2, 1) \cdot (3, 4) / (2^2 + 1^2) = 10/5 = 2$  as before.

A crucial part of such solutions is the general formula for  $\tilde{b} = ax = a(a \cdot b) / |a|^2$ . Geometrically the formula gives the ‘shadow’  $\tilde{b}$  of vector  $b$  when projected by a ‘sun’ high above the line of the vector  $a$ , as illustrated schematically in the margin. As such, the formula is called an orthogonal projection.

**Definition 3.5.19** (orthogonal projection onto 1D). *Let  $u, v \in \mathbb{R}^n$  and vector  $u \neq 0$ , then the **orthogonal projection** of  $v$  onto  $u$  is*

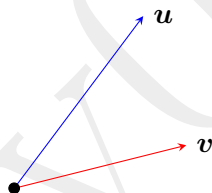
$$\text{proj}_u(v) := u \frac{u \cdot v}{|u|^2}. \quad (3.5a)$$

*In the special but common case when  $\mathbf{u}$  is a unit vector,*

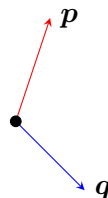
$$\text{proj}_{\mathbf{u}}(\mathbf{v}) := \mathbf{u}(\mathbf{u} \cdot \mathbf{v}). \quad (3.5b)$$

**Example 3.5.20.** For the following pairs of vectors: draw the named orthogonal projection; and for the given inconsistent system, determine whether the ‘best’ approximate solution is in the range  $x < -1$ ,  $-1 < x < 0$ ,  $0 < x < 1$ , or  $1 < x$ .

(a)  $\text{proj}_{\mathbf{u}}(\mathbf{v})$  and  $\mathbf{u}x = \mathbf{v}$



(b)  $\text{proj}_{\mathbf{q}}(\mathbf{p})$  and  $\mathbf{q}x = \mathbf{p}$



**Example 3.5.21.** For the following pairs of vectors: compute the given orthogonal projection; and hence find the ‘best’ approximate solution to the given inconsistent system.

- (a) Find  $\text{proj}_{\mathbf{u}}(\mathbf{v})$  for vectors  $\mathbf{u} = (3, 4)$  and  $\mathbf{v} = (4, 1)$ , and hence best solve  $\mathbf{u}x = \mathbf{v}$ .
- (b) Find  $\text{proj}_{\mathbf{s}}(\mathbf{r})$  for vectors  $\mathbf{r} = (1, 3)$  and  $\mathbf{s} = (2, -2)$ , and hence best solve  $\mathbf{s}x = \mathbf{r}$ .
- (c) Find  $\text{proj}_{\mathbf{p}}(\mathbf{q})$  for vectors  $\mathbf{p} = (\frac{1}{3}, \frac{2}{3}, \frac{2}{3})$  and  $\mathbf{q} = (3, 2, 1)$ , and best solve  $\mathbf{p}x = \mathbf{q}$ .



**Activity 3.5.22.** Use projection to best solve the inconsistent equation  $(1, 4, 8)x = (4, 4, 2)$ . The best answer is which of the following?

- (a)  $x = 4/9$
- (b)  $x = 21/4$
- (c)  $x = 4$
- (d)  $x = 10/13$



VO.410



## Project onto a subspace

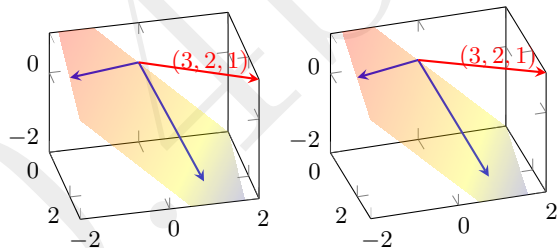
The previous subsection develops a geometric view of the ‘best’ solution to the inconsistent system  $\mathbf{a}x = \mathbf{b}$ . The discussion introduced that the conventional ‘best’ solution—that determined by [Procedure 3.5.4](#)—is to replace  $\mathbf{b}$  by its projection  $\text{proj}_{\mathbf{a}}(\mathbf{b})$ , namely to solve  $\mathbf{a}x = \text{proj}_{\mathbf{a}}(\mathbf{b})$ . The rationale is that this is the *smallest* change to the right-hand side that enables the equation to be solved. This subsection introduces that solving inconsistent equations in more variables involves an analogous projection onto a subspace.

**Definition 3.5.23** (project onto a subspace). *Let  $\mathbb{W}$  be a  $k$ -dimensional subspace of  $\mathbb{R}^n$  with an orthonormal basis  $\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_k\}$ . For every vector  $\mathbf{v} \in \mathbb{R}^n$ , the **orthogonal projection** of vector  $\mathbf{v}$  onto subspace  $\mathbb{W}$  is*

$$\text{proj}_{\mathbb{W}}(\mathbf{v}) = \mathbf{w}_1(\mathbf{w}_1 \cdot \mathbf{v}) + \mathbf{w}_2(\mathbf{w}_2 \cdot \mathbf{v}) + \cdots + \mathbf{w}_k(\mathbf{w}_k \cdot \mathbf{v}).$$

**Example 3.5.24.** (a) Let  $\mathbb{X}$  be the  $xy$ -plane in  $xyz$ -space, find  $\text{proj}_{\mathbb{X}}(3, -4, 2)$ .

(b) For the subspace  $\mathbb{W} = \text{span}\{(2, -2, 1), (2, 1, -2)\}$ , determine  $\text{proj}_{\mathbb{W}}(3, 2, 1)$  (these vectors and subspace are illustrated below).



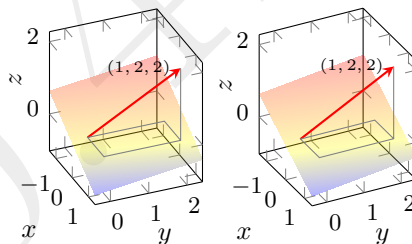
(c) Recall the table tennis ranking Examples 3.3.13 and 3.5.3. To rank the players we seek to solve the matrix-vector system,  $A\mathbf{x} = \mathbf{b}$ ,

$$\begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}.$$

Letting  $\mathbb{A}$  denote the column space of matrix  $A$ , determine

$\text{proj}_{\mathbb{A}}(\mathbf{b})$ .

- (d) Find the projection of the vector  $(1, 2, 2)$  onto the plane  $2x - \frac{1}{2}y + 4z = 6$ .
- (e) Use an SVD to find the projection of the vector  $(1, 2, 2)$  onto the plane  $2x - \frac{1}{2}y + 4z = 0$  (illustrated below).



**Activity 3.5.25.** Determine which of the following is  $\text{proj}_{\mathbb{W}}(1, 1, -2)$  for the subspace  $\mathbb{W} = \text{span}\{(2, 3, 6), (-3, 6, -2)\}$ .

(a)  $(\frac{5}{7}, -\frac{3}{7}, \frac{8}{7})$

(b)  $(-\frac{1}{7}, \frac{9}{7}, \frac{4}{7})$

(c)  $(-\frac{5}{7}, \frac{3}{7}, -\frac{8}{7})$

(d)  $(\frac{1}{7}, -\frac{9}{7}, -\frac{4}{7})$



**Example 3.5.24c** determines the orthogonal projection of the given table tennis results  $\mathbf{b} = (1, 2, 2)$  onto the column space of matrix  $A$  is the vector  $\tilde{\mathbf{b}} = \frac{1}{3}(2, 7, 5)$ . Recall that in **Example 3.5.3**, **Procedure 3.5.4** gives the ‘approximate’ solution of the impossible  $A\mathbf{x} = \mathbf{b}$  to be  $\mathbf{x} = (1, \frac{1}{3}, -\frac{4}{3})$ . Now see that  $A\mathbf{x} = (1 - \frac{1}{3}, 1 - (-\frac{4}{3}), \frac{1}{3} - (-\frac{4}{3})) = (\frac{2}{3}, \frac{7}{3}, \frac{5}{3}) = \tilde{\mathbf{b}}$ . That is, the approximate solution method of **Procedure 3.5.4** solved the problem  $A\mathbf{x} = \text{proj}_{\mathbb{A}}(\mathbf{b})$ . The following theorem confirms this is no accident: orthogonally projecting the right-hand side onto the column space of the matrix in a system of linear equations is equivalent to solving the system with a smallest change to the right-hand side that makes it consistent.

**Theorem 3.5.26.** *The ‘least square’ solution/s of the system  $A\mathbf{x} = \mathbf{b}$  determined by [Procedure 3.5.4](#) is/are the solution/s of  $A\mathbf{x} = \text{proj}_{\mathbb{A}}(\mathbf{b})$  where  $\mathbb{A}$  denotes the column space of  $A$ .*

**Example 3.5.27.** Recall [Example 3.5.1](#) rationalises four apparently contradictory weighings: in kg the weighings are 84.8, 84.1, 84.7 and 84.4. Denoting the ‘uncertain’ weight by  $x$ , we write these weighings as the inconsistent matrix-vector system

$$A\mathbf{x} = \mathbf{b}, \quad \text{namely} \quad \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} x = \begin{bmatrix} 84.8 \\ 84.1 \\ 84.7 \\ 84.4 \end{bmatrix}.$$

Let’s see that the orthogonal projection of the right-hand side onto the column space of  $A$  is the same as the minimal change of [Example 3.5.1](#), which in turn is the well known average.

To find the orthogonal projection, observe matrix  $A$  has one column  $\mathbf{a}_1 = (1, 1, 1, 1)$  so by [Definition 3.5.19](#) the orthogonal projection

$$\text{proj}_{\text{span}\{\mathbf{a}_1\}}(84.8, 84.1, 84.7, 84.4)$$

$$\begin{aligned} &= \mathbf{a}_1 \frac{\mathbf{a}_1 \cdot (84.8, 84.1, 84.7, 84.4)}{|\mathbf{a}_1|^2} \\ &= \mathbf{a}_1 \frac{84.8 + 84.1 + 84.7 + 84.4}{1 + 1 + 1 + 1} \\ &= \mathbf{a}_1 \times 84.5 \\ &= (84.5, 84.5, 84.5, 84.5). \end{aligned}$$

The projected system  $Ax = (84.5, 84.5, 84.5, 84.5)$  is now consistent, with solution  $x = 84.5$  kg. As in [Example 3.5.1](#), this solution is the well-known averaging of the four weights. ■

**Example 3.5.28.** Recall the round robin tournament amongst four players of [Example 3.5.6](#). To estimate the player ratings of the four players from the results of six matches we want to solve the inconsistent system  $A\mathbf{x} = \mathbf{b}$  where

$$A = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 1 \\ 3 \\ 1 \\ -2 \\ 4 \\ -1 \end{bmatrix}.$$

Let's see that the orthogonal projection of  $\mathbf{b}$  onto the column space of  $A$  is the same as the minimal change of [Example 3.5.6](#).

An SVD finds an orthonormal basis for the column space  $\mathbb{A}$  of matrix  $A$ : [Example 3.5.6](#) uses the SVD (2 d.p.)

$$\mathbf{U} = \begin{bmatrix} 0.31 & -0.26 & -0.58 & -0.26 & 0.64 & -0.15 \\ 0.07 & 0.40 & -0.58 & 0.06 & -0.49 & -0.51 \\ -0.24 & 0.67 & 0.00 & -0.64 & 0.19 & 0.24 \end{bmatrix}$$



$$\begin{array}{r}
 \begin{array}{rrrrrr}
 -0.38 & -0.14 & -0.58 & 0.21 & -0.15 & 0.66 \\
 -0.70 & 0.13 & 0.00 & 0.37 & 0.45 & -0.40 \\
 -0.46 & -0.54 & -0.00 & -0.58 & -0.30 & -0.26
 \end{array} \\
 \mathbf{S} = \\
 \begin{array}{rrrr}
 2.00 & 0 & 0 & 0 \\
 0 & 2.00 & 0 & 0 \\
 0 & 0 & 2.00 & 0 \\
 0 & 0 & 0 & 0.00 \\
 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0
 \end{array} \\
 \mathbf{V} = \dots
 \end{array}$$

As there are three non-zero singular values in  $\mathbf{S}$ , the first three columns of  $\mathbf{U}$  are an orthonormal basis for the column space  $\mathbb{A}$ . Letting  $\mathbf{u}_j$  denote the columns of  $\mathbf{U}$ , [Definition 3.5.23](#) gives the orthogonal projection (2 d.p.)

$$\begin{aligned}
 \text{proj}_{\mathbb{A}}(\mathbf{b}) &= \mathbf{u}_1(\mathbf{u}_1 \cdot \mathbf{b}) + \mathbf{u}_2(\mathbf{u}_2 \cdot \mathbf{b}) + \mathbf{u}_3(\mathbf{u}_3 \cdot \mathbf{b}) \\
 &= -1.27 \mathbf{u}_1 + 2.92 \mathbf{u}_2 - 1.15 \mathbf{u}_3 \\
 &= (-0.50, 1.75, 2.25, 0.75, 1.25, -1.00).
 \end{aligned}$$



Compute these three dot products in MATLAB/Octave with  $\mathbf{cs}=\mathbf{U}(:,1:3)$  and then compute the linear combination with  $\mathbf{projb}=\mathbf{U}(:,1:3)*\mathbf{cs}$ . To confirm that [Procedure 3.5.4](#) solves  $A\mathbf{x} = \text{proj}_{\mathbb{A}}(\mathbf{b})$  we check that the ratings found by [Example 3.5.6](#),  $\mathbf{x} = (\frac{1}{2}, 1, -\frac{5}{4}, -\frac{1}{4})$ , satisfy  $A\mathbf{x} = \text{proj}_{\mathbb{A}}(\mathbf{b})$ : in MATLAB/Octave compute  $A*[0.50;1.00;-1.25;-0.25]$  and see the product is  $\text{proj}_{\mathbb{A}}(\mathbf{b})$ . ■

[Section 3.6](#) uses orthogonal projection as an example of a linear transformation. The section shows that a linear transformation always correspond to multiplying by a matrix, which for orthogonal projection is here  $WW^T$ .

There is an useful feature of [Examples 3.5.24e](#) and [3.5.28](#). In both we use MATLAB/Octave to compute the projection in two steps: letting matrix  $W$  denote the matrix of appropriate columns of orthogonal  $U$  (respectively  $W = U(:,2:3)$  and  $W = U(:,1:3)$ ), first the examples compute  $\mathbf{cs}=W'*\mathbf{b}$ , that is, the vector  $\mathbf{c} = W^T\mathbf{b}$ ; and second the examples compute  $\mathbf{proj}=W*\mathbf{cs}$ , that is,  $\text{proj}(\mathbf{b}) = W\mathbf{c}$ . Combining these two steps into one (using associativity) gives

$$\text{proj}_{\mathbb{W}}(\mathbf{b}) = W\mathbf{c} = W(W^T)\mathbf{b} = (WW^T)\mathbf{b}.$$

The interesting feature is that the orthogonal projection formula of [Definition 3.5.23](#) is equivalent to the multiplication by matrix  $(WW^T)$  for an appropriate matrix  $W$ .

**Theorem 3.5.29** (orthogonal projection matrix). *Let  $\mathbb{W}$  be a  $k$ -dimensional subspace of  $\mathbb{R}^n$  with an orthonormal basis  $\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_k\}$ , then for every vector  $\mathbf{v} \in \mathbb{R}^n$ , the orthogonal projection*

$$\text{proj}_{\mathbb{W}}(\mathbf{v}) = (WW^T)\mathbf{v} \quad (3.6)$$

*for the  $n \times k$  matrix  $W = [\mathbf{w}_1 \ \mathbf{w}_2 \ \cdots \ \mathbf{w}_k]$ .*

**Example 3.5.30.** Find the matrices of the following orthogonal projections (from [Example 3.5.21](#)), and use the matrix to find the given projection.

- (a)  $\text{proj}_{\mathbf{u}}(\mathbf{v})$  for vector  $\mathbf{u} = (3, 4)$  and  $\mathbf{v} = (4, 1)$ .
- (b)  $\text{proj}_{\mathbf{s}}(\mathbf{r})$  for vector  $\mathbf{s} = (2, -2)$  and  $\mathbf{r} = (1, 1)$ .
- (c)  $\text{proj}_{\mathbf{p}}(\mathbf{q})$  for vector  $\mathbf{p} = (\frac{1}{3}, \frac{2}{3}, \frac{2}{3})$  and  $\mathbf{q} = (3, 3, 0)$ .



**Activity 3.5.31.** Finding the projection  $\text{proj}_{\mathbf{u}}(\mathbf{v})$  for vectors  $\mathbf{u} = (2, 6, 3)$  and  $\mathbf{v} = (1, 4, 8)$  could be done by premultiplying by which of the following the matrices?

(a) 
$$\begin{bmatrix} \frac{4}{49} & \frac{12}{49} & \frac{6}{49} \\ \frac{12}{49} & \frac{36}{49} & \frac{18}{49} \\ \frac{6}{49} & \frac{18}{49} & \frac{9}{49} \end{bmatrix}$$

(b) 
$$\begin{bmatrix} \frac{1}{81} & \frac{4}{81} & \frac{8}{81} \\ \frac{4}{81} & \frac{16}{81} & \frac{32}{81} \\ \frac{8}{81} & \frac{32}{81} & \frac{64}{81} \end{bmatrix}$$

(c) 
$$\begin{bmatrix} \frac{2}{63} & \frac{8}{63} & \frac{16}{63} \\ \frac{2}{21} & \frac{8}{21} & \frac{16}{21} \\ \frac{1}{21} & \frac{4}{21} & \frac{8}{21} \end{bmatrix}$$

(d) 
$$\begin{bmatrix} \frac{2}{63} & \frac{2}{21} & \frac{1}{21} \\ \frac{8}{63} & \frac{8}{21} & \frac{4}{21} \\ \frac{16}{63} & \frac{16}{21} & \frac{8}{21} \end{bmatrix}$$

**Example 3.5.32.** Find the matrices of the following orthogonal projections (from [Example 3.5.21](#)).

(a)  $\text{proj}_{\mathbb{X}}(\mathbf{v})$  where  $\mathbb{X}$  is the  $xy$ -plane in  $xyz$ -space.

(b)  $\text{proj}_{\mathbb{W}}(\mathbf{v})$  for the subspace  $\mathbb{W} = \text{span}\{(2, -2, 1), (2, 1, -2)\}$ .

(c) The orthogonal projection onto the column space of matrix

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

(d) The orthogonal projection onto the plane  $2x - \frac{1}{2}y + 4z = 0$ .

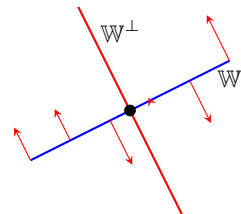


## Orthogonal decomposition separates

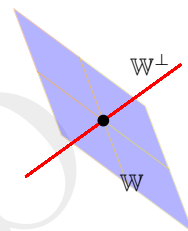
Because orthogonal projection has such a close connection to the geometry underlying important tasks such as ‘least square’ approximation ([Theorem 3.5.26](#)), this section develops further some orthogonal properties.

For any subspace  $\mathbb{W}$  of interest, it is often useful to be able to discuss the set of vectors orthogonal to all those in  $\mathbb{W}$ , called the orthogonal complement. Such a set forms a subspace, called  $\mathbb{W}^\perp$  (read as “ $\mathbb{W}$  perp”), as illustrated below and defined by [Definition 3.5.34](#).

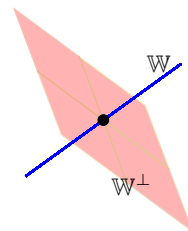
1. Given the blue subspace  $\mathbb{W}$  in  $\mathbb{R}^2$  (the origin is a black dot), consider the set of all vectors at right-angles to  $\mathbb{W}$  (drawn arrows). Move the base of these vectors to the origin, and then they all lie in the red subspace  $\mathbb{W}^\perp$ .



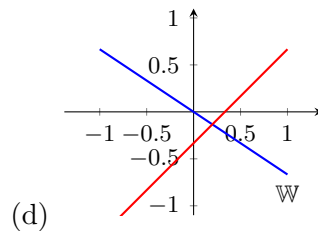
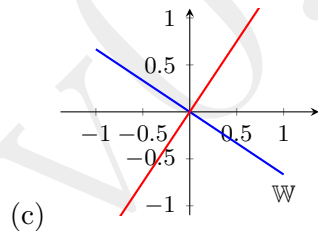
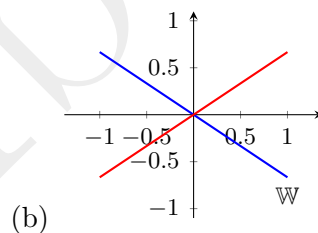
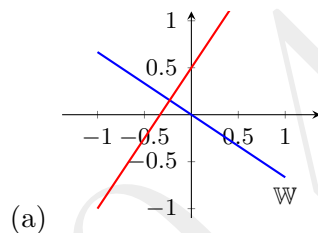
2. Given the blue plane subspace  $\mathbb{W}$  in  $\mathbb{R}^3$  (the origin is a black dot), the red line subspace  $\mathbb{W}^\perp$  contains all vectors orthogonal to  $\mathbb{W}$  (when drawn with their base at the origin).



3. Conversely, given the blue line subspace  $\mathbb{W}$  in  $\mathbb{R}^3$  (the origin is a black dot), the red plane subspace  $\mathbb{W}^\perp$  contains all vectors orthogonal to  $\mathbb{W}$  (when drawn with their base at the origin).



**Activity 3.5.33.** Given the above qualitative description of an orthogonal complement, which of the following red lines is the orthogonal complement to the shown (blue) subspace  $\mathbb{W}$ ?



**Definition 3.5.34** (orthogonal complement). *Let  $\mathbb{W}$  be a  $k$ -dimensional subspace of  $\mathbb{R}^n$ . The set of all vectors  $\mathbf{u} \in \mathbb{R}^n$  (together with  $\mathbf{0}$ ) that are each orthogonal to all vectors in  $\mathbb{W}$  is called the **orthogonal complement**  $\mathbb{W}^\perp$  (“ $W$ -perp”); that is,*

$$\mathbb{W}^\perp = \{\mathbf{u} \in \mathbb{R}^n : \mathbf{u} \cdot \mathbf{w} = 0 \text{ for all } \mathbf{w} \in \mathbb{W}\}.$$

**Example 3.5.35** (orthogonal complement).

- (a) Given the subspace  $\mathbb{W} = \text{span}\{(3, 4)\}$ , find its orthogonal complement  $\mathbb{W}^\perp$ .
- (b) Describe the orthogonal complement  $\mathbb{X}^\perp$  to the subspace  $\mathbb{X} = \text{span}\{(4, -4, 7)\}$ .
- (c) Describe the orthogonal complement of the set  $\mathbb{W} = \{(t, t^2) : t \in \mathbb{R}\}$ .
- (d) Determine the orthogonal complement of the subspace  $\mathbb{W} = \text{span}\{(2, -2, 1), (2, 1, -2)\}$ .





**Activity 3.5.36.** Which of the following vectors are in the orthogonal complement of the vector space spanned by  $(3, -1, 1)$ ?

(a)  $(1, 3, -1)$

(b)  $(6, -2, 2)$

(c)  $(3, 5, -4)$

(d)  $(-1, -1, 1)$



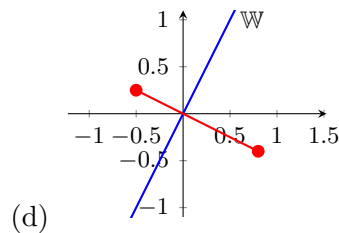
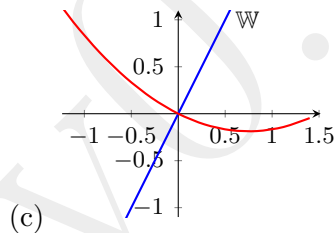
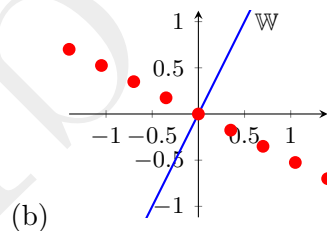
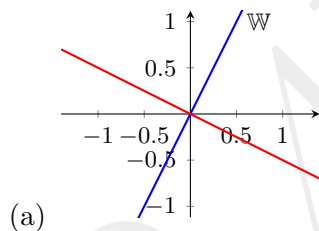
**Example 3.5.37.** Prove  $\{\mathbf{0}\}^\perp = \mathbb{R}^n$  and  $(\mathbb{R}^n)^\perp = \{\mathbf{0}\}$ .



These examples find that orthogonal complements are lines, planes, or the entire space. These indicate that an orthogonal complement is generally a subspace as proved next.

**Theorem 3.5.38** (orthogonal complement is subspace). *For every subspace  $\mathbb{W}$  of  $\mathbb{R}^n$ , the orthogonal complement  $\mathbb{W}^\perp$  is a subspace of  $\mathbb{R}^n$ . Further, the intersection  $\mathbb{W} \cap \mathbb{W}^\perp = \{\mathbf{0}\}$ ; that is, the zero vector is the only vector in both  $\mathbb{W}$  and  $\mathbb{W}^\perp$ .*

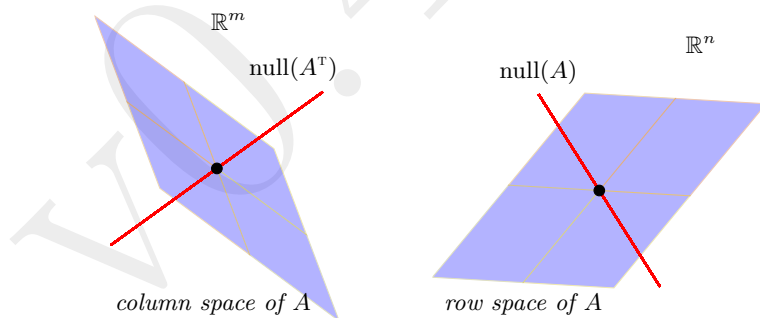
**Activity 3.5.39.** Vectors in which of the following (red) sets form the orthogonal complement to the shown (blue) subspace  $W$ ?



When orthogonal complements arise, they are often usefully written

as the nullspace of a matrix.

**Theorem 3.5.40** (nullspace complementarity). *For every  $m \times n$  matrix  $A$ , the column space of  $A$  has  $\text{null}(A^T)$  as its orthogonal complement in  $\mathbb{R}^m$ . That is, identifying the columns of matrix  $A = [\mathbf{a}_1 \ \mathbf{a}_2 \ \cdots \ \mathbf{a}_n]$ , and denoting the column space by  $\mathbb{A} = \text{span}\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}$ , then the orthogonal complement  $\mathbb{A}^\perp = \text{null}(A^T)$ . Further,  $\text{null}(A)$  in  $\mathbb{R}^n$  is the orthogonal complement of the row space of  $A$ .*



**Example 3.5.41.**

- (a) Let the subspace  $\mathbb{W} = \text{span}\{(2, -1)\}$ . Find the orthogonal complement  $\mathbb{W}^\perp$ .
- (b) Describe the subspace of  $\mathbb{R}^3$  whose orthogonal complement is the plane  $-\frac{1}{2}x - y + 2z = 0$ .
- (c) Find the orthogonal complement to the column space of matrix

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

- (d) Describe the orthogonal complement of the subspace spanned by the four vectors  $(1, 1, 0, 1, 0, 0)$ ,  $(-1, 0, 1, 0, 1, 0)$ ,  $(0, -1, -1, 0, 0, 0)$ , and  $(0, 0, 0, -1, -1, -1)$ .

■

In the previous [Example 3.5.41d](#) there are three non-zero singular

values in the first three rows of  $S$ . These three nonzero singular values determine that the first three columns of  $U$  form a basis for the column space of  $A$ . The example argues that the remaining three columns of  $U$  form a basis for the orthogonal complement of the column space. That is, all six of the columns of the orthogonal  $U$  are used in either the column space or its complement. This is generally true.

**Activity 3.5.42.** A given matrix  $A$  has column space  $\mathbb{W}$  such that  $\dim \mathbb{W} = 4$  and  $\dim \mathbb{W}^\perp = 3$ . What size could the matrix be?

- (a)  $4 \times 3$       (b)  $3 \times 4$       (c)  $7 \times 5$       (d)  $7 \times 3$



**Example 3.5.43.** Recall the cases of [Example 3.5.41](#).

**3.5.41a :**  $\dim \mathbb{W} + \dim \mathbb{W}^\perp = 1 + 1 = 2 = \dim \mathbb{R}^2$ .

**3.5.41b :**  $\dim \mathbb{W} + \dim \mathbb{W}^\perp = 1 + 2 = 3 = \dim \mathbb{R}^3$ .

**3.5.41c :**  $\dim \mathbb{W} + \dim \mathbb{W}^\perp = 2 + 1 = 3 = \dim \mathbb{R}^3$ .

$$3.5.41d : \dim \mathbb{W} + \dim \mathbb{W}^\perp = 3 + 3 = 6 = \dim \mathbb{R}^6.$$



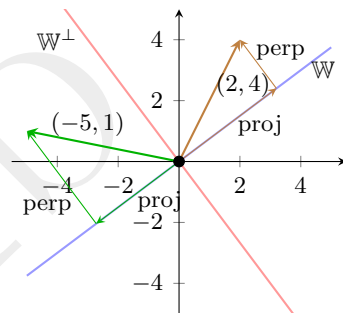
Recall the Rank [Theorem 3.4.39](#) connects the dimension of a space with the dimensions of a nullspace and column space of a matrix. Since a subspace is closely connected to matrices, and its orthogonal complement is connected to nullspaces, then the Rank Theorem should say something general here.

**Theorem 3.5.44.** *Let  $\mathbb{W}$  be a subspace of  $\mathbb{R}^n$ , then  $\dim \mathbb{W} + \dim \mathbb{W}^\perp = n$ ; equivalently,  $\dim \mathbb{W}^\perp = n - \dim \mathbb{W}$ .*

Since the dimension of the whole space is the sum of the dimension of a subspace plus the dimension of its orthogonal complement, surely we must be able to separate vectors into two corresponding components.

**Example 3.5.45.** Recall from [Example 3.5.35a](#) that subspace  $\mathbb{W} = \text{span}\{(3, 4)\}$  has orthogonal complement  $\mathbb{W}^\perp = \text{span}\{(-4, 3)\}$ , as illustrated below.

As shown, for example, write the brown vector  $(2, 4) = (3.2, 2.4) + (-1.2, 1.6) = \text{proj}_{\mathbb{W}}(2, 4) + \text{perp}$ , where here the vector  $\text{perp} = (-1.2, 1.6) \in \mathbb{W}^\perp$ . Indeed, any vector can be written as a component in subspace  $\mathbb{W}$  and a component in the orthogonal complement  $\mathbb{W}^\perp$  (Theorem 3.5.51).



For example, write the green vector  $(-5, 1) = (-2.72, -2.04) + (-2.28, 3.04) = \text{proj}_{\mathbb{W}}(-5, 1) + \text{perp}$ , where in this case the vector  $\text{perp} = (-2.28, 3.04) \in \mathbb{W}^\perp$ . ■

**Activity 3.5.46.** Let subspace  $\mathbb{W} = \text{span}\{(1, 1)\}$  and its orthogonal complement  $\mathbb{W}^\perp = \text{span}\{(1, -1)\}$ . Which of the following writes vector  $(5, -9)$  as a sum of two vectors, one from each of  $\mathbb{W}$  and  $\mathbb{W}^\perp$ ?

(a)  $(7, 7) + (-2, 2)$

(b)  $(-2, -2) + (7, -7)$

(c)  $(9, -9) + (-4, 0)$

(d)  $(5, 5) + (0, -14)$



Further, such a separation can be done for any pair of complementary subspaces  $\mathbb{W}$  and  $\mathbb{W}^\perp$  within any space  $\mathbb{R}^n$ . To proceed, let's define what is meant by “perp” in such a context.

**Definition 3.5.47** (perpendicular component). *Let  $\mathbb{W}$  be a subspace of  $\mathbb{R}^n$ . For every vector  $\mathbf{v} \in \mathbb{R}^n$ , the **perpendicular component** of  $\mathbf{v}$  to  $\mathbb{W}$  is the vector  $\text{perp}_{\mathbb{W}}(\mathbf{v}) := \mathbf{v} - \text{proj}_{\mathbb{W}}(\mathbf{v})$ .*



- Example 3.5.48.** (a) Let the subspace  $\mathbb{W}$  be the span of  $(-2, -3, 6)$ . Find the perpendicular component to  $\mathbb{W}$  of the vector  $(4, 1, 3)$ . Verify the perpendicular component lies in the plane  $-2x - 3y + 6z = 0$ .
- (b) For the vector  $(-5, -1, 6)$  find its perpendicular component to the subspace  $\mathbb{W}$  spanned by  $(-2, -3, 6)$ . Verify the perpendicular component lies in the plane  $-2x - 3y + 6z = 0$ .
- (c) Let the subspace  $\mathbb{X} = \text{span}\{(2, -2, 1), (2, 1, -2)\}$ . Determine the perpendicular component of each of the two vectors  $\mathbf{y} = (3, 2, 1)$  and  $\mathbf{z} = (3, -3, -3)$ .



As seen in all these examples, the perpendicular component of a vector always lies in the orthogonal complement to the subspace (as suggested by the naming).

**Theorem 3.5.49** (perpendicular component is orthogonal). *Let  $\mathbb{W}$  be a subspace of  $\mathbb{R}^n$  and let  $\mathbf{v}$  be any vector in  $\mathbb{R}^n$ , then the perpendicular component  $\text{perp}_{\mathbb{W}}(\mathbf{v}) \in \mathbb{W}^\perp$ .*

**Example 3.5.50.** The previous examples' calculation of the perpendicular component confirm that  $\mathbf{v} = \text{proj}_{\mathbb{W}}(\mathbf{v}) + \text{perp}_{\mathbb{W}}(\mathbf{v})$ , where we now know that  $\text{perp}_{\mathbb{W}}$  is orthogonal to  $\mathbb{W}$ :

$$\begin{aligned} 3.5.45 : \quad (2, 4) &= (3.2, 2.4) + (-1.2, 1.6) \text{ and} \\ (-5, 1) &= (-2.72, -2.04) + (-2.28, 3.04); \end{aligned}$$

$$3.5.48b : \quad (-5, -1, 6) = (-2, -3, 6) + (-3, 2, 0);$$

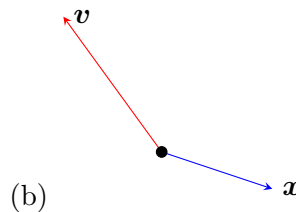
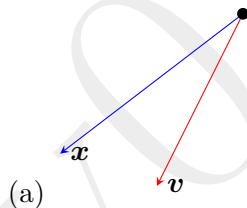
$$\begin{aligned} 3.5.48c : \quad (3, 2, 1) &= (2, 0, -1) + (1, 2, 2) \text{ and} \\ (3, -3, -3) &= (4, -1, -1) + (-1, -2, -2). \end{aligned}$$



Given any subspace  $\mathbb{W}$ , this theorem indicates that every vector can be written as a sum of two vectors: one in the subspace  $\mathbb{W}$ ; and one in its orthogonal complement  $\mathbb{W}^\perp$ .

**Theorem 3.5.51** (orthogonal decomposition). *Let  $\mathbb{W}$  be a subspace of  $\mathbb{R}^n$  and vector  $\mathbf{v} \in \mathbb{R}^n$ , then there exist unique vectors  $\mathbf{w} \in \mathbb{W}$  and  $\mathbf{n} \in \mathbb{W}^\perp$  such that vector  $\mathbf{v} = \mathbf{w} + \mathbf{n}$ ; this particular sum is called an **orthogonal decomposition** of  $\mathbf{v}$ .*

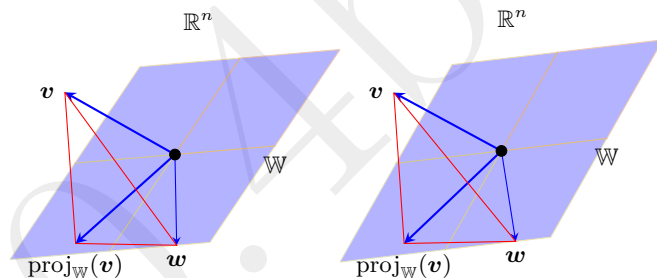
**Example 3.5.52.** For each pair of the shown subspaces  $\mathbb{X} = \text{span}\{\mathbf{x}\}$  and vectors  $\mathbf{v}$ , draw the decomposition of vector  $\mathbf{v}$  into the sum of vectors in  $\mathbb{X}$  and  $\mathbb{X}^\perp$ .



In two or even three dimensions, that a decomposition has such a nice physical picture is appealing. What is powerful is that the

same decomposition works in any number of dimensions: it works no matter how complicated the scenario, no matter how much data. In particular, the next theorem gives a geometric view of the ‘least square’ solution of [Procedure 3.5.4](#): in that procedure the minimal change of the right-hand side  $\mathbf{b}$  to make the linear equation  $A\mathbf{x} = \mathbf{b}$  consistent ([Theorem 3.5.8](#)) is also to be viewed as the projection of the right-hand side  $\mathbf{b}$  to the *closest* point in the columns space of the matrix. That is, the ‘least square’ procedure solves  $A\mathbf{x} = \text{proj}_{\mathbb{A}}(\mathbf{b})$ .

**Theorem 3.5.53** (best approximation). *For every vector  $\mathbf{v}$  in  $\mathbb{R}^n$ , and every subspace  $\mathbb{W}$  in  $\mathbb{R}^n$ ,  $\text{proj}_{\mathbb{W}}(\mathbf{v})$  is the closest vector in  $\mathbb{W}$  to  $\mathbf{v}$ ; that is,  $|\mathbf{v} - \text{proj}_{\mathbb{W}}(\mathbf{v})| \leq |\mathbf{v} - \mathbf{w}|$  for all  $\mathbf{w} \in \mathbb{W}$ .*



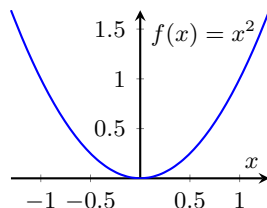
**Discover power laws** Exercises ??–?? use log-log plots as examples of the scientific inference of some surprising patterns in nature. These are simple examples of what, in modern parlance, might be termed ‘data mining’, ‘knowledge discovery’ or ‘artificial intelligence’.

## 3.6 Introducing linear transformations

### Section Contents

3.6.1	Matrices correspond to linear transformations	456
3.6.2	The pseudo-inverse of a matrix . . . . .	463
3.6.3	Function composition connects to matrix inverse . . . . .	472

This optional section unifies the transformation examples seen so far, and forms a foundation for more advanced algebra.



Recall the function notation such as  $f(x) = x^2$  means that for each  $x \in \mathbb{R}$ , the function  $f(x)$  gives a result in  $\mathbb{R}$ , namely the value  $x^2$ , as plotted in the margin. We often write  $f : \mathbb{R} \rightarrow \mathbb{R}$  to denote this functionality: that is,  $f : \mathbb{R} \rightarrow \mathbb{R}$  means function  $f$  transforms any given real number into another real number by some rule.

There is analogous functionality in multiple dimensions with vectors: given any vector

- multiplication by a diagonal matrix stretches and shrinks the vector (Subsection 3.2.2);

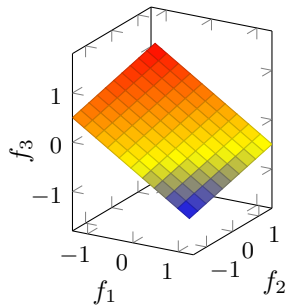
- multiplication by an orthogonal matrix rotates the vector (Subsection 3.2.3); and
- projection finds a vector's components in a subspace (Subsection 3.5.3).

Correspondingly, we use the notation  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  to mean that the function  $f$  transforms a given vector with  $n$  components (in  $\mathbb{R}^n$ ) into another vector with  $m$  components (in  $\mathbb{R}^m$ ) according to some rule. For example, suppose the function  $f(\mathbf{x})$  is to denote multiplication by the matrix

$$A = \begin{bmatrix} 1 & -\frac{1}{3} \\ \frac{1}{2} & -1 \\ -1 & -\frac{1}{2} \end{bmatrix}.$$

Then the function

$$f(\mathbf{x}) = \begin{bmatrix} 1 & -\frac{1}{3} \\ \frac{1}{2} & -1 \\ -1 & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 - x_2/3 \\ x_1/2 - x_2 \\ -x_1 - x_2/2 \end{bmatrix}.$$



That is, here  $f : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ . Given any vector in the 2D-plane, the function  $f$ , also called a transformation, returns a vector in 3D-space. Such a function can be evaluated for every vector  $\mathbf{x} \in \mathbb{R}^2$ , so we ask what is the shape, the structure, of all the possible results of the function. The marginal plot illustrates the subspace formed by this  $f(\mathbf{x})$  for all 2D vectors  $\mathbf{x}$ .

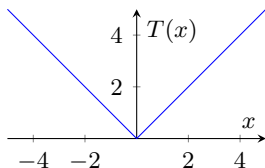
There is a major difference between ‘curvaceous’ functions like the parabola above, and matrix multiplication functions such as rotation and projection. The difference is that linear algebra empowers many practical results in the latter case.

**Definition 3.6.1.** *A transformation/function  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is called a linear transformation if*

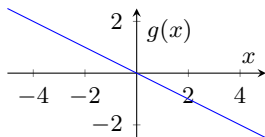
- (a)  $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$  for all  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ , and
- (b)  $T(c\mathbf{v}) = cT(\mathbf{v})$  for all  $\mathbf{v} \in \mathbb{R}^n$  and all scalars  $c$ .



**Example 3.6.2** (1D cases). (a) Show that the parabolic function  $f : \mathbb{R} \rightarrow \mathbb{R}$  where  $f(x) = x^2$  is not a linear transformation.



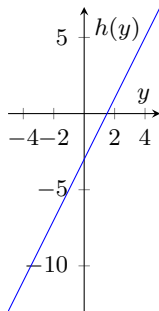
(b) Is the function  $T(x) = |x|$ ,  $T : \mathbb{R} \rightarrow \mathbb{R}$ , a linear transformation?

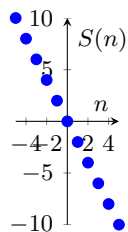


(c) Is the function  $g : \mathbb{R} \rightarrow \mathbb{R}$  such that  $g(x) = -x/2$  a linear transformation?

(d) Show that the function  $h(y) = 2y - 3$ ,  $h : \mathbb{R} \rightarrow \mathbb{R}$ , is not a linear transformation.

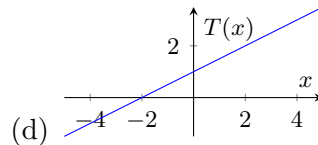
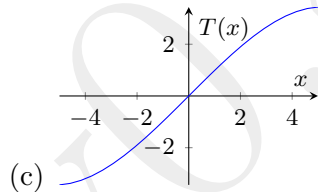
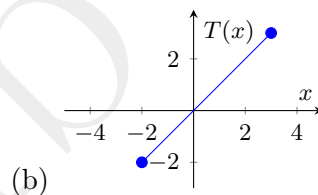
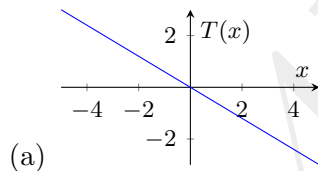
(e) Is the function  $S : \mathbb{Z} \rightarrow \mathbb{Z}$  given by  $S(n) = -2n$  a linear transformation? Here  $\mathbb{Z}$  denotes the set of integers  $\dots, -2, -1, 0, 1, 2, \dots$





**Activity 3.6.3.**  
tion?

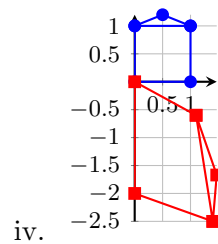
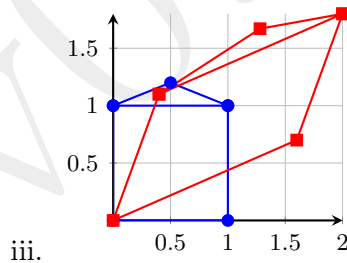
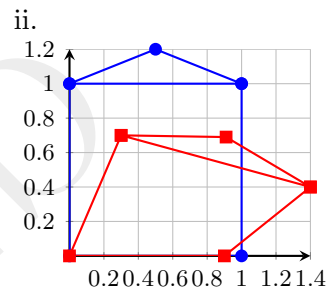
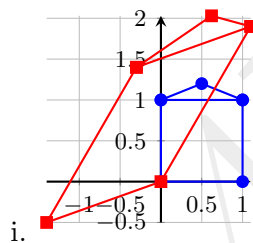
Which of the following is the graph of a linear transformation?

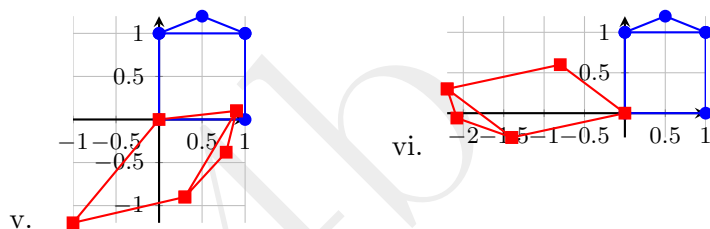


**Example 3.6.4** (higher-D cases). (a) Let function  $T : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be  $T(x, y, z) = (y, z, x)$ . Is  $T$  a linear transformation?

(b) Consider the function  $f(x, y, z) = x + y + 1$ ,  $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ : is  $f$  a linear transformation?

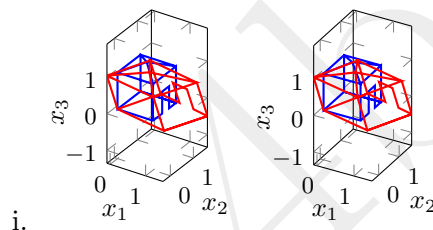
(c) Which of the following illustrated transformations of the plane *cannot* be that of a linear transformation? In each illustration of a transformation  $T$ , the four corners of the blue unit square  $((0, 0), (1, 0), (1, 1)$  and  $(0, 1))$ , are transformed to the four corners of the red figure  $(T(0, 0), T(1, 0), T(1, 1)$  and  $T(0, 1))$ —the ‘roof’ of the unit square clarifies which side goes where).



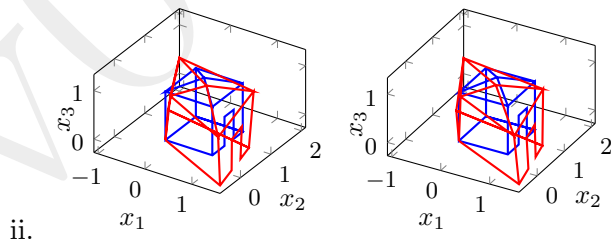


- (d) The previous [Example 3.6.4c](#) illustrated that a linear transformation of the square seems to transform the unit square to a parallelogram: if a function transforms the unit square to something that is not a parallelogram, then the function cannot be a linear transformation. Analogously in higher dimensions: for example, if a function transforms the unit cube to something which is not a parallelepiped, then the function is not a linear transformation. Using this information, which of the following illustrated functions,  $f : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ , *cannot* be a linear transformation? Each of these stereo illustrations plot the unit cube in blue (with a ‘roof’ and ‘door’ to help orientate), and the transform of the unit cube in red (with

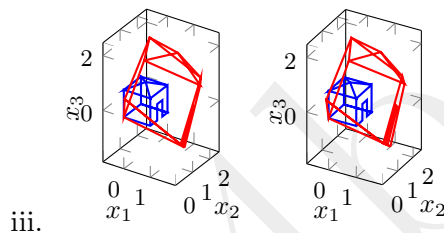
its transformed ‘roof’ and ‘door’).



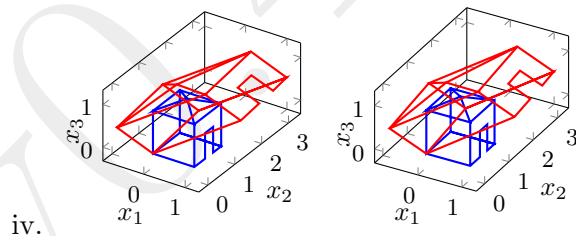
This *may* be a linear transformation as the transform of the unit cube looks like a parallelepiped.



This *cannot* be a linear transformation as the unit cube transforms to something not a parallelepiped.



This *cannot* be a linear transformation as the unit cube transforms to something not a parallelepiped.



This *may* be a linear transformation as the transform of the unit cube looks like a parallelepiped.



**Activity 3.6.5.** Which of the following functions  $f : \mathbb{R}^3 \rightarrow \mathbb{R}^2$  is *not* a linear transformation?

- (a)  $f(x, y, z) = (0, 13x + \pi y)$
- (b)  $f(x, y, z) = (y, x + z)$
- (c)  $f(x, y, z) = (2.7x + 3y, 1 - 2z)$
- (d)  $f(x, y, z) = (0, 0)$

■

**Example 3.6.6.** For any given nonzero vector  $\mathbf{w} \in \mathbb{R}^n$ , prove that the projection  $P : \mathbb{R}^n \rightarrow \mathbb{R}^n$  by  $P(\mathbf{u}) = \text{proj}_{\mathbf{w}}(\mathbf{u})$  is a linear transformation (as a function of  $\mathbf{u}$ ). But, for any given nonzero vector  $\mathbf{u} \in \mathbb{R}^n$ , prove that the projection  $Q : \mathbb{R}^n \rightarrow \mathbb{R}^n$  by  $Q(\mathbf{w}) = \text{proj}_{\mathbf{w}}(\mathbf{u})$  is not a linear transformation (as a function of  $\mathbf{w}$ ). ■



### 3.6.1 Matrices correspond to linear transformations

One important class of linear transformations are the transformations that can be written as matrix multiplications. The reason for the importance is that [Theorem 3.6.10](#) establishes all linear transformations may be written as matrix multiplications! This in turn justifies why we define matrix multiplication to be as it is ([Subsection 3.1.2](#)): *matrix multiplication is defined just so that all linear transformations are encompassed.*

**Example 3.6.7.** But first, the following [Theorem 3.6.8](#) proves, among many other possibilities, that the following transformations we have already met are linear transformations:

- stretching/shrinking along coordinate axes as these are multiplication by a diagonal matrix ([Subsection 3.2.2](#));
- rotations and/or reflections as they arise as multiplications by an orthogonal matrix ([Subsection 3.2.3](#));
- orthogonal projection onto a subspace as all such projections may be expressed as multiplication by a matrix (the

matrix  $WW^T$  in [Theorem 3.5.29](#)).

■

**Theorem 3.6.8.** *Let  $A$  be any given  $m \times n$  matrix and define the transformation  $T_A : \mathbb{R}^n \rightarrow \mathbb{R}^m$  by the matrix multiplication  $T_A(\mathbf{x}) := A\mathbf{x}$  for all  $\mathbf{x} \in \mathbb{R}^n$ . Then  $T_A$  is a linear transformation.*

**Example 3.6.9.** Prove that a matrix multiplication with a nonzero shift  $\mathbf{b}$ ,  $S : \mathbb{R}^n \rightarrow \mathbb{R}^m$  where  $S(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$  for vector  $\mathbf{b} \neq \mathbf{0}$ , is not a linear transformation.

■

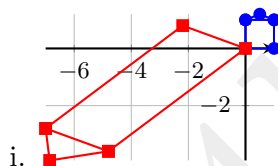
Now let's establish the important converse to [Theorem 3.6.8](#): that every linear transformation can be written as a matrix multiplication.

**Theorem 3.6.10.** *Let  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  be a linear transformation. Then  $T$  is the transformation corresponding to the  $m \times n$  matrix*

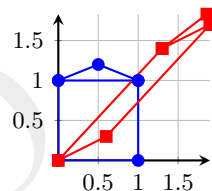
$$A = [T(\mathbf{e}_1) \ T(\mathbf{e}_2) \ \cdots \ T(\mathbf{e}_n)]$$

*where  $\mathbf{e}_j$  are the standard unit vectors in  $\mathbb{R}^n$ . This matrix  $A$ , often denoted  $[T]$ , is called the **standard matrix** of the linear transformation  $T$ .*

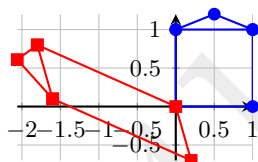
- Example 3.6.11.**
- (a) Find the standard matrix of the linear transformation  $T : \mathbb{R}^3 \rightarrow \mathbb{R}^4$  where  $T(x, y, z) = (y, z, x, 3x - 2y + z)$ .
  - (b) Find the standard matrix of the rotation of the plane by  $60^\circ$  about the origin.
  - (c) Find the standard matrix of the rotation about the point  $(1, 0)$  of the plane by  $45^\circ$ .
  - (d) Estimate the standard matrix for each of the illustrated transformations given they transform the unit square as shown.



*Solution:* Here  
 $T(1, 0) \approx (-2.2, 0.8)$  and  
 $T(0, 1) \approx (-4.8, -3.6)$  so  
 the approximate standard  
 matrix is  $\begin{bmatrix} -2.2 & -4.8 \\ 0.8 & -3.6 \end{bmatrix}$ .



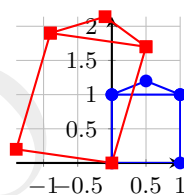
*Solution:* Here  
 $T(1, 0) \approx (0.6, 0.3)$  and  
 $T(0, 1) \approx (1.3, 1.4)$  so the  
 approximate standard  
 matrix is  $\begin{bmatrix} 0.6 & 1.3 \\ 0.3 & 1.4 \end{bmatrix}$ .



iii.

*Solution:* Here $T(1, 0) \approx (0.2, -0.7)$  and $T(0, 1) \approx (-1.8, 0.8)$  so the

approximate standard

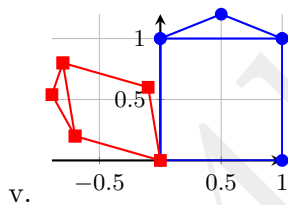
matrix is  $\begin{bmatrix} 0.2 & -1.8 \\ -0.7 & 0.8 \end{bmatrix}$ .

iv.

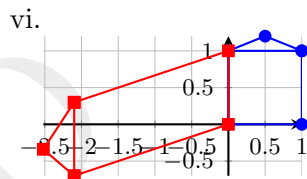
*Solution:* Here $T(1, 0) \approx (-1.4, 0.2)$  and $T(0, 1) \approx (0.5, 1.7)$  so the

approximate standard

matrix is  $\begin{bmatrix} -1.4 & 0.5 \\ 0.2 & 1.7 \end{bmatrix}$ .



*Solution:* Here  
 $T(1, 0) \approx (-0.1, 0.6)$  and  
 $T(0, 1) \approx (-0.7, 0.2)$  so the  
 approximate standard  
 matrix is  $\begin{bmatrix} -0.1 & -0.7 \\ 0.6 & 0.2 \end{bmatrix}$ .



*Solution:* Here  
 $T(1, 0) \approx (0, 1.0)$  and  
 $T(0, 1) \approx (-2.1, -0.7)$  so  
 the approximate standard  
 matrix is  $\begin{bmatrix} 0 & -2.1 \\ 1.0 & -0.7 \end{bmatrix}$ .

**Activity 3.6.12.** Which of the following is the standard matrix for the transformation  $T(x, y, z) = (4.5y - 1.6z, 1.9x - 2z)$ ?

(a)  $\begin{bmatrix} 0 & 4.5 & -1.6 \\ 1.9 & 0 & -2 \end{bmatrix}$

(b)  $\begin{bmatrix} 4.5 & -1.6 \\ 1.9 & -2 \end{bmatrix}$

$$(c) \begin{bmatrix} 4.5 & 1.9 \\ -1.6 & -2 \end{bmatrix}$$

$$(d) \begin{bmatrix} 0 & 1.9 \\ 4.5 & 0 \\ -1.6 & -2 \end{bmatrix}$$

■

**Example 3.6.13.** For a fixed scalar  $a$ , let the function  $H : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be  $H(\mathbf{u}) = a\mathbf{u}$ . Show that  $H$  is a linear transformation, and then find its standard matrix. ■

Consider this last [Example 3.6.13](#) in the case  $a = 1$ : then  $H(\mathbf{u}) = \mathbf{u}$  is the identity and so the example shows that the standard matrix of the identity transformation is  $I_n$ .

### 3.6.2 The pseudo-inverse of a matrix

This subsection is an optional extension.

In solving inconsistent linear equations,  $A\mathbf{x} = \mathbf{b}$  for some given  $A$ , [Procedure 3.5.4](#) finds a solution  $\mathbf{x}$  that depends upon the right-hand side  $\mathbf{b}$ . That is, any given  $\mathbf{b}$  is transformed by the procedure to some result  $\mathbf{x}$ : the result is a function of the given  $\mathbf{b}$ . This section establishes that the resulting solution given by the procedure is a linear transformation of  $\mathbf{b}$ , and hence there must be a matrix, say  $A^+$ , corresponding to the procedure. This matrix gives the resulting solution  $\mathbf{x} = A^+\mathbf{b}$ . We call the matrix  $A^+$  the pseudo-inverse of  $A$ .

**Example 3.6.14.** Find the pseudo-inverse of the matrix  $A = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$ . ■

**Activity 3.6.15.** By finding the smallest magnitude, least-square, solution to  $D\mathbf{x} = \mathbf{b}$  for matrix  $D = \begin{bmatrix} 5 & 0 \\ 0 & 0 \end{bmatrix}$  and arbitrary  $\mathbf{b}$ , determine that the pseudo-inverse of the diagonal matrix  $D$  is which of the following?



$$(a) \begin{bmatrix} 0 & 0 \\ 0 & 0.2 \end{bmatrix}$$

$$(b) \begin{bmatrix} 0 & 0.2 \\ 0 & 0 \end{bmatrix}$$

$$(c) \begin{bmatrix} 0 & 0 \\ 0.2 & 0 \end{bmatrix}$$

$$(d) \begin{bmatrix} 0.2 & 0 \\ 0 & 0 \end{bmatrix}$$



A pseudo-inverse  $A^+$  of a non-invertible matrix  $A$  is only an ‘inverse’ because the pseudo-inverse builds in extra information that you may *sometimes choose* to be desirable. This extra information rationalises all the contradictions encountered in trying to construct an inverse of a non-invertible matrix. Namely, for some applications we *choose* to desire that the pseudo-inverse solves the *nearest* consistent system to the one specified, and we *choose* the smallest of all possibilities then allowed. However, although there are many situations where these choices are useful, beware that there are also many situations where such choices are not appropriate. That is, although sometimes the pseudo-inverse is useful, beware that often the pseudo-inverse is not appropriate.

**Theorem 3.6.16** (pseudo-inverse). Recall that in the context of a system of linear equations  $A\mathbf{x} = \mathbf{b}$  with  $m \times n$  matrix  $A$ , for every  $\mathbf{b} \in \mathbb{R}^m$  [Procedure 3.5.4](#) finds the smallest solution  $\mathbf{x} \in \mathbb{R}^n$  ([Theorem 3.5.13](#)) to the closest consistent system  $A\mathbf{x} = \tilde{\mathbf{b}}$  ([Theorem 3.5.8](#)). [Procedure 3.5.4](#) forms a linear transformation  $T : \mathbb{R}^m \rightarrow \mathbb{R}^n$ ,  $\mathbf{x} = T(\mathbf{b})$ . This linear transformation has an  $n \times m$  standard matrix  $A^+$  called the **pseudo-inverse**, or **Moore–Penrose inverse**, of matrix  $A$ .

**Example 3.6.17.** Find the pseudo-inverse of the matrix  $A = \begin{bmatrix} 5 & 12 \end{bmatrix}$ . ■

**Activity 3.6.18.** Following the steps of [Procedure 3.5.4](#), Find the pseudo-inverse of the matrix  $\begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$  given that this matrix has the SVD

$$\begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} \begin{bmatrix} 10 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}^T$$

The pseudo-inverse is which of these?

(a)  $\begin{bmatrix} 0.1 & 0.2 \\ 0.1 & 0.2 \end{bmatrix}$

(b)  $\begin{bmatrix} 0.1 & 0.1 \\ 0.2 & 0.2 \end{bmatrix}$

(c)  $\begin{bmatrix} 0.1 & -0.1 \\ -0.2 & 0.2 \end{bmatrix}$

(d)  $\begin{bmatrix} 0.1 & -0.2 \\ -0.1 & 0.2 \end{bmatrix}$



**Example 3.6.19.** Recall that [Example 3.5.1](#) explored how to best determine a weight from four apparently contradictory measurements. The exploration showed that [Procedure 3.5.4](#) agrees with the traditional method of simple averaging. Let's see that the pseudo-inverse implements the simple average of the four measurements.

Recall that [Example 3.5.1](#) sought to solve an inconsistent system  $Ax = \mathbf{b}$ , specifically

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} x = \begin{bmatrix} 84.8 \\ 84.1 \\ 84.7 \\ 84.4 \end{bmatrix}.$$

To find the pseudo-inverse of the left-hand side matrix  $A$ , seek to solve the system for arbitrary right-hand side  $\mathbf{b}$ .

(a) As used previously, this matrix  $A$  of ones has an SVD of

$$A = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \end{bmatrix} [1]^T = USV^T.$$

(b) Solve  $U\mathbf{z} = \mathbf{b}$  by computing

$$\mathbf{z} = U^T \mathbf{b} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \mathbf{b}$$

$$= \begin{bmatrix} \frac{1}{2}b_1 + \frac{1}{2}b_2 + \frac{1}{2}b_3 + \frac{1}{2}b_4 \\ \frac{1}{2}b_1 + \frac{1}{2}b_2 - \frac{1}{2}b_3 - \frac{1}{2}b_4 \\ \frac{1}{2}b_1 - \frac{1}{2}b_2 - \frac{1}{2}b_3 + \frac{1}{2}b_4 \\ \frac{1}{2}b_1 - \frac{1}{2}b_2 + \frac{1}{2}b_3 - \frac{1}{2}b_4 \end{bmatrix}.$$

(c) Now try to solve  $Sy = \mathbf{z}$ , that is,

$$\begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \end{bmatrix} y = \begin{bmatrix} \frac{1}{2}b_1 + \frac{1}{2}b_2 + \frac{1}{2}b_3 + \frac{1}{2}b_4 \\ \frac{1}{2}b_1 + \frac{1}{2}b_2 - \frac{1}{2}b_3 - \frac{1}{2}b_4 \\ \frac{1}{2}b_1 - \frac{1}{2}b_2 - \frac{1}{2}b_3 + \frac{1}{2}b_4 \\ \frac{1}{2}b_1 - \frac{1}{2}b_2 + \frac{1}{2}b_3 - \frac{1}{2}b_4 \end{bmatrix}.$$

Instead of seeking an *exact* solution, we *have to* adjust the last three components to zero. Hence we find a solution to a slightly different problem by solving

$$\begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \end{bmatrix} y = \begin{bmatrix} \frac{1}{2}b_1 + \frac{1}{2}b_2 + \frac{1}{2}b_3 + \frac{1}{2}b_4 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

with solution  $y = \frac{1}{4}b_1 + \frac{1}{4}b_2 + \frac{1}{4}b_3 + \frac{1}{4}b_4$ .

(d) Lastly, solve  $V^T x = y$  by computing

$$x = Vy = 1y = \frac{1}{4}b_1 + \frac{1}{4}b_2 + \frac{1}{4}b_3 + \frac{1}{4}b_4 = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix} \mathbf{b}.$$

Hence the pseudo-inverse of matrix  $A$  is  $A^+ = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}$ . Multiplication by this pseudo-inverse implements the traditional answer of averaging measurements. ■

**Example 3.6.20.** Recall that [Example 3.5.3](#) rates three table tennis players, Anne, Bob and Chris. The rating involved solving the inconsistent system  $A\mathbf{x} = \mathbf{b}$  for the particular matrix and vector

$$\begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}.$$

Find the pseudo-inverse of this matrix  $A$ . Use the pseudo-inverse to rate the players in the cases of [Examples 3.3.12](#) and [3.5.3](#). ■

In some common special cases there are alternative formulas for the pseudo-inverse: specifically, the cases are when the rank of the matrix is the same as the number of rows and/or columns.

**Example 3.6.21.** For the matrix  $A = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$ , confirm that  $(A^T A)^{-1} A^T$  is the pseudo-inverse that was found in [Example 3.6.14](#). ■

**Theorem 3.6.22.** *For every  $m \times n$  matrix  $A$  with  $\text{rank } A = n$  (so  $m \geq n$ ), the pseudo-inverse  $A^+ = (A^T A)^{-1} A^T$ .*

**Theorem 3.6.23.** *For every invertible matrix  $A$ , the pseudo-inverse  $A^+ = A^{-1}$ , the inverse.*

**Computer considerations** Except for easy cases, we (almost) never explicitly compute the pseudo-inverse of a matrix. In practical computation, forming  $A^T A$  and then manipulating it is both expensive and error enhancing: for example,  $\text{cond}(A^T A) = (\text{cond } A)^2$  so

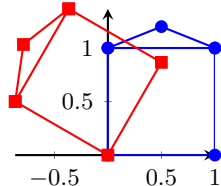
matrix  $A^T A$  typically has a much worse condition number than matrix  $A$ . Computationally there are (almost) always better ways to proceed, such as [Procedure 3.5.4](#). Like an inverse, a pseudo-inverse is a theoretical device, rarely a practical tool.

A main point of this subsection is to illustrate how a complicated procedure is conceptually expressible as a linear transformation, and so has associated matrix properties such as being equivalent to multiplication by some matrix—here the pseudo-inverse.



### 3.6.3 Function composition connects to matrix inverse

To achieve a complex goal we typically decompose the task of attaining the goal into a set of smaller tasks and achieve those tasks one after another. The analogy in linear algebra is that we often apply linear transformations one after another to build up or solve a complex problem. This section certifies how applying a sequence of linear transformations is equivalent to one grand overall linear transformation.



**Example 3.6.24** (simple rotation). Recall [Example 3.6.11b](#) on rotation by  $60^\circ$  (illustrated in the margin) with its standard matrix

$$[R] = \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}.$$

Consider two successive rotations by  $60^\circ$ : show that the standard matrix of the resultant rotation by  $120^\circ$  is the same as the matrix product  $[R][R]$ . ■

**Theorem 3.6.25.** *Let  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  and  $S : \mathbb{R}^m \rightarrow \mathbb{R}^p$  be linear transformations. Recalling the **composition** of functions is  $(S \circ T)(\mathbf{v}) = S(T(\mathbf{v}))$ , then  $S \circ T : \mathbb{R}^n \rightarrow \mathbb{R}^p$  is a linear transformation with standard matrix  $[S \circ T] = [S][T]$ .*

**Example 3.6.26.** Consider the linear transformation  $T : \mathbb{R}^3 \rightarrow \mathbb{R}^2$  defined by  $T(x_1, x_2, x_3) := (3x_1 + x_2, -x_2 - 7x_3)$ , and the linear transformation  $S : \mathbb{R}^2 \rightarrow \mathbb{R}^4$  defined by  $S(y_1, y_2) = (-y_1, -3y_1 + 2y_2, 2y_1 - y_2, 2y_2)$ . Find the standard matrix of the linear transformation  $S \circ T$ , and also that of  $T \circ S$ . ■

**Example 3.6.27.** Find the standard matrix of the transformation of the plane that first rotates by  $45^\circ$  about the origin, and then second reflects in the vertical axis. As an extension, check that although  $R \circ F$  is defined, it is different to  $F \circ R$ : the difference corresponds to the non-commutativity of matrix multiplication ([Subsection 3.1.3](#)). ■

**Activity 3.6.28.** Given the stretching transformation  $S$  with standard matrix  $[S] = \text{diag}(2, 1/2)$ , and the anti-clockwise rotation  $R$  by  $90^\circ$  with standard matrix  $[R] = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ , what is the standard matrix of the transformation composed of first the stretching and then the rotation?

(a)  $\begin{bmatrix} 0 & 2 \\ -\frac{1}{2} & 0 \end{bmatrix}$       (b)  $\begin{bmatrix} 0 & -\frac{1}{2} \\ 2 & 0 \end{bmatrix}$       (c)  $\begin{bmatrix} 0 & -2 \\ \frac{1}{2} & 0 \end{bmatrix}$       (d)  $\begin{bmatrix} 0 & \frac{1}{2} \\ -2 & 0 \end{bmatrix}$

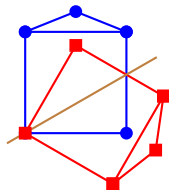


**Invert transformations** Having introduced and characterised the composition of linear transformations, we now discuss when two transformations composed together end up ‘cancelling’ each other out.

**Example 3.6.29** (inverse transformations). (a) Let  $S$  be rotation of the plane by  $60^\circ$ , and  $T$  be rotation of the plane by  $-60^\circ$ . Then  $S \circ T$  is first rotation by  $-60^\circ$  by  $T$ , and second rotation by  $60^\circ$

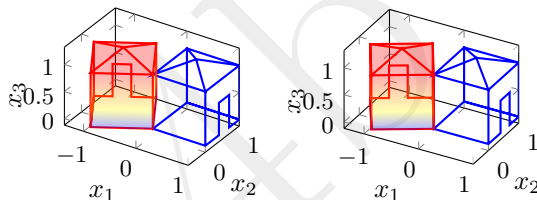
by  $S$ : the result is no change. Because  $S \circ T$  is effectively the identity transformation, we call the rotations  $S$  and  $T$  the inverse transformation of each other.

- (b) Let  $R$  be reflection of the plane in the line at  $30^\circ$  to the horizontal (illustrated in the margin). Then  $R \circ R$  is first reflection in the line at  $30^\circ$  by  $R$ , and second another reflection in the line at  $30^\circ$  by  $R$ : the result is no change. Because  $R \circ R$  is effectively the identity transformation, the reflection  $R$  is its own inverse.

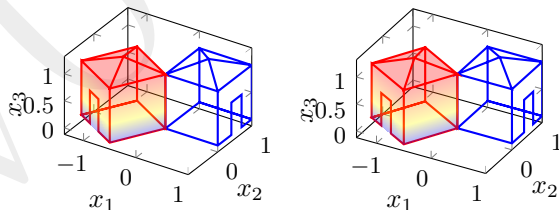


**Definition 3.6.30.** Let  $S$  and  $T$  be linear transformations from  $\mathbb{R}^n$  to  $\mathbb{R}^n$  (the same dimension). If  $S \circ T = T \circ S = I$ , the identity transformation, then  $S$  and  $T$  are **inverse transformations** of each other. Further, we say  $S$  and  $T$  are **invertible**.

**Example 3.6.31.** Let  $S : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be rotation about the vertical axis by  $120^\circ$  (as illustrated in stereo below),

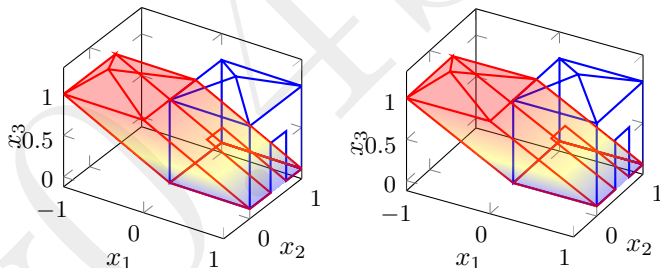


and let  $T : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  be rotation about the vertical axis by  $240^\circ$  (below).



Argue that  $S \circ T = T \circ S = I$  the identity and so  $S$  and  $T$  are inverse transformations of each other. ■

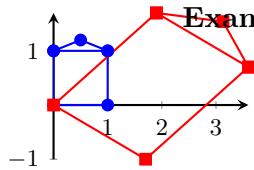
**Example 3.6.32.** In some violent weather a storm passes and the strong winds lean a house sideways as in the shear transformation illustrated below.



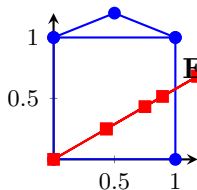
Estimate the standard matrix of the shear transformation shown. To restore the house back upright, we need to shear it an equal amount in the opposite direction: hence write down the standard matrix of the inverse shear. Confirm that the product of the two standard matrices is the standard matrix of the identity. ■

Because of the exact correspondence between linear transformations and matrix multiplication, the inverse of a transformation exactly corresponds to the inverse of a matrix. In the last [Example 3.6.32](#), because  $[R][S] = I_3$  we know that the matrices  $[R]$  and  $[S]$  are inverses of each other. Correspondingly, the transformations  $R$  and  $S$  are inverses of each other.

**Theorem 3.6.33.** *Let  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be an invertible linear transformation. Then its standard matrix  $[T]$  is invertible, and  $[T^{-1}] = [T]^{-1}$ .*



**Example 3.6.34.** Estimate the standard matrix of the linear transformation  $T$  illustrated in the margin. Then use [Theorem 3.2.7](#) to determine the standard matrix of its inverse transformation  $T^{-1}$ . Hence sketch how the inverse transforms the unit square and write a sentence or two about how the sketch confirms it is a reasonable inverse. ■



**Example 3.6.35.** Determine if the orthogonal projection of the plane onto the line at  $30^\circ$  to the horizontal (illustrated in the margin) is an invertible transformation; if it is find its inverse. ■

---

## 4 Eigenvalues and eigenvectors of symmetric matrices

---

### Chapter Contents

4.1	Introduction to eigenvalues and eigenvectors . . . . .	482
4.1.1	Systematically find eigenvalues and eigenvectors	496
4.2	Beautiful properties for symmetric matrices . . . . .	512
4.2.1	Matrix powers maintain eigenvectors . . . . .	513
4.2.2	Symmetric matrices are orthogonally diagonalisable . . . . .	521
4.2.3	Change orthonormal basis to classify quadratics	532

Recall (Subsection 3.1.2) that a symmetric matrix  $A$  is a square matrix such that  $A^T = A$ , that is,  $a_{ij} = a_{ji}$ . For example, of the



following two matrices, the first is symmetric, but the second is not:

$$\begin{bmatrix} -2 & 4 & 0 \\ 4 & 2 & -3 \\ 0 & -3 & 1 \end{bmatrix}; \quad \begin{bmatrix} -1 & 3 & 0 \\ 1 & 1 & 0 \\ 0 & -3 & 1 \end{bmatrix}.$$

**Example 4.0.1.** Compute some SVDs of random symmetric matrices,  $A = USV^T$ , observe in the SVDs that the columns of  $U$  are always  $\pm$  the columns of  $V$  (well, almost always). ■

Why, for symmetric matrices, are the columns of  $U$  (almost) always  $\pm$  the columns of  $V$ ? The answer is connected to the following rearrangement of an SVD. Because  $A = USV^T$ , post-multiplying by  $V$  gives  $AV = USV^TV = US$ , and then the  $j$ th column of the two sides of  $AV = US$  determines  $A\mathbf{v}_j = \sigma_j\mathbf{u}_j$ . [Example 4.0.1](#) observes for symmetric  $A$  that  $\mathbf{u}_j = \pm\mathbf{v}_j$  (almost always) so this last equation becomes  $A\mathbf{v}_j = (\pm\sigma_j)\mathbf{v}_j$ . This equation is of the important form  $A\mathbf{v} = \lambda\mathbf{v}$ . This form is important because it is the mathematical expression of the following geometric question: for

The symbol  $\lambda$  is the Greek letter lambda, and denotes eigenvalues.

what vectors  $\mathbf{v}$  does multiplication by  $A$  just stretch/shrink  $\mathbf{v}$  by some scalar  $\lambda$ ?

**Solid modelling** Lean with a hand on a table/wall: the force changes depending upon the orientation of the surface. Similarly inside any solid: the internal forces  $= A\mathbf{v}$  where  $\mathbf{v}$  is the orthogonal unit vector to the internal ‘surface’. Matrix  $A$  is always symmetric. To know whether a material will break apart under pulling, or to crumble under compression, we need to know where the extreme forces are. They are found as solutions to  $A\mathbf{v} = \lambda\mathbf{v}$  where  $\mathbf{v}$  gives the direction and  $\lambda$  the strength of the force. To understand the potential failure of the material we need to solve equations in the form  $A\mathbf{v} = \lambda\mathbf{v}$ .

## 4.1 Introduction to eigenvalues and eigenvectors

### Section Contents

4.1.1 Systematically find eigenvalues and eigenvectors 496

Compute eigenvalues and eigenvectors . . . . 497

Find eigenvalues and eigenvectors by hand . . 506

This chapter focuses on some marvellous properties of symmetric matrices. Nonetheless it defines some basic concepts which also apply to general matrices. [Chapter 7](#) explores analogous properties for such general matrices. The marvellously useful properties developed here result from asking for which vectors does multiplication by a given matrix simply stretch or shrink the vector.

**Definition 4.1.1.** *Let  $A$  be a square matrix. A scalar  $\lambda$  (lambda) is called an **eigenvalue** of  $A$  if there is a nonzero vector  $\mathbf{x}$  such that  $A\mathbf{x} = \lambda\mathbf{x}$ . Such a vector  $\mathbf{x}$  is called an **eigenvector** of  $A$  corresponding to the eigenvalue  $\lambda$ .*

**Example 4.1.2.** Consider the symmetric matrix

$$A = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}.$$

- (a) Verify an eigenvector is  $(1, 0, -1)$ . What is the corresponding eigenvalue?
- (b) Verify that  $(2, -4, 2)$  is an eigenvector. What is its corresponding eigenvalue.
- (c) Verify that  $(1, 2, 1)$  is not an eigenvector.
- (d) Use inspection to guess and verify another eigenvector (not proportional to either of the above two). What is its eigenvalue?



**Activity 4.1.3.** Which of the following vectors is an eigenvector of the symmetric matrix  $\begin{bmatrix} -1 & 12 \\ 12 & 6 \end{bmatrix}$ ?

(a)  $\begin{bmatrix} 4 \\ -3 \end{bmatrix}$

(b)  $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$

(c)  $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$

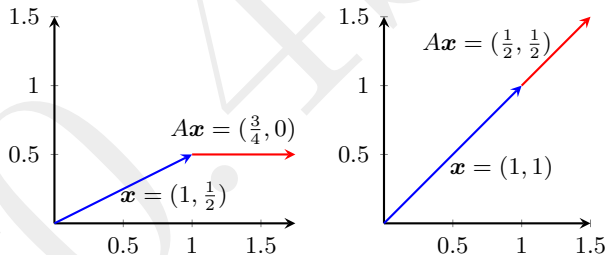
(d)  $\begin{bmatrix} -3 \\ 1 \end{bmatrix}$



Importantly, eigenvectors tell us key directions of a given matrix: the directions in which the multiplication by a matrix is to simply stretch, shrink, or reverse by a factor: the factor being the corresponding eigenvalue. In two dimensional plots we can graphically estimate eigenvectors and eigenvalues. For some examples and exercises we plot a given vector  $\mathbf{x}$  and join onto its head the vector  $A\mathbf{x}$ :

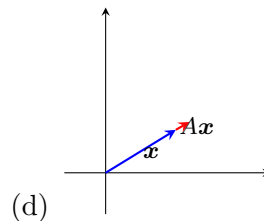
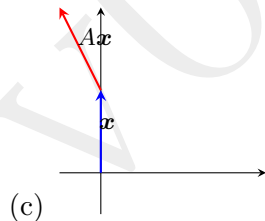
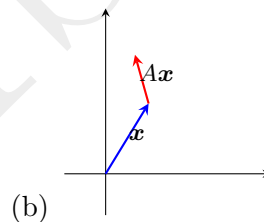
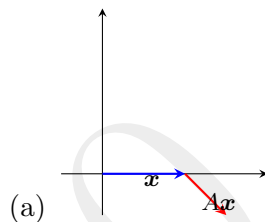
- if both  $\mathbf{x}$  and  $A\mathbf{x}$  are aligned in the same direction, or opposite direction, then  $\mathbf{x}$  is an eigenvector;
- if they form some other angle, then  $\mathbf{x}$  is not an eigenvector.

**Example 4.1.4.** Let the matrix  $A = \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix}$ . The plot below-left shows the vector  $\mathbf{x} = (1, \frac{1}{2})$ , and adjoined to its head the matrix-vector product  $A\mathbf{x} = (\frac{3}{4}, 0)$ : because the two are at an angle,  $(1, \frac{1}{2})$  is not an eigenvector.



However, as plotted above-right, for the vector  $\mathbf{x} = (1, 1)$  the matrix-vector product  $A\mathbf{x} = (\frac{1}{2}, \frac{1}{2})$  and the plot of these vectors head-to-tail illustrates that they are aligned in the same direction. Because of the alignment,  $(1, 1)$  is an eigenvector of this matrix. The constant of proportionality is the corresponding eigenvalue: here  $A\mathbf{x} = (\frac{1}{2}, \frac{1}{2}) = \frac{1}{2}(1, 1) = \frac{1}{2}\mathbf{x}$  so the eigenvalue is  $\lambda = \frac{1}{2}$ . ■

**Activity 4.1.5.** For some matrix  $A$ , the following pictures plot a vector  $\mathbf{x}$  and the corresponding product  $A\mathbf{x}$ , head-to-tail. Which picture indicates that  $\mathbf{x}$  is an eigenvector of the matrix?



**Activity 4.1.6.** Further, for the picture in Activity 4.1.5 that indicates  $\mathbf{x}$  is an eigenvector, is the corresponding eigenvalue  $\lambda$ :

- (a)  $0.5 > \lambda > 0$       (b)  $1 > \lambda > 0.5$       (c)  $\lambda > 1$       (d)  $0 > \lambda$

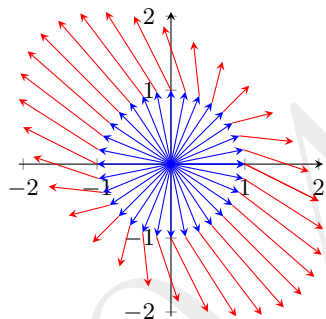


As in the next example, we sometimes plot for many directions  $\mathbf{x}$  a diagram of vector  $A\mathbf{x}$  adjoined head-to-tail to vector  $\mathbf{x}$ . Then inspection estimates the eigenvectors and corresponding eigenvalues (Schonefeld 1995).

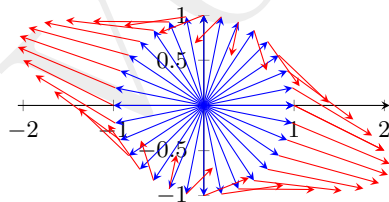


**Example 4.1.7** (graphical eigenvectors one).

The MATLAB function `eigshow(A)` provides an interactive alternative to this static view.



The plot on the left shows many unit vectors  $\mathbf{x}$  (blue), and for some matrix  $A$  the corresponding vectors  $A\mathbf{x}$  (red) adjoined. Estimate which directions  $\mathbf{x}$  are eigenvectors, and for each eigenvector estimate the corresponding eigenvalue.

**Example 4.1.8** (graphical eigenvectors two).

The plot on the left shows many unit vectors  $\mathbf{x}$  (blue), and for some matrix  $A$  the corresponding vectors  $A\mathbf{x}$  (red) adjoined.

Estimate which directions  $\mathbf{x}$  are eigenvectors, and for each eigen-

vector estimate the corresponding eigenvalue. ■

**Example 4.1.9** (diagonal matrix). The eigenvalues of a (square) diagonal matrix are the entries on the diagonal. Consider an  $n \times n$  diagonal matrix

$$D = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix}.$$

Multiply by the standard unit vectors  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$  in turn:

$$D\mathbf{e}_1 = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} d_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = d_1\mathbf{e}_1;$$

$$\begin{aligned}
D\mathbf{e}_2 &= \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ d_2 \\ \vdots \\ 0 \end{bmatrix} = d_2 \mathbf{e}_2; \\
&\vdots \\
D\mathbf{e}_n &= \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ d_n \end{bmatrix} = d_n \mathbf{e}_n.
\end{aligned}$$

By [Definition 4.1.1](#), each diagonal element  $d_j$  is an eigenvalue of the diagonal matrix, and the standard unit vector  $\mathbf{e}_j$  is a corresponding eigenvector. ■

**Eigenvalues** The  $3 \times 3$  matrix of [Example 4.1.2](#) has three eigenvalues. The  $2 \times 2$  matrices underlying [Examples 4.1.7](#) and [4.1.8](#) both have two eigenvalues. [Example 4.1.9](#) shows an  $n \times n$  diagonal

matrix has  $n$  eigenvalues. The next section establishes the general pattern that an  $n \times n$  *symmetric matrix* generally has  $n$  real eigenvalues. However, the eigenvalues of non-symmetric matrices are more complex (in both senses of the word) as explored by [Chapter 7](#).

**Eigenvectors** It is the direction of eigenvectors that is important. In [Example 4.1.2](#) any nonzero multiple of  $(1, -2, 1)$ , positive or negative, is also an eigenvector corresponding to eigenvalue  $\lambda = 3$ . In the diagonal matrices of [Example 4.1.9](#), a straightforward extension of the working shows any nonzero multiple of the standard unit vector  $\mathbf{e}_j$  is an eigenvector corresponding to the eigenvalue  $d_j$ . Let's collect all possible eigenvectors into a subspace.

**Theorem 4.1.10.** *Let  $A$  be a square matrix. A scalar  $\lambda$  is an eigenvalue of  $A$  iff the homogeneous linear system  $(A - \lambda I)\mathbf{x} = \mathbf{0}$  has nonzero solutions  $\mathbf{x}$ . The set of all eigenvectors corresponding to any one eigenvalue  $\lambda$ , together with the zero vector, is a subspace; the subspace is called the **eigenspace** of  $\lambda$  and is denoted by  $\mathbb{E}_\lambda$ .*

Hereafter, “iff” is short for “if and only if”.

**Example 4.1.11.** Reconsider the symmetric matrix

$$A = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}$$

of [Example 4.1.2](#). Find the eigenspaces  $\mathbb{E}_1$ ,  $\mathbb{E}_3$  and  $\mathbb{E}_0$ . ■

**Activity 4.1.12.** Which line, in the  $xy$ -plane, is the eigenspace corresponding to the eigenvalue  $-5$  of the matrix  $\begin{bmatrix} 3 & 4 \\ 4 & -3 \end{bmatrix}$ ?

(a)  $x + 2y = 0$

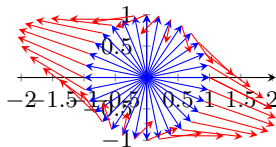
(b)  $y = 2x$

(c)  $x = 2y$

(d)  $2x + y = 0$



**Example 4.1.13** (graphical eigenspaces).



The plot on the left shows unit vectors  $\mathbf{x}$  (blue), and for the matrix  $A$  of [Example 4.1.8](#) the corresponding vectors  $A\mathbf{x}$  (red) adjoined. Estimate and draw the eigenspaces of matrix  $A$ . ■

**Example 4.1.14.** Eigenspaces may be multidimensional. Find the eigenspaces of the diagonal matrix

$$D = \begin{bmatrix} -\frac{1}{3} & 0 & 0 \\ 0 & \frac{3}{2} & 0 \\ 0 & 0 & \frac{3}{2} \end{bmatrix}.$$
 ■

**Definition 4.1.15.** For every real symmetric matrix  $A$ , the **multiplicity** of an eigenvalue  $\lambda$  of  $A$  is the dimension of the corresponding eigenspace  $\mathbb{E}_\lambda$ .

**Example 4.1.16.** The multiplicity of the various eigenvalues in earlier examples are the following.

4.1.11 Recall that in this example:

- the eigenspace  $\mathbb{E}_1 = \text{span}\{(1, 0, -1)\}$  has dimension one, so the multiplicity of eigenvalue  $\lambda = 1$  is one;
- the eigenspace  $\mathbb{E}_3 = \text{span}\{(1, -2, 1)\}$  has dimension one, so the multiplicity of eigenvalue  $\lambda = 3$  is one; and
- the eigenspace  $\mathbb{E}_0 = \text{span}\{(1, 1, 1)\}$  has dimension one, so the multiplicity of eigenvalue  $\lambda = 0$  is one.

4.1.14 Recall that in this example:

- the eigenspace  $\mathbb{E}_{-1/3} = \text{span}\{\mathbf{e}_1\}$  has dimension one, so the multiplicity of eigenvalue  $\lambda = -1/3$  is one; and

- the eigenspace  $\mathbb{E}_{3/2} = \text{span}\{\mathbf{e}_2, \mathbf{e}_3\}$  has dimension two, so the multiplicity of eigenvalue  $\lambda = 3/2$  is two.





### 4.1.1 Systematically find eigenvalues and eigenvectors

Computer packages easily compute eigenvalues and eigenvectors for us. Sometimes we need to explicitly see dependence upon a parameter so this subsection also develops how to find by hand the eigenvalues and eigenvectors of small matrices. We start with computation.

## Compute eigenvalues and eigenvectors

**Compute in Matlab/Octave.** `[V,D]=eig(A)` computes eigenvalues and eigenvectors. The eigenvalues are placed in the diagonal of  $D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ . The  $j$ th column of  $V$  is a unit eigenvector corresponding to the  $j$ th eigenvalue  $\lambda_j$ :  $V = [\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_n]$ . If the matrix  $A$  is real and symmetric, then  $V$  is an orthogonal matrix (Theorem 4.2.19).

**Example 4.1.17.** Reconsider the symmetric matrix of Example 4.1.2:

$$A = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}.$$

Use MATLAB/Octave to find its eigenvalues and corresponding eigenvectors. Confirm that  $AV = VD$  for matrices  $V = [\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_n]$  and  $D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ , and confirm that the computed  $V$  is orthogonal. ■

Table 4.1: As well as the MATLAB/Octave commands and operations listed in Tables 1.2, 2.3, 3.1, 3.2, and 3.3 we need the eigenvector function.

- 
- `[V,D]=eig(A)` computes eigenvectors and the eigenvalues of the  $n \times n$  square matrix  $A$ .
    - The  $n$  eigenvalues of  $A$  (repeated according to their multiplicity, Definition 4.1.15) form the diagonal of  $n \times n$  square matrix  $D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ .
    - Corresponding to the  $j$ th eigenvalue  $\lambda_j$ , the  $j$ th column of  $n \times n$  square matrix  $V$  is an eigenvector (of unit length).
  - `eig(A)` by itself just reports, in a vector, the eigenvalues of square matrix  $A$  (repeated according to their multiplicity, Definition 4.1.15).
  - If the matrix  $A$  is a real symmetric matrix, then the eigenvalues and eigenvectors are all real, and the eigenvector matrix  $V$  is orthogonal.

If the matrix  $A$  is either not symmetric, or is complex valued, then the eigenvalues and eigenvectors may be complex valued.

---

**Activity 4.1.18.** The statement  $[V,D]=\text{eig}(A)$  returns the following result (2 d.p.)

$V =$

0.50	0.50	-0.10	-0.70
0.10	-0.70	0.50	-0.50
-0.70	-0.10	-0.50	-0.50
0.50	-0.50	-0.70	0.10

$D =$

-0.10	0	0	0
0	0.10	0	0
0	0	0.30	0
0	0	0	0.50

Which of the following is *not* an eigenvalue of the matrix  $A$ ?

- (a) 0.5                      (b) 0.1                      (c) -0.5                      (d) -0.1



**Example 4.1.19** (application to vibrations). Consider three masses in a row connected by two springs: on a tiny scale this could represent a molecule of carbon dioxide ( $\text{CO}_2$ ). For simplicity suppose the three masses are equal, and the spring strengths are equal. Define  $y_i(t)$  to be the distance from equilibrium of the  $i$ th mass. Newton's law for bodies says the acceleration of the mass,  $d^2y_i/dt^2$ , is proportional to the forces due to the springs. Hooke's law for springs says the force is proportional to the stretching/compression of the springs,  $y_2 - y_1$  and  $y_3 - y_2$ . For simplicity, suppose the constants of proportionality are all one.

- The left mass ( $y_1$ ) is accelerated by the spring connecting it to the middle mass ( $y_2$ ); that is,  $d^2y_1/dt^2 = y_2 - y_1$ .
- The middle mass ( $y_2$ ) is accelerated by the springs connecting it to the left mass ( $y_1$ ) and to the right mass ( $y_3$ ); that is,  $d^2y_2/dt^2 = (y_1 - y_2) + (y_3 - y_2) = y_1 - 2y_2 + y_3$ .
- The right mass ( $y_3$ ) is accelerated by the spring connecting it to the middle mass ( $y_2$ ); that is,  $d^2y_3/dt^2 = y_2 - y_3$ .

Guess there are solutions oscillating in time, so let's see if we can find

solutions  $y_i(t) = x_i \cos(ft)$  for some as yet unknown frequency  $f$ . Substitute and the three differential equations become

$$\begin{aligned}-f^2 x_1 \cos(ft) &= x_2 \cos(ft) - x_1 \cos(ft), \\ -f^2 x_2 \cos(ft) &= x_1 \cos(ft) - 2x_2 \cos(ft) + x_3 \cos(ft), \\ -f^2 x_3 \cos(ft) &= x_2 \cos(ft) - x_3 \cos(ft).\end{aligned}$$

These are satisfied for all time  $t$  only if the coefficients of the cosine are equal on each side of each equation:

$$\begin{aligned}-f^2 x_1 &= x_2 - x_1, \\ -f^2 x_2 &= x_1 - 2x_2 + x_3, \\ -f^2 x_3 &= x_2 - x_3.\end{aligned}$$

Moving the terms on the left to the right, and all terms on the right to the left, this becomes the eigenproblem  $A\mathbf{x} = \lambda\mathbf{x}$  for symmetric matrix  $A$  of [Example 4.1.17](#) and for eigenvalue  $\lambda = f^2$ , the square of the as yet unknown frequency. The symmetry of matrix  $A$  reflects Newton's law that every action has an equal and opposite reaction: symmetric matrices arise commonly in applications.

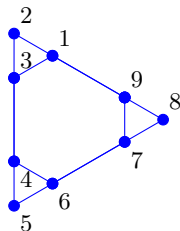
Example 4.1.17 tells us that there are three possible eigenvalue and eigenvector solutions for us to interpret.

- The eigenvalue  $\lambda = 1$  and corresponding eigenvector  $\mathbf{x} \propto (-1, 0, 1)$  corresponds to oscillations of frequency  $f = \sqrt{\lambda} = \sqrt{1} = 1$ . The eigenvector  $(-1, 0, 1)$  shows the middle mass is stationary while the outer two masses oscillate in and out in opposition to each other.
- The eigenvalue  $\lambda = 3$  and corresponding eigenvector  $\mathbf{x} \propto (1, -2, 1)$  corresponds to oscillations of higher frequency  $f = \sqrt{\lambda} = \sqrt{3}$ . The eigenvector  $(1, -2, 1)$  shows the outer two masses oscillate together, and the middle mass moves opposite to them.
- The eigenvalue  $\lambda = 0$  and corresponding eigenvector  $\mathbf{x} \propto (1, 1, 1)$  appears as oscillations of zero frequency  $f = \sqrt{\lambda} = \sqrt{0} = 0$  which is a static displacement. The eigenvector  $(1, 1, 1)$  shows the static displacement is that of all three masses moved all together as a unit.

That these three solutions combine together form a general solution

of the system of differential equations is a topic for a course on differential equations. ■

**Example 4.1.20** (Sierpinski network). Consider three triangles formed into a triangle (as shown in the margin)—perhaps because triangles make strong structures, or perhaps because of a hierarchical computer/social network. Form an matrix  $A = [a_{ij}]$  of ones if node  $i$  is connected to node  $j$ ; set the diagonal  $a_{ii}$  to be minus the number of other nodes to which node  $i$  is connected; and all other components of  $A$  are zero. The symmetry of the matrix  $A$  follows from the symmetry of the connections: construct the matrix, check it is symmetric, and find the eigenvalues and eigenspaces with MATLAB/Octave, and their multiplicity. For the computed matrices  $V$  and  $D$ , check that  $AV = VD$  and also that  $V$  is orthogonal. ■



Challenge: find the two smallest connected networks that have different connectivity and yet the same eigenvalues (unit strength connections).

In 1966 Mark Kac asked “Can one hear the shape of the drum?” That is, from just knowing the eigenvalues of a network such as the one in [Example 4.1.20](#), can one infer the connectivity of the



network? The question for 2D drums was answered “no” in 1992 by Gordon, Webb and Wolpert who constructed two different shaped 2D drums which have the same set of frequencies of oscillation: that is, the same set of eigenvalues.

Why write “the computation may give” in [Example 4.1.20](#)? The reason is associated with the duplicated eigenvalues. What is important is the eigenspace. When an eigenvalue of a symmetric matrix is duplicated (or triplicated) in the diagonal  $D$  then there are many choices of eigenvectors that form an orthonormal basis ([Definition 3.4.18](#)) of the eigenspace (the same holds for singular vectors of a duplicated singular value). Different algorithms may report different orthonormal bases of the same eigenspace. The bases given in [Example 4.1.20](#) are just one possibility for each eigenspace.

**Theorem 4.1.21.** *For every  $n \times n$  square matrix  $A$  (not just symmetric),  $\lambda_1, \lambda_2, \dots, \lambda_m$  are eigenvalues of  $A$  with corresponding eigenvectors  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$ , for some  $m$  (commonly  $m = n$ ), iff  $AV = VD$  for diagonal matrix  $D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m)$  and  $n \times m$  matrix  $V = [\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_m]$  for non-zero  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$ .*

**Example 4.1.22.** Use MATLAB/Octave to compute eigenvectors and the eigenvalues of (symmetric) matrix

$$A = \begin{bmatrix} 2 & 2 & -2 & 0 \\ 2 & -1 & -2 & -3 \\ -2 & -2 & 4 & 0 \\ 0 & -3 & 0 & 1 \end{bmatrix}$$

Confirm  $AV = VD$  for the computed matrices. ■

## Find eigenvalues and eigenvectors by hand

- Recall from previous study ([Theorem 3.2.7](#)) that a  $2 \times 2$  matrix  $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$  has determinant  $\det A = |A| = ad - bc$ , and that  $A$  is not invertible iff  $\det A = 0$ .
- Similarly, although not justified until [Chapter 6](#), a  $3 \times 3$  matrix  $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$  has determinant  $\det A = |A| = aei + bfg + cdh - ceg - afh - bdi$ , and  $A$  is not invertible iff  $\det A = 0$ .

This section shows these two formulas for a determinant are useful for hand calculations on small problems. The formulas are best remembered via the following diagrams where products along the red lines are subtracted from the sum of products along the blue lines, respectively:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \quad \begin{bmatrix} a & b \\ c & d \\ e & f \\ g & h \end{bmatrix} \quad (4.1)$$

Chapter 6 extends the determinant to any size matrix, and explores more useful properties, but for now this is the information we need on determinants.

For hand calculation on small matrices the key is the following. By Definition 4.1.1 eigenvalues and eigenvectors are determined from  $A\mathbf{x} = \lambda\mathbf{x}$ . Rearranging, this equation is equivalent to  $(A - \lambda I)\mathbf{x} = \mathbf{0}$ . Both Theorem 3.2.7 ( $2 \times 2$  matrices) and Theorem 6.1.29 (general matrices) establish that  $(A - \lambda I)\mathbf{x} = \mathbf{0}$  has nonzero solutions  $\mathbf{x}$  iff the determinant  $\det(A - \lambda I) = 0$ . Since eigenvectors must be nonzero, the eigenvalues of a square matrix are precisely the solutions of  $\det(A - \lambda I) = 0$ . This reasoning leads to the following procedure.

**Procedure 4.1.23** (eigenvalues and eigenvectors). *To find by hand eigenvalues and eigenvectors of any (small) square matrix  $A$ :*

1. find all eigenvalues by solving the **characteristic equation** of  $A$ ,  $\det(A - \lambda I) = 0$ ;
2. for each eigenvalue  $\lambda$ , solve the homogeneous  $(A - \lambda I)\mathbf{x} = \mathbf{0}$  to find the corresponding eigenspace  $\mathbb{E}_\lambda$ ;

3. write each eigenspace as the span of a few chosen eigenvectors.

This procedure applies to general matrices  $A$ , as fully established in [Section 7.1](#), but this chapter uses it only for small symmetric matrices. Further, this chapter uses it only as a convenient method to illustrate some properties by hand calculation. None of the beautiful theorems of the next [Section 4.2](#) for symmetric matrices are based upon this ‘by-hand’ procedure.

**Example 4.1.24.** Use [Procedure 4.1.23](#) to find the eigenvalues and eigenvectors of the matrix

$$A = \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix}$$

(this is the matrix illustrated in Examples [4.1.4](#) and [4.1.7](#)). ■

**Activity 4.1.25.** Use the characteristic equation to determine all eigenvalues of the matrix  $A = \begin{bmatrix} 3 & 2 \\ 2 & 0 \end{bmatrix}$ . They are which of the following?

(a) 3, 4

(b) -4, 1

(c) 0, 3

(d) -1, 4



**Example 4.1.26.** Use the determinant to confirm that  $\lambda = 0, 1, 3$  are the *only* eigenvalues of the matrix

$$A = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}.$$

([Example 4.1.11](#) already found the eigenspaces corresponding to these three eigenvalues.)



**Example 4.1.27.** Use [Procedure 4.1.23](#) to find all eigenvalues and the corresponding eigenspaces of the symmetric matrix

$$A = \begin{bmatrix} -2 & 0 & -6 \\ 0 & 4 & 6 \\ -6 & 6 & -9 \end{bmatrix}.$$



General matrices may have complex valued eigenvalues and eigenvectors, as seen in the next example, and for good reasons in some applications. One of the key results of the next [Section 4.2](#) is to prove that real symmetric matrices always have real eigenvalues and eigenvectors. There are many applications where this reality is crucial.

This example aims to recall basic properties of complex numbers as a prelude to the proof of the reality of eigenvalues for every symmetric matrix.

**Example 4.1.28.** Find the eigenvalues and a corresponding eigenvector for the non-symmetric matrix  $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ .



**Example 4.1.28** is a problem that might arise using calculus to describe the dynamics of a mass on a spring. Let the displacement of the mass be  $y_1(t)$  then Newton's law says the acceleration  $d^2y_1/dt^2 \propto -y_1$ , the negative of the displacement; for simplicity, let the constant of proportionality be one. Introduce  $y_2(t) = dy_1/dt$  then Newton's law becomes  $dy_2/dt = -y_1$ . Seek solutions of these two first-order differential equations in the form  $y_j(t) = x_j e^{\lambda t}$  and the differential equations become  $x_2 = \lambda x_1$  and  $\lambda x_2 = -x_1$  respectively. Forming into a matrix-vector problem these are

$$\begin{bmatrix} x_2 \\ -x_1 \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \iff \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \mathbf{x} = \lambda \mathbf{x}.$$

We need to find the eigenvalues and eigenvectors of the matrix: we derive eigenvalues are  $\lambda = \pm\sqrt{-1} = \pm i$ . Physically, such complex eigenvalues represent oscillations in time  $t$  since, for example,  $e^{\lambda t} = e^{it} = \cos t + i \sin t$  by Euler's formula.



## 4.2 Beautiful properties for symmetric matrices

### Section Contents

4.2.1	Matrix powers maintain eigenvectors . . . . .	513
4.2.2	Symmetric matrices are orthogonally diagonalisable . . . . .	521
4.2.3	Change orthonormal basis to classify quadratics	532
	Graph quadratic equations . . . . .	534
	Simplify quadratic forms . . . . .	539

This section starts by exploring two properties for eigenvalues of general matrices, and then proceeds to the special case of real symmetric matrices. Symmetric matrices have the beautifully useful properties of always having real eigenvalues and orthogonal eigenvectors.

### 4.2.1 Matrix powers maintain eigenvectors

Recall that [Section 3.2](#) introduced the inverse of a matrix ([Definition 3.2.2](#)). This first theorem links an eigenvalue of zero to the non-existence of an inverse and hence links a zero eigenvalue to problematic linear equations.

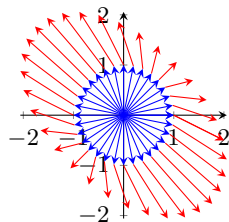
**Theorem 4.2.1.** *A square matrix is invertible iff zero is not an eigenvalue of the matrix.*

**Example 4.2.2.** • The  $3 \times 3$  matrix of [Example 4.1.2](#) (also [4.1.11](#), [4.1.17](#) and [4.1.26](#)) is not invertible as among its eigenvalues of 0, 1 and 3 it has zero as an eigenvalue.

- The plot in the margin shows (unit) vectors  $\mathbf{x}$  (blue), and for some matrix  $A$  the corresponding vectors  $A\mathbf{x}$  (red) adjoined. There are no directions  $\mathbf{x}$  for which  $A\mathbf{x} = \mathbf{0} = 0\mathbf{x}$ . Hence zero cannot be an eigenvalue and the matrix  $A$  must be invertible.

Similarly for [Example 4.1.8](#).

- The  $3 \times 3$  diagonal matrix of [Example 4.1.14](#) has eigenvalues



of only  $-\frac{1}{3}$  and  $\frac{3}{2}$ . Since zero is not an eigenvalue, the matrix is invertible.

- The  $9 \times 9$  matrix of the Sierpinski network in [Example 4.1.20](#) is not invertible as it has zero among its five eigenvalues.
- The  $2 \times 2$  matrix of [Example 4.1.24](#) is invertible as its eigenvalues are  $\lambda = \frac{1}{2}, \frac{3}{2}$ , neither of which are zero. Indeed, the matrix

$$A = \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix}, \quad \text{has inverse } A^{-1} = \begin{bmatrix} \frac{4}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{4}{3} \end{bmatrix}$$

as matrix multiplication confirms.

- The  $2 \times 2$  non-symmetric matrix of [Example 4.1.28](#) is invertible because zero is not among its eigenvalues of  $\lambda = \pm i$ . Indeed, the matrix

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad \text{has inverse } A^{-1} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

as matrix multiplication confirms.



**Example 4.2.3.** The next theorem considers eigenvalues and eigenvectors of powers of a matrix. Two examples are the following.

- Recall the matrix  $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$  has eigenvalues  $\lambda = \pm i$ . The square of this matrix

$$A^2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$$

is diagonal so its eigenvalues are the diagonal elements (Example 4.1.9), namely the only eigenvalue is  $-1$ . Observe that  $A^2$ 's eigenvalue,  $-1 = (\pm i)^2$ , is the square of the eigenvalues of  $A$ . That the eigenvalues of  $A^2$  are the square of those of  $A$  holds generally.

- Also recall matrix

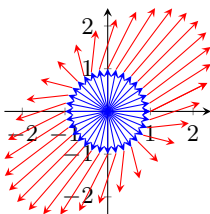
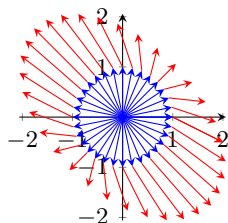
$$A = \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix}, \quad \text{has inverse } A^{-1} = \begin{bmatrix} \frac{4}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{4}{3} \end{bmatrix}.$$

Let's determine the eigenvalues of this inverse. Its characteristic equation (defined in [Procedure 4.1.23](#)) is

$$\det(A^{-1} - \lambda I) = \begin{vmatrix} \frac{4}{3} - \lambda & \frac{2}{3} \\ \frac{2}{3} & \frac{4}{3} - \lambda \end{vmatrix} = \left(\frac{4}{3} - \lambda\right)^2 - \frac{4}{9} = 0.$$

That is,  $(\lambda - \frac{4}{3})^2 = \frac{4}{9}$ . Taking the square-root of both sides gives  $\lambda - \frac{4}{3} = \pm \frac{2}{3}$ ; that is, the two eigenvalues of the inverse  $A^{-1}$  are  $\lambda = \frac{4}{3} \pm \frac{2}{3} = 2, \frac{2}{3}$ . Observe these eigenvalues of the inverse are the reciprocals of the eigenvalues  $\frac{1}{2}, \frac{3}{2}$  of  $A$ . This reciprocal relation also holds generally.

The marginal pictures illustrate the reciprocal relation graphically: the first picture shows  $A\mathbf{x}$  for various  $\mathbf{x}$ , the second picture shows  $A^{-1}\mathbf{x}$ . The eigenvector directions are the same for both matrix and inverse. But in those eigenvector directions where the matrix stretches, the inverse shrinks, and where the matrix shrinks, the inverse stretches. In contrast, in directions which are not eigenvectors, the relationship between  $A\mathbf{x}$  and  $A^{-1}\mathbf{x}$  is somewhat obscure.



**Theorem 4.2.4.** *Let  $A$  be a square matrix with eigenvalue  $\lambda$  and corresponding eigenvector  $\mathbf{x}$ .*

- (a) For every positive integer  $n$ ,  $\lambda^n$  is an eigenvalue of  $A^n$  with corresponding eigenvector  $\mathbf{x}$ .*
- (b) If  $A$  is invertible, then  $1/\lambda$  is an eigenvalue of  $A^{-1}$  with corresponding eigenvector  $\mathbf{x}$ .*
- (c) If  $A$  is invertible, then for every integer  $n$ ,  $\lambda^n$  is an eigenvalue of  $A^n$  with corresponding eigenvector  $\mathbf{x}$ .*

**Example 4.2.5.** Recall from [Example 4.1.24](#) that matrix

$$A = \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix}$$

has eigenvalues  $1/2$  and  $3/2$  with corresponding eigenvectors  $(1, 1)$  and  $(1, -1)$  respectively. Confirm matrix  $A^2$  has eigenvalues which are these squared, and corresponding to the same eigenvectors. ■

**Activity 4.2.6.** You are given that  $-3$  and  $2$  are eigenvalues of the matrix

$$A = \begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix}.$$

- Which of the following matrices has an eigenvalue of  $8$ ?

(a)  $A^{-1}$       (b)  $A^3$       (c)  $A^2$       (d)  $A^{-2}$

- Further, which of the above matrices has eigenvalue  $1/9$ ?



**Example 4.2.7.** Given that the matrix

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

has eigenvalues  $2$ ,  $1$  and  $-1$  with corresponding eigenvectors  $(1, 1, 1)$ ,  $(-1, 0, 1)$  and  $(1, -2, 1)$  respectively. Confirm matrix  $A^2$  has eigenvalues which are these squared, and corresponding to the same

eigenvectors. Given the inverse

$$A^{-1} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

confirm its eigenvalues are the reciprocals of those of  $A$ , and for corresponding eigenvectors. ■

**Example 4.2.8** (long term age structure). Recall [Example 3.1.9](#) introduced how to use a Leslie matrix to predict the future population of an animal. In the example, letting  $\mathbf{x} = (x_1, x_2, x_3)$  be the current number of pups, juveniles, and mature females respectively, then for the Leslie matrix

$$L = \begin{bmatrix} 0 & 0 & 4 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

the predicted population numbers after a year is  $\mathbf{x}' = L\mathbf{x}$ , after two years is  $\mathbf{x}'' = L\mathbf{x}' = L^2\mathbf{x}$ , and so on. Predict what happens after



many generations: does the population die out? grow? oscillate?



### 4.2.2 Symmetric matrices are orthogonally diagonalisable

General matrices may have complex valued eigenvalues (as in Examples 4.1.28 and 4.2.8): that real symmetric matrices always have real eigenvalues (such as in all matrices of Examples 4.2.5 and 4.2.7) is a special property that often reflects the physical reality of many applications.

To establish the reality of eigenvalues (Theorem 4.2.9), the proof invokes contradiction. The contradiction is to assume a complex valued eigenvalue exists, and then prove it cannot. Consequently, the proof of the next Theorem 4.2.9 needs to use some complex numbers and some properties of complex numbers. Recall that any complex number  $z = a + bi$  has a complex conjugate  $\bar{z} = a - bi$  (denoted by the overbar), and that a complex number equals its conjugate only if it is real valued (the imaginary part is zero). Such properties of complex numbers and operations also hold for complex valued vectors, complex valued matrices, and arithmetic operations with complex valued matrices and vectors.

**Theorem 4.2.9.** *For every real symmetric matrix  $A$ , the eigenvalues of  $A$  are all real.*

The other property that we have seen graphically for 2D matrices is that the eigenvectors of symmetric matrices are orthogonal. For [Example 4.2.3](#), both the matrices  $A$  and  $A^{-1}$  in the second part are symmetric and from the marginal illustration their eigenvectors are proportional to  $(1, 1)$  and  $(-1, 1)$  which are orthogonal directions—they are at right angles in the illustration.

**Example 4.2.10.** Recall [Example 4.1.27](#) found the  $3 \times 3$  symmetric matrix

$$\begin{bmatrix} -2 & 0 & -6 \\ 0 & 4 & 6 \\ -6 & 6 & -9 \end{bmatrix}$$

has eigenspaces  $\mathbb{E}_0 = \text{span}\{(-6, -3, 2)\}$ ,  $\mathbb{E}_7 = \text{span}\{(-2, 6, 3)\}$  and  $\mathbb{E}_{-14} = \text{span}\{(3, -2, 6)\}$ . These eigenspaces are orthogonal as evidenced by the dot products of the basis vectors in each span:

$$\mathbb{E}_0, \mathbb{E}_7, \quad (-6, -3, 2) \cdot (-2, 6, 3) = 12 - 18 + 6 = 0;$$

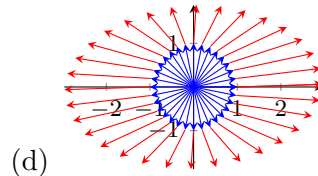
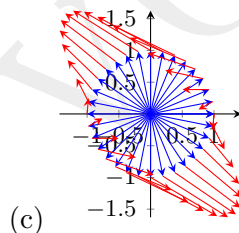
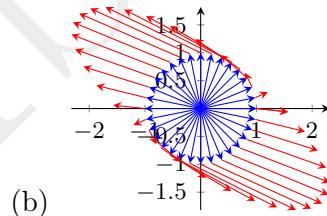
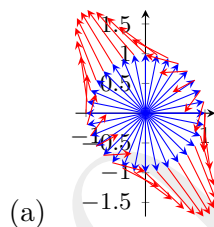
$$\mathbb{E}_7, \mathbb{E}_{-14}, \quad (-2, 6, 3) \cdot (3, -2, 6) = -6 - 12 + 18 = 0;$$

$$\mathbb{E}_{-14}, \mathbb{E}_0, \quad (3, -2, 6) \cdot (-6, -3, 2) = -18 + 6 + 12 = 0.$$



**Theorem 4.2.11.** *Let  $A$  be a real symmetric matrix, then for every two distinct eigenvalues of  $A$ , any corresponding two eigenvectors are orthogonal.*

**Example 4.2.12.** The plots below shows (unit) vectors  $\mathbf{x}$  (blue), and for some matrix  $A$  (different for different plots) the corresponding vectors  $A\mathbf{x}$  (red) adjoined. By estimating eigenvectors determine which cases *cannot* be the plot of a real symmetric matrix.





**Example 4.2.13.** By hand find eigenvectors corresponding to the two distinct eigenvalues of the following matrices. Confirm that symmetric matrix  $A$  has orthogonal eigenvectors, and that non-symmetric matrix  $B$  does not:

$$A = \begin{bmatrix} 1 & \frac{3}{2} \\ \frac{3}{2} & -3 \end{bmatrix}; \quad B = \begin{bmatrix} 0 & -3 \\ -2 & 1 \end{bmatrix}.$$



**Example 4.2.14.** Use MATLAB/Octave to compute eigenvectors of the following matrices. Confirm the eigenvectors are orthogonal for a symmetric matrix.

$$(a) \begin{bmatrix} 0 & 3 & 2 & -1 \\ 0 & 3 & 0 & 0 \\ 3 & 0 & -1 & -1 \\ -3 & 1 & 3 & 0 \end{bmatrix} \quad (b) \begin{bmatrix} -6 & 0 & 1 & 1 \\ 0 & 0 & 2 & 2 \\ 1 & 2 & 2 & -1 \\ 1 & 2 & -1 & -1 \end{bmatrix}$$

■

Recall that to find eigenvalues by hand for  $2 \times 2$  or  $3 \times 3$  matrices we solve a quadratic or cubic characteristic equation, respectively. Thus we find at most two or three eigenvalues, respectively. Further, when we ask MATLAB/Octave to compute eigenvalues of an  $n \times n$  matrix, it always returns  $n$  eigenvalues in an  $n \times n$  diagonal matrix.

**Theorem 4.2.15.** *Every  $n \times n$  real symmetric matrix  $A$  has at most  $n$  distinct eigenvalues.*

The previous theorem establishes there are at most  $n$  distinct eigenvalues (here for symmetric matrices, but [Theorem 7.1.1](#) establishes it is true for general matrices). Now we establish that typically there exist  $n$  distinct eigenvalues of an  $n \times n$  matrix—here symmetric.

[Example 4.0.1](#) started this chapter by observing that in an SVD of a *symmetric* matrix,  $A = USV^T$ , the columns of  $U$  appear to be (almost) always plus/minus the corresponding columns of  $V$ . Exceptions possibly arise in the degenerate cases when two or more singular values are identical. We now prove this close relation between  $U$  and  $V$  in all non-degenerate cases.

**Theorem 4.2.16.** *Let  $A$  be an  $n \times n$  real symmetric matrix with SVD  $A = USV^T$ . If all the singular values are distinct or zero,  $\sigma_1 > \cdots > \sigma_r > \sigma_{r+1} = \cdots = \sigma_n = 0$ , then  $\mathbf{v}_j$  is an eigenvector of  $A$  corresponding to an eigenvalue of either  $\lambda_j = +\sigma_j$  or  $\lambda_j = -\sigma_j$  (not both).*



If non-zero singular values are duplicated, then one can always choose an SVD so the result of this theorem still holds. However, the proof is too involved to give here.

This proof modifies parts of the proof of the SVD [Theorem 3.3.6](#) to the specific case of a symmetric matrix.

Recall that for every real matrix  $A$  an SVD is  $A = USV^T$ . But specifically for symmetric  $A$ , the proof of the previous [Theorem 4.2.16](#) identified that the columns of  $US$ ,  $\sigma_j \mathbf{u}_j$ , are generally the same as  $\lambda_j \mathbf{v}_j$  and hence are the columns of  $VD$  where  $D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ . In which case the SVD becomes  $A = VDV^T$ . This form of an SVD is intimately connected to the following definition.

**Definition 4.2.17.** *A real square matrix  $A$  is **orthogonally diagonalisable** if there exists an orthogonal matrix  $V$  and a diagonal matrix  $D$  such that  $V^T AV = D$ , equivalently  $AV = VD$ , equivalently  $A = VDV^T$  is a factorisation of  $A$ .*

The equivalences in this definition arise immediately from the orthogonality of matrix  $V$  ([Definition 3.2.43](#)): pre-multiply  $V^T AV = D$  by  $V$  gives  $VV^T AV = AV = VD$ ; and so on.

**Example 4.2.18.** (a) Recall from [Example 4.2.13](#) that the symmetric matrix  $A = \begin{bmatrix} 1 & \frac{3}{2} \\ \frac{3}{2} & -3 \end{bmatrix}$  has eigenvalues  $\lambda = -\frac{7}{2}, \frac{3}{2}$  with corresponding orthogonal eigenvectors  $(1, -3)$  and  $(3, 1)$ . Normalise these eigenvectors to unit length as the columns of the orthogonal matrix

$$\begin{aligned} V &= \begin{bmatrix} \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \\ -\frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \end{bmatrix} = \frac{1}{\sqrt{10}} \begin{bmatrix} 1 & 3 \\ -3 & 1 \end{bmatrix} \quad \text{then} \\ V^T A V &= \frac{1}{\sqrt{10}} \begin{bmatrix} 1 & -3 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & \frac{3}{2} \\ \frac{3}{2} & -3 \end{bmatrix} \frac{1}{\sqrt{10}} \begin{bmatrix} 1 & 3 \\ -3 & 1 \end{bmatrix} \\ &= \frac{1}{10} \begin{bmatrix} -\frac{7}{2} & \frac{21}{2} \\ \frac{9}{2} & \frac{3}{2} \end{bmatrix} \begin{bmatrix} 1 & 3 \\ -3 & 1 \end{bmatrix} \\ &= \frac{1}{10} \begin{bmatrix} -35 & 0 \\ 0 & 15 \end{bmatrix} = \begin{bmatrix} -\frac{7}{2} & 0 \\ 0 & \frac{3}{2} \end{bmatrix}. \end{aligned}$$

Hence this matrix is orthogonally diagonalisable.

(b) Recall from [Example 4.2.14](#) that the symmetric matrix

$$B = \begin{bmatrix} -6 & 0 & 1 & 1 \\ 0 & 0 & 2 & 2 \\ 1 & 2 & 2 & -1 \\ 1 & 2 & -1 & -1 \end{bmatrix}$$

has orthogonal eigenvectors computed by MATLAB/Octave into the orthogonal matrix  $V$ . By additionally computing  $V' * B * V$  we get the following diagonal result (2 d.p.)

```
ans =  
-6.45    0.00    0.00    0.00  
 0.00   -3.00    0.00   -0.00  
 0.00    0.00    1.11   -0.00  
-0.00   -0.00   -0.00    3.34
```

and see that this matrix  $B$  is orthogonally diagonalisable.



These examples of orthogonal diagonalisation invoke symmetric



matrices. Also, the connection between an SVD and orthogonal matrices was previously discussed only for symmetric matrices. The next theorem establishes that all real symmetric matrices are orthogonally diagonalisable, and vice versa. That is, eigenvectors of a matrix form an orthogonal set if and only if the matrix is symmetric.

**Theorem 4.2.19** (spectral). *For every real square matrix  $A$ , matrix  $A$  is symmetric iff it is orthogonally diagonalisable.*

### 4.2.3 Change orthonormal basis to classify quadratics

The following preliminary example illustrates the important principle, applicable throughout mathematics, that we often either choose or change to a coordinate system in which the mathematical algebra is simplest.

This optional subsection has many uses—although it is not an application itself as it does not involve real data.

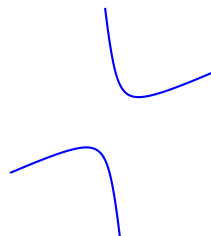
**Example 4.2.20** (choose useful coordinates).

Consider the following two quadratic curves. For each curve draw a coordinate system in which the algebraic description of the curve would be most straightforward.

(a) Ellipse



(b) Hyperbola



Now let's proceed to see how to implement in algebra this geometric idea of choosing good coordinates to fit a given physical curve.

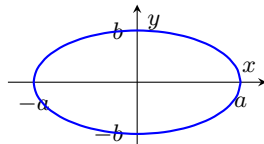
## Graph quadratic equations

**Example 4.2.20** illustrated an ellipse and a hyperbola. These curves are examples of the so-called **conic sections** which arise as solutions of the quadratic equation in two variables, say  $x$  and  $y$ ,

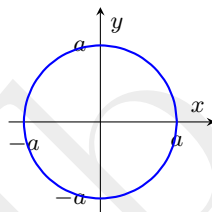
$$ax^2 + bxy + cy^2 + dx + ey + f = 0 \quad (4.2)$$

(where  $a, b, c$  cannot all be zero). As invoked in the example, the canonical simplest algebraic form of such curves are the following. The challenge of this subsection is to choose good new coordinates so that a given quadratic equation (4.2) becomes one of these recognised canonical forms.

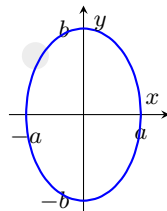
Ellipse or circle :  $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$



- ellipse  $a > b$



- the circle  $a = b$

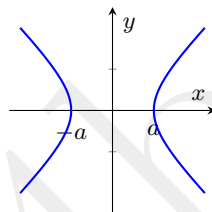


- ellipse  $a < b$

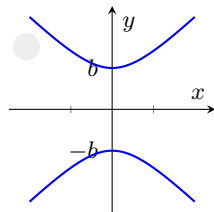
Hyperbola :  $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$  or  $-\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$



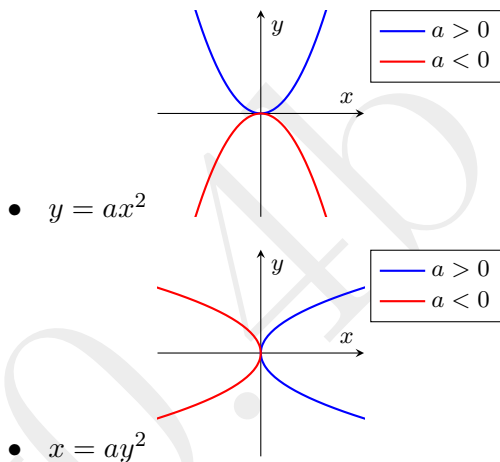
- $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$



- $-\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$



Parabola :  $y = ax^2$  or  $x = ay^2$



[Example 4.2.20](#) implicitly has two steps: first, we decide upon an orientation for the coordinate axes; second, we decide that the coordinate system should be ‘centred’ in the picture. Algebra follows the same two steps.

**Example 4.2.21** (centre coordinates). By shifting coordinates, identify the conic section whose equation is

$$2x^2 + y^2 - 4x + 4y + 2 = 0.$$



**Example 4.2.22** (rotate coordinates). By rotating the coordinate system, identify the conic section whose equation is

$$x^2 + 3xy - 3y^2 - \frac{1}{2} = 0.$$

(There are no terms linear in  $x$  and  $y$  so we do not shift coordinates.)



**Example 4.2.23.** Identify the conic section whose equation is

$$x^2 - xy + y^2 + \frac{5}{2\sqrt{2}}x - \frac{7}{2\sqrt{2}}y + \frac{1}{8} = 0.$$



## Simplify quadratic forms

To understand the response and strength of built structures like bridges, buildings and cars, engineers need to analyse the dynamics of energy distribution in the structure. The potential energy in such structures is expressed and analysed as the following quadratic form. Such quadratic forms are also important in distinguishing maxima from minima in economic optimisation.

**Definition 4.2.24.** A **quadratic form** in variables  $\mathbf{x} \in \mathbb{R}^n$  is a function  $q : \mathbb{R}^n \rightarrow \mathbb{R}$  that may be written as  $q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$  for some real symmetric  $n \times n$  matrix  $A$ .

**Example 4.2.25.** (a) The dot product of a vector with itself is a quadratic form. For all  $\mathbf{x} \in \mathbb{R}^n$  consider

$$\mathbf{x} \cdot \mathbf{x} = \mathbf{x}^T \mathbf{x} = \mathbf{x}^T I_n \mathbf{x},$$

which is the quadratic form associated with the identity matrix  $I_n$ .

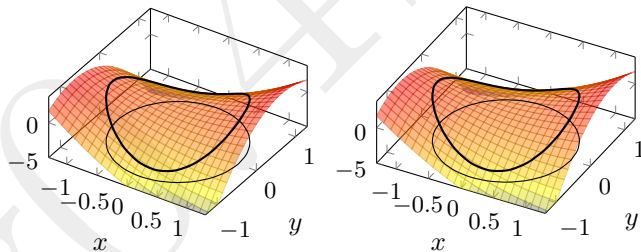
- (b) [Example 4.2.22](#) found the hyperbola satisfying equation  $x^2 + 3xy - 3y^2 - \frac{1}{2} = 0$ . This equation may be written in terms of a quadratic form as  $\mathbf{x}^T A \mathbf{x} - \frac{1}{2} = 0$  for vector  $\mathbf{x} = (x, y)$  and symmetric matrix  $A = \begin{bmatrix} 1 & \frac{3}{2} \\ \frac{3}{2} & -3 \end{bmatrix}$ .
- (c) [Example 4.2.23](#) found the ellipse satisfying the equation  $x^2 - xy + y^2 + \frac{5}{2\sqrt{2}}x - \frac{7}{2\sqrt{2}}y + \frac{1}{8} = 0$  via writing the quadratic part of the equation as  $\mathbf{x}^T A \mathbf{x}$  for vector  $\mathbf{x} = (x, y)$  and symmetric matrix  $A = \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix}$ .



**Theorem 4.2.26** (principal axes theorem). *For every quadratic form, there exists an orthogonal coordinate system that diagonalises the quadratic form. Specifically, for the quadratic form  $\mathbf{x}^T A \mathbf{x}$  find the eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_n$  and orthonormal eigenvectors  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  of symmetric  $A$ , and then in the new coordinate system  $(y_1, y_2, \dots, y_n)$  with unit vectors  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$  the quadratic form has*

the *canonical form*  $\mathbf{x}^T A \mathbf{x} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$ .

**Example 4.2.27.** Consider the quadratic form  $f(x, y) = x^2 + 3xy - 3y^2$ . That is, consider  $f(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$  for  $\mathbf{x} = (x, y)$  and matrix  $A = \begin{bmatrix} 1 & 3/2 \\ 3/2 & -3 \end{bmatrix}$ . The following illustration plots the surface  $f(x, y)$ .



Also plotted in black is the curve of values of  $f(x, y)$  on the unit circle  $x^2 + y^2 = 1$  (also shown); that is,  $f(\mathbf{x})$  for unit vectors  $\mathbf{x}$ . Find the maxima and minima of  $f$  on this unit circle (for unit vectors  $\mathbf{x}$ ). Relate to the eigenvalues of [Example 4.2.13](#). ■

**Theorem 4.2.28.** *Let  $A$  be an  $n \times n$  symmetric matrix with eigenvalues  $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$  (sorted). Then for all unit vectors  $\mathbf{x} \in \mathbb{R}^n$  (that is,  $|\mathbf{x}| = 1$ ), the quadratic form  $\mathbf{x}^T A \mathbf{x}$  has the following properties:*

- (a)  $\lambda_1 \leq \mathbf{x}^T A \mathbf{x} \leq \lambda_n$ ;
- (b) the minimum of  $\mathbf{x}^T A \mathbf{x}$  is  $\lambda_1$ , and occurs when  $\mathbf{x}$  is a unit eigenvector corresponding to  $\lambda_1$ ;
- (c) the maximum of  $\mathbf{x}^T A \mathbf{x}$  is  $\lambda_n$ , and occurs when  $\mathbf{x}$  is a unit eigenvector corresponding to  $\lambda_n$ .

**Activity 4.2.29.** Recall [Example 4.1.27](#) found that the  $3 \times 3$  symmetric matrix

$$A = \begin{bmatrix} -2 & 0 & -6 \\ 0 & 4 & 6 \\ -6 & 6 & -9 \end{bmatrix}$$

has eigenvalues 7, 0 and  $-14$ .

- What is the maximum of the quadratic form  $\mathbf{x}^T A \mathbf{x}$  over unit vectors  $\mathbf{x}$ ?

(a) 7

(b)  $-14$

(c) 0

(d) 14

- Further, what is the minimum of the quadratic form  $\mathbf{x}^T A \mathbf{x}$  over unit vectors  $\mathbf{x}$ ?





---

## 5 Approximate matrices

---

### Chapter Contents

5.1	Measure changes to matrices . . . . .	546
5.1.1	Compress images optimally . . . . .	547
5.1.2	Relate matrix changes to the SVD . . . . .	555
5.1.3	Principal component analysis . . . . .	568
5.2	Regularise linear equations . . . . .	600
5.2.1	The SVD illuminates regularisation . . . . .	603
5.2.2	Tikhonov regularisation . . . . .	611

This chapter could be studied any time after [Chapter 3](#) to help the transition to more abstract linear algebra. Useful as spaced revision of the SVD, rank, orthogonality, and so on.

This chapter develops how concepts associated with length and

distance not only apply to vectors but also apply to matrices. More advanced courses on Linear Algebra place these in a unifying framework that also encompasses much you see both in solving differential equations (and integral equations) and in problems involving complex numbers (such as those in electrical engineering or quantum physics).

## 5.1 Measure changes to matrices

### Section Contents

5.1.1	Compress images optimally . . . . .	547
5.1.2	Relate matrix changes to the SVD . . . . .	555
5.1.3	Principal component analysis . . . . .	568
	Application to latent semantic indexing . . .	579

### 5.1.1 Compress images optimally

Photographs and other images take a lot of storage. Reducing the amount of storage for an image is essential, both for storage and for transmission. The well-known jpeg format for compressing photographs is incredibly useful: the SVD provides a related effective method of compression.

These SVD methods find approximate matrices of the images with the matrices having of various ranks. Recall that a matrix of rank  $k$  ([Definition 3.3.19](#)) means the matrix has precisely  $k$  non-zero singular values, that is, an  $m \times n$  matrix

$$A = USV^T$$

$$= \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_k & \cdots & \mathbf{u}_m \end{bmatrix} \begin{bmatrix} \sigma_1 & \cdots & 0 & & \\ \vdots & \ddots & \vdots & & \\ 0 & \cdots & \sigma_k & & \\ & & & O_{(m-k) \times k} & O_{(m-k) \times (n-k)} \end{bmatrix} \begin{bmatrix} \\ \\ \\ \\ \end{bmatrix} V^T$$

(then multiplying the form of the first two matrices)

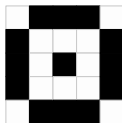
$$= \begin{bmatrix} \sigma_1 \mathbf{u}_1 & \cdots & \sigma_k \mathbf{u}_k & O_{m \times (n-k)} \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_k^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix}$$

(then multiplying the form of these two matrices)

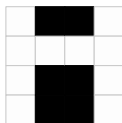
$$= \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T.$$

This last form constructs matrix  $A$ . Further, when the rank  $k$  is low compared to size  $m$  and  $n$ , this last form *has relatively few components*.

**Example 5.1.1.** Invent and write down a rank three representation of the following  $5 \times 5$  ‘bulls eye’ matrix (illustrated in the margin)



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



**Activity 5.1.2.** Which pair of vectors gives a rank one representation,  $\mathbf{uv}^T$  of the matrix

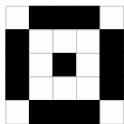
$$\begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix}?$$

- (a)  $\mathbf{u} = (0, 1, 1, 0)$ ,  $\mathbf{v} = (1, 1, 0, 1)$
- (b)  $\mathbf{u} = (0, 1, 1, 0)$ ,  $\mathbf{v} = (1, 0, 1, 1)$
- (c)  $\mathbf{u} = (1, 1, 0, 1)$ ,  $\mathbf{v} = (0, 1, 1, 0)$
- (d)  $\mathbf{u} = (1, 0, 1, 1)$ ,  $\mathbf{v} = (0, 1, 1, 0)$

**Procedure 5.1.3** (approximate images). *Given an image stored as scalars in an  $m \times n$  matrix  $A$ .*

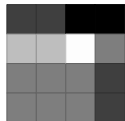
1. *Compute an SVD  $A = USV^T$  with  $[U, S, V] = \text{svd}(A)$ .*
2. *Choose a desired rank  $k$  based upon the singular values (Theorem 5.1.16): typically there will be  $k$  ‘large’ singular values and the rest are ‘small’.*
3. *Then the ‘best’ rank  $k$  approximation to the image matrix  $A$  is (using the subscript  $k$  on the matrix name to denote the rank  $k$  approximation)*

$$\begin{aligned} A_k &:= \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T \\ &= U(:, 1:k) * S(1:k, 1:k) * V(:, 1:k)' \end{aligned}$$



**Example 5.1.4.** Use [Procedure 5.1.3](#) to find the ‘best’ rank two matrix, and also the ‘best’ rank three matrix, to approximate the ‘bulls eye’ image matrix (illustrated in the margin)

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



**Activity 5.1.5.** A given image, shown in the margin, has matrix with SVD (2 d.p.)

$U =$

$$\begin{array}{cccc} -0.72 & 0.48 & 0.50 & -0.00 \\ -0.22 & -0.84 & 0.50 & -0.00 \\ -0.47 & -0.18 & -0.50 & -0.71 \\ -0.47 & -0.18 & -0.50 & 0.71 \end{array}$$



$$S = \begin{bmatrix} 2.45 & 0 & 0 & 0 \\ 0 & 0.37 & 0 & 0 \\ 0 & 0 & 0.00 & 0 \\ 0 & 0 & 0 & 0.00 \end{bmatrix}$$

$$V = \begin{bmatrix} -0.43 & -0.07 & 0.87 & -0.24 \\ -0.43 & -0.07 & -0.44 & -0.78 \\ -0.48 & 0.83 & -0.11 & 0.26 \\ -0.62 & -0.55 & -0.21 & 0.51 \end{bmatrix}$$

What rank representation will exactly reproduce the matrix/image?

- (a) 1                      (b) 3                      (c) 2                      (d) 4



<http://eulerarchive.maa.org/portraits/portraits.html> [Sep 2015]



**Example 5.1.6.** In the margin is a  $326 \times 277$  greyscale image of Euler at 30 years old. As such the image is coded as 90 302 scalar numbers. Let's find a good approximation to the image that uses much fewer numbers, and hence takes less storage. That is, we effectively compress the image for storage or transmission.



Table 5.1: As well as the MATLAB/Octave commands and operations listed in Tables 1.2, 2.3, 3.1, 3.2, 3.3, and 3.7 we may invoke these functions.

- **norm(A)** computes the matrix norm of Definition 5.1.7, namely the largest singular value of the matrix  $A$ .

Also recall that (Table 1.2) **norm(v)** for a vector  $\mathbf{v}$  computes the length  $\sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$ .

- **scatter(x,y,[],c)** draws a 2D scatter plot of points with coordinates in vectors  $\mathbf{x}$  and  $\mathbf{y}$ , each point with a colour determined by the corresponding entry of vector  $\mathbf{c}$ .

Similarly for **scatter3(x,y,z,[],c)** but in 3D.

- **[U,S,V]=svds(A,k)** computes the  $k$  largest singular values of the matrix  $A$  in the diagonal of  $k \times k$  matrix  $S$ , and the  $k$  columns of  $U$  and  $V$  are the corresponding singular vectors.
- **imread('filename')** typically reads an image from a file into an  $m \times n \times 3$  array of red-green-blue values. The values are all 'integers' in the range  $[0, 255]$ .
- **csvread('filename')** reads data from a file into a matrix. When each of the  $m$  lines in the file is  $n$  numbers separated by commas, then the result is an  $m \times n$  matrix.
- **mean(A)** of an  $m \times n$  array computes the  $n$  elements in the row vector of averages (the arithmetic mean) over each column

### 5.1.2 Relate matrix changes to the SVD

We need to define what ‘best’ means in the approximation [Procedure 5.1.3](#) and then show the procedure achieves this best. We need a measure of the magnitude of matrices and distances between matrices.

In linear algebra we use the double vertical bars,  $\|\cdot\|$ , to denote the magnitude of a matrix in order to avoid a notational clash with the well-established use of  $|\cdot|$  for the determinant of a matrix ([Chapter 6](#)).

**Definition 5.1.7.** *Let  $A$  be an  $m \times n$  matrix. Define the **matrix norm** (sometimes called the spectral norm)*

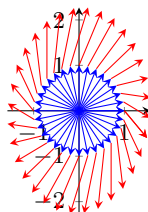
$$\|A\| := \max_{|\mathbf{x}|=1} |A\mathbf{x}|, \quad \text{equivalently } \|A\| = \sigma_1 \quad (5.1)$$

*the largest singular value of the matrix  $A$ .*

The equivalence, that  $\max_{|\mathbf{x}|=1} |A\mathbf{x}| = \sigma_1$ , is due to the definition of the largest singular value in the proof of the existence of an SVD ([Subsection 3.3.3](#)).

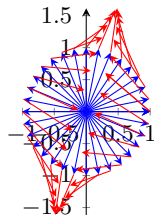
**Example 5.1.8.** The two following  $2 \times 2$  matrices have the product  $A\mathbf{x}$  plotted (red), adjoined to  $\mathbf{x}$  (blue), for a complete range of unit vectors  $\mathbf{x}$  (as in [Section 4.1](#) for eigenvectors). From [Definition 5.1.7](#), the norm of the matrix  $A$  is then the length of the longest such plotted  $A\mathbf{x}$ . As such, this norm is a measure of the magnitude of the matrix. For each matrix, use the plot to roughly estimate their norm.

The MATLAB function `eigshow(A)` provides an interactive alternative to such static views.



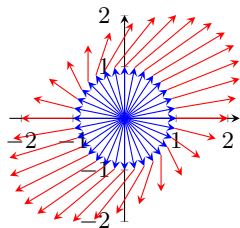
$$(a) \quad A = \begin{bmatrix} 0.5 & 0.5 \\ -0.6 & 1.2 \end{bmatrix}$$

$$(b) \quad B = \begin{bmatrix} -0.7 & 0.4 \\ 0.6 & 0.5 \end{bmatrix}$$



**Example 5.1.9.**

Consider the  $2 \times 2$  matrix  $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ . Algebraically explore products  $A\mathbf{x}$  for unit vectors  $\mathbf{x}$ , as illustrated in the margin, and then find the matrix norm  $\|A\|$ .

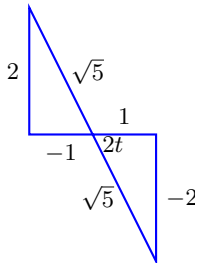


- The standard unit vector  $\mathbf{e}_2 = (0, 1)$  has  $|\mathbf{e}_2| = 1$  and  $A\mathbf{e}_2 = (1, 1)$  has length  $|A\mathbf{e}_2| = \sqrt{2}$ . Since the matrix norm is the maximum of all possible  $|A\mathbf{x}|$ , so  $\|A\| \geq |A\mathbf{e}_2| = \sqrt{2} \approx 1.41$ .
- Another unit vector is  $\mathbf{x} = (\frac{3}{5}, \frac{4}{5})$ . Here  $A\mathbf{x} = (\frac{7}{5}, \frac{4}{5})$  has length  $\sqrt{49 + 16}/5 = \sqrt{65}/5 \approx 1.61$ . Hence the matrix norm  $\|A\| \geq |A\mathbf{x}| \approx 1.61$ .
- To systematically find the norm, recall all unit vectors in 2D are of the form  $\mathbf{x} = (\cos t, \sin t)$ . Then

$$\begin{aligned}
 |A\mathbf{x}|^2 &= |(\cos t + \sin t, \sin t)|^2 \\
 &= (\cos t + \sin t)^2 + \sin^2 t \\
 &= \cos^2 t + 2 \cos t \sin t + \sin^2 t + \sin^2 t \\
 &= \frac{3}{2} + \sin 2t - \frac{1}{2} \cos 2t.
 \end{aligned}$$

This length (squared) is maximised (and minimised) for some  $t$  determined by calculus. Differentiating with respect to  $t$  leads to

$$\frac{d|A\mathbf{x}|^2}{dt} = 2 \cos 2t + \sin 2t = 0 \quad \text{for stationary points.}$$



Rearranging determines we require  $\tan 2t = -2$ . The marginal right-angle triangles illustrate that these stationary points of  $|A\mathbf{x}|^2$  occur for  $\sin 2t = \mp 2/\sqrt{5}$  and correspondingly  $\cos 2t = \pm 1/\sqrt{5}$  (one gives a minimum and one gives the desired maximum). Substituting these two cases gives

$$\begin{aligned} |A\mathbf{x}|^2 &= \frac{3}{2} + \sin 2t - \frac{1}{2} \cos 2t \\ &= \frac{3}{2} \mp \frac{2}{\sqrt{5}} \mp \frac{1}{2} \frac{1}{\sqrt{5}} \\ &= \frac{1}{2} (3 \mp \sqrt{5}) \\ &= \left( \frac{1 \mp \sqrt{5}}{2} \right)^2. \end{aligned}$$

The plus alternative is the larger so gives the maximum, hence

$$\|A\| = \max_{|\mathbf{x}|=1} |A\mathbf{x}| = \frac{1 + \sqrt{5}}{2} = 1.6180.$$

- Confirm with MATLAB/Octave via `svd([1 1;0 1])` which gives the singular values  $\sigma_1 = 1.6180$  and  $\sigma_2 = 0.6180$ . Hence confirming the norm  $\|A\| = \sigma_1 = 1.6180$ .

Alternatively, see [Table 5.1](#), execute `norm([1 1;0 1])` to compute the norm  $\|A\| = 1.6180$ .

**Activity 5.1.10.** A given  $3 \times 3$  matrix  $A$  has the following products

$$A \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad A \begin{bmatrix} -1/3 \\ -2/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 1/3 \\ -7/3 \end{bmatrix}, \quad A \begin{bmatrix} 1 \\ -2 \\ -2 \end{bmatrix} = \begin{bmatrix} 11 \\ 3 \\ 3 \end{bmatrix}.$$

Which of the following is the ‘best’ statement about the norm of matrix  $A$  (best in the sense of giving the largest valid lower bound)?

(a)  $\|A\| \geq 3.9$

(b)  $\|A\| \geq 2.3$

(c)  $\|A\| \geq 1.7$

(d)  $\|A\| \geq 11.7$



**Example 5.1.11.** MATLAB/Octave readily computes the matrix norm either via an SVD or using the `norm()` function directly (Table 5.1). Compute the norm of the following matrices.

$$(a) \ A = \begin{bmatrix} 0.1 & -1.3 & -0.4 & -0.1 & -0.6 \\ 1.9 & 2.4 & -1.8 & 0.2 & 0.8 \\ -0.2 & -0.5 & -0.7 & -2.5 & 1.1 \\ -1.8 & 0.2 & 1.1 & -1.2 & 1.0 \\ -0.0 & 1.2 & 1.1 & -0.1 & 1.7 \end{bmatrix}$$

$$(b) \ B = \begin{bmatrix} 0 & -2 & -1 & -4 & -5 & 0 \\ 2 & 0 & 1 & -2 & -6 & -2 \\ -2 & 0 & 4 & 2 & 3 & -3 \\ 1 & 2 & -4 & 2 & 1 & 3 \end{bmatrix}$$



The Definition 5.1.7 of the magnitude/norm of a matrix may appear a little strange. But, in addition to some marvellously useful properties, it nonetheless has all the familiar properties of a magnitude/length. Recall from Chapter 1 that for vectors:

- $|v| = 0$  if and only if  $v = \mathbf{0}$  (Theorem 1.1.13);
- $|u \pm v| \leq |u| + |v|$  (the triangle inequality of Theorem 1.3.17);
- $|tv| = |t| \cdot |v|$  (Theorem 1.3.17).

Analogous properties hold for the matrix norm as established in the next theorem.

**Theorem 5.1.12** (norm properties). *For every  $m \times n$  real matrix  $A$ :*

- (a)  $\|A\| = 0$  if and only if  $A = O_{m \times n}$ ;
- (b)  $\|I_n\| = 1$ ;
- (c)  $\|A \pm B\| \leq \|A\| + \|B\|$ , for every  $m \times n$  matrix  $B$ , is like a triangle inequality (Theorem 1.3.17c);
- (d)  $\|tA\| = |t|\|A\|$ ;
- (e)  $\|A\| = \|A^T\|$ ;
- (f)  $\|Q_m A\| = \|A\| = \|A Q_n\|$  for every  $m \times m$  orthogonal matrix  $Q_m$  and every  $n \times n$  orthogonal matrix  $Q_n$ ;

- (g)  $|A\mathbf{x}| \leq \|A\|\|\mathbf{x}\|$  for all  $\mathbf{x} \in \mathbb{R}^n$ , is like a Cauchy–Schwarz inequality ([Theorem 1.3.17b](#)), as is the following;
- (h)  $\|AB\| \leq \|A\|\|B\|$  for every  $n \times p$  matrix  $B$ .

Since the matrix norm has the familiar properties of a measure of magnitude, we use the matrix norm to measure the ‘distance’ between matrices.

**Example 5.1.13.** (a) Use the matrix norm to estimate the ‘distance’ between matrices

$$B = \begin{bmatrix} -0.7 & 0.4 \\ 0.6 & 0.5 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} -0.2 & 0.9 \\ 0 & 1.7 \end{bmatrix}.$$

(b) Recall from [Example 3.3.2](#) that the matrix

$$A = \begin{bmatrix} 10 & 2 \\ 5 & 11 \end{bmatrix}$$

has an SVD of

$$USV^T = \begin{bmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{bmatrix} \begin{bmatrix} 10\sqrt{2} & 0 \\ 0 & 5\sqrt{2} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}^T.$$

i. Find  $\|A - B\|$  for the rank one matrix

$$B = \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T = 5\sqrt{2} \begin{bmatrix} -\frac{4}{5} \\ \frac{3}{5} \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 4 & -4 \\ -3 & 3 \end{bmatrix}.$$

ii. Find  $\|A - A_1\|$  for the rank one matrix

$$A_1 = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T = 10\sqrt{2} \begin{bmatrix} \frac{3}{5} \\ \frac{4}{5} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 6 & 6 \\ 8 & 8 \end{bmatrix}.$$

Out of these two matrices,  $A_1$  and  $B$ , the matrix  $A_1$  is ‘closer’ to  $A$  as  $\|A - A_1\| = 5\sqrt{2} < 10\sqrt{2} = \|A - B\|$ .



**Activity 5.1.14.** Which of the following matrices is *not* a distance one from the matrix  $F = \begin{bmatrix} 9 & -1 \\ 1 & 5 \end{bmatrix}$ ?

(a)  $\begin{bmatrix} 8 & -2 \\ 2 & 4 \end{bmatrix}$

(b)  $\begin{bmatrix} 10 & -1 \\ 1 & 6 \end{bmatrix}$

(c)  $\begin{bmatrix} 8 & -1 \\ 1 & 5 \end{bmatrix}$

(d)  $\begin{bmatrix} 9 & -1 \\ 1 & 6 \end{bmatrix}$



**Example 5.1.15.** From [Example 5.1.4](#), recall the ‘bulls eye’ matrix

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

and its rank two and three approximations  $A_2$  and  $A_3$ . Find

$$\|A - A_2\| \text{ and } \|A - A_3\|.$$



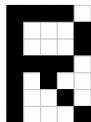
**Theorem 5.1.16** (Eckart–Young). *Let  $A$  be an  $m \times n$  matrix of rank  $r$  with SVD  $A = USV^T$ . Then for every  $k < r$  the matrix*

$$A_k := US_kV^T = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T \quad (5.2)$$

*where  $S_k := \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_k, 0, \dots, 0)$ , is a closest rank  $k$  matrix approximating  $A$ , in the matrix norm. The distance between  $A$  and  $A_k$  is  $\|A - A_k\| = \sigma_{k+1}$ .*

That is, obtain a closest rank  $k$  matrix  $A_k$  by ‘setting’ the singular values  $\sigma_{k+1} = \cdots = \sigma_r = 0$  from an SVD for  $A$ .

**Example 5.1.17** (the letter R). In displays with low resolution, letters and numbers are displayed with noticeable pixel patterns: for example, the letter R is pixellated in the margin. Let’s see how such pixel patterns are best approximated by matrices of different ranks. (This example is illustrative: it is not a practical image compression since the required singular vectors are more complicated than a



small-sized pattern of pixels.)

**Activity 5.1.18.** A given image has singular values 12.74, 8.38, 3.06, 1.96, 1.08, .... What rank approximation has an error of just a little less than 25%?

- (a) 1                      (b) 4                      (c) 3                      (d) 2

**Example 5.1.19.** Recall [Example 5.1.6](#) approximated the image of Euler (1737) with various rank  $k$  approximates from an SVD of the image. Let the image be denoted by matrix  $A$ . From ?? the largest singular value of the image is  $\|A\| = \sigma_1 \approx 40\,000$ .

- From [Theorem 5.1.16](#), the rank 3 approximation in ?? is a distance  $\|A - A_3\| = \sigma_4 \approx 5\,000$  (from ??) away from the image. That is, image  $A_3$  has a relative error roughly  $5\,000/40\,000 = 1/8 \approx 12\%$ .

- From [Theorem 5.1.16](#), the rank 10 approximation in ?? is a distance  $\|A - A_{10}\| = \sigma_{11} \approx 5\,000$  (from ??) away from the image. That is, image  $A_{10}$  has a relative error roughly  $2\,000/40\,000 = 1/20 = 5\%$ .
- From [Theorem 5.1.16](#), the rank 30 approximation in ?? is a distance  $\|A - A_{30}\| = \sigma_{31} \approx 800$  (from ??) away from the image. That is, image  $A_{30}$  has a relative error roughly  $800/40\,000 = 1/50 = 2\%$ .

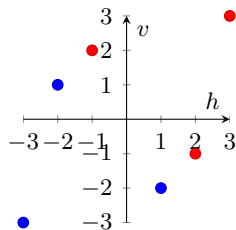




### 5.1.3 Principal component analysis

In its ‘best’ approximation property, [Theorem 5.1.16](#) establishes the effectiveness of an SVD in image compression. Scientists and engineers also use this result for so-called data reduction: often using just a rank two (or three) ‘best’ approximation to high dimensional data, one then plots 2D (or 3D) graphics. Such an approach is often termed a principal component analysis (PCA).

The technique introduced here is so useful that more-or-less the same approach has been invented independently in many fields and so much the same technique has alternative names such as the Karhunen–Loève transform, proper orthogonal decomposition, empirical orthogonal functions, and the Hotelling transform.



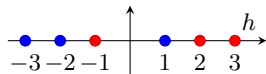
**Example 5.1.20** (toy items). Suppose you are given data about six items, three blue and three red. Suppose each item has two measured properties/attributes called  $h$  and  $v$  as in the following table:

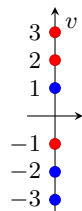
$h$	$v$	colour
-3	-3	blue
-2	1	blue
1	-2	blue
-1	2	red
2	-1	red
3	3	red

The item properties/attributes are the points  $(h, v)$  in 2D as illustrated in the margin. But humans always prefer simple one dimensional summaries: we do it all the time when we rank sport teams, schools, web pages, and so on.

Challenge: is there a one dimensional summary of these six item's data that clearly separates the blue from the red? Using just one of the attributes  $h$  or  $v$  on their own would not suffice:

- using  $h$  alone leads to a 1D view where the red and the blue





are intermingled as shown in the margin;

- similarly, using  $v$  alone leads to a 1D view where the red and the blue are intermingled as shown in the margin.



Although this [Example 5.1.20](#) is just a toy to illustrate concepts, the above steps generalise straightforwardly to be immensely useful on vastly bigger and more challenging data. The next example takes the next step in complexity by introducing how to automatically find a good 2D view of some data in 4D.

**Example 5.1.21** (Iris flower data set). [Table 5.2](#) list part of Edgar Anderson's data on the length and widths of sepals and petals of Iris flowers. There are three species of Irises in the data (Setosa, Versicolor, Virginia). The data is 4D: each instance of thirty Iris flowers is characterised by the four measurements of sepals and petals. Our challenge is to plot a 2D picture of this data in such a way that separates the flowers as best as possible. For high-D data (although 4D is not really that high), simply plotting

Table 5.2: part of Edgar Anderson's Iris data, lengths in centimetres (cm). The measurements come from the flowers of ten each of three different species of Iris.

Sepal length	Sepal width	Petal length	Petal width	Species
4.9	3.0	1.4	0.2	Setosa
4.6	3.4	1.4	0.3	
4.8	3.4	1.6	0.2	
5.4	3.9	1.3	0.4	
5.1	3.7	1.5	0.4	
5.0	3.4	1.6	0.4	
5.4	3.4	1.5	0.4	
5.5	3.5	1.3	0.2	
4.5	2.3	1.3	0.3	
5.1	3.8	1.6	0.2	
6.4	3.2	4.5	1.5	Versicolor
6.3	3.3	4.7	1.6	
5.9	3.0	4.2	1.5	
5.6	3.0	4.5	1.5	
6.1	2.8	4.0	1.3	
6.8	2.8	4.8	1.4	
6.7	3.0	4.7	1.3	

one characteristic against another is rarely useful. For example, [Figure 5.1](#) plots the attributes of sepal widths versus sepal lengths: the plot shows the three species being intermingled together rather than reasonably separated. Our aim is to instead plot ?? which successfully separates the three species.



**Transpose the usual mathematical convention** Perhaps you noticed that the previous [Example 5.1.21](#) flips our usual mathematical convention that vectors are column vectors. The example uses row vectors of the four attributes of each flower: [Table 5.2](#) lists that the first Iris Setosa flower has a row vector of attributes  $[4.9 \ 3.0 \ 1.4 \ 0.2]$  (cm) corresponding to the sepal length and width, and the petal length and width, respectively. Similarly, the last Virginia Iris flower has row vector of attributes of  $[46.3 \ 2.5 \ 5.0 \ 1.9]$  (cm), and the mean vector is the row vector  $[5.81 \ 3.09 \ 3.69 \ 1.22]$  (cm). The reason for this mathematical transposition is that throughout science and engineering, data

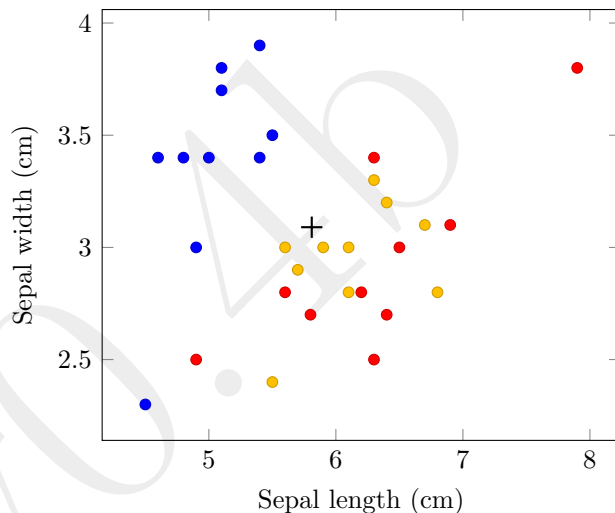


Figure 5.1: scatter plot of sepal widths versus lengths for Edgar Anderson's Iris data of [Table 5.2](#): blue, Setosa; brown, Versicolor; red, Virginia. The black “+” marks the mean sepal width and length.

results are most often presented as rows of different instances of flowers, animals, clients or experiments: each row contains the list of characteristic measured or derived properties/attributes. Table 5.2 has this most common structure. Thus in this sort of application, the mathematics we do needs to reflect this most common structure. Hence many vectors in this subsection appear as row vectors. When they do appear, they are called row vectors: the term vector on its own still means a column vector.

**Definition 5.1.22** (principal components). *Given a  $m \times n$  data matrix  $A$  (usually with zero mean when averaged over all rows) with SVD  $A = USV^T$ , then the  $j$ th column  $\mathbf{v}_j$  of  $V$  is called the  $j$ th **principal vector** and the vector  $\mathbf{x}_j := A\mathbf{v}_j$  is called the  $j$ th **principal components** of the data matrix  $A$ .*

Now what does an SVD tell us for 2D plots of data? We know  $A_2$  is the best rank two approximation to the data matrix  $A$  (Theorem 5.1.16). That is, if we are only to plot two components, those two components are best to come from  $A_2$ . Recall from (5.2) that

$$A_2 = US_2V^T = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T = (\sigma_1 \mathbf{u}_1) \mathbf{v}_1^T + (\sigma_2 \mathbf{u}_2) \mathbf{v}_2^T.$$

That is, in this best rank two approximation of the data, the row vector of attributes of the  $i$ th Iris are the linear combination of row vectors  $(\sigma_1 u_{i1})\mathbf{v}_1^T + (\sigma_2 u_{i2})\mathbf{v}_2^T$ . The vectors  $\mathbf{v}_1$  and  $\mathbf{v}_2$  are orthonormal vectors so we treat them as the horizontal and vertical unit vectors of a scatter plot. That is,  $x_i = \sigma_1 u_{i1}$  and  $y_i = \sigma_2 u_{i2}$  are horizontal and vertical coordinates of the  $i$ th Iris in the best 2D plot. Consequently, in MATLAB/Octave we draw a scatter plot of the components of vectors  $\mathbf{x} = \sigma_1 \mathbf{u}_1$  and  $\mathbf{y} = \sigma_2 \mathbf{u}_2$  (??).

**Theorem 5.1.23.** *Using the matrix norm to measure ‘best’ (Definition 5.1.7), the best  $k$ -dimensional summary of the  $m \times n$  data matrix  $A$  (usually of zero mean) are the first  $k$  principal components in the directions of the first  $k$  principal vectors.*

**Activity 5.1.24.** A given data matrix from some experiment has singular values 12.76, 10.95, 7.62, 0.95, 0.48,  $\dots$ . How many dimensions should you expect to be needed for a good view of the data?

- (a) 4D                      (b) 1D                      (c) 3D                      (d) 2D





**Example 5.1.25** (wine recognition). From the [Lichman \(2013\)](#) repository download the data file `wine.data` and its description file `wine.names`. The wine data has 178 rows of different wine samples, and 14 columns of attributes of which the first column is the cultivar class number and the remaining 13 columns are the amounts of different chemicals measured in the wine. Question: is there a two-dimensional view of these chemical measurements that largely separates the cultivars?



The previous three examples develop the following procedure for ‘best’ viewing data in low dimensions. However, any additional information about the data or preferred results may modify this procedure.

**Procedure 5.1.26** (principal component analysis). *Consider the case when you have data values consisting of  $n$  attributes for each of  $m$  instances, and the aim is to find a good  $k$ -dimensional summary/view of the data.*

1. Form/enter the  $m \times n$  data matrix  $B$ .
2. Scale the data matrix  $B$  to form  $m \times n$  matrix  $A$ :
  - (a) usually make each column have zero mean by subtracting its mean  $\bar{b}_j$ , algebraically  $\mathbf{a}_j = \mathbf{b}_j - \bar{b}_j$ ;
  - (b) but ensure each column has the same ‘physical dimensions’, often by dividing by the standard deviation  $s_j$  of each column, algebraically  $\mathbf{a}_j = (\mathbf{b}_j - \bar{b}_j)/s_j$ .

Compute in MATLAB/Octave with

```
A=bsxfun(@divide,bsxfun(@minus,B,mean(B)),std(B))
```

3. Economically compute an SVD for the best rank  $k$  approximation to the scaled data matrix with  $[\mathbf{U}, \mathbf{S}, \mathbf{V}] = \text{svds}(\mathbf{A}, k)$ .
4. Then the  $j$ th column of  $\mathbf{V}$  is the  $j$ th principal vector, and the principal components are the entries of the  $m \times k$  matrix  $\mathbf{A} \cdot \mathbf{V}$ .

Courses on multivariate statistics prove that, for every (usually zero mean) data matrix  $A$ , the first  $k$  principal vectors  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$  are orthogonal unit vectors that *maximise the total variance*

in the principal components  $\mathbf{x}_j = A\mathbf{v}_j$ ; that is, that maximise  $|\mathbf{x}_1|^2 + |\mathbf{x}_2|^2 + \cdots + |\mathbf{x}_k|^2$ . Indeed, this maximisation of the variance corresponds closely to the constructive proof of the existence of SVDs (Subsection 3.3.3) which successively maximises  $|A\mathbf{v}|$  subject to  $\mathbf{v}$  being orthonormal to the singular/principal vectors already determined. Consequently, when data is approximated in the space of the first  $k$  principal vectors, then the data is the most spread out it can be in  $k$ -dimensions. When the data is most spread out in  $k$ D, then (roughly) it retains the most information possible in  $k$ D.

## Application to latent semantic indexing

This ability to retrieve relevant information based upon meaning rather than literal term usage is the main motivation for using LSI [latent semantic indexing].

*(Berry et al. 1995, p.579)*

Searching for information based upon word matching results in surprisingly poor retrieval of relevant documents (Berry et al. 1995, §5.5). Instead, the so-called method of latent semantic indexing improves retrieval by replacing individual words with nearness of word vectors derived via the singular value decomposition. This section introduces latent semantic indexing via a very small example.

The Society for Industrial and Applied Mathematics (SIAM) reviews many mathematical books. In 2015 six of those books had the following titles:

1. Introduction to Finite and Spectral Element Methods using MATLAB

2. Iterative Methods for Linear Systems: Theory and Applications
3. Singular Perturbations: Introduction to System Order Reduction Methods with Applications
4. Risk and Portfolio Analysis: Principles and Methods
5. Stochastic Chemical Kinetics: Theory and Mostly Systems Biology Applications
6. Quantum Theory for Mathematicians

Consider the capitalised words. For those words that appear in more than one title, let's form a word vector ([Example 1.1.7](#)) for each title, then use principal components to summarise these six books on a 2D plane. This task is part of what is called latent semantic indexing ([Berry et al. 1995](#)). (We should also count words that are used only once, but this example omits for simplicity.)

Follow the principal component analysis [Procedure 5.1.26](#).

1. First find the set of words that are used more than once.

Ignoring pluralisation, they are: Application, Introduction, Method, System, Theory. The corresponding word vector for each book title is then the following:

- $\mathbf{w}_1 = (0, 1, 1, 0, 0)$  *Introduction to Finite and Spectral Element Methods using MATLAB*
- $\mathbf{w}_2 = (1, 0, 1, 1, 1)$  *Iterative Methods for Linear Systems: Theory and Applications*
- $\mathbf{w}_3 = (1, 1, 1, 1, 0)$  *Singular Perturbations: Introduction to System Order Reduction Methods with Applications*
- $\mathbf{w}_4 = (0, 0, 1, 0, 0)$  *Risk and Portfolio Analysis: Principles and Methods*
- $\mathbf{w}_5 = (1, 0, 0, 1, 1)$  *Stochastic Chemical Kinetics: Theory and Mostly Systems Biology Applications*
- $\mathbf{w}_6 = (0, 0, 0, 0, 1)$  *Quantum Theory for Mathematicians*

2. Second, form the data matrix with  $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_6$  as rows (not columns). We could remove the mean word vector, but

choose not to: here the position of each book title relative to an empty title (the origin) is interesting. There is no need to scale each column as each column has the same ‘physical’ dimensions, namely a word count. The data matrix of word vectors is then

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

3. Third, to compute a representation in the 2D plane, principal components uses, as an orthonormal basis, the singular vectors corresponding to the two largest singular values. So compute the economical SVD with `[U,S,V]=svds(A,2)` giving (2 d.p.)

U = ...

S =

3.14      0

0      1.85



$$V = \begin{bmatrix} +0.52 & -0.20 \\ +0.26 & +0.52 \\ +0.50 & +0.57 \\ +0.52 & -0.20 \\ +0.37 & -0.57 \end{bmatrix}$$

4. Columns of  $V$  are word vectors in the 5D space of counts of Application, Introduction, Method, System, and Theory. The two given columns of  $V = [\mathbf{v}_1 \ \mathbf{v}_2]$  are the two orthonormal principal vectors:
- the first  $\mathbf{v}_1$ , from its largest components, mainly identifies the overall direction of Application, Method and System;
  - whereas the second  $\mathbf{v}_2$ , from its largest positive and negative components, mainly distinguishes Introduction and Method from Theory.

The corresponding principal components are the entries of



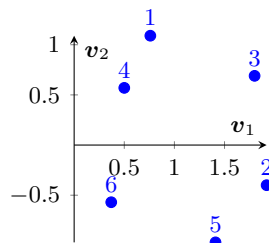
the  $6 \times 2$  matrix

$$AV = \begin{bmatrix} 0.76 & 1.09 \\ 1.92 & -0.40 \\ 1.80 & 0.69 \\ 0.50 & 0.57 \\ 1.41 & -0.97 \\ 0.37 & -0.57 \end{bmatrix} :$$

for each of the six books, the book title has components in the two principal directions given by the corresponding row in this product. We plot the six books on a 2D plane with the MATLAB/Octave command

```
scatter(A*V(:,1),A*V(:,2),[],1:6)
```

to produce a picture like that in the margin. The SVD analysis nicely distributes the six books in this plane.



The above procedure would approximate the original word vector data, formed into a matrix, by the following rank two matrix

(2 d.p.)

$$A_2 = US_2V^T = \begin{bmatrix} 0.18 & 0.77 & 1.01 & 0.18 & -0.33 \\ 1.08 & 0.29 & 0.74 & 1.08 & 0.95 \\ 0.80 & 0.82 & 1.30 & 0.80 & 0.28 \\ 0.15 & 0.43 & 0.58 & 0.15 & -0.14 \\ 0.93 & -0.14 & 0.16 & 0.93 & 1.08 \\ 0.31 & -0.20 & -0.14 & 0.31 & 0.46 \end{bmatrix}.$$

The largest components in each row do correspond to the ones in the original word vector matrix  $A$ . However, in this application we work with the representation in the low dimensional, 2D, subspace spanned by the first two principal vectors  $\mathbf{v}_1$  and  $\mathbf{v}_2$ .

**Angles measure similarity** Recall that [Example 1.3.9](#) introduced using the dot product to measure the similarity between word vectors. We could use the dot product in the 5D space of the word vectors to find the ‘angles’ between the book titles. However, we know that the 2D view just plotted is the ‘best’ 2D summary of the book titles, so we could more economically estimate the angle between book titles using just the 2D summary.

**Example 5.1.27.** What is the ‘angle’ between the first two listed books?

- Introduction to Finite and Spectral Element Methods using MATLAB
- Iterative Methods for Linear Systems: Theory and Applications



We can also use the 2D plane to economically measure similarity between the book titles and any other title or words of interest.

**Example 5.1.28.** Let’s ask which of the six books is ‘closest’ to a book about Applications.



**Search for information from more books** [Berry et al. \(1995\)](#) reviewed the application of the SVD to the problem of searching for information. Let’s explore this further with more data, albeit still

very restricted. [Berry et al. \(1995\)](#) listed some mathematical books including the following fourteen titles.

1. a Course on Integral Equations
2. Automatic Differentiation of Algorithms: Theory, Implementation, and Application
3. Geometrical Aspects of Partial Differential Equations
4. Introduction to Hamiltonian Dynamical Systems and the n-Body Problem
5. Knapsack Problems: Algorithms and Computer Implementations
6. Methods of Solving Singular Systems of Ordinary Differential Equations
7. Nonlinear Systems
8. Ordinary Differential Equations
9. Oscillation Theory of Delay Differential Equations

10. Pseudodifferential Operators and Nonlinear Partial Differential Equations
11. Sinc Methods for Quadrature and Differential Equations
12. Stability of Stochastic Differential Equations with Respect to Semi-Martingales
13. the Boundary Integral Approach to Static and Dynamic Contact Problems
14. the Double Mellin–Barnes Type Integrals and their Applications to Convolution Theory

Principal component analysis summarises and relates these titles. Follow [Procedure 5.1.26](#).

1. The significant (capitalised) words which appear more than once in these titles (ignoring pluralisation) are the fourteen

words

Algorithm, Application, Differential/tion,  
Dynamic/al, Equation, Implementation, Integral, (5.3)  
Method, Nonlinear, Ordinary, Partial, Problem,  
System, and Theory.

With this dictionary of significant words, the titles have the following word vectors.

- $\mathbf{w}_1 = (0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0)$  a Course on Integral Equations
- $\mathbf{w}_2 = (1, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1)$  Automatic Differentiation of Algorithms: Theory, Implementation, and Application
- ...
- $\mathbf{w}_{14} = (0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1)$  the Double Mellin-Barnes Type Integrals and their Applications to Convolution Theory

2. Form the  $14 \times 14$  data matrix with the word count for each title in rows

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

Each row corresponds to a book title, and each column corresponds to a word.

3. To compute a representation of the titles in 3D space, princi-



pal components uses, as an orthonormal basis, the singular vectors corresponding to the three largest singular values. So in MATLAB/Octave compute the economical SVD with  $[U,S,V]=svds(A,3)$  giving (2 d.p.)

```
U = ...
S =
    4.20    0    0
    0    2.65    0
    0    0    2.36
V =
    0.07    0.40    0.14
    0.07    0.38    0.25
    0.65    0.00    0.15
    0.01    0.23   -0.46
    0.64   -0.21   -0.07
    0.07    0.40    0.14
    0.06    0.30   -0.18
    0.19   -0.09   -0.12
    0.10   -0.05   -0.11
```



0.19	-0.09	-0.12
0.17	-0.09	0.02
0.02	0.40	-0.50
0.12	0.05	-0.48
0.16	0.41	0.32

4. The three columns of  $V$  are word vectors in the 14D space of counts of the dictionary words (5.3) Algorithm, Application, Differential, Dynamic, Equation, Implementation, Integral, Method, Nonlinear, Ordinary, Partial, Problem, System, and Theory.
- The first column  $\mathbf{v}_1$  of  $V$ , from its largest components, mainly identifies the two most common words of Differential and Equation.
  - The second column  $\mathbf{v}_2$  of  $V$ , from its largest components, identifies books with Algorithms, Applications, Implementations, Problems, and Theory.
  - The third column  $\mathbf{v}_3$  of  $V$ , from its largest components, largely distinguishes Dynamics, Problems and Systems,

from Differential and Theory.

The corresponding principal components are the entries of the  $14 \times 3$  matrix (2 d.p.)

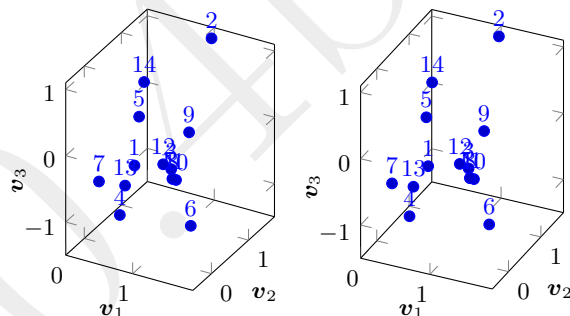
$$AV = \begin{bmatrix} 0.70 & 0.09 & -0.25 \\ 1.02 & 1.59 & 1.00 \\ 1.46 & -0.29 & 0.10 \\ 0.16 & 0.67 & -1.44 \\ 0.16 & 1.19 & -0.22 \\ 1.78 & -0.34 & -0.64 \\ 0.22 & -0.00 & -0.58 \\ 1.48 & -0.29 & -0.04 \\ 1.45 & 0.21 & 0.40 \\ 1.56 & -0.34 & -0.01 \\ 1.48 & -0.29 & -0.04 \\ 1.29 & -0.20 & 0.08 \\ 0.10 & 0.92 & -1.14 \\ 0.29 & 1.09 & 0.39 \end{bmatrix}.$$

Each of the fourteen books is represented in 3D space by the corresponding row of these coordinates. Plot these books in

MATLAB/Octave with

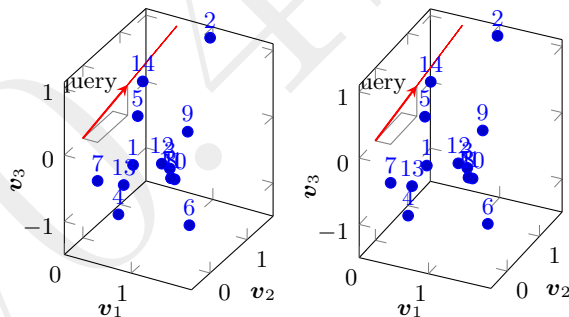
```
scatter3(A*V(:,1),A*V(:,2),A*V(:,3),[],1:14)
```

as shown below in stereo.



There is a cluster of five books near the front along the  $v_1$ -axis (numbered 3, 8, 10, 11 and 12, their focus is Differential Equations), the other nine are spread out.

**Queries** Suppose we search for books on *Application and Theory*. In our dictionary (5.3), the corresponding word vector for this search is  $\mathbf{w} = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1)$ . Project this query into the 3D space of principal components with the product  $\mathbf{w}^T V$  which evaluates to the query vector  $\mathbf{q} = (0.22, 0.81, 0.46)$  whose direction is added to the picture as shown below.



Books 2 and 14 appear close to the direction of the query vector and so should be returned as a match: these books are no surprise as both their titles have both *Application* and *Theory* in their titles. But the above plot also suggests Book 5 is near to the direction

of the query vector, and so is also worth considering despite not having either of the search words in its title! The power of this latent semantic indexing is that it extracts additional titles that are relevant to the query yet share no common words with the query—as commented at the start of this section.

The angles between the query vector and the book title 3D vectors confirm the graphical appearance claimed above. Recall that the dot product determines the angle between vectors ([Theorem 1.3.5](#)).

- From the second row of the above product  $AV$ , Book 2 has the principal component vector  $(1.02, 1.59, 1.00)$  which has length 2.14. Consequently, it is at small angle  $15^\circ$  to the 3D query vector  $\mathbf{q} = (0.22, 0.81, 0.46)$ , of length  $|\mathbf{q}| = 0.96$ , because its cosine

$$\cos \theta = \frac{(1.02, 1.59, 1.00) \cdot \mathbf{q}}{2.14 \cdot 0.96} = 0.97.$$

- Similarly, Book 14 has the principal component vector  $(0.29, 1.09, 0.46)$  which has length 1.20. Consequently, it is at small angle  $10^\circ$

to the 3D query vector  $\mathbf{q} = (0.22, 0.81, 0.46)$  because its cosine

$$\cos \theta = \frac{(0.29, 1.09, 0.39) \cdot \mathbf{q}}{1.20 \cdot 0.96} = 0.99.$$

- Whereas Book 5 has the principal component vector  $(0.16, 1.19, -0.22)$  which has length 1.22. Consequently, it is at moderate angle  $40^\circ$  to the 3D query vector  $\mathbf{q} = (0.22, 0.81, 0.46)$  because its cosine

$$\cos \theta = \frac{(0.16, 1.19, -0.22) \cdot \mathbf{q}}{1.20 \cdot 0.96} = 0.76.$$

Such a significant cosine suggests that Book 5 is also of interest.

If we were to compute the angles in the original 14D space of the full dictionary (5.3), then the title of Book 5 would be orthogonal to the query, because it has no words in common, and so Book 5 would not be flagged as of interest. The principal component analysis reduces the dimensionality to those relatively few directions that are important, and it is in

these important directions that the title of Book 5 appears promising for the query.

- All the other book titles have angles greater than  $62^\circ$  and so are significantly less related to the query.

**Latent semantic indexing in practice** This application of principal components to analysing a few book titles is purely indicative. In practice one would analyse the many thousands of words used throughout hundreds or thousands of documents. Moreover, one would be interested in not just plotting the documents in a 2D plane or 3D space, but in representing the documents in say a 70D space of seventy principal components. [Berry et al. \(1995\)](#) reviews how such statistically derived principal word vectors are a more robust indicator of meaning than individual terms. Hence this SVD analysis of documents becomes an effective way of retrieving information from a search without requiring the results actually match any of the words in the search request—the results just need to match cognate words.

Table 5.3: twenty user reviews of bathrooms in a major chain of hotels. The data comes from the Opinions Opinion/Review in the UCI Machine Learning Repository.

- The room was not overly big, but clean and very comfortable beds, a great shower and very clean bathrooms
- The second room was smaller, with a very inconvenient bathroom layout, but at least it was quieter and we were able to sleep
- Large comfortable room, wonderful bathroom
- The rooms were nice, very comfy bed and very clean bathroom
- Bathroom was spacious too and very clean
- The bathroom only had a single sink, but it was very large
- The room was a standard but nice motel room like any other, bathroom seemed upgraded if I remember
- The room was quite small but perfectly formed with a super bathroom
- You could eat off the bathroom floor it was so clean
- The bathroom door does the same thing, making the bathroom seem slightly larger
- bathroom spotless and nicely appointed
- The rooms are exceptionally clean and also the bathrooms



## 5.2 Regularise linear equations

### Section Contents

5.2.1	The SVD illuminates regularisation . . . . .	603
5.2.2	Tikhonov regularisation . . . . .	611

Singularity is almost invariably a clue.

*Sherlock Holmes, in The Boscombe  
Valley Mystery, by Sir Arthur Co-  
nan Doyle, 1892*

Often we need to approximate the matrix in a linear equation. This is especially so when the matrix itself comes from experimental measurements and so is subject to errors. We do not want such errors to affect results. By avoiding division with small singular values, the procedure developed in this section avoids unwarranted magnification of errors. Sometimes such error magnification is disastrous, so avoiding it is essential.

**Example 5.2.1.** Suppose from measurements in some experiment we want to solve the linear equations

$$0.5x + 0.3y = 1 \quad \text{and} \quad 1.1x + 0.7y = 2,$$

where all the coefficients on *both* the left-hand sides and the right-hand sides are determined from experimental measurements. In particular, suppose they are measured to errors  $\pm 0.05$ . Solve the equations. ■

**Activity 5.2.2.** The coefficients in the following pair of linear equations are obtained from an experiment and so the coefficients have errors of roughly  $\pm 0.05$ :

$$0.8x + 1.1y = 4, \quad 0.6x + 0.8y = 3.$$

By checking how well the equations are satisfied, which of the following *cannot* be a plausible solution  $(x, y)$  of the pair of equations?

(a)  $(5, 0)$

(b)  $(3.6, 1)$

(c)  $(6.6, -1.2)$

(d)  $(5.6, 0.8)$



### 5.2.1 The SVD illuminates regularisation

I think it is much more interesting to live with uncertainty than to live with answers that might be wrong.

*Richard Feynman*

**Procedure 5.2.3** (approximate linear equations). *Suppose the system of linear equation  $A\mathbf{x} = \mathbf{b}$  arises from experiment where both the  $m \times n$  matrix  $A$  and the right-hand side vector  $\mathbf{b}$  are subject to experimental error. Suppose the expected error in the matrix entries are of size  $\epsilon$ .*

*Recall that (Theorem 3.3.29) the symbol  $\epsilon$  is the Greek letter epsilon, and often denotes errors.*

1. *When forming the matrix  $A$  and vector  $\mathbf{b}$ , scale the data so that*
  - *all  $m \times n$  components in  $A$  have the same physical units, and they are of roughly the same size; and*
  - *similarly for the  $m$  components of  $\mathbf{b}$ .*

*Estimate the error  $\epsilon$  corresponding to this matrix  $A$ .*

2. Compute an SVD  $A = USV^T$ .
3. Choose ‘rank’  $k$  to be the number of singular values bigger than the error  $\epsilon$ ; that is,  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_k > \epsilon > \sigma_{k+1} \geq \dots \geq 0$ . Then the rank  $k$  approximation to  $A$  is

$$\begin{aligned} A_k &:= US_kV^T \\ &= \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \dots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T \\ &= \mathbf{U}(:, 1:k) * \mathbf{S}(1:k, 1:k) * \mathbf{V}(:, 1:k)' . \end{aligned}$$

We usually do not construct  $A_k$  as we only need its SVD to solve the system.

4. Solve the approximating linear equation  $A_k \mathbf{x} = \mathbf{b}$  as in Theorems 3.5.8–3.5.13 (often as an inconsistent set of equations). Usually use the SVD  $A_k = US_kV^T$ .
5. Among all the solutions allowed, choose the ‘best’ according to some explicit additional need of the application: often the smallest solution overall; or just as often a solution with the most zero components.

That is, the procedure is to treat as zero all singular values smaller than the expected error in the matrix entries. For example, modern computers have nearly sixteen significant decimal digits accuracy, so even in ‘exact’ computation there is a background relative error of about  $10^{-15}$ . Consequently, in computation on modern computers, every singular value smaller than  $10^{-15}\sigma_1$  must be treated as zero. For safety, even in ‘exact’ computation, every singular value smaller than say  $10^{-8}\sigma_1$  should be treated as zero.

**Activity 5.2.4.** In some system of linear equations the five singular values of the matrix are

$$1.5665, \quad 0.2222, \quad 0.0394, \quad 0.0107, \quad 0.0014.$$

Given the matrix components have errors of about 0.02, what is the effective rank of the matrix?

- (a) 1                      (b) 2                      (c) 3                      (d) 4



The final step in [Procedure 5.2.3](#) arises because in many cases an infinite number of possible solutions are derived. The linear algebra cannot presume which is best for your application. Consequently, you will have to be aware of the freedom, and make a choice based on extra information from your particular application.

- For example, in a CT-scan such as [Example 3.5.17](#) one would usually prefer the grayest result in order to avoid diagnosing artifices.
- For example, in the data mining task of fitting curves or surfaces to data, one would instead usually prefer a curve or surface with fewest non-zero coefficients.

Such extra information from the application is essential.

**Example 5.2.5.** For the following matrices  $A$  and right-hand side vectors  $\mathbf{b}$ , solve  $A\mathbf{x} = \mathbf{b}$ . But suppose the matrix entries come from experiments and are only known to within errors  $\pm 0.05$ , solve  $A'\mathbf{x}' = \mathbf{b}$  for some specific matrices  $A'$  which approximate  $A$  to this error. Finally, use an SVD to find a general solution consistent with the error in matrix  $A$ . Report to two decimal places.

$$(a) \quad A = \begin{bmatrix} -0.2 & -0.6 & 1.8 \\ 0.0 & 0.2 & -0.4 \\ -0.3 & 0.7 & 0.3 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -0.5 \\ 0.1 \\ -0.2 \end{bmatrix}$$

$$(b) \quad A = \begin{bmatrix} -1.1 & 0.1 & 0.7 & -0.1 \\ 0.1 & -0.1 & 1.2 & -0.6 \\ 0.8 & -0.2 & 0.4 & -0.8 \\ 0.8 & 0.1 & -2.0 & 1.0 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} -1.1 \\ -0.1 \\ 1.1 \\ 0.8 \end{bmatrix}$$

Both of these examples gave an infinite number of solutions which are equally valid as far as the linear algebra is concerned. In each example, more information from an application would be needed to choose which to *prefer* among the infinity of solutions. ■

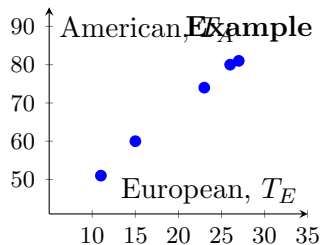
Most often the singular values are spread over a wide range of orders of magnitude. In such cases an assessment of the errors in the matrix is crucial in what one reports as a solution. The following artificial example illustrates the range.



**Example 5.2.6** (various errors). The matrix

$$A = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} \\ \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} & \frac{1}{8} \\ \frac{1}{5} & \frac{1}{6} & \frac{1}{7} & \frac{1}{8} & \frac{1}{9} \end{bmatrix}$$

is an example of a so-called Hilbert matrix. Explore the effects of various assumptions about possible errors in  $A$  upon the solution to  $A\mathbf{x} = \mathbf{1}$  where  $\mathbf{1} := (1, 1, 1, 1, 1)$ . ■



**Example 5.2.7** (translating temperatures). Recall [Example 2.2.12](#) attempts to fit a quartic polynomial to observations (plotted in the margin) of the relation between Celsius and Fahrenheit temperature. The attempt failed because `rcond` is too small. Let's try again now that we can cater for matrices with errors. Recall the data between temperatures reported by a European and an American are the following:

$$\begin{array}{c|ccccc} T_E & 15 & 26 & 11 & 23 & 27 \\ T_A & 60 & 80 & 51 & 74 & 81 \end{array}$$

Example 2.2.12 attempts to fit the data with the quartic polynomial

$$T_A = c_1 + c_2 T_E + c_3 T_E^2 + c_4 T_E^3 + c_5 T_E^4,$$

and deduced the following system of equations for the coefficients

$$\begin{bmatrix} 1 & 15 & 225 & 3375 & 50625 \\ 1 & 26 & 676 & 17576 & 456976 \\ 1 & 11 & 121 & 1331 & 14641 \\ 1 & 23 & 529 & 12167 & 279841 \\ 1 & 27 & 729 & 19683 & 531441 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \end{bmatrix} = \begin{bmatrix} 60 \\ 80 \\ 51 \\ 74 \\ 81 \end{bmatrix}.$$

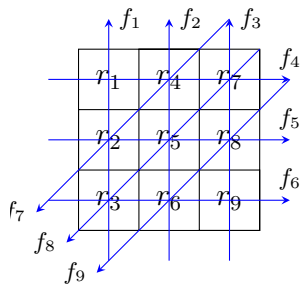
In order to find a robust solution, here let's approximate both the matrix and the right-hand side vector because both the matrix and the vector come from real data with errors of about up to  $\pm 0.5^\circ$ .



Occam's razor: Non sunt multiplicanda entia sine necessitate [Entities must not be multiplied beyond necessity]

*John Punch (1639)*

**Example 5.2.8.** Recall that ?? introduced extra ‘diagonal’ measurements into a 2D CT-scan. As shown in the margin, the 2D region is divided into a  $3 \times 3$  grid of nine blocks. Then measurements taken of the X-rays not absorbed along the shown nine paths: three horizontal, three vertical, and three diagonal. Suppose the measured fractions of X-ray energy are  $\mathbf{f} = (0.048, 0.081, 0.042, 0.020, 0.106, 0.075, 0.177, 0.181, 0.105)$ . Use an SVD to find the ‘grayest’ transmission factors consistent with the measurements and likely errors.



### 5.2.2 Tikhonov regularisation

This optional extension connects to much established practice that graduates may encounter.

Regularisation of poorly-posed linear equations is a widely used practical necessity. Many people have invented alternative techniques. Many have independently re-invented techniques. Perhaps the most common is the so-called Tikhonov regularisation. This section introduces and discusses Tikhonov regularisation.

In statistics, the method is known as ridge regression, and with multiple independent discoveries, it is also variously known as the Tikhonov–Miller method, the Phillips–Twomey method, the constrained linear inversion method, and the method of linear regularization.

*Wikipedia (2015)*

#### Definition 5.2.9.

The greek letter  $\alpha$  is ‘alpha’ (different to the ‘proportional to’ symbol  $\propto$ ).

*In seeking to solve the poorly-posed system  $A\mathbf{x} = \mathbf{b}$  for  $m \times n$  matrix  $A$ , a **Tikhonov regularisation** is the system  $(A^T A + \alpha^2 I_n)\mathbf{x} = A^T \mathbf{b}$  for some chosen regularisation parameter value  $\alpha > 0$ .*

**Example 5.2.10.** Use Tikhonov regularisation to solve the system of  
**Example 5.2.1:**

$$0.5x + 0.3y = 1 \quad \text{and} \quad 1.1x + 0.7y = 2,$$

**Activity 5.2.11.** In the linear system for  $\mathbf{x} = (x, y)$ ,

$$4x - y = -4, \quad -2x + y = 3,$$

the coefficients on the left-hand side are in error by about  $\pm 0.3$ . Tikhonov regularisation should solve which one of the following systems?

- (a)  $\begin{bmatrix} 18.3 & -5 \\ -10 & 3.3 \end{bmatrix} \mathbf{x} = \begin{bmatrix} -19 \\ 11 \end{bmatrix}$       (b)  $\begin{bmatrix} 18.1 & -5 \\ -10 & 3.1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} -19 \\ 11 \end{bmatrix}$
- (c)  $\begin{bmatrix} 20.3 & -6 \\ -6 & 2.3 \end{bmatrix} \mathbf{x} = \begin{bmatrix} -22 \\ 7 \end{bmatrix}$       (d)  $\begin{bmatrix} 20.1 & -6 \\ -6 & 2.1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} -22 \\ 7 \end{bmatrix}$

Do not apply Tikhonov regularisation blindly as it does introduce biases. The following example illustrates the bias.

**Example 5.2.12.** Recall [Example 3.5.1](#) at the start of [Subsection 3.5.1](#) where scales variously reported my weight in kg as 84.8, 84.1, 84.7 and 84.4. To best estimate my weight  $x$  we rewrote the problem in matrix-vector form

$$Ax = \mathbf{b}, \quad \text{namely} \quad \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} x = \begin{bmatrix} 84.8 \\ 84.1 \\ 84.7 \\ 84.4 \end{bmatrix}.$$

A Tikhonov regularisation of this inconsistent system is

$$\left( \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \alpha^2 \right) x = \begin{bmatrix} 84.8 \\ 84.1 \\ 84.7 \\ 84.4 \end{bmatrix}.$$

That is,  $(4 + \alpha^2)x = 338$  kg with solution  $x = 338/(4 + \alpha^2) = 84.5/(1 + \alpha^2/4)$  kg. Because of the division by  $1 + \alpha^2/4$ , this

Tikhonov answer is biased as it is systematically below the average 84.5 kg. For small Tikhonov parameter  $\alpha$  the bias is small, but even so such a bias is unpleasant. ■

**Example 5.2.13.** Use Tikhonov regularisation to solve  $A\mathbf{x} = \mathbf{b}$  for the matrix and vector of [Example 5.2.5a](#). ■

Although [Definition 5.2.9](#) does not look like it, Tikhonov regularisation relates directly to the SVD regularisation of [Subsection 5.2.1](#). The next theorem establishes the connection.

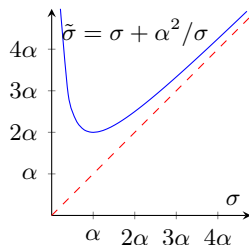
**Theorem 5.2.14** (Tikhonov regularisation). *Solving the Tikhonov regularisation, with parameter  $\alpha$ , of  $A\mathbf{x} = \mathbf{b}$  is equivalent to finding the smallest, least square, solution of the system  $\tilde{A}\mathbf{x} = \mathbf{b}$  where the matrix  $\tilde{A}$  is obtained from  $A$  by replacing each of its non-zero singular values  $\sigma_i$  by  $\tilde{\sigma}_i := \sigma_i + \alpha^2/\sigma_i$ .*

There is another reason to be careful when using Tikhonov regularisation. Yes, it gives a nice, neat, unique solution. However, it

does not hint that there may be an infinite number of equally good nearby solutions (as found through [Procedure 5.2.3](#)). Among those equally good nearby solutions may be ones that you prefer in your application.

### Choose a good regularisation parameter

- One strategy to choose the regularisation parameter  $\alpha$  is that the effective change in the matrix, from  $A$  to  $\tilde{A}$ , should be about the size of errors expected in  $A$ . Since changes in the matrix are largely measured by the singular values we need to consider the relation between  $\tilde{\sigma} = \sigma + \alpha^2/\sigma$  and  $\sigma$ . From the marginal graph the small singular values are changed by a lot, but these are the ones for which we want  $\tilde{\sigma}$  large in order give a ‘least square’ approximation. Significantly, the marginal graph also shows that singular values larger than  $\alpha$  change by less than  $\alpha$ . Thus the parameter  $\alpha$  should not be much larger than the expected error in the elements of the matrix  $A$ .



- Another consideration is the effect of regularisation upon



errors in the right-hand side vector. The condition number of  $A$  may be very bad. However, as the marginal graph shows the smallest  $\tilde{\sigma} \geq 2\alpha$ . Thus, in the regularised system the condition number of the effective matrix  $\tilde{A}$  is approximately  $\sigma_1/(2\alpha)$ . We need to choose the regularisation parameter  $\alpha$  large enough so that  $\frac{\sigma_1}{2\alpha} \times (\text{relative error in } \mathbf{b})$  is an acceptable relative error in the solution  $\mathbf{x}$  ([Theorem 3.3.29](#)). It is only when the regularisation parameter  $\alpha$  is big enough that the regularisation will be effective in finding a least square approximation.

---

## 6 Determinants distinguish matrices

---

### Chapter Contents

6.1	Geometry underlies determinants . . . . .	620
6.2	Laplace expansion theorem for determinants . . . . .	640

Although much of the theoretical role of determinants is usurped by the SVD, nonetheless, determinants aid in establishing forthcoming properties of eigenvalues and eigenvectors, and empower graduates to connect to much extant practice.

Recall from previous study ([Section 4.1.1](#), e.g.)

- a  $2 \times 2$  matrix  $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$  has determinant  $\det A = |A| = ad - bc$ , and that the matrix  $A$  is invertible iff and only if  $\det A \neq 0$ ;
- a  $3 \times 3$  matrix  $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$  has determinant  $\det A = |A| = aei + bfg + cdh - ceg - afh - bdi$ , and that the matrix  $A$  is invertible if and only if  $\det A \neq 0$ .

For hand calculations, these two formulas for a determinant are best remembered via the following diagrams where products along the red lines are subtracted from the products along the blue lines, respectively:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \quad (6.1)$$

This chapter extends these determinants to any size matrix, and explores more of the useful properties of a determinant—especially those properties useful for understanding and developing the general

eigenvalue problems and applications of [Chapter 7](#).

VO.410

## 6.1 Geometry underlies determinants

### Section Contents

Sections 3.2.2, 3.2.3 and 3.6 introduced that multiplication by a matrix transforms areas and volumes. Determinants give precisely how much a square matrix transforms such areas and volumes.

**Example 6.1.1.** Consider the square matrix  $A = \begin{bmatrix} \frac{1}{2} & 1 \\ 0 & 1 \end{bmatrix}$ . Use matrix multiplication to find the image of the unit square under the transformation by  $A$ . How much is the area of the unit square scaled up/down? Compare with the determinant. ■

**Example 6.1.2.** Consider the square matrix  $B = \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$ . Use matrix multiplication to find the image of the unit square under the transformation by  $B$ . How much is the unit area scaled up/down? Compare with the determinant. ■

**Activity 6.1.3.** Upon multiplication by the matrix  $\begin{bmatrix} -2 & 5 \\ -3 & -2 \end{bmatrix}$  the unit square transforms to a parallelogram. Use the determinant of the matrix to find the area of the parallelogram is which of the following.

- (a) 19                      (b) 4                      (c) 11                      (d) 16

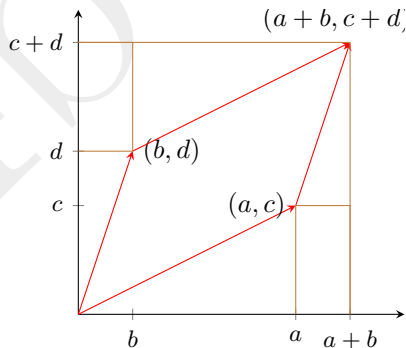


**Example 6.1.4.** Consider the square matrix  $C = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & \frac{3}{2} \end{bmatrix}$ . Use matrix multiplication to find the image of the unit cube under the transformation by  $C$ . How much is the volume of the unit cube scaled up/down? Compare with the determinant.



**Determinants determine area transformation** Consider multiplication by the general  $2 \times 2$  matrix  $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ .

Under multiplication by this matrix  $A$  the unit square becomes the parallelogram shown with four corners at  $(0, 0)$ ,  $(a, c)$ ,  $(b, d)$  and  $(a+b, c+d)$ . Let's determine the area of the parallelogram by that of the containing rectangle (brown) less the two small rectangles and the four small triangles. The two small rectangles have the same area, namely  $bc$ . The two small triangles on the left and the right also have the same area, namely  $\frac{1}{2}bd$ . The two small triangles on the top and the bottom have the same area, namely  $\frac{1}{2}ac$ . Thus, under multiplication by matrix  $A$  the image of the unit square is the parallelogram with

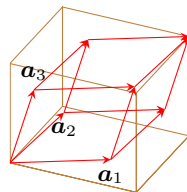


$$\text{area} = (a+b)(c+d) - 2bc - 2 \cdot \frac{1}{2}bd - 2 \cdot \frac{1}{2}ac$$

$$\begin{aligned}
 &= ac + ad + bc + bd - 2bc - bd - ac \\
 &= ad - bc = \det A.
 \end{aligned}$$

This picture is the case when the matrix does not also reflect the image: if the matrix also reflects, as in [Example 6.1.2](#), then the determinant is the negative of the area. In either case, the area of the unit square after transforming by the matrix  $A$  is the magnitude  $|\det A|$ .

Analogous geometric arguments relate determinants of  $3 \times 3$  matrices with transformations of volumes. Under multiplication by a  $3 \times 3$  matrix  $A = [\mathbf{a}_1 \ \mathbf{a}_2 \ \mathbf{a}_3]$ , the image of the unit cube is a parallelepiped with edges  $\mathbf{a}_1$ ,  $\mathbf{a}_2$  and  $\mathbf{a}_3$  as illustrated. By computing the volumes of various rectangular boxes, prisms and tetrahedra, the volume of such a parallelepiped could be expressed as the  $3 \times 3$  determinant formula ([6.1](#)).



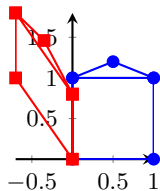
In higher dimensions we want the determinant to behave analogously and so next define the determinant to do so. We use the



terms  **$n$ D-cube** to generalise a square and cube to  $n$  dimensions ( $\mathbb{R}^n$ ),  **$n$ D-volume** to generalise the notion of area and volume to  $n$  dimensions, and so on. When the dimension of the space is unspecified, then we may say **hyper-cube**, **hyper-volume**, and so on.

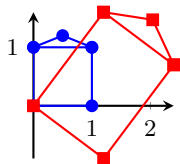
**Definition 6.1.5.** Let  $A$  be an  $n \times n$  square matrix, and let  $C$  be the unit  $n$ D-cube in  $\mathbb{R}^n$ . Transform the  $n$ D-cube  $C$  by  $\mathbf{x} \mapsto A\mathbf{x}$  to its image  $C'$  in  $\mathbb{R}^n$ . Define the **determinant** of  $A$ , denoted either  $\det A$  or sometimes  $|A|$  such that:

- the magnitude  $|\det A|$  is the  $n$ D-volume of  $C'$ ; and
- the sign of  $\det A$  to be negative iff the transformation reflects the orientation of the  $n$ D-cube.



**Example 6.1.6.** Roughly estimate the determinant of the matrix that transforms the unit square to the parallelogram as shown in the margin. ■

**Activity 6.1.7.** Roughly estimate the determinant of the matrix that transforms the unit square to the rectangle as shown in the margin.



(a) 2

(b) 4

(c) 3

(d) 2.5



Basic properties of a determinant follow direct from [Definition 6.1.5](#).

**Theorem 6.1.8.** (a) For every  $n \times n$  diagonal matrix  $D$ , the determinant of  $D$  is the product of the diagonal entries:  $\det D = d_{11}d_{22} \cdots d_{nn}$ .

(b) Every orthogonal matrix  $Q$  has  $\det Q = \pm 1$  (only one alternative, not both). Further,  $\det Q = \det(Q^T)$ .

(c) For every  $n \times n$  matrix  $A$ ,  $\det(kA) = k^n \det A$  for every scalar  $k$ .

**Example 6.1.9.** The determinant of the  $n \times n$  identity matrix is one: that is,  $\det I_n = 1$ . We justify this result in two ways.

- An identity matrix is a diagonal matrix and hence its determinant is the product of the diagonal entries ([Theorem 6.1.8a](#)), here all ones.
- Alternatively, multiplication by the identity does not change the unit  $n$ D-cube and so does not change its  $n$ D-volume ([Definition 6.1.5](#)).



**Activity 6.1.10.** What is the determinant of  $-I_n$ ?

- (a)  $-1$  (b)  $-1$  for odd  $n$ , and  $+1$  for even  $n$   
(c)  $+1$  for odd  $n$ , and  $-1$  for even  $n$  (d)  $+1$



**Example 6.1.11.** Use (6.1) to compute the determinant of the orthogonal matrix

$$Q = \begin{bmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\ -\frac{2}{3} & \frac{1}{3} & \frac{2}{3} \end{bmatrix}.$$

Then use [Theorem 6.1.8](#) to deduce the determinants of the following

matrices:

$$\begin{bmatrix} \frac{1}{3} & \frac{2}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} & \frac{2}{3} \end{bmatrix}, \quad \begin{bmatrix} -1 & 2 & -2 \\ -2 & -2 & -1 \\ 2 & -1 & -2 \end{bmatrix}, \quad \begin{bmatrix} \frac{1}{6} & -\frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \\ -\frac{1}{3} & \frac{1}{6} & \frac{1}{3} \end{bmatrix}.$$

**Activity 6.1.12.**

Given  $\det \begin{bmatrix} 1 & -1 & -2 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & -1 & -2 & -1 \\ -1 & 0 & 0 & -1 \end{bmatrix} = -1$ , what is

$$\det \begin{bmatrix} -2 & 2 & 4 & 0 \\ -2 & -2 & -2 & -2 \\ -2 & 2 & 4 & 2 \\ 2 & 0 & 0 & 2 \end{bmatrix}?$$

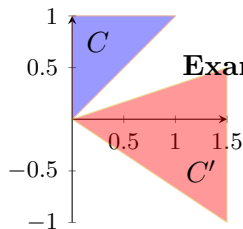
(a)  $-4$

(b)  $8$

(c)  $2$

(d)  $-16$

A consequence of [Theorem 6.1.8c](#) is that a determinant characterises the transformation of any sized hyper-cube. Consider the transformation by a matrix  $A$  of an  $n$ D-cube of side length  $k$  ( $k \geq 0$ ), and hence of volume  $k^n$ . The  $n$ D-cube has edges  $ke_1, ke_2, \dots, ke_n$ . The transformation results in an  $n$ D-parallelepiped with edges  $A(ke_1), A(ke_2), \dots, A(ke_n)$ , which by commutativity and associativity ([Theorem 3.1.25d](#)) are the same edges as  $(kA)e_1, (kA)e_2, \dots, (kA)e_n$ . That is, the resulting  $n$ D-parallelepiped is the same as applying matrix  $(kA)$  to the unit  $n$ D-cube, and so must have  $n$ D-volume  $k^n |\det A|$ . This is a factor of  $|\det A|$  times the original volume. Crucially, this property that matrix multiplication multiplies all sizes of hyper-cubes by the determinant holds for all other shapes and sizes, not just hyper-cubes. Let's see an specific example before proving the general theorem.



**Example 6.1.13.**

Multiplication by some specific matrix transforms the (blue) triangle  $C$  to the (red) triangle  $C'$  as shown in the margin. By finding the ratio of the areas, estimate the magnitude of the determinant of the matrix. ■

**Theorem 6.1.14.** *Consider any bounded smooth  $nD$ -volume  $C$  in  $\mathbb{R}^n$  and its image  $C'$  after multiplication by  $n \times n$  matrix  $A$ . Then*

$$\det A = \pm \frac{nD\text{-volume of } C'}{nD\text{-volume of } C}$$

*with the negative sign when matrix  $A$  changes the orientation.*

A more rigorous proof would involve upper and lower sums for the original and transformed regions, and also explicit restrictions to regions where these upper and lower sums converge to a unique  $nD$ -volume. We do not detail such a more rigorous proof here.

This property of transforming general areas and volumes also establishes the next crucial property of determinants, namely that the determinant of a matrix product is the product of the determinants:  $\det(AB) = \det(A) \det(B)$  for all square matrices  $A$  and  $B$  (of the same size).

**Example 6.1.15.** Recall the two  $2 \times 2$  matrices of Examples 6.1.1 and 6.1.2:

$$A = \begin{bmatrix} \frac{1}{2} & 1 \\ 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}.$$

Check that the determinant of their product is the product of their determinants. ■

**Theorem 6.1.16.** *For every two  $n \times n$  matrices  $A$  and  $B$ ,  $\det(AB) = \det(A) \det(B)$ . Further, for  $n \times n$  matrices  $A_1, A_2, \dots, A_\ell$ ,  $\det(A_1 A_2 \cdots A_\ell) = \det(A_1) \det(A_2) \cdots \det(A_\ell)$ .*

**Activity 6.1.17.** Given that the three matrices

$$\begin{bmatrix} -1 & 0 & -1 \\ 0 & -1 & 1 \\ 0 & 1 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & -2 \\ -1 & -1 & 1 \\ -1 & -2 & -1 \end{bmatrix}, \quad \begin{bmatrix} -1 & 0 & -1 \\ 0 & 1 & 1 \\ 1 & 2 & 0 \end{bmatrix},$$

have determinants 2,  $-4$  and 3, respectively, what is the determinant of the product of the three matrices?



(a) 1

(b) 24

(c) 9

(d) -24



**Example 6.1.18.** (a) Confirm the product rule for determinants, [Theorem 6.1.16](#), for the product

$$\begin{bmatrix} -3 & -2 \\ 3 & -3 \end{bmatrix} = \begin{bmatrix} 3 & 1 & 1 \\ 0 & -3 & 0 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ -1 & 1 \\ 1 & -3 \end{bmatrix}.$$

(b) Given  $\det A = 2$  and  $\det B = \pi$ , what is  $\det(AB)$ ?



**Example 6.1.19.** Use the product theorem to help find the determinant of matrix

$$C = \begin{bmatrix} 45 & -15 & 30 \\ -2\pi & \pi & 2\pi \\ \frac{1}{9} & \frac{2}{9} & -\frac{1}{3} \end{bmatrix}.$$

■

We now proceed to link the determinant of a matrix to the singular values of the matrix.

**Example 6.1.20.** Recall [Example 3.3.4](#) showed that the following matrix has the given SVD:

$$\begin{aligned} A &= \begin{bmatrix} -4 & -2 & 4 \\ -8 & -1 & -4 \\ 6 & 6 & 0 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\ -\frac{2}{3} & \frac{1}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} 12 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} -\frac{8}{9} & -\frac{1}{9} & -\frac{4}{9} \\ -\frac{4}{9} & \frac{4}{9} & \frac{7}{9} \\ -\frac{1}{9} & -\frac{8}{9} & \frac{4}{9} \end{bmatrix}^T. \end{aligned}$$

Use this SVD to find the magnitude  $|\det A|$ . ■

**Theorem 6.1.21.** *For every  $n \times n$  square matrix  $A$ , the magnitude of its determinant  $|\det A| = \sigma_1 \sigma_2 \cdots \sigma_n$ , the product of all its singular values.*

**Example 6.1.22.** Confirm [Theorem 6.1.21](#) for the matrix  $A = \begin{bmatrix} 10 & 2 \\ 5 & 11 \end{bmatrix}$  of [Example 3.3.2](#). ■

**Activity 6.1.23.** The matrix  $A = \begin{bmatrix} -2 & -4 & 5 \\ -6 & 0 & -6 \\ 5 & 4 & -2 \end{bmatrix}$  has an SVD of

$$A = \begin{bmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\ -\frac{2}{3} & \frac{1}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} 9 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -\frac{8}{9} & -\frac{1}{9} & -\frac{4}{9} \\ -\frac{4}{9} & \frac{4}{9} & \frac{7}{9} \\ -\frac{1}{9} & -\frac{8}{9} & \frac{4}{9} \end{bmatrix}^T.$$

What is the the magnitude of the determinant of  $A$ ,  $|\det A|$ ?

(a) 81

(b) 18

(c) 4

(d) 0



**Example 6.1.24.** Use an SVD of the following matrix to find the magnitude of its determinant:

$$A = \begin{bmatrix} -2 & -1 & 4 & -5 \\ -3 & 2 & -3 & 1 \\ -3 & -1 & 0 & 3 \end{bmatrix}.$$



Establishing this connection between determinants and singular values relied on [Theorem 6.1.8b](#) that transposing an orthogonal matrix does not change its determinant,  $\det Q^T = \det Q$ . We now establish that this determinant-transpose property holds for the transpose of all square matrices.

**Example 6.1.25.**      *Example 6.1.18a* determined that  $\det \begin{bmatrix} -3 & -2 \\ 3 & -3 \end{bmatrix} = 15$ .

By (6.1), its transpose has determinant

$$\det \begin{bmatrix} -3 & 3 \\ -2 & -3 \end{bmatrix} = (-3)^2 - 3(-2) = 9 + 6 = 15.$$

The determinants are the same. ■

**Theorem 6.1.26.**      *For every square matrix  $A$ ,  $\det(A^T) = \det A$ .*

**Example 6.1.27.** A general  $3 \times 3$  matrix  $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$  has determinant

$$\det A = |A| = aei + bfg + cdh - ceg - afh - bdi. \text{ Its transpose,}$$

$$A^T = \begin{bmatrix} a & d & g \\ b & e & h \\ c & f & i \end{bmatrix}, \text{ from the rule (6.1)}$$

$$\begin{bmatrix} a & d & g \\ b & e & h \\ c & f & i \end{bmatrix} \begin{matrix} aei \\ dhc \\ gbf \\ gec \\ ahf \\ dbi \end{matrix}$$

has determinant

$$\det A^T = aei + dhc + gbf - gec - ahf - dbi = \det A.$$



One of the main reasons for studying determinants is to establish when solutions to linear equations may exist or not (albeit only

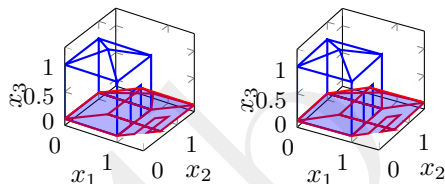
applicable to square matrices when there are  $n$  linear equations in  $n$  unknowns). One example lies in finding eigenvalues by hand (Section 4.1.1) where we solve  $\det(A - \lambda I) = 0$ .

Recall that for  $2 \times 2$  and  $3 \times 3$  matrices we commented that a matrix is invertible only when its determinant is non-zero. Theorem 6.1.29 establishes this in general. The geometric reason for this connection between invertibility and determinants is that when a determinant is zero the action of multiplying by the matrix ‘squashes’ the unit  $n$ D-cube into a  $n$ D-parallelepiped of zero thickness. Such extreme squashing cannot be uniquely undone.

**Example 6.1.28.** Consider multiplication by the matrix

$$A = \begin{bmatrix} 1 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

whose effect on the unit cube is illustrated below:



As illustrated, this matrix squashes the unit cube onto the  $x_1x_2$ -plane ( $x_3 = 0$ ). Consequently the resultant volume is zero and so  $\det A = 0$ . Because many points in 3D space are squashed onto the same point in the  $x_3 = 0$  plane, the action of the matrix cannot be undone. Hence the matrix is not invertible. That the matrix is not invertible and its determinant is zero is not a coincidence. ■

**Theorem 6.1.29.** *A square matrix  $A$  is invertible iff  $\det A \neq 0$ . If a matrix  $A$  is invertible, then  $\det(A^{-1}) = 1/(\det A)$ .*



## 6.2 Laplace expansion theorem for determinants

### Section Contents

This section develops a so-called row/column algebraic expansion for determinants. This expansion is useful for many theoretical purposes. But there are vastly more efficient ways of computing determinants than using a row/column expansion. In MATLAB/Octave one may invoke `det(A)` to compute the determinant of a matrix. You may find this function useful for checking the results of some examples and exercises. However, just like computing an inverse, computing the determinant is expensive and error prone. In medium to large scale problems avoid computing the determinant, something else is almost always better.

The most numerically reliable way to determine whether matrices are singular [not invertible] is to test their singular values. This is far better than trying to compute determinants, which have atrocious scaling properties.

*Cleve Moler, MathWorks, 2006*

Nonetheless, a row/column algebraic expansion for a determinant is useful for small matrix problems, as well as for its beautiful theoretical uses. We start with examples of row properties that underpin a row/column algebraic expansion.

**Example 6.2.1** (Theorem 6.2.5a).      Example 6.1.28 argued geometrically that the determinant is zero for the matrix

$$A = \begin{bmatrix} 1 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Confirm this determinant algebraically. ■

**Example 6.2.2** (Theorem 6.2.5b).      Consider the matrix with two identical rows,

$$A = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{5} \\ 1 & \frac{1}{2} & \frac{1}{5} \\ 0 & \frac{1}{2} & 1 \end{bmatrix}.$$

Confirm algebraically that its determinant is zero. Give a geometric reason for why its determinant has to be zero. ■

**Example 6.2.3** (Theorem 6.2.5c). Consider the two matrices with two rows swapped:

$$A = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \\ \frac{1}{5} & \frac{1}{2} & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 1 & 1 \\ 1 & -1 & 0 \\ \frac{1}{5} & \frac{1}{2} & 1 \end{bmatrix}$$

Confirm algebraically that their determinants are the negative of each other. Give a geometric reason why this should be so. ■

**Example 6.2.4** (Theorem 6.2.5d). Compute the determinant of the matrix

$$B = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \\ 2 & 5 & 10 \end{bmatrix}.$$

Compare  $B$  with matrix  $A$  given in [Example 6.2.3](#), and compare their determinants. ■

The above four examples are specific cases of the four general properties established as the four parts of the following theorem.

**Theorem 6.2.5** (row and column properties of determinants). *For every  $n \times n$  matrix  $A$  the following properties hold.*

- (a) *If  $A$  has a zero row or column, then  $\det A = 0$ .*
- (b) *If  $A$  has two identical rows or columns, then  $\det A = 0$ .*
- (c) *Let  $B$  be obtained by interchanging two rows or columns of  $A$ , then  $\det B = -\det A$ .*
- (d) *Let  $B$  be obtained by multiplying any one row or column of  $A$  by a scalar  $k$ , then  $\det B = k \det A$ .*

**Example 6.2.6.** You are given that  $\det A = -9$  for the matrix

$$A = \begin{bmatrix} 0 & 2 & 3 & 1 & 4 \\ -2 & 2 & -2 & 0 & -3 \\ 4 & -2 & -4 & 1 & 0 \\ 2 & -1 & -4 & 2 & 2 \\ 5 & 4 & 3 & -2 & -5 \end{bmatrix}.$$

Use [Theorem 6.2.5](#) to find the determinant of the following matrices, giving reasons.

$$(a) \begin{bmatrix} 0 & 2 & 3 & 0 & 4 \\ -2 & 2 & -2 & 0 & -3 \\ 4 & -2 & -4 & 0 & 0 \\ 2 & -1 & -4 & 0 & 2 \\ 5 & 4 & 3 & 0 & -5 \end{bmatrix}$$

$$(b) \begin{bmatrix} 0 & 2 & 3 & 1 & 4 \\ -2 & 2 & -2 & 0 & -3 \\ 2 & -1 & -4 & 2 & 2 \\ 4 & -2 & -4 & 1 & 0 \\ 5 & 4 & 3 & -2 & -5 \end{bmatrix}$$

$$(c) \begin{bmatrix} 0 & 2 & 3 & 1 & 4 \\ -2 & 2 & -2 & 0 & -3 \\ 4 & -2 & -4 & 1 & 0 \\ 2 & -1 & -4 & 2 & 2 \\ -2 & 2 & -2 & 0 & -3 \end{bmatrix} \quad (d) \begin{bmatrix} 0 & 1 & 3 & 1 & 4 \\ -2 & 1 & -2 & 0 & -3 \\ 4 & -1 & -4 & 1 & 0 \\ 2 & -\frac{1}{2} & -4 & 2 & 2 \\ 5 & 2 & 3 & -2 & -5 \end{bmatrix}$$

$$(e) \begin{bmatrix} -2 & 2 & -6 & 0 & -3 \\ 4 & -2 & -12 & 1 & 0 \\ 2 & -1 & -12 & 2 & 2 \\ 0 & 2 & 9 & 1 & 4 \\ 5 & 4 & 9 & -2 & -5 \end{bmatrix} \quad (f) \begin{bmatrix} 0 & 3 & 3 & 1 & 4 \\ -2 & 0 & -2 & 0 & -5 \\ 5 & -1 & -4 & 1 & 0 \\ 2 & -1 & -4 & 2 & 2 \\ 5 & 4 & 6 & -2 & -5 \end{bmatrix}$$



**Activity 6.2.7.** Now,  $\det \begin{bmatrix} 2 & -3 & 1 \\ 2 & -5 & -3 \\ -4 & 1 & -3 \end{bmatrix} = -36$ .

- Which of the following matrices has determinant of 18?

(a)  $\begin{bmatrix} 2 & -3 & 1/3 \\ 2 & -5 & -1 \\ -4 & 1 & -1 \end{bmatrix}$

(b)  $\begin{bmatrix} 2 & -3 & 1 \\ 2 & -5 & -3 \\ 2 & -5 & -3 \end{bmatrix}$

(c)  $\begin{bmatrix} -4 & 6 & -2 \\ 2 & -5 & -3 \\ -4 & 1 & -3 \end{bmatrix}$

(d)  $\begin{bmatrix} -1 & -3 & 1 \\ -1 & -5 & -3 \\ 2 & 1 & -3 \end{bmatrix}$

- Further, which has determinant  $-12$ ?  $0$ ?  $72$ ?



**Example 6.2.8.** Without evaluating the determinant, use [Theorem 6.2.5](#) to establish that the determinant equation

$$\begin{vmatrix} 1 & x & y \\ 1 & 2 & 3 \\ 1 & 4 & 5 \end{vmatrix} = 0 \quad (6.2)$$

is the equation of the straight line in the  $xy$ -plane that passes through the two points  $(2, 3)$  and  $(4, 5)$ . ■

**Example 6.2.9.** Without evaluating the determinant, use [Theorem 6.2.5](#) to establish that the determinant equation

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

is, in  $xyz$ -space, the equation of the plane that passes through the origin and the two points  $(-1, -2, 2)$  and  $(3, 5, 2)$ . ■



The next step in developing a general ‘formula’ for a determinant is the special class of matrices for which one column or row is zero except for one element.

**Example 6.2.10.** Find the determinant of  $A = \begin{bmatrix} -2 & -1 & -1 \\ 1 & -3 & -2 \\ 0 & 0 & 2 \end{bmatrix}$  which has two zeros in its last row. ■

**Theorem 6.2.11** (almost zero row/column). *For every  $n \times n$  matrix  $A$ , define the  $(i, j)$ th **minor**  $A_{ij}$  to be the  $(n - 1) \times (n - 1)$  square matrix obtained from  $A$  by omitting the  $i$ th row and  $j$ th column. If, except for the entry  $a_{ij}$ , the  $i$ th row (or  $j$ th column) of  $A$  is all zero, then*

$$\det A = (-1)^{i+j} a_{ij} \det A_{ij} . \quad (6.3)$$

The pattern of signs in this formula,  $(-1)^{i+j}$ , is

$$\begin{array}{cccccc} + & - & + & - & + & \cdots \\ - & + & - & + & - & \cdots \\ + & - & + & - & + & \cdots \\ - & + & - & + & - & \cdots \\ + & - & + & - & + & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{array}$$

**Example 6.2.12.** Use [Theorem 6.2.11](#) to evaluate the determinant of the following matrices.

$$(a) \begin{bmatrix} -3 & -3 & -1 \\ -3 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

$$(b) \begin{bmatrix} 2 & -1 & 7 \\ 0 & 3 & 0 \\ 2 & 2 & 5 \end{bmatrix}$$

$$(c) \begin{bmatrix} 2 & 4 & 3 \\ 8 & 0 & -1 \\ -5 & 0 & -2 \end{bmatrix}$$

$$(d) \begin{bmatrix} 2 & 1 & 3 \\ 0 & -2 & -3 \\ 0 & 2 & 4 \end{bmatrix}$$



**Activity 6.2.13.** Using one of the determinants in the above [Example 6.2.12](#), what is the determinant of the matrix

$$\begin{bmatrix} 2 & 1 & 0 & 3 \\ 5 & -2 & 15 & 2 \\ 0 & -2 & 0 & -3 \\ 0 & 2 & 0 & 4 \end{bmatrix} ?$$

- (a) 60                      (b) 120                      (c) -60                      (d) -120



**Example 6.2.14.** Use [Theorem 6.2.11](#) to evaluate the determinant of the so-called triangular matrix

$$A = \begin{bmatrix} 2 & -2 & 3 & 1 & 0 \\ 0 & 2 & -1 & -1 & -7 \\ 0 & 0 & 5 & -2 & -9 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 3 \end{bmatrix}$$



The relative simplicity of finding the determinant in [Example 6.2.14](#) indicates that there is something special and memorable about matrices with zeros in the entire lower-left ‘triangle’. There is, as expressed by the following definition and theorem.

**Definition 6.2.15.** *A **triangular matrix** is a square matrix where all entries are zero either to the lower-left of the diagonal or to the upper-right:*

- *an upper triangular matrix has the form (although any of the  $a_{ij}$  may also be zero)*

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1\,n-1} & a_{1n} \\ 0 & a_{22} & \cdots & a_{2\,n-1} & a_{2n} \\ \vdots & 0 & \ddots & \vdots & \vdots \\ 0 & \vdots & \ddots & a_{n-1\,n-1} & a_{n-1\,n} \\ 0 & 0 & \cdots & 0 & a_{nn} \end{bmatrix} ;$$

- a lower triangular matrix has the form (although any of the  $a_{ij}$  may also be zero)

$$\begin{bmatrix} a_{11} & 0 & \cdots & 0 & 0 \\ a_{21} & a_{22} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ a_{n-1\ 1} & a_{n-1\ 2} & \cdots & a_{n-1\ n-1} & 0 \\ a_{n\ 1} & a_{n\ 2} & \cdots & a_{n\ n-1} & a_{nn} \end{bmatrix}.$$

Any square diagonal matrix is also an upper triangular matrix, and is also a lower triangular matrix. Thus the following theorem encompasses square diagonal matrices and so generalises [Theorem 6.1.8a](#).

**Theorem 6.2.16** (triangular matrix). *For every  $n \times n$  triangular matrix  $A$ , the determinant of  $A$  is the product of the diagonal entries,  $\det A = a_{11}a_{22} \cdots a_{nn}$ .*

**Activity 6.2.17.** Which of the following matrices is *not* a triangular matrix?

(a) 
$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 3 & 4 & 0 & 0 \\ 4 & -2 & -1 & 0 \\ -1 & -2 & 2 & -3 \end{bmatrix}$$

(b) 
$$\begin{bmatrix} -1 & -1 & 1 & 0 \\ 0 & -5 & 4 & 2 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

(c) 
$$\begin{bmatrix} 0 & 0 & 0 & -2 \\ 0 & 0 & -1 & -1 \\ 0 & 1 & 1 & 4 \\ -1 & -1 & 0 & 3 \end{bmatrix}$$

(d) 
$$\begin{bmatrix} -2 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$



**Example 6.2.18.** Find the determinant of those of the following matrices which are triangular.

(a) 
$$\begin{bmatrix} -1 & -1 & -1 & -5 \\ 0 & -4 & 1 & 4 \\ 0 & 0 & 7 & 0 \\ 0 & 0 & 0 & -3 \end{bmatrix}$$

(b) 
$$\begin{bmatrix} -3 & 0 & 0 & 0 \\ -4 & 2 & 0 & 0 \\ -1 & 1 & 1 & 0 \\ -2 & -3 & 7 & -1 \end{bmatrix}$$

(c) 
$$\begin{bmatrix} -4 & 0 & 0 & 0 \\ 2 & -2 & 0 & 0 \\ -5 & -3 & -2 & 0 \\ -2 & 5 & -2 & 0 \end{bmatrix}$$

(d) 
$$\begin{bmatrix} 0.2 & 0 & 0 & 0 \\ 0 & 1.1 & 0 & 0 \\ 0 & 0 & -0.5 & 0 \\ 0 & 0 & 0 & 0.9 \end{bmatrix}$$



$$(e) \begin{bmatrix} 1 & -1 & 1 & -3 \\ 0 & 0 & 0 & -5 \\ 0 & 0 & -3 & -4 \\ 0 & -2 & 1 & -2 \end{bmatrix}$$

$$(f) \begin{bmatrix} 0 & 0 & 0 & -3 \\ 0 & 0 & 2 & -4 \\ 0 & -1 & 4 & -1 \\ -6 & 1 & 5 & 1 \end{bmatrix}$$

$$(g) \begin{bmatrix} -1 & 0 & 0 & 1 \\ -2 & 0 & 0 & 0 \\ 2 & -2 & -1 & -2 \\ -1 & 0 & 4 & 2 \end{bmatrix}$$



The above case of triangular matrices is a short detour from the main development of this section which is to derive a formula for determinants in general. The following two examples introduce the next property we need before establishing a general formula for

determinants.

**Example 6.2.19.** Let's rewrite the explicit formulas (6.1) for  $2 \times 2$  and  $3 \times 3$  determinants explicitly as the sum of simpler determinants.

- Recall that the  $2 \times 2$  determinant

$$\begin{aligned}\begin{vmatrix} a & b \\ c & d \end{vmatrix} &= ad - bc \\ &= (ad - 0c) + (0d - bc) \\ &= \begin{vmatrix} a & 0 \\ c & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & d \end{vmatrix}.\end{aligned}$$

That is, the original determinant is the same as the sum of two determinants, each with a zero in the first row and the other row unchanged. This identity decomposes the first row as  $[a \ b] = [a \ 0] + [0 \ b]$ , while the other row is unchanged.

- Recall from (6.1) that the  $3 \times 3$  determinant

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei + bfg + cdh - ceg - afh - bdi$$

$$\begin{aligned}
&= +aei + 0fg + 0dh - 0eg - afh - 0di \\
&\quad + 0ei + bfg + 0dh - 0eg - 0fh - bdi \\
&\quad + 0ei + 0fg + cdh - ceg - 0fh - 0di \\
&= \begin{vmatrix} a & 0 & 0 \\ d & e & f \\ g & h & i \end{vmatrix} + \begin{vmatrix} 0 & b & 0 \\ d & e & f \\ g & h & i \end{vmatrix} + \begin{vmatrix} 0 & 0 & c \\ d & e & f \\ g & h & i \end{vmatrix}.
\end{aligned}$$

That is, the original determinant is the same as the sum of three determinants, each with two zeros in the first row and the other rows unchanged. This identity decomposes the first row as  $[a \ b \ c] = [a \ 0 \ 0] + [0 \ b \ 0] + [0 \ 0 \ c]$ , while the other rows are unchanged.

This sort of rearrangement of a determinant makes progress because then [Theorem 6.2.11](#) helps by finding the determinant of the resultant matrices that have an almost all zero row. ■

**Example 6.2.20.** A  $2 \times 2$  example of a more general summation property is furnished by the determinant of matrix  $A = \begin{bmatrix} a_{11} & b_1 + c_1 \\ a_{21} & b_2 + c_2 \end{bmatrix}$ .

$$\begin{aligned} \det A &= a_{11}(b_2 + c_2) - a_{21}(b_1 + c_1) \\ &= a_{11}b_2 + a_{11}c_2 - a_{21}b_1 - a_{21}c_1 \\ &= (a_{11}b_2 - a_{21}b_1) + (a_{11}c_2 - a_{21}c_1) \\ &= \det \begin{bmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{bmatrix} + \det \begin{bmatrix} a_{11} & c_1 \\ a_{21} & c_2 \end{bmatrix} \\ &= \det B + \det C, \end{aligned}$$

where matrices  $B$  and  $C$  have the same first column as  $A$ , and their second columns add up to the second column of  $A$ . ■

**Theorem 6.2.21** (sum formula). *Let  $A$ ,  $B$  and  $C$  be  $n \times n$  matrices. If matrices  $A$ ,  $B$  and  $C$  are identical except for their  $i$ th column, and that the  $i$ th column of  $A$  is the sum of the  $i$ th columns of  $B$  and  $C$ , then  $\det A = \det B + \det C$ . Further, the same sum property holds when “column” is replaced by “row” throughout.*

The sum formula [Theorem 6.2.21](#) leads to the common way to compute determinants by hand for matrices larger than  $3 \times 3$ , albeit not generally practical for matrices significantly larger.

**Example 6.2.22.** Use [Theorems 6.2.21](#) and [6.2.11](#) to evaluate the determinant of matrix

$$A = \begin{bmatrix} -2 & 1 & -1 \\ 1 & -6 & -1 \\ 2 & 1 & 0 \end{bmatrix}.$$



**Activity 6.2.23.** We could compute the determinant of the matrix  $\begin{bmatrix} -3 & 6 & -4 \\ 7 & 4 & 6 \\ 1 & 6 & -3 \end{bmatrix}$  as a particular sum involving three of the following four determinants. Which one of the following would not be used in the sum?

(a)  $\begin{vmatrix} 7 & 6 \\ 1 & -3 \end{vmatrix}$

(b)  $\begin{vmatrix} 4 & 6 \\ 6 & -3 \end{vmatrix}$

(c)  $\begin{vmatrix} 6 & -4 \\ 4 & 6 \end{vmatrix}$

(d)  $\begin{vmatrix} 6 & -4 \\ 6 & -3 \end{vmatrix}$



**Theorem 6.2.24** (Laplace expansion theorem). *For every  $n \times n$  matrix  $A = [a_{ij}]$  ( $n \geq 2$ ), recall the  $(i, j)$ th minor  $A_{ij}$  to be the  $(n-1) \times (n-1)$  matrix obtained from  $A$  by omitting the  $i$ th row and  $j$ th column. Then the determinant of  $A$  can be computed via expansion in any row  $i$  or any column  $j$  as, respectively,*

$$\begin{aligned}\det A &= (-1)^{i+1}a_{i1} \det A_{i1} + (-1)^{i+2}a_{i2} \det A_{i2} \\ &\quad + \cdots + (-1)^{i+n}a_{in} \det A_{in} \\ &= (-1)^{j+1}a_{1j} \det A_{1j} + (-1)^{j+2}a_{2j} \det A_{2j} \\ &\quad + \cdots + (-1)^{j+n}a_{nj} \det A_{nj}.\end{aligned}\tag{6.4}$$

**Example 6.2.25.** Use the Laplace expansion (6.4) to find the determinant of the following matrices.

$$(a) \begin{bmatrix} 0 & 2 & 1 & 2 \\ -1 & 2 & -1 & -2 \\ 1 & 2 & -1 & -1 \\ 0 & -1 & -1 & 1 \end{bmatrix}$$

$$(b) \begin{bmatrix} -3 & -1 & 1 & 0 \\ -2 & 0 & -2 & 0 \\ -3 & -2 & 0 & 0 \\ 1 & -2 & 0 & 3 \end{bmatrix}$$



The Laplace expansion is generally too computationally expensive for all but small matrices. The reason is that computing the determinant of an  $n \times n$  matrix with the Laplace expansion generally takes  $n!$  operations (the next [Theorem 6.2.27](#)), and the factorial  $n! = n(n-1) \cdots 3 \cdot 2 \cdot 1$  grows very quickly even for medium  $n$ . Even for just a  $20 \times 20$  matrix the Laplace expansion has over two quintillion terms ( $2 \cdot 10^{18}$ ). Exceptional matrices are those with lots of zeros, such as triangular matrices ([Theorem 6.2.16](#)). In any case, remember that except for theoretical purposes there is rarely any need to compute a medium to large determinant.

**Example 6.2.26.** The determinant of a  $3 \times 3$  matrix has  $3! = 6$  terms, each a product of three factors: diagram (6.1) gives the determinant

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei + bfg + cdh - ceg - afh - bdi.$$

Further, observe that within each term the factors come from different rows and columns. For example,  $a$  never appears in a term with the entries  $b$ ,  $c$ ,  $d$  or  $g$  (the elements from either the same row or the same column). Similarly,  $f$  never appears in a term with the entries  $d$ ,  $e$ ,  $c$  or  $i$ . ■

**Theorem 6.2.27.** *The determinant of every  $n \times n$  matrix expands to the sum of  $n!$  terms, where each term is  $\pm 1$  times a product of  $n$  factors such that each factor comes from different rows and columns of the matrix.*



---

## 7 Eigenvalues and eigenvectors in general

---

### Chapter Contents

7.1	Find eigenvalues and eigenvectors of matrices . . . .	673
7.1.1	A characteristic equation gives eigenvalues . .	675
7.1.2	Repeated eigenvalues are sensitive . . . . .	689
7.1.3	Application: discrete dynamics of populations	694
7.1.4	Extension: SVDs connect to eigen-problems .	709
7.1.5	Application: Exponential interpolation discovers dynamics . . . . .	711
7.2	Linear independent vectors may form a basis . . . .	725

7.2.1	Linearly (in)dependent sets . . . . .	728
7.2.2	Form a basis for subspaces . . . . .	742
7.3	Diagonalisation identifies the transformation . . . . .	761
7.3.1	Solve systems of differential equations . . . . .	774

**Population modelling** Suppose two species of animals interact: how do their populations evolve in time? Let  $y(t)$  and  $z(t)$  be the number of female animals in each of the species at time  $t$  in years (biologists usually just count females in population models as females usually determine reproduction). Modelling might deduce the populations interact according to the rule that the population one year later is  $y(t+1) = 2y(t) - 4z(t)$  and  $z(t+1) = -y(t) + 2z(t)$ : that is, if it was not for the other species, then for each species the number of females would both double every year (since then  $y(t+1) = 2y(t)$  and  $z(t+1) = 2z(t)$ ); but the other species decreases each of these growths via the  $-4z(t)$  and  $-y(t)$  terms.

Question: can we find special solutions in the form  $(y, z) = \mathbf{x}\lambda^t$  for

some constant  $\lambda$ ? Let's try by substituting  $y = x_1\lambda^t$  and  $z = x_2\lambda^t$  into the equations:

$$\begin{aligned}y(t+1) &= 2y(t) - 4z(t), & z(t+1) &= -y(t) + 2z(t) \\ \iff x_1\lambda^{t+1} &= 2x_1\lambda^t - 4x_2\lambda^t, & x_2\lambda^{t+1} &= -x_1\lambda^t + 2x_2\lambda^t \\ \iff 2x_1 - 4x_2 &= \lambda x_1, & -x_1 + 2x_2 &= \lambda x_2\end{aligned}$$

after dividing by the factor  $\lambda^t$  (assuming constant  $\lambda$  is non-zero). Then form these last two equations as the matrix-vector equation

$$\begin{bmatrix} 2 & -4 \\ -1 & 2 \end{bmatrix} \mathbf{x} = \lambda \mathbf{x}.$$

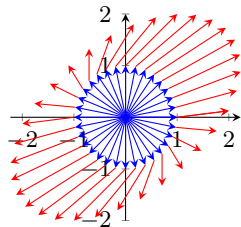
That is, this substitution  $(y, z) = \mathbf{x}\lambda^t$  shows the question about finding solutions of the population equations reduces to solving  $A\mathbf{x} = \lambda\mathbf{x}$ , called an eigen-problem.

This chapter develops linear algebra for such eigen-problems that empowers us to predict that the general solution for the population is, in terms of two constants  $c_1$  and  $c_2$ , that one species has female population  $y(t) = 2c_14^t + 2c_2$  whereas the the second species has female population  $z(t) = -c_14^t + c_2$ .

**The basic eigen-problem** Recall from [Section 4.1](#) that the eigen-problem equation  $A\mathbf{x} = \lambda\mathbf{x}$  is just asking can we find directions  $\mathbf{x}$  such that matrix  $A$  acting on  $\mathbf{x}$  is in the same direction as  $\mathbf{x}$ . That is, when is  $A\mathbf{x}$  the same as  $\lambda\mathbf{x}$  for some proportionality constant  $\lambda$ ? Now  $\mathbf{x} = \mathbf{0}$  is always a solution of the equation  $A\mathbf{x} = \lambda\mathbf{x}$ . Consequently, we are only interested in those values of the eigenvalue  $\lambda$  when non-zero solutions for the eigenvector  $\mathbf{x}$  exist (as it is the directions which are of interest). Rearranging the equation  $A\mathbf{x} = \lambda\mathbf{x}$  as the homogeneous system  $(A - \lambda I)\mathbf{x} = \mathbf{0}$ , let's invoke properties of linear equations to solve the eigen-problem.

- [Procedure 4.1.23](#) establishes that one way to find the eigenvalues  $\lambda$  (albeit *only* suitable for matrices of small size) is to solve the characteristic equation  $\det(A - \lambda I) = 0$ .
- Then for each eigenvalue, solving the homogeneous system  $(A - \lambda I)\mathbf{x} = \mathbf{0}$  gives corresponding eigenvectors  $\mathbf{x}$ .
- The set of eigenvectors for a given eigenvalue forms a subspace called the eigenspace  $\mathbb{E}_\lambda$  ([Theorem 4.1.10](#)).

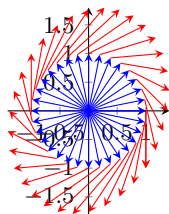
**Three general difficulties in eigen-problems** Recall that [Section 4.1](#) introduced one way to visually estimate eigenvectors and eigenvalues of a given matrix  $A$  ([Schonefeld 1995](#)). The graphical method is to plot many unit vectors  $\mathbf{x}$ , and at the end of each  $\mathbf{x}$  to adjoin the vector  $A\mathbf{x}$ . Since eigenvectors satisfy  $A\mathbf{x} = \lambda\mathbf{x}$  for some scalar eigenvalue  $\lambda$ , we visually identify eigenvectors as those  $\mathbf{x}$  which point in the same (or opposite) direction to  $A\mathbf{x}$ . Let's use this approach to identify three general difficulties.



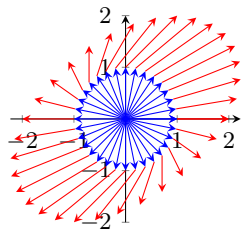
The MATLAB function `eigshow(A)` provides an interactive alternative to this static view.

1. In this first picture, for matrix  $A = \begin{bmatrix} 1 & 1 \\ \frac{1}{8} & 1 \end{bmatrix}$ , the eigenvectors appear to be in directions  $\mathbf{x}_1 \approx \pm(0.9, 0.3)$  and  $\mathbf{x}_2 \approx \pm(0.9, -0.3)$  corresponding to eigenvalues  $\lambda_1 \approx 1.4$  and  $\lambda_2 \approx 0.6$ . (Recall that scalar multiples of an eigenvector are always also eigenvectors, §4.1, so we always see  $\pm$  pairs of eigenvectors in these pictures.) The eigenvectors  $\pm(0.9, 0.3)$  are not orthogonal to the other eigenvectors  $\pm(0.9, -0.3)$ , not at right angles—as happens for symmetric matrices ([Theorem 4.2.11](#)). This lack of orthogonality in general means we soon generalise the concept of orthogonal sets of vectors to a new concept of

linearly independent sets ([Section 7.2](#)).



2. In this second case, for  $A = \begin{bmatrix} 0 & 1 \\ -1 & \frac{1}{2} \end{bmatrix}$ , there appears to be no (red) vector  $A\mathbf{x}$  in the same direction as the corresponding (blue) vector  $\mathbf{x}$ . Thus there appears to be no eigenvectors at all. No eigenvectors and eigenvalues is the answer if we require real answers. However, in most applications we find it sensible to have complex valued eigenvalues and eigenvectors ([Section 7.1](#)), written using  $i = \sqrt{-1}$ . So although we cannot see them graphically, for this matrix there are two complex eigenvalues and two families of complex eigenvectors (analogous to those found in [Example 4.1.28](#)).



3. In this third case, for  $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ , there appears to be only the vectors  $\mathbf{x} = \pm(1, 0)$ , aligned along the horizontal axis, for which  $A\mathbf{x} = \lambda\mathbf{x}$ . Whereas for symmetric matrices there were always two pairs, here we only appear to have one pair of eigenvectors ([Theorem 7.3.14](#)). Such degeneracy occurs for matrices on the border between reality and complexity.

The first problem of the general lack of orthogonality of the eigenvectors is most clearly seen in the case of triangular matrices (Definition 6.2.15). The reason is linked to Theorem 6.2.16 that the determinant of a triangular matrix is simply the product of its diagonal entries.

**Example 7.0.1.** Find by algebra the eigenvalues and eigenvectors of the triangular matrix  $A = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}$ . ■

**Theorem 7.0.2** (triangular matrices). *The diagonal entries of a triangular matrix are the only eigenvalues of the matrix. The corresponding eigenvectors of distinct eigenvalues are generally not orthogonal.*

**Example 7.0.3.** Use Theorem 7.0.2 to find the eigenvalues, corresponding eigenvectors, and corresponding eigenspaces, of the following triangular matrices.

$$(a) \quad A = \begin{bmatrix} -3 & 2 & 0 \\ 0 & -4 & 2 \\ 0 & 0 & 4 \end{bmatrix}$$

$$(b) \quad B = \begin{bmatrix} 3 & 0 & 0 & 0 \\ -2 & -4 & 0 & 0 \\ -3 & 1 & 0 & 0 \\ 0 & 0 & -3 & 1 \end{bmatrix}$$

$$(c) \quad C = \begin{bmatrix} -1 & 1 & -8 & -5 & 5 \\ -3 & 6 & 4 & -3 & 0 \\ 1 & -3 & 1 & 0 & 0 \\ -7 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

**Activity 7.0.4.** What are all the eigenvalues of the matrix

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 2 & 2 & 1 & 0 & 0 \\ 3 & 3 & 1 & 0 & 0 \\ 2 & 2 & 2 & 0 & 1 \end{bmatrix} ?$$



- (a) 0, 1, 2      (b) 0, 1, 2, 3      (c) 0, 1      (d) 1



One consequence of the second part of the proof of [Theorem 7.0.2](#) is that, when counted according to multiplicity, there are precisely  $n$  eigenvalues of an  $n \times n$  triangular matrix. Correspondingly, the next [Section 7.1](#) establishes there are precisely  $n$  eigenvalues of general  $n \times n$  matrices, provided we count the eigenvalues according to multiplicity and allow complex eigenvalues.

## 7.1 Find eigenvalues and eigenvectors of matrices

### Section Contents

7.1.1	A characteristic equation gives eigenvalues . .	675
7.1.2	Repeated eigenvalues are sensitive . . . . .	689
7.1.3	Application: discrete dynamics of populations	694
7.1.4	Extension: SVDs connect to eigen-problems .	709
7.1.5	Application: Exponential interpolation discovers dynamics . . . . .	711
	Generalised eigen-problem . . . . .	715
	General fitting of exponentials . . . . .	718

Given the additional determinant methods of [Chapter 6](#), this section begins exploring the properties and some applications of the eigen-problem  $A\mathbf{x} = \lambda\mathbf{x}$  for general matrices  $A$ . We establish that there are generally  $n$  eigenvalues of an  $n \times n$  matrix, albeit possibly complex valued, and that repeated eigenvalues are sensitive to

errors. Applications include population modelling, connecting to the computation of SVDs, and fitting exponentials to real data.

### 7.1.1 A characteristic equation gives eigenvalues

The Fundamental Theorem of Algebra asserts that every polynomial equation over the complex field has a root. It is almost beneath the dignity of such a majestic theorem to mention that in fact it has precisely  $n$  roots.

*J. H. Wilkinson, 1984 ([Higham 1996](#), p.103)*

Recall that eigenvalues  $\lambda$  and non-zero eigenvectors  $\mathbf{x}$  of a square matrix  $A$  must satisfy  $(A - \lambda I)\mathbf{x} = \mathbf{0}$ . [Theorem 6.1.29](#) then implies the eigenvalues of a square matrix are precisely the solutions of the **characteristic equation**  $\det(A - \lambda I) = 0$ .

**Theorem 7.1.1.** *For every  $n \times n$  square matrix  $A$  we call  $\det(A - \lambda I)$  the **characteristic polynomial** of  $A$ :*

- *the characteristic polynomial of  $A$  is a polynomial of  $n$ th degree in  $\lambda$ ;*
- *there are at most  $n$  distinct eigenvalues of  $A$ .*

**Activity 7.1.2.** A given matrix has eigenvalues of  $-7$ ,  $-1$ ,  $3$ ,  $4$  and  $6$ . The matrix must be of size  $n \times n$  for  $n$  at least which of the following? (Select the smallest valid answer.)

(a) 5

(b) 4

(c) 7

(d) 6



**Example 7.1.3.** Find the characteristic polynomial of each of the following matrices. Where in the coefficients of the polynomial can you see the determinant? and the sum of the diagonal elements?

(a)  $A = \begin{bmatrix} 1 & -1 \\ -2 & 4 \end{bmatrix}$

(b)  $B = \begin{bmatrix} 4 & -2 & 1 \\ 1 & -2 & 0 \\ 8 & 2 & 6 \end{bmatrix}$



These observations about the coefficients in the characteristic polynomials leads to the next theorem.

**Theorem 7.1.4.** *For every  $n \times n$  matrix  $A$ , the product of the eigenvalues equals  $\det A$  and equals the constant term in the characteristic polynomial. The sum of the eigenvalues equals  $(-1)^{n-1}$  times the coefficient of  $\lambda^{n-1}$  in the characteristic polynomial and equals the **trace** of the matrix, defined as the sum of the diagonal elements  $a_{11} + a_{22} + \cdots + a_{nn}$ .*

*This optional theorem helps establish the nature of a characteristic polynomial.*

**Activity 7.1.5.** What is the trace of the matrix

$$\begin{bmatrix} 4 & 5 & -4 & 3 \\ -2 & 2 & -5 & -1 \\ -1 & 2 & 2 & -6 \\ -13 & 4 & 3 & -1 \end{bmatrix}?$$

(a) 8

(b) 7

(c) -13

(d) -12



**Example 7.1.6.** (a) What are the two highest order terms and the constant term in the characteristic polynomial of the matrix

$$A = \begin{bmatrix} -2 & -1 & 3 & -2 \\ -1 & 3 & -2 & 2 \\ 2 & -3 & 0 & 1 \\ 0 & 1 & 0 & -3 \end{bmatrix}.$$

(b) After laborious calculation you find the characteristic polynomial of the matrix

$$B = \begin{bmatrix} -2 & 5 & -3 & -1 & 2 \\ -2 & -5 & -1 & -1 & 3 \\ 1 & 4 & -2 & 1 & -7 \\ 1 & -5 & 1 & 4 & -5 \\ -1 & 0 & 3 & -3 & 1 \end{bmatrix}$$

is  $-\lambda^5 + 2\lambda^4 - 3\lambda^3 + 234\lambda^2 + 884\lambda + 1564$ . Could this polynomial be correct?

(c) After much calculation you find the characteristic polynomial

of the matrix

$$C = \begin{bmatrix} 0 & 0 & 3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & -4 & 0 & 3 & 0 \\ -5 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & -6 & 0 & 0 \end{bmatrix}$$

is  $\lambda^6 + 4\lambda^5 + 5\lambda^4 + 20\lambda^3 + 108\lambda^2 - 540\lambda + 668$ . Could this polynomial be correct?

- (d) What are the two highest order terms and the constant term in the characteristic polynomial of the matrix

$$D = \begin{bmatrix} 0 & 4 & 0 & 0 & 3 & 0 \\ -2 & 0 & 0 & 1 & 0 & -2 \\ 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & -5 & 0 & -4 & 3 \\ 0 & 2 & -3 & 0 & -4 & 0 \\ 0 & -3 & 0 & 0 & 0 & 0 \end{bmatrix}.$$





Recall that an important characteristic of an eigenvalue is their multiplicity. The following definition of *multiplicity* generalises to all matrices the somewhat different [Definition 4.1.15](#) that applies to only symmetric matrices. For symmetric matrices the definitions are equivalent.

**Definition 7.1.7.** *An eigenvalue  $\lambda_0$  of a matrix  $A$  is said to have **multiplicity**  $m$  if the characteristic polynomial factorises to  $\det(A - \lambda I) = (\lambda - \lambda_0)^m g(\lambda)$  with  $g(\lambda_0) \neq 0$ , and  $g(\lambda)$  is a polynomial of degree  $n - m$ . Every eigenvalue of multiplicity  $m \geq 2$  is also called a **repeated eigenvalue**.*

**Activity 7.1.8.** A given matrix  $A$  has characteristic polynomial  $\det(A - \lambda I) = (\lambda + 2)\lambda^2(\lambda - 2)^3(\lambda - 3)^4$ . The eigenvalue 2 has what multiplicity?

- (a) three      (b) four      (c) one      (d) two



**Example 7.1.9.** Use the characteristic polynomials for each of the following matrices to find all eigenvalues and their multiplicity.

(a)  $A = \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$

(b)  $B = \begin{bmatrix} -1 & 1 & -2 \\ -1 & 0 & -1 \\ 0 & -3 & 1 \end{bmatrix}$

(c)  $C = \begin{bmatrix} -1 & 0 & -2 \\ 0 & -3 & 2 \\ 0 & -2 & 1 \end{bmatrix}$

$$(d) \quad D = \begin{bmatrix} 2 & 0 & -1 \\ -5 & 3 & -5 \\ 5 & -2 & -2 \end{bmatrix}$$

$$(e) \quad E = \begin{bmatrix} 0 & 1 \\ -1 & 1 \end{bmatrix}$$



**Example 7.1.10.** Use `eig()` in MATLAB/Octave to find the eigenvalues and their multiplicity for the following matrices. Recall (Table 4.1) that executing just `eig(A)` gives a column vector of eigenvalues of  $A$ , repeated according to their multiplicity.

$$(a) \begin{bmatrix} 2 & 2 & -1 \\ 0 & 1 & -2 \\ 0 & -1 & 0 \end{bmatrix}$$

$$(b) \begin{bmatrix} -2 & -2 & -5 & 0 \\ 0 & -2 & 2 & 1 \\ -1 & 1 & 0 & -1 \\ -2 & 1 & 4 & 0 \end{bmatrix}$$

$$(c) \begin{bmatrix} 3 & -1 & -2 & 1 & -2 \\ 0 & 0 & -2 & -2 & 0 \\ 2 & 1 & 1 & 1 & -1 \\ -1 & -3 & 0 & 1 & 2 \\ 2 & -2 & 1 & 0 & 3 \end{bmatrix}$$

$$(d) \begin{bmatrix} -1 & 0 & 0 & 0 \\ -1 & 2 & -3 & 3 \\ 3 & 1 & -1 & 0 \\ 0 & 3 & -2 & 1 \end{bmatrix}$$



To find eigenvalues and eigenvectors, the following restates [Procedure 4.1.23](#) with a little more information, and now empowered to address larger matrices upon using the determinant tools from [Chapter 6](#).

**Procedure 7.1.11** (eigenvalues and eigenvectors). *To find by hand eigenvalues and eigenvectors of a (small) square matrix  $A$ :*

1. *find all eigenvalues (possibly complex) by solving the **characteristic equation** of  $A$ ,  $\det(A - \lambda I) = 0$ ;*
2. *for each eigenvalue  $\lambda$ , solve the homogeneous linear equation  $(A - \lambda I)\mathbf{x} = \mathbf{0}$  to find the eigenspace  $\mathbb{E}_\lambda$  of all eigenvectors (together with  $\mathbf{0}$ );*

3. write each eigenspace as the span of a few chosen eigenvectors (*Definition 7.2.20* calls such a set a basis).

Since, for an  $n \times n$  matrix, the characteristic polynomial is of  $n$ th degree in  $\lambda$  (*Theorem 7.1.1*), there are  $n$  eigenvalues (when counted according to multiplicity and allowing complex eigenvalues).

Correspondingly, the following restates the computational procedure of [Section 4.1.1](#), but slightly more generally: the extra generality caters for non-symmetric matrices.

**Compute in Matlab/Octave.**

For a given square matrix  $A$ , execute `[V,D]=eig(A)`, then the diagonal entries of  $D$ , `diag(D)`, are the eigenvalues of  $A$ . Corresponding to the eigenvalue  $D(j,j)$  is an eigenvector  $v_j = V(:,j)$ , the  $j$ th column of  $V$ . If an eigenvalue is repeated in the diagonal of  $D$  (multiplicity more than one), then the corresponding columns of  $V$  span the eigenspace (and, as [Section 7.2](#) discusses, when the column vectors have a property called linear independence then they form a so-called basis for the eigenspace).

**Activity 7.1.12.** For the matrix  $A = \begin{bmatrix} 2 & 0 & -1 \\ -5 & 3 & -5 \\ 5 & -2 & -2 \end{bmatrix}$ , which one of the following vectors satisfy  $(A - 3I)\mathbf{x} = \mathbf{0}$  and hence is an eigenvector of  $A$  corresponding to eigenvalue 3?

(a)  $\mathbf{x} = (-1, 0, 1)$

(b)  $\mathbf{x} = (0, 1, 0)$

(c)  $\mathbf{x} = (1, 5, 5)$

(d)  $\mathbf{x} = (1, 5, -1)$



**Example 7.1.13.** Find the eigenspaces corresponding to the eigenvalues found for the first three matrices of [Example 7.1.9](#).

7.1.9a.  $A = \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$

7.1.9b.  $B = \begin{bmatrix} -1 & 1 & -2 \\ -1 & 0 & -1 \\ 0 & -3 & 1 \end{bmatrix}$

$$7.1.9c. \quad C = \begin{bmatrix} -1 & 0 & -2 \\ 0 & -3 & 2 \\ 0 & -2 & 1 \end{bmatrix}$$



The matrices in [Example 7.1.13](#) all have repeated eigenvalues. For these repeated eigenvalues the corresponding eigenspaces happen to be all one dimensional. This contrasts with the case of symmetric matrices where the eigenspaces always have the same dimensionality as the multiplicity of the eigenvalue, as illustrated by [Examples 4.1.14](#) and [4.1.20](#). Subsequent sections work towards [Theorem 7.3.14](#) which establishes that for non-symmetric matrices an eigenspace has dimensionality between one and the multiplicity of the corresponding eigenvalue.



**Example 7.1.14.** By hand, find the eigenvalues and eigenspaces of the matrix

$$A = \begin{bmatrix} 0 & 3 & 0 & 0 & 0 \\ 1 & 0 & 3 & 0 & 0 \\ 0 & 1 & 0 & 3 & 0 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

([Example 7.1.15](#) confirms the answer using `eig()` in MATLAB/Octave.) ■

**Example 7.1.15.** Use MATLAB/Octave to confirm the eigenvalues and eigenvectors found for the matrix of [Example 7.1.14](#). ■

### 7.1.2 Repeated eigenvalues are sensitive

This optional subsection does not prove the sensitivity: it uses examples to introduce and illustrate.

Albeit hidden in [Example 7.1.10](#), repeated eigenvalues are exquisitely sensitive to errors in either the matrix or the computation. If the matrix or the computation has an error  $\epsilon$ , then expect a repeated eigenvalue of multiplicity  $m$  to appear as  $m$  eigenvalues all within about  $\epsilon^{1/m}$  of each other. Consequently, when we find or compute  $m$  eigenvalues all within about  $\epsilon^{1/m}$ , then *suspect* them to be one eigenvalue of multiplicity  $m$ .

**Example 7.1.16.** Explore the eigenvalues of the matrix  $A = \begin{bmatrix} a & 1 \\ 0.0001 & a \end{bmatrix}$  for every parameter  $a$ . ■

Further, since computers work to a relative error of about  $10^{-15}$ , then expect a repeated eigenvalue of multiplicity  $m$  to appear as  $m$  eigenvalues within about  $10^{-15/m}$  of each other—even when there are no experimental errors in the matrix. Repeat some of the previous cases of [Example 7.1.10](#), preceded by the MATLAB/Octave command `format long`, to see that the repeated eigenvalues are

sensitive to computational errors.

**Example 7.1.17.** Use MATLAB/Octave to compute eigenvalues of the following matrices and comment on the effect on repeated eigenvalues of errors in the matrix and/or the computation.

$$(a) \ B = \begin{bmatrix} 3 & 0 & -2 & 0 & 0 \\ -1 & 5 & 0 & -1 & 3 \\ -1 & 2 & 4 & 0 & 1 \\ 5 & -1 & 4 & 1 & -1 \\ 3 & 2 & 1 & -2 & 2 \end{bmatrix}$$

- (b) Suppose the above matrix  $B$  is obtained from some experiment where there are experimental errors in the entries with error about 0.0001. Randomly perturb the entries in matrix  $B$  to see the effects of such errors on the eigenvalues (use `randn()`, Table 3.1).

$$(c) \ C = \begin{bmatrix} -1 & 0 & 0 & 0 \\ -1 & 2 & -3 & 3 \\ 3 & 1 & -1 & 0 \\ 0 & 3 & -2 & 1 \end{bmatrix} \text{ perturbed by errors of size } 10^{-6}$$

**Activity 7.1.18.** In an experiment measurements are made to three decimal place accuracy. Then in analysing the results, a  $5 \times 5$  matrix is formed from the measurements, and its eigenvalues computed by MATLAB/Octave to be

$$-0.9851, \quad 0.1266, \quad 0.9954, \quad 1.0090, \quad 1.0850.$$

What should you suspect is the number of different eigenvalues?

- (a) four      (b) two      (c) three      (d) five

**But symmetric matrices are OK** The eigenvalues of a symmetric matrix are not so sensitive. This lack of sensitivity is fortunate as many applications give rise to symmetric matrices ([Chapter 4](#)). Such symmetry often reflects some symmetry in the

natural world such as Newton's law of every action having an equal and opposite reaction. For symmetric matrices, the eigenvalues and eigenvectors are reasonably robust to both computational perturbations and experimental errors.

**Example 7.1.19.** For perhaps the simplest example, consider the  $2 \times 2$  symmetric matrix  $A = \begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}$ . Being diagonal, matrix  $A$  has eigenvalue  $\lambda = a$  (multiplicity two). Now perturb the matrix by 'experimental' error to  $B = \begin{bmatrix} a & 10^{-4} \\ 10^{-4} & a \end{bmatrix}$ . The characteristic equation of  $B$  is

$$\det(B - \lambda I) = (a - \lambda)^2 - 10^{-8} = 0.$$

Rearrange this equation to  $(\lambda - a)^2 = 10^{-8}$ . Taking square roots gives  $\lambda - a = \pm 10^{-4}$ , that is, the eigenvalues of  $B$  are  $\lambda = a \pm 10^{-4}$ . Because a perturbation to the symmetric matrix of size  $10^{-4}$  only changes the eigenvalues by a similar amount, the eigenvalues are *not* sensitive. ■

**Activity 7.1.20.** What are the eigenvalues of matrix  $\begin{bmatrix} a & 0.01 \\ -0.01 & a \end{bmatrix}$ ?

- (a)  $a \pm 0.1$       (b)  $a \pm 0.01$       (c)  $a \pm 0.01 i$       (d)  $a \pm 0.1 i$



**Example 7.1.21.** Compute the eigenvalues of the symmetric matrix

$$A = \begin{bmatrix} 1 & 1 & 0 & 2 \\ 1 & 0 & 2 & -1 \\ 0 & 2 & 1 & 4 \\ 2 & -1 & 4 & 1 \end{bmatrix}$$

and see matrix  $A$  has an eigenvalue of multiplicity two. Explore the effects on the eigenvalues of errors in the matrix by perturbing the entries by random amounts of size 0.0001.



### 7.1.3 Application: discrete dynamics of populations

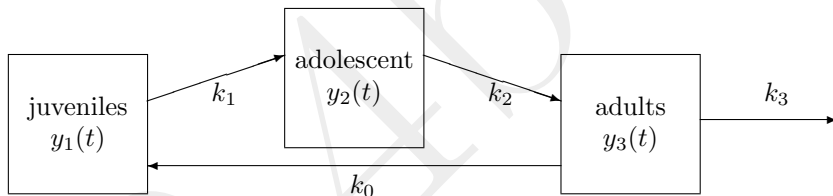
Age structured populations are one case where matrix properties and methods are crucial. The approach of this section is also akin to much mathematical modelling of diseases and epidemics. This section aims to show how to derive and use a matrix-vector model for the change in time  $t$  of interesting properties  $\mathbf{y}$  of a population. Specifically, this subsection derives and analyses the model  $\mathbf{y}(t+1) = A\mathbf{y}(t)$ .

For a given species, let's define

- $y_1(t)$  to be the number of juveniles (including infants),
- $y_2(t)$  the number of adolescents, and
- $y_3(t)$  the number of adults.

Mostly, biologists only count females as females are the determining sex for reproduction. (Although some bacteria/algae have seven sexes!) How do these numbers of females evolve over time? from generation to generation? First we need to choose a basic time interval (the unit of time): it could be one year, one month, one

day, or maybe six months. Whatever we choose as convenient, we then quantify the number of events that happen to the females in each time interval as shown schematically in the diagram below:



Over any one time interval, and only counting females:

- a fraction  $k_1$  of the juveniles become adolescents;
- a fraction  $k_2$  of the adolescents become adults;
- a fraction  $k_3$  of the adults die;
- but adults also give birth to juveniles at rate  $k_0$  per adult.

Model this scenario with a system of discrete dynamical equations which are of the form that the numbers at the next time,  $t + 1$ ,



depend upon the numbers at the time  $t$ :

$$y_1(t+1) = \cdots ,$$

$$y_2(t+1) = \cdots ,$$

$$y_3(t+1) = \cdots .$$

Let's fill in the right-hand sides from the given information about the rate of particular events per time interval.

- A fraction  $k_1$  of the juveniles  $y_1(t)$  becoming adolescents also means a fraction  $(1 - k_1)$  of the juveniles remain juveniles, hence

$$y_1(t+1) = (1 - k_1)y_1(t) + \cdots ,$$

$$y_2(t+1) = +k_1y_1(t) + \cdots ,$$

$$y_3(t+1) = \cdots .$$

- A fraction  $k_2$  of the adolescents  $y_2(t)$  becoming adults also means a fraction  $(1 - k_2)$  of the adolescents remain adolescents, hence additionally

$$y_1(t+1) = (1 - k_1)y_1(t) + \cdots ,$$

$$y_2(t+1) = +k_1y_1(t) + (1 - k_2)y_2(t),$$

$$y_3(t+1) = +k_2y_2(t) + \cdots .$$

- A fraction  $k_3$  of the adults die mean that a fraction  $(1 - k_3)$  of the adults remain adults, hence

$$y_1(t+1) = (1 - k_1)y_1(t) + \cdots ,$$

$$y_2(t+1) = +k_1y_1(t) + (1 - k_2)y_2(t),$$

$$y_3(t+1) = +k_2y_2(t) + (1 - k_3)y_3(t).$$

- But adults also give birth to juveniles at rate  $k_0$  per adult so the number of juveniles increases by  $k_0y_3$  from births:

$$y_1(t+1) = (1 - k_1)y_1(t) + k_0y_3(t),$$

$$y_2(t+1) = +k_1y_1(t) + (1 - k_2)y_2(t),$$

$$y_3(t+1) = +k_2y_2(t) + (1 - k_3)y_3(t).$$

This is our mathematical model of the age structure of the population.

Finally, write the mathematical model as the matrix-vector system

$$\begin{bmatrix} y_1(t+1) \\ y_2(t+1) \\ y_3(t+1) \end{bmatrix} = \begin{bmatrix} 1-k_1 & 0 & k_0 \\ k_1 & 1-k_2 & 0 \\ 0 & k_2 & 1-k_3 \end{bmatrix} \begin{bmatrix} y_1(t) \\ y_2(t) \\ y_3(t) \end{bmatrix},$$

that is,  $\mathbf{y}(t+1) = A\mathbf{y}(t)$ . Such a model empowers predictions.

**Example 7.1.22** (orangutans). From the following extract of the Wikipedia entry on orangutans [20 Mar 2014] derive a mathematical model for the age structure of the orangutans from one year to the next.



Gestation lasts for 9 months, with females giving birth to their first offspring between the ages of 14 and 15 years. Female orangutans have [seven to] eight-year intervals between births, the longest interbirth intervals among the great apes. ... Infant orangutans are completely dependent on their mothers for the first two years of their lives. The mother will carry the infant during travelling, as well as feed it and sleep with it in the same night nest. For the first four months, the infant is carried on its belly and never relieves physical

contact. In the following months, the time an infant spends with its mother decreases. When an orangutan reaches the age of two, its climbing skills improve and it will travel through the canopy holding hands with other orangutans, a behaviour known as “buddy travel”. Orangutans are juveniles from about two to five years of age and will start to temporarily move away from their mothers. Juveniles are usually weaned at about four years of age. Adolescent orangutans will socialize with their peers while still having contact with their mothers. Typically, orangutans live over 30 years in both the wild and captivity.

Suppose the initial population of orangutans in some area at year zero of a study is that of 30 adolescent females and 15 adult females. Use the mathematical model to predict the population for the next five years.



The mathematical model  $\mathbf{y}(t+1) = A\mathbf{y}(t)$  does predict/forecast the future populations. However, to make predictions for many years and for general initial populations we prefer the formula solution given by the upcoming [Theorem 7.1.25](#) and introduced in the next example.

**Example 7.1.23.** A vector  $\mathbf{y}(t) \in \mathbb{R}^2$  changes with time  $t$  according to the model

$$\mathbf{y}(t+1) = A\mathbf{y}(t) = \begin{bmatrix} 1 & -1 \\ -4 & 1 \end{bmatrix} \mathbf{y}(t).$$

First, what is  $\mathbf{y}(3)$  if the initial value  $\mathbf{y}(0) = (0, 1)$ ? Second, find a general formula for  $\mathbf{y}(t)$  from every initial  $\mathbf{y}(0)$ . ■

**Activity 7.1.24.** For [Example 7.1.23](#), what is the particular solution when  $\mathbf{y}(0) = (1, 1)$ ?

(a)  $\mathbf{y} = -\frac{1}{4} \cdot (-1)^t(1, 2) + \frac{3}{4} \cdot 3^t(-1, 2)$

(b)  $\mathbf{y} = 4 \cdot 3^t(-1, 2)$

(c)  $\mathbf{y} = \frac{3}{4} \cdot (-1)^t(1, 2) - \frac{1}{4} \cdot 3^t(-1, 2)$

$$(d) \mathbf{y} = \frac{3}{4} \cdot (-1)^t(-1, 2) - \frac{1}{4} \cdot 3^t(1, 2)$$



Now we establish that the same sort of general solution occurs for all such models.

**Theorem 7.1.25.** *Suppose the  $n \times n$  square matrix  $A$  governs the dynamics of  $\mathbf{y}(t) \in \mathbb{R}^n$  according to  $\mathbf{y}(t+1) = A\mathbf{y}(t)$ .*

(a) *Let  $\lambda_1, \lambda_2, \dots, \lambda_m$  be eigenvalues of  $A$  and  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$  be corresponding eigenvectors, then a solution of  $\mathbf{y}(t+1) = A\mathbf{y}(t)$  is the linear combination*

$$\mathbf{y}(t) = c_1 \lambda_1^t \mathbf{v}_1 + c_2 \lambda_2^t \mathbf{v}_2 + \cdots + c_m \lambda_m^t \mathbf{v}_m \quad (7.1)$$

*for all constants  $c_1, c_2, \dots, c_m$ .*

(b) *Further, if the matrix of eigenvectors  $P = [\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_m]$  is invertible, then the general linear combination (7.1) is a **general solution** in that unique constants  $c_1, c_2, \dots, c_m$  may be found for every given initial value  $\mathbf{y}(0)$ .*

**Activity 7.1.26.** The matrix  $A = \begin{bmatrix} 1 & 1 \\ a^2 & 1 \end{bmatrix}$  has eigenvectors  $(1, a)$  and  $(1, -a)$ . For what value(s) of  $a$  does [Theorem 7.1.25](#) *not* provide a general solution to  $\mathbf{y}(t+1) = A\mathbf{y}(t)$ ?

- (a)  $a = \pm 1$       (b)  $a = -1$       (c)  $a = 1$       (d)  $a = 0$



**Example 7.1.27.** Consider the dynamics of  $\mathbf{y}(t+1) = A\mathbf{y}(t)$  for matrix  $A = \begin{bmatrix} 1 & 3 \\ -1 & 1 \end{bmatrix}$ . First, what is  $\mathbf{y}(3)$  when the initial value  $\mathbf{y}(0) = (1, 0)$ ? Second, find a general solution.



One crucial qualitative aspect we need to know is whether components in the solution [\(7.1\)](#) grow, decay, or stay the same size as time increases. The growth or decay is determined by the eigenvalues: the reason is that  $\lambda_j^t$  is the only place that the time appears in the formula [\(7.1\)](#).

- For example, in the general solution for [Example 7.1.23](#),  $\mathbf{y}(t) = c_1(-1)^t(1, 2) + c_23^t(-1, 2)$ , the  $3^t$  factor grows in time since  $3^1 = 3$ ,  $3^2 = 9$ ,  $3^3 = 27$ , and so on. Whereas the  $(-1)^t$  factor just oscillates in time since  $(-1)^1 = -1$ ,  $(-1)^2 = 1$ ,  $(-1)^3 = -1$ , and so on. Thus for long times, large  $t$ , we know that the term involving the factor  $3^t$  will dominate the solution as it grows.
- In [Example 7.1.27](#) with complex conjugate eigenvalues the situation is more complicated. Let's write every given complex eigenvalue in polar form  $\lambda = r(\cos \theta + i \sin \theta)$  where magnitude  $r = |\lambda|$  and angle  $\theta$  is such that  $\tan \theta = (\Im \lambda)/(\Re \lambda)$ . For example,  $1 - i$  has magnitude  $r = |1 - i| = \sqrt{1^2 + (-1)^2} = \sqrt{2}$  and angle  $\theta = -\frac{\pi}{4}$  since  $\tan(-\frac{\pi}{4}) = -1/1$ .

Question: how does this help understand the solution which has  $\lambda_j^t$  in it? Answer: De Moivre's theorem says that if  $\lambda = r[\cos \theta + i \sin \theta]$ , then  $\lambda^t = r^t [\cos(\theta t) + i \sin(\theta t)]$ . Since the magnitude  $|\cos(\theta t) + i \sin(\theta t)| = \sqrt{\cos^2(\theta t) + \sin^2(\theta t)} = \sqrt{1} = 1$ , the magnitude  $|\lambda^t| = r^t$ . For example, the magnitude  $|(1 - i)^2| = (\sqrt{2})^2 = 2$  which we check by computing  $(1 - i)^2 =$

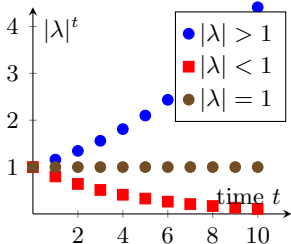


$$1^2 - 2i + i^2 = -2i \text{ and } |-2i| = 2.$$

In [Example 7.1.27](#), the eigenvalue  $\lambda_1 = 1 + i\sqrt{3}$  so its magnitude is  $r_1 = |\lambda_1| = |1 + i\sqrt{3}| = \sqrt{1+3} = 2$ . Hence the magnitude  $|\lambda_1^t| = 2^t$  at every time step  $t$ . Similarly, the magnitude  $|\lambda_2^t| = 2^t$  at every time step  $t$ . Consequently, the general solution

$$\mathbf{y}(t) = c_1 \lambda_1^t \begin{bmatrix} -i\sqrt{3} \\ 1 \end{bmatrix} + c_2 \lambda_2^t \begin{bmatrix} +i\sqrt{3} \\ 1 \end{bmatrix}$$

will grow in magnitude roughly like  $2^t$  as both components grow like  $2^t$ . It is a ‘rough’ growth because the components  $\cos(\theta t)$  and  $\sin(\theta t)$  cause ‘oscillations’ in time  $t$ . Nonetheless the overall growth like  $|\lambda_1|^t = |\lambda_2|^t = 2^t$  is inexorable—and seen previously in the particular solution where we observe  $\mathbf{y}(3)$  is eight times the magnitude of  $\mathbf{y}(0)$ .



In general, for both real or complex eigenvalues  $\lambda$ , a term involving the factor  $\lambda^t$  will, as time  $t$  increases,

- grow to infinity if  $|\lambda| > 1$ ,

- decay to zero if  $|\lambda| < 1$ , and
- remain the same magnitude if  $|\lambda| = 1$ .

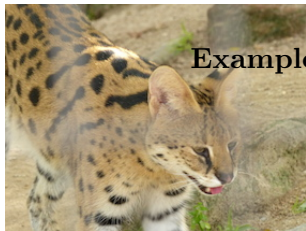
**Activity 7.1.28.** For which of the following values of  $\lambda$ , as time  $t$  increases, will  $\lambda^t$  grow in an oscillatory fashion?

- (a)  $\lambda = -0.8$       (b)  $\lambda = \frac{3}{5} + i\frac{4}{5}$       (c)  $\lambda = -\frac{4}{5} + i\frac{4}{5}$       (d)  $\lambda = 1.5$

**Example 7.1.29** (orangutans over many years). Extend the orangutan analysis of [Example 7.1.22](#). Use [Theorem 7.1.25](#) to predict the population over many years: from an initial population of 30 adolescent females and 15 adult females; and from a general initial population.

In *Guidelines for Assessment and Instruction in Mathematics Modeling Education*, [Bliss et al. \(2016\)](#) discuss mathematical modelling.

- On page 23 they comment “*Modelling (like real life) is open-ended and messy*”: in our two examples here you have to extract the important factors from many unneeded details, and use them in the context of an imperfect model.
- Also on p.23, modellers “*must be making genuine choices*”: in these problems, as in all modelling, there are choices that lead to different models—we have to operate and sensibly predict with such uncertainty.
- Lastly, they recommend to “*focus on the process, not the product*”: depending upon your choices and interpretations you will develop alternative plausible models in these scenarios—it is the process of forming plausible models and interpreting the results that are important.



**Example 7.1.30** (servals grow). The serval is a member of the cat family that lives in Africa. Given next is an extract from Wikipedia of a serval’s Reproduction and Life History.

Kittens are born shortly before the peak breeding period of local rodent populations. A serval is able to give birth

to multiple litters throughout the year, but commonly does so only if the earlier litters die shortly after birth. Gestation lasts from 66 to 77 days and commonly results in the birth of two kittens, although sometimes as few as one or as many as four have been recorded.

The kittens are born in dense vegetation or sheltered locations such as abandoned aardvark burrows. If such an ideal location is not available, a place beneath a shrub may be sufficient. The kittens weigh around 250 gm at birth, and are initially blind and helpless, with a coat of greyish woolly hair. They open their eyes at 9 to 13 days of age, and begin to take solid food after around a month. At around six months, they acquire their permanent canine teeth and begin to hunt for themselves; they leave their mother at about 12 months of age. They may reach sexual maturity from 12 to 25 months of age.

Life expectancy is about 10 years in the wild.

From the information in this extract, create a plausible, age structured, population model of servals: give reasons for estimates of the coefficients in the model. Choose three age categories of kittens, juveniles, sexually mature adults. What does the model predict over long times? Predation, disease, and food shortages are just some processes not included in this model which act to limit the serval's population in ways not included in this model. ■

Crucial in this section—so that we find a solution for all initial values—is that the matrix of eigenvectors is invertible. The next [Section 7.2](#) relates the invertibility of a matrix of eigenvectors to the new concept of ‘linear independence’ ([Theorem 7.2.41](#)).

### 7.1.4 Extension: SVDs connect to eigen-problems

This optional section connects the SVD of a general matrix to a symmetric eigen-problem, in principle.

Recall that [Chapter 4](#) starts by illustrating the close connection between the SVD of a symmetric matrix and the eigenvalues and eigenvectors of that symmetric matrix. This subsection establishes that an SVD of a general matrix is closely connected to the eigenvalues and eigenvectors of a specific matrix of double the size. The connection depends upon determinants and solving linear systems and so, in principle, is an approach to compute an SVD distinct from the inductive maximisation of [Subsection 3.3.3](#).

**Example 7.1.31.** Compute the eigenvalues and eigenvectors of the (symmetric) matrix

$$B = \begin{bmatrix} 0 & 0 & 10 & 2 \\ 0 & 0 & 5 & 11 \\ 10 & 5 & 0 & 0 \\ 2 & 11 & 0 & 0 \end{bmatrix}. \quad \text{For matrix } A = \begin{bmatrix} 10 & 2 \\ 5 & 11 \end{bmatrix},$$

compare with an SVD of  $A$ .



[Procedure 7.1.11](#) computes eigenvalues and eigenvectors by hand (in principle no matter how large the matrix). The procedure is independent of the SVD. Let's now invoke this procedure to establish another method to find an SVD distinct from the inductive maximisation of the proof in [Subsection 3.3.3](#). The following [Theorem 7.1.32](#) is a step towards an efficient numerical computation of an SVD ([Trefethen & Bau 1997](#), p.234).

**Theorem 7.1.32** (SVD as an eigenproblem). *For every real  $m \times n$  matrix  $A$ , the singular values of  $A$  are the non-negative eigenvalues of the  $(m+n) \times (m+n)$  symmetric matrix  $B = \begin{bmatrix} O_m & A \\ A^T & O_n \end{bmatrix}$ . Each corresponding eigenvector  $\mathbf{w} \in \mathbb{R}^{m+n}$  of  $B$  gives corresponding singular vectors of  $A$ , namely  $\mathbf{w} = (\mathbf{u}, \mathbf{v})$  for singular vectors  $\mathbf{u} \in \mathbb{R}^m$  and  $\mathbf{v} \in \mathbb{R}^n$ .*

### 7.1.5 Application: Exponential interpolation discovers dynamics

This optional subsection develops a method useful in many modern applied disciplines.

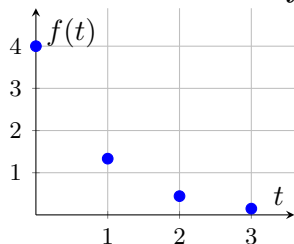
Many applications require identification of rates and frequencies (Pereyra & Scherer 2010, e.g.): as a played musical note decays, what is its frequency? in the observed vibrations of a complicated bridge, what are its natural modes? in measurements of complicated bio-chemical reactions, what rates can be identified? All such tasks require fitting a sum of exponential functions to the data.

**Example 7.1.33.** This example is the simplest case of fitting one exponential to two data points. Suppose we take two measurements of some process:

- at time  $t_1 = 1$  we measure the value  $f_1 = 5$ , and
- at time  $t_2 = 3$  we measure the value  $f_2 = 10$ .

Find an exponential fit to this data of the form  $f(t) = ce^{rt}$  for some as yet unknown coefficients  $c$  and rate  $r$ . ■



**Activity 7.1.34.**

Plotted in the margin is some points from a function  $f(t)$ . Which of the following exponentials best represents the data plotted?

(a)  $f \propto e^{-t/2}$

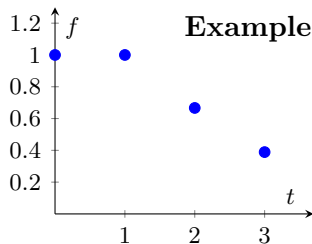
(b)  $f \propto 1/3^t$

(c)  $f \propto e^{-3t}$

(d)  $f \propto 1/2^t$



Now let's develop the approach of [Example 7.1.33](#) to the more complicated and interesting example of fitting the linear combination of two exponentials to four data points.

**Example 7.1.35.**

Suppose in some chemical or biochemical experiment you measure the concentration of a key chemical at four times (as illustrated): at the start of the experiment, time  $t_1 = 0$  you measure concentration  $f_1 = 1$  (in some units); at time  $t_2 = 1$  the measurement is  $f_2 = 1$  (again); at  $t_3 = 2$  the measurement is  $f_2 = \frac{2}{3} = 0.6667$ ; and at  $t_4 = 3$  the measurement is  $f_3 = \frac{7}{18} = 0.3889$  (as plotted in the margin). We generally expect chemical reactions

to decay exponentially in time. So our task is to find a function of the form

$$f(t) = c_1 e^{r_1 t} + c_2 e^{r_2 t}.$$

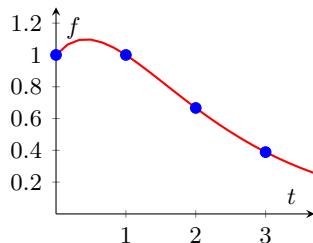
Our aim is for this function to fit the four data points, as plotted. The four unknown coefficients  $c_1$  and  $c_2$  and rates  $r_1$  and  $r_2$  need to be determined from the four data points. That is, let's find these four unknowns from the data that

$$f(0) = 1 \iff c_1 + c_2 = 1,$$

$$f(1) = 1 \iff c_1 e^{r_1} + c_2 e^{r_2} = 1,$$

$$f(2) = \frac{2}{3} \iff c_1 e^{r_1 2} + c_2 e^{r_2 2} = \frac{2}{3},$$

$$f(3) = \frac{7}{18} \iff c_1 e^{r_1 3} + c_2 e^{r_2 3} = \frac{7}{18}.$$



These are nonlinear equations, but some algebraic tricks empower us to use our beautiful linear algebra to find the solution. After solving these four nonlinear equations, we ultimately plot the function  $f(t)$  that interpolates between the data as also shown in the margin.



[Example 7.1.35](#) shows one way that fitting exponentials to data can be done with eigenvalues and eigenvectors. But one undesirable attribute of the example is the need to invert the matrix  $B$  to form matrix  $K = AB^{-1}$ . We avoid this inversion by generalising eigen-problems as introduced by the following reworking of parts of [Example 7.1.35](#).

**Example 7.1.36** (two short-cuts). Recall that [Subsection 7.1.3](#) derived general solutions of dynamic equations such as  $\mathbf{f}_{j+1} = K\mathbf{f}_j$  by seeking solutions of the form  $\mathbf{f}_j = \mathbf{v}\lambda^j$ . For the previous [Example 7.1.35](#) let's instead seek solutions of the form  $\mathbf{f}_j = B\mathbf{w}\lambda^j$ . Substituting this form, the dynamic equation  $\mathbf{f}_{j+1} = K\mathbf{f}_j$  becomes  $B\mathbf{w}\lambda^{j+1} = KB\mathbf{w}\lambda^j$ ; then factoring  $\lambda^j$ , recognising that  $KB = A$ , and swapping sides, this equation becomes  $A\mathbf{w} = \lambda B\mathbf{w}$ . This  $A\mathbf{w} = \lambda B\mathbf{w}$  forms a generalised eigen-problem because it reduces to the standard eigen-problem in cases when the matrix  $B = I$ . Rework parts of [Example 7.1.35](#) via this generalised eigen-problem.



## Generalised eigen-problem

As introduced by [Example 7.1.36](#), let's generalise the [Definition 4.1.1](#) of eigenvalues and eigenvectors. Such generalised eigen-problems also occur in the design analysis of complicated structures, such as buildings and bridges, where the second matrix  $B$  represents the various masses of the various elements making up a structure.

**Definition 7.1.37.** *Let  $A$  and  $B$  be  $n \times n$  square matrices. The **generalised eigen-problem** is to find scalar eigenvalues  $\lambda$  and corresponding nonzero eigenvectors  $\mathbf{v}$  such that  $A\mathbf{v} = \lambda B\mathbf{v}$ .*

**Example 7.1.38.** Given  $A = \begin{bmatrix} -2 & 2 \\ 3 & 1 \end{bmatrix}$  and  $B = \begin{bmatrix} 3 & -3 \\ 2 & 0 \end{bmatrix}$ , what eigenvalue corresponds to the eigenvector  $\mathbf{v}_1 = (1, 1)$  of the generalised eigen-problem  $A\mathbf{v} = \lambda B\mathbf{v}$ ? Also answer for  $\mathbf{v}_2 = (-3, 13)$ . ■

**Activity 7.1.39.** Which of the following vectors is an eigenvector of the generalised eigen-problem

$$\begin{bmatrix} -2 & 0 \\ 1 & 0 \end{bmatrix} \mathbf{v} = \lambda \begin{bmatrix} 3 & -1 \\ -1 & 1 \end{bmatrix} \mathbf{v} ?$$

- (a)  $(0, 0)$       (b)  $(1, -1)$       (c)  $(2, -1)$       (d)  $(0, 1)$



The standard eigen-problem is the case when matrix  $B = I$ . Many of the properties for standard eigenvalues and eigenvectors also hold for generalised eigen-problems, although there are some differences. Most importantly here, albeit without proof, provided matrix  $B$  is invertible, then counted according to multiplicity there are  $n$  eigenvalues of a generalised eigen-problem in  $n \times n$  matrices.

**Example 7.1.40.** Find all eigenvalues and corresponding eigenvectors of the generalised eigen-problem  $A\mathbf{v} = \lambda B\mathbf{v}$  for matrices

$$A = \begin{bmatrix} -1 & 1 \\ -4 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} -1 & 1 \\ -1 & 2 \end{bmatrix}$$

■

**Example 7.1.41.** Find all eigenvalues and corresponding eigenvectors of the generalised eigen-problem  $A\mathbf{v} = \lambda B\mathbf{v}$  for matrices

$$A = \begin{bmatrix} 2 & 2 \\ -1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$$

■

## General fitting of exponentials

Suppose that some experiment or other observation has given us  $2n$  data values  $f_1, f_2, \dots, f_{2n}$  at equi-spaced ‘times’  $t_1, t_2, \dots, t_{2n}$ , where the spacing  $t_{j+1} - t_j = h$ . The general aim is to fit the multi-exponential function (Cuyt 2015, §2.6, e.g.)

$$f(t) = c_1 e^{r_1 t} + c_2 e^{r_2 t} + \dots + c_n e^{r_n t}. \quad (7.2)$$

for some coefficients  $c_1, c_2, \dots, c_n$  and some rates  $r_1, r_2, \dots, r_n$  to be determined (possibly complex valued for oscillations). In general, finding the coefficients and rates is a delicate nonlinear task outside the remit of this book. However, as the previous two examples illustrate, in these circumstances we instead invoke our powerful linear algebra methods.

Because the data is sampled at equi-spaced times,  $h$  apart, then instead of seeking  $r_k$  we seek multipliers  $\lambda_k = e^{r_k h}$ .

**Procedure 7.1.42 (exponential interpolation).** *Given measured data  $f_1, f_2, \dots, f_{2n}$  at  $2n$  equi-spaced times  $t_1, t_2, \dots, t_{2n}$  where time  $t_j = (j-1)h$  for time-spacing  $h$  (and starting from time  $t_1 = 0$  without loss of applicability), we seek to fit the data with a sum of exponentials (7.2).*

1. *From the  $2n$  data points, form two  $n \times n$  (symmetric) Hankel matrices*

$$A = \begin{bmatrix} f_2 & f_3 & \cdots & f_{n+1} \\ f_3 & f_4 & \cdots & f_{n+2} \\ \vdots & \vdots & & \vdots \\ f_{n+1} & f_{n+2} & \cdots & f_{2n} \end{bmatrix},$$

$$B = \begin{bmatrix} f_1 & f_2 & \cdots & f_n \\ f_2 & f_3 & \cdots & f_{n+1} \\ \vdots & \vdots & & \vdots \\ f_n & f_{n+1} & \cdots & f_{2n-1} \end{bmatrix}.$$

*In MATLAB/Octave  $A = \text{hankel}(f(2:n+1), f(n+1:2*n))$  and  $B = \text{hankel}(f(1:n), f(n:2*n-1))$  forms these two matrices*



(this Hankel function is also invoked in exploring El Nino, Example 3.4.27).

2. Find the eigenvalues of the generalised eigen-problem  $A\mathbf{v} = \lambda B\mathbf{v}$  :

- by hand on small problems solve  $\det(A - \lambda B) = 0$  ;
- in MATLAB/Octave invoke `lambda=eig(A,B)` , and then `r=log(lambda)/h` .

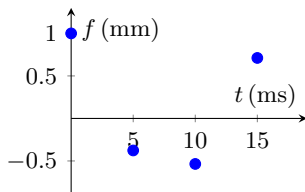
This eigen-problem typically determines  $n$  multipliers  $\lambda_1, \lambda_2, \dots, \lambda_n$ , and thence the  $n$  rates  $r_k = (\log \lambda_k)/h$  .

3. Determine the corresponding  $n$  coefficients  $c_1, c_2, \dots, c_n$  from any  $n$  point subset of the  $2n$  data points. For example, the first  $n$  data points give the linear system

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_n \\ \lambda_1^2 & \lambda_2^2 & \cdots & \lambda_n^2 \\ \vdots & \vdots & & \vdots \\ \lambda_1^{n-1} & \lambda_2^{n-1} & \cdots & \lambda_n^{n-1} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ \vdots \\ f_n \end{bmatrix}$$

In MATLAB/Octave one may construct the matrix  $U$  appearing here with `U=bsxfun(@power,lambd,0:n-1).'` when `lambd` is a column vector of the eigenvalues. Since the eigenvalues  $\lambda$  may be complex valued we need the transpose “.” not the complex conjugate transpose “'” (Table 3.1).

**Example 7.1.43.**



A damped piano string is struck and the sideways displacement of the string is measured at four times, 5 ms apart. The measurements (in mm) are  $f_1 = 1.0000$ ,  $f_2 = -0.3766$ ,  $f_3 = -0.5352$  and  $f_4 = 0.7114$  (as illustrated). Determine, by hand calculation, the frequency and damping of the string.

Recall Euler's formula that  $e^{i\theta} = \cos \theta + i \sin \theta$  so oscillations are here captured by complex valued exponentials.



Table 7.1: As well as the MATLAB/Octave commands and operations listed in Tables 1.2, 2.3, 3.1, 3.2, 3.3, 3.7, and 5.1 this section invokes these functions.

- **hankel(x,y)** for two vectors of the same length,  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  and  $\mathbf{y} = (y_1, y_2, \dots, y_n)$ , forms the  $n \times n$  matrix

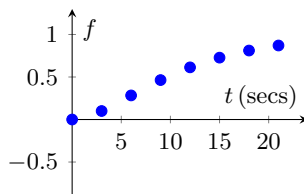
$$\begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_n \\ x_2 & x_3 & \cdots & x_n & y_2 \\ x_3 & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & y_{n-1} \\ x_n & y_2 & \cdots & y_{n-1} & y_n \end{bmatrix}.$$

- **eig(A,B)** for  $n \times n$  matrices  $A$  and  $B$  computes a vector in  $\mathbb{R}^n$  of eigenvalues  $\lambda$  such that  $\det(A - \lambda B) = 0$ . Some of the computed eigenvalues in the vector may be  $\pm \text{Inf}$  (depending upon the nature of  $B$ ) which denotes that a corresponding eigenvalue does not exist.

The command  $[\mathbf{V}, \mathbf{D}] = \text{eig}(\mathbf{A}, \mathbf{B})$  solves the generalised eigenproblem  $\mathbf{A}\mathbf{x} = \lambda\mathbf{B}\mathbf{x}$  for eigenvalues  $\lambda$  returned in the diagonal of matrix  $\mathbf{D}$  (some may be  $\pm \text{Inf}$ ), and corresponding eigenvectors returned in the corresponding column of  $\mathbf{V}$ .

**Example 7.1.44.** For the data of the previous [Example 7.1.43](#), determine the frequency and damping of the piano string using MATLAB/Octave. ■

**Example 7.1.45.** In a biochemical experiment every three seconds we measure the concentration of an output chemical as tabulated below (and illustrated in the margin). Fit a sum of four exponentials to this data.



secs	concentration
0	0.0000
3	0.1000
6	0.2833
9	0.4639
12	0.6134
15	0.7277
18	0.8112
21	0.8705

As with any data fitting, in practical applications be careful about the reliability of the results. Sound statistical analysis needs to supplement [Procedure 7.1.42](#) to inform us about expected errors and sensitivity. This problem of fitting exponentials to data is often sensitive to errors.

The techniques and theory of this subsection generalise to cater for noisy data and for complex system interactions with vast amounts of data. The generalisation, *Dynamic Mode Decomposition* ([Kutz et al. 2016](#), e.g.), is applied across many areas such as fluid mechanics, video processing, epidemiology, neuroscience, and financial trading.

## 7.2 Linear independent vectors may form a basis

### Section Contents

7.2.1	Linearly (in)dependent sets . . . . .	728
7.2.2	Form a basis for subspaces . . . . .	742
	Revisit unique solutions . . . . .	759

In [Chapter 4](#) on symmetric matrices, the eigenvectors from distinct eigenvalues are proved to be always orthogonal—because of the symmetry. For general matrices the eigenvectors are not orthogonal—as introduced at the start of this [Chapter 7](#). But the orthogonal property is extremely useful. Question: is there some analogue of orthogonality that is similarly useful? Answer: yes. We now extend “orthogonal” to the more general concept of “linear independence” which for general matrices replaces orthonormality.

One reason that orthogonal vectors are useful is that they can form an orthonormal basis and hence act as the unit vectors of an orthogonal coordinate systems. Analogously, the concept of linear

independence is closely connected to coordinate systems that are not orthogonal.

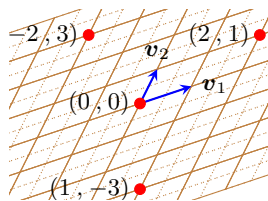
**Subspace coordinate systems** In any given problem we want two things from a general solution:

- firstly, the general solution must encompass every possibility (the solution must span the possibilities); and
- secondly, each possible solution should have a unique algebraic form in the general solution.

For an example of the need for a unique algebraic form, let's suppose we wanted to find solutions to the differential equation  $d^2y/dt^2 - y = 0$ . You might find  $y = 3e^x + 2e^{-x}$ , whereas I find  $y = 5 \cosh x + \sinh x$ , and a friend finds  $y = e^x + 4 \cosh x$ . By looking at these disparate algebraic forms it is apparent that we all disagree. Should we all go and search for errors in the solution process? No. The reason is that all these solutions are the same. The apparent differences arise only because you choose exponentials to represent the solution, whereas I choose hyperbolic functions,

and the friend a mixture: the solutions are the same, it is only the algebraic representation that appears different. In general, when we cannot immediately distinguish identical solutions, all algebraic manipulation becomes immensely more difficult due to algebraic ambiguity. To avoid such immense difficulties, in both calculus and linear algebra, we need to introduce the concept of linear independence.

Linear independence empowers us, often implicitly, to use a non-orthogonal coordinate system in a subspace. We replace the orthonormal standard unit vectors by any suitable set of basis vectors. For example, in the plane any two vectors at an angle to each other suffice to be able to describe uniquely every vector (point) in the plane. As illustrated in the margin, every point in the plane (end point of a vector) is a unique linear combination of the two drawn basis vectors  $\mathbf{v}_1$  and  $\mathbf{v}_2$ . Such a pair of basis vectors, termed a linearly independent pair, avoids the difficulties of algebraic ambiguity.

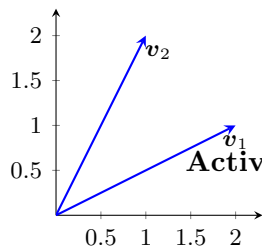




### 7.2.1 Linearly (in)dependent sets

This section defines “linear dependence” and “linear independence”, and then relates the concept to homogeneous linear equations, orthogonality, and sets of eigenvectors.

**Example 7.2.1** (2D non-orthogonal coordinates). Show that every vector in the plane  $\mathbb{R}^2$  can be written uniquely as a linear combination of the two vectors  $\mathbf{v}_1 = (2, 1)$  and  $\mathbf{v}_2 = (1, 2)$  that are shown in the margin. ■



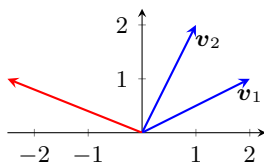
**Activity 7.2.2.** Write the vector shown in the margin as a linear combination of vectors  $\mathbf{v}_1$  and  $\mathbf{v}_2$ .

(a)  $-2.5\mathbf{v}_1 + 2\mathbf{v}_2$

(b)  $-2\mathbf{v}_1 + 1.5\mathbf{v}_2$

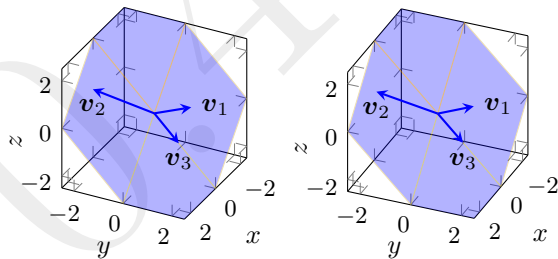
(c)  $-1.5\mathbf{v}_1 + \mathbf{v}_2$

(d)  $-\mathbf{v}_1 + \mathbf{v}_2$



**Example 7.2.3** (3D failure). Show that vectors in  $\mathbb{R}^3$  are not written uniquely as a linear combination of  $\mathbf{v}_1 = (-1, 1, 0)$ ,  $\mathbf{v}_2 = (1, -2, 1)$  and  $\mathbf{v}_3 = (0, 1, -1)$ .

One reason for the failure is that these three vectors only span a plane, as shown below in stereo. The solution here looks at the different issue of unique representation.



**Definition 7.2.4.** A set of vectors  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$  is **linearly dependent** if there are scalars  $c_1, c_2, \dots, c_k$ , at least one of which is nonzero, such that  $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_k\mathbf{v}_k = \mathbf{0}$ . A set of vectors that is not linearly dependent is called **linearly independent** (characterised by only the linear combination with  $c_1 = c_2 = \dots = c_k = 0$  gives the zero vector).

When reading the terms “linearly in/dependent” be very careful: it is all too easy to misread the presence or absence of the crucial “in” syllable. The presence or absence of the “in” syllable makes all the difference between the property and its opposite.

**Example 7.2.5.** Are the following sets of vectors linearly dependent or linearly independent. Give reasons.

(a)  $\{(-1, 1, 0), (1, -2, 1), (0, 1, -1)\}$

(b)  $\{(2, 1), (1, 2)\}$

(c)  $\{(-2, 4, 1, -1, 0)\}$

(d)  $\{(2, 1), (0, 0)\}$

(e)  $\{\mathbf{0}, \mathbf{v}_2, \mathbf{v}_3, \dots, \mathbf{v}_k\}$

(f)  $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ , the set of standard unit vectors in  $\mathbb{R}^3$ .

(g)  $\{(\frac{1}{3}, \frac{2}{3}, \frac{2}{3}), (\frac{2}{3}, \frac{1}{3}, -\frac{2}{3})\}$

These last two cases generalise to the next [Theorem 7.2.8](#) about the linear independence of every orthonormal set of vectors. ■

**Activity 7.2.6.** Which of the following sets of vectors is linearly independent?

(a)  $\{(-1, 2), (-2, 4)\}$

(b)  $\{(0, 0), (-2, 1)\}$

(c)  $\{(0, 1), (0, -1)\}$

(d)  $\{(-1, 1), (0, 1)\}$



**Example 7.2.7** (calculus extension). In calculus the notion of a function corresponds precisely to the notion of a vector in our linear algebra. For the purposes of this example, consider ‘vector’ and ‘function’ to be synonymous, and that ‘all components’ and ‘all  $x$ ’ are synonymous. Show that the set  $\{e^x, e^{-x}, \cosh x, \sinh x\}$  is linearly dependent. What is a subset that is linearly independent? ■

**Theorem 7.2.8.** *Every orthonormal set of vectors ([Definition 3.2.38](#)) is linearly independent.*

In contrast to orthonormal vectors which are always linearly independent, a set of two vectors proportional to each other is always linearly dependent as seen in the following examples. This linear dependence of proportional vectors then generalises in the forthcoming [Theorem 7.2.11](#).

**Example 7.2.9.** Show the following sets are linearly dependent.

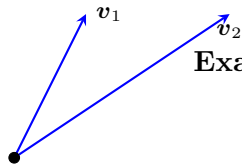
(a)  $\{(1, 2), (3, 6)\}$

(b)  $\{(2.2, -2.1, 0, 1.5), (-8.8, 8.4, 0, -6)\}$

**Activity 7.2.10.** For what value of  $c$  is the set  $\{(-3c, -2 + 2c), (1, 2)\}$  linearly dependent?

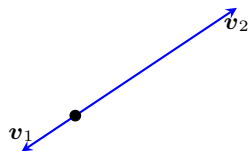
- (a)  $c = \frac{1}{4}$       (b)  $c = 0$       (c)  $c = 1$       (d)  $c = -\frac{1}{3}$

**Theorem 7.2.11.** *A set of vectors  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$  is linearly dependent if and only if at least one of the vectors can be expressed as a linear combination of the other vectors. In particular, a set of two vectors  $\{\mathbf{v}_1, \mathbf{v}_2\}$  is linearly dependent if and only if one of the vectors is a multiple of the other.*



**Example 7.2.12.** Invoke [Theorem 7.2.11](#) to deduce whether the following sets are linearly independent or linearly dependent.

- (a)  $\{(-1, 1, 0), (1, -2, 1), (0, 1, -1)\}$



- (b) The set of two vectors shown in the margin.
- (c) The set of two vectors shown in the margin.
- (d)  $\{(1, 3, 0, -1), (1, 0, -4, 2), (-2, 3, 0, -3), (0, 6, -4, -2)\}$



Recall that [Theorem 4.2.11](#) established that for every two distinct eigenvalues of a symmetric matrix  $A$ , any corresponding two eigenvectors are orthogonal. Consequently, for a symmetric matrix  $A$ , a set of eigenvectors from distinct eigenvalues forms an orthogonal set. The following [Theorem 7.2.13](#) generalises this property to non-symmetric matrices using the concept of linear independence.

**Theorem 7.2.13.** *For every  $n \times n$  matrix  $A$ , let  $\lambda_1, \lambda_2, \dots, \lambda_m$  be distinct eigenvalues of  $A$  with corresponding eigenvectors  $v_1, v_2, \dots, v_m$ . Then the set  $\{v_1, v_2, \dots, v_m\}$  is linearly independent.*

**Activity 7.2.14.** The matrix  $\begin{bmatrix} 2 & 1 \\ a^2 & 2 \end{bmatrix}$  has eigenvectors proportional to  $(1, a)$ , and proportional to  $(1, -a)$ . For what value of  $a$  does the matrix have a repeated eigenvalue?

- (a)  $a = 2$       (b)  $a = 0$       (c)  $a = -1$       (d)  $a = 1$





**Example 7.2.15.** For each of the following matrices, show the eigenvectors from distinct eigenvalues form linearly independent sets.

(a) Consider the matrix  $B = \begin{bmatrix} -1 & 1 & -2 \\ -1 & 0 & -1 \\ 0 & -3 & 1 \end{bmatrix}$  from [Example 7.1.13](#).

(b) [Example 7.1.14](#) found the eigenvalues and eigenvectors of matrix

$$A = \begin{bmatrix} 0 & 3 & 0 & 0 & 0 \\ 1 & 0 & 3 & 0 & 0 \\ 0 & 1 & 0 & 3 & 0 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

In MATLAB/Octave execute

```
A=[0 3 0 0 0
  1 0 3 0 0
  0 1 0 3 0
  0 0 1 0 3
  0 0 0 1 0]
[V,D]=eig(A)
```



to obtain the report (2 d.p.)

$V =$

0.62	-0.62	0.94	-0.85	-0.85
0.62	0.62	-0.00	0.49	-0.49
0.42	-0.42	-0.31	-0.00	0.00
0.21	0.21	-0.00	-0.16	0.16
0.07	-0.07	0.10	0.09	0.09

$D =$

3.00	0	0	0	0
0	-3.00	0	0	0
0	0	-0.00	0	0
0	0	0	-1.73	0
0	0	0	0	1.73

The five eigenvalues are all distinct, so [Theorem 7.2.16](#) asserts a set of corresponding eigenvectors will be linearly independent. The five columns of  $V$ , call them  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_5$ , are a set of corresponding eigenvectors. To confirm their linear independence let's seek a linear combination being zero, that is,  $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_5\mathbf{v}_5 = \mathbf{0}$ . Written as a matrix-vector sys-

tem we seek  $\mathbf{c} = (c_1, c_2, \dots, c_5)$  such that  $V\mathbf{c} = \mathbf{0}$ . Because the five singular values of square matrix  $V$  are all non-zero, obtained from  $\text{svd}(V)$  as

```
ans =  
    1.7703  
    1.1268  
    0.6542  
    0.3625  
    0.1922
```

consequently [Theorem 3.4.43](#) asserts  $V\mathbf{c} = \mathbf{0}$  has only the zero solution. Hence, by [Definition 7.2.4](#) the set of eigenvectors in the columns of  $V$  are linearly independent. ■

This last case of [Example 7.2.15b](#) connects the concept of linear in/dependence to the existence or otherwise of non-zero solutions to a homogeneous system of linear equations,  $V\mathbf{c} = \mathbf{0}$ . So does [Example 7.2.5b](#). The great utility of this connection is that we

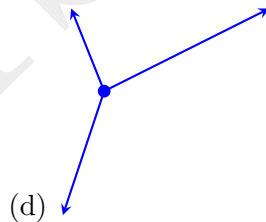
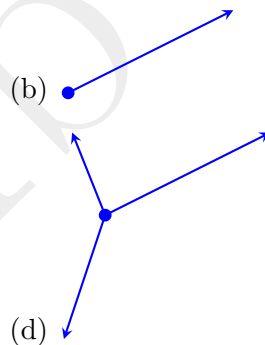
understand a lot about homogeneous systems of linear equations. The next [Theorem 7.2.16](#) establishes this connection in general.

**Theorem 7.2.16.** *Let  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$  be vectors in  $\mathbb{R}^n$ , and let the  $n \times m$  matrix  $V = [\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_m]$ . Then the set  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$  is linearly dependent if and only if the homogeneous system  $V\mathbf{c} = \mathbf{0}$  has a nonzero solution  $\mathbf{c}$ .*

Recall [Theorem 1.3.25](#) that in  $\mathbb{R}^n$  there can be no more than  $n$  vectors in an orthogonal set of vectors. The following theorem is the generalisation: in  $\mathbb{R}^n$  there can be no more than  $n$  vectors in a linearly independent set of vectors.

**Activity 7.2.17.** Which of the following sets of vectors are linearly dependent?

(a) None of these sets.



**Theorem 7.2.18.** *Every set of  $m$  vectors in  $\mathbb{R}^n$  is linearly dependent when the number of vectors  $m > n$ .*

**Example 7.2.19.** Determine if the following sets of vectors are linearly dependent or independent. Give reasons.

(a)  $\{(-1, -2), (-1, 4), (0, 5), (2, 3)\}$

(b)  $\{(-6, -4, -1, -2), (2, 0, 1, -2), (2, -1, -1, 1)\}$

(c)  $\{(-1, -2, 2, -1), (1, 3, 1, -1), (-2, -4, 4, -2)\}$

(d)  $\{(3, 3, -1, -1), (0, -3, -1, -7), (1, 2, 0, 2)\}$

(e)  $\{(10, 3, 3, 1), (2, -3, 0, -1), (1, -1, 2, -1), (2, -1, -3, 0), (-2, 0, 2, 2)\}$

(f)  $\{(-0.4, -1.8, -0.2, 0.7, -0.2), (-1.1, 2.8, 2.7, -3.0, -2.6), (-2.3, -2.6, -5.3, -3.3, -1.3, -4.1), (1.4, 5.2, -6.9, -0.7, 0.6)\}$



## 7.2.2 Form a basis for subspaces

Recall the definition of subspaces and the span, from Sections 2.3 and 3.4: namely that a subspace is a set of vectors closed under addition and scalar multiplication; and a span gives a subspace as all linear combinations of a set of vectors. Also, Definition 3.4.18 defined an “orthonormal basis” for a subspace to be a set of orthonormal vectors that span a subspace. This section generalises the concept of an “orthonormal basis” by relaxing the requirement of orthonormality to result in the concept of a “basis”.

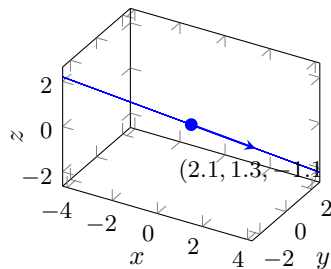
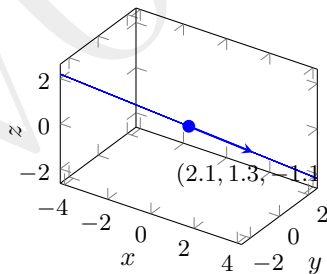
**Definition 7.2.20.** *A **basis** for a subspace  $\mathbb{W}$  of  $\mathbb{R}^n$  is a set of vectors that both span  $\mathbb{W}$  and is linearly independent.*

**Example 7.2.21.** (a) Recall Examples 7.2.5b and 7.2.1 showed that the two vectors  $(2, 1)$  and  $(1, 2)$  are linearly independent and span  $\mathbb{R}^2$ . Hence the set  $\{(2, 1), (1, 2)\}$  is a basis of  $\mathbb{R}^2$ .

(b) Recall that Example 7.2.5a showed the set  $\{(-1, 1, 0), (1, -2, 1), (0, 1, -1)\}$  is linearly dependent so it cannot be a basis.

However, remove one vector, such as the middle one, and consider the set  $\{(-1, 1, 0), (0, 1, -1)\}$ . As the two vectors are not proportional to each other, this set is linearly independent ([Theorem 7.2.11](#)). Also, the plane  $x + y + z = 0$  is a subspace, say  $\mathbb{W}$ . It is characterised by  $y = -x - z$ . So every vector in  $\mathbb{W}$  can be written as  $(x, -x - z, z) = (x, -x, 0) + (0, -z, z) = (-x)(-1, 1, 0) + (-z)(0, 1, -1)$ . That is,  $\text{span}\{(-1, 1, 0), (0, 1, -1)\} = \mathbb{W}$ . Hence  $\{(-1, 1, 0), (0, 1, -1)\}$  is a basis for the plane  $\mathbb{W}$ .

- (c) Find a basis for the line given parametrically as  $x = 2.1t$ ,  $y = 1.3t$  and  $z = -1.1t$  (shown below in stereo).

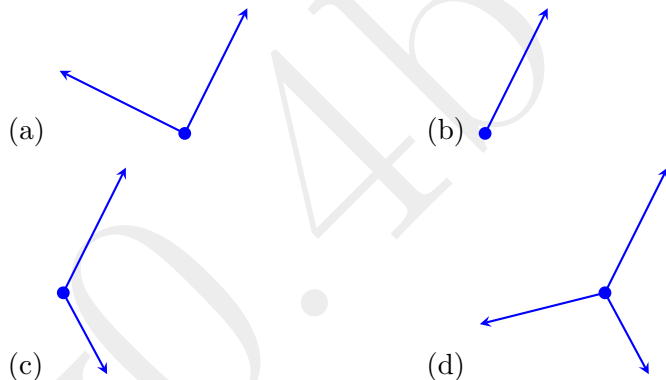




- (d) Find a basis for the line given parametrically as  $x = 5.7t - 0.6$  and  $y = 6.8t + 2.4$ .
- (e) Find a basis for the plane  $3x - 2y + z = 0$ .
- (f) Prove that every orthonormal basis of a subspace  $\mathbb{W}$  is also a basis of  $\mathbb{W}$ .



**Activity 7.2.22.** Which of the following sets of vectors form a basis for  $\mathbb{R}^2$ , but is not an orthonormal basis for  $\mathbb{R}^2$ ?



Recall that [Theorem 3.4.28](#) establishes that an orthonormal basis of a given subspace always has the same number of vectors. The following theorem establishes the same is true for general bases. The proof is direct generalisation of that for [Theorem 3.4.28](#).

**Theorem 7.2.23.** *Any two bases for a given subspace have the same number of vectors.*

**Example 7.2.24.** Consider the plane  $x + y + z = 0$  in  $\mathbb{R}^3$ . Each of the following are a basis for the plane:

- $\{(-1, 1, 0), (1, -2, 1)\};$
- $\{(1, -2, 1), (0, 1, -1)\};$
- $\{(0, 1, -1), (-1, 1, 0)\}.$

The reasons are that all three vectors involved are in the plane, and that each pair are linearly independent (as, in each pair, one is not proportional to the other).

However, consider the set  $\{(-1, 1, 0), (1, -2, 1), (0, 1, -1)\}$ . Although each of the three vectors is in the plane  $x + y + z = 0$ , this set is not a basis because it is not linearly independent ([Example 7.2.5a](#)). Each individual vector, say  $(-1, 1, 0)$ , cannot form a basis for the plane because the span of one vector, such as  $\text{span}\{(-1, 1, 0)\}$ , is a line not the whole plane.

The orthonormal basis  $\{(1, 0, -1)/\sqrt{2}, (1, -2, 1)/\sqrt{6}\}$  is another basis for the plane  $x + y + z = 0$ : both vectors satisfy  $x + y + z = 0$  and are orthogonal and so linearly independent ([Theorem 7.2.8](#)). All these bases possess two vectors. ■

That all bases for a given subspace, including orthonormal bases, have the same number of vectors ([Theorem 7.2.23](#)) leads to the following theorem about the dimensionality.

**Theorem 7.2.25.** *For every subspace  $\mathbb{W}$  of  $\mathbb{R}^n$ , the **dimension** of  $\mathbb{W}$ , denoted  $\dim \mathbb{W}$ , is the number of vectors in any basis for  $\mathbb{W}$ .*

**Activity 7.2.26.** Which of the following sets forms a basis for a subspace of dimension two?

- (a)  $\{(1, 2)\}$
- (b)  $\{(1, 1, -2), (2, 2, -4)\}$
- (c)  $\{(-1, 0, 2), (0, 0, 1), (-1, 2, 0)\}$
- (d)  $\{(1, -2, 1), (1, 0, -1)\}$



**Procedure 7.2.27** (basis for a span). *Find a basis for the subspace  $\mathbb{A} = \text{span}\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}$  given  $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}$  is a set of  $n$  vectors in  $\mathbb{R}^m$ . Recall [Procedure 3.4.23](#) underpins finding an orthonormal basis by the following.*

1. Form  $m \times n$  matrix  $A := [\mathbf{a}_1 \ \mathbf{a}_2 \ \cdots \ \mathbf{a}_n]$ .
2. Factorise  $A$  into its SVD,  $A = USV^T$ , and let  $r = \text{rank } A$  be the number of nonzero singular values (or effectively nonzero when the matrix has experimental errors, [Section 5.2](#)).
3. The set  $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r\}$  (where  $\mathbf{u}_j$  denotes the columns of  $U$ ) is a basis, specifically an orthonormal basis, for the  $r$ -dimensional subspace  $\mathbb{A}$ .

*Alternatively, if the rank  $r = n$ , then the set  $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}$  is linearly independent and span the subspace  $\mathbb{A}$ , and so is also a basis for the  $n$ -dimensional subspace  $\mathbb{A}$ .*

**Example 7.2.28.** Apply [Procedure 7.2.27](#) to find a basis for the following sets.

- (a) Recall [Example 7.2.24](#) identified that every pair of vectors in the set  $\{(-1, 1, 0), (1, -2, 1), (0, 1, -1)\}$  forms a basis for the plane that they span. Find another basis for the plane.
- (b) The span of the three vectors

$$(-2, 0, -4, 1, 1), (7, 1, 2, -1, -5), (-5, -1, 2, 3, -2).$$

- (c) The span of the four vectors  $(1, 0, 3, -4, 0)$ ,  $(-1, -1, 1, 4, 2)$ ,  $(-3, 2, 2, 2, 1)$ ,  $(3, -3, 2, -2, 1)$ .



The procedure is different if the subspace of interest is defined by a system of equations instead of the span of some vectors.

**Example 7.2.29.** Find a basis for the solutions of the system in  $\mathbb{R}^3$  of  $3x + y = 0$  and  $3x + 2y + 3z = 0$ . ■

**Example 7.2.30.** Find a basis for the solutions of  $-2x - y + 3z = 0$  in  $\mathbb{R}^3$ . ■

**Activity 7.2.31.** Which of the following is *not* a basis for the line  $3x + 7y = 0$ ?

(a)  $\{(3, 7)\}$

(b)  $\{(-7, 3)\}$

(c)  $\{(-\frac{7}{3}, 1)\}$

(d)  $\{(1, -\frac{3}{7})\}$

**Procedure 7.2.32** (basis from equations). *Suppose we seek a basis for a subspace  $\mathbb{W}$  specified as the solutions of a system of equations.*

1. *Rewrite the system of equations as the homogeneous system  $A\mathbf{x} = \mathbf{0}$ . Then the subspace  $\mathbb{W}$  is the nullspace of  $m \times n$  matrix  $A$ .*

2. Adapting [Procedure 3.3.15](#) for the specific case of homogeneous systems, first find an SVD factorisation  $A = USV^T$  and let  $r = \text{rank } A$  be the number of nonzero singular values (or effectively nonzero when the matrix has experimental errors, [Section 5.2](#)).
3. Then  $\mathbf{y} = (0, \dots, 0, y_{r+1}, \dots, y_n)$  is a general solution of  $S\mathbf{y} = \mathbf{z} = \mathbf{0}$ . Consequently, all possible solutions  $\mathbf{x} = V\mathbf{y}$  are spanned by the last  $n - r$  columns of  $V$ , which thus form an orthonormal basis for the subspace  $\mathbb{W}$ .

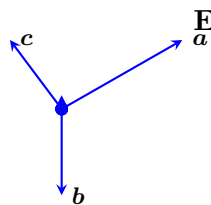
**Example 7.2.33.** Find a basis for all solutions to each of the following systems of equations.

- (a)  $3x + y = 0$  and  $3x + 2y + 3z = 0$  from [Example 7.2.29](#).
- (b)  $7x = 6y + z + 3$  and  $4x + 9y + 2z + 2 = 0$ .
- (c)  $w + x = z$ ,  $3w = x + y + 5z$ ,  $4x + y + 2z = 0$ .

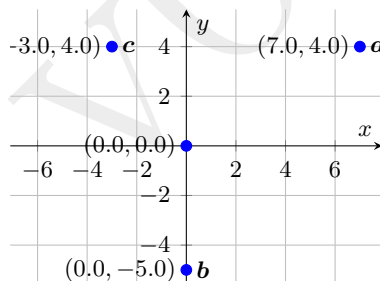




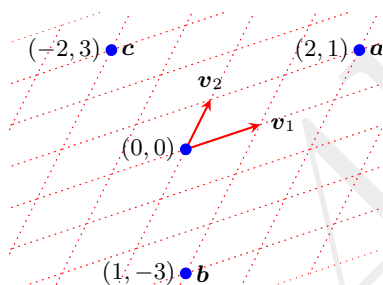
Recall this [Section 7.2](#) started by discussing the need to have a unique representation of solutions to problems. If we do not have uniqueness, then the ambiguity in algebraic representation ruins basic algebra. The forthcoming theorem assures us that the linear independence of a basis ensures the unique representation that we need. In essence it says that every basis, whether orthogonal or not, can be used to form a coordinate system.



**Example 7.2.34** (a tale of two coordinate systems). In the margin are plotted three vectors and the origin. Take the view that these are fixed physically meaningful vectors: the issue of this example is how do we code such vectors in mathematics.



In the standard orthogonal coordinate system these three vectors and the origin have coordinates as plotted on the left by their endpoints. Consequently, we write  $\mathbf{a} = (7, 4)$ ,  $\mathbf{b} = (0, -5)$  and  $\mathbf{c} = (-3, 4)$ .



Now use the (red) basis  $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$  to form a non-orthogonal coordinate system (represented by the dotted grid). Then in this system the three vectors have coordinates  $\mathbf{a} = (2, 1)$ ,  $\mathbf{b} = (1, -3)$  and  $\mathbf{c} = (-2, 3)$ .

But we cannot say both  $\mathbf{a} = (7, 4)$  and  $\mathbf{a} = (2, 1)$ : it appears nonsense. The reason for the different numbers representing the one vector  $\mathbf{a}$  is that the underlying coordinate systems are different. For example, we can say both  $\mathbf{a} = 7\mathbf{e}_1 + 4\mathbf{e}_2$  and  $\mathbf{a} = 2\mathbf{v}_1 + \mathbf{v}_2$  without any apparent contradiction: these two statements explicitly recognise the underlying standard unit vectors in the first expression and the underlying non-orthogonal basis vectors in the second.

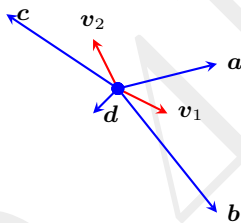
Consequently we invent a new better notation. We write  $[\mathbf{a}]_{\mathcal{B}} = (2, 1)$  to represent that the coordinates of vector  $\mathbf{a}$  in the basis  $\mathcal{B}$  are  $(2, 1)$ . Correspondingly, letting  $\mathcal{E} = \{\mathbf{e}_1, \mathbf{e}_2\}$  denote the basis

of the standard unit vectors, we write  $[\mathbf{a}]_{\mathcal{E}} = (7, 4)$  to represent that the coordinates of vector  $\mathbf{a}$  in the standard basis  $\mathcal{E}$  are  $(7, 4)$ . Similarly,  $[\mathbf{b}]_{\mathcal{E}} = (0, -5)$  and  $[\mathbf{b}]_{\mathcal{B}} = (1, -3)$ ; and  $[\mathbf{c}]_{\mathcal{E}} = (-3, 4)$  and  $[\mathbf{c}]_{\mathcal{B}} = (-2, 3)$ .

The endemic practice of just writing  $\mathbf{a} = (2, 1)$ ,  $\mathbf{b} = (1, -3)$  and  $\mathbf{c} = (-2, 3)$  is rationalised in this new notation by the convention that if no basis is explicitly specified, then the standard basis  $\mathcal{E}$  is assumed. ■

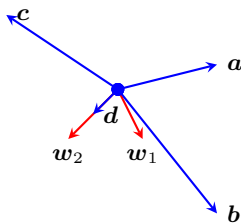
**Theorem 7.2.35.** *For every subspace  $\mathbb{W}$  of  $\mathbb{R}^n$  let  $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$  be a basis for  $\mathbb{W}$ . Then there is exactly one way to write each and every vector  $\mathbf{w} \in \mathbb{W}$  as a linear combination of the basis vectors:  $\mathbf{w} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \dots + c_k\mathbf{v}_k$ . The coefficients  $c_1, c_2, \dots, c_k$  are called the **coordinates of  $\mathbf{w}$  with respect to  $\mathcal{B}$** , and the column vector  $[\mathbf{w}]_{\mathcal{B}} = (c_1, c_2, \dots, c_k)$  is called the **coordinate vector of  $\mathbf{w}$  with respect to  $\mathcal{B}$** .*

**Example 7.2.36.** (a) Consider the diagram of six labelled vectors drawn below.

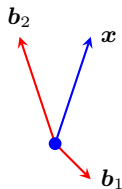


Estimate the coordinates of the four shown vectors  $a$ ,  $b$ ,  $c$  and  $d$  in the shown basis  $\mathcal{B} = \{v_1, v_2\}$ .

(b) Consider the same four vectors but with a pair of different basis vectors: let's see that although the vectors are the same, the coordinates in the different basis are different.



Estimate the coordinates of the four shown vectors  $a$ ,  $b$ ,  $c$  and  $d$  in the shown basis  $\mathcal{W} = \{w_1, w_2\}$ .



**Activity 7.2.37.** For the vector  $x$  shown in the margin, estimate the coordinates of  $x$  in the shown basis  $B = \{b_1, b_2\}$ .

(a)  $[x]_B = (4, 1)$

(b)  $[x]_B = (2, 3)$

(c)  $[x]_B = (3, 2)$

(d)  $[x]_B = (1, 4)$

**Example 7.2.38.** Let the basis  $B = \{v_1, v_2, v_3\}$  for the three given vectors  $v_1 = (-1, 1, -1)$ ,  $v_2 = (1, -2, 0)$  and  $v_3 = (0, 4, 5)$  (each of these are specified in the standard basis  $\mathcal{E}$  of the standard unit vectors  $e_1$ ,  $e_2$  and  $e_3$ ).

(a) What is the vector with coordinates  $[a]_B = (3, -2, 1)$ ?

(b) What is the vector with coordinates  $[b]_B = (-1, 1, 1)$ ?

- (c) What are the coordinates in the basis  $\mathcal{B}$  of the vector  $\mathbf{c}$  where  $[\mathbf{c}]_{\mathcal{E}} = (-1, 3, 3)$  in the standard basis  $\mathcal{E}$ ?
- (d) What are the coordinates in the basis  $\mathcal{B}$  of the vector  $\mathbf{d}$  where  $[\mathbf{d}]_{\mathcal{E}} = (-3, 2, 0)$  in the standard basis  $\mathcal{E}$ ?
- 

**Activity 7.2.39.** What are the coordinates in the basis  $\mathcal{B} = \{(1, 1), (1, -1)\}$  of the vector  $\mathbf{d}$  where  $[\mathbf{d}]_{\mathcal{E}} = (2, -4)$  in the standard basis  $\mathcal{E}$ ?

- (a)  $[\mathbf{d}]_{\mathcal{B}} = (-1, 3)$                       (b)  $[\mathbf{d}]_{\mathcal{B}} = (3, -1)$   
(c)  $[\mathbf{d}]_{\mathcal{B}} = (-2, 6)$                       (d)  $[\mathbf{d}]_{\mathcal{B}} = (1, 3)$
- 

**Example 7.2.40.** You are given a basis  $\mathcal{W} = \{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$  for a 3D subspace  $\mathbb{W}$  of  $\mathbb{R}^5$  where the three basis vectors are  $\mathbf{w}_1 = (1, 3, -4, -3, 3)$ ,  $\mathbf{w}_2 = (-4, 1, -2, -4, 1)$ , and  $\mathbf{w}_3 = (-1, 1, 0, 2, -3)$  (in the standard basis  $\mathcal{E}$ ).

- (a) What are the coordinates in the standard basis of the vector  $\mathbf{a} = 2\mathbf{w}_1 + 3\mathbf{w}_2 + \mathbf{w}_3$ ?
- (b) What are the coordinates in the basis  $\mathcal{W}$  of the vector  $\mathbf{b} = (-1, 2, -6, -11, 10)$  (in the standard coordinates  $\mathcal{E}$ ).



## Revisit unique solutions

Lastly, with all these extra concepts of determinants, eigenvalues, linear independence and a basis, we now revisit the issue of when there is a unique solution to a set of linear equations.

**Theorem 7.2.41** (Unique Solutions: version 3). *For every  $n \times n$  square matrix  $A$ , and extending Theorems 3.3.26 and 3.4.43, the following statements are equivalent:*

- (a)  $A$  is invertible;
- (b)  $A\mathbf{x} = \mathbf{b}$  has a unique solution for every  $\mathbf{b} \in \mathbb{R}^n$ ;
- (c)  $A\mathbf{x} = \mathbf{0}$  has only the zero solution;
- (d) all  $n$  singular values of  $A$  are nonzero;
- (e) the condition number of  $A$  is finite ( $\mathbf{rcond} > 0$ );
- (f)  $\text{rank } A = n$ ;
- (g)  $\text{nullity } A = 0$ ;
- (h) the column vectors of  $A$  span  $\mathbb{R}^n$ ;



- (i) the row vectors of  $A$  span  $\mathbb{R}^n$ .*
- (j)  $\det A \neq 0$ ;*
- (k) 0 is not an eigenvalue of  $A$ ;*
- (l) the  $n$  column vectors of  $A$  are linearly independent;*
- (m) the  $n$  row vectors of  $A$  are linearly independent.*

## 7.3 Diagonalisation identifies the transformation

### Section Contents

7.3.1 Solve systems of differential equations . . . . 774

**Population modelling** Recall that this [Chapter 7](#) started by introducing the dynamics of two interacting species of animals. Recall we let  $y(t)$  and  $z(t)$  be the number of female animals in each of the species at time  $t$  (years). Modelling might deduce the populations interact according to the rule that the population one year later is  $y(t+1) = 2y(t) - 4z(t)$  and  $z(t+1) = -y(t) + 2z(t)$ . Then seeking solutions proportional to  $\lambda^t$  led to the eigen-problem

$$\begin{bmatrix} 2 & -4 \\ -1 & 2 \end{bmatrix} \mathbf{x} = \lambda \mathbf{x}.$$

This section introduces an alternate equivalent approach.

The alternate approach invokes non-orthogonal coordinates. Start by writing the population model as a system in terms of vector

$\mathbf{y}(t) = (y(t), z(t))$ , namely

$$\begin{bmatrix} y(t+1) \\ z(t+1) \end{bmatrix} = \begin{bmatrix} 2y(t) - 4z(t) \\ -y(t) + 2z(t) \end{bmatrix},$$

$$\text{that is, } \mathbf{y}(t+1) = \begin{bmatrix} 2 & -4 \\ -1 & 2 \end{bmatrix} \mathbf{y}(t).$$

Now let's ask if there is a basis  $\mathcal{P} = \{\mathbf{p}_1, \mathbf{p}_2\}$  for the  $yz$ -plane that simplifies this matrix-vector system? In such a basis every vector may be written as  $\mathbf{y} = Y_1\mathbf{p}_1 + Y_2\mathbf{p}_2$  for some components  $Y_1$  and  $Y_2$ —where  $(Y_1, Y_2) = \mathbf{Y} = [\mathbf{y}]_{\mathcal{P}}$ , but to simplify writing we use the symbol  $\mathbf{Y}$  in place of  $[\mathbf{y}]_{\mathcal{P}}$ . Write the relation  $\mathbf{y} = Y_1\mathbf{p}_1 + Y_2\mathbf{p}_2$  as the matrix-vector product  $\mathbf{y} = P\mathbf{Y}$  where matrix  $P = [\mathbf{p}_1 \ \mathbf{p}_2]$  and vector  $\mathbf{Y} = (Y_1, Y_2)$ . The populations  $\mathbf{y}$  depends upon time  $t$ , and hence so does  $\mathbf{Y}$  since  $\mathbf{Y} = [\mathbf{y}]_{\mathcal{P}}$ ; that is,  $\mathbf{y}(t) = P\mathbf{Y}(t)$ . Substitute this identity into the system of equations:

$$\mathbf{y}(t+1) = P\mathbf{Y}(t+1) = \begin{bmatrix} 2 & -4 \\ -1 & 2 \end{bmatrix} P\mathbf{Y}(t).$$

Multiply both sides by  $P^{-1}$  (which exists by linear independence

of the columns, [Theorem 7.2.41](#)) to give

$$\mathbf{Y}(t+1) = \underbrace{P^{-1} \begin{bmatrix} 1 & -4 \\ -1 & 1 \end{bmatrix} P}_{P^{-1}AP} \mathbf{Y}.$$

The question then becomes, for a given square matrix  $A$ , such as this, can we find a matrix  $P$  such that  $P^{-1}AP$  is somehow simple? The answer is yes: using eigenvalues and eigenvectors, in most cases the product  $P^{-1}AP$  can be made into a simple diagonal matrix.

Recall that ([Subsection 4.2.2](#)) for a symmetric matrix  $A$  we could always factor  $A = VDV^T = VDV^{-1}$  for orthogonal matrix  $V$  and diagonal matrix  $D$ : thus a symmetric matrix is always orthogonally diagonalisable ([Definition 4.2.17](#)). For non-symmetric matrices, a diagonalisation mostly (although not always) can be done: the difference being we need an invertible matrix, typically called  $P$ , instead of the orthogonal matrix  $V$ . Such a matrix  $A$  is termed ‘diagonalisable’ instead of ‘orthogonally diagonalisable’.

**Definition 7.3.1.** An  $n \times n$  square matrix  $A$  is **diagonalisable** if there exists a diagonal matrix  $D$  and an invertible matrix  $P$  such that  $A = PDP^{-1}$ , equivalently  $AP = PD$  or  $P^{-1}AP = D$ .

**Example 7.3.2.** (a) Show that  $A = \begin{bmatrix} 0 & 1 \\ 2 & -1 \end{bmatrix}$  is diagonalisable by matrix

$$P = \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix}.$$

(b)  $B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$  is not diagonalisable.

(c) Is matrix  $C = \begin{bmatrix} 1.2 & 3.2 & 2.3 \\ 2.2 & -0.5 & -2.2 \end{bmatrix}$  diagonalisable?



**Example 7.3.3.** [Example 7.3.2a](#) showed that matrix  $P = \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix}$  diagonalises matrix  $A = \begin{bmatrix} 0 & 1 \\ 2 & -1 \end{bmatrix}$  to matrix  $D = \text{diag}(1, -2)$ . As a

prelude to the next [Theorem 7.3.5](#), show that the columns of  $P$  are eigenvectors of  $A$ . ■

**Activity 7.3.4.** Given matrix  $F = \begin{bmatrix} 5 & 8 \\ -4 & -7 \end{bmatrix}$  has eigenvectors  $(-1, 1)$  and  $(2, -1)$  corresponding to respective eigenvalues  $-3$  and  $1$ , what matrix diagonalises  $F$  to  $D = \text{diag}(-3, 1)$ ?

(a)  $\begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$

(b)  $\begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix}$

(c)  $\begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix}$

(d)  $\begin{bmatrix} 2 & -1 \\ 1 & -1 \end{bmatrix}$

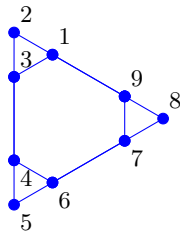
**Theorem 7.3.5.** *For every  $n \times n$  square matrix  $A$ , the matrix  $A$  is diagonalisable if and only if  $A$  has  $n$  linearly independent eigenvectors. If  $A$  is diagonalisable, with diagonal matrix  $D = P^{-1}AP$ , then the*

*diagonal entries of  $D$  are eigenvalues, and the columns of  $P$  are corresponding eigenvectors.*

**Example 7.3.6.** Recall that [Example 7.0.3](#) found the triangular matrix

$$A = \begin{bmatrix} -3 & 2 & 0 \\ 0 & -4 & 2 \\ 0 & 0 & 4 \end{bmatrix}$$

has eigenvalues  $-3$ ,  $-4$  and  $4$  (from its diagonal) and corresponding eigenvectors are proportional to  $(1, 0, 0)$ ,  $(-2, 1, 0)$  and  $(\frac{1}{14}, \frac{1}{4}, 1)$ . Is matrix  $A$  diagonalisable? ■



**Example 7.3.7.** Recall the Sierpinski network of [Example 4.1.20](#) (shown in the margin). Is the  $9 \times 9$  matrix  $A$  encoding the network diagonalisable?

$$A = \begin{bmatrix} -3 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & -2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & -3 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -3 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & -3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -3 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -2 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & -3 \end{bmatrix}.$$

**Example 7.3.8.** Recall [Example 7.1.13](#) found eigenvalues and corresponding eigenspaces for various matrices. Revisit these cases and show none of the matrices are diagonalisable.

(a) Matrix  $A = \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$  had one eigenvalue  $\lambda = 3$  with multiplicity



two and corresponding eigenspace  $\mathbb{E}_3 = \text{span}\{(1, 0)\}$ . This matrix is not diagonalisable as it has only one linearly independent eigenvector, such as  $(1, 0)$  or any non-zero multiple, and it needs two to be diagonalisable.

- (b) Matrix  $B = \begin{bmatrix} -1 & 1 & -2 \\ -1 & 0 & -1 \\ 0 & -3 & 1 \end{bmatrix}$  has eigenvalues  $\lambda = -2$  (multiplicity one) and  $\lambda = 1$  (multiplicity two). The corresponding eigenspaces are  $\mathbb{E}_{-2} = \text{span}\{(1, 1, 1)\}$  and  $\mathbb{E}_1 = \text{span}\{(-1, 0, 1)\}$ . Thus the matrix has only two linearly independent eigenvectors, one from each eigenspace, and it needs three to be diagonalisable.

- (c) Matrix  $C = \begin{bmatrix} -1 & 0 & -2 \\ 0 & -3 & 2 \\ 0 & -2 & 1 \end{bmatrix}$  has only the eigenvalue  $\lambda = -1$  with multiplicity three. The corresponding eigenspace  $\mathbb{E}_{-1} = \text{span}\{(1, 0, 0)\}$ . With only one linearly independent eigenvector, the matrix is not diagonalisable.



**Example 7.3.9.** Use the results of [Example 7.1.14](#) to show the following matrix is diagonalisable:

$$A = \begin{bmatrix} 0 & 3 & 0 & 0 & 0 \\ 1 & 0 & 3 & 0 & 0 \\ 0 & 1 & 0 & 3 & 0 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$



These examples illustrate a widely useful property. The  $5 \times 5$  matrix in [Example 7.3.9](#) has five distinct eigenvalues whose corresponding eigenvectors are necessarily linearly independent ([Theorem 7.2.13](#)) and so diagonalise the matrix ([Theorem 7.3.5](#)). The  $3 \times 3$  matrix in [Example 7.3.6](#) has three distinct eigenvalues whose corresponding eigenvectors are necessarily linearly independent ([Theorem 7.2.13](#)) and so diagonalise the matrix ([Theorem 7.3.5](#)). However, the

matrices of Examples 7.3.7 and 7.3.8 have repeated eigenvalues—eigenvalues of multiplicity two or more—and these matrices may (Example 7.3.7) or may not (Example 7.3.8) be diagonalisable. The following theorem confirms that matrices with as many *distinct* eigenvalues as the size of the matrix are always diagonalisable.

**Theorem 7.3.10.** *For every  $n \times n$  square matrix  $A$ , if  $A$  has  $n$  distinct eigenvalues, then  $A$  is diagonalisable. Consequently, and allowing complex eigenvalues, a real non-diagonalisable matrix must be non-symmetric and must have at least one repeated eigenvalue (an eigenvalue with multiplicity two or more).*

**Example 7.3.11.** From the given information, are the matrices diagonalisable?

- (a) The only eigenvalues of a  $4 \times 4$  matrix are 1.8,  $-3$ , 0.4 and 3.2.
- (b) The only eigenvalues of a  $5 \times 5$  matrix are 1.8,  $-3$ , 0.4 and 3.2.
- (c) The only eigenvalues of a  $3 \times 3$  matrix are 1.8,  $-3$ , 0.4 and 3.2.



**Activity 7.3.12.** A  $3 \times 3$  matrix  $A$  depends upon a parameter  $a$  and has eigenvalues  $6$ ,  $3 - 3a$  and  $2 + a$ . For which of the following values of parameter  $a$  may the matrix be *not* diagonalisable?

- (a)  $a = 1$       (b)  $a = 3$       (c)  $a = 4$       (d)  $a = 2$



**Example 7.3.13.** MATLAB/Octave computes the eigenvalues of matrix

$$A = \begin{bmatrix} -1 & 2 & -2 & 1 & -2 \\ -3 & -1 & -2 & 5 & 6 \\ 3 & 1 & 6 & -2 & -1 \\ 1 & 1 & 2 & 1 & -1 \\ 7 & 5 & -3 & 0 & 0 \end{bmatrix}$$

via `eig(A)` and reports them to be (2 d.p.)

```
ans =  
-3.45 + 3.50i  
-3.45 - 3.50i
```

5.00

5.00

1.91

Is the matrix diagonalisable? ■

Recall that for every symmetric matrix, from [Definition 4.1.15](#), the dimension of an eigenspace,  $\dim \mathbb{E}_{\lambda_j}$ , is equal to the multiplicity of the corresponding eigenvalue  $\lambda_j$ . However, for general matrices this equality is not necessarily so.

**Theorem 7.3.14.** *For every square matrix  $A$ , and for each eigenvalue  $\lambda_j$  of  $A$ , the corresponding eigenspace  $\mathbb{E}_{\lambda_j}$  has dimension less than or equal to the multiplicity of  $\lambda_j$ ; that is,  $1 \leq \dim \mathbb{E}_{\lambda_j} \leq \text{multiplicity of } \lambda_j$ .*


**Example 7.3.15.** Show the following matrix has one eigenvalue of multiplicity three, and the corresponding eigenspace has dimension two:

$$A = \begin{bmatrix} 0 & 5 & 6 \\ -8 & 22 & 24 \\ 6 & -15 & -16 \end{bmatrix}$$



**Example 7.3.16.** Use MATLAB/Octave to find the eigenvalues and the dimension of the eigenspaces of the matrix

$$B = \begin{bmatrix} 344 & -1165 & -149 & -1031 & 1065 & -2816 \\ 90 & -306 & -38 & -272 & 280 & -742 \\ -45 & 140 & 12 & 117 & -115 & 302 \\ 135 & -470 & -70 & -421 & 445 & -1175 \\ -165 & 555 & 67 & 493 & -506 & 1338 \\ -105 & 360 & 48 & 322 & -335 & 886 \end{bmatrix}.$$

MATLAB/Octave may produce for you a quite different matrix  $V$  of eigenvectors (possibly with complex parts). As discussed by [Subsection 7.1.2](#), repeated eigenvalues are very sensitive and this sensitivity means small variations in the hidden MATLAB/Octave algorithm may produce quite large changes in the matrix  $V$  for repeated eigenvalues. However, each eigenspace spanned by the appropriate columns of  $V$  is robust. 

### 7.3.1 Solve systems of differential equations

**Population modelling** The population modelling seen so far (Subsection 7.1.3) expressed the changes of the population over discrete intervals in time via discrete time equations such as  $y(t+1) = \dots$  and  $z(t+1) = \dots$ . One such example is to describe the population numbers year by year. The alternative is to model the changes in the population *continuously* in time. This alternative invokes and analyses differential equations. Such continuous time, differential equation, models are common for exploring the interaction between different species, such as between humans and viruses.

Let's start with a continuous time version of the population modelling discussed at the start of this Chapter 7. Let two species interact continuously in time with populations  $y(t)$  and  $z(t)$  at time  $t$  (years). Suppose they interact according to differential equations  $dy/dt = y - 4z$  and  $dz/dt = -y + z$  (instead of the discrete time equations  $y(t+1) = \dots$  and  $z(t+1) = \dots$ ). Analogous to the start of this Section 7.3, we now ask the following question: is there a matrix transformation to new variables, the vector  $\mathbf{Y}(t)$ , such that  $(y, z) = P\mathbf{Y}$  for some as yet unknown matrix  $P$ , where

the differential equations for  $\mathbf{Y}$  are simple?

- First, form the differential equations into a matrix-vector system:

$$\begin{bmatrix} dy/dt \\ dz/dt \end{bmatrix} = \begin{bmatrix} y - 4z \\ -y + z \end{bmatrix} = \begin{bmatrix} 1 & -4 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix}.$$

So using vector  $\mathbf{y} = (y, z)$ , this system is

$$\frac{d\mathbf{y}}{dt} = A\mathbf{y} \quad \text{for matrix } A = \begin{bmatrix} 1 & -4 \\ -1 & 1 \end{bmatrix}.$$

- Second, see what happens when we transform to some, as yet unknown, new variables  $\mathbf{Y}(t)$  such that  $\mathbf{y} = P\mathbf{Y}$  for some constant invertible matrix  $P$ . Under such a transform:  $\frac{d\mathbf{y}}{dt} = \frac{d}{dt}P\mathbf{Y} = P\frac{d\mathbf{Y}}{dt}$ ; also  $A\mathbf{y} = AP\mathbf{Y}$ . Hence substituting such an assumed transformation into the differential equations leads to

$$P\frac{d\mathbf{Y}}{dt} = AP\mathbf{Y}, \quad \text{that is} \quad \frac{d\mathbf{Y}}{dt} = (P^{-1}AP)\mathbf{Y}.$$



To simplify this system for  $\mathbf{Y}$ , we diagonalise the matrix on the right-hand side. The procedure is to choose the columns of  $P$  to be eigenvectors of the matrix  $A$  ([Theorem 7.3.5](#)).

- Third, find the eigenvectors of  $A$  by hand as it is a  $2 \times 2$  matrix. Here the matrix  $A = \begin{bmatrix} 1 & -4 \\ -1 & 1 \end{bmatrix}$  has characteristic polynomial  $\det(A - \lambda I) = (1 - \lambda)^2 - 4$ . This is zero for  $(1 - \lambda)^2 = 4$ , that is,  $(1 - \lambda) = \pm 2$ . Hence the eigenvalues  $\lambda = 1 \pm 2 = 3, -1$ .
  - For eigenvalue  $\lambda_1 = 3$  the corresponding eigenvectors satisfy
$$(A - \lambda_1 I)\mathbf{p}_1 = \begin{bmatrix} -2 & -4 \\ -1 & -2 \end{bmatrix} \mathbf{p}_1 = \mathbf{0},$$
with general solution  $\mathbf{p}_1 \propto (2, -1)$ .
  - For eigenvalue  $\lambda_2 = -1$  the corresponding eigenvectors satisfy

$$(A - \lambda_2 I)\mathbf{p}_2 = \begin{bmatrix} 2 & -4 \\ -1 & 2 \end{bmatrix} \mathbf{p}_2 = \mathbf{0},$$

with general solution  $\mathbf{p}_2 \propto (2, 1)$ .

Thus setting transformation matrix

$$P = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} \implies \frac{d\mathbf{Y}}{dt} = \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix} \mathbf{Y}$$

(any scalar multiple of the two columns of  $P$  would also work).

- Fourth, having diagonalised the matrix, expand this diagonalised set of differential equations to write this system in terms of components:

$$\frac{dY_1}{dt} = 3Y_1 \quad \text{and} \quad \frac{dY_2}{dt} = -Y_2.$$

Each of these differential equations have well-known exponential solutions, respectively  $Y_1 = c_1 e^{3t}$  and  $Y_2 = c_2 e^{-t}$ , for every constants  $c_1$  and  $c_2$ .

- Lastly, what does this mean for the original problem? From the relation

$$\begin{bmatrix} y \\ z \end{bmatrix} = \mathbf{y} = P\mathbf{Y} = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} c_1 e^{3t} \\ c_2 e^{-t} \end{bmatrix} = \begin{bmatrix} 2c_1 e^{3t} + 2c_2 e^{-t} \\ -c_1 e^{3t} + c_2 e^{-t} \end{bmatrix}.$$

That is, a general solution of the original system of differential equations is  $y(t) = 2c_1e^{3t} + 2c_2e^{-t}$  and  $z(t) = -c_1e^{3t} + c_2e^{-t}$ .

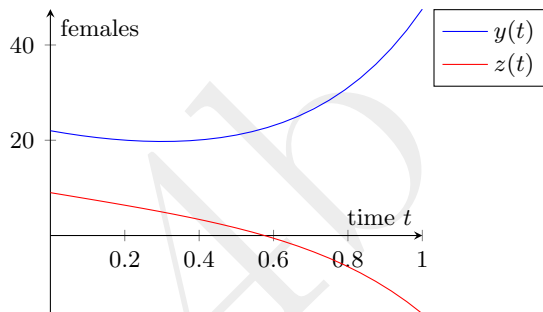
The diagonalisation of the matrix empowers us to solve complicated systems of differential equations as a set of simple systems.

Such a general solution makes predictions. For example, suppose at time zero ( $t = 0$ ) the initial population of female  $y$ -animals is 22 and the population of female  $z$ -animals is 9. From the above general solution we then know that at time  $t = 0$

$$\begin{bmatrix} 22 \\ 9 \end{bmatrix} = \begin{bmatrix} y(0) \\ z(0) \end{bmatrix} = \begin{bmatrix} 2c_1e^{3 \cdot 0} + 2c_2e^{-0} \\ -c_1e^{3 \cdot 0} + c_2e^{-0} \end{bmatrix} = \begin{bmatrix} 2c_1 + 2c_2 \\ -c_1 + c_2 \end{bmatrix}$$

This determines the coefficients:  $2c_1 + 2c_2 = 22$  and  $-c_1 + c_2 = 9$ . Adding the first to twice the second gives  $4c_2 = 40$ , that is,  $c_2 = 10$ . Then either equation determines  $c_1 = 1$ . Consequently, the particular solution from this initial population is

$$\begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} 2 \cdot 1e^{3t} + 2 \cdot 10e^{-t} \\ -1e^{3t} + 10e^{-t} \end{bmatrix} = \begin{bmatrix} 2e^{3t} + 20e^{-t} \\ -e^{3t} + 10e^{-t} \end{bmatrix}.$$



The above graph of this solution shows that the population of  $y$ -animals grows in time, whereas the population of  $z$ -animals crashes and becomes extinct at about time 0.6 years.

The forthcoming [Theorem 7.3.18](#) confirms that the same approach solves general systems of differential equations: it corresponds to [Theorem 7.1.25](#) for discrete dynamics.

**Activity 7.3.17.** A given population model is expressed as the differential equations  $dx/dt = x + y - 3z$ ,  $dy/dt = -2x + z$  and  $dz/dt = -2x + y + 2z$ . This may be written in matrix-vector form  $d\mathbf{x}/dt = A\mathbf{x}$  for vector  $\mathbf{x}(t) = (x, y, z)$  and which of the following matrices?

- (a)  $\begin{bmatrix} 1 & -3 & 1 \\ -2 & 1 & 0 \\ -2 & 2 & 1 \end{bmatrix}$  (b)  $\begin{bmatrix} 1 & -2 & 2 \\ 0 & -2 & 1 \\ 1 & 1 & -3 \end{bmatrix}$
- (c)  $\begin{bmatrix} 1 & 1 & -3 \\ 0 & -2 & 1 \\ 1 & -2 & 2 \end{bmatrix}$  (d)  $\begin{bmatrix} 1 & 1 & -3 \\ -2 & 0 & 1 \\ -2 & 1 & 2 \end{bmatrix}$



**Theorem 7.3.18.** Let  $n \times n$  square matrix  $A$  be diagonalisable by matrix  $P = [\mathbf{p}_1 \ \mathbf{p}_2 \ \cdots \ \mathbf{p}_n]$  whose columns are eigenvectors corresponding to eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_n$ . Then a general solution  $\mathbf{x}(t)$  to the differential equation system  $d\mathbf{x}/dt = A\mathbf{x}$  is the linear combination

$$\mathbf{x}(t) = c_1 \mathbf{p}_1 e^{\lambda_1 t} + c_2 \mathbf{p}_2 e^{\lambda_2 t} + \cdots + c_n \mathbf{p}_n e^{\lambda_n t} \quad (7.3)$$

for arbitrary constants  $c_1, c_2, \dots, c_n$ .

**Activity 7.3.19.** Recall that the differential equations  $dy/dt = y - 4z$  and  $dz/dt = -y + z$  have a general solution  $y(t) = 2c_1e^{3t} + 2c_2e^{-t}$  and  $z(t) = -c_1e^{3t} + c_2e^{-t}$ . What are the values of these constants given that  $y(0) = 2$  and  $z(0) = 3$ ?

(a)  $c_1 = 0, c_2 = -1$

(b)  $c_1 = c_2 = 1$

(c)  $c_1 = -1, c_2 = 2$

(d)  $c_1 = -2, c_2 = 0$



**Example 7.3.20.** Find (by hand) a general solution to the system of differential equations  $\frac{du}{dt} = -2u + 2v$ ,  $\frac{dv}{dt} = u - 2v + w$ , and  $\frac{dw}{dt} = 2v - 2w$ .



**Example 7.3.21.** Use the general solution derived in [Example 7.3.20](#) to predict the solution of the differential equations  $\frac{du}{dt} = -2u + 2v$ ,  $\frac{dv}{dt} = u - 2v + w$ , and  $\frac{dw}{dt} = 2v - 2w$  given the initial conditions that  $u(0) = v(0) = 0$  and  $w(0) = 4$ . ■

**Example 7.3.22.** Use MATLAB/Octave to find a general solution to the system of differential equations

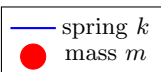
$$\begin{aligned}dx_1/dt &= -\frac{1}{2}x_1 - \frac{1}{2}x_2 + x_3 + 2x_4, \\dx_2/dt &= -\frac{1}{2}x_1 - \frac{1}{2}x_2 + 2x_3 + x_4, \\dx_3/dt &= x_1 + 2x_2 - \frac{1}{2}x_3 - \frac{1}{2}x_4, \\dx_4/dt &= 2x_1 + x_2 - \frac{1}{2}x_3 - \frac{1}{2}x_4.\end{aligned}$$

What is the particular solution that satisfies the initial conditions  $x_1(0) = -5$ ,  $x_2(0) = -1$  and  $x_3(0) = x_4(0) = 0$ ? Record your commands and give reasons. ■

**Example 7.3.23.** Find (by hand) a general solution to the system of differential equations  $\frac{dy}{dt} = z$  and  $\frac{dz}{dt} = -4y$ . ■

**Example 7.3.24.** Further consider [Example 7.3.23](#). Suppose we additionally know that  $y(0) = 3$  and  $z(0) = 0$ . Find the particular solution that satisfies these two initial conditions. ■

**Example 7.3.25.** In a real application the complex numbers of the general solution to [Example 7.3.23](#) are usually inconvenient. Instead we often express the solution solely in terms of real quantities as just done in the previous [Example 7.3.24](#). Use Euler's formula, that  $e^{i\theta} = \cos\theta + i\sin\theta$  for any  $\theta$ , to rewrite the general solution of [Example 7.3.23](#) in terms of real functions. ■





**Example 7.3.26** (oscillating applications). A huge variety of vibrating systems are analogous to the basic oscillations of a mass on a spring, illustrated schematically in the margin. The mass generally will oscillate to and fro. Describe such a system mathematically with two differential equations, and solve the differential equations to confirm it oscillates. ■

---

## Bibliography

---

Berry, M. W., Dumais, S. T. & O'Brien, G. W. (1995), 'Using linear algebra for intelligent information retrieval', *SIAM Review* **37**(4), 573–595.

<http://epubs.siam.org/doi/abs/10.1137/1037127>

Bliss, K., Fowler, K., Galluzzo, B., Garfunkel, S., Giordano, F., Godbold, L., Gould, H., Levy, R., Libertini, J., Long, M., Malkevitch, J., Montgomery, M., Pollak, H., Teague, D., van der Kooij, H. & Zbiek, R. (2016), GAIMME—Guidelines for Assessment and Instruction in Mathematics Modeling Education, Technical report, SIAM and COMAP.

[http://www.siam.org/reports/gaimme.php?\\_ga=1](http://www.siam.org/reports/gaimme.php?_ga=1)

Cuyt, A. (2015), Approximation theory, in N. J. Higham, M. R. Dennis, P. Glendinning, P. A. Martin, F. Santosa & J. Tanner,

eds, 'Princeton Companion to Applied Mathematics', Princeton, chapter IV.9, pp. 248–262.

Higham, N. J. (1996), *Accuracy and stability of numerical algorithms*, SIAM.

Higham, N. J. (2015), Numerical linear algebra and matrix analysis, in N. J. Higham, M. R. Dennis, P. Glendinning, P. A. Martin, F. Santosa & J. Tanner, eds, 'Princeton Companion to Applied Mathematics', Princeton, chapter IV.10, pp. 263–281.

Kutz, J. N., Brunton, S. L., Brunton, B. W. & Proctor, J. L. (2016), *Dynamic Mode Decomposition: Data-driven Modeling of Complex Systems*, number 149 in 'Other titles in applied mathematics', SIAM, Philadelphia.

Lichman, M. (2013), 'UCI machine learning repository', [online].  
<http://archive.ics.uci.edu/ml>

Mandelbrot, B. B. (1982), *The fractal geometry of nature*, W. H. Freeman.

Pereyra, V. & Scherer, G. (2010), *Exponential Data Fitting and its*

*Applications*, Bentham Science.

<http://ebooks.benthamscience.com/book/9781608050482/>

Quarteroni, A. & Saleri, F. (2006), *Scientific Computing with MATLAB and Octave*, Vol. 2 of *Texts in Computational Science and Engineering*, 2nd edn, Springer.

Roulstone, I. & Norbury, J. (2013), *Invisible in the storm: the role of mathematics in understanding weather*, Princeton.

Schonefeld, S. (1995), 'Eigenpictures: Picturing the eigenvector problem', *The College Mathematics Journal* **26**(4), 316–319.

<http://www.jstor.org/stable/2687037>

Trefethen, L. N. & Bau, III, D. (1997), *Numerical linear algebra*, SIAM.