Linear Algebra for 21st C Application

A. J. Roberts *
University of Adelaide, South Australia, 5005

January 25, 2020

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

Contents

1 Vectors

Chapter contents

This chapter provides 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 $1, 2, 3, \ldots$ 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** $\ldots, -2, -1, 0, 1, 2, \ldots$. In the meantime people needed to quantify fractions such as two and a half 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 nonzero). Over two thousand years ago, Pythagoras was forced to recognize 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 $\sqrt{2} = 1.41421\cdots$ and (pi) $\pi = 3.14159\cdots$, mathematicians define the **real numbers** to be all numbers which in principle can be written as a decimal expansion such as $\sqrt{2}$, π,

$$\tfrac{9}{7} = 1.285714285714\cdots \quad \text{or} \quad e = 2.718281828459\cdots .$$

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 number**s which are those of the form $a + b\,\mathrm{i}$ 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 (Chapter 7) we need the marvellous complex numbers.

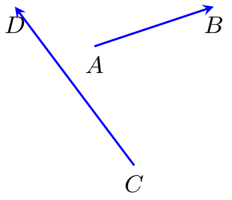
This book uses the term **scalar** to denote a number that could be integer, real, or complex. In this book, and before Chapter 7, a scalar is almost always real valued. The term 'scalar' arises because such numbers are often used to scale the length of a 'vector'.

1.1 Vectors have magnitude and direction

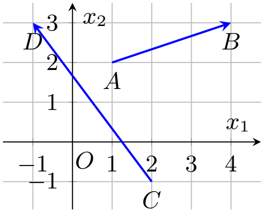
> There are more things in heaven and earth, Horatio,
> than are dreamt of in your philosophy.
>
> *(Hamlet I.5:159–167)*

In the eighteenth century, astronomers needed to describe both the position and the 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 (metres per second) from the south-west has both a magnitude and a 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

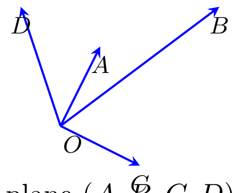


two displacement vectors $\overrightarrow{AB}$ and $\overrightarrow{CD}$ to the right. For example, if your home is at position A and your school at position B, then travelling from home to school is to move by the amount of the displacement vector $\overrightarrow{AB}$.

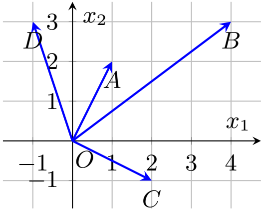


To be able to manipulate vectors we describe them with numbers. For such numbers to have meaning they must be set in the context of a coordinate system. So choose an origin for the coordinate system, usually denoted O, and draw coordinate axes in the plane (or space), as illustrated for the above two 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 $\overrightarrow{CD} = (-3, 4)$. Our choice of the origin O does not affect the number representation of these vectors. $\square$

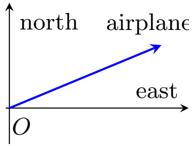
Example 1.1.2 *(position vector)* The next important class of vectors are the **position vectors**. Given some chosen fixed origin in space, usually denoted O, then $\overrightarrow{OA}$ is the position vector of the point A. This picture illustrates the position vectors of four points in the plane (A,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

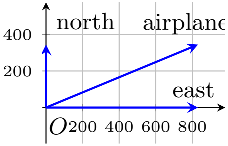


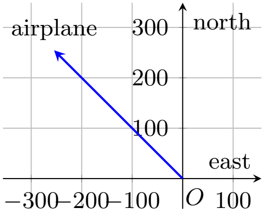
Again, to be able to manipulate such vectors we describe them with numbers, and such numbers have meaning via a coordinate system. So draw coordinate axes in the plane (or space), as illustrated for the above four position vectors. Here the position vector $\overrightarrow{OA}$ goes one unit to the right 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 $\overrightarrow{OD} = (-1, 3)$. Recognize 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 (kilometres per hour) 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° from the east, as illustrated on the right. Trigonometry then tells us that the eastward part of the speed of the airplane is $900\cos(22.5^\circ) = 831.5$ km/hr, whereas the northward part of the speed is $900\sin(22.5^\circ) = 344.4$ km/hr (as indicated quantitatively). 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 km/hr in space.





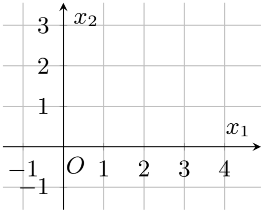


Another airplane takes off from an airport at 360 km/hr to the northwest and climbs at 2 m/s. The direction northwest is 45° to the east-west lines and 45° to the north-south lines. Trigonometry then tells us that the westward speed of the airplane is $360\cos(45°) = 360\cos(\frac{\pi}{4}) = 254.6$ km/hr, whereas the northward speed is $360\sin(45°) = 360\sin(\frac{\pi}{4}) = 254.6$ km/hr as illustrated.

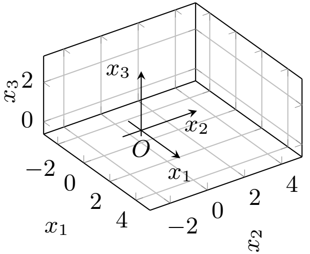
But west is the opposite direction to east, so if the coordinate system treats east as positive, then west must be negative. Consequently, together with the climb in the vertical, the velocity vector is (-254.6 km/hr,254.6 km/hr,2 m/s). But we should avoid mixing units within a vector, so here convert 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), 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), 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 $x_1, x_2, \ldots, x_n$, respectively.

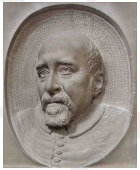
Definition 1.1.4 *Given a standard coordinate system with n coordinate axes, all at right-angles to each other, a* ***vector*** *is an ordered n-tuple of real numbers $x_1, x_2, \ldots, x_n$ equivalently written either as a row in parentheses or as a column in brackets,*

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

(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 $x_1, x_2, \ldots, x_n$ 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 $(x_1, x_2, \ldots, x_n)$ then the position*

vector $\overrightarrow{OX}$ has the same magnitude and direction as the vector denoted $(x_1\,, x_2\,, \ldots, x_n)$.

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 "bicause noe 2 thynges can be moare equalle". He also invented the term "sine" and the method of extracting the square-root by hand.

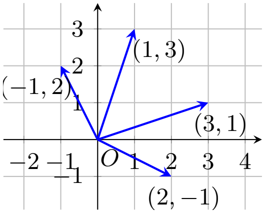
Examples 1.1.1 and 1.1.2 introduced some vectors and wrote them as a row in parentheses, such as $\overrightarrow{AB} = (3, 1)$. In this book, exactly the same thing is meant by the columns in brackets: for example,

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

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

However, as defined subsequently, a row of numbers within brackets is quite different: for two examples, $(3, 1) \neq \begin{bmatrix} 3 & 1 \end{bmatrix}$, and $(831, 344, 0) \neq \begin{bmatrix} 831 & 344 & 0 \end{bmatrix}$.

The *ordering* of the components is very important. For example, as illustrated to the right, 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, $(0, 0, \ldots, 0)$, 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. $\square$

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 just spatial structure. The next example illustrates another use of vectors.

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 each of the sentences is captured simply by counting the number of times each of these three objects and two actions appears in each sentence, and then forming a vector from the counts. Let's use vectors $\boldsymbol{w} = (N_{\text{cat}}\,, N_{\text{dog}}\,, N_{\text{mat}}\,, N_{\text{sat}}\,, N_{\text{scratched}})$ where the various N are the counts of each word ($\boldsymbol{w}$ 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 $\mathbb{R}^5$ as there are five components in each vector $\boldsymbol{w}$. Then

(a) “The dog sat on the mat” is summarized by the vector $\boldsymbol{w} = (0, 1, 1, 1, 0)$.

(b) “The cat scratched the dog” is summarized by the vector $\boldsymbol{w} = (1, 1, 0, 0, 1)$.

(c) “The cat and dog sat on the mat” is summarized by the vector $\boldsymbol{w} = (1, 1, 1, 1, 0)$.

(d) “The dog scratched” is summarized by the vector $\boldsymbol{w} = (0, 1, 0, 0, 1)$.

(e) An empty sentence is the zero vector $\boldsymbol{w} = (0, 0, 0, 0, 0)$.

(f) Together, the two sentences “The dog sat on the mat. The cat scratched the dog.” are summarized by the vector $\boldsymbol{w} = (1, 2, 1, 1, 1)$.

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 number' of components. □

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*

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

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

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

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

*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 $\|\boldsymbol{v}\|$. Either notation is good.)*

Remember that the term *size* denotes the number of components in a vector (Definition 1.1.4) and so "size" should not be used for the length/magnitude of a vector.

Example 1.1.10 Find the lengths of the following vectors: $\boldsymbol{a} = (-3, 4)$; $\boldsymbol{b} = (3, 3)$; $\boldsymbol{c} = (1, -2, 3)$; $\boldsymbol{d} = (1, -1, -1, 1)$. □

Example 1.1.11 Write down three different vectors, all three with the same number of components, for each of the following cases: (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) $\sqrt{11}$ (b) 5 (c) 7 (d) 11

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

1.2 Adding and stretching vectors

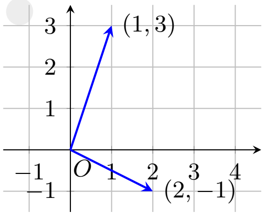
Section contents

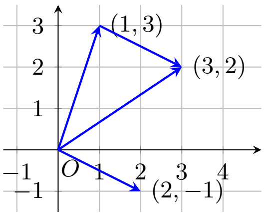
We want to be able to make sense of statements such as "king – man + women = queen". To do so, we need to define operations on vectors. Useful operations on vectors are those that are physically meaningful. Then our algebraic manipulations derive powerful results in applications. The first two vector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

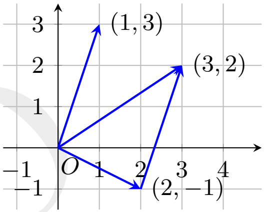
1.2.1 Basic operations

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) Let's add the two vectors shown below-left: $(1,3)+(2,-1) = (1+2, 3+(-1)) = (3,2)$ as illustrated below-middle, where the vector $(2,-1)$ is drawn from the end of $(1,3)$, and the end-point of the result determines the vector addition $(3,2)$.

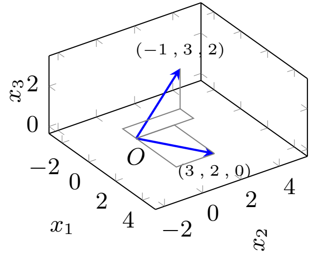


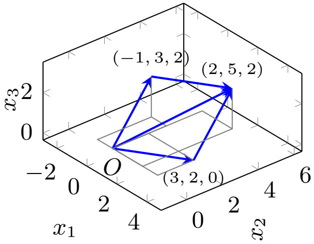


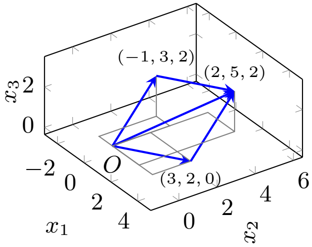


This result $(3,2)$ is the same if the vector $(1,3)$ is drawn from the end of $(2,-1)$ as shown above-right. That is, $(2,-1)+(1,3)=(1,3)+(2,-1)$. That the order of addition is immaterial is the commutative law of vector addition. Theorem 1.2.19(a) establishes this law in general.

(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 to the right) 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 $(-1\,,3\,,2)$.







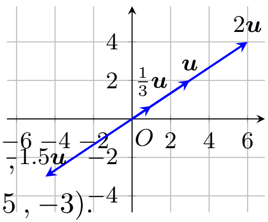
As drawn above, many of the three-D plots in this book are **stereo pair**s, drawing the plot from two slightly different viewpoints: cross your eyes to merge two of the images, and then focus on the pair of plots to see the 3D effect. With practice viewing such 3D stereo pairs becomes less difficult!

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

$\square$

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

(a) Let the vector $\boldsymbol{u} = (3\,,2)$ then, as illustrated to the right,

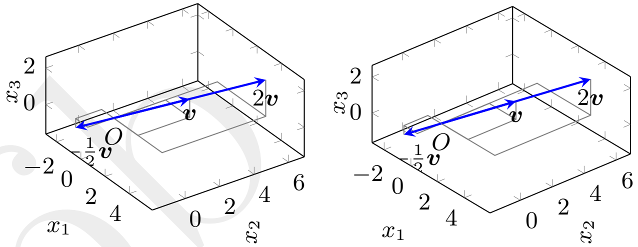


$$2\boldsymbol{u} = 2(3\,,2) = (2\cdot 3\,,2\cdot 2) = (6\,,4)\,,$$
$$\tfrac{1}{3}\boldsymbol{u} = \tfrac{1}{3}(3\,,2) = (\tfrac{1}{3}\cdot 3\,,\tfrac{1}{3}\cdot 2) = (1\,,\tfrac{2}{3})\,,$$
$$(-1.5)\boldsymbol{u} = (-1.5\cdot 3\,,-1.5\cdot 2) = (-4.5\,,-3).$$

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

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

$$(-\tfrac{1}{2})\boldsymbol{v} = -\frac{1}{2}\begin{bmatrix}2\\3\\1\end{bmatrix} = \begin{bmatrix}-\frac{1}{2}\cdot 2\\-\frac{1}{2}\cdot 3\\-\frac{1}{2}\cdot 1\end{bmatrix} = \begin{bmatrix}-1\\-\frac{3}{2}\\-\frac{1}{2}\end{bmatrix}.$$



□

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

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

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

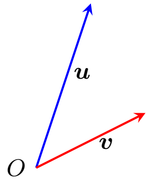
$$\boldsymbol{u}+\boldsymbol{v} := (u_1+v_1, u_2+v_2, \ldots, u_n+v_n).$$

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

$$c\boldsymbol{u} := (cu_1, cu_2, \ldots, cu_n).$$

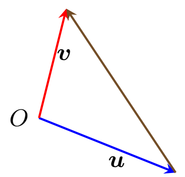
The ***negative*** *of* $\boldsymbol{u}$ *denoted* $-\boldsymbol{u}$*, is defined as the scalar multiple* $-\boldsymbol{u} := (-1)\boldsymbol{u}$*, and is a vector of the same length as* $\boldsymbol{u}$ *but in exactly the opposite direction. The* ***difference*** $\boldsymbol{u} - \boldsymbol{v}$ *is defined as the sum* $\boldsymbol{u} + (-\boldsymbol{v})$ *and is equivalently the vector drawn from the end of* $\boldsymbol{v}$ *to the end of* $\boldsymbol{u}$*.*

Example 1.2.5 For the vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ shown to the right, draw the vectors $\boldsymbol{u}+\boldsymbol{v}$, $\boldsymbol{v}+\boldsymbol{u}$, $\boldsymbol{u}-\boldsymbol{v}$, $\boldsymbol{v}-\boldsymbol{u}$, $\frac{1}{2}\boldsymbol{u}$, and $-\boldsymbol{v}$.

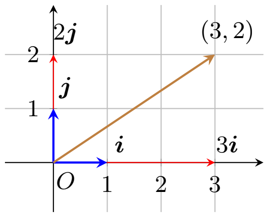


□

Activity 1.2.6 For the vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ shown to the right, what is the result vector, which is also shown?



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



Using vector addition and scalar multiplication, we often write vectors in terms of so-called standard unit vectors. In the plane, as drawn right, are the two unit vectors $\boldsymbol{i}$ and $\boldsymbol{j}$ defined to be of length one and in the direction of the two coordinate axes, respectively. Hence $\boldsymbol{i}=(1,0)$ and $\boldsymbol{j}$=(0,1), as shown. Then, for example,

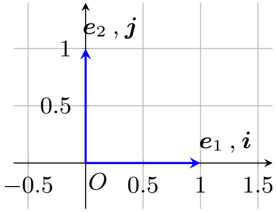
$$\begin{aligned}(3,2) &= (3,0)+(0,2) \quad \text{(by addition)}\\ &= 3(1,0)+2(0,1) \quad \text{(by scalar mult)}\end{aligned}$$

$$= 3\boldsymbol{i} + 2\boldsymbol{j} \quad \text{(by definition of } \boldsymbol{i} \text{ and } \boldsymbol{j}\text{)}.$$

Similarly, in three-dimensional space, we often write vectors in terms of the three vectors $\boldsymbol{i}$, $\boldsymbol{j}$, and $\boldsymbol{k}$, defined to be each of length one, aligned along the three coordinate axes. Hence $\boldsymbol{i} = (1, 0, 0)$, $\boldsymbol{j} = (0, 1, 0)$, and $\boldsymbol{k}$=(0,0,1). For example,

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

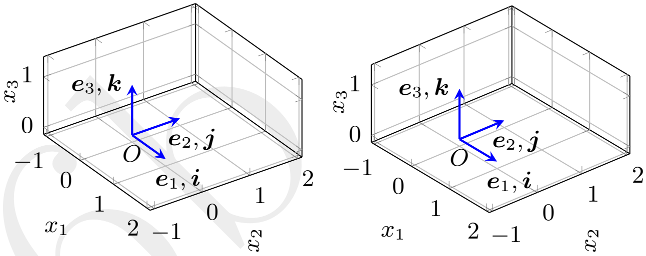
The next definition generalizes these standard unit vectors to vectors in $\mathbb{R}^n$ for every size n.



Definition 1.2.7 *Given a standard coordinate system with n coordinate axes, all at right-angles to each other, the **standard unit vectors** $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ are the vectors of length one in the direction of the corresponding coordinate axis (as illustrated to the right for $\mathbb{R}^2$ and below for $\mathbb{R}^3$). That is,*

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

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



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

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

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

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

(a) $(5\,,0\,,2)$

(b) $(3\,,1)$

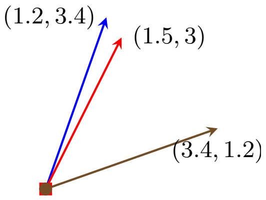
(c) $(0\,,3\,,0\,,0\,,1)$

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

Distance

Defining a 'distance' between vectors empowers us to concisely compare vectors.

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 to the right). Why is the error 0.5? Because the difference between the vectors $(1.5, 3) - (1.2, 3) = (0.3, -0.4)$ has length $\sqrt{0.3^2 + (-0.4)^2} = 0.5$.

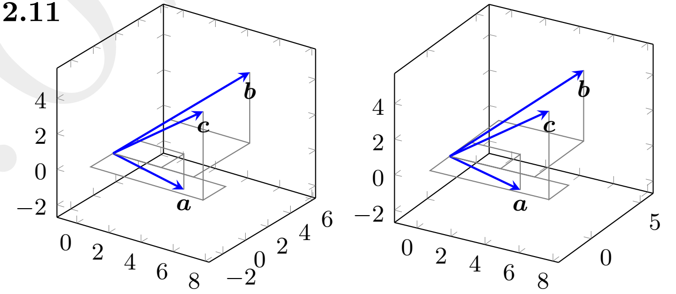


Conversely, we would like to recognize that vectors $(1.2, 3.4)$ and $(3.4, 1.2)$ are very different (as also illustrated)—there is a large 'distance' between them. Why is there a large 'distance'? Because the difference between the vectors $(1.2, 3.4) - (3.4, 1.2) = (-2.2, 2.2)$ has length $\sqrt{(-2.2)^2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

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

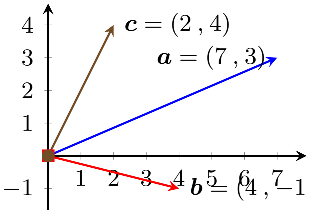
Definition 1.2.10 *The **distance** between vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{R}^n$ is the length of their difference, $|\boldsymbol{u}-\boldsymbol{v}|$.*

Example 1.2.11 Given three vectors $\boldsymbol{a} = 3\boldsymbol{i} + 2\boldsymbol{j} - 2\boldsymbol{k}$, $\boldsymbol{b} = 5\boldsymbol{i}+5\boldsymbol{j}+4\boldsymbol{k}$, and $\boldsymbol{c} = 7\boldsymbol{i} - 2\boldsymbol{j} + 5\boldsymbol{k}$ (shown right in stereo) use the concept of distance between vectors to answer the following: 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? $\boldsymbol{a} = (7\,,3)$, $\boldsymbol{b} = (4\,,-1)$, $\boldsymbol{c} = (2\,,4)$



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

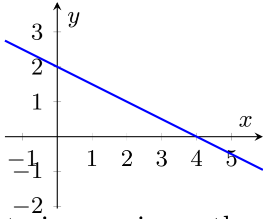
(b) two of the pairs

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

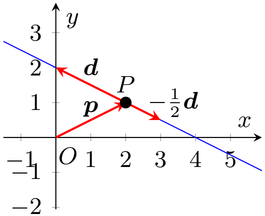
(d) $\boldsymbol{a}$, $\boldsymbol{c}$

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 generalize the descriptions to lines in space, and then in any number of dimensions.

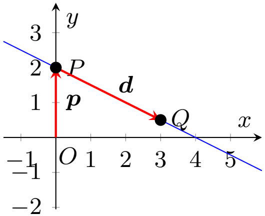


Example 1.2.13 Consider the line drawn to the right in some chosen coordinate system. Recall that one way to find an equation of the line is to find the intercepts with the axes, here at $x = 4$ and at $y = 2$, then write down $\frac{x}{4} + \frac{y}{2} = 1$ as an equation of the line. Algebraic rearrangement 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 to the right. 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 $\boldsymbol{p}$ as drawn. Then, for example, the point $(0, 2)$ on the line has position vector $\boldsymbol{p} + \boldsymbol{d}$ for vector $\boldsymbol{d} = (-2, 1)$ because $\boldsymbol{p} + \boldsymbol{d} = (2, 1) + (-2, 1) = (0, 2)$. Other points on the line are also given using the same vectors, $\boldsymbol{p}$ and $\boldsymbol{d}$: for example, the point $(3, \frac{1}{2})$ has position vector $\boldsymbol{p} - \frac{1}{2}\boldsymbol{d}$ (as drawn) because $\boldsymbol{p} - \frac{1}{2}\boldsymbol{d} = (2, 1) - \frac{1}{2}(-2, 1) = (3, \frac{1}{2})$; and the point $(-2, 3)$ has position vector $\boldsymbol{p} + 2\boldsymbol{d} = (2, 1) + 2(-2, 1)$. In general, every point on the line may be expressed as $\boldsymbol{p} + t\boldsymbol{d}$ for some scalar t.

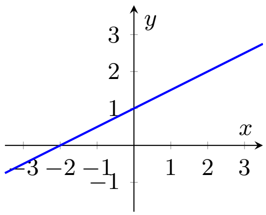
For every given line, there are many possible choices of $\boldsymbol{p}$ and $\boldsymbol{d}$ in such a vector representation. A different looking form, but equally valid, is obtained from any pair of points on the line. For example, one could choose point P to be $(0,2)$ and point Q to be $(3,\frac{1}{2})$,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



- $(2,1)=(0,2)+(2,-1)=(0,2)+\frac{2}{3}(3,-\frac{3}{2})=\boldsymbol{p}+\frac{2}{3}\boldsymbol{d}$;
- $(6,-1)=(0,2)+(6,-3)=(0,2)+2(3,-\frac{3}{2})=\boldsymbol{p}+2\boldsymbol{d}$;
- $(-1,\frac{5}{2})=(0,2)+(-1,\frac{1}{2})=(0,2)-\frac{1}{3}(3,-\frac{3}{2})=\boldsymbol{p}-\frac{1}{3}\boldsymbol{d}$.

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 to the right?



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

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

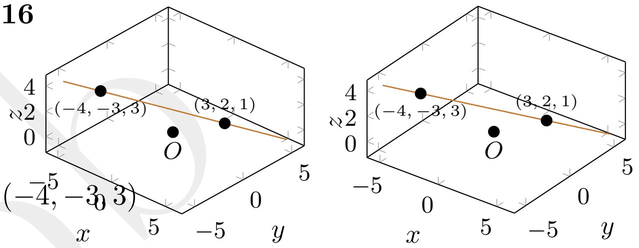
(c) $(2\,,2)+(1\,,1/2)t$

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

Definition 1.2.15 *A **parametric equation** of a line is $\boldsymbol{x}=\boldsymbol{p}+t\boldsymbol{d}$ where $\boldsymbol{p}$ 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 $\boldsymbol{x}$ 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 to the right 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 is $\boldsymbol{x} = (-4 + 2t\ , 3 - t\ , -1 - 4t)$, 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 that the vector properties follow directly from the familiar properties of arithmetic operations on scalars.

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

(a) $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$ (commutative law);

(b) $(\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w} = \boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w})$ (associative law);

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

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

(e) $a(\boldsymbol{u} + \boldsymbol{v}) = a\boldsymbol{u} + a\boldsymbol{v}$ (a distributive law);

(f) $(a + b)\boldsymbol{u} = a\boldsymbol{u} + b\boldsymbol{u}$ (a distributive law);

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

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

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

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

$\square$

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

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

(a) $\boldsymbol{u}+\boldsymbol{v}=\boldsymbol{v}+\boldsymbol{u}$ *(commutative law);*

(b) $(\boldsymbol{u}+\boldsymbol{v})+\boldsymbol{w}=\boldsymbol{u}+(\boldsymbol{v}+\boldsymbol{w})$ *(associative law);*

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

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

(e) $a(\boldsymbol{u}+\boldsymbol{v})=a\boldsymbol{u}+a\boldsymbol{v}$ *(a distributive law);*

(f) $(a+b)\boldsymbol{u}=a\boldsymbol{u}+b\boldsymbol{u}$ *(a distributive law);*

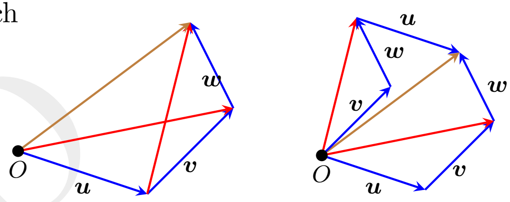
(g) $(ab)\boldsymbol{u}=a(b\boldsymbol{u})$*;*

(h) $1\boldsymbol{u}=\boldsymbol{u}$*;*

(i) $0\boldsymbol{u}=\boldsymbol{0}$*;*

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

Example 1.2.20 Which of these two diagrams best illustrates the associative law 1.2.19(b)? 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 $\boldsymbol{x}$ such that $3\boldsymbol{x} - 2\boldsymbol{u} = 6\boldsymbol{v}$. □

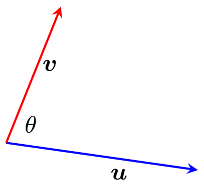
Example 1.2.22 Rearrange $3\boldsymbol{x} - \boldsymbol{a} = 2(\boldsymbol{a} + \boldsymbol{x})$ to write vector $\boldsymbol{x}$ in terms of $\boldsymbol{a}$: give excruciating detail of the justification using Theorem 1.2.19. □

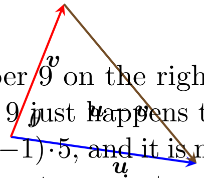
1.3 The dot product determines angles and lengths

Section contents

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. There are at least four ways of multiplying vectors together: each way useful in appropriate circumstances. Often the angle between vectors is denoted by the Greek letter theta, θ. This section introduces one such multiplication, the so-called dot product of two vectors that, among other attributes, gives a valuable way to determine the angle between two vectors.

Example 1.3.1 Consider the two vectors $\boldsymbol{u} = (7, -1)$ and $\boldsymbol{v} = (2, 5)$ plotted first to the right. What is the angle θ between the two vectors? □





The interest in this Example 1.3.1 is the number 9 on the right-hand side of $|\boldsymbol{u}||\boldsymbol{v}|\cos\theta = 9$. 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 $\boldsymbol{u} = (u_1, u_2) = (7, -1)$ and $\boldsymbol{v} = (v_1, v_2) = (2, 5)$. Repeat this example for many pairs of vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ to find that always $|\boldsymbol{u}||\boldsymbol{v}|\cos\theta = u_1v_1 + u_2v_2$ (**??**). This equality suggests that the sum of products of corresponding components of $\boldsymbol{u}$ and $\boldsymbol{v}$ is closely connected to the angle between the vectors.

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

$$\boldsymbol{u} \cdot \boldsymbol{v} := u_1 v_1 + u_2 v_2 + \cdots + u_n v_n .$$

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: "$\boldsymbol{uv}$" is meaningless.

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

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

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

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

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

□

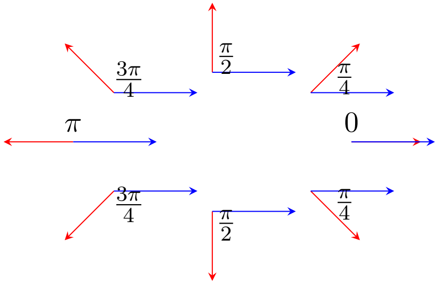
Activity 1.3.4 What is the dot product of the two vectors $\boldsymbol{u} = 2\boldsymbol{i} - \boldsymbol{j}$ and $\boldsymbol{v} = 3\boldsymbol{i} + 4\boldsymbol{j}$?

(a) 2 (b) 5 (c) 8 (d) 10

Theorem 1.3.5 *For every two nonzero vectors* $\boldsymbol{u}$ *and* $\boldsymbol{v}$ *in* $\mathbb{R}^n$, *the **angle*** θ *between the vectors is determined by*

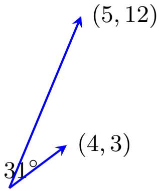
$$\cos\theta = \frac{\boldsymbol{u} \cdot \boldsymbol{v}}{|\boldsymbol{u}||\boldsymbol{v}|}, \quad 0 \le \theta \le \pi \quad (0 \le \theta \le 180^\circ).$$

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 $90°$; when they point in opposite directions the angle is π, or equivalently $180°$. Let's prove this theorem after some examples.

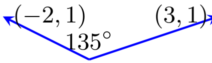


Example 1.3.6 Determine the angle between the following pairs of vectors.

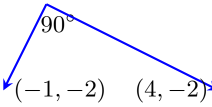
(a) $(4, 3)$ and $(5, 12)$



(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 $(\sqrt{3}\,,1)$?

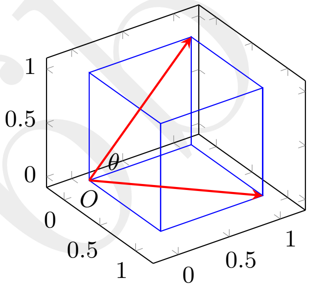
(a) 64.34° (b) 30° (c) 77.50° (d) 60°

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.

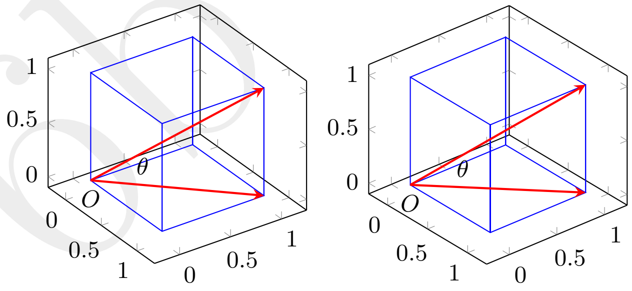
Table 1.1: When a cosine is one of these tabulated special values, then we know the corresponding angle exactly. In other cases, we usually use a calculator (`arccos` or `cos`$^{-1}$) or computer (`acos()`) to compute the angle numerically.

θ	θ	$\cos\theta$	$\cos\theta$
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
π	180°	-1	$-1.$

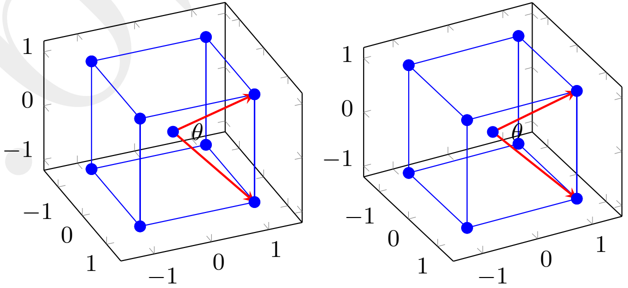
(a) Consider the cube drawn in stereo below, and compute the angle between the diagonals on two adjacent faces.



(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-centred cubic lattice (such as that formed by caesium chloride crystals) has one lattice point in the centre of the unit cell as well as the eight corner points. Consider the body-centred cube of atoms drawn in stereo below with the centre of the cube at the origin: what is the angle between the centre 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,

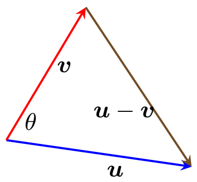
(a) "The dog sat on the mat" has word vector $\boldsymbol{a} = (0, 1, 1, 1, 0)$.

(b) "The cat scratched the dog" has word vector $\boldsymbol{b} = (1, 1, 0, 0, 1)$.

(c) "The cat and dog sat on the mat" has word vector $\boldsymbol{c} = (1, 1, 1, 1, 0)$.

Use the angle between these three word vectors to characterize 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

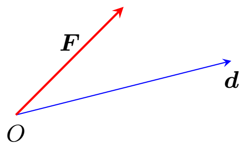


In physics and engineering, "work" has a precise meaning related to energy: when a force of magnitude F acts on a body and that body moves a distance d, then the work done by the force is $W = Fd$. However, this formula applies only for one-dimensional force and displacement, the case when the force and the displacement are 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

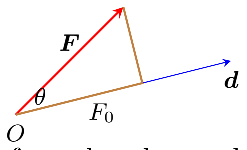
$$W = F \cdot d = (5 \cdot 9.8) \cdot 2 = 98 \text{ joules.}$$

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 to the right). Consider the general case when a vector force $\boldsymbol{F}$ acts on a body which moves a displacement vector $\boldsymbol{d}$.

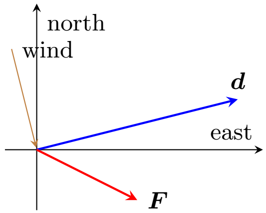


Then the work done by the force on the body is the length of the displacement times the component of the force in the direction of the displacement—the component of the force at right-angles to the displacement does no work.



As illustrated to the right, draw a right-angled triangle to decompose the force $\boldsymbol{F}$ 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 $W = F_0|\boldsymbol{d}|$. As drawn, the force $\boldsymbol{F}$ makes an angle θ to the displacement $\boldsymbol{d}$: the dot product determines this angle via $\cos\theta = (\boldsymbol{F}\cdot\boldsymbol{d})/(|\boldsymbol{F}||\boldsymbol{d}|)$ (Theorem 1.3.5). By basic trigonometry, the adjacent side of the force triangle has length $F_0 = |\boldsymbol{F}|\cos\theta = |\boldsymbol{F}|\frac{\boldsymbol{F}\cdot\boldsymbol{d}}{|\boldsymbol{F}||\boldsymbol{d}|} = \frac{\boldsymbol{F}\cdot\boldsymbol{d}}{|\boldsymbol{d}|}$. Finally, the work done $W = F_0|\boldsymbol{d}| = \frac{\boldsymbol{F}\cdot\boldsymbol{d}}{|\boldsymbol{d}|}|\boldsymbol{d}| = \boldsymbol{F}\cdot\boldsymbol{d}$: that is, the work done is 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 to the right. The wind from abeam, of strength and direction $(1, -4)$ m/s, generates a force $\boldsymbol{F} = (20, -10)$ (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, $9.8\,\mathrm{m/s^2}$. 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) 19.6 joules
(b) 58.8 joules
(c) 98 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 Section 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 $\boldsymbol{u} = (-2, 5, -2)$, $\boldsymbol{v} = (3, 3, -2)$ and $\boldsymbol{w} = (2, 0, -5)$, and scalar $a = 2$, verify that (Theorems 1.3.13(c) and 1.3.13(d))

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

□

Theorem 1.3.13 (dot properties) *For every three vectors $\boldsymbol{u}$, $\boldsymbol{v}$, and $\boldsymbol{w}$ in $\mathbb{R}^n$, and for every scalar a, the following properties hold:*

(a) $\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{v} \cdot \boldsymbol{u}$ *(commutative law);*

(b) $\boldsymbol{u}\cdot\boldsymbol{0}=\boldsymbol{0}\cdot\boldsymbol{u}=0$*;*

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

(d) $(\boldsymbol{u}+\boldsymbol{v})\cdot\boldsymbol{w}=\boldsymbol{u}\cdot\boldsymbol{w}+\boldsymbol{v}\cdot\boldsymbol{w}$ *(distributive law);*

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

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

(a) $(2\boldsymbol{u})\cdot(2\boldsymbol{v})=2(\boldsymbol{u}\cdot\boldsymbol{v})$

(b) $\boldsymbol{u}\cdot\boldsymbol{v}-\boldsymbol{v}\cdot\boldsymbol{u}=0$

(c) $\boldsymbol{u}\cdot(\boldsymbol{v}+\boldsymbol{w})=\boldsymbol{u}\cdot\boldsymbol{v}+\boldsymbol{u}\cdot\boldsymbol{w}$

(d) $(\boldsymbol{u}-\boldsymbol{v})\cdot(\boldsymbol{u}+\boldsymbol{v})=\boldsymbol{u}\cdot\boldsymbol{u}-\boldsymbol{v}\cdot\boldsymbol{v}$

The above proof of Theorem 1.3.13(e), that $\boldsymbol{u}\cdot\boldsymbol{u}=0$ if and only if $\boldsymbol{u}=\boldsymbol{0}$, 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 $\boldsymbol{u} = (3, 4)$ and $\boldsymbol{v} = (2, 1)$ verify the following three properties:

(a) $\sqrt{\boldsymbol{u} \cdot \boldsymbol{u}} = |\boldsymbol{u}|$, the length of $\boldsymbol{u}$;

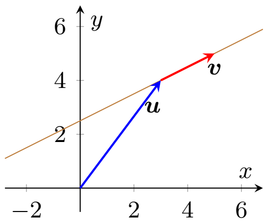
(b) $|\boldsymbol{u} \cdot \boldsymbol{v}| \leq |\boldsymbol{u}||\boldsymbol{v}|$ (Cauchy–Schwarz inequality);

(c) $|\boldsymbol{u} + \boldsymbol{v}| \leq |\boldsymbol{u}| + |\boldsymbol{v}|$ (triangle inequality).

□

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

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 $\boldsymbol{u} = (3, 4)$ and $\boldsymbol{v} = (2, 1)$. Then con-



sider the line given parametrically (Definition 1.2.15) as the position vectors $\boldsymbol{x} = \boldsymbol{u} + t\boldsymbol{v} = (3+2t, 4+t)$ for scalar parameter t—illustrated to the right. The position vector $\boldsymbol{x}$ of any point on the line has length ℓ (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 $\boldsymbol{x}$ 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 $a = 5 = |\boldsymbol{v}|^2$, $c = 25 = |\boldsymbol{u}|^2$, and $b = 20 = 2\cdot 10 = 2(\boldsymbol{u}\cdot\boldsymbol{v})$. So $b^2 - 4ac < 0$, rewritten as $\frac{1}{4}b^2 < ac$, becomes the statement that $\frac{1}{4}[2(\boldsymbol{u}\cdot\boldsymbol{v})]^2 = (\boldsymbol{u}\cdot\boldsymbol{v})^2 < |\boldsymbol{v}|^2|\boldsymbol{u}|^2$. 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* $\boldsymbol{u}$ *and* $\boldsymbol{v}$ *in* $\mathbb{R}^n$ *the following properties hold:*

(a) $\sqrt{\boldsymbol{u} \cdot \boldsymbol{u}} = |\boldsymbol{u}|$, *the length of* $\boldsymbol{u}$*;*

(b) $|\boldsymbol{u} \cdot \boldsymbol{v}| \leq |\boldsymbol{u}||\boldsymbol{v}|$ *(**Cauchy–Schwarz inequality**);*

(c) $|\boldsymbol{u} \pm \boldsymbol{v}| \leq |\boldsymbol{u}| + |\boldsymbol{v}|$ *(**triangle inequality**).*

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

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 when the vectors are aligned with each other, and when the vectors are at right-angles to each other. Recall Theorem 1.3.5 gives the angle θ between two vectors via $\cos\theta = \frac{\boldsymbol{u}\cdot\boldsymbol{v}}{|\boldsymbol{u}||\boldsymbol{v}|}$. For vectors at right-angles $\theta = 90^\circ$, so $\cos\theta = 0$, and hence nonzero vectors are at right-angles only when the dot product $\boldsymbol{u}\cdot\boldsymbol{v} = 0$. We give a special name to vectors at right-angles.

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

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 *nonzero* vectors. Often the requirement that the orthogonal vectors are nonzero 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 $\boldsymbol{i}$, $\boldsymbol{j}$, and $\boldsymbol{k}$ in $\mathbb{R}^3$:

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

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? $\boldsymbol{u} = (-1, 1, -3, 0)$, $\boldsymbol{v} = (2, 4, 2, -6)$, and $\boldsymbol{w} = (-1, 6, -2, 3)$. □

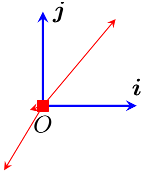
Activity 1.3.22 Which pair of the following three vectors are orthogonal to each other? $\boldsymbol{x} = \boldsymbol{i} - 2\boldsymbol{k}$, $\boldsymbol{y} = -3\boldsymbol{i} - 4\boldsymbol{j}$, $\boldsymbol{z} = -\boldsymbol{i} - 2\boldsymbol{j} + 2\boldsymbol{k}$

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

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

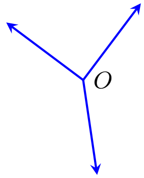
Key properties The next couple of innocuous looking theorems are vital keys to important results in subsequent chapters.

To introduce the first theorem, consider the 2D plane and try to draw a nonzero vector at right-angles to both the two standard unit vectors $\boldsymbol{i}$ and $\boldsymbol{j}$. The red vectors to the right illustrate failed attempts to draw a nonzero vector at right-angles to both $\boldsymbol{i}$ and $\boldsymbol{j}$. 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 nonzero 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 picture to the right 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

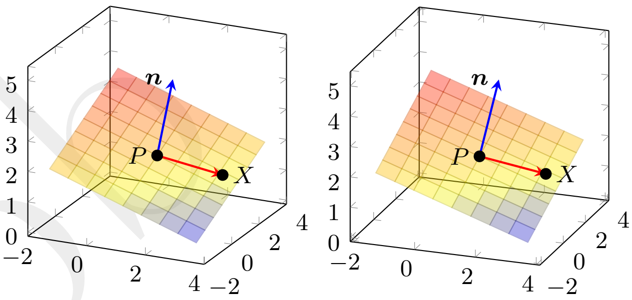
This section uses the dot product to find equations of a plane in 3D. The key is to write points in the plane as all those at right-angles to a certain direction. This direction is perpendicular to the required plane, and is called a normal. Let's start with an example of the idea in 2D.

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).

$\square$

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

(a)	(b)	(c)	(d)
$4x+2y=10$	$x+3y=10$	$4x+y=11$	$2x+3y=11$

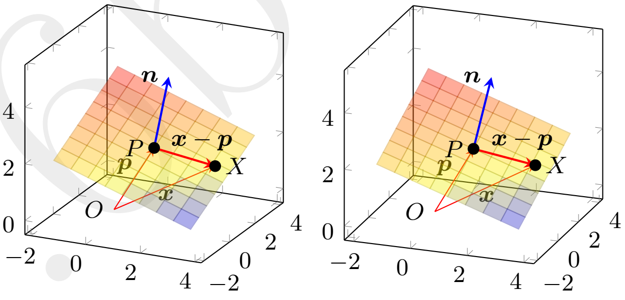


Now use the same approach to finding an equation of a plane in 3D. The problem is to find the equation of the plane that goes through a given point P and is perpendicular to a given vector $\boldsymbol{n}$, called a **normal vector**. As illustrated in stereo, that means to find all points X such that $\overrightarrow{PX}$ is orthogonal to $\boldsymbol{n}$. Denote the position vector of P by $\boldsymbol{p} = (x_0, y_0, z_0)$, the position vector of X by $\boldsymbol{x} = (x, y, z)$, and let the normal vector be $\boldsymbol{n} = (a, b, c)$. Then, as drawn below, the displacement vector $\overrightarrow{PX} = \boldsymbol{x} - \boldsymbol{p} = (x - x_0, y - y_0, z - z_0)$ and so for $\overrightarrow{PX}$ to be orthogonal to $\boldsymbol{n}$ requires $\boldsymbol{n} \cdot (\boldsymbol{x} - \boldsymbol{p}) = 0$; 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 $\boldsymbol{n} = (1, -1, 3)$. (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:

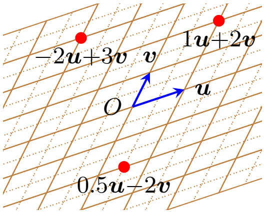
(a) $3x - 6y + 2z = 4$; (b) $z = 0.2x - 3.3y - 1.9$.

□

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

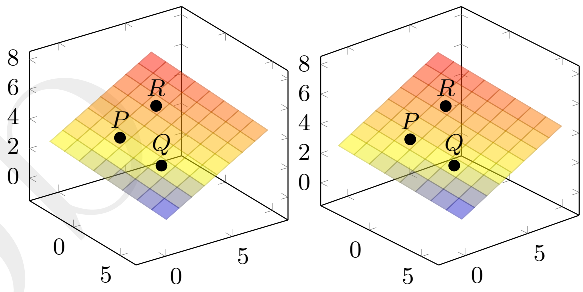
(a) $(1, 2, 4)$
(b) $(-1, 1, 2)$
(c) $(1, 2, 1)$
(d) none of these

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 (Section 1.2.2). Such a parametric representation generalizes to every dimension (Section 2.3).



The basic idea, as illustrated to the right, 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 $\boldsymbol{u}$ and $\boldsymbol{v}$ shown to the right, illustrated are the points $\boldsymbol{u}+2\boldsymbol{v}$, $\frac{1}{2}\boldsymbol{u}-2\boldsymbol{v}$, and $-2\boldsymbol{u}+3\boldsymbol{v}$. Similarly, all points in the plane have a position vector in the form $s\boldsymbol{u}+t\boldsymbol{v}$ for some scalar parameters s and t. The grid shown to the right illustrates the sum of integral and half-integral multiples. The formula $\boldsymbol{x}=s\boldsymbol{u}+t\boldsymbol{v}$ 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 to the right in stereo.

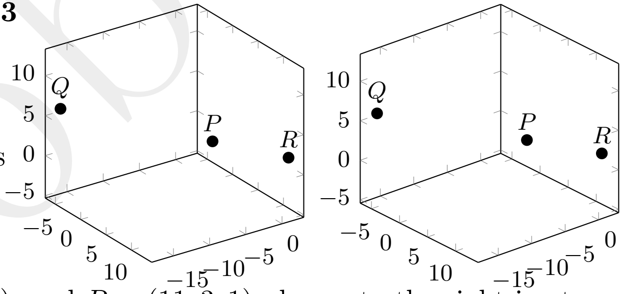


□

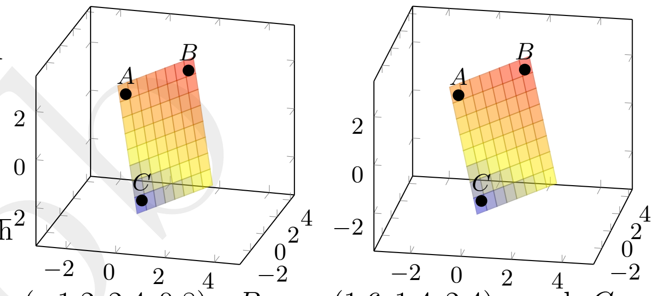
Definition 1.3.32 *A* ***parametric equation*** *of a plane is* $\boldsymbol{x} = \boldsymbol{p} + s\boldsymbol{u} + t\boldsymbol{v}$ *where* $\boldsymbol{p}$ *is the position vector of some point in the plane, the two vectors* $\boldsymbol{u}$ *and* $\boldsymbol{v}$ *are parallel to the plane* ($\boldsymbol{u},\boldsymbol{v} \neq \boldsymbol{0}$ *and are at a nonzero/non-*π *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 use 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 to the right in stereo.



$\square$



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 to the right in stereo.

□

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

(a) $(3s + 2t\,, 4 + 2s + t\,, 4 + 3t)$

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

(c) $\boldsymbol{i} + s\boldsymbol{j} + t\boldsymbol{k}$

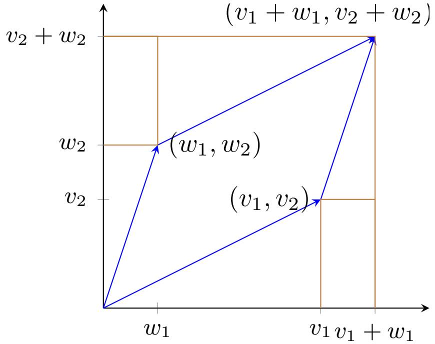
(d) $(-1\,,1\,,-1)s+(4\,,2\,,-1)t$

1.4 The cross product

The dot product of Section 1.3 is not the only way to multiply vectors. In the three dimensions of the world we live in there is a second way to multiply vectors, called the cross product. But for more than three dimensions, qualitatively different techniques are developed in subsequent chapters.

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

Area of a parallelogram

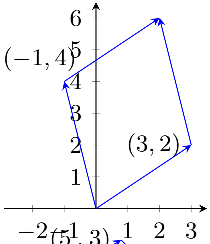


Consider the parallelogram drawn in blue. It has sides given by vectors $\boldsymbol{v} = (v_1, v_2)$ and $\boldsymbol{w} = (w_1, w_2)$, as shown. Let's determine the area of the parallelogram. Its area is the containing rectangle less the two small rectangles and the four small triangles. The two small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

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

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

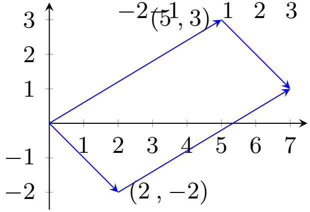
Example 1.4.1 What is the area of the parallelogram (illustrated to the right) whose edges are formed by the vectors $(3, 2)$ and $(-1, 4)$?



□

Activity 1.4.2 What is the area of the parallelogram (illustrated to the right) whose edges are formed by the vectors $(5\,,3)$ and $(2\,,-2)$?

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



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 Section 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 $\boldsymbol{x} = (1, 1, 2) + (1, 1, 0)s + (0, 3, 1)t$. These determine the same plane; they are 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 $\boldsymbol{x} = (1, 1, 2) + (1, 1, 0)s + (0, 3, 1)t$ 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 $\boldsymbol{x} = (4\,, -1\,, -2) + (1\,, -2\,, 1)s + (2\,, -3\,, -2)t$. Which of the following is your computed normal vector?

(a) $(-4\,, 4\,, -10)$

(b) $(5\,, 6\,, 7)$

(c) $(2\,, -2\,, 5)$

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

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 $\boldsymbol{v} = (v_1, v_2, v_3)$ and $\boldsymbol{w} = (w_1, w_2, w_3)$. The procedure computes

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

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

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

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

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

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

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

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

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

Example 1.4.6 Among the standard unit vectors, derive that

(a) $\boldsymbol{i} \times \boldsymbol{j} = \boldsymbol{k}$,

(b) $\boldsymbol{j} \times \boldsymbol{i} = -\boldsymbol{k}$,

(c) $\boldsymbol{j} \times \boldsymbol{k} = \boldsymbol{i}$,

(d) $\boldsymbol{k} \times \boldsymbol{j} = -\boldsymbol{i}$,

(e) $\boldsymbol{k} \times \boldsymbol{i} = \boldsymbol{j}$,

(f) $\boldsymbol{i} \times \boldsymbol{k} = -\boldsymbol{j}$,

(g) $\boldsymbol{i} \times \boldsymbol{i} = \boldsymbol{j} \times \boldsymbol{j} = \boldsymbol{k} \times \boldsymbol{k} = \boldsymbol{0}$.

□

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

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

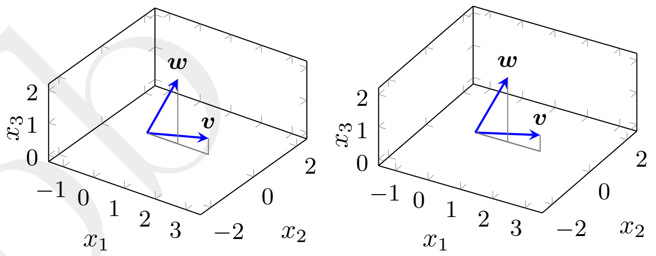
Geometry of a cross product

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 $\boldsymbol{v} = (v_1, v_2, 0)$ and $\boldsymbol{w} = (w_1, w_2, 0)$. 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

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

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

Activity 1.4.9 Using property 1.4.10(b) of the next theorem, in which direction is the cross product $\boldsymbol{v} \times \boldsymbol{w}$ for the two vectors illustrated in stereo to the right?



(a) $+\boldsymbol{j}$

(b) $-\boldsymbol{i}$

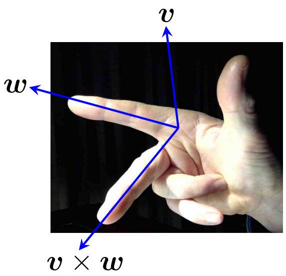
(c) $+\boldsymbol{i}$

(d) $-\boldsymbol{j}$

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

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

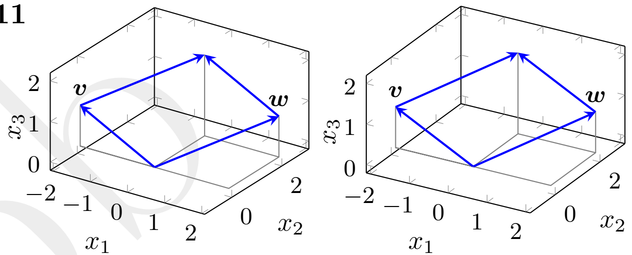
(b) *the direction of $\boldsymbol{v} \times \boldsymbol{w}$ is in the* ***right-hand sense****, in that if $\boldsymbol{v}$ is in the direction of your thumb, and $\boldsymbol{w}$ is in the direction of your straight index finger, then $\boldsymbol{v} \times \boldsymbol{w}$ is in the direction of your bent second/longest finger—all on your right hand as illustrated to the right;*



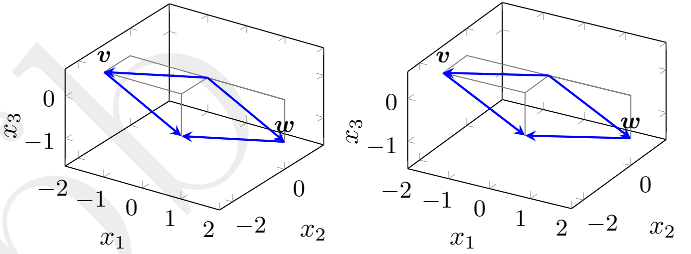
(c) $|\boldsymbol{v} \times \boldsymbol{w}| = |\boldsymbol{v}|\,|\boldsymbol{w}| \sin\theta$ *where θ is the angle between vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ ($0 \leq \theta \leq \pi$, equivalently $0^\circ \leq \theta \leq 180^\circ$); and*

(d) *the length $|\boldsymbol{v} \times \boldsymbol{w}|$ is the area of the*

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



Activity 1.4.12 What is the area of the parallelogram



(in stereo to the right) with edges formed by vectors $\boldsymbol{v} = (-2\,, 1\,, 0)$ and $\boldsymbol{w} = (2\,, 0\,, -1)$?

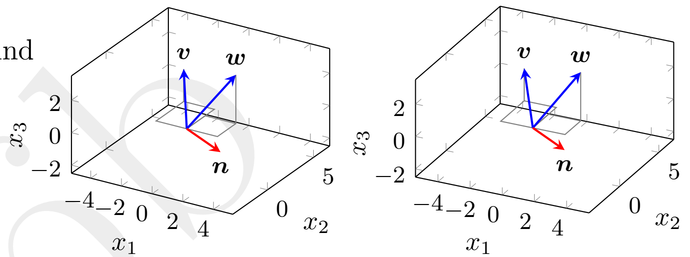
(a) 3

(b) 5

(c) $\sqrt{5}$

(d) 1

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



□

Algebraic properties of a cross product

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

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

(a) $\boldsymbol{v} \times \boldsymbol{v} = \boldsymbol{0}$;

(b) $\boldsymbol{w} \times \boldsymbol{v} = -(\boldsymbol{v} \times \boldsymbol{w})$ *(not commutative);*

(c) $(c\boldsymbol{v}) \times \boldsymbol{w} = c(\boldsymbol{v} \times \boldsymbol{w}) = \boldsymbol{v} \times (c\boldsymbol{w})$;

(d) $\boldsymbol{u} \times (\boldsymbol{v} + \boldsymbol{w}) = \boldsymbol{u} \times \boldsymbol{v} + \boldsymbol{u} \times \boldsymbol{w}$ *(distributive law).*

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

Example 1.4.16 Given $(\boldsymbol{i}+\boldsymbol{j}+\boldsymbol{k})\times(-2\boldsymbol{i}-\boldsymbol{j})=\boldsymbol{i}-2\boldsymbol{j}+\boldsymbol{k}$, what is $(3\boldsymbol{i}+3\boldsymbol{j}+3\boldsymbol{k})\times(-2\boldsymbol{i}-\boldsymbol{j})$? □

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

$$\begin{aligned}\boldsymbol{u}\times\boldsymbol{v}&=-9\boldsymbol{i}+8\boldsymbol{j}-3\boldsymbol{k}\,,\\ \boldsymbol{u}\times\boldsymbol{w}&=-3\boldsymbol{i}-7\boldsymbol{j}-\boldsymbol{k}\,,\\ \boldsymbol{v}\times\boldsymbol{w}&=-8\boldsymbol{i}-9\boldsymbol{j}+7\boldsymbol{k}\,.\end{aligned}$$

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

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

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

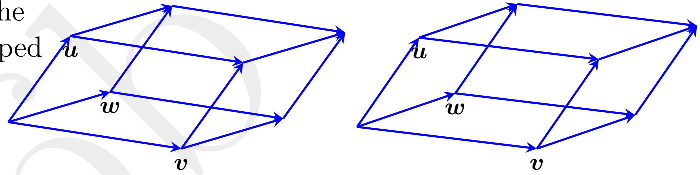
(c) $-11\boldsymbol{i}-16\boldsymbol{j}+6\boldsymbol{k}$

(d) $-12\boldsymbol{i}+\boldsymbol{j}-4\boldsymbol{k}$

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

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 $(\boldsymbol{i} - \boldsymbol{j}) \times (4\boldsymbol{i} + 2\boldsymbol{k})$. $\square$

Volume of a parallelepiped

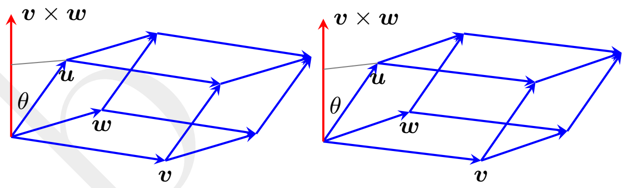


Consider the parallelepiped with edges formed by three vectors $\boldsymbol{u}$, $\boldsymbol{v}$, and $\boldsymbol{w}$ in $\mathbb{R}^3$, as illustrated in stereo to the right. Our challenge is to derive that the volume of the parallelepiped is $|\boldsymbol{u} \cdot (\boldsymbol{v} \times \boldsymbol{w})|$.

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 $\boldsymbol{v}$ and $\boldsymbol{w}$. Hence the base has area $|\boldsymbol{v} \times \boldsymbol{w}|$ (Theorem 1.4.10(d)).

- The height of the parallelepiped is then that part of $\boldsymbol{u}$ in the direction of a normal vector to $\boldsymbol{v}$ and $\boldsymbol{w}$. We know that $\boldsymbol{v}\times\boldsymbol{w}$ is orthogonal to both $\boldsymbol{v}$ and $\boldsymbol{w}$ (Theorem 1.4.10(a)), so by trigonometry the height must be $|\boldsymbol{u}|\cos\theta$ for angle θ between $\boldsymbol{u}$ and $\boldsymbol{v}\times\boldsymbol{w}$, as illustrated.



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

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

The height of the parallelepiped is then

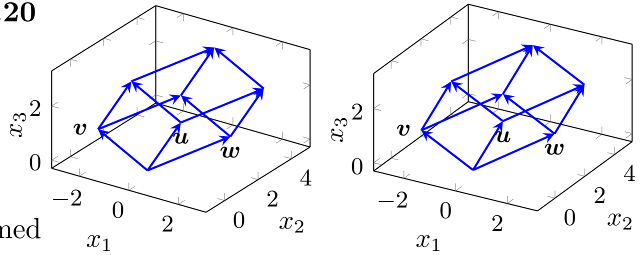
$$|\boldsymbol{u}||\cos\theta| = |\boldsymbol{u}|\frac{|\boldsymbol{u}\cdot(\boldsymbol{v}\times\boldsymbol{w})|}{|\boldsymbol{u}||\boldsymbol{v}\times\boldsymbol{w}|} = \frac{|\boldsymbol{u}\cdot(\boldsymbol{v}\times\boldsymbol{w})|}{|\boldsymbol{v}\times\boldsymbol{w}|}.$$

Consequently, the volume of the parallelepiped equals

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

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

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



□

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

$$
\begin{aligned}
\boldsymbol{v}\cdot(\boldsymbol{u}\times\boldsymbol{w}) &= \begin{vmatrix} -2 & 0 & 2 \\ 0 & 2 & 2 \\ 1 & 1 & 1 \end{vmatrix} \\
&= -2\begin{vmatrix} -2 & 0 & 2 \\ 0 & 2 & 2 \\ 1 & 1 & 1 \end{vmatrix} - 0\begin{vmatrix} -2 & 0 & 2 \\ 0 & 2 & 2 \\ 1 & 1 & 1 \end{vmatrix} + 1\begin{vmatrix} -2 & 0 & 2 \\ 0 & 2 & 2 \\ 1 & 1 & 1 \end{vmatrix} \\
&= -2\begin{vmatrix} 2 & 2 \\ 1 & 1 \end{vmatrix} - 0\begin{vmatrix} 0 & 2 \\ 1 & 1 \end{vmatrix} + 1\begin{vmatrix} 0 & 2 \\ 2 & 2 \end{vmatrix} \\
&= -2\begin{vmatrix} 2 & 2 \\ 1 & 1 \end{vmatrix} - 0\begin{vmatrix} 0 & 2 \\ 1 & 1 \end{vmatrix} + 1\begin{vmatrix} 0 & 2 \\ 2 & 2 \end{vmatrix} \\
&= -2(2\cdot 1 - 1\cdot 2) - 0(0\cdot 1 - 1\cdot 2) + 1(0\cdot 2 - 2\cdot 2) \\
&= -2\cdot 0 - 0(-2) + 1(-4) = -4\,.
\end{aligned}
$$

Hence the parallelepiped formed by $\boldsymbol{u}$, $\boldsymbol{v}$, and $\boldsymbol{w}$ 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 $\boldsymbol{u}$, $\boldsymbol{v}$, and $\boldsymbol{w}$. Chapter 6 shows that this computation of volume generalizes 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

> 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 either of the computer packages Matlab or Octave to perform calculations that would be tedious and error prone if 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 you 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 called `norm()` to compute the length/magnitude of the following vectors (Definition 1.1.9). Generally, computer commands and their output are typeset in the special `fixed width font`, such as this command `norm()`.

(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 summarize* 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 use 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 semicolons). Use* n *numbers separated by semicolons 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*

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

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

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

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)	(b)	(c)	(d)

`2.8000`
`4.3000`
`-7.1000`
	`2.2000`
	`7.4000`
	`-9.0000`
		`Error using *`
		`Inner matrix`
		`dimensions must`
		`agree.`
			`-0.6000`
			`3.1000`
			`1.9000`

Activity 1.5.6 For the vectors of the previous Activity 1.5.5: which is the result of typing the command `2*u`? which is the result of typing the command `u*v`?

Example 1.5.7 Use MATLAB/Octave to compute the angles between the pair of vectors $(4,3)$ and $(5,12)$ (Theorem 1.3.5). □

Example 1.5.8 Verify the distributive law for the dot product $(\boldsymbol{u}+\boldsymbol{v})\cdot\boldsymbol{w}=\boldsymbol{u}\cdot\boldsymbol{w}+\boldsymbol{v}\cdot\boldsymbol{w}$ (Theorem 1.3.13(d)) for vectors

$\boldsymbol{u} = (-0.1\,,\,-3.1\,,\,-2.9\,,\,-1.3)$, $\boldsymbol{v} = (-3\,,\,0.5\,,\,6.4\,,\,-0.9)$, and $\boldsymbol{w} = (-1.5\,,\,-0.2\,,\,0.4\,,\,-3.1)$. □

Activity 1.5.9 Given two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ that have already been typed into Matlab/Octave, which of the following expressions could check the identity that $(\boldsymbol{u}-2\boldsymbol{v})\cdot(\boldsymbol{u}+\boldsymbol{v}) = \boldsymbol{u}\cdot\boldsymbol{u}-\boldsymbol{u}\cdot\boldsymbol{v}-2\boldsymbol{v}\cdot\boldsymbol{v}$?

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

Many other books (e.g., Quarteroni & Saleri 2006, §§1.1–3) 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.

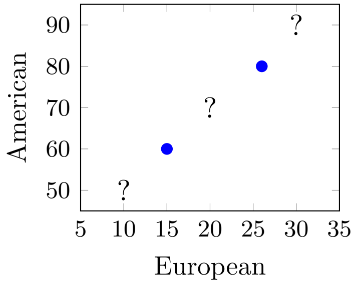
Charles Babbage

2 Systems of linear equations

Chapter contents

Linear relationships are commonly identified in science, engineering, and artificial intelligence. Such relationships 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 'artificial intelligence'—although the algebraic relation is then typically discussed as a computational procedure. But always appearing within such tasks are linear equations to be solved.



Example 2.0.1 *(scientific inference)* Two colleagues, an American and a European, discuss the weather; in particular, they discuss the temperature. (I am sure you can guess where we are going with this example, but let's pretend we do not know.) The American says "yesterday the temperature was 80° but today is much cooler at 60°". The European says, "that's not what I heard, I heard the temperature was 26° and today is 15°". (The graph to the right plots these two data points.) "Hmmmm, 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

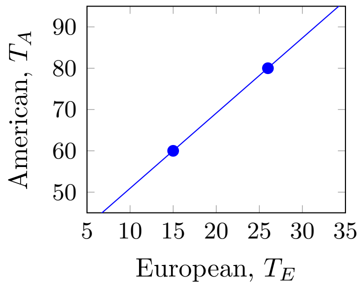
the marginal figure). Let's assume that, in terms of the European temperature T_E, the American temperature $T_A = T_E \cdot c + d$ for some constants c and d that they and we aim to find. The two days of data then give that

$$80 = 26c + d \quad \text{and} \quad 60 = 15c + d\,.$$

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.);
- using 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 $T_A = 1.82\,T_E + 32.73$ (as plotted to the right). 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 $T_A = 1.82\,T_E + 32.73$ 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. □

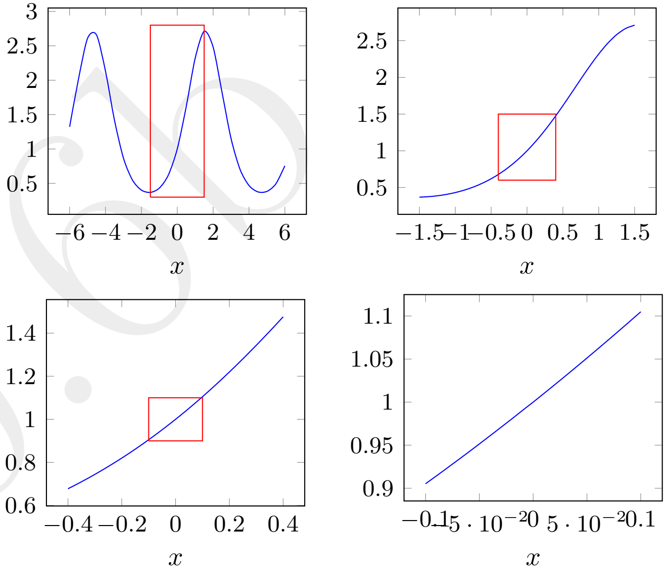


Figure 2.1: Zoom in anywhere on any smooth nonlinear curve, such as the plotted $f(x)$, and we discover that the curve looks like a straight line on the microscale. The (red) rectangles show the region plotted in the next graph in the sequence.

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 may discover that every smooth curve looks like a straight line on its microscale. Thus we may view 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 relationships—for example, microscale linearity underwrites all of calculus.

2.1 Introduction to systems of linear equations

The great aspect of linear equations is that we straightforwardly manipulate them algebraically to deduce results: some results are immensely useful not only 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

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

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

sides to give $-\frac{4}{5}T_C=32$; multiply both sides by $-\frac{5}{4}$, then $T_C=-\frac{5}{4}\times 32=-40$. This algebra discovers that -40°C is the same temperature as -40°F. □

Linear equations are characterized by each unknown never being multiplied or divided by another unknown, or itself, or inside 'curvaceous' functions. Table 2.1 lists examples of both. Generally, problems have many unknown variables. The power of linear algebra is especially important for large numbers of unknown variables. The number n of unknown variables may be two or three as in many examples herein, or may be thousands or millions in many modern applications.

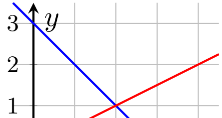
Definition 2.1.2 *A* ***linear equation*** *in the n variables $x_1, x_2, \ldots, x_n$ is an equation that can be written in the form*

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

where the ***coefficients*** *a_1 , a_2 , ..., 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).*

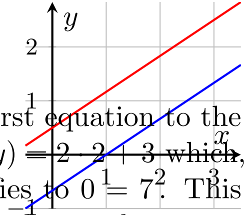
Example 2.1.3 *(two equations in two variables)* Graphically and algebraically solve each of the following systems.

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



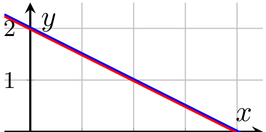
Algebraically, one could add twice the first equation to half of the second equation: $2(x+y)+\frac{1}{2}(2x-4y)=2\cdot 3+\frac{1}{2}\cdot 0$ which simplifies to $3x=6$ as the y terms cancel; hence $x=2$. Then, say, consider the second equation, $2x-4y=0$, which now becomes $2\cdot 2-4y=0$, that is, $y=1$. This algebra gives the same solution $(x,y)=(2,1)$ as graphically. □

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



Algebraically, one could add twice the first equation to the second equation: $2(2x-3y)+(-4x+6y)=2\cdot 2+3$ which, as all the x and y terms cancel, simplifies to $0=7$. This equation is a contradiction, as zero cannot be equal to seven. Thus there are no solutions to the system. □

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

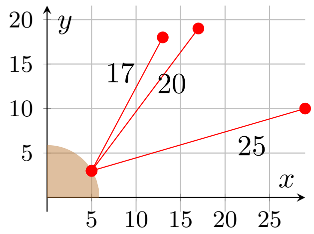


Algebraically, the rearrangement of both equations to exactly the same $y = 2 - x/2$ establishes an infinite number of solutions, here parametrized by x. $\square$

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

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

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. A receiver, say your smart-phone, pick up these signals and, from the time taken for the signals to arrive, knows the distance to all those satellites within 'sight'. Your smart-phone then 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, equivalently 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 to the right). 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 $\tan^{-1} \frac{3}{5} = 0.5404 = 30.96°$N. □

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

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

$\square$

The sequence of graphs in the previous Example 2.1.6 illustrates the equations at each main step in the algebraic manipulations. Apart from keeping the solution intersection point fixed, the sequence of graphs looks rather chaotic. Indeed, for each of these algebraic steps there is no particular geometric pattern or interpretation. In contrast, one feature of the upcoming 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 recognizable in the abstract version. Nonetheless, it is the abstraction by linear algebra that empowers wonderful results for applications.

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

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

into the linear algebra task of
finding the intersection point of three specific planes (in a different space). This task we then do.

Suppose we observe that at some given temperature and humidity 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 Example 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 2.2 establishes the general key property of a system of any number of linear equations in any number of variables:

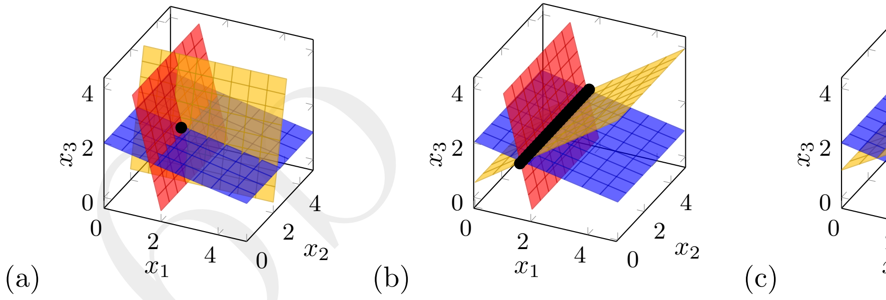


Figure 2.2: Solving three linear equations in three variables finds the intersection point(s) of three planes. The only three possibilities are: (a) a unique solution; (b) infinitely many solutions; or (c) no solution.

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

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 anything from 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 2.1.3(a)

$$\begin{aligned} x + y &= 3 \\ 2x - 4y &= 0 \end{aligned} \qquad \text{is written} \qquad \underbrace{\begin{bmatrix} 1 & 1 \\ 2 & -4 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_{\boldsymbol{x}} = \underbrace{\begin{bmatrix} 3 \\ 0 \end{bmatrix}}_{\boldsymbol{b}}.$$

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

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

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

$\square$

The beauty of the form $A\boldsymbol{x} = \boldsymbol{b}$ is that the numbers involved in the system are abstracted into the matrix A and vector $\boldsymbol{b}$: Matlab/Octave handles such numerical matrices and vectors. For some of you, writing a system in this matrix-vector form $A\boldsymbol{x} = \boldsymbol{b}$ (Definition 2.2.2 below) may 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 $A\boldsymbol{x} = \boldsymbol{b}$ is an expression involving natural operations for matrices and vectors: Section 3.1 defines and explores such useful operations.

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 $A\boldsymbol{x} = \boldsymbol{b}$ 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 $\boldsymbol{x} = (x_1\,, x_2\,, \ldots, x_n)$ and $\boldsymbol{b} = (b_1\,, b_2\,, \ldots, b_m)$. 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)* Write the following systems in matrix-vector form.

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

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

□

Activity 2.2.4 Which of the following systems correspond to the matrix-vector equation

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

(a)
$$\begin{aligned} -u + w &= 1 \\ 3u + 2w &= 0 \end{aligned}$$

(b)
$$\begin{aligned} -x + 3y &= 1 \\ x + 2y &= 0 \end{aligned}$$

(c)
$$\begin{aligned} -x + y &= 1 \\ 3x + 2y &= 0 \end{aligned}$$

(d)
$$\begin{aligned} -u + 3w &= 1 \\ u + 2w &= 0 \end{aligned}$$

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

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

Checking `rcond(A)` avoids gross mistakes. Section 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()` in Procedure 2.2.5 reflect practical realities.

> 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)

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

□

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 which of the following?

(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) $\begin{bmatrix}342\\73.5\end{bmatrix}$

Example 2.2.8 Following the previous Example 2.2.6, solve each of the two systems:

Table 2.3: To realize 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 of A (defined precisely by Definition 3.3.16).*
- `x=A\b` *computes an 'answer' to $A\boldsymbol{x} = \boldsymbol{b}$—but to be a solution requires* `rcond(A)` *to both exist and be not small.*
- *Change one 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)* `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,* $(1\,,1\,,\ldots\,,1)$.

(a) $$\begin{aligned} x_1 + x_2 - x_3 &= -2\,, \\ x_1 + 3x_2 + 5x_3 &= 5\,, \\ x_1 - 3x_2 + x_3 &= 1\,; \end{aligned}$$

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

□

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 is 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,

$$\frac{A}{x-1} + \frac{B}{x+2} = \frac{A(x+2) + B(x-1)}{(x-1)(x+2)}$$

$$= \frac{(A+B)x + (2A-B)}{(x-1)(x+2)}$$

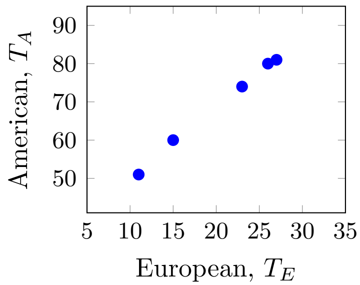
and, by comparing numerators, this equals $\frac{3}{(x-1)(x+2)}$ only if both $A+B=0$ and $2A-B=3$;

- 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 a European compared temperatures and, using measurements from two days, 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 reports that "for the last three days it has been 51°, 74°, and 81°", whereas the European reports "why, I recorded it as 11°, 23°, and 27°". The graph to the right 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)* Now recall 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 microsecond 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 $(6,12,23)\,\text{Mm}$ (megametres) and that the signal was sent at a true time $0.04\,\text{s}$ (seconds) before the phone's time. But the phone's time is different from the true time by some unknown amount, say t. Consequently, the travel time of the signal from the satellite to the phone is actually $t+0.04\,\text{s}$. Given the speed of light is $c = 300\,\text{Mm/s}$, this is a distance of $300(t+0.04) = 300t + 12$ —linear in the discrepancy of the phone's clock to the GPS clock. 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 that 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]*

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

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.

This subsection systematizes 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 minimize 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, a 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,

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

A second step of abstraction omits the symbols "$]\boldsymbol{x} = [$"—often we draw a vertical (dotted) line to show where the symbols "$]\boldsymbol{x} = [$" 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 $A\boldsymbol{x} = \boldsymbol{b}$ is the matrix $\begin{bmatrix} A \vdots \boldsymbol{b} \end{bmatrix}$.*

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) $\begin{bmatrix} 4 & 1 & \vdots & 3 \\ 2 & -1 & \vdots & -2 \end{bmatrix}$

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

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

(d) $\begin{bmatrix} 1 & 4 & \vdots & 3 \\ -1 & 2 & \vdots & -2 \end{bmatrix}$

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 either on 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:

$$x + 2y + z = 1\,,$$
$$2x - 3y = 2\,,$$
$$-3y - z = 2\,.$$

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 reduced row 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) $$\left[\begin{array}{ccc:c} 1 & 0 & -1 & 1 \\ 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 4 \end{array}\right]$$

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

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

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

$\square$

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

(a) $\begin{bmatrix} 0 & 1 & 0 & \vdots & 1 \\ 0 & 0 & 1 & \vdots & 2 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 0 & 1 & \vdots & 2 \\ 0 & 1 & 0 & \vdots & 1 \end{bmatrix}$ (c) $\begin{bmatrix} 0 & 1 & 1 & \vdots & -1 \\ 0 & 0 & 0 & \vdots & 0 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & 1 & 0 & \vdots & 0 \\ 0 & -1 & 1 & \vdots & -1 \end{bmatrix}$

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

$$\begin{bmatrix} 1 & 0 & -0.2 & \vdots & 0.4 \\ 0 & 1 & -1.2 & \vdots & -0.6 \end{bmatrix}?$$

(a) solution does not exist

(b) $(0.2t + 0.4\,, 1.2t - 0.6\,, t)$

(c) $(0.2 + 0.4t\,, 1.2 + 0.6t\,, t)$

(d) $(0.4\,, -0.6\,, 0)$

The previous Example 2.2.21 shows 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.

Generalizing 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.

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 to obtain a* ***general solution****.*

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 $A\boldsymbol{x} = \boldsymbol{b}$, 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 $A\boldsymbol{x} = \mathbf{0}$. 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 $\boldsymbol{x} = \mathbf{0}$ 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 nonzero 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 nonzero 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:

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

Subtract the first row from the second.

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

Divide the second row by (-7).

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

Subtract four times the second row from the first.

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

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 $\boldsymbol{x}=(-2s-\frac{15}{7}t,s,\frac{9}{7}t,t)=(-2,1,0,0)s+(-\frac{15}{7},0,\frac{9}{7},1)t$ for arbitrary s and t. These solutions include $\boldsymbol{x}=\boldsymbol{0}$ 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) $5y=3x$ and $4x=2y$

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

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

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

As Example 2.2.29(d) 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 Example 2.2.29(d)), then a homogeneous system always has an infinite number of solutions.

Theorem 2.2.31 *If $A\boldsymbol{x} = \mathbf{0}$ 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

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

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

The general solution to Example 2.2.21(a) is

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

Such so-called linear combinations occur in many other contexts. Recall that the standard unit vectors in $\mathbb{R}^3$ are $\boldsymbol{e}_1 = (1,0,0)$, $\boldsymbol{e}_2 = (0,1,0)$ and $\boldsymbol{e}_3 = (0,0,1)$ (Definition 1.2.7); so every other vector in $\mathbb{R}^3$ may be written as

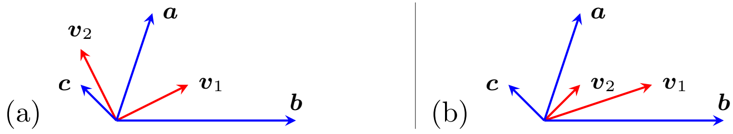
$$\boldsymbol{x} = (x_1, x_2, x_3)$$

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

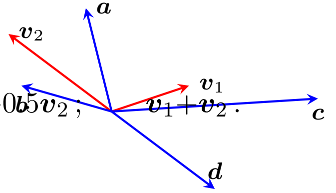
The widespread appearance of such combinations calls for the following definition.

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

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).



□



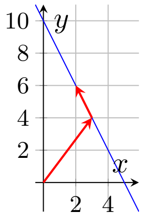
Activity 2.3.3 Choose any one of these linear combinations:

$$2\boldsymbol{v}_1 - 0.5\boldsymbol{v}_2\,; \qquad 0\boldsymbol{v}_1 - \boldsymbol{v}_2\,; \qquad -0.5\boldsymbol{v}_1 + 0.5\boldsymbol{v}_2\,; \qquad \boldsymbol{v}_1 + \boldsymbol{v}_2\,.$$

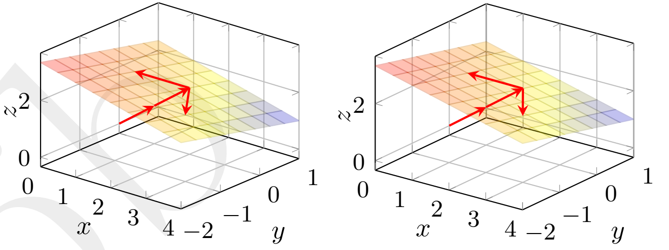
Then in the plot to the right, which vector, $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, or $\boldsymbol{d}$, corresponds to your chosen linear combination?

Example 2.3.4 Parametric descriptions of lines and planes involve linear combinations (Sections 1.2 and 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 to the right. (The line is alternatively described as $2x + y = 10\,$.)

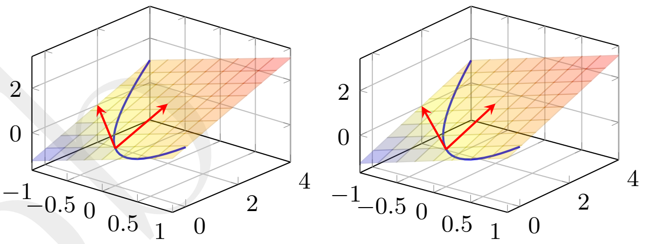


(b) For each value of s and t, the ex-



pression $2(1\,,0\,,1)+s(-1\,,-\frac12\,,\frac12)+t(1\,,-1\,,0)$ is a linear combination of the three vectors $(1\,,0\,,1)$, $(-1\,,-\frac12\,,\frac12)$, and $(1\,,-1\,,0)$. Over all values of the parameters s and t it describes the plane illustrated above-right in stereo. (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 above in stereo, and very soon we will be able to say it is a parabola in the plane spanned by $(-1\,,2\,,0)$ and $(0\,,2\,,1)$.



$\square$

The matrix-vector form $A\boldsymbol{x}=\boldsymbol{b}$ 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 matrix-vector representation of the system of the two equations $-5x + 4y = 1$, and $3x + 2y = -2$. 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. $\square$

Example 2.3.6 Let's repeat the previous example in general. Recall from Definition 2.2.2 that $A\boldsymbol{x} = \boldsymbol{b}$ is our matrix-vector representation 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 (Definition 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, $\boldsymbol{a}_1 = (a_{11}, a_{21}, \dots, a_{m1})$, $\boldsymbol{a}_2 = (a_{12}, a_{22}, \dots, a_{m2})$, …, $\boldsymbol{a}_n = (a_{1n}, a_{2n}, \dots, a_m$ 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 $A\boldsymbol{x} = \boldsymbol{b}$ is identical to the linear combination $x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_n\boldsymbol{a}_n = \boldsymbol{b}$. □

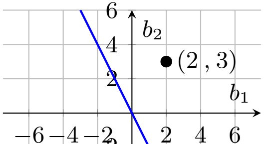
Theorem 2.3.7 *A system of linear equations $A\boldsymbol{x} = \boldsymbol{b}$ is consistent (Procedure 2.2.24) if and only if the right-hand side vector $\boldsymbol{b}$ is a linear combination of the columns of A.*

Be aware of a subtle twist going on in this theorem: for the general Example 2.3.6 this theorem turns a question about the existence of an n variable solution $\boldsymbol{x}$, into a question about vectors with m components, and vice versa.

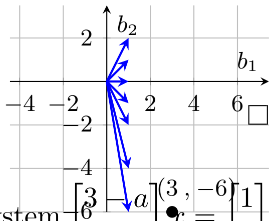
Example 2.3.8 This first example considers the simplest cases when the matrix has only one column, and so any linear combination is only a scalar multiple of that column. Compare the consistency of the equations with the right-hand side being a linear combination of the column of the matrix.

(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?

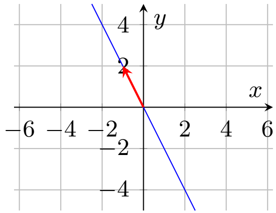
(a) $a = -3$ (b) $a = 2$ (c) $a = 1$ (d) $a = -\frac{1}{2}$

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 $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$, then the set of all linear combinations of $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ is called the **span** of $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$, and is denoted by* $\operatorname{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$ *or* $\operatorname{span} S$.

Example 2.3.11

(a) Let the set $S = \{(-1\,,\,2)\}$ with just one vector. Then $\operatorname{span} S = \operatorname{span}\{(-1\,,\,2)\}$ is the set of all vectors encompassed by the form $t(-1,2)$. From the parametric equation of a line (Definition 1.2.15), $\operatorname{span} S$ is all vectors in the line $y = -2x$, as shown to the right.



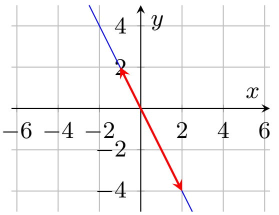
(b) With two vectors in the set, $\operatorname{span}\{(-1,2),(3,4)\} = \mathbb{R}^2$ is the entire 2D plane. To see this, recall that any point in the span must be of the form $s(-1,2)+t(3,4)$. Given any vector (x_1,x_2) in $\mathbb{R}^2$ we choose $s = (-4x_1+3x_2)/10$ and $t = (2x_1+x_2)/10$ and then the linear combination

$$s\begin{bmatrix}-1\\2\end{bmatrix}+t\begin{bmatrix}3\\4\end{bmatrix} = \frac{-4x_1+3x_2}{10}\begin{bmatrix}-1\\2\end{bmatrix}+\frac{2x_1+x_2}{10}\begin{bmatrix}3\\4\end{bmatrix}$$

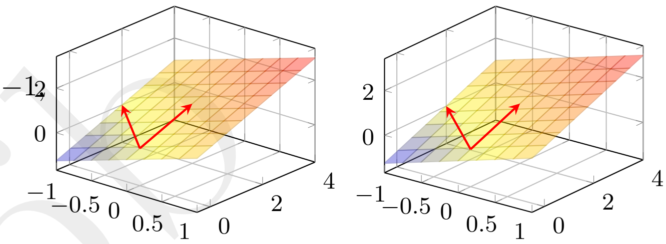
$$
\begin{aligned}
&= x_1\left(\frac{-4}{10}\begin{bmatrix}-1\\2\end{bmatrix}+\frac{2}{10}\begin{bmatrix}3\\4\end{bmatrix}\right)\\
&\quad + x_2\left(\frac{3}{10}\begin{bmatrix}-1\\2\end{bmatrix}+\frac{1}{10}\begin{bmatrix}3\\4\end{bmatrix}\right)\\
&= x_1\begin{bmatrix}1\\0\end{bmatrix}+x_2\begin{bmatrix}0\\1\end{bmatrix}=\begin{bmatrix}x_1\\x_2\end{bmatrix}.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

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

(c) But if two vectors are proportional to each other then their span is a line. For example, $\operatorname{span}\{(-1,2),(2,-4)\}$ 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, $\operatorname{span}\{(-1,2),(2,-4)\} = \operatorname{span}\{(-1,2)\}$ as illustrated to the right.



(d) In 3D, 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 here (Definition 1.3.32). The plane passes through the origin $\mathbf{0}$, obtained when $s=t=0$. We 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 also be expressed as $2x+y-2z=0$.

(e) For the complete set of n standard unit vectors in $\mathbb{R}^n$ (Definition 1.2.7), span$\{\boldsymbol{e}_1\,,\boldsymbol{e}_2\,,\ldots,\boldsymbol{e}_n\}=\mathbb{R}^n$. This is because every vector $\boldsymbol{x}=(x_1\,,x_2\,,\ldots,x_n)$ in $\mathbb{R}^n$ may be written as the linear combination $\boldsymbol{x}=x_1\boldsymbol{e}_1+x_2\boldsymbol{e}_2+\cdots+x_n\boldsymbol{e}_n$, and hence every vector is in span$\{\boldsymbol{e}_1\,,\boldsymbol{e}_2\,,\ldots,\boldsymbol{e}_n\}$.

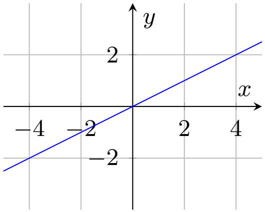
(f) The homogeneous system (Definition 2.2.28) of linear equations from Example 2.2.29(d) has solutions $\boldsymbol{x} = (-2s - \frac{15}{7}t, s, \frac{9}{7}t, t) = (-2, 1, 0, 0)s + (-\frac{15}{7}, 0, \frac{9}{7}, 1)t$ for arbitrary s and t. That is, the set of solutions is $\operatorname{span}\{(-2, 1, 0, 0), (-\frac{15}{7}, 0, \frac{9}{7}, 1)\}$, 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.

$\square$

Activity 2.3.12 To the right is drawn a line: for which one of the following vectors $\boldsymbol{u}$ is $\operatorname{span}\{\boldsymbol{u}\}$ *not* the drawn line?



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

(b) $(-1\,,-0.5)$

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

(d) $(4\,,2)$

Example 2.3.13 Describe in other words $\operatorname{span}\{\boldsymbol{i},\boldsymbol{k}\}$ in $\mathbb{R}^3$. □

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

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

Section 2.2 introduced matrices in the matrix-vector form $A\boldsymbol{x} = \boldsymbol{b}$ 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' $A\boldsymbol{x}$. 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

This section introduces basic matrix concepts, operations, and algebra. You may 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, scalars, written inside **brackets** $[\cdots]$, such as these six examples:

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

- 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, 3×3, 2×2, 3×1, 2×3, 1×3, and 1×1.

  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 (Section 3.1.2), the second special matrix is the **identity matrix**: the symbol I_n denotes an $n \times n$ 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 $I_{m\times n}$. Three examples are

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

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 $\boldsymbol{a}_j$, $A = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \end{bmatrix}$. For example, matrix

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

for the three column vectors

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

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

- 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

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

  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: 1×2 and 2×1 respectively.

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

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

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

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

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

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

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

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

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) none of the others

(b) $A + B$

(c) $B - A$

(d) $A - B$

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

$$\begin{aligned} 3A &= \begin{bmatrix} 3\cdot 5 & 3\cdot 2 \\ 3\cdot(-2) & 3\cdot 3 \end{bmatrix} = \begin{bmatrix} 15 & 6 \\ -6 & 9 \end{bmatrix}, \\ -B = (-1)B &= \begin{bmatrix} (-1)\cdot 1 \\ (-1)\cdot 0 \\ (-1)\cdot(-6) \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 6 \end{bmatrix}, \\ -\pi C = (-\pi)C &= \begin{bmatrix} -5\pi & 6\pi & -4\pi \\ \pi & 3\pi & 3\pi \end{bmatrix}. \end{aligned}$$

□

In general, when A is an $m \times n$ matrix, with entries a_{ij}, then we define the **scalar product** by c, denoted by 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, is written $A\boldsymbol{x} = \boldsymbol{b}$. In this form, $A\boldsymbol{x}$ denotes a matrix-vector product. As implied by Definition 2.2.2, we define the general **matrix-vector product**

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

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

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

This product is only defined when the number of columns of matrix A is the same as the number of components of vector $\boldsymbol{x}$. 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 $\boldsymbol{x}=(2,-3)$ and $\boldsymbol{b}=(1,0,4)$. Then the matrix-vector products

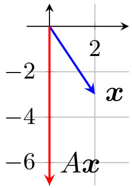
$$\begin{aligned}A\boldsymbol{x}&=\begin{bmatrix}3&2\\-2&1\end{bmatrix}\begin{bmatrix}2\\-3\end{bmatrix}=\begin{bmatrix}3\cdot2+2\cdot(-3)\\(-2)\cdot2+1\cdot(-3)\end{bmatrix}=\begin{bmatrix}0\\-7\end{bmatrix},\\ B\boldsymbol{b}&=\begin{bmatrix}5&-6&4\\-1&-3&-3\end{bmatrix}\begin{bmatrix}1\\0\\4\end{bmatrix}\\ &=\begin{bmatrix}5\cdot1+(-6)\cdot0+4\cdot4\\(-1)\cdot1+(-3)\cdot0+(-3)\cdot4\end{bmatrix}=\begin{bmatrix}21\\-13\end{bmatrix}.\end{aligned}$$

The combinations $A\boldsymbol{b}$ and $B\boldsymbol{x}$ are not defined because the number of columns of each matrix is not equal to the number of components in the multiplying vector.

Further, we do not here define vector-matrix products such as $\boldsymbol{x}A$ or $\boldsymbol{b}B$: the order of multiplication matters with matrices and so these are not in the scope of this definition. □

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

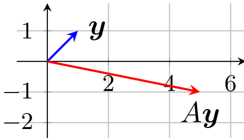
(a) $\begin{bmatrix} 15 \\ 2 \end{bmatrix}$ (b) $\begin{bmatrix} 21 \\ -2 \end{bmatrix}$ (c) $\begin{bmatrix} 14 \\ 5 \end{bmatrix}$ (d) $\begin{bmatrix} 18 \\ -1 \end{bmatrix}$



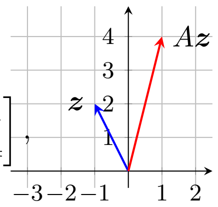
Geometric interpretation Multiplication of a vector by a *square matrix* transforms the vector into another in the same space. The right graph shows the example of $A\boldsymbol{x}$ from Example 3.1.5.

For another vector $\boldsymbol{y} = (1,1)$ and the same matrix A the product

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



as illustrated in the second right-hand picture.



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

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

as illustrated in the third right-hand 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 (Section 3.2.2), rotations (Section 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

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

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

That is, and justifying its name of "identity", the products with an identity matrix give the result that is the vector itself: $I_2\boldsymbol{x} = \boldsymbol{x}$ and $I_3\boldsymbol{b} = \boldsymbol{b}$. Multiplication by the identity matrix leaves the vector unchanged (Theorem 3.1.26(e)). □

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?

Let's count just 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 $x_1' = x_2$.

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 $\boldsymbol{x} = (x_1, x_2)$ and the population one month later be $\boldsymbol{x}' = (x_1', x_2')$. Then our model is that

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

called a Leslie matrix.

- At the start there is one adult pair, one doe, and no female kittens, so the initial population is $\boldsymbol{x} = (0, 1)$.
- After one month, females $\boldsymbol{x}' = L\boldsymbol{x} = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 1 \end{smallmatrix}\right] \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$.
- After two months, females $\boldsymbol{x}'' = L\boldsymbol{x}' = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 1 \end{smallmatrix}\right] \left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right]$.
- After three months, females $\boldsymbol{x}''' = L\boldsymbol{x}'' = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 1 \end{smallmatrix}\right] \left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 2 \\ 3 \end{smallmatrix}\right]$.

- After four months, females $\boldsymbol{x}^{iv} = L\boldsymbol{x}''' = \left[\begin{smallmatrix}0&1\\1&1\end{smallmatrix}\right]\left[\begin{smallmatrix}2\\3\end{smallmatrix}\right] = \left[\begin{smallmatrix}3\\5\end{smallmatrix}\right]$.
- After five months, females $\boldsymbol{x}^{v} = L\boldsymbol{x}^{iv} = \left[\begin{smallmatrix}0&1\\1&1\end{smallmatrix}\right]\left[\begin{smallmatrix}3\\5\end{smallmatrix}\right] = \left[\begin{smallmatrix}5\\8\end{smallmatrix}\right]$.
- After six months, females $\boldsymbol{x}^{vi} = L\boldsymbol{x}^{v} = \left[\begin{smallmatrix}0&1\\1&1\end{smallmatrix}\right]\left[\begin{smallmatrix}5\\8\end{smallmatrix}\right] = \left[\begin{smallmatrix}8\\13\end{smallmatrix}\right]$.

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

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 adult 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 adult

females;

- each adult female breeds and produces four female pups;
- one-third of the adult 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 ecologist's observations to write x_1', x_2', and x_3' as a function of x_1, x_2, and x_3 (this function is called a Markov chain).

(b) Letting vectors $\boldsymbol{x} = (x_1, x_2, x_3)$ and $\boldsymbol{x}' = (x_1', x_2', x_3')$ write down your function as the matrix-vector product $\boldsymbol{x}' = L\boldsymbol{x}$ 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 $\boldsymbol{x} = (60, 70, 20)$, use your matrix to predict the numbers $\boldsymbol{x}'$ at the start of the next year. Continue similarly to predict the numbers after two years ($\boldsymbol{x}''$), and three years ($\boldsymbol{x}'''$).

□

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 generalizes 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

$$\begin{aligned} AB &= A\begin{bmatrix} 5 & -6 & 4 \\ -1 & -3 & -3 \end{bmatrix} = A\left[\begin{array}{c:c:c} 5 & -6 & 4 \\ -1 & -3 & -3 \end{array}\right] = \left[A\begin{bmatrix} 5 \\ -1 \end{bmatrix} \vdots A\begin{bmatrix} -6 \\ -3 \end{bmatrix} \vdots A \right. \\ &= \left[\begin{bmatrix} 13 \\ -11 \end{bmatrix} \vdots \begin{bmatrix} -24 \\ 9 \end{bmatrix} \vdots \begin{bmatrix} 6 \\ -11 \end{bmatrix} \right] = \begin{bmatrix} 13 & -24 & 6 \\ -11 & 9 & -11 \end{bmatrix}. \end{aligned}$$

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

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

and neither of these matrix-vector products can be done because, 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 For the following two matrices, compute, if possible, CD and DC, and compare these products:

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

□

Definition 3.1.12 (matrix product) *Let matrix A be $m \times n$, and matrix B be $n \times p$, then the* ***matrix product*** *$C = AB$, or* ***matrix multiplication****, 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 ith row of A and the jth column of B as illustrated by

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

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

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

(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 adults formed into a vector $\boldsymbol{x} = (x_1, x_2, x_3)$, the number in each age one year later (indicated here by a dash) is $\boldsymbol{x}' = L\boldsymbol{x}$ 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

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

provided that matrix multiplication is associative (established by Theorem 3.1.26(c)) to enable us to write $L(L\boldsymbol{x}) = (LL)\boldsymbol{x}$. 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 threes dashes) is

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

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 scalars. 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

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

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

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

□

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 ith row of A as the ith column of A^{T}, or equivalently by writing the jth column of A to be the jth row of A^{T}. That is, if $B = A^{\mathrm{T}}$, 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) $\begin{bmatrix} 2.9 & -0.5 & 1 \\ -0.2 & -1.4 & -1.4 \\ 1.6 & -2.3 & 0.9 \end{bmatrix}$

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

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

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

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

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

Then the dot product between the two vectors

$$\begin{aligned} \boldsymbol{u} \cdot \boldsymbol{v} &= u_1 v_1 + u_2 v_2 + \cdots + u_n v_n && \text{(Definition 1.3.2 of dot)} \\ &= \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} && \text{(Definition 3.1.12 of mult.)} \\ &= \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} && \text{(transpose Definition 3.1.17)} \end{aligned}$$

$$= \boldsymbol{u}^{\mathrm{T}}\boldsymbol{v}\,.$$

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

Definition 3.1.20 (symmetry) *A (real) matrix A is a* ***symmetric matrix*** *if $A^{\mathrm{T}} = A$; 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 $C \neq C^{\mathrm{T}}$. The following matrix is symmetric:

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

When is the following general 2×2 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) $\begin{bmatrix} 2.2 & -0.9 & -1.2 \\ -0.9 & -1.2 & -3.1 \end{bmatrix}$

(b) $\begin{bmatrix} -2.6 & 0.3 & -1.3 \\ 0.3 & -0.2 & 0 \\ -1.3 & 0 & -2 \end{bmatrix}$

(c) $\begin{bmatrix} 2.3 & -1.3 & -2 \\ -3.2 & -1 & -1.3 \\ -3 & -3.2 & 2.3 \end{bmatrix}$

(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 *Mat*rix *Lab*oratory. Table 3.1 summarizes 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 3×5 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* Tables 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* $\begin{bmatrix} m & n \end{bmatrix}$.
- `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* **prime/quote/dash**, `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.*
- *Predefined matrices include:*
  - `zeros(m,n)` *is the zero matrix* $O_{m\times n}$*;*
  - `eye(m,n)` *is* $m \times n$ *'identity matrix'* $I_{m\times n}$*;*

The answer, "`3    5`", confirms A is 3×5. 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 fifth column vector, here $\begin{bmatrix} 5 \\ -8 \\ -3 \end{bmatrix}$;
- `A(1,:)` is the first row, here $\begin{bmatrix} 0 & 0 & -2 & -11 & 5 \end{bmatrix}$.

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 =
     0     0    -2   -11     5
     0     1    -1     9    -8
    -4     2    10     2    -3
```

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 3.1.24(a)) by seeing that their difference is the 4×4 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 4×3 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 4×3 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 3×4 and 4×2 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 $A\boldsymbol{b}_1$ and $A\boldsymbol{b}_2$ 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, `^`, computes matrix powers in Matlab/Octave, such as the cube `A^3`. But such matrix powers only make sense and work for square matrices A. For example, if matrix A was 3×4, then $A^2 = AA$ would involve multiplying a 3×4 matrix by a 3×4 matrix: since the number of columns of the left A, 4, is not the same as the number of rows of the right A, 3, 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 $(AB)^{\mathrm{T}} = B^{\mathrm{T}}A^{\mathrm{T}}$ (Theorem 3.1.29(d)) 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 3.1.29(f)): for example, generate a random square matrix by first `C=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.

Example 3.1.23 Let matrices $A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix}$. Show that the two products AB and BA are not equal. □

This Example 3.1.23 illustrates that matrix multiplication is not commutative. 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 3.1.24 (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);*

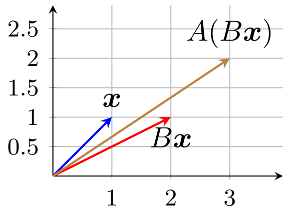
(e) $(c \pm d)A = cA \pm dA$ *(distributivity over scalar addition);*

(f) $c(dA) = (cd)A$ *(associativity of scalar multiplication);*

(g) $1A = A$ *; and*

(h) $0A = O$ *.*

Example 3.1.25 *(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 3.1.26(c). For another example, consider the two matrices and vector

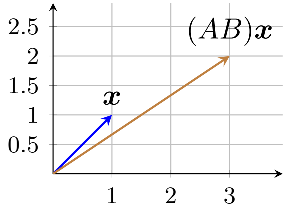


$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 0 \\ 2 & -1 \end{bmatrix}, \quad \boldsymbol{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Now the transform $\boldsymbol{x}' = B\boldsymbol{x} = (2, 1)$, and then transforming with A gives $\boldsymbol{x}'' = A\boldsymbol{x}' = A(B\boldsymbol{x}) = (3, 2)$, as illustrated to the above-right.

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 $(AB)\boldsymbol{x} = (3, 2)$ as also illustrated to the right. Such associativity asserts that $A(B\boldsymbol{x}) = (AB)\boldsymbol{x}$: that is, the geometric transform of $\boldsymbol{x}$ 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 3.1.26(c). □

Theorem 3.1.26 (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) $I_m A = A = AI_n$ *for* $m \times n$ *matrix* A *(multiplicative identity);*

(f) $O_m A = O_{m\times n} = AO_n$ *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 matrices* A *and for positive integers* p *and* q*.*

Example 3.1.27 Show that $(A + B)^2 \neq A^2 + 2AB + B^2$ in general. □

Example 3.1.28 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 3.1.29 (properties of transpose) *Let matrices A and B be of sizes such that the following expressions are defined, then:*

(a) $(A^{\mathrm{T}})^{\mathrm{T}} = A$;

(b) $(A \pm B)^{\mathrm{T}} = A^{\mathrm{T}} \pm B^{\mathrm{T}}$;

(c) $(cA)^{\mathrm{T}} = c(A^{\mathrm{T}})$ *for every scalar* c;

(d) $(AB)^{\mathrm{T}} = B^{\mathrm{T}}A^{\mathrm{T}}$ *—remember the reversed order in this identity;*

(e) $(A^p)^{\mathrm{T}} = (A^{\mathrm{T}})^p$ *for every positive integer exponent* p;

(f) $A + A^{\mathrm{T}}$, $A^{\mathrm{T}}A$ *and* AA^{T} *are symmetric matrices.*

3.2 The inverse of a matrix

Section contents

The previous Section 3.1 introduced addition, subtraction, multiplication, and other operations of matrices. Conspicuously missing from the list is 'division' by a matrix—missing because division is complicated. 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 almost 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 2×2 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 for this restriction 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 the possibility of many inverses. But Theorem 3.2.6 establishes that the inverse is unique. Further, an inverse may not exist for a given matrix A.

Example 3.2.3 Show that matrix

$$B = \begin{bmatrix} 0 & -\frac{1}{4} & -\frac{1}{8} \\ \frac{3}{2} & 1 & \frac{7}{8} \\ \frac{1}{2} & \frac{1}{4} & \frac{3}{8} \end{bmatrix} \text{ is an inverse of } A = \begin{bmatrix} 1 & -1 & 5 \\ -5 & -1 & 3 \\ 2 & 2 & -6 \end{bmatrix}.$$

$\square$

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) 3 (b) -3 (c) -2 (d) 1

But even among square matrices, there are many nonzero 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 1×1 matrices, that is $A = \begin{bmatrix} a_{11} \end{bmatrix}$, the inverse is simply the reciprocal of the entry, that is $A^{-1} = \begin{bmatrix} 1/a_{11} \end{bmatrix}$ provided a_{11} is nonzero. The reason is that $AA^{-1} = \begin{bmatrix} a_{11} \cdot \frac{1}{a_{11}} \end{bmatrix} = \begin{bmatrix} 1 \end{bmatrix} = I_1$ and $A^{-1}A = \begin{bmatrix} \frac{1}{a_{11}} \cdot a_{11} \end{bmatrix} = \begin{bmatrix} 1 \end{bmatrix} = I_1$.

In the case of 2×2 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 (2×2 inverse) *Let 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Then A is invertible if and only if the* ***determinant*** *$ad - bc \neq 0$, in which case*

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}. \tag{3.2}$$

If the determinant $ad - bc = 0$, then A is not invertible (it is a singular matrix).

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 that there is no inverse for matrix

$$A = \begin{bmatrix} 1 & -2 \\ -3 & 6 \end{bmatrix}.$$

Theorem 3.2.7 also establishes that 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) $\begin{bmatrix} 0 & -3 \\ 4 & -2 \end{bmatrix}$

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

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

Computational considerations Except for easy cases such as 2×2 matrices, we rarely explicitly compute the inverse of a matrix. Computationally there are (almost) always better ways such as the Matlab/Octave operation `A\b` of Procedure 2.2.5. The inverse is a crucial theoretical device, but rarely a practical computational tool.

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)*

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, Section 3.3.2 connects solutions to the `rcond` invoked by Procedure 2.2.5. Although in theoretical statements we write expressions like $\boldsymbol{x} = A^{-1}\boldsymbol{b}$, practically, once we know a solution exists (`rcond` is acceptable), we almost always compute a solution without ever constructing the inverse A^{-1}.

Theorem 3.2.10 *If A is an invertible $n \times n$ matrix, then the system of linear equations $A\boldsymbol{x} = \boldsymbol{b}$ has the unique solution $\boldsymbol{x} = A^{-1}\boldsymbol{b}$ for every $\boldsymbol{b}$ in $\mathbb{R}^n$. Equivalently, if there is either no solution or an infinite number of solutions (Theorem 2.2.27), then the matrix A is not invertible (it is singular).*

The forthcoming Theorem 3.3.27 strengthens this theorem to "if and only if": that is, A is invertible if and only if $A\boldsymbol{x} = \boldsymbol{b}$ has a unique solution.

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.

(a) $$\begin{cases} x - y = 4\,, \\ 4x - 3y = 3\,. \end{cases}$$

(b) $$\begin{cases} u - v + 5w = 2\,, \\ -5u - v + 3w = 5\,, \\ 2u + 2v - 6w = 1\,. \end{cases}$$

(c) $$\begin{cases} r - 2s = -1\,, \\ -3r + 6s = 3\,. \end{cases}$$

□

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 $(1, -5, 0, 3)$ (and no other).

(b) A general solution is $(3, -5 + 3t, 3 - t, -1)$ (and no other).

(c) A solution of a system is $(-3/2, -2, -\pi, 2, -4)$ (but there may be others).

(d) A solution of a homogeneous system is $(1, 2, -8)$ (but there may be others).

□

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 this identity;*

(d) *matrix A^{T} is invertible and $(A^{\mathrm{T}})^{-1} = (A^{-1})^{\mathrm{T}}$;*

(e) *for every $p = 1, 2, 3, \ldots$, matrix A^p is invertible 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 then is the inverse of the matrix $\begin{bmatrix}6 & -10\\ 8 & -14\end{bmatrix}$?

(a) $\begin{bmatrix}7 & 4\\ -5 & -3\end{bmatrix}$

(b) $\begin{bmatrix}3.5 & 2\\ -2.5 & -1.5\end{bmatrix}$

(c) $\begin{bmatrix}14 & -10\\ 8 & -3\end{bmatrix}$

(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 3.2.13(e) 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 The previous Example 3.2.16 gives the inverse of a matrix A and determines A^{-2}: what then is A^{-3}?

(a) $\begin{bmatrix}-7 & -12\\ 3 & 5\end{bmatrix}$

(b) $\begin{bmatrix}3 & 5\\ -7 & -12\end{bmatrix}$

(c) $\begin{bmatrix}-7 & 3\\ -12 & 5\end{bmatrix}$

(d) $\begin{bmatrix}3 & -7\\ 5 & -12\end{bmatrix}$

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 $\boldsymbol{x} = (60, 70, 20)$ is the current number of pups, juveniles, and mature females respectively, then our modelling predicts the population numbers after a year is $\boldsymbol{x}' = L\boldsymbol{x}$, after two years is $\boldsymbol{x}'' = L\boldsymbol{x}' = L^2\boldsymbol{x}$, 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 $\boldsymbol{x}^-$ then according to the model the current population $\boldsymbol{x} = L\boldsymbol{x}^-$. Multiply by the inverse of L: $L^{-1}\boldsymbol{x} = L^{-1}L\boldsymbol{x}^- = \boldsymbol{x}^-$; that is, the population a year before the current is $\boldsymbol{x}^- = L^{-1}\boldsymbol{x}$.

- Similarly, letting the population numbers two years ago be denoted by $\boldsymbol{x}^=$ then according to the model $\boldsymbol{x}^- = L\boldsymbol{x}^=$. So multiplication by L^{-1} gives $\boldsymbol{x}^= = L^{-1}\boldsymbol{x}^- = L^{-1}L^{-1}\boldsymbol{x} = L^{-2}\boldsymbol{x}$.

- One more year earlier, letting the population numbers three years ago be denoted by $\boldsymbol{x}^\equiv$, then according to the model $\boldsymbol{x}^= = L\boldsymbol{x}^\equiv$. So multiplication by L^{-1} gives $\boldsymbol{x}^\equiv = L^{-1}\boldsymbol{x}^= = L^{-1}L^{-2}\boldsymbol{x} = L^{-3}\boldsymbol{x}$.

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. $\square$

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 $L\boldsymbol{x}^{-} = \boldsymbol{x}$ for the population $\boldsymbol{x}^{-}$ a year ago, and then similarly solve $L\boldsymbol{x}^{=} = \boldsymbol{x}^{-}$, $L\boldsymbol{x}^{\equiv} = \boldsymbol{x}^{=}$, and so on. □

3.2.2 Diagonal matrices stretch and shrink

This section and the next Section 3.2.3 introduce two classes of matrices whose inverses we can easily write down and use. Importantly, the subsequent Section 3.3 shows that both classes lie at the heart of every matrix via the singular value decomposition of the matrix!

Recall that 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 $a_{11}, a_{22}, \ldots, a_{pp}$ where $p = \min(m, n)$. A matrix whose off-diagonal entries are all zero is called a **diagonal matrix**.*

For brevity, sometimes we write $\operatorname{diag}(v_1, v_2, \ldots, v_n)$ to denote the $n \times n$ square matrix with diagonal entries $v_1, v_2, \ldots, v_n$, or $\operatorname{diag}_{m\times n}(v_1, v_2, \ldots, v_p)$ for an $m \times n$ matrix with diagonal entries $v_1, v_2, \ldots, v_p$.

Example 3.2.22 The five diagonal matrices of Example 3.2.20 could equivalently be written as $\operatorname{diag}(3, 2)$, $\operatorname{diag}(0.58, -1.61, 2.17)$, $\operatorname{diag}(\pi, \sqrt{3}, 0)$, $\operatorname{diag}_{3\times2}(-\sqrt{2}, \frac{1}{2})$, and $\operatorname{diag}_{3\times5}(1, \pi, e)$, respectively. □

Diagonal matrices may also have zeros on the diagonal, as well as the required zeros for the off-diagonal entries.

Activity 3.2.23 Which of the following matrices are not diagonal?

(a) O_n

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

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

(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 size p generates the $p \times p$ matrix*
$$\operatorname{diag}(v_1\,, v_2\,, \ldots, v_p) = \begin{bmatrix} v_1 & 0 & \cdots & 0 \\ 0 & v_2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & v_p \end{bmatrix}.$$
- *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 $(a_{11}\,,a_{22}\,,\ldots,a_{pp})$ of diagonal entries where the result vector size $p = \min(m\,,n)$.*
- *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 For any given numbers b_1, b_2, b_3, 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 $\begin{bmatrix} 0.4 & 0 \\ 0 & 0.1 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 0.1 \\ -0.2 \end{bmatrix}$?

(a) $(4\,,-2)$ (b) $(4\,,-\frac12)$ (c) $(\frac14\,,-2)$ (d) $(\frac14\,,-\frac12)$

Theorem 3.2.27 (inverse of diagonal matrix) *For every $n \times n$ diagonal matrix $D = \operatorname{diag}(d_1, d_2, \ldots, d_n)$, if all the diagonal entries are nonzero, $d_i \neq 0$ for every $i = 1, 2, \ldots, n$, then D is invertible and the inverse $D^{-1} = \operatorname{diag}(1/d_1, 1/d_2, \ldots, 1/d_n)$. Conversely, if a diagonal matrix is invertible, then all its diagonal entries are nonzero.*

Example 3.2.28 The previous Example 3.2.25 gives the inverse of a 3×3 matrix. For the 2×2 matrix $D = \operatorname{diag}(3, 2) = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$ the inverse is $D^{-1} = \operatorname{diag}(\frac{1}{3}, \frac{1}{2}) = \begin{bmatrix} 1/3 & 0 \\ 0 & 1/2 \end{bmatrix}$. Then the solution to

$$\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 2 \\ -5 \end{bmatrix} \quad \text{is } \boldsymbol{x} = \begin{bmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 2 \\ -5 \end{bmatrix} = \begin{bmatrix} 2/3 \\ -5/2 \end{bmatrix}.$$

$\square$

Compute in Matlab/Octave To solve the matrix-vector equation $D\boldsymbol{x} = \boldsymbol{b}$ recognize that this equation means

$$\begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} d_1x_1 \\ d_2x_2 \\ \vdots \\ d_nx_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$
$$\iff \begin{cases} d_1x_1 = b_1 \\ d_2x_2 = b_2 \\ \vdots \\ d_nx_n = b_n \end{cases} \iff \begin{cases} x_1 = b_1/d_1 \\ x_2 = b_2/d_2 \\ \vdots \\ x_n = b_n/d_n \end{cases} \tag{3.3}$$

- 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 $D\boldsymbol{x} = \boldsymbol{b}$ for diagonal $D = \operatorname{diag}(d_1\,, d_2\,, \ldots\,, d_n)$ where $d_n = 0$ (and similarly if any others 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; alternatively
- if $b_n = 0$, then there are an infinite number of solutions as every 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)

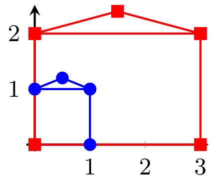
$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{2}{3} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{2}{3} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$$

□

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*

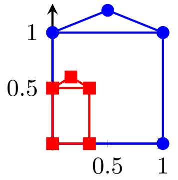
Multiplication by matrices transforms shapes: multiplication by diagonal matrices just stretches or squashes and/or reflects in the direction of the coordinate axes. The next Section 3.2.3 introduces matrices that rotate.



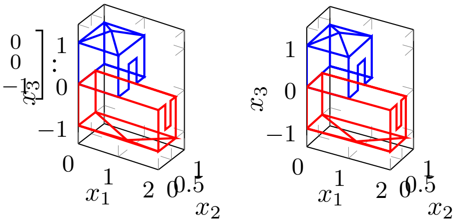
Example 3.2.30 Consider $A = \operatorname{diag}(3,2) = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$. The picture to the right shows this matrix stretches the (blue) unit square (drawn with a 'roof') by a factor of three horizontally and two vertically (to the red). Recall that (x_1, x_2) denotes the corresponding column vector. As seen in the corner points of the graphic to the right, $A \times (1,0) = (3,0)$, $A \times (0,1) = (0,2)$, $A \times (0,0) = (0,0)$, and $A \times (1,1) = (3,2)$. The 'roof' just helps us to track which corner

goes where.

The inverse $A^{-1} = \operatorname{diag}(\frac{1}{3}, \frac{1}{2}) = \begin{bmatrix} 1/3 & 0 \\ 0 & 1/2 \end{bmatrix}$ undoes the stretching of the matrix A by squashing in both the horizontal and vertical directions (from blue to red). □

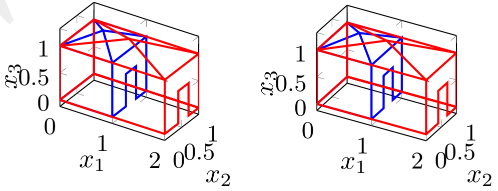


Example 3.2.31 Consider $\operatorname{diag}(2, \frac{2}{3}, -1) = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2/3 & 0 \\ 0 & 0 & -1 \end{bmatrix}$: the stereo pair to the right illustrates how this diagonal matrix stretches in one direction, squashes in another, and reflects in the vertical. By multiplying the matrix by the eight 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. □

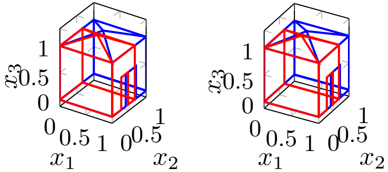


One great aspect of a diagonal matrix is that it is easy to separate its effects into each coordinate direction. For example, the above 3×3 matrix is the same as the combined effects of the following three.

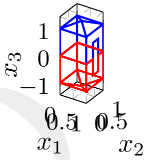
$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Stretch by a factor of two in the x_1 direction.

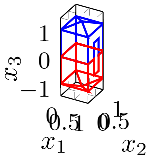


$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2/3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Squash through ‘stretching’ by a factor of $2/3$ in the x_2 direction.

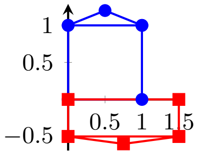


$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$. Reflect in the vertical x_3 direction.



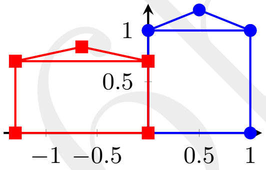


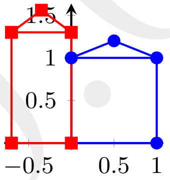
Example 3.2.32 From the illustration to the right, estimate the diagonal matrix that transforms the blue unit square to the red rectangle.

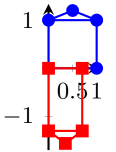


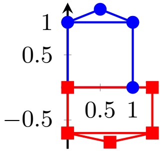
□

Activity 3.2.33 Which of the following diagrams represents the transformation from the (blue) unit square to the (red) rectangle by the matrix $\operatorname{diag}(-1.3\,,0.7)$?

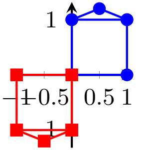
(a)


(b)


(c)


(d)


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 -1s, have the same effect as one rotation, here by 180°, as shown to the right. Matrices that

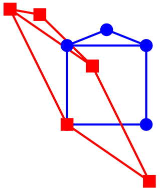


rotate are incredibly useful, and this is the topic of the next Section 3.2.3.

Sketch convenient coordinates

This optional subsubsection is a preliminary to diagonalization.

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 that are convenient for the application. This example illustrates the principle.

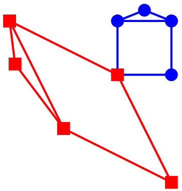


Example 3.2.34 Consider the transformation shown to the right (it might arise from the deformation of some material and we need to know the internal stretching and shrinking to predict failure). The drawing has no coordinate axes shown because it is supposed to be some transformation in nature. Now we 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 $\operatorname{diag}(\frac{1}{2}, 2) = \begin{bmatrix} 1/2 & 0 \\ 0 & 2 \end{bmatrix}$.

□

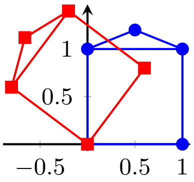
Example 3.2.35 Consider the transformation shown to the right. It has no coordinate axes shown because it is supposed to be some transformation in nature. Now impose on nature our mathematical description. Draw approximate coordinate axes, with origin at the common corner point, so the transformation becomes that of the diagonal matrix $\operatorname{diag}(3, -1) = \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix}$.



□

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 diagonalization 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). This section starts by showing that multiplying by the 'orthogonal matrix' $\begin{bmatrix} 3/5 & -4/5 \\ 4/5 & 3/5 \end{bmatrix}$ rotates by 53.13° as shown to the right. Such orthogonal matrices are the best to compute with, such as in solving linear equations, since they all have `rcond` $= 1$. To see these and related marvellous properties, we must explore the effect of matrix multiplication on the geometry of lengths and angles.

Recall that the dot product determines lengths and angles Section 1.3 introduced the dot product $\boldsymbol{u}\cdot\boldsymbol{v}$ between two vectors (Definition 1.3.2). The dot product is the same as the matrix product (Example 3.1.19): $\boldsymbol{u}\cdot\boldsymbol{v} = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{v} = \boldsymbol{v}^{\mathrm{T}}\boldsymbol{u}$. Further (Theorem 1.3.17(a)), the length of a vector is $|\boldsymbol{v}| = \sqrt{\boldsymbol{v}\cdot\boldsymbol{v}}$. For two nonzero vectors, Theorem 1.3.5 defines the angle θ between the vectors via

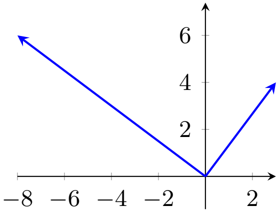
$$\cos\theta = \frac{\boldsymbol{u}\cdot\boldsymbol{v}}{|\boldsymbol{u}||\boldsymbol{v}|}\,,\quad 0\le\theta\le\pi\,.$$

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

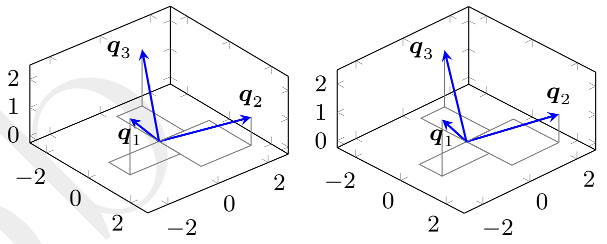
We need sets of orthogonal vectors (nonzero vectors that are all at right-angles to each other). (Recall that a **set** of objects is denoted by a list of the objects enclosed within braces, $\{\ldots\}$.) One example is the set of standard unit vectors $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ that are aligned with the coordinate axes in $\mathbb{R}^n$.

Example 3.2.36 The set of two vectors $\{(3,4),(-8,6)\}$ shown to the right is an orthogonal set as the two vectors have dot product $= 3\cdot(-8)+4\cdot 6 = -24+24 = 0$ and hence are orthogonal.



□

Example 3.2.37 Let vectors $\boldsymbol{q}_1 = (1, -2, 2)$, $\boldsymbol{q}_2 = (2, 2, 1)$, and $\boldsymbol{q}_3 = (-2, 1, 2)$, illustrated in stereo to the right. Is $\{\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3\}$ an orthogonal set?



$\square$

Definition 3.2.38 *A set of nonzero vectors* $\{\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_k\}$ *in* $\mathbb{R}^n$ *is called an* ***orthogonal set*** *if all pairs of distinct vectors in the set are orthogonal: that is,* $\boldsymbol{q}_i \cdot \boldsymbol{q}_j = 0$ *whenever* $i \neq j$ *for* $i, j = 1, 2, \ldots, k$ *. A set of vectors in* $\mathbb{R}^n$ *is called an* ***orthonormal set*** *if it is an orthogonal set of unit vectors.*

A single nonzero 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. $\square$

Example 3.2.40 Let vectors $\boldsymbol{q}_1 = (\frac{1}{3}, -\frac{2}{3}, \frac{2}{3})$, $\boldsymbol{q}_2 = (\frac{2}{3}, \frac{2}{3}, \frac{1}{3})$, and $\boldsymbol{q}_3 = (-\frac{2}{3}, \frac{1}{3}, \frac{2}{3})$. Show that the set $\{\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3\}$ is an orthonormal set. $\square$

Activity 3.2.41 Which one of the following sets of vectors is *not* an orthogonal set?

(a) $\{(-5\,,4)\}$

(b) $\{(2\,,3)\,,(4\,,-1)\}$

(c) $\{(-2\,,3)\,,(6\,,4)\}$

(d) $\{\boldsymbol{i}\,,\boldsymbol{k}\}$

Orthogonal matrices

Now let's see how orthonormal vectors form a so-called orthogonal matrix and underlie its marvellous properties.

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 that $\{(\frac{3}{5},\frac{4}{5}),(-\frac{4}{5},\frac{3}{5})\}$ is an orthonormal set. Form the matrix with these two vectors as its columns: $Q = \begin{bmatrix} 3/5 & -4/5 \\ 4/5 & 3/5 \end{bmatrix}$. Then consider

$$Q^{\mathrm{T}}Q = \begin{bmatrix} \frac{3}{5} & \frac{4}{5} \\ -\frac{4}{5} & \frac{3}{5} \end{bmatrix} \begin{bmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{bmatrix} = \begin{bmatrix} \frac{9+16}{25} & \frac{-12+12}{25} \\ \frac{-12+12}{25} & \frac{16+9}{25} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Similarly $QQ^{\mathrm{T}} = I_2$. Consequently, the transpose Q^{T} is here the inverse of Q (Definition 3.2.2). The transpose being the inverse is no accident here.

Also no accident is that multiplication by this Q gives the rotation illustrated at the start of this section (Section 3.2.3). □

Definition 3.2.43 (orthogonal matrices) *A square $n \times n$ matrix Q is called an **orthogonal matrix** if $Q^{\mathrm{T}}Q = I_n$. 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**.*

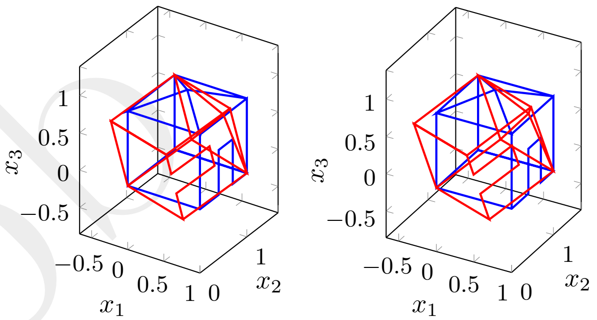
Activity 3.2.44 For which of the following values of p is the matrix $Q = \begin{bmatrix} \frac{1}{2} & p \\ -p & \frac{1}{2} \end{bmatrix}$ orthogonal?

(a) some other value
(b) $p = 3/4$
(c) $p = \sqrt{3}/2$
(d) $p = -1/2$

Example 3.2.45 In the following equation, check that the matrix is orthogonal. Use its orthogonality to solve the equation $Q\boldsymbol{x} = \boldsymbol{b}$:

$$\begin{bmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\ -\frac{2}{3} & \frac{1}{3} & \frac{2}{3} \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 0 \\ 2 \\ -1 \end{bmatrix}.$$

The stereo pair to the right illustrates the rotation of the unit cube under multiplication by this matrix Q: every point $\boldsymbol{x}$ in the (blue) unit cube (with 'roof and door'), is mapped to the point $Q\boldsymbol{x}$ to form the (red) result.



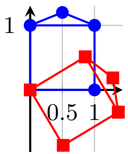
□

Example 3.2.46 Given the matrix is orthogonal, solve the linear equation

$$\begin{bmatrix} \frac12 & \frac12 & \frac12 & \frac12 \\ \frac12 & -\frac12 & \frac12 & -\frac12 \\ \frac12 & -\frac12 & -\frac12 & \frac12 \\ \frac12 & \frac12 & -\frac12 & -\frac12 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 1 \\ -1 \\ 1 \\ 3 \end{bmatrix}.$$

$\square$

Example 3.2.47 The graph to the right 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).



$\square$

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 utilize 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 $Q^{-1} = Q^{\mathrm{T}}$;

(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).*

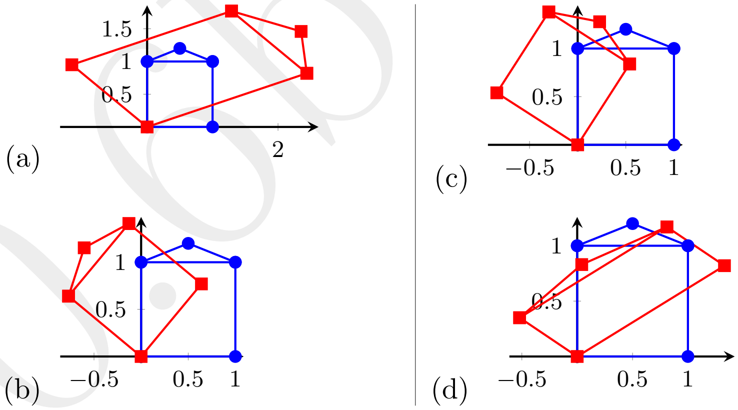
Another important property, proved by **??**, is that the product of orthogonal matrices is also an orthogonal matrix.

Example 3.2.49 Show that these matrices are orthogonal and hence write down their inverses:

$$\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

□

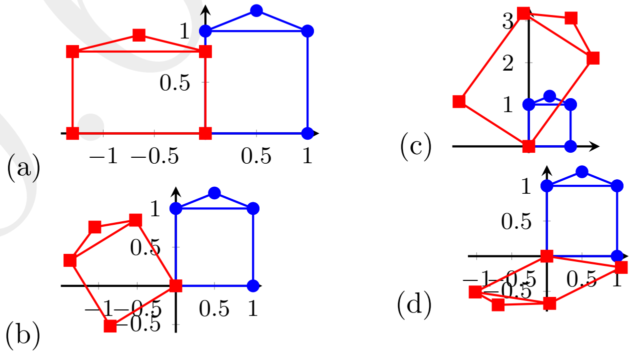
Example 3.2.50 The following graphs illustrate the transformation of the unit square through multiplying by some different matrices. Using Theorem 3.2.48(f), which transformations appear to be that of multiplying by an orthogonal matrix?



□

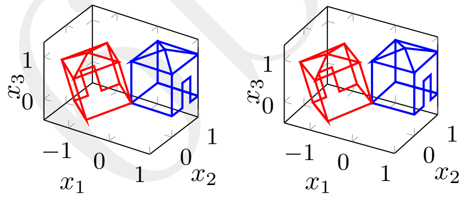
Activity 3.2.51 The following graphs illustrate the transformation of the unit square through multiplying by some different matrices.

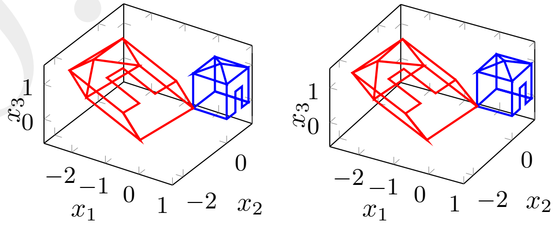
- Which transformation appears to be that of multiplying by an orthogonal matrix?

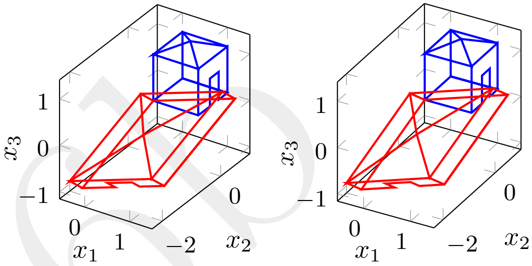


- Further, which one of the above transformations appears to be that of multiplying by a diagonal matrix?

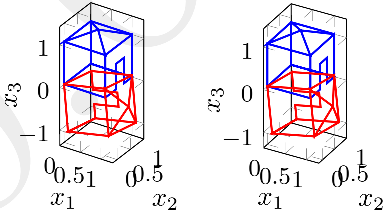
Example 3.2.52 The following stereo pairs illustrate the transformation of the unit cube through multiplying by some different matrices: using Theorem 3.2.48(f), which transformations appear to be that of multiplying by an orthogonal matrix?

(a)


(b)




(c)



(d)

$\square$

3.3 Factorize to the singular value decomposition

Section contents

The singular value decomposition (SVD) is sometimes called the jewel in the crown of linear algebra. The SVD's importance is certified by the many names by which it is invoked in scientific and engineering applications: principal component analysis, singular spectrum analysis, proper 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 $42\,x = 1554$ within twenty seconds,

your time starts now ....

Activity 3.3.1 Given $154 = 2 \cdot 7 \cdot 11$, solve in your head $154\,x = 8008$ or 9856 or 12628 or 13090 or 14322 (teacher to choose): first to answer wins!

Such examples show that factorization can turn a hard problem into several easy problems. We adopt an analogous matrix factorization to solve and understand general linear equations.

To illustrate the procedure to come, let's write the above solution steps in detail: we solve $42\,x = 1554$.

1. Factorize 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 (that is, $z = 3y = 21x$ and $y = 7x$): these are as yet unknown as x is as yet unknown.
2. Solve $2z = 1554$ to determine $z = 777$.
3. Solve $3y = z = 777$ to determine $y = 259$.
4. Solve $7x = y = 259$ to determine $x = 37$ —the answer.

Now let's proceed to small matrix examples. Each example follows analogous solution steps to those above. The following examples introduce the general matrix procedure empowered by a factorization of a matrix. Specifically, we use the matrix factorization called a singular value decomposition (SVD).

Example 3.3.2 Solve the 2×2 system

$$\begin{bmatrix} 10 & 2 \\ 5 & 11 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 18 \\ -1 \end{bmatrix}$$

for $\boldsymbol{x}$ given the matrix factorization

$$\begin{bmatrix} 10 & 2 \\ 5 & 11 \end{bmatrix} = \begin{bmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{bmatrix} \begin{bmatrix} 10\sqrt{2} & 0 \\ 0 & 5\sqrt{2} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}^{\mathsf{T}}$$

(remember the transpose on the last matrix). □

Activity 3.3.3 Let's solve the system $\begin{bmatrix}12 & -41\\ 34 & -12\end{bmatrix}\boldsymbol{x} = \begin{bmatrix}94\\ 58\end{bmatrix}$ using the factorization $\begin{bmatrix}12 & -41\\ 34 & -12\end{bmatrix} = \begin{bmatrix}\frac{4}{5} & -\frac{3}{5}\\ \frac{3}{5} & \frac{4}{5}\end{bmatrix}\begin{bmatrix}50 & 0\\ 0 & 25\end{bmatrix}\begin{bmatrix}\frac{3}{5} & \frac{4}{5}\\ -\frac{4}{5} & \frac{3}{5}\end{bmatrix}^{\mathrm{T}}$ in which the first and third matrices on the right-hand side are orthogonal. After solving $\begin{bmatrix}\frac{4}{5} & -\frac{3}{5}\\ \frac{3}{5} & \frac{4}{5}\end{bmatrix}\boldsymbol{z} = \begin{bmatrix}94\\ 58\end{bmatrix}$, the next step is to solve which of the following?

(a) $\begin{bmatrix}50 & 0\\ 0 & 25\end{bmatrix}\boldsymbol{y} = \begin{bmatrix}10\\ 110\end{bmatrix}$

(b) $\begin{bmatrix}50 & 0\\ 0 & 25\end{bmatrix}\boldsymbol{y} = \begin{bmatrix}\frac{514}{5}\\ -\frac{202}{5}\end{bmatrix}$

(c) $\begin{bmatrix}50 & 0\\ 0 & 25\end{bmatrix}\boldsymbol{y} = \begin{bmatrix}\frac{202}{5}\\ \frac{514}{5}\end{bmatrix}$

(d) $\begin{bmatrix}50 & 0\\ 0 & 25\end{bmatrix}\boldsymbol{y} = \begin{bmatrix}110\\ -10\end{bmatrix}$

Example 3.3.4 Solve the 3×3 system

$$A\boldsymbol{x} = \begin{bmatrix} 10 \\ 2 \\ -2 \end{bmatrix} \quad \text{for matrix } A = \begin{bmatrix} -4 & -2 & 4 \\ -8 & -1 & -4 \\ 6 & 6 & 0 \end{bmatrix}$$

using the following given matrix factorization (note the last matrix is transposed)

$$A = \begin{bmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\ -\frac{2}{3} & \frac{1}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} 12 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} -\frac{8}{9} & -\frac{1}{9} & -\frac{4}{9} \\ -\frac{4}{9} & \frac{4}{9} & \frac{7}{9} \\ -\frac{1}{9} & -\frac{8}{9} & \frac{4}{9} \end{bmatrix}^{\mathsf{T}}.$$

□

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

$$\begin{bmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{bmatrix} \underbrace{\begin{bmatrix} 10\sqrt{2} & 0 \\ 0 & 5\sqrt{2} \end{bmatrix} \overbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}^{\operatorname{T}} \boldsymbol{x}}^{=\boldsymbol{y}}}_{=\boldsymbol{z}} = \begin{bmatrix} 18 \\ -1 \end{bmatrix},$$

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 $\begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}^{\operatorname{T}} \boldsymbol{x} = \begin{bmatrix} 18 \\ -1 \end{bmatrix}$: this matrix is orthogonal, so multiplying by itself (the transpose of the transpose) gives

$$\boldsymbol{x} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 18 \\ -1 \end{bmatrix} = \begin{bmatrix} \frac{19}{\sqrt{2}} \\ \frac{17}{\sqrt{2}} \end{bmatrix}.$$

(c) Inappropriately 'solve' $\begin{bmatrix} 10\sqrt{2} & 0 \\ 0 & 5\sqrt{2} \end{bmatrix} \boldsymbol{y} = \begin{bmatrix} 19/\sqrt{2} \\ 17/\sqrt{2} \end{bmatrix}$: this matrix is diagonal, so dividing by the diagonal elements gives

$$\boldsymbol{y} = \begin{bmatrix} 10\sqrt{2} & 0 \\ 0 & 5\sqrt{2} \end{bmatrix}^{-1} \begin{bmatrix} 19/\sqrt{2} \\ 17/\sqrt{2} \end{bmatrix} = \begin{bmatrix} \frac{19}{20} \\ \frac{17}{10} \end{bmatrix}.$$

(d) Inappropriately 'solve' $\begin{bmatrix} 3/5 & -4/5 \\ 4/5 & 3/5 \end{bmatrix} \boldsymbol{z} = \begin{bmatrix} 19/20 \\ 17/10 \end{bmatrix}$: this matrix is orthogonal, so multiplying by the transpose gives

$$\boldsymbol{z} = \begin{bmatrix} \frac{3}{5} & \frac{4}{5} \\ -\frac{4}{5} & \frac{3}{5} \end{bmatrix} \begin{bmatrix} \frac{19}{20} \\ \frac{17}{10} \end{bmatrix} = \begin{bmatrix} 1.93 \\ 0.26 \end{bmatrix}.$$

And then, since the solution is to be called $\boldsymbol{x}$, we might inappropriately call what we just calculated as the solution $\boldsymbol{x} = (1.93, 0.26)$. □

Avoid this reverse process as it is wrong. Matrix multiplication is *not* commutative (Section 3.1.3). We must use matrix factorization in the correct order: to solve linear equations use the matrices in a factorization 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 factorization is always possible. (http://www.youtube.com/watch?v=JEYLfIVvR9I is an entertaining prelude—their D is our S.)

Theorem 3.3.6 (SVD factorization) *Every $m\times n$ real matrix A can be factored into a product of three matrices*

$$A=USV^{\mathrm{T}}\,, \tag{3.4}$$

*called a **singular value decomposition** (SVD), where*

- *$m\times m$ matrix $U=\begin{bmatrix}\boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_m\end{bmatrix}$ is orthogonal,*
- *$n\times n$ matrix $V=\begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n\end{bmatrix}$ is orthogonal, and*
- *$m\times n$ diagonal matrix S is zero except for non-negative diagonal elements called **singular values** $\sigma_1\,,\sigma_2\,,\ldots,\sigma_{\min(m,n)}$ (the symbol σ is the Greek letter sigma, and denotes singular values), 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 orthonormal sets of vectors $\{\boldsymbol{u}_j\}$ *and* $\{\boldsymbol{v}_j\}$ *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 possible matrices U and V in the SVDs of a matrix, all such SVDs are equivalent in application.

Some people may be disturbed by the non-uniqueness of an SVD. But the non-uniqueness is analogous to the non-uniqueness of Gauss–Jordan elimination upon reordering of equations, and/or reordering the variables in the equations (Section 2.2.2). Do not be disturbed by any non-uniqueness in U and V.

Example 3.3.7 Example 3.3.2 invoked the SVD

$$\begin{bmatrix} 10 & 2 \\ 5 & 11 \end{bmatrix} = \begin{bmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{bmatrix} \begin{bmatrix} 10\sqrt{2} & 0 \\ 0 & 5\sqrt{2} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}^{\mathrm{T}},$$

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

$$\begin{bmatrix} -4 & -2 & 4 \\ -8 & -1 & -4 \\ 6 & 6 & 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\ -\frac{2}{3} & \frac{1}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} 12 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} -\frac{8}{9} & -\frac{1}{9} & -\frac{4}{9} \\ -\frac{4}{9} & \frac{4}{9} & \frac{7}{9} \\ -\frac{1}{9} & -\frac{8}{9} & \frac{4}{9} \end{bmatrix}^{\mathrm{T}},$$

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 $Q = QI_nI_n^{\mathrm{T}}$; that is, $U = Q$, $S = V = I_n$. The S-matrix is the identity. 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)*

- An identity matrix, say I_n, has an SVD $I_n = I_n I_n I_n^{\mathrm{T}}$.
- Additionally, for *every* $n \times n$ orthogonal matrix Q, the identity I_n also has the SVD $I_n = QI_nQ^{\mathrm{T}}$—as this right-hand side $QI_nQ^{\mathrm{T}} = QQ^{\mathrm{T}} = I_n$.
- Further, any constant multiple of an identity, say $sI_n = \operatorname{diag}(s, s, \ldots, s)$, has the same non-uniqueness: an SVD is $sI_n = USV^{\mathrm{T}}$ 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 characterized 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 $QI_nI_n^{\mathrm{T}}$ is an SVD of an orthogonal matrix Q. Which of the following is also an SVD of a given $n \times n$ orthogonal matrix Q?

(a) $I_nQI_n^{\mathrm{T}}$

(b) $I_nI_nQ^{\mathrm{T}}$

(c) $I_nI_n(Q^{\mathrm{T}})^{\mathrm{T}}$

(d) $Q(-I_n)(-I_n)^{\mathrm{T}}$

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 2×2 matrices (Example 3.3.33), 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 3.3.6 asserts that every matrix is the product of two orthogonal matrices and a diagonal matrix. Because, in a matrix's SVD factorization, 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.

This and 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

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]=svd(A)` *computes the three matrices* U, S, *and* V *in a singular value decomposition* (SVD) *of the* $m\times n$ *matrix:* $A=USV^{\mathrm{T}}$ *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).*

  `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 example:*
  - `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* 3×3 *submatrix of the first three columns of the fourth, fifth, and sixth rows.*

to 2 games; Anne beat Chris 3-1; Bob beat Chris 3-2. How good are they all? What are their ratings? That is, we seek to assign a number to each player, called their rating, indicating how good they are at playing: the higher the number the better the player.

□

Compute in Matlab/Octave As seen in the previous example, often we need to compute with a subset of the components of matrices, a submatrix (Table 3.3):

- `b(1:r)` selects the first r entries of vector $\boldsymbol{b}$;
- `S(1:r,1:r)` selects the top-left $r \times r$ submatrix of S;
- `V(:,1:r)` selects the first r columns of matrix 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 best 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):

$$\begin{bmatrix} -9 & -15 & -9 & -15 \\ -10 & 2 & -10 & 2 \\ 8 & 4 & 8 & 4 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} c \\ 8 \\ -5 \end{bmatrix}.$$

□

Procedure 3.3.15 (general solution) *Obtain a general solution of the system $A\boldsymbol{x} = \boldsymbol{b}$ using an* SVD *and via intermediate unknowns.*

1. *Obtain an* SVD *factorization $A = USV^{\mathrm{T}}$.*
2. *Solve $U\boldsymbol{z} = \boldsymbol{b}$ by $\boldsymbol{z} = U^{\mathrm{T}}\boldsymbol{b}$ (unique given U).*
3. *When possible, solve $S\boldsymbol{y} = \boldsymbol{z}$ as follows. Identify the nonzero and the zero singular values: suppose $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ and $\sigma_{r+1} = \cdots = \sigma_{\min(m,n)} = 0$:*

- *if $z_i \neq 0$ for any one (or more) $i = r+1, \ldots, m$, then there is no solution (the equations are* ***inconsistent****);*
- *otherwise (when $z_i = 0$ for every $i = r+1, \ldots, m$) determine the ith component of $\boldsymbol{y}$ by $y_i = z_i/\sigma_i$ for $i = 1, \ldots, r$ (for which $\sigma_i > 0$), whereas for $i = r+1, \ldots, n$ let y_i be a free variable.*

4. *Solve $V^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{y}$ (unique given V and for each $\boldsymbol{y}$) to derive that a general solution is $\boldsymbol{x} = V\boldsymbol{y}$.*

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 $A\boldsymbol{x} = \boldsymbol{b}$ (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, usually).

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:* $\operatorname{cond} A := \sigma_1/\sigma_{\min(m,n)}$ *. By convention: if* $\sigma_{\min(m,n)} = 0$ *, then* $\operatorname{cond} A := \infty$ *(infinity); also, for zero matrices* $\operatorname{cond} O_{m\times n} := \infty$ *.*

Example 3.3.17 Example 3.3.7 gives the singular values of two matrices: for the 2×2 matrix the condition number $\sigma_1/\sigma_2 = (10\sqrt{2})/(5\sqrt{2}) = 2$ (for which $\texttt{rcond} = 0.5$); for the 3×3 matrix the condition number $\sigma_1/\sigma_3 = 12/3 = 4$ (for which $\texttt{rcond} = 0.25$). Example 3.3.8 comments that every $n \times n$ orthogonal matrix has singular values $\sigma_1 = \cdots = \sigma_n = 1$; hence every orthogonal matrix has condition number one ($\texttt{rcond} = 1$). Such condition numbers less than 100 (non-small rcond) indicate that all these 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, $\texttt{rcond} = 0$) 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.) $\square$

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.)

$$\begin{bmatrix} -0.86 & 0.43 & 0.29 \\ -0.29 & -0.86 & 0.43 \\ 0.43 & 0.29 & 0.86 \end{bmatrix} \begin{bmatrix} 28 & 0 & 0 & 0 \\ 0 & 14 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.50 & 0.50 & -0.19 & -0.68 \\ 0.50 & -0.50 & 0.68 & -0.19 \\ 0.50 & 0.50 & 0.19 & 0.68 \\ 0.50 & -0.50 & -0.68 & 0.19 \end{bmatrix}^{\mathsf{T}}$$

(a) 0 (b) 0.5 (c) ∞ (d) 2

In practice, a condition number $> 10^8$ is effectively infinite (equivalently `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 less than $10^{-8}\sigma_1$ are effectively zero.

Definition 3.3.19 *The* ***rank*** *of a matrix A is the number of* nonzero *singular values in an* SVD, $A = USV^{\mathsf T}$ *: letting* $r = \operatorname{rank} A$,

$$S = \begin{bmatrix} \begin{matrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_r \end{matrix} & O_{r\times(n-r)} \\ & \\ O_{(m-r)\times r} & O_{(m-r)\times(n-r)} \end{bmatrix},$$

equivalently $S = \operatorname{diag}_{m\times n}(\sigma_1\,, \sigma_2\,, \ldots, \sigma_r, 0, \ldots, 0)$.

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 that 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.)

$$\begin{bmatrix} -0.86 & 0.43 & 0.29 \\ -0.29 & -0.86 & 0.43 \\ 0.43 & 0.29 & 0.86 \end{bmatrix} \begin{bmatrix} 28 & 0 & 0 & 0 \\ 0 & 14 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.50 & 0.50 & -0.19 & -0.68 \\ 0.50 & -0.50 & 0.68 & -0.19 \\ 0.50 & 0.50 & 0.19 & 0.68 \\ 0.50 & -0.50 & -0.68 & 0.19 \end{bmatrix}^{\text{T}}.$$

(a) 4 (b) 2 (c) 3 (d) 1

Example 3.3.22 Use MATLAB/Octave to find the rank of the matrix $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & -1 \\ 1 & 0 & -1 \\ 2 & 0 & -2 \end{bmatrix}$ □

Example 3.3.23 Use MATLAB/Octave to find the rank of the matrix $\begin{bmatrix} 1 & -2 & -1 & 2 & 1 \\ -2 & -2 & -0 & 2 & -0 \\ -2 & -3 & 1 & -1 & 1 \\ -3 & 0 & 1 & -0 & -1 \\ 2 & 1 & 1 & 2 & -1 \end{bmatrix}$ □

Theorem 3.3.24 *For every matrix A, let an SVD of A be USV^{T}, then the transpose A^{T} has an SVD of $V(S^{\mathrm{T}})U^{\mathrm{T}}$. Further,* $\operatorname{rank}(A^{\mathrm{T}}) = \operatorname{rank} A$ *.*

Example 3.3.25 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 3.3.26 Recall that

$$\begin{bmatrix} -0.86 & 0.43 & 0.29 \\ -0.29 & -0.86 & 0.43 \\ 0.43 & 0.29 & 0.86 \end{bmatrix} \begin{bmatrix} 28 & 0 & 0 & 0 \\ 0 & 14 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.50 & 0.50 & -0.19 & -0.68 \\ 0.50 & -0.50 & 0.68 & -0.19 \\ 0.50 & 0.50 & 0.19 & 0.68 \\ 0.50 & -0.50 & -0.68 & 0.19 \end{bmatrix}^{\mathsf{T}}$$

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?

(a)
$$\begin{bmatrix} -0.86 & 0.43 & 0.29 \\ -0.29 & -0.86 & 0.43 \\ 0.43 & 0.29 & 0.86 \end{bmatrix} \begin{bmatrix} 28 & 0 & 0 \\ 0 & 14 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.50 & 0.50 & -0.19 & -0.68 \\ 0.50 & -0.50 & 0.68 & -0.19 \\ 0.50 & 0.50 & 0.19 & 0.68 \\ 0.50 & -0.50 & -0.68 & 0.19 \end{bmatrix}^{\mathsf{T}}$$

(b)
$$\begin{bmatrix} 0.50 & 0.50 & 0.50 & 0.50 \\ 0.50 & -0.50 & 0.50 & -0.50 \\ \end{bmatrix} \begin{bmatrix} 28 & 0 & 0 \\ 0 & 14 & 0 \\ \end{bmatrix} \begin{bmatrix} -0.86 & -0.29 & 0.43 \\ 0.43 & -0.86 & 0.29 \end{bmatrix}^{\mathsf{T}}$$

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 3.3.27 (Unique Solutions: version 1) *For every $n \times n$ square matrix A, the following statements are equivalent:*

(a) *A is invertible;*

(b) *$A\boldsymbol{x} = \boldsymbol{b}$ has a **unique solution** for every $\boldsymbol{b}$ in $\mathbb{R}^n$;*

(c) *$A\boldsymbol{x} = \boldsymbol{0}$ has* only *the zero solution;*

(d) *all n singular values of A are nonzero;*

(e) *the condition number of A is finite (`rcond` > 0);*

(f) $\operatorname{rank} A = n$.

Practical shades of grey The preceding Unique Solution Theorem 3.3.27 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 $\boldsymbol{b}$ to three significant digits, then solve the linear equations $A\boldsymbol{x} = \boldsymbol{b}$ to estimate quantities of interest $\boldsymbol{x}$: how accurate are your estimates of the interesting quantities $\boldsymbol{x}$? or are your estimates complete nonsense?

This (optional) discussion and theorem reinforces why we must check condition numbers in computation.

Example 3.3.28 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%? (Recall that the *relative error* of an approximation, compared to the exact, is $\big(|(\text{approx}) - (\text{exact})|\big)/|\text{exact}|$.) 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 $\boldsymbol{x} = (p, q, r)$, write the system as $A\boldsymbol{x} = \boldsymbol{b}$ for matrix

$$A = \begin{bmatrix} 0 & -2 & 1 \\ 1 & -5 & 1 \\ -3 & 2 & 3 \end{bmatrix}, \quad \text{and right-hand side } \boldsymbol{b} = \begin{bmatrix} 3 \\ 8 \\ -5 \end{bmatrix}.$$

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;

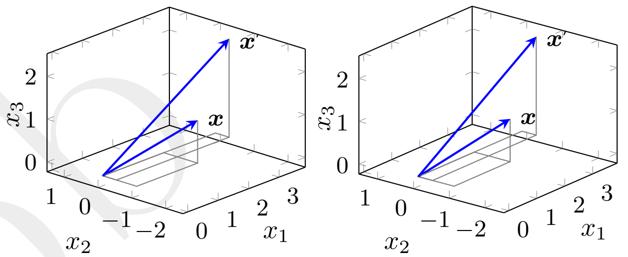
iii. then `x=A\b` gives the solution $\boldsymbol{x} = (2, -1, 1)$ as before.

(b) Now let's recognize that the right-hand side comes from experimental measurements with a 1% error. In Matlab/Octave, `norm(b)` computes the length $|\boldsymbol{b}| = 9.90$ (2 d.p.). Thus a 1% error corresponds to changing $\boldsymbol{b}$ by $0.01 \times 9.90 \approx 0.1$. Let's say the first component of $\boldsymbol{b}$ 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 $\boldsymbol{b}$ and then solves the new system to find $\boldsymbol{x}' = (3.7, -0.4, 2.3)$. This solution is very different to the original solution $\boldsymbol{x} = (2, -1, 1)$!

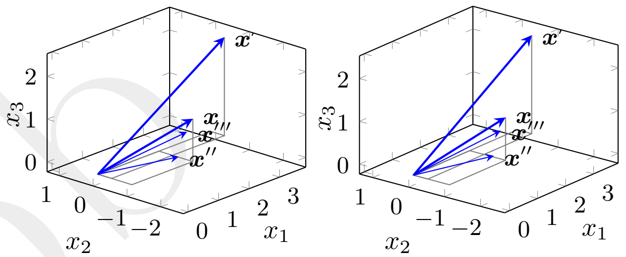
ii. `relerr1=norm(x-x1)/norm(x)` computes its relative error $|\boldsymbol{x} - \boldsymbol{x}'|/|\boldsymbol{x}|$ to be 0.91, that is, 91%—a grossly large relative error.

As illustrated to the right, the large difference



between $\boldsymbol{x}$ and $\boldsymbol{x}'$ indicates that 'the solution' $\boldsymbol{x}$ is almost complete nonsense. How can a 1% error in $\boldsymbol{b}$ turn into the astonishingly large 91% error in solution $\boldsymbol{x}$? Theorem 3.3.30 below shows that it is no accident that the magnification of the error by a factor of 91 (from 1% to 91%)) is of the same order of magnitude as the condition number $= 152.27$ computed via `s=svd(A)` and then `condA=s(1)/s(3)`.

(c) To explore further, let's say the second component



of $\boldsymbol{b}$ is in error by 1% of $\boldsymbol{b}$, 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 $\boldsymbol{x}'' = (1.2\,,-1.3\,,0.4)$ which is quite different to both $\boldsymbol{x}$ and $\boldsymbol{x}'$, as illustrated to the right. Compute its relative error $|\boldsymbol{x}-\boldsymbol{x}''|/|\boldsymbol{x}| = 0.43\,$. At 43%, the relative error in solution $\boldsymbol{x}''$ is also much larger than the 1% error in $\boldsymbol{b}$.

(d) Lastly, let's say the third component of $\boldsymbol{b}$ is in error by 1% of $\boldsymbol{b}$, 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 $\boldsymbol{x}''' = (1.7, -1.1, 0.8)$ which, as illustrated above, is at least roughly $\boldsymbol{x}$. Compute its relative error $|\boldsymbol{x}-\boldsymbol{x}'''|/|\boldsymbol{x}| = 0.15\,$. At 15%, the relative error in solution $\boldsymbol{x}'''$ is significantly larger than the 1% error in $\boldsymbol{b}$.

This example shows that the apparently innocuous matrix A variously multiplies measurement errors in $\boldsymbol{b}$ by factors of at least up to 91 when finding 'the solution' $\boldsymbol{x}$ to $A\boldsymbol{x} = \boldsymbol{b}$. The matrix A must, after all, be a bad matrix. Theorem 3.3.30 shows that this badness is quantified by its condition number 152.27, and its poor (estimated) reciprocal `rcond(A)` $= 0.0031$. □

Example 3.3.29 Consider solving the system of linear equations

$$\begin{bmatrix} 0.4 & 0.4 & -0.2 & 0.8 \\ -0.2 & 0.8 & -0.4 & -0.4 \\ 0.4 & -0.4 & -0.8 & -0.2 \\ -0.8 & -0.2 & -0.4 & 0.4 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} -3 \\ 3 \\ -9 \\ -1 \end{bmatrix}.$$

Use Matlab/Octave to explore the effect on the solution $\boldsymbol{x}$ 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 characterizes 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 3.3.28.

Theorem 3.3.30 (error magnification) *Consider solving $A\boldsymbol{x} = \boldsymbol{b}$ for $n \times n$ matrix A with full $\operatorname{rank} A = n$. Suppose the right-hand side $\boldsymbol{b}$ has relative error of magnitude ϵ (the symbol ϵ is the Greek letter epsilon, and often denotes errors), then the solution $\boldsymbol{x}$ has relative error $\leq \epsilon \operatorname{cond} A$, with equality in the worst case.*

Example 3.3.31 Each of the following cases involves solving a linear system $A\boldsymbol{x} = \boldsymbol{b}$ to determine quantities of interest $\boldsymbol{x}$ from some measured quantities $\boldsymbol{b}$. From the given information, estimate the maximum relative error in $\boldsymbol{x}$, if possible, otherwise say so.

(a) Quantities $\boldsymbol{b}$ are measured to a relative error 0.001, and matrix A has condition number of ten.

(b) Quantities $\boldsymbol{b}$ 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 `rcond(A)` $= 0.002$.

$\square$

Activity 3.3.32 In some experiment the components of $\boldsymbol{b}$, $|\boldsymbol{b}| = 5$, are measured to two decimal places. We compute a vector $\boldsymbol{x}$ by solving $A\boldsymbol{x} = \boldsymbol{b}$. For matrix A, we compute `rcond(A)` $= 0.02$. What is our estimate of the largest possible relative error in $\boldsymbol{x}$?

(a) 2% (b) 0.1% (c) 20% (d) 5%

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 $A\boldsymbol{x} = \boldsymbol{b}$ is the following: it may happen that a small error in the elements of $\boldsymbol{b}$ produces an enormous error in the final $\boldsymbol{x}$. The condition number warns when this happens by characterizing 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*

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 two key parts of the general proof in a definite setting.

Example 3.3.33 *(a 2×2 case)* Recall Example 3.3.2 factorized the matrix

$$A = \begin{bmatrix} 10 & 2 \\ 5 & 11 \end{bmatrix} = \begin{bmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{bmatrix} \begin{bmatrix} 10\sqrt{2} & 0 \\ 0 & 5\sqrt{2} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}^{\mathrm{T}}.$$

Here we find this factorization, $A = USV^{\mathrm{T}}$, by maximizing $|A\boldsymbol{v}|$ over all unit vectors $\boldsymbol{v}$ (all vectors of length one).

$\square$

Example 3.3.34 *(a 3×1 case)* Find the following SVD for the 3×1 matrix

$$A=\begin{bmatrix}1\\1\\1\end{bmatrix}=\begin{bmatrix}\frac{1}{\sqrt3}&\cdot&\cdot\\ \frac{1}{\sqrt3}&\cdot&\cdot\\ \frac{1}{\sqrt3}&\cdot&\cdot\end{bmatrix}\begin{bmatrix}\sqrt3\\0\\0\end{bmatrix}\begin{bmatrix}1\end{bmatrix}^{\mathrm{T}}=USV^{\mathrm{T}},$$

where we do not worry about the elements denoted by dots as they are only multiplied by the zeros in the 'diagonal' $S=(\sqrt3,0,0)$. □

Outline of the general proof We use induction on the size $m\times n$ of the matrix.

- First, zero matrices have a trivial SVD, and, further, $m\times 1$ and $1\times n$ matrices have a straightforward SVD (as in Example 3.3.34).
- For any given $m\times n$ matrix A, choose $\boldsymbol{v}_1$ to maximize $|A\boldsymbol{v}|^2$ among all unit vectors $\boldsymbol{v}$ in $\mathbb{R}^n$.

- Crucially, we then establish that for every vector $\boldsymbol{v}$ orthogonal to $\boldsymbol{v}_1$, the vector $A\boldsymbol{v}$ is orthogonal to $A\boldsymbol{v}_1$.
- Then rotate the standard unit vectors to align one with $\boldsymbol{v}_1$. Similarly for $A\boldsymbol{v}_1$.
- This rotation transforms the matrix A to strip off the leading singular value, and to 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 $USV^{\mathrm{T}}=I_mO_{m\times n}I_n^{\mathrm{T}}=O_{m\times n}=A$.

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 $A=\begin{bmatrix}\boldsymbol{a}_1\end{bmatrix}$ for $\boldsymbol{a}_1=(a_{11},a_{21},\ldots,a_{m1})$. Set the singular value $\sigma_1=|\boldsymbol{a}_1|=\sqrt{a_{11}^2+a_{21}^2+\cdots+a_{m1}^2}$ and unit vector $\boldsymbol{u}_1=\boldsymbol{a}_1/\sigma_1$. Set 1×1 orthogonal matrix $V=\begin{bmatrix}1\end{bmatrix}$; $m\times 1$ diagonal matrix $S=(\sigma_1,0,\ldots,0)$; and $m\times m$ orthogonal matrix $U=\begin{bmatrix}\boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_m\end{bmatrix}$. 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 $\boldsymbol{u}_1$; the other $(m-1)$ unit vectors then become the other $\boldsymbol{u}_j$. Then an SVD for the $m \times 1$ matrix A is

$$USV^{\mathrm{T}} = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_m \end{bmatrix} \begin{bmatrix} \sigma_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} 1^{\mathrm{T}} = \sigma_1 \boldsymbol{u}_1 = \begin{bmatrix} \boldsymbol{a}_1 \end{bmatrix} = A\,.$$

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 with $m, n \geq 2$. Set vector $\boldsymbol{v}_1$ in $\mathbb{R}^n$ to be a *unit vector* that maximizes $|A\boldsymbol{v}|^2$ for unit vectors $\boldsymbol{v}$ in $\mathbb{R}^n$; that is, vector $\boldsymbol{v}_1$ achieves the maximum in $\max_{|\boldsymbol{v}|=1} |A\boldsymbol{v}|^2$.

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 $\boldsymbol{v}$ such that $|A\boldsymbol{v}| > 0$. Since $\boldsymbol{v}_1$ maximizes $|A\boldsymbol{v}|$ it follows that $|A\boldsymbol{v}_1| > 0$.

   The vector $\boldsymbol{v}_1$ is not unique: for example, the negative $-\boldsymbol{v}_1$ 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 $|A\boldsymbol{v}|^2$ is unique, and so the following singular value σ_1 is unique.

2. *Set the singular value $\sigma_1 := |A\boldsymbol{v}_1| > 0$ and unit vector $\boldsymbol{u}_1 := (A\boldsymbol{v}_1)/\sigma_1$ in $\mathbb{R}^m$. For every unit vector $\boldsymbol{v}$ orthogonal to $\boldsymbol{v}_1$ we now prove that the vector $A\boldsymbol{v}$ is orthogonal to $\boldsymbol{u}_1$.* Let $\boldsymbol{u} := A\boldsymbol{v}$ in $\mathbb{R}^m$ and consider $f(t) := |A(\boldsymbol{v}_1 \cos t + \boldsymbol{v} \sin t)|^2$. Since $\boldsymbol{v}_1$ achieves the maximum, and $\boldsymbol{v}_1 \cos t + \boldsymbol{v} \sin t$ 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,

$$\begin{aligned} f(t) &= |A\boldsymbol{v}_1\cos t + A\boldsymbol{v}\sin t|^2 \\ &= |\sigma_1\boldsymbol{u}_1\cos t + \boldsymbol{u}\sin t|^2 \\ &= (\sigma_1\boldsymbol{u}_1\cos t + \boldsymbol{u}\sin t)\cdot(\sigma_1\boldsymbol{u}_1\cos t + \boldsymbol{u}\sin t) \\ &= \sigma_1^2\cos^2 t + \sigma_1\boldsymbol{u}\cdot\boldsymbol{u}_1 2\sin t\,\cos t + |\boldsymbol{u}|^2\sin^2 t\,; \end{aligned}$$

differentiating $f(t)$ and evaluating at zero gives $f'(0) = \sigma_1\boldsymbol{u}\cdot\boldsymbol{u}_1$. As $t=0$ is a maximum, this derivative is zero, so $\sigma_1\boldsymbol{u}\cdot\boldsymbol{u}_1 = 0$. Since the singular value $\sigma_1 > 0$, we must have $\boldsymbol{u}\cdot\boldsymbol{u}_1 = 0$ and so $\boldsymbol{u}_1$ and $\boldsymbol{u}$ 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 $\boldsymbol{v}_1$, 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 $V_1 := \begin{bmatrix}\boldsymbol{v}_1 & \bar V\end{bmatrix}$ which is orthogonal as its columns are orthonormal (Theorem 3.2.48(b)). Similarly, set an $m\times m$ orthogonal matrix $U_1 := \begin{bmatrix}\boldsymbol{u}_1 & \bar U\end{bmatrix}$. Compute the $m\times n$ matrix

$$A_1 := U_1^{\mathrm T}AV_1 = \begin{bmatrix}\boldsymbol{u}_1^{\mathrm T}\\ \bar U^{\mathrm T}\end{bmatrix} A \begin{bmatrix}\boldsymbol{v}_1 & \bar V\end{bmatrix} = \begin{bmatrix}\boldsymbol{u}_1^{\mathrm T}A\boldsymbol{v}_1 & \boldsymbol{u}_1^{\mathrm T}A\bar V\\ \bar U^{\mathrm T}A\boldsymbol{v}_1 & \bar U^{\mathrm T}A\bar V\end{bmatrix}$$

where

- the top-left entry $\boldsymbol{u}_1^{\mathrm{T}} A\boldsymbol{v}_1 = \boldsymbol{u}_1^{\mathrm{T}}\sigma_1\boldsymbol{u}_1 = \sigma_1|\boldsymbol{u}_1|^2 = \sigma_1$,
- the bottom-left column $\bar U^{\mathrm{T}} A\boldsymbol{v}_1 = \bar U^{\mathrm{T}}\sigma_1\boldsymbol{u}_1 = O_{m-1\times 1}$ as the columns of $\bar U$ are orthogonal to $\boldsymbol{u}_1$,
- the top-right row $\boldsymbol{u}_1^{\mathrm{T}} A\bar V = O_{1\times n-1}$ as each column of $\bar V$ is orthogonal to $\boldsymbol{v}_1$ and hence each column of $A\bar V$ is orthogonal to $\boldsymbol{u}_1$,
- and set the bottom-right block $B := \bar U^{\mathrm{T}} A\bar V$ which is an $(m-1)\times(n-1)$ matrix as $\bar U^{\mathrm{T}}$ 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 $A_1 := U_1^{\mathrm{T}} AV_1$ gives $AV_1 = U_1A_1$.

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 $B = \hat U\hat S\hat V^{\text{T}}$ be an SVD for B. Then construct (for appropriately sized zero matrices O)

$$U := U_1\begin{bmatrix}1&O\\O&\hat U\end{bmatrix},\quad V := V_1\begin{bmatrix}1&O\\O&\hat V\end{bmatrix},\quad S := \begin{bmatrix}\sigma_1&O\\O&\hat S\end{bmatrix}.$$

Matrices U and V are orthogonal as each is the product of two orthogonal matrices (**??**). Also, matrix S is diagonal. These form an SVD for matrix A since

$$\begin{aligned}AV &= AV_1\begin{bmatrix}1&O\\O&\hat V\end{bmatrix} = U_1A_1\begin{bmatrix}1&O\\O&\hat V\end{bmatrix} = U_1\begin{bmatrix}\sigma_1&O\\O&B\end{bmatrix}\begin{bmatrix}1&O\\O&\hat V\end{bmatrix}\\ &= U_1\begin{bmatrix}\sigma_1&O\\O&B\hat V\end{bmatrix} = U_1\begin{bmatrix}\sigma_1&O\\O&\hat U\hat S\end{bmatrix} = U_1\begin{bmatrix}1&O\\O&\hat U\end{bmatrix}\begin{bmatrix}\sigma_1&O\\O&\hat S\end{bmatrix} = US\end{aligned}$$

Hence $A = USV^{\text{T}}$ is an SVD for A.

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

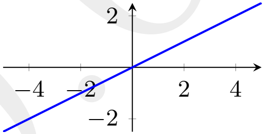
> [Nature] is written in that great book which ever lies before our eyes—I mean the universe—but we cannot understand it if we do not first learn the language and grasp the symbols in which it is written. The book is written in the mathematical language, and the symbols are triangles, circles, and other 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 involves linear combinations such as $(-2,1,0,0)s+(-\frac{15}{7},0,\frac{9}{7},1)t$ (Example 2.2.29(d)) and $y_3\boldsymbol{v}_3+y_4\boldsymbol{v}_4$ (Example 3.3.14): such combinations for all values of the free variables form a plane through the origin (Section 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.

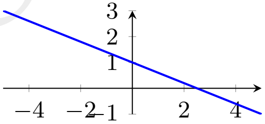
3.4.1 Subspaces are lines, planes, and so on

Let's introduce graphically the concept of a "subspace". Definition 3.4.3 then gives the precise algebraic description.

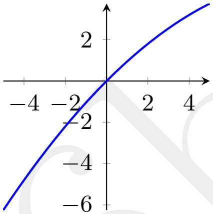
Example 3.4.1 The following graphs illustrate the concept of subspaces through examples (imagine the graphs extend to infinity, as appropriate).

(a) 

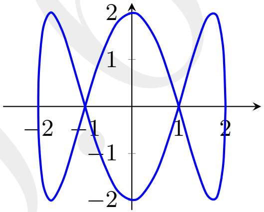
is a subspace as it is a straight line through the origin.

(b) 

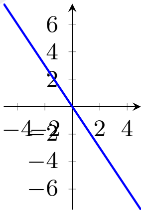
is *not* a subspace as it does not include the origin.

(c) 

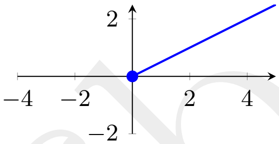
is *not* a subspace as it curves.

(d) 

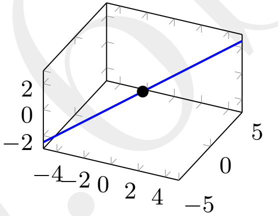
is *not* a subspace as it not only curves, but also does not include the origin.

(e) 

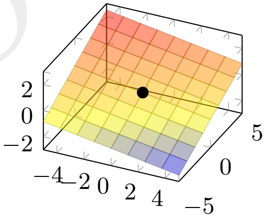
is a subspace.

(f) 

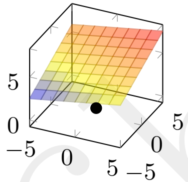
where the disc indicates an end to the line, is *not* a subspace as it does not extend infinitely in both directions.

(g) 

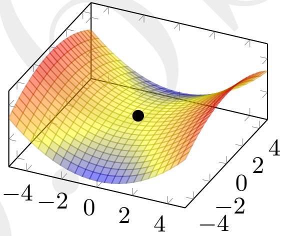
is a subspace as it is a line through the origin (marked • in these 3D plots).

(h) 

is a subspace as it is a plane through the origin.

(i) 

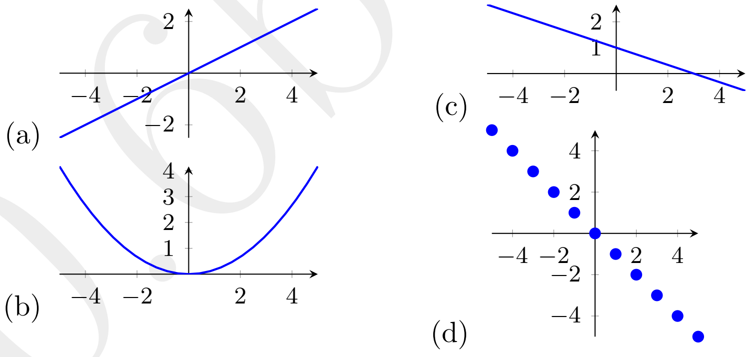
is *not* a subspace as it does not go through the origin.

(j) 

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 "$\boldsymbol{v} \in \mathbb{R}^3$" means "$\boldsymbol{v}$ is a vector with three components". For conciseness, 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 for every $c \in \mathbb{R}$ and for every $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{W}$, then both $\boldsymbol{u} + \boldsymbol{v} \in \mathbb{W}$ and $c\boldsymbol{u} \in \mathbb{W}$. (When these properties hold we often say that $\mathbb{W}$ is **closed** under addition and scalar multiplication.)*

Example 3.4.4 Use Definition 3.4.3 to determine whether each of the following are subspaces, or not.

(a) All vectors (x, y) in the line $y = x/2$ (Example 3.4.1(a)).

(b) All vectors (x, y) with end-points on the curve $y = x - x^2/20$ (Example 3.4.1(c)).

(c) All vectors (x, y) in the line $y = x/2$ for $x, y \geq 0$ (Example 3.4.1(f)).

(d) All vectors (x, y, z) in the plane $z = -x/6 + y/3$ (Example 3.4.1(h)).

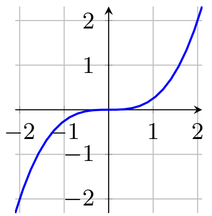
(e) All vectors (x, y, z) in the plane $z = 5 + x/6 + y/3$ (Example 3.4.1(i)).

(f) $\{\mathbf{0}\}$ (the set of the zero vector in some $\mathbb{R}^n$).

(g) $\mathbb{R}^n$.

□

Activity 3.4.5 The following pairs of vectors are all in the set shown to the right (in the sense that their end-points lie on the plotted curve). The sum of which pair proves that the curve plotted to the right is not a subspace?



(a) $(0\,,0)\,,(2\,,2)$

(b) $(2\,,2)\,,(-2\,,-2)$

(c) $(-1\,,-\tfrac{1}{4})\,,(2\,,2)$

(d) $(1\,,\tfrac{1}{4})\,,(0\,,0)$

In summary:

- in two dimensions (denoted $\mathbb{R}^2$), subspaces are the origin $\{\mathbf{0}\}$, any line through $\mathbf{0}$, or the entire plane $\mathbb{R}^2$;

- in three dimensions (denoted $\mathbb{R}^3$), subspaces are the origin $\{\mathbf{0}\}$, any line through $\mathbf{0}$, any plane through $\mathbf{0}$, or the entire space $\mathbb{R}^3$;
- and analogously for higher dimensions (denoted $\mathbb{R}^n$).

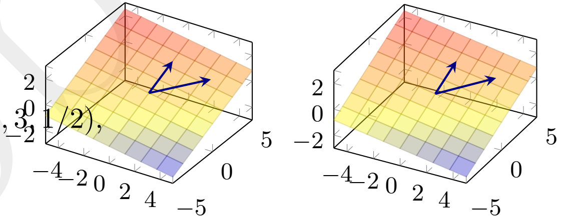
Recall that the set of all linear combinations of a set of vectors, such as $(-2, 1, 0, 0)s + (-\frac{15}{7}, 0, \frac{9}{7}, 1)t$ (Example 2.2.29(d)), is called the span of that set (Definition 2.3.10).

Theorem 3.4.6 *Let $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k$ be k vectors in $\mathbb{R}^n$, then $\operatorname{span}\{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k\}$ is a subspace of $\mathbb{R}^n$.*

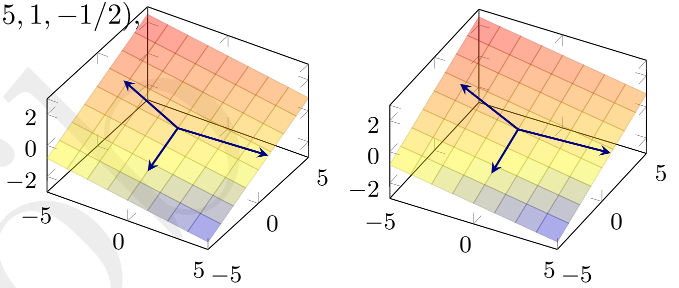
Example 3.4.7 $\operatorname{span}\{(1, \frac{1}{2})\}$ is the subspace $y = x/2$. The reason is that a vector $\boldsymbol{u} \in \operatorname{span}\{(1, \frac{1}{2})\}$ only if there is some constant a_1 such that $\boldsymbol{u} = a_1(1, \frac{1}{2}) = (a_1, a_1/2)$. That is, the y-component is half the x-component and hence it lies on the line $y = x/2$.

$\operatorname{span}\{(1, \frac{1}{2}), (-2, -1)\}$ 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 $\operatorname{span}\{(3,3,1/2), (0,3,1)\}$, as illustrated in stereo to the right, 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).

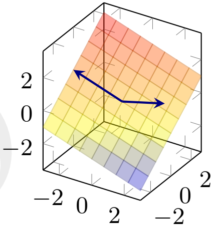


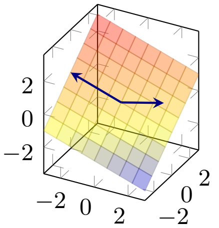
Also, span$\{(5,1,-1/2),(0,-3,-1),(-4,1,1)\}$ is the plane $z=-x/6+y/3$, as illustrated to the right.



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. $\square$

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.

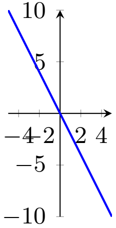
Definition 3.4.10 *For every $m \times n$ matrix A:*

*(a) the **column space** of A is the subspace of $\mathbb{R}^m$ spanned by the n column vectors of A;*

*(b) the **row space** of A is the subspace of $\mathbb{R}^n$ spanned by the m row vectors (transposed) of A.*

Example 3.4.11 Examples 3.4.7 to 3.4.9 provide some cases.

- From Example 3.4.7, the column space of $A = \begin{bmatrix} 1 & -2 \\ 1/2 & -1 \end{bmatrix}$ is the line $y = x/2$.

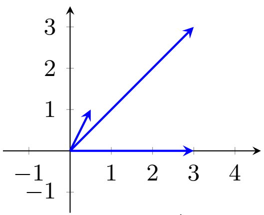
  

  The row space of this matrix A is span$\{(1\,,-2)\,,(\frac{1}{2},-1)\}$. This row space is the set of all vectors of the form $(1,-2)s+(\frac{1}{2},-1)t = (s+t/2,-2s-t) = (1,-2)(s+t/2) = (1,-2)t'$ is the line $y = -2x$ as illustrated to the right. That the row space and the column space are both lines, 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 $\mathbb{R}^3$.

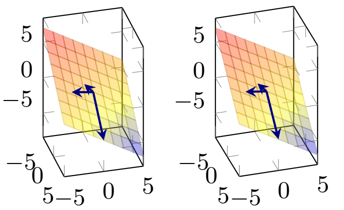


The row space of matrix B is $\operatorname{span}\{(3,0),(3,3),(\frac{1}{2},1)\}$ which is a subspace of $\mathbb{R}^2$—the right-hand plot shows the three vectors. Whereas the column space is a subspace of $\mathbb{R}^3$. Here the span is all of $\mathbb{R}^2$ as for each $(x,y)\in\mathbb{R}^2$ choose 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, 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 to the right 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).

$\square$

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) $\begin{bmatrix} 8 \\ 2 \\ -3 \end{bmatrix}$ (b) $\begin{bmatrix} 8 \\ 5 \\ -2 \end{bmatrix}$ (c) $\begin{bmatrix} 2 \\ 2 \\ -3 \end{bmatrix}$ (d) $\begin{bmatrix} 2 \\ -3 \\ -3 \end{bmatrix}$

Example 3.4.13 Is vector $\boldsymbol{b} = (-0.6, 0, -2.1, 1.9, 1.2)$ 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 $\boldsymbol{c} = (15.2, 5.4, 3.8, -1.9, -3.7)$?

$\square$

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* $\operatorname{null}(A)$ *to be all the solutions $\boldsymbol{x}$ of the homogeneous system $A\boldsymbol{x} = \boldsymbol{0}$. The set* $\operatorname{null}(A)$ *is a subspace of $\mathbb{R}^n$ called the **nullspace** of A.*

Example 3.4.15

- Example 2.2.29(a) showed that the only solution of the homogeneous system $\left\{\begin{smallmatrix} 3x_1-3x_2=0 \\ -x_1-7x_2=0\end{smallmatrix}\right.$ is $\boldsymbol{x} = \boldsymbol{0}$. Thus its set of solutions is $\{\boldsymbol{0}\}$ which is a subspace (Example 3.4.4(f)). Thus $\{\boldsymbol{0}\}$ is the nullspace of matrix $\left[\begin{smallmatrix} 3 & -3 \\ -1 & -7\end{smallmatrix}\right]$.
- Recall that the homogeneous system of linear equations from Example 2.2.29(d) has solutions $\boldsymbol{x} = (-2s - \frac{15}{7}t, s, \frac{9}{7}t, t) = (-2, 1, 0, 0)s + (-\frac{15}{7}, 0, \frac{9}{7}, 1)t$ for arbitrary s and t. That is,

the set of solutions is $\operatorname{span}\{(-2,1,0,0)\,,(-\frac{15}{7},0,\frac{9}{7},1)\}$. 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.

$\square$

Example 3.4.16 Is the vector $\boldsymbol{v}=(-2,6,1)$ in the null space of $A=\begin{bmatrix} 3&1&0\\ -5&-1&-4 \end{bmatrix}$? What about vector $\boldsymbol{w}=(1,-3,2)$? $\square$

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) $\begin{bmatrix} 3 \\ -4 \\ 0 \end{bmatrix}$ (d) $\begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}$

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)

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 challenge.

We prefer to use an orthonormal set of vectors to span a subspace. The virtue is that orthonormal sets have many practically useful properties. Because of their beautiful properties, orthonormal sets underpin JPEG images, our understanding of vibrations, reliable weather forecasting, and much more. Recall that an orthonormal set (Definition 3.2.38) is composed of vectors that are both all at right-angles to each other (their dot products are zero) and all of unit length.

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 3.4.4(g)).

(a) The n standard unit vectors $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ 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 $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) = x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + \cdots + x_n\boldsymbol{e}_n$. Hence the set of standard unit vectors in $\mathbb{R}^n$ is an orthonormal basis for the subspace $\mathbb{R}^n$.

(b) The n columns $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_n$ of an $n\times n$ orthogonal matrix Q also form an orthonormal basis for the subspace $\mathbb{R}^n$. The reasons are: first, Theorem 3.2.48(b) establishes that the column vectors of Q are orthonormal; and second they span the subspace $\mathbb{R}^n$, as for every vector $\boldsymbol{x} \in \mathbb{R}^n$ there exists a linear combination $\boldsymbol{x} = c_1\boldsymbol{q}_1 + c_2\boldsymbol{q}_2 + \cdots + c_n\boldsymbol{q}_n$ obtained by solving $Q\boldsymbol{c} = \boldsymbol{x}$ through calculating $\boldsymbol{c} = Q^{\mathrm{T}}\boldsymbol{x}$ since Q^{T} is the inverse of an orthogonal matrix Q (Theorem 3.2.48(c)).

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) $\{(1\,,1)\,,(1\,,-1)\}$

(b) $\{\frac{1}{5}(3\,,-4)\,,\ \frac{1}{13}(12\,,5)\}$

(c) $\{\mathbf{0}\,,\boldsymbol{i}\,,\boldsymbol{j}\}$

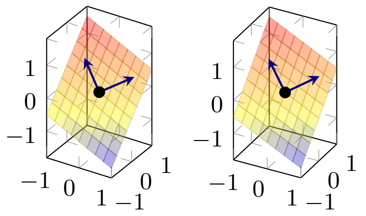
(d) $\{\frac{1}{2}(1\,,\sqrt{3})\,,\ \frac{1}{2}(-\sqrt{3}\,,1)\}$

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 $\{\boldsymbol{u}_1,\boldsymbol{u}_2\}$ where $\boldsymbol{u}_1=(-\frac{2}{3},\frac{1}{3},\frac{2}{3})$, and $\boldsymbol{u}_2=(\frac{2}{3},\frac{2}{3},\frac{1}{3})\}$ as illustrated in stereo to the right.



□

Procedure 3.4.23 (orthonormal basis for a span) *Let $\{\boldsymbol{a}_1,\boldsymbol{a}_2,\ldots,\boldsymbol{a}_n\}$ be a set of n vectors in $\mathbb{R}^m$, then the following procedure finds an orthonormal basis for the subspace* $\operatorname{span}\{\boldsymbol{a}_1,\boldsymbol{a}_2,\ldots,\boldsymbol{a}_n\}$.

1. *Form $m\times n$ matrix $A:=\begin{bmatrix}\boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n\end{bmatrix}$.*
2. *Factorize A into an* SVD, *$A=USV^{\text{T}}$. Let $\boldsymbol{u}_1,\boldsymbol{u}_2,\ldots,\boldsymbol{u}_m$ denote the m columns of U (singular vectors), and let $r=$* $\operatorname{rank}A$ *be the number of nonzero singular values (Definition 3.3.19).*
3. *Then $\{\boldsymbol{u}_1,\boldsymbol{u}_2,\ldots,\boldsymbol{u}_r\}$ is an orthonormal basis for the subspace* $\operatorname{span}\{\boldsymbol{a}_1,\boldsymbol{a}_2,\ldots,\boldsymbol{a}_n\}$.

Example 3.4.24 Compute an orthonormal basis for $\operatorname{span}\{(1, \frac{1}{2}), (-2, -1)\}$. $\square$

Example 3.4.25 Recall that Example 3.4.8 found the plane $z = -x/6 + y/3$ could be written as $\operatorname{span}\{(3, 3, 1/2), (0, 3, 1)\}$ or as $\operatorname{span}\{(5, 1, -1/2), (0, -3, -1), (-4, 1, 1)\}$. Use each of these spans to find two different orthonormal bases for the plane. $\square$

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 by `[U,S,V]=svd(A)` in MATLAB/Octave: what is an orthonormal basis for the column space of the matrix A (2 d.p.)?

```
U =
  -0.67   0.69   0.27
  -0.55  -0.23  -0.80
  -0.49  -0.69   0.53
S =
   9.17      0      0
      0   2.83      0
      0      0   0.00
V =
  -0.75  -0.32  -0.58
  -0.66   0.49   0.58
   0.09   0.81  -0.58
```

Example 3.4.27 *(data reduction)* Every four or five years the phenomenon of El Niño makes a large impact on the world's weather: from drought in Australia to floods in South America. We would like to predict El Niño in advance to save lives and businesses. El Niño 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 Niño.

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.

Let's use a general technique called Singular Spectrum Analysis. The figure shows that the SOI oscillates, to and fro, a couple of times every ten years. This suggests that if we analyse ten year 'snapshots', or 'windows', of the SOI data then there should be some common pattern of oscillations apparent—somehow. Consider a

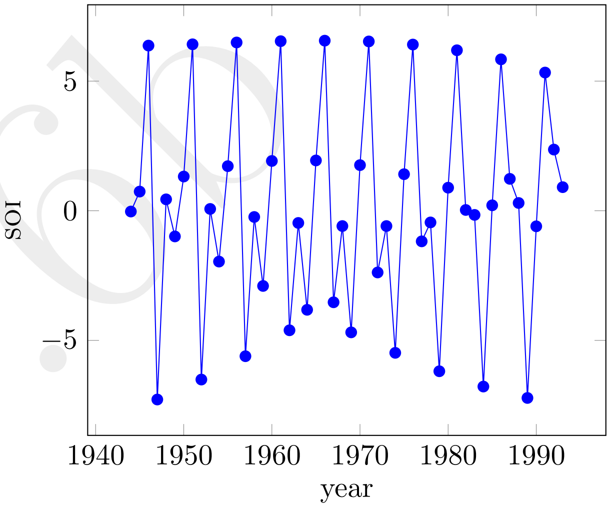


Figure 3.1: Yearly average SOI *over* 50 *years ('smoothed' somewhat for the purposes of the example). The nearly regular behaviour suggests that it should be predictable.*

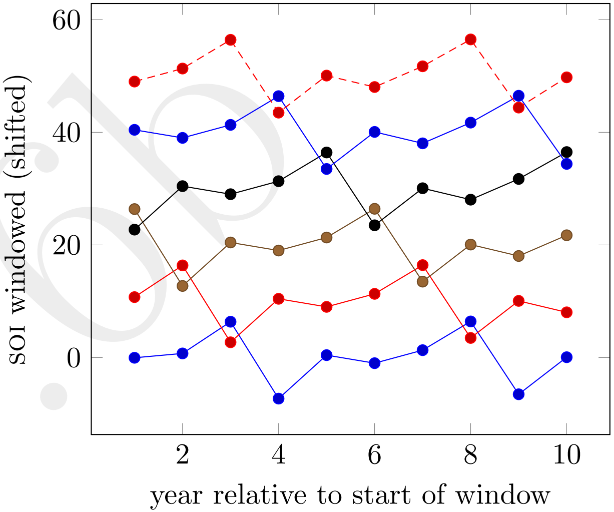


Figure 3.2: The first six windows of the SOI *data of Figure 3.1—displaced vertically for clarity. Each window is of length ten years: lowest, the first window is data 1944–1953; second lowest, covers 1945–1954; third lowest, covers 1946–1955; and so on to the 41st window which is data 1984–1993, not shown.*

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 10×41. 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 10×41 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 'nonzero' 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 that 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.

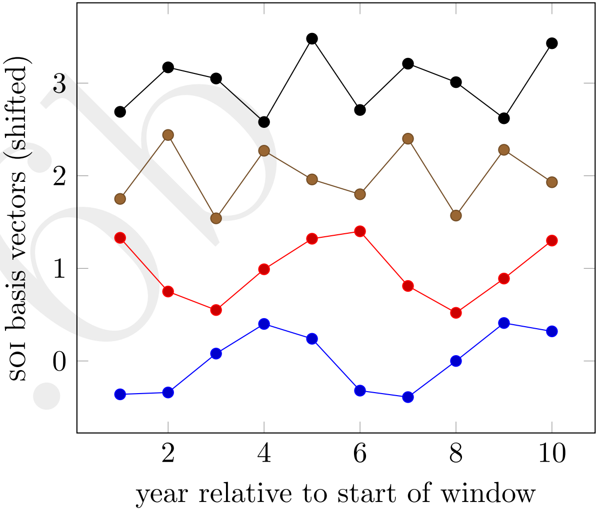


Figure 3.3: The first four singular vectors of the SOI *data—displaced vertically for clarity. The bottom two form a pair to show a five-year cycle. The top two are a pair that show a two–three-year cycle. The combination of these two cycles leads to the structure of the* SOI *in Figure 3.1.*

Theorem 3.4.28 *For every given subspace, any two orthonormal bases have the same number of vectors.*

An existential issue How do we know that every subspace has an orthonormal basis? We know that 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.

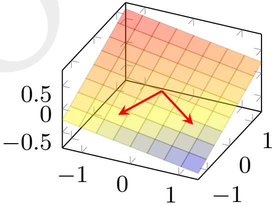
Theorem 3.4.29 (existence of basis) *Let $\mathbb{W}$ be a subspace of $\mathbb{R}^n$, then there exists an orthonormal basis for $\mathbb{W}$.*

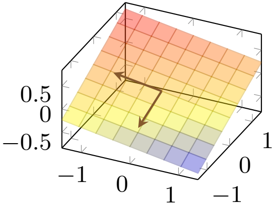
Ensemble simulation makes better weather forecasts Near the end of the twentieth century weather forecasts were becoming amazingly good at predicting the chaotic weather days in advance. However, there were notable failures: occasionally the weather forecast would give no hint of storms that developed (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 magnitude of the necessary changes. Then the storm would be within the possibilities. But the storm would not be forecast if the measurements were, by chance error, the 'other side' of the tipping point (as happened in 1999). 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 $\mathbb{R}^{1,000,000,000}$, 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. To avoid such degeneracy, the meteorologists continuously 'renormalize' 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 that unusual storms are retained in the range of possibilities explored by the ensemble forecast, and hence make weather forecasting much more complete.

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. For example, in a *plane* any two orthonormal basis vectors of the plane could form the two 'standard unit vectors' of a coordinate system in that plane—as suggested by the illustration. Example 3.4.25 found that 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 $\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$ where $\boldsymbol{u}_1 = (-\frac{2}{3}, \frac{1}{3}, \frac{2}{3})$, and $\boldsymbol{u}_2 = (\frac{2}{3}, \frac{2}{3}, \frac{1}{3})$. With two vectors in the basis, the plane is two-dimensional.
- Subspace $\mathbb{W} = \operatorname{span}\{(5, 1, -1/2)\,, (0, -3, -1)\,, (-4, 1, 1)\}$ of Example 3.4.25 is found to have an orthonormal basis of

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 3.4.4(g)) has an orthonormal basis of the n standard unit vectors, $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$, then $\dim \mathbb{R}^n = n$.
- The El Niño 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 $A = USV^{\mathrm{T}}$ and setting $r = \operatorname{rank} A$, an orthonormal basis for the column space is the first r columns of U, and that for the row space is the first r 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) 2 (b) 0 (c) 1 (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* $\operatorname{nullity}(A)$.

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, 2×4, has nullspace written as $\operatorname{span}\{(-2, 1, 0, 0$ $(-\frac{15}{7}, 0, \frac{9}{7}, 1)\}$. 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 nonzero 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 nonzero columns in the diagonal matrix S of an SVD. Put these two facts together in general and we get the following theorem that helps characterize solutions of linear equations.

Theorem 3.4.39 (rank theorem) *For every $m \times n$ matrix A,* $\operatorname{rank} A + \operatorname{nullity} A = n$, *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

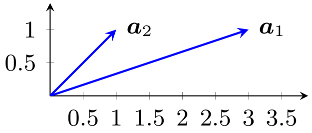
$$\begin{bmatrix} -2 & 1 & 4 & 0 & -4 \\ -1 & 1 & 0 & -2 & 0 \\ -3 & 1 & 3 & 2 & -3 \\ 0 & 0 & 1 & 0 & -1 \end{bmatrix} \text{ has singular values } \begin{array}{l} \texttt{8.1975} \\ \texttt{2.6561} \\ \texttt{1.6572} \\ \texttt{0.0000} \end{array}$$

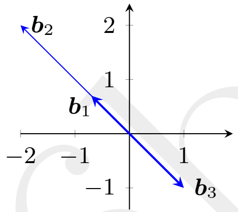
computed with `svd()`. What is its nullity?

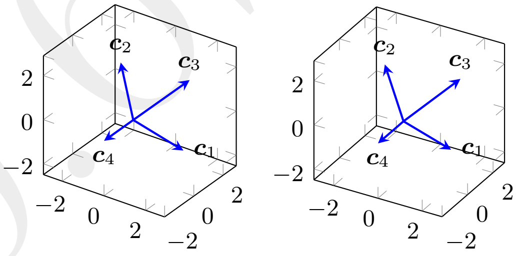
(a) 2 (b) 1 (c) 0 (d) 3

Example 3.4.42 Each of the following graphs plots all the column vectors of a matrix. What is the nullity of each of the matrices? Give reasons.

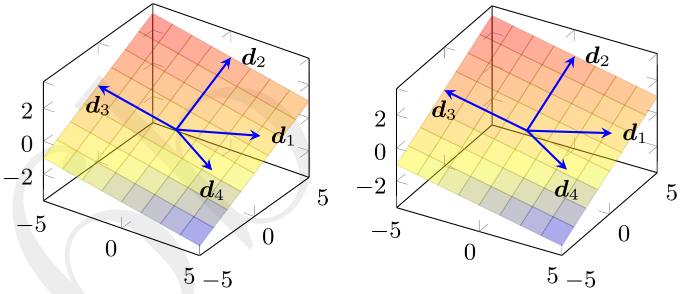
(a)



(b) 

(c) 

a stereo pair.



(d) a stereo pair.

□

Recall the list of exact properties that ensure a system of linear equations has a unique solution, Theorem 3.3.27. The recognition of these new concepts associated with matrices and linear equations lead us to extend this list.

Theorem 3.4.43 (Unique Solutions: version 2) *For every $n\times n$ square matrix A, and extending Theorem 3.3.27, the following statements are equivalent:*

(a) *A is invertible;*

(b) $A\boldsymbol{x} = \boldsymbol{b}$ *has a* ***unique solution*** *for every* $\boldsymbol{b} \in \mathbb{R}^n$*;*

(c) $A\boldsymbol{x} = \boldsymbol{0}$ *has only the zero solution;*

(d) *all* n *singular values of* A *are nonzero;*

(e) *the condition number of* A *is finite* (`rcond` > 0)*;*

(f) $\operatorname{rank} A = n$ *;*

(g) $\operatorname{nullity} A = 0$ *;*

(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

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 huge 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.

As well as being fundamental to engineering, scientific, and computational inference, approximately solving inconsistent equations also introduces the linear transformation of "projection".

3.5.1 Make a minimal change to the problem

Example 3.5.1 *(rationalize 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 kilograms (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 $x \approx (84.8 + 84.1 + 84.7 + 84.4)/4 = 84.5\,\text{kg}$. Let's see this same answer from a new linear algebra justification.

In the linear algebra, view my weight x as an unknown. 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

$$A x = \boldsymbol{b}\,, \quad \text{namely } \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} x = \begin{bmatrix} 84.8 \\ 84.1 \\ 84.7 \\ 84.4 \end{bmatrix}.$$

The linear algebra Procedure 3.3.15 is to 'solve' this system, despite its contradictions, via an SVD and some intermediaries:

$$A x = U \underbrace{S \overbrace{V^{\mathsf{T}} x}^{=y}}_{=\boldsymbol{z}} = \boldsymbol{b}\,.$$

(a) We are given that this particular matrix A of a column of ones has an SVD of

$$A = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac12 & \frac12 & \frac12 & \frac12 \\ \frac12 & \frac12 & -\frac12 & -\frac12 \\ \frac12 & -\frac12 & -\frac12 & \frac12 \\ \frac12 & -\frac12 & \frac12 & -\frac12 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \end{bmatrix}^{\mathsf{T}} = U S V^{\mathsf{T}}$$

(perhaps check the columns of U are orthonormal).

(b) Solve $U\boldsymbol{z} = \boldsymbol{b}$ by computing

$$\boldsymbol{z} = U^{\mathrm{T}}\boldsymbol{b} = \begin{bmatrix} \frac12 & \frac12 & \frac12 & \frac12 \\ \frac12 & \frac12 & -\frac12 & -\frac12 \\ \frac12 & -\frac12 & -\frac12 & \frac12 \\ \frac12 & -\frac12 & \frac12 & -\frac12 \end{bmatrix} \begin{bmatrix} 84.8 \\ 84.1 \\ 84.7 \\ 84.4 \end{bmatrix} = \begin{bmatrix} 169 \\ -0.1 \\ 0.2 \\ 0.5 \end{bmatrix}.$$

(c) Now try to solve $Sy = \boldsymbol{z}$, that is,

$$\begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \end{bmatrix} y = \begin{bmatrix} 169 \\ -0.1 \\ 0.2 \\ 0.5 \end{bmatrix}.$$

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 $\boldsymbol{z} = (169, -0.1, 0.2, 0.5)$ so that we can report a solution to a slightly different problem? Answer: we *have to* adjust the last three components to zero. Moreover, any adjustment to the first component is not needed, it would make the change to $\boldsymbol{z}$ bigger than necessary, and so we do not adjust the first component. Hence we solve a slightly different problem, that of

$$\begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \end{bmatrix} y = \begin{bmatrix} 169 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

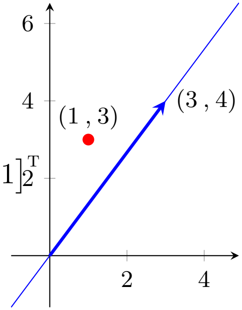
with solution $y = 84.5$. *Let's treat this exact solution to a slightly different problem as an* approximate *solution to the original problem.*

(d) Lastly, solve $V^{\mathrm{T}}x = y$ by computing $x = Vy = 1y = y = 84.5\,\text{kg}$ (upon including the physical units). That is, this linear algebra procedure gives my weight as $x = 84.5\,\text{kg}$ (approximately).

This linear algebra procedure recovers the traditional answer of averaging measurements. □

The methodology of 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 $\boldsymbol{z}$ is equivalent to making the smallest possible change to the original data vector $\boldsymbol{b}$. The reason is that $\boldsymbol{b} = U\boldsymbol{z}$ 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 $\boldsymbol{b}$ is precisely the same magnitude as the smallest possible change to $\boldsymbol{z}$. 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 (illustrated to the right)

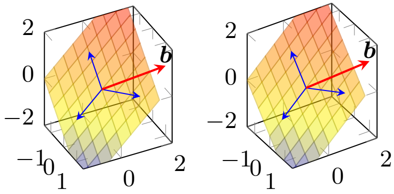


$$\begin{bmatrix}3\\4\end{bmatrix}x = \begin{bmatrix}1\\3\end{bmatrix}, \quad \text{and given} \quad \begin{bmatrix}3\\4\end{bmatrix} = \begin{bmatrix}\frac{3}{5} & \frac{4}{5}\\ \frac{4}{5} & -\frac{3}{5}\end{bmatrix}\begin{bmatrix}5\\0\end{bmatrix}\begin{bmatrix}1\end{bmatrix}^{\text{T}}$$

is an SVD factorization of the 2×1 matrix. Following the procedure of the previous Example 3.5.1, what is the 'best' approximate solution to these inconsistent equations?

(a) $x = 4/7$ (b) $x = 3/4$ (c) $x = 3/5$ (d) $x = 1/3$

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 $\boldsymbol{b}$ is not in the column space of the matrix A (Definition 3.4.10): the stereo picture to the right illustrates the 2D column space spanned by the three columns of A and that the vector $\boldsymbol{b}$, of the results, lies outside the column space.

Now reconsider Step 3 in Example 3.3.13.

(a) We need to interpret and 'solve' $S\boldsymbol{y} = \boldsymbol{z}$ which here is

$$\begin{bmatrix} 1.7321 & 0 & 0 \\ 0 & 1.7321 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{y} = \begin{bmatrix} -2.0412 \\ -2.1213 \\ 0.5774 \end{bmatrix}.$$

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 $\boldsymbol{z} = (-2.0412, -2.1213, 0.5774)$ so that we can report a solution, albeit to a slightly different problem? Answer: we *must* change the last component of $\boldsymbol{z}$ to zero.

Moreover, any change to the first two components is not needed, it would make the change bigger than necessary, and so we do not change the first two components. Hence we find an approximate solution to the player ratings via solving

$$\begin{bmatrix} 1.7321 & 0 & 0 \\ 0 & 1.7321 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{y} = \begin{bmatrix} -2.0412 \\ -2.1213 \\ 0 \end{bmatrix}.$$

Here, via `y=z(1:2)./diag(S(1:2,1:2))`, a general solution is that vector $\boldsymbol{y} = (-1.1785, -1.2247, y_3)$. Varying the free variable y_3 gives equally good approximate solutions.

(b) Lastly, solve $V^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{y}$, via computing `x=V(:,1:2)*y`, to determine

$$\begin{aligned} \boldsymbol{x} = V\boldsymbol{y} &= \begin{bmatrix} 0.0000 & -0.8165 & 0.5774 \\ -0.7071 & 0.4082 & 0.5774 \\ 0.7071 & 0.4082 & 0.5774 \end{bmatrix} \begin{bmatrix} -1.1785 \\ -1.2247 \\ y_3 \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ \frac{1}{3} \\ -\frac{4}{3} \end{bmatrix} + \frac{y_3}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}. \end{aligned}$$

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 $\boldsymbol{x} = (1, \frac{1}{3}, -\frac{4}{3})$ 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 such a 'best' approximate solution.

Procedure 3.5.4 (approximate solution) *Obtain the so-called 'least square' approximate solution(s) of inconsistent equations $A\boldsymbol{x} = \boldsymbol{b}$ using an* SVD *and via intermediate unknowns:*

1. *factorize $A = USV^{\mathrm{T}}$ and set $r = \operatorname{rank} A$ (remembering that relatively small singular values are effectively zero);*
2. *solve $U\boldsymbol{z} = \boldsymbol{b}$ by $\boldsymbol{z} = U^{\mathrm{T}}\boldsymbol{b}$;*
3. *disregard the inconsistent equations for $i = r+1, \ldots, m$ as errors, set $y_i = z_i/\sigma_i$ for $i = 1, \ldots, r$ (as these $\sigma_i > 0$), and otherwise y_i is free for $i = r+1, \ldots, n$;*

4. *solve* $V^{\mathsf{T}}\boldsymbol{x} = \boldsymbol{y}$ *to obtain a general approximate solution as* $\boldsymbol{x} = V\boldsymbol{y}$.

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, 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 summarizes 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

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	—

equal the difference in the scores when they play. □

When rating players or teams based upon results, be clear of the purpose. For example, is the purpose to summarize 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

---

Be aware of Kenneth Arrow's Impossibility Theorem (Arrow 1950)—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 would also be open to criticism by also violating at least one of the criteria.

favourite sport.

Activity 3.5.7 Listed below are four approximate solutions to the system $A\boldsymbol{x} = \boldsymbol{b}$,

$$\begin{bmatrix} 5 & 3 \\ 3 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 9 \\ 2 \\ 10 \end{bmatrix}.$$

Setting vector $\boldsymbol{b}' = A\boldsymbol{x}$ for each, which one minimizes the distance between the original right-hand side $\boldsymbol{b} = (9\,,2\,,10)$ and the approximate $\boldsymbol{b}'$?

(a) $\boldsymbol{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ (b) $\boldsymbol{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ (c) $\boldsymbol{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ (d) $\boldsymbol{x} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$

Theorem 3.5.8 (smallest change) *All approximate solutions obtained by Procedure 3.5.4 solve the linear system $A\boldsymbol{x} = \boldsymbol{b}'$ for the unique consistent right-hand side vector $\boldsymbol{b}'$ that minimizes the distance $|\boldsymbol{b}' - \boldsymbol{b}|$.*

(The dash on $\boldsymbol{b}'$ is to suggest an approximation to $\boldsymbol{b}$.)

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

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 that it gives the same solution as `A\b` in Matlab/Octave.

□

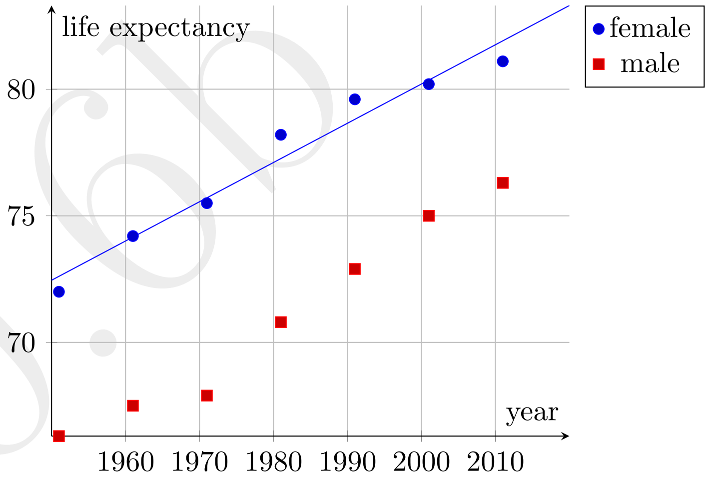


Figure 3.4: The life expectancies in years of females and males born in the given years (Table 3.5). Also plotted is the best straight line fit to the female data obtained by Example 3.5.9.

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) $f_i = x_1 + x_2 v_i$

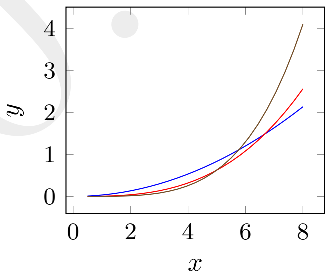
(b) $f_i = x_1$

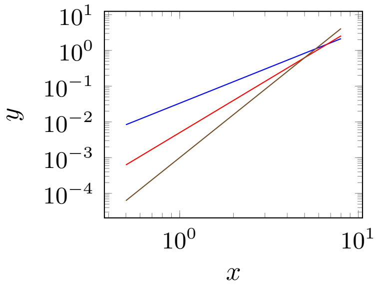
(c) $v_i = x_1 + x_2 f_i$

(d) $v_i = x_1 + x_2 f_i + x_3 f_i^2$

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 best straight line fit in Figure 3.5. Use the data for the planets from Mercury to Uranus to infer the

Power laws and the log-log plot Hundreds of power laws have been identified in engineering, physics, biology, and the social sciences. These laws are 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 a plot of 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, $y \propto x^2$.

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.
[https://en.wikipedia.org/wiki/Orbital_period]

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

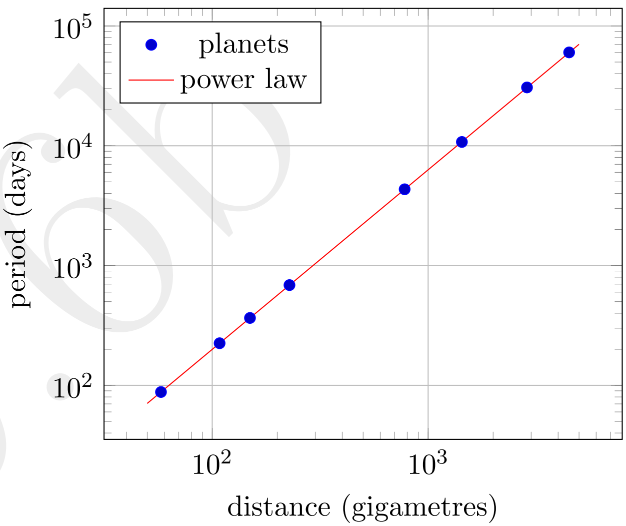


Figure 3.5: The planetary periods as a function of the distance, from the data of Table 3.6: the graph is a log-log plot to show the excellent power law. Also plotted is the power law fit computed by Example 3.5.11.

law with an SVD, confirm that it gives the same solution as `A\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 $A\boldsymbol{x} = \boldsymbol{b}$ by solving the associated normal equation $(A^{\mathrm{T}}A)\boldsymbol{x} = (A^{\mathrm{T}}\boldsymbol{b})$. For *theoretical purposes* this normal equation is very useful. However, in practical computation avoid the normal equation because forming $A^{\mathrm{T}}A$, and then manipulating it, is both expensive and error-enhancing (especially in large problems). For example, $\operatorname{cond}(A^{\mathrm{T}}A) = (\operatorname{cond} A)^2$ (**??**) so matrix $A^{\mathrm{T}}A$ typically has a much worse condition number than matrix A (Procedure 2.2.5). To paraphrase Cleve Molar: Almost anything you can do with $A^{\mathrm{T}}A$ can be done without it [via the SVD].
- The last two examples observe that `A\b` gives an answer that was identical to what the SVD procedure gives. Thus `A\b` can serve as a very useful shortcut to finding a best

approximate solution. For non-square matrices with more rows than columns (more equations than variables), `A\b` generally does this (without comment as Matlab/Octave assumes you know what you are doing). For other scenarios `A\b` does something different, so be wary.

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\b` Examples 3.5.9 and 3.5.11 observe that `A\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\b` is not 'the approximate answer' that you want. Beware.

Example 3.5.12 Use `x=A\b` to 'solve' the problems of Examples 3.5.1, 3.5.3, and 3.5.6.

- With Octave, observe that 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 (R2013b), 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?

□

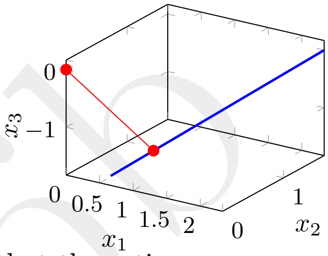
Recall that systems of linear equations may not have unique solutions (as in the rating examples): what does `A\b` compute when there are an infinite number of solutions?

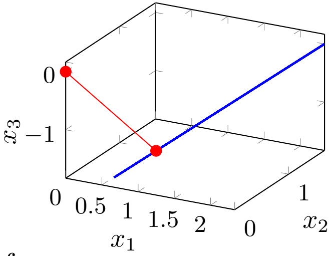
- For systems of equations with the number of equations not equal to the number of variables, $m \neq n$, the Octave operation `A\b` computes for you the *smallest solution* (that is, a solution that is of least magnitude, smallest norm) of all valid solutions (Theorem 3.5.13): often 'exact' when $m < n$, or approximate when $m > n$ (Theorem 3.5.8). Using `A\b` is the most efficient computationally, but using the SVD helps us understand what it does.
- Matlab (R2013b etc.) does something different with `A\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 Procedure 3.3.15 or Procedure 3.5.4, respectively, and then setting to zero the free variables, that is,* $y_{r+1} = \cdots = y_n = 0$.

Example 3.5.14 In the table tennis ratings of Example 3.5.3 the procedure found that the ratings were any of

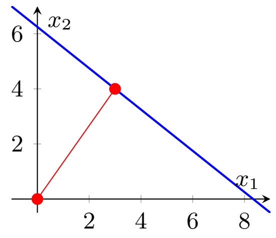




$$\boldsymbol{x} = \begin{bmatrix} 1 \\ \frac{1}{3} \\ -\frac{4}{3} \end{bmatrix} + \frac{y_3}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},$$

as illustrated in stereo to the right (blue). Verify that $|\boldsymbol{x}|$ 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 following SVD:

$$\begin{bmatrix} 2 & 3 & 6 \end{bmatrix} = \begin{bmatrix} 1 \end{bmatrix} \begin{bmatrix} 7 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{2}{7} & -\frac{3}{7} & -\frac{6}{7} \\ \frac{3}{7} & \frac{6}{7} & -\frac{2}{7} \\ \frac{6}{7} & -\frac{2}{7} & \frac{3}{7} \end{bmatrix}^{\mathsf T}.$$

(a) $(-12, -4, 6)$ (b) $(2, 3, 6)$ (c) $(-3, 6, -2)$ (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 'slices') of specific areas of a scanned object, allowing

Table 3.7: As well as the MATLAB*/Octave commands and operations listed in Tables 1.2, 2.3, and 3.1 to 3.3 we may invoke these functions for drawing images—functions which are otherwise not needed.*

- `reshape(A,p,q)` *for an* $m \times n$ *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 greyscale; this command changes the current figure to 64 shades of grey.* *(*`colormap(jet)` *is the default,* `colormap(hot)` *is good for both colour and greyscale reproductions,* `colormap('list')` *lists the available colormaps you can try.)*
- `imagesc(A)` *where* A *is an* $m \times n$ *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., greyscale).*
- `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*

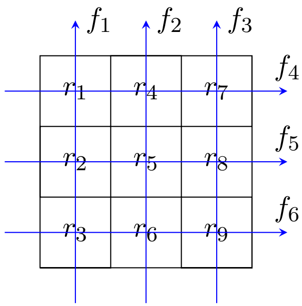
> the user to see inside the object without cutting.
>
> *Wikipedia, 2015*

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, as real-life examples have millions of unknowns and equations. Suppose we divide a cross-section of a body into nine squares (large pixels) in a 3×3 grid. Inside each square the body's material has some unknown density represented by transmission factors, $r_1, r_2, \ldots, r_9$, as shown to the right. 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 jth square: typically, smaller r_i corresponds to higher density in the body.

As indicated next to the right, six X-ray measurements are made through the body where f_1 , f_2 , . . . , 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:



$$r_1 r_2 r_3 = f_1 , \quad r_4 r_5 r_6 = f_2 , \quad r_7 r_8 r_9 = f_3 ,$$
$$r_1 r_4 r_7 = f_4 , \quad r_2 r_5 r_8 = f_5 , \quad r_3 r_6 r_9 = f_6 .$$

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 (computers almost always use “log” to denote the natural logarithm, so we do too. Herein, unsubscripted “log” means the same as “ln”):

$$r_i r_j r_k = f_l \iff (\log r_i) + (\log r_j) + (\log r_k) = (\log f_l).$$

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:

$$\begin{aligned} &x_1 + x_2 + x_3 = b_1\,, \quad && x_4 + x_5 + x_6 = b_2\,, \quad && x_7 + x_8 + x_9 = b_3\,, \\ &x_1 + x_4 + x_7 = b_4\,, \quad && x_2 + x_5 + x_8 = b_5\,, \quad && x_3 + x_6 + x_9 = b_6\,. \end{aligned}$$

This forms the matrix-vector system $A\boldsymbol{x} = \boldsymbol{b}$ for 6×9 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 $\boldsymbol{b} = (-0.91\,, -1.04\,, -1.54\,, -1.52\,, -1.43\,, -0.53)$ (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.)

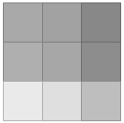
$$\boldsymbol{x} = (-.42, -.39, -.09, -.47, -.44, -.14, -.63, -.60, -.30).$$

These are logarithms so to get the corresponding physical transmission factors compute the exponential of each component, denoted as $\exp(\boldsymbol{x})$,

$$\boldsymbol{r} = \exp(\boldsymbol{x}) = (.66, .68, .91, .63, .65, .87, .53, .55, .74),$$

although it is perhaps more appealing to put these factors into the shape of the 3×3 array of pixels as in (and as illustrated to the right)

$$\begin{bmatrix} r_1 & r_4 & r_7 \\ r_2 & r_5 & r_8 \\ r_3 & r_6 & r_9 \end{bmatrix} = \begin{bmatrix} 0.66 & 0.63 & 0.53 \\ 0.68 & 0.65 & 0.55 \\ 0.91 & 0.87 & 0.74 \end{bmatrix}.$$

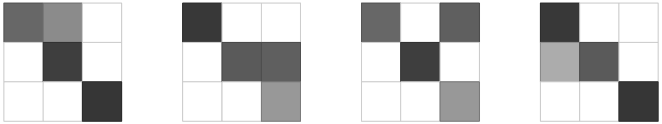


Octave's answer predicts that there is less transmitting, more absorbing, denser, material to the top-right; and more transmitting, less absorbing, less dense, material to the bottom-left.

- However, the answer from MATLAB's `A\b` is (2 d.p.)

$$\boldsymbol{x} = (-0.91, 0, 0, -0.61, -1.43, 1.01, 0, 0, -1.54),$$

as illustrated below in the leftmost picture. This is quite a different answer!



Furthermore, Matlab could give other 'answers', as illustrated in the other pictures above. Reordering the rows in the matrix A and right-hand side $\boldsymbol{b}$ does not change the system of equations. But after such reordering the answer from Matlab's `x=A\b` variously predicts each of the above four pictures.

The reason for such a 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 extra information that we desire the answer that is the 'greyest', the most 'washed out', the answer with fewest features. Finding the answer $\boldsymbol{x}$ that minimizes $|\boldsymbol{x}|$ 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 $A\boldsymbol{x} = \boldsymbol{b}$.

(a) First, find an SVD, $A = USV^{\mathrm{T}}$, 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
     0     0  1.73     0     0     0     0     0     0
     0     0     0  1.73     0     0     0     0     0
     0     0     0     0  1.73     0     0     0     0
     0     0     0     0     0  0.00     0     0     0
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
```

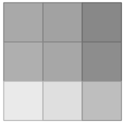
(b) Solve $U\boldsymbol{z} = \boldsymbol{b}$ by `z=U'*b` to find

$$\boldsymbol{z} = (2.85, -0.52, 0.31, 0.05, -0.67, -0.00).$$

(c) Because the sixth singular value is zero, ignore the sixth equation: because $z_6 = 0.00$ (2 d.p.), this is only a small inconsistency error. Now set $y_i = z_i/\sigma_i$ for $i = 1, \ldots, 5$ 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 nonzero values via `y=z(1:5)./diag(S(1:5,1:5))` to find

$$\boldsymbol{y} = (1.16, -0.30, 0.18, 0.03, -0.39, 0, 0, 0, 0)$$

(d) Then, via `x=V(:,1:5)*y`, solve $V^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{y}$ to determine the smallest solution is $\boldsymbol{x} = (-0.42\,, -0.39\,, -0.09\,, -0.47\,, -0.44\,, -0.14\,, -0.63\,, -0.60\,, -0.30)$. This is the same answer as computed by Octave's `A\b` to give the pixel image shown that has minimal artifices.



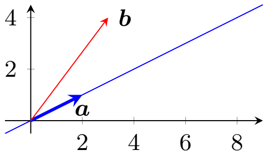
In practice, *each* slice of a real CT-scan would involve finding the absorption of tens of millions of pixels. That is, a CT-scan needs to best solve many systems of tens of millions of equations in tens of millions of unknowns! □

3.5.3 Orthogonal projection resolves vector components

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 that forms a right-angle; that is, it forms an orthogonal vector.

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.

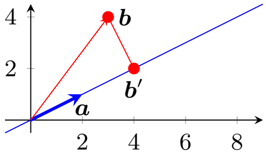
Project onto a direction



Example 3.5.18 Consider 'solving' the inconsistent system $\boldsymbol{a}x = \boldsymbol{b}$ where $\boldsymbol{a} = (2, 1)$ and $\boldsymbol{b} = (3, 4)$; that is, solve

$$\begin{bmatrix} 2 \\ 1 \end{bmatrix} x = \begin{bmatrix} 3 \\ 4 \end{bmatrix}.$$

As illustrated to the right, the impossible task is to find some multiple of the vector $\boldsymbol{a} = (2, 1)$ (all multiples plotted) that equals $\boldsymbol{b} = (3, 4)$. It cannot be done. Question: how may we change the right-hand side vector $\boldsymbol{b}$ so that the task is possible? A partial answer is to replace $\boldsymbol{b}$ by some vector $\boldsymbol{b}'$ which is in the column space of matrix $A = \begin{bmatrix}\boldsymbol{a}\end{bmatrix}$. But we could choose any $\boldsymbol{b}'$ in the column space, so any answer for the multiple x would be possible! Surely any answer is not acceptable.



Instead, often the preferred answer is, out of all vectors in the column space of matrix $A = \begin{bmatrix}\boldsymbol{a}\end{bmatrix}$, find the vector $\boldsymbol{b}'$ in the column space *which is closest to* $\boldsymbol{b}$—as illustrated to the right here, where it looks like $\boldsymbol{b}' = (4,2)$.

The SVD approach of Procedure 3.5.4 to find $\boldsymbol{b}'$ and x is the following.

(a) Use `[U,S,V]=svd([2;1])` to find here the SVD factorization

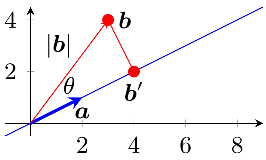
$$A = USV^{\mathrm{T}} = \begin{bmatrix} 0.89 & -0.45 \\ 0.45 & 0.89 \end{bmatrix} \begin{bmatrix} 2.24 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \end{bmatrix}^{\mathrm{T}} \quad \text{(2 d.p.)}.$$

(b) Then $\boldsymbol{z} = U^{\mathrm{T}}\boldsymbol{b} = (4.47, 2.24)$.

(c) Treat the second component of $Sy = \boldsymbol{z}$ as an error—it is the magnitude $|\boldsymbol{b} - \boldsymbol{b}'|$—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 $\boldsymbol{b}' = \boldsymbol{a}x = (2,1)2 = (4,2)$, as is recognizable in the graphs. □

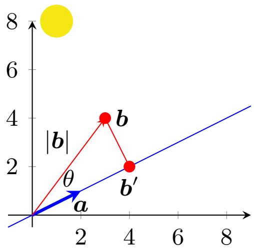


Now let's derive the same result but with two differences: firstly, use more elementary arguments, not the SVD, and secondly, derive the result for general vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ (although continuing to use the same illustration). Start with the crucial observation that the closest point/vector $\boldsymbol{b}'$ in the column space of $A = \begin{bmatrix}\boldsymbol{a}\end{bmatrix}$ is such that $\boldsymbol{b} - \boldsymbol{b}'$ is at right-angles, orthogonal, to $\boldsymbol{a}$. (If $\boldsymbol{b} - \boldsymbol{b}'$ were not orthogonal, then we would be able to slide $\boldsymbol{b}'$ along the line $\operatorname{span}\{\boldsymbol{a}\}$ to reduce the length of $\boldsymbol{b} - \boldsymbol{b}'$.) Thus we form a right-angle triangle with hypotenuse of length $|\boldsymbol{b}|$ and angle θ, as shown above-right. Trigonometry then gives the adjacent length $|\boldsymbol{b}'| = |\boldsymbol{b}|\cos\theta$. But the angle θ is that between the given vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, so the dot product gives the cosine as $\cos\theta = \boldsymbol{a}\cdot\boldsymbol{b}/(|\boldsymbol{a}||\boldsymbol{b}|)$ (Theorem 1.3.5). Hence the adjacent length

$$|\boldsymbol{b}'| = |\boldsymbol{b}|\cos\theta = |\boldsymbol{b}|\frac{\boldsymbol{a}\cdot\boldsymbol{b}}{|\boldsymbol{a}||\boldsymbol{b}|} = \frac{\boldsymbol{a}\cdot\boldsymbol{b}}{|\boldsymbol{a}|}\,.$$

To approximately solve $\boldsymbol{a}x = \boldsymbol{b}$, replace the inconsistent $\boldsymbol{a}x = \boldsymbol{b}$ by the consistent $\boldsymbol{a}x = \boldsymbol{b}'$. Then as x is a scalar, we solve this consistent equation via the ratio of lengths, $x = |\boldsymbol{b}'|/|\boldsymbol{a}| = \boldsymbol{a}\cdot\boldsymbol{b}/|\boldsymbol{a}|^2$. For Example 3.5.18, this gives the '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 $\boldsymbol{b}' = \boldsymbol{a}x = \boldsymbol{a}(\boldsymbol{a}\cdot\boldsymbol{b})/|\boldsymbol{a}|^2$. Geometrically the formula gives the 'shadow' $\boldsymbol{b}'$ of vector $\boldsymbol{b}$ when projected by a 'sun' high above the line of the vector $\boldsymbol{a}$, as illustrated schematically to the right. As such, the formula is called an orthogonal projection.



Definition 3.5.19 (orthogonal projection onto 1D) *Let $\boldsymbol{u},\boldsymbol{v}\in\mathbb{R}^n$ and vector $\boldsymbol{u}\neq\boldsymbol{0}$, then the* ***orthogonal projection*** *of $\boldsymbol{v}$ onto $\boldsymbol{u}$ is*

$$\operatorname{proj}_{\boldsymbol{u}}(\boldsymbol{v}) := \boldsymbol{u}\frac{\boldsymbol{u}\cdot\boldsymbol{v}}{|\boldsymbol{u}|^2}. \tag{3.5a}$$

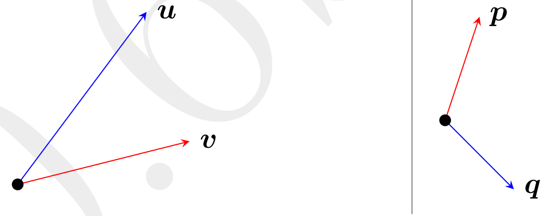
In the special, but common, case when $\boldsymbol{u}$ is a unit vector,

$$\operatorname{proj}_{\boldsymbol{u}}(\boldsymbol{v}) := \boldsymbol{u}(\boldsymbol{u}\cdot\boldsymbol{v}). \tag{3.5b}$$

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) $\operatorname{proj}_{\boldsymbol{u}}(\boldsymbol{v})$ and $\boldsymbol{u}x = \boldsymbol{v}$

(b) $\operatorname{proj}_{\boldsymbol{q}}(\boldsymbol{p})$ and $\boldsymbol{q}x = \boldsymbol{p}$



□

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 $\operatorname{proj}_{\boldsymbol{u}}(\boldsymbol{v})$ for vectors $\boldsymbol{u} = (3, 4)$ and $\boldsymbol{v} = (4, 1)$, and hence best solve $\boldsymbol{u}x = \boldsymbol{v}$.

(b) Find $\operatorname{proj}_{\boldsymbol{s}}(\boldsymbol{r})$ for vectors $\boldsymbol{r} = (1, 3)$ and $\boldsymbol{s} = (2, -2)$, and hence best solve $\boldsymbol{s}x = \boldsymbol{r}$.

(c) Find $\operatorname{proj}_{\boldsymbol{p}}(\boldsymbol{q})$ for vectors $\boldsymbol{p} = (\frac{1}{3}, \frac{2}{3}, \frac{2}{3})$ and $\boldsymbol{q} = (3, 2, 1)$, and best solve $\boldsymbol{p}x = \boldsymbol{q}$.

$\square$

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) $x = 21/4$ (b) $x = 10/13$ (c) $x = 4/9$ (d) $x = 4$

Project onto a subspace

The previous subsection develops a geometric view of the 'best' solution to the inconsistent system $\boldsymbol{a}x = \boldsymbol{b}$. The discussion introduced that the conventional 'best' solution—that determined by Procedure 3.5.4—is to replace $\boldsymbol{b}$ by its projection $\operatorname{proj}_{\boldsymbol{a}}(\boldsymbol{b})$, namely to solve $\boldsymbol{a}x = \operatorname{proj}_{\boldsymbol{a}}(\boldsymbol{b})$. The rationale is that this is the *smallest* change to the right-hand side $\boldsymbol{b}$ that enables the equation to be solved. This subsection introduces that solving inconsistent equations in more variables may be viewed as an analogous projection onto a subspace.

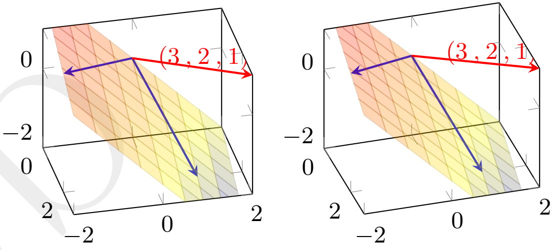
Definition 3.5.23 (project onto a subspace) *Let $\mathbb{W}$ be a k-dimensional subspace of $\mathbb{R}^n$ with an orthonormal basis $\{\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k\}$. For every vector $\boldsymbol{v} \in \mathbb{R}^n$, the* ***orthogonal projection*** *of vector $\boldsymbol{v}$ onto subspace $\mathbb{W}$ is*

$$\operatorname{proj}_{\mathbb{W}}(\boldsymbol{v}) = \boldsymbol{w}_1(\boldsymbol{w}_1 \cdot \boldsymbol{v}) + \boldsymbol{w}_2(\boldsymbol{w}_2 \cdot \boldsymbol{v}) + \cdots + \boldsymbol{w}_k(\boldsymbol{w}_k \cdot \boldsymbol{v}).$$

Example 3.5.24

(a) Let $\mathbb{X}$ be the xy-plane in xyz-space, find $\operatorname{proj}_{\mathbb{X}}(3, -4, 2)$.

(b) For the subspace $\mathbb{W} = \operatorname{span}\{(2,-2,1),(2,1,-2)\}$, determine $\operatorname{proj}_{\mathbb{W}}(3,2,1)$ (illustrated to the right).



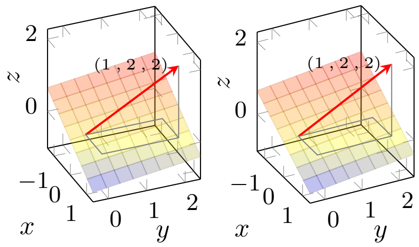
(c) Recall the table tennis ranking Examples 3.5.3 and 3.3.13. To rank the players we seek to solve the matrix-vector system, $A\boldsymbol{x} = \boldsymbol{b}$,

$$\begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}.$$

Letting $\mathbb{A}$ denote the column space of matrix A, determine $\operatorname{proj}_{\mathbb{A}}(\boldsymbol{b})$.

(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 to the right).



□

Activity 3.5.25 Determine which of the following is $\operatorname{proj}_{\mathbb{W}}(1\,,1\,,-2)$ for the subspace $\mathbb{W} = \operatorname{span}\{(2\,,3\,,6)\,,(-3\,,6\,,-2)\}$.

(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 3.5.24(c) determines that the orthogonal projection of the given table tennis results $\boldsymbol{b} = (1, 2, 2)$ onto the column space of matrix A is the vector $\boldsymbol{b}' = \frac{1}{3}(2, 7, 5)$. Recall that Example 3.5.3 invokes Procedure 3.5.4 to find the 'approximate' solution of the impossible $A\boldsymbol{x} = \boldsymbol{b}$ to be $\boldsymbol{x} = (1, \frac{1}{3}, -\frac{4}{3})$. Now see that $A\boldsymbol{x} = \big(1 - \frac{1}{3}, 1 - (-\frac{4}{3}), \frac{1}{3} - (-\frac{4}{3})\big) = (\frac{2}{3}, \frac{7}{3}, \frac{5}{3}) = \boldsymbol{b}'$. That is, the approximate solution method of Procedure 3.5.4 solved the problem $A\boldsymbol{x} = \operatorname{proj}_{\mathbb{A}}(\boldsymbol{b})$. The following Theorem 3.5.26 confirms that 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 the smallest change to the right-hand side that makes it consistent.

Theorem 3.5.26 *The 'least square' solution/s of the system $A\boldsymbol{x} = \boldsymbol{b}$ determined by Procedure 3.5.4 is/are the solution/s of $A\boldsymbol{x} = \operatorname{proj}_{\mathbb{A}}(\boldsymbol{b})$ where $\mathbb{A}$ denotes the column space of A.*

Example 3.5.27 Recall that Example 3.5.1 rationalizes 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

$$A x = \boldsymbol{b}\,, \quad \text{namely } \begin{bmatrix}1\\1\\1\\1\end{bmatrix} x = \begin{bmatrix}84.8\\84.1\\84.7\\84.4\end{bmatrix}.$$

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 that matrix A has one column $\boldsymbol{a}_1 = (1,1,1,1)$ so by Definition 3.5.19 the orthogonal projection

$$\begin{aligned}
&\operatorname{proj}_{\operatorname{span}\{\boldsymbol{a}_1\}}(84.8, 84.1, 84.7, 84.4)\\
&= \boldsymbol{a}_1 \frac{\boldsymbol{a}_1 \cdot (84.8, 84.1, 84.7, 84.4)}{|\boldsymbol{a}_1|^2}\\
&= \boldsymbol{a}_1 \frac{84.8 + 84.1 + 84.7 + 84.4}{1 + 1 + 1 + 1}\\
&= \boldsymbol{a}_1 \cdot 84.5
\end{aligned}$$

$$= (84.5, 84.5, 84.5, 84.5).$$

The projected system $Ax = (84.5, 84.5, 84.5, 84.5)$ is now consistent. Its solution is $x = 84.5\,\text{kg}$. 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 among 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 $A\boldsymbol{x} = \boldsymbol{b}$ where

$$A = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ 3 \\ 1 \\ -2 \\ 4 \\ -1 \end{bmatrix}.$$

Let's see that the orthogonal projection of $\boldsymbol{b}$ 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.)

```
U =
  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
 -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
S =
  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
V = ...
```

As there are three nonzero singular values in `S`, the first three columns of `U` are an orthonormal basis for the column space $\mathbb{A}$.

Letting $\boldsymbol{u}_j$ denote the columns of `U`, Definition 3.5.23 gives the orthogonal projection (2 d.p.)

$$\begin{aligned}\operatorname{proj}_{\mathbb{A}}(\boldsymbol{b}) &= \boldsymbol{u}_1(\boldsymbol{u}_1\cdot\boldsymbol{b}) + \boldsymbol{u}_2(\boldsymbol{u}_2\cdot\boldsymbol{b}) + \boldsymbol{u}_3(\boldsymbol{u}_3\cdot\boldsymbol{b})\\ &= -1.27\,\boldsymbol{u}_1 + 2.92\,\boldsymbol{u}_2 - 1.15\,\boldsymbol{u}_3\\ &= (-0.50, 1.75, 2.25, 0.75, 1.25, -1.00).\end{aligned}$$

Compute these three dot products in Matlab/Octave via `cs=U(:,1:3)` and then compute the linear combination with `projb=U(:,1:3)*cs`. To confirm that Procedure 3.5.4 solves $A\boldsymbol{x} = \operatorname{proj}_{\mathbb{A}}(\boldsymbol{b})$ we check that the ratings that were found by Example 3.5.6, $\boldsymbol{x} = (\frac{1}{2}, 1, -\frac{5}{4}, -\frac{1}{4})$, satisfy $A\boldsymbol{x} = \operatorname{proj}_{\mathbb{A}}(\boldsymbol{b})$: in Matlab/Octave compute `A*[0.50;1.00;-1` and see the product is $\operatorname{proj}_{\mathbb{A}}(\boldsymbol{b})$. □

Section 3.6 uses orthogonal projection as an example of a linear transformation. The section shows that a linear transformation always corresponds to multiplying by a specific matrix, which for orthogonal projection is here WW^{T}.

There is a useful feature of Example 3.5.24(e) and Example 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 `cs=W'*b`, that is, the vector $\boldsymbol{c}=W^{\mathrm{T}}\boldsymbol{b}$; and second the examples compute `proj=W*cs`, that is, $\operatorname{proj}_{\mathbb{W}}(\boldsymbol{b})=W\boldsymbol{c}$. Combining these two steps into one (using associativity) gives

$$\operatorname{proj}_{\mathbb{W}}(\boldsymbol{b})=W\boldsymbol{c}=W(W^{\mathrm{T}})\boldsymbol{b}=(WW^{\mathrm{T}})\boldsymbol{b}.$$

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 $\{\boldsymbol{w}_1,\boldsymbol{w}_2,\ldots,\boldsymbol{w}_k\}$, then for every vector $\boldsymbol{v}\in\mathbb{R}^n$, the orthogonal projection*

$$\operatorname{proj}_{\mathbb{W}}(\boldsymbol{v})=(WW^{\mathrm{T}})\boldsymbol{v} \tag{3.6}$$

for the $n\times k$ matrix $W=\begin{bmatrix}\boldsymbol{w}_1 & \boldsymbol{w}_2 & \cdots & \boldsymbol{w}_k\end{bmatrix}$.

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) $\operatorname{proj}_{\boldsymbol{u}}(\boldsymbol{v})$ for vector $\boldsymbol{u} = (3,4)$ and $\boldsymbol{v} = (4,1)$.

(b) $\operatorname{proj}_{\boldsymbol{s}}(\boldsymbol{r})$ for vector $\boldsymbol{s} = (2,-2)$ and $\boldsymbol{r} = (1,1)$.

(c) $\operatorname{proj}_{\boldsymbol{p}}(\boldsymbol{q})$ for vector $\boldsymbol{p} = (\frac{1}{3}, \frac{2}{3}, \frac{2}{3})$ and $\boldsymbol{q} = (3,3,0)$.

□

Activity 3.5.31 The projection $\operatorname{proj}_{\boldsymbol{u}}(\boldsymbol{v})$ for vectors $\boldsymbol{u} = (2,6,3)$ and $\boldsymbol{v} = (1\,,4\,,8)$ could be done by premultiplying by which of the following matrices?

(a) $\begin{bmatrix} \frac{1}{81} & \frac{4}{81} & \frac{8}{81} \\ \frac{4}{81} & \frac{16}{81} & \frac{32}{81} \\ \frac{8}{81} & \frac{32}{81} & \frac{64}{81} \end{bmatrix}$

(b) $\begin{bmatrix} \frac{2}{63} & \frac{2}{21} & \frac{1}{21} \\ \frac{8}{63} & \frac{8}{21} & \frac{4}{21} \\ \frac{16}{63} & \frac{16}{21} & \frac{8}{21} \end{bmatrix}$

(c) $\begin{bmatrix} \frac{4}{49} & \frac{12}{49} & \frac{6}{49} \\ \frac{12}{49} & \frac{36}{49} & \frac{18}{49} \\ \frac{6}{49} & \frac{18}{49} & \frac{9}{49} \end{bmatrix}$

(d) $\begin{bmatrix} \frac{2}{63} & \frac{8}{63} & \frac{16}{63} \\ \frac{2}{21} & \frac{8}{21} & \frac{16}{21} \\ \frac{1}{21} & \frac{4}{21} & \frac{8}{21} \end{bmatrix}$

Example 3.5.32 Find the matrices of the following orthogonal projections.

(a) $\operatorname{proj}_{\mathbb{X}}(\boldsymbol{v})$ where $\mathbb{X}$ is the xy-plane in xyz-space.

(b) $\operatorname{proj}_{\mathbb{W}}(\boldsymbol{v})$ for the subspace $\mathbb{W} = \operatorname{span}\{(2, -2, 1), (2, 1, -2)\}$.

(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 $2x - \frac{1}{2}y + 4z = 0$.

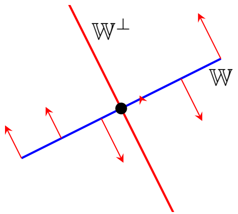
□

Orthogonal decomposition separates

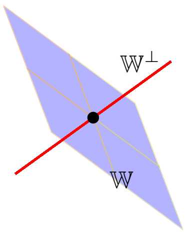
Because orthogonal projection has such a close connection to the geometry underlying important tasks such as 'least square' approximation (Theorem 3.5.26), this section develops further some orthogonal properties.

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 "W perp", as illustrated below and defined by Definition 3.5.34.

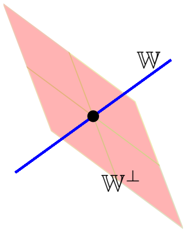
1. Given the blue subspace $\mathbb{W}$ in $\mathbb{R}^2$ (the origin is a black dot), consider the set of all vectors at right-angles to $\mathbb{W}$ (drawn arrows). Move the base of these vectors to the origin, and then they all lie in the red subspace $\mathbb{W}^\perp$.



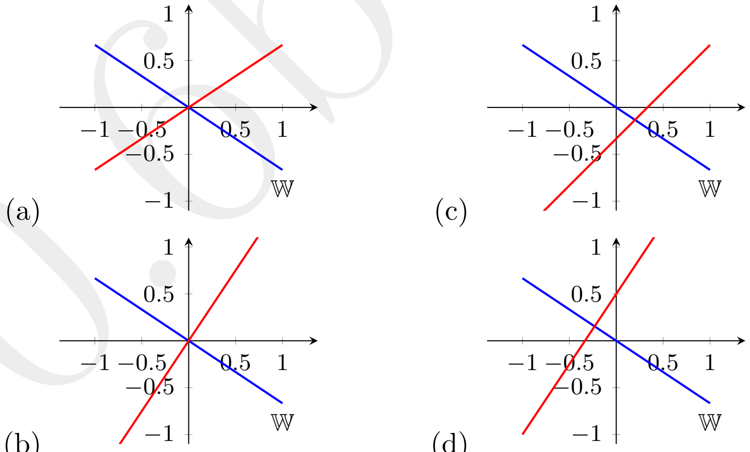
2. Given the blue plane subspace $\mathbb{W}$ in $\mathbb{R}^3$ (the origin is a black dot), the red line subspace $\mathbb{W}^\perp$ contains all vectors orthogonal to $\mathbb{W}$ (when drawn with their base at the origin).



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 $\mathbb{W}$?



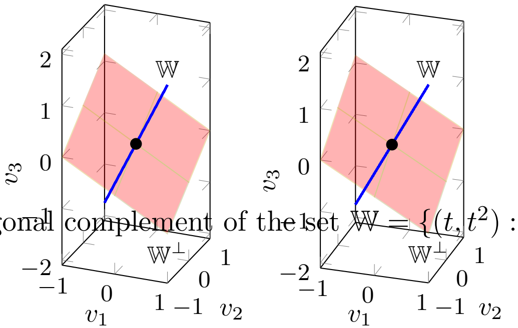
Definition 3.5.34 (orthogonal complement) *Let $\mathbb{W}$ be a k-dimensional subspace of $\mathbb{R}^n$. The set of all vectors $\boldsymbol{u} \in \mathbb{R}^n$ (together with $\boldsymbol{0}$) that are each orthogonal to all vectors in $\mathbb{W}$ is called the* ***orthogonal complement*** *$\mathbb{W}^\perp$ ("W-perp"); that is,*

$$\mathbb{W}^\perp = \{\boldsymbol{u} \in \mathbb{R}^n : \boldsymbol{u} \cdot \boldsymbol{w} = 0 \textit{ for all } \boldsymbol{w} \in \mathbb{W}\}.$$

Example 3.5.35 *(orthogonal complement)*

(a) Given the subspace $\mathbb{W} = \operatorname{span}\{(3,4)\}$, find its orthogonal complement $\mathbb{W}^\perp$.

(b) Describe the orthogonal complement $\mathbb{X}^\perp$ to the subspace $\mathbb{X} = \operatorname{span}\{(4,-4,7)\}$.



(c) Describe the orthogonal complement of the set $\mathbb{W} = \{(t, t^2) : t \in \mathbb{R}\}$.

(d) Given the subspace $\mathbb{W} = \operatorname{span}\{(2, -2, 1), (2, 1, -2)\}$, determine the orthogonal complement of $\mathbb{W}$.

$\square$

Activity 3.5.36 Which of the following vectors are in the orthogonal complement of the vector space spanned by $(3, -1, 1)$?

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

(b) $(3, 5, -4)$

(c) $(1, 3, -1)$

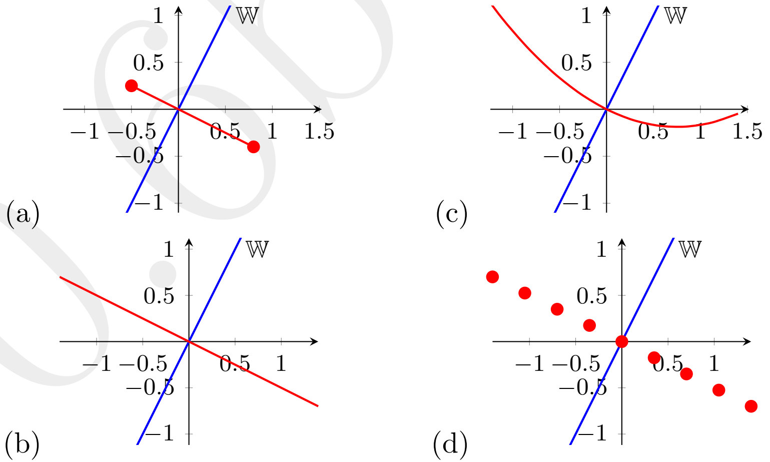
(d) $(6, -2, 2)$

Example 3.5.37 Prove $\{\mathbf{0}\}^\perp = \mathbb{R}^n$ and $(\mathbb{R}^n)^\perp = \{\mathbf{0}\}$. $\square$

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.

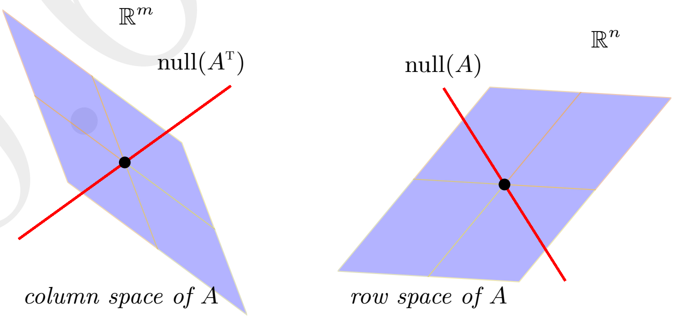
Theorem 3.5.38 (orthogonal complement is a subspace) *For every subspace $\mathbb{W}$ of $\mathbb{R}^n$, the orthogonal complement $\mathbb{W}^\perp$ is a subspace of $\mathbb{R}^n$. Further, the intersection $\mathbb{W} \cap \mathbb{W}^\perp = \{\mathbf{0}\}$; that is, the zero vector is the only vector 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 $\mathbb{W}$?



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 $\operatorname{null}(A^{\mathsf T})$ as its orthogonal complement in $\mathbb{R}^m$. That is, identifying the columns of matrix $A = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \end{bmatrix}$, and denoting the column space by $\mathbb{A} = \operatorname{span}\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n\}$, then the orthogonal complement $\mathbb{A}^\perp = \operatorname{null}(A^{\mathsf T})$. Further, $\operatorname{null}(A)$ in $\mathbb{R}^n$ is the orthogonal complement of the row space of A.*



Example 3.5.41

(a) Let the subspace $\mathbb{W} = \operatorname{span}\{(2, -1)\}$. 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 $(1,1,0,1,0,0)$, $(-1,0,1,0,1,0)$, $(0,-1,-1,0,0$ and $(0,0,0,-1,-1,-1)$.

$\square$

In the previous Example 3.5.41(d) there are three nonzero 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) 3×4 (b) 7×5 (c) 4×3 (d) 7×3

Example 3.5.43 Recall the cases of Example 3.5.41.

3.5.41(a) : $\dim\mathbb{W} + \dim\mathbb{W}^\perp = 1+1 = 2 = \dim\mathbb{R}^2$.

3.5.41(b) : $\dim\mathbb{W} + \dim\mathbb{W}^\perp = 1+2 = 3 = \dim\mathbb{R}^3$.

3.5.41(c) : $\dim\mathbb{W} + \dim\mathbb{W}^\perp = 2+1 = 3 = \dim\mathbb{R}^3$.

3.5.41(d) : $\dim\mathbb{W} + \dim\mathbb{W}^\perp = 3+3 = 6 = \dim\mathbb{R}^6$.

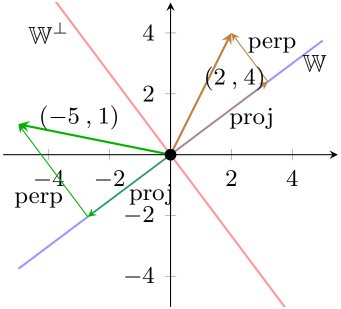
□

Recall that 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 3.5.35(a) that subspace $\mathbb{W} = \operatorname{span}\{(3,4)\}$ has orthogonal complement $\mathbb{W}^\perp = \operatorname{span}\{(-4,3)\}$, as illustrated.



As shown, for example, write the brown vector $(2\,,4) = (3.2\,,2.4) + (-1.2\,,1.6) = \operatorname{proj}_{\mathbb{W}}(2\,,4) + \text{perp}$, 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 another example, write the green vector $(-5\,,1) = (-2.72\,,-2.04) + (-2.28\,,3.04) = \operatorname{proj}_{\mathbb{W}}(-5\,,1) + \text{perp}$, where in this case the vector $\text{perp} = (-2.28\,,3.04) \in \mathbb{W}^\perp$. $\square$

Activity 3.5.46 Let subspace $\mathbb{W} = \operatorname{span}\{(1\,,1)\}$ and its orthogonal complement $\mathbb{W}^\perp = \operatorname{span}\{(1\,,-1)\}$. 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) $(9\,,-9) + (-4\,,0)$
(c) $(-2\,,-2) + (7\,,-7)$
(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 $\boldsymbol{v} \in \mathbb{R}^n$, the* ***perpendicular component*** *of $\boldsymbol{v}$ to $\mathbb{W}$ is the vector* $\operatorname{perp}_{\mathbb{W}}(\boldsymbol{v}) := \boldsymbol{v} - \operatorname{proj}_{\mathbb{W}}(\boldsymbol{v})$.

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 that 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 that the perpendicular component lies in the plane $-2x-3y+6z=0$.

(c) Let the subspace $\mathbb{X} = \operatorname{span}\{(2,-2,1),(2,1,-2)\}$. Determine the perpendicular component of each of the two vectors $\boldsymbol{y}=(3,2,1)$ and $\boldsymbol{z}=(3,-3,-3)$.

□

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 $\boldsymbol{v}$ be any vector in $\mathbb{R}^n$, then the perpendicular component* $\operatorname{perp}_{\mathbb{W}}(\boldsymbol{v}) \in \mathbb{W}^\perp$.

Example 3.5.50 The previous examples' calculation of the perpendicular component confirm that $\boldsymbol{v} = \operatorname{proj}_{\mathbb{W}}(\boldsymbol{v}) + \operatorname{perp}_{\mathbb{W}}(\boldsymbol{v})$, where we now know that $\operatorname{perp}_{\mathbb{W}}$ is orthogonal to $\mathbb{W}$:

Example 3.5.45 : $(2, 4) = (3.2, 2.4) + (-1.2, 1.6)$ and
$(-5, 1) = (-2.72, -2.04) + (-2.28, 3.04)$;

Example 3.5.48(b) : $(-5, -1, 6) = (-2, -3, 6) + (-3, 2, 0)$;

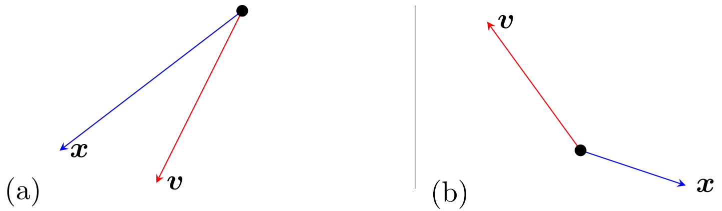
Example 3.5.48(c) : $(3, 2, 1) = (2, 0, -1) + (1, 2, 2)$ and
$(3, -3, -3) = (4, -1, -1) + (-1, -2, -2)$.

□

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 $\boldsymbol{v} \in \mathbb{R}^n$, then there exist unique vectors $\boldsymbol{w} \in \mathbb{W}$ and $\boldsymbol{n} \in \mathbb{W}^\perp$ such that vector $\boldsymbol{v} = \boldsymbol{w} + \boldsymbol{n}$; this particular sum is called an* ***orthogonal decomposition*** *of $\boldsymbol{v}$.*

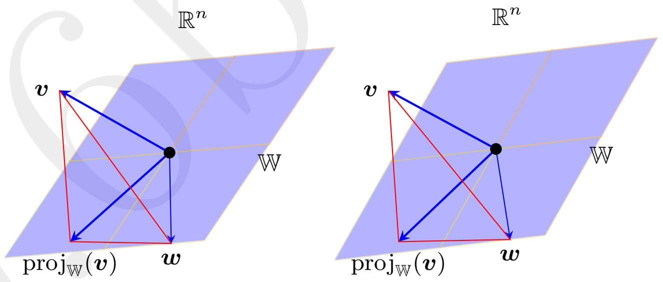
Example 3.5.52 For each pair of the shown subspaces $\mathbb{X} = \operatorname{span}\{\boldsymbol{x}\}$ and vectors $\boldsymbol{v}$, draw the decomposition of vector $\boldsymbol{v}$ into the sum of vectors in $\mathbb{X}$ and $\mathbb{X}^\perp$.



□

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 3.5.53 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 $\boldsymbol{b}$ to make the linear equation $A\boldsymbol{x} = \boldsymbol{b}$ consistent (Theorem 3.5.8) is also to be viewed as the projection of the right-hand side $\boldsymbol{b}$ to the *closest* point in the column space of the matrix. That is, the 'least square' procedure solves $A\boldsymbol{x} = \operatorname{proj}_{\mathbb{A}}(\boldsymbol{b})$.

Theorem 3.5.53 (best approximation) *For every vector* $\boldsymbol{v}$ *in* $\mathbb{R}^n$*, and every subspace* $\mathbb{W}$ *in* $\mathbb{R}^n$*,* $\operatorname{proj}_{\mathbb{W}}(\boldsymbol{v})$ *is the closest vector in* $\mathbb{W}$ *to* $\boldsymbol{v}$*; that is,* $|\boldsymbol{v}-\operatorname{proj}_{\mathbb{W}}(\boldsymbol{v})| \leq |\boldsymbol{v}-\boldsymbol{w}|$ *for every* $\boldsymbol{w} \in \mathbb{W}$*.*

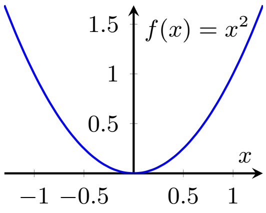


Discover power laws **????????** use log-log plots as examples of the scientific inference of some surprising patterns in nature. These are simple examples of what, in modern parlance, might be termed ‘data mining’, ‘knowledge discovery’, or ‘artificial intelligence’.

3.6 Introducing linear transformations

Section contents

This optional section unifies the transformation examples seen so far, and forms a foundation for more advanced algebra.



Recall function notation. For example, $f(x) = x^2$ means that for every $x \in \mathbb{R}$ the function $f(x)$ gives a result in $\mathbb{R}$, namely the value x^2, as plotted to the right. We often write $f : \mathbb{R} \to \mathbb{R}$ to denote this functionality: that is, $f : \mathbb{R} \to \mathbb{R}$ means that function f transforms any given real number into another real number by some specific rule.

There is analogous functionality in multiple dimensions with vectors: given any vector,

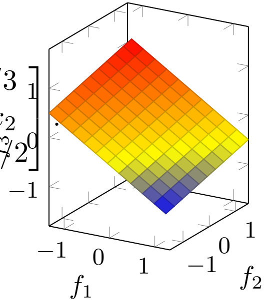
- multiplication by a diagonal matrix stretches and/or shrinks the vector (Section 3.2.2);
- multiplication by an orthogonal matrix rotates and/or reflects the vector (Section 3.2.3); and
- projection finds a vector's components in a subspace (Section 3.5.3).

Correspondingly, we use the notation $f : \mathbb{R}^n \to \mathbb{R}^m$ 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 $f(\boldsymbol{x})$ 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

$$f(\boldsymbol{x}) = A\boldsymbol{x} = \begin{bmatrix} 1 & -\frac{1}{3} \\ \frac{1}{2} & -1 \\ -1 & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 - x_2/3 \\ x_1/2 - x_2 \\ -x_1 - x_2/2 \end{bmatrix}.$$



That is, here $f : \mathbb{R}^2 \to \mathbb{R}^3$. 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 $\boldsymbol{x} \in \mathbb{R}^2$, so we ask what is the shape, the structure, of all the possible results of the function? The plot to the above-right illustrates the subspace formed by this $f(\boldsymbol{x})$ for all 2D vectors $\boldsymbol{x}$.

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 $T : \mathbb{R}^n \to \mathbb{R}^m$ is called a* ***linear transformation*** *if*

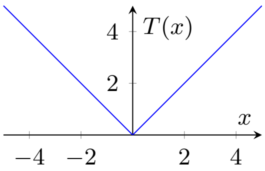
(a) *$T(\boldsymbol{u} + \boldsymbol{v}) = T(\boldsymbol{u}) + T(\boldsymbol{v})$ for every $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$, and*

(b) *$T(c\boldsymbol{v}) = cT(\boldsymbol{v})$ for every $\boldsymbol{v} \in \mathbb{R}^n$ and every scalar c.*

Example 3.6.2 *(1D cases)*

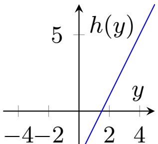
(a) Show that the parabolic function $f : \mathbb{R} \to \mathbb{R}$ where $f(x) = x^2$ is not a linear transformation.

(b) Is the function $T(x) = |x|$, $T : \mathbb{R} \to \mathbb{R}$, a linear transformation?

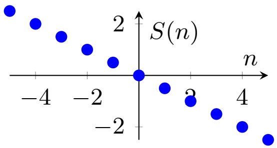


(c) Is the function $g : \mathbb{R} \to \mathbb{R}$ such that $g(x) = -x/2$ a linear transformation?

(d) Show that the function $h(y) = 2y - 3$, $h : \mathbb{R} \to \mathbb{R}$, is not a linear transformation.

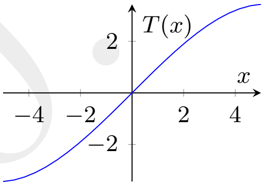


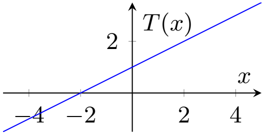
(e) Is the function $S : \mathbb{Z} \to \mathbb{Z}$ given by $S(n) = -n/2$ a linear transformation? Here $\mathbb{Z}$ denotes the set of integers $\ldots, -2, -1, 0, 1, 2, \ldots$.

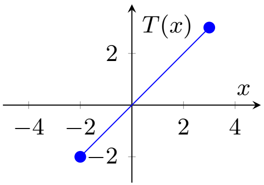


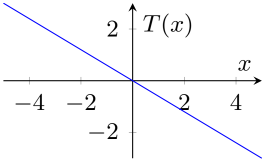
□

Activity 3.6.3 Which of the following is the graph of a linear transformation?

(a) 

(b) 

(c) 

(d) 

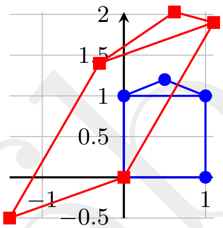
Example 3.6.4 *(higher-D cases)*

(a) Let function $T : \mathbb{R}^3 \to \mathbb{R}^3$ be $T(x, y, z) = (y, z, x)$. Is T a linear transformation?

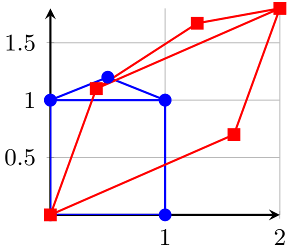
(b) Consider the function $f(x, y, z) = x + y + 1$, $f : \mathbb{R}^3 \to \mathbb{R}$. Is f a linear transformation?

(c) Consider the following illustrated transformations of the plane, $T : \mathbb{R}^2 \to \mathbb{R}^2$. Which *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).

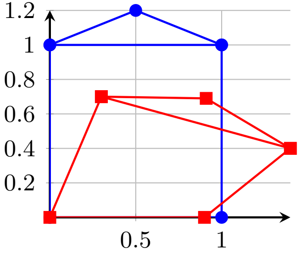
i.



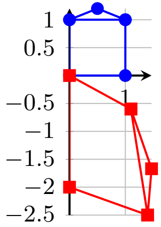
iii.

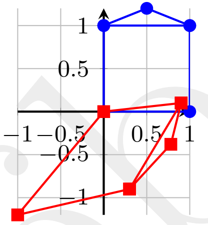


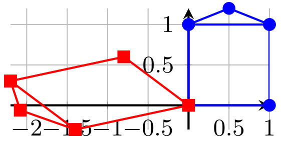
ii.



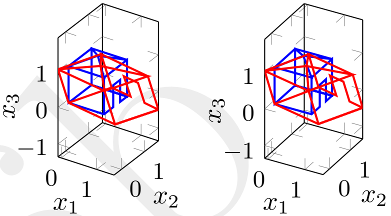
iv.



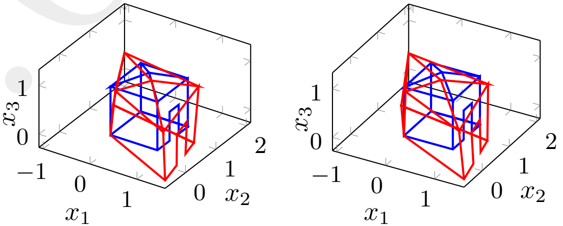
v. 

vi. 

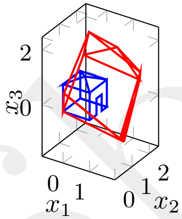
(d) The previous Example 3.6.4(c) illustrated that a linear transformation of the square seems to transform the unit square to a parallelogram: if a function transforms the unit square to something that is not a parallelogram, then the function cannot be a linear transformation. Analogously in higher dimensions: for example, if a function transforms the unit cube to something that is not a parallelepiped, then the function is not a linear transformation. Using this information, which of the following illustrated functions, $f : \mathbb{R}^3 \to \mathbb{R}^3$, *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').

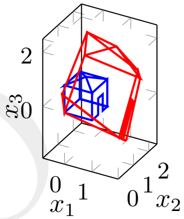
i. 

This *may* be a linear transformation as the transform of the unit cube looks like a parallelepiped, with the origin fixed.

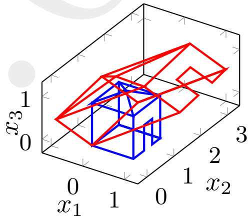
ii. 

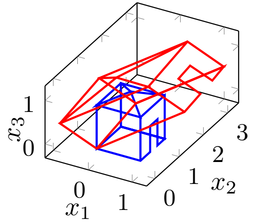
This *cannot* be a linear transformation as the unit cube transforms to something not a parallelepiped.

iii. 



This *cannot* be a linear transformation as the unit cube transforms to something not a parallelepiped.

iv. 



This *may* be a linear transformation as the transform of the unit cube looks like a parallelepiped, with the origin fixed.

□

Activity 3.6.5 Which of the following functions $f : \mathbb{R}^3 \to \mathbb{R}^2$ is *not* a linear transformation?

(a) $f(x\,,y\,,z) = (2.7x+3y\,,1-2z)$

(b) $f(x\,,y\,,z) = (0\,,13x+\pi y)$

(c) $f(x\,,y\,,z) = (y\,,x+z)$

(d) $f(x\,,y\,,z) = (0\,,0)$

Example 3.6.6 For any given nonzero vector $\boldsymbol{w} \in \mathbb{R}^n$, prove that the projection $P : \mathbb{R}^n \to \mathbb{R}^n$ defined by $P(\boldsymbol{u}) := \operatorname{proj}_{\boldsymbol{w}}(\boldsymbol{u})$ is a linear transformation (as a function of $\boldsymbol{u}$). But, for any given nonzero vector $\boldsymbol{u} \in \mathbb{R}^n$, prove that the projection $Q : \mathbb{R}^n \to \mathbb{R}^n$ defined by $Q(\boldsymbol{w}) := \operatorname{proj}_{\boldsymbol{w}}(\boldsymbol{u})$ is not a linear transformation (as a function of $\boldsymbol{w}$). □

3.6.1 Matrices correspond to linear transformations

One important class of linear transformations are the transformations that can be written as matrix multiplications. One reason for the importance is that Theorem 3.6.10 establishes that all linear transformations may be written as matrix multiplications! This in turn justifies why we define matrix multiplication to be as it is (Section 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 (Section 3.2.2);
- rotations and/or reflections, as they arise as multiplications by an orthogonal matrix (Section 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 $T_A : \mathbb{R}^n \to \mathbb{R}^m$ by the matrix multiplication $T_A(\boldsymbol{x}) := A\boldsymbol{x}$ for every $\boldsymbol{x} \in \mathbb{R}^n$. Then T_A is a linear transformation.*

Example 3.6.9 Prove that a matrix multiplication with a nonzero shift $\boldsymbol{b}$, $S : \mathbb{R}^n \to \mathbb{R}^m$ where $S(\boldsymbol{x}) = A\boldsymbol{x} + \boldsymbol{b}$ for some given vector $\boldsymbol{b} \neq \boldsymbol{0}$, 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 $T : \mathbb{R}^n \to \mathbb{R}^m$ be any linear transformation. Then T is the transformation corresponding to the $m \times n$ matrix*

$$A = \begin{bmatrix} T(\boldsymbol{e}_1) & T(\boldsymbol{e}_2) & \cdots & T(\boldsymbol{e}_n) \end{bmatrix}$$

*where $\boldsymbol{e}_j$ 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.*

The matrix $[T]$ is called the *standard matrix* because it is defined in terms of the *standard* unit vectors $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, ... , $\boldsymbol{e}_n$. Later, Theorem 7.2.35 begins to show how to use non-standard vectors to construct and interpret matrices corresponding a given linear transformation.

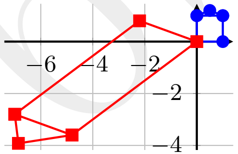
Example 3.6.11

(a) Find the standard matrix of the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^4$ where $T(x, y, z) = (y, z, x, 3x - 2y + z)$.

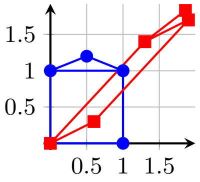
(b) Find the standard matrix of the rotation of the plane by 60° anticlockwise about the origin.

(c) Find the standard matrix of the rotation about the point $(1, 0)$ of the plane by 45° anticlockwise.

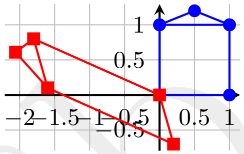
(d) Estimate the standard matrix for each of the illustrated transformations given that they transform the unit square as shown.

i. 

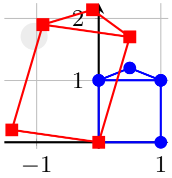
Solution: Here $T(1, 0) \approx (-2.2, 0.8)$ and $T(0, 1) \approx (-4.8, -3.6)$ so the approximate standard matrix is $\begin{bmatrix} -2.2 & -4.8 \\ 0.8 & -3.6 \end{bmatrix}$.

ii. 

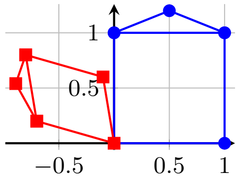
Solution: Here $T(1, 0) \approx (0.6, 0.3)$ and $T(0, 1) \approx (1.3, 1.4)$ so the approximate standard matrix is $\begin{bmatrix} 0.6 & 1.3 \\ 0.3 & 1.4 \end{bmatrix}$.

iii. 

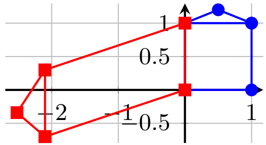
Solution: Here $T(1,0) \approx (0.2, -0.7)$ and $T(0,1) \approx (-1.8, 0.8)$ so the approximate standard matrix is $\begin{bmatrix} 0.2 & -1.8 \\ -0.7 & 0.8 \end{bmatrix}$.

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 $\begin{bmatrix} -1.4 & 0.5 \\ 0.2 & 1.7 \end{bmatrix}$.

v. 

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 $\begin{bmatrix} -0.1 & -0.7 \\ 0.6 & 0.2 \end{bmatrix}$.

vi. 

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) $\begin{bmatrix} 0 & 1.9 \\ 4.5 & 0 \\ -1.6 & -2 \end{bmatrix}$

(b) $\begin{bmatrix} 0 & 4.5 & -1.6 \\ 1.9 & 0 & -2 \end{bmatrix}$

(c) $\begin{bmatrix} 4.5 & -1.6 \\ 1.9 & -2 \end{bmatrix}$

(d) $\begin{bmatrix} 4.5 & 1.9 \\ -1.6 & -2 \end{bmatrix}$

Example 3.6.13 For a fixed scalar a, let the function $H : \mathbb{R}^n \to \mathbb{R}^n$ be $H(\boldsymbol{u}) = a\boldsymbol{u}$. 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 $H(\boldsymbol{u}) = \boldsymbol{u}$ 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 attempting to solve inconsistent linear equations, $A\boldsymbol{x} = \boldsymbol{b}$ for some given A, Procedure 3.5.4 constructs a result $\boldsymbol{x}$, a result that is meant to be a 'solution' of sorts, and that depends upon the right-hand side $\boldsymbol{b}$. That is, any given $\boldsymbol{b}$ is transformed by the procedure to some result $\boldsymbol{x}$: the result is a function of the given $\boldsymbol{b}$. This section establishes that the result given by the procedure is a linear transformation of $\boldsymbol{b}$, and hence there must be a matrix corresponding to the procedure. Let's denote this matrix by A^+. This matrix gives the result $\boldsymbol{x} = A^+\boldsymbol{b}$. We call the matrix A^+ the pseudo-inverse of A.

Example 3.6.14 Find the pseudo-inverse of the matrix $A = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$. $\square$

Activity 3.6.15 By finding the smallest magnitude, least square, solution to $D\boldsymbol{x} = \boldsymbol{b}$ for matrix $D = \begin{bmatrix} 5 & 0 \\ 0 & 0 \end{bmatrix}$ and arbitrary $\boldsymbol{b}$, determine that the pseudo-inverse of the diagonal matrix D is which of the following?

(a) $\begin{bmatrix} 0.2 & 0 \\ 0 & 0 \end{bmatrix}$ (b) $\begin{bmatrix} 0 & 0 \\ 0 & 0.2 \end{bmatrix}$ (c) $\begin{bmatrix} 0 & 0.2 \\ 0 & 0 \end{bmatrix}$ (d) $\begin{bmatrix} 0 & 0 \\ 0.2 & 0 \end{bmatrix}$

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 rationalizes 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 many times the pseudo-inverse is not appropriate.

Theorem 3.6.16 (pseudo-inverse) *In the context of a system of linear equations $A\boldsymbol{x} = \boldsymbol{b}$ with $m \times n$ matrix A vector $\boldsymbol{b} \in \mathbb{R}^m$, recall that Procedure 3.5.4 finds the smallest solution $\boldsymbol{x} \in \mathbb{R}^n$ (Theorem 3.5.13) to the closest consistent system $A\boldsymbol{x} = \boldsymbol{b}'$ (Theorem 3.5.8). Procedure 3.5.4 forms a linear transformation $T : \mathbb{R}^m \to \mathbb{R}^n$, $\boldsymbol{x} = T(\boldsymbol{b})$. 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 $\left[\begin{smallmatrix}1&1\\2&2\end{smallmatrix}\right]$ given that this matrix has the SVD

$$\begin{bmatrix}1&1\\2&2\end{bmatrix}=\begin{bmatrix}\frac{1}{\sqrt5}&-\frac{2}{\sqrt5}\\\frac{2}{\sqrt5}&\frac{1}{\sqrt5}\end{bmatrix}\begin{bmatrix}10&0\\0&0\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt2}&-\frac{1}{\sqrt2}\\\frac{1}{\sqrt2}&\frac{1}{\sqrt2}\end{bmatrix}^{\text{T}}$$

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

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

(a) As used previously, this matrix A of ones has an SVD of

$$A = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac12 & \frac12 & \frac12 & \frac12 \\ \frac12 & \frac12 & -\frac12 & -\frac12 \\ \frac12 & -\frac12 & -\frac12 & \frac12 \\ \frac12 & -\frac12 & \frac12 & -\frac12 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \end{bmatrix}^{\text{T}} = USV^{\text{T}}.$$

(b) Solve $U\boldsymbol{z} = \boldsymbol{b}$ by computing

$$\boldsymbol{z} = U^{\mathrm{T}}\boldsymbol{b} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \boldsymbol{b} = \begin{bmatrix} \frac{1}{2}b_1 + \frac{1}{2}b_2 + \frac{1}{2}b_3 + \frac{1}{2}b_4 \\ \frac{1}{2}b_1 + \frac{1}{2}b_2 - \frac{1}{2}b_3 - \frac{1}{2}b_4 \\ \frac{1}{2}b_1 - \frac{1}{2}b_2 - \frac{1}{2}b_3 + \frac{1}{2}b_4 \\ \frac{1}{2}b_1 - \frac{1}{2}b_2 + \frac{1}{2}b_3 - \frac{1}{2}b_4 \end{bmatrix}.$$

(c) Now try to solve $Sy = \boldsymbol{z}$, 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 *must* 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 $V^{\mathrm{T}} x = y$ by computing

$$x = Vy = 1y = \frac{1}{4}b_1 + \frac{1}{4}b_2 + \frac{1}{4}b_3 + \frac{1}{4}b_4 = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix} \boldsymbol{b}.$$

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 the four 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 $A\boldsymbol{x} = \boldsymbol{b}$ for the particular matrix and vector

$$\begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}.$$

Find the pseudo-inverse of this matrix A. Use the pseudo-inverse to rate the players in the cases of Examples 3.5.3 and 3.3.12. □

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 2×1 matrix $A = \left[\begin{smallmatrix} 3 \\ 4 \end{smallmatrix}\right]$, confirm that $(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ is the pseudo-inverse that was found in Example 3.6.14. □

Theorem 3.6.22 *For every $m \times n$ matrix A with* $\operatorname{rank} A = n$ *(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.*

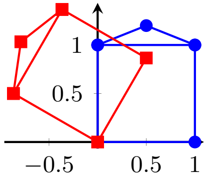
Computer considerations Except for easy cases, we (almost) never explicitly compute the pseudo-inverse of a matrix. In practical computation, forming $A^T A$ and then manipulating it is both expensive and error enhancing: for example, $\operatorname{cond}(A^T A) = (\operatorname{cond} A)^2$ so matrix $A^T A$ 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. As with an inverse, a pseudo-inverse is a theoretical device, rarely a practical tool.

A major 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 complex goal into a set of smaller simpler 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 3.6.11(b) on rotation by 60° (anti-clockwise as positive, as illustrated to the right) 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°: show that the standard matrix of the resultant rotation by 120° is the same as the matrix product $[R][R]$.

□

Theorem 3.6.25 *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ and $S : \mathbb{R}^m \to \mathbb{R}^p$ be any linear transformations. Recall that the* ***composition*** *of functions is $(S \circ T)(\boldsymbol{v}) = S(T(\boldsymbol{v}))$. Then $S \circ T : \mathbb{R}^n \to \mathbb{R}^p$ is a linear transformation with standard matrix $[S \circ T] = [S][T]$.*

Example 3.6.26 Consider two linear transformations: first, $T : \mathbb{R}^3 \to \mathbb{R}^2$ defined by $T(x_1, x_2, x_3) := (3x_1 + x_2, -x_2 - 7x_3)$; and second, the linear transformation $S : \mathbb{R}^2 \to \mathbb{R}^4$ 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° about the origin (anticlockwise as positive), 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 is due to the non-commutativity of matrix multiplication (Section 3.1.3). □

Activity 3.6.28 Given the stretching transformation S with standard matrix $[S] = \operatorname{diag}(2\,,1/2)$, and the anticlockwise rotation R by 90° 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?

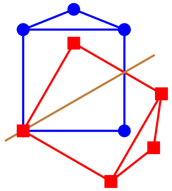
(a) $\begin{bmatrix} 0 & -\frac{1}{2} \\ 2 & 0 \end{bmatrix}$ (b) $\begin{bmatrix} 0 & \frac{1}{2} \\ -2 & 0 \end{bmatrix}$ (c) $\begin{bmatrix} 0 & 2 \\ -\frac{1}{2} & 0 \end{bmatrix}$ (d) $\begin{bmatrix} 0 & -2 \\ \frac{1}{2} & 0 \end{bmatrix}$

Invert transformations Having introduced and characterized 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° (anticlockwise as positive), and T be rotation of the plane by -60° (clockwise as negative). Then $S \circ T$ is first the -60° rotation by T, and second the 60° rotation by S: the overall 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.

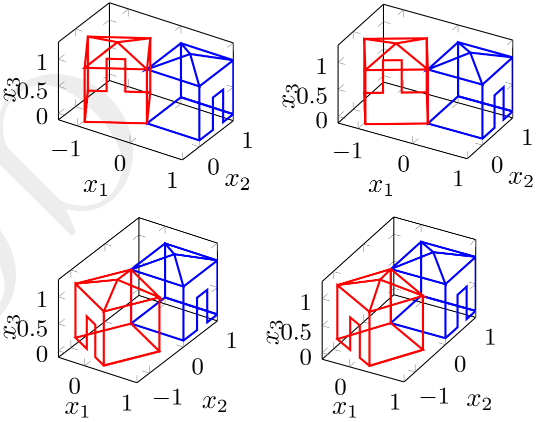
(b) Let R be reflection of the plane in the line at 30° to the horizontal (illustrated to the right). Then $R \circ R$ is first reflection in the line at 30° by R, and second another reflection in the line at 30° by R: the overall result is no change. Because $R \circ R$ is effectively the identity transformation, the reflection R is its own inverse.



□

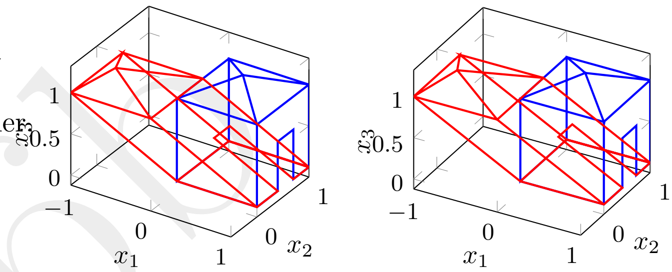
Definition 3.6.30 *Let S and T be linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^n$ (the same dimension). If $S \circ T = T \circ S = I$, the identity transformation, then S and T are **inverse transformations** of each other. Further, we say S and T are **invertible**.*

Example 3.6.31 Let $S : \mathbb{R}^3 \to \mathbb{R}^3$ be rotation about the vertical axis by 120° (as illustrated in stereo above-right), and let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be rotation about the vertical axis by 240° (below-right). Argue that $S \circ T = T \circ S = I$, the identity, and so S and T are inverse transformations of each other.



$\square$

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 to the right. 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. □

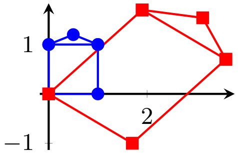


Because of the exact correspondence between linear transformations and matrix multiplication, the inverse of a transformation exactly corresponds to the inverse of a matrix. In the last Example 3.6.32, because $[R][S] = I_3$ we know that the matrices $[R]$

and $[S]$ are inverses of each other. Correspondingly, the transformations R and S are inverses of each other.

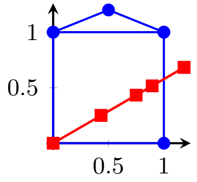
Theorem 3.6.33 *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be an invertible linear transformation. Then its standard matrix $[T]$ is invertible, and $[T^{-1}] = [T]^{-1}$.*

Example 3.6.34 Estimate the standard matrix of the linear transformation T illustrated to the right. 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 that it is a reasonable inverse.



$\square$

Example 3.6.35 Determine if the orthogonal projection of the plane onto the line at 30° to the horizontal (illustrated to the right) is an invertible transformation; if it is, then find its inverse.



□

4 Eigenvalues and eigenvectors of symmetric matrices

Chapter contents

Recall (Section 3.1.2) that a symmetric matrix A is a square matrix such that $A^{\mathrm{T}} = A$, that is, $a_{ij} = a_{ji}$. For example, of the following two matrices, the first is symmetric, but the second is not:

$$\begin{bmatrix} -2 & 4 & 0 \\ 4 & 2 & -3 \\ 0 & -3 & 1 \end{bmatrix}; \qquad \begin{bmatrix} -1 & 3 & 0 \\ 1 & 1 & 0 \\ 0 & -3 & 1 \end{bmatrix}.$$

Example 4.0.1 Compute some SVDs of random symmetric matrices, $A = USV^{\mathrm{T}}$, observe in the SVDs that the columns of U are always $\pm$ the columns of V (well, almost always). $\square$

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 $A = USV^{\mathrm{T}}$, post-multiplying by V gives $AV = USV^{\mathrm{T}}V = US$, and then the jth column of the two sides of $AV = US$ determines $A\boldsymbol{v}_j = \sigma_j \boldsymbol{u}_j$. Example 4.0.1 indicates that for symmetric matrices A we find $\boldsymbol{u}_j = \pm\boldsymbol{v}_j$ (almost always) so this last equation becomes $A\boldsymbol{v}_j = (\pm\sigma_j)\boldsymbol{v}_j$. This equation is of the important form $A\boldsymbol{v} = \lambda\boldsymbol{v}$. (The symbol λ is the Greek letter lambda, and denotes eigenvalues.) This form is important because it is the mathematical expression of the following geometric question: for what vectors $\boldsymbol{v}$ does multiplication by A just stretch/shrink $\boldsymbol{v}$ by some scalar λ?

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 $= A\boldsymbol{v}$ where $\boldsymbol{v}$ is the orthogonal unit vector to the internal 'surface'. In this solid-force scenario, the matrix A is always symmetric. To know whether a material breaks apart under pulling, or crumbles under compression, we need to know where the extreme forces are. The extreme forces are found as solutions to $A\boldsymbol{v} = \lambda\boldsymbol{v}$ where $\boldsymbol{v}$ gives the direction and λ the strength of the extreme force. To understand the potential failure of the material we thus need to solve equations in the form $A\boldsymbol{v} = \lambda\boldsymbol{v}$.

4.1 Introduction to eigenvalues and eigenvectors

Section contents

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, without changing direction?

Definition 4.1.1 *Let A be a square matrix. A scalar λ (lambda) is called an **eigenvalue** of A if there is a nonzero vector $\boldsymbol{x}$ such that $A\boldsymbol{x} = \lambda\boldsymbol{x}$. Such a vector $\boldsymbol{x}$ is called an **eigenvector** of A corresponding to the eigenvalue λ.*

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?

$\square$

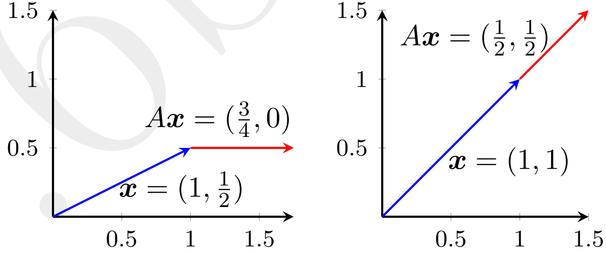
Activity 4.1.3 Which of the following vectors is an eigenvector of the symmetric matrix $\begin{bmatrix} -1 & 12 \\ 12 & 6 \end{bmatrix}$?

(a) $\begin{bmatrix} 4 \\ -3 \end{bmatrix}$ (b) $\begin{bmatrix} -3 \\ 1 \end{bmatrix}$ (c) $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ (d) $\begin{bmatrix} -1 \\ 2 \end{bmatrix}$

Importantly, eigenvectors tell us the 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 $\boldsymbol{x}$ and join onto its head the vector $A\boldsymbol{x}$:

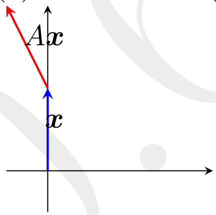
- if both $\boldsymbol{x}$ and $A\boldsymbol{x}$ are aligned in the same direction, or the opposite direction, then $\boldsymbol{x}$ is an eigenvector;
- if they form some other angle, then $\boldsymbol{x}$ is not an eigenvector.

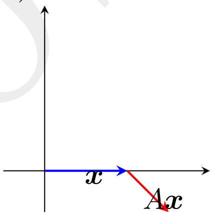
Example 4.1.4 Let the matrix $A = \begin{bmatrix} 1 & -1/2 \\ -1/2 & 1 \end{bmatrix}$. The plot below-left shows the vector $\boldsymbol{x} = (1, \frac{1}{2})$, and adjoined to its head the matrix-vector product $A\boldsymbol{x} = (\frac{3}{4}, 0)$: because the two are at an angle, $(1, \frac{1}{2})$ is not an eigenvector.

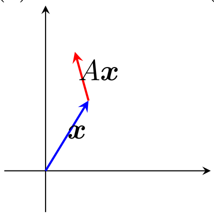


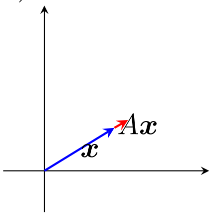
However, as plotted above-right, for the vector $\boldsymbol{x} = (1, 1)$ the matrix-vector product $A\boldsymbol{x} = (\frac{1}{2}, \frac{1}{2})$ and the plot of these vectors head-to-tail illustrates that they are aligned in the same direction. Because of the alignment, $(1, 1)$ is an eigenvector of this matrix. The constant of proportionality is the corresponding eigenvalue: here $A\boldsymbol{x} = (\frac{1}{2}, \frac{1}{2}) = \frac{1}{2}(1, 1) = \frac{1}{2}\boldsymbol{x}$ so the eigenvalue is $\lambda = \frac{1}{2}$. This eigenvalue of $\frac{1}{2}$ is seen graphically by the (red) vector $A\boldsymbol{x}$ being half the length of the (blue) vector $\boldsymbol{x}$. □

Activity 4.1.5 For some matrix A, the following pictures plot a vector $\boldsymbol{x}$ and the corresponding product $A\boldsymbol{x}$, head-to-tail. Which picture indicates that $\boldsymbol{x}$ is an eigenvector of the matrix?

(a) 

(b) 

(c) 

(d) 

Activity 4.1.6 Further, for the picture in Activity 4.1.5 that indicates $\boldsymbol{x}$ is an eigenvector, in which of the following ranges does the corresponding eigenvalue λ lie?

(a) $0.5 > \lambda > 0$

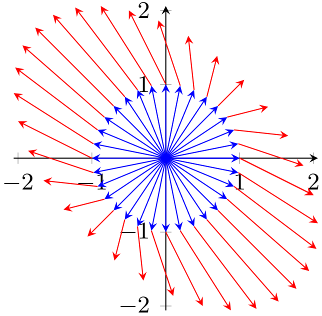
(b) $\lambda > 1$

(c) $1 > \lambda > 0.5$

(d) $0 > \lambda$

As in the next example, we sometimes plot for many directions $\boldsymbol{x}$ a diagram of vector $A\boldsymbol{x}$ adjoined head-to-tail to vector $\boldsymbol{x}$. Then inspection estimates the eigenvectors and corresponding eigenvalues (Schonefeld 1995).

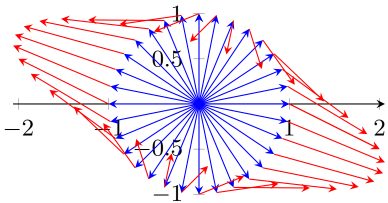
Example 4.1.7 *(graphical eigenvectors one)* The plot on the right shows many unit vectors $\boldsymbol{x}$ (blue), and for some symmetric matrix A the corresponding vectors $A\boldsymbol{x}$ (red) adjoined. Estimate which directions $\boldsymbol{x}$ are eigenvectors, and for each eigenvector estimate the corresponding eigenvalue.



(The Matlab function `eigshow(A)` provides an interactive alternative to this static view. Download from the internet if using a recent version of Matlab.)

□

Example 4.1.8 *(graphical eigenvectors two)* The plot on the right shows many unit vectors $\boldsymbol{x}$ (blue), and for some symmetric matrix A the corresponding vectors $A\boldsymbol{x}$ (red) adjoined. Estimate which directions $\boldsymbol{x}$ are eigenvectors, and for each eigenvector 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 $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$ in turn:

$$\begin{aligned}
D\boldsymbol{e}_1 &= \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} d_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = d_1\boldsymbol{e}_1\,; \\
D\boldsymbol{e}_2 &= \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ d_2 \\ \vdots \\ 0 \end{bmatrix} = d_2\boldsymbol{e}_2\,; \\
&\vdots \\
D\boldsymbol{e}_n &= \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ d_n \end{bmatrix} = d_n\boldsymbol{e}_n\,.
\end{aligned}$$

Thus, by Definition 4.1.1, each diagonal element d_j is an eigenvalue of the diagonal matrix, and the standard unit vector $\boldsymbol{e}_j$ is a corresponding eigenvector. □

Eigenvalues The 3×3 matrix of Example 4.1.2 has three eigenvalues. The 2×2 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 that any nonzero multiple of the standard unit vector $\boldsymbol{e}_j$ is an eigenvector corresponding to the eigenvalue d_j. Let's collect all possible eigenvectors into a subspace.

Hereafter, "iff" is short for "if and only if".

Theorem 4.1.10 *Let A be a square matrix. A scalar λ is an eigenvalue of A iff the homogeneous linear system $(A-\lambda I)\boldsymbol{x}=\boldsymbol{0}$ has nonzero solutions $\boldsymbol{x}$. The set of all eigenvectors corresponding to any one eigenvalue λ, together with the zero vector, is a subspace; the subspace is called the* ***eigenspace*** *of λ and is denoted by $\mathbb{E}_\lambda$.*

Example 4.1.11 Reconsider the symmetric matrix

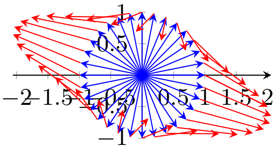
$$A=\begin{bmatrix}1&-1&0\\-1&2&-1\\0&-1&1\end{bmatrix}$$

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) $2x+y=0$
(b) $y=2x$
(c) $x+2y=0$
(d) $x=2y$

Example 4.1.13 *(graphical eigenspaces)* The plot on the above-right shows unit vectors $\boldsymbol{x}$ (blue), and for the matrix A of Example 4.1.8 the corresponding vectors $A\boldsymbol{x}$ (red) adjoined. Estimate and draw the eigenspaces of matrix A.



□

Example 4.1.14 Eigenspaces may be multi-dimensional. 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 λ of A is the dimension of the corresponding eigenspace $\mathbb{E}_\lambda$.*

Be aware that for general matrices (non-symmetric) we subsequently define the multiplicity of an eigenvalue differently (Definition 7.1.7). But for real symmetric matrices the two definitions are the same.

Example 4.1.16 The multiplicity of the various eigenvalues in earlier examples are the following.

Example 4.1.11 : Recall that in this example:

- the eigenspace $\mathbb{E}_1 = \operatorname{span}\{(1, 0, -1)\}$ has dimension one, so the multiplicity of eigenvalue $\lambda = 1$ is one;
- the eigenspace $\mathbb{E}_3 = \operatorname{span}\{(1, -2, 1)\}$ has dimension one, so the multiplicity of eigenvalue $\lambda = 3$ is one; and
- the eigenspace $\mathbb{E}_0 = \operatorname{span}\{(1, 1, 1)\}$ has dimension one, so the multiplicity of eigenvalue $\lambda = 0$ is one.

Example 4.1.14 : Recall that in this example:

- the eigenspace $\mathbb{E}_{-1/3} = \operatorname{span}\{\boldsymbol{e}_1\}$ has dimension one, so the multiplicity of eigenvalue $\lambda = -1/3$ is one; and
- the eigenspace $\mathbb{E}_{3/2} = \operatorname{span}\{\boldsymbol{e}_2, \boldsymbol{e}_3\}$ has dimension two, so the multiplicity of eigenvalue $\lambda = 3/2$ is two. $\square$

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 function `eig()` places eigenvalues in the diagonal of $D = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. Then the jth column of V is a unit eigenvector corresponding to the jth eigenvalue λ_j: $V = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix}$. 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 $V = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix}$ and $D = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, and confirm that the computed V is an orthogonal matrix. □

Table 4.1: As well as the Matlab/Octave commands and operations listed in Tables 1.2, 2.3, and 3.1 to 3.3 we need the eigenvector function `eig()`*.*

- `[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 $D = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ (in no specific order).*
  - *Corresponding to the jth eigenvalue λ_j, the jth 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, Definitions 4.1.15 and 7.1.7).*
- *If the matrix A is a real symmetric matrix, then the eigenvalues and eigenvectors are all real (Theorem 4.2.9), and the eigenvector matrix V is orthogonal (Theorem 4.2.11).*

  If the matrix A is either not symmetric, or is complex valued, then the eigenvalues and eigenvectors may be non-real complex valued.

Activity 4.1.18 The statement `[V,D]=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) -0.1 (b) -0.5 (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 (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 ith 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, here $y_2 - y_1$ and $y_3 - y_2$. For algebraic 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 that 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

$$\begin{aligned}
-f^2x_1\cos(ft) &= x_2\cos(ft) - x_1\cos(ft),\\
-f^2x_2\cos(ft) &= x_1\cos(ft) - 2x_2\cos(ft) + x_3\cos(ft),\\
-f^2x_3\cos(ft) &= x_2\cos(ft) - x_3\cos(ft).
\end{aligned}$$

These are satisfied for all time t only if the coefficients of the cosines are equal on each side of each equation:

$$\begin{aligned}
-f^2x_1 &= x_2 - x_1\,,\\
-f^2x_2 &= x_1 - 2x_2 + x_3\,,\\
-f^2x_3 &= x_2 - x_3\,.
\end{aligned}$$

Moving the terms on the left to the right, and all terms on the right to the left, this becomes the eigen-problem $A\boldsymbol{x} = \lambda\boldsymbol{x}$ 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.

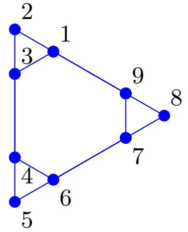
- The eigenvalue $\lambda = 1$ and corresponding eigenvector $\boldsymbol{x} \propto (-1, 0, 1)$ corresponds to oscillations of frequency $f = \sqrt{\lambda} = \sqrt{1} = 1$. The eigenvector $(-1, 0, 1)$ shows that the middle mass is stationary while the outer two masses oscillate in and out, in opposition to each other.

- The eigenvalue $\lambda = 3$ and corresponding eigenvector $\boldsymbol{x} \propto (1, -2, 1)$ 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 $\boldsymbol{x} \propto (1, 1, 1)$ appears as oscillations of zero frequency $f = \sqrt{\lambda} = \sqrt{0} = 0$ which is a static displacement. The eigenvector $(1, 1, 1)$ shows that the static displacement is that of all three masses moved all together as a unit.

That these three solutions combine together to 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 single triangle (as shown to the right)—perhaps because triangles make strong structures, or perhaps because of a hierarchical computer/social network. The marginal picture is called a *network* because it is formed from a set of nodes (the discs) connected by links (the lines). Let's analyse the matrix used to represent such a network.



Form a 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 an orthogonal matrix. □

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. A challenge for an advanced learner is to find the two smallest connected networks that have different connectivity and yet the same eigenvalues (unit strength connections).

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 in the diagonal D (or triplicated), 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), $\lambda_1, \lambda_2, \ldots, \lambda_m$ are eigenvalues of A with corresponding eigenvectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m$, for some m (commonly $m = n$), iff $AV = VD$ for diagonal matrix $D = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_m)$ and $n \times m$ matrix $V = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_m \end{bmatrix}$ for nonzero $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m$.*

Example 4.1.22 Use Matlab/Octave to compute eigenvectors and the eigenvalues of the (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 2×2 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 3×3 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 that 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:

$$\begin{bmatrix} a & b\\ c & d\end{bmatrix} \qquad \left[\begin{matrix} a & b & c\\ d & e & f\\ g & h & i\end{matrix}\right]\begin{matrix} a & b\\ d & e\\ g & h\end{matrix} \tag{4.1}$$

Chapter 6 extends such determinants to any size of matrix, and explores more useful properties, but for now this is all the information we need on determinants.

For hand calculation on small matrices (2×2 or 3×3) the key property is the following. By Definition 4.1.1 eigenvalues and eigenvectors are determined from $A\boldsymbol{x} = \lambda\boldsymbol{x}$. Rearranging, this equation is equivalent to $(A-\lambda I)\boldsymbol{x} = \boldsymbol{0}$. Both Theorem 3.2.7 ($2\times 2$ matrices) and Theorem 6.1.29 (general matrices) establish that $(A-\lambda I)\boldsymbol{x} = \boldsymbol{0}$ has *nonzero* solutions $\boldsymbol{x}$ 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* λ*, solve the homogeneous matrix-vector equation* $(A-\lambda I)\boldsymbol{x} = \boldsymbol{0}$ *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 Consider the matrix $A = \begin{bmatrix} 3 & 2 \\ 2 & 0 \end{bmatrix}$. Use its characteristic equation to determine that all its eigenvalues are which of the following?

(a) $3\,,4$ (b) $0\,,3$ (c) $-4\,,1$ (d) $-1\,,4$

Example 4.1.26 Use the determinant (4.1) 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 non-real 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.

Example 4.1.28 Find the eigenvalues and a corresponding eigenvector for the non-symmetric matrix $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$.

(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 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

$$\begin{bmatrix} x_2 \\ -x_1 \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \iff \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \boldsymbol{x} = \lambda \boldsymbol{x} .$$

We need to find the eigenvalues and eigenvectors of the matrix: we derive that eigenvalues are $\lambda = \pm\sqrt{-1} = \pm\mathrm{i}$. Physically, such complex eigenvalues represent oscillations in time t since, for example, $e^{\lambda t} = e^{\mathrm{i}\,t} = \cos t + \mathrm{i}\sin t$ by Euler's formula.

4.2 Beautiful properties for symmetric matrices

Section contents

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 always having orthogonal eigenvectors.

4.2.1 Matrix powers maintain eigenvectors

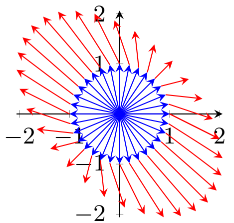
Recall that Section 3.2 introduced the inverse of a matrix (Definition 3.2.2). This first theorem links an eigenvalue of zero to the non-existence of an inverse, and hence links a zero eigenvalue to problematic linear equations.

Theorem 4.2.1 *A square matrix is invertible iff zero is* not *an eigenvalue of the matrix.*

Example 4.2.2

- The 3×3 matrix of Examples 4.1.2, 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 to the right shows (unit) vectors $\boldsymbol{x}$ (blue), and for some matrix A the corresponding vectors $A\boldsymbol{x}$ (red) adjoined. There are no directions $\boldsymbol{x}$ for which $A\boldsymbol{x} = \boldsymbol{0} = 0\boldsymbol{x}$. Hence zero cannot be an eigenvalue, and the matrix A must be invertible.



Similarly for Example 4.1.8.

- The 3×3 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 9×9 matrix of the Sierpinski network in Example 4.1.20 is not invertible as it has zero among its five eigenvalues.
- The 2×2 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, matrix

multiplication confirms that the matrix

$$A = \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix}, \quad \text{has inverse } A^{-1} = \begin{bmatrix} \frac{4}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{4}{3} \end{bmatrix}.$$

- The 2×2 non-symmetric matrix of Example 4.1.28 is invertible because zero is not among its eigenvalues of $\lambda = \pm\mathrm{i}$. Indeed, matrix multiplication confirms that the matrix

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad \text{has inverse } A^{-1} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

□

Example 4.2.3 The next theorem considers eigenvalues and eigenvectors of powers of a matrix. Two examples are the following.

- Recall that the matrix $A = \left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right]$ has eigenvalues $\lambda = \pm\mathrm{i}$.

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 the eigenvalue of A^2, $-1 = (\pm\mathrm{i})^2$, 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 that matrix

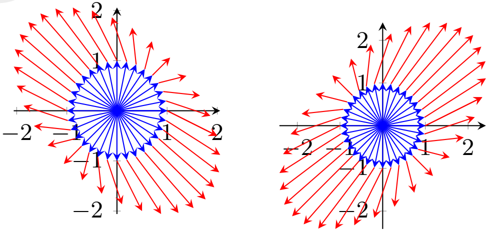
$$A = \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix}, \quad \text{has inverse } A^{-1} = \begin{bmatrix} \frac{4}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{4}{3} \end{bmatrix}.$$

Let's determine the eigenvalues of this inverse. Its characteristic equation (defined in Procedure 4.1.23) is

$$\det(A^{-1} - \lambda I) = \begin{vmatrix} \frac{4}{3} - \lambda & \frac{2}{3} \\ \frac{2}{3} & \frac{4}{3} - \lambda \end{vmatrix} = (\tfrac{4}{3} - \lambda)^2 - \tfrac{4}{9} = 0\,.$$

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 that 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 right-hand pictures illustrates the reciprocal relation graphically: the left picture shows $A\boldsymbol{x}$ for various $\boldsymbol{x}$, the right picture shows $A^{-1}\boldsymbol{x}$. 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 that are not eigenvectors, the relationship between $A\boldsymbol{x}$ and $A^{-1}\boldsymbol{x}$ is somewhat obscure.



□

Theorem 4.2.4 *Let A be a square matrix with eigenvalue λ and corresponding eigenvector $\boldsymbol{x}$.*

(a) For every positive integer n, λ^n is an eigenvalue of A^n with corresponding eigenvector $\boldsymbol{x}$.

(b) If A is invertible, then $1/\lambda$ is an eigenvalue of A^{-1} with corresponding eigenvector $\boldsymbol{x}$.

(c) If A is invertible, then for every integer n (including negative n), λ^n is an eigenvalue of A^n with corresponding eigenvector $\boldsymbol{x}$.

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 that 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^3 (b) A^{-2} (c) A^2 (d) A^{-1}

- Further, which of the above matrices has eigenvalue $1/9$?

Example 4.2.7 Consider the matrix

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

You are given that this matrix has eigenvalues 2, 1, and -1 with corresponding eigenvectors $(1, 1, 1)$, $(-1, 0, 1)$, and $(1, -2, 1)$ respectively. Confirm that matrix A^2 has eigenvalues that are these squared, and corresponding to the same eigenvectors. Given the inverse

$$A^{-1}=\begin{bmatrix}\frac12&\frac12&-\frac12\\ \frac12&-\frac12&\frac12\\ -\frac12&\frac12&\frac12\end{bmatrix}$$

confirm that its eigenvalues are the reciprocals of those of A, and for corresponding eigenvectors. □

Example 4.2.8 *(long-term age-structure)* Recall that Example 3.1.9 introduced how to use a Leslie matrix to predict the future population of an animal. In the example, letting $\boldsymbol{x}=(x_1,x_2,x_3)$ be the current number of pups, juveniles, and mature females respectively, then for the Leslie matrix

$$L=\begin{bmatrix}0&0&4\\ \frac12&0&0\\ 0&\frac13&\frac13\end{bmatrix}$$

the predicted population number after a year is $\boldsymbol{x}' = L\boldsymbol{x}$, after two years is $\boldsymbol{x}'' = L\boldsymbol{x}' = L^2\boldsymbol{x}$, and so on. Predict what happens after many generations: does the population die out? grow? oscillate? $\square$

4.2.2 Symmetric matrices are orthogonally diagonalizable

General real matrices may have non-real complex valued eigenvalues (as in Examples 4.2.8 and 4.1.28). 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 marvellously often reflects the physical reality of many applications.

To establish the reality of eigenvalues (Theorem 4.2.9), we have to eliminate the possibility that they are complex valued with a nonzero imaginary part. 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 $z = a + b\,\mathrm{i}$ has a complex conjugate $\bar{z} = a - b\,\mathrm{i}$ (denoted by the overbar, and similarly for vectors), 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, every eigenvalue of A is real valued.*

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 that Example 4.1.27 found the 3×3 symmetric matrix

$$\begin{bmatrix} -2 & 0 & -6 \\ 0 & 4 & 6 \\ -6 & 6 & -9 \end{bmatrix}$$

has eigenspaces $\mathbb{E}_0 = \operatorname{span}\{(-6, -3, 2)\}$, $\mathbb{E}_7 = \operatorname{span}\{(-2, 6, 3)\}$ and $\mathbb{E}_{-14} = \operatorname{span}\{(3, -2, 6)\}$. These eigenspaces are orthogonal as evidenced by the dot products of the basis vectors in each span:

$$\mathbb{E}_0, \mathbb{E}_7, \quad (-6, -3, 2) \cdot (-2, 6, 3) = 12 - 18 + 6 = 0\,;$$

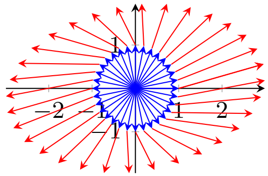
$$\begin{aligned}
&\mathbb{E}_7, \mathbb{E}_{-14}, \quad (-2, 6, 3) \cdot (3, -2, 6) = -6 - 12 + 18 = 0\,; \\
&\mathbb{E}_{-14}, \mathbb{E}_0, \quad (3, -2, 6) \cdot (-6, -3, 2) = -18 + 6 + 12 = 0\,.
\end{aligned}$$

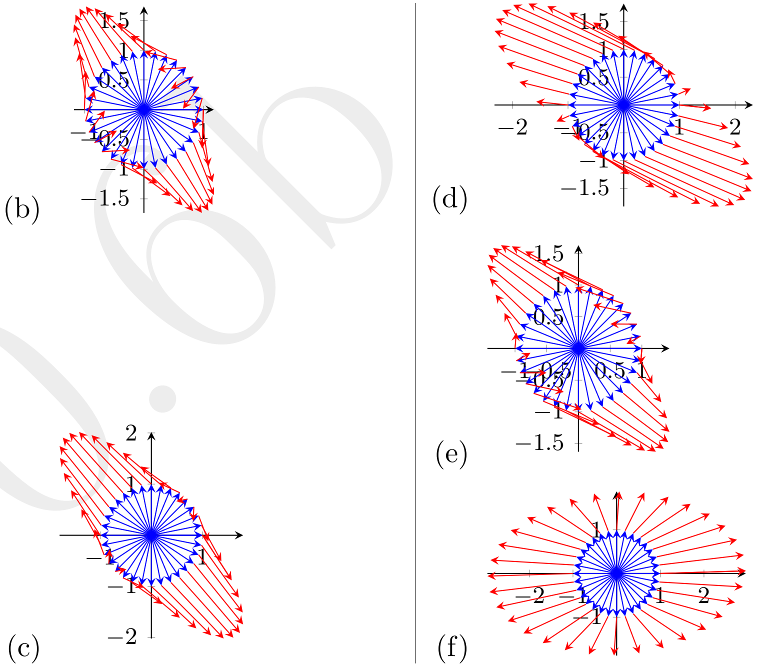
□

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 show (unit) vectors $\boldsymbol{x}$ (blue), and for some matrix A (different for different plots) the corresponding vectors $A\boldsymbol{x}$ (red) adjoined. By estimating eigenvectors, determine which cases *cannot* be the plot of a real symmetric matrix.

(a)




(b)
1.5
1
0.5
−0.5
−1
−1.5
(c)
2
1
−1
−2
(d)
1.5
1
0.5
−2
−1
1
2
−0.5
−1.5
(e)
1.5
1
0.5
0.5
1
−1
−1.5
(f)
1
−2
1
2
−1

□

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 that the eigenvectors are orthogonal for the symmetric matrix.

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

(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 2×2 or 3×3 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 that there are at most n distinct eigenvalues (here for symmetric matrices, but Theorem 7.1.1 establishes that 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, $A = USV^{\mathsf T}$, 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 $A = USV^{\mathsf T}$. If all the singular values are distinct or zero, $\sigma_1 > \cdots > \sigma_r > \sigma_{r+1} = \cdots = \sigma_n = 0$, then $\boldsymbol{v}_j$ is an eigenvector of A corresponding to an eigenvalue of either $\lambda_j = +\sigma_j$ or $\lambda_j = -\sigma_j$ (not both, except for the trivial case of $\sigma_j = 0$).*

If nonzero singular values are duplicated, then one can always choose an SVD such that 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 $A = USV^{\mathrm{T}}$. But specifically for symmetric A, the proof of the previous Theorem 4.2.16 identified that the columns of US, $\sigma_j \boldsymbol{u}_j$, are generally the same as $\lambda_j \boldsymbol{v}_j$ and hence are the columns of VD where $D = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. In which case the SVD becomes $A = VDV^{\mathrm{T}}$. This form of an SVD is intimately connected to the following definition.

Definition 4.2.17 *A real square matrix A is **orthogonally diagonalizable** if there exists an orthogonal matrix V and a diagonal matrix D such that $V^{\mathrm{T}}AV = D$, equivalently $AV = VD$, equivalently $A = VDV^{\mathrm{T}}$ is a factorization of A.*

The equivalences in this definition arise immediately from the orthogonality of matrix V (Definition 3.2.43): pre-multiplying $V^{\mathrm{T}}AV = D$ by V gives $VV^{\mathrm{T}}AV = AV = VD$; and so on.

Example 4.2.18

(a) Recall from Example 4.2.13 that the symmetric matrix $A = \begin{bmatrix} 1 & 3/2 \\ 3/2 & -3 \end{bmatrix}$ has eigenvalues $\lambda = -\frac{7}{2}, \frac{3}{2}$ with corresponding orthogonal eigenvectors $(1,-3)$ and $(3,1)$. Normalize these eigenvectors to unit length as the columns of the orthogonal matrix

$$V=\begin{bmatrix}\frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}}\\ -\frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}}\end{bmatrix}=\frac{1}{\sqrt{10}}\begin{bmatrix}1 & 3\\ -3 & 1\end{bmatrix} \quad \text{then}$$

$$\begin{aligned} V^{\mathrm{T}}AV &= \frac{1}{\sqrt{10}}\begin{bmatrix}1 & -3\\ 3 & 1\end{bmatrix}\begin{bmatrix}1 & \frac{3}{2}\\ \frac{3}{2} & -3\end{bmatrix}\frac{1}{\sqrt{10}}\begin{bmatrix}1 & 3\\ -3 & 1\end{bmatrix}\\ &= \frac{1}{10}\begin{bmatrix}-\frac{7}{2} & \frac{21}{2}\\ \frac{9}{2} & \frac{3}{2}\end{bmatrix}\begin{bmatrix}1 & 3\\ -3 & 1\end{bmatrix}=\frac{1}{10}\begin{bmatrix}-35 & 0\\ 0 & 15\end{bmatrix}=\begin{bmatrix}-\frac{7}{2} & 0\\ 0 & \frac{3}{2}\end{bmatrix}. \end{aligned}$$

Hence this matrix is orthogonally diagonalizable.

(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 diagonalizable.

□

These examples of orthogonal diagonalization 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 diagonalizable, and vice versa. That is, eigenvectors of a matrix form an orthogonal set if and only if the matrix is symmetric.

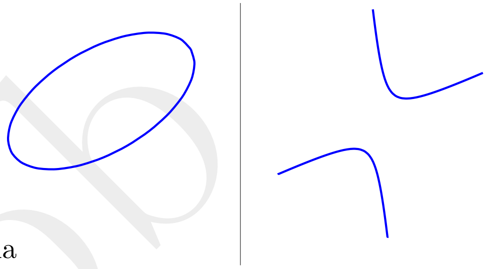
Theorem 4.2.19 (spectral) *For every real square matrix A, matrix A is symmetric iff it is orthogonally diagonalizable.*

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

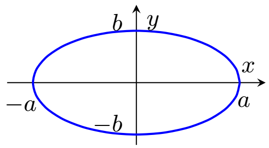
Example 4.2.20 illustrated an ellipse and a hyperbola. These curves are examples of the so-called **conic sections**, which arise as solutions of the quadratic equation in two variables, say x and y,

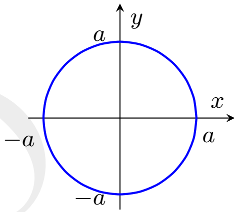
$$ax^2 + bxy + cy^2 + dx + ey + f = 0 \tag{4.2}$$

(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 recognized canonical forms.

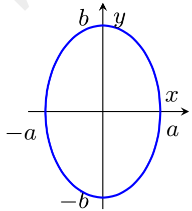
Ellipse or circle : $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$

- ellipse $a > b$





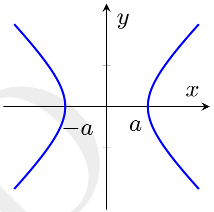
- the circle $a = b$



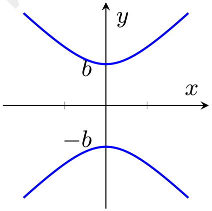
- ellipse $a < b$

Hyperbola : $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$ or $-\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$

- $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$

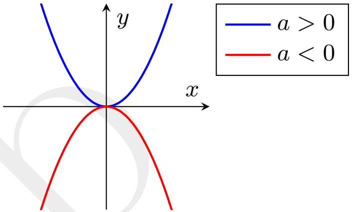


- $-\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$

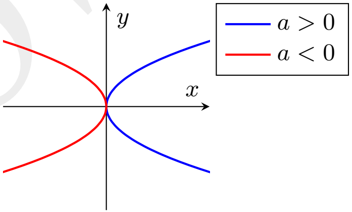


Parabola : $y = ax^2$ or $x = ay^2$

- $y = ax^2$



- $x = ay^2$



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

$$x^2 + 3xy - 3y^2 - \tfrac{1}{2} = 0\,.$$

(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

$$x^2 - xy + y^2 + \tfrac{5}{2\sqrt{2}}x - \tfrac{7}{2\sqrt{2}}y + \tfrac{1}{8} = 0\,.$$

□

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 optimization.

Definition 4.2.24 *A* ***quadratic form*** *in variables $\boldsymbol{x} \in \mathbb{R}^n$ is a function $q : \mathbb{R}^n \to \mathbb{R}$ that may be written as $q(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}} A \boldsymbol{x}$ 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 $\boldsymbol{x} \in \mathbb{R}^n$ consider

$$\boldsymbol{x} \cdot \boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}} \boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}} I_n \boldsymbol{x} \,,$$

which is the quadratic form associated with the identity matrix I_n.

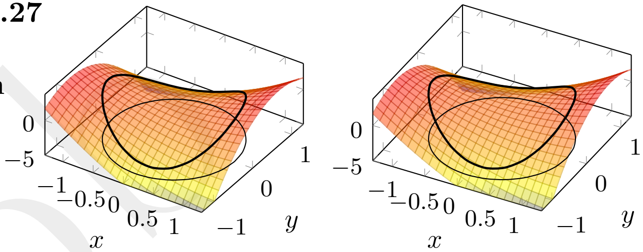
(b) Example 4.2.22 found the hyperbola satisfying equation $x^2 + 3xy - 3y^2 - \frac{1}{2} = 0$. This equation may be written in terms of a quadratic form as $\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x} - \frac{1}{2} = 0$ for vector $\boldsymbol{x} = (x, y)$ and symmetric matrix $A = \begin{bmatrix} 1 & 3/2 \\ 3/2 & -3 \end{bmatrix}$.

(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 $\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x}$ for vector $\boldsymbol{x} = (x, y)$ and symmetric matrix $A = \begin{bmatrix} 1 & -1/2 \\ -1/2 & 1 \end{bmatrix}$.

□

Theorem 4.2.26 (principal axes) *For every quadratic form, there exists an orthogonal coordinate system that diagonalizes the quadratic form. Specifically, for the quadratic form $\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x}$ find the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ and an orthonormal set of eigenvectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ of symmetric A, and then in the new coordinate system $(y_1, y_2, \ldots, y_n)$ with unit vectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ the quadratic form has the **canonical form** $\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$.*

Example 4.2.27 Consider the quadratic form $f(x,y) = x^2+3xy-3y^2$. That is, consider $f(\boldsymbol{x}) = \boldsymbol{x}^{\mathsf T}A\boldsymbol{x}$



for $\boldsymbol{x} = (x,y)$ and matrix $A = \begin{bmatrix} 1 & 3/2 \\ 3/2 & -3 \end{bmatrix}$. The stereo illustration to the right shows 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, $f(\boldsymbol{x})$ for unit vectors $\boldsymbol{x}$. Find the maxima and minima of f on this unit circle (for unit vectors $\boldsymbol{x}$). Relate to the eigenvalues of Example 4.2.13.

□

Theorem 4.2.28 *Let A be an $n \times n$ symmetric matrix with eigenvalues $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$ (sorted). Then for every unit vector $\boldsymbol{x} \in \mathbb{R}^n$ (that is, $|\boldsymbol{x}| = 1$), the quadratic form $\boldsymbol{x}^{\mathrm{T}} A\boldsymbol{x}$ has the following properties:*

(a) $\lambda_1 \le \boldsymbol{x}^{\mathrm{T}} A\boldsymbol{x} \le \lambda_n$ *;*

(b) the minimum of $\boldsymbol{x}^{\mathrm{T}} A\boldsymbol{x}$ is λ_1, and occurs when $\boldsymbol{x}$ is a unit eigenvector corresponding to λ_1;

(c) the maximum of $\boldsymbol{x}^{\mathrm{T}} A\boldsymbol{x}$ is λ_n, and occurs when $\boldsymbol{x}$ is a unit eigenvector corresponding to λ_n.

Activity 4.2.29 Recall that Example 4.1.27 found that the 3×3 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 $\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x}$ over unit vectors $\boldsymbol{x}$?

(a) 0 (b) -14 (c) 14 (d) 7

- Further, what is the minimum of the quadratic form $\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x}$ over unit vectors $\boldsymbol{x}$?

5 Approximate matrices

Chapter contents

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).

This chapter could be studied any time after Chapter 3 to help the transition to more abstract linear algebra. It also is good revision of the SVD, rank, orthogonality, and so on.

5.1 Measure changes to matrices

Section contents

5.1.1 Compress images optimally

Photographs and other images require 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 approximate images by matrices of various ranks. Recall that a matrix of rank k (Definition 3.3.19) means that the matrix has precisely k nonzero singular values, that is, an $m \times n$ matrix

$$\begin{aligned} A &= USV^{\mathrm{T}} \\ &= \begin{bmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_k & \cdots & \boldsymbol{u}_m \end{bmatrix} \left[\begin{array}{ccc|c} \sigma_1 & \cdots & 0 & \\ \vdots & \ddots & \vdots & O_{k\times(n-k)} \\ 0 & \cdots & \sigma_k & \\ & & & \\ & O_{(m-k)\times k} & & O_{(m-k)\times(n-k)} \end{array}\right] V^{\mathrm{T}} \end{aligned}$$

(then multiplying the form of the first two matrices)

$$= \begin{bmatrix} \sigma_1 \boldsymbol{u}_1 & \cdots & \sigma_k \boldsymbol{u}_k & O_{m\times(n-k)} \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{v}_k^{\mathrm{T}} \\ \vdots \\ \boldsymbol{v}_n^{\mathrm{T}} \end{bmatrix}$$

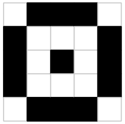
(then multiplying the form of these two matrices)

$$= \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^{\mathrm{T}} + \cdots + \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^{\mathrm{T}} .$$

This last sum precisely constructs matrix A. Further, when the rank k is low compared to size m and n, this last sum *has relatively few components.*

Example 5.1.1 Invent and write down a rank three representation of the following 5×5 'bull's eye' matrix (illustrated to the right)

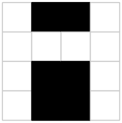
$$A = \begin{bmatrix} 0&1&1&1&0 \\ 1&0&0&0&1 \\ 1&0&1&0&1 \\ 1&0&0&0&1 \\ 0&1&1&1&0 \end{bmatrix} .$$



□

Activity 5.1.2 Which pair of vectors gives a rank one representation, $\boldsymbol{u}\boldsymbol{v}^{\mathrm{T}}$ of the matrix

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



(a) $\boldsymbol{u} = (1\,,1\,,0\,,1)$, $\boldsymbol{v} = (0\,,1\,,1\,,0)$

(b) $\boldsymbol{u} = (1\,,0\,,1\,,1)$, $\boldsymbol{v} = (0\,,1\,,1\,,0)$

(c) $\boldsymbol{u} = (0\,,1\,,1\,,0)$, $\boldsymbol{v} = (1\,,1\,,0\,,1)$

(d) $\boldsymbol{u} = (0\,,1\,,1\,,0)$, $\boldsymbol{v} = (1\,,0\,,1\,,1)$

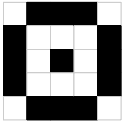
Procedure 5.1.3 (approximate images) *Consider any image stored as scalars in an $m \times n$ matrix A.*

1. *Compute an* SVD $A = USV^{\mathrm{T}}$ *with* `[U,S,V]=svd(A)`.
2. *Choose a desired rank k based upon the singular values (Theorem 5.1.16): typically there are k 'large' singular values and all 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)*

$$\begin{aligned} A_k &:= \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^{\mathrm{T}} + \cdots + \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^{\mathrm{T}} \\ &= \texttt{U(:,1:k)*S(1:k,1:k)*V(:,1:k)'} \end{aligned}$$

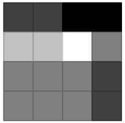
Example 5.1.4 Use Procedure 5.1.3 to find the 'best' rank two matrix, and also the 'best' rank three matrix, to approximate the 'bull's eye' image matrix (illustrated to the right)

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



□

Activity 5.1.5 Consider the image shown to the right. The image has matrix with SVD given below (2 d.p.). What rank representation exactly reproduces the matrix/image?



(a) 1 (b) 4 (c) 2 (d) 3

```
U =
  -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
S =
   2.45      0      0      0
      0   0.37      0      0
      0      0   0.00      0
      0      0      0   0.00
V =
  -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
```

Example 5.1.6 to the left is a 292×277 image of Blaise Pascal (French mathematician, physicist, inventor, and writer, 1623–62). The image is coded as 80 884 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. $\square$

Table 5.1: As well as the Matlab*/Octave commands and operations listed in Tables 1.2, 2.3, 3.1 to 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* `v` *computes the length, or magnitude,* $\sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$.
- `scatter(x,y,[],c)` *draws a 2D scatter plot of points with coordinates in vectors* `x` *and* `y`*, each point with a colour determined by the corresponding entry of vector* `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*

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 of distances between matrices.

In linear algebra we use a pair of double vertical bars, $\|\cdot\|$, to denote the magnitude of a matrix. The double vertical bars 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)*

$$\|A\| := \max_{|\boldsymbol{x}|=1} |A\boldsymbol{x}|, \qquad \text{equivalently } \|A\| = \sigma_1 \tag{5.1}$$

the largest singular value of the matrix A.

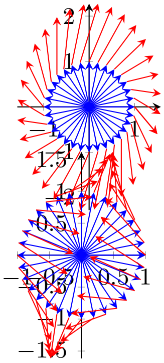
The equivalence, that $\max_{|\boldsymbol{x}|=1}|A\boldsymbol{x}| = \sigma_1$, is due to the definition of the largest singular value in the proof of the existence of an SVD (Section 3.3.3).

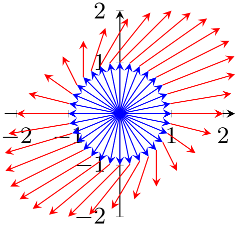
Example 5.1.8 The two following 2×2 matrices have the product $A\boldsymbol{x}$ plotted (red), adjoined to $\boldsymbol{x}$ (blue), for a complete range of unit vectors $\boldsymbol{x}$ (as in Section 4.1 for eigenvectors). From Definition 5.1.7, the norm of the matrix A is then the length of the longest such plotted $A\boldsymbol{x}$ (this norm is a measure of the magnitude of the matrix). (The MATLAB function `eigshow(A)` provides an interactive alternative to such static views.)

For each matrix, use the plot to roughly estimate their norm.

(a) $A = \begin{bmatrix} 0.5 & 0.5 \\ -0.6 & 1.2 \end{bmatrix}$

(b) $B = \begin{bmatrix} -0.7 & 0.4 \\ 0.6 & 0.5 \end{bmatrix}$





Example 5.1.9 Consider the 2×2 matrix $A = \left[\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right]$. Algebraically explore products $A\boldsymbol{x}$ for unit vectors $\boldsymbol{x}$, as illustrated to the right, and then find the matrix norm $\|A\|$.

- The standard unit vector $\boldsymbol{e}_2 = (0,1)$ has $|\boldsymbol{e}_2| = 1$ and $A\boldsymbol{e}_2 = (1,1)$ has length $|A\boldsymbol{e}_2| = \sqrt{2}$. Since the matrix norm is the maximum of all possible $|A\boldsymbol{x}|$, so $\|A\| \geq |A\boldsymbol{e}_2| = \sqrt{2} \approx 1.41$.

- Another unit vector is $\boldsymbol{x} = (\frac{3}{5}, \frac{4}{5})$. Here $A\boldsymbol{x} = (\frac{7}{5}, \frac{4}{5})$ has length $\sqrt{49+16}/5 = \sqrt{65}/5 \approx 1.61$. Hence the matrix norm $\|A\| \geq |A\boldsymbol{x}| \approx 1.61$.

- To systematically find the norm, recall that all unit vectors in 2D are of the form $\boldsymbol{x} = (\cos t, \sin t)$. Then

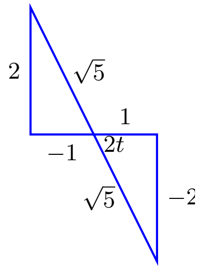
$$\begin{aligned}|A\boldsymbol{x}|^2 &= |(\cos t + \sin t, \sin t)|^2 \\ &= (\cos t + \sin t)^2 + \sin^2 t \\ &= \cos^2 t + 2\cos t\, \sin t + \sin^2 t + \sin^2 t\end{aligned}$$

$$= \tfrac{3}{2} + \sin 2t - \tfrac{1}{2}\cos 2t\,.$$

This length (squared) is maximized (and minimized) for some t determined by calculus. Differentiating with respect to t leads to

$$\frac{d|A\boldsymbol{x}|^2}{dt} = 2\cos 2t + \sin 2t \ = 0 \quad \text{for stationary points.}$$

Rearranging this equation determines that we require $\tan 2t = -2$. The right-angle triangles, drawn to the right, illustrate that these stationary points of $|A\boldsymbol{x}|^2$ 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



$$\begin{aligned}|A\boldsymbol{x}|^2 &= \tfrac{3}{2} + \sin 2t - \tfrac{1}{2}\cos 2t = \tfrac{3}{2} \mp \tfrac{2}{\sqrt{5}} \mp \tfrac{1}{2}\tfrac{1}{\sqrt{5}} \\ &= \tfrac{1}{2}(3 \mp \sqrt{5}) = \left(\frac{1 \mp \sqrt{5}}{2}\right)^2.\end{aligned}$$

The plus alternative is the larger, so gives the maximum, hence

$$\|A\| = \max_{|\boldsymbol{x}|=1} |A\boldsymbol{x}| = \frac{1+\sqrt{5}}{2} = 1.6180\,.$$

- 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 3×3 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) $\|A\| \geq 11.7$ (b) $\|A\| \geq 3.9$ (c) $\|A\| \geq 1.7$ (d) $\|A\| \geq 2.3$

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 4×6 matrix $B = \begin{bmatrix} 0 & -2 & -1 & -4 & -5 & 0 \\ 2 & 0 & 1 & -2 & -6 & -2 \\ -2 & 0 & 4 & 2 & 3 & -3 \\ 1 & 2 & -4 & 2 & 1 & 3 \end{bmatrix}$

$\square$

The Definition 5.1.7 of the magnitude/norm of a matrix may appear a little strange. But, in addition to some other marvellously useful properties, the norm nonetheless has all the familiar properties of a magnitude/length. Recall from Chapter 1 that for vectors:

- $|\boldsymbol{v}| = 0$ if and only if $\boldsymbol{v} = \boldsymbol{0}$ (Theorem 1.1.13);
- $|\boldsymbol{u} \pm \boldsymbol{v}| \leq |\boldsymbol{u}| + |\boldsymbol{v}|$ (the triangle inequality of Theorem 1.3.17);
- $|t\boldsymbol{v}| = |t| \cdot |\boldsymbol{v}|$ (Theorem 1.3.17).

Analogous 'magnitude' properties hold for the matrix norm as established in the next theorem.

Theorem 5.1.12 (matrix 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 1.3.17(c));*

(d) $\|tA\| = |t| \cdot \|A\|$;

(e) $\|A\| = \|A^{\mathrm{T}}\|$;

(f) $\|Q_m A\| = \|A\| = \|AQ_n\|$ *for every* $m \times m$ *orthogonal matrix* Q_m *and every* $n \times n$ *orthogonal matrix* Q_n;

(g) $|A\boldsymbol{x}| \leq \|A\| \cdot |\boldsymbol{x}|$ *for all* $\boldsymbol{x} \in \mathbb{R}^n$, *is like a Cauchy–Schwarz inequality (Theorem 1.3.17(b)), as is the next;*

(h) $\|AB\| \leq \|A\|\|B\|$ *for every* $n \times p$ *matrix* B.

Since the matrix norm has these 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) Find the distances between the matrix

$$A = \begin{bmatrix} 10 & 2 \\ 5 & 11 \end{bmatrix} \quad \text{and } B = \begin{bmatrix} 4 & -4 \\ -3 & 3 \end{bmatrix}, \quad \text{and } A_1 = \begin{bmatrix} 6 & 6 \\ 8 & 8 \end{bmatrix}.$$

Recall from Example 3.3.2 that the matrix A has an SVD of

$$USV^{\mathrm{T}} = \begin{bmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{bmatrix} \begin{bmatrix} 10\sqrt{2} & 0 \\ 0 & 5\sqrt{2} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}^{\mathrm{T}}.$$

i. Find $\|A - B\|$ for the rank one matrix

$$B = \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^{\mathrm{T}} = 5\sqrt{2} \begin{bmatrix} -\frac{4}{5} \\ \frac{3}{5} \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 4 & -4 \\ -3 & 3 \end{bmatrix}.$$

ii. Find $\|A - A_1\|$ for the rank one matrix

$$A_1 = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} = 10\sqrt{2} \begin{bmatrix} \frac{3}{5} \\ \frac{4}{5} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 6 & 6 \\ 8 & 8 \end{bmatrix}.$$

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) $\begin{bmatrix} 10 & -1 \\ 1 & 6 \end{bmatrix}$ (b) $\begin{bmatrix} 8 & -2 \\ 2 & 4 \end{bmatrix}$ (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 'bull's 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 $\|A - A_2\|$ and $\|A - A_3\|$. □

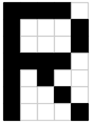
Theorem 5.1.16 (Eckart–Young) *Let A be any $m \times n$ matrix of* rank r *with* SVD $A = USV^{\mathrm{T}}$*. Then for every $k < r$ the matrix*

$$A_k := US_kV^{\mathrm{T}} = \sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + \sigma_2\boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}} + \cdots + \sigma_k\boldsymbol{u}_k\boldsymbol{v}_k^{\mathrm{T}} \tag{5.2}$$

where $S_k :=$ diag$(\sigma_1,\sigma_2,\ldots,\sigma_k,0,\ldots,0)$, 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 digital displays with low resolution, letters and numbers are displayed with noticeable pixel patterns: for example, the letter `R` is pixellated to the right. 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 a relative error of just a little less than 25%?

(a) 4 (b) 3 (c) 2 (d) 1

Example 5.1.19 Recall that Example 5.1.6 approximated the image of Blaise Pascal with various rank k approximations, and that these approximations came from an SVD of the image. Let the image be denoted by matrix A. From **??** the largest singular value of the image is $\|A\| = \sigma_1 \approx 37\,000$.

- From Theorem 5.1.16, the rank 3 approximation in **??** is a distance $\|A - A_3\| = \sigma_4 \approx 3\,200$ (from **??**) away from the image. That is, image A_3 has a relative error roughly $3\,200/37\,000 \approx 9\%$.
- From Theorem 5.1.16, the rank 10 approximation in **??** is a distance $\|A - A_{10}\| = \sigma_{11} \approx 1\,500$ (from **??**) away from the image. That is, image A_{10} has a relative error roughly

$1\,500/37\,000 \approx 4\%$.

- From Theorem 5.1.16, the rank 30 approximation in **??** is a distance $\|A - A_{30}\| = \sigma_{31} \approx 600$ (from **??**) away from the image. That is, image A_{30} has a relative error roughly $600/37\,000 \approx 2\%$.

$\square$

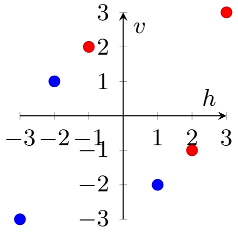
5.1.3 Principal component analysis

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. Consequently, much the same technique has alternative names such as the Karhunen–Loève decomposition, 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:

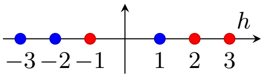
h	v	colour
-3	-3	blue
-2	1	blue
1	-2	blue
-1	2	red
2	-1	red
3	3	red



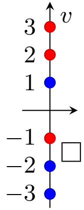
The item properties/attributes are the points (h, v) in 2D as illustrated to the above-right (h for horizontal, and v for vertical). But we humans always prefer simple one-dimensional summaries: we do it all the time when we rank sport teams, schools, web pages, political parties, and so on.

Challenge: is there a one-dimensional summary of these six items' 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 to the right;



- similarly, using v alone leads to a 1D view where the red and the blue are intermingled as shown to the right.



□

Although this Example 5.1.20 is just a toy to illustrate concepts, the above steps generalize 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 lists part of Edgar Anderson's data on the lengths and widths of sepals and petals of iris flowers. There are three species of irises in the data

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. http://archive.ics.uci.edu/ml/datasets/iris *gives the full dataset (Dua & Graff 2019).*

Sepal length	*Sepal width*	*Petal length*	*Petal width*	*Species*
4.9	3.0	1.4	0.2	
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	*Setosa*
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	
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	*Versicolor*

(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 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.

$\square$

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 $\begin{bmatrix}4.9 & 3.0 & 1.4 & 0.2\end{bmatrix}$ (cm) corresponding to the sepal

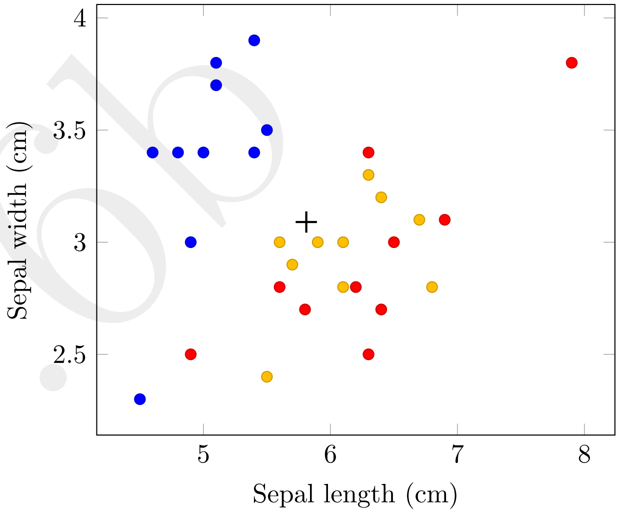


Figure 5.1: Scatter plot of sepal widths versus lengths for Edgar Anderson's iris data of Table 5.2: blue, Setosa; brown, Versicolor; red, Virginia. The black "+" marks the mean sepal width and length.

length and width, and the petal length and width, respectively. Similarly, the last Virginia iris flower has row vector of attributes of $\begin{bmatrix}46.3 & 2.5 & 5.0 & 1.9\end{bmatrix}$ (cm), and the mean vector is the row vector $\begin{bmatrix}5.81 & 3.09 & 3.69 & 1.22\end{bmatrix}$ (cm). The reason for this mathematical transposition is that throughout science and engineering, data 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 an $m\times n$ data matrix A (usually with zero mean when averaged over all rows) with* SVD *$A=USV^{\mathrm T}$, then the jth column $\boldsymbol v_j$ of V is called the jth **principal vector** and the vector $\boldsymbol x_j := A\boldsymbol v_j$ is called the jth **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

$$A_2 = US_2V^{\mathrm{T}} = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^{\mathrm{T}} = (\sigma_1 \boldsymbol{u}_1)\boldsymbol{v}_1^{\mathrm{T}} + (\sigma_2 \boldsymbol{u}_2)\boldsymbol{v}_2^{\mathrm{T}}\,.$$

That is, in this best rank two approximation of the data, the row vector of attributes of the ith iris are the linear combination of row vectors $(\sigma_1 u_{i1})\boldsymbol{v}_1^{\mathrm{T}} + (\sigma_2 u_{i2})\boldsymbol{v}_2^{\mathrm{T}}$. The vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ 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 ith iris in the best 2D plot. Consequently, in Matlab/Octave we draw a scatter plot of the components of vectors $\boldsymbol{x} = \sigma_1 \boldsymbol{u}_1$ and $\boldsymbol{y} = \sigma_2 \boldsymbol{u}_2$ (**??**).

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, ... How many dimensions should you expect to be needed for a good view of the data?

(a) 3D (b) 4D (c) 2D (d) 1D

Example 5.1.25 *(wine recognition)* From the Dua & Graff (2019) 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 about 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 $\boldsymbol{a}_j=\boldsymbol{b}_j-\bar b_j$;*
   (b) *but ensure each column has the same 'physical dimensions', usually by dividing by the standard deviation s_j of each column, algebraically $\boldsymbol{a}_j=(\boldsymbol{b}_j-\bar b_j)/s_j$.*

Compute in Matlab*/Octave using auto-replication:*

```
A=(B-mean(B))./std(B)
```

3. *Economically compute an* svd *for the best rank k approximation to the scaled data matrix with* `[U,S,V]=svds(A,k)`.
4. *Then the jth column of* `V` *is the jth principal vector, and the principal components are the entries of the $m \times k$ matrix* `A*V`.

Courses on multivariate statistics prove that, for every (usually zero mean) data matrix A, the first k principal vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ are orthonormal vectors that *maximize the total variance* in the principal components $\boldsymbol{x}_j = A\boldsymbol{v}_j$; that is, that maximize $|\boldsymbol{x}_1|^2 + |\boldsymbol{x}_2|^2 + \cdots + |\boldsymbol{x}_k|^2$. Indeed, this maximization of the variance corresponds closely to the constructive proof of the existence of svds (Section 3.3.3) which successively maximizes $|A\boldsymbol{v}|$ subject to $\boldsymbol{v}$ 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 kD, then (roughly) it retains the most information possible in kD.

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)*

Information searches 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. The word vectors being derived via the singular value decomposition. This section introduces such 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 capitalized 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 summarize 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 for simplicity this example omits such once-used words.)

Follow the principal component analysis Procedure 5.1.26.

1. First find the set of words that are used more than once. Ignoring pluralization, they are, in alphabetical order, Application, Introduction, Method, System, Theory. The corresponding word vector for each book title is then the following:

   - $\boldsymbol{w}_1 = (0, 1, 1, 0, 0)$ *Introduction* to Finite and Spectral Element *Method*s using MATLAB
   - $\boldsymbol{w}_2 = (1, 0, 1, 1, 1)$ Iterative *Method*s for Linear *System*s: *Theory* and *Application*s
   - $\boldsymbol{w}_3 = (1, 1, 1, 1, 0)$ Singular Perturbations: *Introduction* to *System* Order Reduction *Method*s with *Application*s
   - $\boldsymbol{w}_4 = (0, 0, 1, 0, 0)$ Risk and Portfolio Analysis: Principles and *Method*s
   - $\boldsymbol{w}_5 = (1, 0, 0, 1, 1)$ Stochastic Chemical Kinetics: *Theory* and Mostly *System*s Biology *Application*s
   - $\boldsymbol{w}_6 = (0, 0, 0, 0, 1)$ Quantum *Theory* for Mathematicians

2. Second, form the data matrix with $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_6$ 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 =
  +0.52  -0.20
  +0.26  +0.52
  +0.50  +0.57
  +0.52  -0.20
  +0.37  -0.57
```

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 $V = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 \end{bmatrix}$ are the two orthonormal principal vectors:
   - the first $\boldsymbol{v}_1$, from its largest components, mainly identifies the overall direction of Application, Method, and System;
   - whereas the second $\boldsymbol{v}_2$, from its largest positive and negative components, mainly distinguishes Introduction and Method from Theory.

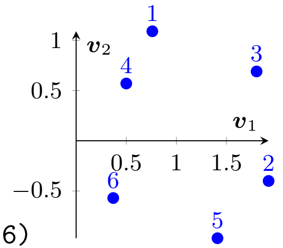
The corresponding principal components are the entries of the 6×2 matrix

$$AV = \begin{bmatrix} 0.76 & 1.09 \\ 1.92 & -0.40 \\ 1.80 & 0.69 \\ 0.50 & 0.57 \\ 1.41 & -0.97 \\ 0.37 & -0.57 \end{bmatrix}:$$

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 to the right. 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.)

$$A_2 = US_2V^{\mathrm{T}} = \begin{bmatrix} 0.18 & 0.77 & 1.01 & 0.18 & -0.33 \\ 1.08 & 0.29 & 0.74 & 1.08 & 0.95 \\ 0.80 & 0.82 & 1.30 & 0.80 & 0.28 \\ 0.15 & 0.43 & 0.58 & 0.15 & -0.14 \\ 0.93 & -0.14 & 0.16 & 0.93 & 1.08 \\ 0.31 & -0.20 & -0.14 & 0.31 & 0.46 \end{bmatrix}.$$

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 $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$.

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 summarizes and relates these titles.

Follow Procedure 5.1.26.

1. The significant (capitalized) words which appear more than once in these titles (ignoring pluralization) are the fourteen words, in alphabetical order,

   $$\begin{array}{l}\text{Algorithm, Application, Differential/tion,}\\ \text{Dynamic/al, Equation, Implementation,}\\ \text{Integral, Method, Nonlinear, Ordinary, Partial,}\\ \text{Problem, System, and Theory.}\end{array} \tag{5.3}$$

   With this dictionary of significant words, the titles have the following word vectors.

   - $\boldsymbol{w}_1 = (0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0)$ a Course on Integral Equations
   - $\boldsymbol{w}_2 = (1, 1, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1)$ Automatic Differentiation of Algorithms: Theory, Implementation, and Application
   - ...

- $\boldsymbol{w}_{14} = (0,1,0,0,0,0,1,0,0,0,0,0,0,1)$ the Double Mellin–Barnes Type Integrals and their Applications to Convolution Theory

2. Form the 14×14 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, principal 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 orthonormal 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 $\boldsymbol{v}_1$ of V, from its largest components, mainly identifies the two most common words of Differential and Equation.
   - The second column $\boldsymbol{v}_2$ of V, from its largest components, identifies books with Algorithms, Applications, Implementations, Problems, and Theory.
   - The third column $\boldsymbol{v}_3$ of V, from its largest components,

largely distinguishes Dynamics, Problems, and Systems, from Differential and Theory.

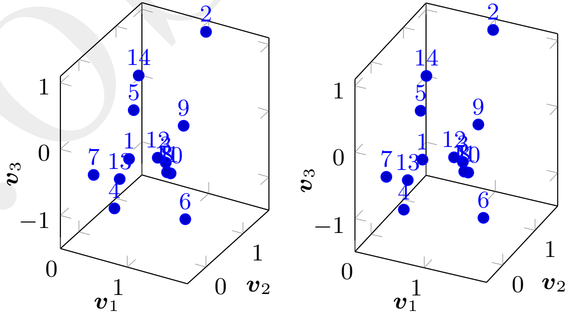
The corresponding principal components are the entries of the 14×3 matrix (2 d.p.)

$$AV = \begin{bmatrix} 0.70 & 0.09 & -0.25 \\ 1.02 & 1.59 & 1.00 \\ 1.46 & -0.29 & 0.10 \\ 0.16 & 0.67 & -1.44 \\ 0.16 & 1.19 & -0.22 \\ 1.78 & -0.34 & -0.64 \\ 0.22 & -0.00 & -0.58 \\ 1.48 & -0.29 & -0.04 \\ 1.45 & 0.21 & 0.40 \\ 1.56 & -0.34 & -0.01 \\ 1.48 & -0.29 & -0.04 \\ 1.29 & -0.20 & 0.08 \\ 0.10 & 0.92 & -1.14 \\ 0.29 & 1.09 & 0.39 \end{bmatrix}.$$

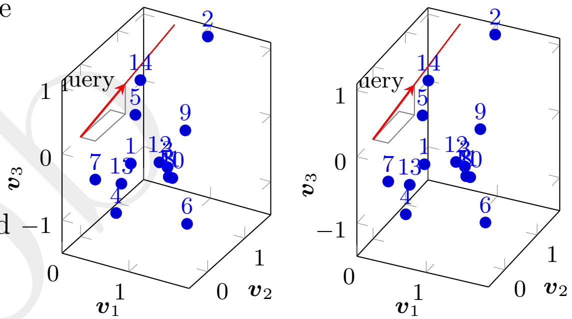
Each of the fourteen books is represented in 3D space by the corresponding row of these coordinates. Plot these books in Matlab/Octave with

```
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 $\boldsymbol{w} = (1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1)$. Project this query into the 3D space of principal components with the product $\boldsymbol{w}^{\mathrm{T}}V$ which evaluates to the query vector $\boldsymbol{q} = (0.22, 0.81, 0.46)$ whose direction is added to the picture as shown to the right.

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 each has both *Application* and *Theory* in its title. But the above plot also suggests that 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, in 3D, between this query vector and the book title 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° to the 3D query vector $\boldsymbol{q} = (0.22, 0.81, 0.46)$, of length $|\boldsymbol{q}| = 0.96$, because its cosine

$$\cos\theta = \frac{(1.02, 1.59, 1.00)\cdot\boldsymbol{q}}{2.14\cdot 0.96} = 0.97\,.$$

- Similarly, Book 14 has the principal component vector $(0.29, 1.09,$ which has length 1.20. Consequently, it is at small angle 10° to the 3D query vector $\boldsymbol{q} = (0.22, 0.81, 0.46)$ because its cosine

$$\cos\theta = \frac{(0.29, 1.09, 0.39)\cdot \boldsymbol{q}}{1.20\cdot 0.96} = 0.99\,.$$

- Whereas Book 5 has the principal component vector $(0.16, 1.19, -0$ which has length 1.22. Consequently, it is at moderate angle 40° to the 3D query vector $\boldsymbol{q} = (0.22, 0.81, 0.46)$ because its cosine

$$\cos\theta = \frac{(0.16, 1.19, -0.22)\cdot \boldsymbol{q}}{1.20\cdot 0.96} = 0.76\,.$$

  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 $62°$ 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 70 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 is part of the Opinosis Opinion/Review (Ganesan et al. 2010) in the UCI Machine Learning Repository (Dua & Graff 2019) [`https://archive.ics.uci.edu/ml/datasets/Opinosis+Opinion+%26frasl%3B+Review`].

- *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 bath-*

5.2 Regularize linear equations

Section contents

Singularity is almost invariably a clue.

Sherlock Holmes, in The Boscombe Valley Mystery, by Sir Arthur Conan Doyle, 1892

Often we need to approximate the matrix in a linear equation. Such approximation is especially likely when the matrix itself comes from experimental measurements and so has 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 two linear equations

$$0.5x + 0.3y = 1 \quad \text{and} \quad 1.1x + 0.7y = 2\,,$$

where all the coefficients on *both* of the left-hand sides and the right-hand sides are determined from experimental measurements. Suppose they are measured to experimental errors ± 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 ± 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) $(5.6\,,0.8)$
(d) $(6.6\,,-1.2)$

5.2.1 The SVD illuminates regularization

> 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 equations $A\boldsymbol{x} = \boldsymbol{b}$ arises from an experiment where both the $m \times n$ matrix A and the right-hand side vector $\boldsymbol{b}$ are subject to experimental error. Suppose the expected error in the matrix and vector entries are of magnitude ϵ. (Recall that (Theorem 3.3.30) the symbol ϵ is the Greek letter epsilon, and often denotes errors.)*

1. *When forming the matrix A and vector $\boldsymbol{b}$, scale the data so that*
   - *all $m \times n$ components in A have the same physical units, and they are of roughly the same magnitude; and*
   - *similarly for the m components of $\boldsymbol{b}$.*

Estimate the error ϵ corresponding to this matrix A.

2. *Compute an* SVD $A = USV^{\mathrm{T}}$.

3. *Choose 'rank' k to be the number of singular values bigger than the error ϵ; that is, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k > \epsilon > \sigma_{k+1} \geq \cdots \geq 0$. Then the rank k approximation to A is*
$$\begin{aligned} A_k &:= US_kV^{\mathrm{T}} \\ &= \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^{\mathrm{T}} + \cdots + \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^{\mathrm{T}} \\ &= \texttt{U(:,1:k)*S(1:k,1:k)*V(:,1:k)'}. \end{aligned}$$
But do not construct A_k, as we only need its SVD *to solve the system.*

4. *Solve the approximating linear equation $A_k\boldsymbol{x} = \boldsymbol{b}$ as in Theorems 3.5.8 and 3.5.13 (often as an inconsistent set of equations). Usually use the* SVD $A_k = US_kV^{\mathrm{T}}$.

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) 2 (b) 4 (c) 3 (d) 1

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, in future applications you will have to be aware of the freedom, and make a choice based on extra information. For two examples:

- in a CT-scan such as Example 3.5.17 one would usually prefer the greyest result in order to avoid diagnosing artifices;
- in the data mining task of fitting curves or surfaces to data, one would instead usually prefer a curve or surface with the fewest nonzero coefficients.

Such extra information from the application is essential.

Example 5.2.5 For the following matrix A and right-hand side vector $\boldsymbol{b}$, solve $A\boldsymbol{x} = \boldsymbol{b}$. But suppose the matrix entries come from experiments and are only known to within errors ± 0.05, solve $A'\boldsymbol{x}' = \boldsymbol{b}$ for some chosen 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.

$$A = \begin{bmatrix} -0.2 & -0.6 & 1.8 \\ 0.0 & 0.2 & -0.4 \\ -0.3 & 0.7 & 0.3 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} -0.5 \\ 0.1 \\ -0.2 \end{bmatrix}.$$

□

This example gives infinitely many solutions which are equally valid as far as the linear algebra is concerned. In such an example, more information from an application is needed to choose which to *prefer* among the infinity of solutions.

Example 5.2.6 Repeat Example 5.2.5 with matrix and right-hand side

$$A = \begin{bmatrix} -1.1 & 0.1 & 0.7 & -0.1 \\ 0.1 & -0.1 & 1.2 & -0.6 \\ 0.8 & -0.2 & 0.4 & -0.8 \\ 0.8 & 0.1 & -2.0 & 1.0 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} -1.1 \\ -0.1 \\ 1.1 \\ 0.8 \end{bmatrix}$$

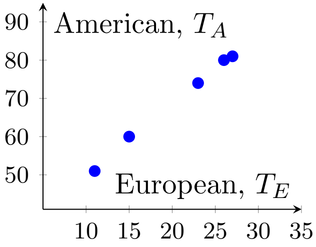
□

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 5.2.7 *(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 $A\boldsymbol{x} = \mathbf{1}$ where $\mathbf{1} := (1, 1, 1, 1, 1)$. □



Example 5.2.8 *(translating temperatures)* Recall that Example 2.2.12 attempts to fit a quartic polynomial to observations (plotted to the right) 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 that the data between temperatures reported by a European and an American are the following:

T_E	15	26	11	23	27
T_A	60	80	51	74	81

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

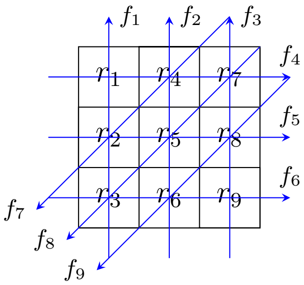
$$\begin{bmatrix} 1 & 15 & 225 & 3375 & 50625 \\ 1 & 26 & 676 & 17576 & 456976 \\ 1 & 11 & 121 & 1331 & 14641 \\ 1 & 23 & 529 & 12167 & 279841 \\ 1 & 27 & 729 & 19683 & 531441 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_3 \\ c_4 \\ c_5 \end{bmatrix} = \begin{bmatrix} 60 \\ 80 \\ 51 \\ 74 \\ 81 \end{bmatrix}.$$

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 $\pm 0.5°$.

□

Occam's razor: Non sunt multiplicanda entia sine necessitate [Entities must not be multiplied beyond necessity] *John Punch (1639)*

Example 5.2.9 Recall that **??** introduced extra 'diagonal' measurements into a 2D CT-scan. As shown to the right, the 2D region is divided into a 3×3 grid of nine blocks. Then measurements are 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 $\boldsymbol{f} = (0.048\,, 0.081\,, 0.042\,, 0.020\,, 0.106\,, 0.075\,, 0.177\,, 0.181\,, 0.105)$. Use an SVD to find the 'greyest' transmission factors consistent with the measurements and likely errors.



□

5.2.2 Tikhonov regularization explained

Regularization of poorly posed linear equations is a widely invoked practical necessity. Many people have invented alternative techniques. Many have independently re-invented techniques. Perhaps the most common is the so-called Tikhonov regularization. This section introduces and discusses Tikhonov regularization.

This optional subsection connects to much established practice that graduates may encounter.

> 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 5.2.10 *In seeking to solve the poorly posed system $A\boldsymbol{x} = \boldsymbol{b}$ for $m \times n$ matrix A, a **Tikhonov regularization** is the system $(A^{\mathrm{T}}A + \alpha^2 I_n)\boldsymbol{x} = A^{\mathrm{T}}\boldsymbol{b}$ for some chosen regularization parameter value $\alpha > 0$. (The greek letter α is 'alpha', and is different to the 'proportional to' symbol $\propto$.)*

Example 5.2.11 Use Tikhonov regularization to solve the system of Example 5.2.1:

$$0.5x + 0.3y = 1 \quad \text{and} \quad 1.1x + 0.7y = 2\,,$$

□

Activity 5.2.12 In the linear system for $\boldsymbol{x} = (x\,,y)$,

$$4x - y = -4 \quad \text{and} \quad -2x + y = 3\,,$$

the coefficients on the left-hand side of each equation are in error by about ± 0.3. Tikhonov regularization should solve which one of the following systems?

(a) $\begin{bmatrix} 20.1 & -6 \\ -6 & 2.1 \end{bmatrix}\boldsymbol{x} = \begin{bmatrix} -22 \\ 7 \end{bmatrix}$

(b) $\begin{bmatrix} 18.3 & -5 \\ -10 & 3.3 \end{bmatrix}\boldsymbol{x} = \begin{bmatrix} -19 \\ 11 \end{bmatrix}$

(c) $\begin{bmatrix} 18.1 & -5 \\ -10 & 3.1 \end{bmatrix}\boldsymbol{x} = \begin{bmatrix} -19 \\ 11 \end{bmatrix}$

(d) $\begin{bmatrix} 20.3 & -6 \\ -6 & 2.3 \end{bmatrix}\boldsymbol{x} = \begin{bmatrix} -22 \\ 7 \end{bmatrix}$

Do not apply Tikhonov regularization blindly as it does introduce biases. The following example illustrates the bias.

Example 5.2.13 Recall Example 3.5.1 at the start of Section 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

$$A x = \boldsymbol{b}\,, \quad \text{namely} \quad \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} x = \begin{bmatrix} 84.8 \\ 84.1 \\ 84.7 \\ 84.4 \end{bmatrix}.$$

A Tikhonov regularization of this inconsistent system is

$$\left(\begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \alpha^2 \right) x = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 84.8 \\ 84.1 \\ 84.7 \\ 84.4 \end{bmatrix}.$$

That is, $(4+\alpha^2)x = 338\,\text{kg}$ with solution $x = 338/(4+\alpha^2) = 84.5/(1+\alpha^2/4)\,\text{kg}$. Because of the division by $1+\alpha^2/4$, this Tikhonov answer is biased as it is systematically below the average $84.5\,\text{kg}$. For small Tikhonov parameter α the bias is small, but even so, such a bias in the methodology is unpleasant. □

Example 5.2.14 Use Tikhonov regularization to solve $A\boldsymbol{x} = \boldsymbol{b}$ for the matrix and vector of Example 5.2.5. □

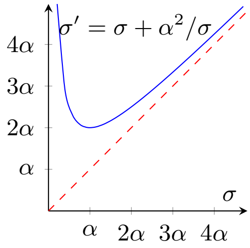
Although Definition 5.2.10 does not look like it, Tikhonov regularization relates directly to the SVD regularization of Section 5.2.1. The next theorem establishes the connection.

Theorem 5.2.15 (Tikhonov regularization) *Solving $A\boldsymbol{x} = \boldsymbol{b}$ by Tikhonov regularization, with parameter $\alpha > 0$, is equivalent to finding the smallest, least square, solution of the system $A'\boldsymbol{x} = \boldsymbol{b}$, where the matrix A' is obtained from A by replacing each of its nonzero singular values σ_i by $\sigma_i' := \sigma_i + \alpha^2/\sigma_i$.*

There is another reason to be careful when using Tikhonov regularization. Yes, it gives a nice, neat, unique solution, but 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 regularization parameter

- One strategy to choose the regularization parameter α is that the effective change in the matrix, from A to A', should be about the magnitude of errors expected in A. (This strategic choice is sometimes called the discrepancy principle (Kress 2015, §7).) Since changes in the matrix are largely measured by the singular values, we need to consider the relation between $\sigma' = \sigma + \alpha^2/\sigma$ and σ. From the graph to the right, the small singular values are changed by a lot, but these are the ones for which we want σ' large, in order to give a 'least square' approximation. Significantly, the marginal graph also shows that singular values larger than α change by less than α. Thus the parameter α should not be much larger than the expected error in the elements of the matrix A.



- Another consideration is the effect of regularization 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 $\sigma' \geq 2\alpha$. Thus, in the regularized system the condition number of the effective matrix A' is approximately $\sigma_1/(2\alpha)$. We need to choose the regularization parameter α large enough so that $\frac{\sigma_1}{2\alpha}\times$(relative error in $\boldsymbol{b}$) is an acceptable relative error in the solution $\boldsymbol{x}$ (Theorem 3.3.30). It is only when the regularization parameter α is big enough that the regularization is effective in finding a least square approximation.

6 Determinants distinguish matrices

Chapter contents

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 2×2 matrix $A=\begin{bmatrix} a&b\\c&d\end{bmatrix}$ has determinant $\det A=|A|=ad-bc$, and that the matrix A is invertible if and only if $\det A\neq 0$;
- a 3×3 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:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{matrix} a & b \\ d & e \\ g & h \end{matrix} \tag{6.1}$$

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.

6.1 Geometry underlies determinants

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 $A = \begin{bmatrix} 1/2 & 1 \\ 0 & 1 \end{bmatrix}$. Use matrix-vector 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-vector 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.

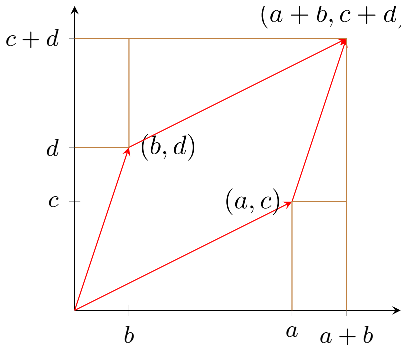
□

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

Example 6.1.4 Let the square matrix $C = \begin{bmatrix}2 & 0 & 0\\ 1 & 1 & 0\\ 0 & 0 & 3/2\end{bmatrix}$. Use matrix-vector 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.

$\square$

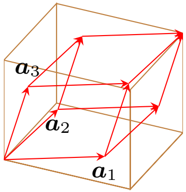


Determinants determine area transformation Consider multiplication by the general 2×2 matrix $A=\begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Under multiplication by this matrix A the unit square becomes the parallelogram, as illustrated to the right, with four corners at $(0,0)$, (a,c), (b,d), and $(a+b,c+d)$. Let's determine the area of the parallelogram by that of the containing rectangle (brown) less the two small rectangles and the four triangles. The two small rectangles have the same area, namely bc. The two triangles on the left and the right also have the same area, namely $\frac{1}{2}bd$. The two 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

$$\begin{aligned} \text{area} &= (a+b)(c+d) - 2\cdot bc - 2\cdot\frac{1}{2}bd - 2\cdot\frac{1}{2}ac \\ &= ac+ad+bc+bd-2bc-bd-ac \end{aligned}$$

$$= ad - bc = \det A\,.$$

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 3×3 matrices with transformations of volumes. Under multiplication by a 3×3 matrix $A = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \boldsymbol{a}_3 \end{bmatrix}$, the image of the unit cube is a parallelepiped with edges $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, and $\boldsymbol{a}_3$ as illustrated. By computing the volumes of various rectangular boxes, prisms, and tetrahedra, the volume of such a parallelepiped could be expressed as the 3×3 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 generalize a square and cube to n dimensions

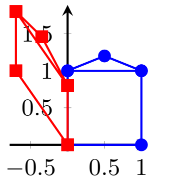
$(\mathbb{R}^n)$, n**D-volume** to generalize the notion of area and volume to n dimensions, and so on. When the dimension of the space is unspecified, then we may say **hyper-cube**, **hyper-volume**, and so on.

Definition 6.1.5 *Let A be an $n \times n$ square matrix, and let C be the unit nD-cube in $\mathbb{R}^n$. Transform the nD-cube C by $\boldsymbol{x} \mapsto A\boldsymbol{x}$ to its image C' in $\mathbb{R}^n$. Define the **determinant** of A, denoted either* $\det A$ *or sometimes $|A|$ such that:*

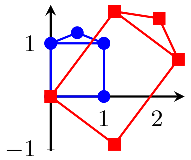
- *the magnitude $|\det A|$ is the nD-volume of C'; and*
- *the sign of* $\det A$ *to be negative iff the transformation reflects the orientation of the nD-cube.*

But what do we mean by "*orientation*" in this definition? Recall that complex-shaped objects come in two versions: humans have a left-hand and a right-hand; bolts have a left-handed thread or a right-handed thread; biological molecules are often either left-handed or right-handed (sometimes one is essential to life, and the other fatal). Such 'two-handedness' holds in every dimension (higher than 1D). Such left-hand and right-hand versions are the mirror image, the reflection, of each other. Here, a negative determinant indicates that the transformation of a geometric object changes a 'left-hand' version to a 'right-hand' version, or vice versa.

Example 6.1.6 Roughly estimate the determinant of the matrix that transforms the unit square to the parallelogram as shown to the right. □



Activity 6.1.7 Roughly estimate the determinant of the matrix that transforms the unit square to the rectangle as shown to the right.



(a) 4 (b) 3 (c) 2.5 (d) 2

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, $\det Q = \det(Q^{\mathrm{T}})$.

(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 either of two ways.

- An identity matrix is a diagonal matrix and hence its determinant is the product of the diagonal entries (Theorem 6.1.8(a)), here all ones.
- Alternatively, multiplication by the identity does not change the unit nD-cube and so does not change its nD-volume or its orientation (Definition 6.1.5). □

Activity 6.1.10 What is the determinant of $-I_n$?

(a) $+1$ for odd n, and -1 for even n
(b) -1
(c) $+1$
(d) -1 for odd n, and $+1$ for even n

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:

$$\begin{bmatrix} \frac{1}{3} & \frac{2}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} & \frac{2}{3} \end{bmatrix}, \quad \begin{bmatrix} -1 & 2 & -2 \\ -2 & -2 & -1 \\ 2 & -1 & -2 \end{bmatrix}, \quad \begin{bmatrix} \frac{1}{6} & -\frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \\ -\frac{1}{3} & \frac{1}{6} & \frac{1}{3} \end{bmatrix}.$$

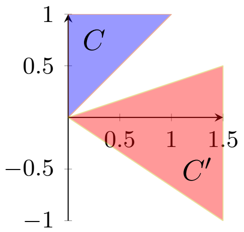
$\square$

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) 8 (b) -4 (c) -16 (d) 2

A consequence of Theorem 6.1.8(c) is that a determinant characterizes the transformation of any sized hyper-cube. Consider the transformation by a matrix A of an nD-cube of side length k $(k \geq 0)$, and hence of volume k^n. The nD-cube has edges $k\boldsymbol{e}_1, k\boldsymbol{e}_2, \ldots, k\boldsymbol{e}_n$. The transformation results in an nD-parallelepiped with edges $A(k\boldsymbol{e}_1), A(k\boldsymbol{e}_2), \ldots, A(k\boldsymbol{e}_n)$, which by commutativity and associativity (Theorem 3.1.26(d)) are the same edges as $(kA)\boldsymbol{e}_1$, $(kA)\boldsymbol{e}_2$, $\ldots$, $(kA)\boldsymbol{e}_n$. That is, the resulting nD-parallelepiped is the same as applying matrix (kA) to the unit nD-cube, and so must have nD-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 a 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 to the right. By finding the ratio of the areas, estimate the magnitude of the determinant of the matrix.



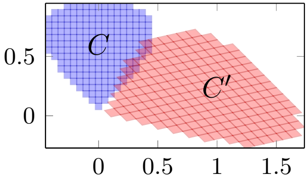
□

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*

$$\det A = \pm \frac{n\text{D-volume of } C'}{n\text{D-volume of } C}$$

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 2×2 matrices of Examples 6.1.1 and 6.1.2: $A = \begin{bmatrix} 1/2 & 1 \\ 0 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$. 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, $\det(AB) = \det(A)\det(B)$. Also, for all $n \times n$ matrices A_1 , A_2 , $\ldots$, A_ℓ, $\det(A_1A_2\cdots A_\ell) = \det(A_1)\det(A_2)\cdots\det(A_\ell)$.*

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 that Example 3.3.4 showed that the following matrix has the given SVD:

$$A=\begin{bmatrix}-4&-2&4\\-8&-1&-4\\6&6&0\end{bmatrix}=\begin{bmatrix}\frac13&-\frac23&\frac23\\\frac23&\frac23&\frac13\\-\frac23&\frac13&\frac23\end{bmatrix}\begin{bmatrix}12&0&0\\0&6&0\\0&0&3\end{bmatrix}\begin{bmatrix}-\frac89&-\frac19&-\frac49\\-\frac49&\frac49&\frac79\\-\frac19&-\frac89&\frac49\end{bmatrix}^{\mathrm{T}}.$$

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 Consider the following matrix and its SVD:

$$A = \begin{bmatrix} -2 & -4 & 5 \\ -6 & 0 & -6 \\ 5 & 4 & -2 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\ -\frac{2}{3} & \frac{1}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} 9 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -\frac{8}{9} & -\frac{1}{9} & -\frac{4}{9} \\ -\frac{4}{9} & \frac{4}{9} & \frac{7}{9} \\ -\frac{1}{9} & -\frac{8}{9} & \frac{4}{9} \end{bmatrix}^{\mathsf{T}}.$$

What is the magnitude of the determinant of A, $|\det A|$?

(a) 4 (b) 81 (c) 0 (d) 18

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 6.1.8(b), that transposing an orthogonal matrix does not change its determinant, $\det Q^{\mathsf T} = \det Q$. We now establish that this determinant-transpose property holds for the transpose of all square matrices.

Example 6.1.25 Example 6.1.18(a) 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, $\det(A^{\mathsf T}) = \det A$.*

Example 6.1.27 Every 3×3 matrix has the form $A = \begin{bmatrix}a&b&c\\d&e&f\\g&h&i\end{bmatrix}$ which has determinant $\det A = |A| = aei + bfg + cdh - ceg - afh -$

bdi. Its transpose, $A^{\mathrm{T}} = \begin{bmatrix} a & d & g \\ b & e & h \\ c & f & i \end{bmatrix}$, from the rule (6.1)

$$\left[\begin{array}{ccc} a & d & g \\ b & e & h \\ c & f & i \end{array}\right]\begin{array}{cc} a & d \\ b & e \\ c & f \end{array}$$

has determinant $\det A^{\mathrm{T}} = aei + dhc + gbf - gec - ahf - dbi = \det A$.

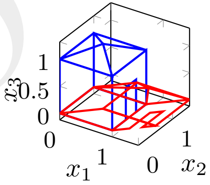
$\square$

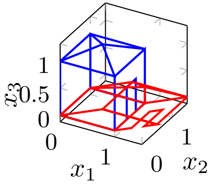
One of the main reasons for studying determinants is to establish when solutions to linear equations may exist or not (albeit only applicable to square matrices when there are n linear equations in n unknowns). One example lies in finding eigenvalues by hand (Section 4.1.1), where we solve $\det(A - \lambda I) = 0$.

Recall that for 2×2 and 3×3 matrices we commented that a matrix is invertible only when its determinant is nonzero. 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

nD-cube into a nD-parallelepiped of zero thickness. Such extreme squashing cannot be uniquely undone.

Example 6.1.28 Consider multiplication by the matrix $A = \begin{bmatrix} 1 & 1/2 & 0 \\ 0 & 1/2 & 1 \\ 0 & 0 & 0 \end{bmatrix}$, whose effect on the unit cube is illustrated to the right. 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

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. In practice, 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. We use these properties to step-by-step develop determinants for matrices in a more and more general form.

Example 6.2.1 *(Theorem 6.2.5(a))* Example 6.1.28 argued geometrically that the determinant is zero for the matrix $A = \begin{bmatrix} 1 & 1/2 & 0 \\ 0 & 1/2 & 1 \\ 0 & 0 & 0 \end{bmatrix}$. Confirm this determinant algebraically. □

Example 6.2.2 *(Theorem 6.2.5(b))* Consider the matrix with two identical rows, $A = \begin{bmatrix} 1 & 1/2 & 1/5 \\ 1 & 1/2 & 1/5 \\ 0 & 1/2 & 1 \end{bmatrix}$. Confirm algebraically that its determinant is zero. Give a geometric reason for why its determinant has to be zero.

□

Example 6.2.3 *(Theorem 6.2.5(c))* Consider the two matrices with two rows swapped (the first two rows): $A = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \\ 1/5 & 1/2 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 1 & 1 \\ 1 & -1 & 0 \\ 1/5 & 1/2 & 1 \end{bmatrix}$ 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 6.2.5(d))* 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 by 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 two identical columns, then* $\det A = 0$ *.*

(c) *Let B be obtained by interchanging two rows or interchanging two 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) $\begin{bmatrix} 0 & 2 & 3 & 1 & 4 \\ -2 & 2 & -2 & 0 & -3 \\ 4 & -2 & -4 & 1 & 0 \\ 2 & -1 & -4 & 2 & 2 \\ -2 & 2 & -2 & 0 & -3 \end{bmatrix}$

(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) $\begin{bmatrix} -2 & 2 & -6 & 0 & -3 \\ 4 & -2 & -12 & 1 & 0 \\ 2 & -1 & -12 & 2 & 2 \\ 0 & 2 & 9 & 1 & 4 \\ 5 & 4 & 9 & -2 & -5 \end{bmatrix}$

(f) $\begin{bmatrix} 0 & 3 & 3 & 1 & 4 \\ -2 & 0 & -2 & 0 & -5 \\ 5 & -1 & -4 & 1 & 0 \\ 2 & -1 & -4 & 2 & 2 \\ 5 & 4 & 6 & -2 & -5 \end{bmatrix}$

$\square$

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

(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) $\begin{bmatrix} 2 & -3 & 1/3 \\ 2 & -5 & -1 \\ -4 & 1 & -1 \end{bmatrix}$

- 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

$$\begin{vmatrix} 1 & x & y \\ 1 & 2 & 3 \\ 1 & 4 & 5 \end{vmatrix} = 0 \tag{6.2}$$

is an equation for 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, an equation of the plane that passes through the origin and the two points $(-1,-2,2)$ and $(3,5,2)$. □

Following Theorem 6.2.5, 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. $\square$

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 ith row and jth column. If, except for the entry a_{ij}, the ith row (or jth column) of A is all zero, then*

$$\det A = (-1)^{i+j} a_{ij} \det A_{ij}\,. \tag{6.3}$$

The pattern of signs in this formula, $(-1)^{i+j}$, is

$$\begin{matrix} + & - & + & - & + & \cdots \\ - & + & - & + & - & \cdots \\ + & - & + & - & + & \cdots \\ - & + & - & + & - & \cdots \\ + & - & + & - & + & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{matrix}$$

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) -120 (d) -60

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 both an upper triangular matrix, and also a lower triangular matrix. Thus the following theorem encompasses square diagonal matrices and so generalizes Theorem 6.1.8(a).

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

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

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

(d) $\begin{bmatrix} -1 & -1 & 1 & 0 \\ 0 & -5 & 4 & 2 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

Example 6.2.18 Find the determinant of those of the following matrices that 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 that we need before establishing a general formula for determinants.

Example 6.2.19 Let's rewrite the explicit formulas (6.1) for 2×2 and 3×3 determinants explicitly as the sum of simpler determinants.

- Recall that the 2×2 determinant

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc = (ad - 0c) + (0d - bc) = \begin{vmatrix} a & 0 \\ c & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & d \end{vmatrix}.$$

  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 $\begin{bmatrix} a & b \end{bmatrix} = \begin{bmatrix} a & 0 \end{bmatrix} + \begin{bmatrix} 0 & b \end{bmatrix}$, while the other row is unchanged.

- Recall from (6.1) that the 3×3 determinant

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = \quad aei + bfg + cdh - ceg - afh - bdi$$

$$
\begin{aligned}
&= +aei + 0fg + 0dh - 0eg - afh - 0di \\
&\quad + 0ei + bfg + 0dh - 0eg - 0fh - bdi \\
&\quad + 0ei + 0fg + cdh - ceg - 0fh - 0di \\
&= \begin{vmatrix} a & 0 & 0 \\ d & e & f \\ g & h & i \end{vmatrix} + \begin{vmatrix} 0 & b & 0 \\ d & e & f \\ g & h & i \end{vmatrix} + \begin{vmatrix} 0 & 0 & c \\ d & e & f \\ g & h & i \end{vmatrix}.
\end{aligned}
$$

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 $\begin{bmatrix} a & b & c \end{bmatrix} = \begin{bmatrix} a & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & b & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & c \end{bmatrix}$, whereas 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 2×2 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 ith column, and that the ith column of A is the sum of the ith 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 of size larger than

3×3, albeit not generally practical for matrices significantly larger in size.

Example 6.2.22 Use Theorems 6.2.11 and 6.2.21 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 specific sum involving three of the following four determinants. Which one of the following would not be used in the sum?

(a) $\begin{vmatrix}6&-4\\4&6\end{vmatrix}$ (b) $\begin{vmatrix}6&-4\\6&-3\end{vmatrix}$ (c) $\begin{vmatrix}7&6\\1&-3\end{vmatrix}$ (d) $\begin{vmatrix}4&6\\6&-3\end{vmatrix}$

Theorem 6.2.24 (Laplace expansion theorem) *For every $n \times n$ matrix $A = \left[a_{ij}\right]$ ($n \geq 2$), recall the (i,j)th minor A_{ij} to be the $(n-1)\times(n-1)$ matrix obtained from A by omitting the ith row and jth 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 20×20 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 3×3 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 three 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 ± 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

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 that the populations interact according to the rule that one year later the population 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 double every year (since then $y(t+1) = 2y(t)$ and $z(t+1) = 2z(t)$); but the presence of the other species causes competition for resources that decreases each of these growths via the $-4z(t)$ and $-y(t)$ terms.

Question: can we find special solutions in the form $(y, z) = \boldsymbol{x}\lambda^t$ for some constant λ and nonzero $\boldsymbol{x} = (x_1, x_2)$? Let's try by substituting $y = x_1\lambda^t$ and $z = x_2\lambda^t$ into the equations:

$$\begin{aligned} & y(t+1) = 2y(t) - 4z(t)\,, \quad z(t+1) = -y(t) + 2z(t) \\ \iff & x_1\lambda^{t+1} = 2x_1\lambda^t - 4x_2\lambda^t\,, \quad x_2\lambda^{t+1} = -x_1\lambda^t + 2x_2\lambda^t \\ \iff & 2x_1 - 4x_2 = \lambda x_1\,, \quad -x_1 + 2x_2 = \lambda x_2 \end{aligned}$$

after dividing by the factor λ^t (assuming constant λ is nonzero). Then form these last two equations as the matrix-vector equation

$$\begin{bmatrix} 2 & -4 \\ -1 & 2 \end{bmatrix} \boldsymbol{x} = \lambda \boldsymbol{x}\,.$$

That is, this substitution $(y,z) = \boldsymbol{x}\lambda^t$ shows the question about finding solutions of the population equations reduces to solving $A\boldsymbol{x} = \lambda\boldsymbol{x}$, called an eigen-problem.

This chapter develops linear algebra for such eigen-problems that empowers us to construct the general solution for the population $y(t)$ and $z(t)$. Here, a general solution is, in terms of two constants c_1 and c_2, that one species has female population $y(t) = 2c_1 4^t + 2c_2$ whereas the second species has female population $z(t) = -c_1 4^t + c_2$.

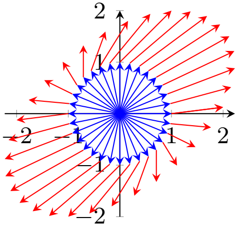
The basic eigen-problem Recall from Section 4.1 that the eigen-problem equation $A\boldsymbol{x} = \lambda\boldsymbol{x}$ is just asking the following: can we find directions $\boldsymbol{x}$ such that matrix A acting on $\boldsymbol{x}$ is in the same direction as $\boldsymbol{x}$? That is, when is $A\boldsymbol{x}$ the same as $\lambda\boldsymbol{x}$ for some proportionality constant λ? Now $\boldsymbol{x} = \boldsymbol{0}$ is always a trivial solution of the equation $A\boldsymbol{x} = \lambda\boldsymbol{x}$. Consequently, we are only interested in those values of the eigenvalue λ when *nonzero* solutions for the eigenvector $\boldsymbol{x}$ exist (as it is the directions which are of interest).

Rearranging the equation $A\boldsymbol{x} = \lambda\boldsymbol{x}$ as the homogeneous system

$(A - \lambda I)\boldsymbol{x} = \boldsymbol{0}$, let's invoke properties of linear equations to solve the eigen-problem.

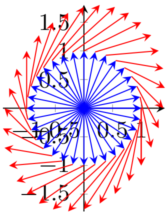
- Procedure 4.1.23 establishes that one way to find the eigenvalues λ (albeit *only* suitable for matrices of small size) is to solve the characteristic equation $\det(A - \lambda I) = 0$.
- Then for each eigenvalue, solving the homogeneous system $(A - \lambda I)\boldsymbol{x} = \boldsymbol{0}$ gives corresponding eigenvectors $\boldsymbol{x}$.
- 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 2×2 matrix A (Schonefeld 1995). The graphical method is to plot many unit vectors $\boldsymbol{x}$, and at the end of each $\boldsymbol{x}$ to adjoin the vector $A\boldsymbol{x}$. Since eigenvectors satisfy $A\boldsymbol{x} = \lambda\boldsymbol{x}$ for some scalar eigenvalue λ, we visually identify eigenvectors as those $\boldsymbol{x}$ for which $A\boldsymbol{x}$ points in the same (or opposite) direction as $\boldsymbol{x}$. Let's use this approach to identify three general difficulties.



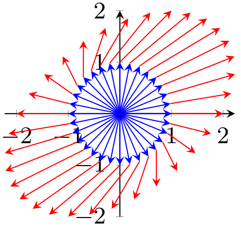
1. In this first picture, for matrix $A = \begin{bmatrix} 1 & 1 \\ 1/8 & 1 \end{bmatrix}$, the eigenvectors appear to be in directions $\boldsymbol{x}_1 \approx \pm(0.9\,,\,0.3)$ and $\boldsymbol{x}_2 \approx \pm(0.9\,,\,-0.3)$ corresponding to eigenvalues $\lambda_1 \approx 1.4$ and $\lambda_2 \approx 0.6$. (Recall that scalar multiples of an eigenvector are always also eigenvectors, Section 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 generalize the concept of orthogonal sets of vectors to a new concept of linearly independent sets (Section 7.2).

2. In this second case, for $A = \begin{bmatrix} 0 & 1 \\ -1 & 1/2 \end{bmatrix}$, there appears to be no (red) vector $A\boldsymbol{x}$ in the same direction as the corresponding (blue) vector $\boldsymbol{x}$. 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 non-real complex valued eigenvalues and eigenvectors (Section 7.1), written using $\mathrm{i} = \sqrt{-1}$. 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).



In this second case the vectors $A\boldsymbol{x}$ all appear to be pointing clockwise. Such a consistent 'rotation' in $A\boldsymbol{x}$ is characteristic of matrices with non-real complex valued eigenvalues and eigenvectors.

3. In this third case, for $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, there appear to be only the vectors $\boldsymbol{x} = \pm(1\,,0)$, aligned along the horizontal axis, for which $A\boldsymbol{x} = \lambda\boldsymbol{x}$. 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) $A = \begin{bmatrix} -3 & 2 & 0 \\ 0 & -4 & 2 \\ 0 & 0 & 4 \end{bmatrix}$

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

(c) $C = \begin{bmatrix} -1 & 1 & -8 & -5 & 5 \\ -3 & 6 & 4 & -3 & 0 \\ 1 & -3 & 1 & 0 & 0 \\ -7 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \end{bmatrix}$

□

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) $0\,,1\,,2\,,3$ (b) 1 (c) $0\,,1$ (d) $0\,,1\,,2$

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 that 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

This section begins exploring the properties and some applications of the eigen-problem $A\boldsymbol{x} = \lambda\boldsymbol{x}$ for *general* matrices A. The section generalises properties established for symmetric matrices, Chapter 4, and relies on the determinant methods of Chapter 6. We establish that there are generally n eigenvalues of an $n \times n$ matrix, albeit possibly non-real complex valued, and also that repeated eigenvalues are sensitive to errors. Applications include population modelling, the computation of SVDs, and fitting exponential functions 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 λ and nonzero eigenvectors $\boldsymbol{x}$ of a square matrix A must satisfy $(A-\lambda I)\boldsymbol{x} = \mathbf{0}$. Theorem 6.1.29 then implies that 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 nth degree in λ;*
- *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) 7 (b) 4 (c) 5 (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 lead to the next (optional) theorem that helps establish the nature of a characteristic polynomial.

Theorem 7.1.4 *For every $n \times n$ matrix A, the product of the eigenvalues equals* $\det A$*, and also equals the constant term in the characteristic polynomial. The sum of the eigenvalues equals* $(-1)^{n-1}$ *times the coefficient of* λ^{n-1} *in the characteristic polynomial, and also equals the* ***trace*** *of the matrix, defined as the sum of the diagonal elements* $a_{11} + a_{22} + \cdots + a_{nn}$ *.*

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) -12 (b) -13 (c) 7 (d) 8

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 its multiplicity. The following definition of *multiplicity* generalizes 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 λ_0 of a matrix A is said to have **multiplicity** m if the characteristic polynomial factorizes 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 $m\geq 2$ may also be 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) one (b) four (c) two (d) three

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) $D = \begin{bmatrix} 2 & 0 & -1 \\ -5 & 3 & -5 \\ 5 & -2 & -2 \end{bmatrix}$

(e) $E = \begin{bmatrix} 0 & 1 \\ -1 & 1 \end{bmatrix}$

□

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 square matrix A (of small size):*

1. *find all eigenvalues (possibly non-real complex) by solving the* ***characteristic equation*** *of A, $\det(A - \lambda I) = 0$;*
2. *for each eigenvalue λ, solve the homogeneous linear equation $(A - \lambda I)\boldsymbol{x} = \mathbf{0}$ 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 nth degree in λ (Theorem 7.1.1), there are n eigenvalues (when counted according to multiplicity and allowing nonreal 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`, namely `diag(D)`, are the eigenvalues of `A`. Corresponding to the eigenvalue `D(j,j)` is an eigenvector $\boldsymbol{v}_j =$ `V(:,j)`, the jth 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 $(A - 3I)\boldsymbol{x} = \boldsymbol{0}$ and hence is an eigenvector of A corresponding to eigenvalue 3?

(a) $\boldsymbol{x} = (1\,,5\,,-1)$ (b) $\boldsymbol{x} = (1\,,5\,,5)$ (c) $\boldsymbol{x} = (0\,,1\,,0)$ (d) $\boldsymbol{x} = (-1\,,0\,,1)$

Example 7.1.13 Find the eigenspaces corresponding to the eigenvalues found for the first three matrices of Example 7.1.9.

7.1.9(a) : $A = \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$

7.1.9(b) : $B = \begin{bmatrix} -1 & 1 & -2 \\ -1 & 0 & -1 \\ 0 & -3 & 1 \end{bmatrix}$

7.1.9(c) : $C = \begin{bmatrix} -1 & 0 & -2 \\ 0 & -3 & 2 \\ 0 & -2 & 1 \end{bmatrix}$

□

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

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 ϵ, 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.

This optional subsection does not prove the sensitivity: it uses examples to introduce and illustrate.

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 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 roughly 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) $C = \begin{bmatrix} -1 & 0 & 0 & 0 \\ -1 & 2 & -3 & 3 \\ 3 & 1 & -1 & 0 \\ 0 & 3 & -2 & 1 \end{bmatrix}$ perturbed by errors of magnitude 10^{-6}.

□

Activity 7.1.18 In an experiment, measurements are made to three decimal place accuracy. Then in analysing the results, a 5×5 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 (distinct) eigenvalues?

(a) five (b) three (c) four (d) two

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). 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 2×2 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 say $B = \begin{bmatrix} a & 10^{-4} \\ 10^{-4} & a \end{bmatrix}$. 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 magnitude 10^{-4} only changes the eigenvalues by a similar amount, the eigenvalues are *not* sensitive. $\square$

Activity 7.1.20 What are the eigenvalues of matrix $\begin{bmatrix} a & 0.01 \\ -0.01 & a \end{bmatrix}$?

(a) $a \pm 0.01$ (b) $a \pm 0.01\,\mathrm{i}$ (c) $a \pm 0.1$ (d) $a \pm 0.1\,\mathrm{i}$

Example 7.1.21 Use Matlab/Octave to 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 that 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 roughly of magnitude 0.0001. $\square$

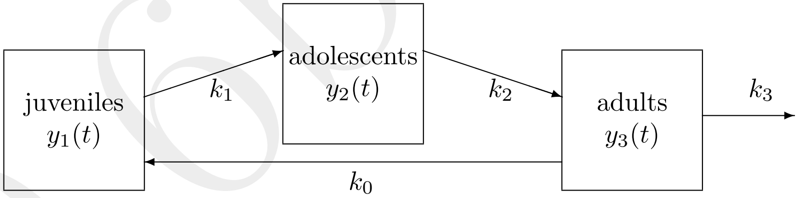
7.1.3 Application: discrete dynamics of populations

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 $\boldsymbol{y}$ of a population. Specifically, this subsection derives and analyses the model $\boldsymbol{y}(t+1) = A\boldsymbol{y}(t)$.

For a given species, and at every time t, 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.

Biologists usually only count the female population as females are the determining sex for reproduction. 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 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:

$$\begin{aligned} y_1(t+1) &= \cdots\,, \\ y_2(t+1) &= \cdots\,, \\ y_3(t+1) &= \cdots\,. \end{aligned}$$

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

$$\begin{aligned} y_1(t+1) &= (1-k_1)y_1(t) + \cdots\,, \\ y_2(t+1) &= +k_1 y_1(t) + \cdots\,, \\ y_3(t+1) &= \cdots\,. \end{aligned}$$

- 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

$$y_1(t+1) = (1-k_1)y_1(t) + \cdots\,,$$

$$\begin{aligned} y_2(t+1) &= +k_1 y_1(t) + (1-k_2) y_2(t), \\ y_3(t+1) &= +k_2 y_2(t) + \cdots . \end{aligned}$$

- A fraction k_3 of the adults die means that a fraction $(1-k_3)$ of the adults remain adults, hence

$$\begin{aligned} y_1(t+1) &= (1-k_1) y_1(t) + \cdots , \\ y_2(t+1) &= +k_1 y_1(t) + (1-k_2) y_2(t), \\ y_3(t+1) &= +k_2 y_2(t) + (1-k_3) y_3(t). \end{aligned}$$

- But adults also give birth to juveniles at rate k_0 per adult so the number of juveniles increases by $k_0 y_3(t)$ from births:

$$\begin{aligned} y_1(t+1) &= (1-k_1) y_1(t) + k_0 y_3(t), \\ y_2(t+1) &= +k_1 y_1(t) + (1-k_2) y_2(t), \\ y_3(t+1) &= +k_2 y_2(t) + (1-k_3) y_3(t). \end{aligned}$$

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, $\boldsymbol{y}(t+1) = A\boldsymbol{y}(t)$. Such a model empowers predictions.

Example 7.1.22 *(orangutans)* From the following extract of the Wikipedia (2014) entry on orangutans (Cawthon Lang 2005) 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 $\boldsymbol{y}(t+1) = A\boldsymbol{y}(t)$ does predict/forecast 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 $\boldsymbol{y}(t) \in \mathbb{R}^2$ changes with time t according to the model

$$\boldsymbol{y}(t+1) = A\boldsymbol{y}(t) = \begin{bmatrix} 1 & -1 \\ -4 & 1 \end{bmatrix} \boldsymbol{y}(t).$$

First, what is $\boldsymbol{y}(3)$ if the initial value $\boldsymbol{y}(0) = (0, 1)$? Second, find a general formula for $\boldsymbol{y}(t)$ from every initial $\boldsymbol{y}(0)$.

□

Activity 7.1.24 For Example 7.1.23, what is the particular solution when $\boldsymbol{y}(0) = (1\,,1)$?

(a) $\boldsymbol{y} = -\frac{1}{4}\cdot(-1)^t(1,2)+\frac{3}{4}\cdot 3^t(-1,2)$

(b) $\boldsymbol{y} = \frac{3}{4}\cdot(-1)^t(1,2)-\frac{1}{4}\cdot 3^t(-1,2)$

(c) $\boldsymbol{y} = \frac{3}{4}\cdot(-1)^t(-1,2)-\frac{1}{4}\cdot 3^t(1,2)$

(d) $\boldsymbol{y} = 4\cdot 3^t(-1\,,2)$

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 $\boldsymbol{y}(t)\in\mathbb{R}^n$ according to $\boldsymbol{y}(t+1) = A\boldsymbol{y}(t)$.*

(a) Let $\lambda_1\,,\lambda_2\,,\ldots,\lambda_m$ be eigenvalues of A and $\boldsymbol{v}_1\,,\boldsymbol{v}_2\,,\ldots,\boldsymbol{v}_m$ be corresponding eigenvectors, then a solution of $\boldsymbol{y}(t+1) = A\boldsymbol{y}(t)$ for all time is the linear combination

$$\boldsymbol{y}(t) = c_1\lambda_1^t\boldsymbol{v}_1 + c_2\lambda_2^t\boldsymbol{v}_2 + \cdots + c_m\lambda_m^t\boldsymbol{v}_m \qquad (7.1)$$

for every value of the constants $c_1\,, c_2\,, \ldots, c_m$.

(b) *Further, if the matrix of eigenvectors* $P = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_m \end{bmatrix}$ *is invertible, then the general linear combination* (7.1) *is a* ***general solution*** *in that unique constants* $c_1, c_2, \ldots, c_m$ *may be found for every given initial value* $\boldsymbol{y}(0)$.

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 $\boldsymbol{y}(t+1) = A\boldsymbol{y}(t)$?

(a) $a = 1$ (b) $a = \pm 1$ (c) $a = -1$ (d) $a = 0$

Example 7.1.27 Consider the dynamics of $\boldsymbol{y}(t+1) = A\boldsymbol{y}(t)$ for matrix $A = \begin{bmatrix} 1 & 3 \\ -1 & 1 \end{bmatrix}$. First, what is $\boldsymbol{y}(3)$ when the initial value $\boldsymbol{y}(0) = (1, 0)$? Second, find a general solution.

□

One crucial qualitative aspect we need to know is whether components in the general solution (7.1) grow, decay, or stay the same magnitude as time increases. The growth or decay is determined by the eigenvalues; the reason is that λ_j^t is the only place that the time appears in the general formula (7.1).

- For example, in the general solution for Example 7.1.23, $\boldsymbol{y}(t) = c_1(-1)^t(1,2) + c_2 3^t(-1,2)$, 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, since it grows rapidly in t, soon dominates the solution.
- In Example 7.1.27, with complex conjugate eigenvalues, the situation is more complicated. Let's write every given complex eigenvalue in polar form $\lambda = r(\cos\theta + \mathrm{i}\sin\theta)$ where magnitude $r = |\lambda|$ and angle θ is such that $\tan\theta = (\text{imag-part }\lambda)/(\text{real

For example, $3+4\,\mathrm{i}$ has magnitude $r = |3+4\,\mathrm{i}\,| = \sqrt{3^2+4^2} = 5$ and angle $\theta = 53.15^\circ$ since $\tan(53.15^\circ) = 4/3$.

> Go to Mr. De Moivre; he knows these things better than I do. *Isaac Newton*

Question: how does this help us to understand the solution which has λ_j^t in it? Answer: De Moivre's theorem says that if $\lambda = r[\cos\theta + \mathrm{i}\sin\theta]$, then $\lambda^t = r^t\big[\cos(\theta t) + \mathrm{i}\sin(\theta t)\big]$. Since the magnitude $|\cos(\theta t) + \mathrm{i}\sin(\theta t)| = \sqrt{\cos^2(\theta t) + \sin^2(\theta t)} = \sqrt{1} = 1$, the magnitude $|\lambda^t| = r^t$. For example, the magnitude $|(3+4\,\mathrm{i})^2| = |3+4\,\mathrm{i}\,|^2 = 5^2 = 25$. We may check this by directly computing $(3+4\,\mathrm{i})^2 = 3^2 + 2\cdot 3\cdot 4\,\mathrm{i} + 4^2\,\mathrm{i}^2 = -7+24\,\mathrm{i}$, and then $|-7+24\,\mathrm{i}\,| = \sqrt{7^2+24^2} = \sqrt{625} = 25$.

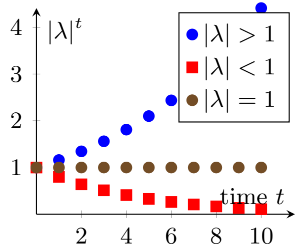
In Example 7.1.27, the eigenvalue $\lambda_1 = 1+\mathrm{i}\sqrt{3}$ so its magnitude is $r_1 = |\lambda_1| = |1+\mathrm{i}\sqrt{3}| = \sqrt{1+3} = 2$. Hence the magnitude $|\lambda_1^t| = 2^t$ at every time step t. Similarly, the magnitude $|\lambda_2^t| = 2^t$ at every time step t. Consequently, the general solution

$$\boldsymbol{y}(t) = c_1\lambda_1^t \begin{bmatrix} -\mathrm{i}\sqrt{3} \\ 1 \end{bmatrix} + c_2\lambda_2^t \begin{bmatrix} +\mathrm{i}\sqrt{3} \\ 1 \end{bmatrix}$$

grows 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 that $\boldsymbol{y}(3)$ is eight times the magnitude of $\boldsymbol{y}(0)$.

In general, for both real or complex eigenvalues λ, a term involving the factor λ^t, as time t increases,

- grows to infinity if $|\lambda| > 1$,
- decays to zero if $|\lambda| < 1$, and
- remains the same magnitude if $|\lambda| = 1$.



Activity 7.1.28 For which of the following values of λ, as time t increases, does λ^t grow in an oscillatory fashion?

(a) $\lambda = -\frac{4}{5} + \mathrm{i}\,\frac{4}{5}$
(b) $\lambda = -0.8$
(c) $\lambda = \frac{3}{5} + \mathrm{i}\,\frac{4}{5}$
(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. □

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 (2014) of a serval's Reproduction and Life History (Sunquist & Sunquist 2002, pp. 142–151).

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, and 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. □

Mathematical modelling in application Examples 7.1.29 and 7.1.30 introduce some of the real-life complexities of mathematical modelling. Bliss et al. (2016) discuss mathematical modelling in *Guidelines for Assessment and Instruction in Mathematics Modeling Education* and some of their comments are relevant here.

- "*Modelling (like real life) is open-ended and messy*" [p.23]:

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.

- 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 may develop alternative plausible models in these scenarios—it is the process of forming plausible models and interpreting the results that is important.

A crucial mathematical feature used in this section and its applications—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

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 computing an SVD distinct from the inductive maximization of Section 3.3.3.

Example 7.1.31 Compute the eigenvalues and eigenvectors of the (symmetric) matrix

$$B=\begin{bmatrix}0&0&10&2\\0&0&5&11\\10&5&0&0\\2&11&0&0\end{bmatrix}.$$

Compare these with an SVD of matrix $A=\left[\begin{smallmatrix}10&2\\5&11\end{smallmatrix}\right]$ from Example 3.3.2. □

Procedure 7.1.11 computes eigenvalues and eigenvectors by hand (albeit not practical for large matrices). 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 maximization of the proof in Section 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 eigen-problem) *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 $B=\begin{bmatrix} O_m & A \\ A^{\mathrm{T}} & O_n \end{bmatrix}$. Each corresponding eigenvector $\boldsymbol{w}\in\mathbb{R}^{m+n}$ of B gives corresponding singular vectors of A, namely $\boldsymbol{w}=(\boldsymbol{u},\boldsymbol{v})$ for singular vectors $\boldsymbol{u}\in\mathbb{R}^m$ and $\boldsymbol{v}\in\mathbb{R}^n$.*

7.1.5 Application: Exponential interpolation discovers dynamics

This optional subsection develops a method useful in many modern applied disciplines.

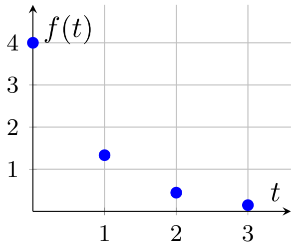
Many applications require identification of rates and frequencies from measured data (e.g., Pereyra & Scherer 2010): as a played musical note decays, what is its frequency? in the observed vibrations of a complicated bridge, what are its natural modes? in measurements of complicated bio-chemical reactions, what rates can be identified? All such tasks require fitting a sum of exponential functions to the data.

Example 7.1.33 This introductory 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 to the right are some points from a function $f(t)$. By inspection, decide which of the following exponentials best represents the data plotted?



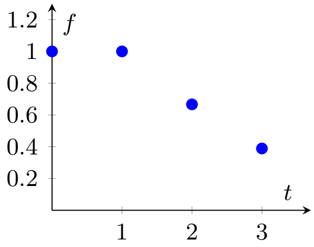
(a) $f \propto e^{-3t}$

(b) $f \propto 1/2^t$

(c) $f \propto e^{-t/2}$

(d) $f \propto 1/3^t$

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 $f_3 = \frac{2}{3} = 0.6667$; and at $t_4 = 3$ the measurement is $f_4 = \frac{7}{18} = 0.3889$ (as plotted to the right). We generally expect chemical reactions to decay exponentially in time. So our task is to find a specific function of the form

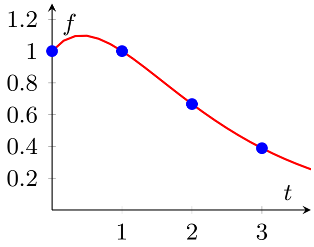
$$f(t) = c_1 e^{r_1 t} + c_2 e^{r_2 t},$$

that explains the data. Our aim is for this function to fit the four known 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

$$f(0) = 1 \iff c_1 + c_2 = 1\,,$$

$$\begin{aligned} f(1) = 1 &\iff c_1 e^{r_1} + c_2 e^{r_2} = 1\,, \\ f(2) = \tfrac{2}{3} &\iff c_1 e^{r_1 2} + c_2 e^{r_2 2} = \tfrac{2}{3}\,, \\ f(3) = \tfrac{7}{18} &\iff c_1 e^{r_1 3} + c_2 e^{r_2 3} = \tfrac{7}{18}\,. \end{aligned}$$

These are four nonlinear equations for the unknown coefficients. However, 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 to the right.



□

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 generalizing eigen-problems as introduced by the following reworking of parts of Example 7.1.35.

Example 7.1.36 *(two shortcuts)* Recall that Section 7.1.3 derived general solutions of dynamic equations such as $\boldsymbol{f}_{j+1} = K\boldsymbol{f}_j$ by seeking solutions of the form $\boldsymbol{f}_j = \boldsymbol{v}\lambda^j$. For the previous Example 7.1.35 let's instead seek solutions of the form $\boldsymbol{f}_j = B\boldsymbol{w}\lambda^j$. Substituting this form, the dynamic equation $\boldsymbol{f}_{j+1} = K\boldsymbol{f}_j$ becomes $B\boldsymbol{w}\lambda^{j+1} = KB\boldsymbol{w}\lambda^j$; then factoring λ^j, recognizing that $KB = A$, and swapping sides, this equation becomes $A\boldsymbol{w} = \lambda B\boldsymbol{w}$. This $A\boldsymbol{w} = \lambda B\boldsymbol{w}$ forms a generalized 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 generalized eigen-problem.

□

Generalized eigen-problem

As introduced by Example 7.1.36, let's generalize the Definition 4.1.1 of eigenvalues and eigenvectors. Such generalized 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* ***generalized eigen-problem*** *is to find scalar eigenvalues λ and corresponding nonzero eigenvectors $\boldsymbol{v}$ such that $A\boldsymbol{v}=\lambda B\boldsymbol{v}$.*

Example 7.1.38 Given $A=\left[\begin{smallmatrix}-2&2\\3&1\end{smallmatrix}\right]$ and $B=\left[\begin{smallmatrix}3&-3\\2&0\end{smallmatrix}\right]$, what eigenvalue corresponds to the eigenvector $\boldsymbol{v}_1=(1,1)$ of the *generalized* eigen-problem $A\boldsymbol{v}=\lambda B\boldsymbol{v}$? Repeat this question for $\boldsymbol{v}_2=(-3,13)$. □

Activity 7.1.39 Which of the following vectors is an eigenvector of the generalized eigen-problem

$$\begin{bmatrix}-2&0\\1&0\end{bmatrix}\boldsymbol{v}=\lambda\begin{bmatrix}3&-1\\-1&1\end{bmatrix}\boldsymbol{v}\,?$$

(a) $(2\,,-1)$ (c) $(0\,,1)$ (d) $(1\,,-1)$
(b) $(0\,,0)$

The standard eigen-problem is the case when matrix $B=I$. Many of the properties for standard eigenvalues and eigenvectors also hold for generalized eigen-problems, although there are some differences. Perhaps the most important familiar property, albeit without proof, is that provided matrix B is invertible, then counted according to multiplicity there are n eigenvalues of a generalized eigen-problem in $n \times n$ matrices.

Example 7.1.40 Find all eigenvalues and corresponding eigenvectors of the generalized eigen-problem $A\boldsymbol{v} = \lambda B\boldsymbol{v}$ for matrices

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

$\square$

Example 7.1.41 Find all eigenvalues and corresponding eigenvectors of the generalized eigen-problem $A\boldsymbol{v} = \lambda B\boldsymbol{v}$ for matrices

$$A = \begin{bmatrix} 2 & 2 \\ -1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$$

$\square$

General fitting of exponentials

Suppose that some experiment or other observation has given us $2n$ data values $f_1, f_2, \ldots, f_{2n}$ at equi-spaced 'times' $t_1, t_2, \ldots, t_{2n}$, 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.)

$$f(t) = c_1 e^{r_1 t} + c_2 e^{r_2 t} + \cdots + c_n e^{r_n t}. \tag{7.2}$$

for some coefficients $c_1, c_2, \ldots, c_n$ and some rates $r_1, r_2, \ldots, r_n$ to be determined (possibly nonreal complex valued for oscillations). In general, finding the coefficients and rates is a delicate nonlinear task outside the remit of this book. However, as Examples 7.1.33 and 7.1.35 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 $\lambda_k = e^{r_k h}$. Then $r_k = (\log \lambda_k)/h$.

Procedure 7.1.42 (exponential interpolation) *Given measured data $f_1, f_2, \ldots, f_{2n}$ at $2n$ equi-spaced times $t_1, t_2, \ldots, t_{2n}$ 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*

$$A = \begin{bmatrix} f_2 & f_3 & \cdots & f_{n+1} \\ f_3 & f_4 & \cdots & f_{n+2} \\ \vdots & \vdots & & \vdots \\ f_{n+1} & f_{n+2} & \cdots & f_{2n} \end{bmatrix}, \quad B = \begin{bmatrix} f_1 & f_2 & \cdots & f_n \\ f_2 & f_3 & \cdots & f_{n+1} \\ \vdots & \vdots & & \vdots \\ f_n & f_{n+1} & \cdots & f_{2n-1} \end{bmatrix}.$$

   `A=hankel(f(2:n+1),f(n+1:2*n))` *and* `B=hankel(f(1:n),f(n:` *form these two matrices in* Matlab*/Octave.*

2. *Find the eigenvalues of the generalized eigen-problem $A\boldsymbol{v} = \lambda B\boldsymbol{v}$:*

   - *by hand on small problems solve* $\det(A - \lambda B) = 0$ *;*

- *invoke* `lambda=eig(A,B)` *, and then* `r=log(lambda)/h` *in* Matlab*/Octave.*

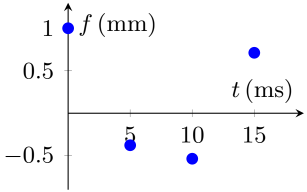
This generalized eigen-problem typically determines n multipliers $\lambda_1, \lambda_2, \ldots, \lambda_n$, and thence the n rates $r_k = (\log\lambda_k)/h$.

3. *Determine the corresponding n coefficients $c_1, c_2, \ldots, c_n$ 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 auto-replication* `U=(lambda.^(0:n-1)).'` *when* `lambda` *is a column vector of the eigenvalues. Since the eigenvalues λ may be nonreal 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 the oscillations here are captured by non-real complex valued exponentials.

□

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.

□

Table 7.1: As well as the MATLAB*/Octave commands and operations listed in Tables 1.2, 2.3, 3.1 to 3.3, 3.7, and 5.1 this section invokes these functions.*

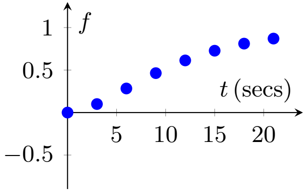
- `hankel(x,y)` *for two vectors of the same size,* `x` $= (x_1, x_2, \ldots, x_n)$ *and* `y` $= (y_1, y_2, \ldots, y_n)$*, when* $x_n = y_1$*, forms the* $n \times n$ *matrix*

$$\begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_n \\ x_2 & x_3 & \cdots & x_n & y_2 \\ x_3 & & \iddots & \iddots & \vdots \\ \vdots & \iddots & \iddots & & y_{n-1} \\ x_n & y_2 & \cdots & y_{n-1} & y_n \end{bmatrix}.$$

- `eig(A,B)` *for* $n \times n$ *matrices* A *and* B *computes a vector in* $\mathbb{R}^n$ *of eigenvalues* λ *such that* $\det(A - \lambda B) = 0$. *Some of the computed eigenvalues in the vector may be* $\pm$`Inf` *(if* B *is not invertible) which denotes that a corresponding eigenvalue does not exist.*
  The command `[V,D]=eig(A,B)` *solves the generalized eigen-problem* $A\boldsymbol{x} = \lambda B\boldsymbol{x}$ *for eigenvalues* λ *returned in the diagonal of matrix* D *(non-existent eigenvalues are denoted*

Example 7.1.45 In a biochemical experiment every three seconds we measure the concentration of an output chemical as tabulated below (and illustrated to the right). 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 (taught in Statistics courses) 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 generalize to cater for noisy data and for complex system interactions with vast amounts of data. The generalization, *Dynamic Mode Decomposition* (e.g., Kutz et al. 2016), applies across many areas such as fluid mechanics, video processing, epidemiology, neuroscience, and financial trading.

7.2 Linearly independent vectors may form a basis

Section contents

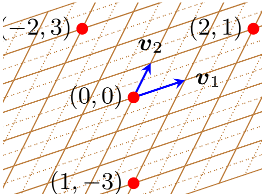
In Chapter 4 on symmetric matrices, the eigenvectors from distinct eigenvalues are proved to be always orthogonal—because of matrix symmetry. For general matrices, the eigenvectors are not orthogonal—as introduced at the start of this Chapter 7. But the orthogonal property is extremely useful. Question: is there some analogue of orthogonality that is similarly useful for general matrices? Answer: yes. We now extend “orthogonal” to the more general concept of “linearly independent”.

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 linearly independent vectors is closely connected to coordinate systems that are not orthogonal.

Subspace coordinate systems In any given mathematical problem, an application wants 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 $y = e^x + 4\cosh x$. It appears from these disparate algebraic forms 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 we need to introduce, in both calculus and linear algebra, the concept of linear independence.

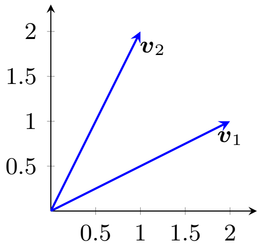


Linear independence empowers us, often implicitly, to use a non-orthogonal coordinate system in a subspace. We replace orthogonal unit vectors by any suitable set of basis vectors. For example, in the plane any two vectors at an angle to each other suffice to be able to describe uniquely every vector (point) in the plane. As illustrated to the right, every point in the plane (end-point of a vector) is a unique linear combination of the two drawn basis vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$. Such a pair of basis vectors, termed a linearly independent pair, avoids the difficulties of algebraic ambiguity.

7.2.1 Linearly (in)dependent sets

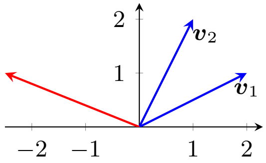
This section defines "linearly dependent" and "linearly independent", and then relates the concept to homogeneous linear equations, orthogonality, and sets of eigenvectors.

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 $\boldsymbol{v}_1 = (2, 1)$ and $\boldsymbol{v}_2 = (1, 2)$ that are shown to the right.



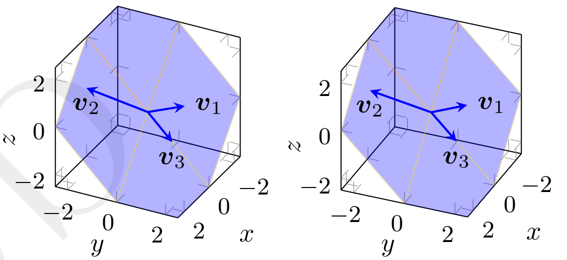
□

Activity 7.2.2 The vector shown to the right is which of the following linear combinations of shown vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$?



(a) $-\boldsymbol{v}_1 + \boldsymbol{v}_2$
(b) $-1.5\boldsymbol{v}_1 + \boldsymbol{v}_2$
(c) $-2.5\boldsymbol{v}_1 + 2\boldsymbol{v}_2$
(d) $-2\boldsymbol{v}_1 + 1.5\boldsymbol{v}_2$

Example 7.2.3 *(3D failure)* Show that vectors in $\mathbb{R}^3$ are not written uniquely as a linear combination of $\boldsymbol{v}_1 = (-1,1,0)$, $\boldsymbol{v}_2 = (1,-2,1)$, and $\boldsymbol{v}_3 = (0,1,-1)$.



One reason for the failure is that these three vectors only span a plane, as shown to the right in stereo. The solution here looks at the different issue of unique representation.

□

Definition 7.2.4 *A set of vectors* $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$ *is* ***linearly dependent*** *if there are scalars* $c_1, c_2, \ldots, c_k$*, at least one of which is nonzero, such that* $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k = \boldsymbol{0}$*. A set of vectors that is not linearly dependent is called* ***linearly***

independent *(characterized by* only *the linear combination with* $c_1 = c_2 = \cdots = c_k = 0$ *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) $\{\mathbf{0}, \boldsymbol{v}_2\,, \boldsymbol{v}_3\,, \ldots, \boldsymbol{v}_k\}$

(f) $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$, 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 generalize 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) $\{(0\,,1)\,,(0\,,-1)\}$

(b) $\{(-1\,,1)\,,(0\,,1)\}$

(c) $\{(-1\,,2)\,,(-2\,,4)\}$

(d) $\{(0\,,0)\,,(-2\,,1)\}$

Example 7.2.7 *(calculus extension)* In calculus the notion of a function corresponds 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 generalizes in the forthcoming Theorem 7.2.11.

Example 7.2.9 Show that the following sets are linearly dependent.

(a) $\{(1,2),\ (3,6)\}$

(b) $\{(2.2,-2.1,0,1.5),\ (-8.8,8.4,0,-6)\}$

$\square$

Activity 7.2.10 For what value of c is the set $\{(-3c\,,-2+2c)\,,\,(1\,,2)\}$ linearly dependent?

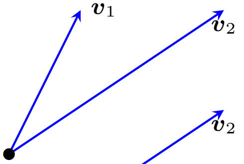
(a) $c=-\frac{1}{3}$ (b) $c=1$ (c) $c=\frac{1}{4}$ (d) $c=0$

Theorem 7.2.11 *A set of vectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m\}$ 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 $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ is linearly dependent if and only if one of the vectors is a scalar multiple of the other.*

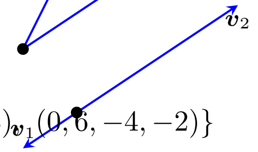
Example 7.2.12 Invoke Theorem 7.2.11 to deduce whether the following sets are linearly independent or linearly dependent.

(a) $\{(-1, 1, 0),\ (1, -2, 1),\ (0, 1, -1)\}$

(b) The set of two vectors shown to the right.



(c) The set of two vectors shown to the right.



(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 generalizes this property to non-symmetric matrices using the concept of linear independence. That the corresponding eigenvalues are all different is crucial.

Theorem 7.2.13 *For every $n \times n$ matrix A, let $\lambda_1, \lambda_2, \ldots, \lambda_m$ be distinct eigenvalues of A with corresponding eigenvectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m$. Then the set $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m\}$ 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 values of a does the matrix have two distinct eigenvalues?

(a) $a \neq -1$ (b) $a \neq 2$ (c) $a \neq 1$ (d) $a \neq 0$

Example 7.2.15 For each of the following matrices, show that 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 that a set of corresponding eigenvectors is linearly independent. The five columns of V, call them $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_5$, are a set of corresponding eigenvectors. To confirm their linear independence let's seek a linear combination being zero, that is, $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_5\boldsymbol{v}_5 = \mathbf{0}$. Written as a matrix-vector system we seek $\boldsymbol{c} = (c_1, c_2, \ldots, c_5)$ such that $V\boldsymbol{c} = \mathbf{0}$.

Because the five singular values of square matrix V are all nonzero, obtained from `svd(V)` as

```
ans =
   1.7703
   1.1268
   0.6542
   0.3625
   0.1922
```

consequently Theorem 3.4.43 asserts that $V\boldsymbol{c} = \boldsymbol{0}$ has only the zero solution. Hence, by Definition 7.2.4 the set of eigenvectors in the columns of V is linearly independent.

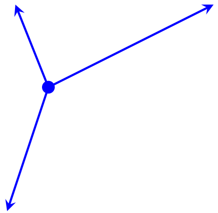
□

This last case of the preceding Example 7.2.15(b) connects the concept of linear in/dependence to the existence or otherwise of nonzero solutions to a homogeneous system of linear equations, $V\boldsymbol{c} = \boldsymbol{0}$. So does Example 7.2.5(b). 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.

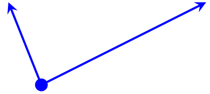
Theorem 7.2.16 *Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m$ be vectors in $\mathbb{R}^n$, and let the $n \times m$ matrix $V = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_m \end{bmatrix}$. Then the set $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m\}$ is linearly dependent if and only if the homogeneous system $V\boldsymbol{c} = \boldsymbol{0}$ has a nonzero solution $\boldsymbol{c}$.*

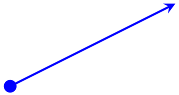
Recall Theorem 1.3.25, that in $\mathbb{R}^n$ there can be no more that n vectors in an orthogonal set of vectors. The following theorem is the generalization: in $\mathbb{R}^n$ there can be no more than n vectors in a linearly independent set of vectors.

Activity 7.2.17 Which of the following sets of vectors are linearly dependent?

(a) 

(b) None of these sets.

(c) 

(d) 

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) $\{(10,3,3,1),(2,-3,0,-1),(1,-1,2,-1),(2,-1,-3,0),(-2,0,2,-$

(f) $\{(-0.4\,,-1.8\,,-0.2\,,0.7\,,-0.2)\,,(-1.1\,,2.8\,,2.7\,,-3.0\,,-2.6)\,,$
$(-2.3\,,-2.3\,,4.1\,,3.4\,,-1.6)\,,(-2.6\,,-5.3\,,-3.3\,,-1.3\,,-4.1)\,,$
$(1.4\,,5.2\,,-6.9\,,-0.7\,,0.6)\}$

$\square$

7.2.2 Form a basis for subspaces

Recall that Sections 2.3 and 3.4 defined subspaces and the span, 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 generalizes 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 such that the set both spans $\mathbb{W}$ and is linearly independent.*

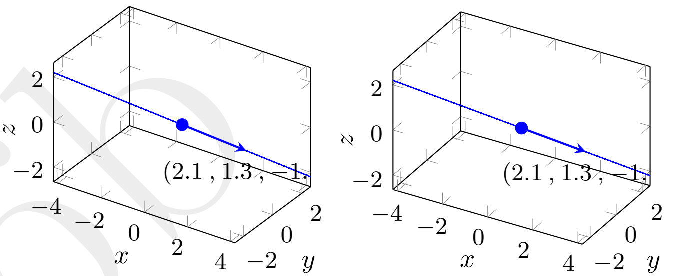
Example 7.2.21

(a) Recall that Example 7.2.5(b) and Example 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 $\mathbb{R}^2$.

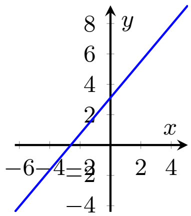
(b) Recall that Example 7.2.5(a) showed that the set $\{(-1,1,0)\,,$ $(1,-2,1)\,,(0,1,-1)\}$ is linearly dependent, so this set cannot be a basis.

However, remove one vector from the set, 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 $\mathbb{W}$. The plane $\mathbb{W}$ is characterized 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, $\operatorname{span}\{(-1,1,0)\ ,(0,1,-1)\}=\mathbb{W}$. Hence $\{(-1,1,0)\ ,(0,1,-1)\}$ is a basis for the plane $\mathbb{W}$.

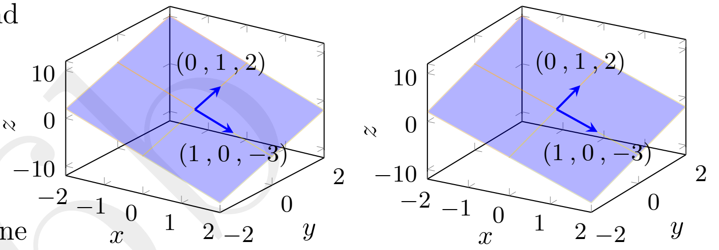
(c) Find a basis for the line given parametrically as $x = 2.1t$, $y = 1.3t$, and $z = -1.1t$ (shown to the right in stereo).



(d) Find a basis for the line given parametrically as $x = 5.7t - 0.6$ and $y = 6.8t + 2.4$.



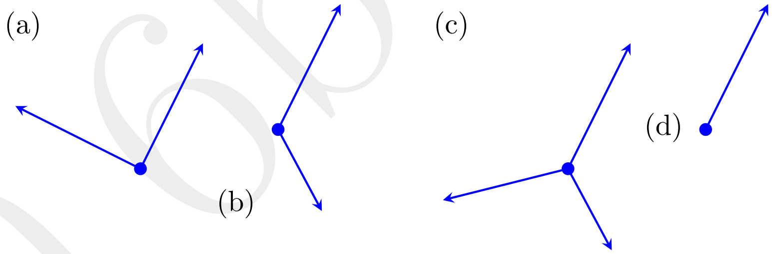
(e) Find a basis for the plane $3x - 2y + z = 0$.



(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 forms 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 that the same is true for general bases. The proof directly generalizes that for Theorem 3.4.28.

Theorem 7.2.23 *Every basis for a given subspace has the same number of vectors.*

Example 7.2.24 Consider the plane $x + y + z = 0$ in $\mathbb{R}^3$. Each of the following is 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 is linearly independent (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 the set is not linearly independent (Example 7.2.5(a)). Each individual vector, say $(-1, 1, 0)$, cannot form a basis for the plane because the span of one vector, such as $\operatorname{span}\{(-1, 1, 0)\}$, is a line not the whole plane.

The orthonormal basis $\left\{(1,0,-1)/\sqrt{2}\,,(1,-2,1)/\sqrt{6}\right\}$ 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) $\{(-1\,,0\,,2)\,,(0\,,0\,,1)\,,(-1\,,2\,,0)\}$

(c) $\{(1\,,1\,,-2)\,,(2\,,2\,,-4)\}$

(d) $\{(1\,,-2\,,1)\,,(1\,,0\,,-1)\}$

Procedure 7.2.27 (basis for a span) *Find a basis for the subspace* $\mathbb{A} = \operatorname{span}\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n\}$ *given that* $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n\}$ *is a set of* n *vectors in* $\mathbb{R}^m$*. Recall that Procedure 3.4.23 underpins finding an orthonormal basis by the following.*

1. *Form* $m \times n$ *matrix* $A := \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n \end{bmatrix}$.
2. *Factorize* A *into its* SVD, $A = USV^{\mathrm{T}}$, *and let* $r = \operatorname{rank} A$ *be the number of nonzero singular values (or effectively nonzero when the matrix has experimental errors, Section 5.2).*
3. *The set* $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_r\}$ *(where* $\boldsymbol{u}_j$ *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* $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n\}$ *is linearly independent and spans 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 that 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) Find a basis for the span of the three vectors

$$(-2,0,-4,1,1),\ (7,1,2,-1,-5),\ (-5,-1,2,3,-2).$$

(c) Find a basis for 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) $\{(-\frac{7}{3}\,,1)\}$ (c) $\{(1\,,-\frac{3}{7})\}$ (d) $\{(-7\,,3)\}$
(b) $\{(3\,,7)\}$

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 $A\boldsymbol{x}=\boldsymbol{0}$. 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 factorization $A=USV^{\mathrm{T}}$ and let $r=\operatorname{rank}A$ be the number of nonzero singular values (or effectively nonzero when the matrix has experimental errors, Section 5.2).*

3. *Then* $\boldsymbol{y} = (0, \ldots, 0, y_{r+1}, \ldots, y_n)$ *is a general solution of* $S\boldsymbol{y} = \boldsymbol{z} = \boldsymbol{0}$. *Consequently, all possible solutions* $\boldsymbol{x} = V\boldsymbol{y}$ *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.

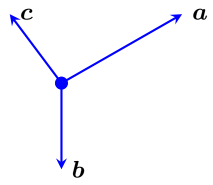
(b) $7x = 6y + z + 3$ and $4x + 9y + 2z + 2 = 0$.

(c) $w + x = z$, $3w = x + y + 5z$, $4x + y + 2z = 0$.

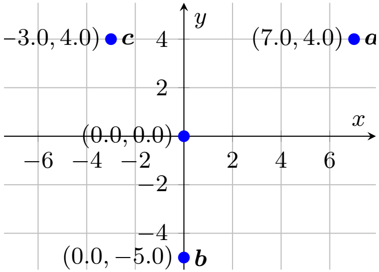
□

Recall that 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.

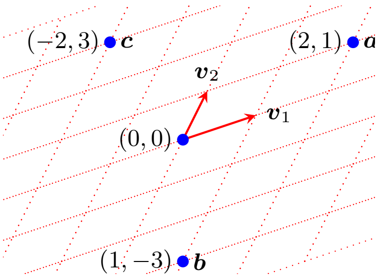
Example 7.2.34 *(a tale of two coordinate systems)* To the right are plotted three vectors and the origin. Take the view that these are fixed physically meaningful vectors; the issue of this example is how do we code such vectors in mathematics?



In the orthogonal standard coordinate system these three vectors and the origin have coordinates as plotted on the right by their endpoints. Consequently, we write $\boldsymbol{a} = (7, 4)$, $\boldsymbol{b} = (0, -5)$, and $\boldsymbol{c} = (-3, 4)$.



Now use the (red) basis $\mathcal{B} = \{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ to form a non-orthogonal coordinate system (represented by the dotted grid). Then in this system the three vectors have coordinates $\boldsymbol{a} = (2, 1)$, $\boldsymbol{b} = (1, -3)$, and $\boldsymbol{c} = (-2, 3)$.



But surely we cannot say both $\boldsymbol{a} = (7, 4)$ and $\boldsymbol{a} = (2, 1)$; it appears nonsense. The reason for the different coordinate numbers representing the one vector $\boldsymbol{a}$ is that the underlying coordinate systems are different. For example, we can say both $\boldsymbol{a} = 7\boldsymbol{e}_1 + 4\boldsymbol{e}_2$ and $\boldsymbol{a} = 2\boldsymbol{v}_1 + \boldsymbol{v}_2$ without any contradiction: these two statements explicitly recognize the underlying standard unit vectors in the first expression and the underlying non-orthogonal basis vectors in the second.

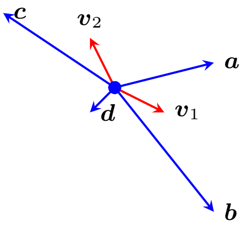
Consequently, mathematicians, scientists, and engineers invented a more precise notation. We write $[\boldsymbol{a}]_{\mathcal{B}} = (2, 1)$ to represent that the coordinates of vector $\boldsymbol{a}$ in the basis $\mathcal{B}$ are $(2, 1)$. Correspondingly, letting $\mathcal{E} = \{\boldsymbol{e}_1, \boldsymbol{e}_2\}$ denote the basis of the standard unit vectors, we write $[\boldsymbol{a}]_{\mathcal{E}} = (7,4)$ to represent that the coordinates of vector $\boldsymbol{a}$ in the standard basis $\mathcal{E}$ are $(7,4)$. Similarly, $[\boldsymbol{b}]_{\mathcal{E}} = (0,-5)$ and $[\boldsymbol{b}]_{\mathcal{B}} = (1,-3)$; and $[\boldsymbol{c}]_{\mathcal{E}} = (-3,4)$ and $[\boldsymbol{c}]_{\mathcal{B}} = (-2,3)$.

The endemic practice of just writing $\boldsymbol{a} = (7,4)$, $\boldsymbol{b} = (0,-5)$ and $\boldsymbol{c} = (-3,4)$ is rationalized in this more precise notation by the convention that if no basis is explicitly specified, then we assume the standard basis $\mathcal{E}$. □

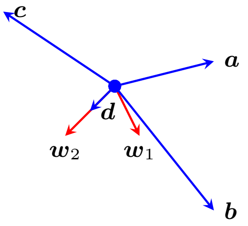
Theorem 7.2.35 *For any given subspace $\mathbb{W}$ of $\mathbb{R}^n$ let $\mathcal{B} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$ be a basis for $\mathbb{W}$. Then there is exactly one way to write each and every vector $\boldsymbol{w} \in \mathbb{W}$ as a linear combination of the basis vectors: $\boldsymbol{w} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k$. The coefficients $c_1, c_2, \ldots, c_k$ are called the* ***coordinates of $\boldsymbol{w}$ with respect to $\mathcal{B}$****, and the column vector $[\boldsymbol{w}]_{\mathcal{B}} = (c_1, c_2, \ldots, c_k)$ is called the* ***coordinate vector of $\boldsymbol{w}$ with respect to $\mathcal{B}$****.*

Example 7.2.36

(a) Consider the diagram to the right of six labelled vectors. Estimate the coordinates of the four shown vectors $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, and $\boldsymbol{d}$ in the shown basis $\mathcal{B} = \{\boldsymbol{v}_1, \boldsymbol{v}_2\}$.

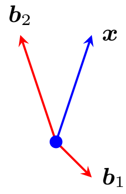


(b) Consider the same four vectors but with a pair of different basis vectors: let's see that although the vectors are the same, the coordinates in the different basis are different. Estimate the coordinates of the four shown vectors $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, and $\boldsymbol{d}$ in the shown basis $\mathcal{W} = \{\boldsymbol{w}_1, \boldsymbol{w}_2\}$.



□

Activity 7.2.37 For the vector $\boldsymbol{x}$ shown to the right, estimate the coordinates of $\boldsymbol{x}$ in the shown basis $\mathcal{B} = \{\boldsymbol{b}_1\,,\boldsymbol{b}_2\}$.



(a) $[\boldsymbol{x}]_{\mathcal{B}} = (4\,,1)$

(b) $[\boldsymbol{x}]_{\mathcal{B}} = (3\,,2)$

(c) $[\boldsymbol{x}]_{\mathcal{B}} = (2\,,3)$

(d) $[\boldsymbol{x}]_{\mathcal{B}} = (1\,,4)$

Example 7.2.38 Let the basis $\mathcal{B} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ for the three given vectors $\boldsymbol{v}_1 = (-1, 1, -1)$, $\boldsymbol{v}_2 = (1, -2, 0)$, and $\boldsymbol{v}_3 = (0, 4, 5)$ (each of these is specified in the standard basis $\mathcal{E}$ of the standard unit vectors $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, and $\boldsymbol{e}_3$).

(a) What is the vector with coordinates $[\boldsymbol{a}]_{\mathcal{B}} = (3, -2, 1)$?

(b) What is the vector with coordinates $[\boldsymbol{b}]_{\mathcal{B}} = (-1, 1, 1)$?

(c) What are the coordinates in the basis $\mathcal{B}$ of the vector $\boldsymbol{c}$ where $[\boldsymbol{c}]_{\mathcal{E}} = (-1, 3, 3)$ in the standard basis $\mathcal{E}$?

(d) What are the coordinates in the basis $\mathcal{B}$ of the vector $\boldsymbol{d}$ where $[\boldsymbol{d}]_{\mathcal{E}} = (-3, 2, 0)$ 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 $\boldsymbol{d}$ where $[\boldsymbol{d}]_{\mathcal{E}} = (2\,,-4)$ in the standard basis $\mathcal{E}$?

(a) $[\boldsymbol{d}]_{\mathcal{B}} = (-2\,,6)$ (b) $[\boldsymbol{d}]_{\mathcal{B}} = (-1\,,3)$ (c) $[\boldsymbol{d}]_{\mathcal{B}} = (1\,,3)$ (d) $[\boldsymbol{d}]_{\mathcal{B}} = (3\,,-1)$

Example 7.2.40 You are given a basis $\mathcal{W} = \{\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3\}$ for a 3D subspace $\mathbb{W}$ of $\mathbb{R}^5$, where the three basis vectors are $\boldsymbol{w}_1 = (1, 3, -4, -3, 3)$, $\boldsymbol{w}_2 = (-4, 1, -2, -4, 1)$, and $\boldsymbol{w}_3 = (-1, 1, 0, 2, -3)$ (in the standard basis $\mathcal{E}$).

(a) What are the coordinates in the standard basis of the vector $\boldsymbol{a} = 2\boldsymbol{w}_1 + 3\boldsymbol{w}_2 + \boldsymbol{w}_3$?

(b) What are the coordinates in the basis $\mathcal{W}$ of the vector $\boldsymbol{b} = (-1\,, 2\,, -6\,, -11\,, 10)$ (in the standard coordinates $\mathcal{E}$).

$\square$

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 3.3.27 and 3.4.43, the following statements are equivalent:*

(a) *A is invertible;*

(b) *$A\boldsymbol{x} = \boldsymbol{b}$ has a **unique solution** for every $\boldsymbol{b} \in \mathbb{R}^n$;*

(c) *$A\boldsymbol{x} = \boldsymbol{0}$ has only the zero solution;*

(d) *all n singular values of A are nonzero;*

(e) *the condition number of A is finite ($\texttt{rcond} > 0$);*

(f) $\operatorname{rank} A = n$ *;*

(g) $\operatorname{nullity} A = 0$ *;*

(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 Diagonalization identifies the transformation

Section contents

Population modelling Recall that this Chapter 7 started by introducing the dynamics of two interacting species of animals. Recall that we let $y(t)$ and $z(t)$ be the number of female animals in each of the two species at time t (years). Modelling might deduce that 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 λ^t led to the eigen-problem

$$\begin{bmatrix} 2 & -4 \\ -1 & 2 \end{bmatrix} \boldsymbol{x} = \lambda \boldsymbol{x} \,.$$

This section introduces an alternate equivalent approach that provides a new view.

The alternate approach invokes non-orthogonal coordinates. Start by writing the population model as a system in terms of vector $\boldsymbol{y}(t) = (y(t), z(t))$, namely

$$\begin{bmatrix} y(t+1) \\ z(t+1) \end{bmatrix} = \begin{bmatrix} 2y(t) - 4z(t) \\ -y(t) + 2z(t) \end{bmatrix}, \quad \text{that is, } \boldsymbol{y}(t+1) = \begin{bmatrix} 2 & -4 \\ -1 & 2 \end{bmatrix} \boldsymbol{y}(t).$$

Now let's ask if there is a basis $\mathcal{P} = \{\boldsymbol{p}_1, \boldsymbol{p}_2\}$ for the yz-plane that simplifies this matrix-vector system. In such a basis every vector may be written as $\boldsymbol{y} = Y_1\boldsymbol{p}_1 + Y_2\boldsymbol{p}_2$ for some scalar components Y_1 and Y_2. That is, $[\boldsymbol{y}]_{\mathcal{P}} = (Y_1, Y_2)$. But to simplify writing we use the vector symbol $\boldsymbol{Y} = (Y_1, Y_2)$ in place of $[\boldsymbol{y}]_{\mathcal{P}}$. Write the relation $\boldsymbol{y} = Y_1\boldsymbol{p}_1 + Y_2\boldsymbol{p}_2$ as the matrix-vector product $\boldsymbol{y} = P\boldsymbol{Y}$ where matrix $P = \begin{bmatrix} \boldsymbol{p}_1 & \boldsymbol{p}_2 \end{bmatrix}$ and vector $\boldsymbol{Y} = (Y_1, Y_2)$. The populations $\boldsymbol{y}$ depends upon time t, and hence so does $\boldsymbol{Y}$ since $\boldsymbol{Y} = [\boldsymbol{y}]_{\mathcal{P}}$; that is, $\boldsymbol{y}(t) = P\boldsymbol{Y}(t)$. Substitute this identity into the system of equations:

$$\boldsymbol{y}(t+1) = P\boldsymbol{Y}(t+1) = \begin{bmatrix} 2 & -4 \\ -1 & 2 \end{bmatrix} P\boldsymbol{Y}(t).$$

Multiply both sides by P^{-1} (which exists by Theorem 7.2.41(l) because of the linear independence of the columns) to give

$$\boldsymbol{Y}(t+1) = \underbrace{P^{-1}\begin{bmatrix} 2 & -4 \\ -1 & 2 \end{bmatrix} P}_{P^{-1}AP}\,\boldsymbol{Y}\,.$$

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, in most cases: using eigenvalues and eigenvector, the product $P^{-1}AP$ can usually be made into a simple diagonal matrix.

7.3.1 Eigenvectors achieve diagonalization

Recall that (Section 4.2.2) for a symmetric matrix A we could always factor $A = VDV^{\mathsf{T}} = VDV^{-1}$ for orthogonal matrix V and diagonal matrix D; thus a symmetric matrix is always orthogonally diagonalizable (Definition 4.2.17). For non-symmetric matrices, a diagonalization can mostly be done (although not always), the difference being that we need an invertible matrix, typically called P, instead of an orthogonal matrix V. Such a matrix A is generally termed 'diagonalizable' instead of the more specific 'orthogonally diagonalizable'.

Definition 7.3.1 *An $n\times n$ square matrix A is **diagonalizable** 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 diagonalizable by matrix $P = \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix}$.

(b) Use a contradiction to prove that $B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ is not diagonalizable.

(c) Is matrix $C = \begin{bmatrix} 1.2 & 3.2 & 2.3 \\ 2.2 & -0.5 & -2.2 \end{bmatrix}$ diagonalizable?

□

Example 7.3.3 Example 7.3.2(a) showed that matrix $P = \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix}$ diagonalizes matrix $A = \begin{bmatrix} 0 & 1 \\ 2 & -1 \end{bmatrix}$ to matrix $D = \operatorname{diag}(1, -2)$. As a prelude to the next Theorem 7.3.5, show that the columns of P are eigenvectors of A. □

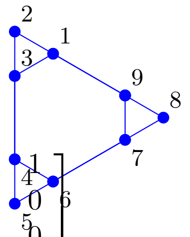
Activity 7.3.4 Given that matrix $F = \begin{bmatrix} 5 & 8 \\ -4 & -7 \end{bmatrix}$ has eigenvectors $(-1\,,1)$ and $(2\,,-1)$ corresponding to respective eigenvalues -3 and 1, what matrix diagonalizes F to $D = \operatorname{diag}(-3\,,1)$?

(a) $\begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$ (b) $\begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix}$ (c) $\begin{bmatrix} -1 & 2 \\ 1 & -1 \end{bmatrix}$ (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 diagonalizable if and only if A has n linearly independent eigenvectors. If A is diagonalizable, 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 proportional to $(1, 0, 0)$, $(-2, 1, 0)$, and $(\frac{1}{14}, \frac{1}{4}, 1)$. Is matrix A diagonalizable? □

Example 7.3.7 Recall the Sierpinski network of Example 4.1.20 (shown to the right). The following 9×9 matrix A encodes the network. Is the matrix diagonalizable?



$$A = \begin{bmatrix} -3 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & -2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & -3 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -3 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & -3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -3 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -2 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & -3 \end{bmatrix}.$$

□

Example 7.3.8 Recall that Example 7.1.13 found eigenvalues and corresponding eigenspaces for various matrices. Revisit these cases and show that none of the matrices are diagonalizable.

(a) Matrix $A = \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$ had one eigenvalue $\lambda = 3$ with multiplicity two and corresponding eigenspace $\mathbb{E}_3 = \operatorname{span}\{(1,0)\}$. This matrix is not diagonalizable as it has only one linearly independent eigenvector, such as $(1,0)$ or any nonzero multiple, and it needs two to be diagonalizable.

(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 $\mathbb{E}_{-2} = \operatorname{span}\{(1,1,1)\}$ and $\mathbb{E}_1 = \operatorname{span}\{(-1,0,1)\}$. Thus the matrix has only two linearly independent eigenvectors, one from each eigenspace, and it needs three to be diagonalizable.

(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 $\mathbb{E}_{-1} = \operatorname{span}\{(1,0,0)\}$. With only one linearly independent eigenvector, the matrix is not diagonalizable. $\square$

Example 7.3.9 Use the results of Example 7.1.14 to show that the following matrix is diagonalizable:

$$A = \begin{bmatrix} 0 & 3 & 0 & 0 & 0 \\ 1 & 0 & 3 & 0 & 0 \\ 0 & 1 & 0 & 3 & 0 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

□

These examples illustrate a widely useful property. The 5×5 matrix in Example 7.3.9 has five distinct eigenvalues whose corresponding eigenvectors are necessarily linearly independent (Theorem 7.2.13) and so diagonalize the matrix (Theorem 7.3.5). The 3×3 matrix in Example 7.3.6 has three distinct eigenvalues whose corresponding eigenvectors are necessarily linearly independent (Theorem 7.2.13) and so diagonalize 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 diagonalizable. The following theorem confirms that matrices with as many *distinct* eigenvalues as the size of the matrix are always diagonalizable.

Theorem 7.3.10 *For every $n \times n$ square matrix A, if A has n distinct eigenvalues, then A is diagonalizable. Consequently, and allowing nonreal complex eigenvalues, a real non-diagonalizable 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 diagonalizable?

(a) The only eigenvalues of a 4×4 matrix are 1.8, -3, 0.4, and 3.2.

(b) The only eigenvalues of a 5×5 matrix are 1.8, -3, 0.4, and 3.2.

(c) The only eigenvalues of a 3×3 matrix are 1.8, -3, 0.4, and 3.2.

□

Activity 7.3.12 A 3×3 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* diagonalizable?

(a) $a=1$ (b) $a=3$ (c) $a=2$ (d) $a=4$

Extension: what are the two other values of a for which the matrix *may not* be diagonalizable?

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 diagonalizable? □

Recall that Definition 7.1.7 defined the multiplicity of every eigenvalue in terms of the characteristic polynomial. Also, the earlier Definition 4.1.15 identified that for every *symmetric* matrix the dimension of an eigenspace, $\dim\mathbb{E}_{\lambda_j}$, is equal to the multiplicity of the corresponding eigenvalue λ_j. However, for general non-symmetric matrices this equality between multiplicity and eigenspace dimension does not necessarily hold.

Theorem 7.3.14 *For every square matrix A, and for each eigenvalue λ_j of A, the corresponding eigenspace $\mathbb{E}_{\lambda_j}$ has dimension less than or equal to the multiplicity of λ_j; that is, $1 \le \dim\mathbb{E}_{\lambda_j} \le$ multiplicity of λ_j .*

Example 7.3.15 Show that the matrix $A = \begin{bmatrix} 0 & 5 & 6 \\ -8 & 22 & 24 \\ 6 & -15 & -16 \end{bmatrix}$ has one eigenvalue of multiplicity three, and that the corresponding eigenspace has dimension two. □

Example 7.3.16 Use MATLAB/Octave to find the eigenvalues and the dimension of the eigenspaces of the matrix

$$B = \begin{bmatrix} 344 & -1165 & -149 & -1031 & 1065 & -2816 \\ 90 & -306 & -38 & -272 & 280 & -742 \\ -45 & 140 & 12 & 117 & -115 & 302 \\ 135 & -470 & -70 & -421 & 445 & -1175 \\ -165 & 555 & 67 & 493 & -506 & 1338 \\ -105 & 360 & 48 & 322 & -335 & 886 \end{bmatrix}.$$

The MATLAB/Octave function `eig()` may produce for you a quite different matrix `V` of eigenvectors (possibly with complex parts). As discussed by Section 7.1.2, repeated eigenvalues are very sensitive, and this sensitivity means that 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. □

7.3.2 Solve systems of differential equations

Population modelling The population modelling seen so far (Section 7.1.3) expressed the changes of the population over discrete intervals in time via discrete time equations such as $y(t+1) = \cdots$ and $z(t+1) = \cdots$. 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. Such differential equations are also common throughout engineering and science.

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 $y(t+1) = \cdots$ and $z(t+1) = \cdots$). 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 $\boldsymbol{Y}(t)$, such that $(y,z)=P\boldsymbol{Y}$ for some as yet unknown matrix P, where the differential equations for $\boldsymbol{Y}$ are simple? Equivalently, is there a different basis $\mathcal{P}=\{\boldsymbol{p}_1,\boldsymbol{p}_2\}$ for the yz-plane in which the differential equations for $\boldsymbol{Y}=[\boldsymbol{y}]_{\mathcal{P}}$ 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 $\boldsymbol{y}=(y,z)$, this system is

$$\frac{d\boldsymbol{y}}{dt}=A\boldsymbol{y} \quad \text{for matrix } A=\begin{bmatrix} 1 & -4 \\ -1 & 1 \end{bmatrix}.$$

- Second, let's see what happens when we transform to some, as yet unknown, new vector variable $\boldsymbol{Y}(t)$ such that $\boldsymbol{y}=P\boldsymbol{Y}$ for some constant invertible matrix P. Under such a transform: $\frac{d\boldsymbol{y}}{dt}=\frac{d}{dt}P\boldsymbol{Y}=P\frac{d\boldsymbol{Y}}{dt}$; also $A\boldsymbol{y}=AP\boldsymbol{Y}$. Hence substituting such an assumed transformation into the matrix-vector form of the differential equation leads to

$$P\frac{d\boldsymbol{Y}}{dt} = AP\boldsymbol{Y}\,, \quad \text{that is} \quad \frac{d\boldsymbol{Y}}{dt} = \left(P^{-1}AP\right)\boldsymbol{Y}.$$

To simplify this system for $\boldsymbol{Y}$, we diagonalize 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 2×2 matrix. Here the matrix $A = \left[\begin{smallmatrix}1 & -4\\ -1 & 1\end{smallmatrix}\right]$ 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

    $$(A-\lambda_1 I)\boldsymbol{p}_1 = \begin{bmatrix}-2 & -4\\ -1 & -2\end{bmatrix}\boldsymbol{p}_1 = \boldsymbol{0}\,,$$

    with general solution $\boldsymbol{p}_1 \propto (2,-1)$.

- For eigenvalue $\lambda_2 = -1$ the corresponding eigenvectors satisfy

$$(A - \lambda_2 I)\boldsymbol{p}_2 = \begin{bmatrix} 2 & -4 \\ -1 & 2 \end{bmatrix} \boldsymbol{p}_2 = \boldsymbol{0}\,,$$

with general solution $\boldsymbol{p}_2 \propto (2, 1)$.

Thus setting transformation matrix

$$P = \begin{bmatrix} \boldsymbol{p}_1 & \boldsymbol{p}_2 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} \implies \frac{d\boldsymbol{Y}}{dt} = \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix} \boldsymbol{Y}$$

(every nonzero scalar multiple of the two columns of P also work).

- Fourth, having diagonalized the matrix, expand this diagonalized set of differential equations to write this system in terms of components $\boldsymbol{Y} = (Y_1, Y_2)$:

$$\frac{dY_1}{dt} = 3Y_1 \quad \text{and} \quad \frac{dY_2}{dt} = -Y_2\,.$$

Each of these differential equations has well-known exponential solutions, respectively $Y_1 = c_1 e^{3t}$ and $Y_2 = c_2 e^{-t}$, for every value of the constants c_1 and c_2.

- Lastly, what does this mean for the original problem? Recall the relation

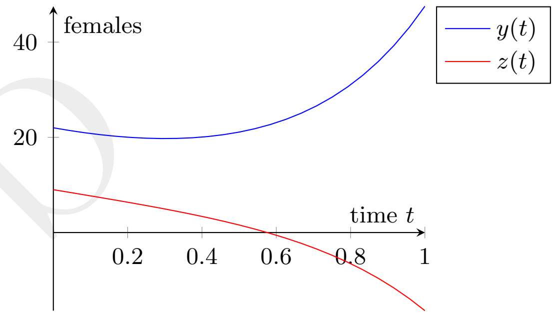
$$\begin{bmatrix} y \\ z \end{bmatrix} = \boldsymbol{y} = P\boldsymbol{Y} = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} c_1 e^{3t} \\ c_2 e^{-t} \end{bmatrix} = \begin{bmatrix} 2c_1 e^{3t} + 2c_2 e^{-t} \\ -c_1 e^{3t} + c_2 e^{-t} \end{bmatrix}.$$

That is, a general solution of the original system of differential equations is $y(t) = 2c_1 e^{3t} + 2c_2 e^{-t}$ and $z(t) = -c_1 e^{3t} + c_2 e^{-t}$.

The diagonalization 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 (female) of y-animals is 22 and the population of z-animals is 9. From the above general solution we then know that at time $t = 0$

$$\begin{bmatrix} 22 \\ 9 \end{bmatrix} = \begin{bmatrix} y(0) \\ z(0) \end{bmatrix} = \begin{bmatrix} 2c_1 e^{3\cdot 0} + 2c_2 e^{-0} \\ -c_1 e^{3\cdot 0} + c_2 e^{-0} \end{bmatrix} = \begin{bmatrix} 2c_1 + 2c_2 \\ -c_1 + c_2 \end{bmatrix}$$



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 graph, above-right, 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 is analogous 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 $d\boldsymbol{x}/dt = A\boldsymbol{x}$ for vector $\boldsymbol{x}(t) = (x, y, z)$ and which of the following matrices?

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

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

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

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

Theorem 7.3.18 *Let $n \times n$ square matrix A be diagonalizable by matrix $P = \begin{bmatrix} \boldsymbol{p}_1 & \boldsymbol{p}_2 & \cdots & \boldsymbol{p}_n \end{bmatrix}$ whose columns are eigenvectors corresponding to eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of matrix A. Then a general solution $\boldsymbol{x}(t)$ to the differential equation system $d\boldsymbol{x}/dt = A\boldsymbol{x}$ is the linear combination*

$$\boldsymbol{x}(t) = c_1\boldsymbol{p}_1 e^{\lambda_1 t} + c_2\boldsymbol{p}_2 e^{\lambda_2 t} + \cdots + c_n\boldsymbol{p}_n e^{\lambda_n t} \tag{7.3}$$

for arbitrary constants $c_1, c_2, \ldots, c_n$.

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) $c_1 = -2$, $c_2 = 0$ (b) $c_1 = -1$, $c_2 = 2$ (c) $c_1 = 0$, $c_2 = -1$ (d) $c_1 = 1$, $c_2 = 1$

Example 7.3.20 Use matrix analysis to 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

$$\begin{aligned} dx_1/dt &= -\tfrac{1}{2}x_1 - \tfrac{1}{2}x_2 + x_3 + 2x_4\,, \\ dx_2/dt &= -\tfrac{1}{2}x_1 - \tfrac{1}{2}x_2 + 2x_3 + x_4\,, \\ dx_3/dt &= \quad x_1 + 2x_2 - \tfrac{1}{2}x_3 - \tfrac{1}{2}x_4\,, \\ dx_4/dt &= \quad 2x_1 + x_2 - \tfrac{1}{2}x_3 - \tfrac{1}{2}x_4\,. \end{aligned}$$

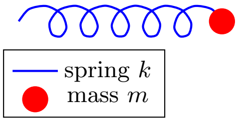
What is the particular solution that satisfies the initial conditions $x_1(0) = -5$, $x_2(0) = -1$, and $x_3(0) = x_4(0) = 0$? Record your commands and give reasons. □

Example 7.3.23 Use matrix analysis to 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 typically express the solution solely in terms of real quantities as just done in the previous Example 7.3.24. Use Euler's formula, that $e^{\mathrm{i}\,\theta} = \cos\theta + \mathrm{i}\sin\theta$ for every θ, 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 to the right. The mass generally oscillates to and fro. Describe such a system mathematically with two differential equations, and solve the differential equations to confirm it oscillates.

□

Bibliography

Arrow, K. J. (1950), 'A difficulty in the concept of social welfare', *Journal of Political Economy* **58**(4), 328–346.

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

Cawthon Lang, K. A. (2005), 'Primate factsheets: Orangutan (pongo) taxonomy, morphology, & ecology', http://pin.primate.wisc.edu/factsheets/entry/orangutan.

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.

Dua, D. & Graff, C. (2019), 'UCI machine learning repository'. http://archive.ics.uci.edu/ml

Ganesan, K., Zhai, C. & Han, J. (2010), A graph based approach to abstractive summarization of highly redundant opinions, *in* 'Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010)', Beijing, pp. 340–348.

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.

Kress, R. (2015), Integral equations, *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.4, pp. 200–208.

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.

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

Sunquist, M. & Sunquist, F. (2002), *Wild Cats of the World*, University of Chicago Press.

Trefethen, L. N. & Bau, III, D. (1997), *Numerical linear algebra*, SIAM.