Linear Algebra Reformed for 21st C Application

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

September 1, 2016

Please inform me of typographical glitches, writing infelicities, and errors.

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

v0.2a

Contents

1 Vectors **10**

1.1 Vectors have magnitude and direction 12

1.2 Adding and stretching vectors 22

1.3 The dot product determines angles and lengths . . . 37

1.4 The cross product 62

1.5 Use MATLAB/Octave for vector computation 79

2 Systems of linear equations **91**

2.1 Introduction to systems of linear equations 94

2.2 Directly solve linear systems 103

2.3 Linear combinations span sets 132

3 Matrices encode system interactions **143**

3.1 Matrix operations and algebra 145

3.2 The inverse of a matrix 177

3.3 Factorise to the singular value decomposition 221

3.4 Subspaces, basis and dimension 259

3.5 Project to solve inconsistent equations 299

3.6 Introducing linear transformations 364

4 Eigenvalues and eigenvectors of symmetric matrices **407**

4.1 Introduction to eigenvalues and eigenvectors 409

4.2 Beautiful properties for symmetric matrices 435

5 Approximate matrices **469**

5.1 Measure changes to matrices 470

5.2 Regularise linear equations 514

v0.2a

6 Determinants distinguish matrices 538

6.1 Geometry underlies determinants 539

6.2 Laplace expansion theorem for determinants 556

7 Eigenvalues and eigenvectors in general 585

7.1 Find eigenvalues and eigenvectors of matrices 593

7.2 Linear independent vectors may form a basis 639

7.3 Diagonalisation identifies the transformation 672

v0.2a

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Preface

> Traditional courses in linear algebra make considerable use of the reduced row echelon form (RREF), but the RREF is an unreliable tool for computation in the face of inexact data and arithmetic. The [Singular Value Decomposition] SVD can be regarded as a modern, computationally powerful replacement for the RREF.[1]
>
> *Cleve Moler, MathWorks (2006)*

The Singular Value Decomposition (SVD) is sometimes called the *jewel in the crown* of linear algebra. Traditionally the SVD is introduced and explored at the end of several linear algebra courses. Question: Why were students required to wait until the end of the course, if at all, to be introduced to beauty and power of this jewel? Answer: limitations of hand calculation.

This book establishes a new route through linear algebra, one that reaches the SVD jewel in linear algebra's crown very early, in Section 3.3. Thereafter its beautiful power both explores many modern applications and also develops traditional linear algebra concepts, theory, and methods. No rigour is lost in this new route: indeed, this book demonstrates that most theory is better proved with an SVD rather than with the traditional RREF. This new route through linear algebra becomes available by the ready availability of ubiquitous computing in the 21st century.

> As so many other disciplines use the SVD, it is not only important that mathematicians understand what it is, but also teach it thoroughly in linear algebra and matrix analysis courses. *(Turner et al. 2015, p.30)*

[1] http://au.mathworks.com/company/newsletters/articles/professor-svd.html [9 Jan 2015]

Aims for students

> How should mathematical sciences departments reshape their curricula to suit the needs of a well-educated workforce in the twenty-first century?
>
> ... The mathematical sciences themselves are changing as the needs of big data and the challenges of modeling complex systems reveal the limits of traditional curricula. *Bressoud et al. (2014)*

Linear algebra is packed with compelling results for application in science, engineering and computing, and with answers for the twenty-first century needs of big data and complex systems. This book provides the conceptual understanding of the essential linear algebra of vectors and matrices for modern engineering and science. The traditional linear algebra course has been reshaped herein to meet modern demands.

Crucial is to inculcate the terms and corresponding relationships that you will most often encounter later in professional life, often when using professional software. For example, the manual for the engineering software package Fluent most often invokes the linear algebra terms of diagonal, dot product, eigenvalue, least square, orthogonal, projection, principal axes, symmetric, unit vector. Engineers need to know these terms. What such useful terms mean, their relationships, and use in applications are central to the mathematical development in this book: you will see them introduced early and used often.

For those who proceed on to do higher mathematics, the development also provides a great solid foundation of key concepts, relationships and transformations necessary for higher mathematics.

Important for all is to develop facility in manipulating, interpreting and transforming between visualisations, algebraic forms, and vector-matrix representations—of problems, working, solutions and interpretation. In particular, one overarching aim of the book is to encourage your formulation, thinking and operation at the crucial system-wide level of matrix/vector operations.

In view of ubiquitous computing, throughout this book explicitly integrates computer support for developing concepts and their relations. The central computational tools to understand are the operation `A\` for solving straightforward linear equations; the function `svd()` for difficult linear equation; the function `svd()` for approximation; and function `eig()` for probing structures. This provides a framework to understand key computational tools to effectively utilise the so-called third arm of science: namely, computation.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Throughout the book examples, many graphical, introduce and illustrate the concepts and relationships between the concepts. Working through these will help form the mathematical relationships essential for application. Interspersed throughout the text are questions labelled "Activity": these aim to help form and test your understanding of the concepts being developed.

Also included are many varied applications, described to varying levels of details. These applications indicate how the mathematics will empower you to answer many practical challenges in engineering and science.

> The main contribution of mathematics to the natural sciences is not in formal computations . . . , but in the investigation of those non-formal questions where the exact setting of the question (what are we searching for and what specific models must be used) usually constitute half the matter.
>
> . . . Examples teach no less than rules, and errors, more than correct but abstruse proofs. Looking at the pictures in this book, the reader will understand more than learning by rote dozens of axioms
>
> *(Arnold 2014, p.xiii)*

Background for teachers

Depending upon the background of your students, your class should pick up the story somewhere in the first two or three chapters. Some students will have previously learnt some vector material in just 2D and 3D, in which case refresh the concepts in nD as presented in the first chapter.

As a teacher you can use this book in several ways.

- One way is as a reasonably rigorous mathematical development of concepts and interconnections by invoking its definitions, theorems, and proofs, all interleaved with examples.
- Another way is as the development of practical techniques and insight for application orientated science and engineering students via the motivating examples to appropriate definitions, theorems and applications.
- Or any mix of these two.

One of the aims of this book is to organise the development of linear algebra so that if a student only studies part of the material, then s/he still obtains a powerful and useful body of knowledge for science or engineering.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

The book typically introduces concepts in low dimensional cases, and subsequently develops the general theory of the concept. This is to help focus the learning, empowered by visualisation, while also making a preformal connection to be strengthened subsequently. People are not one dimensional; knowledge is not linear. Cross-references make many connections, explicitly recalling earlier learning (although sometimes forward to material not yet 'covered').

> information that is recalled grows stronger with each retrieval ... spaced practice is preferable to massed practice. *Halpern & Hakel (2003) [p.38]*

One characteristic of the development is that the concept of linear independence does not appear until relatively late, namely in Chapter 7. This is good for several reasons. First, orthogonality is much more commonly invoked in science and engineering than is linear independence. Second, it is well documented that students struggle with linear independence:

> there is ample talk in the math ed literature of classes hitting a 'brick wall', when linear (in)dependence is studied in the middle of such a course *Uhlig (2002)*

Consequently, here we learn the more specific orthogonality before the more abstract linear independence. Many modern applications are made available by the relatively early introduction of orthogonality.

In addition to many exercises at the end of every section, throughout the book are questions labelled "Activity". These are for the students to do to help form the concepts being introduced. These activities may be used in class to foster active participation by students (perhaps utilising clickers or web tools such as that provided by `http://www.quizsocket.com`). Such active learning has positive effects (Pashler et al. 2007).

On visualisation

> All Linear Algebra courses should stress visualization and geometric interpretation of theoretical ideas in 2- and 3-dimensional spaces. Doing so highlights "algebraic and geometric" as "contrasting but complementary points of view," *(Schumacher et al. 2015, p.38)*

Throughout, this book also integrates visualisation. This visualisation reflects the fundamentally geometric nature of linear algebra. It also empowers learners to utilise different parts of their brain and integrate the knowledge together from the different perspectives. Visualisation also facilitates greater skills at interpretation

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

and modelling so essential in applications. Lastly, a visual exercise question develops learner's understanding without them being able to defer the challenge to online tools, as yet.

> Visual representations are effective because they tap into the capabilities of the powerful and highly parallel human visual system. We like receiving information in visual form and can process it very efficiently: around a quarter of our brains are devoted to vision, more than all our other senses combined . . . researchers (especially those from mathematic backgrounds) see visual notations as being informal, and that serious analysis can only take place at the level of their semantics. However, this is a misconception: visual languages are no less formal than textual ones *Moody (2009)*

On integrated computation

> Cowen argued that because "no serious application of linear algebra happens without a computer," computation should be part of every beginning Linear Algebra course. . . . While the increasing applicability of linear algebra does not require that we stop teaching theory, Cowen argues that "it should encourage us to see the role of the theory in the subject as it is applied."
> *(Schumacher et al. 2015, p.38)*

We need to empower students to use computers to improve their understanding, learning and application of mathematics; not only integrated in their study but also in their later professional career.

One often expects it should be easy to sprinkle a few computational tips and tools throughout a mathematics course. This is not so—extra computing is difficult. There are two reasons for the difficulty: first, the number of computer language details that have to be learned is surprisingly large; second, for students it is a genuine intellectual overhead to learn and relate both the mathematics and the computations.

Consequently, this book chooses a computing language where it is as simple as reasonably possible to perform linear algebra operations: MATLAB/Octave appears to answer this criteria.[2] Further, we are as ruthless as possible in invoking herein the smallest feasible set of commands and functions from MATLAB/Octave so that students

[2] To compare popular packages, just look at the length of expressions students have to type in order to achieve core computations: MATLAB/Octave is almost always the shortest (Nakos & Joyner 1998, e.g.). (Of course be wary of this metric: e.g., APL would surely be too concise!)

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

have the minimum to learn. Most teachers will find many of their favourite commands are missing—this omission is all to the good in focussing upon useful mathematical development aided by only essential integrated computation.

This book does not aim to teach computer programming: there is no flow control, no looping, no recursion, nor function definitions. The aim herein is to use short sequences of declarative assignment statements, coupled with the power of vector and matrix data structures, to learn core mathematical concepts, applications and their relationships in linear algebra.

The internet is now ubiquitous and pervasive. So too is computing power: students can execute MATLAB/Octave not only on laptops, but also on tablets and smart phones, perhaps using university or public servers, `octave-online.net`, MATLAB-Online or MATLAB-Mobile. We no longer need separate computer laboratories. Instead, expect students to access computational support simply by reaching into their pockets or bags.

> long after Riemann had passed away, historians discovered that he had developed advanced techniques for calculating the Riemann zeta function and that his formulation of the Riemann hypothesis—often depicted as a triumph of pure thought—was actually based on painstaking numerical work.
>
> *Donoho & Stodden (2015)*

Acknowledgements

I acknowledge with thanks the work of many others who inspired much design and details here, including the stimulating innovations of calculus reform (Hughes-Hallett et al. 2013, e.g.), the comprehensive efforts behind recent reviews of undergraduate mathematics teaching (Alpers et al. 2013, Bressoud et al. 2014, Turner et al. 2015, Schumacher et al. 2015, Bliss et al. 2016, e.g.), and the books of Anton & Rorres (1991), Davis & Uhl (1999), Holt (2013), Larson (2013), Lay (2012), Nakos & Joyner (1998), Poole (2015), Will (2004). I also thank the entire LaTeX team, especially Knuth, Lamport, Feuersänger, and the AMS.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

1 Vectors

Chapter Contents

1.1 Vectors have magnitude and direction 12

1.1.1 Exercises 19

1.2 Adding and stretching vectors 22

1.2.1 Basic operations 22

1.2.2 Parametric equation of a line 27

1.2.3 Manipulation requires algebraic properties . . 30

1.2.4 Exercises 34

1.3 The dot product determines angles and lengths . . . 37

1.3.1 Work done involves the dot product 43

1.3.2 Algebraic properties of the dot product . . . 44

1.3.3 Orthogonal vectors are at right-angles 50

1.3.4 Normal vectors and equations of a plane . . . 52

1.3.5 Exercises 58

1.4 The cross product 62

1.4.1 Exercises 74

1.5 Use MATLAB/Octave for vector computation 79

1.5.1 Exercises 87

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

Mathematics started with counting. The natural numbers $1, 2, 3, \ldots$ quantify how many objects have been counted. Eventually, via many existential arguments over centuries about whether meaningful, negative numbers and the zero were included to form the **integer**s $\ldots, -2, -1, 0, 1, 2, \ldots$. In the mean time people needed to quantify fractions such as two and half a bags, or a third of a cup which led to the rational numbers such as $2\frac{1}{2} = \frac{5}{2}$ or $\frac{1}{3}$, and now defined as all numbers writeable in the form $\frac{p}{q}$ for integers p and q (q non-zero). Roughly two thousand years ago, Pythagoras was forced to recognise that for many triangles the length of a side could not be rational, and hence there must be more numbers in the world about us than rationals could provide. To cope with non-rational numbers such as $\sqrt{2} = 1.41421\cdots$ and $\pi = 3.14159\cdots$, mathematicians

define the **real numbers** to be all numbers which in principle can be written as a decimal expansion such as

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

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[1] 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. Consequently, many areas of modern science and engineering use **complex numbers** which are those of the form $a + bi$ for real numbers a and b. The symbol $\mathbb{C}$ denotes the set of all possible complex numbers. This book mostly uses integers and real numbers, but eventually we need the marvellous complex numbers.

In some places this book uses the term **scalar** to denote a number that could be integer, real or complex. The term 'scalar' arises because such numbers are often used to scale the length of a 'vector'.

[1] Considered one of the great mathematicians of the Renaissance, Cardano was one of the key figures in the foundation of probability and the earliest introducer of the binomial coefficients and the binomial theorem in the western world. He wrote more than 200 works on medicine, mathematics, physics, chemistry, biology, astronomy, philosophy, religion, and music. ... he is well-known for his achievements in algebra. He made the first systematic use of negative numbers, published with attribution the solutions of other mathematicians for the cubic and quartic equations, and acknowledged the existence of imaginary numbers. (*Wikipedia, 2015*)

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

1.1 Vectors have magnitude and direction

Section Contents

1.1.1 Exercises 19

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

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

Example 1.1.1 (displacement vector)**.** An important class of vectors are the so-called **displacement vectors**. Given two points in space, say A and B, the displacement vector $\overrightarrow{AB}$ is the directed line segment from the point A to the point B—as illustrated by the two displacement vectors $\overrightarrow{AB}$ and $\overrightarrow{CD}$ in the margin. For example, if your home is at position A and your school at position B, then travelling from home to school is moving 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 defined via a coordinate system. So choose an origin for the coordinate system, usually denoted O, and draw coordinate axes in the plane (or space), as illustrated for the above two displacement vectors. Here the displacement vector $\overrightarrow{AB}$ goes three units to the right and one unit up, so we denote it by the ordered pair of numbers $\overrightarrow{AB} = (3, 1)$. Whereas the displacement vector $\overrightarrow{CD}$ goes three units to the left and four units up, so we denote it by the ordered pair of numbers $\overrightarrow{CD} = (-3, 4)$. ■

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

Again, to be able to manipulate such vectors we describe them with numbers, and such numbers have meaning via a coordinate system. So draw coordinate axes in the plane (or space), as illustrated for the above four position vectors. Here the position vector $\overrightarrow{OA}$ goes one

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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 $\overrightarrow{OB} = (-1, 3)$. Recognise that the ordered pairs of numbers in the position vectors are exactly the coordinates of each of the specified end-points. ■

Example 1.1.3 (velocity vector)**.** Consider an airplane in level flight at 900 km/hr to the east-north-east. Choosing coordinate axes oriented to the East and the North, the direction of the airplane is at an angle 22.5° from the East, as illustrated in the margin. Trigonometry then tells us that the Eastward part of the speed of the airplane is $900\cos(22.5°) = 831.5$ km/hr, whereas the Northward part of the speed is $900\sin(22.5°) = 344.4$ km/hr (as indicated in the margin). Further, the airplane is in level flight, not going up or down, so in the third direction of space (vertically) its speed component is zero. Putting these together forms the velocity vector $(831.5, 344.4, 0)$ in 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 in the margin. But West is the opposite direction to East, so if the coordinate system treats East as positive, then West must be negative. Consequently, together with the climb in the vertical, the velocity vector is (−254.6 km/hr,254.6 km/hr,2 m/s). But it is best to avoid mixing units within a vector, so here convert all speeds to m/s: here 360 km/hr upon dividing by 3600 secs/hr and multiplying by 1000 m/km gives 360 km/hr = 100 m/s. Then the North and West speeds are both $100\cos(\frac{\pi}{4}) = 70.71$ m/s. Consequently, the velocity vector of the climbing airplane should be described as $(-70.71, 70.71, 2)$ in m/s. ■

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

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

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

- In the three dimensions of space the standard coordinate system has three coordinate axes, two horizontal and one vertical all at right-angles to each other, often labelled x_1, x_2 and x_3 respectively (as illustrated in the margin), although labels x, y and z are also common.
- Correspondingly, in so-called 'n dimensions' the standard coordinate system has n coordinate axes, all at right-angles to each other, and often labelled $x_1, x_2, \dots, x_n$, respectively.

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

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

(they mean the same, it is just more convenient to usually use a row in parentheses in text, and a column in brackets in displayed mathematics). The real numbers $x_1, x_2, \dots, x_n$ are called the ***components*** *of the vector, and the number of components is termed its* ***size*** *(here n). The components are determined such that letting X be the point with coordinates $(x_1, x_2, \dots, x_n)$ then the position vector $\overrightarrow{OX}$ has the same magnitude and direction as the vector denoted $(x_1, x_2, \dots, x_n)$. Two vectors of the same size are* ***equal***, *$=$, if all their corresponding components are equal (vectors with different sizes are never equal).*

Robert Recorde invented the equal sign circa 1557 "bicause noe 2 thynges can be moare equalle".

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

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

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

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

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

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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

Example 1.1.6.

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

■

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

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

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

(a) The dog sat on the mat.

(b) The cat scratched the dog.

(c) The cat and dog sat on the mat.

(d) The dog scratched.

These four sentences involve up to three objects, cat, dog and mat, and two actions, sat and scratched. Some characteristic of the sentences is captured simply by counting the number of times each of these three objects and two actions appear in each sentence, and then forming a vector from the counts. Let's use vectors $\boldsymbol{w} = (N_{\text{cat}}, N_{\text{dog}}, N_{\text{mat}}, N_{\text{sat}}, N_{\text{scratched}})$ 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 $\mathbb{R}^5$. Then

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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

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

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

(d) "The dog scratched" is summarised 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 summarised 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).[2]

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. ■

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) "The dog scratched the cat on the mat."

(b) "A dog sat. A cat scratched the dog. The cat sat."

(c) "A dog and cat both sat on the mat which the dog had scratched."

(d) "Which cat sat by the dog on the mat, and then scratched the dog."

■

[2] Look up Latent Semantic Indexing, such as at https://en.wikipedia.org/wiki/Latent_semantic_indexing [April 2015]

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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. [a] *Technology Review, 2015*

[a] http://www.technologyreview.com/view/541356 [Oct 2015]

Definition 1.1.9 (Pythagoras)**.** *Given any vector* $\boldsymbol{v} = (v_1, v_2, \dots, v_n)$ *in* $\mathbb{R}^n$*, define the* ***length*** *(or* ***magnitude****) of 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****. (Be aware that many books denote the length of a vector with a pair of double lines, as in* $\|\boldsymbol{v}\|$*.)*

Example 1.1.10. Find the lengths of the following vectors.

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

Solution: $|\boldsymbol{a}| = \sqrt{(-3)^2 + 4^2} = \sqrt{25} = 5.$

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

Solution: $|\boldsymbol{b}| = \sqrt{3^2 + 3^2} = \sqrt{18} = 3\sqrt{2}.$

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

Solution: $|\boldsymbol{c}| = \sqrt{1^2 + (-2)^2 + 3^2} = \sqrt{14}.$

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

Solution: $|\boldsymbol{d}| = \sqrt{1^2 + (-1)^2 + (-1)^2 + 1^2} = \sqrt{4} = 2.$

■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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

Solution: (a) Humans knew of the $3:4:5$ right-angled triangle thousands of years ago, so perhaps one answer could be $(3,4)$, $(-4,3)$ and $(5,0)$.

(b) One answer might be $(3,0,0)$, $(0,3,0)$ and $(0,0,3)$. A more interesting answer might arise from knowing $1^2+2^2+2^2=3^2$ leading to an answer of $(1,2,2)$, $(2,-1,2)$ and $(-2,2,1)$.

(c) Since the length of a vector is $\sqrt{\cdots}$ which is always positive or zero, the length cannot be negative, so there is no possible answer to this last request.

■

Activity 1.1.12. What is the length of the vector $(1,-2,2)$?

(a) 1 (b) $\sqrt{5}$ (c) 3 (d) 5

■

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

Proof. First establish the zero vector has length zero. From Definition 1.1.9, in $\mathbb{R}^n$,

$$|\mathbf{0}|=\sqrt{0^2+0^2+\cdots+0^2}=\sqrt{0}=0\,.$$

Second, if a vector has length zero then it must be the zero vector. Let vector $\boldsymbol{v}=(v_1\,,v_2\,,\dots,v_n)$ in $\mathbb{R}^n$ have zero length. By squaring both sides of the Definition 1.1.9 for length we then know that

$$\underbrace{v_1^2}_{\geq 0}+\underbrace{v_2^2}_{\geq 0}+\cdots+\underbrace{v_n^2}_{\geq 0}=0\,.$$

Being squares, all terms on the left are non-negative, so the only way they can all add to zero is if they are all zero. That is, $v_1=v_2=\cdots=v_n=0$. Hence, the vector $\boldsymbol{v}$ must be the zero vector $\mathbf{0}$. □

v0.2a

1.1.1 Exercises

Exercise 1.1.1. For each case: on the plot draw the displacement vectors $\overrightarrow{AB}$ and $\overrightarrow{CD}$, and the position vectors of the points A and D.

Exercise 1.1.2. For each case: roughly estimate (to say ± 0.2) each of the two components of the four position vectors of the points A, B, C and D.

v0.2a

(c)

(d)

(e)

(f)

Exercise 1.1.3. For each case plotted in Exercise 1.1.2: from your estimated components of each of the four position vectors, calculate the length (or magnitude) of the four vectors. Also use a ruler (or otherwise) to directly measure an estimate of the length of each vector. Confirm your calculated lengths reasonably approximate your measured lengths.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Exercise 1.1.4. Below are the titles of eight books that The Society of Industrial and Applied Mathematics (SIAM) reviewed recently.

(a) Introduction to Finite and Spectral Element Methods using MATLAB

(b) Derivative Securities and Difference Methods

(c) Iterative Methods for Linear Systems: Theory and Applications

(d) Singular Perturbations: Introduction to System Order Reduction Methods with Applications

(e) Risk and Portfolio Analysis: Principles and Methods

(f) Differential Equations: Theory, Technique, and Practice

(g) Contract Theory in Continuous-Time Models

(h) Stochastic Chemical Kinetics: Theory and Mostly Systems Biology Applications

Make a list of the five significant words that appear more than once in this list (not including the common nontechnical words such as "and" and "for", and not distinguishing between words with a common root). Being consistent about the order of words, represent each of the eight titles by a word vector in $\mathbb{R}^7$.

VU.2a

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

1.2 Adding and stretching vectors

Section Contents

1.2.1 Basic operations 22

Distance 26

1.2.2 Parametric equation of a line 27

1.2.3 Manipulation requires algebraic properties . . 30

1.2.4 Exercises 34

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

1.2.1 Basic operations

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

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

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

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

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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

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 three-D effect. With practice viewing such three-D stereo pairs becomes less difficult.

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

■

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

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

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

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

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

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

■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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) $(3\,, -2)$ (d) $(5\,, -8)$

■

Definition 1.2.4. *Let two vectors in $\mathbb{R}^n$ be $\boldsymbol{u} = (u_1, u_2, \dots, u_n)$ and $\boldsymbol{v} = (v_1, v_2, \dots, v_n)$, 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*

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

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*

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

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 $\boldsymbol{u} + (-\boldsymbol{v})$ 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 in the margin, 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 $-\boldsymbol{v}$.

Solution: Drawn below.

(a) $\boldsymbol{u} + \boldsymbol{v}$

(b) $\boldsymbol{v} + \boldsymbol{u}$

(c) $\boldsymbol{u} - \boldsymbol{v}$

(d) $\boldsymbol{v} - \boldsymbol{u}$

(e) $\frac{1}{2}\boldsymbol{u}$

(f) $-\boldsymbol{v}$

■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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

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

■

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

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

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

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

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

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

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

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 also illustrated).

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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) $(3\,,1)$ (b) $(5\,,0\,,2)$

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

■

Distance

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

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

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

This concept of distance between two vectors $\boldsymbol{u}$ 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, $|\boldsymbol{u} - \boldsymbol{v}|$.*

v0.2a

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

Solution: Compute the distances between each pair.

- $|\boldsymbol{b} - \boldsymbol{a}| = |2\boldsymbol{i} + 3\boldsymbol{j} + 6\boldsymbol{k}| = \sqrt{2^2 + 3^2 + 6^2} = \sqrt{49} = 7$.
- $|\boldsymbol{c} - \boldsymbol{a}| = |4\boldsymbol{i} - 4\boldsymbol{j} + 7\boldsymbol{k}| = \sqrt{4^2 + (-4)^2 + 7^2} = \sqrt{81} = 9$.
- $|\boldsymbol{c} - \boldsymbol{b}| = |2\boldsymbol{i} - 7\boldsymbol{j} - \boldsymbol{k}| = \sqrt{2^2 + (-7)^2 + (-1)^2} = \sqrt{54} = 7.3485$.

The smallest distance of 7 is between $\boldsymbol{a}$ and $\boldsymbol{b}$ so these two are the closest pair of vectors. The largest distance of 9 is between $\boldsymbol{a}$ and $\boldsymbol{c}$ so these two are the furthest pair of vectors.

■

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

(c) $\boldsymbol{b}$, $\boldsymbol{c}$ (d) two pairs

■

1.2.2 Parametric equation of a line

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

Example 1.2.13. Consider the line drawn in the margin in some chosen coordinate system. Recall one way to find an equation of the line is to find the intercepts with the axes, here at $x = 4$ and $y = 2$, 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 $y = 2 - x/2$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

The alternative is to describe the line with vectors. Choose any point P on the line, such as $(2,1)$ as drawn in the margin. Then view any 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 any given line, there are many possible choices of $\boldsymbol{p}$ and $\boldsymbol{d}$ in such a vector representation. A different looking, but equally valid form is obtained from any pair of points on the line. For example, one could choose point P to be $(0,2)$ and point Q to be $(3,\frac{1}{2})$, as drawn in the margin. Let position vector $\boldsymbol{p}=\overrightarrow{OP}=(0,2)$ 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:

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

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

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

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

(c) $(-1\,,1/2)+(2\,,-1)t$ (d) $(-2\,,0)+(-4\,,-2)t$

■

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** $\boldsymbol{d}$ is parallel to the line ($\boldsymbol{d}\neq\boldsymbol{0}$), and the scalar **parameter** t varies over all real values to give all position vectors on the line.*

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

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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

Solution: Let's call the points $(-4,-3,3)$ and $(3,2,1)$ as P and Q, respectively, and as shown below. First, choose a point on the line, say P, and set its position vector $\boldsymbol{p} = \overrightarrow{OP} = (-4,-3,3) = -4\boldsymbol{i} - 3\boldsymbol{j} + 3\boldsymbol{k}$, as drawn. Second, choose a direction vector to be, say, $\boldsymbol{d} = \overrightarrow{PQ} = (3,2,1) - (-4,-3,3) = 7\boldsymbol{i} + 5\boldsymbol{j} - 2\boldsymbol{k}$, also drawn. A parametric equation of the line is then $\boldsymbol{x} = \boldsymbol{p} + t\boldsymbol{d}$, specifically

$$\begin{aligned}\boldsymbol{x} &= (-4\boldsymbol{i} - 3\boldsymbol{j} + 3\boldsymbol{k}) + t(7\boldsymbol{i} + 5\boldsymbol{j} - 2\boldsymbol{k}) \\ &= (-4+7t)\boldsymbol{i} + (-3+5t)\boldsymbol{j} + (3-2t)\boldsymbol{k}.\end{aligned}$$

■

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 $(-6, 4, 4)$.

Solution:

- For the point $(-1.6, 1.8, -5.8)$ we need to find the parameter value t such that $-4+2t = -1.6$, $3-t = 1.8$ and $-1-4t = -5.8$. The first of these requires $t = (-1.6+4)/2 = 1.2$, the second requires $t = 3 - 1.8 = 1.2$, and the third requires $t = (-1+5.8)/4 = 1.2$. All three agree that choosing parameter $t = 1.2$ gives the required point.

- For the point $(-3, 2.5, -3)$ we need to find the parameter value t such that $-4+2t = -3$, $3-t = 2.5$ and $-1-4t = -3$. The first of these requires $t = (-3+4)/2 = 0.5$, the second requires $t = 3 - 2.5 = 0.5$, and the third requires $t = (-1+3)/4 = 0.5$. All three agree that choosing parameter $t = 0.5$ gives the required point.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

- For the point $(-6, 4, 4)$ we need to find the parameter value t such that $-4 + 2t = -6$, $3 - t = 4$ and $-1 - 4t = 4$. The first of these requires $t = (-6 + 4)/2 = -1$, the second requires $t = 3 - 4 = -1$, and the third requires $t = (-1 - 4)/4 = -1.25$. Since these three require different values of t, namely -1 and -1.25, it means that there is no single value of the parameter t that gives the required point. That is, the point $(-6, 4, 4)$ cannot be on the line. Consequently the task is impossible. [3]

■

1.2.3 Manipulation requires algebraic properties

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

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

Example 1.2.18. Let vectors $\boldsymbol{u} = (1, 2)$, $\boldsymbol{v} = (3, 1)$, 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);

Solution: $\boldsymbol{u} + \boldsymbol{v} = (1, 2) + (3, 1) = (1 + 3, 2 + 1) = (4, 3)$, whereas $\boldsymbol{v} + \boldsymbol{u} = (3, 1) + (1, 2) = (3 + 1, 1 + 2) = (4, 3)$ is the same.

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

Solution: $(\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w} = (4, 3) + (-2, 3) = (2, 6)$, whereas $\boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w}) = \boldsymbol{u} + ((3, 1) + (-2, 3)) = (1, 2) + (1, 4) = (2, 6)$ is the same.

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

Solution: $\boldsymbol{u} + \boldsymbol{0} = (1, 2) + (0, 0) = (1 + 0, 2 + 0) = (1, 2) = \boldsymbol{u}$.

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

Solution: Recall $-\boldsymbol{u} = (-1)\boldsymbol{u} = (-1)(1, 2) = (-1, -2)$, and so $\boldsymbol{u} + (-\boldsymbol{u}) = (1, 2) + (-1, -2) = (1 - 1, 2 - 2) = (0, 0) = \boldsymbol{0}$.

[3] Section 3.5 develops how to treat such inconsistent information in order to 'best solve' such impossible tasks.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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

Solution: $a(\boldsymbol{u}+\boldsymbol{v}) = -\frac{1}{2}(4,3) = (-\frac{1}{2}\cdot 4, -\frac{1}{2}\cdot 3) = (-2, -\frac{3}{2})$, whereas $a\boldsymbol{u} + a\boldsymbol{v} = -\frac{1}{2}(1,2) + (-\frac{1}{2})(3,1) = (-\frac{1}{2}, -1) + (-\frac{3}{2}, -\frac{1}{2}) = (-\frac{1}{2} - \frac{3}{2}, -1 - \frac{1}{2}) = (-2, -\frac{3}{2})$ which is the same.

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

Solution: $(a+b)\boldsymbol{u} = (-\frac{1}{2}+\frac{5}{2})(1,2) = 2(1,2) = (2\cdot 1, 2\cdot 2) = (2,4)$, wheras $a\boldsymbol{u} + b\boldsymbol{u} = (-\frac{1}{2})(1,2) + \frac{5}{2}(1,2) = (-\frac{1}{2}, -1) + (\frac{5}{2}, 5) = (-\frac{1}{2} + \frac{5}{2}, -1 + 5) = (-2, 4)$ which is the same.

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

Solution: $(ab)\boldsymbol{u} = (-\frac{1}{2}\cdot 52)(1,2) = (-\frac{5}{4})(1,2) = (-\frac{5}{4}, -\frac{5}{2})$, whereas $a(b\boldsymbol{u}) = a(\frac{5}{2}(1,2)) = (-\frac{1}{2})(\frac{5}{2}, 5) = (-\frac{5}{4}, -\frac{5}{2})$ which is the same.

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

Solution: $1\boldsymbol{u} = 1(1,2) = (1\cdot 1, 1\cdot 2) = (1,2) = \boldsymbol{u}$.

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

Solution: $0\boldsymbol{u} = 0(1,2) = (0\cdot 1, 0\cdot 2) = (0,0) = \boldsymbol{0}$.

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

Solution: Now $|a| = |-\frac{1}{2}| = \frac{1}{2}$, and the length $|\boldsymbol{u}| = \sqrt{1^2+2^2} = \sqrt{5}$ (Definition 1.1.9). Consequently, $|a\boldsymbol{u}| = |(-\frac{1}{2})(1,2)| = |(-\frac{1}{2}, -1)| = \sqrt{(-\frac{1}{2})^2 + (-1)^2} = \sqrt{\frac{1}{4}+1} = \sqrt{\frac{5}{4}} = \frac{1}{2}\sqrt{5} = |a|\cdot|\boldsymbol{u}|$ as required.

■

Theorem 1.2.19. *Let $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{w}$ be any vectors with n components (in $\mathbb{R}^n$), and let a and b be any scalars. Then 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) $|a\boldsymbol{u}| = |a|\cdot|\boldsymbol{u}|$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Proof. We prove property 1.2.19a, and leave the proof of other properties as exercises. The approach is to establish the properties of vector operations using the known properties of scalar operations.

Property 1.2.19a is the commutativity of vector addition. Example 1.2.1a shows graphically how the equality $\boldsymbol{u}+\boldsymbol{v}=\boldsymbol{v}+\boldsymbol{u}$ in one case, and the margin here shows another case. In general, let vectors $\boldsymbol{u}=(u_1\,,u_2\,,\dots,u_n)$ and $\boldsymbol{v}=(v_1\,,v_2\,,\dots,v_n)$ then

$$\begin{aligned}
&\boldsymbol{u}+\boldsymbol{v}\\
&=(u_1\,,u_2\,,\dots,u_n)+(v_1\,,v_2\,,\dots,v_n)\\
&=(u_1+v_1\,,u_2+v_2\,,\dots,u_n+v_n)\quad\text{(by Defn. 1.2.4)}\\
&=(v_1+u_1\,,v_2+u_2\,,\dots,v_n+u_n)\quad\text{(commutative scalar add)}\\
&=(v_1\,,v_2\,,\dots,v_n)+(u_1\,,u_2\,,\dots,u_n)\quad\text{(by Defn. 1.2.4)}\\
&=\boldsymbol{v}+\boldsymbol{u}\,.
\end{aligned}$$

□

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

Solution: The left diagram.

- In the left diagram, the two red vectors represent $\boldsymbol{u}+\boldsymbol{v}$ (left) and $\boldsymbol{v}+\boldsymbol{w}$ (right). Thus the left-red followed by the blue $\boldsymbol{w}$ represents $(\boldsymbol{u}+\boldsymbol{v})+\boldsymbol{w}$, whereas the $\boldsymbol{u}$ followed by the right-red represents $\boldsymbol{u}+(\boldsymbol{v}+\boldsymbol{w})$. The brown vector shows they are equal: $(\boldsymbol{u}+\boldsymbol{v})+\boldsymbol{w}=\boldsymbol{u}+(\boldsymbol{v}+\boldsymbol{w})$.
- The right-hand diagram invokes the commutative law as well. The top-left part of the diagram shows $(\boldsymbol{v}+\boldsymbol{w})+\boldsymbol{u}$, whereas the bottom-right part shows $(\boldsymbol{u}+\boldsymbol{v})+\boldsymbol{w}$. That these are equal, the brown vector, requires both the commutative and associative laws.

■

We frequently use these algebraic properties in rearranging and solving vector equations.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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

Solution: Using Theorem 1.2.19, all the following equations are equivalent:

$$\begin{aligned}
3\boldsymbol{x} - 2\boldsymbol{u} &= 6\boldsymbol{v}; \\
(3\boldsymbol{x} - 2\boldsymbol{u}) + 2\boldsymbol{u} &= 6\boldsymbol{v} + 2\boldsymbol{u} \quad \text{(add } 2\boldsymbol{u} \text{ to both sides);} \\
3\boldsymbol{x} + (-2\boldsymbol{u} + 2\boldsymbol{u}) &= 6\boldsymbol{v} + 2\boldsymbol{u} \quad \text{(by 1.2.19b, associativity);} \\
3\boldsymbol{x} + \boldsymbol{0} &= 6\boldsymbol{v} + 2\boldsymbol{u} \quad \text{(by 1.2.19d);} \\
3\boldsymbol{x} &= 6\boldsymbol{v} + 2\boldsymbol{u} \quad \text{(by 1.2.19c);} \\
\tfrac{1}{3}(3\boldsymbol{x}) &= \tfrac{1}{3}(6\boldsymbol{v} + 2\boldsymbol{u}) \quad \text{(multiply both sides by } \tfrac{1}{3}\text{);} \\
\tfrac{1}{3}(3\boldsymbol{x}) &= \tfrac{1}{3}(6\boldsymbol{v}) + \tfrac{1}{3}(2\boldsymbol{u}) \quad \text{(by 1.2.19e, distributivity);} \\
(\tfrac{1}{3} \cdot 3)\boldsymbol{x} &= (\tfrac{1}{3} \cdot 6)\boldsymbol{v} + (\tfrac{1}{3} \cdot 2)\boldsymbol{u} \quad \text{(by 1.2.19g);} \\
1\boldsymbol{x} &= 2\boldsymbol{v} + \tfrac{2}{3}\boldsymbol{u} \quad \text{(by scalar operations);} \\
\boldsymbol{x} &= 2\boldsymbol{v} + \tfrac{2}{3}\boldsymbol{u} \quad \text{(by 1.2.19h).}
\end{aligned}$$

Generally we do not write down all such details. Generally the following shorter derivation is acceptable. The following are equivalent:

$$\begin{aligned}
3\boldsymbol{x} - 2\boldsymbol{u} &= 6\boldsymbol{v}; \\
3\boldsymbol{x} &= 6\boldsymbol{v} + 2\boldsymbol{u} \quad \text{(adding } 2\boldsymbol{u} \text{ to both sides);} \\
\boldsymbol{x} &= 2\boldsymbol{v} + \tfrac{2}{3}\boldsymbol{u} \quad \text{(dividing both sides by 3).}
\end{aligned}$$

But exercises and examples in this section often explicitly require full details and justification.

■

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.

Solution: Using Theorem 1.2.19, the following statements are equivalent:

$$\begin{aligned}
3\boldsymbol{x} - \boldsymbol{a} &= 2(\boldsymbol{a} + \boldsymbol{x}) \\
3\boldsymbol{x} - \boldsymbol{a} &= 2\boldsymbol{a} + 2\boldsymbol{x} \quad \text{(by 1.2.19e, distributivity);} \\
(3\boldsymbol{x} - \boldsymbol{a}) + \boldsymbol{a} &= (2\boldsymbol{a} + 2\boldsymbol{x}) + \boldsymbol{a} \quad \text{(adding } \boldsymbol{a} \text{ to both sides);} \\
3\boldsymbol{x} + (-\boldsymbol{a} + \boldsymbol{a}) &= 2\boldsymbol{a} + (2\boldsymbol{x} + \boldsymbol{a}) \quad \text{(by 1.2.19b, associativity);} \\
3\boldsymbol{x} + \boldsymbol{0} &= 2\boldsymbol{a} + (\boldsymbol{a} + 2\boldsymbol{x}) \quad \text{(by 1.2.19d and 1.2.19a);} \\
3\boldsymbol{x} &= (2\boldsymbol{a} + \boldsymbol{a}) + 2\boldsymbol{x} \quad \text{(by 1.2.19c and 1.2.19b);} \\
3\boldsymbol{x} &= (2\boldsymbol{a} + 1\boldsymbol{a}) + 2\boldsymbol{x} \quad \text{(by 1.2.19h);} \\
3\boldsymbol{x} &= (2 + 1)\boldsymbol{a} + 2\boldsymbol{x} \quad \text{(by 1.2.19f, distributivity);} \\
3\boldsymbol{x} + (-2)\boldsymbol{x} &= 3\boldsymbol{a} + 2\boldsymbol{x} + (-2)\boldsymbol{x} \quad \text{(sub. } 2\boldsymbol{x} \text{ from both sides);} \\
(3 + (-2))\boldsymbol{x} &= 3\boldsymbol{a} + (2 + (-2))\boldsymbol{x} \quad \text{(by 1.2.19f, distributivity);} \\
1\boldsymbol{x} &= 3\boldsymbol{a} + 0\boldsymbol{x} \quad \text{(by scalar arithmetic);}
\end{aligned}$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$\begin{aligned} \boldsymbol{x} &= 3\boldsymbol{a} + \boldsymbol{0} \quad \text{(by 1.2.19h and 1.2.19i);} \\ \boldsymbol{x} &= 3\boldsymbol{a} \quad \text{(by 1.2.19c).} \end{aligned}$$

If the question had not requested full details, then the following would be enough. The following statements are equivalent:

$$\begin{aligned} 3\boldsymbol{x} - \boldsymbol{a} &= 2(\boldsymbol{a} + \boldsymbol{x}) \quad \text{(distribute the mupltiplication)} \\ 3\boldsymbol{x} &= 2\boldsymbol{a} + 2\boldsymbol{x} + \boldsymbol{a} \quad \text{(adding } \boldsymbol{a} \text{ to both sides);} \\ \boldsymbol{x} &= 3\boldsymbol{a} \quad \text{(subtracting } 2\boldsymbol{x} \text{ from both sides).} \end{aligned}$$

■

1.2.4 Exercises

Exercise 1.2.1. For each of the pairs of vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ shown below, 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 $-\boldsymbol{v}$.

Exercise 1.2.2. For each of the following pairs of vectors shown below, use a ruler (or other measuring stick) to directly measure the distance between the pair of vectors.

(e)

(f)

Exercise 1.2.3. For each of the following groups of vectors, use the distance between vectors to find which pair in the group are closest to each other, and which pair in the group are furthest from each other.

(a) $\boldsymbol{u} = (-5\,,0\,,3)$, $\boldsymbol{v} = (1\,,-6\,,10)$, $\boldsymbol{w} = (-4\,,4\,,11)$

(b) $\boldsymbol{u} = (2\,,2\,,-1)$, $\boldsymbol{v} = (3\,,6\,,-9)$, $\boldsymbol{w} = (1\,,-2\,,-9)$

(c) $\boldsymbol{u} = (1\,,1\,,-3)$, $\boldsymbol{v} = (7\,,7\,,-10)$, $\boldsymbol{w} = (-1\,,4\,,-9)$

(d) $\boldsymbol{u} = 3\boldsymbol{i}$, $\boldsymbol{v} = 4\boldsymbol{i} - 2\boldsymbol{j} + 2\boldsymbol{k}$, $\boldsymbol{w} = 4\boldsymbol{i} + 2\boldsymbol{j} + 2\boldsymbol{k}$

(e) $\boldsymbol{u} = (-5\,,3\,,5\,,6)$, $\boldsymbol{v} = (-6\,,1\,,3\,,10)$, $\boldsymbol{w} = (-4\,,6\,,2\,,15)$

(f) $\boldsymbol{u} = (-4,-1,-1,2)$, $\boldsymbol{v} = (-5,-2,-2,1)$, $\boldsymbol{w} = (-3,-2,-2,1)$

(g) $\boldsymbol{u} = 5\boldsymbol{e}_1+\boldsymbol{e}_3+5\boldsymbol{e}_4$, $\boldsymbol{v} = 6\boldsymbol{e}_1-2\boldsymbol{e}_2+3\boldsymbol{e}_3+\boldsymbol{e}_4$, $\boldsymbol{w} = 7\boldsymbol{e}_1-2\boldsymbol{e}_2-3\boldsymbol{e}_3$

(h) $\boldsymbol{u} = 2\boldsymbol{e}_1 + 4\boldsymbol{e}_1 - \boldsymbol{e}_3 + 5\boldsymbol{e}_4$, $\boldsymbol{v} = -2\boldsymbol{e}_1 + 8\boldsymbol{e}_2 - 6\boldsymbol{e}_3 - 3\boldsymbol{e}_4$, $\boldsymbol{w} = -6\boldsymbol{e}_3 + 11\boldsymbol{e}_4$

Exercise 1.2.4. Find a parametric equation of the line through the given two points.

(a) $(-11\,,0\,,3)$, $(-3\,,-2\,,2)$

(b) $(-4\,,1\,,-2)$, $(3\,,-5\,,5)$

(c) $(2.4\,,5.5\,,-3.9)$, $(1.5\,,-5.4\,,-0.5)$

(d) $(0.2\,,-7.2\,,-4.6\,,-2.8)$, $(3.3\,,-1.1\,,-0.4\,,-0.3)$

(e) $(2.2\,,5.8\,,4\,,3\,,2)$, $(-1.1\,,2.2\,,-2.4\,,-3.2\,,0.9)$

(f) $(1.8\,,-3.1\,,-1\,,-1.3\,,-3.3)$, $(-1.4\,,0.8\,,-2.6\,,3.1\,,-0.8)$

Exercise 1.2.5. Verify the algebraic properties of Theorem 1.2.19 for each of the following sets of vectors and scalars.

(a) $\boldsymbol{u} = 2.4\boldsymbol{i} - 0.3\boldsymbol{j}$, $\boldsymbol{v} = -1.9\boldsymbol{i} + 0.5\boldsymbol{j}$, $\boldsymbol{w} = -3.5\boldsymbol{i} - 1.8\boldsymbol{j}$, $a = 0.4$ and $b = 1.4$.

(b) $\boldsymbol{u} = (1/3, 14/3)$, $\boldsymbol{v} = (4, 4)$, $\boldsymbol{w} = (2/3, -10/3)$, $a = -2/3$ and $b = -1$.

(c) $\boldsymbol{u} = -\frac{1}{2}\boldsymbol{j} + \frac{3}{2}\boldsymbol{k}$, $\boldsymbol{v} = 2\boldsymbol{i} - \boldsymbol{j}$, $\boldsymbol{w} = 2\boldsymbol{i} - \boldsymbol{k}$, $a = -3$ and $b = \frac{1}{2}$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(d) $\boldsymbol{u} = (2, 1, 4, -2)$, $\boldsymbol{v} = (-3, -2, 0, -1)$, $\boldsymbol{w} = (-6, 5, 4, 2)$, $a = -4$ and $b = 3$.

Exercise 1.2.6. Prove in detail some algebraic properties chosen from Theorem 1.2.19b–1.2.19j on vector addition and scalar multiplication.

Exercise 1.2.7. For each of the following vectors equations, rearrange the equations to get vector $\boldsymbol{x}$ in terms of the other vectors. Give excruciating detail of the justification using Theorem 1.2.19.

(a) $\boldsymbol{x} + \boldsymbol{a} = \boldsymbol{0}$.

(b) $2\boldsymbol{x} - \boldsymbol{b} = 3\boldsymbol{b}$.

(c) $3(\boldsymbol{x} + \boldsymbol{a}) = \boldsymbol{x} + (\boldsymbol{a} - 2\boldsymbol{x})$.

(d) $-4\boldsymbol{b} = \boldsymbol{x} + 3(\boldsymbol{a} - \boldsymbol{x})$.

v0.2a

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

1.3 The dot product determines angles and lengths

Section Contents

1.3.1 Work done involves the dot product 43

1.3.2 Algebraic properties of the dot product . . . 44

1.3.3 Orthogonal vectors are at right-angles 50

1.3.4 Normal vectors and equations of a plane . . . 52

1.3.5 Exercises 58

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 scalars. This section introduces the so-called dot product of two vectors that, among other attributes, gives us a way of determining the angle between the two vectors.

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

Solution: Form a triangle with the vector $\boldsymbol{u} - \boldsymbol{v} = (5, -6)$ going from the tip of $\boldsymbol{v}$ to the tip of $\boldsymbol{u}$, as shown in the margin. The sides of the triangles are of length $|\boldsymbol{u}| = \sqrt{7^2 + (-1)^2} = \sqrt{50} = 5\sqrt{2}$, $|\boldsymbol{v}| = \sqrt{2^2 + 5^2} = \sqrt{29}$, and $|\boldsymbol{u} - \boldsymbol{v}| = \sqrt{5^2 + (-6)^2} = \sqrt{61}$. By the cosine rule for triangles

$$|\boldsymbol{u} - \boldsymbol{v}|^2 = |\boldsymbol{u}|^2 + |\boldsymbol{v}|^2 - 2|\boldsymbol{u}||\boldsymbol{v}| \cos\theta\,.$$

Here this rule rearranges to

$$\begin{aligned} |\boldsymbol{u}||\boldsymbol{v}| \cos\theta &= \tfrac{1}{2}(|\boldsymbol{u}|^2 + |\boldsymbol{v}|^2 - |\boldsymbol{u} - \boldsymbol{v}|^2) \\ &= \tfrac{1}{2}(50 + 29 - 61) \\ &= 9\,. \end{aligned}$$

Dividing by the product of the lengths then gives $\cos\theta = 9/(5\sqrt{58}) = 0.2364$ so the angle $\theta = \arccos(0.2364) = 1.3322 = 76.33^\circ$ as is reasonable from the plots.

■

The interest in this Example 1.3.1 is the number nine 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 any pair of vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ to find that always $|\boldsymbol{u}||\boldsymbol{v}| \cos\theta = u_1v_1 + u_2v_2$ (Exercise 1.3.1). 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.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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

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

The dot product of two vectors gives a scalar result, 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) $\boldsymbol{u} = (-2, 5, -2)$, $\boldsymbol{v} = (3, 3, -2)$

Solution: $\boldsymbol{u} \cdot \boldsymbol{v} = (-2)3 + 5 \cdot 3 + (-2)(-2) = 13$. Alternatively, $\boldsymbol{v} \cdot \boldsymbol{u} = 3(-2) + 3 \cdot 5 + (-2)(-2) = 13$. That these give the same result is a consequence of the general commutative law, Theorem 1.3.13a, and so in the following we compute the dot product only one way around.

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

Solution: There is no answer: the dot product cannot be computed here as the two vectors are of different sizes.

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

Solution: $\boldsymbol{a} \cdot \boldsymbol{b} = (-7)(-3) + 3 \cdot 4 + 0(-4) + 2 \cdot 2 + 2 \cdot 0 = 37$.

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

Solution: $\boldsymbol{p} \cdot \boldsymbol{q} = (-0.1)(-1.6) + (-2.5)1.1 + (-3.3)(-3.4) + 0.2 \cdot 2.2 = 9.07$.

■

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 any two non-zero vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{R}^n$, the* ***angle*** *θ between the vectors is determined by*

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

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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 $180°$.

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

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

Solution: These vectors (shown in the margin) have length $\sqrt{4^2 + 3^2} = \sqrt{25} = 5$ and $\sqrt{5^2 + 12^2} = \sqrt{169} = 13$, respectively. Their dot product $(4, 3) \cdot (5, 12) = 20 + 36 = 56$. Hence $\cos\theta = 56/(5 \cdot 13) = 0.8615$ and so angle $\theta = \arccos(0.8615) = 0.5325 = 30.51°$.

(b) $(3, 1)$ and $(-2, 1)$

Solution: These vectors (shown in the margin) have length $\sqrt{3^2 + 1^2} = \sqrt{10}$ and $\sqrt{(-2)^2 + 1^2} = \sqrt{5}$, respectively. Their dot product $(3, 1) \cdot (-2, 1) = -6 + 1 = -5$. Hence $\cos\theta = -5/(\sqrt{10} \cdot \sqrt{5}) = -1/\sqrt{2} = -0.7071$ and so angle $\theta = \arccos(-1/\sqrt{2}) = 2.3562 = \frac{3}{4}\pi = 135°$ (Table 1.1).

(c) $(4, -2)$ and $(-1, -2)$

Solution: These vectors (shown in the margin) have length $\sqrt{4^2 + (-2)^2} = \sqrt{20} = 2\sqrt{5}$ and $\sqrt{(-1)^2 + (-2)^2} = \sqrt{5}$, respectively. Their dot product $(4, -2) \cdot (-1, -2) = -4 + 4 = 0$. Hence $\cos\theta = 0/(2\sqrt{5} \cdot \sqrt{5}) = 0$ and so angle $\theta = \frac{1}{2}\pi = 90°$ (Table 1.1).

■

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

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

■

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

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

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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	$-1.$

Solution: Draw two vectors along adjacent diagonals: the above pair of vectors are $(1,1,0)$ and $(0,1,1)$. They both have the same length as $|(1,1,0)| = \sqrt{1^2+1^2+0^2} = \sqrt{2}$ and $|(0,1,1)| = \sqrt{0^2+1^2+1^2} = \sqrt{2}$. The dot product is $(1,1,0)\cdot(0,1,1) = 0+1+0 = 1$. Hence the cosine $\cos\theta = 1/(\sqrt{2}\cdot\sqrt{2}) = 1/2$. Table 1.1 gives the angle $\theta = \frac{\pi}{3} = 60°$.

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

Solution: Draw two vectors along the diagonals: the above pair of vectors are $(1,1,0)$ and $(1,1,1)$. The face-diagonal has length $|(1,1,0)| = \sqrt{1^2+1^2+0^2} = \sqrt{2}$ whereas the cube diagonal has length $|(1,1,1)| = \sqrt{1^2+1^2+1^2} = \sqrt{3}$. The

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

dot product is $(1,1,0)\cdot(1,1,1) = 1+1+0 = 2$. Hence $\cos\theta = 2/(\sqrt{2}\cdot\sqrt{3}) = \sqrt{2/3} = 0.8165$. Then a calculator (or MATLAB/Octave, see Section 1.5) gives the angle $\theta = \arccos(0.8165) = 0.6155 = 35.26°$.

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

Solution: Draw two corresponding vectors from the center atom: the above pair of vectors are $(1,1,1)$ and $(1,1,-1)$. These have the same length $|(1,1,1)| = \sqrt{1^2+1^2+1^2} = \sqrt{3}$ and $|(1,1,-1)| = \sqrt{1^2+1^2+(-1)^2} = \sqrt{3}$. The dot product is $(1,1,1)\cdot(1,1,-1) = 1+1-1 = 1$. Hence $\cos\theta = 1/(\sqrt{3}\cdot\sqrt{3}) = 1/3 = 0.3333$. Then a calculator (or MATLAB/Octave, see Section 1.5) gives the angle $\theta = \arccos(1/3) = 1.2310 = 70.53°$.

■

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 characterise the similarity of sentences: a small angle means the sentences are somehow close; a large angle means the sentences are disparate.

Solution: First, these word vectors have lengths $|\boldsymbol{a}| = |\boldsymbol{b}| = \sqrt{3}$ and $|\boldsymbol{c}| = 2$. Second, the 'angles' between these sentences are the following.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

- The angle θ_{ab} between "The dog sat on the mat" and "The cat scratched the dog" satisfies

$$\cos\theta_{ab} = \frac{\boldsymbol{a}\cdot\boldsymbol{b}}{|\boldsymbol{a}||\boldsymbol{b}|} = \frac{0+1+0+0+0}{\sqrt{3}\cdot\sqrt{3}} = \frac{1}{3}\,.$$

  A calculator (or MATLAB/Octave, see Section 1.5) then gives the angle $\theta_{ab} = \arccos(1/3) = 1.2310 = 70.53^\circ$ so the sentences are quite dissimilar.

- The angle θ_{ac} between "The dog sat on the mat" and "The cat and dog sat on the mat" satisfies

$$\cos\theta_{ac} = \frac{\boldsymbol{a}\cdot\boldsymbol{c}}{|\boldsymbol{a}||\boldsymbol{c}|} = \frac{0+1+1+1+0}{\sqrt{3}\cdot 2} = \frac{3}{2\sqrt{3}} = \frac{\sqrt{3}}{2}\,.$$

  Table 1.1 gives the angle $\theta_{ac} = \frac{\pi}{6} = 30^\circ$ so the sentences are roughly similar.

- The angle θ_{bc} between "The cat scratched the dog" and "The cat and dog sat on the mat" satisfies

$$\cos\theta_{bc} = \frac{\boldsymbol{b}\cdot\boldsymbol{c}}{|\boldsymbol{b}||\boldsymbol{c}|} = \frac{1+1+0+0+0}{\sqrt{3}\cdot 2} = \frac{2}{2\sqrt{3}} = \frac{1}{\sqrt{3}}\,.$$

  A calculator (or MATLAB/Octave, see Section 1.5) then gives the angle $\theta_{bc} = \arccos(1/\sqrt{3}) = 0.9553 = 54.74^\circ$ so the sentences are moderately dissimilar.

The following stereo plot schematically draws these three vectors at the correct angles from each other, and with correct lengths, in some abstract coordinate system (Section 3.4 gives the techniques to do such plots systematically).

■

Proof. To prove the angle Theorem 1.3.5, form a triangle from vectors $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{u}-\boldsymbol{v}$ as illustrated in the margin. Recall and apply the cosine rule for triangles

$$|\boldsymbol{u}-\boldsymbol{v}|^2 = |\boldsymbol{u}|^2 + |\boldsymbol{v}|^2 - 2|\boldsymbol{u}||\boldsymbol{v}|\cos\theta\,.$$

In $\mathbb{R}^n$ this rule rearranges to

$$2|\boldsymbol{u}||\boldsymbol{v}|\cos\theta = |\boldsymbol{u}|^2 + |\boldsymbol{v}|^2 - |\boldsymbol{u}-\boldsymbol{v}|^2$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$
\begin{aligned}
&= u_1^2 + u_2^2 + \cdots + u_n^2 + v_1^2 + v_2^2 + \cdots + v_n^2 \\
&\quad - (u_1 - v_1)^2 - (u_2 - v_2)^2 - \cdots - (u_n - v_n)^2 \\
&= u_1^2 + u_2^2 + \cdots + u_n^2 + v_1^2 + v_2^2 + \cdots + v_n^2 \\
&\quad - u_1^2 + 2u_1v_1 - v_1^2 - u_2^2 + 2u_2v_2 - v_2^2 \\
&\quad - \cdots - u_n^2 + 2u_nv_n - v_n^2 \\
&= 2u_1v_1 + 2u_2v_2 + \cdots + 2u_nv_n \\
&= 2(u_1v_1 + u_2v_2 + \cdots + u_nv_n) \\
&= 2\boldsymbol{u} \cdot \boldsymbol{v}\,.
\end{aligned}
$$

Dividing both sides by $2|\boldsymbol{u}||\boldsymbol{v}|$ gives $\cos\theta = \frac{\boldsymbol{u}\cdot\boldsymbol{v}}{|\boldsymbol{u}||\boldsymbol{v}|}$ as required. □

1.3.1 Work done involves the dot product

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

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

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

In general, the applied force and the displacement are not in the same direction (as illustrated in the margin). Consider the general case when a vector force $\boldsymbol{F}$ 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 in the margin, 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 $W = F_0|\boldsymbol{d}| = \frac{\boldsymbol{F}\cdot\boldsymbol{d}}{|\boldsymbol{d}|}|\boldsymbol{d}| = \boldsymbol{F}\cdot\boldsymbol{d}$, the dot product of the vector force and vector displacement.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Example 1.3.10. A sailing boat travels a distance of 40 m East and 10 m North, as drawn in the margin. The wind from abeam of strength and direction $(1, -4)$ m/s generates a force $\boldsymbol{F} = (20, -10)$ (newtons) on the sail, as drawn. What is the work done by the wind?

Solution: The direction of the wind is immaterial except for the force it generates. The displacement vector $\boldsymbol{d} = (40, 10)$ m. Then the work done is $W = \boldsymbol{F} \cdot \boldsymbol{d} = (40, 10) \cdot (20, -10) = 800 - 100 = 700$ joules.

■

Activity 1.3.11. Recall the force of gravity on an object is the mass of the object time 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) 29.4 joules
(c) 58.8 joules
(d) 98 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 (properties 1.3.13c and 1.3.13d)

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

Solution:

- For the first:

$$\begin{aligned}
a(\boldsymbol{u} \cdot \boldsymbol{v}) &= 2\big((-2, 5, -2) \cdot (3, 3, -2)\big) \\
&= 2\big((-2)3 + 5 \cdot 3 + (-2)(-2)\big) \\
&= 2 \cdot 13 = 26\,; \\
(a\boldsymbol{u}) \cdot \boldsymbol{v} &= (-4, 10, -4) \cdot (3, 3, -2) \\
&= (-4)3 + 10 \cdot 3 + (-4)(-2) \\
&= 26\,; \\
\boldsymbol{u} \cdot (a\boldsymbol{v}) &= (-2, 5, -2) \cdot (6, 6, -4)
\end{aligned}$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$= (-2)6 + 5 \cdot 6 + (-2)(-4)$$
$$= 26\,.$$

These three are equal.

- For the second:

$$\begin{aligned}(\boldsymbol{u}+\boldsymbol{v})\cdot\boldsymbol{w} &= (1,8,-4)\cdot(2,0,-5)\\ &= 1\cdot 2 + 8\cdot 0 + (-4)(-5)\\ &= 22\,;\\ \boldsymbol{u}\cdot\boldsymbol{w}+\boldsymbol{v}\cdot\boldsymbol{w} &= (-2,5,-2)\cdot(2,0,-5)\\ &\quad + (3,3,-2)\cdot(2,0,-5)\\ &= \big[(-2)2 + 5\cdot 0 + (-2)(-5)\big]\\ &\quad + \big[3\cdot 2 + 3\cdot 0 + (-2)(-5)\big]\\ &= 6 + 16 = 22\,.\end{aligned}$$

These are both equal.

■

Theorem 1.3.13 (dot properties). *Let $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{w}$ be any vectors in $\mathbb{R}^n$, and let a be any scalar. 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* $\boldsymbol{u} = \boldsymbol{0}$*.*

Proof. Here prove only the commutative law 1.3.13a and the inequality 1.3.13e. Exercise 1.3.6 asks you to analogously prove the other properties. At the core of each proof is the definition of the dot product which empowers us to deduce a property via the corresponding property for scalars.

- To prove the commutative law 1.3.13a consider

$$\begin{aligned}\boldsymbol{u}\cdot\boldsymbol{v} &= u_1v_1 + u_2v_2 + \cdots + u_nv_n \quad \text{(by Defn. 1.3.2)}\\ &\qquad \text{(using each scalar mult. is commutative)}\\ &= v_1u_1 + v_2u_2 + \cdots + v_nu_n\\ &= \boldsymbol{v}\cdot\boldsymbol{u} \quad \text{(by Defn. 1.3.2)}.\end{aligned}$$

- To prove the inequality 1.3.13e consider

$$\boldsymbol{u}\cdot\boldsymbol{u} = u_1u_1 + u_2u_2 + \cdots + u_nu_n \quad \text{(by Defn. 1.3.2)}$$

v0.2a

$$
\begin{aligned}
&= u_1^2 + u_2^2 + \cdots + u_n^2 \\
&\geq 0 + 0 + \cdots + 0 \quad \text{(as each scalar term is } \geq 0) \\
&= 0\,.
\end{aligned}
$$

To prove the "moreover" part, first consider the zero vector. From Definition 1.3.2, in $\mathbb{R}^n$,

$$\mathbf{0} \cdot \mathbf{0} = 0^2 + 0^2 + \cdots + 0^2 = 0\,.$$

Second, let vector $\boldsymbol{u} = (u_1\,, u_2\,, \ldots, u_n)$ in $\mathbb{R}^n$ satisfy $\boldsymbol{u} \cdot \boldsymbol{u} = 0$. Then we know that

$$\underbrace{u_1^2}_{\geq 0} + \underbrace{u_2^2}_{\geq 0} + \cdots + \underbrace{u_n^2}_{\geq 0} = 0\,.$$

Being squares, all terms on the left are non-negative, so the only way they can all add to zero is if they are all zero. That is, $u_1 = u_2 = \cdots = u_n = 0$. Hence, the vector $\boldsymbol{u}$ must be the zero vector $\mathbf{0}$.

□

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

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

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

(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 1.3.13e, that $\boldsymbol{u} \cdot \boldsymbol{u} = 0$ if and only if $\boldsymbol{u} = \mathbf{0}$, may look uncannily familiar. The reason is that this last part is essentially the same as the proof of Theorem 1.1.13 that the zero vector is the only vector of length zero. The upcoming Theorem 1.3.17 establishes that this connection between dot products and lengths is no coincidence.

Example 1.3.15. For the two vectors $\boldsymbol{u} = (3, 4)$ 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).

Solution: (a) Here $\sqrt{\boldsymbol{u} \cdot \boldsymbol{u}} = \sqrt{3 \cdot 3 + 4 \cdot 4} = \sqrt{25} = 5$, whereas the length $|\boldsymbol{u}| = \sqrt{3^2 + 4^2} = \sqrt{25} = 5$ (Definition 1.1.9). These expressions are equal.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(b) Here $|\boldsymbol{u}\cdot\boldsymbol{v}| = |3\cdot 2+4\cdot 1| = 10$, whereas $|\boldsymbol{u}||\boldsymbol{v}| = 5\sqrt{2^2+1^2} = 5\sqrt{5} = 11.180$. Hence $|\boldsymbol{u}\cdot\boldsymbol{v}| = 10 \le 11.180 = |\boldsymbol{u}||\boldsymbol{v}|$.

(c) Here $|\boldsymbol{u}+\boldsymbol{v}| = |(5,5)| = \sqrt{5^2+5^2} = \sqrt{50} = 7.071$, whereas $|\boldsymbol{u}|+|\boldsymbol{v}| = 5+\sqrt{5} = 7.236$. Hence $|\boldsymbol{u}+\boldsymbol{v}| = 7.071 \le 7.236 = |\boldsymbol{u}|+|\boldsymbol{v}|$. This is called the triangle inequality because the vectors $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{u}+\boldsymbol{v}$ may be viewed as forming a triangle, as illustrated in the margin, and this inequality follows because the length of any side of a triangle must be less than the sum of the other two sides.

■

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

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 $\boldsymbol{v} = (2,1)$ and consider 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 in the margin. 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 $b^2-4ac < 0$, written 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) $|\boldsymbol{u}\cdot\boldsymbol{v}| \le |\boldsymbol{u}||\boldsymbol{v}|$ (**Cauchy–Schwarz inequality**);*

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

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Proof. Each property depends upon the previous.

1.3.17a

$$\begin{aligned}\sqrt{\boldsymbol{u}\cdot\boldsymbol{u}} &= \sqrt{u_1u_1+u_2u_2+\cdots+u_nu_n} \quad \text{(by Defn. 1.3.2)}\\ &= \sqrt{u_1^2+u_2^2+\cdots+u_n^2}\\ &= |\boldsymbol{u}| \quad \text{(by Defn. 1.1.9).}\end{aligned}$$

1.3.17b To prove the Cauchy–Schwarz inequality between vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ first consider the trivial case when $\boldsymbol{v}=\boldsymbol{0}$: then the left-hand side $|\boldsymbol{u}\cdot\boldsymbol{v}| = |\boldsymbol{u}\cdot\boldsymbol{0}| = |\boldsymbol{0}| = 0$; whereas the right-hand side $|\boldsymbol{u}||\boldsymbol{v}| = |\boldsymbol{u}||\boldsymbol{0}| = |\boldsymbol{u}|0 = 0$; and so the inequality $|\boldsymbol{u}\cdot\boldsymbol{v}| \le |\boldsymbol{u}||\boldsymbol{v}|$ is satisfied in this case.

Second, for the case when $\boldsymbol{v}\ne\boldsymbol{0}$, consider the line given parametrically by $\boldsymbol{x}=\boldsymbol{u}+t\boldsymbol{v}$ for (real) scalar parameter t, as illustrated in the margin. The distance ℓ of a point on the line from the origin is the length of its position vector, and by property 1.3.17a

$$\begin{aligned}\ell^2 &= \boldsymbol{x}\cdot\boldsymbol{x}\\ &= (\boldsymbol{u}+t\boldsymbol{v})\cdot(\boldsymbol{u}+t\boldsymbol{v})\\ &\qquad \text{(then using distibutivity 1.3.13d)}\\ &= \boldsymbol{u}\cdot(\boldsymbol{u}+t\boldsymbol{v}) + (t\boldsymbol{v})\cdot(\boldsymbol{u}+t\boldsymbol{v})\\ &\qquad \text{(again using distibutivity 1.3.13d)}\\ &= \boldsymbol{u}\cdot\boldsymbol{u} + \boldsymbol{u}\cdot(t\boldsymbol{v}) + (t\boldsymbol{v})\cdot\boldsymbol{u} + (t\boldsymbol{v})\cdot(t\boldsymbol{v})\\ &\qquad \text{(using scalar mult. property 1.3.13c)}\\ &= \boldsymbol{u}\cdot\boldsymbol{u} + t(\boldsymbol{u}\cdot\boldsymbol{v}) + t(\boldsymbol{v}\cdot\boldsymbol{u}) + t^2(\boldsymbol{v}\cdot\boldsymbol{v})\\ &\qquad \text{(using 1.3.17a and commutativity 1.3.13a)}\\ &= |\boldsymbol{u}|^2 + 2(\boldsymbol{u}\cdot\boldsymbol{v})t + |\boldsymbol{v}|^2t^2\\ &= at^2+bt+c,\end{aligned}$$

a quadratic in t, with coefficients $a=|\boldsymbol{v}|^2>0$, $b=2(\boldsymbol{u}\cdot\boldsymbol{v})$, and $c=|\boldsymbol{u}|^2$. Since $\ell^2\ge 0$ (it may be zero if the line goes through the origin), then this quadratic in t has either no zeros or just one zero. By the properties of quadratic equations, the discriminant $b^2-4ac\le 0$, that is, $\frac{1}{4}b^2\le ac$. Substituting the particular coefficients here gives $\frac{1}{4}\big[2(\boldsymbol{u}\cdot\boldsymbol{v})\big]^2 = (\boldsymbol{u}\cdot\boldsymbol{v})^2 \le |\boldsymbol{v}|^2|\boldsymbol{u}|^2$. Taking the square-root of both sides then establishes the Cauchy–Schwarz inequality $|\boldsymbol{u}\cdot\boldsymbol{v}|\le|\boldsymbol{u}||\boldsymbol{v}|$.

1.3.17c To prove the triangle inequality between vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ first observe the Cauchy–Schwarz inequality implies $(\boldsymbol{u}\cdot\boldsymbol{v})\le|\boldsymbol{u}||\boldsymbol{v}|$ (since the left-hand side has magnitude $\le$ the right-hand side). Then consider (analogous to the $t=1$ case of the above)

$$\begin{aligned}|\boldsymbol{u}+\boldsymbol{v}|^2 &= (\boldsymbol{u}+\boldsymbol{v})\cdot(\boldsymbol{u}+\boldsymbol{v})\\ &\qquad \text{(then using distibutivity 1.3.13d)}\end{aligned}$$

$$
\begin{aligned}
&= \boldsymbol{u}\cdot(\boldsymbol{u}+\boldsymbol{v}) + \boldsymbol{v}\cdot(\boldsymbol{u}+\boldsymbol{v}) \\
&\qquad \text{(again using distibutivity 1.3.13d)} \\
&= \boldsymbol{u}\cdot\boldsymbol{u} + \boldsymbol{u}\cdot\boldsymbol{v} + \boldsymbol{v}\cdot\boldsymbol{u} + \boldsymbol{v}\cdot\boldsymbol{v} \\
&\qquad \text{(using 1.3.17a and commutativity 1.3.13a)} \\
&= |\boldsymbol{u}|^2 + 2(\boldsymbol{u}\cdot\boldsymbol{v}) + |\boldsymbol{v}|^2 \\
&\qquad \text{(using Cauchy–Schwarz inequality)} \\
&\le |\boldsymbol{u}|^2 + 2|\boldsymbol{u}||\boldsymbol{v}| + |\boldsymbol{v}|^2 \\
&\quad = (|\boldsymbol{u}| + |\boldsymbol{v}|)^2 .
\end{aligned}
$$

Take the square-root of both sides to establish the triangle inequality $|\boldsymbol{u}+\boldsymbol{v}| \le |\boldsymbol{u}| + |\boldsymbol{v}|$.

The minus case follows because $|\boldsymbol{u}-\boldsymbol{v}| = |\boldsymbol{u}+(-\boldsymbol{v})| \le |\boldsymbol{u}| + |-\boldsymbol{v}| = |\boldsymbol{u}| + |\boldsymbol{v}|$.

□

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

Solution: We need the length of the vectors:

$$
\begin{aligned}
|\boldsymbol{a}| &= \sqrt{(-1)^2 + (-2)^2 + 1^2 + 3^2 + (-2)^2} \\
&= \sqrt{19} = 4.3589, \\
|\boldsymbol{b}| &= \sqrt{(-3)^2 + (-2)^2 + 10^2 + 2^2 + 2^2} \\
&= \sqrt{121} = 11\,.
\end{aligned}
$$

also, the dot product

$$
\begin{aligned}
\boldsymbol{a}\cdot\boldsymbol{b} &= (-1)(-3) + (-2)(-2) + 1\cdot 10 + 3\cdot 2 + (-2)2 \\
&= 19\,.
\end{aligned}
$$

Hence $|\boldsymbol{a}\cdot\boldsymbol{b}| = 19 < 47.948 = |\boldsymbol{a}||\boldsymbol{b}|$, which verifies the Cauchy–Schwarz inequality.

Now, the length of the sum

$$
\begin{aligned}
|\boldsymbol{a}+\boldsymbol{b}| &= |(-4,-4,11,5,0)| \\
&= \sqrt{(-4)^2 + (-4)^2 + 11^2 + 5^2 + 0^2} \\
&= \sqrt{178} = 13.342\,.
\end{aligned}
$$

Here $|\boldsymbol{a}+\boldsymbol{b}| = 13.342$ whereas $|\boldsymbol{a}|+|\boldsymbol{b}| = 11+\sqrt{19} = 15.359$. Hence, here $|\boldsymbol{a}+\boldsymbol{b}| \le |\boldsymbol{a}| + |\boldsymbol{b}|$ which verifies the triangle inequality.

■

v0.2a

1.3.3 Orthogonal vectors are at right-angles

Of all the angles that vectors can make with each other, the two most important angles are, firstly, when the vectors are aligned with each other, and secondly, when the vectors are at right-angles to each other. Recall Theorem 1.3.5 gives the angle θ between two vectors via $\cos\theta = \frac{\boldsymbol{u}\cdot\boldsymbol{v}}{|\boldsymbol{u}||\boldsymbol{v}|}$. For vectors at right-angles $\theta = 90^\circ$ and so $\cos\theta = 0$ and hence non-zero 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.

The term 'orthogonal' derives from the Greek for 'right-angled'.

Definition 1.3.19. *Two vectors $\boldsymbol{u}$ 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 *non-zero* vectors. Often the requirement that the orthogonal vectors are non-zero is explicitly made, but beware that sometimes the requirement may be implicit in the problem.

Example 1.3.20. The standard unit vectors (Definition 1.2.7) are orthogonal to each other. For example, consider the standard unit vectors $\boldsymbol{i}$, $\boldsymbol{j}$ 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 $\boldsymbol{w} = (-1,6,-2,3)$.

Solution: Use the dot product.

- $\boldsymbol{u}\cdot\boldsymbol{v} = (-1,1,-3,0)\cdot(2,4,2,-6) = -2+4-6+0 = -4 \neq 0$ so this pair are not perpendicular.
- $\boldsymbol{u}\cdot\boldsymbol{w} = (-1,1,-3,0)\cdot(-1,6,-2,3) = 1+6+6+0 = 13 \neq 0$ so this pair are not perpendicular.
- $\boldsymbol{v}\cdot\boldsymbol{w} = (2,4,2,-6)\cdot(-1,6,-2,3) = -2+24-4-18 = 0$ so this pair are the only two vectors perpendicular to each other.

■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Activity 1.3.22. Which pair of the following 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) $\boldsymbol{y}$, $\boldsymbol{z}$ (b) $\boldsymbol{x}$, $\boldsymbol{z}$ (c) $\boldsymbol{x}$, $\boldsymbol{y}$ (d) no pair

■

Example 1.3.23. Find the 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.

Solution: For vectors to be at right-angles, their dot product must be zero. Hence find b such that

$$0 = \boldsymbol{a} \cdot \boldsymbol{b} = (\boldsymbol{i} + 4\boldsymbol{j} + 2\boldsymbol{k}) \cdot (\boldsymbol{i} + b\boldsymbol{j} - 3\boldsymbol{k}) = 1 + 4b - 6 = 4b - 5\,.$$

Solving $0 = 4b - 5$ gives $b = 5/4$. That is, $\boldsymbol{i} + \frac{5}{4}\boldsymbol{j} - 3\boldsymbol{k}$ is at right-angles to $\boldsymbol{i} + 4\boldsymbol{j} + 2\boldsymbol{k}$.

■

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

To introduce the first theorem, consider the 2D plane and try to draw a non-zero vector at right-angles to both the two standard unit vectors $\boldsymbol{i}$ and $\boldsymbol{j}$. The red vectors in the margin illustrate three failed attempts to draw a 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 non-zero vector orthogonal to all n standard unit vectors in $\mathbb{R}^n$.*

Proof. Let $\boldsymbol{u} = (u_1\,, u_2\,, \dots, u_n)$ be a vector in $\mathbb{R}^n$ that is orthogonal to all n standard unit vectors. Then by Definition 1.3.19 of orthogonality:

- $0 = \boldsymbol{u}\cdot\boldsymbol{e}_1 = (u_1,u_2,\dots,u_n)\cdot(1,0,\dots,0) = u_1+0+\cdots+0 = u_1$, and so the first component must be zero;
- $0 = \boldsymbol{u}\cdot\boldsymbol{e}_2 = (u_1,u_2,\dots,u_n)\cdot(1,0,\dots,0) = 0+u_2+0+\cdots+0 = u_2$, and so the second component must be zero; and so on to
- $0 = \boldsymbol{u}\cdot\boldsymbol{e}_n = (u_1,u_2,\dots,u_n)\cdot(0,0,\dots,1) = 0+0+\cdots+u_n = u_n$, and so the last component must be zero.

Since $u_1 = u_2 = \cdots = u_n = 0$ the only vector that is orthogonal to all the standard unit vectors is $\boldsymbol{u} = \boldsymbol{0}$, the zero vector. □

To introduce the second theorem, imagine trying to draw three unit vectors in any orientation in the 2D plane such that all three are at

right-angles to each other. The margin illustrates one attempt. It cannot be done. There are at most two vectors in 2D that are all at right-angles to each other.

Theorem 1.3.25 (orthogonal completeness)**.** *In a set of orthogonal unit vectors in $\mathbb{R}^n$, there can be no more than n vectors in the set.* [4]

Proof. Use contradiction. Suppose there are more than n orthogonal unit vectors in the set. Define a coordinate system for $\mathbb{R}^n$ using the first n of the given unit vectors as the n standard unit vectors (as illustrated for $\mathbb{R}^2$ in the margin). Theorem 1.3.24 then says there cannot be any more non-zero vectors orthogonal than these n standard unit vectors. This contradicts there being more than n orthogonal unit vectors. To avoid this contradiction the supposition must be wrong; that is, there cannot be more than n orthogonal unit vectors in $\mathbb{R}^n$. □

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 those at right-angles to the direction perpendicular to the plane, 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).

Solution: Recall that vectors at right-angles have a zero dot product (section 1.3.3). Thus the position vector $\boldsymbol{x}$ of any point in the line satisfies the dot product $\boldsymbol{x} \cdot (2, 3) = 0$. For $\boldsymbol{x} = (x, y)$, as illustrated in the margin, $\boldsymbol{x} \cdot (2, 3) = 2x + 3y$ so the equation of the line is $2x + 3y = 0$.

When the line goes through $(4, 1)$ (instead of the origin), then it is the displacement $\boldsymbol{x} - (4, 1)$ that must be orthogonal to $(2, 3)$, as illustrated. That is, the equation of the line is $(x-4, y-1)\cdot(2, 3) = 0$. Evaluating the dot product gives $2(x - 4) + 3(y - 1) = 0$; that is, $2x + 3y = 2 \cdot 4 + 3 \cdot 1 = 11$ is an equation of the line.

■

[4] For the pure at heart, this property forms part of the definition of what we mean by $\mathbb{R}^n$. The representation of a vector in $\mathbb{R}^n$ by n components (here Definition 1.1.4) then follows as a consequence, instead of vice-versa.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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

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

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

■

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 below, 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, and 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.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Solution: Letting $\boldsymbol{x} = (x, y, z)$ be the coordinates of a point in the plane, the above argument asserts an equation of the plane is $\boldsymbol{n} \cdot (\boldsymbol{x} - \overrightarrow{OP}) = 0$ which becomes $1(x-1) - 1(y-1) + 3(z-2) = 0$; that is, $x - 1 - y + 1 + 3z - 6 = 0$, which rearranged is $x - y + 3z = 6$.

To find some points in the plane, rearrange this equation to $z = 2 - x/3 + y/3$ and then substitute any values for x and y: $x = y = 0$ gives $z = 2$ so $(0, 0, 2)$ is on the plane; $x = 3$ and $y = 0$ gives $z = 1$ so $(3, 0, 1)$ is on the plane; $x = 2$ and $y = -2$ gives $z = 2/3$ so $(2, -2, \frac{2}{3})$ is on the plane; and so on.

■

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

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

Solution:

(a) In this standard form $3x - 6y + 2z = 4$ a normal vector is the coefficients of the variables, $\boldsymbol{n} = (3, -6, 2)$ (or any scalar multiple).

(b) Rearrange $z = 0.2x - 3.3y - 1.9$ to standard form $-0.2x + 3.3y + z = -1.9$ then a normal is $\boldsymbol{n} = (-0.2, 3.3, 1)$ (or any multiple).

■

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

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

(c) $(1\,,2\,,4)$ (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 generalises to any dimensions (Section 2.3).

The basic idea, as illustrated in the margin, is that given any plane (through the origin for the moment), then choosing almost any two vectors in the plane allows us to write all points in the plane as a sum of multiples of the two vectors. With the given vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ shown in the margin, 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

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

in the form $s\boldsymbol{u} + t\boldsymbol{v}$ for some scalar parameters s and t. The grid shown in the margin illustrates the sum of integral and half-integral multiples. The formula $\boldsymbol{x} = s\boldsymbol{u} + t\boldsymbol{v}$ 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 below in stereo.

Solution: This plane does not pass through the origin, so we first choose a point and make the description relative to that point: say choose point P with position vector $\boldsymbol{p} = \overrightarrow{OP} = -\boldsymbol{i} + 2\boldsymbol{j} + 3\boldsymbol{k}$. Then, as illustrated below, two vectors parallel to the required plane are

$$\begin{aligned}
\boldsymbol{u} &= \overrightarrow{PQ} = \overrightarrow{OQ} - \overrightarrow{OP} \\
&= (2\boldsymbol{i} + 3\boldsymbol{j} + 2\boldsymbol{k}) - (-\boldsymbol{i} + 2\boldsymbol{j} + 3\boldsymbol{k}) \\
&= 3\boldsymbol{i} + \boldsymbol{j} - \boldsymbol{k}, \\
\boldsymbol{v} &= \overrightarrow{PR} = \overrightarrow{OR} - \overrightarrow{OP} \\
&= (4\boldsymbol{j} + 5\boldsymbol{k}) - (-\boldsymbol{i} + 2\boldsymbol{j} + 3\boldsymbol{k}) \\
&= \boldsymbol{i} + 2\boldsymbol{j} + 2\boldsymbol{k}.
\end{aligned}$$

Lastly, every point in the plane is the sum of the displacement vector $\boldsymbol{p}$ and arbitrary multiples of the parallel vectors $\boldsymbol{u}$ and $\boldsymbol{v}$. That is, a parametric equation of the plane is $\boldsymbol{x} = \boldsymbol{p} + s\boldsymbol{u} + t\boldsymbol{v}$ which here is

$$\begin{aligned}
\boldsymbol{x} &= (-\boldsymbol{i} + 2\boldsymbol{j} + 3\boldsymbol{k}) + s(3\boldsymbol{i} + \boldsymbol{j} - \boldsymbol{k}) + t(\boldsymbol{i} + 2\boldsymbol{j} + 2\boldsymbol{k}) \\
&= (-1 + 3s + t)\boldsymbol{i} + (2 + s + 2t)\boldsymbol{j} + (3 - s + 2t)\boldsymbol{k}.
\end{aligned}$$

■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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 non-zero 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 by using vectors with the corresponding number of components.

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

Solution: First choose a point and make the description relative to that point: say choose point P with position vector $\boldsymbol{p} = \overrightarrow{OP} = 6\boldsymbol{i} - 4\boldsymbol{j} + 3\boldsymbol{k}$. Then, as illustrated below, two vectors parallel to the required plane are

$$\begin{aligned}
\boldsymbol{u} &= \overrightarrow{PQ} = \overrightarrow{OQ} - \overrightarrow{OP} \\
&= (-4\boldsymbol{i} - 18\boldsymbol{j} + 7\boldsymbol{k}) - (6\boldsymbol{i} - 4\boldsymbol{j} + 3\boldsymbol{k}) \\
&= -10\boldsymbol{i} - 14\boldsymbol{j} + 4\boldsymbol{k}, \\
\boldsymbol{v} &= \overrightarrow{PR} = \overrightarrow{OR} - \overrightarrow{OP} \\
&= (11\boldsymbol{i} + 3\boldsymbol{j} + \boldsymbol{k}) - (6\boldsymbol{i} - 4\boldsymbol{j} + 3\boldsymbol{k}) \\
&= 5\boldsymbol{i} + 7\boldsymbol{j} - 2\boldsymbol{k}.
\end{aligned}$$

Oops: notice that $\boldsymbol{u} = -2\boldsymbol{v}$ so the vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are not at a nontrivial angle; instead they are aligned along a line because the three points P, Q and R are collinear. There are an infinite number of planes passing through such collinear points. Hence we cannot answer the question which requires "the plane".

■

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

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Solution: First choose a point and make the description relative to that point: say choose point A with position vector $\boldsymbol{a} = \overrightarrow{OA} = -1.2\boldsymbol{i} + 2.4\boldsymbol{j} + 0.8\boldsymbol{k}$. Then, as illustrated below, two vectors parallel to the required plane are

$$\begin{aligned}
\boldsymbol{u} &= \overrightarrow{AB} = \overrightarrow{OB} - \overrightarrow{OA} \\
&= (1.6\boldsymbol{i} + 1.4\boldsymbol{j} + 2.4\boldsymbol{k}) - (-1.2\boldsymbol{i} + 2.4\boldsymbol{j} + 0.8\boldsymbol{k}) \\
&= 2.8\boldsymbol{i} - \boldsymbol{j} + 1.6\boldsymbol{k}\,, \\
\boldsymbol{v} &= \overrightarrow{AC} = \overrightarrow{OC} - \overrightarrow{OA} \\
&= (0.2\boldsymbol{i} - 0.4\boldsymbol{j} - 2.5\boldsymbol{k}) - (-1.2\boldsymbol{i} + 2.4\boldsymbol{j} + 0.8\boldsymbol{k}) \\
&= 1.4\boldsymbol{i} - 2.8\boldsymbol{j} - 3.3\boldsymbol{k}\,.
\end{aligned}$$

Lastly, every point in the plane is the sum of the displacement vector $\boldsymbol{a}$, and arbitrary multiples of the parallel vectors $\boldsymbol{u}$ and $\boldsymbol{v}$. That is, a parametric equation of the plane is $\boldsymbol{x} = \boldsymbol{a} + s\boldsymbol{u} + t\boldsymbol{v}$ which here is

$$\begin{aligned}
\boldsymbol{x} &= \begin{bmatrix} -1.2 \\ 2.4 \\ 0.8 \end{bmatrix} + s \begin{bmatrix} 2.8 \\ -1 \\ 1.6 \end{bmatrix} + t \begin{bmatrix} 1.4 \\ -2.8 \\ -3.3 \end{bmatrix} \\
&= \begin{bmatrix} -1.2 + 2.8s + 1.4t \\ 2.4 - s - 2.8t \\ 0.8 + 1.6s - 3.3t \end{bmatrix}.
\end{aligned}$$

■

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

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

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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

(c) $(3s+2t, 4+2s+t, 4+3t)$

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

■

1.3.5 Exercises

Exercise 1.3.1. Following Example 1.3.1, use the cosine rule for triangles to find the angle between the following pairs of vectors. Confirm that $|\boldsymbol{u}||\boldsymbol{v}|\cos\theta = \boldsymbol{u}\cdot\boldsymbol{v}$ in each case.

(a) $(6\,,5)$ and $(-3\,,1)$

(b) $(6\,,2\,,2)$ and $(-1\,,-2\,,5)$

(c) $(2\,,2.9)$ and $(-1.4\,,0.8)$

(d) $(-3.6\,,0\,,-0.7)$ and $(1.2\,,-0.9\,,-0.6)$

Exercise 1.3.2. Which of the following pairs of vectors appear orthogonal?

(a)

(b)

(c)

(d)

(e)

(f)

v0.2a

(g)

(h)

Exercise 1.3.3. Recall that Example 1.1.7 represented the following sentences by word vectors $\boldsymbol{w} = (N_{\text{cat}}, N_{\text{dog}}, N_{\text{mat}}, N_{\text{sat}}, N_{\text{scratched}})$.

- "The cat and dog sat on the mat" is summarised by the vector $\boldsymbol{a} = (1, 1, 1, 1, 0)$.
- "The dog scratched" is summarised by the vector $\boldsymbol{b} = (0, 1, 0, 0, 1)$.
- "The dog sat on the mat; the cat scratched the dog." is summarised by the vector $\boldsymbol{c} = (1, 2, 1, 1, 1)$.

Find the similarity between pairs of these sentences by calculating the angle between each pair of word vectors. What is the most similar pair of sentences?

Exercise 1.3.4. Recall Exercise 1.1.4 found word vectors in $\mathbb{R}^7$ for the titles of eight books that The Society of Industrial and Applied Mathematics (SIAM) reviewed recently. The following four titles have more than one word counted in the word vectors.

(a) Introduction to Finite and Spectral Element Methods using MATLAB

(b) Iterative Methods for Linear Systems: Theory and Applications

(c) Singular Perturbations: Introduction to System Order Reduction Methods with Applications

(d) Stochastic Chemical Kinetics: Theory and Mostly Systems Biology Applications

Find the similarity between pairs of these titles by calculating the angle between each pair of corresponding word vectors in $\mathbb{R}^7$. What is the most similar pair of titles? What is the most dissimilar titles?

Exercise 1.3.5. Suppose two non-zero word vectors are orthogonal. Explain what such orthogonality means in terms of the words of the original sentences.

Exercise 1.3.6. For the properties of the dot product, Theorem 1.3.13, prove some properties chosen from 1.3.13b–1.3.13d.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Exercise 1.3.7. Verify the Cauchy–Schwarz inequality (+ case) and also the triangle inequality for the following pairs of vectors.

(a) $(2\,,-4\,,4)$ and $(6\,,7\,,6)$

(b) $(1\,,-2\,,2)$ and $(-3\,,6\,,-6)$

(c) $(-2\,,-3\,,6)$ and $(3\,,1\,,2)$

(d) $(3\,,-5\,,-1\,,-1)$ and $(1\,,-1\,,-1\,,-1)$

(e) $\begin{bmatrix}-0.2\\0.8\\-3.8\\-0.3\end{bmatrix}$ and $\begin{bmatrix}2.4\\-5.2\\5.0\\1.9\end{bmatrix}$

(f) $\begin{bmatrix}0.8\\0.8\\6.6\\-1.5\end{bmatrix}$ and $\begin{bmatrix}4.4\\-0.6\\2.1\\2.2\end{bmatrix}$

Exercise 1.3.8. Find an equation of the plane with the given normal vector $\boldsymbol{n}$ and through the given point P.

(a) $P=(1\,,2\,,-3)$, $\boldsymbol{n}=(2\,,-5\,,-2)$.

(b) $P=(5\,,-4\,,-13)$, $\boldsymbol{n}=(-1\,,0\,,-1)$.

(c) $P=(10\,,-4\,,-1)$, $\boldsymbol{n}=(-2\,,4\,,5)$.

(d) $P=(2\,,-5\,,-1)$, $\boldsymbol{n}=(4\,,9\,,-4)$.

(e) $P=(1.7\,,-4.2\,,2.2)$, $\boldsymbol{n}=(1\,,0\,,4)$.

(f) $P=(3\,,5\,,-2)$, $\boldsymbol{n}=(-2.5\,,-0.5\,,0.4)$.

(g) $P=(-7.3\,,-1.6\,,5.8)$, $\boldsymbol{n}=(-2.8\,,-0.8\,,4.4)$.

(h) $P=(0\,,-1.2\,,2.2)$, $\boldsymbol{n}=(-1.4\,,-8.1\,,-1.5)$.

Exercise 1.3.9. Write down a normal vector to the plane described by each of the following equations.

(a) $2x+3y+2z=6$

(b) $-7x-2y+4=-5z$

(c) $-12x_1+2x_2+2x_3-8=0$

(d) $2x_3=8x_1+5x_2+1$

(e) $0.1x=1.5y+1.1z+0.7$

(f) $-5.5x_1+1.6x_2=6.7x_3-1.3$

Exercise 1.3.10. For each case, find a parametric equation of the plane through the three given points.

(a) $(0\,,5\,,-4)$, $(-3\,,-2\,,2)$, $(5\,,1\,,-3)$.

(b) $(0\,,-1\,,-1)$, $(-4\,,1\,,-5)$, $(0\,,-3\,,-2)$.

(c) $(2\,,2\,,3)$, $(2\,,3\,,3)$, $(3\,,1\,,0)$.

(d) $(-1\,,2\,,2)$, $(0\,,1\,,-1)$, $(1\,,0\,,-4)$.

(e) $(0.4\,,-2.2\,,8.7)$, $(-2.2\,,1.3\,,-4.9)$, $(-1.4\,,3.2\,,-0.4)$.

(f) $(2.2\,,-6.7\,,2)$, $(-2.6\,,-1.6\,,-0.5)$, $(2.9\,,5.4\,,-0.6)$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(g) $(-5.6\,,-2.2\,,-6.8)$,
$(-1.8\,,4.3\,,-3.9)$,
$(2.5\,,-3.5\,,-1.7)$,

(h) $(1.8\,,-0.2\,,-0.7)$,
$(-1.6\,,2\,,-3.7)$,
$(1.4\,,-0.5\,,0.5)$,

Exercise 1.3.11. For each case of Exercise 1.3.10 that you have done, find two other parametric equations of the plane.

v0.2a

1.4 The cross product

Section Contents

Area of a parallelogram 62
Normal vector to a plane 63
Definition of a cross product 64
Geometry of a cross product 66
Algebraic properties of a cross product 70
Volume of a parallelepiped 71
1.4.1 Exercises 74

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

In the three dimensions of the world we live in, as well as the dot product there is another way to multiply vectors, called the cross product. For more than three dimensions, qualitatively different techniques are developed in subsequent chapters.

Area of a parallelogram

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

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

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

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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

Solution: The parallelogram area $= |3\cdot 4 - 2\cdot(-1)| = |12+2| = 14$. The illustration indicates that this area must be about right as with imagination one could cut the area and move it about to form a rectangle roughly 3 by 5, and hence the area should be roughly 15.

■

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

(a) -16 (b) -4 (c) 4

(d) 11 (e) 16 (f) 19

■

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, 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 $x-y+3z = 6$.

Solution: They key to deriving the normal equation is to find that a normal vector to the plane is $(1,-1,3)$. This normal vector comes from the two vectors that multiply the parameters in the parametric form, $(1,1,0)$ and $(0,3,1)$. (Those who have seen 3×3 determinants will recognise the following has the same pattern—see Chapter 6.) Write the vectors as two consecutive columns, following a first column of the *symbols* of the standard unit vectors $\boldsymbol{i}$, $\boldsymbol{j}$ and $\boldsymbol{k}$, in

$$\boldsymbol{n} = \begin{vmatrix} \boldsymbol{i} & 1 & 0 \\ \boldsymbol{j} & 1 & 3 \\ \boldsymbol{k} & 0 & 1 \end{vmatrix}$$

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

$$= \boldsymbol{i}\begin{vmatrix} \cancel{\boldsymbol{i}} & \cancel{1} & \cancel{0} \\ \cancel{\boldsymbol{j}} & 1 & 3 \\ \cancel{\boldsymbol{k}} & 0 & 1 \end{vmatrix} - \boldsymbol{j}\begin{vmatrix} \boldsymbol{i} & 1 & 0 \\ \cancel{\boldsymbol{j}} & \cancel{1} & \cancel{3} \\ \cancel{\boldsymbol{k}} & 0 & 1 \end{vmatrix} + \boldsymbol{k}\begin{vmatrix} \boldsymbol{i} & 1 & 0 \\ \boldsymbol{j} & 1 & 3 \\ \cancel{\boldsymbol{k}} & \cancel{0} & \cancel{1} \end{vmatrix}$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$= \boldsymbol{i}\begin{vmatrix}1 & 3\\0 & 1\end{vmatrix} - \boldsymbol{j}\begin{vmatrix}1 & 0\\0 & 1\end{vmatrix} + \boldsymbol{k}\begin{vmatrix}1 & 0\\1 & 3\end{vmatrix}$$

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

$$= \boldsymbol{i}\cancel{\begin{vmatrix}1 & 3\\0 & 1\end{vmatrix}} - \boldsymbol{j}\cancel{\begin{vmatrix}1 & 0\\0 & 1\end{vmatrix}} + \boldsymbol{k}\cancel{\begin{vmatrix}1 & 0\\1 & 3\end{vmatrix}}$$

$$= \boldsymbol{i}(1\cdot 1 - 0\cdot 3) - \boldsymbol{j}(1\cdot 1 - 0\cdot 0) + \boldsymbol{k}(1\cdot 3 - 1\cdot 0)$$

$$= \boldsymbol{i} - \boldsymbol{j} + 3\boldsymbol{k}\,.$$

Using this normal vector, the equation of the plane must be of the form $x - y + 3z =$ constant. Since the plane goes through point $(1, 1, 2)$, the constant $= 1 - 1 + 3\cdot 2 = 6$; that is, the plane is $x - y + 3z = 6$ (as given).

■

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) $(5\,,6\,,7)$ (b) $(7\,,4\,,1)$

(c) $(-4\,,4\,,-10)$ (d) $(2\,,-2\,,5)$

(e) $(-7\,,-4\,,-1)$ (f) $(-4\,,-3\,,3)$

■

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

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

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

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

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

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

Solution: Using Definition 1.4.5:

$$
\begin{aligned}
\boldsymbol{i} \times \boldsymbol{j} &= (1,0,0) \times (0,1,0) \\
&= \boldsymbol{i}(0\cdot 0 - 0\cdot 1) + \boldsymbol{j}(0\cdot 0 - 1\cdot 0) + \boldsymbol{k}(1\cdot 1 - 0\cdot 0) \\
&= \boldsymbol{k}\,; \\
\boldsymbol{j} \times \boldsymbol{i} &= (0,1,0) \times (1,0,0) \\
&= \boldsymbol{i}(1\cdot 0 - 0\cdot 0) + \boldsymbol{j}(0\cdot 1 - 0\cdot 0) + \boldsymbol{k}(1\cdot 1 - 1\cdot 1) \\
&= -\boldsymbol{k}\,; \\
\boldsymbol{i} \times \boldsymbol{i} &= (1,0,0) \times (1,0,0) \\
&= \boldsymbol{i}(0\cdot 0 - 0\cdot 0) + \boldsymbol{j}(0\cdot 1 - 1\cdot 0) + \boldsymbol{k}(1\cdot 0 - 0\cdot 1) \\
&= \boldsymbol{0}\,.
\end{aligned}
$$

Exercise 1.4.1 asks you to correspondingly establish the other six identities.

These example cross products most clearly demonstrate the orthogonality of a cross product to its two argument vectors (Theorem 1.4.10a) and that the direction is in the right-hand sense (Theorem 1.4.10b). ■

Activity 1.4.7. Use the 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) $(3\,, 6\,, -6)$.

■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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 1.4.10d). (Also the direction of the cross product, $\pm\boldsymbol{k}$, is orthogonal to the x_1x_2-plane containing the two vectors—Theorem 1.4.10a). ■

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

(a) $+\boldsymbol{i}$ (b) $+\boldsymbol{j}$ (c) $+\boldsymbol{k}$ (d) $-\boldsymbol{i}$ (e) $-\boldsymbol{j}$ (f) $-\boldsymbol{k}$

■

Theorem 1.4.10 (cross product geometry)**.** *Let $\boldsymbol{v}$ and $\boldsymbol{w}$ be any 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 in the margin;

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

(d) the length $|\boldsymbol{v} \times \boldsymbol{w}|$ is the area of the parallelogram with edges $\boldsymbol{v}$ and $\boldsymbol{w}$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Proof. Let $\boldsymbol{v} = (v_1\,, v_2\,, v_3)$ and $\boldsymbol{w} = (w_1\,, w_2\,, w_3)$.

1.4.10a Recall that two vectors are orthogonal if their dot product is zero (Definition 1.3.19). To determine orthogonality between $\boldsymbol{v}$ and the cross product $\boldsymbol{v} \times \boldsymbol{w}$, consider

$$\begin{aligned}
\boldsymbol{v} \cdot (\boldsymbol{v} \times \boldsymbol{w}) &= (v_1\boldsymbol{i} + v_2\boldsymbol{j} + v_3\boldsymbol{k}) \cdot \big[\boldsymbol{i}(v_2w_3 - v_3w_2) \\
&\quad + \boldsymbol{j}(v_3w_1 - v_1w_3) + \boldsymbol{k}(v_1w_2 - v_2w_1)\big] \\
&= v_1(v_2w_3 - v_3w_2) + v_2(v_3w_1 - v_1w_3) \\
&\quad + v_3(v_1w_2 - v_2w_1) \\
&= v_1v_2w_3 - v_1v_3w_2 + v_2v_3w_1 \\
&\quad - v_1v_2w_3 + v_1v_3w_2 - v_2v_3w_1 \qquad = 0
\end{aligned}$$

as each term in the penultimate line cancels with the term underneath in the last line. Since the dot product is zero, the cross product $\boldsymbol{v} \times \boldsymbol{w}$ is orthogonal to vector $\boldsymbol{v}$.

Similarly, $\boldsymbol{v} \times \boldsymbol{w}$ is orthogonal to $\boldsymbol{w}$ (Exercise 1.4.5).

1.4.10b This right-handed property follows from the convention that the standard unit vectors $\boldsymbol{i}$, $\boldsymbol{j}$ and $\boldsymbol{k}$ are right-handed: that if $\boldsymbol{i}$ is in the direction of your thumb, and $\boldsymbol{j}$ is in the direction of your straight index finger, then $\boldsymbol{k}$ is in the direction of your bent second/longest finger—all on your right-hand.

We prove only for the case of vectors in the x_1x_2-plane, in which case $\boldsymbol{v} = (v_1\,, v_2\,, 0)$ and $\boldsymbol{w} = (w_1\,, w_2\,, 0)$, and when both $v_1\,, w_1 > 0$. One example is in stereo below.

Example 1.4.8 derived the cross product $\boldsymbol{v} \times \boldsymbol{w} = \boldsymbol{k}(v_1w_2 - v_2w_1)$. Consequently, this cross product is in the $+\boldsymbol{k}$ direction only when $v_1w_2 - v_2w_1 > 0$ (it is in the $-\boldsymbol{k}$ direction in the complementary case when $v_1w_2 - v_2w_1 < 0$). This inequality for $+\boldsymbol{k}$ rearranges to $v_1w_2 > v_2w_1$. Dividing by the positive v_1w_1 requires $\frac{w_2}{w_1} > \frac{v_2}{v_1}$. That is, in the x_1x_2-plane the 'slope' of vector $\boldsymbol{w}$ must greater than the 'slope' of vector $\boldsymbol{v}$. In this case, if $\boldsymbol{v}$ is in the direction of your thumb on your right-hand, and $\boldsymbol{w}$ is in the direction of your straight index finger, then your bent second/longest finger is in the direction $+\boldsymbol{k}$ as required by the cross-product $\boldsymbol{v} \times \boldsymbol{w}$.

1.4.10c Exercise 1.4.6 establishes the identity $|\boldsymbol{v} \times \boldsymbol{w}|^2 = |\boldsymbol{v}|^2|\boldsymbol{w}|^2 - (\boldsymbol{v} \cdot \boldsymbol{w})^2$. From Theorem 1.3.5 substitute $\boldsymbol{v} \cdot \boldsymbol{w} = |\boldsymbol{v}||\boldsymbol{w}|\cos\theta$

into this identity:

$$\begin{aligned}|\boldsymbol{v}\times\boldsymbol{w}|^2 &= |\boldsymbol{v}|^2|\boldsymbol{w}|^2-(\boldsymbol{v}\cdot\boldsymbol{w})^2\\ &= |\boldsymbol{v}|^2|\boldsymbol{w}|^2-(|\boldsymbol{v}||\boldsymbol{w}|\cos\theta)^2\\ &= |\boldsymbol{v}|^2|\boldsymbol{w}|^2-|\boldsymbol{v}|^2|\boldsymbol{w}|^2\cos^2\theta\\ &= |\boldsymbol{v}|^2|\boldsymbol{w}|^2(1-\cos^2\theta)\\ &= |\boldsymbol{v}|^2|\boldsymbol{w}|^2\sin^2\theta\,.\end{aligned}$$

Take the square-root of both sides to determine $|\boldsymbol{v}\times\boldsymbol{w}| = \pm|\boldsymbol{v}||\boldsymbol{w}|\sin\theta$. But $\sin\theta\ge 0$ since the angle $0\le\theta\le\pi$, and all the lengths are also ≥ 0, so only the plus case applies. That is, the length $|\boldsymbol{v}\times\boldsymbol{w}| = |\boldsymbol{v}||\boldsymbol{w}|\sin\theta$ as required.

1.4.10d Consider the plane containing the vectors $\boldsymbol{v}$ and $\boldsymbol{w}$, and hence containing the parallelogram formed by these vectors—as illustrated in the margin. Using vector $\boldsymbol{v}$ as the base of the parallelogram, with length $|\boldsymbol{v}|$, by basic trigonometry the height of the parallelogram is then $|\boldsymbol{w}|\sin\theta$. Hence the area of the parallelogram is the product base$\times$height $= |\boldsymbol{v}||\boldsymbol{w}|\sin\theta = |\boldsymbol{v}\times\boldsymbol{w}|$ by the previous part 1.4.10c.

□

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 in stereo below.

Solution: The area is the length of the cross product

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

Then the parallelogram area $|\boldsymbol{v}\times\boldsymbol{w}| = \sqrt{(-2)^2+4^2+(-4)^2} = \sqrt{4+16+16} = \sqrt{36} = 6$.

■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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

(a) 1 (b) $\sqrt{5}$ (c) 3 (d) 5

■

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 $\boldsymbol{x} = -2\boldsymbol{i} - \boldsymbol{j} + 3\boldsymbol{k} + (-2\boldsymbol{i} + 3\boldsymbol{j} + 2\boldsymbol{k})s + (2\boldsymbol{i} + 2\boldsymbol{j} + 3\boldsymbol{k})t$.

Solution: Use Definition 1.4.5 of the cross-product to find a normal vector:

$$\begin{aligned}\boldsymbol{v} \times \boldsymbol{w} &= \boldsymbol{i}(3 \cdot 3 - 2 \cdot 2) + \boldsymbol{j}(2 \cdot 2 - (-2) \cdot 3) + \boldsymbol{k}((-2) \cdot 2 - 3 \cdot 2) \\ &= 5\boldsymbol{i} + 10\boldsymbol{j} - 10\boldsymbol{k}\,.\end{aligned}$$

A normal vector is any vector proportional to this, so we could divide by five and choose normal vector $\boldsymbol{n} = \boldsymbol{i} + 2\boldsymbol{j} - 2\boldsymbol{k}$ (as illustrated above).

An equation of the plane through $-2\boldsymbol{i} - \boldsymbol{j} + 3\boldsymbol{k}$ is then given by the dot product

$$\begin{aligned}&(\boldsymbol{i} + 2\boldsymbol{j} - 2\boldsymbol{k}) \cdot [(x+2)\boldsymbol{i} + (y+1)\boldsymbol{j} + (z-3)\boldsymbol{k}] = 0\,,\\ \text{that is,}\quad &x + 2 + 2y + 2 - 2z + 6 = 0\,,\\ \text{that is,}\quad &x + 2y - 2z + 10 = 0\end{aligned}$$

is the required normal equation of the plane.

■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Algebraic properties of a cross product

Exercises 1.4.9–1.4.11 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 any vectors in $\mathbb{R}^3$, and c be any 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).*

Proof. Let's prove property 1.4.14a two ways—algebraically and geometrically. Exercises 1.4.9–1.4.11 ask you to prove the other properties.

- Algebraically: with vector $\boldsymbol{v} = (v_1, v_2, v_3)$, Definition 1.4.5 gives

$$\begin{aligned}\boldsymbol{v} \times \boldsymbol{v} &= \boldsymbol{i}(v_2v_3 - v_3v_2) + \boldsymbol{j}(v_3v_1 - v_1v_3) + \boldsymbol{k}(v_1v_2 - v_2v_1)\\ &= 0\boldsymbol{i} + 0\boldsymbol{j} + 0\boldsymbol{k} = \boldsymbol{0}\,.\end{aligned}$$

- Geometrically: from Theorem 1.4.10d, $|\boldsymbol{v}\times\boldsymbol{v}|$ is the area of the parallelogram with edges $\boldsymbol{v}$ and $\boldsymbol{v}$. But such a parallelogram has zero area, so $|\boldsymbol{v} \times \boldsymbol{v}| = 0$. Since the only vector of length zero is the zero vector (Theorem 1.1.13), $\boldsymbol{v} \times \boldsymbol{v} = \boldsymbol{0}$.

□

Example 1.4.15. As an example of Theorem 1.4.14b, Example 1.4.6 showed 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 $\boldsymbol{w} \times \boldsymbol{v}$?

Solution: By Theorem 1.4.14b, $\boldsymbol{w} \times \boldsymbol{v} = -(\boldsymbol{v} \times \boldsymbol{w}) = -5\boldsymbol{i} - 10\boldsymbol{j} + 10\boldsymbol{k}$.

■

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 $(3\boldsymbol{i} + 3\boldsymbol{j} + 3\boldsymbol{k}) \times (-2\boldsymbol{i} - \boldsymbol{j})$?

Solution: The first vector is $3(\boldsymbol{i} + \boldsymbol{j} + \boldsymbol{k})$ so by Theorem 1.4.14c,

$$\begin{aligned}&(3\boldsymbol{i} + 3\boldsymbol{j} + 3\boldsymbol{k}) \times (-2\boldsymbol{i} - \boldsymbol{j})\\ &= [3(\boldsymbol{i} + \boldsymbol{j} + \boldsymbol{k})] \times (-2\boldsymbol{i} - \boldsymbol{j})\\ &= 3[(\boldsymbol{i} + \boldsymbol{j} + \boldsymbol{k}) \times (-2\boldsymbol{i} - \boldsymbol{j})]\\ &= 3[\boldsymbol{i} - 2\boldsymbol{j} + \boldsymbol{k}] = 3\boldsymbol{i} - 6\boldsymbol{j} + 3\boldsymbol{k}\,.\end{aligned}$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

■

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

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

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

(a) $5\boldsymbol{i} - 2\boldsymbol{j} + 8\boldsymbol{k}$ (b) $\boldsymbol{i} - 17\boldsymbol{j} + 10\boldsymbol{k}$

(c) $-6\boldsymbol{i} + 15\boldsymbol{j} - 2\boldsymbol{k}$ (d) $-11\boldsymbol{i} - 16\boldsymbol{j} + 6\boldsymbol{k}$

(e) $-12\boldsymbol{i} + \boldsymbol{j} - 4\boldsymbol{k}$ (f) $-17\boldsymbol{i} - \boldsymbol{j} + 4\boldsymbol{k}$

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

Solution: In full detail:

$$\begin{aligned}&(\boldsymbol{i} - \boldsymbol{j}) \times (4\boldsymbol{i} + 2\boldsymbol{k})\\ &= (\boldsymbol{i} - \boldsymbol{j}) \times (4\boldsymbol{i}) + (\boldsymbol{i} - \boldsymbol{j}) \times (2\boldsymbol{k}) \quad \text{(by Thm 1.4.14d)}\\ &= 4(\boldsymbol{i} - \boldsymbol{j}) \times \boldsymbol{i} + 2(\boldsymbol{i} - \boldsymbol{j}) \times \boldsymbol{k} \quad \text{(by Thm 1.4.14c)}\\ &= -4\boldsymbol{i} \times (\boldsymbol{i} - \boldsymbol{j}) - 2\boldsymbol{k} \times (\boldsymbol{i} - \boldsymbol{j}) \quad \text{(by Thm 1.4.14b)}\\ &= -4[\boldsymbol{i} \times \boldsymbol{i} + \boldsymbol{i} \times (-\boldsymbol{j})] - 2[\boldsymbol{k} \times \boldsymbol{i} + \boldsymbol{k} \times (-\boldsymbol{j})] \quad \text{(by Thm 1.4.14d)}\\ &= -4[\boldsymbol{i} \times \boldsymbol{i} - \boldsymbol{i} \times \boldsymbol{j}] - 2[\boldsymbol{k} \times \boldsymbol{i} - \boldsymbol{k} \times \boldsymbol{j}] \quad \text{(by Thm 1.4.14c)}\\ &= -4[\boldsymbol{0} - \boldsymbol{k}] - 2[\boldsymbol{j} - (-\boldsymbol{i})] \quad \text{(by Example 1.4.6)}\\ &= -2\boldsymbol{i} - 2\boldsymbol{j} + 4\boldsymbol{k}\,.\end{aligned}$$

■

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 below. Our challenge is to derive that the volume of the parallelepiped is $|\boldsymbol{u} \cdot (\boldsymbol{v} \times \boldsymbol{w})|$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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 1.4.10d).
- 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 1.4.10a), 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 below.

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 any 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 area 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 below.

Solution: Example 1.4.11 found the cross product $\boldsymbol{v}\times\boldsymbol{w} = -2\boldsymbol{i}+4\boldsymbol{j}-4\boldsymbol{k}$. So the scalar triple product $\boldsymbol{u}\cdot(\boldsymbol{v}\times\boldsymbol{w}) = (2\boldsymbol{j}+\boldsymbol{k})\cdot(-2\boldsymbol{i}+4\boldsymbol{j}-4\boldsymbol{k}) = 8-4 = 4$. Hence the volume of the parallelepiped is 4 (cubic units).

The order of the vectors in a scalar triple product only affects the sign of the result. For example, we also find the volume of this parallelepiped via $\boldsymbol{v}\cdot(\boldsymbol{u}\times\boldsymbol{w})$. Returning to the procedure of Example 1.4.3 to find the cross product gives

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

Then the triple product $\boldsymbol{v}\cdot(\boldsymbol{u}\times\boldsymbol{w}) = (-2\boldsymbol{i}+\boldsymbol{k})\cdot(2\boldsymbol{j}-4\boldsymbol{k}) = 0+0-4 = -4$. Hence the volume of the parallelepiped is $|-4| = 4$ as before.

■

Using the procedure of Example 1.4.3 to find a scalar triple product establishes a strong connection to the determinants of Chapter 6. In the second solution to the previous Example 1.4.20, in finding $\boldsymbol{u}\times\boldsymbol{w}$, 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 $\boldsymbol{v}$:

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

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

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 generalises to determining, via analogous matrices of vectors, the 'volume' of objects formed by vectors with any number of components.

1.4.1 Exercises

Exercise 1.4.1. Use Definition 1.4.5 to establish some of the standard unit vector identities in Example 1.4.6:

(a) $\boldsymbol{j}\times\boldsymbol{k} = \boldsymbol{i}\,,\ \boldsymbol{k}\times\boldsymbol{j} = -\boldsymbol{i}\,,\ \boldsymbol{j}\times\boldsymbol{j} = \boldsymbol{0}\,;$

(b) $\boldsymbol{k}\times\boldsymbol{i} = \boldsymbol{j}\,,\ \boldsymbol{i}\times\boldsymbol{k} = -\boldsymbol{j}\,,\ \boldsymbol{k}\times\boldsymbol{k} = \boldsymbol{0}\,.$

Exercise 1.4.2. Use Definition 1.4.5, perhaps via the procedure used in Example 1.4.3, to determine the following cross products. Confirm each cross product is orthogonal to the two vectors in the given product. Show your details.

(a) $(3\boldsymbol{i}+\boldsymbol{j})\times(3\boldsymbol{i}-3\boldsymbol{j}-2\boldsymbol{k})$

(b) $(3\boldsymbol{i}+\boldsymbol{k})\times(5\boldsymbol{i}+6\boldsymbol{k})$

(c) $(2\boldsymbol{i}-\boldsymbol{j}-3\boldsymbol{k})\times(3\boldsymbol{i}+2\boldsymbol{k})$

(d) $(\boldsymbol{i}-\boldsymbol{j}+2\boldsymbol{k})\times(3\boldsymbol{i}+3\boldsymbol{k})$

(e) $(-1\,,3\,,2)\times(3\,,-5\,,1)$

(f) $(3\,,0\,,4)\times(5\,,1\,,2)$

(g) $(4\,,1\,,3)\times(3\,,2\,,-1)$

(h) $(3\,,-7\,,3)\times(2\,,1\,,0)$

Exercise 1.4.3. For each of the stereo pictures below, estimate the area of the pictured parallelogram by estimating the edge vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ (all components are integers), then computing their cross product.

(a)

v0.2a

(b)

(c)

(d)

(e)

(f)

Exercise 1.4.4. Each of the following equations describes a plane in 3D. Find a normal vector to each of the planes.

(a) $\boldsymbol{x} = (-1, 0, 1) + (-5, 2, -1)s + (2, -4, 0)t$

(b) $2x + 2y + 4z = 20$

(c) $x_1 - x_2 + x_3 + 2 = 0$

(d) $\boldsymbol{x} = 6\boldsymbol{i} - 3\boldsymbol{j} + (3\boldsymbol{i} - 3\boldsymbol{j} - 2\boldsymbol{k})s - (\boldsymbol{i} + \boldsymbol{j} + \boldsymbol{k})$

(e) $\boldsymbol{x} = \boldsymbol{j} + 2\boldsymbol{k} + (\boldsymbol{i} - \boldsymbol{k})s + (-5\boldsymbol{i} + \boldsymbol{j} - 3\boldsymbol{k})t$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(f) $3y = x + 2z + 4$

(g) $3p + 8q - 9 = 4r$

(h) $\boldsymbol{x} = (-2, 2, -3) + (-3, 2, 0)s + (-1, 3, 2)t$

Exercise 1.4.5. Use Definition 1.4.5 to prove that for all vectors $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^3$, the cross product $\boldsymbol{v} \times \boldsymbol{w}$ is orthogonal to $\boldsymbol{w}$.

Exercise 1.4.6. For all vectors $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^3$ prove the identity $|\boldsymbol{v} \times \boldsymbol{w}|^2 = |\boldsymbol{v}|^2|\boldsymbol{w}|^2 - (\boldsymbol{v} \cdot \boldsymbol{w})^2$ (an identity invoked in the proof of Theorem 1.4.10c). Use the algebraic Definitions 1.3.2 and 1.4.5 of the dot and cross products to expand both sides of the identity and show both sides expand to the same complicated expression.

Exercise 1.4.7. Using Theorem 1.4.14, and the identities among standard unit vectors of Example 1.4.6, compute the following cross products. Record and justify each step in detail.

(a) $\boldsymbol{i} \times (3\boldsymbol{j})$

(b) $(4\boldsymbol{j} + 3\boldsymbol{k}) \times \boldsymbol{k}$

(c) $(4\boldsymbol{k}) \times (\boldsymbol{i} + 6\boldsymbol{j})$

(d) $\boldsymbol{j} \times (3\boldsymbol{i} + 2\boldsymbol{k})$

(e) $(2\boldsymbol{i} + 2\boldsymbol{k}) \times (\boldsymbol{i} + \boldsymbol{j})$

(f) $(\boldsymbol{i} - 5\boldsymbol{j}) \times (-\boldsymbol{j} + 3\boldsymbol{k})$

Exercise 1.4.8. You are given that three specific vectors $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{w}$ in $\mathbb{R}^3$ have the following cross products:

$$\begin{aligned} \boldsymbol{u} \times \boldsymbol{v} &= -\boldsymbol{j} + \boldsymbol{k}\,, \\ \boldsymbol{u} \times \boldsymbol{w} &= \boldsymbol{i} - \boldsymbol{k}\,, \\ \boldsymbol{v} \times \boldsymbol{w} &= -\boldsymbol{i} + 2\boldsymbol{j}\,. \end{aligned}$$

Use Theorem 1.4.14 to compute the following cross products. Record and justify each step in detail.

(a) $(\boldsymbol{u} + \boldsymbol{v}) \times \boldsymbol{w}$

(b) $(3\boldsymbol{u} + \boldsymbol{w}) \times (2\boldsymbol{u})$

(c) $(3\boldsymbol{v}) \times (\boldsymbol{u} + \boldsymbol{v})$

(d) $(2\boldsymbol{v} + \boldsymbol{w}) \times (\boldsymbol{u} + 3\boldsymbol{v})$

(e) $(2\boldsymbol{v} + 3\boldsymbol{w}) \times (\boldsymbol{u} + 2\boldsymbol{w})$

(f) $(\boldsymbol{u} + 4\boldsymbol{v} + 2\boldsymbol{w}) \times \boldsymbol{w}$

Exercise 1.4.9. Use Definition 1.4.5 to algebraically prove Theorem 1.4.14b—the property that $\boldsymbol{w} \times \boldsymbol{v} = -(\boldsymbol{v} \times \boldsymbol{w})$. Explain how this property also follows from the basic geometry of the cross product (Theorem 1.4.10).

Exercise 1.4.10. Use Definition 1.4.5 to algebraically prove Theorem 1.4.14c—the property that $(c\boldsymbol{v}) \times \boldsymbol{w} = c(\boldsymbol{v} \times \boldsymbol{w}) = \boldsymbol{v} \times (c\boldsymbol{w})$. Explain how this property also follows from the basic geometry of the cross product (Theorem 1.4.10)—consider $c > 0$, $c = 0$ and $c < 0$ seperately.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Exercise 1.4.11. Use Definition 1.4.5 to algebraically prove Theorem 1.4.14d—the distributive property that $\boldsymbol{u} \times (\boldsymbol{v} + \boldsymbol{w}) = \boldsymbol{u} \times \boldsymbol{v} + \boldsymbol{u} \times \boldsymbol{w}$.

Exercise 1.4.12. For each of the following illustrated parallelepipeds: estimate the edge vectors $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{w}$ (all components are integers); then use the scalar triple product to estimate the volume of the parallelepiped.

(a)

(b)

(c)

(d)

(e)

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(f)

v0.2a

1.5 Use Matlab/Octave for vector computation

Section Contents

1.5.1 Exercises 87

> It is the science of *calculation*,—which becomes continually more necessary at each step of our progress, and which must ultimately govern the whole of the applications of science to the arts of life.
>
> *Charles Babbage, 1832*

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

- MATLAB is commercial software available from Mathworks.[5] It is also useable over the internet as MATLAB-Online or MATLAB-Mobile.
- Octave is free software, that for our limited purposes is almost the same as MATLAB, and downloadable over the internet.[6] Octave is also freely useable over the internet.[7]
- Alternatively, your home institution may provide MATLAB/Octave via a web service that is useable from smart phones, tablets and computers.

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

(a) $(2, -1)$

Solution: Start MATLAB/Octave. After a prompt, "`>>`" in MATLAB or "`octave:1>`" in Octave, type a command, followed by the Return/Enter key to get it executed. As indicated by Table 1.2 the numbers with brackets separated by semi-colons forms a vector, and the `=` character assigns the result to variable for subsequent use.

[5] http://mathworks.com
[6] https://www.gnu.org/software/octave/
[7] http://octave-online.net

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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

- Real numbers are limited to being zero or of magnitude from 10^{-323} to 10^{+308}, both positive and negative (called the **floating point** numbers). Real numbers are computed and stored to a maximum precision of nearly sixteen significant digits.[a]
- MATLAB/Octave potentially uses complex numbers ($\mathbb{C}$), but mostly we stay within real numbers ($\mathbb{R}$).
- Each MATLAB/Octave command is usually typed on one line by itself.
- `[ . ; . ; . ]` where each dot denotes a number, forms vectors in $\mathbb{R}^3$ (or use newlines instead of the semi-colons). Use n numbers separated by semi-colons for vectors in $\mathbb{R}^n$.
- `=` assigns the result of the expression on the right 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` (denoted by "`ans =`" in MATLAB/Octave).
- Variable names are alphanumeric starting with a letter.
- `size(v)` returns the number of components of the vector (Definition 1.1.4): if the vector $\boldsymbol{v}$ is in $\mathbb{R}^m$, then `size(v)` returns $\begin{bmatrix} m & 1 \end{bmatrix}$.
- `norm(v)` computes the length/magnitude of the vector `v` (Definition 1.1.9).
- `+,-,*` is vector/scalar addition, subtraction, and multiplication, but only provided the sizes of the two vectors are the same. Parentheses `()` control the order of operations.
- `/x` divides a vector/scalar by a scalar x. However, be warned that `/v` for a vector $\boldsymbol{v}$ typically gives strange results as MATLAB/Octave interprets it to mean you want to approximately solve some linear equation.
- `x^y` for scalars x and y computes x^y.
- `dot(u,v)` computes the dot product of vectors `u` and `v` (Definition 1.3.2)—if they have the same size.
- `acos(q)` computes the arc-cos, the inverse cosine, of the scalar `q` in radians. To find the angle in degrees use `acos(q)*180/pi`.
- `quit` terminates the MATLAB/Octave session.

[a] If desired, 'computer algebra' software provides us with an arbitrary level of precision, even exact. Current computer algebra software includes the free Sage, Maxima and Reduce, and the commercial Maple, Mathematica and (via Matlab) MuPad.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Assign the vector to a variable `a` by the command `a=[2;-1]`. Then executing `norm(a)` reports `ans = 2.2361` as shown in the dialogue to the right.

```
>> a=[2;-1]
a =
   2
  -1
>> norm(a)
ans =  2.2361
```

This computes the answer $|(2,-1)| = \sqrt{2^2+(-1)^2} = \sqrt{5} = 2.2361$ (to five significant digits which we take to be practically exact).

The qr-code appearing in the margin here encodes these MATLAB/Octave commands. You may scan such qr-codes with your favourite app[8], and then copy and paste the code direct into a MATLAB/Octave client. Alternatively, if reading an electronic version of this book, then you may copy and paste the commands; however, be warned that the quote character ' usually needs correcting. Although here the saving in typing is negligible, later you can save considerable typing

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

Solution: In MATLAB/Octave:

Assign the vector to a variable with `b=[-1;1;-5;4]` as shown to the right. Then execute `norm(b)` and find that MATLAB/Octave reports `ans = 6.5574`

```
>> b=[-1;1;-5;4]
b =
  -1
   1
  -5
   4
>> norm(b)
ans =  6.5574
```

Hence 6.5574 is the length of $(-1, 1, -5, 4)$ (to five significant digits which we take to be practically exact).

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

Solution: In MATLAB/Octave:

i. assign the vector with the command `c=[-0.3;4.3;-2.5;-2.8;7;-1.9]`

ii. execute `norm(c)` and find that MATLAB/Octave reports `ans = 9.2347`

Hence the length of vector $(-0.3, 4.3, -2.5, -2.8, 7, -1.9)$ is 9.2347 (to five significant digits).

[8] At the time of writing, qr-code scanning applications for smart-phones include *WaspScan* and *QRReader*—but I have no expertise to assess their quality.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

■

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).

Solution: In MATLAB/Octave type the commands, each followed by Return/Enter key.

Assign the named vectors with the commands `u=[-4.1;1.7;4.1]` and `v=[2.9;0.9;-2.4]` to see the two steps in the dialogue to the right.

```
>> u=[-4.1;1.7;4.1]
u =
  -4.1000
   1.7000
   4.1000
>> v=[2.9;0.9;-2.4]
v =
   2.9000
   0.9000
  -2.4000
```

Execute `u+v` to find from the dialogue on the right that the sum $\boldsymbol{u}+\boldsymbol{v}=(-1.2\,,2.6\,,1.7)$.

```
>> u+v
ans =
  -1.2000
   2.6000
   1.7000
```

Execute `u-v` to find from the dialogue on the right that the difference $\boldsymbol{u}-\boldsymbol{v}=(-7\,,0.8\,,6.5)$.

```
>> u-v
ans =
  -7.0000
   0.8000
   6.5000
```

Execute `3*u` to find from the dialogue on the right that the scalar multiple $3\boldsymbol{u}=(-12.3\,,5.1\,,12.3)$ (the asterisk is essential to compute multiplication).

```
>> 3*u
ans =
  -12.3000
    5.1000
   12.3000
```

■

v0.2a

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 $\boldsymbol{q} = (2, 4, 3, 5)$.

Solution: In MATLAB/Octave

Assign the first vector with `p=[1;0;2;-6]` as shown to the right.

```
>> p=[1;0;2;-6]
p =
   1
   0
   2
  -6
```

Assign the other vector with `q=[2;4;3;5]`.

```
>> q=[2;4;3;5]
q =
   2
   4
   3
   5
```

Compute $2(2\boldsymbol{p} - 3\boldsymbol{q}) + 6(\boldsymbol{q} - \boldsymbol{p})$ with the command `2*(2*p-3*q)+6*(q-p)` as shown to the right, and see the result is evidently $-2\boldsymbol{p}$.

```
>> 2*(2*p-3*q)+6*(q-p)
ans =
   -2
    0
   -4
   12
```

Confirm it is $-2\boldsymbol{p}$ by adding $2\boldsymbol{p}$ to the above result with the command `ans+2*p` as shown to the right, and see the zero vector result.

```
>> ans+2*p
ans =
   0
   0
   0
   0
```

■

v0.2a

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

Solution: In MATLAB/Octave

Assign $\boldsymbol{u}=(8\,,-6\,,-4\,,-2)$ with command `u=[8;-6;-4;-2]` as shown to the right.

```
>> u=[8;-6;-4;-2]
u =
   8
  -6
  -4
  -2
```

Assign $\boldsymbol{v}=(4\,,3\,,-1)$ with the command `v=[4;3;-1]`.

```
>> v=[4;3;-1]
v =
   4
   3
  -1
```

Compute $\boldsymbol{u}+\boldsymbol{v}$ with the command `u+v` as shown to the right. MATLAB prints an error message because the vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are of different sizes and so cannot be added together.

```
>> u+v
Error using  +
Matrix dimensions must agree.
```

Check the sizes of the vectors in the sum using `size(u)` and `size(v)` to confirm $\boldsymbol{u}$ is in $\mathbb{R}^4$ whereas $\boldsymbol{v}$ is in $\mathbb{R}^3$. Hence the two vectors cannot be added (Definition 1.2.4).

```
>> size(u)
ans =
   4   1
>> size(v)
ans =
   3   1
```

Alternatively, Octave gives the following error message in which "nonconformant arguments" means of wrong sizes.

```
error: operator +: nonconformant arguments
(op1 is 4x1, op2 is 3x1)
```

■

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`?

v0.2a

(a)
```
   2.8000
   4.3000
  -7.1000
```

(b)
```
  -0.6000
   3.1000
  -1.9000
```

(c)
```
Error using *
Inner matrix
dimensions must
agree.
```

(d)
```
   2.2000
   7.4000
  -9.0000
```

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

■

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

Solution: In MATLAB/Octave

Because each vector is used twice in the formula $\cos\theta = (\boldsymbol{u}\cdot\boldsymbol{v})/(|\boldsymbol{u}||\boldsymbol{v}|)$, give each a name as shown to the right.

```
>> u=[4;3]
u =
   4
   3
>> v=[5;12]
v =
   5
  12
```

Then invoke `dot()` to compute the dot product in the formula for the cosine of the angle.

```
>> cost=dot(u,v)/norm(u)/norm(v)
cost =  0.8615
```

Lastly, invoke `acos()` for the arc-cosine, then convert the radians to degrees to find the angle $\theta = 30.510°$.

```
>> theta=acos(cost)*180/pi
theta =  30.510
```

■

Example 1.5.7. 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 1.3.13d) for vectors $\boldsymbol{u} = (-0.1, -3.1, -2.9, -1.3)$, $\boldsymbol{v} = (-3, 0.5, 6.4, -0.9)$ and $\boldsymbol{w} = (-1.5, -0.2, 0.4, -3.1)$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Solution: In MATLAB/Octave

Assign vector $\boldsymbol{u} = (-0.1\,,-3.1\,,-2.9\,,-1.3)$ with the command `u=[-0.1;-3.1;-2.9;-1.3]` as shown to the right.

```
>> u=[-0.1;-3.1;-2.9;-1.3]
u =
  -0.1000
  -3.1000
  -2.9000
  -1.3000
```

Assign vector $\boldsymbol{v} = (-3\,,0.5\,,6.4\,,-0.9)$ with the command `v=[-3;0.5;6.4;-0.9]` as shown to the right.

```
>> v=[-3;0.5;6.4;-0.9]
v =
  -3.0000
   0.5000
   6.4000
  -0.9000
```

Assign vector $\boldsymbol{w} = (-1.5\,,-0.2\,,0.4\,,-3.1)$ with the command `w=[-1.5;-0.2;0.4;-3.1]` as shown to the right.

```
>> w=[-1.5;-0.2;0.4;-3.1]
w =
  -1.5000
  -0.2000
   0.4000
  -3.1000
```

Compute the dot product $(\boldsymbol{u}+\boldsymbol{v})\cdot\boldsymbol{w}$ with the command `dot(u+v,w)` to find the answer is 13.390.

```
>> dot(u+v,w)
ans =  13.390
```

Compare this with the dot product expression $\boldsymbol{u}\cdot\boldsymbol{w}+\boldsymbol{v}\cdot\boldsymbol{w}$ via the command `dot(u,w)+dot(v,w)` to find the answer is 13.390.

```
>> dot(u,w)+dot(v,w)
ans =  13.390
```

That the two answers agree verifies the distributive law for dot products.

■

Activity 1.5.8. 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) `(u-2*v)*(u+v)-u*u+u*v+2*v*v`

(b) `dot(u-2*v,u+v)-dot(u,u)+dot(u,v)+2*dot(v,v)`

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(c) `dot(u-2v,u+v)-dot(u,u)+dot(u,v)+2dot(v,v)`

(d) None of the others.

■

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

1.5.1 Exercises

Exercise 1.5.1. Use MATLAB/Octave to compute the length of each of the following vectors (the first five have integer lengths).

(a) $\begin{bmatrix}2\\3\\6\end{bmatrix}$ (b) $\begin{bmatrix}4\\-4\\7\end{bmatrix}$ (c) $\begin{bmatrix}-2\\6\\9\end{bmatrix}$ (d) $\begin{bmatrix}0.5\\0.1\\-0.5\\0.7\end{bmatrix}$ (e) $\begin{bmatrix}8\\-4\\5\\-8\end{bmatrix}$

(f) $\begin{bmatrix}1.1\\1.7\\-4.2\\-3.8\\0.9\end{bmatrix}$ (g) $\begin{bmatrix}2.6\\-0.1\\3.2\\-0.6\\-0.2\end{bmatrix}$ (h) $\begin{bmatrix}1.6\\-1.1\\-1.4\\2.3\\-1.6\end{bmatrix}$

Exercise 1.5.2. Use MATLAB/Octave to determine which are wrong out of the following identities and relations for vectors $\boldsymbol{p} = (0.8, -0.3, 1.1, 2.6, 0.1)$ and $\boldsymbol{q} = (1, 2.8, 1.2, 2.3, 2.3)$.

(a) $3(\boldsymbol{p}-\boldsymbol{q}) = 3\boldsymbol{p} - 3\boldsymbol{q}$ (b) $2(\boldsymbol{p}-3\boldsymbol{q}) + 3(2\boldsymbol{q}-\boldsymbol{p}) = \boldsymbol{p}$

(c) $\frac{1}{2}(\boldsymbol{p}-\boldsymbol{q}) + \frac{1}{2}(\boldsymbol{p}+\boldsymbol{q}) = \boldsymbol{p}$ (d) $|\boldsymbol{p}+\boldsymbol{q}| \le |\boldsymbol{p}| + |\boldsymbol{q}|$

(e) $|\boldsymbol{p}-\boldsymbol{q}| \le |\boldsymbol{p}| + |\boldsymbol{q}|$ (f) $|\boldsymbol{p}\cdot\boldsymbol{q}| \le |\boldsymbol{p}||\boldsymbol{q}|$

Exercise 1.5.3. Use MATLAB/Octave to find the angles between pairs of vectors in each of the following groups.

(a) $\boldsymbol{p} = (2\,,3\,,6)$, $\boldsymbol{q} = (6\,,2\,,-3)$, $\boldsymbol{r} = (3\,,-6\,,2)$

v0.2a

(b) $\boldsymbol{u} = \begin{bmatrix} -1 \\ -7 \\ -1 \\ -7 \end{bmatrix}$, $\boldsymbol{v} = \begin{bmatrix} -1 \\ 4 \\ 4 \\ 4 \end{bmatrix}$, $\boldsymbol{w} = \begin{bmatrix} 1 \\ -4 \\ -4 \\ -4 \end{bmatrix}$

(c) $\boldsymbol{u} = \begin{bmatrix} 5 \\ 1 \\ -1 \\ 3 \end{bmatrix}$, $\boldsymbol{v} = \begin{bmatrix} -6 \\ 4 \\ 2 \\ -5 \end{bmatrix}$, $\boldsymbol{w} = \begin{bmatrix} -4 \\ 2 \\ 1 \\ -2 \end{bmatrix}$

(d) $\boldsymbol{u} = \begin{bmatrix} -9 \\ -8 \\ 4 \\ 8 \end{bmatrix}$, $\boldsymbol{v} = \begin{bmatrix} 9 \\ 3 \\ 1 \\ -3 \end{bmatrix}$, $\boldsymbol{w} = \begin{bmatrix} -6 \\ 5 \\ -2 \\ -4 \end{bmatrix}$

(e) $\boldsymbol{a} = \begin{bmatrix} -4.1 \\ 9.8 \\ 0.3 \\ 1.4 \\ 2.7 \end{bmatrix}$, $\boldsymbol{b} = \begin{bmatrix} -0.6 \\ 2.6 \\ -1.2 \\ -0.2 \\ -0.9 \end{bmatrix}$, $\boldsymbol{c} = \begin{bmatrix} -2.8 \\ -0.9 \\ -6.2 \\ -2.3 \\ -4.7 \end{bmatrix}$, $\boldsymbol{d} = \begin{bmatrix} 1.8 \\ -3.4 \\ -8.6 \\ 1.4 \\ 1.8 \end{bmatrix}$

(f) $\boldsymbol{a} = \begin{bmatrix} -0.5 \\ 2.0 \\ -3.4 \\ 1.8 \\ 0.1 \end{bmatrix}$, $\boldsymbol{b} = \begin{bmatrix} 5.4 \\ 7.4 \\ 0.5 \\ 0.7 \\ 1.3 \end{bmatrix}$, $\boldsymbol{c} = \begin{bmatrix} -0.2 \\ -1.5 \\ -0.3 \\ 1.1 \\ 2.5 \end{bmatrix}$, $\boldsymbol{d} = \begin{bmatrix} 1.0 \\ 2.0 \\ -1.3 \\ -4.4 \\ -2.0 \end{bmatrix}$

v0.2a

Answers to selected exercises

1.2.2b : 1.3

1.2.2d : 6.5

1.2.2f : 3.5

1.2.3b : $\boldsymbol{u}$ and $\boldsymbol{w}$ are closest; both the other pairs are equal furthest.

1.2.3d : $\boldsymbol{v}$ and $\boldsymbol{w}$ are furthest; both the other pairs are equal closest.

1.2.3f : all pairs are the same distance apart.

1.2.3h : $\boldsymbol{u}$ and $\boldsymbol{w}$ are closest; $\boldsymbol{v}$ and $\boldsymbol{w}$ are furthest.

1.2.4b : One possibility is $\boldsymbol{x} = (-4+7t)\boldsymbol{i} + (1-6t)\boldsymbol{j} + (-2+7t)\boldsymbol{k}$

1.2.4d : One possibility is $\boldsymbol{x} = (0.2+3.1t)\boldsymbol{e}_1 + (-7.2+6.1t)\boldsymbol{e}_2 + (-4.6+4.2t)\boldsymbol{e}_3 + (-2.8+2.5t)\boldsymbol{e}_4$

1.2.4f : One possibility is $\boldsymbol{x} = (1.8-3.2t)\boldsymbol{e}_1 + (-3.1+3.9t)\boldsymbol{e}_2 - (1+1.6t)\boldsymbol{e}_3 + (-1.3+4.1t)\boldsymbol{e}_4 + (-3.3+2.5t)\boldsymbol{e}_5$

1.3.2b : Not orthogonal.

1.3.2d : Orthogonal.

1.3.2f : Not orthogonal.

1.3.2h : Not orthogonal.

1.3.4 : Angles are $\theta_{ab} = 73.22^\circ$, $\theta_{ac} = 45^\circ$, $\theta_{ad} = 90^\circ$, $\theta_{bc} = 52.24^\circ$, $\theta_{bd} = 45^\circ$, $\theta_{cd} = 54.74^\circ$. So the pair $\boldsymbol{a}$ and $\boldsymbol{c}$, and the pair $\boldsymbol{b}$ and $\boldsymbol{d}$ are both closest pairs. The most dissimilar titles are $\boldsymbol{a}$ and $\boldsymbol{d}$.

1.3.10b : $(-4\,,1\,,-5) + (4\,,-2\,,4)s + (4\,,-4\,,3)t$

1.3.10d : There is no "the plane" as the three points are collinear.

1.3.10f : $(-2.6\,,-1.6\,,-0.5) + (4.8\,,-5.1\,,2.5)s + (5.5\,,7\,,-0.1)t$

1.3.10h : $(-1.6\,,2\,,-3.7) + (3.4\,,-2.2\,,3)s + (3\,,-2.5\,,4.2)t$

1.4.2b : $-13\boldsymbol{j}$

1.4.2d : $-3\boldsymbol{i} + 3\boldsymbol{j} + 3\boldsymbol{k}$

1.4.2f : $(-4\,,14\,,3)$

1.4.2h : $(-3\,,6\,,17)$

1.4.3b : $\sqrt{13} = 3.606$

1.4.3d : $\sqrt{235} = 15.33$

1.4.3f : $\sqrt{94} = 9.695$

1.4.4b : $\propto (1\,,1\,,2)$

1.4.4d : $\propto \boldsymbol{i} + 5\boldsymbol{j} - 6\boldsymbol{k}$

v0.2a

1.4.4f : $\propto -\boldsymbol{i} + 3\boldsymbol{j} - 2\boldsymbol{k}$

1.4.4h : $\propto (4\,,6\,,-7)$

1.4.7b : $4\boldsymbol{i}$

1.4.7d : $2\boldsymbol{i} - 3\boldsymbol{k}$

1.4.7f : $-15\boldsymbol{i} - 3\boldsymbol{j} - \boldsymbol{k}$

1.4.8b : $-2\boldsymbol{i} + 2\boldsymbol{k}$

1.4.8d : $2\boldsymbol{i} - 4\boldsymbol{j} - \boldsymbol{k}$

1.4.8f : $-3\boldsymbol{i} + 8\boldsymbol{j} - \boldsymbol{k}$

1.4.12b : 1

1.4.12d : 12

1.4.12f : 2

1.5.1b : 9

1.5.1d : 1

1.5.1f : 6.0819

1.5.1h : 3.6851

1.5.3b : $\theta_{uv} = 147.44^\circ$, $\theta_{uw} = 32.56^\circ$, $\theta_{vw} = 180^\circ$

1.5.3d : $\theta_{uv} = 146.44^\circ$, $\theta_{uw} = 101.10^\circ$, $\theta_{vw} = 108.80^\circ$

1.5.3f : $\theta_{ab} = 73.11^\circ$, $\theta_{ac} = 88.54^\circ$, $\theta_{ad} = 90.48^\circ$, $\theta_{bc} = 106.56^\circ$, $\theta_{bd} = 74.20^\circ$, $\theta_{cd} = 137.36^\circ$

v0.2a

2 Systems of linear equations

Chapter Contents

2.1 Introduction to systems of linear equations 94

2.1.1 Exercises 100

2.2 Directly solve linear systems 103

2.2.1 Compute a system's solution 103

2.2.2 Algebraic manipulation solves systems 114

2.2.3 Three possible numbers of solutions 123

2.2.4 Exercises 126

2.3 Linear combinations span sets 132

2.3.1 Exercises 138

Linear relationships are commonly identified in science and engineering, and are commonly expressed as linear equations. One of the reasons is that scientists and engineers can do amazingly powerful algebraic transformations with linear equations. Such transformations and their practical implications are the subject of this book.

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

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

Example 2.0.1 (scientific inference)**.** Two colleagues, and American and a European, discuss the weather; in particular, they discuss the temperature. 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 marginal figure 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 $T_A = cT_E + d$ for some constants c and d they and we aim to find. The two days of data then give that

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

To find c and d:

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

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

You may quite rightly object that the two colleagues *assumed* a linear relation, they do *not know* it is linear. You may also object that the predicted relation is erroneous as it should be $T_A = \frac{9}{5}T_E + 32$ (the relation between Celsius and Fahrenheit). Absolutely, you should object. Scientifically, the deduced relation $T_A = 1.82\,T_E + 32.73$ 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. This is fundamental scientific methodology—and central to it is the algebra of linear equations. ■

Linear algebra and equations are also crucial for nonlinear relationships. Figure 2.1 shows four plots of the same nonlinear function, but on successively smaller scales. Zooming in on the point $(0, 1)$ we see the curve looks straighter and straighter until on the microscale it is effectively a straight line. The same is true for anywhere on any smooth nonlinear curve: we discover that the curve looks like a straight line on the microscale. Thus we may view any smooth nonlinear function as roughly being made up of lots of microscale straight line segments. Linear equations and their algebra on this microscale empower our understanding of nonlinear relations—for example, microscale linearity underwrites all of calculus.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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

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

2.1 Introduction to systems of linear equations

Section Contents

2.1.1 Exercises 100

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

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

- What is a formula for the Celsius temperature as a function of the temperature in Fahrenheit? Answer by rearranging the equation: subtract 32 from both sides, $T_F - 32 = \frac{9}{5}T_C$; multiply both sides by $\frac{5}{9}$, then $\frac{5}{9}(T_F - 32) = T_C$; that is, $T_C = \frac{5}{9}T_F - \frac{160}{9}$.
- What temperature has the same *numerical value* in the two scales? That is, when is $T_F = T_C$? Answer by algebra: we want $T_C = T_F = \frac{9}{5}T_C + 32$; subtract $\frac{9}{5}T_C$ from both sides to give $-\frac{4}{5}T_C = 32$; multiply both sides by $-\frac{5}{4}$, then $T_C = -\frac{5}{4} \times 32 = -40$.

■

Linear equations are characterised by each unknown never being multiplied or divided by another unknown, or itself, nor inside 'curvaceous' functions. Table 2.1 lists examples of both. The power of linear algebra is especially important for large numbers of unknown variables: generally we say there are n variables.

v0.2a

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

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

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

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

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

Solution: To draw the graphs seen in the marginal plot, rearrange the linear equations as $y = 3 - x$ and $y = x/2$. They intersect at the point $(2, 1)$ so $x = 2$ and $y = 1$ is the unique solution.

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

Solution: To draw the graphs seen in the marginal plot, rearrange the linear equations as $y = \frac{2}{3}x - \frac{2}{3}$ and $y = \frac{2}{3}x + \frac{1}{2}$. Evidently these lines never intersection, they are parallel, so there appears to be no solution.

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

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

Solution: To draw the graphs seen in the marginal plot, rearrange the linear equations as $y = 2 - x/2$ and $y = 2 - x/2$. They are the same line so every point on this line is a solution of the system. There are an infinite number of possible solutions.

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

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

■

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

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

■

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

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

Solution: From these three sources of information, Pythagoras and the length of displacement vectors gives the three equations

$$\begin{aligned}(x-29)^2 + (y-10)^2 &= 25^2\,,\\ (x-17)^2 + (y-19)^2 &= 20^2\,,\\ (x-13)^2 + (y-18)^2 &= 17^2\,.\end{aligned}$$

These three equations constrain your as yet unknown location (x, y). Expanding the squares in these equations gives the equivalent system

$$\begin{aligned}x^2 - 58x + 841 + y^2 - 20y + 100 &= 625\,,\\ x^2 - 34x + 289 + y^2 - 38y + 361 &= 400\,,\\ x^2 - 26x + 169 + y^2 - 36y + 324 &= 289\,.\end{aligned}$$

Involving squares of the unknowns, these are a nonlinear system of equations and so appear to lie outside the remit of this book. However, straightforward algebra transforms these three nonlinear equations into a system of two linear equations which we solve.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Let's subtract the third equation from each of the other two, then the nonlinear squared terms cancel giving a system of two linear equations in two variables:

$$\begin{aligned} -32x + 672 + 16y - 224 = 336 &\iff -2x + y = -7\,, \\ -8x + 120 - 2y + 37 = 111 &\iff -4x - y = -23\,. \end{aligned}$$

Graphically, include these two lines to the picture, namely $y = -7 + 2x$ and $y = 23 - 4x$, and then their intersection gives your location.

Algebraically, one could add the two equations together: $(-2x+y)+(-4x-y) = -7-23$ which reduces to $-6x = -30$, that is, $x = 5$. Then either equation, say the first, determines $y = -7 + 2x = -7 + 2 \cdot 5 = 3$. That is, your location is $(x, y) = (5, 3)$ (in Mm), as drawn.

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 $\tan^{-1}\frac{3}{5} = 30.96°$N. ■

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

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

Solution: The marginal plot shows the three planes represented by the given equations (in the order blue, brown, red), and plots the (black) point we seek of intersection of all three planes.

Algebraically we combine and manipulate the equations in a sequence of steps designed to simplify the form of the system. *By doing the same manipulation to the whole of each of the equations, we ensure the validity of the result.*

(a) Subtract the first equation from each of the other two equations to deduce (as illustrated)

$$\begin{aligned} x_1 + x_2 - x_3 &= -2\,, \\ 2x_2 + 6x_3 &= 10\,, \\ x_2 + 2x_3 &= 3\,. \end{aligned}$$

(b) Divide the second equation by two:

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

v0.2a

(c) Subtract the second equation from each of the other two (as illustrated):

$$\begin{aligned} x_1 \qquad - 4x_3 &= -7\,, \\ x_2 + 3x_3 &= 5\,, \\ -x_3 &= -2\,. \end{aligned}$$

(d) Multiply the third equation by (-1):

$$\begin{aligned} x_1 \qquad - 4x_3 &= -7\,, \\ x_2 + 3x_3 &= 5\,, \\ x_3 &= 2\,. \end{aligned}$$

(e) Add four times the third equation to the first, and subtract three times it from the second (as illustrated):

$$\begin{aligned} x_1 \qquad\qquad &= 1\,, \\ x_2 \qquad &= -1\,, \\ x_3 &= 2\,. \end{aligned}$$

Thus the only solution to this system of three linear equations in three variables is $(x_1, x_2, x_3) = (1, -1, 2)$.

■

The sequence of marginal graphs in the previous Example 2.1.6 illustrate the equations at each main step in the algebraic manipulations. Apart from keeping the solution point fixed, the sequence of graphs looks rather chaotic. Indeed there is no particular geometric pattern or interpretation of the steps in this algebra. One feature of Section 3.3 is that we discover how the so-called 'singular value decomposition' solves linear equations via a great method with a strong geometric interpretation. This geometric interpretation then empowers further methods useful in applications.

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

v0.2a

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

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

Suppose we observe that at some given temperature and 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’).

Solution: To infer a relation to hold generally—to fill in the gaps between the known measurements, seek ‘rainfall’ to be predicted by the linear formula

$$(\text{‘rain’}) = x_1 + x_2(\text{‘temp’}) + x_3(\text{‘humid’}),$$

for some coefficients x_1, x_2 and x_3 to be determined. The measured data of Table 2.2 constrains and determines these coefficients: substitute each triple of measurements to require

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

The previous Example 2.1.6 solves this set of three linear equations in three unknowns to determine the solution that the coefficients $(x_1, x_2, x_3) = (1, -1, 2)$. That is, the requisite formula to infer rain from any given temperature and humidity is

$$(\text{‘rain’}) = 1 - (\text{‘temp’}) + 2(\text{‘humid’}).$$

This example illustrates that the geometry of fitting a plane to three points (as plotted) translates into the abstract geometry of finding the intersection of three planes (plotted in the previous example). The linear algebra procedure for this latter abstract problem then gives the required ‘physical’ solution.

■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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

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 establishes the general key property of a system of any number of linear equations in any number of variables: the system has either

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

2.1.1 Exercises

Exercise 2.1.1. Graphically and algebraically solve each of the following systems.

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

(b) $\begin{aligned} x + 2y &= 5 \\ 6x - 2y &= 2 \end{aligned}$

(c) $\begin{aligned} x - y &= 2 \\ -2x + 7y &= -4 \end{aligned}$

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

(e) $\begin{aligned} 3x - 2y &= 1 \\ 6x - 4y &= -2 \end{aligned}$

(f) $\begin{aligned} 4x - 3y &= -1 \\ -5x + 4y &= 1 \end{aligned}$

(g) $\begin{aligned} p + q &= 3 \\ -p - q &= 2 \end{aligned}$

(h) $\begin{aligned} p - q &= 1 \\ -3p + 5q &= -4 \end{aligned}$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(i) $\begin{aligned} 3u - v &= 0 \\ u - v &= -1 \end{aligned}$

(j) $\begin{aligned} 4u + 4v &= -2 \\ -u - v &= 1 \end{aligned}$

(k) $\begin{aligned} -3s + 4t &= 0 \\ -3s + 3t &= -\tfrac{3}{2} \end{aligned}$

(l) $\begin{aligned} -4s + t &= -2 \\ 4s - t &= 2 \end{aligned}$

Exercise 2.1.2. For each of the following graphs: estimate the equations of the pair of lines; solve the pair of equations algebraically; and confirm the algebraic solution is reasonably close to the intersection of the pair of lines.

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

Exercise 2.1.3. Graphically and algebraically solve each of the following systems of three equations for the two unknowns.

(a) $\begin{aligned} 4x + y &= 8 \\ 3x - 3y &= -\tfrac{3}{2} \\ -4x + 2y &= -2 \end{aligned}$

(b) $\begin{aligned} -4x + 3y &= \tfrac{7}{2} \\ 7x + y &= -3 \\ x - 2y &= \tfrac{3}{2} \end{aligned}$

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

(d) $\begin{aligned} -2x - 4y &= 3 \\ x + 2y &= 3 \\ -4x - 8y &= -6 \end{aligned}$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(e) $$\begin{aligned} 3x + 2y &= 4 \\ -2x - 4y &= -4 \\ 4x + 2y &= 5 \end{aligned}$$

(f) $$\begin{aligned} -2x + 3y &= -3 \\ -5x + 2y &= -9 \\ 3x + 3y &= 6 \end{aligned}$$

Exercise 2.1.4 (Global Positioning System in 2D)**.** For each case below, and in two dimensions, suppose you know from three GPS satellites that you and your GPS receiver are given distances away from the given locations of each of the three satellites (locations and distance are in Mm). Following Example 2.1.5, determine your position.

(a) 25 from (7 , 30)
26 from (10 , 30)
29 from (20 , 27)

(b) 25 from (11 , 29)
26 from (28 , 15)
20 from (16 , 21)

(c) 20 from (22 , 12)
26 from (16 , 24)
29 from (26 , 21)

(d) 17 from (12 , 21)
25 from (10 , 29)
26 from (27 , 15)

In which of these cases: are you at the 'North Pole'? flying high above the Earth? the measurement data is surely in error?

v0.2a

2.2 Directly solve linear systems

Section Contents

2.2.1 Compute a system's solution 103
2.2.2 Algebraic manipulation solves systems 114
2.2.3 Three possible numbers of solutions 123
2.2.4 Exercises 126

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

2.2.1 Compute a system's solution

> It is unworthy of excellent persons to lose hours like slaves in the labour of calculation.
> *Gottfried Wilhelm von Leibniz*

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

Example 2.2.1. The first system of Example 2.1.3a

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

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 $A = \begin{bmatrix} 1 & 1 \\ 2 & -4 \end{bmatrix}$, right-hand side vector $\boldsymbol{b} = (3, 0)$, and vector of variables $\boldsymbol{x} = (x, y)$. ■

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

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) will appear to be just some mystic rearrangement of symbols—such an interpretation is sufficient for this chapter. However, those of you who have met matrix multiplication will recognise that $A\boldsymbol{x} = \boldsymbol{b}$ is an expression involving natural operations

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

for matrices and vectors: Section 3.1 defines and explores such useful operations.

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

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

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

Solution: (a) The first system, that of Example 2.1.6, is of three equations in three variables ($m = n = 3$) and is written in the form $A\boldsymbol{x} = \boldsymbol{b}$ as

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

for square matrix A.

(b) The second system has three variables called r, s and t and two equations. Variables 'missing' from an equation are represented as zero times that variable, thus the system

$$\begin{aligned} -2r + 3s + 0t &= 6\,,\\ 0r + s - 4t &= -\pi\,, \end{aligned} \quad \text{is} \quad \underbrace{\begin{bmatrix} -2 & 3 & 0 \\ 0 & 1 & -4 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} r \\ s \\ t \end{bmatrix}}_{\boldsymbol{x}} = \underbrace{\begin{bmatrix} 6 \\ -\pi \end{bmatrix}}_{\boldsymbol{b}}$$

for 2×3 matrix A.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

■

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

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

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

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

(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 Table 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()` avoids gross mistakes such as "Lies My Computer Told Me" (Poole 2015, p.83). Section 3.3.2 discovers what `rcond()` is, and why `rcond()` avoids mistakes.[1] In practice, decisions about acceptability are rarely black and white, and so the qualitative ranges of `rcond()` reflects practical reality.

> In theory, there is no difference between theory and practice. But, in practice, there is.
>
> *Jan L. A. van de Snepscheut*

Example 2.2.6. Use MATLAB/Octave to solve the system (from Example 2.1.6)

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

[1] Interestingly, there are incredibly rare pathological matrices for which `rcond()` and `A\` fails us (Driscoll & Maki 2007). For example, among 32×32 matrices the probability is about 10^{-22} of encountering a matrix for which `rcond()` misleads us by more than a factor of a hundred in using `A\`.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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

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

Solution: Begin by writing the system in the abstract matrix-vector form $A\boldsymbol{x} = \boldsymbol{b}$ as already done by Example 2.2.3. Then the three steps of Procedure 2.2.5 are the following.

Beware that the symbol "=" in MATLAB/Octave is a procedural assignment of a value—quite different in nature to the "=" in algebra which denotes equality.

(a) Form matrix A and column vector $\boldsymbol{b}$ with the MATLAB/Octave assignments

```
A=[1 1 -1; 1 3 5; 1 2 1]
b=[-2;8;1]
```

Table 2.3 summarises that in MATLAB/Octave: each line is one command; the `=` symbol assigns the value of the right-hand expression to the variable name of the left-hand side; and the brackets `[ ]` construct matrices and vectors.

(b) Check the value of `rcond(A)`: here it is 0.018 which is in the good range.

(c) Since `rcond(A)` is acceptable, execute `x=A\b` to compute the solution vector $\boldsymbol{x} = (1, -1, 2)$ (and assign it to the variable `x`, see Table 2.3).

All together that is the four commands

```
A=[1 1 -1; 1 3 5; 1 2 1]
b=[-2;8;1]
rcond(A)
```

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

```
x=A\b
```

Such qr-codes in the margin encodes these commands for you to possibly scan, copy and paste into MATLAB/Octave.

■

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

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

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

Solution: Begin by writing, or at least by imaging, each system in matrix-vector form:

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

As the matrices and vectors are modifications of the previous Example 2.2.6 we reduce typing by modifying the matrix and vector of the previous example (using the ability to change a each element in a matrix, see Table 2.3).

(a) For the first system execute `A(3,2)=-3` and `b(2)=5` to see the matrix and vector are now

```
A =
     1     1    -1
     1     3     5
     1    -3     1
b =
    -2
     5
     1
```

Check the value of `rcond(A)`, here 0.14 is good. Then obtain the solution from `x=A\b` with the result

```
x =
   -0.6429
```

```
  -0.1429
   1.2143
```

That is, the solution $\boldsymbol{x} = (-0.64, -0.14, 1.21)$ to two decimal places (2 d.p.).[2]

(b) For the second system now execute `A(2,3)=-2` to see the new matrix is the required

```
A =
     1     1    -1
     1     3    -2
     1    -3     1
```

Check: find that `rcond(A)` is zero which is classified as terrible. Consequently we cannot compute a solution of this second system of linear equations (as in Figure 2.2(c)).

If we were to try `x=A\b` in this second system, then MATLAB/Octave would report

```
Warning: Matrix is singular to working precision.
```

However, we cannot rely on MATLAB/Octave producing such useful messages: we must use `rcond()` to almost always avoid mistakes.

■

Example 2.2.9. Use MATLAB/Octave to solve the system

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

Solution: Following Procedure 2.2.5, form the corresponding matrix and vector as

```
A=[1  -2   3   1   2
  -2  -6  -3  -2   2
   2   3   0   0  -2
  -2   1   0   0   0
  -2  -2   1   1  -2 ]
b=[7;-1;-9;-3;5]
```

Check: find `rcond(A)` is acceptably 0.020, so compute the solution via `x=A\b` to find the result

[2] The four or five significant digits printed by MATLAB/Octave is effectively exact for most practical purposes. This book often reports two significant digits as two is enough for most human readable purposes. When a numerical result is reported to two decimal places, the book indicates this truncation with "(2 d.p.)".

v0.2a

```
x =
    0.8163
   -1.3673
   -6.7551
   17.1837
    3.2653
```

that is, the solution $\boldsymbol{x} = (0.82, -1.37, -6.76, 17.18, 3.27)$ (2 d.p.).

■

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

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

Solution: The corresponding system of linear equations is

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

Invoking Procedure 2.2.5:

(a) form matrix A and column vector $\boldsymbol{b}$ with

```
A=[-7 3; 7 -5; 1 -2]
b=[3;-2;1]
```

(b) check `rcond(A)`: MATLAB/Octave gives the message

```
Error using rcond
Input must be a square matrix.
```

As `rcond(A)` does not exist, the procedure cannot give a solution.

The reason for the procedure not leading to a solution is that a system of three equations in two variables, as here, generally does not have a solution. [3]

■

[3] If one were to execute `x=A\b`, then you would find MATLAB/Octave gives the 'answer' $\boldsymbol{x} = (-0.77\,, -0.73)$ (2 d.p.). But this answer is not a solution. Instead this answer has another meaning, often sensibly useful, which is explained by Section 3.5. Using `rcond()` helps us to avoid confusing such an answer with a solution.

v0.2a

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 $\frac{3}{(x-1)(x+2)}$ needs to be written as $\frac{1}{x-1}-\frac{1}{x+2}$. Solving linear equations helps:

- here pose that $\frac{3}{(x-1)(x+2)}=\frac{A}{x-1}+\frac{B}{x+2}$ for some unknown A and B;
- then write the right-hand side over the common denominator,

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

  and this equals $\frac{3}{(x-1)(x+2)}$ only if $A+B=0$ 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 $\frac{-4x^3+8x^2-5x+2}{x^2(x-1)^2}$.

Solution: Recalling that repeated factors require extra terms in the decomposition, seek a decomposition of the form

$$\begin{aligned}&\frac{A}{x^2}+\frac{B}{x}+\frac{C}{(x-1)^2}+\frac{D}{x-1}\\ =&\frac{A(x-1)^2+Bx(x-1)^2+Cx^2+Dx^2(x-1)}{x^2(x-1)^2}\\ =&\frac{(B+D)x^3+(A-2B+C-D)x^2+(-2A+B)x+(A)}{x^2(x-1)^2}\\ =&\frac{-4x^3+8x^2-5x+2}{x^2(x-1)^2}.\end{aligned}$$

For this last equality to hold for all x the coefficients of various powers of x must be equal: this leads to the linear equation system

$$\begin{aligned}&B+D=-4\\ &A-2B+C-D=8\\ &-2A+B=-5\\ &A=2\end{aligned}\iff\begin{bmatrix}0&1&0&1\\1&-2&1&-1\\-2&1&0&0\\1&0&0&0\end{bmatrix}\begin{bmatrix}A\\B\\C\\D\end{bmatrix}=\begin{bmatrix}-4\\8\\-5\\2\end{bmatrix}.$$

Either solve by hand or by computer.

- By hand, the last equation gives $A=2$ so the third equation then gives $B=-1$. Then the first gives $D=-3$. Lastly, the second then gives $C=8-2+2(-1)+(-3)=1$. That is, the decomposition is

$$\frac{-4x^3+8x^2-5x+2}{x^2(x-1)^2}=\frac{2}{x^2}-\frac{1}{x}+\frac{1}{(x-1)^2}-\frac{3}{x-1}$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

- Using MATLAB/Octave, form the matrix and right-hand side with

```
a=[0 1 0 1
   1 -2 1 -1
   -2 1 0 0
   1 0 0 0]
b=[-4;8;-5;2]
```

Then solve by checking `rcond(a)`, which at 0.04 is good, and so `ABCD=a\b` finds the answer $(2, -1, 1, -3)$. As before, these coefficients give the decomposition as

$$\frac{-4x^3 + 8x^2 - 5x + 2}{x^2(x-1)^2} = \frac{2}{x^2} - \frac{1}{x} + \frac{1}{(x-1)^2} - \frac{3}{x-1}$$

■

Example 2.2.12 (`rcond` avoids disaster)**.** In Example 2.0.1 an American and European compared temperatures and using two days temperatures discovered the approximation that the American temperature $T_A = 1.82\,T_E + 32.73$ where T_E denotes the European temperature. Continuing the story, three days later they again meet and compare the temperatures they experienced: the American reporting that "for the last three days it has been 51°, 74° and 81°", whereas the European reports "why, I recorded it as 11°, 23° and 27°". The marginal figure plots this data with the original two data points, apparently confirming a reasonable linear relationship between the two temperature scales.

Let's fit a polynomial to this temperature data.

Solution: There are five data points. Each data point gives us an equation to be satisfied. This suggests we use linear algebra to determine five coefficients in a formula. Let's fit the data with the quartic polynomial

$$T_A = c_1 + c_2 T_E + c_3 T_E^2 + c_4 T_E^3 + c_5 T_E^4, \tag{2.1}$$

and use the data to determine the coefficients $c_1, c_2, \dots, c_5$. Substituting each of the five pairs of T_E and T_A into this equation gives the five linear equations

$$\begin{aligned} 60 &= c_1 + 15c_2 + 225c_3 + 3375c_4 + 50625c_5\,, \\ &\vdots \\ 81 &= c_1 + 27c_2 + 729c_3 + 19683c_4 + 531441c_5\,. \end{aligned}$$

In MATLAB/Octave, form these into matrix-vector equation $A\boldsymbol{c} = \boldsymbol{t}_A$ for the unknown coefficients $\boldsymbol{c} = (c_1, c_2, c_3, c_4, c_5)$, the vectors of American temperatures $\boldsymbol{t}_A$, and the 5×5 matrix A constructed below (recall from Table 2.3 that `te.^p` computes the `p`th power of each element in the column vector `te`).

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

```
te=[15;26;11;23;27]
ta=[60;80;51;74;81]
plot(te,ta,'o')
A=[ones(5,1) te te.^2 te.^3 te.^4]
```

Then solve for the coefficients using `c=A\ta` to get

```
A =
       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
c =
 -163.5469
   46.5194
   -3.6920
    0.1310
   -0.0017
```

Job done—or is it? To check, let's plot the predictions of the quartic polynomial (2.1) with these coefficients. In MATLAB/Octave we may plot a graph with the following

```
t=linspace(5,35);
plot(t,c(1)+c(2)*t+c(3)*t.^2+c(4)*t.^3+c(5)*t.^4)
```

and see a graph like the marginal one. Disaster: the quartic polynomial relationship is clearly terrible as it is too wavy and nothing like the straight line we know it should be ($T_A = \frac{9}{5}T_E + 32$).

The problem is we forgot `rcond`. In MATLAB/Octave execute `rcond(A)` and discover `rcond` is $3 \cdot 10^{-9}$. This value is in the 'terrible' range classified by Procedure 2.2.5. Thus the solution of the linear equations must not be used: here the marginal plot indeed shows the solution coefficients are not acceptable. Always use `rcond` to check for bad systems of linear equations.

■

The previous Example 2.2.12 illustrates one of the 'rules of thumb' in science and engineering: *for data fitting, avoid using polynomials of degree higher than cubic.*

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

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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

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

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

Given your phone also detects that four other satellites broadcast the following position and time information: $(13, 20, 12)$ time shift 0.04 s before; $(17, 14, 10)$ time shift $0.033\cdots$ s before; $(8, 21, 10)$ time shift $0.033\cdots$ s before; and $(22, 9, 8)$ time shift 0.04 s before. Adapting the approach of Example 2.1.5, use linear algebra to determine your phone's location in space.

Solution: Let your unknown position be (x, y, z) and the unknown time shift to the phone's clock t be found from $w = 300t$. Then the five equations from the five satellites are, respectively,

$$\begin{aligned}
&(x-6)^2 + (y-12)^2 + (z-23)^2 = (300t+12)^2 = (w+12)^2,\\
&(x-13)^2 + (y-20)^2 + (z-12)^2 = (300t+12)^2 = (w+12)^2,\\
&(x-17)^2 + (y-14)^2 + (z-10)^2 = (300t+10)^2 = (w+10)^2,\\
&(x-8)^2 + (y-21)^2 + (z-10)^2 = (300t+10)^2 = (w+10)^2,\\
&(x-22)^2 + (y-9)^2 + (z-8)^2 = (300t+12)^2 = (w+12)^2.
\end{aligned}$$

Expand all the squares in these equations:

$$\begin{aligned}
&x^2 - 12x + 36 + y^2 - 24y + 144 + z^2 - 46z + 529\\
&= w^2 + 24w + 144\,,\\
&x^2 - 26x + 169 + y^2 - 40y + 400 + z^2 - 24z + 144\\
&= w^2 + 24w + 144\,,\\
&x^2 - 34x + 289 + y^2 - 28y + 196 + z^2 - 20z + 100\\
&= w^2 + 20w + 100\,,\\
&x^2 - 16x + 64 + y^2 - 42y + 441 + z^2 - 20z + 100\\
&= w^2 + 20w + 100\,,\\
&x^2 - 44x + 484 + y^2 - 18y + 81 + z^2 - 16z + 64\\
&= w^2 + 24w + 144\,.
\end{aligned}$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

These are a system of nonlinear equations and so outside the remit of the course, but a little algebra brings them within. Subtract the last equation, say, from each of the first four equations: then *all* of the nonlinear squares of variables cancel leaving a linear system. Combining the constants on the right-hand side, and moving the w terms to the left gives the system of four linear equations

$$\begin{aligned}
32x - 6y - 30z + 0w &= -80\,,\\
18x - 22y - 8z + 0w &= -84\,,\\
10x - 10y - 4z + 4w &= 0\,,\\
28x - 24y - 4z + 4w &= -20\,.
\end{aligned}$$

Following Procedure 2.2.5, solve this system by forming the corresponding matrix and vector as

```
A=[32  -6 -30  0
   18 -22  -8  0
   10 -10  -4  4
   28 -24  -4  4 ]
b=[-80;-84;0;-20]
```

Check `rcond(A)`: it is acceptably 0.023 so compute the solution via `x=A\b` to find

```
x =
   2
   4
   4
   9
```

Hence your phone is at location $(x, y, z) = (2, 4, 4)\,\text{Mm}$. Further, the time discrepancy between your phone and the GPS satellites' time is proportional to $w = 9\,\text{Mm}$. Since $w = 300t$, where $300\,\text{Mm/s}$ is the speed of light, the time discrepancy is $t = \frac{9}{300} = 0.03\,\text{s}$.

■

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 much more details about linear algebra.

v0.2a

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

In hand calculations we often want to minimise writing, so the discussion here uses two forms side-by-side for the linear equations: one form with all symbols recorded for best clarity; and beside it, one form where only coefficients are recorded for quickest writing. Translating from one to the other is crucial even in a computing era as the computer also primarily deals with arrays of numbers, and we must interpret what those arrays of numbers mean in terms of linear equations.

Example 2.2.14. Recall the system of linear equations of Example 2.1.6:

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

The first crucial level of abstraction is to write this in the matrix-vector form, Example 2.2.3,

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

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* $\left[A \,\vdots\, \boldsymbol{b}\right]$.

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

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Solution:

$$\begin{cases} -2r + 3s = 6 \\ s - 4t = -\pi \end{cases} \iff \left[\begin{array}{ccc:c} -2 & 3 & 0 & 6 \\ 0 & 1 & -4 & -\pi \end{array}\right]$$

$$\begin{cases} -7y + 3z = 3 \\ 7y - 5z = -2 \\ y - 2z = 1 \end{cases} \iff \left[\begin{array}{cc:c} -7 & 3 & 3 \\ 7 & -5 & -2 \\ 1 & -2 & 1 \end{array}\right]$$

An augmented matrix is not unique: it depends upon the order of the equations, and also the order you choose for the variables in $\boldsymbol{x}$. The first example implicitly chose $\boldsymbol{x} = (r, s, t)$; if instead we choose to order the variables as $\boldsymbol{x} = (s, t, r)$, then

$$\begin{cases} 3s - 2r = 6 \\ s - 4t = -\pi \end{cases} \iff \left[\begin{array}{ccc:c} 3 & 0 & -2 & 6 \\ 1 & -4 & 0 & -\pi \end{array}\right]$$

Such variations to the augmented matrix are valid, but you must remember the corresponding chosen order of the variables.

■

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

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

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

■

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

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

(a) interchange two equations/rows; or

(b) multiply an equation/row by a nonzero constant; or

(c) add a multiple of an equation/row to another.

Proof. We just address the system of equations form as the augmented matrix form is equivalent but more abstract.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

1. Swapping the order of two equations does not change the system of equations, the set of relations between the variables, so does not change the solution.

2. Let vector $\boldsymbol{x}$ satisfy $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$. Then $ca_1x_1 + ca_2x_2 + \cdots + ca_nx_n = c(a_1x_1 + a_2x_2 + \cdots + a_nx_n) = cb$ and so $\boldsymbol{x}$ satisfies c times the equation. When the constant c is non-zero, the above can be reversed through dividing by c. Hence multiplying an equation by a non-zero constant c does not change the possible solutions.

3. Let vector $\boldsymbol{x}$ satisfy both $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$ and $a'_1x_1 + a'_2x_2 + \cdots + a'_nx_n = b'$. Then

$$\begin{aligned}
&(a'_1 + ca_1)x_1 + (a'_2 + ca_2)x_2 + \cdots + (a'_n + ca_n)x_n \\
&= a'_1x_1 + ca_1x_1 + a'_2x_2 + ca_2x_2 + \cdots + a'_nx_n + ca_nx_n \\
&= a'_1x_1 + a'_2x_2 + \cdots + a'_nx_n + c(a_1x_1 + a_2x_2 + \cdots + a_nx_n) \\
&= b' + cb\,.
\end{aligned}$$

That is, $\boldsymbol{x}$ also satisfies the equation formed by adding c times the first to the second. Conversely, every vector $\boldsymbol{x}$ that satisfies both $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$ and $(a'_1 + ca_1)x_1 + (a'_2 + ca_2)x_2 + \cdots + (a'_n + ca_n)x_n = b' + cb$, by adding $(-c)$ times the first to the second as above, also satisfies $a'_1x_1 + a'_2x_2 + \cdots + a'_nx_n = b'$. Hence adding a multiple of an equation to another does not change the possible solutions.

□

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

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

Confirm with MATLAB/Octave.

Solution: In order to know what the row operations should find, let's first solve the system with MATLAB/Octave via Procedure 2.2.5. In matrix-vector form the system is

$$\begin{bmatrix} 1 & 2 & 1 \\ 2 & -3 & 0 \\ 0 & -3 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix};$$

hence in MATLAB/Octave execute

```
A=[1 2 1;2 -3 0;0 -3 -1]
b=[1;2;2]
rcond(A)
x=A\b
```

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

`rcond(A)` is just good, 0.0104, so the computed answer $\boldsymbol{x} = (x, y, z) = (7, 4, -14)$ is the solution.

Second, use elementary row operations. Let's write the working in both full symbolic equations and in augmented matrix form in order to see the correspondence between the two—you would not have to do both, either one would suffice.

$$\begin{cases} x + 2y + z = 1 \\ 2x - 3y + 0z = 2 \\ 0x - 3y - z = 2 \end{cases} \iff \left[\begin{array}{ccc:c} 1 & 2 & 1 & 1 \\ 2 & -3 & 0 & 2 \\ 0 & -3 & -1 & 2 \end{array}\right]$$

Add (-2) times the first equation/row to the second.

$$\begin{cases} x + 2y + z = 1 \\ 0x - 7y - 2z = 0 \\ 0x - 3y - z = 2 \end{cases} \iff \left[\begin{array}{ccc:c} 1 & 2 & 1 & 1 \\ 0 & -7 & -2 & 0 \\ 0 & -3 & -1 & 2 \end{array}\right]$$

This makes the first column have a leading one (Definition 2.2.20). Start on the second column by dividing the second equation/row by (-7).

$$\begin{cases} x + 2y + z = 1 \\ 0x + y + \frac{2}{7}z = 0 \\ 0x - 3y - z = 2 \end{cases} \iff \left[\begin{array}{ccc:c} 1 & 2 & 1 & 1 \\ 0 & 1 & \frac{2}{7} & 0 \\ 0 & -3 & -1 & 2 \end{array}\right]$$

Now subtract twice the second equation/row from the first, and add three times the second to the third.

$$\begin{cases} x + 0y + \frac{3}{7}z = 1 \\ 0x + y + \frac{2}{7}z = 0 \\ 0x + 0y - \frac{1}{7}z = 2 \end{cases} \iff \left[\begin{array}{ccc:c} 1 & 0 & \frac{3}{7} & 1 \\ 0 & 1 & \frac{2}{7} & 0 \\ 0 & 0 & -\frac{1}{7} & 2 \end{array}\right]$$

This makes the second column have the second leading one (Definition 2.2.20). Start on the third column by multiplying the third equation/row by (-7).

$$\begin{cases} x + 0y + \frac{3}{7}z = 1 \\ 0x + y + \frac{2}{7}z = 0 \\ 0x + 0y + z = -14 \end{cases} \iff \left[\begin{array}{ccc:c} 1 & 0 & \frac{3}{7} & 1 \\ 0 & 1 & \frac{2}{7} & 0 \\ 0 & 0 & 1 & -14 \end{array}\right]$$

Now subtract 3/7 of the third equation/row from the first, and 2/7 from the second.

$$\begin{cases} x + 0y + 0z = 7 \\ 0x + y + 0z = 4 \\ 0x + 0y + z = -14 \end{cases} \iff \left[\begin{array}{ccc:c} 1 & 0 & 0 & 7 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & -14 \end{array}\right]$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

This completes the transformation of the equations/augmented matrix into a so-called reduced row echelon form (Definition 2.2.20). From this form we read off the solution: the system of equation on the left directly gives $x = 7$, $y = 4$ and $z = -14$, that is, the solution vector $\boldsymbol{x} = (x, y, z) = (7, 4, -14)$ (as computed by MATLAB/Octave); the transformed augmented matrix on the right tells us exactly the same thing because (Definition 2.2.15) it means the same as the matrix-vector

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 7 \\ 4 \\ -14 \end{bmatrix},$$

which is the same as the system on the above-left and tells us the solution $\boldsymbol{x} = (x, y, z) = (7, 4, -14)$.

■

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

(a) any equations with all zero coefficients, or rows of the matrix consisting entirely of zeros, are at the bottom;

(b) in each nonzero equation/row, the first nonzero coefficient/entry is a one (called the ***leading one****), and is in a variable/column to the left of any leading ones below it; and*

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

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

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

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

Solution: This is in RREF with leading ones on the variables x_1 and x_3. Let the other variables be free by say setting $x_2 = s$ and $x_4 = t$ for arbitrary parameters s and t. Then the two equations give $x_1 = -2 - s + 2t$ and $x_3 = 5 - 4t$. Consequently, the most general solution is $\boldsymbol{x} = (x_1, x_2, x_3, x_4) = (-2 - s + 2t, s, 5 - 4t, t)$ for arbitrary s and t.

(b) $\left[\begin{array}{ccc|c} 1 & 0 & -1 & 1 \\ 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 4 \end{array}\right]$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Solution: This augmented matrix is in RREF with leading ones in the first and second columns. To find solutions, explicitly write down the corresponding system of linear equations. But we do not know the variables! If the context does not give variable names, then use the generic $x_1, x_2, \dots, x_n$. Thus here the corresponding system is

$$x_1 - x_3 = 1\,, \quad x_2 - x_3 = -2\,, \quad 0 = 4\,.$$

The first two equations are valid, but the last is contradictory as $0 \neq 4$. Hence there are no solutions to the system.

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

Solution: This augmented matrix is the same as the previous except for a zero in the bottom right entry. It is in RREF with leading ones in the first and second columns. Explicitly, the corresponding system of linear equations is

$$x_1 - x_3 = 1\,, \quad x_2 - x_3 = -2\,, \quad 0 = 0\,.$$

Now the last equation, $0 = 0$, is is always satisfied. Hence the first two equations determine solutions to the system: letting free variable $x_3 = s$ for arbitrary s the two equations give solutions $\boldsymbol{x} = (1 + s, -2 + s, s)$.

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

Solution: This system is not in RREF: although there are two leading ones multiplying x and y in the first and the second equation respectively, the variable y does not have zero coefficients in the first equation. (A solution to this system exists, shown graphically in the margin, but the question does not ask for it.)

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

Solution: This augmented matrix is not in RREF as there are not even any leading ones.

■

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

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

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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

■

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

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

(a) $(0.4\,,-0.6\,,0)$ (b) $(0.2t+0.4\,,1.2t+0.6\,,t)$

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

■

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

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

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

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

Example 2.2.25. Use Gauss–Jordan elimination, Procedure 2.2.24, to find all possible solutions to the system

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

depending upon the parameter c.

Solution: Here write both the full symbolic equations and the augmented matrix form—you would only have to do one.

$$\begin{cases} -x-y=-3 \\ x+4y=-1 \\ 2x+4y=c \end{cases} \iff \left[\begin{array}{cc:c} -1 & -1 & -3 \\ 1 & 4 & -1 \\ 2 & 4 & c \end{array}\right]$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Multiply the first by (-1).

$$\begin{cases} x+y=3 \\ x+4y=-1 \\ 2x+4y=c \end{cases} \iff \left[\begin{array}{cc:c} 1 & 1 & 3 \\ 1 & 4 & -1 \\ 2 & 4 & c \end{array}\right]$$

Subtract the first from the second, and twice the first from the third.

$$\begin{cases} x+y=3 \\ 0x+3y=-4 \\ 0x+2y=c-6 \end{cases} \iff \left[\begin{array}{cc:c} 1 & 1 & 3 \\ 0 & 3 & -4 \\ 0 & 2 & c-6 \end{array}\right]$$

Divide the second by three.

$$\begin{cases} x+y=3 \\ 0x+y=-\frac{4}{3} \\ 0x+2y=c-6 \end{cases} \iff \left[\begin{array}{cc:c} 1 & 1 & 3 \\ 0 & 1 & -\frac{4}{3} \\ 0 & 2 & c-6 \end{array}\right]$$

Subtract the second from the first, and twice the second from the third.

$$\begin{cases} x+0y=\frac{13}{3} \\ 0x+y=-\frac{4}{3} \\ 0x+0y=c-\frac{10}{3} \end{cases} \iff \left[\begin{array}{cc:c} 1 & 0 & \frac{13}{3} \\ 0 & 1 & -\frac{4}{3} \\ 0 & 0 & c-\frac{10}{3} \end{array}\right]$$

The system is now in reduced row echelon form. The last row immediately tells us that there is no solution for parameter $c \neq \frac{10}{3}$ as the equation would then be inconsistent. If parameter $c = \frac{10}{3}$, then the system is consistent and the first two rows give that the only solution is $(x, y) = (\frac{13}{3}, -\frac{4}{3})$.

■

Example 2.2.26. Use Gauss–Jordan elimination, Procedure 2.2.24, to find all possible solutions to the system

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

Solution: Here write both the full symbolic equations and the augmented matrix form—you would choose one or the other.

$$\begin{cases} 0u-2v+3w=-1 \\ 2u+v+w=-1 \end{cases} \iff \left[\begin{array}{ccc:c} 0 & -2 & 3 & -1 \\ 2 & 1 & 1 & -1 \end{array}\right]$$

Swap the two rows to get a non-zero top-left entry.

$$\begin{cases} 2u+v-w=-1 \\ 0u-2v+3w=-1 \end{cases} \iff \left[\begin{array}{ccc:c} 2 & 1 & -1 & -1 \\ 0 & -2 & 3 & -1 \end{array}\right]$$

Divide the first row by two.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$\begin{cases} u + \frac{1}{2}v - \frac{1}{2}w = -\frac{1}{2} \\ 0u - 2v + 3w = -1 \end{cases} \iff \left[\begin{array}{ccc:c} 1 & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ 0 & -2 & 3 & -1 \end{array}\right]$$

Divide the second row by (-2).

$$\begin{cases} u + \frac{1}{2}v - \frac{1}{2}w = -\frac{1}{2} \\ 0u + v - \frac{3}{2}w = \frac{1}{2} \end{cases} \iff \left[\begin{array}{ccc:c} 1 & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ 0 & 1 & -\frac{3}{2} & \frac{1}{2} \end{array}\right]$$

Subtract half the second row from the first.

$$\begin{cases} u + 0v + \frac{1}{4}w = -\frac{3}{4} \\ 0u + v - \frac{3}{2}w = \frac{1}{2} \end{cases} \iff \left[\begin{array}{ccc:c} 1 & 0 & \frac{1}{4} & -\frac{3}{4} \\ 0 & 1 & -\frac{3}{2} & \frac{1}{2} \end{array}\right]$$

The system is now in reduced row echelon form. The third column is that of a free variable so set the third component $w = t$ for arbitrary t. Then the first row gives $u = -\frac{3}{4} - \frac{1}{4}t$, and the second row gives $v = \frac{1}{2} + \frac{3}{2}t$. That is, the solutions are $(u, v, w) = (-\frac{3}{4} - \frac{1}{4}t, \frac{1}{2} + \frac{3}{2}t, t)$ for arbitrary t.

■

2.2.3 Three possible numbers of solutions

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

Theorem 2.2.27. *For any 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.*

Proof. First, if there is exactly none or one solution to $A\boldsymbol{x} = \boldsymbol{b}$, then the theorem holds. Second, suppose there are two distinct solutions; let them be $\boldsymbol{y}$ and $\boldsymbol{z}$ so $A\boldsymbol{y} = \boldsymbol{b}$ and $A\boldsymbol{z} = \boldsymbol{b}$. Then consider $\boldsymbol{x} = t\boldsymbol{y} + (1-t)\boldsymbol{z}$ for all t (a parametric description of the line through $\boldsymbol{y}$ and $\boldsymbol{z}$, section 1.2.2). Consider the first row of $A\boldsymbol{x}$: by Definition 2.2.2 it is

$$\begin{aligned} & a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ = {} & a_{11}[ty_1 + (1-t)z_1] + a_{12}[ty_2 + (1-t)z_2] \\ & + \cdots + a_{1n}[ty_n + (1-t)z_n] \\ & \quad \text{(then rearrange this scalar expression)} \\ = {} & t[a_{11}y_1 + a_{12}y_2 + \cdots + a_{1n}y_n] \\ & + (1-t)[a_{11}z_1 + a_{12}z_2 + \cdots + a_{1n}z_n] \end{aligned}$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$
\begin{aligned}
&= t[\text{first row of } A\boldsymbol{y}] + (1-t)[\text{first row of } A\boldsymbol{z}] \\
&= tb_1 + (1-t)b_1 \quad (\text{as } A\boldsymbol{y} = \boldsymbol{b} \text{ and } A\boldsymbol{z} = \boldsymbol{b}) \\
&= b_1 \,.
\end{aligned}
$$

Similarly for all rows of $A\boldsymbol{x}$: that is, each row in $A\boldsymbol{x}$ equals the corresponding row of $\boldsymbol{b}$. Consequently, $A\boldsymbol{x} = \boldsymbol{b}$. Hence, if there are ever two distinct solutions, then there are an infinite number of solutions: $\boldsymbol{x} = t\boldsymbol{y} + (1-t)\boldsymbol{z}$ for all t. □

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 $A\boldsymbol{x} = \boldsymbol{0}$. 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 $\boldsymbol{x} = \boldsymbol{0}$ as a solution as the system is homogeneous.

(b) $\begin{cases} 2r + s - t = 0 \\ r + s + 2t = 0 \\ -2r + s = 3 \\ 2r + 4s - t = 0 \end{cases}$ is not homogeneous because there is a non-zero constant on the right-hand side.

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

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

$$
\begin{cases} x_1 + 2x_2 + 4x_3 - 3x_4 = 0 \\ x_1 + 2x_2 - 3x_3 + 6x_4 = 0 \end{cases} \iff \left[\begin{array}{cccc:c} 1 & 2 & 4 & -3 & 0 \\ 1 & 2 & -3 & 6 & 0 \end{array}\right]
$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Subtract the first row from the second.

$$\begin{cases} x_1 + 2x_2 + 4x_3 - 3x_4 = 0 \\ 0x_1 + 0x_2 - 7x_3 + 9x_4 = 0 \end{cases} \iff \left[\begin{array}{cccc:c} 1 & 2 & 4 & -3 & 0 \\ 0 & 0 & -7 & 9 & 0 \end{array}\right]$$

Divide the second row by (-7).

$$\begin{cases} x_1 + 2x_2 + 4x_3 - 3x_4 = 0 \\ 0x_1 + 0x_2 + x_3 - \frac{9}{7}x_4 = 0 \end{cases} \iff \left[\begin{array}{cccc:c} 1 & 2 & 4 & -3 & 0 \\ 0 & 0 & 1 & -\frac{9}{7} & 0 \end{array}\right]$$

Subtract four times the second row from the first.

$$\begin{cases} x_1 + 2x_2 + 0x_3 + \frac{15}{7}x_4 = 0 \\ 0x_1 + 0x_2 + x_3 - \frac{9}{7}x_4 = 0 \end{cases} \iff \left[\begin{array}{cccc:c} 1 & 2 & 0 & \frac{15}{7} & 0 \\ 0 & 0 & 1 & -\frac{9}{7} & 0 \end{array}\right]$$

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

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

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

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

■

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

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

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Proof. The zero vector, $\boldsymbol{x} = \mathbf{0}$ in $\mathbb{R}^n$, is a solution of $A\boldsymbol{x} = \mathbf{0}$ so a homogeneous system is always consistent. In the reduced row echelon form there at most m leading variables—one for each row. Here $n > m$ and so the number of free variables is at least $n-m > 0$. Hence there is at least one free variable and consequently 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 2015a, §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.

2.2.4 Exercises

Exercise 2.2.1. For each of the following systems, write down two different matrix-vector forms of the equations. For each system: how many different possible matrix-vector forms could be written down?

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

(b) $\begin{aligned} -2p - q + 1 &= 0 \\ p - 6q &= 2 \end{aligned}$

(c) $\begin{aligned} 7.9x - 4.7y &= -1.7 \\ 2.4x - 0.1y &= 1.00 \\ -3.1x + 2.7y &= 2.30 \end{aligned}$

(d) $\begin{aligned} 3a + 4b - \tfrac{5}{2}c &= 0 \\ -\tfrac{7}{2}a + b - \tfrac{9}{2}c - \tfrac{9}{2} &= 0 \end{aligned}$

(e) $\begin{aligned} u + v - 2w &= -1 \\ -2u - v + 2w &= 3 \\ u + v + 5w &= 2 \end{aligned}$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Exercise 2.2.2. Use Procedure 2.2.5 in MATLAB/Octave to try to solve each of the systems of Exercise 2.2.1.

Exercise 2.2.3. Use Procedure 2.2.5 in MATLAB/Octave to try to solve each of the following systems.

(a) $$\begin{aligned} -5x - 3y + 5z &= -3 \\ 2x + 3y - z &= -5 \\ -2x + 3y + 4z &= -3 \end{aligned}$$

(b) $$\begin{aligned} -p + 2q - r &= -2 \\ -2p + q + 2z &= 1 \\ -3p + 4q &= 4 \end{aligned}$$

(c) $$\begin{aligned} u + 3v + 2w &= -1 \\ 3v + 5w &= 1 \\ -u + 3w &= 2 \end{aligned}$$

(d) $$\begin{aligned} -4a - b - 3c &= -2 \\ 2a - 4c &= 4 \\ a - 7c &= -2 \end{aligned}$$

Exercise 2.2.4. Use elementary row operations (Theorem 2.2.18) to solve the systems in Exercise 2.2.3.

Exercise 2.2.5. Use Procedure 2.2.5 in MATLAB/Octave to try to solve each of the following systems.

(a) $$\begin{aligned} 2.2x_1 - 2.2x_2 - 3.5x_3 - 2.2x_4 &= 2.9 \\ 4.8x_1 + 1.8x_2 - 3.1x_3 - 4.8x_4 &= -1.6 \\ -0.8x_1 + 1.9x_2 - 3.2x_3 + 4.1x_4 &= -5.1 \\ -9x_1 + 3.5x_2 - 0.7x_3 + 1.6x_4 &= -3.3 \end{aligned}$$

(b) $$\begin{aligned} 0.7c_1 + 0.7c_2 + 4.1c_3 - 4.2c_4 &= -0.70 \\ c_1 + c_2 + 2.1c_3 - 5.1c_4 &= -2.8 \\ 4.3c_1 + 5.4c_2 + 0.5c_3 + 5.5c_4 &= -6.1 \\ -0.6c_1 + 7.2c_2 + 1.9c_3 - 0.6c_4 &= -0.3 \end{aligned}$$

Exercise 2.2.6. Each of the following show some MATLAB/Octave commands and their results. Write down a possible problem that these commands aim to solve, and interpret what the results mean for the problem.

(a)
```
>> A=[1.1 2 5.6; 0.4 5.4 0.5; 2 -0.2 -2.8]
A =
    1.1000    2.0000    5.6000
    0.4000    5.4000    0.5000
    2.0000   -0.2000   -2.8000
>> b=[-3;2.9;1]
b =
   -3.0000
    2.9000
```

v0.2a

```
    1.0000
>> rcond(A)
ans =
    0.2936
>> x=A\b
x =
   -0.3936
    0.6294
   -0.6832
```

Exercise 2.2.7. Which of the following systems are in reduced row echelon form? For those that are, determine all solutions, if any.

(a) $$\begin{aligned} x_1 &= -194 \\ x_2 &= 564 \\ x_3 &= -38 \\ x_4 &= 275 \end{aligned}$$

(b) $$\begin{aligned} y_1 - 13.3y_4 &= -13.1 \\ y_2 + 6.1y_4 &= 5.7 \\ y_3 + 3.3y_4 &= 3.1 \end{aligned}$$

(c) $$\begin{aligned} z_1 - 13.3z_3 &= -13.1 \\ z_2 + 6.1z_3 &= 5.7 \\ 3.3z_3 + z_4 &= 3.1 \end{aligned}$$

(d) $$\begin{aligned} a - d &= -4 \\ b - \tfrac{7}{2}d &= -29 \\ c - \tfrac{1}{4}d &= -\tfrac{7}{2} \end{aligned}$$

(e) $$\begin{aligned} x + 0y &= 0 \\ 0x + y &= 0 \\ 0x + 0y &= 1 \\ 0x + 0y &= 0 \end{aligned}$$

(f) $$\begin{aligned} x + 0y &= -5 \\ 0x + y &= 1 \\ 0x + 0y &= 3 \end{aligned}$$

Exercise 2.2.8. For the following rational expressions, express the task of finding the partial fraction decomposition as a system of linear equations. Solve the system to find the decomposition. Record your working.

(a) $$\frac{-x^2 + 2x - 5}{x^2(x-1)}$$

(b) $$\frac{-4x^3 + 2x^2 - x + 2}{(x+1)^2(x-1)^2}$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(c)

$$\frac{5x^4 - x^3 + 3x^2 + 10x - 1}{x^2(x+2)(x-1)^2}$$

(d)

$$\frac{4x^4 + 2x^3 - x^2 - 7x - 2}{(x+1)^3(x-1)^2}$$

Exercise 2.2.9. For each of the following tables of data, use a system of linear equations to determine the nominated polynomial that finds the second column as a function of the first column. Sketch a graph of your fitted polynomial and the data points. Record your working.

(a) linear

x	y
2	-4
3	4

(b) quadratic

x	y
-2	-1
1	0
2	5

(c) quadratic

p	q
0	-1
2	3
3	4

(d) cubic

r	t
-3	-4
-2	0
-1	-3
0	-6

Exercise 2.2.10. In three consecutive years a company sells goods to the value of \$51M, \$81M and \$92M (in millions of dollars). Find a quadratic that fits this data, and use the quadratic to predict the value of sales in the fourth year.

Exercise 2.2.11. In 2011 there were 98 wolves in Yellowstone National Park; in 2012 there were 83 wolves; and in 2013 there were 95 wolves. Find a quadratic that fits this data, and use the quadratic to predict the number of wolves in 2014. To keep the coefficients manageable, write the quadratic in terms of the number of years from the starting year of 2011.

Exercise 2.2.12. Table 2.4 lists the time taken by a planet to orbit the Sun and a typical distance of the planet from the Sun. Analogous to Example 2.2.12, fit a quadratic polynomial $T = c_1 + c_2R + c_3R^2$ for the period T as a function of distance R. Use the data for Mercury, Venus and Earth. Then use the quadratic to predict the period

v0.2a

Table 2.4: orbital periods for four 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].

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

of Mars: what is the error in your prediction? (Example 3.5.10 shows a power law fit is better, and that the power law agrees with Kepler's law.)

Exercise 2.2.13 (Global Positioning System in space-time)**.** For each case below, and in space-time, suppose you know from five GPS satellites that you and your GPS receiver are the given measured time shift away from the given locations of each of the three satellites (locations are in Mm). Following Example 2.2.13, determine both your position and the discrepancy in time between your GPS receiver and the satellites GPS time. Which case needs another satellite?

(a) $\begin{cases} 0.03\,s & \text{time shift before } (17,11,17) \\ 0.03\,s & \text{time shift before } (11,20,14) \\ 0.0233\cdots s & \text{time shift before } (20,10,9) \\ 0.03\,s & \text{time shift before } (9,13,21) \\ 0.03\,s & \text{time shift before } (7,24,8) \end{cases}$

(b) $\begin{cases} 0.1\,s & \text{time shift before } (11,12,18) \\ 0.1066\cdots s & \text{time shift before } (18,6,19) \\ 0.1\,s & \text{time shift before } (11,19,9) \\ 0.1066\cdots s & \text{time shift before } (9,10,22) \\ 0.1\,s & \text{time shift before } (23,3,9) \end{cases}$

(c) $\begin{cases} 0.03\,s & \text{time shift before } (17,11,17) \\ 0.03\,s & \text{time shift before } (19,12,14) \\ 0.0233\cdots s & \text{time shift before } (20,10,9) \\ 0.03\,s & \text{time shift before } (9,13,21) \\ 0.03\,s & \text{time shift before } (7,24,8) \end{cases}$

Exercise 2.2.14. Formulate the following two thousand year old Chinese puzzle as a system of linear equations. Use algebraic manipulation to solve the system.

> There are three classes of grain, of which three bundles of the first class, two of the second, and one of the third make 39 measure. Two of the first, three of the second, and one of the third make 34 measures. And one of the

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

first, two of the second, and three of the third make 26 measures. How many measures of grain are contained in one bundle of each class?

*Jiuzhang Suanshu, 200*BC *(Chartier 2015, p.3)*

Exercise 2.2.15. Suppose you are given data at n points, equi-spaced in x. Say the known data points are $(1, y_1)$, $(2, y_2)$, $\dots$, (n, y_n) for some given $y_1, y_2, \dots, y_n$. Seek a polynomial fit to the data of degree $(n-1)$; that is, seek the fit $y = c_1 + c_2 x + c_3 x^2 + \cdots + c_n x^{n-1}$. In MATLAB/Octave, form the matrix of the linear equations that need to be solved for the coefficients $c_1, c_2, \dots, c_n$. According to Procedure 2.2.5, for what number n of data points is `rcond` good? poor? bad? terrible?

Exercise 2.2.16 (rational functions)**.** Sometimes we wish to fit rational functions to data. This fit also reduces to solving linear equations for the coefficients of the rational function. For example, to fit the rational function $y = a/(1 + bx)$ to data points $(x, y) = (-7, -\frac{1}{9})$ and $(x, y) = (2, \frac{1}{3})$ we need to satisfy the two equations

$$-\frac{1}{9} = \frac{a}{1 - 7b} \quad \text{and} \quad \frac{1}{3} = \frac{a}{1 + 2b}\,.$$

Multiply both sides of each by their denominator to require

$$\begin{aligned} & -\tfrac{1}{9}(1 - 7b) = a \quad \text{and} \quad \tfrac{1}{3}(1 + 2b) = a \\ \iff\ & a - \tfrac{7}{9}b = -\tfrac{1}{9} \quad \text{and} \quad a - \tfrac{2}{3}b = \tfrac{1}{3}\,. \end{aligned}$$

By hand (or MATLAB/Octave) solve this pair of linear equations to find $a = 3$ and $b = 4$. Hence the required rational function is $y = 3/(1 + 4x)$.

(a) Similarly, use linear equations to fit the rational function $y = (a_0 + a_1 x)/(1 + b_1 x)$ to the three data points $(1, \frac{7}{2})$, $(3, \frac{19}{4})$ and $(4, 5)$.

(b) Similarly, use linear equations to fit the rational function $y = (a_0 + a_1 x + a_2 x^2)/(1 + b_1 x + b_2 x^2)$ to the five data points $(-\frac{5}{2}, \frac{69}{44})$, $(-1, \frac{3}{2})$, $(\frac{1}{2}, \frac{9}{8})$, $(1, \frac{5}{4})$ and $(2, \frac{15}{11})$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

2.3 Linear combinations span sets

Section Contents

2.3.1 Exercises 138

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

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

The general solution to Example 2.2.21a 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 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): any other vector in $\mathbb{R}^3$ may be written as

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

The wide-spread appearance of such 'linear 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 10% error).

(a)

Solution: By visualising various combinations:
$\boldsymbol{a} \approx \boldsymbol{v}_1 + \boldsymbol{v}_2 = 1\boldsymbol{v}_1 + 1\boldsymbol{v}_2$;
$\boldsymbol{b} \approx 2\boldsymbol{v}_1 - \boldsymbol{v}_2 = 2\boldsymbol{v}_1 + (-1)\boldsymbol{v}_2$;
$\boldsymbol{c} \approx -0.2\boldsymbol{v}_1 + 0.6\boldsymbol{v}_2$.

(b)

Solution: By visualising various combinations:
$\boldsymbol{a} \approx -1\boldsymbol{v}_1 + 4\boldsymbol{v}_2$;
$\boldsymbol{b} \approx 2\boldsymbol{v}_1 - 2\boldsymbol{v}_2$;
$\boldsymbol{c} \approx -1\boldsymbol{v}_1 + 2\boldsymbol{v}_2$.

■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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

■

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

(a) For each value of t, the expression $(3,4) + t(-1,2)$ is a linear combination of the two vectors $(3,4)$ and $(-1,2)$. Over all values of parameter t it describes the line illustrated in the margin. (The line is alternatively described as $2x + y = 10$.)

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

(c) The expression $t(-1,2,0) + t^2(0,2,1)$ is a linear combination of the two vectors $(-1,2,0)$ and $(0,2,1)$ as the vectors are multiplied by scalars and then added. That a coefficient is a nonlinear function of some parameter is irrelevant to the property of linear combination. This expression is the parametric description of a parabola in $\mathbb{R}^3$, as illustrated below, and very soon we will be able to say it is a parabola in the plane spanned by $(-1,2,0)$ and $(0,2,1)$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

■

Example 2.3.5. The matrix-vector form $A\boldsymbol{x} = \boldsymbol{b}$ of a system of linear equations involves a linear combination on the left-hand side. Recall from Definition 2.2.2 that $A\boldsymbol{x} = \boldsymbol{b}$ is our abstract abbreviation for the system of m equations

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

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, $\boldsymbol{a}_1 = (a_{11}, a_{21}, \dots, a_{m1})$, $\boldsymbol{a}_2 = (a_{12}, a_{22}, \dots, a_{m2})$, $\dots$, $\boldsymbol{a}_n = (a_{1n}, a_{2n}, \dots, a_{mn})$, then the left-hand side is a linear combination of these vectors, with the coefficients of the linear combination being $x_1, x_2, \dots, x_n$. ■

Theorem 2.3.6. *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 here: this theorem turns a question about the existence of n variable solution $\boldsymbol{x}$, into a question about vectors with m components; and vice-versa.

Proof. Example 2.3.5 establishes that if a solution $\boldsymbol{x}$ exists, then $\boldsymbol{b}$ is a linear combination of the columns. Conversely, if $\boldsymbol{b}$ is a linear combination of the columns, then a solution $\boldsymbol{x}$ exists with components of $\boldsymbol{x}$ set to the coefficients in the linear combination. □

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Example 2.3.7. 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) $\begin{bmatrix}-1\\2\end{bmatrix}x=\begin{bmatrix}-2\\4\end{bmatrix}$.

Solution: The system is consistent because $x=2$ is a solution (Procedure 2.2.24). Also, the right-hand side $\boldsymbol{b}=(-2,4)$ is the linear combination $2(-1,2)$ of the column of the matrix.

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

Solution: The system is inconsistent as the first equation requires $x=-2$ whereas the second requires $x=\frac{3}{2}$ and these cannot hold simultaneously (Procedure 2.2.24). Also, there is no multiple of $(-1,2)$ that gives the right-hand side $\boldsymbol{b}=(2,3)$ so the right-hand side cannot be a linear combination of the column of the matrix—as illustrated in the margin.

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

Solution: The first equation requires $x=3$ whereas the second equation requires $ax=-6$; that is, $a\cdot 3=-6$, that is, $a=-2$. Thus it is only for $a=-2$ that the system is consistent; for $a\neq -2$ the system is inconsistent. Also, plotted in the margin are vectors $(1,a)$ for various a. It is only for $a=-2$ that the vector is aligned towards the given $(3,-6)$. Hence it is only for $a=-2$ that a linear combination of $(1,a)$ can give the required $(3,-6)$.

■

Activity 2.3.8. 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) $a=-1/2$ (c) $a=1$ (d) $a=2$

■

In the Examples 2.3.4 and 2.3.5 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.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Definition 2.3.9. *Let a set of k vectors in $\mathbb{R}^n$ be $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k\}$, then the set of all linear combinations of $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k$ is called the* ***span*** *of $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k$, and is denoted by $\operatorname{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k\}$ or $\operatorname{span} S$.*[4]

Example 2.3.10. (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 in the margin.

(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

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

Since every vector in $\mathbb{R}^2$ can be expressed as $s(-1, 2) + t(3, 4)$, 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 in the margin.

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

[4] In the degenerate case of the set S being the empty set, we take its span to be just the zero vector; that is, by convention $\operatorname{span}\{\} = \{\mathbf{0}\}$. But we rarely need this degenerate case.

v0.2a

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

(e) For the complete set of n standard unit vectors in $\mathbb{R}^n$ (Definition 1.2.7), $\operatorname{span}\{\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n\} = \mathbb{R}^n$. This is because any vector $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$ 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 is in $\operatorname{span}\{\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n\}$.

(f) The homogeneous system (Definition 2.2.28) of linear equations from Example 2.2.29d 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 some vectors.

■

Activity 2.3.11. In the margin is drawn a line: for which one of the following vectors $\boldsymbol{u}$ is $\operatorname{span}\{\boldsymbol{u}\}$ *not* the drawn line?

(a) $(2\,,1)$ (b) $(-1\,,-0.5)$

(c) $(-1\,,-2)$ (d) $(4\,,2)$

■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Example 2.3.12. Describe in other words $\operatorname{span}\{\boldsymbol{i},\boldsymbol{k}\}$ in $\mathbb{R}^3$.

Solution: All vectors in $\operatorname{span}\{\boldsymbol{i},\boldsymbol{k}\}$ are of the form $c_1\boldsymbol{i}+c_2\boldsymbol{k} = c_1(1,0,0)+c_2(0,0,1) = (c_1,0,c_2)$. Hence the span is all vectors with second component zero—the plane $y=0$ in (x,y,z) coordinates.

■

Example 2.3.13. 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 $\operatorname{span} T = \{(-a-2b-2, -b+1, -3b-1) : a,b \text{ scalars}\}$.

Solution: Because vectors $(3b, a+b, -2a-4b) = a(0,1,-2) + b(3,1,-4)$ for all scalars a and b, a suitable set is $S = \{(0,1,-2), (3,1,-4)\}$.

Second, vectors $(-a-2b-2, -b+1, 3b-1) = a(-1,0,0)+b(-2,-1,3)+(-2,1,-1)$ which are linear combinations for all a and b. However, the vectors cannot form a span due to the constant vector $(-2,1,-1)$ because a span requires *all* linear combinations of its component vectors. The given set cannot be expressed as a span.

■

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).

2.3.1 Exercises

Exercise 2.3.1. For each of the following, express vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ as a linear combination of vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$. Estimate the coefficients roughly (to say 10%).

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(e)

(f)

(g)

(h)

Exercise 2.3.2. For each of the following lines in 2D, write down a parametric equation of the line as a linear combination of two vectors, one of which is multiplied by the parameter.

(a)

(b)

(c)

(d)

(e)

(f)

Exercise 2.3.3. Write each of the following systems of linear equations as one vector equation involving a linear combination of vectors.

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

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

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(c)
$$\begin{aligned} x_1 + 3x_2 + x_3 - 2x_4 &= 2 \\ 2x_1 + x_2 + 4x_3 - 2x_4 &= -1 \\ -x_1 + 2x_2 - 2x_3 - x_4 &= 3 \end{aligned}$$

(d)
$$\begin{aligned} -2p - 2q &= -1 \\ q &= 2 \\ 3p - q &= 1 \end{aligned}$$

Exercise 2.3.4. For each of the cases in Exercise 2.3.3, by attempting to solve the system, determine if the right-hand side vector is in the span of the vectors on the left-hand side.

Exercise 2.3.5. For each of the following sets, write the set as a span, if possible. Give reasons.

(a) $\{(p - 4q, p + 2q, p + 2q) : p, q \text{ scalars}\}$

(b) $\{(-p + 2r, 2p - 2q, p + 2q + r, -q - 3r) : p, q, r \text{ scalars}\}$

(c) The line $y = 2x + 1$ in $\mathbb{R}^2$.

(d) The line $x = y = z$ in $\mathbb{R}^3$.

(e) The set of vectors $\boldsymbol{x}$ in $\mathbb{R}^4$ with component $x_3 = 0$.

Exercise 2.3.6. Show the following identities hold for any given vectors $\boldsymbol{u}$,$\boldsymbol{v}$ and $\boldsymbol{w}$:

(a) $\operatorname{span}\{\boldsymbol{u}, \boldsymbol{v}\} = \operatorname{span}\{\boldsymbol{u} - \boldsymbol{v}, \boldsymbol{u} + \boldsymbol{v}\}$;

(b) $\operatorname{span}\{\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}\} = \operatorname{span}\{\boldsymbol{u}, \boldsymbol{u} - \boldsymbol{v}, \boldsymbol{u} + \boldsymbol{v} + \boldsymbol{w}\}$.

Exercise 2.3.7. Suppose $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_s$ are any s vectors in $\mathbb{R}^n$. Let set $R = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_r\}$ for some $r < s$, and set $S = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_r, \boldsymbol{u}_{r+1}, \dots, \boldsymbol{u}_s\}$.

(a) Prove that $\operatorname{span} R \subseteq \operatorname{span} S$.

(b) Hence deduce that if $\operatorname{span} R = \mathbb{R}^n$, then $\operatorname{span} S = \mathbb{R}^n$.

Exercise 2.3.8. Suppose $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_r$ and $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_s$ are all vectors in $\mathbb{R}^n$.

(a) Prove that if every vector $\boldsymbol{u}_j$ is a linear combination of $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_s$, then $\operatorname{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_r\} \subseteq \operatorname{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_s\}$.

(b) Prove that if, additionally, every vector $\boldsymbol{v}_j$ is a linear combination of $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_r$, then $\operatorname{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_r\} = \operatorname{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_s\}$.

Exercise 2.3.9. Let $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_s\}$ be a set of vectors in $\mathbb{R}^n$ such that vector $\boldsymbol{v}_1$ is a linear combination of $\boldsymbol{v}_2, \boldsymbol{v}_3, \dots, \boldsymbol{v}_s$. Prove that $\operatorname{span} S = \operatorname{span}\{\boldsymbol{v}_2, \boldsymbol{v}_3, \dots, \boldsymbol{v}_s\}$.

v0.2a

Answers to selected exercises

2.1.1b : $(x\,,y)=(1\,,2)$

2.1.1d : line $y=\frac{3}{2}x-1$

2.1.1f : $(x\,,y)=(-1\,,-1)$

2.1.1h : $(p\,,q)=(\frac{1}{2}\,,-\frac{1}{2})$

2.1.1j : no solution

2.1.1l : line $t=4s-2$

2.1.2b : $(1.8\,,1.04)$

2.1.2d : $(1.25\,,0.41)$

2.1.2f : $(2.04\,,0.88)$

2.1.2h : $(2.03\,,0.34)$

2.1.3b : no solution

2.1.3d : no solution

2.1.3f : no solution

2.1.4b : $(4\,,5)$

2.1.4d : $(-1\,,11)/9$ surely indicates an error.

2.2.1b : e.g. $\begin{bmatrix}-2&-1\\1&-6\end{bmatrix}\begin{bmatrix}p\\q\end{bmatrix}=\begin{bmatrix}-1\\2\end{bmatrix}$, $\begin{bmatrix}1&-6\\-2&-1\end{bmatrix}\begin{bmatrix}p\\q\end{bmatrix}=\begin{bmatrix}2\\-1\end{bmatrix}$. Four: two orderings of rows, and two orderings of the variables.

2.2.1d : Twelve possibilities: two orderings of rows, and $3!=6$ orderings of the variables.

2.2.2 :

1. $(x\,,y)=(-0.4\,,-1.2)$
2. $(p\,,q)=(0.6154\,,-0.2308)$
3. No solution as `rcond` requires a square matrix.
4. No solution as `rcond` requires a square matrix.
5. $(u\,,v\,,w)=(-2\,,1.8571\,,0.4286)$

2.2.3b : $(p\,,q\,,r)=(32\,,25\,,20)$

2.2.3d : $(a\,,b\,,c)=(3.6\,,-14.8\,,0.8)$

2.2.5a : $\boldsymbol{x}=(-0.26\,,-1.33\,,0.18\,,-0.54)$ (2 d.p.)

2.2.6a : Solve the system $1.1x_1+2x_2+5.6x_3=-3$, $0.4x_1+5.4x_2+0.5x_3=2.9$, $2x_1-0.2x_2-2.8x_3=1$. Since `rcond` is good, the solution is $x_1=-0.39$, $x_2=0.63$ and $x_3=-0.68$ (2 d.p.).

2.2.7b : Yes, $\boldsymbol{y}=(-13.1+13.3t\,,5.7-6.1t\,,3.1-3.3t\,,t)$ for all t

2.2.7d : Yes. $(a\,,b\,,c\,,d)=(-4+t\,,-29+\frac{7}{2}t\,,-\frac{7}{2}+\frac{1}{4}t\,,t)$ for all t

v0.2a

2.2.7f : Not in RREF as the last does not have a leading one.

2.2.8b : $-\frac{2}{x+1} + \frac{9/4}{(x+1)^2} - \frac{2}{x-1} - \frac{1}{(x-1)^2}$

2.2.8d : $\frac{19/8}{x-1} - \frac{1/2}{(x-1)^2} + \frac{13/8}{x+1} - \frac{9/4}{(x+1)^2} + \frac{3/2}{(x+1)^3}$

2.2.9b : $y = -\frac{8}{3} + \frac{3}{2}x + \frac{7}{6}x^2$

2.2.9d : $t = -6 - \frac{2}{3}t + \frac{7}{2}t^2 + \frac{7}{6}t^3$

2.2.11 : 134 wolves

2.2.13a : `rcond` $= 0.014$, $(3\,,4\,,3)$, shift $= 12/300 = 0.04\,\mathrm{s}$

2.2.13c : `rcond` $= 0$, singular, needs another satellite.

2.2.15 : Good, $\{1\,,2\}$; poor, $\{3\,,4\}$; bad, $\{5\,,6\,,7\}$; terrible, $\{8\,,9\,,\ldots\}$.

2.2.16b : $y = (1 + x + 3x^2)/(1 + x + 2x^2)$

2.3.1b : $\boldsymbol{a} = -1\boldsymbol{v}_1 + 1.5\boldsymbol{v}_2$, $\boldsymbol{b} = 1\boldsymbol{v}_1 + 3\boldsymbol{v}_2$

2.3.1d : $\boldsymbol{a} = 0.3\boldsymbol{v}_1 - 1.7\boldsymbol{v}_2$, $\boldsymbol{b} = -1.4\boldsymbol{v}_1 + 3.3\boldsymbol{v}_2$

2.3.1f : $\boldsymbol{a} = 2.3\boldsymbol{v}_1 - 2.6\boldsymbol{v}_2$, $\boldsymbol{b} = -1.4\boldsymbol{v}_1 - 0.4\boldsymbol{v}_2$

2.3.1h : $\boldsymbol{a} = -1.8\boldsymbol{v}_1 + 1.5\boldsymbol{v}_2$, $\boldsymbol{b} = -1.5\boldsymbol{v}_1 - 0.6\boldsymbol{v}_2$

2.3.2b : e.g. $(0\,,3.5) + t(-2.5\,,2)$

2.3.2d : e.g. $(-1.5\,,0.5) + t(1\,,-1.5)$

2.3.2f : e.g. $(0.5\,,1) + t(0.5\,,1)$

2.3.3b : $\begin{bmatrix}-3\\0\\1\end{bmatrix} x + \begin{bmatrix}2\\1\\-3\end{bmatrix} y + \begin{bmatrix}-3\\1\\0\end{bmatrix} z = \begin{bmatrix}0\\0\\0\end{bmatrix}$

2.3.3d : $\begin{bmatrix}-2\\0\\3\end{bmatrix} p + \begin{bmatrix}-2\\1\\-1\end{bmatrix} q = \begin{bmatrix}-1\\2\\1\end{bmatrix}$

2.3.5a : e.g. $\operatorname{span}\{(1\,,1\,,1)\,,(-4\,,2\,,2)\}$

2.3.5c : Not a span.

2.3.5e : e.g. $\operatorname{span}\{\boldsymbol{e}_1\,,\boldsymbol{e}_2\,,\boldsymbol{e}_4\}$.

v0.2a

3 Matrices encode system interactions

Chapter Contents

3.1 Matrix operations and algebra 145

3.1.1 Basic matrix terminology 145

3.1.2 Addition, subtraction and multiplication with matrices 148

3.1.3 Familiar algebraic properties of matrix operations 165

3.1.4 Exercises 171

3.2 The inverse of a matrix 177

3.2.1 Introducing the unique inverse 177

3.2.2 Diagonal matrices stretch and shrink 188

3.2.3 Orthogonal matrices rotate 197

3.2.4 Exercises 206

3.3 Factorise to the singular value decomposition 221

3.3.1 Introductory examples 221

3.3.2 The SVD solves general systems 226

3.3.3 Prove the SVD Theorem 3.3.6 244

3.3.4 Exercises 249

3.4 Subspaces, basis and dimension 259

3.4.1 Subspaces are lines, planes, and so on 260

3.4.2 Orthonormal bases form a foundation 270

3.4.3 Is it a line? a plane? The dimension answers 282

3.4.4 Exercises 290

3.5 Project to solve inconsistent equations 299

3.5.1 Make a minimal change to the problem . . . 299

3.5.2 Compute the smallest appropriate solution . 314

3.5.3 Orthogonal projection resolves vector components 321

3.5.4 Exercises 349

v0.2a

3.6 Introducing linear transformations 364
3.6.1 Matrices correspond to linear transformations 370
3.6.2 The pseudo-inverse of a matrix 374
3.6.3 Function composition connects to matrix inverse 383
3.6.4 Exercises 390

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)*

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

3.1 Matrix operations and algebra

Section Contents

3.1.1 Basic matrix terminology 145
3.1.2 Addition, subtraction and multiplication with matrices 148
Matrix addition and subtraction 148
Scalar multiplication of matrices 149
Matrix-vector multiplication transforms . . . 150
Matrix-matrix multiplication 153
The transpose of a matrix 157
Compute in MATLAB/Octave 159
3.1.3 Familiar algebraic properties of matrix operations 165
3.1.4 Exercises 171

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

3.1.1 Basic matrix terminology

Let's start with some basic definitions of terminology.

- A **matrix** is a rectangular array of real numbers, written inside **brackets** $[\cdot]$, such as the following six: [1]

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

- 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 for a matrix 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.

[1] Chapter 7 starts using complex numbers in a matrix, but until then we stay within the realm of real numbers. Some books use parentheses, $(\cdot)$, around matrices: we do not as here parentheses denote a vector (most often used when written in text).

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

- 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 matrix horizontally within **parentheses** ( ). The third matrix in (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 in (3.1).

- 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.[2] 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: $O_{m\times n}$ 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**: I_n denotes a $n \times n$ square matrix which has zero entries except for the diagonal from the top-left to the bottom-right which are all ones. Occasionally we invoke non-square 'identity' matrices denoted $I_{m\times n}$. For examples,

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

[2] Some books use the capital letter subscripted for its entries: that is, some use A_{ij} to denote the entry in the ith row and jth column of matrix A.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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 2.3.5, 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 we could write these column vectors 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

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

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

(c) $\begin{bmatrix} \sqrt{9} & -1 & 2^2 \\ -2 & 0 & \cos 0 \end{bmatrix}$ (d) $\begin{bmatrix} 3 & 1-2 & \sqrt{16} \\ 3-2 & 0 & e^0 \end{bmatrix}$

■

v0.2a

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 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 implications of its associated operations.

Matrix addition and subtraction

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

Example 3.1.2. Let matrices

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

Then the addition and subtraction

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

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, suppose 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,

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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,

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

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

For example, letting O denote the zero matrix of the appropriate size, then

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

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

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

■

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 matrices

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

Then the scalar multiplications

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

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

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

■

v0.2a

In general, suppose A is an $m \times n$ matrix, with entries a_{ij}, then we define the **scalar product** by c, either cA or Ac, as the $m \times n$ matrix whose (i,j)th entry is ca_{ij}. [3] That is,

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

Matrix-vector multiplication transforms

Recall that the matrix-vector form of a system of linear equations, Definition 2.2.2, wrote $A\boldsymbol{x} = \boldsymbol{b}$. 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 are the same as the number of components of vector $\boldsymbol{x}$. [4]

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

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

[3] Be aware that MATLAB/Octave reasonably treats multiplication by a '1×1 matrix' as a scalar multiplication. Strictly speaking products such as '$[0.35]\,A$' are not defined because strictly speaking $[0.35]$ is not a scalar but is a 1×1 matrix. However, MATLAB/Octave do not make the distinction.

[4] Some of you who have studied calculus may wonder about what might be called 'continuous matrices' $A(x\,,y)$ which multiply a function $f(x)$ according to the integral $\int_a^b A(x\,,y)f(y)\,dy$. Then you might wonder about solving problems such as find the unknown $f(x)$ such that $\int_0^1 A(x\,,y)f(y)\,dy = \sin \pi x$ for given 'continuous matrix' $A(x\,,y) := \min(x\,,y)\big[1 - \max(x\,,y)\big]$; you may check that here the solution is $f = \pi^2 \sin \pi x$. Such notions are a useful generalisation of our linear algebra: they are called integral equations; the main structures and patterns developed by this course also apply.

© A J Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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

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

Further, we do not here define 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 the definition. ■

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

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

■

Geometric interpretation Multiplication of a vector by a square matrix transforms the vector into another in the same space as shown in the margin for $A\boldsymbol{x}$ 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 marginal 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 marginal picture. Such a geometric interpretation underlies the use of matrix multiplication in video and picture processing, for example. Such video/picture processing employs stretching and shrinking (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

$$I_2\boldsymbol{x} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ -3 \end{bmatrix} = \begin{bmatrix} 1\cdot 2 + 0\cdot(-3) \\ 0\cdot 2 + 1\cdot(-3) \end{bmatrix} = \begin{bmatrix} 2 \\ -3 \end{bmatrix},$$
$$I_3\boldsymbol{b} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 4 \end{bmatrix} = \begin{bmatrix} 1\cdot 1 + 0\cdot 0 + 0\cdot 4 \\ 0\cdot 1 + 1\cdot 0 + 0\cdot 4 \\ 0\cdot 1 + 0\cdot 0 + 1\cdot 4 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 4 \end{bmatrix}.$$

v0.2a

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 3.1.24e). ■

Example 3.1.8 (age structured population)**.** An ecologist studies an isolated population of a species of animal. The growth of the population depends primarily upon the females so it is only these that are counted. The females are grouped into three ages: female pups (in their first year), juvenile females (one year old), and mature females (two years or older). During the study, the ecologist observes the following happens over the period of a year:

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

(a) Let x_1, x_2 and x_3 be the number of females at the start of a year, of ages zero, one and two+ respectively, and let x_1', x_2' and x_3' be their number at the start of the next year. Use the observations to write x_1', x_2' and x_3' as a function of x_1, x_2 and x_3 (this is called a Markov chain).

(b) Letting vectors $\boldsymbol{x} = (x_1, x_2, x_3)$ 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 ($\boldsymbol{x}'''$)?

Solution: (a) Since mature females breed and produce four female pups, $x_1' = 4x_3$. Since half of the female pups survive and become juvenile females, $x_2' = \frac{1}{2}x_1$. Since one-third of the juvenile females survive and become mature females, $\frac{1}{3}x_2$ contribute to x_3', but additionally one-third of the mature females survive to breed in the following year, so $x_3' = \frac{1}{3}x_2 + \frac{1}{3}x_3$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(b) Writing these equations into vector form

$$\boldsymbol{x}' = \begin{bmatrix} x_1' \\ x_2' \\ x_3' \end{bmatrix} = \begin{bmatrix} 4x_3 \\ \frac{1}{2}x_1 \\ \frac{1}{3}x_2 + \frac{1}{3}x_3 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 0 & 4 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} \end{bmatrix}}_{L} \boldsymbol{x}\,.$$

(c) Given the initial numbers of female animals is $\boldsymbol{x} = (60, 70, 20)$, the number of females after one year is then predicted by the matrix-vector product

$$\boldsymbol{x}' = L\boldsymbol{x} = \begin{bmatrix} 0 & 0 & 4 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 60 \\ 70 \\ 20 \end{bmatrix} = \begin{bmatrix} 80 \\ 30 \\ 30 \end{bmatrix}.$$

That is, the predicted numbers of females are 80 pups, 30 juveniles, and 30 mature.

After a second year the number of females is then predicted by the matrix-vector product $\boldsymbol{x}'' = L\boldsymbol{x}'$. Here

$$\boldsymbol{x}'' = L\boldsymbol{x}' = \begin{bmatrix} 0 & 0 & 4 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 80 \\ 30 \\ 30 \end{bmatrix} = \begin{bmatrix} 120 \\ 40 \\ 20 \end{bmatrix}.$$

After a third year the number of females is predicted by the matrix-vector product $\boldsymbol{x}''' = L\boldsymbol{x}''$. Here

$$\boldsymbol{x}''' = L\boldsymbol{x}'' = \begin{bmatrix} 0 & 0 & 4 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 120 \\ 40 \\ 20 \end{bmatrix} = \begin{bmatrix} 80 \\ 60 \\ 20 \end{bmatrix}.$$

■

Matrix-matrix multiplication

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

Example 3.1.9. Let 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

$$AB = A \begin{bmatrix} 5 & -6 & 4 \\ -1 & -3 & -3 \end{bmatrix}$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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

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

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

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

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

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

Example 3.1.10. Let matrices

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

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

Solution:

- On the one hand,

$$
\begin{aligned}
CD &= C\begin{bmatrix} -2 & -1 & -3 \\ 1 & 3 & 0 \end{bmatrix} \\
&= \left[C\begin{bmatrix} -2 \\ 1 \end{bmatrix} \vdots C\begin{bmatrix} -1 \\ 3 \end{bmatrix} \vdots C\begin{bmatrix} -3 \\ 0 \end{bmatrix} \right] \\
&= \begin{bmatrix} 7 & 1 & 12 \\ 7 & 1 & 12 \\ 2 & 11 & -3 \end{bmatrix}.
\end{aligned}
$$

- Conversely,

$$
DC = D\begin{bmatrix} -4 & -1 \\ -4 & -1 \\ 1 & 4 \end{bmatrix}
$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$= \left[D \begin{bmatrix} -4 \\ -4 \\ 1 \end{bmatrix} \vdots D \begin{bmatrix} -1 \\ -1 \\ 4 \end{bmatrix} \right]$$
$$= \begin{bmatrix} 9 & -9 \\ -16 & -4 \end{bmatrix} .$$

Interestingly, $CD \neq DC$.

■

Definition 3.1.11 (matrix product)**.** *Let matrix A be $m \times n$, and matrix B be $n \times p$, then the* ***matrix product*** *$C = AB$ is the $m \times p$ matrix whose (i, j)th entry is*

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

This formula looks like a dot product of two vectors (1.3.2): 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

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

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

Activity 3.1.12. 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 3.1.13. 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

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

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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 3.1.14 (age structured population)**.** Matrix powers occur naturally in modelling populations by ecologists such as the animals of Example 3.1.8. Recall that given the numbers of female pups, juveniles and mature aged 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 (Theorem 3.1.24c) 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. ■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

The transpose of a matrix

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

Example 3.1.15. 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 3.1.16 (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 3.1.17. Which of the following matrices is the transpose of matrix

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

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

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

■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Example 3.1.18 (transpose and dot product). Consider two vectors in $\mathbb{R}^n$, say $\boldsymbol{u} = (u_1, u_2, \dots, u_n)$ and $\boldsymbol{v} = (v_1, v_2, \dots, v_n)$; 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

$$\begin{aligned} \boldsymbol{u} \cdot \boldsymbol{v} &= u_1 v_1 + u_2 v_2 + \cdots + u_n v_n \quad \text{(Defn. 1.3.2 of dot)} \\ &= \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \quad \text{(Defn. 3.1.11 of mult.)} \\ &= \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \quad \text{(transpose Defn. 3.1.16)} \\ &= \boldsymbol{u}^{\mathrm{T}} \boldsymbol{v}\,. \end{aligned}$$

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

Definition 3.1.19 (symmetry). *A 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 3.1.20. None of the three matrices in Example 3.1.15 are symmetric: the first two matrices are not square so cannot be symmetric, and the third $C \neq C^{\mathrm{T}}$. 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?

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

Solution: Consider the transpose

$$E^{\mathrm{T}} = \begin{bmatrix} a & c \\ b & d \end{bmatrix} \quad \text{compared with } E = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

The top-left and bottom-right elements are always the same. The top-right and bottom-left elements will be the same if and only if (iff) $b = c$. That is, the 2×2 matrix E is symmetric iff $b = c$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

■

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

Activity 3.1.21. Which of the following matrices are a symmetric matrix?

(a) $\begin{bmatrix} 0 & -3.2 & -0.8 \\ 3.2 & 0 & 3.2 \\ 0.8 & -3.2 & 0 \end{bmatrix}$ (b) $\begin{bmatrix} 2.3 & -1.3 & -2 \\ -3.2 & -1 & -1.3 \\ -3 & -3.2 & 2.3 \end{bmatrix}$

(c) $\begin{bmatrix} -2.6 & 0.3 & -1.3 \\ 0.3 & -0.2 & 0 \\ -1.3 & 0 & -2 \end{bmatrix}$ (d) $\begin{bmatrix} 2.2 & -0.9 & -1.2 \\ -0.9 & -1.2 & -3.1 \end{bmatrix}$

■

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 summarises the Matlab/Octave version of the operations introduced so far, and used in the rest of this book.

Matrix size and elements Let the matrix

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

We readily see this is a 3×5 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)
```

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 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

v0.2a

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

- `size(A)` returns the number of rows and columns of matrix A: if A is $m \times n$, then `size(A)` returns $\begin{bmatrix} m & n \end{bmatrix}$.
- `A(i,j)` is the $(i\,,j)$th entry of a matrix A, `A(:,j)` is the jth column, `A(i,:)` is the ith 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 pth power of square matrix A (in contrast to `A.^p` which computes the pth power of *each element* of A, Table 2.3).
- The character single quote, `A'`, transposes the matrix A.
- Predefined matrices include:
  - `zeros(m,n)` is the zero matrix $O_{m\times n}$;
  - `eye(m,n)` is $m \times n$ 'identity matrix' $I_{m\times n}$;
  - `ones(m,n)` is the $m \times n$ matrix where all entries are one;
  - `randn(m,n)` is a $m \times n$ matrix with random entries (distributed Normally, mean zero, standard deviation one).

  A single argument gives the square matrix version:
  - `zeros(n)` is $O_n = O_{n\times n}$;
  - `eye(n)` is the $n \times n$ identity matrix $I_n = I_{n\times n}$;
  - `ones(n)` is the $n \times n$ matrix of all ones;
  - `randn(n)` is an $n \times n$ matrix with random entries.

  With no argument, these functions return the corresponding scalar: for example, `randn` computes a single random number.
- Very large and small magnitide numbers are printed in MATLAB/Octave like the following:
  - `4.852e+08` denotes the large $4.852 \cdot 10^8$; whereas
  - `3.469e-16` denotes the small $3.469 \cdot 10^{-16}$.

```
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
```

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

```
    -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 3.1.22a) by seeing that their difference is the 3×5 zero matrix: execute `(A+B)-(B+A)` (the parentheses control the order of evaluation). However, expressions such as `B+C` and `A-C` give an error, because the matrices are of incompatible sizes, reported by MATLAB as

v0.2a

```
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.[5] 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,

[5] Although in some contexts such mathematical expressions are routinely accepted, be careful of their meaning.

v0.2a

```
>> 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

v0.28

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 it only makes sense and works for square matrices A.[6] 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 is not the same as the number of rows of the right A such a multiplication is not allowed.

The transpose and symmetry In MATLAB/Octave the single apostrophe denotes matrix transpose. For example, see it transpose a couple of random matrices with

```
A=randn(3,4)
B=randn(4,2)
A'
B'
```

giving here for example (2 d.p.)

```
A =
   0.80   0.30  -0.12  -0.57
   0.07  -0.51  -0.81   1.95
   0.29  -0.10   0.17   0.70
B =
  -0.71  -0.34
  -0.33  -0.73
   1.11  -0.21
   0.41   0.33

>> A'
ans =
   0.80   0.07   0.29
   0.30  -0.51  -0.10
  -0.12  -0.81   0.17
```

[6] We define matrix powers for only integer power. MATLAB/Octave will compute the power of a square matrix for any real/complex exponent, but its meaning involves matrix exponentials and logarithms that we do not explore here.

v0.2a

```
  -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 3.1.27d) by verifying `(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 3.1.27f): 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.

3.1.3 Familiar algebraic properties of matrix operations

Almost all of the familiar algebraic properties of scalar addition, subtraction and multiplication—namely commutativity, associativity and distributivity—hold for matrix addition, subtraction and multiplication.

The one outstanding exception is that matrix multiplication is *not* commutative: for matrices A and B the products AB and BA are usually not equal. We are used to 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

v0.2a

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 3.1.22 (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) $0A = O$ *.*

Proof. The proofs directly match those of the corresponding vector properties and are set as exercises. □

Example 3.1.23 (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 3.1.14 invoked the associativity Theorem 3.1.24c. 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 in the margin. This is the same results as forming the product

$$AB = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 2 & -1 \end{bmatrix} = \begin{bmatrix} 4 & -1 \\ 2 & 0 \end{bmatrix}$$

and then computing $(AB)\boldsymbol{x} = (3, 2)$ as also illustrated in the margin. 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 3.1.24c. ■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Theorem 3.1.24 (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) $I_mA = A = AI_n$ *for $m\times n$ matrix A (multiplicative identity);*

(f) $O_mA = O_{m\times n} = AO_n$ *for $m\times n$ matrix A;*

(g) $A^pA^q = A^{p+q}$, $(A^p)^q = A^{pq}$ *and* $(cA)^p = c^pA^p$ *for square A and for positive integers p and q.*[7]

Proof. Let's document a few proofs, others are exercises.

3.1.24a : The direct proof involves some long expressions involving the entries of $m\times n$ matrix A, and $n\times p$ matrices B and C. Let $(\cdot)_{ij}$ denote the $(i\,,j)$th entry of whatever matrix expression is inside the parentheses. By Definition 3.1.11 of matrix multiplication

$$\begin{aligned}
&(A(B\pm C))_{ij}\\
&= a_{i1}(B\pm C)_{1j} + a_{i2}(B\pm C)_{2j} + \cdots + a_{in}(B\pm C)_{nj}\\
&\qquad\text{(by definition of matrix addition)}\\
&= a_{i1}(b_{1j}\pm c_{1j}) + a_{i2}(b_{2j}\pm c_{2j}) + \cdots + a_{in}(b_{nj}\pm c_{nj})\\
&\qquad\text{(distributing the scalar multiplications)}\\
&= a_{i1}b_{1j}\pm a_{i1}c_{1j} + a_{i2}b_{2j}\pm a_{i2}c_{2j} + \cdots + a_{in}b_{nj}\pm a_{in}c_{nj}\\
&\qquad\text{(upon reordering terms in the sum)}\\
&= a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}\\
&\quad\pm(a_{i1}c_{1j} + a_{i2}c_{2j} + \cdots + a_{in}c_{nj})\\
&\qquad\text{(using Definition 3.1.11 for matrix products)}\\
&= (AB)_{ij}\pm(AC)_{ij}\,.
\end{aligned}$$

Since this identity holds for all indices i and j, the matrix identity $A(B\pm C) = AB\pm AC$ holds, proving Theorem 3.1.24a.

3.1.24c : Associativity involves some longer expressions involving the entries of $m\times n$ matrix A, $n\times p$ matrix B, and $p\times q$ matrix C. By Definition 3.1.11 of matrix multiplication

$$\begin{aligned}
(A(BC))_{ij} &= a_{i1}(BC)_{1j} + a_{i2}(BC)_{2j} + \cdots + a_{in}(BC)_{nj}\\
&\qquad\text{(then using Definition 3.1.11 for } BC)\\
&= a_{i1}(b_{11}c_{1j} + b_{12}c_{2j} + \cdots + b_{1p}c_{pj})
\end{aligned}$$

[7] Generically, these exponent properties hold for all scalar p and q, although we have to be very careful with non-integer exponents.

$$
\begin{aligned}
&+ a_{i2}(b_{21}c_{1j} + b_{22}c_{2j} + \cdots + b_{2p}c_{pj}) \\
&+ \cdots \\
&+ a_{in}(b_{n1}c_{1j} + b_{n2}c_{2j} + \cdots + b_{np}c_{pj}) \\
&\quad \text{(distributing the scalar multiplications)} \\
= \quad & a_{i1}b_{11}c_{1j} + a_{i1}b_{12}c_{2j} + \cdots + a_{i1}b_{1p}c_{pj} \\
&+ a_{i2}b_{21}c_{1j} + a_{i2}b_{22}c_{2j} + \cdots + a_{i2}b_{2p}c_{pj} \\
&+ \cdots \\
&+ a_{in}b_{n1}c_{1j} + a_{in}b_{n2}c_{2j} + \cdots + a_{in}b_{np}c_{pj} \\
&\quad \text{(reordering the terms—transpose)} \\
= \quad & a_{i1}b_{11}c_{1j} + a_{i2}b_{21}c_{1j} + \cdots + a_{in}b_{n1}c_{1j} \\
&+ a_{i1}b_{12}c_{2j} + a_{i2}b_{22}c_{2j} + \cdots + a_{in}b_{n2}c_{2j} \\
&+ \cdots \\
&+ a_{i1}b_{1p}c_{pj} + a_{i2}b_{2p}c_{pj} + \cdots + a_{in}b_{np}c_{pj} \\
&\quad \text{(factoring } c_{1j}\,, c_{2j}\,, \ldots c_{pj}) \\
= \quad & (a_{i1}b_{11} + a_{i2}b_{21} + \cdots + a_{in}b_{n1})c_{1j} \\
&+ (a_{i1}b_{12} + a_{i2}b_{22} + \cdots + a_{in}b_{n2})c_{2j} \\
&+ \cdots \\
&+ (a_{i1}b_{1p} + a_{i2}b_{2p} + \cdots + a_{in}b_{np})c_{pj} \\
&\quad \text{(recognising the entries for } (AB)_{ik}) \\
&= (AB)_{i1}c_{1j} + (AB)_{i2}c_{2j} + \cdots + (AB)_{ip}c_{pj} \\
&\quad \text{(again using Definition 3.1.11)} \\
&= ((AB)C)_{ij}.
\end{aligned}
$$

Since this identity holds for all indices i and j, the matrix identity $A(BC) = (AB)C$ holds, proving Theorem 3.1.24c.

3.1.24g : Other proofs develop from previous parts of the theorem. For example, to establish $A^pA^q = A^{p+q}$ start from the definition of matrix powers:

$$
\begin{aligned}
A^pA^q &= \underbrace{AA\cdots A}_{p \text{ times}}\underbrace{AA\cdots A}_{q \text{ times}} \\
&\quad \text{(using associativity, Thm. 3.1.24c)} \\
&= \underbrace{AA\cdots A}_{p+q \text{ times}} \\
&= A^{p+q}.
\end{aligned}
$$

□

Example 3.1.25. Show that $(A+B)^2 \neq A^2 + 2AB + B^2$ in general.

Solution: Consider

$$
\begin{aligned}
(A+B)^2 &= (A+B)(A+B) \quad \text{(matrix power)} \\
&= A(A+B) + B(A+B) \quad \text{(Thm. 3.1.24b)}
\end{aligned}
$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$= AA + AB + BA + BB \quad \text{(Thm. 3.1.24a)}$$
$$= A^2 + AB + BA + B^2 \quad \text{(matrix power)}.$$

This expression is only equal to $A^2 + 2AB + B^2$ if we can replace BA by AB. But this requires $BA = AB$ which is generally not true. That is, $(A+B)^2 = A^2 + 2AB + B^2$ only if $BA = AB$.

■

Example 3.1.26. 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).

Solution: Among many other ways to show J is not a multiplicative identity, let's invoke a general 3×3 matrix

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix},$$

and evaluate the product

$$JA = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \cdots = \begin{bmatrix} g & h & i \\ d & e & f \\ a & b & c \end{bmatrix} \neq A.$$

Since $JA \neq A$ then matrix J cannot be a multiplicative identity (the multiplicative identity is only when the ones are along the diagonal from top-left to bottom-right).

■

Theorem 3.1.27 (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 any scalar* c;

The reversed order is important in the identity $(AB)^{\mathrm{T}} = B^{\mathrm{T}}A^{\mathrm{T}}$.

(d) $(AB)^{\mathrm{T}} = B^{\mathrm{T}}A^{\mathrm{T}}$;

(e) $(A^p)^{\mathrm{T}} = (A^{\mathrm{T}})^p$ *for all positive integer exponents* p;[8]

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

[8] With care, this property also holds for all scalar exponents p.

v0.2a

Proof. Let's document a few proofs, others are exercises. Some proofs use primitive definitions—usually using $(\cdot)_{ij}$ to denote the (i,j)th entry of whatever matrix expression is inside the parentheses—others invoke earlier proved parts.

3.1.27b : Recall from Definition 3.1.16 of the transpose that

$$\begin{aligned}((A\pm B)^{\mathrm{T}})_{ij} &= (A\pm B)_{ji} \\ &\qquad \text{(by definition of addition)} \\ &= a_{ji}\pm b_{ji} \\ &\qquad \text{(by Defn. 3.1.16 of transpose)} \\ &= (A^{\mathrm{T}})_{ij}\pm(B^{\mathrm{T}})_{ij}\,.\end{aligned}$$

Since this identity holds for all indices i and j, then $(A\pm B)^{\mathrm{T}} = A^{\mathrm{T}}\pm B^{\mathrm{T}}$.

3.1.27d : The transpose of matrix multiplication is more involved. Let matrices A and B be of sizes $m\times n$ and $n\times p$ respectively. Then from Definition 3.1.16 of the transpose

$$\begin{aligned}&((AB)^{\mathrm{T}})_{ij} = (AB)_{ji} \\ &\qquad \text{(by Defn. 3.1.11 of multiplication)} \\ &= a_{j1}b_{1i}+a_{j2}b_{2i}+\cdots+a_{jn}b_{ni} \\ &\qquad \text{(commuting the products)} \\ &= b_{1i}a_{j1}+b_{2i}a_{j2}+\cdots+b_{ni}a_{jn} \\ &\qquad \text{(by Defn. 3.1.16 of transpose)} \\ &= (B^{\mathrm{T}})_{i1}(A^{\mathrm{T}})_{1j}+(B^{\mathrm{T}})_{i2}(A^{\mathrm{T}})_{2j}+\cdots+(B^{\mathrm{T}})_{in}(A^{\mathrm{T}})_{nj} \\ &\qquad \text{(by Defn. 3.1.11 of multiplication)} \\ &= (B^{\mathrm{T}}A^{\mathrm{T}})_{ij}\,.\end{aligned}$$

Since this identity holds for all indices i and j, then $(AB)^{\mathrm{T}} = B^{\mathrm{T}}A^{\mathrm{T}}$.

3.1.27f : To prove the second, that $A^{\mathrm{T}}A$ is equal to its transpose, we invoke earlier parts. Consider the transpose

$$\begin{aligned}(A^{\mathrm{T}}A)^{\mathrm{T}} &= (A)^{\mathrm{T}}(A^{\mathrm{T}})^{\mathrm{T}} \quad \text{(by 3.1.27d)} \\ &= A^{\mathrm{T}}A \quad \text{(by 3.1.27a).}\end{aligned}$$

Since $A^{\mathrm{T}}A$ equals its transpose, it is symmetric.

□

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

3.1.4 Exercises

Exercise 3.1.1. Consider the following six matrices: $A = \begin{bmatrix} -1 & 3 \\ 0 & -5 \\ 0 & -7 \end{bmatrix}$; $B = \begin{bmatrix} -4 & -3 & -3 & 1 \\ -3 & -2 & 0 & -1 \end{bmatrix}$; $C = \begin{bmatrix} -3 & 1 \end{bmatrix}$; $D = \begin{bmatrix} 0 & 6 & 6 & 3 \\ 2 & 2 & 0 & -5 \end{bmatrix}$; $E = \begin{bmatrix} 0 & 1 & 1 & -2 \\ -1 & 5 & 4 & -1 \\ 1 & -3 & 7 & 3 \\ -6 & -3 & 0 & 2 \end{bmatrix}$; $F = \begin{bmatrix} 4 & 1 & 0 \\ -1 & 1 & 6 \\ -4 & 5 & -2 \end{bmatrix}$.

(a) What is the size of each of these matrices?

(b) Which pairs of matrices may be added or subtracted?

(c) Which matrix multiplications can be performed between two of the matrices?

Exercise 3.1.2. Consider the following six matrices: $A = \begin{bmatrix} 3 & \frac{17}{6} \\ -\frac{5}{3} & \frac{1}{2} \\ -\frac{1}{6} & -\frac{1}{6} \\ \frac{5}{3} & 1 \end{bmatrix}$; $B = \begin{bmatrix} \frac{7}{6} & \frac{1}{3} & \frac{17}{3} \end{bmatrix}$; $C = \begin{bmatrix} -\frac{11}{3} & -\frac{7}{3} \\ \frac{2}{3} & \frac{4}{3} \\ \frac{3}{2} & -\frac{17}{6} \end{bmatrix}$; $D = \begin{bmatrix} 0 & -\frac{13}{6} & 0 & \frac{13}{3} \\ \frac{20}{3} & 2 & -\frac{8}{3} & -\frac{7}{2} \\ \frac{5}{6} & \frac{1}{3} & \frac{13}{6} & -\frac{16}{3} \end{bmatrix}$; $E = \begin{bmatrix} \frac{13}{6} & -\frac{1}{6} \\ -\frac{7}{3} & -5 \end{bmatrix}$; $F = \begin{bmatrix} -\frac{1}{3} \\ \frac{13}{3} \end{bmatrix}$.

(a) What is the size of each of these matrices?

(b) Which pairs of matrices may be added or subtracted?

(c) Which matrix multiplications can be performed between two of the matrices?

Exercise 3.1.3. Given the matrix

$$A = \begin{bmatrix} -0.3 & 2.1 & -4.8 \\ -5.9 & 3.6 & -1.3 \end{bmatrix} :$$

write down its column vectors; what are the values of elements a_{13} and a_{21}?

Exercise 3.1.4. Given the matrix

$$B = \begin{bmatrix} 7.6 & -1.1 & -0.7 & -4.5 \\ -1.1 & -9.3 & 0.1 & 8.2 \\ 2.6 & 6.9 & 1.2 & -3.6 \\ -1.5 & -7.5 & 3.7 & 2.6 \\ -0.2 & 5.5 & -0.9 & 2.4 \end{bmatrix} :$$

write down its column vectors; what are the values of entries b_{13}, b_{31}, b_{42}?

v0.2a

Exercise 3.1.5. Write down the column vectors of the identity I_4. What do we call these column vectors?

Exercise 3.1.6. For the following pairs of matrices, calculate their sum and difference.

(a) $A = \begin{bmatrix} 2 & 1 & -1 \\ -4 & 1 & -3 \\ -2 & 2 & -1 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 1 & 0 \\ 4 & -6 & -6 \\ -6 & 4 & 0 \end{bmatrix}$

(b) $C = \begin{bmatrix} -2 & -2 & -7 \end{bmatrix}$, $D = \begin{bmatrix} 4 & 2 & -2 \end{bmatrix}$

(c) $P = \begin{bmatrix} -2 & 5 & 1 \\ 3 & -3 & 2 \\ -3 & 3 & -3 \end{bmatrix}$, $Q = \begin{bmatrix} -1 & -3 & -1 \\ 6 & -4 & -2 \\ 3 & -3 & 1 \end{bmatrix}$

(d) $R = \begin{bmatrix} -2.5 & -0.4 \\ -1.0 & -3.5 \\ -3.3 & 1.8 \end{bmatrix}$, $S = \begin{bmatrix} -0.9 & 4.9 \\ -1.2 & -0.7 \\ -4.0 & -5.4 \end{bmatrix}$

Exercise 3.1.7. For the given matrix, evaluate the following matrix-scalar products.

(a) $A = \begin{bmatrix} -3 & -2 \\ 4 & -2 \\ 2 & -4 \end{bmatrix}$: $-2A$, $2A$, and $3A$.

(b) $B = \begin{bmatrix} 4 & 0 \\ -1 & -1 \end{bmatrix}$: $1.9B$, $2.6B$, and $-6.9B$.

(c) $U = \begin{bmatrix} -3.9 & -0.3 & -2.9 \\ 3.1 & -3.9 & -1. \\ 3.1 & -6.5 & 0.9 \end{bmatrix}$: $-4U$, $2U$, and $4U$.

(d) $V = \begin{bmatrix} -2.6 & -3.2 \\ 3.3 & -0.8 \\ -0.3 & 0.3 \end{bmatrix}$: $1.3V$, $-3.7V$, and $2.5V$.

Exercise 3.1.8. Use MATLAB/Octave to generate some random matrices of a suitable size of your choice, and some random scalars (see Table 3.1). Then confirm the addition and scalar multiplication properties of Theorem 3.1.22. Record all your commands and the output from MATLAB/Octave.

Exercise 3.1.9. Use the definition of matrix addition and scalar multiplication to prove the basic properties of Theorem 3.1.22.

Exercise 3.1.10. For each of the given matrices, calculate the specifed matrix-vector products.

(a) For $A = \begin{bmatrix} 4 & -3 \\ -2 & 5 \end{bmatrix}$ and vectors $\boldsymbol{p} = \begin{bmatrix} -6 \\ -5 \end{bmatrix}$, $\boldsymbol{q} = \begin{bmatrix} -2 \\ -4 \end{bmatrix}$, and $\boldsymbol{r} = \begin{bmatrix} -3 \\ 1 \end{bmatrix}$, calculate $A\boldsymbol{p}$, $A\boldsymbol{q}$ and $A\boldsymbol{r}$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(b) For $B = \begin{bmatrix} 1 & 6 \\ 4 & -5 \end{bmatrix}$ and vectors $\boldsymbol{p} = \begin{bmatrix} -3 \\ -3 \end{bmatrix}$, $\boldsymbol{q} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, and $\boldsymbol{r} = \begin{bmatrix} -5 \\ 2 \end{bmatrix}$, calculate $B\boldsymbol{p}$, $B\boldsymbol{q}$ and $B\boldsymbol{r}$.

(c) For $C = \begin{bmatrix} -3 & 0 & -3 \\ -1 & -1 & 1 \end{bmatrix}$ and vectors $\boldsymbol{u} = \begin{bmatrix} -4 \\ 3 \\ 2 \end{bmatrix}$, $\boldsymbol{v} = \begin{bmatrix} -3 \\ 1 \\ 2 \end{bmatrix}$, and $\boldsymbol{w} = \begin{bmatrix} -4 \\ 5 \\ -4 \end{bmatrix}$, calculate $C\boldsymbol{u}$, $C\boldsymbol{v}$ and $C\boldsymbol{w}$.

(d) For $D = \begin{bmatrix} 0 & 4 \\ 1 & 2 \\ -1 & 1 \end{bmatrix}$ and vectors $\boldsymbol{u} = \begin{bmatrix} 3 \\ -0.9 \end{bmatrix}$, $\boldsymbol{v} = \begin{bmatrix} 0.9 \\ 6.8 \end{bmatrix}$, and $\boldsymbol{w} = \begin{bmatrix} 0.3 \\ 7.3 \end{bmatrix}$, calculate $D\boldsymbol{u}$, $D\boldsymbol{v}$ and $D\boldsymbol{w}$.

Exercise 3.1.11. For each of the given matrices and vectors, calculate the matrix-vector products. Plot in 2D, and label, the vectors and the specified matrix-vector products.

(a) $A = \begin{bmatrix} 3 & 2 \\ -3 & -1 \end{bmatrix}$, $\boldsymbol{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $\boldsymbol{v} = \begin{bmatrix} 0 \\ -3 \end{bmatrix}$, and $\boldsymbol{w} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$.

(b) $B = \begin{bmatrix} 3 & -2 \\ 3 & 2 \end{bmatrix}$, $\boldsymbol{p} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $\boldsymbol{q} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$, and $\boldsymbol{r} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$.

(c) $C = \begin{bmatrix} -2.1 & 1.1 \\ 4.6 & -1 \end{bmatrix}$, $\boldsymbol{x}_1 = \begin{bmatrix} 2.1 \\ 0 \end{bmatrix}$, $\boldsymbol{x}_2 = \begin{bmatrix} -0.1 \\ 1.1 \end{bmatrix}$, and $\boldsymbol{x}_3 = \begin{bmatrix} -0.3 \\ -1 \end{bmatrix}$.

(d) $D = \begin{bmatrix} 0.1 & 3.4 \\ 3.9 & 5.1 \end{bmatrix}$, $\boldsymbol{a} = \begin{bmatrix} 0.2 \\ 0.5 \end{bmatrix}$, $\boldsymbol{b} = \begin{bmatrix} -0.3 \\ 0.3 \end{bmatrix}$, and $\boldsymbol{c} = \begin{bmatrix} -0.2 \\ -0.6 \end{bmatrix}$.

Exercise 3.1.12. For each of the given matrices and vectors, calculate the matrix-vector products. Plot in 2D, and label, the vectors and the specified matrix-vector products. For each of the matrices, interpret the matrix multiplication of the vectors as either a rotation, a reflection, a stretch, or none of these.

(a) $P = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$, $\boldsymbol{u} = \begin{bmatrix} 1 \\ -1.4 \end{bmatrix}$, $\boldsymbol{v} = \begin{bmatrix} -3.6 \\ -1.7 \end{bmatrix}$, and $\boldsymbol{w} = \begin{bmatrix} 0.1 \\ 2.3 \end{bmatrix}$.

(b) $Q = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$, $\boldsymbol{p} = \begin{bmatrix} 2.1 \\ 1.9 \end{bmatrix}$, $\boldsymbol{q} = \begin{bmatrix} 2.8 \\ -1.1 \end{bmatrix}$, and $\boldsymbol{r} = \begin{bmatrix} 0.8 \\ 3.3 \end{bmatrix}$.

(c) $R = \begin{bmatrix} 0.8 & -0.6 \\ 0.6 & 0.8 \end{bmatrix}$, $\boldsymbol{x}_1 = \begin{bmatrix} -4 \\ 2 \end{bmatrix}$, $\boldsymbol{x}_2 = \begin{bmatrix} 4 \\ -3 \end{bmatrix}$, and $\boldsymbol{x}_3 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$.

(d) $S = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, $\boldsymbol{a} = \begin{bmatrix} -1.1 \\ 0 \end{bmatrix}$, $\boldsymbol{b} = \begin{bmatrix} -4.6 \\ -1.5 \end{bmatrix}$, and $\boldsymbol{c} = \begin{bmatrix} -3.1 \\ 0.9 \end{bmatrix}$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Exercise 3.1.13. Using the matrix-vector products you calculated for Exercise 3.1.10, write down the results of the following matrix-matrix products.

(a) For $A = \begin{bmatrix} 4 & -3 \\ -2 & 5 \end{bmatrix}$, write down the matrix products

i. $A \begin{bmatrix} -6 & -2 \\ -5 & -4 \end{bmatrix}$, ii. $A \begin{bmatrix} -6 & -3 \\ -5 & 1 \end{bmatrix}$,

iii. $A \begin{bmatrix} -2 & -3 \\ -4 & 1 \end{bmatrix}$, iv. $A \begin{bmatrix} -6 & -2 & -3 \\ -5 & -4 & 1 \end{bmatrix}$.

(b) For $B = \begin{bmatrix} 1 & 6 \\ 4 & -5 \end{bmatrix}$, write down the matrix products

i. $B \begin{bmatrix} -3 & 2 \\ -3 & 1 \end{bmatrix}$, ii. $B \begin{bmatrix} -5 & 2 \\ 2 & 1 \end{bmatrix}$,

iii. $B \begin{bmatrix} -5 & -3 \\ 2 & -3 \end{bmatrix}$, iv. $B \begin{bmatrix} -5 & 2 & -3 \\ 2 & 1 & -3 \end{bmatrix}$.

(c) For $C = \begin{bmatrix} -3 & 0 & -3 \\ -1 & -1 & 1 \end{bmatrix}$, write down the matrix products

i. $C \begin{bmatrix} -4 & -3 \\ 3 & 1 \\ 2 & 2 \end{bmatrix}$, ii. $C \begin{bmatrix} -4 & -3 \\ 5 & 1 \\ -4 & 2 \end{bmatrix}$,

iii. $C \begin{bmatrix} -4 & -4 \\ 5 & 3 \\ -4 & 2 \end{bmatrix}$, iv. $C \begin{bmatrix} -4 & -3 & -4 \\ 5 & 1 & 3 \\ -4 & 2 & 2 \end{bmatrix}$.

(d) For $D = \begin{bmatrix} 0 & 4 \\ 1 & 2 \\ -1 & 1 \end{bmatrix}$, write down the matrix products

i. $D \begin{bmatrix} 0.9 & 0.3 \\ 6.8 & 7.3 \end{bmatrix}$, ii. $D \begin{bmatrix} 0.9 & 3 \\ 6.8 & -0.9 \end{bmatrix}$,

iii. $D \begin{bmatrix} 0.3 & 3 \\ 7.3 & -0.9 \end{bmatrix}$, iv. $D \begin{bmatrix} 0.9 & 0.3 & 3 \\ 6.8 & 7.3 & -0.9 \end{bmatrix}$.

Exercise 3.1.14. Use MATLAB/Octave to generate some random matrices of a suitable size of your choice, and some random scalars (see Table 3.1). Choose some suitable exponents. Then confirm the matrix multiplication properties of Theorem 3.1.24. Record all your commands and the output from MATLAB/Octave.

In checking some properties you may get matrices with elements such as `2.2204e-16`: recall from Table 3.1 that this denotes the very small number $2.2204 \cdot 10^{-16}$. When adding and subtracting numbers of size one or so, the result `2.2204e-16` is effectively zero (due to the sixteen digit precision of MATLAB/Octave, Table 1.2).

v0.2

Exercise 3.1.15. Use Definition 3.1.11 of matrix-matrix multiplication to prove multiplication properties of Theorem 3.1.24. Prove parts: 3.1.24b, distributivity; 3.1.24d, scalar associativity; 3.1.24e, identity; 3.1.24f, zeros.

Exercise 3.1.16. Use the other parts of Theorem 3.1.24 to prove part 3.1.24g that $(A^p)^q = A^{pq}$ and $(cA)^p = c^p A^p$ for square matrix A, scalar c, and for positive integer exponents p and q.

Exercise 3.1.17 (Tasmanian Devils)**.** Ecologists studying a colony of Tasmanian Devils, an Australian marsupial, observed the following: two-thirds of the female newborns survive to be one year old; two-thirds of female one year olds survive to be two years old; one-half of female two year olds survive to be three years old; each year, each female aged two or three years gives birth to two female offspring; female Tasmanian Devils survive for four years, at most.

Analogous to Example 3.1.8 define a vector $\boldsymbol{x}$ in $\mathbb{R}^4$ to be the number of females of specified ages. Use the above information to write down the Leslie matrix L that predicts the number in the next year, $\boldsymbol{x}'$, from the number in any year, $\boldsymbol{x}$. Given the observed initial female numbers of 18 newborns, 9 one year olds, 18 two year olds, and 18 three year olds, use matrix multiplication to predict the numbers of female Tasmanian Devils one, two and three years later. Does the population appear to be increasing? or decreasing?

Exercise 3.1.18. Write down the transpose of each of the following matrices. Which of the following matrices are a symmetric matrix?

(a) $\begin{bmatrix} -2 & 3 \\ 3 & 0 \\ -8 & 2 \\ -2 & -4 \end{bmatrix}$

(b) $\begin{bmatrix} 3 & -4 & -2 & 2 \\ -5 & 2 & -3 & 3 \end{bmatrix}$

(c) $\begin{bmatrix} 14 & 5 & 3 & 2 \\ 5 & 0 & -1 & 1 \\ 3 & -1 & -6 & -4 \\ 2 & 1 & -4 & 4 \end{bmatrix}$

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

(e) $\begin{bmatrix} 5 & -1 & -2 & 2 \\ 1 & -2 & -2 & 0 \\ -1 & -5 & 4 & -1 \\ 5 & 2 & -1 & -2 \end{bmatrix}$

(f) $\begin{bmatrix} -4 & -5.1 & 0.3 \\ -5.1 & -7.4 & -3. \\ 0.3 & -3 & 2.6 \end{bmatrix}$

(g) $\begin{bmatrix} -1.5 & -0.6 & -1.7 \\ -1 & -0.4 & -5.6 \end{bmatrix}$

(h) $\begin{bmatrix} 1.7 & -0.2 & -0.4 \\ 0.7 & -0.3 & -0.4 \\ 0.6 & 3 & -2.2 \end{bmatrix}$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Exercise 3.1.19. Are the following matrices symmetric? I_4, $I_{3\times 4}$, O_3, and $O_{3\times 1}$.

Exercise 3.1.20. Use MATLAB/Octave to generate some random matrices of a suitable size of your choice, and some random scalars (see Table 3.1). Choose some suitable exponents. Recalling that in MATLAB/Octave the dash ' performs the transpose, confirm the matrix transpose properties of Theorem 3.1.27. Record all your commands and the output from MATLAB/Octave.

Exercise 3.1.21. Use Definition 3.1.16 of the matrix transpose to prove properties 3.1.27a and 3.1.27c of Theorem 3.1.27.

Exercise 3.1.22. Use the other parts of Theorem 3.1.27 to prove parts 3.1.27e and 3.1.27f.

v0.2a

3.2 The inverse of a matrix

Section Contents

3.2.1 Introducing the unique inverse 177

3.2.2 Diagonal matrices stretch and shrink 188

Solve systems whose matrix is diagonal . . . 189

But do not divide by zero 192

Stretch or squash the unit square 193

Sketch convenient coordinates 195

3.2.3 Orthogonal matrices rotate 197

Orthogonal set of vectors 197

Orthogonal matrices 199

3.2.4 Exercises 206

The previous Section 3.1 introduced addition, subtraction, multiplication, and other operations of matrices. Conspicuously missing from the list is 'division' by a matrix: this section addresses 'division' by a matrix as multiplication by the inverse of a matrix. The analogue in ordinary arithmetic is that division by ten is the same as multiplying by its reciprocal, one-tenth. But the inverse of a matrix looks nothing like a reciprocal.

3.2.1 Introducing the unique inverse

Let's start with an example that illustrates an analogy with the reciprocal/inverse of a scalar number.

Example 3.2.1. Recall that a crucial property is that a number multiplied by its reciprocal/inverse is one: for example, $2 \times 0.5 = 1$ so 0.5 is the reciprocal/inverse of 2. Similarly, show that matrix

$$B = \begin{bmatrix} -3 & 1 \\ -4 & 1 \end{bmatrix} \text{ is an inverse of } A = \begin{bmatrix} 1 & -1 \\ 4 & -3 \end{bmatrix}$$

by showing their product is the 2×2 identity matrix I_2.

Solution: Multiply

$$AB = \begin{bmatrix} 1 & -1 \\ 4 & -3 \end{bmatrix} \begin{bmatrix} -3 & 1 \\ -4 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2$$

the multiplicative identity. But matrix multiplication is not commutative (section 3.1.3), so also consider

$$BA = \begin{bmatrix} -3 & 1 \\ -4 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 4 & -3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2 \,.$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

That these products are the identity—analogous to the number one in scalar arithmetic—means that the matrix A has the same relation to the matrix B as a number has to its reciprocal/inverse.

Being the inverse, matrix B 'undoes' the action of matrix A—as illustrated in the margin. The first picture shows multiplication by A transforms the vector $(2,1)$ to the vector $(1,5)$: $A\begin{bmatrix}2\\1\end{bmatrix}=\begin{bmatrix}1\\5\end{bmatrix}$. The second picture shows that multiplication by B undoes the transform by A because $B\begin{bmatrix}1\\5\end{bmatrix}=\begin{bmatrix}2\\1\end{bmatrix}$ the original vector.

■

The previous example 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 because we need to be able to compute both AB and BA.

Definition 3.2.2 (inverse). *If A is an $n\times n$ square matrix, then 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 A is called **invertible**.*

Example 3.2.3. Show that matrix

$$B=\begin{bmatrix}0&-\frac14&-\frac18\\\frac32&1&\frac78\\\frac12&\frac14&\frac38\end{bmatrix}\text{ is an inverse of }A=\begin{bmatrix}1&-1&5\\-5&-1&3\\2&2&-6\end{bmatrix}.$$

Solution: First compute

$$\begin{aligned}AB&=\begin{bmatrix}1&-1&5\\-5&-1&3\\2&2&-6\end{bmatrix}\begin{bmatrix}0&-\frac14&-\frac18\\\frac32&1&\frac78\\\frac12&\frac14&\frac38\end{bmatrix}\\&=\begin{bmatrix}1\cdot0-1\cdot\frac32+5\cdot\frac12&1\cdot(-\frac14)-1\cdot1+5\cdot\frac14&1\cdot(-\frac18)-1\cdot\frac78+5\cdot\frac38\\-5\cdot0-1\cdot\frac32+3\cdot\frac12&-5\cdot(-\frac14)-1\cdot1+3\cdot\frac14&-5\cdot(-\frac18)-1\cdot\frac78+3\cdot\frac38\\2\cdot0+2\cdot\frac32-6\cdot\frac12&2\cdot(-\frac14)+2\cdot1-6\cdot\frac14&2\cdot(-\frac18)+2\cdot\frac78-6\cdot\frac38\end{bmatrix}\\&=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}=I_3\,.\end{aligned}$$

Second compute

$$BA=\begin{bmatrix}0&-\frac14&-\frac18\\\frac32&1&\frac78\\\frac12&\frac14&\frac38\end{bmatrix}\begin{bmatrix}1&-1&5\\-5&-1&3\\2&2&-6\end{bmatrix}$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$= \begin{bmatrix} 0\cdot 1 - \frac{1}{4}\cdot(-5) - \frac{1}{8}\cdot 2 & 0\cdot(-1) - \frac{1}{4}\cdot(-1) - \frac{1}{8}\cdot 2 & 0\cdot 5 - \frac{1}{4}\cdot 3 - \frac{1}{8}\cdot(-6) \\ \frac{3}{2}\cdot 1 + 1\cdot(-5) + \frac{7}{8}\cdot 2 & \frac{3}{2}\cdot(-1) + 1\cdot(-1) + \frac{7}{8}\cdot 2 & \frac{3}{2}\cdot 5 + 1\cdot 3 + \frac{7}{8}\cdot(-6) \\ \frac{1}{2}\cdot 1 + \frac{1}{4}\cdot(-5) + \frac{3}{8}\cdot 2 & \frac{1}{2}\cdot(-1) + \frac{1}{4}\cdot(-1) + \frac{3}{8}\cdot 2 & \frac{1}{2}\cdot 5 + \frac{1}{4}\cdot 3 + \frac{3}{8}\cdot(-6) \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I_3 \,.$$

Since both of these products are the identity, then matrix A is invertible, and B is an inverse of A.

■

Activity 3.2.4. What value of b makes matrix $\begin{bmatrix} -1 & b \\ 1 & 2 \end{bmatrix}$ the inverse of $\begin{bmatrix} 2 & 3 \\ -1 & -1 \end{bmatrix}$?

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

■

But even among square matrices, there are many non-zero matrices which do not have inverses! The next Section 3.3 further explores why some matrices do not have an inverse: the reason is associated with both `rcond` $= 0$ (Procedure 2.2.5) and 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.

Solution: Assume there is an inverse matrix

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

Then by Definition 3.2.2 the product $AB = I_2$; that is,

$$\begin{aligned} AB &= \begin{bmatrix} 1 & -2 \\ -3 & 6 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} \\ &= \begin{bmatrix} a - 2c & b - 2d \\ -3a + 6c & -3b + 6d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \end{aligned}$$

The bottom-left entry in this matrix equality asserts $-3a + 6c = 0$ which is $-3(a - 2c) = 0$, that is, $a - 2c = 0$. But the top-left entry in the matrix equality asserts $a - 2c = 1$. Both of these equations involving a and c cannot be true simultaneously; therefore the assumption of an inverse must be incorrect. The matrix A does not have an inverse.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

■

Theorem 3.2.6 (unique inverse). *If A is an invertible matrix, then its inverse is unique (and denoted by A^{-1}).*

Proof. We suppose there are two inverses, say B_1 and B_2, and proceed to show they must be the same. Since they are inverses, by Definition 3.2.2 both $AB_1 = B_1A = I_n$ and $AB_2 = B_2A = I_n$. Consequently, using associativity of matrix multiplication (Theorem 3.1.24c),

$$B_1 = B_1I_n = B_1(AB_2) = (B_1A)B_2 = I_nB_2 = B_2\,.$$

That is, $B_1 = B_2$ and so the inverse is unique. □

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 non-zero. The reason is that $AA^{-1} = \left[a_{11} \cdot \frac{1}{a_{11}}\right] = \begin{bmatrix} 1 \end{bmatrix} = I_1$ and $A^{-1}A = \left[\frac{1}{a_{11}} \cdot a_{11}\right] = \begin{bmatrix} 1 \end{bmatrix} = I_1$.

In the case of 2×2 matrices the inverse is a little more complicated, but should be remembered.

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 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.

Example 3.2.8. (a) Recall Example 3.2.1 verified that

$$B = \begin{bmatrix} -3 & 1 \\ -4 & 1 \end{bmatrix} \text{ is an inverse of } A = \begin{bmatrix} 1 & -1 \\ 4 & -3 \end{bmatrix}.$$

Formula (3.2) gives this inverse from the matrix A: its elements are $a = 1$, $b = -1$, $c = 4$ and $d = -3$ so the determinant $ad - bc = 1 \cdot (-3) - (-1) \cdot 4 = 1$ and hence formula (3.2) derives the inverse

$$A^{-1} = \frac{1}{1} \begin{bmatrix} -3 & -(-1) \\ -4 & 1 \end{bmatrix} = \begin{bmatrix} -3 & 1 \\ -4 & 1 \end{bmatrix} = B\,.$$

(b) Further, recall Example 3.2.5 proved there is no inverse for matrix

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

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Theorem 3.2.7 also establishes this matrix is not invertible because the matrix determinant $ad-bc = 1\cdot 6-(-2)\cdot(-3) = 6-6=0$.

■

Activity 3.2.9. Which of the following matrices is invertible?

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

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

■

Proof. To prove Theorem 3.2.7, first show the given A^{-1} satisfies Definition 3.2.2 when the determinant $ad-bc \neq 0$ (and using associativity of scalar-matrix multiplication, Theorem 3.1.24d). For the proposed A^{-1}, on the one hand,

$$\begin{aligned} A^{-1}A &= \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} \\ &= \frac{1}{ad-bc}\begin{bmatrix} da-bc & db-bd \\ -ca+ac & -cb+ad \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2 . \end{aligned}$$

On the other hand,

$$\begin{aligned} AA^{-1} &= \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}\frac{1}{ad-bc} \\ &= \begin{bmatrix} ad-bc & -ab+ba \\ cd-dc & -cb+da \end{bmatrix}\frac{1}{ad-bc} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2 . \end{aligned}$$

By uniqueness (Theorem 3.2.6), equation (3.2) is the only inverse when $ad-bc \neq 0$.

Now eliminate the case when $ad-bc=0$. If an inverse exists, say X, then it must satisfy $AX = I_2$. The top-left entry of this matrix equality requires $ax_{11}+bx_{21}=1$, whereas the bottom-left equality requires $cx_{11}+dx_{21}=0$. Regard these as a system of two linear equations for the as yet unknowns x_{11} and x_{21}: from $d\times$ the first subtract $b\times$ the second to deduce that an inverse requires $dax_{11}+dbx_{21}-bcx_{11}-bdx_{21} = d\cdot 1 - b\cdot 0$. By cancellation and factorising x_{11}, this equation then requires $(ad-bc)x_{11}=d$. But the determinant is zero, so this equation requires $0\cdot x_{11}=d$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

- If element d is non-zero, then this equation cannot be satisfied and hence no inverse can be found.
- Otherwise, if d is zero, then any x_{11} satisfies this equation so if there is an inverse then there would be an infinite number of inverses (through the free variable x_{11}). But this contradicts the uniqueness Theorem 3.2.6 so an inverse cannot exist in this case either.

That is, if the determinant $ad - bc = 0$, then the 2×2 matrix is not invertible. □

> Almost anything you can do with A^{-1} can be done without it. *G. E. Forsythe and C. B. Moler, 1967 (Higham 1996, p.261)*

Computer considerations Except for easy cases such as 2×2 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, rarely a practical computational tool.

The following Theorem 3.2.10 is an example: for a system of linear equations the theorem connects the existence of a unique solution to the invertibility of the matrix of coefficients. Further, 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 generally compute the solution without ever constructing A^{-1}.

Theorem 3.2.10. *If A is an invertible $n \times n$ matrix, then the system of linear equations $A\boldsymbol{x} = \boldsymbol{b}$ has the unique solution $\boldsymbol{x} = A^{-1}\boldsymbol{b}$ for any $\boldsymbol{b}$ in $\mathbb{R}^n$.*

One consequence is the following: if a system of linear equations has no solution or an infinite number of solutions (Theorem 2.2.27), then this theorem establishes that the matrix of the system is not invertible.

Proof. The proof has two parts: first showing $\boldsymbol{x} = A^{-1}\boldsymbol{b}$ is a solution, and second showing that there are no others. First, try $\boldsymbol{x} = A^{-1}\boldsymbol{b}$ and use associativity (Theorem 3.1.24c) and the inverse Definition 3.2.2:

$$A\boldsymbol{x} = A(A^{-1}\boldsymbol{b}) = (AA^{-1})\boldsymbol{b} = I_n\boldsymbol{b} = \boldsymbol{b}\,.$$

Second, suppose $\boldsymbol{y}$ is any solution, that is, $A\boldsymbol{y} = \boldsymbol{b}$. Multiply both sides by the inverse A^{-1}, and again use associativity and the definition of the inverse, to deduce

$$A^{-1}(A\boldsymbol{y}) = A^{-1}\boldsymbol{b} \implies (A^{-1}A)\boldsymbol{y} = \boldsymbol{x}$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$\implies I_n \boldsymbol{y} = \boldsymbol{x}$$
$$\implies \boldsymbol{y} = \boldsymbol{x}\,.$$

Since any solution $\boldsymbol{y}$ has to be the same as $\boldsymbol{x}$, $\boldsymbol{x} = A^{-1}\boldsymbol{b}$ is the unique solution. □

Example 3.2.11. Use the matrices of Examples 3.2.1, 3.2.25 and 3.2.5 to decide whether 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) $\begin{cases} r - 2s = -1\,, \\ -3r + 6s = 3\,. \end{cases}$

Solution: (a) A matrix for this system is $\begin{bmatrix} 1 & -1 \\ 4 & -3 \end{bmatrix}$ which Example 3.2.1 shows has an inverse. Theorem 3.2.10 then assures us the system has a unique solution.

(b) A matrix for this system is $\begin{bmatrix} 1 & -1 & 5 \\ -5 & -1 & 3 \\ 2 & 2 & -6 \end{bmatrix}$ which Example 3.2.25 shows has an inverse. Theorem 3.2.10 then assures us the system has a unique solution.

(c) A matrix for this system is $\begin{bmatrix} 1 & -2 \\ -3 & 6 \end{bmatrix}$ which Example 3.2.5 shows is not invertible. Theorem 3.2.10 then assures us the system does not have a unique solution. By Theorem 2.2.27 there may be either no solution or an infinite number of solutions—the matrix alone does not tell us which.

■

Example 3.2.12. Given the following information about solutions of systems of linear equations, write down if the matrix associated with each system is invertible, or not, or there is not enough given information to decide. Give reasons.

(a) A general solution is $(1\,, -5\,, 0\,, 3)$.

(b) A general solution is $(3\,, -5 + 3t\,, 3 - t\,, -1)$.

(c) A solution of a system is $(-3/2\,, -2\,, -\pi\,, 2\,, -4)$.

(d) A solution of a homogeneous system is $(1\,, 2\,, -8)$.

Solution: (a) Since the solution is unique, the matrix in the system must be invertible.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(b) This solution has an apparent free parameter, t, and so there are many solutions which implies the matrix is not invertible.

(c) Not enough information is given as we do not know whether there are any more solutions.

(d) Since a homogeneous system always has $\mathbf{0}$ as a solution (section 2.2.3), then we know that there are at least two solutions to the system, and hence the matrix is not invertible.

■

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:*

The reversed order is important in the identity $(AB)^{-1} = B^{-1}A^{-1}$.

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

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

(e) matrices A^p are invertible for all $p = 1, 2, 3, \dots$ and $(A^p)^{-1} = (A^{-1})^p$.

Proof. Three parts are proved, and two are left as exercises.

3.2.13a : By Definition 3.2.2 the matrix A^{-1} satisfies $A^{-1}A = AA^{-1} = I$. But also by Definition 3.2.2 this is exactly the identities we need to assert that matrix A is the inverse of matrix (A^{-1}). Hence $A = (A^{-1})^{-1}$.

3.2.13c : Test that $B^{-1}A^{-1}$ has the required properties for the inverse of AB. First, by associativity (Theorem 3.1.24c) and multiplication by the identity (Theorem 3.1.24e)

$$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}IB = B^{-1}B = I\,.$$

Second, and similarly

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

Hence by Definition 3.2.2 and the uniqueness Theorem 3.2.6, matrix AB is invertible and $B^{-1}A^{-1}$ is the inverse.

3.2.13e : Prove by induction and use 3.2.13c.

- For the case of exponent $p = 1$, $(A^1)^{-1} = (A)^{-1} = A^{-1} = (A^{-1})^1$ and so the identity holds.

- For any integer exponent $p \geq 2$, assume the identity $(A^{p-1})^{-1} = (A^{-1})^{p-1}$. Consider

$$\begin{aligned}(A^p)^{-1} &= (AA^{p-1})^{-1} \quad \text{(by power law Thm. 3.1.24g)}\\ &= (A^{p-1})^{-1}A^{-1} \quad \text{(by Thm. 3.2.13c, } B = A^{p-1})\\ &= (A^{-1})^{p-1}A^{-1} \quad \text{(by inductive assumption)}\\ &= (A^{-1})^p \quad \text{(by power law Thm. 3.1.24g).}\end{aligned}$$

- By induction, the identity $(A^p)^{-1} = (A^{-1})^p$ holds for exponents $p = 1, 2, 3, \ldots$.

□

Activity 3.2.14. Given the matrix $\begin{bmatrix}3 & -5\\4 & -7\end{bmatrix}$ has inverse $\begin{bmatrix}7 & -5\\4 & -3\end{bmatrix}$, what is the inverse of matrix $\begin{bmatrix}6 & -10\\8 & -14\end{bmatrix}$?

(a) $\begin{bmatrix}14 & -10\\8 & -3\end{bmatrix}$ (b) $\begin{bmatrix}3.5 & -2.5\\2 & -1.5\end{bmatrix}$

(c) $\begin{bmatrix}3.5 & 2\\-2.5 & -1.5\end{bmatrix}$ (d) $\begin{bmatrix}7 & 4\\-5 & -3\end{bmatrix}$

Further, which of the above is the inverse of $\begin{bmatrix}3 & 4\\-5 & -7\end{bmatrix}$? ■

Definition 3.2.15 (non-positive powers)**.** *If A is an invertible matrix, then define $A^0 = I$ and for any positive integer p define $A^{-p} = (A^{-1})^p$ (or by Theorem 3.2.13e equivalently as $(A^p)^{-1}$).*

Example 3.2.16. Recall from Example 3.2.1 that matrix

$$A = \begin{bmatrix}1 & -1\\4 & -3\end{bmatrix} \text{ has inverse } A^{-1} = \begin{bmatrix}-3 & 1\\-4 & 1\end{bmatrix}.$$

Compute A^{-2} and A^{-4}.

Solution: From Definition 3.2.15,

$$\begin{aligned}A^{-2} &= (A^{-1})^2 = \begin{bmatrix}-3 & 1\\-4 & 1\end{bmatrix}\begin{bmatrix}-3 & 1\\-4 & 1\end{bmatrix} = \begin{bmatrix}5 & -2\\8 & -3\end{bmatrix},\\ A^{-4} &= (A^{-1})^4 = \left[(A^{-1})^2\right]^2\\ &= \begin{bmatrix}5 & -2\\8 & -3\end{bmatrix}\begin{bmatrix}5 & -2\\8 & -3\end{bmatrix} = \begin{bmatrix}9 & -4\\16 & -7\end{bmatrix},\end{aligned}$$

upon using one of the power laws of Theorem 3.1.24g.

■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Activity 3.2.17. Example 3.2.16 gives the inverse of a matrix A and determines A^{-2}: what is A^{-3}?

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

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

■

Example 3.2.18 (predict the past)**.** Recall Example 3.1.8 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, for the Leslie matrix

$$L = \begin{bmatrix} 0 & 0 & 4 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} \end{bmatrix}, \quad \text{which has inverse } L^{-1} = \begin{bmatrix} 0 & 2 & 0 \\ -\frac{1}{4} & 0 & 3 \\ \frac{1}{4} & 0 & 0 \end{bmatrix},$$

by the modelling the predicted 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. Assume the same rule applies for earlier years.

- Letting the population numbers a year ago be denoted by $\boldsymbol{x}^-$ then by the modelling 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 by the modelling $\boldsymbol{x}^- = L\boldsymbol{x}^=$ and 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 two years ago be denoted by $\boldsymbol{x}^\equiv$ then by the modelling $\boldsymbol{x}^= = L\boldsymbol{x}^\equiv$ and 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.

Solution: Verify the given inverse by evaluating (showing only non-zero terms in a sum)

$$\begin{aligned} LL^{-1} &= \begin{bmatrix} 0 & 0 & 4 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 0 & 2 & 0 \\ -\frac{1}{4} & 0 & 3 \\ \frac{1}{4} & 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 4 \cdot \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{2} \cdot 2 & 0 \\ \frac{1}{3} \cdot (-\frac{1}{4}) + \frac{1}{3} \cdot \frac{1}{4} & 0 & \frac{1}{3} \cdot 3 \end{bmatrix} \end{aligned}$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I_3 \,,$$

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

$$= \begin{bmatrix} 2\frac{1}{2} & 0 & 0 \\ 0 & 3\cdot\frac{1}{3} & -\frac{1}{4}\cdot 4 + 3\cdot\frac{1}{3} \\ 0 & 0 & \frac{1}{4}\cdot 4 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I_3 \,.$$

Hence the given L^{-1} is indeed the inverse. For the current population $\boldsymbol{x} = (60, 70, 20)$, now use the inverse to compute earlier populations.

- The population of females one year ago was

$$\boldsymbol{x}^{-} = L^{-1}\boldsymbol{x} = \begin{bmatrix} 0 & 2 & 0 \\ -\frac{1}{4} & 0 & 3 \\ \frac{1}{4} & 0 & 0 \end{bmatrix} \begin{bmatrix} 60 \\ 70 \\ 20 \end{bmatrix} = \begin{bmatrix} 140 \\ 45 \\ 15 \end{bmatrix} .$$

  That is, there were 140 pups, 45 juveniles, and 15 mature females.

- Computing the square of the inverse

$$L^{-2} = (L^{-1})^2 = \begin{bmatrix} 0 & 2 & 0 \\ -\frac{1}{4} & 0 & 3 \\ \frac{1}{4} & 0 & 0 \end{bmatrix}^2 = \begin{bmatrix} -\frac{1}{2} & 0 & 6 \\ \frac{3}{4} & -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 \end{bmatrix} ,$$

  we predict the population of females two years ago was

$$\boldsymbol{x}^{=} = L^{-2}\boldsymbol{x} = \begin{bmatrix} -\frac{1}{2} & 0 & 6 \\ \frac{3}{4} & -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 \end{bmatrix} \begin{bmatrix} 60 \\ 70 \\ 20 \end{bmatrix} = \begin{bmatrix} 90 \\ 10 \\ 35 \end{bmatrix} .$$

- Similarly, computing the cube of the inverse

$$L^{-3} = L^{-2}L^{-1} = \cdots = \begin{bmatrix} \frac{3}{2} & -1 & 0 \\ \frac{1}{8} & \frac{3}{2} & -\frac{3}{2} \\ -\frac{1}{8} & 0 & \frac{3}{2} \end{bmatrix} ,$$

  we predict the population of females three years ago was

$$\boldsymbol{x}^{\equiv} = L^{-3}\boldsymbol{x} = \begin{bmatrix} \frac{3}{2} & -1 & 0 \\ \frac{1}{8} & \frac{3}{2} & -\frac{3}{2} \\ -\frac{1}{8} & 0 & \frac{3}{2} \end{bmatrix} \begin{bmatrix} 60 \\ 70 \\ 20 \end{bmatrix} = \begin{bmatrix} 20 \\ 82.5 \\ 22.5 \end{bmatrix} .$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(Predicting half animals in this last calculation is because the modelling only deals with average numbers, not exact numbers.)

■

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.

Solution: Execute

```
L=[0 0 4;1/2 0 0;0 1/3 1/3]
x=[60;70;20]
rcond(L)
xm=L\x
xmm=L\xm
xmmm=L\xmm
```

Since `rcond` is 0.08 (good), this code uses `L\` to solve the linear systems and confirm the population of females in previous years is as determined by Example 3.2.18, namely

$$\texttt{xm} = \begin{bmatrix}140\\45\\15\end{bmatrix}, \quad \texttt{xmm} = \begin{bmatrix}90\\10\\35\end{bmatrix}, \quad \texttt{xmmm} = \begin{bmatrix}20\\82.5\\22.5\end{bmatrix}.$$

■

3.2.2 Diagonal matrices stretch and shrink

Recall that the identity matrices are zero except for a diagonal of ones from the top-left to the bottom-right of the matrix. Because of the nature of matrix multiplication it is this diagonal that is special. Because of the special nature of this diagonal, this section explores matrices which are zero except for the numbers (not generally ones) in the top-left to bottom-right diagonal.

Example 3.2.20. That is, this section explores the nature of so-called diagonal matrices such as

$$\begin{bmatrix}3&0\\0&2\end{bmatrix}, \quad \begin{bmatrix}0.58&0&0\\0&-1.61&0\\0&0&2.17\end{bmatrix}, \quad \begin{bmatrix}\pi&0&0\\0&\sqrt{3}&0\\0&0&0\end{bmatrix}.$$

We use the term diagonal matrix to also include non-square matrices such as

$$\begin{bmatrix}-\sqrt{2}&0\\0&\frac{1}{2}\\0&0\end{bmatrix}, \quad \begin{bmatrix}1&0&0&0&0\\0&\pi&0&0&0\\0&0&e&0&0\end{bmatrix}.$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

The term diagonal matrix does *not* apply to say

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 2 & 0 \\ -\frac{1}{2} & 0 & 0 \end{bmatrix}, \quad \text{and } not \begin{bmatrix} -0.17 & 0 & 0 & 0 \\ 0 & -4.22 & 0 & 0 \\ 0 & 0 & 0 & 3.05 \end{bmatrix}.$$

■

Amazingly, the singular value decomposition of Section 3.3 proves that diagonal matrices lie at the very heart of the action of *all* matrices.

Definition 3.2.21 (diagonal matrix)**.** *Given an $m \times n$ matrix A, the **diagonal entries** of A are $a_{11}, a_{22}, \dots, a_{pp}$ where $p = \min(m, n)$. A matrix whose non-diagonal entries are all zero is called a **diagonal matrix**.*

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

Example 3.2.22. The five diagonal matrices of Example 3.2.20 could equivalently be written as $\operatorname{diag}(3\,,\,2)$, $\operatorname{diag}(0.58\,,\,-1.61\,,\,2.17)$ $\operatorname{diag}(\pi\,,\sqrt{3}\,,0)$, $\operatorname{diag}_{3\times 2}(-\sqrt{2}\,,\frac{1}{2})$ and $\operatorname{diag}_{3\times 5}(1\,,\pi\,,e)$, respectively. ■

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

Activity 3.2.23. Which of the following matrices are not diagonal?

(a) I_n

(b) O_n

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

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

■

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).

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Table 3.2: As well as the basics of MATLAB/Octave listed in Tables 1.2, 2.3 and 3.1, we need these matrix operations.

- `diag(v)` where `v` is a row/column vector of length p generates the $p \times p$ matrix

$$\operatorname{diag}(v_1\,,v_2\,,\dots,v_p) = \begin{bmatrix} v_1 & 0 & \cdots & 0 \\ 0 & v_2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & v_p \end{bmatrix}.$$

- In MATLAB/Octave (but not usually in algebra), `diag` also does the opposite: for an $m\times n$ matrix A such that both $m,n \geq 2$, `diag(A)` returns the (column) vector $(a_{11}\,,a_{22}\,,\dots,a_{pp})$ of diagonal entries where the result vector length $p = \min(m\,,n)$.
- The dot operators `./` and `.*` do element-by-element division and multiplication of two matrices/vectors of the same size. For example,
  `[5 14 33]./[5 7 3]=[1 2 11]`
- Section 3.5 also needs to compute the logarithm of data: `log10(v)` finds the logarithm to base 10 of each component of `v` and returns the results in a vector of the same size; `log(v)` does the same but for the natural logarithm (not `ln(v)`).

Example 3.2.24. Solve

$$\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 \\ -5 \end{bmatrix}$$

Solution: Algebraically this matrix-vector equation means

$$\begin{cases} 3x_1 + 0x_2 = 2 \\ 0x_1 + 2x_2 = -5 \end{cases} \iff \begin{cases} 3x_1 = 2 \\ 2x_2 = -5 \end{cases} \iff \begin{cases} x_1 = 2/3 \\ x_2 = -5/2 \end{cases}$$

The solution is $\boldsymbol{x} = (2/3, -5/2)$. Interestingly, the two components of this solution are firstly the 2 on the right-hans side divided by the 3 in the matrix, and secondly the -5 on the right-hand side divided by the 2 in the matrix.

■

Example 3.2.25. Solve

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{2}{3} & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

Solution: Algebraically this equation means

$$\begin{cases} 2x_1 + 0x_2 + 0x_3 = b_1 \\ 0x_1 + \frac{2}{3}x_2 + 0x_3 = b_2 \\ 0x_1 + 0x_2 - 1x_3 = b_3 \end{cases} \iff \begin{cases} 2x_1 = b_1 \\ \frac{2}{3}x_2 = b_2 \\ -x_3 = b_3 \end{cases} \iff \begin{cases} x_1 = \frac{1}{2}b_1 \\ x_2 = \frac{3}{2}b_2 \\ x_3 = -b_3 \end{cases}.$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

The solution is

$$\boldsymbol{x} = \begin{bmatrix} \frac{1}{2}b_1 \\ \frac{3}{2}b_2 \\ -b_3 \end{bmatrix}, \quad \text{rewritten as } \boldsymbol{x} = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{3}{2} & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}.$$

Consequently, by its uniqueness (Theorem 3.2.6), the inverse of the given diagonal matrix must be

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

which interestingly is the diagonal of reciprocals of the given matrix.

■

Activity 3.2.26. What is the solution to the system $\begin{bmatrix} 0.4 & 0 \\ 0 & 0.1 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 0.1 \\ -0.2 \end{bmatrix}$?

(a) $(4\,,-2)$ (b) $(\frac{1}{4}\,,-\frac{1}{2})$ (c) $(4\,,-\frac{1}{2})$ (d) $(\frac{1}{4}\,,-2)$

■

Theorem 3.2.27 (inverse of diagonal matrix)**.** *Let D be the $n \times n$ diagonal matrix $D = \operatorname{diag}(d_1, d_2, \dots, d_n)$. If all the diagonal entries are nonzero, $d_i \neq 0$ for $i = 1, 2, \dots, n$, then D is invertible and the inverse $D^{-1} = \operatorname{diag}(1/d_1, 1/d_2, \dots, 1/d_n)$.*

Proof. Consider the matrix product

$$\begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix} \begin{bmatrix} \frac{1}{d_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{d_2} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{d_n} \end{bmatrix}$$

$$= \begin{bmatrix} d_1\frac{1}{d_1} + 0 + \cdots + 0 & d_1 0 + 0\frac{1}{d_1} + \cdots + 0 & \cdots & d_1 0 + 0 + \cdots + 0\frac{1}{d_n} \\ 0\frac{1}{d_1} + d_2 0 + \cdots + 0 & 0 + d_2\frac{1}{d_2} + \cdots + 0 & & 0 + d_2 0 + \cdots + 0\frac{1}{d_n} \\ \vdots & & \ddots & \vdots \\ 0\frac{1}{d_1} + 0 + \cdots + d_n 0 & 0 + 0\frac{1}{d_2} + \cdots + d_n 0 & \cdots & 0 + 0 + \cdots + d_n\frac{1}{d_n} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = I_n.$$

Similarly for the reverse product. By Definition 3.2.2, D is invertible with the given inverse. □

v0.2a

Example 3.2.28. The previous Example 3.2.25 gave 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 $D^{-1} = \operatorname{diag}(\frac{1}{3}, \frac{1}{2}) = \begin{bmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$. Then the solution to

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

■

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

$$\begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} d_1 x_1 \\ d_2 x_2 \\ \vdots \\ d_n x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

$$\iff \begin{cases} d_1 x_1 = b_1 \\ d_2 x_2 = b_2 \\ \vdots \\ d_n x_n = b_n \end{cases} \iff \begin{cases} x_1 = b_1/d_1 \\ x_2 = b_2/d_2 \\ \vdots \\ x_n = b_n/d_n \end{cases} \tag{3.3}$$

- 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 aways 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 for any others that are zero). From (3.3) we need to solve $d_n x_n = b_n$, which here is $0 \cdot x_n = b_n$, that is, $0 = b_n$. There are two cases:

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

- if $b_n \neq 0$, then there is no solution; conversely
- if $b_n = 0$, then there is an infinite number of solutions as any x_n satisfies $0 \cdot x_n = 0$.

Example 3.2.29. Solve the two systems (the only difference is the last component on the RHS)

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

Solution: Algebraically, the first system means

$$\begin{array}{l} 2x_1 + 0x_2 + 0x_3 = 1 \\ 0x_1 + \frac{2}{3}x_2 + 0x_3 = 2 \\ 0x_1 + 0x_2 + 0x_3 = 3 \end{array} \iff \begin{array}{l} 2x_1 = 1 \\ \frac{2}{3}x_2 = 2 \\ 0x_3 = 3 \end{array}.$$

There is no solution in the first case as there is no choice of x_3 such that $0x_3 = 3$.

Algebraically, the second system means

$$\begin{array}{l} 2x_1 + 0x_2 + 0x_3 = 1 \\ 0x_1 + \frac{2}{3}x_2 + 0x_3 = 2 \\ 0x_1 + 0x_2 + 0x_3 = 0 \end{array} \iff \begin{array}{l} 2x_1 = 1 \\ \frac{2}{3}x_2 = 2 \\ 0x_3 = 0 \end{array}.$$

In this second case we satisfy the equation $0x_3 = 0$ with any x_3. Hence there are an infinite number of solutions, namely $\boldsymbol{x} = (\frac{1}{2}, 3, t)$ for all t—a free variable just as in Gauss–Jordan elimination (Procedure 2.2.24).

■

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.

v0.2a

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

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

Example 3.2.31. Consider $\operatorname{diag}(2, \frac{2}{3}, -1) = \begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{2}{3} & 0 \\ 0 & 0 & -1 \end{bmatrix}$: the stereo pair below illustrates how this diagonal matrix stretches in one direction, squashes in another, and reflects in the vertical. By multiplying the matrix by corner vectors $(1,0,0)$, $(0,1,0)$, $(0,0,1)$, and so on, we see that the blue unit cube (with 'roof' and 'door') maps to the red.

■

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

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

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

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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

Example 3.2.32. What diagonal matrix transforms the blue unit square to the red in the illustration in the margin?

Solution: In the illustration, the horizontal is stretched by a factor of 3/2, whereas the vertical is squashed by a factor of 1/2, and reflected (minus sign). Hence the matrix is $\operatorname{diag}(\frac{3}{2}, -\frac{1}{2}) = \begin{bmatrix} \frac{3}{2} & 0 \\ 0 & -\frac{1}{2} \end{bmatrix}$.

■

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

■

Some diagonal matrices rotate Now consider the transformation of multiplying by matrix $\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$: the two reflections of this diagonal matrix, the two −1s, have the same effect as one rotation, here by 180°, as shown to the left. Matrices that rotate are incredibly useful and is the topic of the next section 3.2.3.

Sketch convenient coordinates

Optional preliminary to diagonalisation.

One of the fundamental principles of applying mathematics in science and engineering is that the real world—nature—does its thing irrespective of our mathematical description. Hence we often simplify our mathematical description of real world applications by choosing a coordinate system to suit its nature. That is, although this book (almost) always draws the x or x_1 axis horizontally, and

v0.2a

the y or x_2 axis vertically, in applications it is often better to draw the axes in some other directions—directions which are convenient for the application. This example illustrates the principle.

Example 3.2.34. Consider the transformation shown in the margin (it might arise from the deformation of some material and we need to know the internal stretching and shrinking to predict failure). The drawing has no coordinate axes shown because it 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 $\operatorname{diag}(\frac{1}{2}, 2) = \begin{bmatrix} 0.5 & 0 \\ 0 & 2 \end{bmatrix}$.

Solution: From the diagonal matrix we first look for a direction in which the transformation squashes by a factor of 1/2: from the marginal graph, this direction must be towards the top-right. Second, from the diagonal matrix we look for a direction in which the transformation stretches by a factor of two: from the marginal picture this direction must be aligned to the top-left. Because the top-right corner of the square is stretched a little in this second direction, the first direction must be aimed a little lower than this corner. Hence, coordinate axes that make the transformation the given diagonal matrix are as shown in the margin.

■

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

Solution: From the diagonal matrix we first look for a direction in which the transformation stretches by a factor of three: from the marginal graph, this direction must be aligned along the diagonal top-left to bottom-right. Second, from the diagonal matrix we look for a direction in which the transformation reflects: from the marginal picture this direction must be aligned along the top-right to bottom-left. Hence, coordinate axes that make the transformation the given diagonal matrix are as shown in the margin.

■

Finding such coordinate systems in which a given real world transformation is diagonal is important in science, engineering, and

v0.2

computer science. Systematic methods for such diagonalisation are developed in Section 3.3, and Chapters 4 and 7. These rely on understanding the algebra and geometry of not only diagonal matrices, but also rotations, which is our next topic.

3.2.3 Orthogonal matrices rotate

Whereas diagonal matrices stretch and squash, the so-called 'orthogonal matrices' represent just rotations (and/or reflection). For example, this section shows that multiplying by the 'orthogonal matrix' $\begin{bmatrix} 3/5 & -4/5 \\ 4/5 & 3/5 \end{bmatrix}$ rotates by 53.13° as shown in the marginal picture. Orthogonal matrices are the best to compute with, such as to solve linear equations, since they all have `rcond` $= 1$. To see these and related marvellous properties, we must invoke the geometry of lengths and angles.

Recall the dot product determines lengths and angles Section 1.3 introduced the dot product between two vectors (Definition 1.3.2). For any two vectors in $\mathbb{R}^n$, $\boldsymbol{u} = (u_1, \dots, u_n)$ and $\boldsymbol{v} = (v_1, \dots, v_n)$, define the **dot product**

$$\begin{aligned} \boldsymbol{u} \cdot \boldsymbol{v} &= (u_1, \dots, u_n) \cdot (v_1, \dots, v_n) \\ &= u_1 v_1 + u_2 v_2 + \cdots + u_n v_n \,. \end{aligned}$$

Considering the two vectors as column matrices, the dot product is the same as the matrix product (Example 3.1.18)

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

Also (Theorem 1.3.17a), the length of a vector $\boldsymbol{v} = (v_1, v_2, \dots, v_n)$ in $\mathbb{R}^n$ is the real number

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

and that unit vectors are those of length one. For two non-zero vectors $\boldsymbol{u}, \boldsymbol{v}$ in $\mathbb{R}^n$, 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 (non-zero vectors which are all at right-angles to each other). One example is the set of standard unit vectors $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n\}$ aligned with the coordinate axes in $\mathbb{R}^n$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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

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

Solution: Yes, because all the pairwise dot products are zero: $\boldsymbol{q}_1\cdot\boldsymbol{q}_2 = 2-4+2 = 0$; $\boldsymbol{q}_1\cdot\boldsymbol{q}_3 = -2-2+4 = 0$; $\boldsymbol{q}_2\cdot\boldsymbol{q}_3 = -4+2+2 = 0$.

■

Definition 3.2.38. *A set of non-zero vectors $\{\boldsymbol{q}_1,\boldsymbol{q}_2,\dots,\boldsymbol{q}_k\}$ 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 $i,j = 1,2,\dots,k$. A set of vectors in $\mathbb{R}^n$ is called an* ***orthonormal set*** *if it is an orthogonal set of unit vectors.*

A single non-zero vector always forms an orthogonal set. A single unit vector always forms an orthonormal set.

Example 3.2.39. Any set, or subset, of standard unit vectors in $\mathbb{R}^n$ (Definition 1.2.7) are an orthonormal set as they are all at right-angles (orthogonal), and all of length one. ■

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

Solution: These vectors are all 1/3 of the vectors in Example 3.2.37 and so are orthogonal. They all have length one: $|\boldsymbol{q}_1|^2 = \frac{1}{9}+\frac{4}{9}+\frac{4}{9} = 1$; $|\boldsymbol{q}_2|^2 = \frac{4}{9}+\frac{4}{9}+\frac{1}{9} = 1$; $|\boldsymbol{q}_3|^2 = \frac{4}{9}+\frac{1}{9}+\frac{4}{9} = 1$. Hence $\{\boldsymbol{q}_1,\boldsymbol{q}_2,\boldsymbol{q}_3\}$ is an orthonormal set in $\mathbb{R}^3$.

■

v0.2a

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

(a) $\{(2\,,3)\,,(4\,,-1)\}$ (b) $\{(-5\,,4)\}$

(c) $\{(-2\,,3)\,,(4\,,-6)\}$ (d) $\{\boldsymbol{i}\,,\boldsymbol{k}\}$

■

Orthogonal matrices

Example 3.2.42. Example 3.2.36 showed $\{(3,4),(-8,6)\}$ is an orthogonal set. The vectors have lengths five and ten, respectively, so dividing each by their length means $\{(\frac{3}{5},\frac{4}{5}),(-\frac{4}{5},\frac{3}{5})\}$ is an orthonormal set. Form the matrix Q with these two vectors as its columns:

$$Q=\begin{bmatrix}\frac{3}{5} & -\frac{4}{5}\\ \frac{4}{5} & \frac{3}{5}\end{bmatrix}.$$

Then consider

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

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

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

Definition 3.2.43 (orthogonal matrices)**.** *A square $n\times n$ matrix Q is called an* ***orthogonal matrix*** *if $Q^{\mathrm{T}}Q=I_n$. Because of its special properties (Theorem 3.2.48), multiplication by an orthogonal matrix is called a* ***rotation and/or reflection****;* [9] *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) $p=-1/2$ (b) $p=\sqrt{3}/2$ (c) $p=3/4$ (d) some other

■

[9] Although herein we term multiplication by an orthogonal matrix as a 'rotation' it generally is not a single rotation about a single axis. Instead, generally the 'rotation' characterised by any one orthogonal matrix may be composed of a sequence of several rotations about different axes—each axis with a different orientation.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Example 3.2.45. In the following equation, check that the matrix is orthogonal, and hence 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 below illustrates the rotation of the unit cube under multiplication by the matrix Q: every point $\boldsymbol{x}$ in the (blue) unit cube, is mapped to the point $Q\boldsymbol{x}$ to form the (red) result.

Solution: In MATLAB/Octave (recall the single quote (prime) gives the transpose, Table 3.1),

```
Q=[1,-2,2;2,2,1;-2,1,2]/3
Q'*Q
```

Since the product $Q^{\mathrm{T}}Q$ is I_3, the matrix is orthogonal. Multiplying by Q^{T} both sides of the equation $Q\boldsymbol{x} = \boldsymbol{b}$ gives $Q^{\mathrm{T}}Q\boldsymbol{x} = Q^{\mathrm{T}}\boldsymbol{b}$; that is, $I_3\boldsymbol{x} = Q^{\mathrm{T}}\boldsymbol{b}$, equivalently, $\boldsymbol{x} = Q^{\mathrm{T}}\boldsymbol{b}$. Here,

```
x=Q'*[0;2;-1]
```

gives the solution $\boldsymbol{x} = (2, 1, 0)$.

■

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

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

Solution: Denote the matrix by Q. Given the matrix Q is orthogonal, we know $Q^{\mathrm{T}}Q = I_4$. Just multiply the equation $Q\boldsymbol{x} = \boldsymbol{b}$ by the transpose Q^{T} to deduce $Q^{\mathrm{T}}Q\boldsymbol{x} = Q^{\mathrm{T}}\boldsymbol{b}$, that is, $\boldsymbol{x} = Q^{\mathrm{T}}\boldsymbol{b}$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Here this equation gives the solution to be

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

■

Example 3.2.47. The marginal graph shows a rotation of the unit square. From the graph estimate roughly the matrix Q such that multiplication by Q performs the rotation. Confirm that your estimated matrix is orthogonal (approximately).

Solution: Consider what happens to the standard unit vectors: multiplying Q by $(1,0)$ gives $(0.5,-0.9)$ to roughly (one decimal place); whereas multiplying Q by $(0,1)$ gives $(0.9,0.5)$ roughly. To do this the matrix Q must have these two vectors as its two columns, that is,

$$Q \approx \begin{bmatrix} 0.5 & 0.9 \\ -0.9 & 0.5 \end{bmatrix}.$$

Let's check what happens to the corner point $(1,1)$: $Q(1,1) \approx (1.4,-0.4)$ which looks approximately correct. To confirm orthogonality of Q, find

$$Q^{\mathrm{T}}Q = \begin{bmatrix} 0.5 & -0.9 \\ 0.9 & 0.5 \end{bmatrix} \begin{bmatrix} 0.5 & 0.9 \\ -0.9 & 0.5 \end{bmatrix} = \begin{bmatrix} 1.06 & 0 \\ 0 & 1.06 \end{bmatrix} \approx I_2\,,$$

and so Q is approximately orthogonal.

■

Because orthogonal matrices represent rotations, they arise frequently in engineering and scientific mechanics of bodies. Also, the ease in solving equations with orthogonal matrices puts orthogonal matrices at the heart of coding and decoding photographs (jpeg), videos (mpeg), signals (Fourier transforms), and so on. Furthermore, an extension of orthogonal matrices to complex valued matrices, the so-called unitary matrices, is at the core of quantum physics and quantum computing. Moreover, the next Section 3.3 establishes that orthogonal matrices express the orientation of the action of *every* matrix and hence are a vital component of solving linear equations in general. But to utilise orthogonal matrices across the wide range of applications we need to establish the following properties.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Theorem 3.2.48. *Let Q be an $n \times n$ matrix, then 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).

Proof. Write matrix $Q = \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \cdots & \boldsymbol{q}_n \end{bmatrix}$ in terms of its n columns $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_n$. So then the transpose Q^{T} has the vectors $\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_n$ as its rows—the ith row being the row vector $\boldsymbol{q}_i^{\mathrm{T}}$.

3.2.48a $\iff$ 3.2.48b : Consider

$$
\begin{aligned}
Q^{\mathrm{T}}Q &= \begin{bmatrix} \boldsymbol{q}_1^{\mathrm{T}} \\ \boldsymbol{q}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{q}_n^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \cdots & \boldsymbol{q}_n \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{q}_1 & \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{q}_2 & \cdots & \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{q}_n \\ \boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{q}_1 & \boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{q}_2 & \cdots & \boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{q}_n \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{q}_n^{\mathrm{T}}\boldsymbol{q}_1 & \boldsymbol{q}_n^{\mathrm{T}}\boldsymbol{q}_2 & \cdots & \boldsymbol{q}_n^{\mathrm{T}}\boldsymbol{q}_n \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{q}_1 \cdot \boldsymbol{q}_1 & \boldsymbol{q}_1 \cdot \boldsymbol{q}_2 & \cdots & \boldsymbol{q}_1 \cdot \boldsymbol{q}_n \\ \boldsymbol{q}_2 \cdot \boldsymbol{q}_1 & \boldsymbol{q}_2 \cdot \boldsymbol{q}_2 & \cdots & \boldsymbol{q}_2 \cdot \boldsymbol{q}_n \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{q}_n \cdot \boldsymbol{q}_1 & \boldsymbol{q}_n \cdot \boldsymbol{q}_2 & \cdots & \boldsymbol{q}_n \cdot \boldsymbol{q}_n \end{bmatrix}
\end{aligned}
$$

Matrix Q is orthogonal iff this product is the identity (Definition 3.2.43) which is iff $\boldsymbol{q}_i \cdot \boldsymbol{q}_j = 0$ for $i \neq j$ and $|\boldsymbol{q}_i|^2 = \boldsymbol{q}_i \cdot \boldsymbol{q}_i = 1$, that is, iff the columns are orthonormal (Definition 3.2.38).

3.2.48b $\implies$ 3.2.48c : First, consider the homogeneous system $Q^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{0}$ for $\boldsymbol{x}$ in $\mathbb{R}^n$. We establish $\boldsymbol{x} = \boldsymbol{0}$ is the only solution. The system $Q^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{0}$, written in terms of the orthonormal columns of Q, is

$$
\begin{bmatrix} \boldsymbol{q}_1^{\mathrm{T}} \\ \boldsymbol{q}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{q}_n^{\mathrm{T}} \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{x} \\ \boldsymbol{q}_2^{\mathrm{T}}\boldsymbol{x} \\ \vdots \\ \boldsymbol{q}_n^{\mathrm{T}}\boldsymbol{x} \end{bmatrix} = \begin{bmatrix} \boldsymbol{q}_1 \cdot \boldsymbol{x} \\ \boldsymbol{q}_2 \cdot \boldsymbol{x} \\ \vdots \\ \boldsymbol{q}_n \cdot \boldsymbol{x} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.
$$

Since the dot products are all zero, either $\boldsymbol{x} = \boldsymbol{0}$ or $\boldsymbol{x}$ is orthogonal (at right angles) to all of the n orthonormal vectors $\{\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_n\}$. In $\mathbb{R}^n$ we cannot have $(n+1)$ non-zero vectors all

at right angles to each other (Theorem 1.3.25), consequently $\boldsymbol{x} = \boldsymbol{0}$ is the only possibility as the solution of $Q^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{0}$.

Second, let $n \times n$ matrix $X = I_n - QQ^{\mathrm{T}}$. Pre-multiply by Q^{T}: $Q^{\mathrm{T}}X = Q^{\mathrm{T}}(I_n - QQ^{\mathrm{T}}) = Q^{\mathrm{T}}I_n - (Q^{\mathrm{T}}Q)Q^{\mathrm{T}} = Q^{\mathrm{T}}I_n - I_nQ^{\mathrm{T}} = Q^{\mathrm{T}} - Q^{\mathrm{T}} = O_n$. That is, $Q^{\mathrm{T}}X = O_n$. But each column of $Q^{\mathrm{T}}X = O_n$ is of the form $Q^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{0}$ which we know requires $\boldsymbol{x} = \boldsymbol{0}$, hence $X = O_n$. Then $X = I_n - QQ^{\mathrm{T}} = O_n$ which rearranged gives $QQ^{\mathrm{T}} = I_n$. Put $QQ^{\mathrm{T}} = I_n$ together with $Q^{\mathrm{T}}Q = I_n$ (Definition 3.2.43), then by Definition 3.2.2 Q is invertible with inverse Q^{T}.

3.2.48c $\implies$ 3.2.48a,3.2.48d : Part 3.2.48c asserts Q is invertible with inverse Q^{T}: by Definition 3.2.2 of the inverse, $Q^{\mathrm{T}}Q = QQ^{\mathrm{T}} = I_n$. Since $Q^{\mathrm{T}}Q = I_n$, matrix Q is orthogonal.

Since $I_n = QQ^{\mathrm{T}} = (Q^{\mathrm{T}})^{\mathrm{T}}Q^{\mathrm{T}}$, by Definition 3.2.43 Q^{T} is orthogonal.

3.2.48d $\iff$ 3.2.48e : The proof is similar to that for 3.2.48a $\iff$ 3.2.48b, but for the rows of Q and $QQ^{\mathrm{T}} = I_n$.

3.2.48e $\implies$ 3.2.48c : Similar to that for 3.2.48b $\implies$ 3.2.48c, but for the rows of Q, $Q\boldsymbol{x} = \boldsymbol{0}$ and $X = I_n - Q^{\mathrm{T}}Q$.

3.2.48a $\implies$ 3.2.48f : We prove that multiplication by orthogonal Q preserves all lengths and angles, as illustrated in Examples 3.2.45 and 3.2.47, by comparing the properties of transformed vectors $Q\boldsymbol{u}$ with the properties of the original $\boldsymbol{u}$. For any vectors $\boldsymbol{u}$, $\boldsymbol{v}$ in $\mathbb{R}^n$, consider the dot product $(Q\boldsymbol{u})\cdot(Q\boldsymbol{v}) = (Q\boldsymbol{u})^{\mathrm{T}}Q\boldsymbol{v} = \boldsymbol{u}^{\mathrm{T}}Q^{\mathrm{T}}Q\boldsymbol{v} = \boldsymbol{u}^{\mathrm{T}}I_n\boldsymbol{v} = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{v} = \boldsymbol{u}\cdot\boldsymbol{v}$. We use this identity $(Q\boldsymbol{u})\cdot(Q\boldsymbol{v}) = \boldsymbol{u}\cdot\boldsymbol{v}$ below.

- Firstly, the length $|Q\boldsymbol{u}| = \sqrt{(Q\boldsymbol{u})\cdot(Q\boldsymbol{u})} = \sqrt{\boldsymbol{u}\cdot\boldsymbol{u}} = |\boldsymbol{u}|$ is preserved, and correspondingly for $\boldsymbol{v}$.
- Secondly, let θ be the angle between $\boldsymbol{u}$ and $\boldsymbol{v}$ and θ' be the angle between $Q\boldsymbol{u}$ and $Q\boldsymbol{v}$ (recall $0 \le \text{angle} \le \pi$), then

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

  Since $\cos\theta' = \cos\theta$ and the cosine is 1-1 for $0 \le \text{angles} \le \pi$, all angles are preserved.

3.2.48f $\implies$ 3.2.48b : Look at the consequences of matrix Q preserving all lengths and angles when applied to the standard unit vectors $\boldsymbol{e}_1$, $\boldsymbol{e}_2, \dots, \boldsymbol{e}_n$. Observe $Q\boldsymbol{e}_j = \boldsymbol{q}_j$, the jth column of matrix Q. Then for all j, the length of the jth column $|\boldsymbol{q}_j| = |Q\boldsymbol{e}_j| = |\boldsymbol{e}_j| = 1$ by the preservation of the length of the standard unit vector. Also, for all $i \ne j$ the dot product of columns $\boldsymbol{q}_i\cdot\boldsymbol{q}_j = |\boldsymbol{q}_i||\boldsymbol{q}_j|\cos\theta' = 1\cdot 1\cdot\cos\frac{\pi}{2} = 0$ where θ' is the angle between $\boldsymbol{q}_i$ and $\boldsymbol{q}_j$ which is the angle between $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$ by

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

preservation, namely the angle $\frac{\pi}{2}$. That is, the columns of Q form an orthonormal set.

□

Another important property, proved by Exercise 3.2.19, 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:

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

Solution: For the first matrix each column is a unit vector, and each column is orthogonal to each other: since the matrix has orthonormal columns, then the matrix is orthogonal (Theorem 3.2.48b). Its inverse is the transpose (Theorem 3.2.48c)

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

For the second matrix the two columns are unit vectors as $|(\cos\theta, \sin\theta)|^2 = \cos^2\theta + \sin^2\theta = 1$ and $|(-\sin\theta, \cos\theta)|^2 = \sin^2\theta + \cos^2\theta = 1$. The two columns are orthogonal as the dot product $(\cos\theta, \sin\theta) \cdot (-\sin\theta, \cos\theta) = -\cos\theta\sin\theta + \sin\theta\cos\theta = 0$. Since the matrix has orthonormal columns, then the matrix is orthogonal (Theorem 3.2.48b). Its inverse is the transpose (Theorem 3.2.48c)

$$\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}.$$

■

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

Solution:

(a) No—the square is stretched and angles changed.

(b) Yes—the square is just rotated.

(c) Yes—the square is rotated and reflected.

(d) No—the square is squashed and angles changed.

■

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

- Which transformations appear to be that of multiplying by an orthogonal matrix?

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

■

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

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(c)

(d)

Solution: (a) Yes—the cube is just rotated.

(b) No—the cube is stretched and angles changed.

(c) No—the cube is stretched and angles changed.

(d) Yes—the cube is just rotated and reflected.

■

3.2.4 Exercises

Exercise 3.2.1. By direct multiplication, both ways, confirm that for each of the following pairs, matrix B is an inverse of matrix A, or not.

(a) $A = \begin{bmatrix} 0 & -4 \\ 4 & 4 \end{bmatrix}$, $B = \begin{bmatrix} 1/4 & 1/4 \\ -1/4 & 0 \end{bmatrix}$

(b) $A = \begin{bmatrix} -3 & 3 \\ -3 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 1/6 & -1/2 \\ 1/2 & -1/2 \end{bmatrix}$

(c) $A = \begin{bmatrix} 5 & -1 \\ 3 & -2 \end{bmatrix}$, $B = \begin{bmatrix} 3/7 & -1/7 \\ 3/7 & -5/7 \end{bmatrix}$

(d) $A = \begin{bmatrix} -1 & 1 \\ -5 & -1 \end{bmatrix}$, $B = \begin{bmatrix} -1/6 & -1/6 \\ 5/6 & -1/6 \end{bmatrix}$

(e) $A = \begin{bmatrix} -2 & 4 & 2 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/6 & 1/3 & 0 \\ 1/6 & -1/6 & 1/2 \end{bmatrix}$

(f) $A = \begin{bmatrix} -3 & 0 & -1 \\ 1 & 4 & 2 \\ 3 & -4 & -1 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 1 & 1 \\ 7/4 & 3/2 & 5/4 \\ -4 & -3 & -3 \end{bmatrix}$

(g) $A = \begin{bmatrix} -3 & -3 & 3 \\ 4 & 3 & -3 \\ -2 & -1 & 4 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 1 & 0 \\ -1 & -2/3 & 1/3 \\ 2/9 & 1/3 & 1/3 \end{bmatrix}$

v0.2a

(h) $A = \begin{bmatrix} -1 & 3 & 4 & -1 \\ 2 & 2 & -2 & 0 \\ 1 & 2 & -2 & 0 \\ 4 & 2 & 4 & -1 \end{bmatrix}$, $B = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 1 & 5 & -5 & -1 \\ 1 & 11/2 & -6 & -1 \\ 6 & 36 & -38 & -7 \end{bmatrix}$, use MATLAB/Octave

(i) $A = \begin{bmatrix} 0 & -2 & -2 & -3 \\ -4 & -5 & 0 & 1 \\ -2 & 0 & 4 & 5 \\ -1 & 1 & 0 & -2 \end{bmatrix}$, $B = \begin{bmatrix} 2/3 & -1/3 & 1/3 & -1/3 \\ -2/3 & 1/9 & -1/3 & 2/9 \\ 7/6 & -4/9 & 5/6 & 1/9 \\ -2/3 & 2/9 & -1/3 & -2/9 \end{bmatrix}$, use MATLAB/Octave

(j) $A = \begin{bmatrix} 0 & 0 & 7 & -1 \\ 2 & 4 & 0 & 4 \\ -2 & -2 & 1 & -2 \\ -3 & -3 & 1 & -3 \end{bmatrix}$, $B = \begin{bmatrix} 0 & -1/2 & 2 & -2 \\ 1 & 1/2 & -21 & 15 \\ 0 & 0 & 3 & -2 \\ -1 & 0 & 21 & -14 \end{bmatrix}$, use MATLAB/Octave

(k) $A = \begin{bmatrix} 1 & -7 & -4 & 3 \\ 0 & 2 & 1 & -1 \\ 0 & -2 & -3 & 2 \\ 3 & -2 & -4 & 1 \end{bmatrix}$, $B = \begin{bmatrix} -2 & -7 & -1 & 1 \\ -1 & -8/3 & 0 & 1/3 \\ -2 & -22/3 & -1 & 2/3 \\ -4 & -41/3 & -1 & 4/3 \end{bmatrix}$, use MATLAB/Octave

Exercise 3.2.2. Use Theorem 3.2.7 to calculate the inverse, when it exists, of the following 2×2 matrices.

(a) $\begin{bmatrix} -2 & 2 \\ -1 & 4 \end{bmatrix}$ (b) $\begin{bmatrix} -5 & -10 \\ -1 & -2 \end{bmatrix}$

(c) $\begin{bmatrix} -2 & -4 \\ 5 & 2 \end{bmatrix}$ (d) $\begin{bmatrix} -3 & 2 \\ -1 & -2 \end{bmatrix}$

(e) $\begin{bmatrix} 2 & -4 \\ 3 & 0 \end{bmatrix}$ (f) $\begin{bmatrix} -0.6 & -0.9 \\ 0.8 & -1.4 \end{bmatrix}$

(g) $\begin{bmatrix} 0.3 & 0 \\ 0.9 & 1.9 \end{bmatrix}$ (h) $\begin{bmatrix} 0.6 & 0.5 \\ -0.3 & 0.5 \end{bmatrix}$

Exercise 3.2.3. Given the inverses of Exercises 3.2.1, solve each of the following systems of linear equations with a matrix-vector multiply (Theorem 3.2.10).

(a) $\begin{cases} -4y = 1 \\ 4x + 4y = -5 \end{cases}$ (b) $\begin{cases} -3p + 3q = 3 \\ -3p + q = -1 \end{cases}$

(c) $\begin{cases} m - x = 1 \\ -m - 5x = -1 \end{cases}$ (d) $\begin{cases} -3x - z = 3 \\ x + 4y + 2z = -3 \\ 3x - 4y - z = 2 \end{cases}$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(e) $\begin{cases} 2p - 2q + 4r = -1 \\ -p + q + r = -2 \\ p + q - r = 1 \end{cases}$

(f) $\begin{cases} -x_1 + 3x_2 + 4x_3 - x_4 = 0 \\ 2x_1 + 2x_2 - 2x_3 = -1 \\ x_1 + 2x_2 - 2x_3 = 3 \\ 4x_1 + 2x_2 + 4x_3 - x_4 = -5 \end{cases}$

(g) $\begin{cases} p - 7q - 4r + 3s = -1 \\ 2q + r - s = -5 \\ -2q - 3r + 2s = 3 \\ 3p - 2q - 4r + s = -1 \end{cases}$

(h) $\begin{cases} -3b - 2c - 2d = 4 \\ -4a + b - 5d = -3 \\ -2a + 5b + 4c = -2 \\ -a - 2b + d = 0 \end{cases}$

Exercise 3.2.4. 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) The general solution is $(-2, 1, 2, 0, 2)$.

(b) A solution of a system is $(2.4, -2.8, -3.6, -2.2, -3.8)$.

(c) A solution of a homogeneous system is $(0.8, 0.4, -2.3, 2.5)$.

(d) The general solution of a system is $(4, 1, 0, 2)t$ for all t.

(e) The general solution of a homogeneous system is $(0, 0, 0, 0)$.

(f) A solution of a homogeneous system is $(0, 0, 0, 0, 0)$.

Exercise 3.2.5. Use MATLAB/Octave to generate some random matrices of a suitable size of your choice, and some random scalar exponents (see Table 3.1). Then confirm the properties of inverse matrices given by Theorem 3.1.22. For the purposes of this exercise, use the MATLAB/Octave function `inv(A)` that computes the inverse of the matrix A if it exists (as commented, computing the inverse of a matrix is generally not appropriate—the inverse is primarily a theoretical device—this exercise only computes the inverse for educational purposes). Record all your commands and the output from MATLAB/Octave.

Exercise 3.2.6. Consider Theorem 3.2.13 on the properties of the inverse. Invoking properties of matrix operations from section 3.1.3,

(a) prove Part 3.2.13b using associativity, and

(b) prove Part 3.2.13d using the transpose.

v0.2a

Exercise 3.2.7. Using the inverses identified in Exercise 3.2.1, and matrix multiplication, calculate the following matrix powers.

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

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

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

(d) $\begin{bmatrix} -1/6 & -1/6 \\ 5/6 & -1/6 \end{bmatrix}^{-4}$

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

(f) $\begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/6 & 1/3 & 0 \\ 1/6 & -1/6 & 1/2 \end{bmatrix}^{-2}$

(g) $\begin{bmatrix} -1 & 3 & 4 & -1 \\ 2 & 2 & -2 & 0 \\ 1 & 2 & -2 & 0 \\ 4 & 2 & 4 & -1 \end{bmatrix}^{-2}$ use MATLAB/Octave

(h) $\begin{bmatrix} -2 & -7 & -1 & 1 \\ -1 & -8/3 & 0 & 1/3 \\ -2 & -22/3 & -1 & 2/3 \\ -4 & -41/3 & -1 & 4/3 \end{bmatrix}^{-3}$ use MATLAB/Octave.

Exercise 3.2.8. Which of the following matrices are diagonal? For those that are diagonal, write down how they may be represented with the diag function (algebraic, not MATLAB/Octave).

(a) $\begin{bmatrix} 9 & 0 & 0 \\ 0 & -5 & 0 \\ 0 & 0 & 4 \end{bmatrix}$

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

(c) $\begin{bmatrix} -5 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 9 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

(d) $\begin{bmatrix} 6 & 0 & 0 & 0 \\ 0 & 1 & 0 & -9 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

(e) $\begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -5 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

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

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

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

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

(j) $\begin{bmatrix} -3c & 0 & 0 & -4d \\ 0 & 5b & 0 & 0 \\ a & 0 & 0 & 0 \end{bmatrix}$

v0.2

Exercise 3.2.9. Write down the individual algebraic equations represented by each of the following diagonal matrix-vector equations. Hence, where possible solve each system.

(a) $\begin{bmatrix} -4 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 6 \\ -1 \\ -4 \\ -2 \\ 4 \end{bmatrix}$

(b) $\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -8 \\ 3 \\ -1 \end{bmatrix}$

(c) $\begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 5 \\ 0 \end{bmatrix}$

(d) $\begin{bmatrix} -6 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix}$

(e) $\begin{bmatrix} -3 & 0 & 0 & 0 & 0 \\ 0 & -4 & 0 & 0 & 0 \\ 0 & 0 & -6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} -5 \\ -5 \\ 1 \\ -1 \end{bmatrix}$

(f) $\begin{bmatrix} 0.5 & 0 & 0 & 0 \\ 0 & 0.2 & 0 & 0 \end{bmatrix} \begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$

(g) $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -3 & 0 & 0 \\ 0 & 0 & -3 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} p \\ q \\ r \\ s \end{bmatrix} = \begin{bmatrix} -2 \\ -6 \\ 8 \\ 0 \end{bmatrix}$

(h) $\begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \ \end{bmatrix} = \begin{bmatrix} -3 \\ 0 \\ 3 \end{bmatrix}$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Exercise 3.2.10. In each of the following illustrations, the unit square (blue) is transformed by a matrix multiplication to some shape (red). Which of these transformations correspond to multiplication by a diagonal matrix? For those that are, estimate the elements of the diagonal matrix.

v0.2a

Exercise 3.2.11. In each of the following stereo illustrations, the unit cube (blue) is transformed by a matrix multiplication to some shape (red). Which of these transformations correspond to multiplication by a diagonal matrix? For those that are, estimate the elements of the diagonal matrix.

(a)

(b)

(c)

(d)

(e)

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(f)

Exercise 3.2.12. Consider each of the transformations shown below that transform the from blue unit square to the red parallelogram. They each have no coordinate axes shown because it is supposed to be some transformation in nature. Now impose on nature our mathematical description. Draw approximate orthogonal coordinate axes, with origin at the common corner point, so the transformation becomes that of multiplication by the specified diagonal matrix.

(a) $\operatorname{diag}(1/2\,,2)$,

(b) $\operatorname{diag}(3/2\,,1/2)$,

(c) $\operatorname{diag}(3\,,0)$,

(d) $\operatorname{diag}(-1/2\,,-3/2)$,

(e) $\operatorname{diag}(1/2\,,-5/2)$,

(f) $\operatorname{diag}(-1\,,1)$,

(g) $\operatorname{diag}(-1\,,-2)$,

(h) $\operatorname{diag}(-1/2\,,1)$,

v0.2a

Exercise 3.2.13. Which of the following pairs of vectors appear to form an orthogonal set of two vectors? Which appear to form an orthonormal set of two vectors?

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

Exercise 3.2.14. Which of the following sets of vectors, drawn as stereo pairs, appear to form an orthogonal set? Which appear to form an orthonormal set?

(a)

v0.2a

(b)

(c)

(d)

(e)

(f)

Exercise 3.2.15. Use the dot product to determine which of the following sets of vectors are orthogonal sets. For the orthogonal sets, scale the vectors to form an orthonormal set.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(a) $\{(2,3,6),\ (3,-6,2),\ (6,2,-3)\}$

(b) $\{(4,4,7),\ (1,-8,4),\ (-8,1,4)\}$

(c) $\{(2,6,9),\ (9,-6,2)\}$

(d) $\{(6,3,2),\ (3,-6,2),\ (2,-3,6)\}$

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

(f) $\{(1,2,2,4),\ (2,-1,-4,2)\}$

(g) $\{(1,2,2,4),\ (2,-1,-4,2),\ (-4,2,2,-1)\}$

(h) $\{(5,6,2,4),\ (-2,6,-5,-4),\ (6,-5,-4,2)\}$

Exercise 3.2.16. Using Definition 3.2.43 determine which of the following matrices are orthogonal matrices. For those matrices which are orthogonal, confirm Theorem Part 3.2.48c.

(a) $\begin{bmatrix} \frac{5}{13} & \frac{12}{13} \\ -\frac{12}{13} & \frac{5}{13} \end{bmatrix}$

(b) $\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$

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

(d) $\begin{bmatrix} \frac{2}{7} & \frac{3}{7} & \frac{6}{7} \\ \frac{3}{7} & -\frac{6}{7} & \frac{2}{7} \\ \frac{6}{7} & \frac{2}{7} & -\frac{3}{7} \end{bmatrix}$

(e) $\begin{bmatrix} \frac{2}{11} & \frac{9}{11} \\ \frac{6}{11} & -\frac{6}{11} \\ \frac{9}{11} & \frac{2}{11} \end{bmatrix}$

(f) $\begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix}$

(g) $\frac{1}{9}\begin{bmatrix} 4 & 1 & 8 \\ 4 & -8 & -1 \\ 7 & 4 & -4 \end{bmatrix}$

(h) $\frac{1}{7}\begin{bmatrix} 1 & 4 & 4 & 4 \\ 4 & -5 & 2 & 2 \\ 4 & 2 & -5 & 2 \\ 4 & 2 & 2 & -5 \end{bmatrix}$

(i) $\begin{bmatrix} 0.2 & 0.4 & 0.4 \\ 0.4 & -0.2 & 0.8 \\ 0.4 & -0.8 & -0.2 \\ 0.8 & 0.4 & -0.4 \end{bmatrix}$

(j) $\frac{1}{6}\begin{bmatrix} 1 & 1 & 3 & 5 \\ -5 & -3 & 1 & 1 \\ 3 & -5 & -1 & 1 \\ 1 & -1 & 5 & 3 \end{bmatrix}$

(k) $\begin{bmatrix} 0.1 & 0.5 & 0.5 & 0.7 \\ 0.5 & -0.1 & -0.7 & 0.5 \\ 0.5 & 0.7 & -0.1 & -0.5 \\ 0.7 & -0.5 & 0.5 & -0.1 \end{bmatrix}$

v0.2a

Exercise 3.2.17. Each part gives an orthogonal matrix Q and two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$. For each part calculate the lengths of $\boldsymbol{u}$ and $\boldsymbol{v}$, and the angle between $\boldsymbol{u}$ and $\boldsymbol{v}$. Confirm these are the same as the lengths of $Q\boldsymbol{u}$ and $Q\boldsymbol{v}$, and the angle between $Q\boldsymbol{u}$ and $Q\boldsymbol{v}$, respectively.

(a) $Q = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, $\boldsymbol{u} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$, $\boldsymbol{v} = \begin{bmatrix} 12 \\ 5 \end{bmatrix}$

(b) $Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$, $\boldsymbol{u} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$, $\boldsymbol{v} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$

(c) $Q = \begin{bmatrix} -\frac{3}{5} & \frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{bmatrix}$, $\boldsymbol{u} = \begin{bmatrix} 2 \\ -3 \end{bmatrix}$, $\boldsymbol{v} = \begin{bmatrix} 0 & -4 \end{bmatrix}$

(d) $Q = \begin{bmatrix} \frac{3}{5} & \frac{4}{5} \\ -\frac{4}{5} & \frac{3}{5} \end{bmatrix}$, $\boldsymbol{u} = \begin{bmatrix} 7 \\ -4 \end{bmatrix}$, $\boldsymbol{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

(e) $Q = \frac{1}{11} \begin{bmatrix} 7 & 6 & 6 \\ 6 & 2 & -9 \\ 6 & -9 & 2 \end{bmatrix}$, $\boldsymbol{u} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$, $\boldsymbol{v} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$

(f) $Q = \frac{1}{9} \begin{bmatrix} 7 & 4 & 4 \\ 4 & -8 & 1 \\ 4 & 1 & -8 \end{bmatrix}$, $\boldsymbol{u} = \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}$, $\boldsymbol{v} = \begin{bmatrix} -1 \\ 2 \\ -3 \end{bmatrix}$

(g) $Q = \begin{bmatrix} 0.1 & 0.1 & 0.7 & 0.7 \\ 0.1 & -0.1 & -0.7 & 0.7 \\ 0.7 & 0.7 & -0.1 & -0.1 \\ 0.7 & -0.7 & 0.1 & -0.1 \end{bmatrix}$, $\boldsymbol{u} = \begin{bmatrix} 0 \\ 0 \\ -2 \\ -3 \end{bmatrix}$, $\boldsymbol{v} = \begin{bmatrix} -1 \\ -1 \\ -2 \\ -1 \end{bmatrix}$

(h) $Q = \begin{bmatrix} 0.1 & 0.3 & 0.3 & 0.9 \\ 0.3 & -0.1 & -0.9 & 0.3 \\ 0.3 & 0.9 & -0.1 & -0.3 \\ 0.9 & -0.3 & 0.3 & -0.1 \end{bmatrix}$, $\boldsymbol{u} = \begin{bmatrix} 3 \\ 1 \\ 2 \\ 1 \end{bmatrix}$, $\boldsymbol{v} = \begin{bmatrix} -2 \\ -1 \\ 0 \\ 2 \end{bmatrix}$

Exercise 3.2.18. Using one or other of the orthogonal matrices appearing in Exercise 3.2.17, solve each of the following systems of linear equations by a matrix-vector multiplication.

(a) $\begin{cases} \frac{3}{5}x + \frac{4}{5}y = 5 \\ -\frac{4}{5}x + \frac{3}{5}y = 2 \end{cases}$

(b) $\begin{cases} -\frac{3}{5}x + \frac{4}{5}y = 1.6 \\ \frac{4}{5}x + \frac{3}{5}y = -3.5 \end{cases}$

(c) $\begin{cases} \frac{1}{\sqrt{2}}(x + y) = 3 \\ \frac{1}{\sqrt{2}}(x - y) = 2 \end{cases}$

(d) $\begin{cases} 3x + 4y = 20 \\ -4x + 3y = 5 \end{cases}$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(e) $\begin{cases} \frac{7}{9}p + \frac{4}{9}q + \frac{4}{9}r = 2 \\ \frac{4}{9}p - \frac{8}{9}q + \frac{1}{9}r = 3 \\ \frac{4}{9}p + \frac{1}{9}q - \frac{8}{9}r = 7 \end{cases}$

(f) $\begin{cases} \frac{7}{9}u + \frac{4}{9}v + \frac{4}{9}w = 1 \\ \frac{4}{9}u + \frac{1}{9}v - \frac{8}{9}w = 2 \\ \frac{4}{9}u - \frac{8}{9}v + \frac{1}{9}w = 0 \end{cases}$

(g) $\begin{cases} 7a + 6b + 6c = 22 \\ 6a + 2b - 9c = 11 \\ 6a - 9b + 2c = -22 \end{cases}$

(h) $\begin{cases} 0.1x_1 + 0.1x_2 + 0.7x_3 + 0.7x_4 = 1 \\ 0.1x_1 - 0.1x_2 - 0.7x_3 + 0.7x_4 = -1 \\ 0.7x_1 + 0.7x_2 - 0.1x_3 - 0.1x_4 = 0 \\ 0.7x_1 - 0.7x_2 + 0.1x_3 - 0.1x_4 = -2 \end{cases}$

(i) $\begin{cases} 0.1y_1 + 0.1y_2 + 0.7y_3 + 0.7y_4 = 1 \\ 0.7y_1 + 0.7y_2 - 0.1y_3 - 0.1y_4 = -2.5 \\ 0.7y_1 - 0.7y_2 + 0.1y_3 - 0.1y_4 = 2 \\ 0.1y_1 - 0.1y_2 - 0.7y_3 + 0.7y_4 = 2.5 \end{cases}$

(j) $\begin{cases} z_1 + 3z_2 + 3z_3 + 9z_4 = 5 \\ 3z_1 - z_2 - 9z_3 + 3z_4 = 0 \\ 3z_1 + 9z_2 - z_3 - 3z_4 = -1 \\ 9z_1 - 3z_2 + 3z_3 - z_4 = -3 \end{cases}$

Exercise 3.2.19 (product of orthogonal matrices)**.** Use Definition 3.2.43 to prove that if Q_1 and Q_2 are orthogonal matrices of the same size, then so is the product Q_1Q_2. Consider $(Q_1Q_2)^{\mathrm{T}}(Q_1Q_2)$.

Exercise 3.2.20. Fill in details of the proof for Theorem 3.2.48 to establish that a matrix Q^{T} is orthogonal iff the row vectors of Q form an orthonormal set.

Exercise 3.2.21. Fill in details of the proof for Theorem 3.2.48 to establish that if the row vectors of Q form an orthonormal set, then Q is invertible and $Q^{-1} = Q^{\mathrm{T}}$.

Exercise 3.2.22. The following graphs illustrate the transformation of the unit square through multiplying by some different matrices. Using Theorem 3.2.48f, which transformations appear to be that of multiplying by an orthogonal matrix?

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Exercise 3.2.23. The following stereo pairs illustrate the transformation of the unit cube through multiplying by some different matrices. Using Theorem 3.2.48f, which transformations appear to be that of multiplying by an orthogonal matrix?

v0.2a

(c)

(d)

(e)

(f)

(g)

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

3.3 Factorise to the singular value decomposition

Section Contents

3.3.1 Introductory examples 221
3.3.2 The SVD solves general systems 226
Computers empower use of the SVD 228
Condition number and rank determine possibilities 234
3.3.3 Prove the SVD Theorem 3.3.6 244
Prelude to the proof 244
Detailed proof of the SVD Theorem 3.3.6 . . 246
3.3.4 Exercises 249

Beltrami first derived the SVD in 1873. The first reliable method for computing an SVD was developed by Golub and Kahan in 1965, and only thereafter did applications proliferate.

The singular value decomposition (SVD) is sometimes called the jewel in the crown of linear algebra. Its importance is certified by the many names by which it is invoked in scientific and engineering applications: principal component analysis, singular spectrum analysis, principal orthogonal decomposition, latent semantic indexing, Schmidt decomposition, correspondence analysis, Lanczos methods, dimension reduction, and so on. Let's start seeing what it can do for us.

3.3.1 Introductory examples

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 . . . . . .

Solution: Long division is hopeless in the time available. However, recognise $42 = 2\cdot 3\cdot 7$ and so divide 1554 by 2 to get 777, divide 777 by 3 to get 259, and divide 259 by 7 to get 37, and win the prize.

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

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

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

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

1. Factorise the coefficient $42 = 2 \cdot 3 \cdot 7$ so the equation becomes

$$2 \cdot \underbrace{3 \cdot \overbrace{7x}^{=y}}_{=z} = 1554\,,$$

   and introduce two intermediate unknowns y and z as shown above.

2. Solve $2z = 1554$ to get $z = 777$.

3. Solve $3y = z = 777$ to get $y = 259$.

4. Solve $7x = y = 259$ to get $x = 37$—the answer.

Now let's proceed to small matrix examples to introduce the general matrix procedure empowered by the factorisation of a singular value decomposition (SVD).

Example 3.3.2. Solve the 2×2 system

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

given the matrix factorisation

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

(note the transpose on the last matrix).

Solution: Optionally check the factorisation if you like:

$$\begin{bmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{bmatrix} \begin{bmatrix} 10\sqrt{2} & 0 \\ 0 & 5\sqrt{2} \end{bmatrix} = \begin{bmatrix} 6\sqrt{2} & -4\sqrt{2} \\ 8\sqrt{2} & 3\sqrt{2} \end{bmatrix};$$

$$\text{then} \quad \begin{bmatrix} 6\sqrt{2} & -4\sqrt{2} \\ 8\sqrt{2} & 3\sqrt{2} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 10 & 2 \\ 5 & 11 \end{bmatrix}.$$

Given the factorisation, the following four steps forms the general procedure.

(a) Write the system using the factorisation, and with two intermediate unknowns $\boldsymbol{y}$ and $\boldsymbol{z}$:

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

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(b) Solve $\begin{bmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{bmatrix} \boldsymbol{z} = \begin{bmatrix} 18 \\ -1 \end{bmatrix}$: recall that the matrix appearing here is orthogonal (and this orthogonality is no accident), so multiplying by the transpose gives the intermediary

$$\boldsymbol{z} = \begin{bmatrix} \frac{3}{5} & \frac{4}{5} \\ -\frac{4}{5} & \frac{3}{5} \end{bmatrix} \begin{bmatrix} 18 \\ -1 \end{bmatrix} = \begin{bmatrix} 10 \\ -15 \end{bmatrix}.$$

(c) Now solve $\begin{bmatrix} 10\sqrt{2} & 0 \\ 0 & 5\sqrt{2} \end{bmatrix} \boldsymbol{y} = \boldsymbol{z} = \begin{bmatrix} 10 \\ -15 \end{bmatrix}$: the matrix appearing here is diagonal (and this is no accident), so dividing by the respective diagonal elements gives the intermediary

$$\boldsymbol{y} = \begin{bmatrix} 10\sqrt{2} & 0 \\ 0 & 5\sqrt{2} \end{bmatrix}^{-1} \begin{bmatrix} 10 \\ -15 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} \\ -3/\sqrt{2} \end{bmatrix}.$$

(d) Finally solve $\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}^{\mathrm{T}} \boldsymbol{x} = \boldsymbol{y} = \begin{bmatrix} 1/\sqrt{2} \\ -3/\sqrt{2} \end{bmatrix}$: now the matrix appearing here is also orthogonal (this orthogonality is also no accident), so multiplying by itself (the transpose of the transpose) gives the solution

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

■

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 factorisation

$$\begin{bmatrix} 12 & -41 \\ 34 & -12 \end{bmatrix} = \begin{bmatrix} \frac{4}{5} & -\frac{3}{5} \\ \frac{3}{5} & \frac{4}{5} \end{bmatrix} \begin{bmatrix} 50 & 0 \\ 0 & 25 \end{bmatrix} \begin{bmatrix} \frac{3}{5} & \frac{4}{5} \\ -\frac{4}{5} & \frac{3}{5} \end{bmatrix}^{\mathrm{T}}$$

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

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

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

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

■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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 factorisation (note the last is transposed)

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

Solution: Use MATLAB/Octave. Enter the matrices and the right-hand side, and check the factorisation (and the typing):

```
U=[1,-2,2;2,2,1;-2,1,2]/3
S=[12,0,0;0,6,0;0,0,3]
V=[-8,-1,-4;-4,4,7;-1,-8,4]/9
b=[10;2;-2]
A=U*S*V'
```

(a) Write the system $A\boldsymbol{x} = \boldsymbol{b}$ using the factorisation, and with two intermediate unknowns $\boldsymbol{y}$ and $\boldsymbol{z}$:

$$\begin{bmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\ -\frac{2}{3} & \frac{1}{3} & \frac{2}{3} \end{bmatrix} \underbrace{\begin{bmatrix} 12 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 3 \end{bmatrix} \overbrace{\begin{bmatrix} -\frac{8}{9} & -\frac{1}{9} & -\frac{4}{9} \\ -\frac{4}{9} & \frac{4}{9} & \frac{7}{9} \\ -\frac{1}{9} & -\frac{8}{9} & \frac{4}{9} \end{bmatrix}^{\mathrm{T}} \boldsymbol{x}}^{=\boldsymbol{y}}}_{=\boldsymbol{z}} = \begin{bmatrix} 10 \\ 2 \\ -2 \end{bmatrix} .$$

(b) Solve $\begin{bmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\ -\frac{2}{3} & \frac{1}{3} & \frac{2}{3} \end{bmatrix} \boldsymbol{z} = \begin{bmatrix} 10 \\ 2 \\ -2 \end{bmatrix}$. Now this matrix, called `U`, is orthogonal—check by computing `U'*U`—so multiplying by the transpose gives the intermediary: `z=U'*b` $= (6, -6, 6)$.

(c) Then solve $\begin{bmatrix} 12 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 3 \end{bmatrix} \boldsymbol{y} = \boldsymbol{z} = \begin{bmatrix} 6 \\ -6 \\ 6 \end{bmatrix}$: this matrix, called `S`, is diagonal, so dividing by the respective diagonal elements gives the intermediary `y=z./diag(S)` $= (\frac{1}{2}, -1, 2)$.

(d) Finally solve $\begin{bmatrix} -\frac{8}{9} & -\frac{1}{9} & -\frac{4}{9} \\ -\frac{4}{9} & \frac{4}{9} & \frac{7}{9} \\ -\frac{1}{9} & -\frac{8}{9} & \frac{4}{9} \end{bmatrix}^{\mathrm{T}} \boldsymbol{x} = \boldsymbol{y} = \begin{bmatrix} \frac{1}{2} \\ -1 \\ 2 \end{bmatrix}$. This matrix, called `V`, is also orthogonal—check by computing `V'*V`—so multiplying by itself (the transpose of the transpose) gives the final solution `x=V*y` $= (-\frac{11}{9}, \frac{8}{9}, \frac{31}{18})$.

v0.2a

■

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

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

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

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

(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' $\begin{bmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{bmatrix} \boldsymbol{z} = \begin{bmatrix} \frac{19}{20} \\ \frac{17}{10} \end{bmatrix}$: 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 multiplicative is *not* commutative (section 3.1.3). We must use an SVD factorisation in the correct order: to solve linear equations use the matrices in an SVD from left to right.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

3.3.2 The SVD solves general systems

The previous examples depended upon a matrix being factored into a product of three matrices: two orthogonal and one diagonal. Amazingly, such factorisation is always possible.

http://www.youtube.com/watch?v=JEYLfIVvR9I is an entertaining prelude

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

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

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*** *$\sigma_1, \sigma_2, \ldots, \sigma_{\min(m,n)}$, 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 greek letter σ (sigma) denotes the singular values.

The orthonormal vectors $\boldsymbol{u}_j$ and $\boldsymbol{v}_j$ are called ***singular vectors****.* [10]

Proof. Detailed in Section 3.3.3. Importantly, the singular values are unique (when ordered), although the orthogonal matrices U and V are not unique (e.g., one may change the sign of any column in U together with its corresponding column in V). Nonetheless, although there are many SVDs of a matrix, all SVDs are equivalent in application. □

Some may be disturbed by the non-uniqueness of an SVD. But the non-uniqueness is analogous to the non-uniqueness of row reduction upon arbitrary re-ordering of equations, and/or re-ordering the variables in the equations.

Example 3.3.7. Example 3.3.2 invoked the SVD

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

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

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

[10] This enormously useful theorem also generalises from $m \times n$ matrices to analogues in 'infinite' dimensions: an SVD exists for all compact linear operators (Kress 2015, §7).

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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, $U = Q$, $S = V = I_n$. 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 $I_n = I_nI_nI_n^{\mathrm{T}}$.
- Additionally, for *any* $n \times n$ 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 $sI_n = \operatorname{diag}(s, s, \dots, s)$, has the same non-uniqueness: an SVD is $sI_n = USV^{\mathrm{T}}$ for matrices $U = Q$, $S = sI_n$ and $V = Q$ for any $n \times n$ orthogonal Q (provided $s \geq 0$).

The matrices in this example are characterised by all their singular values having an identical value. In general, analogous non-uniqueness 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 an orthogonal matrix Q?

(a) $Q(-I_n)(-I_n)^{\mathrm{T}}$ (b) $I_nQI_n^{\mathrm{T}}$

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

■

Example 3.3.11 (positive ordering)**.** Find an SVD of the diagonal matrix

$$D = \begin{bmatrix} 2.7 & 0 & 0 \\ 0 & -3.9 & 0 \\ 0 & 0 & -0.9 \end{bmatrix}.$$

Solution: Singular values cannot be negative so a factorisation is

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 2.7 & 0 & 0 \\ 0 & 3.9 & 0 \\ 0 & 0 & 0.9 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{\mathrm{T}},$$

where the (-1)s in the first matrix encode the signs of the corresponding diagonal elements (one could alternatively use the rightmost matrix to encode the pattern of signs). However, Theorem 3.3.6 requires that singular values be ordered in decreasing

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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 examples:
  - `V(:,1:r)` selects the first r columns of V;
  - `A([2 3 5],:)` selects the second, third and fifth row of matrix A;
  - `B(4:6,1:3)` selects the 3×3 submatrix of the first three columns of the fourth, fifth and sixth rows.

magnitude, so sort the diagonal of the middle matrix into order and correspondingly permute the columns of the outer two matrices to obtain the following SVD:

$$D = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 3.9 & 0 & 0 \\ 0 & 2.7 & 0 \\ 0 & 0 & 0.9 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}^{\mathrm{T}} .$$

■

Computers empower use of the SVD

Except for simple cases such as 2×2 matrices (Example 3.3.32), constructing an SVD is usually far too laborious by hand. [11] Typically, this book either gives an SVD (as in the earlier two examples) or asks you to compute an SVD in MATLAB/Octave with `[U,S,V]=svd(A)` (Table 3.3).

The SVD theorem asserts that every matrix is the product of two orthogonal matrices and a diagonal matrix. Because the rotations (and/or reflection) by the two orthogonal matrices are so 'nice', any 'badness' or 'trickiness' in a matrix is represented in the diagonal matrix S of the singular values in its SVD factorisation.

The following examples illustrate the cases of either no or infinite solutions, to complement the case of unique solutions of the first two examples.

[11] For those interested advanced students, Trefethen & Bau (1997) [p.234] discusses how the standard method of numerically computing an SVD is based upon first transforming to bidiagonal form, and then using an iteration based upon a so-called QR factorisation.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Example 3.3.12 (rate sport teams/players). Consider three table tennis players, Anne, Bob and Chris: Anne beat Bob 3 games to 2 games; Anne beat Chris 3-1; Bob beat Chris 3-2. How good are they? What is their rating?

Solution: Denote Anne's rating by x_1, Bob's rating by x_2, and Chris' rating by x_3. The ratings should predict the results of matches, so from the above three match results, surely

- Anne beat Bob 3 games to 2 $\leftrightarrow x_1 - x_2 = 3 - 2 = 1$;
- Anne beat Chris 3-1 $\leftrightarrow x_1 - x_3 = 3 - 1 = 2$; and
- Bob beat Chris 3-2 $\leftrightarrow x_2 - x_3 = 3 - 2 = 1$.

In matrix-vector form, $A\boldsymbol{x} = \boldsymbol{b}$,

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

In MATLAB/Octave, we might try Procedure 2.2.5:

```
A=[1,-1,0;1,0,-1;0,1,-1]
b=[1;2;1]
rcond(A)
```

but find `rcond=0` which is extremely terrible so we cannot use `A\b` to solve the system $A\boldsymbol{x} = \boldsymbol{b}$. *Whenever difficulties arise, use an* SVD.

(a) Compute an SVD $A = USV^{\mathrm{T}}$ with `[U,S,V]=svd(A)` (Table 3.3): here

```
U =
    0.4082   -0.7071    0.5774
   -0.4082   -0.7071   -0.5774
   -0.8165   -0.0000    0.5774
S =
    1.7321         0         0
         0    1.7321         0
         0         0    0.0000
V =
    0.0000   -0.8165    0.5774
   -0.7071    0.4082    0.5774
    0.7071    0.4082    0.5774
```

so the singular values are $\sigma_1 = \sigma_2 = 1.7321 = \sqrt{3}$ and $\sigma_3 = 0$ (different computers may give different U and V, but any deductions will be equivalent). The system of equations for the ratings becomes

$$A\boldsymbol{x} = U \underbrace{S \overbrace{V^{\mathrm{T}}\boldsymbol{x}}^{=\boldsymbol{y}}}_{=\boldsymbol{z}} = \boldsymbol{b} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}.$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(b) As U is orthogonal, $U\boldsymbol{z} = \boldsymbol{b}$ has unique solution $\boldsymbol{z} = U^{\mathrm{T}}\boldsymbol{b}$ computed by `z=U'*b`:

```
z =
   -1.2247
   -2.1213
         0
```

(c) Now solve $S\boldsymbol{y} = \boldsymbol{z}$. But S has a troublesome zero on the diagonal. So interpret the equation $S\boldsymbol{y} = \boldsymbol{z}$ in detail as

$$\begin{bmatrix} 1.7321 & 0 & 0 \\ 0 & 1.7321 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{y} = \begin{bmatrix} -1.2247 \\ -2.1213 \\ 0 \end{bmatrix} :$$

i. the first line implies $y_1 = -1.2247/1.7321$;

ii. the second line implies $y_2 = -2.1213/1.7321$;

iii. the third line is $0y_3 = 0$ which is satisfied for all y_3.

In using MATLAB/Octave you must notice $\sigma_3 = 0$, check that the corresponding $z_3 = 0$, and then compute a *particular solution* from the first two components to give the first two components of $\boldsymbol{y}$:

```
y=z(1:2)./diag(S(1:2,1:2))
y =
   -0.7071
   -1.2247
```

The third component, involving the free variable y_3, we omit from this numerical computation.

(d) Finally, as V is orthogonal, $V^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{y}$ has the solution $\boldsymbol{x} = V\boldsymbol{y}$ (unique for each valid $\boldsymbol{y}$): in MATLAB/Octave, compute a particular solution with `x=V(:,1:2)*y`

```
x =
    1.0000
    0.0000
   -1.0000
```

Then for a general solution remember to add an arbitrary multiple, y_3, of `V(:,3)` $= (0.5774, 0.5774, 0.5774) = (1,1,1)/\sqrt{3}$.

Thus the three player ratings may be any one from the general solution

$$(x_1, x_2, x_3) = (1, 0, -1) + y_3(1,1,1)/\sqrt{3}\,.$$

In this application we only care about relative ratings, not absolute ratings, so here adding any multiple of $(1,1,1)$ is immaterial. This solution for the ratings indicates Anne is the best player, and Chris the worst.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

■

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

- `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?

Solution: The only change to the problem is the new right-hand side $\boldsymbol{b} = (1, 2, 2)$.

(a) An SVD of matrix A remains the same.

(b) $U\boldsymbol{z} = \boldsymbol{b}$ has unique solution `z=U'*b` of

```
z =
   -2.0412
   -2.1213
    0.5774
```

(c) We need to interpret $S\boldsymbol{y} = \boldsymbol{z}$,

$$\begin{bmatrix} 1.7321 & 0 & 0 \\ 0 & 1.7321 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{y} = \begin{bmatrix} -2.0412 \\ -2.1213 \\ 0.5774 \end{bmatrix}.$$

The third line of this system says $0y_3 = 0.5774$ which is impossible for any y_3.

In this case there is no solution of the system of equations. It would appear we cannot assign ratings to the players!

■

Section 3.5 further explores systems with no solution and uses the SVD to determine a good approximate solution (Example 3.5.3).

Example 3.3.14. Find the value(s) of the parameter c such that the following system has a solution, and find a general solution for that (those) parameter value(s):

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

Solution: Because the matrix is not square, we cannot use Procedure 2.2.5: instead use an SVD.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(a) In MATLAB/Octave, compute an SVD of this 3×4 matrix with

```
A=[-9 -15 -9 -15; -10 2 -10 2; 8 4 8 4]
[U,S,V]=svd(A)
U =
    0.8571    0.4286    0.2857
    0.2857   -0.8571    0.4286
   -0.4286    0.2857    0.8571
S =
   28.0000         0         0         0
         0   14.0000         0         0
         0         0    0.0000         0
V =
   -0.5000    0.5000   -0.1900   -0.6811
   -0.5000   -0.5000    0.6811   -0.1900
   -0.5000    0.5000    0.1900    0.6811
   -0.5000   -0.5000   -0.6811    0.1900
```

Depending upon MATLAB/Octave you may get different alternatives for the last two columns for `V`, and different signs for columns of `U` and `V`—adjust accordingly.

The singular values are $\sigma_1 = 28$, $\sigma_2 = 14$ and the problematic $\sigma_3 = 0$ (it is computed as the negligible 10^{-15}).

(b) To solve $U\boldsymbol{z} = \boldsymbol{b}$ we compute $\boldsymbol{z} = U^{\mathrm{T}}\boldsymbol{b}$. But for the next step we must have the third component of $\boldsymbol{z}$ to be zero as otherwise there is no solution. Now $z_3 = \boldsymbol{u}_3^{\mathrm{T}}\boldsymbol{b}$ (where $\boldsymbol{u}_3$ is the third column of U); that is, $z_3 = 0.2857 \times c + 0.4286 \times 8 + 0.8571 \times (-5)$ needs to be zero, which requires $c = -(0.4286 \times 8 + 0.8571 \times (-5))/0.2857$. Recognise this expression is equivalent to $c = -(0.2857 \times 0 + 0.4286 \times 8 + 0.8571 \times (-5))/0.2857 = \boldsymbol{u}_3 \cdot (0, 8, -5)/0.2857$ and so compute

```
c=-U(:,3)'*[0;8;-5]/U(1,3)
```

Having found $c = 3$, compute $\boldsymbol{z}$ from `z=U'*[3;8;-5]` to find $\boldsymbol{z} = (7, -7, 0)$.

(c) Find a general solution of the diagonal system $S\boldsymbol{y} = \boldsymbol{z}$:

$$\begin{bmatrix} 28 & 0 & 0 & 0 \\ 0 & 14 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \boldsymbol{y} = \begin{bmatrix} 7 \\ -7 \\ 0 \end{bmatrix}.$$

The first line gives $y_1 = 7/28 = 1/4$, the second line gives $y_2 = -7/14 = -1/2$, and the third line is $0y_3 + 0y_4 = 0$ which is satisfied for all y_3 and y_4 (because we chose c correctly). Thus $\boldsymbol{y} = (\frac{1}{4}, -\frac{1}{2}, y_3, y_4)$ is a general solution for this intermediary. Compute the particular solution with $y_3 = y_4 = 0$ via

```
y=z(1:2)./diag(S(1:2,1:2))
```

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(d) Finally solve $V^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{y}$ as $\boldsymbol{x} = V\boldsymbol{y}$, namely

$$\boldsymbol{x} = \begin{bmatrix} -0.5 & 0.5 & -0.1900 & -0.6811 \\ -0.5 & -0.5 & 0.6811 & -0.1900 \\ -0.5 & 0.5 & 0.1900 & 0.6811 \\ -0.5 & -0.5 & -0.6811 & 0.1900 \end{bmatrix} \begin{bmatrix} 1/4 \\ -1/2 \\ y_3 \\ y_4 \end{bmatrix}$$

Obtain a particular solution with `x=V(:,1:2)*y` of $\boldsymbol{x} = (-3, 1, -3, 1)/8$, and then add the free components:

$$\boldsymbol{x} = \begin{bmatrix} -\frac{3}{8} \\ \frac{1}{8} \\ -\frac{3}{8} \\ \frac{1}{8} \end{bmatrix} + \begin{bmatrix} -0.1900 \\ 0.6811 \\ 0.1900 \\ -0.6811 \end{bmatrix} y_3 + \begin{bmatrix} -0.6811 \\ -0.1900 \\ 0.6811 \\ 0.1900 \end{bmatrix} y_4 \,.$$

■

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 *factorisation $A = USV^{\mathrm{T}}$.*
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.* [12] *Identify the non-zero 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 $i = r+1, \ldots, m$, then there is no solution (the equations are **inconsistent**);*
   - *otherwise (when $z_i = 0$ for all $i = r+1, \ldots, m$) determine the ith component of $\boldsymbol{y}$ by $y_i = z_i/\sigma_i$ for $i = 1, \ldots, r$ (for which $\sigma_i > 0$), and let y_i be a free variable for $i = r+1, \ldots, n$.*
4. *Solve $V^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{y}$ (unique given V and for each $\boldsymbol{y}$) to obtain a general solution as $\boldsymbol{x} = V\boldsymbol{y}$.*

Proof. Given an SVD $A = USV^{\mathrm{T}}$ (Theorem 3.3.6), consider each and every solution of $A\boldsymbol{x} = \boldsymbol{b}$:

$$\begin{aligned} A\boldsymbol{x} = \boldsymbol{b} &\iff USV^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{b} \quad \text{(by step 1)} \\ &\iff S(V^{\mathrm{T}}\boldsymbol{x}) = U^{\mathrm{T}}\boldsymbol{b} \\ &\iff S\boldsymbol{y} = \boldsymbol{z} \quad \text{(by steps 2 and 4)}, \end{aligned}$$

and step 3 determines all possible $\boldsymbol{y}$ satisfying $S\boldsymbol{y} = \boldsymbol{z}$. Hence Procedure 3.3.15 determines all possible solutions of $A\boldsymbol{x} = \boldsymbol{b}$.[13] □

[12] Being diagonal, S is in a special row reduced echelon form (Definition 2.2.20).
[13] Any non-uniqueness in the orthogonal U and V just gives rise to equivalent different algebraic expressions for the set of possibilities.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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 no or infinite solutions.

Condition number and rank determine possibilities

> The expression 'ill-conditioned' is sometimes used merely as a term of abuse . . . It is characteristic of ill-conditioned sets of equations that small percentage errors in the coefficients given may lead to large percentage errors in the solution. *Alan Turing, 1934 (Higham 1996, p.131)*

The MATLAB/Octave function `rcond()` roughly estimates the reciprocal of what is called the condition number (estimates it to within a factor of two or three).

Definition 3.3.16. *For any $m \times n$ matrix A, the* ***condition number*** *is the 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* $\operatorname{cond} A = \infty$ *; 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 $\sigma_1/\sigma_2 = (10\sqrt{2})/(5\sqrt{2}) = 2$ (`rcond` $= 0.5$); for the 3×3 matrix the condition number $\sigma_1/\sigma_3 = 12/3 = 4$ (`rcond` $= 0.25$). Example 3.3.8 comments that every $n \times n$ orthogonal matrix has singular values $\sigma_1 = \cdots = \sigma_n = 1$; hence an orthogonal matrix has condition number one (`rcond` $= 1$). Such small condition numbers (non-small `rcond`) indicate all orthogonal matrices are "good" matrices (as classified by Procedure 2.2.5).

However, the matrix in the sports ranking Example 3.3.12 has singular values $\sigma_1 = \sigma_2 = \sqrt{3}$ and $\sigma_3 = 0$ so its condition number $\sigma_1/\sigma_3 = \sqrt{3}/0 = \infty$ which suggests the equations are likely to be unsolvable. (In MATLAB/Octave, see that $\sigma_3 = 2\cdot 10^{-17}$ so a numerical calculation would give condition number $1.7321/\sigma_3 = 7 \cdot 10^{16}$ which is effectively infinite.) ■

Activity 3.3.18. What is the condition number of the matrix of Example 3.3.14,

$$\begin{bmatrix} -9 & -15 & -9 & -15 \\ -10 & 2 & -10 & 2 \\ 8 & 4 & 8 & 4 \end{bmatrix},$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

given it has an SVD (2 d.p.)

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

(a) 0 (b) 0.5 (c) 2 (d) ∞

■

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 $< 10^{-8}\sigma_1$ are effectively zero.

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

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

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 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.)

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

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

■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Example 3.3.22. Use MATLAB/Octave to find the ranks of the two matrices

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

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

Solution: (a) Enter the matrix into MATLAB/Octave and compute its singular values with `svd(A)`: [14]

```
A=[0 1 0
 1 1 -1
 1 0 -1
 2 0 -2 ]
svd(A)
```

The singular values are 3.49, 1.34 and $1.55 \cdot 10^{-16} \approx 0$ (2 d.p.). Since two singular values are nonzero, the rank of the matrix is two.

(b) Enter the matrix into MATLAB/Octave and compute its singular values with `svd(A)`:

```
A=[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 ]
svd(A)
```

The singular values are 5.58, 4.17, 3.13, 1.63 and $2.99 \cdot 10^{-16} \approx 0$ (2 d.p.). Since four singular values are nonzero, the rank of the matrix is four.

■

Theorem 3.3.23. *For any 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,* $\operatorname{rank}(A^{\mathrm{T}}) = \operatorname{rank} A$.

Proof. Let $m \times n$ matrix A have SVD USV^{T}. Using the properties of the matrix transpose (Theorem 3.1.27),

$$A^{\mathrm{T}} = (USV^{\mathrm{T}})^{\mathrm{T}} = (V^{\mathrm{T}})^{\mathrm{T}} S^{\mathrm{T}} U^{\mathrm{T}} = V(S^{\mathrm{T}})U^{\mathrm{T}}$$

[14] Some advanced students will know that MATLAB/Octave provides the `rank()` function to directly compute the rank. However, this example is to reinforce its meaning in terms of singular values.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

which is an SVD for A^{T} since U and V are orthogonal, and S^{T} has the necessary diagonal structure. Since the number of non-zero values along the diagonal of S^{T} is precisely the same as that of the diagonal of S, $\operatorname{rank}(A^{\mathrm{T}}) = \operatorname{rank} A$. □

Example 3.3.24. From earlier examples, write down an SVD of the matrices

$$\begin{bmatrix} 10 & 5 \\ 2 & 11 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -4 & -8 & 6 \\ -2 & -1 & 6 \\ 4 & -4 & 0 \end{bmatrix}.$$

Solution: These matrices are the transpose of the two matrices whose SVDs are given in Example 3.3.7. Hence their SVDs are the transpose of the SVDs in that example (remembering that the transpose of a product is the product of the transpose but in reverse order):

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

and

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

■

Activity 3.3.25. Recall the matrix of Example 3.3.14,

$$\begin{bmatrix} -9 & -15 & -9 & -15 \\ -10 & 2 & -10 & 2 \\ 8 & 4 & 8 & 4 \end{bmatrix},$$

has an SVD (2 d.p.)

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

Which of the following is an SVD of the transpose of the matrix?

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

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

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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

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

■

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.

Theorem 3.3.26 (Unique Solutions: version 1)**.** *Let A be an $n \times n$ square matrix. 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) $\operatorname{rank} A = n$ *.*

Proof. Prove a circular chain of implications.

3.3.26a $\implies$ 3.3.26b : Established by Theorem 3.2.10.

3.3.26b $\implies$ 3.3.26c : Now $\boldsymbol{x} = \boldsymbol{0}$ is always a solution of $A\boldsymbol{x} = \boldsymbol{0}$. If property 3.3.26b holds, then this is the only solution.

3.3.26c $\implies$ 3.3.26d : Use contradiction. Assume a singular value is zero. Then Procedure 3.3.15 finds an infinite number of solutions to the homogeneous system $A\boldsymbol{x} = \boldsymbol{0}$, which contradicts 3.3.26c. Hence the assumption is wrong.

3.3.26d $\implies$ 3.3.26e : Property 3.3.26e is direct from Definition 3.3.19.

3.3.26e $\implies$ 3.3.26a : Find an SVD $A = USV^{\mathrm{T}}$. The inverse of the $n \times n$ diagonal matrix S exists as its diagonal elements are the n non-zero singular values (Theorem 3.2.27). Let $B = VS^{-1}U^{\mathrm{T}}$. Then $AB = USV^{\mathrm{T}}VS^{-1}U^{\mathrm{T}} = USS^{-1}U^{\mathrm{T}} = UU^{\mathrm{T}} = I_n$. Similarly $BA = I_n$. From Definition 3.2.2, A is invertible (with $B = VS^{-1}U^{\mathrm{T}}$ as its inverse).

□

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Practical shades of grey The Unique Solution Theorem 3.3.26 is 'black-and-white': either a solution exists, or it does not. 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?

Optional: this discussion and theorem reinforces why we must check condition numbers in computation.

Example 3.3.27. Consider the following innocuous looking system of linear equations

$$\begin{cases} -2q + r = 3 \\ p - 5q + r = 8 \\ -3p + 2q + 3r = -5 \end{cases}$$

Solve by hand (Procedure 2.2.24) to find the unique solution is $(p, q, r) = (2, -1, 1)$.

But, and it is a big but in practical applications, what happens if the right-hand side comes from experimental measurements with a relative error of 1%? Let's explore by writing the system in matrix-vector form and using MATLAB/Octave to solve with various example errors.

(a) First solve the system as stated. Denoting the unknowns by vector $\boldsymbol{x} = (p, q, r)$, 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;

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

(b) Now recognise 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 $\boldsymbol{x} = (2, -1, 1)$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

ii. `relerr1=norm(x-x1)/norm(x)` computes its relative error $|\boldsymbol{x}-\boldsymbol{x}'|/|\boldsymbol{x}|$ to be 0.91—rather large.

As illustrated below, the large difference between $\boldsymbol{x}$ and $\boldsymbol{x}'$ indicates '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 3.3.29 below shows it is no accident that the magnification of the error by a factor of 91 is of the same order of magnitude as the condition number $= 152.27$ computed via `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 below. Compute its relative error $|\boldsymbol{x}-\boldsymbol{x}''|/|\boldsymbol{x}| = 0.43$. At 43%, the relative error in solution $\boldsymbol{x}''$ is still 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 is 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 still significantly larger than the 1% error in $\boldsymbol{b}$.

This example shows that the apparently innocuous matrix A variously multiples measurement errors in $\boldsymbol{b}$ by factors of 91, 41 or 15 when finding 'the solution' $\boldsymbol{x}$ to $A\boldsymbol{x} = \boldsymbol{b}$. The matrix A must, after all, be a bad matrix. Theorem 3.3.29 shows this badness is quantified by its condition number 152.27, and its estimated reciprocal `rcond`. ■

v0.2

Example 3.3.28. Consider solving the system of linear equations

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

Use MATLAB/Octave to explore the effect on the solution $\boldsymbol{x}$ of 1% errors in the right-hand side vector.

Solution: Enter the matrix and right-hand side vector into MATLAB/Octave, then solve with Procedure 2.2.5:

```
Q=[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]
b=[-3;3;-9;-1]
rcond(Q)
x=Q\b
```

to find the solution $\boldsymbol{x} = (-4.6, 5, 7, -2.2)$.

Now see the effect on this solution of 1% errors in $\boldsymbol{b}$. Since the length $|\boldsymbol{b}| =$ `norm(b)` $= 10$ we find the solution for various changes to $\boldsymbol{b}$ of size 0.1.

- For example, adding the 1% error $(0.1, 0, 0, 0)$ to $\boldsymbol{b}$, the MATLAB/Octave commands

  ```
  x1=Q\(b+[0.1;0;0;0])
  relerr1=norm(x-x1)/norm(x)
  ```

  show the changed solution is $\boldsymbol{x}' = (-4.56, 5.04, 6.98, -2.12)$ which here is reasonably close to $\boldsymbol{x}$. Indeed its relative error $|\boldsymbol{x} - \boldsymbol{x}'|/|\boldsymbol{x}|$ is computed to be $0.0100 = 1\%$. Here the relative error in solution $\boldsymbol{x}$ is exactly the same as the relative error in $\boldsymbol{b}$.

- Exploring further, upon adding the 1% error $(0, 0.1, 0, 0)$ to $\boldsymbol{b}$, analogous commands show the changed solution is $\boldsymbol{x}'' = (-4.62, 5.08, 6.96, -2.24)$ which has relative error $|\boldsymbol{x} - \boldsymbol{x}''|/|\boldsymbol{x}| = 0.0100 = 1\%$ again.

- Whereas, upon adding 1% error $(0, 0, 0.1, 0)$ to $\boldsymbol{b}$, analogous commands show the changed solution is $\boldsymbol{x}''' = (-4.56, 4.96, 6.92, -2.22)$ which has relative error $|\boldsymbol{x} - \boldsymbol{x}'''|/|\boldsymbol{x}| = 0.0100 = 1\%$ again.

- Lastly, upon adding 1% error $(0, 0, 0, 0.1)$ to $\boldsymbol{b}$, analogous commands show the changed solution is $\boldsymbol{x}'''' = (-4.68, 4.98, 6.96, -2.16)$ which has relative error $|\boldsymbol{x} - \boldsymbol{x}''''|/|\boldsymbol{x}| = 0.0100 = 1\%$ yet again.

In this example, and in contrast to the previous example, throughout the relative error in solution $\boldsymbol{x}$ is exactly the same as the relative error in $\boldsymbol{b}$. The reason is that here the matrix Q is an orthogonal

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

matrix—check by computing `Q'*Q` (Definition 3.2.43). Being orthogonal, multiplication by Q only rotates or reflects, and so it never stretches or distorts (Theorem 3.2.48c). Consequently errors remain the same size when multiplied by such orthogonal matrices, as seen in this example, and as reflected in the condition number of Q being one (as computed via `s=svd(Q)` and then `condQ=s(1)/s(4)`).

■

The condition number determines the reliability of the solution of a system of linear equations. This is why we should always precede the computation of a solution with an estimate of the condition number such as that provided by the reciprocal `rcond()` (Procedure 2.2.5). The next theorem establishes that the condition number characterises the amplification of errors that occurs in solving a linear system. Hence solving a system of linear equations with a large condition number (small `rcond`) means that errors are amplified by a large factor as happens in Example 3.3.27.

The greek letter ϵ (epsilon) often denotes errors.

Theorem 3.3.29 (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 size ϵ, then the solution $\boldsymbol{x}$ has relative error $\leq \epsilon \operatorname{cond} A$, with equality in the worst case.*

Proof. Let the length of the right-hand side vector be $b = |\boldsymbol{b}|$. Then the error in $\boldsymbol{b}$ has size ϵb since ϵ is the relative error. Following Procedure 3.3.15, let $A = USV^{\mathrm{T}}$ be an SVD for matrix A. Compute $\boldsymbol{z} = U^{\mathrm{T}}\boldsymbol{b}$: recall that multiplication by orthogonal U preserves lengths (Theorem 3.2.48), so not only is $|\boldsymbol{z}| = b$, but also $\boldsymbol{z}$ will be in error by an amount ϵb since $\boldsymbol{b}$ has this error Consider solving $S\boldsymbol{y} = \boldsymbol{z}$: the diagonals of S stretch and shrink both 'the signal and the noise'. The *worst case* is when $\boldsymbol{z} = (b\,,0\,,\dots,0\,,\epsilon b)$; that is, when all the 'signal' happens to be in the first component of $\boldsymbol{z}$, and all the 'noise', the error, is in the last component. Then the intermediary $\boldsymbol{y} = (b/\sigma_1\,,0\,,\dots,\epsilon b/\sigma_n)$. Consequently, the intermediary has relative error $(\epsilon b/\sigma_n)/(b/\sigma_1) = \epsilon(\sigma_1/\sigma_n) = \epsilon \operatorname{cond} A$. Again because multiplication by orthogonal V preserves lengths, the solution $\boldsymbol{x} = V\boldsymbol{y}$ has the same relative error: in the worst case of $\epsilon \operatorname{cond} A$. □

Example 3.3.30. 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.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(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)` $= 0.002$.

Solution: (a) The relative error in $\boldsymbol{x}$ could be as big as $0.001 \times 10 = 0.01$.

(b) Measuring to three significant digits means the relative error is 0.0005, while with `rcond(A)` $= 0.025$, matrix A has condition number of roughly 40, so the relative error of $\boldsymbol{x}$ is less than $0.0005 \times 40 = 0.02$; that is, up to 2%.

(c) There is not enough information as we cannot determine the *relative* error in measurements $\boldsymbol{b}$.

(d) Two significant digits means the relative error is 0.005, while matrix A has condition number of roughly $1/0.002 = 500$ so the relative error of $\boldsymbol{x}$ could be as big as $0.005 \times 500 = 2.5$; that is, the estimate $\boldsymbol{x}$ is possibly complete rubbish.

■

Activity 3.3.31. 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 $A\boldsymbol{x} = \boldsymbol{b}$ with a matrix A for which we compute `rcond(A)` $= 0.02$. What is our estimate of the relative error in $\boldsymbol{x}$?

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

■

This issue of the amplification of errors occurs in other contexts. The eminent mathematician Henri Poincaré (1854–1912) was the first to detect possible chaos in the orbits of the planets.

> If we knew exactly the laws of nature and the situation of the universe at the initial moment, we could predict exactly the situation of that same universe at a succeeding moment. But even if it were the case that the natural laws had no longer any secret for us, we could still only know the initial situation approximately. If that enabled us to predict the succeeding situation with the same approximation, that is all we require, and we should say that the phenomenon had been predicted, that it is governed by laws. But it is not always so; it may happen that small differences in the initial conditions produce very great ones in the final phenomena.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

> 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 $\boldsymbol{b}$ will produce an enormous error in the final $\boldsymbol{x}$. The condition number warns when this happens by characterising the amplification.

3.3.3 Prove the SVD Theorem 3.3.6

> When doing maths there's this great feeling. You start with a problem that just mystifies you. You can't understand it, it's so complicated, you just can't make head nor tail of it. But then when you finally resolve it, you have this incredible feeling of how beautiful it is, how it all fits together so elegantly. *Andrew Wiles, C1993*

This proof may be delayed until the last week of a semester. It may be given together with the closely related classic proof of Theorem 4.2.15 on the eigenvectors of symmetric matrices.

Two preliminary examples introduce the structure of the general proof that an SVD exists. As in this example prelude, the proof of a general singular value decomposition is similarly constructive.

Prelude to the proof

These first two examples are optional: their purpose is to introduce key parts of the general proof in a definite setting.

Example 3.3.32 (a 2×2 case)**.** Recall Example 3.3.2 factorised the matrix

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

Find this factorisation, $A = USV^{\mathrm{T}}$, by maximising $|A\boldsymbol{v}|$ over all unit vectors $\boldsymbol{v}$.

Solution: In 2D, all unit vectors are of the form $\boldsymbol{v} = (\cos t, \sin t)$ for $-\pi < t \le \pi$. The marginal picture plots these unit vectors $\boldsymbol{v}$ in blue for 32 angles t. Plotted in red from the end of each $\boldsymbol{v}$ is the vector $A\boldsymbol{v}$ (scaled down by a factor of ten for clarity). Our aim is to find the $\boldsymbol{v}$ that maximises the length of the corresponding adjoined $A\boldsymbol{v}$. By inspection, the longest red vectors $A\boldsymbol{v}$ occur towards the top-right or the bottom-left, either of these directions $\boldsymbol{v}$ are what we first find.

The MATLAB function `eigshow(A)` provides an interactive alternative to this static view (click on the `eig/(svd)` button to make it show `svd/(eig)`.

Maximising $|A\boldsymbol{v}|$ is the same as maximising $|A\boldsymbol{v}|^2$ which is what the following considers: since

$$A\boldsymbol{v} = \begin{bmatrix} 10 & 2 \\ 5 & 11 \end{bmatrix} \begin{bmatrix} \cos t \\ \sin t \end{bmatrix} = \begin{bmatrix} 10\cos t + 2\sin t \\ 5\cos t + 11\sin t \end{bmatrix} ,$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$
\begin{aligned}
|A\boldsymbol{v}|^2 &= (10\cos t + 2\sin t)^2 + (5\cos t + 11\sin t)^2 \\
&= 100\cos^2 t + 40\cos t\,\sin t + 4\sin^2 t \\
&\quad + 25\cos^2 t + 110\cos t\,\sin t + 121\sin^2 t \\
&= 125(\cos^2 t + \sin^2 t) + 150\sin t\,\cos t \\
&= 125 + 75\sin 2t \qquad \text{(shown in the margin).}
\end{aligned}
$$

Since the sine function has maximum of one at angle $\frac{\pi}{2}$, the maximum of $|A\boldsymbol{v}|^2$ is $125 + 75 = 200$ for $2t = \frac{\pi}{2}$, that is, for $t = \frac{\pi}{4}$ corresponding to unit vector $\boldsymbol{v}_1 = (\cos\frac{\pi}{4}, \sin\frac{\pi}{4}) = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$—this vector point to the top-right as identified from the marginal figure. This vector is the first column of V.

Now multiply to find $A\boldsymbol{v}_1 = (6\sqrt{2}, 8\sqrt{2})$. The length of this vector is $\sqrt{72+128} = \sqrt{200} = 10\sqrt{2} = \sigma_1$ the leading singular value. Normalise the vector $A\boldsymbol{v}_1$ by $A\boldsymbol{v}_1/\sigma_1 = (6\sqrt{2}, 8\sqrt{2})/(10\sqrt{2}) = (\frac{3}{5}, \frac{4}{5}) = \boldsymbol{u}_1$, the first column of U.

The other column of V must be orthogonal (at right-angles) to $\boldsymbol{v}_1$ in order for matrix V to be orthogonal. Thus set $\boldsymbol{v}_2 = (-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ as shown in the marginal graph. Now multiply to find $A\boldsymbol{v}_2 = (-4\sqrt{2}, 3\sqrt{2})$: magically, and a crucial part of the general proof, this vector is orthogonal to $\boldsymbol{u}_1$. The length of $A\boldsymbol{v}_2 = (-4\sqrt{2}, 3\sqrt{2})$ is $\sqrt{32+18} = \sqrt{50} = 5\sqrt{2} = \sigma_2$ the other singular value. Normalise the vector to $A\boldsymbol{v}_2/\sigma_2 = (-4\sqrt{2}, 3\sqrt{2})/(2\sqrt{2}) = (-\frac{4}{5}, \frac{3}{5}) = \boldsymbol{u}_2$, the second column of U.

This construction establishes that here $AV = US$. Then post-multiply each side by V^{T} to find an SVD is $A = USV^{\mathrm{T}}$.

In this example we could have chosen the negative of $\boldsymbol{v}_1$ (angle $t = -\frac{3\pi}{4}$), and/or chosen the negative of $\boldsymbol{v}_2$. The result would still be a valid SVD of the matrix A. The orthogonal matrices in an SVD are not unique, and need not be. The singular values are unique.

■

Example 3.3.33 (a 3×1 case)**.** Find the following SVD for the 3×1 matrix

$$
A = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{3}} & \cdot & \cdot \\ \frac{1}{\sqrt{3}} & \cdot & \cdot \\ \frac{1}{\sqrt{3}} & \cdot & \cdot \end{bmatrix} \begin{bmatrix} \sqrt{3} \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \end{bmatrix}^{\mathrm{T}} = USV^{\mathrm{T}},
$$

where we do not worry about the elements denoted by dots as they are multiplied by the zeros in $S = (\sqrt{3}, 0, 0)$.

Solution: We seek to maximise $|A\boldsymbol{v}|^2$ but here vector $\boldsymbol{v}$ is in $\mathbb{R}^1$. Being of unit magnitude, there are two alternatives: $\boldsymbol{v} = (\pm 1)$. Each alternative gives the same $|A\boldsymbol{v}|^2 = |(\pm 1, \pm 1, \pm 1)| = 3$. Choose one alternative, say $\boldsymbol{v}_1 = (1)$, fixes the matrix $V = \begin{bmatrix} 1 \end{bmatrix}$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Then $A\boldsymbol{v}_1 = (1,1,1)$ which is of length $\sqrt{3}$. This length is the singular value $\sigma_1 = \sqrt{3}$. Dividing $A\boldsymbol{v}_1$ by its length gives the unit vector $\boldsymbol{u}_1 = (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}})$, the first column of U. To find the other columns of U, consider the three standard unit vectors in $\mathbb{R}^3$ (red in the illustration below), rotate them all together so that one lines up with $\boldsymbol{u}_1$, and then the other two rotated unit vectors form the other two columns of U (blue vectors below). Since the columns of U are then orthonormal, U is an orthogonal matrix (Theorem 3.2.48).

■

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

- First zero matrices have trivial SVD, and $m \times 1$ and $1 \times n$ matrices have straightforward SVD (as in Example 3.3.33).
- Choose $\boldsymbol{v}_1$ to maximise $|A\boldsymbol{v}|^2$ for 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 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 4.2.15 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 $m \le n$, one for $m \ge n$, 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 $USV^{\mathrm{T}} = I_m O_{m\times n} I_n^{\mathrm{T}} = O_{m\times n} = A$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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 $\boldsymbol{a}_1 = (a_{11}, a_{21}, \dots, a_{m1})$. 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 $S = (\sigma_1, 0, \dots, 0)$; 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

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

Case $1 \times n$ ($m = 1$): use an exactly complementary argument to the preceding $m \times 1$ case.

Induction Assume an SVD exists for all $(m-1) \times (n-1)$ matrices: we proceed to prove that consequently an SVD must exist for all $m \times n$ matrices. Consider any $m \times n$ nonzero matrix A for $m, n \geq 2$. Set vector $\boldsymbol{v}_1$ in $\mathbb{R}^n$ to be a *unit vector* that maximises $|A\boldsymbol{v}|^2$ 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.* Proved in higher level analysis.

   As matrix A is nonzero, there exists $\boldsymbol{v}$ such that $|A\boldsymbol{v}| > 0$. Since $\boldsymbol{v}_1$ maximises $|A\boldsymbol{v}|$ 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 $\mathbb{R}^m$. Let $\boldsymbol{v}$ be any unit vector orthogonal to $\boldsymbol{v}_1$: we prove that then 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 (Exercise 3.3.16), 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 $f'(0) = \sigma_1 \boldsymbol{u} \cdot \boldsymbol{u}_1$. But from the 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 3.2.48b). 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

$$\begin{aligned} A_1 := U_1^{\mathrm{T}} A V_1 &= \begin{bmatrix} \boldsymbol{u}_1^{\mathrm{T}} \\ \bar{U}^{\mathrm{T}} \end{bmatrix} A \begin{bmatrix} \boldsymbol{v}_1 & \bar{V} \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{u}_1^{\mathrm{T}} A \boldsymbol{v}_1 & \boldsymbol{u}_1^{\mathrm{T}} A \bar{V} \\ \bar{U}^{\mathrm{T}} A \boldsymbol{v}_1 & \bar{U}^{\mathrm{T}} A \bar{V} \end{bmatrix} \end{aligned}$$

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 $A_1 := U_1^{\mathrm{T}} A V_1$ gives $A V_1 = U_1 A_1$.

4. *By induction assumption,* $(m-1) \times (n-1)$ *matrix* B *has an* SVD*, and so we now construct an* SVD *for* $m \times n$ *matrix* A*.* Let $B = \hat{U} \hat{S} \hat{V}^{\mathrm{T}}$ be an SVD for B. Then construct

$$U := U_1 \begin{bmatrix} 1 & 0 \\ 0 & \hat{U} \end{bmatrix}, \quad V := V_1 \begin{bmatrix} 1 & 0 \\ 0 & \hat{V} \end{bmatrix}, \quad S := \begin{bmatrix} \sigma_1 & 0 \\ 0 & \hat{S} \end{bmatrix}.$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Matrices U and V are orthogonal as each are the product of two orthogonal matrices (Exercise 3.2.19), and matrix S is diagonal. These form an SVD for matrix A since

$$\begin{aligned}
AV &= AV_1\begin{bmatrix}1&0\\0&\hat V\end{bmatrix} = U_1A_1\begin{bmatrix}1&0\\0&\hat V\end{bmatrix}\\
&= U_1\begin{bmatrix}\sigma_1&0\\0&B\end{bmatrix}\begin{bmatrix}1&0\\0&\hat V\end{bmatrix} = U_1\begin{bmatrix}\sigma_1&0\\0&B\hat V\end{bmatrix}\\
&= U_1\begin{bmatrix}\sigma_1&0\\0&\hat U\hat S\end{bmatrix} = U_1\begin{bmatrix}1&0\\0&\hat U\end{bmatrix}\begin{bmatrix}\sigma_1&0\\0&\hat S\end{bmatrix}\\
&= US.
\end{aligned}$$

Hence $A = USV^{\mathrm{T}}$. By induction, an SVD exists for all $m \times n$ matrices.

This argument establishes the SVD Theorem 3.3.6.

3.3.4 Exercises

Exercise 3.3.1. Using a factorisation of the left-hand side coefficient, quickly solve by hand the following equations.

(a) $18x = 1134$
(b) $42x = 2226$
(c) $66x = 3234$
(d) $70x = 3150$
(e) $99x = 8118$
(f) $154x = 7854$
(g) $175x = 14350$
(h) $242x = 20086$
(i) $245x = 12495$
(j) $363x = 25047$
(k) $385x = 15785$
(l) $539x = 28028$

Exercise 3.3.2. Find a general solution, if a solution exists, of each of the following systems of linear equations using Procedure 3.3.15. Calculate by hand using the given SVD factorisation; record your working.

(a) $\underbrace{\begin{bmatrix}-\frac{9}{5}&\frac{12}{5}\\-4&-3\end{bmatrix}}_{=A}\boldsymbol{x} = \begin{bmatrix}-\frac{9}{5}\\\frac{17}{2}\end{bmatrix}$ given the SVD

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

(b) $\underbrace{\begin{bmatrix}\frac{15}{13}&\frac{36}{13}\\\frac{36}{13}&-\frac{15}{13}\end{bmatrix}}_{=B}\boldsymbol{x} = \begin{bmatrix}\frac{54}{13}\\-\frac{45}{26}\end{bmatrix}$ given the SVD

$$B = \begin{bmatrix}-\frac{12}{13}&\frac{5}{13}\\\frac{5}{13}&\frac{12}{13}\end{bmatrix}\begin{bmatrix}3&0\\0&3\end{bmatrix}\begin{bmatrix}0&1\\-1&0\end{bmatrix}^{\mathrm{T}}$$

v0.2a

(c) $\underbrace{\begin{bmatrix}-0.96 & 1.28\\ -0.72 & 0.96\end{bmatrix}}_{=C}\boldsymbol{x}=\begin{bmatrix}2.88\\ 2.16\end{bmatrix}$ given the SVD

$$C=\begin{bmatrix}\frac{4}{5} & \frac{3}{5}\\ \frac{3}{5} & -\frac{4}{5}\end{bmatrix}\begin{bmatrix}2 & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}-\frac{3}{5} & -\frac{4}{5}\\ \frac{4}{5} & -\frac{3}{5}\end{bmatrix}^{\mathrm{T}}$$

(d) $\underbrace{\begin{bmatrix}-\frac{5}{26} & -\frac{6}{13}\\ -\frac{12}{13} & \frac{5}{13}\end{bmatrix}}_{=D}\boldsymbol{x}=\begin{bmatrix}-\frac{7}{13}\\ \frac{34}{13}\end{bmatrix}$ given the SVD

$$D=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & \frac{1}{2}\end{bmatrix}\begin{bmatrix}-\frac{12}{13} & -\frac{5}{13}\\ \frac{5}{13} & -\frac{12}{13}\end{bmatrix}^{\mathrm{T}}$$

(e) $\underbrace{\begin{bmatrix}-\frac{2}{3} & \frac{23}{51} & \frac{22}{51}\\ \frac{1}{6} & \frac{7}{51} & -\frac{31}{51}\end{bmatrix}}_{=E}\boldsymbol{x}=\begin{bmatrix}-\frac{115}{102}\\ -\frac{35}{102}\end{bmatrix}$ given the SVD

$$E=\begin{bmatrix}\frac{15}{17} & -\frac{8}{17}\\ -\frac{8}{17} & -\frac{15}{17}\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\ 0 & \frac{1}{2} & 0\end{bmatrix}\begin{bmatrix}-\frac{2}{3} & \frac{1}{3} & -\frac{2}{3}\\ \frac{1}{3} & -\frac{2}{3} & -\frac{2}{3}\\ \frac{2}{3} & \frac{2}{3} & -\frac{1}{3}\end{bmatrix}^{\mathrm{T}}$$

(f) $\underbrace{\begin{bmatrix}\frac{3}{35} & \frac{9}{35} & \frac{9}{70}\\ -\frac{4}{35} & -\frac{12}{35} & -\frac{6}{35}\end{bmatrix}}_{=F}\boldsymbol{x}=\begin{bmatrix}\frac{3}{8}\\ -\frac{1}{2}\end{bmatrix}$ given the SVD

$$F=\begin{bmatrix}-\frac{3}{5} & \frac{4}{5}\\ \frac{4}{5} & \frac{3}{5}\end{bmatrix}\begin{bmatrix}\frac{1}{2} & 0 & 0\\ 0 & 0 & 0\end{bmatrix}\begin{bmatrix}-\frac{2}{7} & -\frac{6}{7} & \frac{3}{7}\\ -\frac{6}{7} & \frac{3}{7} & \frac{2}{7}\\ -\frac{3}{7} & -\frac{2}{7} & -\frac{6}{7}\end{bmatrix}^{\mathrm{T}}$$

(g) $\underbrace{\begin{bmatrix}\frac{7}{39} & -\frac{17}{39}\\ -\frac{22}{39} & -\frac{19}{39}\\ -\frac{4}{39} & -\frac{53}{78}\end{bmatrix}}_{=G}\boldsymbol{x}=\begin{bmatrix}-\frac{1}{3}\\ -\frac{2}{3}\\ -\frac{2}{3}\end{bmatrix}$ given the SVD

$$G=\begin{bmatrix}-\frac{1}{3} & \frac{2}{3} & \frac{2}{3}\\ -\frac{2}{3} & -\frac{2}{3} & \frac{1}{3}\\ -\frac{2}{3} & \frac{1}{3} & -\frac{2}{3}\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & \frac{1}{2}\\ 0 & 0\end{bmatrix}\begin{bmatrix} & \frac{5}{13} & \frac{12}{13}\\ \frac{12}{13} & -\frac{5}{13} & \end{bmatrix}^{\mathrm{T}}$$

v0.2a

(h) $\underbrace{\begin{bmatrix} \frac{36}{119} & -\frac{11}{17} \\ \frac{164}{119} & -\frac{18}{17} \\ -\frac{138}{119} & -\frac{6}{17} \end{bmatrix}}_{=H} \boldsymbol{x} = \begin{bmatrix} \frac{11}{17} \\ \frac{9}{17} \\ \frac{3}{17} \end{bmatrix}$ given the SVD

$$H = \begin{bmatrix} \frac{2}{7} & -\frac{3}{7} & -\frac{6}{7} \\ \frac{6}{7} & -\frac{2}{7} & \frac{3}{7} \\ -\frac{3}{7} & -\frac{6}{7} & \frac{2}{7} \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{15}{17} & \frac{8}{17} \\ -\frac{8}{17} & \frac{15}{17} \end{bmatrix}^{\mathrm{T}}$$

(i) $\underbrace{\begin{bmatrix} -\frac{17}{18} & -\frac{8}{9} & -\frac{8}{9} \\ 1 & \frac{2}{3} & -\frac{2}{3} \\ -\frac{11}{9} & \frac{8}{9} & -\frac{7}{9} \end{bmatrix}}_{=\mathcal{I}} \boldsymbol{x} = \begin{bmatrix} -\frac{17}{18} \\ \frac{5}{3} \\ -\frac{7}{18} \end{bmatrix}$ given the SVD

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

(j) $\underbrace{\begin{bmatrix} -\frac{10}{27} & -\frac{2}{27} & \frac{31}{54} \\ -\frac{4}{27} & \frac{10}{27} & \frac{17}{27} \\ -\frac{8}{27} & -\frac{7}{27} & \frac{7}{27} \end{bmatrix}}_{=J} \boldsymbol{x} = \begin{bmatrix} \frac{83}{54} \\ \frac{49}{54} \\ \frac{17}{54} \end{bmatrix}$ given the SVD

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

(k) $\underbrace{\begin{bmatrix} \frac{4}{33} & \frac{4}{11} & \frac{6}{11} \\ \frac{4}{33} & \frac{4}{11} & \frac{6}{11} \\ \frac{2}{33} & \frac{2}{11} & \frac{3}{11} \end{bmatrix}}_{=K} \boldsymbol{x} = \begin{bmatrix} -\frac{7}{3} \\ -\frac{7}{3} \\ -\frac{7}{6} \end{bmatrix}$ given the SVD

$$K = \begin{bmatrix} \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \\ \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{2}{11} & -\frac{9}{11} & \frac{6}{11} \\ \frac{6}{11} & \frac{6}{11} & \frac{7}{11} \\ \frac{9}{11} & -\frac{2}{11} & -\frac{6}{11} \end{bmatrix}^{\mathrm{T}}$$

v0.2a

(l) $\underbrace{\begin{bmatrix} -\frac{6}{11} & -\frac{1}{11} & \frac{81}{22} \\ \frac{7}{11} & \frac{3}{11} & \frac{27}{11} \\ -\frac{6}{11} & \frac{9}{22} & -\frac{9}{11} \end{bmatrix}}_{=L} \boldsymbol{x} = \begin{bmatrix} -\frac{35}{2} \\ -\frac{41}{4} \\ \frac{15}{8} \end{bmatrix}$ given the SVD

$$L = \begin{bmatrix} \frac{9}{11} & \frac{6}{11} & \frac{2}{11} \\ \frac{6}{11} & -\frac{7}{11} & -\frac{6}{11} \\ -\frac{2}{11} & \frac{6}{11} & -\frac{9}{11} \end{bmatrix} \begin{bmatrix} \frac{9}{2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ 1 & 0 & 0 \end{bmatrix}^{\mathrm{T}}$$

Exercise 3.3.3. Find a general solution, if a solution exists, of each of the following systems of linear equations. Calculate by hand using the given SVD factorisation; check the SVD and confirm your calculations with MATLAB/Octave (Procedure 3.3.15); then compare and contrast the two methods.

(a) $\underbrace{\begin{bmatrix} \frac{7}{180} & \frac{8}{45} & \frac{41}{180} \\ \frac{19}{180} & -\frac{22}{45} & \frac{101}{180} \\ -\frac{19}{180} & \frac{4}{45} & \frac{133}{180} \\ \frac{59}{60} & -\frac{2}{15} & \frac{91}{60} \end{bmatrix}}_{=A} \boldsymbol{x} = \begin{bmatrix} -\frac{13}{40} \\ -\frac{9}{8} \\ -\frac{17}{20} \\ -\frac{15}{4} \end{bmatrix}$ given the SVD

$$A = \begin{bmatrix} -\frac{1}{10} & \frac{3}{10} & \frac{3}{10} & -\frac{9}{10} \\ -\frac{3}{10} & -\frac{9}{10} & -\frac{1}{10} & -\frac{3}{10} \\ -\frac{3}{10} & -\frac{1}{10} & \frac{9}{10} & \frac{3}{10} \\ -\frac{9}{10} & \frac{3}{10} & -\frac{3}{10} & \frac{1}{10} \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -\frac{4}{9} & \frac{4}{9} & -\frac{7}{9} \\ \frac{1}{9} & \frac{8}{9} & \frac{4}{9} \\ -\frac{8}{9} & -\frac{1}{9} & \frac{4}{9} \end{bmatrix}^{\mathrm{T}}$$

(b) $\underbrace{\begin{bmatrix} \frac{79}{66} & \frac{7}{33} & -\frac{29}{33} \\ -\frac{65}{66} & -\frac{13}{33} & \frac{35}{33} \\ \frac{31}{66} & -\frac{29}{33} & -\frac{37}{33} \\ \frac{17}{66} & -\frac{23}{33} & -\frac{43}{33} \end{bmatrix}}_{=B} \boldsymbol{x} = \begin{bmatrix} -\frac{22}{6} \\ \frac{20}{6} \\ -\frac{1}{6} \\ \frac{1}{6} \end{bmatrix}$ given the SVD

$$B = \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{8}{3} & 0 & 0 \\ 0 & \frac{4}{3} & 0 \\ 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -\frac{6}{11} & -\frac{6}{11} & -\frac{7}{11} \\ \frac{2}{11} & -\frac{9}{11} & \frac{6}{11} \\ \frac{9}{11} & -\frac{2}{11} & -\frac{6}{11} \end{bmatrix}^{\mathrm{T}}$$

(c) $\underbrace{\begin{bmatrix} \frac{14}{15} & -\frac{14}{15} & \frac{7}{15} & -\frac{28}{15} \\ \frac{2}{5} & -\frac{8}{5} & -\frac{4}{5} & \frac{4}{5} \\ -\frac{6}{5} & -\frac{6}{5} & \frac{12}{5} & \frac{3}{5} \end{bmatrix}}_{=C} \boldsymbol{x} = \begin{bmatrix} 0 \\ 4 \\ 27 \end{bmatrix}$ given the SVD

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$C = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & \frac{7}{3} & 0 & 0 \\ 0 & 0 & 2 & 0 \end{bmatrix} \begin{bmatrix} \frac{2}{5} & -\frac{2}{5} & -\frac{1}{5} & \frac{4}{5} \\ \frac{2}{5} & \frac{2}{5} & \frac{4}{5} & \frac{1}{5} \\ -\frac{4}{5} & -\frac{1}{5} & \frac{2}{5} & \frac{2}{5} \\ -\frac{1}{5} & \frac{4}{5} & -\frac{2}{5} & \frac{2}{5} \end{bmatrix}^{\mathrm{T}}$$

(d) $$\underbrace{\begin{bmatrix} \frac{57}{22} & -\frac{3}{22} & -\frac{45}{22} & -\frac{9}{22} \\ -\frac{14}{11} & \frac{32}{11} & -\frac{4}{11} & -\frac{14}{11} \\ -\frac{9}{22} & -\frac{3}{22} & -\frac{45}{22} & \frac{57}{22} \end{bmatrix}}_{=D} \boldsymbol{x} = \begin{bmatrix} 117 \\ -72 \\ 63 \end{bmatrix} \text{ given the SVD}$$

$$D = \begin{bmatrix} -\frac{6}{11} & \frac{9}{11} & \frac{2}{11} \\ \frac{7}{11} & \frac{6}{11} & -\frac{6}{11} \\ -\frac{6}{11} & -\frac{2}{11} & -\frac{9}{11} \end{bmatrix} \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 3 & 0 \end{bmatrix} \begin{bmatrix} -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \end{bmatrix}^{\mathrm{T}}$$

(e) $$\underbrace{\begin{bmatrix} -\frac{2}{5} & -\frac{2}{5} & -\frac{26}{45} & \frac{26}{45} \\ \frac{11}{9} & \frac{11}{9} & -\frac{1}{3} & \frac{1}{3} \\ \frac{31}{90} & \frac{31}{90} & \frac{17}{90} & -\frac{17}{90} \\ \frac{4}{9} & \frac{4}{9} & -\frac{2}{9} & \frac{2}{9} \end{bmatrix}}_{=E} \boldsymbol{x} = \begin{bmatrix} 3 \\ -6 \\ -3 \\ -2 \end{bmatrix} \text{ given the SVD}$$

$$E = \begin{bmatrix} \frac{2}{9} & \frac{8}{9} & \frac{1}{3} & \frac{2}{9} \\ -\frac{8}{9} & \frac{2}{9} & -\frac{2}{9} & \frac{1}{3} \\ -\frac{2}{9} & -\frac{1}{3} & \frac{8}{9} & \frac{2}{9} \\ -\frac{1}{3} & \frac{2}{9} & \frac{2}{9} & -\frac{8}{9} \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -\frac{7}{10} & -\frac{1}{10} & \frac{1}{10} & -\frac{7}{10} \\ -\frac{7}{10} & -\frac{1}{10} & -\frac{1}{10} & \frac{7}{10} \\ \frac{1}{10} & -\frac{7}{10} & -\frac{7}{10} & -\frac{1}{10} \\ -\frac{1}{10} & \frac{7}{10} & -\frac{7}{10} & -\frac{1}{10} \end{bmatrix}^{\mathrm{T}}$$

(f) $$\underbrace{\begin{bmatrix} \frac{5}{14} & -\frac{41}{14} & \frac{1}{2} & -\frac{3}{2} \\ \frac{22}{7} & -\frac{4}{7} & 0 & 2 \\ -\frac{9}{14} & -\frac{13}{14} & \frac{5}{2} & -\frac{1}{2} \\ -\frac{2}{7} & -\frac{6}{7} & 2 & 2 \end{bmatrix}}_{=F} \boldsymbol{x} = \begin{bmatrix} -45 \\ -50 \\ 18 \\ 20 \end{bmatrix} \text{ given the SVD}$$

$$F = \begin{bmatrix} \frac{4}{7} & \frac{2}{7} & -\frac{5}{7} & \frac{2}{7} \\ -\frac{4}{7} & \frac{5}{7} & -\frac{2}{7} & -\frac{2}{7} \\ \frac{4}{7} & \frac{2}{7} & \frac{2}{7} & -\frac{5}{7} \\ \frac{1}{7} & \frac{4}{7} & \frac{4}{7} & \frac{4}{7} \end{bmatrix} \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix}^{\mathrm{T}}$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Exercise 3.3.4. Find a general solution, if possible, of each of the following systems of linear equations with MATLAB/Octave and using Procedure 3.3.15.

(a) $\begin{bmatrix} 2.4 & 1.6 & 1 & -0.8 \\ -1.2 & 3.2 & -2 & -0.4 \\ -1.2 & -0.8 & 2 & -1.6 \\ 0.6 & -1.6 & -4 & -0.8 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} -29.4 \\ -12.4 \\ 13.2 \\ -0.8 \end{bmatrix}$

(b) $\begin{bmatrix} -0.7 & -0.7 & -2.5 & -0.7 \\ 1 & -2.2 & 0.2 & -0.2 \\ -1 & 1.4 & -1.4 & -2.6 \\ 2.6 & -1.4 & -1 & -1.4 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} -4 \\ 2.4 \\ 3.2 \\ 0 \end{bmatrix}$

(c) $\begin{bmatrix} -3.14 & -1.18 & 0.46 & -0.58 \\ 0.66 & 0.18 & -0.06 & 2.22 \\ -1.78 & -2.54 & -1.82 & -5.26 \\ 0.58 & 1.06 & -0.82 & 0.26 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} -17.38 \\ -1.14 \\ 5.22 \\ 12.26 \end{bmatrix}$

(d) $\begin{bmatrix} 1.38 & 0.50 & 3.30 & 0.34 \\ -0.66 & -0.70 & 1.50 & -2.38 \\ -0.90 & 2.78 & -0.54 & 0.10 \\ 0.00 & 1.04 & -0.72 & -1.60 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} -7.64 \\ -7.72 \\ -20.72 \\ -20.56 \end{bmatrix}$

(e) $\begin{bmatrix} 1.32 & 1.40 & 1.24 & -0.20 \\ 1.24 & 3.00 & 2.68 & 1.00 \\ 1.90 & -1.06 & -1.70 & 2.58 \\ -1.30 & 0.58 & 0.90 & -0.94 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} -5.28 \\ 2.04 \\ 6.30 \\ 2.50 \end{bmatrix}$

(f) $\begin{bmatrix} 2.16 & 0.82 & -2.06 & 0.72 \\ -0.18 & -0.56 & 1.84 & -0.78 \\ 1.68 & -0.14 & 0.02 & -0.24 \\ -1.14 & -0.88 & -2.48 & 0.66 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} -12.6 \\ 13.8 \\ 0.2 \\ -32.6 \end{bmatrix}$

(g) $\begin{bmatrix} 0.00 & -0.54 & -0.72 & 0.90 \\ 0.40 & 0.74 & 0.32 & -0.10 \\ 1.20 & 2.22 & 0.96 & -0.30 \\ -0.00 & -0.18 & -0.24 & 0.30 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} -1.8 \\ -3.2 \\ -9.6 \\ -0.6 \end{bmatrix}$

(h) $\begin{bmatrix} 7 & 1 & -1 & 4 \\ 2 & 4 & -4 & 0 \\ 0 & 4 & 0 & -1 \\ -4 & 1 & 1 & -1 \\ -1 & 0 & -1 & 3 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 22.4 \\ 11.2 \\ -6.1 \\ -8.3 \\ 17.8 \end{bmatrix}$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(i) $\begin{bmatrix} 7 & 1 & -1 & 4 \\ 2 & 4 & -4 & 0 \\ 0 & 4 & 0 & -1 \\ -4 & 1 & 1 & -1 \\ -1 & 0 & -1 & 3 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} -2.1 \\ 2.2 \\ 4.6 \\ -0.7 \\ 5.5 \end{bmatrix}$

(j) $\begin{bmatrix} -1 & 0 & -6 & 0 & 5 \\ 0 & -3 & 2 & 1 & 7 \\ 0 & 2 & -3 & -2 & 2 \\ 0 & -3 & 7 & -5 & 0 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 30.7 \\ -17.0 \\ 21.3 \\ -45.7 \end{bmatrix}$

(k) $\begin{bmatrix} 1 & 6 & 1 & 1 & -4 \\ 3 & -2 & 0 & -4 & 7 \\ 1 & -3 & -1 & -5 & -2 \\ -1 & 4 & -2 & -1 & -2 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 4 \\ -7 \\ 2 \\ -3 \end{bmatrix}$

Exercise 3.3.5. Recall Theorems 2.2.27 and 2.2.31 on the existence of none, one, or an infinite number of solutions to linear equations. Use procedure 3.3.15 to provide an alternative proof to each of these two theorems.

Exercise 3.3.6. Write down the condition number and the rank of each of the matrices $A, \dots, L$ in Exercise 3.3.2 using the given SVDs.

Exercise 3.3.7. Write down the condition number and the rank of each of the matrices $A, \dots, F$ in Exercise 3.3.3 using the given SVDs. For each square matrix, compute `rcond` and comment on its relation to the condition number.

Exercise 3.3.8. In MATLAB/Octave, use `randn()` to generate some random matrices A of chosen sizes, and some correspondingly sized random right-hand side vectors $\boldsymbol{b}$. For each, find a general solution, if possible, of the system $A\boldsymbol{x} = \boldsymbol{b}$ with MATLAB/Octave and using Procedure 3.3.15. Record each step, the condition number and rank of A, and comment on what is interesting about the sizes you choose.

Exercise 3.3.9. Let $m \times n$ matrix A have the SVD $A = USV^{\mathrm{T}}$. Derive that the matrix $A^{\mathrm{T}}A$ has an SVD $A^{\mathrm{T}}A = V\bar{S}V^{\mathrm{T}}$, for what matrix $\bar{S}$? Derive that the matrix AA^{T} has an SVD $AA^{\mathrm{T}} = U\tilde{S}U^{\mathrm{T}}$, for what matrix $\tilde{S}$?

Exercise 3.3.10. Consider the problems (a)–(l) in Exercise 3.3.2 and problems (a)–(f) in Exercise 3.3.3. For each of these problems comment on the applicability of the Unique Solution Theorem 3.3.26, and comment on how the solution(s) illustrate the theorem.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Exercise 3.3.11. Recall Definition 3.2.2 says that a square matrix A is invertible if there exists a matrix B such that *both* $AB = I$ *and* $BA = I$. We now see that we need only one of these to ensure the matrix is invertible.

(a) Use Theorem 3.3.26c to now prove that a square matrix A is invertible if there exists a matrix B such that $BA = I$.

(b) Use the transpose and Theorems 3.3.26e and 3.3.23 to then prove that a square matrix A is invertible if there exists a matrix B such that $AB = I$.

Exercise 3.3.12. For each of the following systems, explore the effect on the solution of 1% errors in the right-hand side, and comment on the relation to the given condition number of the matrix.

(a) $\begin{bmatrix} 1 & 0 \\ -4 & 1 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} -2 \\ 10 \end{bmatrix}$, cond $= 17.94$

(b) $\begin{bmatrix} 2 & -4 \\ -2 & -1 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 10 \\ 0 \end{bmatrix}$, cond $= 2$

(c) $\begin{bmatrix} -3 & 1 \\ -4 & 2 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 6 \\ 8 \end{bmatrix}$, cond $= 14.93$

(d) $\begin{bmatrix} -1 & 1 \\ 4 & -5 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} -2 \\ 10 \end{bmatrix}$, cond $= 42.98$

(e) $\begin{bmatrix} -1 & -2 \\ 3 & 1 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} -5 \\ 9 \end{bmatrix}$, cond $= 2.618$

Exercise 3.3.13. For each of the following systems, use MATLAB/Octave to explore the effect on the solution of 0.1% errors in the right-hand side. Record your commands and output, and comment on the relation to the condition number of the matrix.

(a) $\begin{bmatrix} 1 & 2 & 2 \\ -1 & -1 & 0 \\ 0 & 3 & 1 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} -7 \\ 2 \\ -7 \end{bmatrix}$

(b) $\begin{bmatrix} -1 & 6 & -1 \\ 0 & 1 & 3 \\ -1 & 7 & 3 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 6 \\ 2 \\ 8 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 3 & 4 & 0 \\ 0 & 0 & -5 & 5 \\ 3 & 1 & 0 & 8 \\ 1 & 2 & 1 & 5 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 0 \\ 5 \\ 7 \\ 4 \end{bmatrix}$

v0.2a

(d) $\begin{bmatrix} -3 & -2 & -2 & -2 \\ 2 & 1 & -5 & -7 \\ 2 & 4 & 3 & 3 \\ 2 & 1 & 1 & 1 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} -3 \\ -8 \\ 5 \\ 2 \end{bmatrix}$

(e) $\begin{bmatrix} -1 & 6 & -6 & 2 & 7 \\ -7 & 4 & 3 & 1 & -8 \\ 7 & 6 & 4 & 0 & 5 \\ -8 & 3 & 3 & 2 & 4 \\ 2 & 0 & -3 & 1 & 0 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 5 \\ -7 \\ 5 \\ -2 \\ 1 \end{bmatrix}$

(f) $\begin{bmatrix} 9 & 0 & -10 & -8 & -1 \\ 9 & 3 & -5 & -4 & 4 \\ -1 & 0 & -3 & -6 & -6 \\ 4 & 6 & 0 & -5 & -14 \\ -2 & -1 & -4 & -7 & 5 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 7 \\ 4 \\ 3 \\ 5 \\ 1 \end{bmatrix}$

Exercise 3.3.14. For any $m \times n$ matrix A, use an SVD $A = USV^{\mathrm{T}}$ to prove that $\operatorname{rank}(A^{\mathrm{T}}A) = \operatorname{rank} A$ and that $\operatorname{cond}(A^{\mathrm{T}}A) = \operatorname{cond}(A)^2$ (see Exercise 3.3.9).

Exercise 3.3.15. Recall Example 3.3.32 introduced that finding a singular vector and singular value of a matrix A came from maximising $|A\boldsymbol{v}|$. Each of the following matrices, say A for discussion, has plotted $A\boldsymbol{v}$ (red) adjoined the corresponding unit vector $\boldsymbol{v}$ (blue). For each case:

(i) by inspection of the plot, estimate a singular vector $\boldsymbol{v}_1$ that appears to maximise $|A\boldsymbol{v}_1|$ (to one decimal place say);

(ii) estimate the corresponding singular value σ_1 by measuring $|A\boldsymbol{v}_1|$ on the plot;

(iii) set the second singular vector $\boldsymbol{v}_2$ to be orthogonal to $\boldsymbol{v}_1$ by swapping components, and making one negative;

(iv) estimate the corresponding singular value σ_2 by measuring $|A\boldsymbol{v}_2|$ on the plot;

(v) compute the matrix-vector products $A\boldsymbol{v}_1$ and $A\boldsymbol{v}_2$, and confirm they are orthogonal (approximately).

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(a) $A = \begin{bmatrix} 1 & 1 \\ 0.2 & 1.4 \end{bmatrix}$

(b) $B = \begin{bmatrix} 0 & -1.3 \\ 0.4 & 1.1 \end{bmatrix}$

(c) $C = \begin{bmatrix} 1.3 & 0.9 \\ 1.4 & 0.9 \end{bmatrix}$

(d) $D = \begin{bmatrix} 1.4 & -0.4 \\ -1.6 & 0.9 \end{bmatrix}$

Exercise 3.3.16. Use properties of the dot product to prove that when $\boldsymbol{v}_1$ and $\boldsymbol{v}$ are orthogonal unit vectors the vector $\boldsymbol{v}_1 \cos t + \boldsymbol{v} \sin t$ is also a unit vector for all t (used in the proof of the SVD in section 3.3.3).

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

3.4 Subspaces, basis and dimension

Section Contents

3.4.1 Subspaces are lines, planes, and so on 260

3.4.2 Orthonormal bases form a foundation 270

3.4.3 Is it a line? a plane? The dimension answers 282

3.4.4 Exercises 290

> [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*

One of the most fundamental geometric structures in linear algebra are the lines or planes through the origin, and higher dimensional analogues. For example, a general solution of linear equations often involve linear combinations such as $(-2,1,0,0)s + (-\frac{15}{7},0,\frac{9}{7},1)t$ (Example 2.2.29d) and $y_3\boldsymbol{v}_3 + y_4\boldsymbol{v}_4$ (Example 3.3.14): such combinations for all values of the free variables forms 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. Each such geometric structure is called a subspace.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

3.4.1 Subspaces are lines, planes, and so on

Example 3.4.1. The following graphs illustrate the concept of subspaces with structures (and imagined to infinitely extend 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 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.

v0.2a

(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?

(a)

(b)

(c)

(d)

■

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 "$\boldsymbol{v} \in \mathbb{R}^n$" means "$\boldsymbol{v}$ is a vector in $\mathbb{R}^n$". Hereafter, this book uses "$\in$" extensively.

Definition 3.4.3. *A **subspace** $\mathbb{W}$ of $\mathbb{R}^n$, is a set of vectors such that $\mathbf{0} \in \mathbb{W}$ and such that $\mathbb{W}$ is closed under addition and scalar multiplication: that is, for all $c \in \mathbb{R}$ and $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{W}$, then both $\boldsymbol{u} + \boldsymbol{v} \in \mathbb{W}$ and $c\boldsymbol{u} \in \mathbb{W}$.*

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

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

Solution: The origin $\mathbf{0}$ is in the line $y = x/2$ as $x = y = 0$ satifies the equation. The line $y = x/2$ is composed of vectors

in the form $\boldsymbol{u} = (1, \frac{1}{2})t$ for some parameter t. Then for any $c \in \mathbb{R}$, $c\boldsymbol{u} = c(1, \frac{1}{2})t = (1, \frac{1}{2})(ct) = (1, \frac{1}{2})t'$ for new parameter $t' = ct$; hence $c\boldsymbol{u}$ is in the line. Let $\boldsymbol{v} = (1, \frac{1}{2})s$ be another vector in the line for some parameter s, then $\boldsymbol{u} + \boldsymbol{v} = (1, \frac{1}{2})t + (1, \frac{1}{2})s = (1, \frac{1}{2})(t + s) = (1, \frac{1}{2})t'$ for new parameter $t' = t + s$; hence $\boldsymbol{u} + \boldsymbol{v}$ is in the line. The three requirements of Definition 3.4.3 are met, and so this line is a subspace.

(b) All vectors (x, y) such that $y = x - x^2/20$ (Example 3.4.1c).

Solution: A vector is 'in the set' when its end-point lies on a plot of the set, as in the margin. To show something is not a subspace, we only need to give one instance when one of the properties fail. One instance is that the vector $(20, 0)$ is in the curve as $20 - 20^2/20 = 0$, but the scalar multiple of half of this vector $\frac{1}{2}(20, 0) = (10, 0)$ is not as $10 - 10^2/20 = 5 \neq 0$. That is, the curve is not closed under scalar multiplication and hence is not a subspace.

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

Solution: A vector is 'in the set' when its end-point lies on a plot of the set, as in the margin. Although vectors (x, y) in the line $y = x/2$ for $x, y \geq 0$ includes the origin and is closed under addition, it fails the scalar multiplication test. For example, $\boldsymbol{u} = (2, 1)$ is in the line, but the scalar multiple $(-1)\boldsymbol{u} = (-2, -1)$ is not. Hence it is not a subspace.

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

Solution: The origin $\mathbf{0}$ is in the plane $z = -x/6 + y/3$ as $x = y = z = 0$ satisfies the equation. A vector $\boldsymbol{u} = (u_1, u_2, u_3)$ is in the plane provided $-u_1 + 2u_2 - 6u_3 = 0$. Consider $c\boldsymbol{u} = (cu_1, cu_2, cu_3)$ for which $-(cu_1) + 2(cu_2) - 6(cu_3) = c(-u_1 + 2u_2 - 6u_3) = c \times 0 = 0$ and hence must also be in the plane. Also let vector $\boldsymbol{v} = (v_1, v_2, v_3)$ be in the plane and consider $\boldsymbol{u} + \boldsymbol{v} = (u_1 + v_1, u_2 + v_2, u_3 + v_3)$ for which $-(u_1 + v_1) + 2(u_2 + v_2) - 6(u_3 + v_3) = -u_1 - v_1 + 2u_2 + 2v_2 - 6u_3 - 6v_3 = (-u_1 + 2u_2 - 6u_3) + (-v_1 + 2v_2 - 6v_3) = 0 + 0 = 0$ and hence must also be in the plane. The three requirements of Definition 3.4.3 are met, and so this plane is a subspace.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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

Solution: In this case, consider the vector $\boldsymbol{u} = (0, 0, 5)$ (shown in the above picture): any scalar multiple, say $2\boldsymbol{u} = (0, 0, 10)$, is not in the plane. That is, vectors in the plane are not closed under scalar multiplication, and hence the plane is not a subspace.

(f) $\{\mathbf{0}\}$.

Solution: The zero vector forms a trivial subspace, $\mathbb{W} = \{\mathbf{0}\}$: firstly, $\mathbf{0} \in \mathbb{W}$; secondly, the only vector in $\mathbb{W}$ is $\boldsymbol{u} = \mathbf{0}$ for which every scalar multiple $c\boldsymbol{u} = c\mathbf{0} = \mathbf{0} \in \mathbb{W}$; and thirdly, a second vector $\boldsymbol{v}$ in $\mathbb{W}$ can only be $\boldsymbol{v} = \mathbf{0}$ so $\boldsymbol{u} + \boldsymbol{v} = \mathbf{0} + \mathbf{0} = \mathbf{0} \in \mathbb{W}$. The three requirements of Definition 3.4.3 are met, and so $\{\mathbf{0}\}$ is always a subspace.

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

Solution: Lastly, $\mathbb{R}^n$ also is a subspace: firstly, $\mathbf{0} = (0\,, 0\,, \dots, 0) \in \mathbb{R}^n$; secondly, for $\boldsymbol{u} = (u_1\,, u_2\,, \dots, u_n) \in \mathbb{R}^n$, the scalar multiplication $c\boldsymbol{u} = c(u_1\,, u_2\,, \dots, u_n) = (cu_1\,, cu_2\,, \dots, cu_n) \in \mathbb{R}^n$; and thirdly, for $\boldsymbol{v} = (v_1\,, v_2\,, \dots, v_n) \in \mathbb{R}^n$, the vector addition $\boldsymbol{u} + \boldsymbol{v} = (u_1\,, u_2\,, \dots, u_n) + (v_1\,, v_2\,, \dots, v_n) = (u_1 + v_1\,, u_2 + v_2\,, \dots, u_n + v_n) \in \mathbb{R}^n$. The three requirements of Definition 3.4.3 are met, and so $\mathbb{R}^n$ is always a subspace.

■

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

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

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

(c) $(1\,, \frac{1}{4})\,,\ (0\,, 0)$

(d) $(-1\,, -\frac{1}{4})\,,\ (2\,, 2)$

■

In summary:

v0.2a

- in two dimensions ($\mathbb{R}^2$), subspaces are the origin $\mathbf{0}$, a line, through $\mathbf{0}$, or the entire plane $\mathbb{R}^2$;
- in three dimensions ($\mathbb{R}^3$), subspaces are the origin $\mathbf{0}$, a line through $\mathbf{0}$, a plane through $\mathbf{0}$, or the entire space $\mathbb{R}^3$;
- and analogously for higher dimensions ($\mathbb{R}^n$).

Recall that the set of all linear combinations of a set of vectors, such as $(-2,1,0,0)s + (-\frac{15}{7},0,\frac{9}{7},1)t$ (Example 2.2.29d), is called the span of that set (Definition 2.3.9).

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

Proof. Denote $\operatorname{span}\{\boldsymbol{v}_1\,,\boldsymbol{v}_2\,,\dots,\boldsymbol{v}_k\}$ by $\mathbb{W}$; we aim to prove it is a subspace (Definition 3.4.3). First, $\mathbf{0} = 0\boldsymbol{v}_1 + 0\boldsymbol{v}_2 + \cdots + 0\boldsymbol{v}_k$ which is a linear combination of $\boldsymbol{v}_1\,,\boldsymbol{v}_2\,,\dots,\boldsymbol{v}_k$, and so the zero vector $\mathbf{0}\in\mathbb{W}$. Now let $\boldsymbol{u}\,,\boldsymbol{v}\in\mathbb{W}$ then by Definition 2.3.9 there are coefficients $a_1\,,a_2\,,\dots,a_n$ and $b_1\,,b_2\,,\dots,b_n$ such that

$$\begin{aligned}\boldsymbol{u} &= a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + \cdots + a_k\boldsymbol{v}_k\ ,\\ \boldsymbol{v} &= b_1\boldsymbol{v}_1 + b_2\boldsymbol{v}_2 + \cdots + b_k\boldsymbol{v}_k\,.\end{aligned}$$

Secondly, consequently

$$\begin{aligned}\boldsymbol{u}+\boldsymbol{v} &= a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + \cdots + a_k\boldsymbol{v}_k\\ &\quad + b_1\boldsymbol{v}_1 + b_2\boldsymbol{v}_2 + \cdots + b_k\boldsymbol{v}_k\\ &= (a_1+b_1)\boldsymbol{v}_1 + (a_2+b_2)\boldsymbol{v}_2 + \cdots + (a_k+b_k)\boldsymbol{v}_k\,,\end{aligned}$$

which is a linear combination of $\boldsymbol{v}_1\,,\boldsymbol{v}_2\,,\dots,\boldsymbol{v}_k$, and so is in $\mathbb{W}$. Thirdly, for any scalar c,

$$\begin{aligned}c\boldsymbol{u} &= c(a_1\boldsymbol{v}_1 + a_2\boldsymbol{v}_2 + \cdots + a_k\boldsymbol{v}_k)\\ &= ca_1\boldsymbol{v}_1 + ca_2\boldsymbol{v}_2 + \cdots + ca_k\boldsymbol{v}_k\,,\end{aligned}$$

which is a linear combination of $\boldsymbol{v}_1\,,\boldsymbol{v}_2\,,\dots,\boldsymbol{v}_k$, and so is in $\mathbb{W}$. Hence $\mathbb{W} = \operatorname{span}\{\boldsymbol{v}_1\,,\boldsymbol{v}_2\,,\dots,\boldsymbol{v}_k\}$ is a subspace. □

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 $y = x/2$. ■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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 below, since every linear combination of these two vectors fills out the plane: $a_1(3,3,1/2) + a_2(0,3,1) = (3a_1, 3a_1 + 3a_2, a_1/2 + a_2)$ and so lies in the plane as $-x/6+y/3-z = -\frac{1}{6}3a_1 + \frac{1}{3}(3a_1+3a_2) - (a_1/2+a_2) = -\frac{1}{2}a_1 + a_1 + a_2 - \frac{1}{2}a_1 - a_2 = 0$ for all a_1 and a_2 (although such arguments do not establish that the linear combinations cover the whole plane—we need Theorem 3.4.14).

Also, $\operatorname{span}\{(5,1,-1/2),(0,-3,-1),(-4,1,1)\}$ is the plane $z = -x/6 + y/3$, as illustrated below. The reason is that every linear combination of these three vectors fills out the plane: $a_1(5,1,-1/2) + a_2(0,-3,-1) + a_3(-4,1,1) = (5a_1 - 4a_3, a_1 - 3a_2 + a_3, -a_1/2 - a_2 + a_3)$ and so lies in the plane as $-x/6 + y/3 - z = -\frac{1}{6}(5a_1 - 4a_3) + \frac{1}{3}(a_1 - 3a_2 + a_3) - (-a_1/2 - a_2 + a_3) = -\frac{5}{6}a_1 + \frac{2}{3}a_3 + \frac{1}{3}a_1 - a_2 + \frac{1}{3}a_3 + \frac{1}{2}a_1 + a_2 - a_3 = 0$ for all a_1, a_2 and a_3.

■

Example 3.4.9. Find a set of two vectors that spans the plane $x - 2y + 3z = 0$.

Solution: Write the equation for this plane as $x = 2y - 3z$, say, then vectors in the plane are all of the form $\boldsymbol{u} = (x,y,z) = (2y - 3z, y, z) = (2,1,0)y + (-3,0,1)z$. That is, all vectors in the plane may be written as a linear combination of the two vectors $(2,1,0)$ and $(-3,0,1)$, hence the plane is $\operatorname{span}\{(2,1,0),(-3,0,1)\}$ as illustrated in stereo below.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

■

Such subspaces connect with matrices. The connection is via a matrix whose columns are the vectors appearing within the span. Although sometimes we also use the rows of the matrix to be the vectors in the span.

Definition 3.4.10. *(a) The **column space** of any $m \times n$ matrix A is the subspace of $\mathbb{R}^m$ spanned by the n column vectors of A.* [15]

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

Example 3.4.11. Examples 3.4.7–3.4.9 provide some cases.

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

  The row space of this matrix A is $\operatorname{span}\{(1,-2),(\frac{1}{2},-1)\}$. This row space is the set of all vectors of the form $(1,-2)s + (\frac{1}{2},-1)t = (s+t/2, -2s-t) = (1,-2)(s+t/2) = (1,-2)t'$ is the line $y = -2x$ as illustrated in the margin. That the row space and the column space are both lines, albeit different lines, is not a coincidence (Theorem 3.4.32).

- Example 3.4.8 shows that the column space of matrix

$$B = \begin{bmatrix} 3 & 0 \\ 3 & 3 \\ \frac{1}{2} & 1 \end{bmatrix}$$

  is the plane $z = -x/6 + y/3$ in $\mathbb{R}^3$.

  The row space of matrix B is $\operatorname{span}\{(3,0),(3,3),(\frac{1}{2},1)\}$ which is a subspace of $\mathbb{R}^2$, whereas the column space is a subspace of $\mathbb{R}^3$. Here the span is all of $\mathbb{R}^2$ as for each $(x,y) \in \mathbb{R}^2$ choose the linear combination $\frac{x-y}{3}(3,0) + \frac{y}{3}(3,3) + 0(\frac{1}{2},1) = (x-y+y+0, 0+y+0) = (x,y)$ so each (x,y) is in the span, and hence all of the $\mathbb{R}^2$ plane is the span. That the column space and the row space are both planes is no coincidence (Theorem 3.4.32).

- Example 3.4.8 also shows that the column space of matrix

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

  is also the plane $z = -x/6 + y/3$ in $\mathbb{R}^3$.

[15] Some of you will know that the column space is also called the range, but for the moment we just use column space.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Now, $\operatorname{span}\{(5,0,-4),(1,-3,1),(-\frac{1}{2},-1,1)\}$ is the row space of matrix C. It is not readily apparent but we can check that this space is the plane $4x+3y+5z=0$ as illustrated below in stereo. To see this, consider all linear combinations $a_1(5,0,-4)+a_2(1,-3,1)+a_3(-\frac{1}{2},-1,1)=(5a_1+a_2-a_3/2,-3a_2-a_3,-4a_1+a_2+a_3)$ satisfy $4x+3y+5z=4(5a_1+a_2-a_3/2)+3(-3a_2-a_3)+5(-4a_1+a_2+a_3)=20a_1+4a_2-2a_3-9a_2-3a_3-20a_1+5a_2+5a_3=0$.

Again, it is no coincidence that the row and column spaces of C are both planes (Theorem 3.4.32).

■

Activity 3.4.12. Which one of the following vectors is in the column space of the matrix

$$\begin{bmatrix} 6 & 2 \\ -3 & 5 \\ -2 & -1 \end{bmatrix}?$$

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

■

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

Solution: The question is: can we find a linear combination of the columns of A which equals vector $\boldsymbol{b}$? That is, can we find some vector $\boldsymbol{x}$ such that $A\boldsymbol{x}=\boldsymbol{b}$? Answer using our knowledge of linear equations.

Let's use Procedure 3.3.15 in MATLAB/Octave.

(a) Compute an SVD of this 5×3 matrix with

v0.2a

```
A=[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]
[U,S,V]=svd(A)
```

to find (2 d.p.)

```
U =
  -0.58   0.49   0.53  -0.07   0.37
  -0.17   0.69  -0.65  -0.04  -0.25
  -0.56  -0.28  -0.10   0.74  -0.22
   0.57   0.43   0.21   0.66   0.14
  -0.04  -0.15  -0.49   0.10   0.85
S =
   7.52      0      0
      0   4.91      0
      0      0   3.86
      0      0      0
      0      0      0
V = ...
```

(b) Then solve $U\boldsymbol{z} = \boldsymbol{b}$ with `z=U'*[-0.6;0;-2.1;1.9;1.2]` to find (2 d.p.) $\boldsymbol{z} = (2.55, 0.92, -0.29, -0.15, 1.54)$.

(c) Now the diagonal matrix S has three non-zero singular values, and the last two rows are zero. So to be able to solve $S\boldsymbol{y} = \boldsymbol{z}$ we need the last two components of $\boldsymbol{z}$ to be zero. But, at -0.15 and 1.54, they are not zero, so the system is not solvable. Hence there is no linear combination of the columns of A that gives us vector $\boldsymbol{b}$. Consequently, vector $\boldsymbol{b}$ is not in the column space of A.

(d) For the case of vector $\boldsymbol{c} = (15.2, 5.4, 3.8, -1.9, -3.7)$ solve $U\boldsymbol{z} = \boldsymbol{c}$ with

```
z=U'*[15.2;5.4;3.8;-1.9;-3.7]
```

to find $\boldsymbol{z} = (-12.800, 9.876, 5.533, 0.000, 0.000)$. Since the last two entries in vector $\boldsymbol{z}$ are zero, corresponding to the zero rows of S, a solution exists to $S\boldsymbol{y} = \boldsymbol{z}$. Hence a solution exists to $A\boldsymbol{x} = \boldsymbol{c}$. Consequently, vector $\boldsymbol{c}$ is in the column space of A.

(Incidentally, you may check that $\boldsymbol{c} = 2\boldsymbol{a}_1 - 2\boldsymbol{a}_2 + \boldsymbol{a}_3$.)

■

Another subspace associated with matrices is the set of possible solutions to a homogeneous system of linear equations.

v0.2a

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

Proof. First, $A\mathbf{0} = \mathbf{0}$ so $\mathbf{0} \in \operatorname{null} A$. Let $\boldsymbol{u}, \boldsymbol{v} \in \operatorname{null} A$; that is, $A\boldsymbol{u} = \mathbf{0}$ and $A\boldsymbol{v} = \mathbf{0}$. Second, by the distributivity of matrix-vector multiplication (Theorem 3.1.24), $A(\boldsymbol{u} + \boldsymbol{v}) = A\boldsymbol{u} + A\boldsymbol{v} = \mathbf{0} + \mathbf{0} = \mathbf{0}$ and so $\boldsymbol{u} + \boldsymbol{v} \in \operatorname{null} A$. Third, by the associativity and commutativity of *scalar* multiplication (Theorem 3.1.22), for any $c \in \mathbb{R}$, $A(c\boldsymbol{u}) = Ac\boldsymbol{u} = cA\boldsymbol{u} = c(A\boldsymbol{u}) = c\mathbf{0} = \mathbf{0}$ and so $c\boldsymbol{u} \in \operatorname{null} A$. Hence $\operatorname{null} A$ is a subspace (Definition 3.4.3). □

Example 3.4.15.

- Example 2.2.29a showed that the only solution of the homogeneous system $\begin{cases} 3x_1 - 3x_2 = 0 \\ -x_1 - 7x_2 = 0 \end{cases}$ is $\boldsymbol{x} = \mathbf{0}$. Thus its set of solutions is $\{\mathbf{0}\}$ which is a subspace (Example 3.4.4f). Thus $\{\mathbf{0}\}$ is the nullspace of matrix $\begin{bmatrix} 3 & -3 \\ -1 & -7 \end{bmatrix}$.
- Recall the homogeneous system of linear equations from Example 2.2.29d has solutions $\boldsymbol{x} = (-2s - \frac{15}{7}t, s, \frac{9}{7}t, t) = (-2, 1, 0, 0)s + (-\frac{15}{7}, 0, \frac{9}{7}, 1)t$ 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. In general, a *non*-homogeneous system does not form a subspace of solutions.

■

Example 3.4.16. Given the matrix

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

is vector $\boldsymbol{v} = (-2, 6, 1)$ in the null space of A? What about vector $\boldsymbol{w} = (1, -3, 2)$?

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Solution: To test a given vector, just multiply by the matrix and see if the result is zero.

- $A\boldsymbol{v} = \begin{bmatrix} 3\cdot(-2)+1\cdot 6+0\cdot 1 \\ -5\cdot(-2)-1\cdot 6-4\cdot 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \mathbf{0}$, so $\boldsymbol{v} \in \operatorname{null} A$.
- $A\boldsymbol{w} = \begin{bmatrix} 3\cdot 1+1\cdot(-3)+0\cdot 2 \\ -5\cdot 1-1\cdot(-3)-4\cdot 2 \end{bmatrix} = \begin{bmatrix} 0 \\ -10 \end{bmatrix} \neq \mathbf{0}$, so $\boldsymbol{w}$ is not in the nullspace.

■

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

■

Summary Three ways that subspaces arise from a matrix are as the column space, row space, and nullspace.

3.4.2 Orthonormal bases form a foundation

> The importance of orthogonal basis functions in interpolation and approximation cannot be overstated.
> *(Cuyt 2015, §5.3)*

Given that subspaces arise frequently in linear algebra, and that there are many ways of representing the same subspace (as seen in some previous examples), is there a 'best' way of representing subspaces? The next definition and theorems largely answer this question.

The aim is to mainly use an orthonormal set of vectors to span a subspace. The virtue is that orthonormal sets have many practically useful properties. For example, orthonormal sets underpin JPEG images, understanding of vibration, and reliable weather forecasting. Recall that an orthonormal set is composed of vectors that are both at right-angles to each other (their dot products are zero) and all of unit length (Definition 3.2.38).

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Definition 3.4.18. *An **orthonormal basis** for a subspace $\mathbb{W}$ of $\mathbb{R}^n$ is an orthonormal set of vectors that span $\mathbb{W}$.*

Example 3.4.19. Recall that $\mathbb{R}^n$ is itself a subspace of $\mathbb{R}^n$ (Example 3.4.4g).

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

(b) The n columns $\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_n$ of an $n \times n$ orthogonal matrix Q also form an orthonormal basis for the subspace $\mathbb{R}^n$. The reasons are: first, Theorem 3.2.48b establishes 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 3.2.48c).

This example also illustrates that generally there are many different orthonormal bases for a given subspace. ■

Activity 3.4.20. Which of the following sets is an orthonormal basis for $\mathbb{R}^2$?

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

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

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

(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 $\mathbb{R}^3$.

Solution: This line is a subspace as it passes through $\mathbf{0}$. A parametric description of the line is $\boldsymbol{x} = (x, y, z) = (t, t, t) = (1, 1, 1)t$ for every t. So the subspace is spanned by $\{(1, 1, 1)\}$. But this is not an orthonormal basis as it is not of unit length, so divide by its length $|(1, 1, 1)| = \sqrt{1^2 + 1^2 + 1^2} = \sqrt{3}$. That is, $\{(1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})\}$ is an orthonormal basis for the subspace, as illustrated in stereo below.

Another orthonormal basis is the unit vector in the opposite direction, $\{(-1/\sqrt{3}, -1/\sqrt{3}, -1/\sqrt{3})\}$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

■

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 below.

Solution: First, the given set is of unit vectors as the lengths are $|\boldsymbol{u}_1|=\sqrt{\frac{4}{9}+\frac{1}{9}+\frac{4}{9}}=1$ and $|\boldsymbol{u}_2|=\sqrt{\frac{4}{9}+\frac{4}{9}+\frac{1}{9}}=1$. Second, the set is orthonormal as their dot product is zero: $\boldsymbol{u}_1\cdot\boldsymbol{u}_2=-\frac{4}{9}+\frac{2}{9}+\frac{2}{9}=0$. Third, they both lie in the plane as we check by substituting their components in the equation: for $\boldsymbol{u}_1$, $-x+2y-2z=\frac{2}{3}+2(\frac{1}{3})-2(\frac{2}{3})=\frac{2}{3}+\frac{2}{3}-\frac{4}{3}=0$; and for $\boldsymbol{u}_2$, $-x+2y-2z=-\frac{2}{3}+2(\frac{2}{3})-2(\frac{2}{3})=-\frac{2}{3}+\frac{4}{3}-\frac{2}{3}=0$. Lastly, from the parametric form of an equation for a plane (section 1.3.4) we know that all linear combinations of $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ will span the plane.

■

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

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

Proof. The argument corresponds to that for Procedure 3.3.15. Consider any point $\boldsymbol{b}\in\operatorname{span}\{\boldsymbol{a}_1,\boldsymbol{a}_2,\dots,\boldsymbol{a}_n\}$. Because $\boldsymbol{b}$ is in the span, there exist coefficients $x_1,x_2,\dots,x_n$ such that

$$\begin{aligned}\boldsymbol{b}&=\boldsymbol{a}_1x_1+\boldsymbol{a}_2x_2+\cdots+\boldsymbol{a}_nx_n\\&=A\boldsymbol{x}\quad\text{(by matrix-vector product §3.1.2)}\\&=USV^{\mathrm{T}}\boldsymbol{x}\quad\text{(by the SVD of }A\text{)}\end{aligned}$$

$$
\begin{aligned}
&= USy \quad (\text{for } y = V^{\mathrm{T}}x) \\
&= Uz \quad (\text{for } z = (z_1, z_2, \dots, z_r, 0, \dots, 0) = Sy) \\
&= u_1 z_1 + u_2 z_2 + \cdots + u_r z_r \quad (\text{by matrix-vector product}) \\
&\in \operatorname{span}\{u_1, u_2, \dots, u_r\}.
\end{aligned}
$$

These equalities also hold in reverse due to the invertibility of U and V, and with $y_i = z_i/\sigma_i$ for $i = 1, 2, \dots, r$. Hence a point is in $\operatorname{span}\{a_1, a_2, \dots, a_n\}$ if and only if it is in $\operatorname{span}\{u_1, u_2, \dots, u_r\}$. Lastly, U is an orthogonal matrix and so the set of columns $\{u_1, u_2, \dots, u_r\}$ is an orthonormal set. Hence $\{u_1, u_2, \dots, u_r\}$ forms an orthonormal basis for $\operatorname{span}\{a_1, a_2, \dots, a_n\}$. □

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

Solution: Form the matrix whose columns are the given vectors

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

then ask MATLAB/Octave for an SVD and interpret.

```
A=[1 -2; 1/2 -1]
[U,S,V]=svd(A)
```

The computed SVD is (V is immaterial here)

```
U =
   -0.8944   -0.4472
   -0.4472    0.8944
S =
    2.5000         0
         0    0.0000
V = ...
```

There is one non-zero singular value—the matrix has rank one—so an orthonormal basis for the span is the first column of matrix U, namely the set $\{(-0.89, -0.45)\}$ (2 d.p.). That is, every vector in $\operatorname{span}\{(1, \frac{1}{2}), (-2, -1)\}$ can be written as $(-0.89, -0.45)t$ for some t: hence the span is a line.

■

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.

Solution:

- Form the matrix whose columns are the given vectors

$$A = \begin{bmatrix} 3 & 0 \\ 3 & 3 \\ \frac{1}{2} & 1 \end{bmatrix},$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

then ask MATLAB/Octave for the SVD and interpret. It is often easier to form the matrix in MATLAB/Octave by entering the vectors as rows and then transposing:

```
A=[3 3 1/2;0 3 1]'
[U,S,V]=svd(A)
```

The computed SVD is (2 d.p.)

```
U =
  -0.51   0.85   0.16
  -0.84  -0.44  -0.31
  -0.20  -0.29   0.94
S =
   4.95      0
      0   1.94
      0      0
V = ...
```

There are two non-zero singular values—the matrix has rank two—so an orthonormal basis for the plane is the set of the first two columns of matrix U, namely $(-0.51, -0.84, -0.20)$ and $(0.85, -0.44, -0.29)$. These basis vectors are illustrated as the pair of red vectors in stereo below.

- Similarly, form the matrix

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

then ask MATLAB/Octave for the SVD and interpret. Form the matrix in MATLAB/Octave by entering the vectors as rows and then transposing:

```
B=[5 1 -1/2; 0 -3 -1; -4 1 1]'
[U,S,V]=svd(B)
```

The computed SVD is (2 d.p.)

```
U =
  -0.99  -0.04   0.16
  -0.01  -0.95  -0.31
   0.16  -0.31   0.94
S =
   6.49      0      0
```

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

```
       0    3.49       0
       0       0    0.00
V = ...
```

There are two non-zero singular values—the matrix has rank two—so an orthonormal basis for the plane spanned by the three vectors is the set of the first two columns of matrix U, namely the vectors $(-0.99, -0.01, 0.16)$ and $(-0.04, -0.95, -0.31)$. These are the pair of brown vectors in the above stereo illustration.

■

Activity 3.4.26. The matrix

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

has the following SVD computed via MATLAB/Octave command `[U,S,V]=svd(A)`: 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
```

(a) $\{(-0.67, 0.69, 0.27),\ (-0.55, -0.23, -0.80)\}$

(b) $\{(-0.67, -0.55, -0.49),\ (0.69, -0.23, -0.69)\}$

(c) $\{(-0.75, -0.32, -0.58),\ (-0.66, 0.49, 0.58)\}$

(d) $\{(-0.75, -0.66, 0.09),\ (-0.32, 0.49, 0.81)\}$

Extension: recalling Theorem 3.3.23, which of the above is an orthonormal basis for the row space of A? ■

v0.2a

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

Example 3.4.27 (data reduction). Every four or five years the phenomenon of El Nino makes a large impact on the world's weather: from drought in Australia to floods in South America. We would like to predict El Nino in advance to save lives and economies. El Nino is correlated significantly with the difference in atmospheric pressure between Darwin and Tahiti—the so-called Southern Oscillation Index (SOI). This example seeks patterns in the SOI in order to be able to predict the SOI and hence predict El Nino.

Figure 3.1 plots the yearly average SOI each year for fifty years up to 1993. A strong regular structure is apparent, but there are significant variations and complexities in the year-to-year signal. The challenge of this example is to explore the full details of this signal.

Let's use a general technique called a Singular Spectrum Analysis. Consider a window of ten years of the SOI, and let the window 'slide' across the data to give us many 'local' pictures of the evolution in time. For example, Figure 3.2 plots six windows (each displaced vertically for clarity) each of length ten years. As the 'window' slides across the fifty year data of Figure 3.1 there are 41 such local views of the data of length ten years. Let's invoke the concept of subspaces to detect regularity in the data via these windows.

The fundamental property is that if the data has regularities, then it should lie in some subspace. We detect such subspaces using the SVD of a matrix.

- First form the 41 data windows of length ten into a matrix of size 10×41. The numerical values of the SOI data of Figure 3.1 are the following:

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

```
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()`

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

function:

```
A=hankel(soi(1:10),soi(10:50))
```

In MATLAB/Octave the command `hankel(s(1:w),s(w:n))` forms the $w \times (n-w+1)$ so-called Hankel matrix

$$\begin{bmatrix} s_1 & s_2 & s_3 & \cdots & s_{n-w} & s_{n-w+1} \\ s_2 & s_3 & \vdots & & s_{n-w+1} & \vdots \\ s_3 & \vdots & s_w & & \vdots & \vdots \\ \vdots & s_w & s_{w+1} & & \vdots & s_{n-1} \\ s_w & s_{w+1} & s_{w+2} & \cdots & s_{n-1} & s_n \end{bmatrix}$$

- Lastly, compute the SVD of the matrix of these windows:

```
[U,S,V]=svd(A);
singValues=diag(S)
plot(U(:,1:4))
```

The computed singular values are 44.63, 43.01, 39.37, 36.69, 0.03, 0.03, 0.02, 0.02, 0.02, 0.01. In practice, treat the six small singular values as zero. Since there are four non-zero singular values, the windows of data lie in a subspace spanned by the first four columns of U.

That is, all the structure seen in the fifty year SOI data of Figure 3.1 can be expressed in terms of the orthonormal basis of four ten-year vectors plotted in Figure 3.3. This analysis implies the SOI data is

v0.2a

composed of two cycles of different frequencies.[16] ■

Example 3.4.25 obtained two different orthonormal bases for the one plane. Although the bases are different, they both had the same number of vectors. The next theorem establishes that this same number always occurs.

Theorem 3.4.28. *Any two orthonormal bases for a given subspace have the same number of vectors.*

Proof. Let $\mathcal{U} = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_r\}$, with r vectors, and $\mathcal{V} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_s\}$, with s vectors, be any two orthonormal bases for a subspace in $\mathbb{R}^n$. Prove the number of vectors $r = s$ by contradiction. First assume $r < s$ ($\mathcal{U}$ has less vectors than $\mathcal{V}$). Since $\mathcal{U}$ is an orthonormal basis for the subspace every vector in $\mathcal{V}$ can be written as a linear combination of vectors in $\mathcal{U}$ with some coefficients a_{ij}:

$$\begin{aligned}
\boldsymbol{v}_1 &= \boldsymbol{u}_1 a_{11} + \boldsymbol{u}_2 a_{21} + \cdots + \boldsymbol{u}_r a_{r1} \ , \\
\boldsymbol{v}_2 &= \boldsymbol{u}_1 a_{12} + \boldsymbol{u}_2 a_{22} + \cdots + \boldsymbol{u}_r a_{r2} \ , \\
&\vdots \\
\boldsymbol{v}_s &= \boldsymbol{u}_1 a_{1s} + \boldsymbol{u}_2 a_{2s} + \cdots + \boldsymbol{u}_r a_{rs} \, .
\end{aligned}$$

Write each of these, such as the first one, in the form

$$\boldsymbol{v}_1 = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_r \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{r1} \end{bmatrix} = U\boldsymbol{a}_1 \, ,$$

where matrix $U = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_r \end{bmatrix}$. Then the $n \times s$ matrix

$$\begin{aligned}
V &= \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_s \end{bmatrix} \\
&= \begin{bmatrix} U\boldsymbol{a}_1 & U\boldsymbol{a}_2 & \cdots & U\boldsymbol{a}_s \end{bmatrix} \\
&= U \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_s \end{bmatrix} = UA
\end{aligned}$$

for the $r \times s$ matrix $A = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_s \end{bmatrix}$. By assumption $r < s$ and so Theorem 2.2.31 assures that the homogeneous system $A\boldsymbol{x} = \boldsymbol{0}$ has infinitely many solutions, choose any one non-trivial solution $\boldsymbol{x} \neq \boldsymbol{0}$. Consider

$$\begin{aligned}
V\boldsymbol{x} &= UA\boldsymbol{x} \quad \text{(from above)} \\
&= U\boldsymbol{0} \quad (\text{since } A\boldsymbol{x} = \boldsymbol{0}) \\
&= \boldsymbol{0} \, .
\end{aligned}$$

[16] However, I 'smoothed' the SOI data for the purposes of this example. The real SOI data is considerably noisier. Also we would use 600 monthly averages not 50 yearly averages: so a ten year window would be a window of 120 months, and the matrix would be considerably larger 120×481. Nonetheless, the conclusions with the real data, and justified by Chapter 5, are much the same.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Since $\mathcal{V}$ is an orthonormal set, the matrix V is orthogonal (Theorem 3.2.48). Then premultiplying $V\boldsymbol{x} = \boldsymbol{0}$ by V^{T} gives $V^{\mathrm{T}}V\boldsymbol{x} = V^{\mathrm{T}}\boldsymbol{0}$ which simplifies to $I_s\boldsymbol{x} = \boldsymbol{0}$; that is, $\boldsymbol{x} = \boldsymbol{0}$. But this is a contradiction, so we cannot have $r < s$.

Second, a corresponding argument establishes we cannot have $s < r$. Hence $r = s$, and so all orthonormal bases of a given subspace must have the same number of vectors. □

The following optional theorem and proof settles an existential issue.

An existential issue How do we know that every subspace has an orthonormal basis? We know many subspaces, such as row and column spaces, have an orthonormal basis because they are the span of rows and columns of a matrix, and then Procedure 3.4.23 assures us they have an orthonormal basis. But do all subspaces have an orthonormal basis? The following theorem certifies that they do.

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

Proof. If the subspace is $\mathbb{W} = \{\boldsymbol{0}\}$, then $\mathbb{W} = \operatorname{span}\emptyset$ gives a basis and this trivial case is done.

Then for other subspaces $\mathbb{W} \neq \{\boldsymbol{0}\}$ there exists a non-zero vector $\boldsymbol{w} \in \mathbb{W}$. Normalise the vector by setting $\boldsymbol{u}_1 = \boldsymbol{w}/|\boldsymbol{w}|$. Then all scalar multiples $c\boldsymbol{u}_1 = c\boldsymbol{w}/|\boldsymbol{w}| = (c/|\boldsymbol{w}|)\boldsymbol{w} \in \mathbb{W}$ by closure of $\mathbb{W}$ under scalar multiplication (Definition 3.4.3). Hence $\operatorname{span}\{\boldsymbol{u}_1\} \subseteq \mathbb{W}$ (as illustrated below for an example). Consequently, either $\operatorname{span}\{\boldsymbol{u}_1\} = \mathbb{W}$ and we are done, or we repeat the following step until the space $\mathbb{W}$ is spanned.

Given orthonormal vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k$ such that the set $\operatorname{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k\} \subset \mathbb{W}$, so $\operatorname{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k\} \neq \mathbb{W}$. Then there must exist a vector $\boldsymbol{w} \in \mathbb{W}$ which is not in the set $\operatorname{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k\}$. By the closure of subspace $\mathbb{W}$ under addition and scalar multiplication (Definition 3.4.3), the set $\operatorname{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k, \boldsymbol{w}\} \subseteq \mathbb{W}$. Procedure 3.4.23, on the orthonormal basis for a span, then assures us that an SVD gives an orthonormal basis $\{\boldsymbol{u}'_1, \boldsymbol{u}'_2, \ldots, \boldsymbol{u}'_{k+1}\}$ for the set $\operatorname{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k, \boldsymbol{w}\} \subseteq \mathbb{W}$ (as illustrated for an example). Consequently, either $\operatorname{span}\{\boldsymbol{u}'_1, \boldsymbol{u}'_2, \ldots, \boldsymbol{u}'_{k+1}\} = \mathbb{W}$

and we are done, or we repeat the process of this paragraph with k bigger by one.

The process must terminate because $\mathbb{W} \subseteq \mathbb{R}^n$. If the process ever repeats until $k = n$, then we know $\mathbb{W} = \mathbb{R}^n$ and we are done as $\mathbb{R}^n$ is spanned by the n standard unit vectors. □

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[17]) Why?

Occasionally the weather is both near a 'tipping point' where small changes may cause a storm, and where the errors in measuring the current weather are of the size of the necessary changes. Then the storm would be within the possibilities, but it would not be forecast if the measurements were, by chance error, the 'other side' of the tipping point. Meteorologists now mostly overcome this problem by executing on their computers an ensemble of simulations, perhaps an ensemble of a hundred different forecast simulations (Roulstone & Norbury 2013, pp.274–80, e.g.). Such a set of 100 simulations essentially lie in a subspace spanned by 100 vectors in the vastly larger space, say $\mathbb{R}^{1000,000,000}$, of the maybe billion variables in the weather model. But what happens in the computational simulations is that the ensemble of simulations degenerate in time: so the meteorologists continuously 'renormalise' the ensemble of simulations by rewriting the ensemble in terms of an *orthonormal basis* of 100 vectors. Such an orthonormal basis for the ensemble reasonably ensures unusual storms are retained in the range of possibilities explored by the ensemble forecast.

[17] http://en.wikipedia.org/wiki/1999_Sydney_hailstorm [April 2015]

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

3.4.3 Is it a line? a plane? The dimension answers

> *physical dimension.* It is an intuitive notion that appears to go back to an archaic state before Greek geometry, yet deserves to be taken up again. *Mandelbrot (1982)*

One of the beauties of an orthonormal basis is that, being orthonormal, they look just like a rotated version of the standard unit vectors. That is, the two orthonormal basis of a plane could form the two 'standard unit vectors' of a coordinate system in that plane. Example 3.4.25 found the plane $z = -x/6 + y/3$ could have the following two orthonormal bases: either of these orthonormal bases, or indeed any other pair of orthonormal vectors, could act as a pair of 'standard unit vectors' of the given planar subspace.

Similarly in other dimensions for other subspaces. Just as $\mathbb{R}^n$ is called n-dimensional and has n standard unit vectors, so we analogously define the dimension of any subspace.

Definition 3.4.30. *Let $\mathbb{W}$ be a subspace of $\mathbb{R}^n$. The number of vectors in an orthonormal basis for $\mathbb{W}$ is called the* ***dimension*** *of $\mathbb{W}$, denoted $\dim \mathbb{W}$. By convention, $\dim\{\mathbf{0}\} = 0$.*

Example 3.4.31.

- Example 3.4.21 finds that the linear subspace $x = y = z$ is spanned by the orthonormal basis $\{(1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3})\}$. With one vector in the basis, the line is one dimensional.
- Example 3.4.22 finds that the planar subspace $-x+2y-2z = 0$ is spanned by the orthonormal basis $\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$ 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 the vectors $(-0.99, -0.01, 0.16)$ and $(-0.04, -0.95, -0.31)$. With two vectors in the basis, the subspace is two dimensional; that is, $\dim \mathbb{W} = 2$.
- Since the subspace $\mathbb{R}^n$ (Example 3.4.4g) has an orthonormal basis of the n standard unit vectors, $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n\}$, then $\dim \mathbb{R}^n = n$.
- The El Nino windowed data of Example 3.4.27 is effectively spanned by four orthonormal vectors. Despite the apparent

complexity of the signal, the data effectively lies in a subspace of dimension four (that of two oscillators).

■

Theorem 3.4.32. *The row space and column space of a matrix A have the same dimension. Further, given an* SVD *of the matrix, say $A = USV^{\mathrm{T}}$, an orthonormal basis for the column space is the first* $\operatorname{rank} A$ *columns of U, and that for the row space is the first* $\operatorname{rank} A$ *columns of V.*

Proof. From Definition 3.1.16 of the transpose, the rows of A are the same as the columns of A^{T}, and so the row space of A is the same as the column space of A^{T}. Hence,

$$\begin{aligned}
&\text{dimension of the row space of } A\\
&= \text{dimension of the column space of } A^{\mathrm{T}}\\
&= \operatorname{rank}(A^{\mathrm{T}}) \quad \text{(by Proc. 3.4.23)}\\
&= \operatorname{rank} A \quad \text{(by Thm. 3.3.23)}\\
&= \text{dimension of the column space of } A \quad \text{(by Proc. 3.4.23).}
\end{aligned}$$

Let $m \times n$ matrix A have an SVD $A = USV^{\mathrm{T}}$ and $r = \operatorname{rank} A$. Then Procedure 3.4.23 establishes that an orthonormal basis for the column space of A is the first r columns of U. Recall that $A^{\mathrm{T}} = (USV^{\mathrm{T}})^{\mathrm{T}} = VS^{\mathrm{T}}U^{\mathrm{T}}$ is an SVD for A^{T} (Theorem 3.3.23) and so an orthonormal basis for the column space of A^{T} is the first r columns of V (Procedure 3.4.23). Since the row space of A is the column space of A^{T}, an orthonormal basis for the row space of A 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.

Solution: Ask MATLAB/Octave for an SVD and interpret:

```
A=[1 -4; 1/2 -4]
[U,S,V]=svd(A)
```

computes the SVD

```
U =
  -0.8944  -0.4472
  -0.4472   0.8944
S =
   4.6098        0
        0   0.0000
V =
  -0.2425   0.9701
   0.9701   0.2425
```

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

There is one non-zero singular value—the matrix has rank one—so an orthonormal basis for the column space is the first column of matrix U, namely $(-0.89, -0.45)$ (2 d.p.).

Complementing this, as there is one non-zero singular value—the matrix has rank one—so an orthonormal basis for the row space is the first column of matrix V, namely $(-0.24, 0.97)$. As illustrated in the margin, the two subspaces, the row space (red) and the column space (blue), are different but of the same dimension. (As in general, the row and column spaces are not orthogonal.)

■

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

■

Example 3.4.35. Use the SVD of the matrix B in Example 3.4.25 to compare the column space and the row space of matrix B.

Solution: Recall that there are two non-zero singular values—the matrix has rank two—so an orthonormal basis for the column space is the first two columns of matrix U, namely the vectors $(-0.99, -0.01, 0.16)$ and $(-0.04, -0.95, -0.31)$.

Complementing this, as there are two non-zero singular values—the matrix has rank two—so an orthonormal basis for the row space is the set of the first two columns of matrix V, namely the vectors $(-0.78, -0.02, 0.63)$ and $(-0.28, 0.91, -0.32)$. As illustrated below in stereo, the two subspaces, the row space (red) and the column space (blue), are different but of the same dimension.

■

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 non-zero singular values, there are two orthonormal vectors spanning the subspace, the nullspace, hence its dimension, the nullity, is two.

■

Example 3.4.38. For the matrix

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

find an orthonormal basis for its nullspace and hence determine its nullity.

Solution: To find the nullspace construct a general solution to the homogeneous system $C\boldsymbol{x} = \mathbf{0}$ with Procedure 3.3.15.

(a) Enter into MATLAB/Octave the matrix C and compute an SVD via `[U,S,V]=svd(C)` to find (2 d.p.)

```
U =
   0.24   0.78  -0.58
  -0.55   0.60   0.58
   0.80   0.18   0.58
S =
   6.95      0      0      0
      0   3.43      0      0
      0      0   0.00      0
V =
   0.43  -0.65   0.63  -0.02
```

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

```
-0.88  -0.19   0.42   0.10
 0.11   0.68   0.62  -0.37
 0.15   0.28   0.21   0.92
```

(b) Since the right-hand side of $C\boldsymbol{x} = \mathbf{0}$ is zero the solution to $U\boldsymbol{z} = \mathbf{0}$ is $\boldsymbol{z} = \mathbf{0}$.

(c) Then, because the rank of the matrix is two, the solution to $S\boldsymbol{y} = \boldsymbol{z} = \mathbf{0}$ is $\boldsymbol{y} = (0, 0, y_3, y_4)$ for free variables y_3 and y_4.

(d) The solution to $V^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{y}$ is $\boldsymbol{x} = V\boldsymbol{y} = \boldsymbol{v}_3 y_3 + \boldsymbol{v}_4 y_4 = y_3(0.63, 0.42, 0.62, 0.21) + y_4(-0.02, 0.10, -0.37, 0.92)$.

Hence $\operatorname{span}\{(0.63, 0.42, 0.62, 0.21), (-0.02, 0.10, -0.37, 0.92)\}$ is the nullspace of matrix C. Because the columns of V are orthonormal, the two vectors appearing in this span are orthonormal and so form an orthonormal basis for the nullspace. Hence $\operatorname{nullity} C = 2$.

■

This Example 3.4.38 indicates that the nullity is determined by the number of zero columns in the diagonal matrix S of an SVD. Conversely, the rank of a matrix is determined by the number of non-zero columns in the diagonal matrix S of an SVD. Put these two facts together in general and we get the following theorem that helps characterise solutions of linear equations.

Theorem 3.4.39 (rank theorem). *Let A be an $m \times n$ matrix, then $\operatorname{rank} A + \operatorname{nullity} A = n$, the number of columns of A.*

Proof. Set $r = \operatorname{rank} A$. By Procedure 3.3.15 a general solution to the homogeneous system $A\boldsymbol{x} = \mathbf{0}$ involves $n - r$ free variables $y_{r+1}, \dots, y_n$ in the linear combination form $\boldsymbol{v}_{r+1} y_{r+1} + \cdots + \boldsymbol{v}_n y_n$. Hence the nullspace is $\operatorname{span}\{\boldsymbol{v}_{r+1}, \dots, \boldsymbol{v}_n\}$. Because matrix V is orthogonal, the vectors $\boldsymbol{v}_{r+1}, \dots, \boldsymbol{v}_n$ are orthonormal; that is, they form an orthonormal basis for the nullspace, and so the nullspace is of dimension $n - r$. Consequently, $\operatorname{rank} A + \operatorname{nullity} A = r + (n - r) = n$. □

Example 3.4.40. Compute SVDs to determine the rank and nullity of each of the given matrices.

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

Solution: Enter the matrix into MATLAB/Octave and compute the singular values:

```
A=[1 -1 2
2 -2 4]
svd(A)
```

The resultant singular values are

```
5.4772
0.0000
```

The one non-zero singular value indicates rank $A = 1$. Since the matrix has three columns, the nullity—the dimension of the nullspace—is $3 - 1 = 2$.

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

Solution: Enter the matrix into MATLAB/Octave and compute the singular values:

```
B=[1 -1 -1
1 0 -1
-1 3 1]
svd(B)
```

The resultant singular values are

```
3.7417
1.4142
0.0000
```

The two non-zero singular values indicate rank $B = 2$. Since the matrix has three columns, the nullity—the dimension of the nullspace—is $3 - 2 = 1$.

(c) $\begin{bmatrix} 0 & 0 & -1 & -3 & 2 \\ -2 & -2 & 1 & 0 & 1 \\ 1 & -1 & 2 & 8 & -2 \\ -1 & 1 & 0 & -2 & -2 \\ -3 & -1 & 0 & -5 & 1 \end{bmatrix}$

Solution: Enter the matrix into MATLAB/Octave and compute the singular values:

```
C=[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]
svd(C)
```

The resultant singular values are

```
10.8422
 4.0625
 3.1532
 0.0000
 0.0000
```

v0.2a

Three non-zero singular values indicate $\operatorname{rank} C = 3$. Since the matrix has five columns, the nullity—the dimension of the nullspace—is $5 - 3 = 2$.

■

Activity 3.4.41. The matrix

$$\begin{bmatrix} -2 & 1 & 4 & 0 & -4 \\ -1 & 1 & 0 & -2 & 0 \\ -3 & 1 & 3 & 2 & -3 \\ 0 & 0 & 1 & 0 & -1 \end{bmatrix} \text{ has singular values } \begin{matrix} \texttt{8.1975} \\ \texttt{2.6561} \\ \texttt{1.6572} \\ \texttt{0.0000} \end{matrix}$$

computed with `svd()`. What is its nullity?

(a) 0 (b) 1 (c) 2 (d) 3

■

Example 3.4.42. Each of the following graphs plot all the column vectors of a matrix. What is the nullity of each of the matrices? Give reasons.

(a)

Solution: Zero. These two column vectors in the plane must come from a 2×2 matrix A. Since the two columns are at a non-trivial angle, every point in the plane may be written as a linear combination of $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$, hence the column space of A is $\mathbb{R}^2$. Consequently, $\operatorname{rank} A = 2$. From the rank theorem: $\operatorname{nullity} A = n - \operatorname{rank} A = 2 - 2 = 0$.

(b)

Solution: Two. These three column vectors in the plane must come from a 2×3 matrix B. The three vectors are all in a line, so the column space of matrix B is a line. Consequently, $\operatorname{rank} B = 1$. From the rank theorem: $\operatorname{nullity} B = n - \operatorname{rank} B = 3 - 1 = 2$.

v0.2a

(c)

a stereo pair.

Solution: One. These four column vectors in 3D space must come from a 3×4 matrix C. Since the four columns do not all lie in a line or plane, every point in space may be written as a linear combination of $\boldsymbol{c}_1, \boldsymbol{c}_2, \dots, \boldsymbol{c}_4$, hence the column space of C is $\mathbb{R}^3$. Consequently, $\operatorname{rank} C = 3$. From the rank theorem: $\operatorname{nullity} C = n - \operatorname{rank} C = 4 - 3 = 1$.

(d)

a stereo pair.

Solution: Two. These four column vectors in 3D space must come from a 3×4 matrix D. Since the four columns all lie in a plane (as suggested by the drawn plane), and linear combinations can give every point in the plane, hence the column space of D has dimension two. Consequently, $\operatorname{rank} D = 2$. The rank theorem gives $\operatorname{nullity} D = n - \operatorname{rank} D = 4 - 2 = 2$.

■

The recognition of these new concepts associated with matrices and linear equations, then empowers us to extend the list of exact properties that ensure a system of linear equations has a unique solution.

Theorem 3.4.43 (Unique Solutions: version 2)**.** *Let A be an $n \times n$ square matrix. Extending Theorem 3.3.26, 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;

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(e) $\operatorname{rank} A = n$;

(f) $\operatorname{nullity} A = 0$;

(g) the column vectors of A span $\mathbb{R}^n$;

(h) the row vectors of A span $\mathbb{R}^n$.

Proof. Theorem 3.3.26 establishes the equivalence of the statements 3.4.43a–3.4.43e. We prove the equivalence of these with the statements 3.4.43f–3.4.43h.

3.4.43e $\iff$ 3.4.43f : The Rank Theorem 3.4.39 assures us that $\operatorname{nullity} A = 0$ iff $\operatorname{rank} A = n$.

3.4.43b $\implies$ 3.4.43g : By 3.4.43b every $\boldsymbol{b} \in \mathbb{R}^n$ can be written as $\boldsymbol{b} = A\boldsymbol{x}$. But $A\boldsymbol{x}$ is a linear combination of the columns of A and so $\boldsymbol{b}$ is in the span of the columns. Hence the column vectors of A span $\mathbb{R}^n$.

3.4.43g $\implies$ 3.4.43e : Suppose $\operatorname{rank} A = r$ reflecting r non-zero singular values in an SVD $A = USV^{\mathrm{T}}$. Procedure 3.4.23 assures us the column space of A has orthonormal basis $\{\boldsymbol{u}_1 , \boldsymbol{u}_2 , \dots , \boldsymbol{u}_r\}$. But the column space is $\mathbb{R}^n$ (statement 3.4.43g) which also has the orthonormal basis of the n standard unit vectors. Theorem 3.4.28 assures us that the number of basis vectors must be the same; that is, $\operatorname{rank} A = r = n$.

3.4.43e $\iff$ 3.4.43h : Theorem 3.3.23 asserts $\operatorname{rank}(A^{\mathrm{T}}) = \operatorname{rank} A$, so the statement 3.4.43e implies $\operatorname{rank}(A^{\mathrm{T}}) = n$, and so statement 3.4.43g asserts the columns of A^{T} span $\mathbb{R}^n$. But the columns of A^{T} are the rows of A so the rows of A span $\mathbb{R}^n$. Conversely, if the rows of A span $\mathbb{R}^n$, then so do the columns of A^{T}, hence $\operatorname{rank}(A^{\mathrm{T}}) = n$ which by Theorem 3.3.23 implies $\operatorname{rank} A = n$.

□

3.4.4 Exercises

Exercise 3.4.1. Use your intuitive notion of a subspace to decide whether each of the following drawn sets (3D in stereo pair) is a subspace, or not.

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

(i)

(j)

(k)

v0.2a

(l)

(m)

(n)

(o)

(p)

Exercise 3.4.2. Use Definition 3.4.3 to decide whether each of the following is a subspace, or not. Give reasons.

(a) All vectors in the line $y = 2x$.

(b) All vectors in the line $3.2y = 0.8x$.

(c) All vectors $(x, y) = (t, 2 + t)$ for all real t.

(d) All vectors $(1.3n, -3.4n)$ for all integer n.

(e) All vectors $(x, y) = (-3.3 - 0.3t, 2.4 - 1.8t)$ for all real t.

(f) $\operatorname{span}\{(6, -1), (1, 2)\}$

(g) All vectors $(x, y) = (6 - 3t, t - 2)$ for all real t.

(h) The vectors $(2,1,-3)t+(5,-\frac{1}{2},2)s$ for all real s,t.

(i) The vectors $(0.9,2.4,1)t-(0.2,0.6,0.3)s$ for all real s,t.

(j) All vectors (x,y) such that $y=x^3$.

(k) All vectors (x,y,z) such that $x=2t$, $y=t^2$ and $z=t/2$ for all t.

(l) The vectors $(t,n,2t+3n)$ for real t and integer n.

(m) $\operatorname{span}\{(0,-1,1),(-1,0,2)\}$

(n) The vectors $(2.7,2.6,-0.8,2.1)s+(0.5,0.1,-1,3.3)t$ for all real s,t.

(o) The vectors $(1.4,2.3,1.5,4)+(1.2,-0.8,-1.2,2)t$ for all real t.

(p) The vectors $(t^3,2t^3)$ for all real t (tricky).

(q) The vectors $(t^2,3t^2)$ for all real t (tricky).

Exercise 3.4.3. Let $\mathbb{W}_1$ and $\mathbb{W}_2$ be any two subspaces of $\mathbb{R}^n$ (Definition 3.4.3).

(a) Use the definition to prove that the intersection of $\mathbb{W}_1$ and $\mathbb{W}_2$ is also a subspace of $\mathbb{R}^n$.

(b) Give an example to prove that the union of $\mathbb{W}_1$ and $\mathbb{W}_2$ is not necessarily a subspace of $\mathbb{R}^n$.

Exercise 3.4.4. For each of the following matrices, partially solve linear equations to determine whether the given vector $\boldsymbol{b}_j$ is in the column space, and to determine if the given vector $\boldsymbol{r}_j$ is in the row space of the matrix. Work small problems by hand, and address larger problems with MATLAB/Octave. Record your working or MATLAB/Octave commands and output.

(a) $A=\begin{bmatrix}2&1\\5&4\end{bmatrix}$, $\boldsymbol{b}_1=\begin{bmatrix}3\\-2\end{bmatrix}$, $\boldsymbol{r}_1=\begin{bmatrix}1\\0\end{bmatrix}$

(b) $B=\begin{bmatrix}-2&1\\4&-2\end{bmatrix}$, $\boldsymbol{b}_2=\begin{bmatrix}1\\-3\end{bmatrix}$, $\boldsymbol{r}_2=\begin{bmatrix}0\\0\end{bmatrix}$

(c) $C=\begin{bmatrix}1&-1\\-3&4\\-3&5\end{bmatrix}$, $\boldsymbol{b}_3=\begin{bmatrix}5\\0\\-1\end{bmatrix}$, $\boldsymbol{r}_3=\begin{bmatrix}0\\1\end{bmatrix}$

(d) $D=\begin{bmatrix}-2&-4&-5\\-6&-2&1\end{bmatrix}$, $\boldsymbol{b}_4=\begin{bmatrix}3\\1\end{bmatrix}$, $\boldsymbol{r}_4=\begin{bmatrix}2\\-6\\-11\end{bmatrix}$

(e) $E=\begin{bmatrix}3&2&4\\1&6&0\\1&-2&2\end{bmatrix}$, $\boldsymbol{b}_5=\begin{bmatrix}10\\2\\4\end{bmatrix}$, $\boldsymbol{r}_5=\begin{bmatrix}0\\-1\\2\end{bmatrix}$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(f) $F = \begin{bmatrix} 0 & -1 & -4 \\ -2 & 0 & 4 \\ 7 & -1 & -3 \\ 1 & 1 & 3 \end{bmatrix}$, $\boldsymbol{b}_6 = \begin{bmatrix} 2 \\ 3 \\ 1 \\ 3 \end{bmatrix}$, $\boldsymbol{r}_6 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$

(g) $G = \begin{bmatrix} -2 & -1 & 5 & 4 \\ 0 & -3 & 1 & -1 \\ 3 & 3 & 4 & 3 \end{bmatrix}$, $\boldsymbol{b}_7 = \begin{bmatrix} 1 \\ 2 \\ -4 \end{bmatrix}$, $\boldsymbol{r}_7 = \begin{bmatrix} 0 \\ -3 \\ 1 \\ -1 \end{bmatrix}$

(h) $H = \begin{bmatrix} -2 & 1 & 1 & -1 \\ 2 & -2 & -1 & 0 \\ -2 & 1 & 1 & -2 \\ 2 & 5 & -1 & -1 \end{bmatrix}$, $\boldsymbol{b}_8 = \begin{bmatrix} 2 \\ -1 \\ 3 \\ 0 \end{bmatrix}$, $\boldsymbol{r}_8 = \begin{bmatrix} -1 \\ -2 \\ 0 \\ -1 \end{bmatrix}$

(i) $\mathcal{I} = \begin{bmatrix} 1.0 & 0.8 & 2.1 & 1.4 \\ 0.6 & -0.1 & 2.1 & 1.8 \\ 0.1 & -0.1 & 2.1 & 1.2 \\ 1.7 & -1.1 & -1.9 & 2.9 \end{bmatrix}$, $\boldsymbol{b}_9 = \begin{bmatrix} 4.3 \\ 1.2 \\ 0.5 \\ 0.3 \end{bmatrix}$, $\boldsymbol{r}_9 = \begin{bmatrix} 0.0 \\ 1.5 \\ -0.5 \\ 1.0 \end{bmatrix}$

Exercise 3.4.5. In each of the following, is the given vector in the nullspace of the given matrix?

(a) $A = \begin{bmatrix} -11 & -2 & 5 \\ -1 & 1 & 1 \end{bmatrix}$, $\boldsymbol{p} = \begin{bmatrix} -7 \\ 6 \\ -13 \end{bmatrix}$

(b) $B = \begin{bmatrix} 3 & -3 & 2 \\ 1 & 1 & -3 \end{bmatrix}$, $\boldsymbol{q} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

(c) $C = \begin{bmatrix} -5 & 0 & -2 \\ -6 & 2 & -2 \\ 0 & -5 & -1 \end{bmatrix}$, $\boldsymbol{r} = \begin{bmatrix} -2 \\ -1 \\ 5 \end{bmatrix}$

(d) $D = \begin{bmatrix} -3 & -2 & 0 & 2 \\ 5 & 0 & 1 & -2 \\ 4 & -4 & 4 & 2 \end{bmatrix}$, $\boldsymbol{s} = \begin{bmatrix} 6 \\ 1 \\ -10 \\ 10 \end{bmatrix}$

(e) $E = \begin{bmatrix} -3 & 2 & 3 & 1 \\ -3 & -2 & -1 & 4 \\ 6 & 1 & -1 & -1 \end{bmatrix}$, $\boldsymbol{t} = \begin{bmatrix} 2 \\ -4 \\ 1 \\ -2 \end{bmatrix}$

(f) $F = \begin{bmatrix} -4 & -2 & 2 & -2 \\ 2 & -1 & -2 & 1 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & -8 & -2 \end{bmatrix}$, $\boldsymbol{u} = \begin{bmatrix} 11 \\ -2 \\ 4 \\ -16 \end{bmatrix}$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Exercise 3.4.6. Given the SVDs of Exercises 3.3.2 and 3.3.3, write down an orthonormal basis for the span of the following sets of vectors.

(a) $(-\frac{9}{5}, -4)$, $(\frac{12}{5}, -3)$

(b) $(-0.96, -0.72)$, $(1.28, 0.96)$

(c) $(7, -22, -4)/39$, $(-34, -38, -53)/78$

(d) $(4, 4, 2)/33$, $(4, 4, 2)/11$, $(6, 6, 3)/11$

(e) $(-\frac{2}{5}, \frac{11}{9}, \frac{31}{90}, \frac{4}{9})$, $(-\frac{2}{5}, \frac{11}{9}, \frac{31}{90}, \frac{4}{9})$, $(-\frac{26}{45}, -\frac{1}{3}, \frac{17}{90}, -\frac{2}{9})$, $(\frac{26}{45}, \frac{1}{3}, -\frac{17}{90}, \frac{2}{9})$

Exercise 3.4.7. Given any $m \times n$ matrix A.

(a) Explain how Procedure 3.4.23 uses an SVD to find an orthonormal basis for the column space of A.

(b) How does the same SVD give the orthonormal basis $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_r\}$ for the row space of A? Justify your answer.

(c) Why does the same SVD also give the orthonormal basis $\{\boldsymbol{v}_{r+1}, \ldots, \boldsymbol{v}_n\}$ for the nullspace of A? Justify.

Exercise 3.4.8. For each of the following matrices, compute an SVD with MATLAB/Octave, and then use the properties of Exercise 3.4.7 to write down an orthonormal basis for the column space, the row space, and the nullspace of the matrix. (The bases, especially for the nullspace, may differ in detail depending upon your version of MATLAB/Octave.)

(a) $\begin{bmatrix} 19 & -36 & -18 \\ -3 & 12 & 6 \\ -17 & 48 & 24 \end{bmatrix}$

(b) $\begin{bmatrix} -12 & 0 & -4 \\ -30 & -6 & 4 \\ 34 & 22 & 8 \\ -50 & -10 & 12 \end{bmatrix}$

(c) $\begin{bmatrix} 0 & 0 & 0 & 0 \\ 4 & 10 & 1 & -3 \\ 2 & 6 & 0 & -2 \\ -2 & -4 & -1 & 1 \end{bmatrix}$

(d) $\begin{bmatrix} -13 & 9 & 10 & -4 & -6 \\ -7 & 27 & -2 & 4 & -10 \\ -4 & 0 & 4 & 4 & -4 \\ -4 & -18 & 10 & -8 & 5 \end{bmatrix}$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(e) $\begin{bmatrix} 1 & -2 & 3 & 9 \\ -1 & 5 & 0 & 0 \\ 0 & 3 & 3 & 9 \\ 2 & -9 & 1 & 3 \\ 1 & -7 & -2 & -6 \end{bmatrix}$

(f) $\begin{bmatrix} 9 & 3 & 0 & -9 \\ 15 & 1 & 24 & -15 \\ -12 & -4 & 0 & 12 \\ 9 & 3 & 0 & -9 \\ -3 & -1 & 0 & 3 \\ 11 & 5 & -8 & -11 \end{bmatrix}$

(g) $\begin{bmatrix} -8 & 17 & 7 & -51 & 20 \\ 5 & -2 & -1 & 15 & -2 \\ 15 & -30 & -15 & 75 & -30 \\ -2 & -1 & -5 & -33 & 8 \end{bmatrix}$

(h) $\begin{bmatrix} -128 & 6 & 55 & -28 & -1 \\ 20 & 12 & -31 & 18 & -3 \\ -12 & -30 & 39 & -24 & 7 \\ -1 & 6 & -1 & 7 & -3 \end{bmatrix}$

Exercise 3.4.9. For each of the matrices in Exercise 3.4.8, from your computed bases write down the dimension of the column space, the row space and the nullspace. Comment on how these confirm the rank theorem 3.4.39.

Exercise 3.4.10. What are the possible values for nullity(A) in the following cases?

(a) A is a 2×5 matrix.

(b) A is a 3×3 matrix.

(c) A is a 3×2 matrix.

(d) A is a 4×6 matrix.

(e) A is a 4×4 matrix.

(f) A is a 6×5 matrix.

Exercise 3.4.11 (Cowen (1997))**.** Alice and Bob are taking linear algebra. One of the problems in their homework assignment is to find the nullspace of a 4×5 matrix A. In each of the following cases: are their answers consistent with each other? Give reasons.

(a) Alice's answer is that the nullspace is spanned by $(-2,-2,0,2,-6)$, $(1,5,4,-3,11)$, $(3,5,2,-4,13)$, and $(0,-2,-2,1,-4)$. Bob's answer is that the nullspace is spanned by $(1,1,0,-1,3)$, $(-2,0,2,1,-2)$, and $(-1,3,4,1,5)$.

(b) Alice's answer is that the nullspace is spanned by $(2,-3,1,-2,-5)$, $(2,-7,2,-1,-6)$, $(1,-2,1,1,0)$, $(3,-6,3,3,0)$.

v0.2a

Figure 3.4: vertical acceleration of the ankle of a person walking normally, a person who has Parkinson's Disease. The data is recorded 0.125 s apart (here subsampled and smoothed for the purposes of the exercise).

Bob's answer is that the nullspace is spanned by $(1,-2,1,1,0)$, $(0,4,-1,-1,1)$, $(1,-1,0,-3,-5)$.

(c) Alice's answer is that the nullspace is spanned by $(-2,0,-2,4,-5)$, $(0,2,-2,2,-2)$, $(0,-2,2,-2,2)$, $(-4,-12,8,-4,2)$. Bob's answer is that the nullspace is spanned by $(0,2,-2,2,-2)$, $(-2,-4,2,0,-1)$, $(1,0,-1,1,-3)$.

(d) Alice's answer is that the nullspace is spanned by $(-1,0,0,0,0)$, $(5,3,-2,5,1)$, $(-5,1,0,-6,-2)$, $(4,-2,0,1,8)$. Bob's answer is that the nullspace is spanned by $(1,-2,0,-3,4)$, $(2,-1,1,3,2)$, $(3,0,-1,2,3)$.

Exercise 3.4.12. Prove that if the columns of a matrix A are orthonormal, then they must form an orthonormal basis for the column space of A.

Exercise 3.4.13. Let A be any $m \times n$ matrix. Use an SVD to prove that every vector in the row space of A is orthogonal to every vector in the nullspace.

Exercise 3.4.14. Bachlin et al. [*IEEE Transactions on Information Technology in Biomedicine*, 14(2), 2010] explored the walking gait of people with Parkinson's Disease. Among many measurements, they measured the vertical ankle acceleration of the people when they walked. Figure 3.4 shows ten seconds of just one example: use the so-called Singular Spectrum Analysis to find the regular structures in this complex data.

Following Example 3.4.27:

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(a) enter the data into MATLAB/Octave;

```
time=(0.0625:0.125:9.85)'
acc=[5.34; 0.85; -1.90; -1.39; -0.99; 5.64;
-1.76; -5.90; 4.74; 1.85; -2.09; -1.16; -1.58;
5.19; -0.27; -6.54; 3.94; 2.70; -2.36; -0.94;
-1.68; 4.61; 0.79; -6.90; 3.23; 3.59; -2.65;
-0.99; -1.65; 4.09; 1.78; -7.11; 2.26; 4.27;
-2.53; -0.84; -1.84; 3.34; 2.62; -6.74; 1.54;
4.16; -2.29; -0.50; -1.97; 2.80; 2.92; -6.37;
1.09; 4.17; -2.05; -0.44; -2.03; 2.08; 3.91;
-5.84; -0.78; 4.98; -1.28; -0.94; -1.86; 0.50;
5.40; -4.19; -3.88; 5.45; 0.44; -1.71; -1.59;
-0.90; 5.86; -1.95; -5.95; 4.75; 1.90; -2.06;
-1.21; -1.61; 5.16]
```

(b) use the `hankel()` function to form a matrix whose 66 columns are 66 'windows' of accelerations, each of length fourteen data points (of length 1.75 s);

(c) compute an SVD of the matrix, and explain why the windows of measured accelerations are close to lying in a four dimensional subspace;

(d) plot orthonormal basis vectors for the four dimensional subspace of the windowed accelerations.

Exercise 3.4.15. Consider $m \times n$ matrices bordered by zeros in the block form

$$E = \begin{bmatrix} F & O_{k\times n-\ell} \\ O_{m-k\times \ell} & O_{m-k\times n-\ell} \end{bmatrix}$$

where F is some $k \times \ell$ matrix. Given matrix F has an SVD, find an SVD of matrix E, and hence prove that $\operatorname{rank} E = \operatorname{rank} F$.

Exercise 3.4.16. For compatibly sized matrices A and B, use their SVDs, and the result of the previous Exercise 3.4.15 (applied to the matrix $S_A V_A^{\mathrm{T}} U_B S_B$), to prove that $\operatorname{rank}(AB) \le \operatorname{rank} A$ and that $\operatorname{rank}(AB) \le \operatorname{rank} B$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

3.5 Project to solve inconsistent equations

Section Contents

3.5.1 Make a minimal change to the problem . . . 299
3.5.2 Compute the smallest appropriate solution . 314
3.5.3 Orthogonal projection resolves vector components . . . 321
Project onto a direction . . . 321
Project onto a subspace . . . 324
Orthogonal decomposition separates . . . 335
3.5.4 Exercises . . . 349

> Agreement with experiment is the sole criterion of truth for a physical theory. *Pierre Duhem, 1906*

As well as being fundamental to engineering, scientific and computational inference, approximately solving inconsistent equations also introduces the linear transformation of projection.

The scientific method is to infer general laws from data and then validate the laws. This section addresses the inference of general laws from data. A big challenge is that data is typically corrupted by noise and errors. So this section shows how the singular value decomposition (SVD) leads to understanding the so-called least square methods.

3.5.1 Make a minimal change to the problem

Example 3.5.1 (rationalise contradictory weights)**.** I weighed myself the other day. I weighed myself four times, each time separated by a few minutes: the scales reported my weight in kg as 84.8, 84.1, 84.7 and 84.4 . The measurements give four different weights! What sense can we make of this apparently contradictory data? Traditionally we just average and say my weight is $x \approx (84.8+84.1+84.7+84.4)/4 = 84.5\,\text{kg}$. Let's see this same answer from a new linear algebra view.

In the linear algebra view my weight x is an unknown and the four experimental measurements give four equations for this one unknown:

$$x = 84.8\,, \quad x = 84.1\,, \quad x = 84.7\,, \quad x = 84.4\,.$$

Despite being manifestly impossible to satisfy all four equations, let's see what linear algebra can do for us. Linear algebra writes these four equations as the matrix-vector system

$$Ax = \boldsymbol{b}\,, \quad \text{namely} \quad \begin{bmatrix}1\\1\\1\\1\end{bmatrix} x = \begin{bmatrix}84.8\\84.1\\84.7\\84.4\end{bmatrix}.$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

The linear algebra Procedure 3.3.15 is to 'solve' this system, despite its contradictions, via an SVD and some intermediaries:

$$Ax = U\underbrace{S\overbrace{V^{\mathrm{T}}x}^{=y}}_{=\boldsymbol{z}} = \boldsymbol{b}\,.$$

(a) You are given that this matrix A of ones has an SVD of (perhaps check the columns of U are orthonormal)

$$A = \begin{bmatrix}1\\1\\1\\1\end{bmatrix} = \begin{bmatrix}\frac12&\frac12&\frac12&\frac12\\\frac12&\frac12&-\frac12&-\frac12\\\frac12&-\frac12&-\frac12&\frac12\\\frac12&-\frac12&\frac12&-\frac12\end{bmatrix}\begin{bmatrix}2\\0\\0\\0\end{bmatrix}\begin{bmatrix}1\end{bmatrix}^{\mathrm{T}} = USV^{\mathrm{T}}.$$

(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,

$$\begin{bmatrix}2\\0\\0\\0\end{bmatrix} y = \begin{bmatrix}169\\-0.1\\0.2\\0.5\end{bmatrix}.$$

But we cannot because the last three components in the equation are impossible: we cannot satisfy any of

$$0y = -0.1\,,\quad 0y = 0.2\,,\quad 0y = 0.5\,.$$

Instead of seeking an *exact* solution, ask what is the *smallest change* we can make to $\boldsymbol{z} = (169, -0.1, 0.2, 0.5)$ so that we can report a solution to a slightly different problem? Answer: we *have to* adjust the last three components to zero. Also, any adjustment to the first component is unnecessary, 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 $y = 84.5$. We treat this exact solution to a slightly different problem, as an *approximate* solution to the original problem.

v0.2a

(d) Lastly, solve $V^{\mathrm{T}}x = y$ by computing $x = Vy = 1y = y = 84.5\,\mathrm{kg}$ (upon including the physical units). That is, this linear algebra procedure gives my weight as $x = 84.5\,\mathrm{kg}$ (approximately).

This linear algebra procedure recovers the traditional answer of averaging measurements. ■

The answer to the previous Example 3.5.1 illustrates how traditional averaging emerges from trying to make sense of apparently inconsistent information. Importantly, the principle of making the smallest possible change to the intermediary $\boldsymbol{z}$ 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 the same 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

$$\begin{bmatrix}3\\4\end{bmatrix} x = \begin{bmatrix}1\\3\end{bmatrix}, \quad \text{and where} \quad \begin{bmatrix}3\\4\end{bmatrix} = \begin{bmatrix}\frac{3}{5} & \frac{4}{5}\\ \frac{4}{5} & -\frac{3}{5}\end{bmatrix}\begin{bmatrix}5\\0\end{bmatrix}\begin{bmatrix}1\end{bmatrix}^{\mathrm{T}}$$

is an SVD factorisation of the 2×1 matrix. Following the procedure of the previous example, what is the 'best' approximate solution to these inconsistent equations?

(a) $x = 1/3$ (b) $x = 4/7$ (c) $x = 3/5$ (d) $x = 3/4$

■

Example 3.5.3. Recall the table tennis player rating Example 3.3.13. There we found that we could not solve the equations to find some ratings because the equations were inconsistent. In our new terminology of the previous Section 3.4, the right-hand side vector $\boldsymbol{b}$ is not in the column space of the matrix A (Definition 3.4.10): the stereo picture below illustrates the 2D column space spanned by the three columns of A and that the vector $\boldsymbol{b}$ lies outside the column space.

Now reconsider step 3 in Example 3.3.13.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(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 to zero. But any change to the first two components is unnecessary, would make the change bigger than necessary, and so we do not change the first two components. Hence find an approximate solution to the player ratings via solving

$$\begin{bmatrix} 1.7321 & 0 & 0 \\ 0 & 1.7321 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{y} = \begin{bmatrix} -2.0412 \\ -2.1213 \\ 0 \end{bmatrix}.$$

Here a general soution is $\boldsymbol{y} = (-1.1785, -1.2247, y_3)$ from `y=z(1:2)./diag(S(1:2,1:2))`. Varying the free variable y_3 gives equally good solutions as approximations.

(b) Lastly, solve $V^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{y}$, via computing `x=V(:,1:2)*y`, to determines

$$\begin{aligned} \boldsymbol{x} = V\boldsymbol{y} &= \begin{bmatrix} 0.0000 & -0.8165 & 0.5774 \\ -0.7071 & 0.4082 & 0.5774 \\ 0.7071 & 0.4082 & 0.5774 \end{bmatrix} \begin{bmatrix} -1.1785 \\ -1.2247 \\ y_3 \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ \frac{1}{3} \\ -\frac{4}{3} \end{bmatrix} + \frac{y_3}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}. \end{aligned}$$

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 the '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. *factorise $A = USV^{\mathrm{T}}$ and set $r = \operatorname{rank} A$ (remembering that relatively small singular values are effectively zero);*

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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	-

2. *solve* $U\boldsymbol{z} = \boldsymbol{b}$ *by* $\boldsymbol{z} = U^{\mathrm{T}}\boldsymbol{b}$*;*

3. *disregard the equations for* $i = r+1, \dots, m$ *as errors, set* $y_i = z_i/\sigma_i$ *for* $i = 1, \dots, r$ *(as these* $\sigma_i > 0$*), and otherwise* y_i *is free for* $i = r+1, \dots, n$ *;*

4. *solve* $V^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{y}$ *to obtain a general approximate solution as* $\boldsymbol{x} = V\boldsymbol{y}$*.*

Example 3.5.5 (round robin tournament)**.** Consider four players (or teams) that play in a round robin sporting event: Anne, Bob, Chris and Dee. Table 3.4 summarises the results of the six games played. From these results estimate the relative player ratings of the four players. As in many real-life situations, the information appears contradictory such as Anne beats Bob, who beats Dee, who in turn beats Anne. Assume that the rating x_i of player i is to reflect, as best we can, the difference in scores upon playing player j: that is, pose the difference in ratings, $x_i - x_j$, should equal the difference in the scores when they play.

Solution: The first stage is to model the results by idealised mathematical equations. From Table 3.4 six games were played with the following scores. Each game then generates the shown ideal equation for the difference between two ratings.

- Anne beats Bob 3-2, so $x_1 - x_2 = 3 - 2 = 1$.
- Anne beats Chris 3-0, so $x_1 - x_3 = 3 - 0 = 3$.
- Bob beats Chris 2-1, so $x_2 - x_3 = 2 - 1 = 1$.
- Anne is beaten by Dee 1-3, so $x_1 - x_4 = 1 - 3 = -2$.
- Bob beats Dee 4-0, so $x_2 - x_4 = 4 - 0 = 4$.
- Chris is beaten by Dee 2-3, so $x_3 - x_4 = 2 - 3 = -1$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

These six equations form the linear system $A\boldsymbol{x} = \boldsymbol{b}$ where

$$A = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ 3 \\ 1 \\ -2 \\ 4 \\ -1 \end{bmatrix}.$$

We cannot satisfy all these equations exactly, so we have to accept an approximate solution that estimates the ratings as best we can. The second stage uses an SVD to 'best' solve the equations.

(a) Enter the matrix A and vector $\boldsymbol{b}$ into MATLAB/Octave with

```
A=[1  -1   0   0
   1   0  -1   0
   0   1  -1   0
   1   0   0  -1
   0   1   0  -1
   0   0   1  -1 ]
b=[1;3;1;-2;4;-1]
```

Then factorise matrix $A = USV^{\mathrm{T}}$ with `[U,S,V]=svd(A)` (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 =
  0.00  0.00 -0.87 -0.50
 -0.62  0.53  0.29 -0.50
 -0.14 -0.80  0.29 -0.50
  0.77  0.28  0.29 -0.50
```

Although the first three columns of `U` and `V` may be different for you (because the first three singular values are all the same), the eventual solution is the same. The system of

v0.2a

equations $A\boldsymbol{x} = \boldsymbol{b}$ for the ratings becomes

$$U \underbrace{S \overbrace{V^{\mathrm{T}}\boldsymbol{x}}^{=\boldsymbol{y}}}_{=\boldsymbol{z}} = \boldsymbol{b}.$$

(b) Solve $U\boldsymbol{z} = \boldsymbol{b}$ by $\boldsymbol{z} = U^{\mathrm{T}}\boldsymbol{b}$ via computing `z=U'*b` to get the $\mathbb{R}^6$ vector

```
z =
  -1.27
   2.92
  -1.15
   0.93
   1.76
  -4.07
```

(c) Now solve $S\boldsymbol{y} = \boldsymbol{z}$. But the last three rows of the diagonal matrix S are zero, whereas the last three components of $\boldsymbol{z}$ are non-zero: hence there is no exact solution. Instead we approximate by setting the last three components of $\boldsymbol{z}$ to zero. This approximation is the *smallest change* we can make to the data of the game results that will make the results consistent.

That is, since $\operatorname{rank} A = 3$ from the three non-zero singular values, so we approximately solve the system in MATLAB/Octave by `y=z(1:3)./diag(S(1:3,1:3))`:

```
y =
  -0.63
   1.46
  -0.58
```

The fourth component y_4 is arbitrary.

(d) Lastly, solve $V^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{y}$ as $\boldsymbol{x} = V\boldsymbol{y}$. Obtain a particular solution in MATLAB/Octave by computing `x=V(:,1:3)*y`:

```
x =
   0.50
   1.00
  -1.25
  -0.25
```

Add an arbitrary multiple of the fourth column of `V` to get a general solution

$$\boldsymbol{x} = \begin{bmatrix} \frac{1}{2} \\ 1 \\ -\frac{5}{4} \\ -\frac{1}{4} \end{bmatrix} + y_4 \begin{bmatrix} -\frac{1}{2} \\ -\frac{1}{2} \\ -\frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}.$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Be aware of Kenneth Arrow's Impossibility Theorem (one of the great theorems of the 20th century): *all 1D ranking systems are flawed!* Wikipedia [2014] describes the theorem this way (in the context of voting systems): that among

> three or more distinct alternatives (options), no rank order voting system can convert the ranked preferences of individuals into a community-wide (complete and transitive) ranking while also meeting [four sensible] criteria ... called unrestricted domain, non-dictatorship, Pareto efficiency, and independence of irrelevant alternatives.

In rating sport players/teams:

- the "distinct alternatives" are the players/teams;
- the "ranked preferences of individuals" are the individual results of each game played; and
- the "community-wide ranking" is the assumption that we can rate each player/team by a one-dimensional numerical rating.

Arrow's theorem assures us that every such scheme must violate at least one of four sensible criteria. Every ranking scheme is thus open to criticism. But every alternative scheme will also be open to criticism by also violating one of the criteria.

The final stage is to interpret the solution for the application. In this application the absolute ratings are not important, so we ignore y_4 (consider it zero). From the game results of Table 3.4 this analysis indicates the players' rankings are, in decreasing order, Bob, Anne, Dee, and Chris.

■

When rating players or teams based upon results, be clear the purpose. For example, is the purpose to summarise past performance? or to predict future contests? If the latter, then my limited experience suggests that one should fit the win-loss record instead of the scores.

Activity 3.5.6. 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 $\tilde{\boldsymbol{b}} = A\boldsymbol{x}$ for each, which one minimises the distance between the original right-hand side $\boldsymbol{b} = (9, 2, 10)$ and the approximate $\tilde{\boldsymbol{b}}$?

(a) $\boldsymbol{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ (b) $\boldsymbol{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ (c) $\boldsymbol{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ (d) $\boldsymbol{x} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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 3.5.8.

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

■

Theorem 3.5.7 (smallest change). *All approximate solutions obtained by Procedure 3.5.4 solve the linear system $A\boldsymbol{x} = \tilde{\boldsymbol{b}}$ for the consistent right-hand side vector $\tilde{\boldsymbol{b}}$ that minimises the distance $|\tilde{\boldsymbol{b}} - \boldsymbol{b}|$.*

The over-tilde on $\boldsymbol{b}$ is to suggest an approximation.

Proof. Find an SVD $A = USV^{\mathrm{T}}$ of $m \times n$ matrix A. Then Procedure 3.5.4 computes $\boldsymbol{z} = U^{\mathrm{T}}\boldsymbol{b} \in \mathbb{R}^m$, that is, $\boldsymbol{b} = U\boldsymbol{z}$ as U is orthogonal. For any $\tilde{\boldsymbol{b}} \in \mathbb{R}^m$ let $\tilde{\boldsymbol{z}} = U^{\mathrm{T}}\tilde{\boldsymbol{b}} \in \mathbb{R}^m$, that is, $\tilde{\boldsymbol{b}} = U\tilde{\boldsymbol{z}}$. Then $|\tilde{\boldsymbol{b}} - \boldsymbol{b}| = |U\tilde{\boldsymbol{z}} - U\boldsymbol{z}| = |U(\tilde{\boldsymbol{z}} - \boldsymbol{z})| = |\tilde{\boldsymbol{z}} - \boldsymbol{z}|$ as multiplication by orthogonal U preserves distances (Theorem 3.2.48). Thus minimising $|\tilde{\boldsymbol{b}} - \boldsymbol{b}|$ is equivalent to minimising $|\tilde{\boldsymbol{z}} - \boldsymbol{z}|$. Procedure 3.5.4 seeks to solve the diagonal system $S\boldsymbol{y} = \boldsymbol{z}$ for $\boldsymbol{y} \in \mathbb{R}^n$. That is, for a matrix of rank $A = r$

$$\begin{bmatrix} \sigma_1 & \cdots & 0 & \\ \vdots & \ddots & \vdots & O_{r\times(n-r)} \\ 0 & \cdots & \sigma_r & \\ & & & \\ O_{(m-r)\times r} & & & O_{(m-r)\times(n-r)} \end{bmatrix} \boldsymbol{y} = \begin{bmatrix} z_1 \\ \vdots \\ z_r \\ z_{r+1} \\ \vdots \\ z_m \end{bmatrix}.$$

Procedure 3.5.4 approximately solves this inconsistent system by adjusting the right-hand side to $\tilde{\boldsymbol{z}} = (z_1, \dots, z_r, 0, \dots, 0) \in \mathbb{R}^m$. This change makes $|\boldsymbol{z} - \tilde{\boldsymbol{z}}|$ as small as possible because we must zero the last $(m - r)$ components of $\boldsymbol{z}$ in order to obtain a consistent set of equations, and because any adjustment to the first r components of $\boldsymbol{z}$ would only increase $|\boldsymbol{z} - \tilde{\boldsymbol{z}}|$. Hence the solution computed by Procedure 3.5.4 solves the consistent system $A\boldsymbol{x} = \tilde{\boldsymbol{b}}$ (with $\tilde{\boldsymbol{b}} = U\tilde{\boldsymbol{z}}$) such that $|\boldsymbol{b} - \tilde{\boldsymbol{b}}|$ is minimised. □

Example 3.5.8 (life expectancy). Table 3.5 lists life expectancies of people born in a given year; Figure 3.5 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 3.5, the best straight line to the female life expectancy. Solve the approximation problem with an SVD and confirm it gives the same solution as `A\b` in MATLAB/Octave.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Figure 3.5: 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 3.5.8.

Solution: Start by posing a mathematical model: let's suppose that the life expectancy ℓ is a straight line function of year of birth: $\ell = x_1 + x_2 t$ where we need to find the coefficients x_1 and x_2, and where t counts the number of decades since 1951, the start of the data. Table 3.5 then gives seven ideal equations to solve for x_1 and x_2:

$$\begin{aligned}
(1951)\quad & x_1 + 0x_2 = 72.0\,,\\
(1961)\quad & x_1 + 1x_2 = 74.2\,,\\
(1971)\quad & x_1 + 2x_2 = 75.5\,,\\
(1981)\quad & x_1 + 3x_2 = 78.2\,,\\
(1991)\quad & x_1 + 4x_2 = 79.6\,,\\
(2001)\quad & x_1 + 5x_2 = 80.2\,,\\
(2011)\quad & x_1 + 6x_2 = 81.1\,.
\end{aligned}$$

Form these into the matrix-vector system $A\boldsymbol{x} = \boldsymbol{b}$ where

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \\ 1 & 6 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 72.0 \\ 74.2 \\ 75.5 \\ 78.2 \\ 79.6 \\ 80.2 \\ 81.1 \end{bmatrix}.$$

Procedure 3.5.4 then determines a best approximate solution.

(a) Enter the matrix A and vector $\boldsymbol{b}$ into MATLAB/Octave, and compute an SVD of $A = USV^{\mathrm{T}}$ via `[U,S,V]=svd(A)` (2 d.p.):

v0.2a

```
U =
  0.02  0.68 -0.38 -0.35 -0.32 -0.30 -0.27
  0.12  0.52 -0.14  0.06  0.26  0.45  0.65
  0.22  0.36  0.89 -0.09 -0.08 -0.07 -0.05
  0.32  0.20 -0.10  0.88 -0.13 -0.15 -0.16
  0.42  0.04 -0.10 -0.14  0.81 -0.23 -0.28
  0.52 -0.12 -0.09 -0.16 -0.24  0.69 -0.39
  0.62 -0.28 -0.09 -0.19 -0.29 -0.40  0.50
S =
  9.80     0
     0  1.43
     0     0
     0     0
     0     0
     0     0
     0     0
V =
  0.23  0.97
  0.97 -0.23
```

(b) Solve $U\boldsymbol{z} = \boldsymbol{b}$ to give this first intermediary $\boldsymbol{z} = U^{\mathrm{T}}\boldsymbol{b}$ via the command `z=U'*b`:

```
z =
   178.19
   100.48
    -0.05
     1.14
     1.02
     0.10
    -0.52
```

(c) Now solve approximately $S\boldsymbol{y} = \boldsymbol{z}$. From the two non-zero singular values in S the matrix A has rank 2. So the approximation is to discard/zero (as 'errors') all but the first two elements of $\boldsymbol{z}$ and find the best approximate $\boldsymbol{y}$ via `y=z(1:2)./diag(S(1:2,1:2))`:

```
y =
   18.19
   70.31
```

(d) Solve $V^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{y}$ by $\boldsymbol{x} = V\boldsymbol{y}$ via `x=V*y`:

```
x =
   72.61
    1.55
```

Compute `x=A\b` to find it gives exactly the same answer: we soon discuss why `A\b` gives exactly the same 'best' approximate solution.

Lastly, interpret the answer. The approximation gives $x_1 = 72.61$ and $x_2 = 1.55$ (2 d.p.). Since the ideal model was life expectancy

v0.2a

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 3.5.10

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

$\ell = x_1 + x_2 t$ we determine a 'best' approximate model is $\ell \approx 72.61 + 1.55\,t$ years where t is the number of decades since 1951: this is the straight line drawn in Figure 3.5. That is, females tend to live an extra 1.55 years for every decade born after 1951. For example, for females born in 2021, some seven decades after 1951, this model predicts a life expectancy of $\ell \approx 72.61 + 1.55 \times 7 = 83.46$ years.

■

Activity 3.5.9. 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?

(a) $f_i = x_1$

(b) $v_i = x_1 + x_2 f_i$

(c) $f_i = x_1 + x_2 v_i$

(d) $v_i = x_1 + x_2 f_i + x_3 f_i^2$

■

Example 3.5.10 (planetary orbital periods)**.** Table 3.6 lists each orbital period of the planets of the solar system; Figure 3.6 plots the data points as a function of the distance of the planets from the sun. Let's infer Kepler's law that the period grows as the distance to the power 3/2: shown by the straight line fit in Figure 3.6. Use the data for Mercury to Uranus to infer the law with an SVD, confirm it gives the same solution as `A\b` in MATLAB/Octave, and use the fit to predict Neptune's period from its distance.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Figure 3.6: 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 3.5.10.

Solution: Start by posing a mathematical model: Kepler's law is a power law that the ith period $p_i = c_1 d_i^{c_2}$ for some unknown coefficient c_1 and exponent c_2. Take logarithms (to any base so let's use base 10) and seek that $\log_{10} p_i = \log_{10} c_1 + c_2 \log_{10} d_i$; that is, seek unknowns x_1 and x_2 such that $\log_{10} p_i = x_1 + x_2 \log_{10} d_i$. The first seven rows of Table 3.6 then gives seven ideal linear equations to solve for x_1 and x_2:

$$\begin{aligned}
x_1 + \log_{10} 57.91\, x_2 &= \log_{10} 87.97\,,\\
x_1 + \log_{10} 108.21\, x_2 &= \log_{10} 224.70\,,\\
x_1 + \log_{10} 149.60\, x_2 &= \log_{10} 365.26\,,\\
x_1 + \log_{10} 227.94\, x_2 &= \log_{10} 686.97\,,\\
x_1 + \log_{10} 778.55\, x_2 &= \log_{10} 4332.59\,,\\
x_1 + \log_{10} 1433.45\, x_2 &= \log_{10} 10759.22\,,\\
x_1 + \log_{10} 2870.67\, x_2 &= \log_{10} 30687.15\,.
\end{aligned}$$

Form these into the matrix-vector system $A\boldsymbol{x} = \boldsymbol{b}$: for simplicity recorded here to two decimal places albeit computed more accurately,

$$A = \begin{bmatrix} 1 & 1.76 \\ 1 & 2.03 \\ 1 & 2.17 \\ 1 & 2.36 \\ 1 & 2.89 \\ 1 & 3.16 \\ 1 & 3.46 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 1.94 \\ 2.35 \\ 2.56 \\ 2.84 \\ 3.64 \\ 4.03 \\ 4.49 \end{bmatrix}.$$

Procedure 3.5.4 then determines a best approximate solution.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(a) Enter these matrices in MATLAB/Octave by the commands, for example,

```
d=[   57.91
     108.21
     149.60
     227.94
     778.55
    1433.45
    2870.67];
p=[ 87.97
   224.70
   365.26
   686.97
  4332.59
 10759.22
 30687.15];
A=[ones(7,1) log10(d)]
b=log10(p)
```

since the MATLAB/Octave function `log10()` computes the logarithm to base 10 of each component in its argument (Table 3.2). Then compute an SVD of $A = USV^{\mathrm{T}}$ via `[U,S,V]=svd(A)` (2 d.p.):

```
U =
 -0.27 -0.57 -0.39 -0.38 -0.34 -0.32 -0.30
 -0.31 -0.40 -0.21 -0.09  0.27  0.45  0.65
 -0.32 -0.31  0.88 -0.10 -0.06 -0.04 -0.02
 -0.35 -0.19 -0.11  0.90 -0.10 -0.09 -0.09
 -0.41  0.14 -0.08 -0.11  0.80 -0.25 -0.30
 -0.45  0.31 -0.06 -0.11 -0.26  0.67 -0.41
 -0.49  0.51 -0.04 -0.11 -0.31 -0.41  0.47
S =
  7.38     0
     0  0.55
     0     0
     0     0
     0     0
     0     0
     0     0
V =
 -0.35 -0.94
 -0.94  0.35
```

(b) Solve $U\boldsymbol{z} = \boldsymbol{b}$ to give this first intermediary $\boldsymbol{z} = U^{\mathrm{T}}\boldsymbol{b}$ via the command `z=U'*b`:

```
z =
  -8.5507
   0.6514
```

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

```
  0.0002
  0.0004
  0.0005
 -0.0018
  0.0012
```

(c) Now solve approximately $S\boldsymbol{y} = \boldsymbol{z}$. From the two non-zero singular values in S the matrix A has rank two. So the approximation is to discard/zero all but the first two elements of $\boldsymbol{z}$ (as an error, here all small in value). Then find the best approximate $\boldsymbol{y}$ via `y=z(1:2)./diag(S(1:2,1:2))`:

```
y =
  -1.1581
   1.1803
```

(d) Solve $V^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{y}$ by $\boldsymbol{x} = V\boldsymbol{y}$ via `x=V*y`:

```
x =
  -0.6980
   1.4991
```

Also check that computing `x=A\b` gives exactly the same 'best' approximate solution.

Lastly, interpret the answer. The approximation gives $x_1 = -0.6980$ and $x_2 = 1.4991$. Since the ideal model was the log of the period $\log_{10} p = x_1 + x_2 \log_{10} d$ we determine a 'best' approximate model is $\log_{10} p \approx -0.6980 + 1.4991 \log_{10} d$. Raising ten to the power of both sides gives the power law that the period $p \approx 0.2005\, d^{1.4991}$ days: this is the straight line drawn in Figure 3.6. The exponent 1.4991 is within 0.1% of the exponent 3/2 that is Kepler's law.

For example, for Neptune with a semi-major axis distance of 4498.542 Gm, using the 'best' model predicts Neptune's period

$$10^{-0.6980+1.4991\log_{10} 4498.542} = 60019 \text{ days.}$$

This prediction is pleasingly close to the observed period of 60190 days.

■

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, $\operatorname{cond}(A^{\mathrm{T}}A) = (\operatorname{cond} A)^2$ (Exercise 3.3.14) so matrix $A^{\mathrm{T}}A$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

typically has a much worse condition number than matrix A (Procedure 2.2.5).

- 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 short-cut 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 assume you know what you are doing).

3.5.2 Compute the smallest appropriate solution

> I'm thinking of two numbers. Their average is three. What are the numbers? *Cleve Molar,* The world's simplest impossible problem *(1990)*

The Matlab/Octave operation `A\b` Recall that Examples 3.5.8 and 3.5.10 observed 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 3.5.11. Use `x=A\b` to 'solve' the problems of Examples 3.5.1, 3.5.3 and 3.5.5.

- With Octave, observe the answer returned is the *particular* solution determined by the SVD Procedure 3.5.4 (whether approximate or exact): respectively 84.5 kg; ratings $(1, \frac{1}{3}, -\frac{4}{3})$; and ratings $(\frac{1}{2}, 1, -\frac{5}{4}, -\frac{1}{4})$.
- With MATLAB, the computed answers are often different: respectively 84.5 kg (the same); ratings (`NaN`, `Inf`, `Inf`) with a warning; and ratings $(0.75, 1.25, -1, 0)$ 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* of all valid solutions (Theorem 3.5.12): often 'exact' when $m < n$, or approximate when $m > n$ (Theorem 3.5.7). Using `A\b` is the most efficient computationally, but using the SVD helps us understand what it does.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

- MATLAB (R2013b) 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 3.5.12 (smallest solution). *Obtain the smallest solution, whether exact or as an approximation, to a system of linear equations by setting to zero the free variables, $y_{r+1} = \cdots = y_n = 0$, in Procedures 3.3.15 and 3.5.4.*

Proof. We obtain all possible solutions, whether exact (Procedure 3.3.15) or approximate (Procedure 3.5.4), from solving $\boldsymbol{x} = V\boldsymbol{y}$. Since multiplication by orthogonal V preserves lengths (Theorem 3.2.48), the lengths of $\boldsymbol{x}$ and $\boldsymbol{y}$ are the same: consequently, $|\boldsymbol{x}|^2 = |\boldsymbol{y}|^2 = y_1^2 + \cdots + y_r^2 + y_{r+1}^2 + \cdots + y_n^2$. Now variables $y_1, y_2, \ldots, y_r$ are fixed by Procedures 3.3.15 and 3.5.4, but $y_{r+1}, \ldots, y_n$ are free to vary. Hence the smallest $|\boldsymbol{y}|^2$ is obtained by setting $y_{r+1} = \cdots = y_n = 0$. Then this gives the particular solution $\boldsymbol{x} = V\boldsymbol{y}$ of smallest $|\boldsymbol{x}|$. □

Example 3.5.13. In the table tennis ratings of Example 3.5.3 the procedure found 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 below (blue). Verify $|\boldsymbol{x}|$ is a minimum only when the free variable $y_3 = 0$ (a disc in the plot).

Solution:

$$\begin{aligned} |\boldsymbol{x}|^2 &= \boldsymbol{x} \cdot \boldsymbol{x} \\ &= \left(\begin{bmatrix} 1 \\ \frac{1}{3} \\ -\frac{4}{3} \end{bmatrix} + \frac{y_3}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right) \cdot \left(\begin{bmatrix} 1 \\ \frac{1}{3} \\ -\frac{4}{3} \end{bmatrix} + \frac{y_3}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right) \\ &= \begin{bmatrix} 1 \\ \frac{1}{3} \\ -\frac{4}{3} \end{bmatrix} \cdot \begin{bmatrix} 1 \\ \frac{1}{3} \\ -\frac{4}{3} \end{bmatrix} + \frac{2y_3}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ \frac{1}{3} \\ -\frac{4}{3} \end{bmatrix} + \frac{y_3^2}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \end{aligned}$$

v0.2a

$$= \tfrac{26}{9} + 0y_3 + y_3^2$$

Because of the $0y_3$ term, this quadratic is minimised for $y_3 = 0$. Hence the length $|\boldsymbol{x}|$ is minimised by the free variable $y_3 = 0$.

■

Example 3.5.14 (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 3.5.12. Confirm the Octave computation `A\b` gives this same closest point, but MATLAB gives a different answer.

Solution: The point on the line $3x_1 + 4x_2 = 25$ closest to the origin, is the smallest solution of $3x_1 + 4x_2 = 25$. Rephrase as the matrix vector system $A\boldsymbol{x} = b$ for matrix $A = \begin{bmatrix}3 & 4\end{bmatrix}$ and $b = 25$, and apply Procedure 3.3.15.

(a) Factorise $A = USV^{\mathrm{T}}$ in MATLAB/Octave via the command `[U,S,V]=svd([3 4])`:

```
U =  1
S =
   5   0
V =
   0.6000  -0.8000
   0.8000   0.6000
```

(b) Solve $Uz = b = 25$ which here gives $z = 25$.

(c) Solve $S\boldsymbol{y} = z = 25$ with general solution here of $\boldsymbol{y} = (5, y_2)$. Obtain the smallest solution with free variable $y_2 = 0$.

(d) Solve $V^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{y}$ by $\boldsymbol{x} = V\boldsymbol{y} = V(5, 0) = (3, 4)$.

This is the smallest solution and hence the point on the line closest to the origin (as plotted).

Computing `x=A\b`, which here is simply `x=[3 4]\25`, gives answer $\boldsymbol{x} = (3, 4)$ in Octave; as determined by the SVD, this point is the closest on the line to the origin. In MATLAB, `x=[3 4]\25` gives $\boldsymbol{x} = (0, 6.25)$ which the marginal graph shows is a valid solutions, but not the smallest solution.

■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Table 3.7: As well as the MATLAB/Octave commands and operations listed in Tables 1.2, 2.3, 3.1, 3.2, and 3.3 we may invoke these functions for drawing images—functions which are otherwise not needed.

- `reshape(A,p,q)` for a $m \times n$ matrix/vector A, provided $mn = pq$, generates a $p \times q$ matrix with entries taken column-wise from A. Either p or q can be `[]` in which case MATLAB/Octave uses $p = mn/q$ or $q = mn/p$ respectively.
- `colormap(gray)` MATLAB/Octave usually draws graphs with colour, but for many images we need grayscale; this command changes the current figure to 64 shades of gray.
  (`colormap(jet)` is the default, `colormap(hot)` is good for both colour and grayscale reproductions, `colormap('list')` lists the available colormaps you can try.)
- `imagesc(A)` where A is a $m \times n$ matrix of values draws an $m \times n$ image in the current figure window using the values of A (scaled to fit) to determine the colour from the current colormap (e.g., grayscale).
- `log(x)` where x is a matrix, vector or scalar computes the natural logarithm to the base e of each element, and returns the result(s) as a correspondingly sized matrix, vector or scalar.
- `exp(x)` where x is a matrix, vector or scalar computes the exponential of each element, and returns the result(s) as a correspondingly sized matrix, vector or scalar.

Activity 3.5.15. What is the closest point to the origin of the plane $2x + 3y + 6z = 98$? given the SVD

$$\begin{bmatrix} 2 & 3 & 6 \end{bmatrix} = \begin{bmatrix} 1 \end{bmatrix} \begin{bmatrix} 7 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{2}{7} & -\frac{3}{7} & -\frac{6}{7} \\ \frac{3}{7} & \frac{6}{7} & -\frac{2}{7} \\ \frac{6}{7} & -\frac{2}{7} & \frac{3}{7} \end{bmatrix}^{\mathrm{T}} .$$

(a) $(2\,,3\,,6)$ (b) $(4\,,6\,,12)$

(c) $(-3\,,6\,,-2)$ (d) $(-12\,,-4\,,6)$

■

Example 3.5.16 (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 the user to see inside the object without cutting.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Wikipedia, 2015

Importantly for medical diagnosis and industrial purposes, the computed 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

Computers almost always use "log" to denote the natural logarithm, so we do too. Herein, unsubscripted "log" means the same as "ln".

Let's analyse a toy example.[18] 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, $r_1, r_2, \dots, r_9$ as shown in the margin, that the CT-scan is to find: the fraction r_j of the incident X-ray emerges after passing through the jth square.

As indicated next in the margin, six X-ray measurements are made through the body where $f_1, f_2, \dots, f_6$ denote the fraction of energy in the measurements relative to the power of the X-ray beam. Thus we need to solve six equations for the nine unknown transmission factors:

$$\begin{aligned} r_1r_2r_3 &= f_1\,, \quad & r_4r_5r_6 &= f_2\,, \quad & r_7r_8r_9 &= f_3\,, \\ r_1r_4r_7 &= f_4\,, \quad & r_2r_5r_8 &= f_5\,, \quad & r_3r_6r_9 &= f_6\,. \end{aligned}$$

Turn such nonlinear equations into linear equations that we can handle by taking the logarithm (to any base, but here say the natural logarithm to base e) of both sides of all equations:

$$r_ir_jr_k = f_l \iff (\log r_i) + (\log r_j) + (\log r_k) = (\log f_l).$$

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 densities 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)

[18] For those interested in reading further, Kress (2015) [§8] introduces the advanced, highly mathematical, approach to computerized tomography.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

```
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 in the margin)

$$\begin{bmatrix} r_1 & r_4 & r_7 \\ r_2 & r_5 & r_8 \\ r_3 & r_6 & r_9 \end{bmatrix} = \begin{bmatrix} 0.66 & 0.63 & 0.53 \\ 0.68 & 0.65 & 0.55 \\ 0.91 & 0.87 & 0.74 \end{bmatrix}.$$

Octave's answer predicts that there is less transmitting, more absorbing, denser, material to the top-right; and more transmitting, less absorbing, less dense, material to the bottom-left.

- However, the answer from MATLAB's `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—the leftmost picture—which is quite different! [19]

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.

[19] MATLAB does give a warning in this instance (`Warning: Rank deficient, ...`), but it does not always. For example, it does not warn of issues when you ask it to solve $\frac{1}{2}(x_1 + x_2) = 3$ via `[0.5 0.5]\3`: it simply computes the 'answer' $\boldsymbol{x} = (6\,, 0)$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

The reason for such multiplicity of mathematically valid answers is that the problem is underdetermined. There are nine unknowns but only six equations, so in linear algebra there are typically an infinity of valid answers (as in Theorem 2.2.31): just five of these are illustrated above. *In this application to* CT-*scans* we add the additional information that we desire the answer that is the 'greyest', the most 'washed out', the answer with fewest features. Finding the answer $\boldsymbol{x}$ that minimises $|\boldsymbol{x}|$ is a reasonable way to quantify this desire. [20]

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 $z_6 = 0.00$ this is only a small inconsistency

[20] Another possibility is to increase the number of measurements in order to increase the number of equations to match the number of unknown pixels. However, measurements are often prohibitively expensive. Further, increasing the number of measurements may tempt us to increase the resolution by having more smaller pixels: in which case we again have to deal with the same issue of more variables than known equations.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

error. Now set $y_i = z_i/\sigma_i$ for $i = 1, \dots, 5$ and *for the smallest magnitude answer set the free variables* $y_6 = y_7 = y_8 = y_9 = 0$ (Theorem 3.5.12). Obtain the non-zero 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 solve $V^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{y}$ to determine the smallest solution via `x=V(:,1:5)*y` 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

This optional section does usefully support least square approximation, and provides examples of transformations for the next section 3.6. These orthogonal projections are extensively used in applications.

Reconsider the task of making a minimal change to the right-hand side of a system of linear equations, and let's connect it to the so-called orthogonal projection. This important connection occurs because of the geometry that the closest point on a line or plane to another given point is the one which forms a right-angle; that is, is forms an orthogonal vector.

Project onto a direction

Example 3.5.17. 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 in the margin, 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 $\tilde{\boldsymbol{b}}$ which is in the column space of matrix $A = \begin{bmatrix}\boldsymbol{a}\end{bmatrix}$. But we could choose any $\tilde{\boldsymbol{b}}$ in the column space, so any answer is possible! Surely any answer is not acceptable. Instead, the preferred answer is to find the vector $\tilde{\boldsymbol{b}}$ in the column space of matrix $A = \begin{bmatrix}\boldsymbol{a}\end{bmatrix}$ *and which is closest to* $\boldsymbol{b}$, as illustrated in the margin here.

The SVD approach of Procedure 3.5.4 to find $\tilde{\boldsymbol{b}}$ and x is the following.

(a) Use `[U,S,V]=svd([2;1])` to find here the SVD factorisation $A = USV^{\mathrm{T}} = \begin{bmatrix}0.89 & -0.45\\0.45 & 0.89\end{bmatrix}\begin{bmatrix}2.24\\0\end{bmatrix}[1]^{\mathrm{T}}$ (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 $|\boldsymbol{b} - \tilde{\boldsymbol{b}}|$—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, vector $\tilde{\boldsymbol{b}} = \boldsymbol{a}x = (2, 1)2 = (4, 2)$, as is recognisable in the graphs. ■

Now let's derive the same result but with two differences: firstly, use more elementary arguments, not the SVD; and secondly, derive the result for general vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ (although continuing to use the same illustration). Start with the crucial observation that the closest point/vector $\tilde{\boldsymbol{b}}$ in the column space of $A = [\boldsymbol{a}]$ is such that $\boldsymbol{b} - \tilde{\boldsymbol{b}}$ is at right-angles, orthogonal, to $\boldsymbol{a}$. (If $\boldsymbol{b} - \tilde{\boldsymbol{b}}$ were not orthogonal, then we would be able to slide $\tilde{\boldsymbol{b}}$ along the line span$\{\boldsymbol{a}\}$ to reduce the length of $\boldsymbol{b} - \tilde{\boldsymbol{b}}$.) Thus we form a right-angle triangle with hypotenuse of length $|\boldsymbol{b}|$ and angle θ as shown in the margin. Trigonometry then gives the adjacent length $|\tilde{\boldsymbol{b}}| = |\boldsymbol{b}|\cos\theta$. 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 $|\tilde{\boldsymbol{b}}| = |\boldsymbol{b}|\boldsymbol{a}\cdot\boldsymbol{b}/(|\boldsymbol{a}||\boldsymbol{b}|) = \boldsymbol{a}\cdot\boldsymbol{b}/|\boldsymbol{a}|$. To approximately solve $\boldsymbol{a}x = \boldsymbol{b}$, replace the inconsistent $\boldsymbol{a}x = \boldsymbol{b}$ by the consistent $\boldsymbol{a}x = \tilde{\boldsymbol{b}}$. Then as x is a scalar we solve this consistent equation via the ratio of lengths, $x = |\tilde{\boldsymbol{b}}|/|\boldsymbol{a}| = \boldsymbol{a}\cdot\boldsymbol{b}/|\boldsymbol{a}|^2$. For Example 3.5.17, this gives 'solution' $x = (2, 1)\cdot(3, 4)/(2^2 + 1^2) = 10/5 = 2$ as before.

A crucial part of such solutions is the general formula for $\tilde{\boldsymbol{b}} = \boldsymbol{a}x = \boldsymbol{a}(\boldsymbol{a}\cdot\boldsymbol{b})/|\boldsymbol{a}|^2$. Geometrically the formula gives the 'shadow' $\tilde{\boldsymbol{b}}$ of vector $\boldsymbol{b}$ when projected by a 'sun' high above the line of the vector $\boldsymbol{a}$, as illustrated schematically in the margin. As such, the formula is called an orthogonal projection.

Definition 3.5.18 (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 3.5.19. For the following pairs of vectors: draw the named orthogonal projection; and for the given inconsistent system, determine whether the 'best' approximate solution is in the range $x < -1$, $-1 < x < 0$, $0 < x < 1$, or $1 < x$.

(a) $\text{proj}_{\boldsymbol{u}}(\boldsymbol{v})$ and $\boldsymbol{u}x = \boldsymbol{v}$

(b) $\text{proj}_{\boldsymbol{q}}(\boldsymbol{p})$ and $\boldsymbol{q}x = \boldsymbol{p}$

Solution:

(a) Draw a line perpendicular to $\boldsymbol{u}$ that passes through the tip of $\boldsymbol{v}$. Then $\text{proj}_{\boldsymbol{u}}(\boldsymbol{v})$ is as shown. To 'best solve' $\boldsymbol{u}x = \boldsymbol{v}$, approximate the equation $\boldsymbol{u}x = \boldsymbol{v}$ by $\boldsymbol{u}x = \text{proj}_{\boldsymbol{u}}(\boldsymbol{v})$. Since $\text{proj}_{\boldsymbol{u}}(\boldsymbol{v})$ is smaller than $\boldsymbol{u}$ and the same direction, $0 < x < 1$.

(b) Vector $\boldsymbol{q}$ in $\text{proj}_{\boldsymbol{q}}(\boldsymbol{p})$ gives the direction of a line, so we can and do project onto the negative direction of $\boldsymbol{q}$. To 'best solve' $\boldsymbol{q}x = \boldsymbol{p}$, approximate the equation $\boldsymbol{q}x = \boldsymbol{p}$ by $\boldsymbol{q}x = \text{proj}_{\boldsymbol{q}}(\boldsymbol{p})$. Since $\text{proj}_{\boldsymbol{q}}(\boldsymbol{p})$ is smaller than $\boldsymbol{q}$ and in the opposite direction, $-1 < x < 0$.

■

Example 3.5.20. For the following pairs of vectors: compute the given orthogonal projection; and hence find the 'best' approximate solution to the given inconsistent system.

(a) Find $\text{proj}_{\boldsymbol{u}}(\boldsymbol{v})$ for vectors $\boldsymbol{u} = (3,4)$ and $\boldsymbol{v} = (4,1)$, and hence best solve $\boldsymbol{u}x = \boldsymbol{v}$.

Solution:

$$\text{proj}_{\boldsymbol{u}}(\boldsymbol{v}) = (3,4)\frac{(3,4)\cdot(4,1)}{|(3,4)|^2} = (3,4)\frac{16}{25} = (\tfrac{48}{25},\tfrac{64}{25}).$$

Approximate equation $\boldsymbol{u}x = \boldsymbol{v}$ by $\boldsymbol{u}x = \text{proj}_{\boldsymbol{u}}(\boldsymbol{v})$, that is, $(3,4)x = (\frac{48}{25},\frac{64}{25})$ with solution $x = \frac{16}{25}$ (from either component, or the ratio of lengths).

(b) Find $\text{proj}_{\boldsymbol{s}}(\boldsymbol{r})$ for vectors $\boldsymbol{r} = (1,3)$ and $\boldsymbol{s} = (2,-2)$, and hence best solve $\boldsymbol{s}x = \boldsymbol{r}$.

Solution:

$$\text{proj}_{\boldsymbol{s}}(\boldsymbol{r}) = (2,-2)\frac{(2,-2)\cdot(1,3)}{|(2,-2)|^2} = (2,-2)\frac{-4}{8} = (-1,1).$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Approximate equation $\boldsymbol{s}x = \boldsymbol{r}$ by $\boldsymbol{s}x = \operatorname{proj}_{\boldsymbol{s}}(\boldsymbol{r})$, that is, $(2,-2)x = (-1,1)$ with solution $x = -1/2$ (from either component, or the ratio of lengths).

(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 $\boldsymbol{p}x = \boldsymbol{q}$.

Solution: Vector $\boldsymbol{r}$ is a unit vector, so we use the simpler formula that

$$\begin{aligned}\operatorname{proj}_{\boldsymbol{r}}(\boldsymbol{q}) &= (\tfrac{1}{3},\tfrac{2}{3},\tfrac{2}{3})\big[(\tfrac{1}{3},\tfrac{2}{3},\tfrac{2}{3})\cdot(3,2,1)\big]\\ &= (\tfrac{1}{3},\tfrac{2}{3},\tfrac{2}{3})\big[1+\tfrac{4}{3}+\tfrac{2}{3}\big]\\ &= (\tfrac{1}{3},\tfrac{2}{3},\tfrac{2}{3})3 = (1,2,2).\end{aligned}$$

Then 'best solve' equation $\boldsymbol{p}x = \boldsymbol{q}$ by the approximation $\boldsymbol{p}x = \operatorname{proj}_{\boldsymbol{p}}(\boldsymbol{q})$, that is, $(\frac{1}{3},\frac{2}{3},\frac{2}{3})x = (1,2,2)$ with solution $x = 3$ (from any component, or the ratio of lengths).

■

Activity 3.5.21. Use projection to best solve the inconsistent equation $(1,4,8)x = (4,4,2)$. The best answer is which of the following?

(a) $x = 4/9$ (b) $x = 10/13$

(c) $x = 4$ (d) $x = 21/4$

■

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 that enables the equation to be solved. This subsection introduces that solving inconsistent equations in more variables involves an analogous projection onto a subspace.

Definition 3.5.22 (project onto a subspace)**.** *Let $\mathbb{W}$ be a k-dimensional subspace of $\mathbb{R}^n$ with an orthonormal basis $\{\boldsymbol{w}_1,\boldsymbol{w}_2,\dots,\boldsymbol{w}_k\}$. For any vector $\boldsymbol{v}\in\mathbb{R}^n$, the* ***orthogonal projection*** *of vector $\boldsymbol{v}$ onto subspace $\mathbb{W}$ is*

$$\operatorname{proj}_{\mathbb{W}}(\boldsymbol{v}) = \boldsymbol{w}_1(\boldsymbol{w}_1\cdot\boldsymbol{v}) + \boldsymbol{w}_2(\boldsymbol{w}_2\cdot\boldsymbol{v}) + \cdots + \boldsymbol{w}_k(\boldsymbol{w}_k\cdot\boldsymbol{v}).$$

v0.2a

Example 3.5.23. (a) Let $\mathbb{X}$ be the xy-plane in xyz-space, find $\operatorname{proj}_{\mathbb{X}}(3,-4,2)$.

Solution: An orthogonal basis for the xy-plane (blue plane in the stereo picture below) are the two unit vectors $\boldsymbol{i}=(1,0,0)$ and $\boldsymbol{j}=(0,1,0)$. Hence

$$\begin{aligned}\operatorname{proj}_{\mathbb{W}}(3,-4,2) &= \boldsymbol{i}(\boldsymbol{i}\cdot(3,-4,2))+\boldsymbol{j}(\boldsymbol{j}\cdot(3,-4,2))\\ &= \boldsymbol{i}(3+0+0)+\boldsymbol{j}(0-4+0)\\ &= (3,-4,0) \qquad \text{(shown in brown).}\end{aligned}$$

That is, just set the third component of $(3,-4,2)$ to zero.

(b) For the subspace $\mathbb{W}=\operatorname{span}\{(2,-2,1),(2,1,-2)\}$, determine $\operatorname{proj}_{\mathbb{W}}(3,2,1)$ (these vectors and subspace are illustrated below).

Solution: Although the two vectors in the span are orthogonal (blue in the stereo picture above), they are not unit vectors. Normalise the vectors by dividing by their length $\sqrt{2^2+(-2)^2+1^2}=\sqrt{2^2+1^2+(-2)^2}=3$ to find the vectors $\boldsymbol{w}_1=(\frac{2}{3},-\frac{2}{3},\frac{1}{3})$ and $\boldsymbol{w}_2=(\frac{2}{3},\frac{1}{3},-\frac{2}{3})$ are an orthonormal basis for $\mathbb{W}$ (plane). Hence

$$\begin{aligned}&\operatorname{proj}_{\mathbb{W}}(3,2,1)\\ &= \boldsymbol{w}_1(\boldsymbol{w}_1\cdot(3,2,1))+\boldsymbol{w}_2(\boldsymbol{w}_2\cdot(3,2,1))\\ &= \boldsymbol{w}_1(2-\tfrac{4}{3}+\tfrac{1}{3})+\boldsymbol{w}_2(2+\tfrac{2}{3}-\tfrac{2}{3})\\ &= \boldsymbol{w}_1+2\boldsymbol{w}_2\\ &= (\tfrac{2}{3},-\tfrac{2}{3},\tfrac{1}{3})+2(\tfrac{2}{3},\tfrac{1}{3},-\tfrac{2}{3})\\ &= (2,0,-1) \qquad \text{(shown in brown below).}\end{aligned}$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(c) Recall the table tennis ranking Examples 3.3.13 and 3.5.3. To rank the players we seek to solve the matrix-vector system, $A\boldsymbol{x} = \boldsymbol{b}$,

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

Letting $\mathbb{A}$ denote the column space of matrix A, determine $\operatorname{proj}_{\mathbb{A}}(\boldsymbol{b})$.

Solution: We need to find an orthonormal basis for the column space (the illustrated plane spanned by the three shown column vectors)—an SVD gives it to us.

Example 3.3.12 found an SVD $A = USV^{\mathrm{T}}$, in MATLAB/Octave via `[U,S,V]=svd(A)`, to be

```
U =
   0.4082   -0.7071    0.5774
  -0.4082   -0.7071   -0.5774
  -0.8165   -0.0000    0.5774
S =
   1.7321         0         0
        0    1.7321         0
        0         0    0.0000
V =
   0.0000   -0.8165    0.5774
  -0.7071    0.4082    0.5774
   0.7071    0.4082    0.5774
```

Since there are only two non-zero singular values, the column space $\mathbb{A}$ is 2D and spanned by the first two orthonormal columns of matrix U: that is, an orthonormal basis for $\mathbb{A}$ is the two vectors (as illustrated below)

$$\boldsymbol{u}_1 = \begin{bmatrix} 0.4082 \\ -0.4082 \\ -0.8165 \end{bmatrix} = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix},$$
$$\boldsymbol{u}_2 = \begin{bmatrix} -0.7071 \\ -0.7071 \\ -0.0000 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ -1 \\ 0 \end{bmatrix}.$$

Hence

$$
\begin{aligned}
&\operatorname{proj}_{\mathbb{A}}(1,2,2)\\
&= \boldsymbol{u}_1(\boldsymbol{u}_1\cdot(1,2,2)) + \boldsymbol{u}_2(\boldsymbol{u}_2\cdot(1,2,2))\\
&= \boldsymbol{u}_1(1-2-4)/\sqrt{6} + \boldsymbol{u}_2(-1-2+0)/\sqrt{2}\\
&= -\tfrac{5}{\sqrt{6}}\boldsymbol{u}_1 - \tfrac{3}{\sqrt{2}}\boldsymbol{u}_2\\
&= \tfrac{1}{6}(-5,5,10) + \tfrac{1}{2}(3,3,0)\\
&= \tfrac{1}{3}(2,7,5) \qquad \text{(shown in brown below).}
\end{aligned}
$$

(d) Find the projection of the vector $(1,2,2)$ onto the plane $2x - \frac{1}{2}y + 4z = 6$.

Solution: This plane is not a subspace as it does not pass through the origin. Definition 3.5.22 only defines projection onto a subspace so we cannot answer this problem (as yet).

(e) Use an SVD to find the projection of the vector $(1,2,2)$ onto the plane $2x - \frac{1}{2}y + 4z = 0$ (illustrated below).

Solution: This plane does pass through the origin so it forms a subspace, call it $\mathbb{P}$ (illustrated above). To project we need two orthonormal basis vectors. Recall that a normal to the plane is its vectors of coefficients, here $(2,-\frac{1}{2},4)$, so we need to find two orthonormal vectors which are orthogonal to $(2,-\frac{1}{2},4)$. Further, recall that the columns of an orthogonal matrix are orthonormal (Theorem 3.2.48b), so use an SVD to

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

find orthonormal vectors to $(2, -\frac{1}{2}, 4)$. In MATLAB/Octave, compute an SVD with `[U,S,V]=svd([2;-1/2;4])` to find

```
U =
  -0.4444   0.1111  -0.8889
   0.1111   0.9914   0.0684
  -0.8889   0.0684   0.4530
S =
   4.5000
        0
        0
V = -1
```

The first column $\boldsymbol{u}_1 = (-4, 1, -8)/9$ of orthogonal matrix U is in the direction of a normal to the plane as it must since it must be in the span of $(2, -\frac{1}{2}, 4)$. Since matrix U is orthogonal, the last two columns (say $\boldsymbol{u}_2$ and $\boldsymbol{u}_3$, drawn in blue below) are not only orthonormal, but also orthogonal to $\boldsymbol{u}_1$ and hence an orthonormal basis for the plane $\mathbb{P}$.

Hence

$$\begin{aligned}
\operatorname{proj}_{\mathbb{P}}(1,2,2) &= \boldsymbol{u}_2(\boldsymbol{u}_2 \cdot (1,2,2)) + \boldsymbol{u}_3(\boldsymbol{u}_3 \cdot (1,2,2)) \\
&= 2.2308\,\boldsymbol{u}_2 + 0.1539\,\boldsymbol{u}_3 \\
&= 2.2308(0.1111, 0.9914, 0.0684) \\
&\quad + 0.1539(-0.8889, 0.0684, 0.4530) \\
&= (0.1111, 2.2222, 0.2222) \\
&= \tfrac{1}{9}(1, 10, 2) \qquad \text{(shown in brown below).}
\end{aligned}$$

This answer may be computed in MATLAB/Octave via the two dot products `cs=U(:,2:3)'*[1;2;2]`, giving the two coefficients 2.2308 and 0.1539, and then the linear combination `proj=U(:,2:3)*cs`.

■

Activity 3.5.24. Determine which of the following is $\text{proj}_{\mathbb{W}}(1,1,-2)$ for the subspace $\mathbb{W} = \text{span}\{(2,3,6),(-3,6,-2)\}$.

(a) $(-\frac{5}{7}, \frac{3}{7}, -\frac{8}{7})$ (b) $(-\frac{1}{7}, \frac{9}{7}, \frac{4}{7})$

(c) $(\frac{1}{7}, -\frac{9}{7}, -\frac{4}{7})$ (d) $(\frac{5}{7}, -\frac{3}{7}, \frac{8}{7})$

■

Example 3.5.23c determines the orthogonal projection of the given table tennis results $\boldsymbol{b} = (1,2,2)$ onto the column space of matrix A is the vector $\tilde{\boldsymbol{b}} = \frac{1}{3}(2,7,5)$. Recall that in Example 3.5.3, Procedure 3.5.4 gives the 'approximate' solution of the impossible $A\boldsymbol{x} = \boldsymbol{b}$ to be $\boldsymbol{x} = (1, \frac{1}{3}, -\frac{4}{3})$. Now see that $A\boldsymbol{x} = \big(1 - \frac{1}{3}, 1 - (-\frac{4}{3}), \frac{1}{3} - (-\frac{4}{3})\big) = (\frac{2}{3}, \frac{7}{3}, \frac{5}{3}) = \tilde{\boldsymbol{b}}$. That is, the approximate solution method of Procedure 3.5.4 solved the problem $A\boldsymbol{x} = \text{proj}_{\mathbb{A}}(\boldsymbol{b})$. The following theorem confirms this is no accident: orthogonally projecting the right-hand side onto the column space of the matrix in a system of linear equations is equivalent to solving the system with a smallest change to the right-hand side that makes it consistent.

Theorem 3.5.25. *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 $A\boldsymbol{x} = \text{proj}_{\mathbb{A}}(\boldsymbol{b})$ where $\mathbb{A}$ denotes the column space of A.*

Proof. For any $m \times n$ matrix A, Procedure 3.5.4 first finds an SVD $A = USV^{\mathrm{T}}$ and sets $r = \text{rank}\, A$. Second, it computes $\boldsymbol{z} = U^{\mathrm{T}}\boldsymbol{b}$ but disregards z_i for $i = r+1, \dots, m$ as errors. That is, instead of using $\boldsymbol{z} = U^{\mathrm{T}}\boldsymbol{b}$ Procedure 3.5.4 solves the equations with $\tilde{\boldsymbol{z}} = (z_1, z_2, \dots, z_r, 0, \dots, 0)$. This vector $\tilde{\boldsymbol{z}}$ corresponds to a modified right-hand side $\tilde{\boldsymbol{b}}$ satisfying $\tilde{\boldsymbol{z}} = U^{\mathrm{T}}\tilde{\boldsymbol{b}}$; that is, $\tilde{\boldsymbol{b}} = U\tilde{\boldsymbol{z}}$ as matrix U is orthogonal. Recalling $\boldsymbol{u}_i$ denotes the ith column of U and that components $z_i = \boldsymbol{u}_i \cdot \boldsymbol{b}$ from $\boldsymbol{z} = U^{\mathrm{T}}\boldsymbol{b}$, the matrix-vector product $\tilde{\boldsymbol{b}} = U\tilde{\boldsymbol{z}}$ is the linear combination (Example 2.3.5)

$$\begin{aligned}\tilde{\boldsymbol{b}} &= \boldsymbol{u}_1\tilde{z}_1 + \boldsymbol{u}_2\tilde{z}_2 + \cdots + \boldsymbol{u}_r\tilde{z}_r + \boldsymbol{u}_{r+1}0 + \cdots + \boldsymbol{u}_m 0\\ &= \boldsymbol{u}_1(\boldsymbol{u}_1\cdot\boldsymbol{b}) + \boldsymbol{u}_2(\boldsymbol{u}_2\cdot\boldsymbol{b}) + \cdots + \boldsymbol{u}_r(\boldsymbol{u}_r\cdot\boldsymbol{b})\\ &= \text{proj}_{\text{span}\{\boldsymbol{u}_1,\boldsymbol{u}_2,\dots,\boldsymbol{u}_r\}}(\boldsymbol{b}),\end{aligned}$$

by Definition 3.5.22 since the columns $\boldsymbol{u}_i$ of U are orthonormal (Theorem 3.2.48). Theorem 3.4.32 establishes that this span is the column space $\mathbb{A}$ of matrix A. Hence, $\tilde{\boldsymbol{b}} = \text{proj}_{\mathbb{A}}(\boldsymbol{b})$ and so Procedure 3.5.4 solves the system $A\boldsymbol{x} = \text{proj}_{\mathbb{A}}(\boldsymbol{b})$. □

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Example 3.5.26. Recall Example 3.5.1 rationalises four apparently contradictory weighings: in kg the weighings are 84.8, 84.1, 84.7 and 84.4. Denoting the 'uncertain' weight by x, we write these weighings as the inconsistent matrix-vector system

$$Ax = \boldsymbol{b}\,, \quad \text{namely} \quad \begin{bmatrix}1\\1\\1\\1\end{bmatrix} x = \begin{bmatrix}84.8\\84.1\\84.7\\84.4\end{bmatrix}.$$

Let's see that the orthogonal projection of the right-hand side onto the column space of A is the same as the minimal change of Example 3.5.1, which in turn is the well known average.

To find the orthogonal projection, observe matrix A has one column $\boldsymbol{a}_1 = (1,1,1,1)$ so by Definition 3.5.18 the orthogonal projection

$$\begin{aligned}
&\operatorname{proj}_{\operatorname{span}\{\boldsymbol{a}_1\}}(84.8, 84.1, 84.7, 84.4)\\
&= \boldsymbol{a}_1 \frac{\boldsymbol{a}_1 \cdot (84.8, 84.1, 84.7, 84.4)}{|\boldsymbol{a}_1|^2}\\
&= \boldsymbol{a}_1 \frac{84.8 + 84.1 + 84.7 + 84.4}{1+1+1+1}\\
&= \boldsymbol{a}_1 \times 84.5\\
&= (84.5, 84.5, 84.5, 84.5).
\end{aligned}$$

The projected system $Ax = (84.5, 84.5, 84.5, 84.5)$ is now consistent, with solution $x = 84.5\,\text{kg}$. As in Example 3.5.1, this solution is the well-known averaging of the four weights. ■

Example 3.5.27. Recall the round robin tournament amongst four players of Example 3.5.5. 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 3.5.5.

An SVD finds an orthonormal basis for the column space $\mathbb{A}$ of matrix A: Example 3.5.5 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
```

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

```
 -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 non-zero 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 3.5.22 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 with `cs=U(:,1:3)'*b`, 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 found by Example 3.5.5, $\boldsymbol{x} = (\frac{1}{2}, 1, -\frac{5}{4}, -\frac{1}{4})$, satisfy $A\boldsymbol{x} = \operatorname{proj}_{\mathbb{A}}(\boldsymbol{b})$: in MATLAB/Octave compute `A*[0.50;1.00;-1.25;-0.25]` 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 correspond to multiplying by a matrix, which for orthogonal projection is here WW^{T}.

There is an useful feature of Examples 3.5.23e and 3.5.27. 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, $\operatorname{proj}(\boldsymbol{b}) = W\boldsymbol{c}$. 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 useful feature is that the orthogonal projection formula of Definition 3.5.22 is equivalent to the multiplication by matrix (WW^{T}) for an appropriate matrix W.

Theorem 3.5.28 (orthogonal projection matrix)**.** *Let $\mathbb{W}$ be a k-dimensional subspace of $\mathbb{R}^n$ with an orthonormal basis $\{\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_k\}$, then for any 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 $W = \begin{bmatrix}\boldsymbol{w}_1 & \boldsymbol{w}_2 & \cdots & \boldsymbol{w}_k\end{bmatrix}$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Proof. Directly from Definition 3.5.22,

$$\begin{aligned}\operatorname{proj}_{\mathbb{W}}(\boldsymbol{v}) &= \boldsymbol{w}_1(\boldsymbol{w}_1\cdot\boldsymbol{v}) + \boldsymbol{w}_2(\boldsymbol{w}_2\cdot\boldsymbol{v}) + \cdots + \boldsymbol{w}_k(\boldsymbol{w}_k\cdot\boldsymbol{v}) \\ &\qquad \text{(using that } \boldsymbol{w}\cdot\boldsymbol{v} = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{v}\text{, Example 3.1.18)} \\ &= \boldsymbol{w}_1\boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{v} + \boldsymbol{w}_2\boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{v} + \cdots + \boldsymbol{w}_k\boldsymbol{w}_k^{\mathrm{T}}\boldsymbol{v} \\ &= (\boldsymbol{w}_1\boldsymbol{w}_1^{\mathrm{T}} + \boldsymbol{w}_2\boldsymbol{w}_2^{\mathrm{T}} + \cdots + \boldsymbol{w}_k\boldsymbol{w}_k^{\mathrm{T}})\,\boldsymbol{v}.\end{aligned}$$

Let the components of the vector $\boldsymbol{w}_j = (w_{1j}\,, w_{2j}\,, \dots, w_{nj})$, then from the matrix product Definition 3.1.11, the k products in the sum

$$\begin{aligned}&\boldsymbol{w}_1\boldsymbol{w}_1^{\mathrm{T}} + \boldsymbol{w}_2\boldsymbol{w}_2^{\mathrm{T}} + \cdots + \boldsymbol{w}_k\boldsymbol{w}_k^{\mathrm{T}} \\ &= \begin{bmatrix} w_{11}w_{11} & w_{11}w_{21} & \cdots & w_{11}w_{n1} \\ w_{21}w_{11} & w_{21}w_{21} & \cdots & w_{21}w_{n1} \\ \vdots & \vdots & & \vdots \\ w_{n1}w_{11} & w_{n1}w_{21} & \cdots & w_{n1}w_{n1} \end{bmatrix} \\ &\quad + \begin{bmatrix} w_{12}w_{12} & w_{12}w_{22} & \cdots & w_{12}w_{n2} \\ w_{22}w_{12} & w_{22}w_{22} & \cdots & w_{22}w_{n2} \\ \vdots & \vdots & & \vdots \\ w_{n2}w_{12} & w_{n2}w_{22} & \cdots & w_{n2}w_{n2} \end{bmatrix} \\ &\quad + \cdots \\ &\quad + \begin{bmatrix} w_{1k}w_{1k} & w_{1k}w_{2k} & \cdots & w_{1k}w_{nk} \\ w_{2k}w_{1k} & w_{2k}w_{2k} & \cdots & w_{2k}w_{nk} \\ \vdots & \vdots & & \vdots \\ w_{nk}w_{1k} & w_{nk}w_{2k} & \cdots & w_{nk}w_{nk} \end{bmatrix}.\end{aligned}$$

So the $(i\,,j)$th entry of this sum is

$$\begin{aligned}&w_{i1}w_{j1} + w_{i2}w_{j2} + \cdots + w_{ik}w_{jk} \\ &= w_{i1}(W^{\mathrm{T}})_{1j} + w_{i2}(W^{\mathrm{T}})_{2j} + \cdots + w_{ik}(W^{\mathrm{T}})_{kj}\,,\end{aligned}$$

which, from Definition 3.1.11 again, is the $(i\,,j)$th entry of the product WW^{T}. Hence $\operatorname{proj}_{\mathbb{W}}(\boldsymbol{v}) = (WW^{\mathrm{T}})\boldsymbol{v}$. □

Example 3.5.29. Find the matrices of the following orthogonal projections (from Example 3.5.20), and use the matrix to find the given projection.

(a) $\operatorname{proj}_{\boldsymbol{u}}(\boldsymbol{v})$ for vector $\boldsymbol{u} = (3,4)$ and $\boldsymbol{v} = (4,1)$.

Solution: First, normalise $\boldsymbol{u}$ to the unit vector $\boldsymbol{w} = \boldsymbol{u}/|\boldsymbol{u}| = (3,4)/5$. Second, the matrix is

$$WW^{\mathrm{T}} = \boldsymbol{w}\boldsymbol{w}^{\mathrm{T}} = \begin{bmatrix} \frac{3}{5} \\ \frac{4}{5} \end{bmatrix} \begin{bmatrix} \frac{3}{5} & \frac{4}{5} \end{bmatrix} = \begin{bmatrix} \frac{9}{25} & \frac{12}{25} \\ \frac{12}{25} & \frac{16}{25} \end{bmatrix}.$$

Then the projection

$$\operatorname{proj}_{\boldsymbol{u}}(\boldsymbol{v}) = (WW^{\mathrm{T}})\boldsymbol{v} = \begin{bmatrix} \frac{9}{25} & \frac{12}{25} \\ \frac{12}{25} & \frac{16}{25} \end{bmatrix} \begin{bmatrix} 4 \\ 1 \end{bmatrix} = \begin{bmatrix} 48/25 \\ 64/25 \end{bmatrix}$$

v0.2a

(b) $\text{proj}_{\boldsymbol{s}}(\boldsymbol{r})$ for vector $\boldsymbol{s} = (2, -2)$ and $\boldsymbol{r} = (1, 1)$.

Solution: Normalise $\boldsymbol{s}$ to the unit vector $\boldsymbol{w} = \boldsymbol{s}/|\boldsymbol{s}| = (2, -2)/(2\sqrt{2}) = (1, -1)/\sqrt{2}$, then the matrix is

$$WW^{\mathrm{T}} = \boldsymbol{w}\boldsymbol{w}^{\mathrm{T}} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Consequently the projection

$$\text{proj}_{\boldsymbol{s}}(\boldsymbol{r}) = (WW^{\mathrm{T}})\boldsymbol{r} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \boldsymbol{0}.$$

(c) $\text{proj}_{\boldsymbol{p}}(\boldsymbol{q})$ for vector $\boldsymbol{p} = (\frac{1}{3}, \frac{2}{3}, \frac{2}{3})$ and $\boldsymbol{q} = (3, 3, 0)$.

Solution: Vector $\boldsymbol{p}$ is already a unit vector so the matrix is

$$WW^{\mathrm{T}} = \boldsymbol{p}\boldsymbol{p}^{\mathrm{T}} = \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix} \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \end{bmatrix} = \begin{bmatrix} \frac{1}{9} & \frac{2}{9} & \frac{2}{9} \\ \frac{2}{9} & \frac{4}{9} & \frac{4}{9} \\ \frac{2}{9} & \frac{4}{9} & \frac{4}{9} \end{bmatrix}.$$

Then the projection

$$\text{proj}_{\boldsymbol{p}}(\boldsymbol{q}) = (WW^{\mathrm{T}})\boldsymbol{q} = \begin{bmatrix} \frac{1}{9} & \frac{2}{9} & \frac{2}{9} \\ \frac{2}{9} & \frac{4}{9} & \frac{4}{9} \\ \frac{2}{9} & \frac{4}{9} & \frac{4}{9} \end{bmatrix} \begin{bmatrix} 3 \\ 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}.$$

■

Activity 3.5.30. Finding the projection $\text{proj}_{\boldsymbol{u}}(\boldsymbol{v})$ for vectors $\boldsymbol{u} = (2, 6, 3)$ and $\boldsymbol{v} = (1, 4, 8)$ could be done by premultiplying by which of the following the matrices?

(a) $\begin{bmatrix} \frac{4}{49} & \frac{12}{49} & \frac{6}{49} \\ \frac{12}{49} & \frac{36}{49} & \frac{18}{49} \\ \frac{6}{49} & \frac{18}{49} & \frac{9}{49} \end{bmatrix}$

(b) $\begin{bmatrix} \frac{2}{63} & \frac{8}{63} & \frac{16}{63} \\ \frac{2}{21} & \frac{8}{21} & \frac{16}{21} \\ \frac{1}{21} & \frac{4}{21} & \frac{8}{21} \end{bmatrix}$

(c) $\begin{bmatrix} \frac{2}{63} & \frac{2}{21} & \frac{1}{21} \\ \frac{8}{63} & \frac{8}{21} & \frac{4}{21} \\ \frac{16}{63} & \frac{16}{21} & \frac{8}{21} \end{bmatrix}$

(d) $\begin{bmatrix} \frac{1}{81} & \frac{4}{81} & \frac{8}{81} \\ \frac{4}{81} & \frac{16}{81} & \frac{32}{81} \\ \frac{8}{81} & \frac{32}{81} & \frac{64}{81} \end{bmatrix}$

■

v0.2a

Example 3.5.31. Find the matrices of the following orthogonal projections (from Example 3.5.20).

(a) $\text{proj}_{\mathbb{X}}(\boldsymbol{v})$ where $\mathbb{X}$ is the xy-plane in xyz-space.

Solution: The two unit vectors $\boldsymbol{i} = (1,0,0)$ and $\boldsymbol{j} = (0,1,0)$ form an orthogonal basis, so matrix

$$W = \begin{bmatrix} \boldsymbol{i} & \boldsymbol{j} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix},$$

hence the matrix of the projection is

$$WW^{\mathrm{T}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

(b) $\text{proj}_{\mathbb{W}}(\boldsymbol{v})$ for the subspace $\mathbb{W} = \text{span}\{(2,-2,1),(2,1,-2)\}$.

Solution: Now $\boldsymbol{w}_1 = (\frac{2}{3}, -\frac{2}{3}, \frac{1}{3})$ and $\boldsymbol{w}_2 = (\frac{2}{3}, \frac{1}{3}, -\frac{2}{3})$ form an orthonormal basis for $\mathbb{W}$, so matrix

$$W = \begin{bmatrix} \boldsymbol{w}_1 & \boldsymbol{w}_2 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 2 & 2 \\ -2 & 1 \\ 1 & -2 \end{bmatrix},$$

hence the matrix of the projection is

$$\begin{aligned} WW^{\mathrm{T}} &= \frac{1}{3} \begin{bmatrix} 2 & 2 \\ -2 & 1 \\ 1 & -2 \end{bmatrix} \frac{1}{3} \begin{bmatrix} 2 & -2 & 1 \\ 2 & 1 & -2 \end{bmatrix} \\ &= \frac{1}{9} \begin{bmatrix} 8 & -2 & -2 \\ -2 & 5 & -4 \\ -2 & -4 & 5 \end{bmatrix}. \end{aligned}$$

(c) The orthogonal projection onto the column space of matrix

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

Solution: The SVD of Example 3.5.23c determines an orthonormal basis is $\boldsymbol{u}_1 = (1,-1,-2)/\sqrt{6}$ and $\boldsymbol{u}_2 = (-1,-1,0)/\sqrt{2}$. Hence the matrix of the projection is

$$\begin{aligned} WW^{\mathrm{T}} &= \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \\ -\frac{2}{\sqrt{6}} & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{bmatrix} \\ &= \begin{bmatrix} \frac{2}{3} & \frac{1}{3} & -\frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} & \frac{1}{3} \\ -\frac{1}{3} & \frac{1}{3} & \frac{2}{3} \end{bmatrix}. \end{aligned}$$

v0.2a

Alternatively, recall the SVD of matrix A from Example 3.3.12, and recall that the first two columns of `U` are the orthonormal basis vectors. Hence matrix $W =$ `U(:,1:2)` and so MATLAB/Octave computes the matrix of the projection, WW^{T}, via `WWT=U(:,1:2)*U(:,1:2)'` to give the answer

```
WWT =
    0.6667    0.3333   -0.3333
    0.3333    0.6667    0.3333
   -0.3333    0.3333    0.6667
```

(d) The orthogonal projection onto the plane $2x - \frac{1}{2}y + 4z = 0$.

Solution: The SVD of Example 3.5.23e determines an orthonormal basis is the last two columns of

```
U =
  -0.4444   0.1111  -0.8889
   0.1111   0.9914   0.0684
  -0.8889   0.0684   0.4530
```

Hence MATLAB/Octave computes the matrix of the projection with `WWT=U(:,2:3)*U(:,2:3)'` giving the answer

```
WWT =
   0.8025   0.0494  -0.3951
   0.0494   0.9877   0.0988
  -0.3951   0.0988   0.2099
```

■

Orthogonal decomposition separates

Because orthogonal projection has such a close connection to the geometry underlying important tasks such as 'least square' approximation (Theorem 3.5.25), 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 3.5.33.

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 $\mathbb{W}^{\perp}$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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 3.5.32. Given the above qualitative description of an orthogonal complement, which of the following red lines is the orthogonal complement to the shown subspace $\mathbb{W}$?

(a)

(b)

(c)

(d)

■

Definition 3.5.33 (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 $\mathbf{0}$) that are each orthogonal to all vectors in $\mathbb{W}$ is called the* ***orthogonal complement*** *$\mathbb{W}^\perp$ ("W-perp"); that is,*

$$\mathbb{W}^\perp = \{\boldsymbol{u} \in \mathbb{R}^n : \boldsymbol{u} \cdot \boldsymbol{w} = 0 \text{ for all } \boldsymbol{w} \in \mathbb{W}\}.$$

Example 3.5.34 (orthogonal complement)**.**

(a) Given the subspace $\mathbb{W} = \operatorname{span}\{(3, 4)\}$, find its orthogonal complement $\mathbb{W}^\perp$.

Solution: Every vector in $\mathbb{W}$ is of the form $\boldsymbol{w} = (3c, 4c)$. For any vector $\boldsymbol{v} = (u, v) \in \mathbb{R}^2$ the dot product

$$\boldsymbol{w} \cdot \boldsymbol{v} = (3c, 4c) \cdot (u, v) = c(3u + 4v).$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

This dot product is zero for all c if and only if $3u + 4v = 0$. That is, when $u = -4v/3$. Hence $\boldsymbol{v} = (-\frac{4}{3}v, v) = (-\frac{4}{3}, 1)v$, for every v, and so $\mathbb{W}^\perp = \text{span}\{(-\frac{4}{3}, 1)\}$.

(b) Describe the orthogonal complement $\mathbb{X}^\perp$ to the subspace $\mathbb{X} = \text{span}\{(4, -4, 7)\}$.

Solution: Every vector in $\mathbb{W}$ is of the form $\boldsymbol{w} = (4, -4, 7)c$. Seek all vectors $\boldsymbol{v}$ such that $\boldsymbol{w} \cdot \boldsymbol{v} = 0$. For vectors $\boldsymbol{v} = (v_1, v_2, v_3)$ the inner product

$$\boldsymbol{w} \cdot \boldsymbol{v} = c(4, -4, 7) \cdot (v_1, v_2, v_3) = c(4v_1 - 4v_2 + 7v_3)$$

is zero for all c if and only if $4v_1 - 4v_2 + 7v_3 = 0$. That is, the orthogonal complement is all vectors $\boldsymbol{v}$ in the plane $4v_1 - 4v_2 + 7v_3 = 0$ (illustrated in stereo below).

(c) Describe the orthogonal complement of the set $\mathbb{W} = \{(t, t^2) : t \in \mathbb{R}\}$.

Solution: It does not exist as an orthogonal complement is only defined for a subspace, and the parabola (t, t^2) is not a subspace.

(d) Determine the orthogonal complement of the subspace $\mathbb{W} = \text{span}\{(2, -2, 1), (2, 1, -2)\}$.

Solution: Let $\boldsymbol{w}_1 = (2, -2, 1)$ and $\boldsymbol{w}_2 = (2, 1, -2)$ then all vectors $\boldsymbol{w} \in \mathbb{W}$ are of the form $\boldsymbol{w} = c_1\boldsymbol{w}_1 + c_2\boldsymbol{w}_2$ for all c_1 and c_2. Every vector $\boldsymbol{v} \in \mathbb{W}^\perp$ must satisfy, for all c_1 and c_2,

$$\boldsymbol{w} \cdot \boldsymbol{v} = (c_1\boldsymbol{w}_1 + c_2\boldsymbol{w}_2) \cdot \boldsymbol{v} = c_1\boldsymbol{w}_1 \cdot \boldsymbol{v} + c_2\boldsymbol{w}_2 \cdot \boldsymbol{v} = 0\,.$$

The only way to be zero for all c_1 and c_2 is for *both* $\boldsymbol{w}_1 \cdot \boldsymbol{v} = 0$ and $\boldsymbol{w}_2 \cdot \boldsymbol{v} = 0$. For vectors $\boldsymbol{v} = (v_1, v_2, v_3)$ these two equations become the pair

$$2v_1 - 2v_2 + v_3 = 0 \quad \text{and} \quad 2v_1 + v_2 - 2v_3 = 0\,.$$

Adding twice the second to the first, and subtracting the first from the second give the equivalent pair

$$6v_1 - 3v_3 = 0 \quad \text{and} \quad 3v_2 - 3v_3 = 0\,.$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Both are satisfied for all $v_3 = t$ with $v_1 = t/2$ and $v_2 = t$. Therefore all possible $\boldsymbol{v}$ in the complement $\mathbb{W}^\perp$ are those in the form of the line $\boldsymbol{v} = (\frac{1}{2}t, t, t)$. That is, $\mathbb{W}^\perp = \operatorname{span}\{(\frac{1}{2}, 1, 1)\}$ (as illustrated below in stereo).

■

Activity 3.5.35. Which of the following vectors are in the orthogonal complement of the vector space spanned by $(3, -1, 1)$?

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

(c) $(3\,,5\,,-4)$ (d) $(6\,,-2\,,2)$

■

Example 3.5.36. Prove $\{\mathbf{0}\}^\perp = \mathbb{R}^n$ and $(\mathbb{R}^n)^\perp = \{\mathbf{0}\}$.

Solution:

- The only vector in $\{\mathbf{0}\}$ is $\boldsymbol{w} = \mathbf{0}$. Since all vectors $\boldsymbol{v} \in \mathbb{R}^n$ satisfy $\boldsymbol{w} \cdot \boldsymbol{v} = \mathbf{0} \cdot \boldsymbol{v} = 0$, by Definition 3.5.33 $\{\mathbf{0}\}^\perp = \mathbb{R}^n$.
- Certainly, $\mathbf{0} \in (\mathbb{R}^n)^\perp$ as $\boldsymbol{w} \cdot \mathbf{0} = 0$ for all vectors $\boldsymbol{w} \in \mathbb{R}^n$. Establish there are no others by contradiction. Assume a non-zero vector $\boldsymbol{v} \in (\mathbb{R}^n)^\perp$. Now set $\boldsymbol{w} = \boldsymbol{v} \in \mathbb{R}^n$, then $\boldsymbol{w} \cdot \boldsymbol{v} = \boldsymbol{v} \cdot \boldsymbol{v} = |\boldsymbol{v}|^2 \neq 0$ as $\boldsymbol{v}$ is non-zero. Consequently, a non-zero $\boldsymbol{v}$ cannot be in the complement. Thus $(\mathbb{R}^n)^\perp = \{\mathbf{0}\}$.

■

These examples find orthogonal complements that are lines, planes, or the entire space. These indicate that an orthogonal complement is generally a subspace as proved next.

Theorem 3.5.37 (orthogonal complement is subspace)**.** *For any 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 $\mathbb{W}^\perp$.*

Proof. Recall the Definition 3.4.3 of a subspace: we need to establish $\mathbb{W}^\perp$ has the zero vector, and is closed under addition and scaler multiplication.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

- For all $\boldsymbol{w} \in \mathbb{W}$, $\mathbf{0} \cdot \boldsymbol{w} = 0$ and so $\mathbf{0} \in \mathbb{W}^\perp$.
- Let $\boldsymbol{v}_1$, $\boldsymbol{v}_2 \in \mathbb{W}^\perp$, then for all $\boldsymbol{w} \in \mathbb{W}$ the dot product $(\boldsymbol{v}_1+\boldsymbol{v}_2)\cdot\boldsymbol{w} = \boldsymbol{v}_1\cdot\boldsymbol{w}+\boldsymbol{v}_2\cdot\boldsymbol{w} = 0+0 = 0$ and so $\boldsymbol{v}_1+\boldsymbol{v}_2 \in \mathbb{W}^\perp$.
- Let scalar $c \in \mathbb{R}$ and $\boldsymbol{v} \in \mathbb{W}^\perp$, then for all $\boldsymbol{w} \in \mathbb{W}$ the dot product $(c\boldsymbol{v}) \cdot \boldsymbol{w} = c(\boldsymbol{v} \cdot \boldsymbol{w}) = c0 = 0$ and so $c\boldsymbol{v} \in \mathbb{W}^\perp$.

Hence, by Definition 3.4.3, $\mathbb{W}^\perp$ is a subspace.

Further, as they are both subspaces, the zero vector is in both $\mathbb{W}$ and $\mathbb{W}^\perp$. Let vector $\boldsymbol{u}$ be any vector in both $\mathbb{W}$ and $\mathbb{W}^\perp$. As $\boldsymbol{u} \in \mathbb{W}^\perp$, by Definition 3.5.33 $\boldsymbol{u} \cdot \boldsymbol{w} = 0$ for all $\boldsymbol{w} \in \mathbb{W}$. But $\boldsymbol{u} \in \mathbb{W}$ also, so using this for $\boldsymbol{w}$ in the previous equation gives $\boldsymbol{u} \cdot \boldsymbol{u} = 0$; that is, $|\boldsymbol{u}|^2 = 0$. Hence vector $\boldsymbol{u}$ has to be the zero vector (Theorem 1.1.13). That is, $\mathbb{W} \cap \mathbb{W}^\perp = \{\mathbf{0}\}$. □

Activity 3.5.38. 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 3.5.39 (nullspace complementarity)**.** *For any $m \times n$ matrix A, the column space of A has* $\operatorname{null}(A^{\mathrm{T}})$ *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 $\mathbb{A}^\perp = \operatorname{null}(A^{\mathrm{T}})$. Further,* $\operatorname{null}(A)$ *in $\mathbb{R}^n$ is the orthogonal complement of the row space of A.*

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Proof. First, by Definition 3.5.33, any vector $\boldsymbol{v} \in \mathbb{A}^{\perp}$ is orthogonal to all vectors in the column space of A, in particular it is orthogonal to the columns of A:

$$\begin{aligned}
& \boldsymbol{a}_1 \cdot \boldsymbol{v} = 0\,,\ \boldsymbol{a}_2 \cdot \boldsymbol{v} = 0\,,\ \dots,\ \boldsymbol{a}_k \cdot \boldsymbol{v} = 0 \\
\iff\ & \boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{v} = 0\,,\ \boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{v} = 0\,,\ \dots,\ \boldsymbol{a}_k^{\mathrm{T}}\boldsymbol{v} = 0 \\
\iff\ & \begin{bmatrix} \boldsymbol{a}_1^{\mathrm{T}} \\ \boldsymbol{a}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{a}_k^{\mathrm{T}} \end{bmatrix} \boldsymbol{v} = \boldsymbol{0} \\
\iff\ & A^{\mathrm{T}}\boldsymbol{v} = \boldsymbol{0} \\
\iff\ & \boldsymbol{v} \in \text{null}(A^{\mathrm{T}}).
\end{aligned}$$

That is, $\mathbb{A}^{\perp} \subseteq \text{null}(A^{\mathrm{T}})$. Second, for any $\boldsymbol{v} \in \text{null}(A^{\mathrm{T}})$, recall that by Definition 3.4.10 for any vector $\boldsymbol{w}$ in the column space of A, there exists a linear combination $\boldsymbol{w} = c_1\boldsymbol{a}_1 + c_2\boldsymbol{a}_2 + \cdots + c_n\boldsymbol{a}_n$. Then

$$\begin{aligned}
\boldsymbol{w} \cdot \boldsymbol{v} &= (c_1\boldsymbol{a}_1 + c_2\boldsymbol{a}_2 + \cdots + c_n\boldsymbol{a}_n) \cdot \boldsymbol{v} \\
&= c_1(\boldsymbol{a}_1 \cdot \boldsymbol{v}) + c_2(\boldsymbol{a}_2 \cdot \boldsymbol{v}) + \cdots + c_n(\boldsymbol{a}_n \cdot \boldsymbol{v}) \\
&= c_1 0 + c_2 0 + \cdots + c_n 0 \quad (\text{from above} \iff) \\
&= 0\,,
\end{aligned}$$

and so by Definition 3.5.33 vector $\boldsymbol{v} \in \mathbb{A}^{\perp}$; that is, $\text{null}(A^{\mathrm{T}}) \subseteq \mathbb{A}^{\perp}$. Putting these two together, $\text{null}(A^{\mathrm{T}}) = \mathbb{A}^{\perp}$.

Lastly, that the $\text{null}(A)$ in $\mathbb{R}^n$ is the orthogonal complement of the row space of A follows from applying the above result to the matrix A^{T}. □

Example 3.5.40.

(a) Let the subspace $\mathbb{W} = \text{span}\{(2, -1)\}$. Find the orthogonal complement $\mathbb{W}^{\perp}$.

Solution: Here the subspace $\mathbb{W}$ is the column space of the matrix

$$W = \begin{bmatrix} 2 \\ -1 \end{bmatrix}.$$

To find $\mathbb{W}^\perp = \text{null}(W^{\mathrm{T}})$, solve $W^{\mathrm{T}}\boldsymbol{v} = \mathbf{0}$, that is, for vectors $\boldsymbol{v} = (u, v)$

$$\begin{bmatrix} 2 & -1 \end{bmatrix} \boldsymbol{v} = 2u - v = 0\,.$$

All solutions are $v = 2u$ (as illustrated). Hence $\boldsymbol{v} = (u, 2u) = (1, 2)u$, and so $\mathbb{W}^\perp = \text{span}\{(1, 2)\}$.

(b) Describe the subspace of $\mathbb{R}^3$ whose orthogonal complement is the plane $-\frac{1}{2}x - y + 2z = 0$.

Solution: The equation of the plane in $\mathbb{R}^3$ may be written

$$\begin{bmatrix} -\frac{1}{2} & -1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 0\,, \quad \text{that is } W^{\mathrm{T}}\boldsymbol{v} = 0$$

for matrix $W = \begin{bmatrix} \boldsymbol{w}_1 \end{bmatrix}$ and vectors $\boldsymbol{w}_1 = (-\frac{1}{2}, -1, 2)$ and $\boldsymbol{v} = (x, y, z)$. Since the plane is the nullspace of matrix W^{T}, the plane must be the orthogonal complement of the line $\mathbb{W} = \text{span}\{\boldsymbol{w}_1\}$ (as illustrated below).

(c) Find the orthogonal complement to the column space of matrix

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

Solution: The required orthogonal complement is the nullspace of A^{T}. Recall from Section 2.1 that for such small problems we find all solutions of $A^{\mathrm{T}}\boldsymbol{v} = \mathbf{0}$ by algebraic elimination; that is,

$$\begin{bmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & -1 \end{bmatrix} \boldsymbol{v} = \mathbf{0} \iff \begin{cases} v_1 + v_2 = 0\,, \\ -v_1 + v_3 = 0\,, \\ -v_2 - v_3 = 0\,, \end{cases}$$

$$\iff \begin{cases} v_2 = -v_1\,, \\ v_3 = v_1\,, \\ -v_2 - v_3 = v_1 - v_1 = 0\,. \end{cases}$$

Therefore all solutions of $A^{\mathrm{T}}\boldsymbol{v} = \mathbf{0}$ are of the form $v_1 = t$, $v_2 = -v_1 = -t$ and $v_3 = v_1 = t$; that is, $\boldsymbol{v} = (1, -1, 1)t$. Hence the orthogonal complement is $\text{span}\{(1, -1, 1)\}$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(d) Describe the orthogonal complement of the subspace spanned by the four vectors $(1,1,0,1,0,0)$, $(-1,0,1,0,1,0)$, $(0,-1,-1,0,0,1)$ and $(0,0,0,-1,-1,-1)$.

Solution: Arrange these vectors as the four columns of a matrix, say

$$A = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix},$$

then seek $\operatorname{null}(A^{\mathrm{T}})$, the solutions of $A^{\mathrm{T}}\boldsymbol{x} = \mathbf{0}$. Adapt Procedure 3.3.15 to solve $A^{\mathrm{T}}\boldsymbol{x} = \mathbf{0}$:

i. Example 3.5.5 computed an SVD $A = USV^{\mathrm{T}}$ for this matrix A, which gives the SVD $A^{\mathrm{T}} = VS^{\mathrm{T}}U^{\mathrm{T}}$ for the transpose where (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 = ...
```

ii. $V\boldsymbol{z} = \mathbf{0}$ determines $\boldsymbol{z} = \mathbf{0}$.

iii. $S^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{z} = \mathbf{0}$ determines $y_1 = y_2 = y_3 = 0$ as there are three non-zero singular values, and y_4, y_5 and y_6 are free variables; that is, $\boldsymbol{y} = (0,0,0,y_4,y_5,y_6)$.

iv. Denoting the columns of U by $\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_6$, the solutions of $U^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{y}$ are $\boldsymbol{x} = U\boldsymbol{y} = \boldsymbol{u}_4 y_4 + \boldsymbol{u}_5 y_5 + \boldsymbol{u}_6 y_6$.

That is, the orthogonal complement is the three dimensional subspace $\operatorname{span}\{\boldsymbol{u}_4, \boldsymbol{u}_5, \boldsymbol{u}_6\}$ in $\mathbb{R}^6$, where (2 d.p.)

$$\begin{aligned} \boldsymbol{u}_4 &= (-0.26, 0.06, -0.64, 0.21, 0.37, -0.58), \\ \boldsymbol{u}_5 &= (0.64, -0.49, 0.19, -0.15, 0.45, -0.30), \\ \boldsymbol{u}_6 &= (-0.15, -0.51, 0.24, 0.66, -0.40, -0.26). \end{aligned}$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

■

In the previous Example 3.5.40d there are three non-zero singular values in the first three rows of S. These three nonzero singular values determine that the first three columns of U form a basis for the column space of A. The example argues that the remaining three columns of U form a basis for the orthogonal complement of the column space. That is, all six of the columns of the orthogonal U are used in either the column space or its complement. This is generally true.

Activity 3.5.41. 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) 4×3 (c) 7×3 (d) 7×5

■

Example 3.5.42. Recall the cases of Example 3.5.40.

3.5.40a : $\dim \mathbb{W} + \dim \mathbb{W}^\perp = 1 + 1 = 2 = \dim \mathbb{R}^2$.

3.5.40b : $\dim \mathbb{W} + \dim \mathbb{W}^\perp = 1 + 2 = 3 = \dim \mathbb{R}^3$.

3.5.40c : $\dim \mathbb{W} + \dim \mathbb{W}^\perp = 2 + 1 = 3 = \dim \mathbb{R}^3$.

3.5.40d : $\dim \mathbb{W} + \dim \mathbb{W}^\perp = 3 + 3 = 6 = \dim \mathbb{R}^6$.

■

Recall the Rank Theorem 3.4.39 connects the dimension of a space with the dimensions of a nullspace and column space of a matrix. Since a subspace is closely connected to matrices, and its orthogonal complement is connected to nullspaces, then the Rank Theorem should say something general here.

Theorem 3.5.43. *Let $\mathbb{W}$ be a subspace of $\mathbb{R}^n$, then $\dim \mathbb{W} + \dim \mathbb{W}^\perp = n$; equivalently, $\dim \mathbb{W}^\perp = n - \dim \mathbb{W}$.*

Proof. Let the columns of a matrix W form an orthonormal basis for the subspace $\mathbb{W}$ (Theorem 3.4.29 asserts a basis exists). Theorem 3.5.39 establishes that $\mathbb{W}^\perp = \operatorname{null}(W^{\mathrm{T}})$. Equating dimensions of both sides,

$$\begin{aligned}
\dim \mathbb{W}^\perp &= \operatorname{nullity}(W^{\mathrm{T}}) \quad \text{(from Defn. 3.4.36)} \\
&= n - \operatorname{rank}(W^{\mathrm{T}}) \quad \text{(from Rank Thm. 3.4.39)} \\
&= n - \operatorname{rank}(W) \quad \text{(from Thm. 3.3.23)} \\
&= n - \dim \mathbb{W} \quad \text{(from Proc. 3.4.23)},
\end{aligned}$$

as required. □

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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 3.5.44. Recall from Example 3.5.34a that subspace $\mathbb{W} = \operatorname{span}\{(3,4)\}$ has orthogonal complement $\mathbb{W}^\perp = \operatorname{span}\{(-4,3)\}$, as illustrated below.

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 3.5.50).

For 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 $\text{perp} = (-2.28,3.04) \in \mathbb{W}^\perp$. ■

Activity 3.5.45. 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) $(-2\,,-2) + (7\,,-7)$ (b) $(5\,,5) + (0\,,-14)$

(c) $(7\,,7) + (-2\,,2)$ (d) $(9\,,-9) + (-4\,,0)$

■

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 3.5.46 (perpendicular component)**.** *Let $\mathbb{W}$ be a subspace of $\mathbb{R}^n$. For any 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 3.5.47. (a) Let the subspace $\mathbb{W}$ be the span of $(-2,-3,6)$. Find the perpendicular component to $\mathbb{W}$ of the vector $(4,1,3)$. Verify the perpendicular component lies in the plane $-2x - 3y + 6z = 0$.

Solution: Projection is easiest with a unit vector. Obtain a unit vector to span $\mathbb{W}$ by normalising the basis vector to $\boldsymbol{w}_1 = (-2,-3,6)/\sqrt{2^2+3^2+6^2} = (-2,-3,6)/7$. Then

$$\operatorname{perp}_{\mathbb{W}}(4,1,3) = (4,1,3) - \boldsymbol{w}_1(\boldsymbol{w}_1 \cdot (4,1,3))$$

$$= (4,1,3) - \boldsymbol{w}_1(-8-3+18)/7$$
$$= (4,1,3) - \boldsymbol{w}_1 = (30,10,15)/7\,.$$

For $(x,y,z) = (30,10,15)/7$ we find

$$-2x - 3y + 6z = \tfrac{1}{7}(-60-30+90) = \tfrac{1}{7}0 = 0\,.$$

Hence $\text{perp}_{\mathbb{W}}(4,1,3)$ lies in the plane $-2x-3y+6z=0$ (which is the orthogonal complement $\mathbb{W}^\perp$, as illustrated in in stereo below).

(b) For the vector $(-5,-1,6)$ find its perpendicular component to the subspace $\mathbb{W}$ spanned by $(-2,-3,6)$. Verify the perpendicular component lies in the plane $-2x-3y+6z=0$.

Solution: As in the previous case, use the basis vector $\boldsymbol{w}_1 = (-2,-3,6)/7$. Then

$$\begin{aligned}\text{perp}_{\mathbb{W}}(-5,-1,6) &= (-5,-1,6) - \boldsymbol{w}_1(\boldsymbol{w}_1\cdot(-5,-1,6))\\ &= (-5,-1,6) - \boldsymbol{w}_1(10+3+36)/7\\ &= (-5,-1,6) - \boldsymbol{w}_1 7 = (-3,2,0).\end{aligned}$$

For $(x,y,z) = (-3,2,0)$ we find $-2x-3y+6z = 6-6+0 = 0$. Hence $\text{perp}_{\mathbb{W}}(-5,-1,6)$ lies in the plane $-2x-3y+6z=0$ (which is the orthogonal complement $\mathbb{W}^\perp$, as illustrated below in stereo).

(c) Let the subspace $\mathbb{X} = \text{span}\{(2,-2,1),(2,1,-2)\}$. Determine the perpendicular component of each of the two vectors $\boldsymbol{y} = (3,2,1)$ and $\boldsymbol{z} = (3,-3,-3)$.

Solution: Computing $\text{proj}_{\mathbb{X}}$ needs an orthonormal basis for $\mathbb{X}$ (Definition 3.5.22). The two vectors in the span are orthogonal, so normalise them to $\boldsymbol{w}_1 = (2,-2,1)/3$ and $\boldsymbol{w}_2 = (2,1,-2)/3$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

- Then for the first vector $\boldsymbol{y} = (3, 2, 1)$,

$$\begin{aligned}
\operatorname{perp}_{\mathbb{X}}(\boldsymbol{y}) &= \boldsymbol{y} - \operatorname{proj}_{\mathbb{X}}(\boldsymbol{y}) \\
&= \boldsymbol{y} - \boldsymbol{w}_1(\boldsymbol{w}_1 \cdot \boldsymbol{y}) - \boldsymbol{w}_2(\boldsymbol{w}_2 \cdot \boldsymbol{y}) \\
&= \boldsymbol{y} - \boldsymbol{w}_1(6 - 4 + 1)/3 - \boldsymbol{w}_2(6 + 2 - 2)/3 \\
&= \boldsymbol{y} - \boldsymbol{w}_1 - 2\boldsymbol{w}_2 \\
&= (3, 2, 1) - (2, -2, 1)/3 - (4, 2, -4)/3 \\
&= (1, 2, 2)
\end{aligned}$$

(as illustrated below in brown).

- For the second vector $\boldsymbol{z} = (3, -3, -3)$ (in green in the picture above),

$$\begin{aligned}
\operatorname{perp}_{\mathbb{X}}(\boldsymbol{z}) &= \boldsymbol{z} - \operatorname{proj}_{\mathbb{X}}(\boldsymbol{z}) \\
&= \boldsymbol{z} - \boldsymbol{w}_1(\boldsymbol{w}_1 \cdot \boldsymbol{z}) - \boldsymbol{w}_2(\boldsymbol{w}_2 \cdot \boldsymbol{z}) \\
&= \boldsymbol{z} - \boldsymbol{w}_1(6 + 6 - 3)/3 - \boldsymbol{w}_2(6 - 3 + 6)/3 \\
&= \boldsymbol{z} - 3\boldsymbol{w}_1 - 3\boldsymbol{w}_2 \\
&= (2, -2, -2) - (2, -2, 1) - (2, 1, -2) \\
&= (-1, -2, -2).
\end{aligned}$$

■

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 3.5.48 (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* $\operatorname{perp}_{\mathbb{W}}(\boldsymbol{v}) \in \mathbb{W}^{\perp}$.

Proof. Let vectors $\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_k$ form an orthonormal basis for the subspace $\mathbb{W}$ (a basis exists by Theorem 3.4.29). Let the $n \times k$ matrix $W = \begin{bmatrix} \boldsymbol{w}_1 & \boldsymbol{w}_2 & \cdots & \boldsymbol{w}_k \end{bmatrix}$ so subspace $\mathbb{W}$ is the column space of matrix W, then Theorem 3.5.39 asserts we just need to check that $W^{\mathrm{T}} \operatorname{perp}_{\mathbb{W}}(\boldsymbol{v}) = \boldsymbol{0}$. Consider

$$\begin{aligned}
W^{\mathrm{T}} \operatorname{perp}_{\mathbb{W}}(\boldsymbol{v}) &= W^{\mathrm{T}} [\boldsymbol{v} - \operatorname{proj}_{\mathbb{W}}(\boldsymbol{v})] && \text{(from Defn. 3.5.46)} \\
&= W^{\mathrm{T}} [\boldsymbol{v} - (WW^{\mathrm{T}})\boldsymbol{v}] && \text{(from Thm. 3.5.28)}
\end{aligned}$$

$$
\begin{aligned}
&= W^{\mathrm{T}}\boldsymbol{v} - W^{\mathrm{T}}(WW^{\mathrm{T}})\boldsymbol{v} \quad \text{(by distributivity)}\\
&= W^{\mathrm{T}}\boldsymbol{v} - (W^{\mathrm{T}}W)W^{\mathrm{T}}\boldsymbol{v} \quad \text{(by associativity)}\\
&= W^{\mathrm{T}}\boldsymbol{v} - I_kW^{\mathrm{T}}\boldsymbol{v} \quad \text{(only if } W^{\mathrm{T}}W = I_k\text{)}\\
&= W^{\mathrm{T}}\boldsymbol{v} - W^{\mathrm{T}}\boldsymbol{v} = \boldsymbol{0}\,.
\end{aligned}
$$

Hence $\operatorname{perp}_{\mathbb{W}}(\boldsymbol{v}) \in \operatorname{null}(W^{\mathrm{T}})$ and so is in $\mathbb{W}^{\perp}$ (by Theorem 3.5.39).

But this proof only holds if $W^{\mathrm{T}}W = I_k$. To establish this identity, use the same argument as in the proof of Theorem 3.2.48a $\iff$ 3.2.48b:

$$
\begin{aligned}
W^{\mathrm{T}}W &= \begin{bmatrix} \boldsymbol{w}_1^{\mathrm{T}} \\ \boldsymbol{w}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{w}_k^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \boldsymbol{w}_1 & \boldsymbol{w}_2 & \cdots & \boldsymbol{w}_k \end{bmatrix}\\
&= \begin{bmatrix} \boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{w}_1 & \boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{w}_2 & \cdots & \boldsymbol{w}_1^{\mathrm{T}}\boldsymbol{w}_k \\ \boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{w}_1 & \boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{w}_2 & \cdots & \boldsymbol{w}_2^{\mathrm{T}}\boldsymbol{w}_k \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{w}_k^{\mathrm{T}}\boldsymbol{w}_1 & \boldsymbol{w}_k^{\mathrm{T}}\boldsymbol{w}_2 & \cdots & \boldsymbol{w}_k^{\mathrm{T}}\boldsymbol{w}_k \end{bmatrix}\\
&= \begin{bmatrix} \boldsymbol{w}_1\cdot\boldsymbol{w}_1 & \boldsymbol{w}_1\cdot\boldsymbol{w}_2 & \cdots & \boldsymbol{w}_1\cdot\boldsymbol{w}_k \\ \boldsymbol{w}_2\cdot\boldsymbol{w}_1 & \boldsymbol{w}_2\cdot\boldsymbol{w}_2 & \cdots & \boldsymbol{w}_2\cdot\boldsymbol{w}_k \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{w}_k\cdot\boldsymbol{w}_1 & \boldsymbol{w}_k\cdot\boldsymbol{w}_2 & \cdots & \boldsymbol{w}_k\cdot\boldsymbol{w}_k \end{bmatrix}\\
&= I_k
\end{aligned}
$$

as vectors $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_k$ are an orthonormal set (from Definition 3.2.38, $\boldsymbol{w}_i\cdot\boldsymbol{w}_j = 0$ for $i \neq j$ and $|\boldsymbol{w}_i|^2 = \boldsymbol{w}_i\cdot\boldsymbol{w}_i = 1$). □

Example 3.5.49. 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 $\mathbb{W}$:

3.5.44 : $(2,4) = (3.2, 2.4) + (-1.2, 1.6)$ and
$(-5,1) = (-2.72, -2.04) + (-2.28, 3.04)$;

3.5.47b : $(-5,-1,6) = (-2,-3,6) + (-3,2,0)$;

3.5.47c : $(3,2,1) = (2,0,-1) + (1,2,2)$ 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 $\mathbb{W}^{\perp}$.

v0.2a

Theorem 3.5.50 (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 $\boldsymbol{v}$.*

Proof.

- First establish existence. By Definition 3.5.46, $\operatorname{perp}_{\mathbb{W}}(\boldsymbol{v}) = \boldsymbol{v} - \operatorname{proj}_{\mathbb{W}}(\boldsymbol{v})$, so it follows that $\boldsymbol{v} = \operatorname{proj}_{\mathbb{W}}(\boldsymbol{v}) + \operatorname{perp}_{\mathbb{W}}(\boldsymbol{v}) = \boldsymbol{w} + \boldsymbol{n}$ when we set $\boldsymbol{w} = \operatorname{proj}_{\mathbb{W}}(\boldsymbol{v}) \in \mathbb{W}$ and $\boldsymbol{n} = \operatorname{perp}_{\mathbb{W}}(\boldsymbol{v}) \in \mathbb{W}^\perp$.
- Second establish uniqueness by contradiction. Suppose there is another decomposition $\boldsymbol{v} = \boldsymbol{w}' + \boldsymbol{n}'$ where $\boldsymbol{w}' \in \mathbb{W}$ and $\boldsymbol{n}' \in \mathbb{W}^\perp$. Then $\boldsymbol{w} + \boldsymbol{n} = \boldsymbol{v} = \boldsymbol{w} + \boldsymbol{n}'$. Rearranging gives $\boldsymbol{w} - \boldsymbol{w}' = \boldsymbol{n}' - \boldsymbol{n}$. By closure of a subspace under vector addition (Definition 3.4.3), the left-hand side is in $\mathbb{W}$ and the right-hand side is in $\mathbb{W}^\perp$, so the two sides must be both in $\mathbb{W}$ and $\mathbb{W}^\perp$. The zero vector is the only common vector to the two subspaces (Theorem 3.5.37), so $\boldsymbol{w} - \boldsymbol{w}' = \boldsymbol{n}' - \boldsymbol{n} = \boldsymbol{0}$, and hence both $\boldsymbol{w} = \boldsymbol{w}'$ and $\boldsymbol{n} = \boldsymbol{n}'$. That is, the decomposition must be unique.

□

Example 3.5.51. 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 $\mathbb{X}^\perp$.

Solution: In each case, the two brown vectors shown are the decomposition, with proj $\in \mathbb{X}$ and perp $\in \mathbb{X}^\perp$.

■

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.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

In particular, the next theorem gives a geometric view of the 'least square' solution of Procedure 3.5.4: in that procedure the minimal change of the right-hand side $\boldsymbol{b}$ to make the linear equation $A\boldsymbol{x} = \boldsymbol{b}$ consistent (Theorem 3.5.7) is also to be viewed as the projection of the right-hand side $\boldsymbol{b}$ to the *closest* point in the columns space of the matrix. That is, the 'least square' procedure solves $A\boldsymbol{x} = \text{proj}_{\mathbb{A}}(\boldsymbol{b})$.

Theorem 3.5.52 (best approximation). *Given any vector $\boldsymbol{v}$ in $\mathbb{R}^n$, and any subspace $\mathbb{W}$ in $\mathbb{R}^n$, then $\text{proj}_{\mathbb{W}}(\boldsymbol{v})$ is the closest vector in $\mathbb{W}$ to $\boldsymbol{v}$; that is, $|\boldsymbol{v} - \text{proj}_{\mathbb{W}}(\boldsymbol{v})| \le |\boldsymbol{v} - \boldsymbol{w}|$ for all $\boldsymbol{w} \in \mathbb{W}$.*

Proof. For any vector $\boldsymbol{w} \in \mathbb{W}$, consider the triangle formed by the three vectors $\boldsymbol{v} - \text{proj}_{\mathbb{W}}(\boldsymbol{v})$, $\boldsymbol{v} - \boldsymbol{w}$ and $\boldsymbol{w} - \text{proj}_{\mathbb{W}}(\boldsymbol{v})$ (the stereo illustration above schematically plots this triangle in red). This is a right-angle triangle as $\boldsymbol{w} - \text{proj}_{\mathbb{W}}(\boldsymbol{v}) \in \mathbb{W}$ by closure of the subspace $\mathbb{W}$, and as $\boldsymbol{v} - \text{proj}_{\mathbb{W}}(\boldsymbol{v}) = \text{perp}_{\mathbb{W}}(\boldsymbol{v}) \in \mathbb{W}^{\perp}$. Then Pythagoras tells us

$$\begin{aligned} |\boldsymbol{v} - \boldsymbol{w}|^2 &= |\boldsymbol{v} - \text{proj}_{\mathbb{W}}(\boldsymbol{v})|^2 + |\boldsymbol{w} - \text{proj}_{\mathbb{W}}(\boldsymbol{v})|^2 \\ &\ge |\boldsymbol{v} - \text{proj}_{\mathbb{W}}(\boldsymbol{v})|^2. \end{aligned}$$

Hence $|\boldsymbol{v} - \boldsymbol{w}| \ge |\boldsymbol{v} - \text{proj}_{\mathbb{W}}(\boldsymbol{v})|$ for all $\boldsymbol{w} \in \mathbb{W}$. □

3.5.4 Exercises

Exercise 3.5.1. During an experiment on the strength of beams, you and your partner measure the length of a crack in the beam. With vernier callipers you measure the crack as 17.8 mm long, whereas your partner measures it as 18.4 mm long.

- Write this information as a simple matrix-vector equation for the as yet to be decided length x, and involving the matrix $A = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.
- Confirm that an SVD of the matrix is

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

- Use the SVD to 'best' solve the inconsistent equations and estimate the length of the crack is $x \approx 18.1\,\text{mm}$—the average of the two measurements.

Exercise 3.5.2. In measuring the amount of butter to use in cooking a recipe you weigh a container to have 207 g (grams), then a bit later weigh it at 211 g. Wanting to be more accurate you weigh the butter container a third time and find 206 g.

- Write this information as a simple matrix-vector equation for the as yet to be decided weight x, and involving the matrix $B = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.
- Confirm that an SVD of the matrix is

$$B = \begin{bmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} \sqrt{3} \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \end{bmatrix}^{\mathrm{T}}.$$

- Use the SVD to 'best' solve the inconsistent equations and estimate the butter container weighs $x \approx 208\,\text{g}$—the average of the three measurements.

Exercise 3.5.3. Consider three sporting teams that play each other in a round robin event: Newark, Yonkers, and Edison: Yonkers beat Newark, 2 to 0; Edison beat Newark 5 to 2; and Edison beat Yonkers 3 to 2. Assuming the teams can be rated, and based upon the scores, write three equations that ideally relate the team ratings. Use Procedure 3.5.4 to estimate the ratings.

Exercise 3.5.4. Consider three sporting teams that play each other in a round robin event: Adelaide, Brisbane, and Canberra: Adelaide beat Brisbane, 5 to 1; Canberra beat Adelaide 5 to 0; and Brisbane beat Canberra 2 to 1. Assuming the teams can be rated, and based upon the scores, write three equations that ideally relate the team ratings. Use Procedure 3.5.4 to estimate the ratings.

Exercise 3.5.5. Consider four sporting teams that play each other in a round robin event: Acton, Barbican, Clapham, and Dalston. Table 3.8 summarises the results of the six matches played. Assuming the teams can be rated, and based upon the scores, write six equations that ideally relate the team ratings. Use Procedure 3.5.4 to estimate the ratings.

v0.2a

Table 3.8: the results of six matches played in a round robin: the scores are games/goals/points scored by each when playing the others. For example, Clapham beat Acton 4 to 2. Exercise 3.5.5 rates these teams.

	Acton	Barbican	Clapham	Dalston
Acton	-	2	2	6
Barbican	2	-	2	6
Clapham	4	4	-	5
Dalston	3	1	0	-

Table 3.9: the results of ten matches played in a round robin: the scores are games/goals/points scored by each when playing the others. For example, Atlanta beat Concord 3 to 2. Exercise 3.5.6 rates these teams.

	Atlanta	Boston	Concord	Denver	Frankfort
Atlanta	-	3	3	2	5
Boston	2	-	2	3	8
Concord	2	7	-	6	1
Denver	2	2	1	-	5
Frankfort	2	3	6	7	-

Exercise 3.5.6. Consider five sporting teams that play each other in a round robin event: Atlanta, Boston, Concord, Denver, and Frankfort. Table 3.9 summarises the results of the ten matches played. Assuming the teams can be rated, and based upon the scores, write ten equations that ideally relate the team ratings. Use Procedure 3.5.4 to estimate the ratings.

Exercise 3.5.7. Consider six sporting teams in a weekly competition: Algeria, Botswana, Chad, Djibouti, Ethiopia, and Gabon. In the first week of competition Algeria beat Botswana 3 to 0, Chad and Djibouti drew 3 all, and Ethiopia beat Gabon 4 to 2. In the second week of competition Chad beat Algeria 4 to 2, Botswana beat Ethiopia 4 to 2, Djibouti beat Gabon 4 to 3. In the third week of competition Algeria beat Ethiopia 4 to 1, Botswana beat Djibouti 3 to 1, Chad drew with Gabon 2 all. Assuming the teams can be rated, and based upon the scores after the first three weeks, write nine equations that ideally relate the ratings of the six teams. Use Procedure 3.5.4 to estimate the ratings.

Exercise 3.5.8. In calibrating a vortex flowmeter the following flow rates were obtained for various applied voltages.[21]

voltage (V)	0.97	1.29	1.81	2.41	2.85	3.09	3.96
flow rate (litre/s)	0.01	0.27	0.59	0.94	1.21	1.36	2.14

[21] Adapted from https://www.che.udel.edu/pdf/FittingData.pdf, 2016

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Table 3.10: the body weight and heat production of various mammals (Kleiber 1947). Recall that numbers written as xEn denote the number $x \cdot 10^n$.

animal	body weight (kg)	heat prod. (kcal/day)
mouse	1.95E−2	3.06E+0
rat	2.70E−1	2.61E+1
cat	3.62E+0	1.56E+2
dog	1.28E+1	4.35E+2
goat	2.58E+1	7.50E+2
sheep	5.20E+1	1.14E+3
cow	5.34E+2	7.74E+3
elephant	3.56E+3	4.79E+4

Use Procedure 3.5.4 to find the best straight line that gives the flow rate as a function of the applied voltage. Plot both the data and the fitted straight line.

Discover power laws Exercises 3.5.9–3.5.12 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'.

Exercise 3.5.9. Table 3.10 lists data on the body weight and heat production of various mammals. As in Example 3.5.10, use this data to discover Kleiber's power law that (heat) $\propto$ (weight)$^{3/4}$. Graph the data on a log-log plot, fit a straight line, check the correspondence between neglected parts of the right-hand side and the quality of the graphical fit, describe the power law.

Exercise 3.5.10. Table 3.11 lists data on river lengths and basin areas of some Russian rivers. As in Example 3.5.10, use this data to discover Hack's exponent in the power law that (length) $\propto$ (area)$^{0.58}$. Graph the data on a log-log plot, fit a straight line, check the correspondence between neglected parts of the right-hand side and the quality of the graphical fit, describe the power law.

Exercise 3.5.11. Find for another country some river length and basin area data akin to that of Exercise 3.5.10. Confirm, or otherwise, Hack's exponent for your data. Write a short report.

Exercise 3.5.12. The area-length relationship of a river is expected to be (length) $\propto$ (area)$^{1/2}$, so it is a puzzle as to why one consistently finds Hack's exponent (e.g., Exercise 3.5.10). The puzzle may be answered by the surprising notion that rivers do not have a well

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Table 3.11: river length and basin area for some Russian rivers (Arnold 2014, p.154).

river	basin area (km^2)	length (km)
Moscow	17640	502
Protva	4640	275
Vorya	1160	99
Dubna	5474	165
Istra	2120	112
Nara	2170	156
Pakhra	2720	129
Skhodnya	259	47
Volgusha	265	40
Pekhorka	513	42
Setun	187	38
Yauza	452	41

Table 3.12: given a measuring stick of some length, compute the length of the west coast of Britain (Mandelbrot 1982, Plate 33).

stick length (km)	coast length (km)
10.4	2845
30.2	2008
99.6	1463
202.	1138
532.	929
933.	914

defined length! L. F. Richardson first established this remarkable notion for coastlines.

Table 3.12 lists data on the length of the west coast of Britain computed by using measuring sticks of various lengths: as one uses a smaller and smaller measuring stick, more and more bays and inlets are resolved and measured which increases the computed coast length. As in Example 3.5.10, use this data to discover the power law that the coast length $\propto$ (stick)$^{-1/4}$. Hence as the measuring stick length goes to 'zero', the coast length goes to 'infinity'! Graph the data on a log-log plot, fit a straight line, check the correspondence between neglected parts of the right-hand side and the quality of the graphical fit, describe the power law.

I do not condone nor endorse such naive one dimensional ranking of complex multi-faceted institutions. This exercise simply illustrates a technique that deconstructs such a credulous endeavour.

Exercise 3.5.13. Table 3.13 lists nine of the US universities ranked by an organisation in 2013, in the order they list. The table also lists three of the attributes used to generate the ranked list. Find a

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Table 3.13: a selection of nine of the US universities ranked in 2013 by *The Center for Measuring University Performance* [http://mup.asu.edu/research_data.html]. Among others, these particular nine universities are listed by the Center in the following order. The other three columns give just three of the attributes used to create their ranked list.

Institution	Research fund(M$)	Faculty awards	Median SAT U/G
Stanford University	868	45	1455
Yale University	654	45	1500
University of California, San Diego	1004	35	1270
University of Pittsburgh, Pittsburgh	880	22	1270
Vanderbilt University	535	19	1440
Pennsylvania State University, University Park	677	20	1195
Purdue University, West Lafayette	520	22	1170
University of Utah	410	12	1110
University of California, Santa Barbara	218	11	1205

formula that approximately reproduces the listed ranking from the three given attributes.

(a) Pose the rank of the ith institution is a linear function of the attributes and a constant, say the rank $i = x_1 f_i + x_2 a_i + x_3 s_i + x_4$ where f_i denotes the funding, a_i denotes the awards, and s_i denotes the SAT.

(b) Form a system of nine equations that we would ideally solve to find the coefficients $\boldsymbol{x} = (x_1, x_2, x_3, x_4)$.

(c) Enter the data into MATLAB/Octave and find a best approximate solution (you should find the formula is roughly that rank $\approx 97 - 0.01 f_i - 0.07 a_i - 0.01 s_i$).

(d) Discuss briefly how well the approximation reproduces the ranking of the list.

Exercise 3.5.14. For each of the following lines and planes, use an SVD to find the point closest to the origin in the line or plane. For the lines in 2D, draw a graph to show the answer is correct.

(a) $5x_1 - 12x_2 = 169$

(b) $x_1 - 2x_2 = 5$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(c) $-x + y = -1$

(d) $-2p - 3q = 5$

(e) $2x_1 - 3x_2 + 6x_3 = 7$

(f) $x_1 + 4x_2 - 8x_3 = 27$

(g) $2u_2 - 5u_2 - 3u_3 = -2$

(h) $q_1 + q_2 - 5q_3 = 2$

Exercise 3.5.15. Following the computed tomography Example 3.5.16, predict the densities in the body if the fraction of X-ray energy measured in the six paths is $\boldsymbol{f} = (0.9, 0.2, 0.8, 0.9, 0.8, 0.2)$ respectively. Draw an image of your predictions. Which region is the most absorbing (least transmitting)?

Exercise 3.5.16. In an effort to remove the need for requiring the 'smallest', most washed out, CT-scan, you make three more measurements, as illustrated in the margin, so that you obtain nine equations for the nine unknowns.

(a) Write down the nine equations for the transmission factors in terms of the fraction of X-ray energy measured after passing through the body. Take logarithms to form a system of linear equations.

(b) Encode the matrix A of the system and check `rcond(A)`: curses, `rcond` is terrible, so we must still use an SVD.

(c) Suppose the measured fractions of X-ray energy are $\boldsymbol{f} = (0.05, 0.35, 0.33, 0.31, 0.05, 0.36, 0.07, 0.32, 0.51)$. Use an SVD to find the 'grayest' transmission factors consistent with the measurements.

(d) Which part of the body is predicted to be the most absorbing?

Exercise 3.5.17. Use a little higher resolution in computed tomography: suppose the two dimensional 'body' is notionally divided into sixteen regions as illustrated in the margin. Suppose a CT-scan takes thirteen measurements of the intensity of an X-ray after passing through the shown paths, and that the fraction of the X-ray energy that is measured is $\boldsymbol{f} = (0.29, 0.33, 0.07, 0.35, 0.36, 0.07, 0.31, 0.32, 0.62, 0.40, 0.06, 0.47, 0.58)$.

(a) Write down the thirteen equations for the sixteen transmission factors in terms of the fraction of X-ray energy measured after passing through the body. Take logarithms to form a system of linear equations.

(b) Encode the matrix A of the system and find it has rank twelve.

(c) Use an SVD to find the 'grayest' transmission factors consistent with the measurements.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(d) In which square pixel is the 'lump' of dense material?

Exercise 3.5.18. This exercise is for those who, in Calculus courses, have studied constrained optimisation with Lagrange multipliers. The aim is to derive how to use the SVD to find the vector $\boldsymbol{x}$ that minimises $|A\boldsymbol{x} - \boldsymbol{b}|$ such that the length $|\boldsymbol{x}| \le \alpha$ for some given magnitude α.

- Given vector $\boldsymbol{z} \in \mathbb{R}^n$ and $n \times n$ diagonal matrix $S = \operatorname{diag}(\sigma_1, \sigma_2, \dots, \sigma_n)$, with $\sigma_1, \sigma_2, \dots, \sigma_n > 0$. In one of two cases, use a Lagrange multiplier λ to find the vector $\boldsymbol{y}$ (as a function of λ and $\boldsymbol{z}$) that minimises $|S\boldsymbol{y} - \boldsymbol{z}|^2$ such that $|\boldsymbol{y}|^2 \le \alpha^2$ for some given magnitude α: show that the multiplier λ satisfies a polynomial equation of degree n.
- What can be further deduced if one or more $\sigma_j = 0$?
- Use an SVD of $n \times n$ matrix A to find the vector $\boldsymbol{x} \in \mathbb{R}^n$ that minimises $|A\boldsymbol{x} - \boldsymbol{b}|$ such that the length $|\boldsymbol{x}| \le \alpha$ for some given magnitude α. Use that multiplication by orthogonal matrices preserves lengths.

Exercise 3.5.19. For each pair of vectors, draw the orthogonal projection $\operatorname{proj}_{\boldsymbol{u}}(\boldsymbol{v})$.

v0.2a

Exercise 3.5.20. For the following pairs of vectors: compute the orthogonal projection $\text{proj}_{\boldsymbol{u}}(\boldsymbol{v})$; and hence find the 'best' approximate solution to the inconsistent system $\boldsymbol{u}\,x = \boldsymbol{v}$.

(a) $\boldsymbol{u} = (2\,,1)$, $\boldsymbol{v} = (2\,,0)$

(b) $\boldsymbol{u} = (4\,,-1)$, $\boldsymbol{v} = (-1\,,1)$

(c) $\boldsymbol{u} = (6\,,0)$, $\boldsymbol{v} = (-1\,,-1)$

(d) $\boldsymbol{u} = (2\,,-2)$, $\boldsymbol{v} = (-1\,,2)$

(e) $\boldsymbol{u} = (4\,,5\,,-1)$, $\boldsymbol{v} = (-1\,,2\,,-1)$

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

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

(h) $\boldsymbol{u} = (-1\,,-7\,,5)$, $\boldsymbol{v} = (1\,,1\,,-1)$

(i) $\boldsymbol{u} = (2\,,4\,,0\,,-1)$, $\boldsymbol{v} = (0\,,2\,,-1\,,0)$

(j) $\boldsymbol{u} = (3\,,-6\,,-3\,,-2)$, $\boldsymbol{v} = (-1\,,1\,,0\,,1)$

(k) $\boldsymbol{u} = (1\,,2\,,1\,,-1\,,-4)$, $\boldsymbol{v} = (1\,,-1\,,2\,,-2\,,1)$

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

Exercise 3.5.21. For each of the following subspaces $\mathbb{W}$ (given as the span of orthogonal vectors), and the given vectors $\boldsymbol{v}$, find the orthogonal projection $\text{proj}_{\mathbb{W}}(\boldsymbol{v})$.

(a) $\mathbb{W} = \text{span}\{(-6\,,-6\,,7)\,,(2\,,-9\,,-6)\}$, $\boldsymbol{v} = (0\,,1\,,-2)$

(b) $\mathbb{W} = \text{span}\{(4\,,-7\,,-4)\,,(1\,,-4\,,8)\}$, $\boldsymbol{v} = (0\,,-4\,,-1)$

(c) $\mathbb{W} = \text{span}\{(-6\,,-3\,,-2)\,,(-2\,,6\,,-3)\}$, $\boldsymbol{v} = (3\,,-2\,,-3)$

(d) $\mathbb{W} = \text{span}\{(1\,,8\,,-4)\,,(-8\,,-1\,,-4)\}$, $\boldsymbol{v} = (-2\,,2\,,0)$

(e) $\mathbb{W} = \text{span}\{(-1\,,2\,,-2)\,,(-2\,,1\,,2)\,,(2\,,2\,,1)\}$, $\boldsymbol{v} = (3\,,-1\,,1)$

(f) $\mathbb{W} = \text{span}\{(-2\,,4\,,-2\,,5)\,,(-5\,,-2\,,-4\,,-2)\}$, $\boldsymbol{v} = (1\,,-2\,,-1\,,-3)$

(g) $\mathbb{W} = \text{span}\{(6\,,2\,,-4\,,5)\,,(-5\,,2\,,-4\,,2)\}$, $\boldsymbol{v} = (3\,,3\,,2\,,7)$

(h) $\mathbb{W} = \text{span}\{(-1\,,3\,,1\,,5)\,,(-3\,,-1\,,-5\,,1)\}$, $\boldsymbol{v} = (-3\,,2\,,3\,,-2)$

(i) $\mathbb{W} = \text{span}\{(-1\,,5\,,3\,,-1)\,,(-1\,,-1\,,1\,,-1)\}$, $\boldsymbol{v} = (0\,,-2\,,-5\,,-5)$

(j) $\mathbb{W} = \text{span}\{(-1\,,1\,,-1\,,1)\,,(-1\,,1\,,1\,,-1)\,,(1\,,1\,,1\,,1)\}$, $\boldsymbol{v} = (0\,,1\,,1\,,2)$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(k) $\mathbb{W} = \operatorname{span}\{(2\,,-2\,,-4\,,-5)\,,(-4\,,4\,,1\,,-4)\,,(2\,,-1\,,4\,,-2)\}$, $\boldsymbol{v} = (2\,,-3\,,1\,,0)$

(l) $\mathbb{W} = \operatorname{span}\{(1\,,4\,,-2\,,-2)\,,(-4\,,1\,,4\,,-4)\,,(-2\,,4\,,2\,,5)\}$, $\boldsymbol{v} = (-2\,,-4\,,3\,,-1)$

Exercise 3.5.22. For each of the following matrices, compute an SVD in MATLAB/Octave to find an orthonormal basis for the column space of the matrix, and then compute the matrix of the orthogonal projection onto the column space.

(a) $A = \begin{bmatrix} 0 & -2 & 4 \\ 4 & -1 & -14 \\ 1 & -1 & -2 \end{bmatrix}$

(b) $B = \begin{bmatrix} -3 & 4 \\ -1 & 5 \\ -3 & -1 \end{bmatrix}$

(c) $C = \begin{bmatrix} -3 & 11 & 6 \\ 12 & 19 & 3 \\ -30 & 5 & 15 \end{bmatrix}$

(d) $D = \begin{bmatrix} -8 & 4 & -2 \\ -24 & 12 & -6 \\ -16 & 8 & -4 \end{bmatrix}$

(e) $E = \begin{bmatrix} -3 & 0 & -5 \\ -1 & -4 & 1 \end{bmatrix}$

(f) $F = \begin{bmatrix} -5 & 5 & 5 \\ 4 & -4 & -4 \\ -1 & 1 & 1 \\ 5 & -5 & -5 \end{bmatrix}$

(g) $G = \begin{bmatrix} 12 & 0 & 10 & 5 \\ -26 & -5 & 5 & 0 \\ -1 & -2 & -16 & 1 \\ -29 & -9 & 29 & 8 \end{bmatrix}$

(h) $H = \begin{bmatrix} -12 & 4 & 8 & 16 & 8 \\ 15 & -5 & -10 & -20 & -10 \end{bmatrix}$

(i) $\mathcal{I} = \begin{bmatrix} 1 & 26 & -13 & 10 \\ -13 & 2 & 9 & 10 \\ -4 & -2 & 4 & 2 \\ -21 & 32 & 1 & 28 \\ -1 & -9 & 5 & -3 \end{bmatrix}$

(j) $J = \begin{bmatrix} 51 & -15 & -19 & -35 & 11 \\ -7 & 2 & 5 & 6 & -5 \\ 14 & -17 & -2 & -8 & -4 \\ 10 & -12 & -2 & -6 & -2 \\ -40 & 30 & 14 & 27 & -4 \end{bmatrix}$

Exercise 3.5.23. Generally, each of the following systems of equations are inconsistent. Use your answers to the previous Exercise 3.5.22 to find the right-hand side vector $\boldsymbol{b}'$ that is the closest vector to the given right-hand side among all the vectors in the column space of the matrix. What is the magnitude of the difference between $\boldsymbol{b}'$ and the given right-hand side? Hence write down a system of *consistent* equations that best approximates the original system.

(a) $\begin{bmatrix} 0 & -2 & 4 \\ 4 & -1 & -14 \\ 1 & -1 & -2 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 6 \\ -19 \\ -3 \end{bmatrix}$

(b) $\begin{bmatrix} 0 & -2 & 4 \\ 4 & -1 & -14 \\ 1 & -1 & -2 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 2 \\ -8 \\ -1 \end{bmatrix}$

v0.2a

(c) $\begin{bmatrix} -3 & 4 \\ -1 & 5 \\ -3 & -1 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 9 \\ 11 \\ -1 \end{bmatrix}$ (d) $\begin{bmatrix} -3 & 4 \\ -1 & 5 \\ -3 & -1 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} -1 \\ 2 \\ -3 \end{bmatrix}$

(e) $\begin{bmatrix} -3 & 11 & 6 \\ 12 & 19 & 3 \\ -30 & 5 & 15 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 3 \\ 5 \\ -3 \end{bmatrix}$ (f) $\begin{bmatrix} -3 & 11 & 6 \\ 12 & 19 & 3 \\ -30 & 5 & 15 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 5 \\ 27 \\ -14 \end{bmatrix}$

(g) $\begin{bmatrix} -3 & 0 & -5 \\ -1 & -4 & 1 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} -9 \\ 10 \end{bmatrix}$ (h) $\begin{bmatrix} -3 & 0 & -5 \\ -1 & -4 & 1 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 6 \\ 3 \end{bmatrix}$

(i) $\begin{bmatrix} -5 & 5 & 5 \\ 4 & -4 & -4 \\ -1 & 1 & 1 \\ 5 & -5 & -5 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 5 \\ -6 \\ 1 \\ -6 \end{bmatrix}$ (j) $\begin{bmatrix} -5 & 5 & 5 \\ 4 & -4 & -4 \\ -1 & 1 & 1 \\ 5 & -5 & -5 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} -6 \\ 6 \\ -2 \\ 5 \end{bmatrix}$

(k) $\begin{bmatrix} 12 & 0 & 10 & 5 \\ -26 & -5 & 5 & 0 \\ -1 & -2 & -16 & 1 \\ -29 & -9 & 29 & 8 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 4 \\ -45 \\ 27 \\ -98 \end{bmatrix}$ (l) $\begin{bmatrix} 12 & 0 & 10 & 5 \\ -26 & -5 & 5 & 0 \\ -1 & -2 & -16 & 1 \\ -29 & -9 & 29 & 8 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} -11 \\ -4 \\ 18 \\ -37 \end{bmatrix}$

Exercise 3.5.24. Theorems 3.5.7 and 3.5.25, examples and Exercise 3.5.23 solve an inconsistent system of equations by some specific 'best approximation' that forms a consistent system of equations to solve. Describe briefly the key idea of this 'best approximation'. Discuss other possibilities for a 'best approximation' that might be developed.

Exercise 3.5.25. For any matrix A, suppose you know an orthonormal basis for the column space of A. Form the matrix W from all the vectors of the orthonormal basis. What is the result of the product $(WW^{\mathrm{T}})A$? Explain why.

Exercise 3.5.26. For each of the following subspaces, draw its orthogonal complement on the plot.

(a)

(b)

v0.2a

(c)

(d)

Exercise 3.5.27. Describe the orthogonal complement of each of the sets given below, if the set has one.

(a) $\mathbb{A} = \operatorname{span}\{(-1, 2)\}$

(b) $\mathbb{B} = \operatorname{span}\{(5, -1)\}$

(c) $\mathbb{C} = \operatorname{span}\{(1, 9, -9)\}$

(d) $\mathbb{D}$ is the plane $-4x_1 + 4x_2 + 5x_3 = 0$

(e) $\mathbb{E}$ is the plane $5x + 2y + 3z = 3$

(f) $\mathbb{F} = \operatorname{span}\{(-5, 5, -3), (-2, 1, 1)\}$

(g) $\mathbb{G} = \operatorname{span}\{(-2, 2, 8), (5, 3, 5)\}$

(h) $\mathbb{H} = \operatorname{span}\{(6, 5, 1, -3)\}$

Exercise 3.5.28. Compute, using MATLAB/Octave when necessary, an orthonormal basis for the orthogonal complement, if it exists, to each of the following sets. Use that the an orthogonal complement is the nullspace of the transpose of a matrix of column vectors (Theorem 3.5.39).

(a) The $\mathbb{R}^3$ vectors in the plane $-6x + 2y - 3z = 0$

(b) The $\mathbb{R}^3$ vectors in the plane $x + 4y + 8z = 0$

(c) The $\mathbb{R}^3$ vectors in the plane $3x + 3y + 2z = 9$

(d) The span of vectors $(-3\,,11\,,-25)$, $(24\,,32\,,-40)$, $(-8\,,-8\,,8)$.

(e) The span of vectors $(3\,,-2\,,1)$, $(-3\,,2\,,-1)$, $(-9\,,6\,,-3)$, $(-6\,,4\,,-2)$

(f) The span of vectors $(26,-2,-4,20)$, $(23,-3,2,6)$, $(2,-2,8,-16)$, $(21\,,-5\,,12\,,-16)$

(g) The span of vectors $(7\,,-5\,,1\,,-6\,,-4)$, $(6\,,-4\,,-2\,,-8\,,-4)$, $(-5\,,5\,,-15\,,-10\,,0)$, $(8\,,-6\,,4\,,-4\,,-4)$

(h) The column space of matrix $\begin{bmatrix} 2 & -1 & 2 & 6 \\ -9 & 11 & -12 & -22 \\ -7 & -6 & -15 & -46 \\ 7 & -23 & 2 & -14 \\ 0 & -2 & 2 & 0 \end{bmatrix}$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(i) The intersection in $\mathbb{R}^4$ of the two hyper-planes $4x_1 + x_2 - 2x_3 + 5x_4 = 0$ and $-4x_1 - x_2 - 7x_3 + 2x_4 = 0$.

(j) The intersection in $\mathbb{R}^4$ of the two hyper-planes $-3x_1 + x_2 + 4x_3 - 7x_4 = 0$ and $-6x_2 - x_3 - 2x_4 = 0$.

Exercise 3.5.29. For the subspace $\mathbb{X} = \text{span}\{\boldsymbol{x}\}$ and the vector $\boldsymbol{v}$, draw the decomposition of $\boldsymbol{v}$ into the sum of vectors in $\mathbb{X}$ and $\mathbb{X}^\perp$.

Exercise 3.5.30. For each of the following vectors, find the perpendicular component to the subspace $\mathbb{W} = \text{span}\{(4, -4, 7)\}$. Verify that the perpendicular component lies in the plane $4x - 4y + 7z = 0$.

(a) $(4\,,2\,,4)$

(b) $(0\,,1\,,-2)$

(c) $(0\,,-2\,,-2)$

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

(e) $(5\,,1\,,5)$

(f) $(p\,,q\,,r)$

v0.2a

Exercise 3.5.31. For each of the following vectors, find the perpendicular component to the subspace $\mathbb{W} = \text{span}\{(1,5,5,7),,(-5,1,-7,5)\}$.

(a) $(1\,,2\,,-1\,,-1)$ (b) $(-2\,,4\,,5\,,0)$

(c) $(2\,,-6\,,1\,,-3)$ (d) $(p\,,q\,,r\,,s)$

Exercise 3.5.32. Let $\mathbb{W}$ be a subspace of $\mathbb{R}^n$ and let $\boldsymbol{v}$ be any vector in $\mathbb{R}^n$. Prove that $\text{perp}_{\mathbb{W}}(\boldsymbol{v}) = (I_n - WW^{\mathrm{T}})\boldsymbol{v}$ where the columns of the matrix W are an orthonormal basis for $\mathbb{W}$.

Exercise 3.5.33. For each of the following vectors in $\mathbb{R}^2$, write the vector as the orthogonal decomposition with respect to the subspace $\mathbb{W} = \text{span}\{(3,4)\}$.

(a) $(-2\,,4)$ (b) $(-3\,,3)$

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

Exercise 3.5.34. For each of the following vectors in $\mathbb{R}^3$, write the vector as the orthogonal decomposition with respect to the subspace $\mathbb{W} = \text{span}\{(3,-6,2)\}$.

(a) $(-5\,,4\,,-5)$ (b) $(0\,,5\,,-1)$

(c) $(1\,,-1\,,-2)$ (d) $(-3\,,1\,,-1)$

Exercise 3.5.35. For each of the following vectors in $\mathbb{R}^4$, write the vector as the orthogonal decomposition with respect to the subspace $\mathbb{W} = \text{span}\{(3,-1,9,3),(-9,3,3,1)\}$.

(a) $(5\,,-5\,,1\,,-3)$ (b) $(-4\,,-2\,,5\,,5)$

(c) $(2\,,-1\,,-4\,,-3)$ (d) $(5\,,4\,,0\,,3)$

Exercise 3.5.36. The vector $(-3,4)$ has an orthogonal decomposition $(1,2)+(-4,2)$. Draw in $\mathbb{R}^2$ the possibilities for the subspace $\mathbb{W}$ and its orthogonal complement.

Exercise 3.5.37. The vector $(2,0,-3)$ in $\mathbb{R}^3$ has an orthogonal decomposition $(2,0,0)+(0,0,-3)$. Describe the possibilities for the subspace $\mathbb{W}$ and its orthogonal complement.

v0.2a

Exercise 3.5.38. The vector $(0, -2, 5, 0)$ in $\mathbb{R}^4$ has an orthogonal decomposition $(0, -2, 0, 0) + (0, 0, 5, 0)$. Describe the possibilities for the subspace $\mathbb{W}$ and its orthogonal complement.

v0.2a

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

3.6 Introducing linear transformations

Section Contents

3.6.1 Matrices correspond to linear transformations 370

3.6.2 The pseudo-inverse of a matrix 374

3.6.3 Function composition connects to matrix inverse 383

3.6.4 Exercises 390

This optional section unifies the transformation examples seen so far, and forms a foundation for more advanced algebra.

Recall the function notation such as $f(x) = x^2$ means that for each $x \in \mathbb{R}$, the function $f(x)$ gives a result in $\mathbb{R}$, namely the value x^2, as plotted in the margin. We often write $f : \mathbb{R} \to \mathbb{R}$ to denote this functionality: that is, $f : \mathbb{R} \to \mathbb{R}$ means function f transforms any given real number into another real number by some rule.

There is analogous functionality in multiple dimensions with vectors: given any vector

- multiplication by a diagonal matrix stretches and shrinks the vector (section 3.2.2);
- multiplication by an orthogonal matrix rotate 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 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

$$f(\boldsymbol{x}) = \begin{bmatrix} 1 & -\frac{1}{3} \\ \frac{1}{2} & -1 \\ -1 & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 - x_2/3 \\ x_1/2 - x_2 \\ -x_1 - x_2/2 \end{bmatrix}.$$

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. The marginal plot 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 all* $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$*, and*

(b) $T(c\boldsymbol{v}) = cT(\boldsymbol{v})$ *for all* $\boldsymbol{v} \in \mathbb{R}^n$ *and all scalars* 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.

Solution: To test Property 3.6.1a, consider $f(x+y) = (x+y)^2 = x^2+2xy+y^2 = f(x)+2xy+f(y) \neq f(x)+f(y)$ for all x and y (it is equal if either are zero, but the test requires equality to hold for all x and y). Alternatively one could test Property 3.6.1b and consider $f(cx) = (cx)^2 = c^2x^2 = c^2f(x) \neq cf(x)$ for all c. Either of these prove that f is not a linear transformation.

(b) Is the function $T(x) = |x|$, $T : \mathbb{R} \to \mathbb{R}$, a linear transformation?

Solution: To prove not it is sufficient to find just one instance when Definition 3.6.1 fails. Let $u = -1$ and $v = 2$, then $T(u+v) = |-1+2| = |1| = 1$ whereas $T(u)+T(v) = |-1|+|2| = 1+2 = 3 \neq T(u+v)$ so the function T fails the additivity and so is not a linear transformation.

(c) Is the function $g : \mathbb{R} \to \mathbb{R}$ such that $g(x) = -x/2$ a linear transformation?

Solution: Because the graph of g is a straight line (as in the marginal picture) we suspect it is a linear transformation. Thus check the properties in full generality:

3.6.1a : for all $u, v \in \mathbb{R}$, $g(u+v) = -(u+v)/2 = -u/2 - v/2 = (-u/2) + (-v/2) = g(u) + g(v)$;

3.6.1b : for all $u, c \in \mathbb{R}$, $g(cu) = -(cu)/2 = c(-u/2) = cg(u)$.

Hence g is a linear transformation.

(d) Show that the function $h(y) = 2y - 3$, $h : \mathbb{R} \to \mathbb{R}$, is not a linear transformation.

Solution: Because the graph of $h(y)$ is a straight line we suspect it may be a linear transformation (as shown in the margin). To prove not it is enough to find one instance when Definition 3.6.1 fails. Let $u = 0$ and $c = 2$, then $h(cu) = h(2 \cdot 0) = h(0) = -3$ whereas $ch(u) = 2h(0) = 2 \cdot (-3) = -6 \neq h(cu)$ so the function g fails the multiplication rule and hence is not a linear transformation. (This function fails because linear transformations have to pass through the origin.)

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(e) Is the function $S : \mathbb{Z} \to \mathbb{Z}$ given by $S(n) = -2n$ a linear transformation? Here $\mathbb{Z}$ denotes the set of integers $\dots, -2, -1, 0, 1, 2, \dots$.

Solution: No, because the function S is here only defined for integers $\mathbb{Z}$ (as plotted in the margin) whereas Definition 3.6.1 requires the function to be defined for all reals. [22]

■

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?

Solution: Yes, because it satisfies the two properties:

3.6.1a : for all $\boldsymbol{u} = (x, y, z)$ and $\boldsymbol{v} = (x', y', z')$ in $\mathbb{R}^3$ consider $T(\boldsymbol{u} + \boldsymbol{v}) = T(x + x', y + y', z + z') = (y + y', z + z', x + x') = (y, z, x) + (y', z', x') = T(x, y, z) + T(x', y', z') = T(\boldsymbol{u}) + T(\boldsymbol{v})$;

3.6.1b : for all $\boldsymbol{u} = (x, y, z)$ and scalars c consider $T(c\boldsymbol{u}) = T(cx, cy, cz) = (cy, cz, cx) = c(y, z, x) = cT(x, y, z) = cT(\boldsymbol{u})$.

Hence, T is 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?

[22] More advanced linear algebra generalises the definition of a linear transformation to non-reals, but not here.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Solution: No. For example, choose $\boldsymbol{u} = \mathbf{0}$ and scalar $c = 2$ then $f(c\boldsymbol{u}) = f(2 \cdot \mathbf{0}) = f(\mathbf{0}) = 1$ whereas $cf(\boldsymbol{u}) = 2f(\mathbf{0}) = 2 \cdot 1 = 2$. Hence f fails the scalar multiplication property 3.6.1b.

(c) Which of the following illustrated transformations of the plane *cannot* be that of a linear transformation? In each illustration of a transformation T, the four corners of the blue unit square ($(0,0)$, $(1,0)$, $(1,1)$ and $(0,1)$), are transformed to the four corners of the red figure ($T(0,0)$, $T(1,0)$, $T(1,1)$ and $T(0,1)$—the 'roof' of the unit square clarifies which side goes where).

Solution: To *test* we check the addition property 3.6.1a. First, with $\boldsymbol{u} = \boldsymbol{v} = \mathbf{0}$ Definition 3.6.1a requires $T(\mathbf{0} + \mathbf{0}) = T(\mathbf{0}) + T(\mathbf{0})$, but the left-hand side is just $T(\mathbf{0})$ which cancels with one on the right-hand side to leave that a linear transformation has to satisfy $T(\mathbf{0}) = \mathbf{0}$: all the shown transformations satisfy $T(\mathbf{0}) = \mathbf{0}$ as the (blue) origin point is transformed to the (red) origin point. Second, with $\boldsymbol{u} = (1,0)$, $\boldsymbol{v} = (1,0)$ and $\boldsymbol{u} + \boldsymbol{v} = (1,1)$ Definition 3.6.1a requires $T(1,1) = T(1,0) + T(0,1)$: let's see which do not pass this test.

i. Here $T(1,1) \approx (-0.3, 1.4)$, whereas $T(1,0) + T(0,1) \approx (-1.4, -0.5) + (1.1, 1.9) = (-0.3, 1.4) \approx T(1,1)$ so this may be a linear transformation.

ii. Here $T(1,1) \approx (1.4, 0.4)$, whereas $T(1,0) + T(0,1) \approx (0.9, 0) + (0.3, 0.7) = (1.2, 0.7) \not\approx T(1,1)$ so this *cannot* be a linear transformation.

iii. Here $T(1,1) \approx (2.0, 1.8)$, whereas $T(1,0) + T(0,1) \approx (1.6, 0.7) + (0.4, 1.1) = (2.0, 1.8) \approx T(1,1)$ so this may be a linear transformation.

iv. Here $T(1,1) \approx (1.4, -2.5)$, whereas $T(1,0) + T(0,1) \approx (0, -2) + (1.1, -0.6) = (1.1, -2.6) \not\approx T(1,1)$ so this *cannot* be a linear transformation.

v. Here $T(1,1) \approx (0.3, -0.9)$, whereas $T(1,0) + T(0,1) \approx (-1, -1.2) + (0.9, 0.1) = (-0.1, -1.1) \not\approx T(1,1)$ so this *cannot* be a linear transformation.

vi. Here $T(1,1) \approx (-2.2, 0.3)$, whereas $T(1,0) + T(0,1) \approx (-0.8, 0.6) + (-1.4, -0.3) = (-2.2, 0.3) \approx T(1,1)$ so this may be a linear transformation.

The ones that pass this test may fail other tests: all we are sure of is that those that fail such tests *cannot* be linear transformations.

(d) The previous Example 3.6.4c illustrated that a linear transformation of the square seems to transform the unit square to a parallelogram: if a function transforms the unit square to something that is not a parallelogram, then the function cannot be a linear transformation. Analogously in higher dimensions: for example, if a function transforms the unit cube to something which is not a parallelepiped, then the function is not a linear transformation. Using this information, which of the following illustrated functions, $f : \mathbb{R}^3 \to \mathbb{R}^3$, *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.

v0.2a

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.

■

Activity 3.6.5. Which of the following functions $f : \mathbb{R}^3 \to \mathbb{R}^2$ is not a linear transformation?

(a) $f(x, y, z) = (0, 0)$

(b) $f(x, y, z) = (y, x + z)$

(c) $f(x, y, z) = (2.7x + 3y, 1 - 2z)$

(d) $f(x, y, z) = (0, 13x + \pi y)$

■

Example 3.6.6. For any given nonzero vector $\boldsymbol{w} \in \mathbb{R}^n$, prove that the projection $P : \mathbb{R}^n \to \mathbb{R}^n$ 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 $Q : \mathbb{R}^n \to \mathbb{R}^n$ by $Q(\boldsymbol{w}) = \operatorname{proj}_{\boldsymbol{w}}(\boldsymbol{u})$ is not a linear transformation (as a function of $\boldsymbol{w}$).

Solution:

- Consider the two properties of Definition 3.6.1 for the function P.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

3.6.1a : for all $\boldsymbol{u},\boldsymbol{v}\in\mathbb{R}^n$, from Defintion 3.5.18 for a projection, $P(\boldsymbol{u}+\boldsymbol{v})=\operatorname{proj}_{\boldsymbol{w}}(\boldsymbol{u}+\boldsymbol{v})=\boldsymbol{w}[\boldsymbol{w}\cdot(\boldsymbol{u}+\boldsymbol{v})]/|\boldsymbol{w}|^2=\boldsymbol{w}[(\boldsymbol{w}\cdot\boldsymbol{u})+(\boldsymbol{w}\cdot\boldsymbol{v})]/|\boldsymbol{w}|^2=\boldsymbol{w}(\boldsymbol{w}\cdot\boldsymbol{u})/|\boldsymbol{w}|^2+\boldsymbol{w}(\boldsymbol{w}\cdot\boldsymbol{v})/|\boldsymbol{w}|^2=\operatorname{proj}_{\boldsymbol{w}}(\boldsymbol{u})+\operatorname{proj}_{\boldsymbol{w}}(\boldsymbol{v})=P(\boldsymbol{u})+P(\boldsymbol{v})$;

3.6.1b : for all $\boldsymbol{u}\in\mathbb{R}^n$ and scalars c, $P(c\boldsymbol{u})=\operatorname{proj}_{\boldsymbol{w}}(c\boldsymbol{u})=\boldsymbol{w}[\boldsymbol{w}\cdot(c\boldsymbol{u})]/|\boldsymbol{w}|^2=\boldsymbol{w}[c(\boldsymbol{w}\cdot\boldsymbol{u})]/|\boldsymbol{w}|^2=c\big[\boldsymbol{w}(\boldsymbol{w}\cdot\boldsymbol{u})/|\boldsymbol{w}|^2\big]=c\operatorname{proj}_{\boldsymbol{w}}(\boldsymbol{u})=cP(\boldsymbol{u})$.

Hence, the projection P is a linear transformation.

- Now consider $Q(\boldsymbol{w})=\operatorname{proj}_{\boldsymbol{w}}(\boldsymbol{u})$. For any $\boldsymbol{u},\boldsymbol{w}\in\mathbb{R}^n$ let's check $Q(2\boldsymbol{w})=\operatorname{proj}_{(2\boldsymbol{w})}(\boldsymbol{u})=(2\boldsymbol{w})[(2\boldsymbol{w})\cdot\boldsymbol{u}]/|2\boldsymbol{w}|^2=4\boldsymbol{w}(\boldsymbol{w}\cdot\boldsymbol{u})/(4|\boldsymbol{u}|^2)=\boldsymbol{w}(\boldsymbol{w}\cdot\boldsymbol{u})/|\boldsymbol{u}|^2=\operatorname{proj}_{\boldsymbol{w}}(\boldsymbol{u})=Q(\boldsymbol{w})\neq 2Q(\boldsymbol{w})$ and so the projection is not a linear transformation when considered as a function of the direction of the transform $\boldsymbol{w}$ for some given $\boldsymbol{u}$.

■

3.6.1 Matrices correspond to linear transformations

One important class of linear transformations are the transformations that can be written as matrix multiplications. The reason for the importance is that Theorem 3.6.10 establishes all linear transformations may be written as matrix multiplications! This in turn justifies why we define matrix multiplication to be as it is (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 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 3.5.28).

■

Theorem 3.6.8. *Let A be a 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 all $\boldsymbol{x}\in\mathbb{R}^n$. Then T_A is a linear transformation.*

Proof. Let vectors $\boldsymbol{u}\,,\boldsymbol{v}\in\mathbb{R}^n$ and scalar $c\in\mathbb{R}$ and consider the two properties of Definition 3.6.1.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

3.6.1a. By the distributivity of matrix-vector multiplication (Theorem 3.1.22), $T_A(\boldsymbol{u}+\boldsymbol{v}) = A(\boldsymbol{u}+\boldsymbol{v}) = A\boldsymbol{u}+A\boldsymbol{v} = T_A(\boldsymbol{u})+T_A(\boldsymbol{v})$.

3.6.1b. By commutativity of scalar multiplication (Theorem 3.1.24), $T_A(c\boldsymbol{u}) = A(c\boldsymbol{u}) = c(A\boldsymbol{u}) = cT_A(\boldsymbol{u})$.

Hence 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 vector $\boldsymbol{b}\neq\boldsymbol{0}$, is not a linear transformation.

Solution: Just consider the addition property 3.6.1a for the zero vectors $\boldsymbol{u}=\boldsymbol{v}=\boldsymbol{0}$: on the one hand, $S(\boldsymbol{u}+\boldsymbol{v}) = S(\boldsymbol{0}+\boldsymbol{0}) = S(\boldsymbol{0}) = A\boldsymbol{0}+\boldsymbol{b} = \boldsymbol{b}$; on the other hand $S(\boldsymbol{u})+S(\boldsymbol{v}) = S(\boldsymbol{0})+S(\boldsymbol{0}) = A\boldsymbol{0}+\boldsymbol{b}+A\boldsymbol{0}+\boldsymbol{b} = 2\boldsymbol{b}$. Hence when the shift $\boldsymbol{b}$ is nonzero, there are vectors for which $S(\boldsymbol{u}+\boldsymbol{v})\neq S(\boldsymbol{u})+S(\boldsymbol{v})$ and so S is not a linear transformation.

■

Now let's establish the important converse to Theorem 3.6.8: every linear transformation can be written as a matrix multiplication.

Theorem 3.6.10. *Let $T:\mathbb{R}^n\to\mathbb{R}^m$ be a 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.*

Proof. Let $\boldsymbol{x}$ be any vector in $\mathbb{R}^n$: then $\boldsymbol{x} = (x_1,x_2,\dots,x_n) = x_1\boldsymbol{e}_1+x_2\boldsymbol{e}_2+\cdots+x_n\boldsymbol{e}_n$ for standard unit vectors $\boldsymbol{e}_1,\boldsymbol{e}_2,\dots,\boldsymbol{e}_n$. Then

$$\begin{aligned}
T(\boldsymbol{x}) &= T(x_1\boldsymbol{e}_1+x_2\boldsymbol{e}_2+\cdots+x_n\boldsymbol{e}_n)\\
&\qquad\text{(using the identity of Exercise 3.6.6)}\\
&= x_1T(\boldsymbol{e}_1)+x_2T(\boldsymbol{e}_2)+\cdots+x_nT(\boldsymbol{e}_n)\\
&= \begin{bmatrix} T(\boldsymbol{e}_1) & T(\boldsymbol{e}_2) & \cdots & T(\boldsymbol{e}_n)\end{bmatrix}\begin{bmatrix} x_1\\ x_2\\ \vdots\\ x_n\end{bmatrix}\\
&= A\boldsymbol{x}
\end{aligned}$$

for matrix A of the theorem. Since $T(\boldsymbol{e}_1),T(\boldsymbol{e}_2),\dots,T(\boldsymbol{e}_n)$ are n (column) vectors in $\mathbb{R}^m$, the matrix A is $m\times n$. □

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Example 3.6.11. (a) Find the standard matrix of the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^4$ where $T(x, y, z) = (y, z, x, 3x - 2y + z)$.

Solution: We need to find the transform of the three standard unit vectors in $\mathbb{R}^3$:

$$\begin{aligned} T(\boldsymbol{e}_1) &= T(1,0,0) = (0,0,1,3); \\ T(\boldsymbol{e}_2) &= T(0,1,0) = (1,0,0,-2); \\ T(\boldsymbol{e}_3) &= T(0,0,1) = (0,1,0,1). \end{aligned}$$

Form the standard matrix with these as its three columns, in order,

$$[T] = \begin{bmatrix} T(\boldsymbol{e}_1) & T(\boldsymbol{e}_2) & T(\boldsymbol{e}_3) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 3 & -2 & 1 \end{bmatrix}.$$

(b) Find the standard matrix of the rotation of the plane by 60° about the origin.

Solution: Denote the rotation of the plane by the function $R : \mathbb{R}^2 \to \mathbb{R}^2$. Since $60° = \frac{\pi}{3}$ then, as illustrated in the margin,

$$\begin{aligned} R(\boldsymbol{e}_1) &= (\cos\tfrac{\pi}{3}, \sin\tfrac{\pi}{3}) = (\tfrac{1}{2}, \tfrac{\sqrt{3}}{2}), \\ R(\boldsymbol{e}_2) &= (-\sin\tfrac{\pi}{3}, \cos\tfrac{\pi}{3}) = (-\tfrac{\sqrt{3}}{2}, \tfrac{1}{2}). \end{aligned}$$

Form the standard matrix with these as its columns, in order,

$$[R] = \begin{bmatrix} R(\boldsymbol{e}_1) & R(\boldsymbol{e}_2) \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}.$$

(c) Find the standard matrix of the rotation about the point $(1, 0)$ of the plane by 45°.

Solution: Since the origin $(0, 0)$ is transformed by the rotation to $(1, -1)$ which is nonzero, this transformation cannot be of the form $A\boldsymbol{x}$, so cannot have a standard matrix, and hence is not a linear transformation.

(d) Estimate the standard matrix for each of the illustrated transformations given they transform the unit square as shown.

v0.2a

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) $\begin{bmatrix} 4.5 & -1.6 \\ 1.9 & -2 \end{bmatrix}$

(b) $\begin{bmatrix} 4.5 & 1.9 \\ -1.6 & -2 \end{bmatrix}$

(c) $\begin{bmatrix} 0 & 1.9 \\ 4.5 & 0 \\ -1.6 & -2 \end{bmatrix}$

(d) $\begin{bmatrix} 0 & 4.5 & -1.6 \\ 1.9 & 0 & -2 \end{bmatrix}$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

■

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.

Solution: Let $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ and c be any scalar. Function H is a linear transformation because

- $H(\boldsymbol{u}+\boldsymbol{v}) = a(\boldsymbol{u}+\boldsymbol{v}) = a\boldsymbol{u} + a\boldsymbol{v} = H(\boldsymbol{u}) + H(\boldsymbol{v})$, and
- $H(c\boldsymbol{u}) = a(c\boldsymbol{u}) = (ac)\boldsymbol{u} = (ca)\boldsymbol{u} = c(a\boldsymbol{u}) = cH(\boldsymbol{u})$.

To find the standard matrix consider

$$\begin{aligned} H(\boldsymbol{e}_1) &= a\boldsymbol{e}_1 = (a,0,0,\dots,0), \\ H(\boldsymbol{e}_2) &= a\boldsymbol{e}_2 = (0,a,0,\dots,0), \\ &\vdots \\ H(\boldsymbol{e}_n) &= a\boldsymbol{e}_n = (0,0,\dots,0,a). \end{aligned}$$

Hence the standard matrix $[H] = \operatorname{diag}(a,a,\dots,a) = aI_n$.

$$\begin{aligned} [H] &= \begin{bmatrix} H(\boldsymbol{e}_1) & H(\boldsymbol{e}_2) & \cdots & H(\boldsymbol{e}_n) \end{bmatrix} \\ &= \begin{bmatrix} a & 0 & \cdots & 0 \\ 0 & a & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & a \end{bmatrix} = aI_n\,. \end{aligned}$$

■

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 solving inconsistent linear equations, $A\boldsymbol{x} = \boldsymbol{b}$ for some given A, Procedure 3.5.4 finds a solution $\boldsymbol{x}$ 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 the resulting solution given by the procedure is a linear transformation of $\boldsymbol{b}$, and hence there must be a matrix, say A^+, corresponding to the procedure. This matrix gives the resulting solution $\boldsymbol{x} = A^+\boldsymbol{b}$. We call the matrix A^+ the pseudo-inverse of A.

v0.2a

Example 3.6.14. Find the pseudo-inverse of the matrix $A = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$.

Solution: Apply Procedure 3.5.4 to solve $Ax = \boldsymbol{b}$ for any right-hand side $\boldsymbol{b}$.

(a) This matrix has an SVD

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

(b) Hence $\boldsymbol{z} = U^{\mathrm{T}}\boldsymbol{b} = \begin{bmatrix} \frac{3}{5}b_1 + \frac{4}{5}b_2 \\ -\frac{4}{5}b_1 + \frac{3}{5}b_2 \end{bmatrix}$.

(c) Then the diagonal system $Sy = \boldsymbol{z}$ is $\begin{bmatrix} 5 \\ 0 \end{bmatrix} y = \begin{bmatrix} \frac{3}{5}b_1 + \frac{4}{5}b_2 \\ -\frac{4}{5}b_1 + \frac{3}{5}b_2 \end{bmatrix}$: approximately solve by neglecting the second component in the equations, and so from the first component just set $y = \frac{3}{25}b_1 + \frac{4}{25}b_2$.

(d) Then the procedure's solution is $x = Vy = 1(\frac{3}{25}b_1 + \frac{4}{25}b_2) = \frac{3}{25}b_1 + \frac{4}{25}b_2$.

That is, for all right-hand side vectors $\boldsymbol{b}$, this least square solution

$$x = A^{+}\boldsymbol{b} \quad \text{for pseudo-inverse } A^{+} = \begin{bmatrix} \frac{3}{25} & \frac{4}{25} \end{bmatrix}.$$

■

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) $\begin{bmatrix} 0 & 0.2 \\ 0 & 0 \end{bmatrix}$

(c) $\begin{bmatrix} 0 & 0 \\ 0.2 & 0 \end{bmatrix}$ (d) $\begin{bmatrix} 0 & 0 \\ 0 & 0.2 \end{bmatrix}$

■

A pseudo-inverse A^{+} of a non-invertible matrix A is only an 'inverse' because the pseudo-inverse builds in extra information that we *sometimes choose* to be desirable. This extra information rationalises all the contradictions encountered in trying to construct an inverse of a non-invertible matrix. Namely, for some applications we *choose* to desire that the pseudo-inverse solves the *nearest* consistent system to the one specified, and we *choose* the smallest of all possibilities then allowed. However, although there are many

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

situations where these choices are useful, beware that there are also many situations where such choices are not appropriate. That is, although sometimes the pseudo-inverse is useful, beware that often the pseudo-inverse is not appropriate.

Theorem 3.6.16 (pseudo-inverse)**.** *For a given $m \times n$ matrix A and in the context of wanting to solve $A\boldsymbol{x} = \boldsymbol{b}$, Procedure 3.5.4 forms a linear transformation $T : \mathbb{R}^m \to \mathbb{R}^n$ to find the smallest solution $\boldsymbol{x}$ (Theorem 3.5.12) to the closest consistent system $A\boldsymbol{x} = \tilde{\boldsymbol{b}}$ (Theorem 3.5.7). This linear transformation has an $n \times m$ standard matrix A^+ called the **pseudo-inverse**, or **Moore–Penrose inverse**, of matrix A.*

Proof. First, for each right-hand side vector $\boldsymbol{b}$ in $\mathbb{R}^m$, the procedure gives a result $\boldsymbol{x}$ in $\mathbb{R}^n$ and so is some function $T : \mathbb{R}^m \to \mathbb{R}^n$. We proceed to confirm that Procedure 3.5.4 satisfies the two defining properties of a linear transformation (Definition 3.6.1). For each of any two right-hand side vectors $\boldsymbol{b}', \boldsymbol{b}'' \in \mathbb{R}^m$ let Procedure 3.5.4 generate similarly dashed intermediaries $(\boldsymbol{z}', \boldsymbol{z}'', \boldsymbol{y}', \boldsymbol{y}'')$ through to two corresponding least square solutions $\boldsymbol{x}', \boldsymbol{x}'' \in \mathbb{R}^n$. That is, $\boldsymbol{x}' = T(\boldsymbol{b}')$ and $\boldsymbol{x}'' = T(\boldsymbol{b}'')$. Throughout let the matrix A have SVD $A = USV^{\mathrm{T}}$ and set $r = \operatorname{rank} A$.

3.6.1a : To check $T(\boldsymbol{b}' + \boldsymbol{b}'') = T(\boldsymbol{b}') + T(\boldsymbol{b}'')$, apply the procedure when the right-hand side is $\boldsymbol{b} = \boldsymbol{b}' + \boldsymbol{b}''$:

2. solve $U\boldsymbol{z} = \boldsymbol{b}$ by $\boldsymbol{z} = U^{\mathrm{T}}\boldsymbol{b} = U^{\mathrm{T}}(\boldsymbol{b}'+\boldsymbol{b}'') = U^{\mathrm{T}}\boldsymbol{b}'+U^{\mathrm{T}}\boldsymbol{b}'' = \boldsymbol{z}' + \boldsymbol{z}''$;
3.
   - for $i = 1, \dots, r$ set $y_i = z_i/\sigma_i = (z_i' + z_i'')/\sigma_i = z_i'/\sigma_i + z_i''/\sigma_i = y_i' + y_i''$, and
   - for $i = r+1, \dots, n$ set the free variables $y_i = 0 = 0 + 0 = y_i' + y_i''$ to obtain the smallest solution (Theorem 3.5.12),;

   and hence $\boldsymbol{y} = \boldsymbol{y}' + \boldsymbol{y}''$;
4. solve $\boldsymbol{x} = V^{\mathrm{T}}\boldsymbol{y}$ with $\boldsymbol{x} = V\boldsymbol{y} = V(\boldsymbol{y}' + \boldsymbol{y}'') = V\boldsymbol{y}' + V\boldsymbol{y}'' = \boldsymbol{x}' + \boldsymbol{x}''$.

Since result $\boldsymbol{x} = \boldsymbol{x}' + \boldsymbol{x}''$, thus $T(\boldsymbol{b}' + \boldsymbol{b}'') = T(\boldsymbol{b}') + T(\boldsymbol{b}'')$.

3.6.1b : To check $T(c\boldsymbol{b}') = cT(\boldsymbol{b}')$ for any scalar c, apply the procedure when the right-hand side is $\boldsymbol{b} = c\boldsymbol{b}'$:

2. solve $U\boldsymbol{z} = \boldsymbol{b}$ by $\boldsymbol{z} = U^{\mathrm{T}}\boldsymbol{b} = U^{\mathrm{T}}(c\boldsymbol{b}') = cU^{\mathrm{T}}\boldsymbol{b}' = c\boldsymbol{z}'$;
3.
   - for $i = 1,\dots,r$ set $y_i = z_i/\sigma_i = (cz_i')/\sigma_i = c(z_i'/\sigma_i) = cy_i'$, and
   - for $i = r+1,\dots,n$ set the free variables $y_i = 0 = c0 = cy_i'$ to obtain the smallest solution (Theorem 3.5.12),

   and hence $\boldsymbol{y} = c\boldsymbol{y}'$;

4. solve $\boldsymbol{x} = V^{\mathrm{T}}\boldsymbol{y}$ with $\boldsymbol{x} = V\boldsymbol{y} = V(c\boldsymbol{y}') = cV\boldsymbol{y}' = c\boldsymbol{x}'$.

Since the result $\boldsymbol{x} = c\boldsymbol{x}'$, consequently $T(c\boldsymbol{b}') = cT(\boldsymbol{b}')$.

Since Procedure 3.5.4, denoted by T, is a linear transformation $T : \mathbb{R}^m \to \mathbb{R}^n$, Theorem 3.6.10 assures us T has a corresponding $n \times m$ standard matrix which we denote A^+ and call the pseudo-inverse. □

Example 3.6.17. Find the pseudo-inverse of the matrix $A = \begin{bmatrix} 5 & 12 \end{bmatrix}$.

Solution: Apply Procedure 3.5.4 to solve $A\boldsymbol{x} = b$ for any right-hand side b.

(a) This matrix has an SVD

$$A = \begin{bmatrix} 5 & 12 \end{bmatrix} = \begin{bmatrix} 1 \end{bmatrix} \begin{bmatrix} 13 & 0 \end{bmatrix} \begin{bmatrix} \frac{5}{13} & -\frac{12}{13} \\ \frac{12}{13} & \frac{5}{13} \end{bmatrix}^{\mathrm{T}} = USV^{\mathrm{T}}.$$

(b) Hence $z = U^{\mathrm{T}}b = 1b = b$.

(c) The diagonal system $S\boldsymbol{y} = z$ becomes $\begin{bmatrix} 13 & 0 \end{bmatrix}\boldsymbol{y} = b$ with general solution $\boldsymbol{y} = (b/13, y_2)$. The smallest of these solutions is $\boldsymbol{y} = (b/13, 0)$.

(d) Then the procedure's result is

$$\boldsymbol{x} = V\boldsymbol{y} = \begin{bmatrix} \frac{5}{13} & -\frac{12}{13} \\ \frac{12}{13} & \frac{5}{13} \end{bmatrix} \begin{bmatrix} b/13 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{5}{169}b \\ -\frac{12}{169}b \end{bmatrix}.$$

That is, for all right-hand sides b, this procedure's result is

$$\boldsymbol{x} = A^+b \quad \text{for pseudo-inverse } A^+ = \begin{bmatrix} \frac{5}{169} \\ -\frac{12}{169} \end{bmatrix}.$$

■

Activity 3.6.18. Following the steps of Procedure 3.5.4, Find the pseudo-inverse of the matrix $\begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$ given that the matrix has the SVD

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

The pseudo-inverse is which of these?

(a) $\begin{bmatrix} 0.1 & 0.1 \\ 0.2 & 0.2 \end{bmatrix}$

(b) $\begin{bmatrix} 0.1 & -0.1 \\ -0.2 & 0.2 \end{bmatrix}$

(c) $\begin{bmatrix} 0.1 & 0.2 \\ 0.1 & 0.2 \end{bmatrix}$

(d) $\begin{bmatrix} 0.1 & -0.2 \\ -0.1 & 0.2 \end{bmatrix}$

■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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

$$A = \begin{bmatrix}1\\1\\1\\1\end{bmatrix} = \begin{bmatrix}\frac12&\frac12&\frac12&\frac12\\\frac12&\frac12&-\frac12&-\frac12\\\frac12&-\frac12&-\frac12&\frac12\\\frac12&-\frac12&\frac12&-\frac12\end{bmatrix}\begin{bmatrix}2\\0\\0\\0\end{bmatrix}\begin{bmatrix}1\end{bmatrix}^{\mathrm{T}} = USV^{\mathrm{T}}.$$

(b) Solve $U\boldsymbol{z} = \boldsymbol{b}$ by computing

$$\begin{aligned}\boldsymbol{z} = U^{\mathrm{T}}\boldsymbol{b} &= \begin{bmatrix}\frac12&\frac12&\frac12&\frac12\\\frac12&\frac12&-\frac12&-\frac12\\\frac12&-\frac12&-\frac12&\frac12\\\frac12&-\frac12&\frac12&-\frac12\end{bmatrix}\boldsymbol{b}\\ &= \begin{bmatrix}\frac12b_1+\frac12b_2+\frac12b_3+\frac12b_4\\\frac12b_1+\frac12b_2-\frac12b_3-\frac12b_4\\\frac12b_1-\frac12b_2-\frac12b_3+\frac12b_4\\\frac12b_1-\frac12b_2+\frac12b_3-\frac12b_4\end{bmatrix}.\end{aligned}$$

(c) Now try to solve $Sy = \boldsymbol{z}$, that is,

$$\begin{bmatrix}2\\0\\0\\0\end{bmatrix} y = \begin{bmatrix}\frac12b_1+\frac12b_2+\frac12b_3+\frac12b_4\\\frac12b_1+\frac12b_2-\frac12b_3-\frac12b_4\\\frac12b_1-\frac12b_2-\frac12b_3+\frac12b_4\\\frac12b_1-\frac12b_2+\frac12b_3-\frac12b_4\end{bmatrix}.$$

Instead of seeking an *exact* solution, we *have to* adjust the last three components to zero. Hence we find a solution to a slightly different problem by solving

$$\begin{bmatrix}2\\0\\0\\0\end{bmatrix} y = \begin{bmatrix}\frac12b_1+\frac12b_2+\frac12b_3+\frac12b_4\\0\\0\\0\end{bmatrix},$$

v0.2a

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 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 3.3.12 and 3.5.3.

Solution: To find the pseudo-inverse, follow Procedure 3.5.4 with a general right-hand side vector $\boldsymbol{b}$.

(a) Compute an SVD $A = USV^{\mathrm{T}}$ in MATLAB/Octave with `[U,S,V]=svd(A)`:

```
U =
    0.4082   -0.7071    0.5774
   -0.4082   -0.7071   -0.5774
   -0.8165   -0.0000    0.5774
S =
    1.7321         0         0
         0    1.7321         0
         0         0    0.0000
V =
    0.0000   -0.8165    0.5774
   -0.7071    0.4082    0.5774
    0.7071    0.4082    0.5774
```

Upon recognising various square-roots, these matrices are

$$U = \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \\ -\frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \end{bmatrix} ,$$

$$S = \begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & \sqrt{3} & 0 \\ 0 & 0 & 0 \end{bmatrix} ,$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$V = \begin{bmatrix} 0 & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix}.$$

The system of equations for the ratings becomes

$$A\boldsymbol{x} = U \underbrace{S \overbrace{V^{\mathrm{T}}\boldsymbol{x}}^{=\boldsymbol{y}}}_{=\boldsymbol{z}} = \boldsymbol{b}.$$

(b) As U is orthogonal, $U\boldsymbol{z} = \boldsymbol{b}$ has unique solution

$$\boldsymbol{z} = U^{\mathrm{T}}\boldsymbol{b} = \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix} \boldsymbol{b}.$$

(c) Now solve $S\boldsymbol{y} = \boldsymbol{z}$. But S has a troublesome zero on the diagonal. So interpret the equation $S\boldsymbol{y} = \boldsymbol{z}$ in detail as

$$\begin{bmatrix} \sqrt{3} & 0 & 0 \\ 0 & \sqrt{3} & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{y} = \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix} \boldsymbol{b}:$$

i. the first line requires $y_1 = \frac{1}{\sqrt{3}} \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} \end{bmatrix} \boldsymbol{b} = \frac{1}{3} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & -\sqrt{2} \end{bmatrix}$;

ii. the second line requires $y_2 = \frac{1}{\sqrt{3}} \begin{bmatrix} -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{bmatrix} \boldsymbol{b} = \begin{bmatrix} -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & 0 \end{bmatrix} \boldsymbol{b}$;

iii. the third line requires $0y_3 = \begin{bmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix} \boldsymbol{b}$ which generally cannot be satisfied, so we set $y_3 = 0$ to get the *smallest solution of the system after projecting* $\boldsymbol{b}$ *onto the column space of* A.

(d) Finally, as V is orthogonal, $V^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{y}$ has the solution $\boldsymbol{x} = V\boldsymbol{y}$ (unique for each valid $\boldsymbol{y}$):

$$\begin{aligned} \boldsymbol{x} &= V\boldsymbol{y} \\ &= \begin{bmatrix} 0 & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} \frac{1}{3\sqrt{2}} & -\frac{1}{3\sqrt{2}} & -\frac{\sqrt{2}}{3} \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{b} \\ &= \frac{1}{3} \begin{bmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & -1 \end{bmatrix} \boldsymbol{b} \end{aligned}$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Hence the pseudo-inverse of A is

$$A^+ = \frac{1}{3}\begin{bmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & -1 \end{bmatrix}.$$

- In Example 3.3.12, Anne beat Bob 3-2 games; Anne beat Chris 3-1; Bob beat Chris 3-2 so the right-hand side vector is $\boldsymbol{b} = (1, 2, 1)$. The procedure's ratings are then, as before,

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

- In Example 3.5.3, Bob instead beat Chris 3-1 so the right-hand side vector is $\boldsymbol{b} = (1, 2, 2)$. The procedure's ratings are then, as before,

$$\boldsymbol{x} = A^+\boldsymbol{b} = \frac{1}{3}\begin{bmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & -1 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{1}{3} \\ -\frac{4}{3} \end{bmatrix}.$$

■

In some common special cases there are alternative formulas for the pseudo-inverse: specifically, the cases are when the rank of the matrix is the same as the number of rows and/or columns.

Example 3.6.21. For the matrix $A = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$, confirm that $(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ is the pseudo-inverse that was found in Example 3.6.14.

Solution: Here $A^{\mathrm{T}}A = 3^2 + 4^2 = 25$, so its inverse is $(A^{\mathrm{T}}A)^{-1} = 1/25$. Then, $(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}} = \frac{1}{25}\begin{bmatrix} 3 & 4 \end{bmatrix} = \begin{bmatrix} \frac{3}{25} & \frac{4}{25} \end{bmatrix}$ as found in Example 3.6.14 for the pseudo-inverse.

■

Theorem 3.6.22. *In the case of an $m \times n$ matrix A with* $\operatorname{rank} A = n$ *(so $m \geq n$), the pseudo-inverse $A^+ = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$.*

Proof. Apply Procedure 3.5.4 to the system $A\boldsymbol{x} = \boldsymbol{b}$ for an arbitrary $\boldsymbol{b} \in \mathbb{R}^m$.

1. Let $m \times n$ matrix A have SVD $A = USV^{\mathrm{T}}$. Since $\operatorname{rank} A = n$ there are n nonzero singular values on the diagonal of $m \times n$ matrix S, and so $m \geq n$.
2. Solve $U\boldsymbol{z} = \boldsymbol{b}$ with $\boldsymbol{z} = U^{\mathrm{T}}\boldsymbol{b} \in \mathbb{R}^m$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

3. Approximately solve $S\boldsymbol{y} = \boldsymbol{z}$ by setting $y_i = z_i/\sigma_i$ for $i = 1, \dots, n$ and neglecting the last $(m-n)$ equations. This is identical to setting $\boldsymbol{y} = S^+\boldsymbol{z} = S^+U^{\mathrm{T}}\boldsymbol{b}$ after defining the $n \times m$ matrix

$$S^+ := \begin{bmatrix} 1/\sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1/\sigma_2 & & 0 & 0 & \cdots & 0 \\ \vdots & & \ddots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1/\sigma_n & 0 & \cdots & 0 \end{bmatrix}.$$

4. Solve $V^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{y}$ with $\boldsymbol{x} = V\boldsymbol{y} = VS^+U^{\mathrm{T}}\boldsymbol{b}$.

Hence the pseudo-inverse is $A^+ = VS^+U^{\mathrm{T}}$.

Let's find $(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ is the same expression. First, since $A^{\mathrm{T}} = (USV^{\mathrm{T}})^{\mathrm{T}} = VS^{\mathrm{T}}U^{\mathrm{T}}$,

$$A^{\mathrm{T}}A = VS^{\mathrm{T}}U^{\mathrm{T}}USV^{\mathrm{T}} = VS^{\mathrm{T}}SV^{\mathrm{T}} = V(S^{\mathrm{T}}S)V^{\mathrm{T}}$$

where $n \times n$ matrix $(S^{\mathrm{T}}S) = \operatorname{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2)$. Since $\operatorname{rank} A = n$, the singular values $\sigma_1, \sigma_2, \dots, \sigma_n > 0$ and so matrix $(S^{\mathrm{T}}S)$ is invertible as it is square and diagonal with all nonzero elements in the diagonal, and the inverse is $(S^{\mathrm{T}}S)^{-1} = \operatorname{diag}(1/\sigma_1^2, 1/\sigma_2^2, \dots, 1/\sigma_n^2)$ (Theorem 3.2.27). Second, the $n \times n$ matrix $(A^{\mathrm{T}}A)$ is invertible as it has SVD $V(S^{\mathrm{T}}S)V^{\mathrm{T}}$ with n nonzero singular values (Theorem 3.3.26d), and so

$$\begin{aligned} (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}} &= (V(S^{\mathrm{T}}S)V^{\mathrm{T}})^{-1}VS^{\mathrm{T}}U^{\mathrm{T}} \\ &= (V^{\mathrm{T}})^{-1}(S^{\mathrm{T}}S)^{-1}V^{-1}VS^{\mathrm{T}}U^{\mathrm{T}} \\ &= V(S^{\mathrm{T}}S)^{-1}V^{\mathrm{T}}VS^{\mathrm{T}}U^{\mathrm{T}} \\ &= V(S^{\mathrm{T}}S)^{-1}S^{\mathrm{T}}U^{\mathrm{T}} \\ &= VS^+U^{\mathrm{T}}, \end{aligned}$$

where the last equality follows because

$$\begin{aligned} &(S^{\mathrm{T}}S)^{-1}S^{\mathrm{T}} \\ &= \operatorname{diag}(1/\sigma_1^2, 1/\sigma_2^2, \dots, 1/\sigma_n^2)\operatorname{diag}_{n\times m}(\sigma_1, \sigma_2, \dots, \sigma_n) \\ &= \operatorname{diag}_{n\times m}(1/\sigma_1, 1/\sigma_2, \dots, 1/\sigma_n) = S^+. \end{aligned}$$

Hence the pseudo-inverse $A^+ = VS^+U^{\mathrm{T}} = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$. □

Theorem 3.6.23. *If A is an invertible matrix, then the pseudo-inverse $A^+ = A^{-1}$, the inverse.*

Proof. If A is invertible it must be square, say $n \times n$, and of rank $A = n$ (Theorem 3.3.26e). Further, A^{T} is invertible with inverse $(A^{-1})^{\mathrm{T}}$ (Theorem 3.2.13d). Then the expression from Theorem 3.6.22 for the pseudo-inverse gives

$$A^+ = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}} = A^{-1}(A^{\mathrm{T}})^{-1}A^{\mathrm{T}} = A^{-1}I_n = A^{-1}.$$

□

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Computer considerations Except for easy cases, we (almost) never explicitly compute the pseudo-inverse of a matrix. In practical computation, forming $A^{\mathrm{T}}A$ and then manipulating it is both expensive and error enhancing: 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. Computationally there are (almost) always better ways to proceed, such as Procedure 3.5.4. Like an inverse, a pseudo-inverse is a theoretical device, rarely a practical tool.

A main point of this subsection is to illustrate how a complicated procedure is conceptually expressible as a linear transformation, and so has associated matrix properties such as being equivalent to multiplication by some matrix—here the pseudo-inverse.

3.6.3 Function composition connects to matrix inverse

To achieve a complex goal we typically decompose the task of attaining the goal into a set of smaller tasks and achieve those tasks one after another. The analogy in linear algebra is that we often apply linear transformations one after another to build up or solve a complex problem. This section introduces how applying a sequence of linear transformations is equivalent to one grand overall linear transformation.

Example 3.6.24 (simple rotation)**.** Recall Example 3.6.11b on rotation by 60° (illustrated in the margin) with its standard matrix

$$[R] = \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}.$$

Consider two successive rotations by 60°: show that the standard matrix of the resultant rotation by 120° is the same as the matrix product $[R][R]$.

Solution: On the one hand, rotation by $120° = 2\pi/3$, call it S, transforms the unit vectors as (illustrated in the margin)

$$S(\boldsymbol{i}) = (\cos\tfrac{2\pi}{3}, \sin\tfrac{2\pi}{3}) = (-\tfrac{1}{2}, \tfrac{\sqrt{3}}{2}),$$
$$S(\boldsymbol{j}) = (-\sin\tfrac{2\pi}{3}, \cos\tfrac{2\pi}{3}) = (-\tfrac{\sqrt{3}}{2}, -\tfrac{1}{2}).$$

Form the standard matrix with these as its columns, in order,

$$[S] = \begin{bmatrix} S(\boldsymbol{i}) & S(\boldsymbol{j}) \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix}.$$

On the other hand, the matrix multiplication

$$\begin{aligned} [R][R] &= \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{4} - \frac{3}{4} & -\frac{\sqrt{3}}{4} - \frac{\sqrt{3}}{4} \\ \frac{\sqrt{3}}{4} + \frac{\sqrt{3}}{4} & -\frac{3}{4} + \frac{1}{4} \end{bmatrix} \end{aligned}$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$= \begin{bmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix} = [S].$$

That is, multiplying the two matrices is equivalent to performing the two rotations in succession: the next theorem confirms this is generally true.

■

Theorem 3.6.25. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ and $S : \mathbb{R}^m \to \mathbb{R}^p$ be linear transformations. Recalling 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 $[S \circ T] = [S][T]$.*

Proof. For given transformations S and T, let matrix $A = [S]$ $(p \times m)$ and $B = [T]$ $(m \times n)$. Let $\boldsymbol{u}$ be any vector in $\mathbb{R}^n$, then $(S \circ T)(\boldsymbol{u}) = S\big(T(\boldsymbol{u})\big) = S(B\boldsymbol{u}) = A(B\boldsymbol{u}) = (AB)\boldsymbol{u}$ (using associativity, Theorem 3.1.24c). Hence the effect of $S \circ T$ is identical to multiplication by the $p \times n$ matrix (AB). It is thus a matrix transformation, which is consequently linear (Theorem 3.6.8), and its standard matrix $[S \circ T] = AB = [S][T]$. □

Example 3.6.26. Consider the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^2$ defined by $T(x_1, x_2, x_3) := (3x_1 + x_2, -x_2 - 7x_3)$, and 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 $T \circ S$.

Solution: From the given formulas for the two given linear transformations we write down the standard matrices

$$[T] = \begin{bmatrix} 3 & 1 & 0 \\ 0 & -1 & -7 \end{bmatrix} \quad \text{and} \quad [S] = \begin{bmatrix} -1 & 0 \\ -3 & 2 \\ 2 & -1 \\ 0 & 2 \end{bmatrix}.$$

First, Theorem 3.6.25 assures us the standard matrix of the composition

$$\begin{aligned} [S \circ T] &= [S][T] \\ &= \begin{bmatrix} -1 & 0 \\ -3 & 2 \\ 2 & -1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 3 & 1 & 0 \\ 0 & -1 & -7 \end{bmatrix} \\ &= \begin{bmatrix} -3 & -1 & 0 \\ -9 & -5 & -14 \\ 6 & 3 & 7 \\ 0 & -2 & -14 \end{bmatrix}. \end{aligned}$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

However, second, the standard matrix of $T \circ S$ does not exist because it would require the multiplication of a 2×3 matrix by a 4×2 matrix, and such a multiplication is not defined. The failure is rooted earlier in the question because $S : \mathbb{R}^2 \to \mathbb{R}^4$ and $T : \mathbb{R}^3 \to \mathbb{R}^2$ so a result of S, which is in $\mathbb{R}^4$, cannot be used as an argument to T, which must be in $\mathbb{R}^3$: the lack of a defined multiplication is a direct reflection of this incompatibility in '$T \circ S$' which means $T \circ S$ cannot exist.

■

Example 3.6.27. Find the standard matrix of the transformation of the plane that first rotates by 45° about the origin, and then second reflects in the vertical axis.

Solution: Two possible solutions are the following.

- Let R denote the rotation about the origin of the plane by 45°, $R : \mathbb{R}^2 \to \mathbb{R}^2$ (illustrated in the margin). Its standard matrix is
$$[R] = \begin{bmatrix} R(\boldsymbol{i}) & R(\boldsymbol{j}) \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$
Let F denote the reflection in the vertical axis of the plane, $F : \mathbb{R}^2 \to \mathbb{R}^2$ (illustrated in the margin). Its standard matrix is
$$[F] = \begin{bmatrix} F(\boldsymbol{i}) & F(\boldsymbol{j}) \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}.$$
Then the standard matrix of the composition
$$[F \circ R] = [F][R] = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

- Alternatively, just consider the action of the two component transformations on the standard unit vectors.
  - $(F \circ R)(\boldsymbol{i}) = F(R(\boldsymbol{i}))$ which first rotates $\boldsymbol{i} = (1, 0)$ to point to the top-right, then reflects in the vertical axis to point to the top-left and thus $(F \circ R)(\boldsymbol{i}) = (-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$.
  - $(F \circ R)(\boldsymbol{j}) = F(R(\boldsymbol{j}))$ which first rotates $\boldsymbol{j} = (0, 1)$ to point to the top-left, then reflects in the vertical axis to point to the top-right and thus $(F \circ R)(\boldsymbol{j}) = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$.

Then the standard matrix of the composition (as illustrated in the margin)
$$[F \circ R] = \begin{bmatrix} (F \circ R)(\boldsymbol{i}) & (F \circ R)(\boldsymbol{j}) \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

v0.2a

As an extension, check that although $R \circ F$ is defined, it is different to $F \circ R$: the difference corresponds to the non-commutativity of matrix multiplication (section 3.1.3). ■

Activity 3.6.28. Given the stretching transformation S with standard matrix $[S] = \operatorname{diag}(2, 1/2)$, and the anti-clockwise 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) $\begin{bmatrix} 0 & 2 \\ -\frac{1}{2} & 0 \end{bmatrix}$ (c) $\begin{bmatrix} 0 & -\frac{1}{2} \\ 2 & 0 \end{bmatrix}$ (d) $\begin{bmatrix} 0 & -2 \\ \frac{1}{2} & 0 \end{bmatrix}$

■

Invert transformations Having introduced and characterised the composition of linear transformations, we now discuss when two transformations composed together end up 'cancelling' each other out.

Example 3.6.29 (inverse transformations)**.** (a) Let S be rotation of the plane by $60°$, and T be rotation of the plane by $-60°$. Then $S \circ T$ is first rotation by $-60°$ by T, and second rotation by $60°$ by S: the result is no change. Because $S \circ T$ is effectively the identity transformation, we call the rotations S and T the inverse transformation of each other.

(b) Let R be reflection of the plane in the line at $30°$ to the horizontal (illustrated in the margin). 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 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****.*

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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 below),

and let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be rotation about the vertical axis by 240° (below).

Argue that $S \circ T = T \circ S = I$ the identity and so S and T are inverse transformations of each other.

Solution: A basic argument is that rotation by 120° together with a rotation of 240° about the same axis, in either order, is the same as a rotation by 360° about the axis. But a 360° rotation leaves everything unchanged and so must be the identity.

Alternatively one could dress up the argument with some algebra as in the following. First consider $T \circ S$:

- the vertical unit vector $\boldsymbol{e}_3$ is unchanged by both S and T so $(T \circ S)(\boldsymbol{e}_3) = T(S(\boldsymbol{e}_3)) = \boldsymbol{e}_3$;
- the unit vector $\boldsymbol{e}_1$ is rotated 120° by S, and then by 240° by T which is a total of 360°, that is, it is rotated back to itself so $(T \circ S)(\boldsymbol{e}_1) = T(S(\boldsymbol{e}_1)) = \boldsymbol{e}_1$; and
- the unit vector $\boldsymbol{e}_2$ is rotated 120° by S, and then by 240° by T which is a total of 360°, that is, it is rotated back to itself so $(T \circ S)(\boldsymbol{e}_2) = T(S(\boldsymbol{e}_2)) = \boldsymbol{e}_2$.

Form these results into a matrix to deduce the standard matrix

$$[T \circ S] = \begin{bmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \boldsymbol{e}_3 \end{bmatrix} = I_3$$

which is the standard matrix of the identity transformation. Hence $T \circ S = I$ the identity.

Second, an exactly corresponding argument gives $S \circ T = I$. By Definition 3.6.30, S and T are inverse transformations of each other.

■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Example 3.6.32. In some violent weather a storm passes and the strong winds lean a house sideways as in the shear transformation illustrated below.

Estimate the standard matrix of the shear transformation shown. To restore the house back upright, we need to shear it an equal amount in the opposite direction: hence write down the standard matrix of the inverse shear. Confirm that the product of the two standard matrices is the standard matrix of the identity.

Solution: As shown, the unit vectors $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ are unchanged by the storm S. However, the vertical unit vector $\boldsymbol{e}_3$ is sheared by the storm to $S(\boldsymbol{e}_3) = (-1, -0.5, 1)$. Hence the standard matrix of the storm S is

$$[S] = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix} .$$

To restore, R, the house upright we need to shear in the opposite direction, so the restoration shear has $R(\boldsymbol{e}_3) = (1, \frac{1}{2}, 1)$; that is, the standard matrix of the inverse is

$$[R] = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix} .$$

Multiplying these matrices together gives

$$\begin{aligned} [R \circ S] &= [R][S] \\ &= \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1+0+0 & 0+0+0 & -1+0+1 \\ 0+0+0 & 0+1+0 & 0-\frac{1}{2}+\frac{1}{2} \\ 0+0+0 & 0+0+0 & 0+0+1 \end{bmatrix} \\ &= I_3 . \end{aligned}$$

This is the standard matrix of the identity transformation.

■

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 $[T^{-1}] = [T]^{-1}$.*

Proof. As T is invertible, let the symbol T^{-1} denote its inverse. Since both are linear transformations, they both have standard matrices, $[T]$ and $[T^{-1}]$. Then $[T \circ T^{-1}] = [T][T^{-1}]$; but also $[T \circ T^{-1}] = [I] = I_n$; so $[T][T^{-1}] = I_n$. Similarly, $[T^{-1}][T] = [T^{-1} \circ T] = [I] = I_n$. Consequently the matrices $[T]$ and $[T^{-1}]$ are the inverses of each other. □

Example 3.6.34. Estimate the standard matrix of the linear transformation T illustrated in the margin. Then use Theorem 3.2.7 to determine the standard matrix of its inverse transformation T^{-1}. Hence sketch how the inverse transforms the unit square and write a sentence or two about how the sketch confirms it is a reasonable inverse.

Solution: The illustrated transformation shows $T(\boldsymbol{i}) \approx (1.7, -1)$ and $T(\boldsymbol{j}) \approx (1.9, 1.7)$ hence its standard matrix is

$$[T] \approx \begin{bmatrix} 1.7 & 1.9 \\ -1 & 1.7 \end{bmatrix}.$$

Using Theorem 3.2.7 the inverse of this matrix is, since its determinant $= 1.7 \cdot 1.7 - (-1) \cdot 1.9 = 4.79$,

$$[T^{-1}] = [T]^{-1} = \frac{1}{4.79} \begin{bmatrix} 1.7 & -1.9 \\ 1 & 1.7 \end{bmatrix} \approx \begin{bmatrix} 0.35 & -0.40 \\ 0.21 & 0.35 \end{bmatrix}$$

This matrix determines that the inverse transforms the corners of the unit square as $T^{-1}(1, 0) = (0.35, 0.21)$, $T^{-1}(0, 1) = (-0.40, 0.35)$ and $T^{-1}(1, 1) = (-0.05, 0.56)$. Hence the unit square is transformed as shown in the margin. The original transformation, roughly, rotated the unit square clockwise and stretched it: the sketch shows the inverse roughly rotates the unit square anti-clockwise and shrinks it. Thus the inverse does appear to undo the action of the original transformation.

■

Example 3.6.35. Determine if the orthogonal projection of the plane onto the line at 30° to the horizontal (illustrated in the margin) is an invertible transformation; if it is find its inverse.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Solution: Recall that Theorem 3.5.28 gives the matrix of an orthogonal projection as WW^{T} where columns of W are an orthonormal basis for the projected space. Here the projected space is the line at 30° to the horizontal (illustrated in the margin) which has orthonormal basis of the one vector $\boldsymbol{w} = (\cos 30°, \sin 30°) = (\frac{\sqrt{3}}{2}, \frac{1}{2})$. Hence the standard matrix of the projection is

$$\boldsymbol{w}\boldsymbol{w}^{\mathrm{T}} = \begin{bmatrix} \frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{3}{4} & \frac{\sqrt{3}}{4} \\ \frac{\sqrt{3}}{4} & \frac{1}{4} \end{bmatrix}.$$

From Theorem 3.2.7 this matrix is invertible only if the determinant (det $= ad - bc$) is nonzero, but here

$$\det(\boldsymbol{w}\boldsymbol{w}^{\mathrm{T}}) = \frac{3}{4} \cdot \frac{1}{4} - \frac{\sqrt{3}}{4} \cdot \frac{\sqrt{3}}{4} = \frac{3}{16} - \frac{3}{16} = 0\,.$$

Since its standard matrix is not invertible, the given orthogonal projection is also not invertible (as the illustration shows, the projection 'squashes' the plane onto the line, which cannot be uniquely undone, and hence is not invertible).

■

3.6.4 Exercises

Exercise 3.6.1. Which of the following illustrated transformations of the plane *cannot* be that of a linear transformation? In each illustration of a transformation T, the four corners of the blue unit square ($(0,0)$, $(1,0)$, $(1,1)$ and $(0,1)$), are transformed to the four corners of the red figure ($T(0,0)$, $T(1,0)$, $T(1,1)$ and $T(0,1)$—the 'roof' of the unit square clarifies which side goes where).

Exercise 3.6.2. Consider the transformations of Example 3.6.1: for those transformations that *may* be linear transformations, assume they are and so estimate roughly the standard matrix of each such linear transformation.

Exercise 3.6.3. Consider the following illustrated transformations of $\mathbb{R}^3$. Which *cannot* be that of a linear transformation?

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Exercise 3.6.4. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Prove from the Definition 3.6.1 that $T(\mathbf{0}) = \mathbf{0}$ and $T(\boldsymbol{u} - \boldsymbol{v}) = T(\boldsymbol{u}) - T(\boldsymbol{v})$ for all $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$.

Exercise 3.6.5 (equivalent definition)**.** Consider a function $T : \mathbb{R}^n \to \mathbb{R}^m$. Prove that T is a linear transformation (Definition 3.6.1) if and only if $T(c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2) = c_1T(\boldsymbol{v}_1) + c_2T(\boldsymbol{v}_2)$ for all $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathbb{R}^n$ and all scalars c_1, c_2.

Exercise 3.6.6. Given $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, use induction to prove that for all k

$$T(c_1\boldsymbol{u}_1 + c_2\boldsymbol{u}_2 + \cdots + c_k\boldsymbol{u}_k) = c_1T(\boldsymbol{u}_1) + c_2T(\boldsymbol{u}_2) + \cdots + c_kT(\boldsymbol{u}_k)$$

for all scalars $c_1, c_2, \ldots, c_k$ and all vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k$ in $\mathbb{R}^n$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Exercise 3.6.7. Consider each of the following vector transformations: if it is a linear transformation, then write down its standard matrix.

(a) $A(x,y,z) = (-3x+2y, -z)$

(b) $B(x,y) = (0, 3x+7y, -2y, -3y)$

(c) $C(x_1,x_2,\dots,x_5) = (2x_1+x_2-2x_3, 7x_1+7x_4)$

(d) $D(x,y) = (2x+3, -5y+3, 2x-4y+3, 0, -6x)$

(e) $E(p,q,r,s) = (-3p-4r, -s, 0, p+r+6s, 5p+6q-s)$

(f) $F(x,y) = (5x+4y, x^2, 2y, -4x, 0)$

(g) $G(x_1,x_2,x_3,x_4) = (-x_1+4x_2+e^{x_3}, 8x_4)$

(h) $H(u_1,u_2,\dots,u_5) = (7u_1-9u_3-u_5, 3u_1-3u_4)$

Exercise 3.6.8. Use Procedure 3.5.4 to derive that the pseudo-inverse of the general 2×1 matrix $A = \begin{bmatrix} a \\ b \end{bmatrix}$ is the 1×2 matrix $A^+ = \begin{bmatrix} \frac{a}{a^2+b^2} & \frac{b}{a^2+b^2} \end{bmatrix}$. Further, what is the pseudo-inverse of the general 1×2 matrix $\begin{bmatrix} a & b \end{bmatrix}$?

Exercise 3.6.9. Consider the general $m\times n$ diagonal matrix of rank r,

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

equivalently $S = \operatorname{diag}_{m\times n}(\sigma_1,\sigma_2,\dots,\sigma_r,0,\dots,0)$. Derive that, based upon Procedure 3.5.4, the pseudo-inverse of S is the $n\times m$ diagonal matrix of rank r,

$$S^+ = \begin{bmatrix} 1/\sigma_1 & \cdots & 0 & \\ \vdots & \ddots & \vdots & O_{r\times(m-r)} \\ 0 & \cdots & 1/\sigma_r & \\ & & & \\ O_{(n-r)\times r} & & & O_{(n-r)\times(m-r)} \end{bmatrix},$$

equivalently $S^+ = \operatorname{diag}_{n\times m}(1/\sigma_1, 1/\sigma_2, \dots, 1/\sigma_r, 0, \dots, 0)$.

Exercise 3.6.10. Given any $m\times n$ matrix A with SVD $A = USV^{\mathrm{T}}$, use the result of Exercise 3.6.9 to establish that the pseudo-inverse $A^+ = VS^+U^{\mathrm{T}}$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Exercise 3.6.11. Use MATLAB/Octave and the identity of Exercise 3.6.10 to compute the pseudo-inverse of each of the following matrices.

(a) $A = \begin{bmatrix} 0.2 & 0.3 \\ 0.2 & 1.5 \\ 0.1 & -0.3 \end{bmatrix}$

(b) $B = \begin{bmatrix} 2.5 & 0.6 & 0.3 \\ 0.5 & 0.4 & 0.2 \end{bmatrix}$

(c) $C = \begin{bmatrix} 0.1 & -0.5 \\ 0.6 & -3.0 \\ 0.4 & -2.0 \end{bmatrix}$

(d) $D = \begin{bmatrix} 0.3 & -0.3 & -1.2 \\ 0.1 & 0.3 & 1.4 \\ -3.1 & -0.5 & -3.8 \\ 1.5 & -0.3 & -0.6 \end{bmatrix}$

(e) $E = \begin{bmatrix} 4.1 & 1.8 & -0.4 & 0.0 & -0.1 & -1.4 \\ -3.3 & -3.9 & 0.6 & -2.2 & 0.5 & 0.1 \\ -0.9 & -1.9 & 0.6 & -2.2 & 0.1 & 0.5 \\ -4.3 & -3.6 & 0.8 & -2.0 & 0.3 & 1.2 \end{bmatrix}$

(f) $F = \begin{bmatrix} -0.6 & -1.3 & -1.2 & -1.9 & 1.6 & 1.6 \\ -0.7 & 0.6 & -0.2 & 0.9 & -0.6 & -0.7 \\ 0.0 & 0.2 & 0.7 & 1.1 & -0.4 & -0.6 \\ 0.8 & 0.1 & -1.6 & -0.9 & -0.5 & -0.8 \\ -0.5 & 0.9 & 1.3 & 1.7 & -0.3 & -0.6 \end{bmatrix}$

(g) $G = \begin{bmatrix} 0.0 & -1.6 & 1.2 & -0.4 & -0.4 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 2.0 & -1.5 & 0.5 & 0.5 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & -1.6 & 1.2 & -0.4 & -0.4 \end{bmatrix}$

(h) $H = \begin{bmatrix} -1.9 & -1.8 & -1.9 & -1.8 & -1.0 & -1.9 \\ 0.4 & 2.5 & -0.5 & 1.9 & 0.8 & 2.4 \\ -0.3 & -0.3 & 1.3 & -0.5 & -0.1 & -0.4 \\ 0.7 & 0.5 & 1.1 & 0.5 & 0.4 & 0.6 \end{bmatrix}$

Exercise 3.6.12. Prove that in the case of an $m \times n$ matrix A with $\operatorname{rank} A = m$ (so $m \le n$), the pseudo-inverse is the $n \times m$ matrix $A^+ = A^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}$.

Exercise 3.6.13. Use Theorem 3.6.22 and the identity in Exercise 3.6.12 to prove that $(A^+)^+ = A$ in the case when $m \times n$ matrix A has $\operatorname{rank} A = n$. (Be careful as many plausible looking steps are incorrect.)

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Exercise 3.6.14. Confirm that the composition of the two linear transformations in $\mathbb{R}^2$ has a standard matrix that is the same as multiplying the two standard matrices of the specified linear transformations.

(a) Rotation by $30°$ followed by rotation by $60°$.

(b) Rotation by $120°$ followed by rotation by $-60°$ (clockwise $60°$.

(c) Reflection in the x-axis followed by reflection in the line $y = x$.

(d) Reflection in the line $y = x$ followed by reflection in the x-axis.

(e) Reflection in the line $y = x$ followed by rotation by $90°$.

(f) Reflection in the line $y = \sqrt{3}x$ followed by rotation by $-30°$ (clockwise $30°$).

Exercise 3.6.15. For each of the following pairs of linear transformations S and T, if possible determine the standard matrices of the compositions $S \circ T$ and $T \circ S$.

(a) $S(x) = (-5x, 2x, -x, -x)$ and $T(y_1, y_2, y_3, y_4) = -4y_1 - 3y_2 - 4y_3 + 5y_4$

(b) $S(x_1, x_2, x_3, x_4) = (-2x_2 - 3x_3 + 6x_4, -3x_1 + 2x_2 - 4x_3 + 3x_4)$ and $T(y) = (-4y, 0, 0, -y)$

(c) $S(x, y) = (-5x - 3y, -5x + 5y)$ and $T(z_1, z_2, z_3, z_4) = (3z_1 + 3z_2 - 2z_3 - 2z_4, 4z_1 + 7z_2 + 4z_3 + 3z_4)$

(d) $S(x, y, z) = 5x - y + 4z$ and $T(p) = (6p, -3p, 3p, p)$

(e) $S(u_1, u_2, u_3, u_4) = (-u_1 + 2u_2 + 2u_3 - 3u_4, -3u_1 + 3u_2 + 4u_3 + 3u_4, -u_2 - 2u_3)$ and $T(x, y, z) = (-2y - 4z, -4x + 2y + 2z, x + 3y + 4z, 2x - 2z)$

(f) $S(p, q, r, s) = (5p - r - 2s, q - r + 2s, 7p + q + s)$ and $T(x, y) = (y, 2x + 3y, -2x + 2y, -5x - 4y)$

Exercise 3.6.16. For each of the illustrated transformations, estimate the standard matrix of the linear transformation. Then use Theorem 3.2.7 to determine the standard matrix of its inverse transformation. Hence sketch how the inverse transforms the unit square and write a sentence or two about how the sketch confirms it is a reasonable inverse.

(a)

(b)

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(c)

(d)

(e)

v0.2a

Answers to selected exercises

3.1.1b : Only B and D.

3.1.2a : A, 4×2; B, 1×3; C, 3×2; D, 3×4; E, 2×2; F, 2×1.

3.1.2c : AE, AF, BC, BD, CE, CF, DA, E^2, EF, FB.

3.1.4 : $\boldsymbol{b}_1 = (7.6,-1.1,2.6,-1.5,-0.2)$, $\boldsymbol{b}_2 = (-1.1,-9.3,6.9,-7.5,5.5)$, $\boldsymbol{b}_3 = (-0.7,0.1,1.2,3.7,-0.9)$, $\boldsymbol{b}_4 = (-4.5,8.2,-3.6,2.6,2.4)$; $b_{13} = -0.7$, $b_{31} = 2.6$, $b_{42} = -7.5$.

3.1.6a : $A+B = \begin{bmatrix} 3 & 2 & -1 \\ 0 & -5 & -9 \\ -8 & 6 & -1 \end{bmatrix}$, $A-B = \begin{bmatrix} 1 & 0 & -1 \\ -8 & 7 & 3 \\ 4 & -2 & -1 \end{bmatrix}$

3.1.6c : $P+Q = \begin{bmatrix} -3 & 2 & 0 \\ 9 & -7 & 0 \\ 0 & 0 & -2 \end{bmatrix}$, $P-Q = \begin{bmatrix} -1 & 8 & 2 \\ -3 & 1 & 4 \\ -6 & 6 & -4 \end{bmatrix}$

3.1.7a : $-2A = \begin{bmatrix} 6 & 4 \\ -8 & 4 \\ -4 & 8 \end{bmatrix}$, $2A = \begin{bmatrix} -6 & -4 \\ 8 & -4 \\ 4 & -8 \end{bmatrix}$, $3A = \begin{bmatrix} -9 & -6 \\ 12 & -6 \\ 6 & -12 \end{bmatrix}$.

3.1.7c : $-4U = \begin{bmatrix} 15.6 & 1.2 & 11.6 \\ -12.4 & 15.6 & 4. \\ -12.4 & 26. & -3.6 \end{bmatrix}$, $2U = \begin{bmatrix} 7.8 & 0.6 & 5.8 \\ -6.2 & 7.8 & 2. \\ -6.2 & 13. & -1.8 \end{bmatrix}$, $4U = \begin{bmatrix} -15.6 & -1.2 & -11.6 \\ 12.4 & -15.6 & -4. \\ 12.4 & -26. & 3.6 \end{bmatrix}$.

3.1.10a : $A\boldsymbol{p} = \begin{bmatrix} -9 \\ -13 \end{bmatrix}$, $A\boldsymbol{q} = \begin{bmatrix} 4 \\ -16 \end{bmatrix}$, $A\boldsymbol{r} = \begin{bmatrix} -15 \\ 11 \end{bmatrix}$.

3.1.10c : $C\boldsymbol{u} = \begin{bmatrix} 6 \\ 3 \end{bmatrix}$, $C\boldsymbol{v} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$, $C\boldsymbol{w} = \begin{bmatrix} 24 \\ -5 \end{bmatrix}$.

3.1.11a : $A\boldsymbol{u} = (7\,,-5)$, $A\boldsymbol{v} = (-6\,,3)$, $A\boldsymbol{w} = (9\,,-6)$.

3.1.11c : $C\boldsymbol{x}_1 = (-4.41\,,9.66)$, $C\boldsymbol{x}_2 = (1.42\,,-1.56)$, $C\boldsymbol{x}_3 = (-0.47\,,-0.38)$.

3.1.12a : $P\boldsymbol{u} = (1,1.4)$, $P\boldsymbol{v} = (-3.6,1.7)$, $P\boldsymbol{w} = (0.1,-2.3)$. Reflection in the horizontal axis.

3.1.12c : $R\boldsymbol{x}_1 = (-4.4\,,-0.8)$, $R\boldsymbol{x}_2 = (5\,,0)$, $R\boldsymbol{x}_3 = (-0.2\,,3.6)$. Rotation (by 36.87°).

3.1.13a : $\begin{bmatrix} -9 & 4 \\ -13 & -16 \end{bmatrix}$, $\begin{bmatrix} -9 & -15 \\ -13 & 11 \end{bmatrix}$, $\begin{bmatrix} 4 & -15 \\ -16 & 11 \end{bmatrix}$, $\begin{bmatrix} -9 & 4 & -15 \\ -13 & -16 & 11 \end{bmatrix}$.

3.1.13c : $\begin{bmatrix} 6 & 3 \\ 3 & 4 \end{bmatrix}$, $\begin{bmatrix} 24 & 3 \\ -5 & 4 \end{bmatrix}$, $\begin{bmatrix} 24 & 6 \\ -5 & 3 \end{bmatrix}$, $\begin{bmatrix} 24 & 3 & 6 \\ -5 & 4 & 3 \end{bmatrix}$.

3.1.17 : Let x_j be the number of females of age $(j-1)$ years. $L =$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$\begin{bmatrix} 0 & 0 & 2 & 2 \\ \frac{2}{3} & 0 & 0 & 0 \\ 0 & \frac{2}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \end{bmatrix}$. After one year $\boldsymbol{x}' = (72\,,12\,,6\,,9)$, two years $\boldsymbol{x}'' = (30,48,8,3)$, three years $\boldsymbol{x}''' = (22,20,32,4)$. Increasing.

3.1.18b : $\begin{bmatrix} 3 & -5 \\ -4 & 2 \\ -2 & -3 \\ 2 & 3 \end{bmatrix}$

3.1.18d : $\begin{bmatrix} 3 \\ 1 \\ -2 \\ -3 \end{bmatrix}$

3.1.18f : $\begin{bmatrix} -4 & -5.1 & 0.3 \\ -5.1 & -7.4 & -3 \\ 0.3 & -3. & 2.6 \end{bmatrix}$, symmetric

3.1.18h : $\begin{bmatrix} 1.7 & 0.7 & 0.6 \\ -0.2 & -0.3 & 3 \\ -0.4 & -0.4 & -2.2 \end{bmatrix}$

3.2.1b : Inverse.

3.2.1d : Inverse.

3.2.1f : Inverse.

3.2.1h : Inverse.

3.2.1j : Not inverse.

3.2.2a : $\begin{bmatrix} -2/3 & 1/3 \\ -1/6 & 1/3 \end{bmatrix}$

3.2.2c : $\begin{bmatrix} 1/8 & 1/4 \\ -5/16 & -1/8 \end{bmatrix}$

3.2.2e : $\begin{bmatrix} 0 & 1/3 \\ -1/4 & 1/6 \end{bmatrix}$

3.2.2g : $\begin{bmatrix} 3.3333 & 0 \\ -1.5789 & 0.5263 \end{bmatrix}$

3.2.3a : $(x\,,y) = (-1\,,-1/4)$

3.2.3c : $x = 0\,,\ m = 1$

3.2.3e : $(q\,,r\,,p) = (-1/2\,,-5/6\,,2/3)$

3.2.3g : $(p\,,q\,,r\,,s) = (33\,,14\,,35\,,68)$

3.2.4a : Invertible

3.2.4c : Not invertible.

v0.2a

3.2.4e : Invertible.

3.2.7a : $\begin{bmatrix} 0 & 1/16 \\ -1/16 & -1/16 \end{bmatrix}$

3.2.7c : $\begin{bmatrix} -2/9 & 1/6 \\ -1/6 & 0 \end{bmatrix}$

3.2.7e : $\begin{bmatrix} -5/4 & -1/2 & -3/4 \\ -5/8 & 1/4 & -1/8 \\ 11/4 & 1/2 & 5/4 \end{bmatrix}$

3.2.7g : $\begin{bmatrix} 0 & -1/2 & 1 & 0 \\ -6 & -75/2 & 42 & 7 \\ -13/2 & -81/2 & 91/2 & 15/2 \\ -44 & -275 & 308 & 51 \end{bmatrix}$

3.2.8a : $\operatorname{diag}(9\,,-5\,,4)$

3.2.8c : $\operatorname{diag}(-5\,,1\,,9\,,1\,,0)$

3.2.8e : $\operatorname{diag}_{5\times 3}(0\,,-1\,,-5)$

3.2.8g : $\operatorname{diag}(2\,,1\,,0)$

3.2.8i : Diagonal only when $c=0$.

3.2.9a : $\boldsymbol{x}=(-3/2\,,-1\,,2\,,-2\,,-4/3)$

3.2.9c : $(x\,,y)=(5/4\,,t)$ for all t

3.2.9e : No solution.

3.2.9g : $(p\,,q\,,r\,,s)=(-2\,,2\,,-8/3\,,t)$ for all t

3.2.10a : $\operatorname{diag}(2\,,1)$

3.2.10c : $\operatorname{diag}(3\,,2)$

3.2.10e : Not diagonal.

3.2.10g : $\operatorname{diag}(0.8\,,-0.5)$

3.2.10i : $\operatorname{diag}(0.4\,,1.5)$

3.2.11a : $\operatorname{diag}(2\,,1.5\,,1)$

3.2.11c : $\operatorname{diag}(0.9\,,-0.5\,,0.6)$

3.2.11e : Not diagonal.

3.2.13a : Not orthogonal.

3.2.13c : Not orthogonal.

3.2.13e : Orthogonal.

3.2.13g : Orthonormal.

3.2.14a : Not orthogonal.

3.2.14c : Not orthogonal.

v0.2a

3.2.14e : Orthonormal.

3.2.15a : Orthogonal set, divide each by seven.

3.2.15c : Orthogonal set, divide each by eleven.

3.2.15e : Orthogonal set, divide each by two.

3.2.15g : Not orthogonal set.

3.2.16a : Orthogonal matrix.

3.2.16c : Not orthogonal matrix.

3.2.16e : Not orthogonal matrix as not square.

3.2.16g : Orthogonal matrix.

3.2.16i : Not orthogonal matrix as not square.

3.2.16k : Orthogonal matrix.

3.2.17b : $\theta = 90^\circ$

3.2.17d : $\theta = 93.18^\circ$

3.2.17f : $\theta = 72.02^\circ$

3.2.17h : $\theta = 115.49^\circ$

3.2.18b : $(x\,,y) = (-3.6\,,-0.7)$

3.2.18d : $(x\,,y) = (1.6\,,3.8)$

3.2.18f : $(u\,,v\,,w) = (5/3\,,2/3\,,-4/3)$

3.2.18h : $\boldsymbol{x} = (-1.4\,,1.6\,,1.2\,,0.2)$

3.2.18j : $\boldsymbol{z} = (-0.25\,,0.15\,,0.07\,,0.51)$

3.2.22b : Yes—the square is rotated.

3.2.22d : Yes—the square is rotated and reflected.

3.2.22f : Yes—the square is rotated.

3.2.22h : No—the square is squashed.

3.2.23b : No—the cube is deformed.

3.2.23d : No—the cube is deformed.

3.2.23f : Yes—the cube appears rotated and reflected.

3.3.2b : $\boldsymbol{x} = (0\,,\frac{3}{2})$

3.3.2d : $\boldsymbol{x} = (-2\,,2)$

3.3.2f : $\boldsymbol{x} = (\frac{5}{14}\,,\frac{15}{14}\,,\frac{15}{28}) + (-\frac{6}{7}\,,\frac{3}{7}\,,-\frac{2}{7})s + (\frac{3}{7}\,,\frac{2}{7}\,,-\frac{6}{7})t$

3.3.2h : No solution.

3.3.2j : No solution.

3.3.2l : $\boldsymbol{x} = (2\,,-\frac{7}{4}\,,-\frac{9}{2})$

v0.2a

3.3.3b : No solution.

3.3.3d : $\boldsymbol{x} = (27\,,-12\,,-24\,,9) + (\frac{1}{2}\,,\frac{1}{2}\,,\frac{1}{2}\,,\frac{1}{2})t$

3.3.3f : $\boldsymbol{x} = (-\frac{33}{2}\,,\frac{25}{2}\,,\frac{17}{2}\,,\frac{9}{2})$

3.3.4b : $\boldsymbol{x} = (-1\,,-1\,,3\,,-3)$

3.3.4d : $\boldsymbol{x} = (-4\,,-9\,,0\,,7)$

3.3.4f : $\boldsymbol{x} = (0.6\,,8.2\,,9.8\,,-0.6) + (-0.1\,,0.3\,,-0.3\,,-0.9)t$

3.3.4h : $\boldsymbol{x} = (0\,,-0.3\,,-3.1\,,4.9)$

3.3.4j : $\boldsymbol{x} = (0.18\,,3.35\,,-4.86\,,0.33\,,0.35)$
$+ (0.91\,,-0.34\,,-0.21\,,-0.09\,,-0.07)t$ (2 d.p.)

3.3.6 :

1. cond $= 5/3$, rank $= 2$;
2. cond $= 1$, rank $= 2$;
3. cond $= \infty$, rank $= 1$;
4. cond $= 2$, rank $= 2$;
5. cond $= 2$, rank $= 2$;
6. cond $= \infty$, rank $= 1$;
7. cond $= 2$, rank $= 2$;
8. cond $= 2$, rank $= 2$;
9. cond $= 2$, rank $= 3$;
10. cond $= \infty$, rank $= 2$;
11. cond $= \infty$, rank $= 1$;
12. cond $= 9$, rank $= 3$;

3.3.10 : The theorem applies to the square matrix systems of Exercise 3.3.2 (a)–(d), (i)–(l), and of Exercise 3.3.3 (e) and (f). The cases with no zero singular value, full rank, have a unique solution. The cases with a zero singular value, rank less than n, either have no solution or an infinite number.

3.3.15b : $\boldsymbol{v}_1 \approx (0.1\,,1.0)$, $\sigma_1 \approx 1.7$, $\boldsymbol{v}_2 \approx (1.0\,,-0.1)$, $\sigma_1 \approx 0.3$.

3.3.15d : $\boldsymbol{v}_1 \approx (0.9\,,-0.4)$, $\sigma_1 \approx 2.3$, $\boldsymbol{v}_2 \approx (0.4\,,0.9)$, $\sigma_1 \approx 0.3$.

3.4.1b : Not a subspace.

3.4.1d : Not a subspace.

3.4.1f : Not a subspace.

3.4.1h : Subspace.

3.4.1j : Not a subspace.

3.4.1l : Subspace.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

3.4.1n : Not a subspace.

3.4.1p : Not a subspace.

3.4.2b : Subspace.

3.4.2d : Not a subspace.

3.4.2f : Subspace.

3.4.2h : Subspace.

3.4.2j : Not a subspace.

3.4.2l : Not a subspace.

3.4.2n : Subspace.

3.4.2p : Subspace.

3.4.4a : $\boldsymbol{b}_5$ is in column space; $\boldsymbol{r}_5$ is in row space.

3.4.4c : $\boldsymbol{b}_5$ is not in column space; $\boldsymbol{r}_5$ is in row space.

3.4.4e : $\boldsymbol{b}_5$ is in column space; $\boldsymbol{r}_5$ is not in row space.

3.4.4g : $\boldsymbol{b}_5$ is in column space; $\boldsymbol{r}_5$ is in row space.

3.4.4i : $\boldsymbol{b}_5$ is in column space; $\boldsymbol{r}_5$ is in row space.

3.4.5b : no

3.4.5d : yes

3.4.5f : yes

3.4.6b : $(\frac{4}{5}, \frac{3}{5})$

3.4.6d : $(2, 2, 1)/3$

3.4.8a : (2 d.p.) column space $\{(-0.61, 0.19, 0.77), (-0.77, -0.36, -0.52)\}$; row space $\{(-0.35, 0.84, 0.42), (-0.94, -0.31, -0.15)\}$; nullspace $\{(0, 0.45, -0.89)\}$.

3.4.8c : (2 d.p.) column space $\{(0.00, -0.81, -0.48, 0.33), (0.00, -0.08, 0.66, 0.74)\}$; row space $\{(-0.35, -0.89, -0.08, 0.27), (-0.48, 0.17, -0.80, -0.32)\}$; nullspace $\{(-0.78, 0.36, 0.43, 0.28), (-0.19, -0.22, 0.41, -0.86)\}$.

3.4.8e : (2 d.p.) column space $\{(-0.53, -0.11, -0.64, 0.01, 0.54), (0.40, -0.37, 0.03, 0.76, 0.35)\}$; row space $\{(0.01, -0.34, -0.30, -0.89), (0.21, -0.92, 0.11, 0.32)\}$; nullspace $\{(-0.06, -0.01, 0.95, -0.31), (0.98, 0.20, 0.04, -0.08)\}$.

3.4.8g : (2 d.p.) column space $\{(0.52, -0.14, -0.79, 0.28), (-0.02, 0.18, -0.37, -0.91), (-0.31, -0.94, -0.09, -0.14)\}$; row space $\{(-0.16, 0.29, 0.13, -0.87, 0.33), (-0.15, 0.67, 0.58, 0.40, 0.17), (-0.75, -0.12, 0.19, -0.11, -0.61)\}$; nullspace $\{(0.62, 0.05, 0.45, -0.25, -0.59), (-0.05, -0.67, 0.63, 0.02, 0.38)\}$.

3.4.9 : 1, 1, 1; 3, 3, 0; 2, 2, 2; 3, 3, 2; 2, 2, 2; 2, 2, 2; 3, 3, 2; 4, 4, 1.

v0.2a

3.4.10b : 0,1,2,3

3.4.10d : 2,3,4,5,6

3.4.10f : 0,1,2,3,4,5

3.4.11b : yes.

3.4.11d : no.

3.5.4 : To within an arbitrary constant: Adelaide, -0.33; Brisbane, -1.00; Canberra, 1.33.

3.5.6 : To within an arbitrary constant: Atlanta, 1.0; Boston, 0.0; Concord, 0.8; Denver, -1.6; Frankfort, -0.2.

3.5.8 : flow-rate $= -0.658 + 0.679\,\text{voltage}$ (3.d.p)

3.5.14b : $\boldsymbol{x} = (1\,,-2)$

3.5.14d : $(p\,,q) = (-0.7692\,,-1.1539)$

3.5.14f : $\boldsymbol{x} = (\frac{1}{3}\,,\frac{4}{3}\,,-\frac{8}{3})$

3.5.14h : $\boldsymbol{q} = (0.0741\,,0.0741\,,-0.3704)$

3.5.16 : $\boldsymbol{r} = (0.70\,,0.14\,,0.51\,,0.50\,,0.70\,,1.00\,,0.90\,,0.51\,,0.71)$ so the middle left is the most absorbing.

3.5.20a : $\operatorname{proj}_{\boldsymbol{u}}(\boldsymbol{v}) = (1.6\,,0.8)$, $x = 0.8$

3.5.20c : (2 d.p.) $\operatorname{proj}_{\boldsymbol{u}}(\boldsymbol{v}) = -\boldsymbol{e}_1$, $x = -0.17$

3.5.20e : (2 d.p.) $\operatorname{proj}_{\boldsymbol{u}}(\boldsymbol{v}) = (0.67\,,0.83\,,-0.17)$, $x = 0.17$

3.5.20g : $\operatorname{proj}_{\boldsymbol{u}}(\boldsymbol{v}) = \boldsymbol{e}_2$, $x = 0.5$

3.5.20i : (2 d.p.) $\operatorname{proj}_{\boldsymbol{u}}(\boldsymbol{v}) = (0.76\,,1.52\,,0\,,-0.38)$, $x = 0.38$

3.5.20k : (2 d.p.) $\operatorname{proj}_{\boldsymbol{u}}(\boldsymbol{v}) = (-0.04\,,-0.09\,,-0.04\,,0.04\,,0.17)$, $x = -0.04$

3.5.21a : (2 d.p.) $\operatorname{proj}_{\mathbb{W}}(\boldsymbol{v}) = (1.04\,,0.77\,,-1.31)$

3.5.21c : (2 d.p.) $\operatorname{proj}_{\mathbb{W}}(\boldsymbol{v}) = (1.10\,,-0.73\,,0.80)$

3.5.21e : $\operatorname{proj}_{\mathbb{W}}(\boldsymbol{v}) = \boldsymbol{v}$

3.5.21g : (2 d.p.) $\operatorname{proj}_{\mathbb{W}}(\boldsymbol{v}) = (4.08\,,1.14\,,-2.27\,,3.03)$

3.5.21i : (2 d.p.) $\operatorname{proj}_{\mathbb{W}}(\boldsymbol{v}) = (0.06\,,-3.28\,,-1.17\,,0.06)$

3.5.21k : (2 d.p.) $\operatorname{proj}_{\mathbb{W}}(\boldsymbol{v}) = (2.68\,,-2.24\,,0.88\,,0.06)$

3.5.22a : (2 d.p.) $\begin{bmatrix} 0.88 & -0.08 & 0.31 \\ -0.08 & 0.95 & 0.21 \\ 0.31 & 0.21 & 0.17 \end{bmatrix}$

3.5.22c : (2 d.p.) $\begin{bmatrix} 0.25 & 0.37 & 0.22 \\ 0.37 & 0.81 & -0.11 \\ 0.22 & -0.11 & 0.93 \end{bmatrix}$

v0.2a

3.5.22e : I_2

3.5.22g : (2 d.p.) $\begin{bmatrix} 0.63 & -0.42 & 0.05 & 0.23 \\ -0.42 & 0.51 & 0.05 & 0.26 \\ 0.05 & 0.05 & 0.99 & -0.03 \\ 0.23 & 0.26 & -0.03 & 0.86 \end{bmatrix}$

3.5.22i : (2 d.p.) $\begin{bmatrix} 0.66 & -0.30 & -0.16 & 0.21 & -0.25 \\ -0.30 & 0.39 & 0.15 & 0.33 & 0.13 \\ -0.16 & 0.15 & 0.06 & 0.08 & 0.06 \\ 0.21 & 0.33 & 0.08 & 0.79 & -0.06 \\ -0.25 & 0.13 & 0.06 & -0.06 & 0.09 \end{bmatrix}$

3.5.23a : (2 d.p.) $\boldsymbol{b}' = (5.84\,,-19.10\,,-2.58)$, difference 0.46

3.5.23c : (2 d.p.) $\boldsymbol{b}' = (9.27\,,10.75\,,-1.18)$, difference 0.41

3.5.23e : (2 d.p.) $\boldsymbol{b}' = (1.96\,,5.52\,,-2.69)$, difference 1.21

3.5.23g : (2 d.p.) $\boldsymbol{b}' = (-9\,,10)$, difference 0

3.5.23i : (2 d.p.) $\boldsymbol{b}' = (7.00\,,-5.50\,,1.38\,,-7.00)$, difference 2.32

3.5.23k : (2 d.p.) $\boldsymbol{b}' = (0.71\,,-48.77\,,27.40\,,-96.00)$, difference 5.40

3.5.27a : The line $x = 2y$

3.5.27c : The plane $x + 9y - 9z = 0$

3.5.27e : It is not a subspace as it does not include $\mathbf{0}$, and so does not have an orthogonal complement.

3.5.27g : The line span{(7,-25,8)}

3.5.28a : $\{(-\frac{6}{7}\,,\frac{2}{7}\,,-\frac{3}{7})\}$ is one possibility.

3.5.28c : This plane is not a subspace (does not include $\mathbf{0}$), so does not have an orthogonal complement.

3.5.28e : $\{(-0.53\,,-0.43\,,0.73)\,,(0.28\,,0.73\,,0.63)\}$ is one possibility (2 d.p.).

3.5.28g : $\{(-0.73,-0.30,0.29,-0.22,-0.50),(-0.18,-0.33,0.24,-0.43,0.79)\,,(-0.06\,,-0.74\,,-0.51\,,0.43\,,0.06)\}$ is one possibility (2 d.p.).

3.5.28i : $\{(-0.05-0.910.250.32),(-0.580.350.450.58)\}$ is one possibility (2 d.p.).

3.5.30a : $\frac{1}{9}(20\,,34\,,8)$

3.5.30c : $\frac{1}{27}(8\,,-62\,,-40)$

3.5.30e : $\frac{1}{27}(67\,,95\,,16)$

3.5.31a : $(0.96\,,2.06\,,-1.02\,,-0.88)$

3.5.31c : $(0.54\,,-3.42\,,0.54\,,1.98)$

3.5.33a : $(-2\,,4) = (\frac{6}{5}\,,\frac{8}{5}) + (-\frac{16}{5}\,,\frac{12}{5})$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

3.5.33c : $(0\,,0) = (0\,,0) + (0\,,0)$

3.5.34a : $(-3\,,6\,,-2) + (-2\,,-2\,,-3)$

3.5.34c : $(0.31\,,-0.61\,,0.20) + (0.69\,,-0.39\,,-2.20)$ (2 d.p.)

3.5.35a : $(6\,,-2\,,0\,,0) + (-1\,,-3\,,1\,,-3)$

3.5.35c : $(2.1\,,-0.7\,,-4.5\,,-1.5) + (-0.1\,,-0.3\,,0.5\,,-1.5)$

3.5.36 : Either $\mathbb{W} = \operatorname{span}\{(1\,,2)\}$ and $\mathbb{W}^\perp = \operatorname{span}\{(-2\,,1)\}$, or vice-versa.

3.5.38 : Use $x_1x_2x_3x_4$-space. Either $\mathbb{W}$ is x_2-axis, or $x_1x_2x_4$-space, or any plane in $x_1x_2x_4$-space that contains the x_2-axis, and $\mathbb{W}^\perp$ corresponding complement, or vice-versa.

3.6.1b : Not a LT.

3.6.1d : Maybe a LT, with standard matrix $\begin{bmatrix} 0.8 & -0.3 \\ 0.4 & 1.0 \end{bmatrix}$

3.6.1f : Maybe a LT, with standard matrix $\begin{bmatrix} 0.9 & -2.2 \\ -1.7 & -1.6 \end{bmatrix}$

3.6.1h : Not a LT.

3.6.1j : Maybe a LT, with standard matrix $\begin{bmatrix} -2.6 & -0.5 \\ 1.4 & -0.1 \end{bmatrix}$

3.6.3b : Not a LT.

3.6.3d : Not a LT.

3.6.3f : Maybe a LT.

3.6.7b : $[E] = \begin{bmatrix} 0 & 0 \\ 3 & 7 \\ 0 & -2 \\ 0 & -3 \end{bmatrix}$

3.6.7d : Not a LT.

3.6.7f : Not a LT.

3.6.7h : $[E] = \begin{bmatrix} 7 & 0 & -9 & 0 & -1 \\ 3 & 0 & 0 & -3 & 0 \end{bmatrix}$

3.6.11b : $E^+ = \begin{bmatrix} 0.57 & -0.86 \\ -0.57 & 2.86 \\ -0.29 & 1.43 \end{bmatrix}$ (2 d.p.)

3.6.11d : $E^+ = \begin{bmatrix} 0.17 & -0.12 & -0.15 & 0.31 \\ -0.05 & 0.04 & -0.01 & -0.08 \\ -0.16 & 0.14 & -0.12 & -0.20 \end{bmatrix}$ (2 d.p.)

v0.2a

3.6.11f : $E^+ = \begin{bmatrix} -0.07 & -0.89 & 0.28 & 0.45 & 0.33 \\ -0.24 & -0.25 & -2.02 & 0.36 & 1.32 \\ -0.33 & -0.35 & -0.27 & -0.40 & -0.07 \\ 0.36 & 0.07 & 1.10 & 0.15 & 0.13 \\ 1.16 & -1.07 & 0.61 & 1.24 & 2.11 \\ -0.57 & 0.36 & -1.05 & -0.89 & -0.82 \end{bmatrix}$ (2 d.p.)

3.6.11h : $E^+ = \begin{bmatrix} -0.80 & -0.54 & -0.60 & -0.89 \\ -2.11 & -0.02 & 1.45 & -5.31 \\ -0.34 & -0.02 & 0.73 & -0.56 \\ -1.96 & -0.29 & 0.73 & -4.43 \\ 3.43 & 0.51 & -1.83 & 8.34 \\ 2.66 & 0.58 & -1.23 & 6.29 \end{bmatrix}$ (2 d.p.)

3.6.14b : Equivalent to rotation by 60° with matrix $\begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}$.

3.6.14d : Equivalent to rotation by −90° (clockwise 90°) with matrix $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$.

3.6.14f : Equivalent to reflection in the line $y = x$ with matrix $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

3.6.15b : $[S \circ T] = \begin{bmatrix} -6 \\ 9 \end{bmatrix}$ and $[T \circ S]$ does not exist.

3.6.15d : $[S \circ T]$ does not exist, and $[T \circ S] = \begin{bmatrix} 30 & -6 & 24 \\ -15 & 3 & -12 \\ 15 & -3 & 12 \\ 5 & -1 & 4 \end{bmatrix}$

3.6.15f : $[S \circ T] = \begin{bmatrix} 12 & 11 \\ -6 & -7 \\ -3 & 6 \end{bmatrix}$ and $[T \circ S]$ does not exist.

3.6.16b : Inverse $\approx \begin{bmatrix} 1.43 & -0.44 \\ -0.00 & 0.77 \end{bmatrix}$

3.6.16d : Inverse $\approx \begin{bmatrix} -0.23 & -0.53 \\ -0.60 & 0.26 \end{bmatrix}$

v0.2a

4 Eigenvalues and eigenvectors of symmetric matrices

Chapter Contents

4.1 Introduction to eigenvalues and eigenvectors 409

4.1.1 Systematically find eigenvalues and eigenvectors 416

4.1.2 Exercises 429

4.2 Beautiful properties for symmetric matrices 435

4.2.1 Matrix powers maintain eigenvectors 435

4.2.2 Symmetric matrices are orthogonally diagonalisable 441

4.2.3 Change orthonormal basis to classify quadratics 450

4.2.4 Exercises 457

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, observe the columns of U are always $\pm$ the columns of V (well, almost always).

Solution: Repeat as often us you like for any size of square matrix that you like (one example is recorded here to two decimal places).

(a) Generate in MATLAB/Octave some random symmetric matrix by adding a random matrix to its transpose with `A=randn(5); A=A+A'` (Table 3.1):

```
A =
  -0.45  -0.18   1.59  -0.96  -0.54
  -0.18  -0.24  -1.04   0.14   0.80
   1.59  -1.04  -2.87  -0.40   1.11
  -0.96   0.14  -0.40  -0.26  -1.90
  -0.54   0.80   1.11  -1.90   1.64
```

This matrix is symmetric as $a_{ij} = a_{ji}$.

(b) Find its SVD via `[U,S,V]=svd(A)`

```
U =
  -0.41  -0.09  -0.28  -0.67   0.55
   0.25  -0.11  -0.05   0.53   0.80
   0.82  -0.19  -0.40  -0.36  -0.07
  -0.15   0.51  -0.80   0.27  -0.11
  -0.27  -0.83  -0.36   0.25  -0.22
S =
   4.28      0      0      0      0
      0   3.12      0      0      0
      0      0   1.65      0      0
      0      0      0   1.14      0
      0      0      0      0   0.51
V =
   0.41  -0.09   0.28  -0.67  -0.55
  -0.25  -0.11   0.05   0.53  -0.80
  -0.82  -0.19   0.40  -0.36   0.07
   0.15   0.51   0.80   0.27   0.11
   0.27  -0.83   0.36   0.25   0.22
```

Observe the second and fourth columns of U and V are identical, and the other columns have opposite signs.

Repeat and observe $\boldsymbol{u}_j = \pm\boldsymbol{v}_j$ for all columns j.

■

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 $AV = US$ is $A\boldsymbol{v}_j = \sigma_j\boldsymbol{u}_j$. Example 4.0.1 observes for symmetric A that $\boldsymbol{u}_j = \pm\boldsymbol{v}_j$ (almost always) so this last equation becomes $A\boldsymbol{v}_j = (\pm\sigma_j)\boldsymbol{v}_j$. This equation is of the important form $A\boldsymbol{v} = \lambda\boldsymbol{v}$. 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 $= A\boldsymbol{n}$ where $\boldsymbol{n}$ is the normal unit vector to the internal 'surface'. Matrix A is always symmetric. To know whether a material will break apart under pulling, or to crumble under compression, we need to know where the extreme forces are. They are found as solutions to $A\boldsymbol{n} = \lambda\boldsymbol{n}$ where $\boldsymbol{n}$ gives the direction and λ the strength of the force.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

4.1 Introduction to eigenvalues and eigenvectors

Section Contents

4.1.1 Systematically find eigenvalues and eigenvectors 416
Compute eigenvalues and eigenvectors 416
Find eigenvalues and eigenvectors by hand . . 423
4.1.2 Exercises 429

This chapter focuses on some marvellous properties of symmetric matrices. Nonetheless it defines basic concepts in general. Chapter 7 explores analogous properties for general matrices. The marvellously useful properties result from asking for which vectors does multiplication by a given matrix simply stretch or shrink the vector.

Definition 4.1.1. *Let A be a square matrix. A scalar λ is called an* ***eigenvalue*** *of A if there is a nonzero vector $\boldsymbol{x}$, called an* ***eigenvector*** *corresponding to λ, such that $A\boldsymbol{x} = \lambda\boldsymbol{x}$.*

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?

Solution: The simplest approach is to multiply the matrix by the given vector and see what happens.

(a) For vector $\boldsymbol{x} = (1, 0, -1)$,

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

Hence $(1, 0, -1)$ is an eigenvector of A corresponding to the eigenvalue $\lambda = 1$.

(b) For vector $\boldsymbol{x} = (2, -4, 2)$,

$$A\boldsymbol{x} = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ -4 \\ 2 \end{bmatrix} = \begin{bmatrix} 6 \\ -12 \\ 6 \end{bmatrix} = 3 \cdot \boldsymbol{x} .$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Hence $(2,-4,2)$ is an eigenvector of A corresponding to the eigenvalue $\lambda = 3$.

(c) For vector $\boldsymbol{x} = (1,2,1)$,

$$A\boldsymbol{x} = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix} \not\propto \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}.$$

If there was a constant of proportionality (an eigenvalue), then the first component would require the constant $\lambda = -1$ but the second component would require $\lambda = +1$ which is a contradiction. Hence $(1,2,1)$ is not an eigenvector of A.

(d) Inspection is useful if it is quick: here one might see that the elements in each row of A sum to the same thing, namely zero, so try vector $\boldsymbol{x} = (1,1,1)$:

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

Hence $(1,1,1)$ is an eigenvector of A corresponding to the eigenvalue $\lambda = 0$.

■

Importantly, eigenvectors tell us key directions of a given matrix: the directions in which the multiplication by a matrix is to simply stretch, shrink, or reverse by a factor: the factor being the corresponding eigenvalue. In two dimensional plots we can graphically estimate eigenvectors and eigenvalues.

We might 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 opposite direction), then $\boldsymbol{x}$ is an eigenvector; if they form a non-trivial angle, then $\boldsymbol{x}$ is not an eigenvector.

Example 4.1.3. Let the matrix $A = \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix}$. 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 $\lambda = \frac{1}{2}$. ■

As in the next example, we can 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 4.1.4 (graphical eigenvectors one)**.**

The MATLAB function `eigshow(A)` provides an interactive alternative to this static view.

The plot on the left shows (unit) vectors $\boldsymbol{x}$ (blue), and for some 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.

Solution: We seek vectors $\boldsymbol{x}$ such that $A\boldsymbol{x}$ is in the same direction (to graphical accuracy). It appears that vectors at 45° to the axes are the only ones for which $A\boldsymbol{x}$ is in the same direction as $\boldsymbol{x}$:

- the two (blue) vectors $\pm(0.7,0.7)$ appear to be shrunk to length a half (red) so we estimate two eigenvectors are $\boldsymbol{x} \approx \pm(0.7,0.7)$ and the corresponding eigenvalue $\lambda \approx 0.5$;
- the two (blue) vectors $\pm(0.7,-0.7)$ appear to be stretched by a factor about 1.5 (red) so we estimate two eigenvectors are $\boldsymbol{x} \approx \pm(0.7,-0.7)$ and the corresponding eigenvalue is $\lambda \approx 1.5$;
- and for no other (unit) vector $\boldsymbol{x}$ is $A\boldsymbol{x}$ aligned with $\boldsymbol{x}$.

Any multiple of these eigenvectors will also be eigenvectors so we may report the directions more simply, perhaps $(1,1)$ and $(1,-1)$ respectively.

■

Example 4.1.5 (graphical eigenvectors two)**.**

The plot on the left shows (unit) vectors $\boldsymbol{x}$ (blue), and for some 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.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Solution: We seek vectors $\boldsymbol{x}$ such that $A\boldsymbol{x}$ is in the same direction (to graphical accuracy):

- the two (blue) vectors $\pm(0.9, -0.3)$ appear stretched a little by a factor about 1.2 (red) so we estimate eigenvectors are $\boldsymbol{x} \propto (0.9, -0.3)$ and the corresponding eigenvalue is $\lambda \approx 1.2$;
- the two (blue) vectors $\pm(0.3, 0.9)$ appear shrunk and *reversed* by a factor about 0.4 (red) so we estimate eigenvectors are $\boldsymbol{x} \propto (0.3, 0.9)$ and the corresponding eigenvalue is $\lambda \approx -0.4$ — negative because the direction is reversed;
- and for no other (unit) vector $\boldsymbol{x}$ is $A\boldsymbol{x}$ aligned with $\boldsymbol{x}$.

If this matrix arose in the description of forces inside a solid, then the forces would be compressive in directions $\pm(0.3, 0.9)$, and the forces would be (tension) 'ripping apart' the solid in directions $\pm(0.9, -0.3)$.

■

Example 4.1.6 (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 $\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n$ in turn:

$$D\boldsymbol{e}_1 = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} d_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = d_1\boldsymbol{e}_1\,;$$

$$D\boldsymbol{e}_2 = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ d_2 \\ \vdots \\ 0 \end{bmatrix} = d_2\boldsymbol{e}_2\,;$$

$$\vdots$$

$$D\boldsymbol{e}_n = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ d_n \end{bmatrix} = d_n\boldsymbol{e}_n\,.$$

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. ■

v0.2a

Eigenvalues The 3×3 matrix of Example 4.1.2 has three eigenvalues. The 2×2 matrices underlying Examples 4.1.4–4.1.5 both have two eigenvalues. 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 4.1.6, a straightforward extension of the working shows 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.

Theorem 4.1.7. *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 λ, together with the zero vector, is a subspace; the subspace is called the* ***eigenspace*** *of λ and is denoted by $\mathbb{E}_\lambda$.*

Proof. From Definition 4.1.1, $A\boldsymbol{x} = \lambda\boldsymbol{x} \iff A\boldsymbol{x} - \lambda\boldsymbol{x} = \boldsymbol{0} \iff A\boldsymbol{x} - \lambda I\boldsymbol{x} = \boldsymbol{0} \iff (A - \lambda I)\boldsymbol{x} = \boldsymbol{0}$. Also, eigenvectors $\boldsymbol{x}$ must be nonzero, so the homogeneous system $(A - \lambda I)\boldsymbol{x} = \boldsymbol{0}$ must have nonzero solutions. Theorem 3.4.14 assures us that the set of solutions to a homogeneous system, such as $(A - \lambda I)\boldsymbol{x} = \boldsymbol{0}$ for any given λ, is a subspace. Hence the set of eigenvectors for any given eigenvalue λ, nonzero solutions, together with $\boldsymbol{0}$, form a subspace. □

Example 4.1.8. 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 $\mathbb{E}_0$.

Solution:

- The eigenspace $\mathbb{E}_1$ is the set of solutions of

$$(A - 1I)\boldsymbol{x} = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 1 & -1 \\ 0 & -1 & 0 \end{bmatrix} \boldsymbol{x} = \boldsymbol{0}.$$

  That is, $-x_2 = 0$, $-x_1 + x_2 - x_3 = 0$ and $-x_2 = 0$. Hence, $x_2 = 0$ and $x_1 = -x_3$. A general solution is $\boldsymbol{x} = (t, 0, -t)$ so the eigenspace $\mathbb{E}_1 = \{(t, 0, -t) : t \in \mathbb{R}\} = \operatorname{span}\{(1, 0, -1)\}$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

- The eigenspace $\mathbb{E}_3$ is the set of solutions of

$$(A-3I)\boldsymbol{x} = \begin{bmatrix} -2 & -1 & 0 \\ -1 & -1 & -1 \\ 0 & -1 & -2 \end{bmatrix} \boldsymbol{x} = \boldsymbol{0}\,.$$

  That is, $-2x_1 - x_2 = 0$, $-x_1 - x_2 - x_3 = 0$ and $-x_2 - 2x_3 = 0$. From the first $x_2 = -2x_1$ which substituted into the third gives $2x_3 = -x_2 = 2x_1$. This suggests we try $x_1 = t$, $x_2 = -2t$ and $x_3 = t$; that is, $\boldsymbol{x} = (t, -2t, t)$. This also satisfies the second equation and so is a general solution. So the eigenspace $\mathbb{E}_3 = \{(t, -2t, t) : t \in \mathbb{R}\} = \operatorname{span}\{(1, -2, 1)\}$.

- The eigenspace $\mathbb{E}_0$ is the set of solutions of

$$(A-0I)\boldsymbol{x} = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \boldsymbol{x} = \boldsymbol{0}\,.$$

  That is, $x_1 - x_2 = 0$, $-x_1 + 2x_2 - x_3 = 0$ and $-x_2 + x_3 = 0$. The first and third of these require $x_1 = x_2 = x_3$ which also satisfies the second. Thus a general solution is $\boldsymbol{x} = (t, t, t)$ so the eigenspace $\mathbb{E}_0 = \{(t, t, t) : t \in \mathbb{R}\} = \operatorname{span}\{(1, 1, 1)\}$.

■

Example 4.1.9 (graphical eigenspaces)**.**

The plot on the left shows (unit) vectors $\boldsymbol{x}$ (blue), and for the matrix A of Example 4.1.5 the corresponding vectors $A\boldsymbol{x}$ (red) adjoined. Estimate and draw the eigenspaces of matrix A.

Solution:

Example 4.1.5 found directions in which $A\boldsymbol{x}$ is aligned with $\boldsymbol{x}$. Then the corresponding eigenspace is all vectors in the line aligned with that direction, including the opposite direction.

■

Example 4.1.10. Eigenspaces may be multidimensional. Find the eigenspaces of the diagonal matrix

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

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Solution: Example 4.1.6 shows this matrix has two distinct eigenvalues $\lambda = -\frac{1}{3}$ and $\lambda = \frac{3}{2}$.

- Eigenvectors corresponding to eigenvalue $\lambda = -\frac{1}{3}$ satisfy

$$(D + \tfrac{1}{3}I)\boldsymbol{x} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{11}{6} & 0 \\ 0 & 0 & \frac{11}{6} \end{bmatrix} \boldsymbol{x} = \boldsymbol{0}\,.$$

  Hence $\boldsymbol{x} = t\boldsymbol{e}_1$ are eigenvectors, for nonzero t. The eigenspace $\mathbb{E}_{-1/3} = \{t\boldsymbol{e}_1 : t \in \mathbb{R}\} = \operatorname{span}\{\boldsymbol{e}_1\}$.

- Eigenvectors corresponding to eigenvalue $\lambda = \frac{3}{2}$ satisfy

$$(D - \tfrac{3}{2}I)\boldsymbol{x} = \begin{bmatrix} -\frac{11}{6} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{x} = \boldsymbol{0}\,.$$

  Hence $\boldsymbol{x} = t\boldsymbol{e}_2 + s\boldsymbol{e}_3$ are eigenvectors, for nonzero t and s. Then the eigenspace $\mathbb{E}_{3/2} = \{t\boldsymbol{e}_2 + s\boldsymbol{e}_3 : t, s \in \mathbb{R}\} = \operatorname{span}\{\boldsymbol{e}_2, \boldsymbol{e}_3\}$ is two dimensional.

■

Definition 4.1.11. *For a symmetric matrix, the* ***multiplicity*** *of an eigenvalue λ is the dimension of the corresponding eigenspace $\mathbb{E}_\lambda$.*[1]

Example 4.1.12. The multiplicity of the various eigenvalues in earlier examples are the following.

4.1.8 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.

4.1.10 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.

■

[1] Section 7.3 discusses that for non-symmetric matrices the dimension of an eigenspace may be less than the multiplicity of an eigenvalue, Theorem 7.3.12, but for symmetric matrices they are the same.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Table 4.1: As well as the MATLAB/Octave commands and operations listed in Tables 1.2, 2.3, 3.1, 3.2, and 3.3 we need the eigenvector function.

- `[V,D]=eig(A)` computes eigenvectors and the eigenvalues of the $n \times n$ square matrix A.
  - The n eigenvalues of A (repeated according to their multiplicity, Definition 4.1.11) form the diagonal of $n \times n$ square matrix $D = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$.
  - 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, Definition 4.1.11).
- If the matrix A is a real symmetric matrix, then the eigenvalues and eigenvectors are all real, and the eigenvector matrix V is orthogonal.
  If the matrix A is either not symmetric, or is complex valued, then the eigenvalues and eigenvectors may be complex valued.

4.1.1 Systematically find eigenvalues and eigenvectors

Computer packages easily compute eigenvalues and eigenvectors for us. However, sometimes we need to explicitly see dependence upon a parameter so we also explore small systems by hand.

Compute eigenvalues and eigenvectors

Compute in Matlab/Octave. `[V,D]=eig(A)` computes eigenvalues and eigenvectors. The eigenvalues are placed in the diagonal of $D = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$. 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 4.2.18).

Example 4.1.13. 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 $D = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$, and confirm that the computed V is orthogonal.

Solution: Enter the matrix into MATLAB/Octave and execute `eig()`:

```
A=[1 -1 0;-1 2 -1;0 -1 1]
[V,D]=eig(A)
```

v0.2a

The output is

```
A =
     1    -1     0
    -1     2    -1
     0    -1     1
V =
   -0.5774   -0.7071    0.4082
   -0.5774    0.0000   -0.8165
   -0.5774    0.7071    0.4082
D =
    0.0000         0         0
         0    1.0000         0
         0         0    3.0000
```

- The first diagonal element of `D` is zero (actually 10^{-16}) so eigenvalue $\lambda_1 = 0$. A corresponding eigenvector is the first column of `V`, namely $\boldsymbol{v}_1 = -0.5774(1,1,1)$; since eigenvectors can be scaled by a constant, we could also say an eigenvector is $\boldsymbol{v}_1 = (1,1,1)$.
- The second diagonal element of `D` is one so eigenvalue $\lambda_2 = 1$. A corresponding eigenvector is the second column of `V`, namely $\boldsymbol{v}_2 = 0.7071(-1,0,1)$; we could also say an eigenvector is $\boldsymbol{v}_2 = (-1,0,1)$.
- The third diagonal element of `D` is three so eigenvalue $\lambda_3 = 3$. A corresponding eigenvector is the third column of `V`, namely $\boldsymbol{v}_3 = 0.4082(1,-2,1)$; we could also say an eigenvector is $\boldsymbol{v}_3 = (1,-2,1)$.

Confirm $AV = VD$ simply by computing `A*V-V*D` and seeing it is zero (to numerical error of circa 10^{-16}):

```
ans =
   5.7715e-17  -1.1102e-16   4.4409e-16
   1.6874e-16   1.2490e-16   0.0000e+00
  -5.3307e-17  -1.1102e-16  -2.2204e-16
```

To verify the computed matrix `V` is orthogonal (Definition 3.2.43), check `V'*V` gives the identity:

```
ans =
    1.0000   -0.0000    0.0000
   -0.0000    1.0000    0.0000
    0.0000    0.0000    1.0000
```

■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Example 4.1.14 (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, $y_2 - y_1$ and $y_3 - y_2$. For simplicity, suppose the constants of proportionality are all one.

- The left mass (y_1) is accelerated by the spring connecting it to the middle mass (y_2); that is, $d^2y_1/dt^2 = y_2 - y_1$.
- The middle mass (y_2) is accelerated by the springs connecting it to the left mass (y_1) and to the right mass (y_3); that is, $d^2y_2/dt^2 = (y_1 - y_2) + (y_3 - y_2) = y_1 - 2y_2 + y_3$.
- The right mass (y_3) is accelerated by the spring connecting it to the middle mass (y_2); that is, $d^2y_3/dt^2 = y_2 - y_3$.

Guess there are solutions oscillating in time, so let's see if we can find solutions $y_i(t) = x_i \cos(ft)$ for some as yet unknown frequency f. Substitute and the three differential equations become

$$\begin{aligned}
-f^2x_1\cos(ft) &= x_2\cos(ft) - x_1\cos(ft),\\
-f^2x_2\cos(ft) &= x_1\cos(ft) - 2x_2\cos(ft) + x_3\cos(ft),\\
-f^2x_3\cos(ft) &= x_2\cos(ft) - x_3\cos(ft).
\end{aligned}$$

These are satisfied for all time t only if the coefficients of the cosine 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 eigenproblem $A\boldsymbol{x} = \lambda\boldsymbol{x}$ for symmetric matrix A of Example 4.1.13 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 4.1.13 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 the middle mass is stationary while the outer two masses oscillate in and out in opposition to each other.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

- 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 the static displacement is that of all three masses moved all together as a unit.

That these three solutions combine together form a general solution of the system of differential equations is a topic for a course on differential equations. ■

Example 4.1.15 (Sierpinski network)**.** Consider three triangles formed into a triangle (as shown in the margin)—perhaps because triangles make strong structures, or perhaps because of a hierarchical computer/social network. Form an matrix $A = [a_{ij}]$ of ones if node i is connected to node j; and set the diagonal a_{ii} to be minus the number of other nodes to which node i is connected. The symmetry of the matrix A follows from the symmetry of the connections: construct the matrix, check it is symmetric, and find the eigenvalues and eigenspaces with MATLAB/Octave, and their multiplicity. For the computed matrices V and D, check that $AV = VD$ and also that V is orthogonal.

Solution: In MATLAB/Octave use

```
A=[-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 ]
A-A'
[V,D]=eig(A)
```

To two decimal places so that it fits the page, the computation may give

```
V =
 -0.41  0.51 -0.16 -0.21 -0.45  0.18 -0.40  0.06  0.33
  0.00 -0.13  0.28  0.63  0.13 -0.18 -0.58 -0.08  0.33
  0.41 -0.20 -0.49 -0.42  0.32  0.01 -0.36 -0.17  0.33
 -0.41 -0.11  0.52 -0.42  0.32  0.01  0.14 -0.37  0.33
 -0.00 -0.18 -0.26  0.37 -0.22  0.51  0.36 -0.46  0.33
  0.41  0.53  0.07  0.05 -0.10 -0.51  0.33 -0.23  0.33
```

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

```
 -0.41 -0.39 -0.36  0.05 -0.10 -0.51  0.25  0.31  0.33
  0.00  0.31 -0.03  0.16  0.55  0.34  0.22  0.55  0.33
  0.41 -0.33  0.42 -0.21 -0.45  0.18  0.03  0.40  0.33
D =
 -5.00     0     0     0     0     0     0     0     0
     0 -4.30     0     0     0     0     0     0     0
     0     0 -4.30     0     0     0     0     0     0
     0     0     0 -3.00     0     0     0     0     0
     0     0     0     0 -3.00     0     0     0     0
     0     0     0     0     0 -3.00     0     0     0
     0     0     0     0     0     0 -0.70     0     0
     0     0     0     0     0     0     0 -0.70     0
     0     0     0     0     0     0     0     0 -0.00
```

The five eigenvalues are -5.00, -4.30, -3.00, -0.70 and 0.00 (to two decimal places). Three of the eigenvalues are repeated as a consequence of the additional geometric symmetry in the network. The following are the eigenspaces.

- Corresponding to eigenvalue $\lambda = -5$ are eigenvectors $\boldsymbol{x} \propto (-0.41, 0, 0.41, -0.41, 0, 0.41, -0.41, 0, 0.41)$; that is, the eigenspace $\mathbb{E}_{-5} = \operatorname{span}\{(-1, 0, 1, -1, 0, 1, -1, 0, 1)\}$. From Definition 4.1.11 the multiplicity of eigenvalue $\lambda = -5$ is one.

- Corresponding to eigenvalue $\lambda = -4.30$ there are two eigenvectors computed by MATLAB/Octave. These two eigenvectors are orthogonal (you should check). Because these arise as the solutions of the homogeneous system $(A - \lambda I)\boldsymbol{x} = \mathbf{0}$, any (nonzero) linear combination of these is also an eigenvector corresponding to the same eigenvalue. That is, the eigenspace

$$\mathbb{E}_{-4.30} = \operatorname{span}\left\{ \begin{bmatrix} 0.51 \\ -0.13 \\ -0.20 \\ -0.11 \\ -0.18 \\ 0.53 \\ -0.39 \\ 0.31 \\ -0.33 \end{bmatrix}, \begin{bmatrix} -0.16 \\ 0.28 \\ -0.49 \\ 0.52 \\ -0.26 \\ 0.07 \\ -0.36 \\ -0.03 \\ 0.42 \end{bmatrix} \right\}.$$

  Hence the eigenvalue $\lambda = -4.30$ has multiplicity two.

- Corresponding to eigenvalue $\lambda = -3$ there are three eigenvectors computed by MATLAB/Octave. These three eigenvectors

are orthogonal (you should check). Thus the eigenspace

$$\mathbb{E}_{-3} = \operatorname{span}\left\{ \begin{bmatrix} -0.21 \\ 0.63 \\ -0.42 \\ -0.42 \\ 0.37 \\ 0.05 \\ 0.05 \\ 0.16 \\ -0.21 \end{bmatrix}, \begin{bmatrix} -0.45 \\ 0.13 \\ 0.32 \\ 0.32 \\ -0.22 \\ -0.10 \\ -0.10 \\ 0.55 \\ -0.45 \end{bmatrix}, \begin{bmatrix} 0.18 \\ -0.18 \\ 0.01 \\ 0.01 \\ 0.51 \\ -0.51 \\ -0.51 \\ 0.34 \\ 0.18 \end{bmatrix} \right\},$$

and so eigenvalue $\lambda = -3$ has multiplicity three.

- Corresponding to eigenvalue $\lambda = -0.70$ there are two eigenvectors computed by MATLAB/Octave. These two eigenvectors are orthogonal (you should check). Thus the eigenspace

$$\mathbb{E}_{-0.70} = \operatorname{span}\left\{ \begin{bmatrix} -0.40 \\ -0.58 \\ -0.36 \\ 0.14 \\ 0.36 \\ 0.33 \\ 0.25 \\ 0.22 \\ 0.03 \end{bmatrix}, \begin{bmatrix} 0.06 \\ -0.08 \\ -0.17 \\ -0.37 \\ -0.46 \\ -0.23 \\ 0.31 \\ 0.55 \\ 0.40 \end{bmatrix} \right\},$$

and so eigenvalue $\lambda = -0.70$ has multiplicity two.

- Lastly, corresponding to eigenvalue $\lambda = 0$ are eigenvectors $\boldsymbol{x} \propto (0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33, 0.33)$; that is, the eigenspace $\mathbb{E}_0 = \operatorname{span}\{(1,1,1,1,1,1,1,1,1)\}$, and so eigenvalue $\lambda = 0$ has multiplicity one.

Then check `A*V-V*D` is zero (2 d.p.),

```
ans =
  0.00  0.00  0.00  0.00  0.00 -0.00 -0.00  0.00 -0.00
  0.00  0.00 -0.00  0.00 -0.00  0.00 -0.00 -0.00  0.00
  0.00  0.00  0.00  0.00 -0.00 -0.00 -0.00  0.00 -0.00
 -0.00  0.00 -0.00 -0.00  0.00 -0.00  0.00 -0.00  0.00
 -0.00 -0.00  0.00 -0.00  0.00 -0.00 -0.00  0.00 -0.00
  0.00  0.00  0.00 -0.00  0.00  0.00 -0.00  0.00  0.00
 -0.00 -0.00  0.00 -0.00 -0.00  0.00 -0.00 -0.00 -0.00
  0.00  0.00  0.00 -0.00  0.00 -0.00  0.00 -0.00  0.00
  0.00  0.00  0.00  0.00 -0.00 -0.00  0.00 -0.00 -0.00
```

and confirm V is orthogonal by checking `V'*V` is the identity (2 d.p.)

```
ans =
  1.00 -0.00 -0.00  0.00  0.00  0.00 -0.00  0.00  0.00
 -0.00  1.00 -0.00 -0.00  0.00  0.00 -0.00 -0.00 -0.00
 -0.00 -0.00  1.00 -0.00  0.00  0.00 -0.00  0.00  0.00
  0.00 -0.00 -0.00  1.00  0.00 -0.00 -0.00  0.00 -0.00
```

v0.2a

```
 0.00  0.00  0.00  0.00  1.00  0.00 -0.00  0.00 -0.00
 0.00  0.00  0.00 -0.00  0.00  1.00 -0.00  0.00 -0.00
-0.00 -0.00 -0.00 -0.00 -0.00 -0.00  1.00 -0.00  0.00
 0.00 -0.00  0.00  0.00  0.00  0.00 -0.00  1.00  0.00
 0.00 -0.00  0.00 -0.00 -0.00 -0.00  0.00  0.00  1.00
```

■

Challenge: find the two smallest connected networks that have different connectivity and yet the same eigenvalues (unit strength connections).

In 1966 Mark Kac asked "Can one hear the shape of the drum?" That is, from just knowing the eigenvalues of a network such as the one in Example 4.1.15, can one infer the connectivity of the network? The question for 2D drums was answered "no" in 1992 by Gordon, Webb and Wolpert who constructed two different shaped 2D drums which have the same set of frequencies of oscillation: that is, the same set of eigenvalues.

Why write "the computation may give" in Example 4.1.15? The reason is associated with the duplicated eigenvalues. What is important is the eigenspace. When an eigenvalue of a symmetric matrix is duplicated (or triplicated) 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: those given in Example 4.1.15 are just one possibility for each eigenspace.

Theorem 4.1.16. *Consider any $n \times n$ matrix A (not just symmetric). Then $\lambda_1, \lambda_2, \dots, \lambda_m$ are eigenvalues of A with corresponding eigenvectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_m$, for some m (commonly $m = n$), iff $AV = VD$ for diagonal matrix $D = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_m)$ and $n \times m$ matrix $V = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_m \end{bmatrix}$ for non-zero $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_m$.*

Proof. From Definition 4.1.1, λ_j are eigenvalues and non-zero $\boldsymbol{v}_j$ are eigenvectors iff $A\boldsymbol{v}_j = \lambda_j \boldsymbol{v}_j$. Form into the matrix equation

$$\begin{aligned}
& \begin{bmatrix} A\boldsymbol{v}_1 & A\boldsymbol{v}_2 & \cdots & A\boldsymbol{v}_m \end{bmatrix} = \begin{bmatrix} \lambda_1\boldsymbol{v}_1 & \lambda_2\boldsymbol{v}_2 & \cdots & \lambda_m\boldsymbol{v}_m \end{bmatrix} \\
\iff\ & A \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_m \end{bmatrix} \\
& = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_m \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_m \end{bmatrix} \\
\iff\ & AV = VD\,.
\end{aligned}$$

□

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Example 4.1.17. Use MATLAB/Octave to compute eigenvectors and the eigenvalues of (symmetric) matrix

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

Confirm $AV = VD$ for the computed matrices.

Solution: First compute

```
A=[2 2 -2 0
  2 -1 -2 -3
 -2 -2 4 0
  0 -3 0 1]
[V,D]=eig(A)
```

The output is (2 d.p.)

```
V =
  -0.23   0.83   0.08   0.50
   0.82   0.01  -0.40   0.42
   0.15   0.52  -0.42  -0.72
   0.51   0.20   0.81  -0.23
D =
  -3.80      0      0      0
      0   0.77      0      0
      0      0   2.50      0
      0      0      0   6.53
```

Hence the eigenvalues are (2 d.p.) $\lambda_1 = -3.80$, $\lambda_2 = 0.77$, $\lambda_3 = 2.50$ and $\lambda_4 = 6.53$. From the columns of `V`, corresponding eigenvectors are (2 d.p.) $\boldsymbol{v}_1 \propto (-0.23, 0.82, 0.15, 0.51)$, $\boldsymbol{v}_2 \propto (0.83, 0.01, 0.52, 0.20)$, $\boldsymbol{v}_3 \propto (0.08, -0.40, -0.42, 0.81)$, and $\boldsymbol{v}_4 \propto (0.50, 0.42, -0.72, -0.23)$. Then confirm `A*V-V*D` is zero:

```
ans =
   0.00   0.00   0.00   0.00
   0.00   0.00  -0.00   0.00
  -0.00   0.00  -0.00  -0.00
  -0.00  -0.00  -0.00   0.00
```

■

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 A is not invertible iff $\det A = 0$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

- 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 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:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{matrix} a & b \\ d & e \\ g & h \end{matrix} \tag{4.1}$$

Chapter 6 extends the determinant to any size matrix, and explores more useful properties, but for now this is the information we need on determinants.

For hand calculation on small matrices the key is the following. By Definition 4.1.1 eigenvalues and eigenvectors are determined from $A\boldsymbol{x} = \lambda\boldsymbol{x}$. Rearranging, this equation is equivalent to $(A - \lambda I)\boldsymbol{x} = \boldsymbol{0}$. Recall 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 4.1.18 (eigenvalues and eigenvectors)**.** *To find by hand eigenvalues and eigenvectors of a (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 $(A - \lambda I)\boldsymbol{x} = \boldsymbol{0}$ to find the 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 symmetric matrices. Further, this chapter uses it only as a convenient method to illustrate properties by some hand calculation. None of the beautiful theorems of the next Section 4.2 for symmetric matrices are based upon this procedure.

Example 4.1.19. Use Procedure 4.1.18 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 4.1.3 and 4.1.4).

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Solution: Follow the two steps of Procedure 4.1.18.

(a) Solve $\det(A-\lambda I)=0$ for the eigenvalues λ. Using (4.1),

$$\det(A-\lambda I)=\begin{vmatrix}1-\lambda & -\frac{1}{2}\\ -\frac{1}{2} & 1-\lambda\end{vmatrix}=(1-\lambda)^2-\tfrac{1}{4}=0\,.$$

That is, $(\lambda-1)^2=\frac{1}{4}$. Taking account of both square roots this quadratic gives $\lambda-1=\pm\frac{1}{2}$; that is, $\lambda=1\pm\frac{1}{2}=\frac{1}{2},\frac{3}{2}$ are the only two eigenvalues.

(b) Consider the two eigenvalues in turn.

i. For eigenvalue $\lambda=\frac{1}{2}$ solve $(A-\lambda I)\boldsymbol{x}=\mathbf{0}$. That is,

$$(A-\tfrac{1}{2}I)\boldsymbol{x}=\begin{bmatrix}1-\frac{1}{2} & -\frac{1}{2}\\ -\frac{1}{2} & 1-\frac{1}{2}\end{bmatrix}\boldsymbol{x}=\begin{bmatrix}\frac{1}{2} & -\frac{1}{2}\\ -\frac{1}{2} & \frac{1}{2}\end{bmatrix}\boldsymbol{x}=\mathbf{0}\,.$$

The first component of this system says $x_1-x_2=0$; that is, $x_2=x_1$. The second component of this system says $-x_1+x_2=0$; that is, $x_2=x_1$ (the same). So a general solution for a corresponding eigenvector is $\boldsymbol{x}=(1,1)t$ for any nonzero t. That is, the eigenspace $\mathbb{E}_{1/2}=\operatorname{span}\{(1,1)\}$.

ii. For eigenvalue $\lambda=\frac{3}{2}$ solve $(A-\lambda I)\boldsymbol{x}=\mathbf{0}$. That is,

$$(A-\tfrac{3}{2}I)\boldsymbol{x}=\begin{bmatrix}1-\frac{3}{2} & -\frac{1}{2}\\ -\frac{1}{2} & 1-\frac{3}{2}\end{bmatrix}\boldsymbol{x}=\begin{bmatrix}-\frac{1}{2} & -\frac{1}{2}\\ -\frac{1}{2} & -\frac{1}{2}\end{bmatrix}\boldsymbol{x}=\mathbf{0}\,.$$

The first component of this system says $x_1+x_2=0$, as does the second component; that is, $x_2=-x_1$. So a general solution for a corresponding eigenvector is $\boldsymbol{x}=(1,-1)t$ for any nonzero t. That is, the eigenspace $\mathbb{E}_{3/2}=\operatorname{span}\{(1,-1)\}$.

■

Example 4.1.20. Use the determinant to confirm that $\lambda=0,1,3$ are the only eigenvalues of the matrix

$$A=\begin{bmatrix}1 & -1 & 0\\ -1 & 2 & -1\\ 0 & -1 & 1\end{bmatrix}.$$

(Example 4.1.8 already found the eigenspaces corresponding to these three eigenvalues.)

Solution: To find all eigenvalues, find all solutions of the characteristic equation $\det(A-\lambda I)=0$. Using (4.1),

$$\det(A-\lambda I)$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$
\begin{aligned}
&= \begin{vmatrix} 1-\lambda & -1 & 0 \\ -1 & 2-\lambda & -1 \\ 0 & -1 & 1-\lambda \end{vmatrix} \\
&= (1-\lambda)^2(2-\lambda) + 0 + 0 - 0 - (1-\lambda) - (1-\lambda) \\
&= (1-\lambda)\left[(1-\lambda)(2-\lambda) - 2\right] \\
&= (1-\lambda)\left[2 - 3\lambda + \lambda^2 - 2\right] \\
&= (1-\lambda)\left[-3\lambda + \lambda^2\right] \\
&= (1-\lambda)(-3+\lambda)\lambda\,.
\end{aligned}
$$

So the characteristic equation is $(1-\lambda)(-3+\lambda)\lambda = 0$ whose only solutions are the three eigenvalues $\lambda = 0, 1, 3$ as previously identified.

■

Example 4.1.21. Use Procedure 4.1.18 to find all eigenvalues and the corresponding eigenspaces of the symmetric matrix

$$
A = \begin{bmatrix} -2 & 0 & -6 \\ 0 & 4 & 6 \\ -6 & 6 & -9 \end{bmatrix}.
$$

Solution: Follow the steps of Procedure 4.1.18.

(a) Solve $\det(A - \lambda I) = 0$ for the eigenvalues. Using (4.1),

$$
\begin{aligned}
&\det(A - \lambda I) \\
&= \begin{vmatrix} -2-\lambda & 0 & -6 \\ 0 & 4-\lambda & 6 \\ -6 & 6 & -9-\lambda \end{vmatrix} \\
&= (-2-\lambda)(4-\lambda)(-9-\lambda) + 0\cdot 6\cdot(-6) + (-6)\cdot 0\cdot 6 \\
&\quad - (-6)(4-\lambda)(-6) - (-2-\lambda)\cdot 6\cdot 6 - 0\cdot 0\cdot(-9-\lambda) \\
&= (2+\lambda)(4-\lambda)(9+\lambda) + 36(-4+\lambda) + 36(2+\lambda) \\
&= -\lambda^3 - 7\lambda^2 + 98\lambda \\
&= -\lambda(\lambda^2 - 7\lambda + 98) \\
&= -\lambda(\lambda - 14)(\lambda + 7).
\end{aligned}
$$

This determinant is zero only for the three eigenvalues $\lambda = 0, -7, 14$.

(b) Consider the three eigenvalues in turn.

i. For eigenvalue $\lambda = 0$ solve $(A - \lambda I)\boldsymbol{v} = \boldsymbol{0}$. That is,

$$
(A - 0I)\boldsymbol{v} = \begin{bmatrix} -2 & 0 & -6 \\ 0 & 4 & 6 \\ -6 & 6 & -9 \end{bmatrix} \boldsymbol{v}
$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$= \begin{bmatrix} -2v_1 - 6v_3 \\ 4v_2 + 6v_3 \\ -6v_1 + 6v_2 - 9v_3 \end{bmatrix} = \mathbf{0}\,.$$

The first row says $v_1 = -3v_3$, the second row says $v_2 = -\frac{3}{2}v_3$. Substituting these into the left-hand side of the third row gives $-6v_1+6v_2-9v_3 = 18v_3-9v_3-9v_3 = 0$ for all v_3 which confirms there are non-zero solutions to form eigenvectors. Eigenvectors may be written in the form $\boldsymbol{v} = (-3v_3, -\frac{3}{2}v_3, v_3)$; that is, the eigenspace $\mathbb{E}_0 = \operatorname{span}\{(-6,-3,2)\}$.

ii. For eigenvalue $\lambda = 7$ solve $(A - \lambda I)\boldsymbol{v} = \mathbf{0}$. That is,

$$(A - 7I)\boldsymbol{v} = \begin{bmatrix} -9 & 0 & -6 \\ 0 & -3 & 6 \\ -6 & 6 & -16 \end{bmatrix} \boldsymbol{v}$$
$$= \begin{bmatrix} -9v_1 - 6v_3 \\ -3v_2 + 6v_3 \\ -6v_1 + 6v_2 - 16v_3 \end{bmatrix} = \mathbf{0}\,.$$

The first row says $v_1 = -\frac{2}{3}v_3$, the second row says $v_2 = 2v_3$. Substituting these into the left-hand side of the third row gives $-6v_1+6v_2-16v_3 = 4v_3+12v_3-16v_3 = 0$ for all v_3 which confirms there are non-zero solutions to form eigenvectors. Eigenvectors may be written in the form $\boldsymbol{v} = (-\frac{2}{3}v_3, 2v_3, v_3)$; that is, the eigenspace $\mathbb{E}_7 = \operatorname{span}\{(-2,6,3)\}$.

iii. For eigenvalue $\lambda = -14$ solve $(A - \lambda I)\boldsymbol{v} = \mathbf{0}$. That is,

$$(A + 14I)\boldsymbol{v} = \begin{bmatrix} 12 & 0 & -6 \\ 0 & 18 & 6 \\ -6 & 6 & 5 \end{bmatrix} \boldsymbol{v}$$
$$= \begin{bmatrix} 12v_1 - 6v_3 \\ 18v_2 + 6v_3 \\ -6v_1 + 6v_2 + 5v_3 \end{bmatrix} = \mathbf{0}\,.$$

The first row says $v_1 = \frac{1}{2}v_3$, the second row says $v_2 = -\frac{1}{3}v_3$. Substituting these into the left-hand side of the third row gives $-6v_1+6v_2+5v_3 = -3v_3-2v_3+5v_3 = 0$ for all v_3 which confirms there are non-zero solutions to form eigenvectors. Eigenvectors may be written in the form $\boldsymbol{v} = (\frac{1}{2}v_3, -\frac{1}{3}v_3, v_3)$; that is, the eigenspace $\mathbb{E}_{-14} = \operatorname{span}\{(3,-2,6)\}$.

■

General matrices may have complex valued eigenvalues and eigenvectors, as seen in the next example, and for good reasons in some

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

applications. One of the key results of the next Section 4.2 is to prove that real symmetric matrices always have real eigenvalues and eigenvectors. There are many applications where this reality is crucial.

This example aims to recall basic properties of complex numbers as a prelude to the proof of reality of eigenvalues for symmetric matrices.

Example 4.1.22. Find the eigenvalues and a corresponding eigenvector for the non-symmetric matrix $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$.

Solution: (a) Solve $\det(A - \lambda I) = 0$ for the eigenvalues λ. Using (4.1),

$$\det(A - \lambda I) = \begin{vmatrix} -\lambda & 1 \\ -1 & -\lambda \end{vmatrix} = \lambda^2 + 1 = 0\,.$$

That is, $\lambda^2 = -1$. Taking square roots we find there are two complex eigenvalues $\lambda = \pm\sqrt{-1} = \pm i$. Despite the appearance of complex numbers, all our arithmetic, algebra and properties continue to hold. Thus we proceed to find complex valued eigenvectors.

(b) Consider the two eigenvalues in turn.

i. For eigenvalue $\lambda = i$ solve $(A - \lambda I)\boldsymbol{x} = \boldsymbol{0}$. That is,

$$(A - iI)\boldsymbol{x} = \begin{bmatrix} -i & 1 \\ -1 & -i \end{bmatrix} \boldsymbol{x} = \boldsymbol{0}\,.$$

The first component of this system says $-ix_1 + x_2 = 0$; that is, $x_2 = ix_1$. The second component of this system says $-x_1 - ix_2 = 0$; that is, $x_2 = ix_1$ (the same). So a general corresponding eigenvector is $\boldsymbol{x} = (1, i)t$ for any nonzero t.

ii. For eigenvalue $\lambda = -i$ solve $(A - \lambda I)\boldsymbol{x} = \boldsymbol{0}$. That is,

$$(A + iI)\boldsymbol{x} = \begin{bmatrix} i & 1 \\ -1 & i \end{bmatrix} \boldsymbol{x} = \boldsymbol{0}\,.$$

The first component of this system says $ix_1 + x_2 = 0$; that is, $x_2 = -ix_1$. The second component of this system says $-x_1 + ix_2 = 0$; that is, $x_2 = -ix_1$ (the same). So a general corresponding eigenvector is $\boldsymbol{x} = (1, -i)t$ for any nonzero t.

■

Example 4.1.22 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 $y_2(t) = dy_1/dt$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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 eigenvalues are $\lambda = \pm\sqrt{-1} = \pm i$. Such complex eigenvalues represent oscillations in time t since, for example, $e^{\lambda t} = e^{it} = \cos t + i \sin t$ by Euler's formula.

4.1.2 Exercises

Exercise 4.1.1. Each plot below shows (unit) vectors $\boldsymbol{x}$ (blue), and for some matrix the corresponding vectors $A\boldsymbol{x}$ (red) adjoined. Estimate which directions $\boldsymbol{x}$ are eigenvectors of matrix A, and for each eigenvector estimate the corresponding eigenvalue.

(a)

(b)

(c)

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(d)

(e)

(f)

(g)

(h)

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(i)

Exercise 4.1.2. In each case use the matrix-vector product to determine which of the given vectors are eigenvectors of the given matrix? and for each eigenvector what is the corresponding eigenvalue?

(a) $\begin{bmatrix} 6 & 2 \\ 3 & 1 \end{bmatrix}$, $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 1 \\ -3 \end{bmatrix}$, $\begin{bmatrix} 3 \\ -2 \end{bmatrix}$, $\begin{bmatrix} -\frac{1}{2} \\ -\frac{1}{4} \end{bmatrix}$, $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 2 \\ -3 \end{bmatrix}$

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

(c) $\begin{bmatrix} 3 & 3 & 4 \\ 0 & -4 & 0 \\ -2 & -1 & -6 \end{bmatrix}$, $\begin{bmatrix} 4 \\ 0 \\ -1 \end{bmatrix}$, $\begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}$, $\begin{bmatrix} -2 \\ 6 \\ -1 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $\begin{bmatrix} \frac{1}{3} \\ -1 \\ \frac{1}{6} \end{bmatrix}$, $\begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$

(d) $\begin{bmatrix} 3 & 0 & 0 \\ 2 & -5 & -3 \\ -4 & -2 & 0 \end{bmatrix}$, $\begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$, $\begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix}$, $\begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ -1 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ -1 \\ -\frac{1}{3} \end{bmatrix}$

(e) $\begin{bmatrix} -2 & 0 & 0 & 0 \\ 1 & -2 & 3 & -2 \\ -5 & -3 & 4 & -2 \\ -5 & 1 & -1 & 0 \end{bmatrix}$, $\begin{bmatrix} 2 \\ 0 \\ 4 \\ 7 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 2 \\ 2 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 1 \\ -2 \\ 2 \\ -1 \end{bmatrix}$, $\begin{bmatrix} 0 \\ 2 \\ 2 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 0 \\ -1 \\ -1 \\ 0 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 0 \\ 2 \\ 1 \end{bmatrix}$

(f) $\begin{bmatrix} 5 & -4 & -1 & 0 \\ 1 & 0 & -1 & 0 \\ -5 & 8 & 2 & 5 \\ 4 & -4 & 1 & 1 \end{bmatrix}$, $\begin{bmatrix} 2 \\ 0 \\ 4 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 1 \\ 1 \\ -1 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 1 \\ 0 \\ -\frac{3}{5} \end{bmatrix}$, $\begin{bmatrix} 1 \\ -\frac{1}{2} \\ 3 \\ 3 \end{bmatrix}$, $\begin{bmatrix} 0 \\ \frac{1}{2} \\ 1 \\ 1 \end{bmatrix}$, $\begin{bmatrix} -1 \\ -1 \\ 2 \\ 1 \end{bmatrix}$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Exercise 4.1.3. Use MATLAB/Octave function `eig()` to determine the eigenvalues and corresponding eigenspaces of the following symmetric matrices.

(a) $\begin{bmatrix} -1 & -\frac{4}{5} & \frac{12}{5} & -2 \\ -\frac{4}{5} & \frac{1}{5} & -\frac{8}{5} & -\frac{12}{5} \\ \frac{12}{5} & -\frac{8}{5} & -\frac{11}{5} & -\frac{14}{5} \\ -2 & -\frac{12}{5} & -\frac{14}{5} & 2 \end{bmatrix}$

(b) $\begin{bmatrix} \frac{7}{5} & -\frac{2}{5} & 0 & -\frac{2}{5} \\ -\frac{2}{5} & 2 & \frac{2}{5} & 0 \\ 0 & \frac{2}{5} & \frac{13}{5} & \frac{2}{5} \\ -\frac{2}{5} & 0 & \frac{2}{5} & 2 \end{bmatrix}$

(c) $\begin{bmatrix} \frac{30}{7} & \frac{16}{7} & \frac{16}{7} & \frac{4}{7} \\ \frac{16}{7} & \frac{30}{7} & \frac{16}{7} & \frac{4}{7} \\ \frac{16}{7} & \frac{16}{7} & \frac{30}{7} & \frac{4}{7} \\ \frac{4}{7} & \frac{4}{7} & \frac{4}{7} & \frac{15}{7} \end{bmatrix}$

(d) $\begin{bmatrix} -\frac{36}{7} & -\frac{8}{7} & \frac{20}{7} & \frac{12}{7} \\ -\frac{8}{7} & -\frac{6}{7} & -\frac{12}{7} & \frac{20}{7} \\ \frac{20}{7} & -\frac{12}{7} & -3 & -\frac{32}{7} \\ \frac{12}{7} & \frac{20}{7} & -\frac{32}{7} & -3 \end{bmatrix}$

(e) $\begin{bmatrix} -2.6 & -2.7 & 5.2 & 2.1 \\ -2.7 & 4.6 & 9.9 & 5.2 \\ 5.2 & 9.9 & -2.6 & 2.7 \\ 2.1 & 5.2 & 2.7 & 4.6 \end{bmatrix}$

(f) $\begin{bmatrix} 1.4 & -7.1 & -0.7 & 6.2 \\ -7.1 & -1.0 & -2.2 & -2.5 \\ -0.7 & -2.2 & -3.4 & -4.1 \\ 6.2 & -2.5 & -4.1 & -1.0 \end{bmatrix}$

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

(h) $\begin{bmatrix} 5 & -1 & 3 & 0 & 0 \\ -1 & -1 & 0 & 0 & -2 \\ 3 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & -2 & 1 & 0 & -1 \end{bmatrix}$

Exercise 4.1.4. For each of the given symmetric matrices, determine all eigenvalues by finding and solving the characteristic equation of the matrix.

(a) $\begin{bmatrix} 2 & 3 \\ 3 & 2 \end{bmatrix}$

(b) $\begin{bmatrix} 6 & \frac{11}{2} \\ \frac{11}{2} & 6 \end{bmatrix}$

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

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

(e) $\begin{bmatrix} 5 & -4 \\ -4 & -1 \end{bmatrix}$

(f) $\begin{bmatrix} -2 & \frac{9}{2} \\ \frac{9}{2} & 10 \end{bmatrix}$

(g) $\begin{bmatrix} 6 & 0 & -4 \\ 0 & 6 & 3 \\ -4 & 3 & 6 \end{bmatrix}$

(h) $\begin{bmatrix} -2 & 4 & 6 \\ 4 & 0 & 4 \\ 6 & 4 & -2 \end{bmatrix}$

(i) $\begin{bmatrix} 2 & -3 & -3 \\ -3 & 2 & -3 \\ -3 & -3 & 2 \end{bmatrix}$

(j) $\begin{bmatrix} 4 & -4 & 3 \\ -4 & -2 & 6 \\ 3 & 6 & -8 \end{bmatrix}$

(k) $\begin{bmatrix} 8 & 4 & 2 \\ 4 & 0 & 0 \\ 2 & 0 & 0 \end{bmatrix}$

(l) $\begin{bmatrix} 0 & 0 & -3 \\ 0 & 2 & 0 \\ -3 & 0 & 0 \end{bmatrix}$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Example 4.1.23. For each symmetric matrix, find the eigenspace of the given 'eigenvalues' by hand solution of linear equations, or determine from your solution that the given value cannot be an eigenvalue.

(a) $\begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$, 4, −2

(b) $\begin{bmatrix} 4 & -2 \\ -2 & 7 \end{bmatrix}$, 3, 6

(c) $\begin{bmatrix} -7 & 0 & 2 \\ 0 & -7 & -2 \\ 2 & -2 & -0 \end{bmatrix}$, −8, −7, 1

(d) $\begin{bmatrix} 0 & 6 & -3 \\ 6 & 0 & 7 \\ -3 & 7 & 3 \end{bmatrix}$, −6, 4, 9

(e) $\begin{bmatrix} 0 & -4 & 2 \\ -4 & 1 & -0 \\ 2 & -0 & 1 \end{bmatrix}$, −4, 1, 2

(f) $\begin{bmatrix} 7 & -4 & -2 \\ -4 & 9 & -4 \\ -2 & -4 & 7 \end{bmatrix}$, 1, 4, 9

■

Example 4.1.24. For each symmetric matrix, find by hand all eigenvalues and an orthogonal basis for the corresponding eigenspace. What is the multiplicity of each eigenvalue?

(a) $\begin{bmatrix} -8 & 3 \\ 3 & 0 \end{bmatrix}$

(b) $\begin{bmatrix} 6 & -5 \\ -5 & 6 \end{bmatrix}$

(c) $\begin{bmatrix} -2 & -2 \\ -2 & -5 \end{bmatrix}$

(d) $\begin{bmatrix} 2 & -3 \\ -3 & -6 \end{bmatrix}$

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

(f) $\begin{bmatrix} -5 & 2 & -2 \\ 2 & -2 & 1 \\ -2 & 1 & -10 \end{bmatrix}$

(g) $\begin{bmatrix} 11 & 4 & -2 \\ 4 & 5 & 4 \\ -2 & 4 & 11 \end{bmatrix}$

(h) $\begin{bmatrix} -7 & 2 & 2 \\ 2 & -6 & 0 \\ 2 & 0 & -8 \end{bmatrix}$

(i) $\begin{bmatrix} 6 & 10 & -5 \\ 10 & 13 & -2 \\ -5 & -2 & -2 \end{bmatrix}$

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

■

v0.2a

Exercise 4.1.5 (Seven Bridges of Königsberg)**.** The marginal picture shows the abstract graph of the seven bridges of Königsberg during the time of Euler: the small disc nodes represent the islands of Königsberg; the lines between represent the seven different bridges. This abstract graph is famous for its role in founding the theory of such networks, but this exercise addresses an aspect relevant to well used web search software. Number the nodes from 1 to 4. Form the 4×4 symmetric matrix of the number of lines from each node to the other nodes (and zero for the number of lines from a node to itself). Use MATLAB/Octave function `eig()` to find the eigenvalues and eigenvectors for this matrix. Analogous to well known web search software, identify the largest eigenvalue and a corresponding eigenvector: then rank the importance of each node in order of the size of the component in the corresponding eigenvector.

Exercise 4.1.6. For each of the following networks:

- label the nodes;
- construct the symmetric adjacency matrix A such that a_{ij} is one if node i is linked to node j, and a_{ij} is zero otherwise (and zero on the diagonal);
- in MATLAB/Octave use `eig()` to find all eigenvalues and eigenvectors;
- rank the 'importance' of the nodes from the magnitude of their component in the eigenvector of the largest (most positive) eigenvalue.

Although a well known web search engine does much the same thing for web pages, it uses an approximate iterative algorithm more suited to the mind-boggingly vast size of the internet.

(a)

(b)

(c)

(d)

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

4.2 Beautiful properties for symmetric matrices

Section Contents

4.2.1 Matrix powers maintain eigenvectors 435
4.2.2 Symmetric matrices are orthogonally diagonalisable 441
4.2.3 Change orthonormal basis to classify quadratics 450
Graph quadratic equations 451
Simplify quadratic forms 455
4.2.4 Exercises 457

This section starts with two properties for eigenvalues of general matrices, and then proceeds to the special case of real symmetric matrices which have the beautifully useful properties of always having real eigenvalues and orthogonal eigenvectors.

4.2.1 Matrix powers maintain eigenvectors

Recall that section 3.2 introduced the inverse of a matrix (Definition 3.2.2). The first theorem links an eigenvalue of zero to the non-existence of an inverse and hence to more problematic linear equations.

Theorem 4.2.1. *A square matrix is invertible iff zero is* not *an eigenvalue of the matrix.*

Proof. From Definition 4.1.1, zero is an eigenvalue, $\lambda = 0$, iff $A\boldsymbol{x} = 0\boldsymbol{x}$ has nonzero solutions $\boldsymbol{x}$; that is, iff the homogeneous system $A\boldsymbol{x} = \boldsymbol{0}$ has nonzero solutions $\boldsymbol{x}$. But the Unique Solution Theorems 3.3.26 or 3.4.43 assure us that this occurs iff matrix A is not invertible. Consequently a matrix is invertible iff zero is not an eigenvalue. □

Example 4.2.2.

- The 3×3 matrix of Example 4.1.2 (also 4.1.8, 4.1.13 and 4.1.20) is not invertible as among its eigenvalues of 0, 1 and 3 it has zero as an eigenvalue.
- The plot in the margin shows (unit) vectors $\boldsymbol{x}$ (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 4.1.5.

- The 3×3 diagonal matrix of Example 4.1.10 has eigenvalues of only $-\frac{1}{3}$ and $\frac{3}{2}$. Since zero is not an eigenvalue, the matrix is invertible.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

- The 9×9 matrix of the Sierpinski network in Example 4.1.15 is not invertible as it has zero among its nine eigenvalues.
- The 2×2 matrix of Example 4.1.19 is invertible as its eigenvalues are $\lambda = \frac{1}{2}, \frac{3}{2}$, neither of which are zero. Indeed, the matrix

$$A = \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix}, \quad \text{has inverse } A^{-1} = \begin{bmatrix} \frac{4}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{4}{3} \end{bmatrix}$$

as matrix multiplication confirms.

- The 2×2 non-symmetric matrix of Example 4.1.22 is invertible because zero is not among its eigenvalues of $\lambda = \pm i$. Indeed, the matrix

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad \text{has inverse } A^{-1} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

as matrix multiplication confirms.

■

Example 4.2.3. The next theorem considers eigenvalues and eigenvectors of powers of a matrix. Two examples are the following.

- Recall the matrix $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ has eigenvalues $\lambda = \pm i$. The square of this matrix

$$A^2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$$

is diagonal so its eigenvalues are the diagonal elements (Example 4.1.6), namely the only eigenvalue is -1. Observe that A^2s eigenvalue, $-1 = (\pm i)^2$, is the square of the eigenvalues of A. That the eigenvalues of A^2 are the square of those of A holds generally.

- Also recall matrix

$$A = \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix}, \quad \text{has inverse } A^{-1} = \begin{bmatrix} \frac{4}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{4}{3} \end{bmatrix}.$$

Let's determine the eigenvalues of this inverse. Its characteristic equation (defined in Procedure 4.1.18) 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 these eigenvalues

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

of the inverse are the reciprocals of the eigenvalues $\frac{1}{2}, \frac{3}{2}$ of A. This relation also holds generally.

The marginal pictures illustrates the reciprocal relation graphically: the first picture shows $A\boldsymbol{x}$ for various $\boldsymbol{x}$, the second picture shows $A^{-1}\boldsymbol{x}$. The eigenvector directions are the same for both matrix and inverse. But in those directions where the matrix stretches, the inverse shrinks, and where the matrix shrinks, the inverse stretches. The relationship is obscure in directions which are not eigenvectors.

■

Theorem 4.2.4. *Let A be a square matrix with eigenvalue λ and corresponding eigenvector $\boldsymbol{x}$.*

(a) For any 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 any integer n, λ^n is an eigenvalue of A^n with corresponding eigenvector $\boldsymbol{x}$.

Proof. Consider each property in turn.

4.2.4a. Firstly, the result hold for $n = 1$ by Definition 4.1.1, that $A\boldsymbol{x} = \lambda\boldsymbol{x}$. Secondly, for the case $n = 2$ consider $A^2\boldsymbol{x} = (AA)\boldsymbol{x} = A(A\boldsymbol{x}) = A(\lambda\boldsymbol{x}) = \lambda(A\boldsymbol{x}) = \lambda(\lambda\boldsymbol{x}) = (\lambda^2)\boldsymbol{x}$. Hence by definition 4.1.1 λ^2 is an eigenvalue of A^2 corresponding to eigenvector $\boldsymbol{x}$. Third, use induction to extend to any power: assume the result for $n = k$ (and proceed to prove it for $n = k + 1$). Consider $A^{k+1}\boldsymbol{x} = (A^kA)\boldsymbol{x} = A^k(A\boldsymbol{x}) = A^k(\lambda\boldsymbol{x}) = \lambda(A^k\boldsymbol{x}) = \lambda(\lambda^k\boldsymbol{x}) = \lambda^{k+1}\boldsymbol{x}$. Hence by Definition 4.1.1 λ^{k+1} is an eigenvalue of A^{k+1} corresponding to eigenvector $\boldsymbol{x}$. By induction the property 4.2.4a. holds for all integer $n \geq 1$.

4.2.4b. For invertible A we know none of the eigenvalues are zero: thus $1/\lambda$ exists. Pre-multiply $A\boldsymbol{x} = \lambda\boldsymbol{x}$ by $\frac{1}{\lambda}A^{-1}$ to deduce $\frac{1}{\lambda}A^{-1}A\boldsymbol{x} = \frac{1}{\lambda}A^{-1}\lambda\boldsymbol{x}$, which gives $\frac{1}{\lambda}I\boldsymbol{x} = \frac{1}{\lambda}\lambda A^{-1}\boldsymbol{x}$, that is, $\frac{1}{\lambda}\boldsymbol{x} = A^{-1}\boldsymbol{x}$. Consequently, $1/\lambda$ is an eigenvalue of A^{-1} with corresponding eigenvector $\boldsymbol{x}$.

4.2.4c. Proved by Exercise 4.2.11.

□

v0.2a

Example 4.2.5. Recall from Example 4.1.19 that matrix

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

has eigenvalues $1/2$ and $3/2$ with corresponding eigenvectors $(1,1)$ and $(1,-1)$ respectively. Confirm matrix A^2 has eigenvalues which are these squared, and corresponding to the same eigenvectors.

Solution: Compute

$$A^2 = \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix} = \begin{bmatrix} \frac{5}{4} & -1 \\ -1 & \frac{5}{4} \end{bmatrix}.$$

Then

$$A^2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{4} \\ \frac{1}{4} \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$
$$A^2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} \frac{9}{4} \\ -\frac{9}{4} \end{bmatrix} = \frac{9}{4} \begin{bmatrix} 1 \\ -1 \end{bmatrix},$$

and so A^2 has eigenvalues $1/4 = (1/2)^2$ and $9/4 = (3/2)^2$ with the same corresponding eigenvectors $(1,1)$ and $(1,-1)$ respectively.

■

Example 4.2.6. Given that the matrix

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

has eigenvalues 2, 1 and -1 with corresponding eigenvectors $(1,1,1)$, $(-1,0,1)$ and $(1,-2,1)$ respectively. Confirm matrix A^2 has eigenvalues which are these squared, and corresponding to the same eigenvectors. Given the inverse

$$A^{-1} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

confirm its eigenvalues are the reciprocals of those of A, and for corresponding eigenvectors.

Solution:

- Compute

$$A^2 = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}.$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Then

$$A^2\begin{bmatrix}1\\1\\1\end{bmatrix}=\begin{bmatrix}4\\4\\4\end{bmatrix}=4\begin{bmatrix}1\\1\\1\end{bmatrix},$$
$$A^2\begin{bmatrix}-1\\0\\1\end{bmatrix}=\begin{bmatrix}-1\\0\\1\end{bmatrix}=1\begin{bmatrix}-1\\0\\1\end{bmatrix},$$
$$A^2\begin{bmatrix}1\\-2\\1\end{bmatrix}=\begin{bmatrix}1\\-2\\1\end{bmatrix}=1\begin{bmatrix}1\\-2\\1\end{bmatrix},$$

has eigenvalues $4=2^2$ and $1=(\pm1)^2$ with corresponding eigenvectors $(1,1,1)$, and the pair $(-1,0,1)$ and $(1,-2,1)$. Thus here $\operatorname{span}\{(-1,0,1),(1,-2,1)\}$ is the eigenspace of A^2 corresponding to eigenvalue one.

- For the inverse

$$A^{-1}\begin{bmatrix}1\\1\\1\end{bmatrix}=\begin{bmatrix}\frac12\\\frac12\\\frac12\end{bmatrix}=\frac12\begin{bmatrix}1\\1\\1\end{bmatrix},$$
$$A^{-1}\begin{bmatrix}-1\\0\\1\end{bmatrix}=\begin{bmatrix}-1\\0\\1\end{bmatrix}=1\begin{bmatrix}-1\\0\\1\end{bmatrix},$$
$$A^{-1}\begin{bmatrix}1\\-2\\1\end{bmatrix}=\begin{bmatrix}-1\\2\\-1\end{bmatrix}=(-1)\begin{bmatrix}1\\-2\\1\end{bmatrix},$$

has eigenvalues $1/2$, $1=1/1$ and $-1=1/(-1)$ with corresponding eigenvectors $(1,1,1)$, $(-1,0,1)$ and $(1,-2,1)$.

■

Example 4.2.7 (long term age structure)**.** Recall Example 3.1.8 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 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. Predict what happens after many generations: does the population die out? grow? oscillate?

Solution: Consider what happens after n generations for large n, say $n=10$ or 100. The predicted population is $\boldsymbol{x}^{(n)}=L^n\boldsymbol{x}$;

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

that is, the matrix L^n transforms the current population to that after n generations. The stretching and/or shrinking of matrix L^n is summarised by its eigenvectors and eigenvalues (section 4.1). By Theorem 4.2.4 the eigenvalues of L^n are λ^n in terms of the eigenvalues λ of L. By hand (Procedure 4.1.18), the characteristic equation of L is

$$\begin{aligned}
\det(L-\lambda I) &= \begin{vmatrix} -\lambda & 0 & 4 \\ \frac{1}{2} & -\lambda & 0 \\ 0 & \frac{1}{3} & \frac{1}{3}-\lambda \end{vmatrix} \\
&= \lambda^2(\tfrac{1}{3}-\lambda)+0+\tfrac{2}{3}-0-0-0 \\
&= (1-\lambda)(\lambda^2+\tfrac{2}{3}\lambda+\tfrac{2}{3}) \\
&= (1-\lambda)\big[(\lambda+\tfrac{1}{3})^2+\tfrac{5}{9}\big]=0 \\
\implies \quad & \lambda = 1, (-1\pm i\sqrt{5})/3\,.
\end{aligned}$$

Such complex valued eigenvalues may arise in real applications when the matrix is not symmetric, as here—the next Theorem 4.2.8 proves such complexities do not arise for symmetric matrices.

But the algebra still works with complex eigenvalues (Chapter 7). Here, the eigenvalues of L^n are λ^n (Theorem 4.2.4) namely $1^n = 1$ for all n and $[(-1\pm i\sqrt{5})/3]^n$. Because the absolute value $|(-1\pm i\sqrt{5})/3| = |-1\pm i\sqrt{5}|/3 = \sqrt{1+5}/3 = \sqrt{6}/3 = \sqrt{2/3} = 0.8165$, then the absolute value of $[(-1\pm i\sqrt{5})/3]^n$ is 0.8165^n which becomes negligibly small for large n; for example, $0.8165^{34} \approx 0.001$. Since the eigenvectors of L^n are the same as those of L (Theorem 4.2.4), these negligibly small eigenvalues of L^n imply that any component in the initial population in the direction of the corresponding eigenvectors is shrunk to zero by L^n. For large n, it is only the component in the eigenvector corresponding to eigenvalue $\lambda = 1$ that remains. Find the eigenvector by solving $(L-I)\boldsymbol{x} = \mathbf{0}$, namely

$$\begin{bmatrix} -1 & 0 & 4 \\ \frac{1}{2} & -1 & 0 \\ 0 & \frac{1}{3} & -\frac{2}{3} \end{bmatrix}\boldsymbol{x} = \begin{bmatrix} -x_1+4x_3 \\ \frac{1}{2}x_1 - x_2 \\ \frac{1}{3}x_2 - \frac{2}{3}x_3 \end{bmatrix} = \mathbf{0}\,.$$

The first row gives that $x_1 = 4x_3$, the third row that $x_2 = 2x_3$, and the second row confirms these are correct as $\frac{1}{2}x_1 - x_2 = \frac{1}{2}4x_3 - 2x_3 = 0$. Eigenvectors corresponding to $\lambda = 1$ are then of the form $(4x_3, 2x_3, x_3) = (4,2,1)x_3$. Because the corresponding eigenvalue of $L^n = 1^n = 1$ the component of $\boldsymbol{x}$ in this direction remains in $L^n\boldsymbol{x}$ whereas all other components decay to zero. Thus the model predicts that after many generations the population reaches a steady state of the pups, juveniles, and mature females being in the ratio of $4:2:1$.

■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

4.2.2 Symmetric matrices are orthogonally diagonalisable

General matrices may have complex valued eigenvalues (as in Examples 4.1.22 and 4.2.7): that real symmetric matrices always have real eigenvalues (such as in all matrices of Examples 4.2.5 and 4.2.6) is a special property that often reflects the physical reality of many applications.

To establish the reality of eigenvalues (Theorem 4.2.8), the proof invokes contradiction. The contradiction is to assume a complex valued eigenvalue exists and then prove it cannot. Thus the proof of Theorem 4.2.8 needs to use some complex numbers and some properties of complex numbers. Recall: that any complex number $z = a + bi$ has a complex conjugate $\bar{z} = a - bi$ (denoted by the overbar); a complex number equals its conjugate only if it is real valued (the imaginary part is zero); and properties of complex numbers and operations also hold for complex valued vectors, complex valued matrices, and arithmetic operations with complex matrices and vectors.

Theorem 4.2.8. *Let A be a real symmetric matrix, then the eigenvalues of A are all real.*

Proof. Let λ be any eigenvalue of A with corresponding eigenvector $\boldsymbol{x}$; that is, $A\boldsymbol{x} = \lambda\boldsymbol{x}$ for $\boldsymbol{x} \neq \boldsymbol{0}$. To establish a contradiction: assume the eigenvalue λ is complex valued, and so correspondingly is the eigenvector $\boldsymbol{x}$. First, the complex conjugate $\bar{\lambda}$ must be another eigenvalue of A, corresponding to an eigenvector which is the complex conjugate $\bar{\boldsymbol{x}}$. To see this just take the complex conjugate of both sides of $A\boldsymbol{x} = \lambda\boldsymbol{x}$:

$$\overline{A\boldsymbol{x}} = \overline{\lambda\boldsymbol{x}} \implies \bar{A}\bar{\boldsymbol{x}} = \bar{\lambda}\bar{\boldsymbol{x}} \implies A\bar{\boldsymbol{x}} = \bar{\lambda}\bar{\boldsymbol{x}}$$

as matrix A is real ($\bar{A} = A$). Second, consider $\boldsymbol{x}^{\mathrm{T}}A\bar{\boldsymbol{x}}$ in two ways:

$$\begin{aligned}
\boldsymbol{x}^{\mathrm{T}}A\bar{\boldsymbol{x}} &= \boldsymbol{x}^{\mathrm{T}}(A\bar{\boldsymbol{x}}) = \boldsymbol{x}^{\mathrm{T}}(\bar{A}\bar{\boldsymbol{x}}) \quad \text{(as } A \text{ is real)} \\
&= \boldsymbol{x}^{\mathrm{T}}(\overline{A\boldsymbol{x}}) = \boldsymbol{x}^{\mathrm{T}}(\overline{\lambda\boldsymbol{x}}) = \boldsymbol{x}^{\mathrm{T}}(\bar{\lambda}\bar{\boldsymbol{x}}) = \bar{\lambda}\boldsymbol{x}^{\mathrm{T}}\bar{\boldsymbol{x}}\,; \\
\boldsymbol{x}^{\mathrm{T}}A\bar{\boldsymbol{x}} &= (\boldsymbol{x}^{\mathrm{T}}A)\bar{\boldsymbol{x}} = (A^{\mathrm{T}}\boldsymbol{x})^{\mathrm{T}}\bar{\boldsymbol{x}} = (A\boldsymbol{x})^{\mathrm{T}}\bar{\boldsymbol{x}} \quad \text{(by symmetry)} \\
&= (\lambda\boldsymbol{x})^{\mathrm{T}}\bar{\boldsymbol{x}} = \lambda\boldsymbol{x}^{\mathrm{T}}\bar{\boldsymbol{x}}\,.
\end{aligned}$$

Equating the two ends of this identity gives $\bar{\lambda}\boldsymbol{x}^{\mathrm{T}}\bar{\boldsymbol{x}} = \lambda\boldsymbol{x}^{\mathrm{T}}\bar{\boldsymbol{x}}$. Rearrange to $\bar{\lambda}\boldsymbol{x}^{\mathrm{T}}\bar{\boldsymbol{x}} - \lambda\boldsymbol{x}^{\mathrm{T}}\bar{\boldsymbol{x}} = (\bar{\lambda} - \lambda)\boldsymbol{x}^{\mathrm{T}}\bar{\boldsymbol{x}} = 0$. Because this product is zero, either $\bar{\lambda} - \lambda = 0$ or $\boldsymbol{x}^{\mathrm{T}}\bar{\boldsymbol{x}} = 0$. But we next prove the second is impossible, hence the first must hold; that is, $\bar{\lambda} - \lambda = 0$, equivalently $\bar{\lambda} = \lambda$. Consequently, the eigenvalue λ must be real—it cannot be complex.

Lastly confirm $\boldsymbol{x}^{\mathrm{T}}\bar{\boldsymbol{x}} \neq 0$. The nonzero eigenvector $\boldsymbol{x}$ will be generally complex, say

$$\boldsymbol{x} = (a_1 + b_1 i\,, a_2 + b_2 i\,, \dots, a_n + b_n i)$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$\implies \bar{\boldsymbol{x}} = (a_1 - b_1 i\,, a_2 - b_2 i\,, \dots, a_n - b_n i).$$

Then the product

$$\begin{aligned}
\boldsymbol{x}^{\mathrm{T}}\bar{\boldsymbol{x}} &= \begin{bmatrix} a_1 + b_1 i & a_2 + b_2 i & \cdots & a_n + b_n i \end{bmatrix} \begin{bmatrix} a_1 - b_1 i \\ a_2 - b_2 i \\ \vdots \\ a_n - b_n i \end{bmatrix} \\
&= (a_1 + b_1 i)(a_1 - b_1 i) + (a_2 + b_2 i)(a_2 - b_2 i) \\
&\quad + \cdots + (a_n + b_n i)(a_n - b_n i) \\
&= (a_1^2 + b_1^2) + (a_2^2 + b_2^2) + \cdots + (a_n^2 + b_n^2) \\
&> 0
\end{aligned}$$

since $\boldsymbol{x}$ is an eigenvector which necessarily is nonzero and so at least one term in the sum is positive. □

The other property that we have seen graphically for 2D matrices is that the eigenvectors of symmetric matrices are orthogonal. In 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 4.2.9. Recall Example 4.1.21 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 seen by the dot products

$$\begin{aligned}
(-6, -3, 2) \cdot (-2, 6, 3) &= 12 - 18 + 6 = 0\,, \\
(-2, 6, 3) \cdot (3, -2, 6) &= -6 - 12 + 18 = 0\,, \\
(3, -2, 6) \cdot (-6, -3, 2) &= -18 + 6 + 12 = 0\,.
\end{aligned}$$

■

Theorem 4.2.10. *Let A be a real symmetric matrix, then for every two distinct eigenvalues of A, any corresponding two eigenvectors are orthogonal.*

Proof. Let eigenvalues $\lambda_1 \neq \lambda_2$, and let $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ be corresponding eigenvectors, respectively; that is, $A\boldsymbol{x}_1 = \lambda_1 \boldsymbol{x}_1$ and $A\boldsymbol{x}_2 = \lambda_2 \boldsymbol{x}_2$. Consider $\boldsymbol{x}_1^{\mathrm{T}} A \boldsymbol{x}_2$ in two ways:

$$\boldsymbol{x}_1^{\mathrm{T}} A \boldsymbol{x}_2 = \boldsymbol{x}_1^{\mathrm{T}}(A\boldsymbol{x}_2) = \boldsymbol{x}_1^{\mathrm{T}}(\lambda_2 \boldsymbol{x}_2) = \lambda_2 \boldsymbol{x}_1^{\mathrm{T}} \boldsymbol{x}_2 = \lambda_2 \boldsymbol{x}_1 \cdot \boldsymbol{x}_2\,;$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$\begin{aligned}\boldsymbol{x}_1^{\mathrm{T}}A\boldsymbol{x}_2 &= \boldsymbol{x}_1^{\mathrm{T}}A^{\mathrm{T}}\boldsymbol{x}_2 \quad \text{(as } A \text{ is symmetric)}\\ &= (\boldsymbol{x}_1^{\mathrm{T}}A^{\mathrm{T}})\boldsymbol{x}_2 = (A\boldsymbol{x}_1)^{\mathrm{T}}\boldsymbol{x}_2\\ &= (\lambda_1\boldsymbol{x}_1)^{\mathrm{T}}\boldsymbol{x}_2 = \lambda_1\boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{x}_2 = \lambda_1\boldsymbol{x}_1\cdot\boldsymbol{x}_2\,.\end{aligned}$$

Equating the two ends of this identity gives $\lambda_2\boldsymbol{x}_1\cdot\boldsymbol{x}_2 = \lambda_1\boldsymbol{x}_1\cdot\boldsymbol{x}_2$. Rearrange to $\lambda_2\boldsymbol{x}_1\cdot\boldsymbol{x}_2 - \lambda_1\boldsymbol{x}_1\cdot\boldsymbol{x}_2 = (\lambda_2-\lambda_1)(\boldsymbol{x}_1\cdot\boldsymbol{x}_2) = 0$. Since $\lambda_1 \neq \lambda_2$ it follows that the dot product $\boldsymbol{x}_1\cdot\boldsymbol{x}_2 = 0$. Hence the two eigenvectors are orthogonal. □

Example 4.2.11. The plots below shows (unit) vectors $\boldsymbol{x}$ (blue), and for some matrix A the corresponding vectors $A\boldsymbol{x}$ (red) adjoined. By estimating eigenvectors determine which cases *cannot* be the plot of a real symmetric matrix.

(a)

Solution: Estimate eigenvectors $(0.8\,,0.5)$ and $(-0.5\,,0.8)$ which are orthogonal, so may be a symmetric matrix

(b)

Solution: Estimate eigenvectors $(1\,,0.1)$ and $(1\,,-0.3)$ which are not orthogonal, so cannot be from a symmetric matrix

(c)

Solution: Estimate eigenvectors $(1\,,0.2)$ and $(0.8\,,-0.7)$ which are not orthogonal, so cannot be from a symmetric matrix

(d)

Solution: Estimate eigenvectors $(0.1\,,-1)$ and $(1\,,0.1)$ which are orthogonal, so may be a symmetric matrix

■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Example 4.2.12. By hand find eigenvectors corresponding to the two distinct eigenvalues of the following matrices and confirm that symmetric matrix A has orthogonal eigenvectors, and that non-symmetric matrix B does not:

$$A = \begin{bmatrix} 1 & \frac{3}{2} \\ \frac{3}{2} & -3 \end{bmatrix}; \quad B = \begin{bmatrix} 0 & -3 \\ -2 & 1 \end{bmatrix}.$$

Solution:

- For matrix A, the eigenvalues come from the characteristic equation

$$\begin{aligned} \det(A - \lambda I) &= (1-\lambda)(-3-\lambda) - \tfrac{9}{4} \\ &= \lambda^2 + 2\lambda - \tfrac{21}{4} = (\lambda+1)^2 - \tfrac{25}{4} = 0\,, \end{aligned}$$

  so eigenvalues are $\lambda = -1 \pm \frac{5}{2} = -\frac{7}{2}, \frac{3}{2}$.

  - Corresponding to eigenvalue $\lambda = -7/2$, eigenvectors $\boldsymbol{v}$ satisfy $(A + \frac{7}{2}I)\boldsymbol{v} = \boldsymbol{0}$, that is

$$\begin{bmatrix} \frac{9}{2} & \frac{3}{2} \\ \frac{3}{2} & \frac{1}{2} \end{bmatrix} \boldsymbol{v} = \frac{1}{2} \begin{bmatrix} 9v_1 + 3v_2 \\ 3v_1 + v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

    giving $v_2 = -3v_1$. Eigenvectors must be $\boldsymbol{v} \propto (1, -3)$.

  - Corresponding to eigenvalue $\lambda = 3/2$, eigenvectors $\boldsymbol{v}$ satisfy $(A - \frac{3}{2}I)\boldsymbol{v} = \boldsymbol{0}$, that is

$$\begin{bmatrix} -\frac{1}{2} & \frac{3}{2} \\ \frac{3}{2} & -\frac{9}{2} \end{bmatrix} \boldsymbol{v} = \frac{1}{2} \begin{bmatrix} -v_1 + 3v_2 \\ 3v_1 - 9v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

    giving $v_1 = 3v_2$. Eigenvectors must be $\boldsymbol{v} \propto (3, 1)$.

  The dot product of the two basis eigenvectors is $(1, -3) \cdot (3, 1) = 3 - 3 = 0$ and hence they are orthogonal.

- For matrix B, the eigenvalues come from the characteristic equation

$$\begin{aligned} \det(B - \lambda I) &= (-\lambda)(1-\lambda) - 6 \\ &= \lambda^2 - \lambda - 6 = (\lambda - 3)(\lambda + 2) = 0\,, \end{aligned}$$

  so eigenvalues are $\lambda = 3, -2$.

  - Corresponding to eigenvalue $\lambda = -2$, eigenvectors $\boldsymbol{v}$ satisfy $(B + 2I)\boldsymbol{v} = \boldsymbol{0}$, that is

$$\begin{bmatrix} 2 & -3 \\ -2 & 3 \end{bmatrix} \boldsymbol{v} = \begin{bmatrix} 2v_1 - 3v_2 \\ -2v_1 + 3v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

    giving $v_2 = \frac{2}{3}v_1$. Eigenvectors must be $\boldsymbol{v} \propto (3, 2)$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

– Corresponding to eigenvalue $\lambda = 3$, eigenvectors $\boldsymbol{v}$ satisfy $(B - 3I)\boldsymbol{v} = \boldsymbol{0}$, that is

$$\begin{bmatrix} -3 & -3 \\ -2 & -2 \end{bmatrix} \boldsymbol{v} = \begin{bmatrix} -3v_1 - 3v_2 \\ -2v_1 - 2v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

giving $v_1 = -v_2$. Eigenvectors must be $\boldsymbol{v} \propto (-1, 1)$.

The dot product of the two basis eigenvectors is $(3, 2) \cdot (-1, 1) = -3 + 2 = -1 \neq 0$ and hence they are not orthogonal.

■

Example 4.2.13. Use MATLAB/Octave to compute eigenvectors of the following matrices, and to confirm the eigenvectors are orthogonal for a symmetric matrix.

(a) $\begin{bmatrix} 0 & 3 & 2 & -1 \\ 0 & 3 & 0 & 0 \\ 3 & 0 & -1 & -1 \\ -3 & 1 & 3 & 0 \end{bmatrix}$ (b) $\begin{bmatrix} -6 & 0 & 1 & 1 \\ 0 & 0 & 2 & 2 \\ 1 & 2 & 2 & -1 \\ 1 & 2 & -1 & -1 \end{bmatrix}$

Solution: For each matrix, enter the matrix as say `A`, then execute `[V,D]=eig(A)` to give eigenvectors as the columns of `V`. Then confirm orthogonality of all pairs of eigenvectors by computing `V'*V` and confirming the off-diagonal dot products are zero, or confirm the lack of orthogonality if non-zero. (In the case of repeated eigenvalues, MATLAB/Octave generates an orthonormal basis for the corresponding eigenspace so the returned matrix `V` of eigenvectors is still orthogonal for symmetric `A`.)

(a) The MATLAB/Octave code

```
A=[0 3 2 -1
 0 3 0 0
 3 0 -1 -1
 -3 1 3 0]
[V,D]=eig(A)
V'*V
```

gives the following (2 d.p.)

```
V =
  -0.49  -0.71   0.41   0.74
   0.00   0.00   0.00   0.34
   0.32  -0.71   0.41   0.57
  -0.81  -0.00   0.82  -0.06
D =
  -3.00      0      0      0
      0   2.00      0      0
      0      0   0.00      0
      0      0      0   3.00
```

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

so eigenvectors corresponding to the four distinct eigenvalues are $(-0.49,0,0.32,-0.81)$, $(-0.71,0,-0.71,0)$, $(0.41,0,0.41,0.82)$ and $(0.74\,,0.34\,,0.57\,,-0.6)$. Then `V'*V` is (2 d.p.)

```
   1.00   0.11  -0.73  -0.13
   0.11   1.00  -0.58  -0.93
  -0.73  -0.58   1.00   0.49
  -0.13  -0.93   0.49   1.00
```

As the off-diagonal elements are non-zero, the pairs of dot products are non-zero indicating the column vectors are not orthogonal. Hence the matrix `A` cannot be symmetric.

(b) The MATLAB/Octave code

```
B=[-6 0 1 1
 0 0 2 2
 1 2 2 -1
 1 2 -1 -1]
[V,D]=eig(B)
V'*V
```

gives the following (2 d.p.)

```
V =
   0.94   0.32  -0.04  -0.10
   0.13  -0.63  -0.53  -0.55
  -0.17   0.32   0.43  -0.83
  -0.25   0.63  -0.73  -0.08
D =
  -6.45      0      0      0
      0  -3.00      0      0
      0      0   1.11      0
      0      0      0   3.34
```

so eigenvectors corresponding to the four distinct eigenvalues are $(0.94\,,0.13\,,-0.17\,,-0.25)$, $(0.32\,,-0.63\,,0.32\,,0.63)$, $(-0.04\,,-0.53\,,0.43\,,-0.73)$ and $(-0.10\,,-0.55\,,-0.83\,,-0.08)$. Then `V'*V` is (2 d.p.)

```
   1.00  -0.00  -0.00  -0.00
  -0.00   1.00   0.00  -0.00
  -0.00   0.00   1.00  -0.00
  -0.00  -0.00  -0.00   1.00
```

As the off-diagonal elements are zero, the pairs of dot products are zero indicating the column vectors are orthogonal. The symmetry of this matrix `B` requires such orthogonality.

■

Recall that when we find eigenvalues by hand for 2×2 or 3×3 matri-

v0.2a

ces we solve a quadratic or cubic characteristic equation, respectively. Thus we get 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 values.

Theorem 4.2.14. *Let A be an $n \times n$ real symmetric matrix, then A has at most n distinct eigenvalues.*

Proof. Invoke the pigeonhole principle and contradiction. Assume there are more than n distinct eigenvalues, then there would be more than n eigenvectors corresponding to distinct eigenvalues. Theorem 4.2.10 asserts all such eigenvectors are orthogonal. But there cannot be more than n vectors in an orthogonal set in $\mathbb{R}^n$ (Theorem 1.3.25). Hence the assumption is wrong and there cannot be any more than n distinct eigenvalues. □

The previous theorem establishes there are at most n distinct eigenvalues (for symmetric matrices but it is true more generally). 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, $A = USV^{\mathrm{T}}$, 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 4.2.15. *Let A be an $n \times n$ real symmetric matrix with* SVD *$A = USV^{\mathrm{T}}$. 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).*

If non-zero singular values are duplicated, then one can always choose an SVD so the result of this theorem still holds. However, the proof is too involved to give here.

This proof modifies parts of the proof of the SVD Theorem 3.3.6 to the specific case of a symmetric matrix.

Proof. First, for any zero singular value, $\sigma_j = 0$, then the result is immediate as from the SVD, $AV = US$ the jth column gives $A\boldsymbol{v}_j = 0\boldsymbol{u}_j = \boldsymbol{0} = 0\boldsymbol{v}_j$ for the nonzero $\boldsymbol{v}_j$.

Second, for the singular values $\sigma_j > 0$ use a form of induction allied with a contradiction to prove $\boldsymbol{u}_j = \pm\boldsymbol{v}_j$. By contradiction, suppose $\boldsymbol{u}_1 \neq \pm\boldsymbol{v}_1$, then we can write $\boldsymbol{u}_1 = \boldsymbol{v}_1 \cos t + \boldsymbol{w} \sin t$ for some vector $\boldsymbol{w}$ orthogonal to $\boldsymbol{v}_1$ ($\boldsymbol{w} := \operatorname{perp}_{\boldsymbol{v}_1} \boldsymbol{u}_1$) and for angle $0 < t < \pi$ (as illustrated in the margin). Multiply by A giving the identity $A\boldsymbol{u}_1 = A\boldsymbol{v}_1 \cos t + A\boldsymbol{w} \sin t$. Now the first column of $AV = US$ gives $A\boldsymbol{v}_1 = \sigma_1 \boldsymbol{u}_1$. Also for the symmetric matrix $A = A^{\mathrm{T}} = (USV^{\mathrm{T}})^{\mathrm{T}} = VS^{\mathrm{T}}U^{\mathrm{T}} = VSU^{\mathrm{T}}$ is an alternative SVD of A: so $AU = VS$ giving

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

in its first column $A\boldsymbol{u}_1 = \sigma_1\boldsymbol{v}_1$. That is, the identity becomes $\sigma_1\boldsymbol{v}_1 = \sigma_1\boldsymbol{u}_1\cos t + (A\boldsymbol{w})\sin t$. Further, $A\boldsymbol{w}$ is orthogonal to $\boldsymbol{u}_1$ by the proof of the SVD Theorem 3.3.6 (since $\boldsymbol{w}$ is orthogonal to $\boldsymbol{v}_1$). Equate the lengths of both sides: $\sigma_1^2 = \sigma_1^2\cos^2 t + |A\boldsymbol{w}|^2\sin^2 t$ which rearranging implies $(\sigma_1^2 - |A\boldsymbol{w}|^2)\sin^2 t = 0$. For angles $0 < t < \pi$ this implies $|A\boldsymbol{w}| = \sigma_1$ for a vector orthogonal to $\boldsymbol{v}_1$ which implies the singular value σ_1 is repeated. This contradicts the supposition; hence $\boldsymbol{u}_1 = \pm\boldsymbol{v}_1$ (for one of the signs, not both).

Recall the induction proof for the SVD Theorem 3.3.6 (section 3.3.3). Here, since $\boldsymbol{u}_1 = \pm\boldsymbol{v}_1$ we can and do choose $\bar{U} = \bar{V}$. Hence $B = \bar{U}^{\mathrm{T}}A\bar{V} = \bar{V}^{\mathrm{T}}A\bar{V}$ is symmetric as $B^{\mathrm{T}} = \bar{V}^{\mathrm{T}}A\bar{V}^{\mathrm{T}} = \bar{V}^{\mathrm{T}}A^{\mathrm{T}}\bar{V} = \bar{V}^{\mathrm{T}}A\bar{V} = B$. Consequently, the same argument applies at all steps in the induction for the proof of an SVD and hence establishes $\boldsymbol{u}_j = \pm\boldsymbol{v}_j$ (each for one of the signs, not both).

Third and lastly, from the jth column of $AV = US$, $A\boldsymbol{v}_j = \sigma_j\boldsymbol{u}_j = \sigma_j(\pm\boldsymbol{v}_j) = \lambda_j\boldsymbol{v}_j$ for eigenvalue λ_j one of $\pm\sigma_j$ but not both. □

Recall that for any real matrix A an SVD is $A = USV^{\mathrm{T}}$. But specifically for symmetric A, the proof of the previous theorem 4.2.15 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 $A = VDV^{\mathrm{T}}$ and so is intimately connected to the following definition.

Definition 4.2.16. *A real square matrix A is **orthogonally diagonalisable** 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 factorisation of A.*

The equivalences in this definition arise immediately from the orthogonality of matrix V (Definition 3.2.43): pre-multiply $V^{\mathrm{T}}AV = D$ by V gives $VV^{\mathrm{T}}AV = AV = VD$; and so on.

Example 4.2.17. (a) Recall from Example 4.2.12 that the symmetric matrix $A = \begin{bmatrix} 1 & \frac{3}{2} \\ \frac{3}{2} & -3 \end{bmatrix}$ has eigenvalues $\lambda = -\frac{7}{2}, \frac{3}{2}$ with corresponding orthogonal eigenvectors $(1, -3)$ and $(3, 1)$. Normalise these eigenvectors to unit length as the columns of the orthogonal matrix

$$V = \begin{bmatrix} \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \\ -\frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \end{bmatrix} = \frac{1}{\sqrt{10}}\begin{bmatrix} 1 & 3 \\ -3 & 1 \end{bmatrix} \quad \text{then}$$

$$\begin{aligned} V^{\mathrm{T}}AV &= \frac{1}{\sqrt{10}}\begin{bmatrix} 1 & -3 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} 1 & \frac{3}{2} \\ \frac{3}{2} & -3 \end{bmatrix}\frac{1}{\sqrt{10}}\begin{bmatrix} 1 & 3 \\ -3 & 1 \end{bmatrix} \\ &= \frac{1}{10}\begin{bmatrix} -\frac{7}{2} & \frac{21}{2} \\ \frac{9}{2} & \frac{3}{2} \end{bmatrix}\begin{bmatrix} 1 & 3 \\ -3 & 1 \end{bmatrix} \end{aligned}$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$= \frac{1}{10}\begin{bmatrix} -35 & 0 \\ 0 & 15 \end{bmatrix} = \begin{bmatrix} -\frac{7}{2} & 0 \\ 0 & \frac{3}{2} \end{bmatrix}.$$

Hence this matrix is orthogonally diagonalisable.

(b) Recall from Example 4.2.13 that the symmetric matrix

$$B = \begin{bmatrix} -6 & 0 & 1 & 1 \\ 0 & 0 & 2 & 2 \\ 1 & 2 & 2 & -1 \\ 1 & 2 & -1 & -1 \end{bmatrix}$$

has orthogonal eigenvectors computed by MATLAB/Octave into the orthogonal matrix `V`. By additionally computing `V'*B*V` we get the following diagonal result (2 d.p.)

```
ans =
  -6.45   0.00   0.00   0.00
   0.00  -3.00   0.00  -0.00
   0.00   0.00   1.11  -0.00
  -0.00  -0.00  -0.00   3.34
```

and see that this matrix B is orthogonally diagonalisable.

■

These examples of orthogonal diagonalisation involve symmetric matrices. Also, the connection between an SVD and orthogonal matrices was previously discussed only for symmetric matrices. The next theorem establishes that all real symmetric matrices are orthogonally diagonalisable, and vice versa. That is, eigenvectors form an orthogonal set if and only if the matrix is symmetric.

Theorem 4.2.18 (spectral)**.** *Let A be real square matrix. Then matrix A is symmetric iff it is orthogonally diagonalisable.*

Proof. The "if" and the "only if" lead to two parts in the proof.

- If matrix A is orthogonally diagonalisable, then $A = VDV^{\mathrm{T}}$ for orthogonal V and diagonal D (and recall $D^{\mathrm{T}} = D$). Consider

$$A^{\mathrm{T}} = (VDV^{\mathrm{T}})^{\mathrm{T}} = V^{\mathrm{T}\mathrm{T}}D^{\mathrm{T}}V^{\mathrm{T}} = VDV^{\mathrm{T}} = A\,.$$

  Consequently the matrix A is symmetric.

- Theorem 4.2.15 establishes the converse for the generic case of distinct singular values. If matrix A is symmetric, then Theorem 4.2.15 asserts an SVD $A = USV^{\mathrm{T}}$ has matrix U such that columns $\boldsymbol{u}_j = \pm\boldsymbol{v}_j$. That is, we can write $U = VR$ for diagonal matrix $R = \operatorname{diag}(\pm 1\,, \pm 1\,, \dots, \pm 1)$ for appropriately chosen signs. Then by the SVD $A = USV^{\mathrm{T}} = VRSV^{\mathrm{T}} =$

VDV^{T} for diagonal matrix $D = RS = \operatorname{diag}(\pm\sigma_1,\pm\sigma_2,\ldots,\pm\sigma_n)$ for the same pattern of signs. Hence matrix A is orthogonally diagonalisable.

We omit proving the degenerate case when non-zero singular values are repeated. □

4.2.3 Change orthonormal basis to classify quadratics

An optional subsection which has many uses—although not an application itself as it does not involve real data.

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.

Example 4.2.19 (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

Solution: Among various possibilities are the following.

(a) Ellipse

In this coordinate system the ellipse is algebraically $(x/2)^2 + y^2 = 1$.

(b) Hyperbola

In this coordinate system the hyperbola is algebraically $y^2 = 1 + 2x^2$.

■

Now let's proceed to see how this geometric idea of choosing good coordinates may be implemented in algebra.

v0.2a

Graph quadratic equations

Example 4.2.19 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 new coordinates so that the quadratic equation (4.2) becomes one of these recognised canonical forms.

Ellipse or circle : $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$

- ellipse $a > b$

- the circle $a = b$

- ellipse $a < b$

Hyperbola : $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$ or $-\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$

- $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$

- $-\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$

Parabola : $y = ax^2$ or $x = ay^2$

v0.2a

- $y = ax^2$

- $x = ay^2$

Example 4.2.19 implicitly had two steps: first, we decided upon an orientation for the coordinate axes; second, we decided that the coordinate system should be 'centred' in the picture. Algebra follows the same two steps.

Example 4.2.20 (centre coordinates)**.** By shifting coordinates, identify the conic section whose equation is

$$2x^2 + y^2 - 4x + 4y + 2 = 0\,.$$

Solution: Group the linear terms with corresponding quadratic powers and seek to rewrite as a perfect square: the equation is

$$\begin{aligned} & (2x^2 - 4x) + (y^2 + 4y) + 2 = 0 \\ \iff\ & 2(x^2 - 2x) + (y^2 + 4y) = -2 \\ \iff\ & 2(x^2 - 2x + 1) + (y^2 + 4y + 4) = -2 + 2 + 4 \\ \iff\ & 2(x-1)^2 + (y+2)^2 = 4\,. \end{aligned}$$

Thus changing to a new (dashed) coordinate system $x' = x - 1$ and $y' = y + 2$, that is choosing the origin of the dashed coordinate system at $(x, y) = (1, -2)$, the quadratic equation becomes

$$2x'^2 + y'^2 = 4\,, \quad \text{that is } \frac{x'^2}{2} + \frac{y'^2}{4} = 1\,.$$

In this new coordinate system the equation is that of an ellipse with horizontal axis of half-length $\sqrt{2}$ and vertical axis of half-length $\sqrt{4} = 2$ (as illustrated in the margin).

■

Example 4.2.21 (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.)

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Solution: The equation contains the product xy. To identify the conic we must eliminate the xy term. To use matrix algebra, and in terms of the vector $\boldsymbol{x} = (x, y)$, recognise that the quadratic terms may be written as $\boldsymbol{x}^{\mathrm{T}} A \boldsymbol{x}$ for symmetric matrix

$$A = \begin{bmatrix} 1 & \frac{3}{2} \\ \frac{3}{2} & -3 \end{bmatrix} \quad \text{as then}$$

$$\begin{aligned} \boldsymbol{x}^{\mathrm{T}} A \boldsymbol{x} &= \boldsymbol{x}^{\mathrm{T}} \begin{bmatrix} 1 & \frac{3}{2} \\ \frac{3}{2} & -3 \end{bmatrix} \boldsymbol{x} \\ &= \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} x + \frac{3}{2}y \\ \frac{3}{2}x - 3y \end{bmatrix} \\ &= x(x + \tfrac{3}{2}y) + y(\tfrac{3}{2}x - 3y) \\ &= x^2 + 3xy - 3y^2. \end{aligned}$$

(The matrix form $\boldsymbol{x}^{\mathrm{T}} A \boldsymbol{x}$ splits the cross-product term $3xy$ into two equal halves represented by the two off-diagonal elements $\frac{3}{2}$ in matrix A). Suppose we change to some new (dashed) coordinate system with its standard unit vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ as illustrated in the margin. The vectors in the plane will be written as the linear combination $\boldsymbol{x} = \boldsymbol{v}_1 x' + \boldsymbol{v}_2 y'$. That is, $\boldsymbol{x} = V\boldsymbol{x}'$ for new coordinate vector $\boldsymbol{x}' = (x', y')$ and matrix $V = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 \end{bmatrix}$.

In the new coordinate system, related to the old by $\boldsymbol{x} = V\boldsymbol{x}'$, the quadratic terms

$$\boldsymbol{x}^{\mathrm{T}} A \boldsymbol{x} = (V\boldsymbol{x}')^{\mathrm{T}} A (V\boldsymbol{x}') = \boldsymbol{x}'^{\mathrm{T}} V^{\mathrm{T}} A V \boldsymbol{x}' = \boldsymbol{x}'^{\mathrm{T}} (V^{\mathrm{T}} A V) \boldsymbol{x}'.$$

Thus choose V to simplify $V^{\mathrm{T}} A V$. Because matrix A is symmetric, Theorem 4.2.18 asserts it is orthogonally diagonalisable (using eigenvectors). Indeed, Example 4.2.17a orthogonally diagonalised this particular matrix A, via its eigenvalues and eigenvectors, using the orthogonal matrix

$$V = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{10}} & \frac{3}{\sqrt{10}} \\ -\frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \end{bmatrix}.$$

Using this V, in the new dashed coordinate system (illustrated) the quadratic terms in the equation become

$$\boldsymbol{x}'^{\mathrm{T}} (V^{\mathrm{T}} A V) \boldsymbol{x}' = \boldsymbol{x}'^{\mathrm{T}} D \boldsymbol{x}' = \boldsymbol{x}'^{\mathrm{T}} \begin{bmatrix} -\frac{7}{2} & 0 \\ 0 & \frac{3}{2} \end{bmatrix} \boldsymbol{x}' = -\tfrac{7}{2}x'^2 + \tfrac{3}{2}y'^2$$

Hence the quadratic equation becomes

$$-\tfrac{7}{2}x'^2 + \tfrac{3}{2}y'^2 - \tfrac{1}{2} = 0 \iff -7x'^2 + 3y'^2 = 1$$

which is the equation of a hyperbola intersecting the y'-axis at $y' = \pm 1/\sqrt{3}$, as illustrated in the margin.

■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Example 4.2.22. Identify the conic section whose equation is

$$x^2 - xy + y^2 + \tfrac{5}{2\sqrt{2}}x - \tfrac{7}{2\sqrt{2}}y + \tfrac{1}{8} = 0\,.$$

Solution: When there are both the cross-product xy and linear terms, it is easier to first rotate coordinates, and second shift coordinates.

(a) Rewrite the quadratic terms using vector $\boldsymbol{x} = (x, y)$ and splitting the cross-product into two equal halfs:

$$\begin{aligned} x^2 - xy + y^2 &= x(x - \tfrac{1}{2}y) + y(-\tfrac{1}{2}x + y) \\ &= \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} x - \frac{1}{2}y \\ -\frac{1}{2}x + y \end{bmatrix} \\ &= \boldsymbol{x}^{\mathrm{T}} \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\ &= \boldsymbol{x}^{\mathrm{T}} A \boldsymbol{x} \quad \text{for matrix } A = \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix}. \end{aligned}$$

Recall that Example 4.1.19 found the eigenvalues of this matrix are $\lambda = \frac{1}{2}, \frac{3}{2}$ with corresponding orthonormal eigenvectors $\boldsymbol{v}_1 = (1, 1)/\sqrt{2}$ and $\boldsymbol{v}_2 = (-1, 1)/\sqrt{2}$, respectively. Let's change to a new (dashed) coordinate system (x', y') with $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ as its standard unit vectors (as illustrated in the margin). Then throughout the 2D-plane every vector/position

$$\boldsymbol{x} = \boldsymbol{v}_1 x' + \boldsymbol{v}_2 y' = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 \end{bmatrix} \begin{bmatrix} x' \\ y' \end{bmatrix} = V\boldsymbol{x}'$$

for orthogonal matrix

$$V = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}.$$

In the new coordinates:

- the quadratic terms

$$\begin{aligned} x^2 - xy + y^2 &= \boldsymbol{x}^{\mathrm{T}} A \boldsymbol{x} \\ &= (V\boldsymbol{x}')^{\mathrm{T}} A (V\boldsymbol{x}') \\ &= \boldsymbol{x}'^{\mathrm{T}} V^{\mathrm{T}} A V \boldsymbol{x}' \\ &= \boldsymbol{x}'^{\mathrm{T}} \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{3}{2} \end{bmatrix} \boldsymbol{x}' \quad (\text{as } V^{\mathrm{T}} A V = D) \\ &= \tfrac{1}{2} x'^2 + \tfrac{3}{2} y'^2; \end{aligned}$$

- whereas the linear terms

$$\begin{aligned} \tfrac{5}{2\sqrt{2}}x - \tfrac{7}{2\sqrt{2}}y &= \begin{bmatrix} \frac{5}{2\sqrt{2}} & -\frac{7}{2\sqrt{2}} \end{bmatrix} \boldsymbol{x} \\ &= \begin{bmatrix} \frac{5}{2\sqrt{2}} & -\frac{7}{2\sqrt{2}} \end{bmatrix} V\boldsymbol{x}' \\ &= \begin{bmatrix} -\frac{1}{2} & -3 \end{bmatrix} \boldsymbol{x}' \\ &= -\tfrac{1}{2}x' - 3y'; \end{aligned}$$

v0.2a

- so the quadratic equation transforms to

$$\tfrac{1}{2}x'^2 + \tfrac{3}{2}y'^2 - \tfrac{1}{2}x' - 3y' + \tfrac{1}{8} = 0\,.$$

(b) The second step is to shift coordinates via completing the squares:

$$\begin{aligned}
& \tfrac{1}{2}x'^2 + \tfrac{3}{2}y'^2 - \tfrac{1}{2}x' - 3y' + \tfrac{1}{8} = 0 \\
\iff\ & \tfrac{1}{2}(x'^2 - x') + \tfrac{3}{2}(y'^2 - 2y') = -\tfrac{1}{8} \\
\iff\ & \tfrac{1}{2}(x'^2 - x' + \tfrac{1}{4}) + \tfrac{3}{2}(y'^2 - 2y' + 1) = -\tfrac{1}{8} + \tfrac{1}{8} + \tfrac{3}{2} \\
\iff\ & \tfrac{1}{2}(x' - \tfrac{1}{2})^2 + \tfrac{3}{2}(y' - 1)^2 = \tfrac{3}{2}
\end{aligned}$$

Thus let's change to a new (double dashed) coordinate system $x'' = x' - \frac{1}{2}$ and $y'' = y' - 1$ (equivalently, choose the origin of a new coordinate system to be at $(x', y') = (\frac{1}{2}, 1)$ as illustrated in the margin). In this new coordinate system the quadratic equation becomes

$$\tfrac{1}{2}x''^2 + \tfrac{3}{2}y''^2 = \tfrac{3}{2}, \quad \text{that is } \frac{x''^2}{3} + \frac{y''^2}{1} = 1\,.$$

In this new coordinate system the equation is that of an ellipse with x''-axis of half-length $\sqrt{3}$ and y''-axis of half-length 1 (as illustrated in the margin).

■

Simplify quadratic forms

Definition 4.2.23. *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 4.2.24. (a) The dot product in $\mathbb{R}^n$ 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.

■

Theorem 4.2.25 (principal axes theorem)**.** *For every quadratic form, there exists an orthogonal coordinate system that diagonalises 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 orthonormal 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** $\boldsymbol{x}^{\mathrm{T}} A \boldsymbol{x} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$.*

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Proof. In the new coordinate system $(y_1, y_2, \dots, y_n)$ the orthonormal vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n$ (called the **principal axes**) act as the standard unit vectors. Hence any vector $\boldsymbol{x} \in \mathbb{R}^n$ may be written as a linear combination

$$\boldsymbol{x} = y_1\boldsymbol{v}_1 + y_2\boldsymbol{v}_2 + \cdots + y_n\boldsymbol{v}_n = \begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n\end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = V\boldsymbol{y}$$

for orthogonal matrix $V = \begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n\end{bmatrix}$ and vector $\boldsymbol{y} = (y_1, y_2, \dots, y_n)$. Then the quadratic form

$$\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x} = (V\boldsymbol{y})^{\mathrm{T}}A(V\boldsymbol{y}) = \boldsymbol{y}^{\mathrm{T}}V^{\mathrm{T}}AV\boldsymbol{y} = \boldsymbol{y}^{\mathrm{T}}D\boldsymbol{y}\,,$$

since $V^{\mathrm{T}}AV = D = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ by Theorem 4.2.18. Consequently,

$$\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x} = \boldsymbol{y}^{\mathrm{T}}D\boldsymbol{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2.$$

□

Theorem 4.2.26 (extreme values)**.** *Let A be an $n \times n$ symmetric matrix with eigenvalues $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$ (sorted). Then for all unit vectors $\boldsymbol{x} \in \mathbb{R}^n$ (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.

Proof. Change to an orthogonal coordinate system that diagonalises the matrix A (Theorem 4.2.25): say $\boldsymbol{y} = V\boldsymbol{x}$ for orthogonal matrix V whose columns are orthogonal eigenvectors of A in order so that $D = V^{\mathrm{T}}AV = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$. Then the quadratic form

$$\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x} = (V\boldsymbol{y})^{\mathrm{T}}A(V\boldsymbol{y}) = \boldsymbol{y}^{\mathrm{T}}V^{\mathrm{T}}AV\boldsymbol{y} = \boldsymbol{y}^{\mathrm{T}}D\boldsymbol{y}\,.$$

Since V is orthogonal it preserves lengths (Theorem 3.2.48f) so the unit vector condition $|\boldsymbol{x}| = 1$ is the same as $|\boldsymbol{y}| = 1$.

1. To prove the lower bound, consider

$$\begin{aligned}\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x} &= \boldsymbol{y}^{\mathrm{T}}D\boldsymbol{y} \\ &= \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2 \\ &= \lambda_1 y_1^2 + \lambda_1 y_2^2 + \cdots + \lambda_1 y_n^2 \\ &\qquad + \underbrace{(\lambda_2 - \lambda_1)y_2^2}_{\ge 0} + \cdots + \underbrace{(\lambda_n - \lambda_1)y_n^2}_{\ge 0}\end{aligned}$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$
\begin{aligned}
&\geq \lambda_1 y_1^2 + \lambda_1 y_2^2 + \cdots + \lambda_1 y_n^2 \\
&\quad = \lambda_1 (y_1^2 + y_2^2 + \cdots + y_n^2) \\
&\quad = \lambda_1 |\boldsymbol{y}|^2 \\
&\quad = \lambda_1 \,.
\end{aligned}
$$

Similarly for the upper bound (Exercise 4.2.22). Thus $\lambda_1 \leq \boldsymbol{x}^{\mathrm{T}} A \boldsymbol{x} \leq \lambda_n$ for all unit vectors $\boldsymbol{x}$.

2. Let $\boldsymbol{v}_1$ be a unit eigenvector of A corresponding to the smallest eigenvalue λ_1; that is, $A\boldsymbol{v}_1 = \lambda_1 \boldsymbol{v}_1$ and $|\boldsymbol{v}_1| = 1$. Then, setting $\boldsymbol{x} = \boldsymbol{v}_1$, the quadratic form

$$
\boldsymbol{x}^{\mathrm{T}} A \boldsymbol{x} = \boldsymbol{v}_1^{\mathrm{T}} A \boldsymbol{v}_1 = \boldsymbol{v}_1^{\mathrm{T}} \lambda_1 \boldsymbol{v}_1 = \lambda_1 (\boldsymbol{v}_1^{\mathrm{T}} \boldsymbol{v}_1) = \lambda_1 |\boldsymbol{v}_1|^2 = \lambda_1 \,.
$$

Thus the quadratic form $\boldsymbol{x}^{\mathrm{T}} A \boldsymbol{x}$ takes on the minimum value λ_1 and it occurs when $\boldsymbol{x} = \boldsymbol{v}_1$ (at least).

3. Exercise 4.2.22 proves the maximum value occurs.

□

4.2.4 Exercises

Exercise 4.2.1. Each plot below shows (unit) vectors $\boldsymbol{x}$ (blue), and for some 2×2 matrix A the corresponding vectors $A\boldsymbol{x}$ (red) adjoined. By assessing whether there are any zero eigenvalues, estimate if the matrix A is invertible or not.

(e)

(f)

Exercise 4.2.2. For each of the following symmetric matrices: from a hand derivation of the characteristic equation (defined in Procedure 4.1.18), determine whether each matrix has a zero eigenvalue or not, and hence determine whether it is invertible or not.

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

(c) $\begin{bmatrix} 0 & -2/5 \\ -2/5 & 3/5 \end{bmatrix}$ (d) $\begin{bmatrix} 2 & 1 & -2 \\ 1 & 3 & -1 \\ -2 & -1 & 2 \end{bmatrix}$

(e) $\begin{bmatrix} -2 & -1 & 1 \\ -1 & -0 & -1 \\ 1 & -1 & -2 \end{bmatrix}$ (f) $\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$

(g) $\begin{bmatrix} -1/2 & 3/2 & 1 \\ 3/2 & -3 & -3/2 \\ 1 & -3/2 & -1/2 \end{bmatrix}$ (h) $\begin{bmatrix} 1 & -1 & -1 \\ -1 & -1/2 & 1/2 \\ -1 & 1/2 & -1/2 \end{bmatrix}$

Exercise 4.2.3. For each of the following matrices, find by hand the eigenvalues and eigenvectors. Using these eigenvectors, confirm that the eigenvalues of the matrix squared are the square of its eigenvalues. If the matrix has an inverse, what are the eigenvalues of the inverse?

(a) $A = \begin{bmatrix} 0 & -2 \\ -2 & 3 \end{bmatrix}$ (b) $B = \begin{bmatrix} 5/2 & -2 \\ -2 & 5/2 \end{bmatrix}$

(c) $C = \begin{bmatrix} 3 & 8 \\ 8 & -9 \end{bmatrix}$ (d) $D = \begin{bmatrix} -2 & 1 \\ 1 & 14/5 \end{bmatrix}$

(e) $E = \begin{bmatrix} -1 & -2 & 0 \\ -2 & 0 & 2 \\ 0 & 2 & 1 \end{bmatrix}$ (f) $F = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 0 & -1 \\ 3 & -1 & 2 \end{bmatrix}$

(g) $G = \begin{bmatrix} 0 & -1 & -1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}$ (h) $H = \begin{bmatrix} -1 & 3/2 & 3/2 \\ 3/2 & -3 & -1/2 \\ 3/2 & -1/2 & 3 \end{bmatrix}$

v0.2a

Exercise 4.2.4. Each plot below shows (unit) vectors $\boldsymbol{x}$ (blue), and for some 2×2 matrix A the corresponding vectors $A\boldsymbol{x}$ (red) adjoined. For each plot of a matrix A there is a companion plot of the inverse matrix A^{-1}. By roughly estimating eigenvalues and eigenvectors by eye, identify the pairs of plots corresponding to each matrix and its inverse.

Exercise 4.2.5. For the symmetric matrices of Exercise 4.2.3, confirm that eigenvectors corresponding to distinct eigenvalues are orthogonal (Theorem 4.2.10). Show your working.

Exercise 4.2.6. For an $n \times n$ symmetric matrix,

- eigenvectors corresponding to different eigenvalues are orthogonal (Theorem 4.2.10), and

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

- there are generally n eigenvalues.

Which of the illustrated 2D examples of Exercise 4.1.1 appear to come from symmetric matrices, and which appear to come from non-symmetric matrices?

Exercise 4.2.7. For each of the following *non-symmetric* matrices, confirm that eigenvectors corresponding to distinct eigenvalues are *not* orthogonal. Show and comment on your working. Find eigenvectors by hand for 2×2 and 3×3 matrices, and compute with MATLAB/Octave for 3×3 matrices and larger (using `eig()` and `V'*V`).

(a) $A=\begin{bmatrix}-2&2\\3&-3\end{bmatrix}$ (b) $B=\begin{bmatrix}1&-3\\-2&2\end{bmatrix}$

(c) $C=\begin{bmatrix}1&-2&6\\4&3&2\\-3&1&-8\end{bmatrix}$ (d) $D=\begin{bmatrix}-1&-2&9\\3&-6&3\\0&0&3\end{bmatrix}$

(e) $E=\begin{bmatrix}2&1&-2&1\\1&-3&4&-4\\3&2&4&-5\\-3&-1&-3&0\end{bmatrix}$ (f) $F=\begin{bmatrix}-4&0&-5&7\\2&3&1&-3\\1&4&-2&4\\1&-3&4&2\end{bmatrix}$

(g) $G=\begin{bmatrix}1&4&3&1&-1\\5&1&6&-0&1\\0&4&-3&1&4\\-3&2&1&4&-1\\-2&2&2&2&1\end{bmatrix}$ (h) $H=\begin{bmatrix}2&0&2&-1&-1\\2&1&-1&2&-0\\5&-2&6&2&-1\\4&0&-2&6&-5\\2&0&-5&-3&-5\end{bmatrix}$

Exercise 4.2.8. For the symmetric matrices of Exercise 4.2.3, use MATLAB/Octave to compute an SVD (USV^{T}) of each matrix. Confirm that each column of V is an eigenvector of the matrix (that is, proportional to what the exercise found) and the corresponding singular value is the magnitude of the corresponding eigenvalue (Theorem 4.2.15). Show and discuss your working.

Exercise 4.2.9. To complement the previous exercise, for each of the *non-symmetric* matrices of Exercise 4.2.7, use MATLAB/Octave to compute an SVD (USV^{T}) of each matrix. Confirm that each column of V is *not* an eigenvector of the matrix, and the singular values do *not* appear closely related to the eigenvalues. Show and discuss your working.

Exercise 4.2.10. Let A be an $m\times n$ matrix with SVD $A=USV^{\mathrm{T}}$. Prove that for any $j=1,2,\dots,n$, the jth column of V, $\boldsymbol{v}_j$, is an eigenvector of the $n\times n$ symmetric matrix $A^{\mathrm{T}}A$ corresponding to the eigenvalue $\lambda_j=\sigma_j^2$ (or $\lambda_j=0$ if $m<j\le n$).

v0.2

Exercise 4.2.11. Prove Theorem part 4.2.4c using parts 4.2.4a and 4.2.4b: that if matrix A is invertible, then for any integer n, λ^n is an eigenvalue of A^n with corresponding eigenvector $\boldsymbol{x}$.

Exercise 4.2.12. For each of the following matrices, give reasons as to whether the matrix is orthogonally diagonalisable, and if it is then find an orthogonal matrix V that does so and the corresponding diagonal matrix D. Use MATLAB/Octave for the larger matrices.

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

(b) $B = \begin{bmatrix} 2/3 & 2 \\ 2 & 7/3 \end{bmatrix}$

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

(d) $D = \begin{bmatrix} 2 & 0 & 2 \\ 1 & -1 & 0 \\ -1 & 0 & -1 \end{bmatrix}$

(e) $E = \begin{bmatrix} 0 & -2 & -2 & 0 \\ -2 & 0 & -2 & 0 \\ -2 & -2 & 2 & 2 \\ 0 & 0 & 2 & -2 \end{bmatrix}$

(f) $F = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 3 & -2 & -2 \\ 0 & -2 & 0 & 1 \\ 0 & -2 & 1 & 0 \end{bmatrix}$

(g) $G = \begin{bmatrix} 3 & 1 & 1 & 1 & -1 \\ -1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & -1 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$

(h) $H = \begin{bmatrix} 3 & 0 & 2 & 0 & 0 \\ 0 & 3 & 1 & 0 & -1 \\ 2 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 3 & 2 \\ 0 & -1 & -1 & 2 & 1 \end{bmatrix}$

Exercise 4.2.13. Let matrix A be invertible and orthogonally diagonalisable. Show that the inverse A^{-1} is orthogonally diagonalisable.

Exercise 4.2.14. Suppose matrices A and B are orthogonally diagonalisable by the same orthogonal matrix V. Show that $AB = BA$ and that the product AB is orthogonally diagonalisable.

Exercise 4.2.15. For each of the given symmetric matrices, say A, find a symmetric matrix X such that $X^2 = A$. That is, find a square-root of the matrix.

(a) $A = \begin{bmatrix} 5/2 & 3/2 \\ 3/2 & 5/2 \end{bmatrix}$

(b) $B = \begin{bmatrix} 6 & -5 & -5 \\ -5 & 10 & 1 \\ -5 & 1 & 10 \end{bmatrix}$

(c) $C = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 2 & 3 \\ 3 & 3 & 6 \end{bmatrix}$

(d) How many possible answers are there for each of the given matrices? Why?

v0.2a

(e) For all symmetric matrices A, show that if every eigenvalue of A is non-negative, then there exists a symmetric matrix X such that $A = X^2$.

(f) Continuing the previous part, how many such matrices X exist? Justify your answer.

Exercise 4.2.16. For each of the following conic sections, draw a pair of coordinate axes for a coordinate system in which the algebraic description of the curves should be simplest.

(a)

(b)

(c)

(d)

(f)

(e)

Exercise 4.2.17. By shifting to a new coordinate axes, find the canonical form of each of the following quadratic equations, and hence describe each curve.

(a) $-4x^2 + 5y^2 + 4x + 4y - 1 = 0$

(b) $-4y^2 - 6x - 4y + 2 = 0$

(c) $5x^2 - y^2 - 2y + 4 = 0$

(d) $3x^2 + 5y^2 + 6x + 4y + 2 = 0$

(e) $-2x^2 - 4y^2 + 7x - 2y + 1 = 0$

(f) $-9x^2 - y^2 - 3y + 6 = 0$

(g) $-x^2 - 4x - 8y + 4 = 0$

(h) $-8x^2 - y^2 - 2x + 2y - 3 = 0$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Exercise 4.2.18. By rotating to new coordinate axes, identify each of the following conic sections. Write the quadratic terms in the form $\boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x}$, and use the eigenvalues and eigenvectors of matrix A.

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

(b) $3x^2 + 8xy - 3y^2 = 0$

(c) $2x^2 - 3xy + 6y^2 - 5 = 0$

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

(e) $-4x^2 + 3xy - 4y^2 + 11 = 0$

(f) $3x^2 - 2xy + 3y^2 - 6 = 0$

(g) $-x^2 + 2xy - y^2 + 5 = 0$

(h) $-x^2 - 4xy + 2y^2 - 6 = 0$

Exercise 4.2.19. By rotating and shifting to new coordinate axes, identify each of the following conic sections from its equation.

(a) $-2x^2 - 5xy - 2y^2 - \frac{33}{2}x - 15y - 32 = 0$

(b) $-7x^2 + 3xy - 3y^2 - 52x + \frac{33}{2}y - \frac{381}{4} = 0$

(c) $-4xy - 3y^2 - 18x - 11y + \frac{37}{4} = 0$

(d) $2x^2 - y^2 + 10x + 6y + \frac{11}{2} = 0$

(e) $-2x^2 + 5xy - 2y^2 - \frac{13}{2}x + \frac{5}{2}y + \frac{31}{4} = 0$

(f) $-4xy + 3y^2 + 18x - 13y + \frac{3}{4} = 0$

(g) $6x^2 + 6y^2 - 12x - 42y + \frac{155}{2} = 0$

(h) $2x^2 - 4xy + 5y^2 + 34x - 52y + \frac{335}{2} = 0$

Exercise 4.2.20. For each of the following matrices, say A, consider the quadratic form $q(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}A\boldsymbol{x}$. Find coordinate axes, the principal axes, such that the quadratic has the canonical form in the new coordinates $y_1, y_2, \ldots, y_n$. Use eigenvalues and eigenvectors, and use MATLAB/Octave for the larger matrices. Over all unit vectors, what is the maximum value of $q(\boldsymbol{x})$? and what is the minimum value?

(a) $A = \begin{bmatrix} 0 & -5 \\ -5 & 0 \end{bmatrix}$

(b) $B = \begin{bmatrix} 3 & 2 \\ 2 & 0 \end{bmatrix}$

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

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

(e) $E = \begin{bmatrix} 6 & 0 & -3 & 1 \\ 0 & -1 & 5 & -3 \\ -3 & 5 & -4 & -7 \\ 1 & -3 & -7 & 0 \end{bmatrix}$

(f) $F = \begin{bmatrix} -5 & -1 & 1 & 2 \\ -1 & 2 & 4 & 1 \\ 1 & 4 & -7 & 7 \\ 2 & 1 & 7 & 2 \end{bmatrix}$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(g)

$$G = \begin{bmatrix} 1 & 1 & -1 & -6 & -7 \\ 1 & 0 & 3 & 3 & -6 \\ -1 & 3 & 12 & 4 & -4 \\ -6 & 3 & 4 & 1 & -2 \\ -7 & -6 & -4 & -2 & 3 \end{bmatrix}$$

(h)

$$H = \begin{bmatrix} 12 & -3 & -3 & -4 & -6 \\ -3 & 0 & 0 & 2 & -5 \\ -3 & 0 & -4 & 1 & -3 \\ -4 & 2 & 1 & -5 & 2 \\ -6 & -5 & -3 & 2 & 0 \end{bmatrix}$$

Exercise 4.2.21. For any given $n \times n$ symmetric matrix A consider the quadratic form $q(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}} A \boldsymbol{x}$. For general vectors $\boldsymbol{x}$, not necessarily unit vectors, what is the maximum value of $q(\boldsymbol{x})$ in terms of $|\boldsymbol{x}|$? and what is the minimum value of $q(\boldsymbol{x})$? Justify your answer.

Exercise 4.2.22. Complete the proof of Theorem 4.2.26 by detailing the proof of the upper bound, and that the upper bound is achieved for an appropriate unit eigenvector.

Exercise 4.2.23. For a symmetric matrix, discuss the similarities and differences between the SVD and the diagonalisation factorisation, the singular values and the eigenvalues, and the singular vectors and the eigenvectors.

v0.2a

Answers to selected exercises

4.1.1b : $\boldsymbol{v}_1 \propto \pm(-0.5,0.9)$, $\lambda_1 \approx -0.3$; and $\boldsymbol{v}_2 \propto \pm(0.6,0.8)$, $\lambda_2 \approx 1.1$.

4.1.1d : $\boldsymbol{v}_1 \propto \pm(1\,,-0.2)$, $\lambda_1 \approx -0.7$; and $\boldsymbol{v}_2 \propto \pm(-0.2\,,1)$, $\lambda_2 \approx 1.1$.

4.1.1f : $\boldsymbol{v}_1 \propto \pm(0.5,0.9)$, $\lambda_1 \approx -0.3$; and $\boldsymbol{v}_2 \propto \pm(-0.9,0.5)$, $\lambda_2 \approx 0.8$.

4.1.1h : $\boldsymbol{v}_1 \propto \pm(0\,,1)$, $\lambda_1 \approx -0.6$; and $\boldsymbol{v}_2 \propto \pm(0.9\,,0.5)$, $\lambda_2 \approx 1.1$.

4.1.2a : Corresponding eigenvalues are: 7, 0, n/a, 7, n/a, n/a

4.1.2c : Corresponding eigenvalues are: 2, -5, -4, n/a, -4, n/a

4.1.2e : Corresponding eigenvalues are: -2, 0, n/a, 1, 1, n/a

4.1.3a : Eigenvalues $-3\,,-5\,,2\,,5$.

4.1.3c : Eigenvalues $2\,,9$, eigenspace $\mathbb{E}_2$ is 3D.

4.1.3e : Eigenvalues $2\,,-13\,,15\,,0$.

4.1.3g : Eigenvalues -5.1461, -1.6639, -0.7427, 0.7676, 7.7851.

4.1.4a : $-1\,,5$

4.1.4c : $-6\,,-1$

4.1.4e : $-3\,,7$

4.1.4g : $1\,,6\,,11$

4.1.4i : $-4\,,5$(twice)

4.1.4k : $-2\,,0\,,10$

4.1.23a : $\mathbb{E}_4 = \operatorname{span}\{(-1\,,1)\}$, $\mathbb{E}_{-2} = \operatorname{span}\{(1\,,1)\}$

4.1.23c : $\mathbb{E}_{-8} = \operatorname{span}\{(2\,,-2\,,-1)\}$, $\mathbb{E}_{-7} = \operatorname{span}\{(-1\,,-1\,,0)\}$, $\mathbb{E}_1 = \operatorname{span}\{(1\,,-1\,,4)\}$

4.1.23e : $\mathbb{E}_{-4} = \operatorname{span}\{(5\,,4\,,-2)\}$, $\mathbb{E}_1 = \operatorname{span}\{(0\,,1\,,2)\}$, 2 is not an eigenvalue

4.1.24a : $\mathbb{E}_{-9} = \operatorname{span}\{(3\,,-1)\}$, $\mathbb{E}_1 = \operatorname{span}\{(1\,,3)\}$. Both eigenvalues have multiplicity one.

4.1.24c : $\mathbb{E}_{-6} = \operatorname{span}\{(1\,,2)\}$, $\mathbb{E}_{-1} = \operatorname{span}\{(-2\,,1)\}$. Both eigenvalues have multiplicity one.

4.1.24e : $\mathbb{E}_{-7} = \operatorname{span}\{(1\,,3\,,-1)\}$, $\mathbb{E}_{-4} = \operatorname{span}\{(1\,,0\,,1)\}$, $\mathbb{E}_4 = \operatorname{span}\{(-3\,,2\,,3)\}$. All eigenvalues have multiplicity one.

4.1.24g : $\mathbb{E}_1 = \operatorname{span}\{(-1\,,2\,,-1)\}$, $\mathbb{E}_{13} = \operatorname{span}\{(-1\,,2\,,5)\,,(2\,,1\,,0)\}$. Eigenvalue $\lambda = 1$ has multiplicity one, whereas $\lambda = 13$ has multiplicity two.

4.1.24i : $\mathbb{E}_{-5} = \operatorname{span}\{(5\,,-2\,,7)\}$, $\mathbb{E}_1 = \operatorname{span}\{(1\,,-1\,,-1)\}$, $\mathbb{E}_{21} = \operatorname{span}\{(3\,,4\,,-1)\}$. All eigenvalues have multiplicity one.

v0.2a

4.1.5 : Number the nodes as you like, say 1=top, 2=centre, 3=right, and 4=bottom. Then the matrix is $\begin{bmatrix} 0 & 2 & 1 & 0 \\ 2 & 0 & 1 & 2 \\ 1 & 1 & 0 & 1 \\ 0 & 2 & 1 & 0 \end{bmatrix}$. Eigenvalues are -2.86 , -0.77 , 0.00 , 3.63 (2 d.p.) and an eigenvector corresponding to the largest 3.63 is $(0.46 , 0.63 , 0.43 , 0.46)$. Thus rank the centre left node as the most important, the right node is least important, and the top and bottom nodes equal second importance.

4.2.1b : not invertible

4.2.1d : not invertible

4.2.1f : invertible

4.2.2b : eigenvalues $0 , 5$ so not invertible.

4.2.2d : eigenvalues $0 , 2 , 5$ so not invertible.

4.2.2f : eigenvalues $1 , 4$ so invertible.

4.2.2h : eigenvalues $-1 , 2$ so invertible.

4.2.3b : Eigenvalues $1/2 , 9/2$, and corresponding eigenvectors proportional to $(1 , 1)$, $(-1 , 1)$. The inverse has eigenvalues $2 , 2/9$.

4.2.3d : Eigenvalues $-11/5 , 3$, and corresponding eigenvectors proportional to $(-5 , 1)$, $(1 , 5)$. The inverse has eigenvalues $-5/11 , 1/3$.

4.2.3f : Eigenvalues $-2 , 1 , 5$, and corresponding eigenvectors proportional to $(-1 , 1 , 1)$, $(1 , 2 , -1)$, $(1 , 0 , 1)$. The inverse has eigenvalues $-1/2 , 1 , 1/5$.

4.2.3h : Eigenvalues $-4 , -1/2 , 7/2$, and corresponding eigenvectors proportional to $(-3 , 5 , 1)$, $(3 , 2 , -1)$, $(1 , 0 , 3)$. The inverse has eigenvalues $-1/4 , -2 , 2/7$.

4.2.7a : Eigenvectors proportional to $(1 , 1)$, $(-2 , 3)$.

4.2.7c : Eigenvectors proportional to $(-2,3,1)$, $(2,-2,-1)$, $(-13,3,14)$.

4.2.7e : Eigenvectors proportional to $(-.39 , .44 , .76 , -.28)$, $(.58 , -.41 , -.68 , .18)$, $(.22 , -.94 , .23 , .11)$, $(.21 , -.53 , .53 , .62)$ (2 d.p.).

4.2.7g : Eigenvectors proportional to $(.01 , -.63 , .78 , .04 , -.05)$, $(-.59 , .49 , .21 , -.46 , -.39)$, $(.57 , .74 , .33 , -.01 , .15)$, $(-.46 , -.46 , .07 , .62 , .43)$, $(-.52 , -.07 , .32 , -.53 , .59)$ (2 d.p.).

4.2.12a : Not symmetric, so not orthogonally diagonalisable.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

4.2.12c : $V = \begin{bmatrix} -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \end{bmatrix}$, and $D = \operatorname{diag}(1\,,2\,,4)$.

4.2.12e : $V = \begin{bmatrix} -.50 & -.41 & .71 & -.29 \\ -.50 & -.41 & -.71 & -.29 \\ -.50 & .00 & .00 & .87 \\ .50 & -.82 & .00 & .29 \end{bmatrix}$ (2 d.p.), and $D = \operatorname{diag}(-4\,,-2\,,2\,,4)$.

4.2.12g : Not symmetric, so not orthogonally diagonalisable.

4.2.15a : $X = \begin{bmatrix} 3/2 & 1/2 \\ 1/2 & 3/2 \end{bmatrix}$

4.2.15c : $X = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix}$

4.2.17a : hyperbola, centred $(1/2\,,-2/5)$, $-\frac{x'^2}{1/5} + \frac{y'^2}{4/25} = 1$

4.2.17c : hyperbola, centred $(0\,,-1)$, $-\frac{x'^2}{1} + \frac{y'^2}{5} = 1$

4.2.17e : ellipse, centred $(7/4\,,-1/4)$, $\frac{x'^2}{59/16} + \frac{y'^2}{59/32} = 1$

4.2.17g : parabola, base $(-2\,,1)$, $y' = -\frac{1}{8}x'^2$

4.2.18a : With axes $\boldsymbol{i}' = (-\frac{2}{\sqrt{5}}\,,\frac{1}{\sqrt{5}})$ and $\boldsymbol{j}' = (\frac{1}{\sqrt{5}}\,,\frac{2}{\sqrt{5}})$, $-\frac{x'^2}{3} + \frac{y'^2}{3/4} = 1$ is hyperbola.

4.2.18c : In axes $\boldsymbol{i}' = (\frac{3}{\sqrt{10}}\,,\frac{1}{\sqrt{10}})$ and $\boldsymbol{j}' = (-\frac{1}{\sqrt{10}}\,,\frac{3}{\sqrt{10}})$, $\frac{x'^2}{10/3} + \frac{y'^2}{10/13} = 1$ is ellipse.

4.2.18e : In axes $\boldsymbol{i}' = (\frac{1}{\sqrt{2}}\,,-\frac{1}{\sqrt{2}})$ and $\boldsymbol{j}' = (\frac{1}{\sqrt{2}}\,,\frac{1}{\sqrt{2}})$, $\frac{x'^2}{2} + \frac{y'^2}{22/5} = 1$ is ellipse.

4.2.18g : In axes $\boldsymbol{i}' = (\frac{1}{\sqrt{2}}\,,-\frac{1}{\sqrt{2}})$ and $\boldsymbol{j}' = (\frac{1}{\sqrt{2}}\,,\frac{1}{\sqrt{2}})$, $\frac{x'^2}{5/2} = 1$ is pair of parallel lines.

4.2.19a : hyperbola centred $(-1\,,-5/2)$ at angle 45°

4.2.19c : hyperbola centred $(4\,,-9/2)$ at angle -27°

4.2.19e : hyperbola centred $(3/2\,,5/2)$ at angle 45°

4.2.19g : circle centred $(1\,,7/2)$

4.2.20a : $q = -5y_1^2 + 5y_2^2$, max$= 5$, min$=-5$

4.2.20c : $q = -4y_1^2 - y_2^2 + 2y_3^2$, max$= 2$, min$=-4$

4.2.20e : (2 d.p.) $q = -10.20y_1^2 - 3.56y_2^2 + 4.81y_3^2 + 9.95y_4^2$, max$= 9.95$, min$=-10.20$

v0.2a

4.2.20g : $q = -8.82y_1^2 - 4.71y_2^2 + 3.04y_3^2 + 10.17y_4^2 + 17.03y_5^2$, max= 17.03, min=$-8.82$

v0.2a

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

5 Approximate matrices

Chapter Contents

5.1 Measure changes to matrices 470

5.1.1 Compress images optimally 470

5.1.2 Relate matrix changes to the SVD 477

5.1.3 Principal component analysis 488

5.1.4 Exercises 507

5.2 Regularise linear equations 514

5.2.1 The SVD illuminates regularisation 515

5.2.2 Tikhonov regularisation 530

5.2.3 Exercises 534

This chapter could be studied any time after Chapter 3 to help the transition to more abstract linear algebra. Useful to time as spaced revision of earlier material on the SVD, rank, orthogonality, and so on.

This chapter develops how concepts associated with length and distance applies to matrices as well as vectors. Further courses on Linear Algebra place these in a unifying framework that also encompasses much you see both in solving differential equations (and integral equations) and in problems involving complex numbers (such as those in electrical engineering or quantum physics).

5.1 Measure changes to matrices

Section Contents

5.1.1 Compress images optimally 470

5.1.2 Relate matrix changes to the SVD 477

5.1.3 Principal component analysis 488

Application to latent semantic indexing . . . 498

5.1.4 Exercises 507

5.1.1 Compress images optimally

Photographs and other images take a lot of storage. Reducing the amount of storage taken, both for storage and for transmission, is essential. The well-known jpeg format for compressing photographs is incredibly useful: the SVD provides another effective means of compression.

These methods find approximate matrices of the images with the matrices having of various ranks. Recall that a matrix of rank k (Definition 3.3.19) means the matrix has precisely k non-zero singular values, that is, an $m \times n$ matrix

$$\begin{aligned}
A &= USV^{\mathrm{T}} \\
&= \begin{bmatrix} \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_k & \cdots & \boldsymbol{u}_m \end{bmatrix} \begin{bmatrix} \begin{matrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_k \end{matrix} & O_{k\times(n-k)} \\ O_{(m-k)\times k} & O_{(m-k)\times(n-k)} \end{bmatrix} V^{\mathrm{T}} \\
&= \begin{bmatrix} \sigma_1\boldsymbol{u}_1 & \cdots & \sigma_k\boldsymbol{u}_k & O_{m\times(n-k)} \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_1^{\mathrm{T}} \\ \vdots \\ \boldsymbol{v}_k^{\mathrm{T}} \\ \vdots \\ \boldsymbol{v}_n^{\mathrm{T}} \end{bmatrix} \\
&= \sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + \sigma_2\boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}} + \cdots + \sigma_k\boldsymbol{u}_k\boldsymbol{v}_k^{\mathrm{T}} .
\end{aligned}$$

This last form constructs matrix A and *has relatively few components when the rank k is low* compared to m and n.

Example 5.1.1. Invent and write down a rank three representation of the 5×5 'bulls eye' matrix (illustrated in the margin)

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

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Solution: Here we set all coefficients $\sigma_1 = \sigma_2 = \sigma_3 = 1$ and let the vectors set the magnitude (subsequently, the vectors will be unit vectors and σ_j will set the magnitude).

(a) Arbitrarily start by addressing together the first and last rows of the image: they can be computed by choosing $\boldsymbol{u}_1 = (1,0,0,0,1)$ and $\boldsymbol{v}_1 = (0,1,1,1,0)$ as then (as illustrated) a rank one matrix

$$\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} = \begin{bmatrix}1\\0\\0\\0\\1\end{bmatrix}\begin{bmatrix}0 & 1 & 1 & 1 & 0\end{bmatrix} = \begin{bmatrix}0 & 1 & 1 & 1 & 0\\0 & 0 & 0 & 0 & 0\\0 & 0 & 0 & 0 & 0\\0 & 0 & 0 & 0 & 0\\0 & 1 & 1 & 1 & 0\end{bmatrix}.$$

(b) Next choose to address together the first and last columns of the image: they can be computed by choosing $\boldsymbol{u}_2 = (0,1,1,1,0)$ and $\boldsymbol{v}_2 = (1,0,0,0,1)$ as then (as illustrated) a rank two matrix

$$\begin{aligned}\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + \boldsymbol{u}_2\boldsymbol{v}_2 &= \boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + \begin{bmatrix}0\\1\\1\\1\\0\end{bmatrix}\begin{bmatrix}1 & 0 & 0 & 0 & 1\end{bmatrix}\\ &= \boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + \begin{bmatrix}0 & 0 & 0 & 0 & 0\\1 & 0 & 0 & 0 & 1\\1 & 0 & 0 & 0 & 1\\1 & 0 & 0 & 0 & 1\\0 & 0 & 0 & 0 & 0\end{bmatrix}\\ &= \begin{bmatrix}0 & 1 & 1 & 1 & 0\\1 & 0 & 0 & 0 & 1\\1 & 0 & 0 & 0 & 1\\1 & 0 & 0 & 0 & 1\\0 & 1 & 1 & 1 & 0\end{bmatrix}.\end{aligned}$$

(c) Lastly put the dot in the middle of the image: choose $\boldsymbol{u}_3 = \boldsymbol{v}_3 = (0,0,1,0,0)$ as then (to form the original) a rank three matrix

$$\begin{aligned}&\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + \boldsymbol{u}_2\boldsymbol{v}_2 + \boldsymbol{u}_3\boldsymbol{v}_3\\ &= \boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + \boldsymbol{u}_2\boldsymbol{v}_2 + \begin{bmatrix}0\\0\\1\\0\\0\end{bmatrix}\begin{bmatrix}0 & 0 & 1 & 0 & 0\end{bmatrix}\\ &= \boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + \begin{bmatrix}0 & 0 & 0 & 0 & 0\\0 & 0 & 0 & 0 & 0\\0 & 0 & 1 & 0 & 0\\0 & 0 & 0 & 0 & 0\\0 & 0 & 0 & 0 & 0\end{bmatrix}\end{aligned}$$

v0.2a

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

■

Procedure 5.1.2 (approximate images). *For an image stored as numbers 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 5.1.12): typically there will be k 'large' singular values and the rest are 'small'.*
3. *Then the 'best' rank k approximation to the image matrix A is (using the subscript k on the matrix name to denote the rank k approximation)*

$$\begin{aligned} A_k &= \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^{\mathrm{T}} + \cdots + \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^{\mathrm{T}} \\ &= \texttt{U(:,1:k)*S(1:k,1:k)*V(:,1:k)'} \end{aligned}$$

Example 5.1.3. Use Procedure 5.1.2 to find the 'best' rank two and three matrices to approximate the 'bulls eye' image matrix (illustrated in the margin)

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

Solution: Enter the matrix into MATLAB/Octave and compute an SVD, $A = USV^{\mathrm{T}}$, with `[U,S,V]=svd(A)` to find (2 d.p.)

```
U =
  -0.47   0.51   0.14   0.71   0.05
  -0.35  -0.44   0.43  -0.05   0.71
  -0.56  -0.31  -0.77   0.00   0.00
  -0.35  -0.44   0.43   0.05  -0.71
  -0.47   0.51   0.14  -0.71  -0.05
S =
   2.68      0      0      0      0
      0   2.32      0      0      0
      0      0   0.64      0      0
      0      0      0   0.00      0
      0      0      0      0   0.00
V =
  -0.47  -0.51   0.14  -0.68  -0.18
```

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

```
-0.35   0.44   0.43  -0.18   0.68
-0.56   0.31  -0.77   0.00  -0.00
-0.35   0.44   0.43   0.18  -0.68
-0.47  -0.51   0.14   0.68   0.18
```

- For this matrix there are three 'large' singular values of 2.68, 2.32 and 0.64, and two 'small' singular values of 0.00 (they are precisely zero), thus construct a rank three approximation to the image matrix as

$$A_3 = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^{\mathrm{T}} + \sigma_3 \boldsymbol{u}_3 \boldsymbol{v}_3^{\mathrm{T}},$$

computed with `A3=U(:,1:3)*S(1:3,1:3)*V(:,1:3)'`, giving (2 d.p.)

```
A3 =
   0.00   1.00   1.00   1.00   0.00
   1.00   0.00   0.00   0.00   1.00
   1.00   0.00   1.00   0.00   1.00
   1.00   0.00   0.00   0.00   1.00
  -0.00   1.00   1.00   1.00  -0.00
```

The rank 3 matrix A_3 exactly reproduces the image matrix A. This exactness is due to the fourth and fifth singular values being precisely zero (to numerical error).

- Alternatively, in the context of some application, we could subjectively decide that there are two 'large' singular values of 2.68 and 2.32, and three 'small' singular values of 0.64 and 0.00. In such a case, construct a rank two approximation to the image matrix as (illustrated in the margin)

$$A_2 = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^{\mathrm{T}},$$

computed with `A2=U(:,1:2)*S(1:2,1:2)*V(:,1:2)'`, giving (2 d.p.)

```
A2 =
  -0.01   0.96   1.07   0.96  -0.01
   0.96  -0.12   0.21  -0.12   0.96
   1.07   0.21   0.62   0.21   1.07
   0.96  -0.12   0.21  -0.12   0.96
  -0.01   0.96   1.07   0.96  -0.01
```

This rank two approximation A_2 is indeed roughly the same as the image matrix A, albeit with errors of 20% or so. Subsequent theory confirms that the relative error is characterised by $\sigma_3/\sigma_1 = 0.24$ here.

■

Euler, 1737, by Vasilij Sokolov

http://eulerarchive.maa.org/portraits/portraits.html [Sep 2015]

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Figure 5.1: four approximate images of Euler ranging from the poor rank 3, via the adequate rank 10, the good rank 30, to the original rank 277.

Example 5.1.4. In the margin is a 326×277 greyscale image of Euler at 30 years old. As such the image is coded as $90,302$ numbers. Let's find a good approximation to the image that uses much fewer numbers, and hence takes less storage. That is, we will effectively compress the image.

Solution: We use an SVD to approximate the image of Euler to controllable levels of approximation. For example, Figure 5.1 shows four approximations to the image of Euler, ranging from the hopeless (labelled "rank 3") to the original (labelled "rank 277").

Procedure 5.1.2 is as follows.

- First download the image from the website, and then read the image into MATLAB/Octave using

  ```
  rgb=imread('euler1737.png');
  A=mean(rgb,3);
  ```

  The `imread` command sets the $326 \times 277 \times 3$ array `rgb` to the red-green-blue values of the image data. Then convert into a grayscale image matrix `A` by averaging the red-green-blue values via the function `mean(rgb,3)` which computes the mean over the third dimension of the array (over the three colours).

- Compute an SVD, $A = USV^{\mathrm{T}}$, of the matrix `A` with the usual command `[U,S,V]=svd(A)`. Here orthogonal `U` is 326×326, diagonal $\mathtt{S} = \operatorname{diag}(41\,422, 8\,309, \dots, 0.79, 0)$ is 326×277, and

Figure 5.2: singular values of the image of Euler, 1737.

orthogonal `V` is 277×277. These matrices are far too big to record in this text.

- Figure 5.2 plots the non-zero singular values from largest to smallest: they cover a range of five orders of magnitude. Choose some number k of singular vectors to use: k is the rank of the approximate images in Figure 5.1. The choice may be guided by the decrease of these singular values: as discussed later, for a say 1% error choose k such that $\sigma_k \approx 0.01\sigma_1$ which from the index j axis of Figure 5.2 is around $k \approx 25$.

- Construct the approximate rank k image

$$\begin{aligned} A_k &= \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^{\mathrm{T}} + \cdots + \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^{\mathrm{T}} \\ &= \texttt{U(:,1:k)*S(1:k,1:k)*V(:,1:k)'} \end{aligned}$$

- Let's say the rank 30 image of Figure 5.1 is the desired good approximation. To reconstruct it we need 30 singular values $\sigma_1, \sigma_2, \ldots, \sigma_{30}$, 30 columns $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_{30}$ of U, 30 columns $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_{30}$ of V making a total of

$$30 + 30 \times 326 + 30 \times 277 = 18,120 \text{ numbers.}$$

These $18,120$ numbers are much fewer than (one fifth) the $326 \times 277 = 90,302$ numbers of the original image. The SVD provides an effective flexible data compression.

■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Table 5.1: As well as the MATLAB/Octave commands and operations listed in Tables 1.2, 2.3, 3.1, 3.2, 3.3, and 3.7 we may invoke these functions.

- `norm(A)` computes the matrix norm of Definition 5.1.5, namely the largest singular value of the matrix A.
  Also, and consistent with the matrix norm, recall that `norm(v)` for a vector `v` computes the length $\sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$.
- `scatter(x,y,[],c)` draws a 2D scatter plot of points with coordinates in vectors `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 $[0\,,255]$.
- `mean(A)` of an $m \times n$ array computes the n elements in the row vector of averages (the arithmetic mean) over each column of A.
  Whereas `mean(A,p)` for an ℓ-dimensional array A of dimension $m_1 \times m_2 \times \cdots \times m_\ell$, computes the mean over the `p`th index to give an array of size $m_1 \times \cdots \times m_{p-1} \times m_{p+1} \times \cdots \times m_\ell$.
- `std(A)` of an $m \times n$ array computes the n elements in the row vector of the standard deviation over each column of A (close to the root-mean-square from the mean).
- `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.
- `semilogy(x,y,'o')` draws a point plot of y versus x with the vertical axis being logarithmic.
- `axis` sets some properties of a drawn figure:
  - `axis equal` ensures horizontal and vertical directions are scaled the same—so here there is no distortion of the image;
  - `axis off` means that the horizontal and vertical axes are not drawn—so here the image is unadorned.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

5.1.2 Relate matrix changes to the SVD

We need to define what 'best' means in the approximation Procedure 5.1.2 and then show the procedure achieves this best. To proceed to understand image compression, we need a measure of the magnitude of matrices and distances between matrices.

In linear algebra we use the double vertical bars, $\|\cdot\|$, to denote the magnitude of a matrix in order to avoid a notational clash with the well-established use of $|\cdot|$ for the determinant of a matrix (Chapter 6).

Definition 5.1.5. *Let A be an $m \times n$ matrix. Define the* ***matrix norm*** *(sometimes called the spectral norm)*

$$\|A\| := \max_{|\boldsymbol{x}|=1} |A\boldsymbol{x}|, \quad \textit{equivalently } \|A\| = \sigma_1 \tag{5.1}$$

the largest singular value of the matrix A. [1]

Proof. The equivalence $\max_{|\boldsymbol{x}|=1} |A\boldsymbol{x}| = \sigma_1$ is due to the definition of the largest singular value in the proof of the existence of an SVD (section 3.3.3). □

Example 5.1.6. 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 5.1.5, the norm of the matrix A is then the length of the longest such plotted $A\boldsymbol{x}$. For each matrix, use the plot to roughly estimate their norm.

(a) $A = \begin{bmatrix} 0.5 & 0.5 \\ -0.6 & 1.2 \end{bmatrix}$

The MATLAB function `eigshow(A)` provides an interactive alternative to this static view.

Solution: The longest $A\boldsymbol{x}$ appear to be near the top and bottom of the plot, and appear to be a little longer than one, so estimate $\|A\| \approx 1.3$.

(b) $B = \begin{bmatrix} -0.7 & 0.4 \\ 0.6 & 0.5 \end{bmatrix}$

Solution: Near the top and bottom of the plot, $B\boldsymbol{x}$ appears to be of length 0.6. But the vectors pointing inwards from the right and left appear longer at about 0.9. So estimate $\|B\| \approx 0.9$.

■

[1] Sometimes this matrix norm is more specifically called a 2-norm and correspondingly denoted by $\|A\|_2$: but not in this book because at other times and places $\|A\|_2$ denotes something slightly different.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Example 5.1.7. Consider the 2×2 matrix $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. Explore products $A\boldsymbol{x}$ for unit vectors $\boldsymbol{x}$, 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 all unit vectors in 2D are of the form $\boldsymbol{x} = (\cos t, \sin t)$. Then

$$\begin{aligned} |A\boldsymbol{x}|^2 &= |(\cos t + \sin t, \sin t)|^2 \\ &= (\cos t + \sin t)^2 + \sin^2 t \\ &= \cos^2 t + 2\cos t\,\sin t + \sin^2 t + \sin^2 t \\ &= \tfrac{3}{2} + \sin 2t - \tfrac{1}{2}\cos 2t\,. \end{aligned}$$

This length (squared) is maximised (and minimised) for some t determined by calculus. Differentiating with respect to t leads to

$$\frac{d|A\boldsymbol{x}|^2}{dt} = 2\cos 2t + \sin 2t \;= 0 \quad \text{for stationary points.}$$

Rearranging determines we require $\tan 2t = -2$. The marginal right-angle triangles illustrate that a stationary point of $|A\boldsymbol{x}|^2$ occurs 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 $\|A\| = 1.6180$.

■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Example 5.1.8. For larger matrices, MATLAB/Octave readily computes the norm either via an SVD or using the `norm` function directly (Table 5.1). Compute the norm of the following matrices.

(a) $A = \begin{bmatrix} 0.1 & -1.3 & -0.4 & -0.1 & -0.6 \\ 1.9 & 2.4 & -1.8 & 0.2 & 0.8 \\ -0.2 & -0.5 & -0.7 & -2.5 & 1.1 \\ -1.8 & 0.2 & 1.1 & -1.2 & 1.0 \\ -0.0 & 1.2 & 1.1 & -0.1 & 1.7 \end{bmatrix}$

Solution: Enter the matrix into MATLAB/Octave then executing `svd(A)` returns the vector of singular values

$$(4.0175, 3.5044, 2.6568, 0.8571, 0.1618),$$

so $\|A\| = \sigma_1 = 4.0175$. Alternatively, executing `norm(A)` directly gives $\|A\| = 4.0175$.

(b) $B = \begin{bmatrix} 0 & -2 & -1 & -4 & -5 & 0 \\ 2 & 0 & 1 & -2 & -6 & -2 \\ -2 & 0 & 4 & 2 & 3 & -3 \\ 1 & 2 & -4 & 2 & 1 & 3 \end{bmatrix}$

Solution: Enter the matrix into MATLAB/Octave then executing `svd(B)` returns the vector of singular values

$$(10.1086, 7.6641, 3.2219, 0.8352),$$

so $\|B\| = \sigma_1 = 10.1086$. Alternatively, executing `norm(B)` directly gives $\|B\| = 10.1086$.

■

The Definition 5.1.5 of the magnitude/norm of a matrix may appear a little strange, but, in addition to some marvellously useful properties, it nonetheless has all the familiar properties of a magnitude/length. Recall from Chapter 1 that for vectors:

- $|\boldsymbol{v}| = 0$ if and only if $\boldsymbol{v} = \boldsymbol{0}$ (Theorem 1.1.13);
- $|\boldsymbol{u} \pm \boldsymbol{v}| \le |\boldsymbol{u}| + |\boldsymbol{v}|$ (the triangle inequality of Theorem 1.3.17);
- $|t\boldsymbol{v}| = |t| \cdot |\boldsymbol{v}|$ (Theorem 1.3.17).

Analogous properties hold for the matrix norm as established in the next theorem.

Theorem 5.1.9 (norm properties). *Let A be an $m \times n$ real matrix:*

(a) $\|A\| = 0$ *if and only if* $A = O_{m\times n}$ *;*

(b) $\|I_n\| = 1$ *;*

(c) $\|A \pm B\| \le \|A\| + \|B\|$*, for any* $m \times n$ *matrix* B*, like a triangle inequality (Theorem 1.3.17c);*

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(d) $\|tA\| = |t|\|A\|$ *;*

(e) $\|A\| = \|A^{\mathrm{T}}\|$ *;*

(f) $\|Q_m A\| = \|A\| = \|AQ_n\|$ *for any* $m\times m$ *orthogonal matrix* Q_m *and any* $n \times n$ *orthogonal matrix* Q_n*;*

(g) $|A\boldsymbol{x}| \le \|A\||\boldsymbol{x}|$ *for all* $\boldsymbol{x} \in \mathbb{R}^n$*, like a Cauchy–Schwarz inequality (Theorem 1.3.17b), as is the following;*

(h) $\|AB\| \le \|A\|\|B\|$ *for any* $n \times p$ *matrix* B*.*

Proof. Alternative proofs to the following may be invoked (Exercise 5.1.3). Where necessary in the following, let matrix A have the SVD $A = USV^{\mathrm{T}}$.

5.1.9a. If $A = O_{m\times n}$, then from Definition 5.1.5

$$\|A\| = \max_{|\boldsymbol{x}|=1} |O\boldsymbol{x}| = \max_{|\boldsymbol{x}|=1} |\boldsymbol{0}| = \max_{|\boldsymbol{x}|=1} 0 = 0\,.$$

Conversely, if $\|A\| = 0$, then the largest singular value $\sigma_1 = 0$ (Definition 5.1.5), which implies that all singular values are zero, so the matrix A has an SVD of the form $A = UO_{m\times n}V^{\mathrm{T}}$, which evaluates to $A = O_{m\times n}$.

5.1.9b. From Definition 5.1.5,

$$\|I_n\| = \max_{|\boldsymbol{x}|=1} |I_n\boldsymbol{x}| = \max_{|\boldsymbol{x}|=1} |\boldsymbol{x}| = \max_{|\boldsymbol{x}|=1} 1 = 1\,.$$

5.1.9c. Using Definition 5.1.5 at the first and last steps:

$$\begin{aligned}
\|A \pm B\| &= \max_{|\boldsymbol{x}|=1} |(A \pm B)\boldsymbol{x}| \\
&= \max_{|\boldsymbol{x}|=1} |A\boldsymbol{x} \pm B\boldsymbol{x}| \quad \text{(by distributivity)} \\
&\le \max_{|\boldsymbol{x}|=1} (|A\boldsymbol{x}| + |B\boldsymbol{x}|) \quad \text{(by triangle inequality)} \\
&\le \max_{|\boldsymbol{x}|=1} |A\boldsymbol{x}| + \max_{|\boldsymbol{x}|=1} |B\boldsymbol{x}| \\
&= \|A\| + \|B\|\,.
\end{aligned}$$

5.1.9d. Using Definition 5.1.5,

$$\begin{aligned}
\|tA\| &= \max_{|\boldsymbol{x}|=1} |(tA)\boldsymbol{x}| \\
&= \max_{|\boldsymbol{x}|=1} |t(A\boldsymbol{x})| \quad \text{(by associativity)} \\
&= \max_{|\boldsymbol{x}|=1} |t||A\boldsymbol{x}| \quad \text{(by Theorem 1.3.17)} \\
&= |t| \max_{|\boldsymbol{x}|=1} |A\boldsymbol{x}| \\
&= |t|\|A\|\,.
\end{aligned}$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

5.1.9e. Recall that then A^{T} has an SVD $A^{\mathrm{T}} = (USV^{\mathrm{T}})^{\mathrm{T}} = VS^{\mathrm{T}}U^{\mathrm{T}}$. So the largest singular value of A^{T} is the same as that of A. Hence $\|A\| = \|A^{\mathrm{T}}\|$.

5.1.9f. Recall that multiplication by an orthogonal matrix is a rotation/reflection and so does not change lengths (Theorem 3.2.48f): correspondingly, it also does not change the norm of a matrix as established here.

Now $Q_mA = Q_m(USV^{\mathrm{T}}) = (Q_mU)SV^{\mathrm{T}}$. But Q_mU is an orthogonal matrix (Exercise 3.2.19), so $(Q_mU)SV^{\mathrm{T}}$ is an SVD for Q_mA. From the singular values in S, $\|Q_mA\| = \sigma_1 = \|A\|$.

Also, using 5.1.9e twice: $\|AQ_n\| = \|(AQ_n)^{\mathrm{T}}\| = \|Q_n^{\mathrm{T}}A^{\mathrm{T}}\| = \|A^{\mathrm{T}}\| = \|A\|$.

5.1.9g If $\boldsymbol{x} = \mathbf{0}$, then $|A\boldsymbol{x}| = |A\mathbf{0}| = |\mathbf{0}| = 0$ whereas $\|A\||\boldsymbol{x}| = \|A\||\mathbf{0}| = \|A\|0 = 0$, so $|A\boldsymbol{x}| \le \|A\||\boldsymbol{x}|$. Alternatively, if $\boldsymbol{x} \ne \mathbf{0}$, then we write $\boldsymbol{x} = \hat{\boldsymbol{x}}|\boldsymbol{x}|$ for unit vector $\hat{\boldsymbol{x}} = \boldsymbol{x}/|\boldsymbol{x}|$ so that

$$\begin{aligned}
|A\boldsymbol{x}| &= \big|A\hat{\boldsymbol{x}}|\boldsymbol{x}|\big| \\
&= |A\hat{\boldsymbol{x}}||\boldsymbol{x}| \quad \text{(as } |\boldsymbol{x}| \text{ is a scalar)} \\
&\le \max_{|\hat{\boldsymbol{x}}|=1} |A\hat{\boldsymbol{x}}||\boldsymbol{x}| \quad \text{(as } \hat{\boldsymbol{x}} \text{ is a unit vector)} \\
&= \|A\||\boldsymbol{x}| \quad \text{(by Definition 5.1.5)}
\end{aligned}$$

5.1.9h See Exercise 5.1.3e.

□

Since the matrix norm has the familiar properties of a measure of magnitude, we use the matrix norm to measure the 'distance' between matrices.

Example 5.1.10. (a) Use the matrix norm to estimate the 'distance' between matrices

$$B = \begin{bmatrix} -0.7 & 0.4 \\ 0.6 & 0.5 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} -0.2 & 0.9 \\ 0 & 1.7 \end{bmatrix}.$$

Solution: The 'distance' is

$$\|C - B\| = \left\| \begin{bmatrix} 0.5 & 0.5 \\ -0.6 & 1.2 \end{bmatrix} \right\| = \|A\| \approx 1.3$$

for the matrix A of Example 5.1.6a and its estimated norm.

(b) Recall from Example 3.3.2 that the matrix

$$A = \begin{bmatrix} 10 & 2 \\ 5 & 11 \end{bmatrix}$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

has an SVD of

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

- Find $\|A - B\|$ for the rank one matrix

$$B = \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^{\mathrm{T}} = 5\sqrt{2} \begin{bmatrix} -\frac{4}{5} \\ \frac{3}{5} \end{bmatrix} \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 4 & -4 \\ -3 & 3 \end{bmatrix} .$$

Solution: Let's write matrix

$$\begin{aligned} B &= \begin{bmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 5\sqrt{2} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}^{\mathrm{T}} \\ &= U \begin{bmatrix} 0 & 0 \\ 0 & 5\sqrt{2} \end{bmatrix} V^{\mathrm{T}} . \end{aligned}$$

Then the difference is

$$\begin{aligned} A - B &= U \begin{bmatrix} 10\sqrt{2} & 0 \\ 0 & 5\sqrt{2} \end{bmatrix} V^{\mathrm{T}} - U \begin{bmatrix} 0 & 0 \\ 0 & 5\sqrt{2} \end{bmatrix} V^{\mathrm{T}} \\ &= U \begin{bmatrix} 10\sqrt{2} & 0 \\ 0 & 0 \end{bmatrix} V^{\mathrm{T}} . \end{aligned}$$

This is an SVD for $A - B$ with singular values $10\sqrt{2}$ and 0, so by Definition 5.1.5 its norm $\|A - B\| = \sigma_1 = 10\sqrt{2}$.

- Find $\|A - A_1\|$ for the rank one matrix

$$A_1 = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} = 10\sqrt{2} \begin{bmatrix} \frac{3}{5} \\ \frac{4}{5} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 6 & 6 \\ 8 & 8 \end{bmatrix} .$$

Solution: Let's write matrix

$$\begin{aligned} A_1 &= \begin{bmatrix} \frac{3}{5} & -\frac{4}{5} \\ \frac{4}{5} & \frac{3}{5} \end{bmatrix} \begin{bmatrix} 10\sqrt{2} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}^{\mathrm{T}} \\ &= U \begin{bmatrix} 10\sqrt{2} & 0 \\ 0 & 0 \end{bmatrix} V^{\mathrm{T}} . \end{aligned}$$

Then the difference is

$$\begin{aligned} A - A_1 &= U \begin{bmatrix} 10\sqrt{2} & 0 \\ 0 & 5\sqrt{2} \end{bmatrix} V^{\mathrm{T}} - U \begin{bmatrix} 10\sqrt{2} & 0 \\ 0 & 0 \end{bmatrix} V^{\mathrm{T}} \\ &= U \begin{bmatrix} 0 & 0 \\ 0 & 5\sqrt{2} \end{bmatrix} V^{\mathrm{T}} . \end{aligned}$$

This is an SVD for $A - A_1$ with singular values $5\sqrt{2}$ and 0, albeit out of order, so by Definition 5.1.5 its norm $\|A - A_1\|$ is the largest singular value which here is $5\sqrt{2}$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Out of these two matrices, A_1 and B, the matrix A_1 is 'closer' to A as $\|A - A_1\| = 5\sqrt{2} < 10\sqrt{2} = \|A - B\|$.

■

Example 5.1.11. From Example 5.1.3, recall the 'bulls eye' matrix

$$A = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix},$$

and its rank two and three approximations A_2 and A_3. Find $\|A - A_2\|$ and $\|A - A_3\|$.

Solution:

- Example 5.1.3 found $A_3 = A$ hence $\|A - A_3\| = \|O_5\| = 0$.
- Although $\|A - A_2\|$ is nontrivial, finding it is straightforward using SVDs. Recall that, from the given SVD $A = USV^{\mathrm{T}}$,

$$\begin{aligned} A_2 &= \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^{\mathrm{T}} \\ &= \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^{\mathrm{T}} + 0 \boldsymbol{u}_3 \boldsymbol{v}_3^{\mathrm{T}} + 0 \boldsymbol{u}_4 \boldsymbol{v}_4^{\mathrm{T}} + 0 \boldsymbol{u}_5 \boldsymbol{v}_5^{\mathrm{T}} \\ &= U \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} V^{\mathrm{T}}. \end{aligned}$$

Hence the difference

$$\begin{aligned} A - A_2 &= U \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 \\ 0 & 0 & 0 & \sigma_4 & 0 \\ 0 & 0 & 0 & 0 & \sigma_5 \end{bmatrix} V^{\mathrm{T}} \\ &\quad - U \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} V^{\mathrm{T}} \\ &= U \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 \\ 0 & 0 & 0 & \sigma_4 & 0 \\ 0 & 0 & 0 & 0 & \sigma_5 \end{bmatrix} V^{\mathrm{T}}. \end{aligned}$$

This is an SVD for $A - A_2$, albeit irregular with the singular values out of order, with singular values of 0, 0, $\sigma_3 = 0.64$,

v0.2a

and $\sigma_4 = \sigma_5 = 0$. The largest singular value gives the norm $\|A - A_2\| = 0.64$ (2 d.p.).

One might further comment that the relative error in the approximate A_2 is $\|A - A_2\|/\|A\| = 0.64/2.68 = 0.24 = 24\%$ (2 d.p.).

■

Theorem 5.1.12 (Eckart–Young). *Let A be an $m \times n$ matrix of* rank r *with* SVD $A = USV^{\mathrm{T}}$. *Then for any $k < r$ a closest rank k matrix approximating A, in the matrix norm, is*

$$A_k := US_kV^{\mathrm{T}} = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^{\mathrm{T}} + \cdots + \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^{\mathrm{T}} \tag{5.2}$$

where $S_k := \operatorname{diag}(\sigma_1, \sigma_2, \dots, \sigma_k, 0, \dots, 0)$. The distance between A and A_k is $\|A - A_k\| = \sigma_{k+1}$.

That is, obtain A_k by 'setting' the singular values $\sigma_{k+1} = \cdots = \sigma_r = 0$ from an SVD for A.

Proof. As a prelude, let's establish the distance between A and A_k. Using their SVDs,

$$\begin{aligned} A - A_k &= USV^{\mathrm{T}} - US_kV^{\mathrm{T}} = U(S - S_k)V^{\mathrm{T}} \\ &= U \operatorname{diag}(0\,, \dots, 0\,, \sigma_{k+1}\,, \dots, \sigma_r\,, 0\,, \dots, 0)V^{\mathrm{T}}, \end{aligned}$$

and so $A - A_k$ has largest singular value σ_{k+1}. Then from Definition 5.1.5, $\|A - A_k\| = \sigma_{k+1}$.

Use contradiction to prove there is no matrix of rank k closer to A when using $\|\cdot\|$ to measure matrix distances (Trefethen & Bau 1997, p.36). Assume there is some $m \times n$ matrix B with $\operatorname{rank} B \le k$ *and* closer to A than is A_k, that is, $\|A - B\| < \|A - A_k\|$. First, the Rank Theorem 3.4.39 asserts the null space of B has dimension nullity $B = n - \operatorname{rank} B \ge n - k$ as $\operatorname{rank} B \le k$. For every $\boldsymbol{w} \in \operatorname{null} B$, as $B\boldsymbol{w} = \boldsymbol{0}$, $A\boldsymbol{w} = A\boldsymbol{w} - B\boldsymbol{w} = (A - B)\boldsymbol{w}$. Then

$$\begin{aligned} |A\boldsymbol{w}| &= |(A - B)\boldsymbol{w}| \\ &\le \|A - B\||\boldsymbol{w}| \quad \text{(by Theorem 5.1.9g)} \\ &< \|A - A_k\||\boldsymbol{w}| \quad \text{(by assumption)} \\ &= \sigma_{k+1}|\boldsymbol{w}| \end{aligned}$$

That is, under the assumption there exists an (at least) $(n - k)$-dimensional subspace in which $|A\boldsymbol{w}| < \sigma_{k+1}|\boldsymbol{w}|$.

Second, consider *any* vector $\boldsymbol{v}$ in the $(k + 1)$-dimensional subspace $\operatorname{span}\{\boldsymbol{v}_1\,, \boldsymbol{v}_2\,, \dots, \boldsymbol{v}_{k+1}\}$. Say $\boldsymbol{v} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_{k+1}\boldsymbol{v}_{k+1} = V\boldsymbol{c}$

v0.2a

for any vector of coefficients $\boldsymbol{c} = (c_1\,, c_2\,, \dots, c_{k+1}\,, 0\,, \dots, 0) \in \mathbb{R}^n$. Then

$$\begin{aligned}
|A\boldsymbol{v}| &= |USV^{\mathrm{T}}V\boldsymbol{c}| \\
&= |US\boldsymbol{c}| \quad (\text{as } V^{\mathrm{T}}V = I) \\
&= |S\boldsymbol{c}| \quad (\text{as } U \text{ is orthogonal}) \\
&= \big|(\sigma_1 c_1\,, \sigma_2 c_2\,, \dots, \sigma_{k+1}c_{k+1}\,, 0\,, \dots, 0)\big| \\
&= \sqrt{\sigma_1^2 c_1^2 + \sigma_2^2 c_2^2 + \cdots + \sigma_{k+1}^2 c_{k+1}^2} \\
&\geq \sqrt{\sigma_{k+1}^2 c_1^2 + \sigma_{k+1}^2 c_2^2 + \cdots + \sigma_{k+1}^2 c_{k+1}^2} \\
&= \sigma_{k+1}\sqrt{c_1^2 + c_2^2 + \cdots + c_{k+1}^2} \\
&= \sigma_{k+1}|\boldsymbol{c}| \\
&= \sigma_{k+1}|V\boldsymbol{c}| \quad (\text{as } V \text{ is orthogonal}) \\
&= \sigma_{k+1}|\boldsymbol{v}|.
\end{aligned}$$

That is, there exists a $(k+1)$-dimensional subspace in which $|A\boldsymbol{v}| \geq \sigma_{k+1}|\boldsymbol{v}|$.

Lastly, since the sum of the dimensions of these two subspaces of $\mathbb{R}^n$ is at least $(n-k)+(k+1) > n$, there must be a nonzero vector, say $\boldsymbol{u}$, lying in both. So for this $\boldsymbol{u}$, simultaneously $|A\boldsymbol{u}| < \sigma_{k+1}|\boldsymbol{u}|$ and $|A\boldsymbol{u}| \geq \sigma_{k+1}|\boldsymbol{u}|$. These two deductions contradict each other. Hence the assumption is wrong: there is no rank k matrix more closely approximating A than A_k. □

Example 5.1.13 (the letter `R`). In displays with low resolution, letters and numbers are displayed with noticeable pixel patterns: for example, the letter `R` is pixellated in the margin. Let's see how such pixel patterns are best approximated by matrices of different ranks. (This example is illustrative: it is not a practical image compression since the required singular vectors are more complicated than a small-sized pattern of pixels.)

Solution: Use Procedure 5.1.2. First, form and enter into MATLAB/Octave the 7×5 matrix of the pixel pattern as illustrated in the margin

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

Second, compute an SVD via `[U,S,V]=svd(R)` to find (2 d.p.)

```
U =
  -0.53   0.38  -0.00  -0.29  -0.70  -0.06  -0.07
  -0.28  -0.49   0.00  -0.13   0.10  -0.69  -0.42
```

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

```
  -0.28  -0.49  -0.00  -0.13  -0.02   0.72  -0.39
  -0.53   0.38  -0.00  -0.29   0.70   0.06   0.07
  -0.32   0.03  -0.71   0.63  -0.00  -0.00  -0.00
  -0.32   0.03   0.71   0.63  -0.00   0.00   0.00
  -0.28  -0.49  -0.00  -0.13  -0.08  -0.02   0.81
S =
   3.47      0      0      0      0
      0   2.09      0      0      0
      0      0   1.00      0      0
      0      0      0   0.75      0
      0      0      0      0   0.00
      0      0      0      0      0
      0      0      0      0      0
V =
  -0.73  -0.32   0.00   0.40  -0.45
  -0.30   0.36  -0.00  -0.76  -0.45
  -0.40   0.37  -0.71   0.07   0.45
  -0.40   0.37   0.71   0.07   0.45
  -0.24  -0.70  -0.00  -0.50   0.45
```

The singular values are $\sigma_1 = 3.47$, $\sigma_2 = 2.09$, $\sigma_3 = 1.00$, $\sigma_4 = 0.75$ and $\sigma_5 = 0$. Four successively better approximations to the image are the following.

- The coarsest approximation is $R_1 = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}}$, that is

$$R_1 = 3.47 \begin{bmatrix} -0.53 \\ -0.28 \\ -0.28 \\ -0.53 \\ -0.32 \\ -0.32 \\ -0.28 \end{bmatrix} \begin{bmatrix} -0.73 & -0.30 & -0.40 & -0.40 & -0.24 \end{bmatrix}.$$

  Compute with `R1=U(:,1)*S(1,1)*V(:,1)'` to find (2 d.p.), as illustrated,

```
R1 =
   1.34   0.55   0.72   0.72   0.44
   0.71   0.29   0.39   0.39   0.24
   0.71   0.29   0.39   0.39   0.24
   1.34   0.55   0.72   0.72   0.44
   0.83   0.34   0.45   0.45   0.27
   0.83   0.34   0.45   0.45   0.27
   0.71   0.29   0.39   0.39   0.24
```

  This has difference $\|R - R_1\| = \sigma_2 = 2.09$ which at 60% of σ_1 is large: indeed the letter `R` is not recognisable.

- The second approximation is $R_2 = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^{\mathrm{T}}$. Compute via `R2=U(:,1:2)*S(1:2,1:2)*V(:,1:2)'` to find (2 d.p.), as illustrated,

```
R2 =
```

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

```
   1.09   0.83   1.02   1.02  -0.11
   1.04  -0.07   0.01   0.01   0.95
   1.04  -0.07   0.01   0.01   0.95
   1.09   0.83   1.02   1.02  -0.11
   0.81   0.36   0.47   0.47   0.24
   0.81   0.36   0.47   0.47   0.24
   1.04  -0.07   0.01   0.01   0.95
```

This has difference $\|R - R_2\| = \sigma_3 = 1.00$ which at 29% of σ_1 is large: but one can begin to imagine the letter `R` in the image.

- The third approximation is

$$R_3 = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^{\mathrm{T}} + \sigma_3 \boldsymbol{u}_3 \boldsymbol{v}_3^{\mathrm{T}} .$$

Compute with `R3=U(:,1:3)*S(1:3,1:3)*V(:,1:3)'` to find (2 d.p.), as illustrated,

```
R3 =
   1.09   0.83   1.02   1.02  -0.11
   1.04  -0.07   0.01   0.01   0.95
   1.04  -0.07   0.01   0.01   0.95
   1.09   0.83   1.02   1.02  -0.11
   0.81   0.36   0.97  -0.03   0.24
   0.81   0.36  -0.03   0.97   0.24
   1.04  -0.07   0.01   0.01   0.95
```

This has difference $\|R - R_3\| = \sigma_4 = 0.75$ which at 22% of σ_1 is moderate and one can see the letter `R` emerging.

- The fourth approximation is

$$R_4 = \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^{\mathrm{T}} + \cdots + \sigma_4 \boldsymbol{u}_4 \boldsymbol{v}_4^{\mathrm{T}} .$$

Compute with `R4=U(:,1:4)*S(1:4,1:4)*V(:,1:4)'` to find (2 d.p.), as illustrated,

```
R4 =
   1.00   1.00   1.00   1.00  -0.00
   1.00   0.00  -0.00   0.00   1.00
   1.00   0.00  -0.00  -0.00   1.00
   1.00   1.00   1.00   1.00  -0.00
   1.00  -0.00   1.00   0.00  -0.00
   1.00  -0.00  -0.00   1.00   0.00
   1.00   0.00  -0.00  -0.00   1.00
```

This has difference $\|R - R_4\| = \sigma_5 = 0.00$ and R_4 exactly reproduces `R`.

■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Example 5.1.14. Recall Example 5.1.4 approximated the image of Euler (1737) with various rank k approximates from an SVD of the image. Let the image be denoted by matrix A. From Figure 5.2 the largest singular value of the image is $\|A\| = \sigma_1 \approx 40,000$.

- From Theorem 5.1.12, the rank 3 approximation in Figure 5.1 is a distance $\|A - A_3\| = \sigma_4 \approx 5,000$ (from Figure 5.2) away from the image. That is, image A_3 has a relative error roughly $5000/40000 = 1/8 \approx 12\%$.
- From Theorem 5.1.12, the rank 10 approximation in Figure 5.1 is a distance $\|A - A_{10}\| = \sigma_{11} \approx 5,000$ (from Figure 5.2) away from the image. That is, image A_{10} has a relative error roughly $2000/40000 = 1/20 = 5\%$.
- From Theorem 5.1.12, the rank 30 approximation in Figure 5.1 is a distance $\|A - A_{30}\| = \sigma_{31} \approx 800$ (from Figure 5.2) away from the image. That is, image A_{30} has a relative error roughly $800/40000 = 1/50 = 2\%$.

■

5.1.3 Principal component analysis

In its 'best' approximation property, Theorem 5.1.12 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).

Example 5.1.15 (toy items)**.** Suppose you are given data about six items, three blue and three red. Suppose each item has two measured properties/attributes called h and v as in the following table:

h	v	colour
-3	-3	blue
-2	1	blue
1	-2	blue
-1	2	red
2	-1	red
3	3	red

The item properties/attributes are the points (h, v) in 2D as illustrated in the margin. But humans always prefer simple one dimensional summaries: we do it all the time when we rank sport teams, schools, web pages, and so on.

Challenge: is there a one dimensional summary of these six item's data that clearly separates the blue from the red? Using just one of the attributes h or v on their own would not suffice:

- using h alone leads to a 1D view where the red and the blue are intermingled as shown in the margin;
- similarly, using v alone leads to a 1D view where the red and the blue are intermingled as shown in the margin.

Solution: Use an SVD to automatically find the best 1D view of the data.

(a) Enter the 6×2 matrix of data into MATLAB/Octave with

```
A=[-3 -3
-2 1
1 -2
-1 2
2 -1
3 3 ]
```

(b) Then `[U,S,V]=svd(A)` computes an SVD, $A = USV^{\mathrm{T}}$, of the data (2 d.p.):

```
U =
  -0.69   0.00  -0.09   0.09   0.14   0.70
  -0.11  -0.50   0.50  -0.50   0.48  -0.08
  -0.11   0.50   0.82   0.18  -0.18   0.01
   0.11  -0.50   0.18   0.82   0.18  -0.01
   0.11   0.50  -0.14   0.14   0.83  -0.09
   0.69  -0.00   0.12  -0.12   0.02   0.70
S =
   6.16      0
      0   4.24
      0      0
      0      0
      0      0
      0      0
V =
   0.71   0.71
   0.71  -0.71
```

(c) Now what does such an SVD tell us? Recall from the proof of the SVD (section 3.3.3) that $A\boldsymbol{v}_1 = \sigma_1\boldsymbol{u}_1$. Further recall from the proof that $\boldsymbol{v}_1$ is the unit vector that maximises $|A\boldsymbol{v}_1|$ so in some sense it is the direction in which the data in A is most spread out ($\boldsymbol{v}_1$ is called the principal vector). We find here (2 d.p.)

$$A\boldsymbol{v}_1 = \sigma_1\boldsymbol{u}_1 = (-4.24, -0.71, -0.71, 0.71, 0.71, 4.24)$$

which neatly separates the blue items (negative) from the red (positive). In essence, the product $A\boldsymbol{v}_1$ orthogonally projects (Section 3.5.3) the items' (h, v) data onto the subspace span$\{\boldsymbol{v}_1\}$ as illustrated in the margin.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Table 5.2: part of Edgar Anderson's Iris data, lengths in cm. The measurements come from the flowers of ten each of three different species of Iris.

Sepal length	Sepal width	Petal length	Petal width	Species
4.9	3.0	1.4	0.2	
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
6.8	2.8	4.8	1.4	
5.5	2.4	3.7	1.0	
6.7	3.1	4.7	1.5	
6.1	3.0	4.6	1.4	
5.7	2.9	4.2	1.3	
5.8	2.7	5.1	1.9	
4.9	2.5	4.5	1.7	
6.4	2.7	5.3	1.9	
6.5	3.0	5.5	1.8	
5.6	2.8	4.9	2.0	Virginia
6.2	2.8	4.8	1.8	
7.9	3.8	6.4	2.0	
6.3	3.4	5.6	2.4	
6.9	3.1	5.1	2.3	
6.3	2.5	5.0	1.9	

■

Although this Example 5.1.15 is just a toy to illustrate concepts, the above steps generalise straightforwardly to be immensely useful on vastly bigger and more challenging data. The next example takes the next step in complexity by introducing how to automatically find a good 2D view of some data in 4D.

v0.2a

Figure 5.3: 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.

Example 5.1.16 (Iris flower data set)**.** Table 5.2 list part of Edgar Anderson's data on the length and widths of sepals and petals of Iris flowers.[2] There are three species of Irises in the data. The data is 4D: each instance of thirty Iris flowers is characterised by the four lengths. 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 5.3 plots the attributes of sepal widths versus sepal lengths and results in the three species being intermingled together rather than reasonably separated. Our aim is to instead plot Figure 5.4 which successfully separates the three species.

Solution: Use an SVD to find a best low-rank view of the data.

(a) Enter the 30×5 matrix of Iris data into MATLAB/Octave with a complete version of

```
iris=[
4.9 3.0 1.4 0.2 1
4.6 3.4 1.4 0.3 1
...
6.3 2.5 5.0 1.9 3
]
```

where the fifth column of $1, 2, 3$ corresponds to the species Setosa, Versicolor or Virginia, respectively. Then a scatter plot such as Figure 5.3 may be drawn with the command

[2] http://archive.ics.uci.edu/ml/datasets/Iris gives the full dataset (Lichman 2013).

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

```
scatter(iris(:,1),iris(:,2),[],iris(:,5))
```

The above `scatter(x,y,[],s)` plots a scatter plot of points with colour depending upon `s` which here corresponds to each different species.

(b) If we were on a walk to a scenic lookout to get a view of the countryside, then the scenic lookout would be in the countryside: it is no good going to a lookout a long way away from the scene we wish to view. Correspondingly, to best view a dataset we typically look it at from the very centre of the data, namely its mean. That is, here we use an SVD of the data matrix only after subtracting the mean of each attribute so that the SVD analyses the variations from the mean. Here the mean Iris sepal length and width is 5.81 cm and 3.09 cm (the black + in Figure 5.3), and the mean petal length and width is 3.69 cm and 1.22 cm. In MATLAB/Octave execute the following to form a matrix A of the variations from the mean, and compute an SVD:

```
meaniris=mean(iris(:,1:4))
A=iris(:,1:4)-ones(30,1)*meaniris
[U,S,V]=svd(A)
```

The resulting SVD is (2 d.p.)

```
U = ...
S =
   10.46       0       0       0
       0    2.86       0       0
       0       0    1.47       0
       0       0       0    0.85
      ...     ...     ...     ...
V =
   0.34   0.72  -0.56  -0.20
  -0.07   0.65   0.74   0.14
   0.87  -0.17   0.14   0.45
   0.36  -0.15   0.33  -0.86
```

where a ... indicates information that is not directly of interest.

(c) As justified shortly, the two most important components of a flower's shape are those in the directions of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ (called the two principal vectors). Because $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are orthonormal, the first component for each Iris flower is $\boldsymbol{x} = A\boldsymbol{v}_1$ and the second component for each is $\boldsymbol{y} = A\boldsymbol{v}_2$. The beautiful Figure 5.4 is a scatter plot of the components of $\boldsymbol{y}$ versus the components of $\boldsymbol{x}$ that untangles the three species. Obtain Figure 5.4 in MATLAB/Octave with the command

```
scatter(A*V(:,1),A*V(:,2),[],iris(:,5))
```

Figure 5.4: best 2D scatter plot of Edgar Anderson's iris data: blue, Setosa; brown, Versicolor; red, Virginia.

Figure 5.4 shows our SVD based analysis largely separates the three species using these two different combinations of the flowers' attributes.

■

Transpose the usual mathematical convention Perhaps you noticed that the previous Example 5.1.16 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 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, and 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 are row vectors.

Definition 5.1.17 (principal components). *Given a $m \times n$ data matrix A (usually with zero mean when averaged over all rows), with* SVD $A = USV^{\mathrm{T}}$ *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 5.1.12). 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 $\boldsymbol{y} = \sigma_2\boldsymbol{u}_2$ (Figure 5.4).

Theorem 5.1.18. *Using the matrix norm to measure 'best', 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.*

Proof. Let $A = USV^{\mathrm{T}}$ be an SVD of matrix A. For any $k < \operatorname{rank} A$, Theorem 5.1.12 establishes that

$$A_k := US_kV^{\mathrm{T}} = \sigma_1\boldsymbol{u}_1\boldsymbol{v}_1^{\mathrm{T}} + \sigma_2\boldsymbol{u}_2\boldsymbol{v}_2^{\mathrm{T}} + \cdots + \sigma_k\boldsymbol{u}_k\boldsymbol{v}_k^{\mathrm{T}}$$

is the best rank k approximation to A in the matrix norm. Letting matrix $U = \big[u_{ij}\big]$, write the ith row of A_k as $(\sigma_1 u_{i1})\boldsymbol{v}_1^{\mathrm{T}} + (\sigma_2 u_{i2})\boldsymbol{v}_2^{\mathrm{T}} + \cdots + (\sigma_k u_{ik})\boldsymbol{v}_k^{\mathrm{T}}$ and hence the transpose of each row of A_k lies in the kD subspace $\operatorname{span}\{\boldsymbol{v}_1\,,\boldsymbol{v}_2\,,\ldots,\boldsymbol{v}_k\}$. This establishes that these are principal vectors.

Since $\{\boldsymbol{v}_1\,,\boldsymbol{v}_2\,,\ldots,\boldsymbol{v}_k\}$ is an orthonormal set, we now use them as standard unit vectors of a coordinate system for the kD subspace. From the above linear combination, the components of the ith data point approximation in this subspace coordinate system are $\sigma_1 u_{i1}\,,\ \sigma_2 u_{i2}\,,\ldots,\sigma_k u_{ik}$. That is, the jth coordinate for all data points, the principal components, is $\sigma_j\boldsymbol{u}_j$. By post-multiplying the SVD $A = USV^{\mathrm{T}}$ by orthogonal V, recall that $AV = US$ which written in terms of columns is

$$\begin{bmatrix} A\boldsymbol{v}_1 & A\boldsymbol{v}_2 & \cdots & A\boldsymbol{v}_r \end{bmatrix} = \begin{bmatrix} \sigma_1\boldsymbol{u}_1 & \sigma_2\boldsymbol{u}_2 & \cdots & \sigma_r\boldsymbol{u}_m r \end{bmatrix}$$

where $r = \operatorname{rank} A$. Consequently, the vector $\sigma_j\boldsymbol{u}_j$ of jth coordinates in the subspace are equal to $A\boldsymbol{v}_j$, the principal components. □

v0.2a

Figure 5.5: for the wine data of Example 5.1.19, a plot of the measured malic acid versus measured alcohol, and coloured depending upon the cultivar, shows these measurements alone cannot effectively discriminate between the cultivars.

Example 5.1.19 (wine recognition). From the Lichman (2013) repository[3] 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?

Solution: Use and SVD to find the best two-dimensional, rank two, view of the data.

(a) Read in the 178×14 matrix of data into MATLAB/Octave with the commands

```
wine=csvread('wine.data')
[m,n]=size(wine)
scatter(wine(:,2),wine(:,3),[],wine(:,1))
```

The scatter plot, Figure 5.5, shows that if we just plot the first two chemicals, alcohol and malic acid, then the three cultivars are inextricably intermingled. Our aim is to automatically draw Figure 5.6 in which the three cultivars are largely separated.

(b) To find the principal components of the wine chemicals it is best to remove the mean with

```
meanw=mean(wine(:,2:14))
A=wine(:,2:14)-ones(m,1)*meanw;
```

[3] http://archive.ics.uci.edu/ml/datasets/Wine

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

where the `mean(X)` computes the mean/average of each column of `X`.

But now a further issue arises: the values in the columns are of widely different magnitudes; moreover, each column has different physical units (in contrast, the Iris flower measurements were all cm). In practice we *must not* mix together quantities with different physical units. The general rule, after making each column zero mean, is to scale each column by dividing by its standard deviation, equivalently by its root-mean-square. This scaling does two practically useful things:

- since the standard deviation measures the spread of data in a column, it has the same physical units as the column of data, so dividing by it renders the results dimensionless, and so suitable for mixing with other scaled columns;
- also the spread of data in each column is now comparable to each other, namely around about size one, instead of some columns being of the order of one-tenths and other columns being in the hundreds.

Consequently, form the 178×13 matrix to analyse by commands

```
meanw=mean(wine(:,2:14))
stdw=std(wine(:,2:14))
A=(wine(:,2:14)-ones(m,1)*meanw)*diag(1./stdw);
```

where the `std(X)` computes the standard deviation of each column of `X`.

(c) Now compute and use an SVD $A = USV^{\mathrm{T}}$. But for low rank approximations we only ever use the first few singular values and first few singular vectors. Thus it is pointless computing a full SVD which here has 178×178 matrix U and 13×13 matrix V.[4] Consequently, use `[U,S,V]=svds(A,4)` to economically compute only the first four singular values and singular vectors (change the four to suit your purpose) to find (2 d.p.)

```
U = ...
S =
   28.86       0       0       0
       0   21.02       0       0
       0       0   16.00       0
       0       0       0   12.75
V =
```

[4] Yes, on modern computers this is here done within a millisecond. But for modern datasets with thousands to billions of rows a full SVD is infeasible so let's see how to analyse modern large datasets.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Figure 5.6: for the wine data of Example 5.1.19, a plot of the first two principal components almost entirely separates the three cultivars.

```
-0.14   0.48  -0.21  -0.02
 0.25   0.22   0.09   0.54
 0.00   0.32   0.63  -0.21
 0.24  -0.01   0.61   0.06
-0.14   0.30   0.13  -0.35
-0.39   0.07   0.15   0.20
-0.42  -0.00   0.15   0.15
 0.30   0.03   0.17  -0.20
-0.31   0.04   0.15   0.40
 0.09   0.53  -0.14   0.07
-0.30  -0.28   0.09  -0.43
-0.38  -0.16   0.17   0.18
-0.29   0.36  -0.13  -0.23
```

where the ... indicates we do not here need to know U.

(d) Recall that the columns of this orthogonal V are the principal vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_4$, and the jth principal components of the data are $\boldsymbol{x}_j = A\boldsymbol{v}_j$. We form a 2D plotted view of the data, Figure 5.6, by drawing a scatter plot of the first two principal components with

```
scatter(A*V(:,1),A*V(:,2),[],wine(:,1))
```

Figure 5.6 shows these two principal components do an amazingly good job of almost completely disentangling the three wine cultivars (use `scatter3` to explore the first three principal components).

■

The previous three examples develop the following procedure for

'best' viewing data in low dimensions. The procedure is automatic. However, additional information about the data or preferred results would lead to modifications.

Procedure 5.1.20 (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', often by dividing by the standard deviation s_j of each column, algebraically $\boldsymbol{a}_j = (\boldsymbol{b}_j - \bar{b}_j)/s_j$.*

   Compute `A=(B-ones(m,1)*mean(B))*diag(1./std(B))` *in* MATLAB/*Octave.*
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 any (usually zero mean) data matrix A, the first k principal vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k$ are orthogonal unit vectors that *maximise the total variance* in the principal components $\boldsymbol{x}_j = A\boldsymbol{v}_j$; that is, that maximise $|\boldsymbol{x}_1|^2 + |\boldsymbol{x}_2|^2 + \cdots + |\boldsymbol{x}_k|^2$. Indeed, this maximisation of the variance is closely tied to the constructive proof of the existence of SVDs (section 3.3.3) which successively maximises $|A\boldsymbol{v}|$ 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.

Application to latent semantic indexing

> This ability to retrieve relevant information based upon meaning rather than literal term usage is the main motivation for using LSI [latent semantic indexing].
>
> *(Berry et al. 1995, p.579)*

Searching for information based upon word matching results in surprisingly poor retrieval of relevant documents (Berry et al. 1995, §5.5). Instead, the so-called latent semantic indexing improves

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

retrieval by replacing individual words with nearness of word vectors derived via the singular value decomposition. This section introduces latent semantic indexing via a very small example.

The Society for Industrial and Applied Mathematics (SIAM) reviews many mathematical books. In 2015 six of those books had the following titles:

1. Introduction to Finite and Spectral Element Methods using MATLAB
2. Iterative Methods for Linear Systems: Theory and Applications
3. Singular Perturbations: Introduction to System Order Reduction Methods with Applications
4. Risk and Portfolio Analysis: Principles and Methods
5. Stochastic Chemical Kinetics: Theory and Mostly Systems Biology Applications
6. Quantum Theory for Mathematicians

Consider the capitalised words. For those words that appear in more than one title, let's form a word vector (Example 1.1.7) for each title, then use principal components to summarise these six books on a 2D plane. This task is part of what is called latent semantic indexing (Berry et al. 1995). (We should also count words that are used only once, but this example omits for simplicity.)

Follow the principal component analysis Procedure 5.1.20.

1. First find the set of words that are used more than once. Ignoring pluralisation, they are: Application, Introduction, Method, System, Theory. The corresponding word vector for each book title is then the following:
   - $\boldsymbol{w}_1 = (0,1,1,0,0)$ Introduction to Finite and Spectral Element Methods using MATLAB
   - $\boldsymbol{w}_2 = (1,0,1,1,1)$ Iterative Methods for Linear Systems: Theory and Applications
   - $\boldsymbol{w}_3 = (1,1,1,1,0)$ Singular Perturbations: Introduction to System Order Reduction Methods with Applications
   - $\boldsymbol{w}_4 = (0,0,1,0,0)$ Risk and Portfolio Analysis: Principles and Methods
   - $\boldsymbol{w}_5 = (1,0,0,1,1)$ Stochastic Chemical Kinetics: Theory and Mostly Systems Biology Applications
   - $\boldsymbol{w}_6 = (0,0,0,0,1)$ Quantum Theory for Mathematicians
2. Second, form the data matrix with $\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_6$ as rows (not columns). We could remove the mean word vector, but

v0.2a

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

v0.2a

in this product. We plot the six books on a 2D plane with the MATLAB/Octave command

```
scatter(A*V(:,1),A*V(:,2),[],1:6)
```

to produce a picture like that in the margin. The SVD analysis nicely distributes the books.

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 5.1.21. 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

Solution: Find the angle two ways.

(a) First, the corresponding 5D word vectors are $\boldsymbol{w}_1 = (0,1,1,0,0)$ and $\boldsymbol{w}_2 = (1,0,1,1,1)$, with lengths $|\boldsymbol{w}_1| = \sqrt{2}$ and $|\boldsymbol{w}_2| = \sqrt{4} = 2$. The dot product then determines

$$\cos\theta = \frac{\boldsymbol{w}_1 \cdot \boldsymbol{w}_2}{|\boldsymbol{w}_1|\,|\boldsymbol{w}_2|} = \frac{0+0+1+0+0}{2\sqrt{2}} = 0.3536\,.$$

Hence the angle $\theta = 69.30°$.

(b) Secondly, estimate the angle using the 2D view. For these two books the principal component vectors are $(0.76, 1.09)$ and $(1.92, -0.40)$, respectively, with lengths 1.33 and 1.96 (2 d.p.). The dot product gives

$$\cos\theta \approx \frac{(0.76, 1.09)\cdot(1.92, -0.40)}{1.33 \cdot 1.96} = \frac{1.02}{2.61} = 0.39\,.$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Hence the angle $\theta \approx 67^\circ$ which is effectively the same as the first exact calculation.

Because of the relatively large 'angle' between these two book titles, we deduce that the two books are quite dissimilar.

■

We can also use the 2D plane to economically measure similarity between the book titles and any other title or words of interest.

Example 5.1.22. Let's ask which of the six books is 'closest' to a book about Applications.

Solution: The word Application has word vector $\boldsymbol{w} = (1, 0, 0, 0, 0)$. So we could do some computations in the original 5D space of word vectors finding precise angles between this word vector and the word vectors of all titles. Alternatively, let's draw a picture in 2D. The Application word vector $\boldsymbol{w}$ projects onto the 2D plane of principal components by computing $\boldsymbol{w} \cdot \boldsymbol{v}_1 = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{v}_1$ and $\boldsymbol{w} \cdot \boldsymbol{v}_2 = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{v}_2$, that is, $\boldsymbol{w}^{\mathrm{T}}V$. Here the Application word vector $\boldsymbol{w} = (1, 0, 0, 0, 0)$, so $\boldsymbol{w}^{\mathrm{T}}V = \begin{bmatrix} 0.52 & -0.20 \end{bmatrix}$, as plotted in the margin. Which of the six books makes the smallest angle with the line through $(0.52, -0.20)$? Visually, books 2 and 5 are closest, and book 2 appears to have slightly smaller angle to the line than book 5. On this data, we deduce that closest to "Application" is book 2: "Iterative Methods for Linear Systems: Theory and 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

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

7. Nonlinear Systems
8. Ordinary Differential Equations
9. Oscillation Theory of Delay Differential Equations
10. Pseudodifferential Operators and Nonlinear Partial Differential Equations
11. Sinc Methods for Quadrature and Differential Equations
12. Stability of Stochastic Differential Equations with Respect to Semi-Martingales
13. the Boundary Integral Approach to Static and Dynamic Contact Problems
14. the Double Mellin–Barnes Type Integrals and their Applications to Convolution Theory

Principal component analysis summarises and relates these titles.

Follow Procedure 5.1.20.

1. The significant (capitalised) words which appear more than once in these titles (ignoring pluralisation) are the fourteen words

$$\text{Algorithm, Application, Differential/tion, Dynamic/al, Equation, Implementation, Integral, Method, Nonlinear, Ordinary, Partial, Problem, System, and Theory.} \tag{5.3}$$

   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

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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 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.

v0.2a

- 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 $\boldsymbol{v}_1$-axis (numbered 3, 8, 10, 11 and 12, their focus is Differential Equations), the other nine are spread out.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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 below.

Books 2 and 14 appear close to the direction of the query vector and so should be returned as a match: these books are no surprise as both their titles have both *Application* and *Theory* in their titles. But the above plot also suggests Book 5 is near to the direction of the query vector, and so is also worth considering despite not having either of the search words in its title! The power of this latent semantic indexing is that it extracts additional titles that are relevant to the query yet share no common words with the query—as commented at the start of this section.

The angles between the query vector and the book title 3D vectors confirm the graphical appearance claimed above. Recall that the dot product determines the angle between vectors (Theorem 1.3.5).

- From the second row of the above product AV, Book 2 has the principal component vector $(1.02, 1.59, 1.00)$ which has length 2.14. Consequently, it is at small angle 15° 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 $(0.29, 1.09, 0.39)$ 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 $(0.16, 1.19, -0.22)$ 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
$$\cos\theta = \frac{(0.16, 1.19, -0.22)\cdot\boldsymbol{q}}{1.20\cdot 0.96} = 0.76\,.$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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 seventy principal components. Berry et al. (1995) reviews how such statistically derived principal word vectors are a more robust indicator of meaning than individual terms. Hence this SVD analysis of documents becomes an effective way of retrieving information from a search without requiring the results actually match any of the words in the search request—the results just need to match cognate words.

5.1.4 Exercises

Exercise 5.1.1. For some 2×2 matrices A the following plots adjoin the product $A\boldsymbol{x}$ to $\boldsymbol{x}$ for a complete range of unit vectors $\boldsymbol{x}$. Use each plot to roughly estimate the norm of the underlying matrix for that plot.

(e)

(f)

Exercise 5.1.2. For the following matrices, use few unit vectors $\boldsymbol{x}$ to determine how the matrix-vector product varies with $\boldsymbol{x}$. Using calculus to find the appropriate maximum, find from definition the norm of the matrices (hint: all norms here are integers).

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

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

(e) $\begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix}$ (f) $\begin{bmatrix} 6 & -1 \\ 4 & 6 \end{bmatrix}$

(g) $\begin{bmatrix} 2 & 4 \\ -7 & -2 \end{bmatrix}$ (h) $\begin{bmatrix} 2 & -4 \\ 1 & -2 \end{bmatrix}$

Exercise 5.1.3. Many properties of a norm may be proved in other ways than those given for Theorem 5.1.9.

(a) Use an SVD factorisation of I_n to prove Theorem 5.1.9b.

(b) Use the existence of an SVD factorisation of A to prove Theorem 5.1.9d.

(c) Use the definition of a norm as a maximum to prove Theorem 5.1.9f.

(d) Use the existence of an SVD factorisation of A to prove Theorem 5.1.9g.

(e) Prove Theorem 5.1.9h using 5.1.9g and the definition of a norm as a maximum.

Exercise 5.1.4. Let $m \times n$ matrix A have in each row and column at most one non-zero element. Argue that there exists an SVD which establishes that the norm $\|A\| = \max_{i,j} |a_{ij}|$.

Exercise 5.1.5. The margin shows a 7×5 pixel image of the letter G. Compute an SVD of the pixel image. By inspecting various rank approximations from this SVD, determine the rank of the approximation to G shown to the right.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Exercise 5.1.6. Write down two different rank two representations of the pixel image of the letter L, as shown in the margin. Compute SVD representations of the letter L. Compare and comment on the various representations.

Exercise 5.1.7 (Sierpinski triangle)**.** Whereas mostly we deal with the smooth geometry of lines, planes and curves, the subject of fractal geometry recognises that there is much in the world around us that has a rich fractured structure: from clouds, and rocks, to the cardiovascular structure within each of us. The Sierpinski triangle, illustrated in the margin, is a simple fractal. Generate such fractals using recursion as in the following MATLAB/Octave code[5]: the recursion is that the next generation image A is computed from the previous generation A.

```
A=1
A=[A 0*A;A A]
A=[A 0*A;A A]
A=[A 0*A;A A]
A=[A 0*A;A A]
imagesc(1-A)
colormap('gray')
axis equal,axis off
```

(a) Add code to this recursion to compute and print the singular values of each generation of the Sierpinski triangle image. What do you conjecture about the number of distinct singular values as a function of generation number k? Test your conjecture for more iterations in the recursion.

(b) Returning to the 16×16 Sierpinski triangle formed after four iterations, use an SVD to form the best rank five approximation of the Sierpinski triangle (as illustrated in the margin, it has a beautiful structure). Comment on why such a rank five approximation may be a reasonable one to draw. What is the next rank for an approximation that is reasonable to draw? Justify.

(c) Modify the code to compute and draw the 256×256 image of the Sierpinski triangle. Use an SVD to generate and draw the best rank nine approximation to the image.

Exercise 5.1.8 (Sierpinski carpet)**.** The Sierpinski carpet is another fractal easily generated by recursion (as illustrated in the margin after three iterations in the recursion).

(a) Modify the recursive MATLAB/Octave code of the previous Exercise 5.1.7 to generate such an image.

[5] Many students would know that a for-loop would more concisely compute the recursion.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(b) For a range of generations of the image, compute the singular values and comment on the apparent patterns in the singular values.

Exercise 5.1.9 (another fractal). Illustrated in the margin is another fractal easily generated by recursion.

(a) Modify the recursive MATLAB/Octave code of Exercise 5.1.7 to generate such an image.

(b) For a range of generations of the image, compute the singular values and comment on the apparent patterns in the singular values.

Exercise 5.1.10.

Ada Lovelace (1815–52)

This is an image of Countess Ada Lovelace: the first computer programmer, she invented, developed and wrote programs for Charles Babbage's analytical engine. Download the 249×178 image. Using an SVD draw various rank approximations to the image. Using the matrix norm to measures errors, what is the smallest rank to reproduce the image to an error of 5%? and of 1%?

http://www.maa.org/sites/default/files/images/upload_library/46/Portraits/Lovelace_Ada.jpg [Sep 2015]

Exercise 5.1.11 (hiding information project). Steganographic methods embed secret messages in images. Some methods are based upon the singular value decomposition (Gorodetski et al. 2001, e.g.). Perhaps the easiest method is to use the unimportant small singular values. For example, Exercise 5.1.10 indicates a rank 32 approximation of the image of Ada Lovelace accurately reproduces the image. Using the SVD $A = USV^{\mathrm{T}}$, let's keep the singular values $\sigma_1, \sigma_2, \ldots, \sigma_{32}$ the same to produce a good image. The unimportant singular values σ_{33}, σ_{34}, $\sigma_{35}, \ldots$ can hide a small message in the image. Suppose the message is forty binary digits such as

$$0000001000000100000110001000011010001111$$

The scheme is to set the unimportant singular values recursively

$$\sigma_{32+k} = \begin{cases} 0.99\sigma_{31+k}, & \text{if } k\text{th digit is one,} \\ 0.96\sigma_{31+k}, & \text{if } k\text{th digit is zero.} \end{cases}$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

These ratios achieve two things: first, the singular values are decreasing as is necessary to maintain the order of the singular vectors in the standard computation of an SVD; second the ratios are sufficiently different to be robustly detected in an image.

- Download the 249×178 image A.
- Compute an SVD, $A = USV^{\mathrm{T}}$.
- Change the singular values matrix S to S' with the first 32 singular values unchanged, and the next 40 singular values encoding the message. The MATLAB/Octave function `cumprod()` neatly computes the recursive product.
- Using the first 72 columns of U and V, compute and draw the new image $B = \texttt{round}(US'V^{\mathrm{T}})$ (as shown in the margin it is essentially the same as the original), where `round()` is the MATLAB/Octave function that rounds the real values to the nearest integer (greyscale values should be in the range zero to 255).

  It is this image that contains the hidden message.

- To check the hidden message is recoverable, compute an SVD of the new image, $B = QDR^{\mathrm{T}}$. Compare the singular values in D, say $\delta_1, \delta_2, \ldots, \delta_{72}$, with those of S': comment on the effect of the rounding in the computation of B. Invent and test MATLAB/Octave commands to extract the hidden message: perhaps use `diff(log( ))` to undo much of the recursive product in the singular values.
- Report on all code, its role, and the results.

Exercise 5.1.12. As in Example 5.1.22, consider the 2D plot of the six books. Add to the plot the word vector corresponding to a query for books relevant to "Introduction". Using angle to measure closeness, which book is closest to "Introduction"? Confirm by computing the angles, in the 2D plane, between "Introduction" and all six books.

Exercise 5.1.13. As in Example 5.1.22, consider the 2D plot of the six books. Add to the plot the word vector corresponding to a query for books relevant to "Application and Method". Using angle to measure closeness, which book is closest to "Application and Method"? Confirm by computing, in the 2D plane, the angles between "Application and Method" and all six books.

Exercise 5.1.14. Reconsider the word vectors of the group of fourteen mathematical books listed in the text. Instead of computing a representation of these books in 3D space, here use principal component analysis to compute a representation in a 2D plane.

(a) Plot the titles of the fourteen books in the 2D plane of the two principal vectors.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(b) Add the vector corresponding to the query of *Differential and Equation*: which books appear to have a small angle to this query in this 2D representation?

(c) Add the vector corresponding to the query of *Application and Theory*: which books appear to have a small angle to this query in this 2D representation?

Exercise 5.1.15. The discussion on latent semantic indexing only considered queries which were "and", such as a search for books relevant to *Application and Theory*. What if we wanted to search for books relevant to *either Application or Theory*? Discuss how such an "or" query might be phrased in terms of the angle between vectors and a multi-D subspace.

Exercise 5.1.16. Table 5.3 lists twenty short reviews about bathrooms of a major chain of hotels.[6] There are about 17 meaningful words common to more than one review: create a list of such words. Then form corresponding word vectors for each review. Use an SVD to best plot these reviews in 2D. Discuss any patterns in the results.

[6] From http://archive.ics.uci.edu/ml

v0.2a

Table 5.3: twenty user reviews of bathrooms in a major chain of hotels. The data comes from the Opinosis Opinion/Review in the UCI Machine Learning Repository.

- The room was not overly big, but clean and very comfortable beds, a great shower and very clean bathrooms
- The second room was smaller, with a very inconvenient bathroom layout, but at least it was quieter and we were able to sleep
- Large comfortable room, wonderful bathroom
- The rooms were nice, very comfy bed and very clean bathroom
- Bathroom was spacious too and very clean
- The bathroom only had a single sink, but it was very large
- The room was a standard but nice motel room like any other, bathroom seemed upgraded if I remember
- The room was quite small but perfectly formed with a super bathroom
- You could eat off the bathroom floor it was so clean
- The bathroom door does the same thing, making the bathroom seem slightly larger
- bathroom spotless and nicely appointed
- The rooms are exceptionally clean and also the bathrooms
- The bathroom was clean and the bed was comfy
- They provide you with great aveda products in the bathroom
- Also, the bathroom was a bit dirty , brown water came out of the bath tub faucet initially and the sink wall by the toilet was dirty
- If your dog tends to be a little disruptive or on the noisy side, there is a bathroom fan that you can keep on to make noise
- The bathroom was big and clean as well
- Also, the bathrooms were quite well set up, with a separate toilet shower to basin, so whilst one guest is showering another can use the basin
- The bathroom was marble and we had luxurious bathrobes and really, every detail attended to
- It was very clean, had a beautiful bathroom, and was comfortable

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

5.2 Regularise linear equations

Section Contents

5.2.1 The SVD illuminates regularisation 515
5.2.2 Tikhonov regularisation 530
5.2.3 Exercises 534

Singularity is almost invariably a clue.
Sherlock Holmes, in The Boscombe Valley Mystery, by Sir Arthur Conan Doyle, 1892

Often we need to approximate matrices involved in solving linear equations. This is especially so when the matrix itself comes from experimental measurements and so is subject to errors. We do not want such errors to affect results. By avoiding division with small singular values, the procedure developed in this section avoids unwarranted magnification of errors. Sometimes such error magnification is disastrous, so avoiding it is essential.

Example 5.2.1. Suppose from some experiment we to solve the linear equations

$$0.5x + 0.3y = 1 \quad \text{and} \quad 1.1x + 0.7y = 2\,,$$

where all the coefficients on *both* the left-hand sides and the right-hand sides are determined from experimental measurements. In particular, suppose they are measured to errors ± 0.05. Solve the equations.

Solution:

- Following Procedure 2.2.5 in MATLAB/Octave, form the matrix and the right-hand side

```
A=[0.5 0.3;1.1 0.7]
b=[1.0;2.0]
```

  Then check the condition number with `rcond(A)` to find it is 0.007 which previously we would call only just outside the 'good' range (Procedure 2.2.5). So proceed to compute the solution with `A\b` to find $(x, y) = (5, -5)$ (as illustrated in the margin).

- Is this solution reasonable? No. Not when the matrix itself has errors. Let's perturb the matrix A by amounts consistent with its experimental error of ± 0.05 and explore the predicted solutions (the first two illustrated):

$$A = \begin{bmatrix} 0.47 & 0.29 \\ 1.06 & 0.68 \end{bmatrix} \implies \boldsymbol{x} = \texttt{A\textbackslash b} = (8.2, -9.8);$$

$$A = \begin{bmatrix} 0.45 & 0.32 \\ 1.05 & 0.67 \end{bmatrix} \implies \boldsymbol{x} = \texttt{A\textbackslash b} = (-0.9, 4.3);$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$A = \begin{bmatrix} 0.46 & 0.31 \\ 1.06 & 0.73 \end{bmatrix} \implies \boldsymbol{x} = \texttt{A\b} = (15.3, -19.4).$$

For equally valid matrices A, the predicted solutions are all over the place!

- If the matrix itself has errors, then we must reconsider Procedure 2.2.5. The SVD empowers a sensible resolution. Compute an SVD of the matrix, $A = USV^{\mathrm{T}}$, with `[U,S,V]=svd(A)` to find (2 d.p.)

$$A = \begin{bmatrix} -0.41 & -0.91 \\ -0.91 & 0.41 \end{bmatrix} \begin{bmatrix} 1.43 & 0 \\ 0 & 0.01 \end{bmatrix} \begin{bmatrix} -0.85 & -0.53 \\ -0.53 & 0.85 \end{bmatrix}^{\mathrm{T}}$$

  Because the matrix A has errors ± 0.05, the small singular value of 0.01 might as well be zero: it is zero to experimental error. The appropriate solution algorithm is then Procedure 3.5.4 for inconsistent equations (not Procedure 2.2.5). Thus (2 d.p.)

  (a) $\boldsymbol{z} = U^{\mathrm{T}}\boldsymbol{b} = (-2.23, -0.10)$;

  (b) due to the approximately zero singular value, neglect $z_2 = -0.10$ as an error, and solve $\begin{bmatrix} 1.43 & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{y} = \begin{bmatrix} -2.23 \\ 0 \end{bmatrix}$ to deduce $\boldsymbol{y} = (-1.56, y_2)$;

  (c) consequently, we find reasonable solutions are $\boldsymbol{x} = V\boldsymbol{y} = (1.32, 0.83) + y_2(-0.53, 0.85)$.

- To choose between this infinitude of solutions, extra information must be provided by the context/application/modelling. For example, often one prefers the solution of smallest length/magnitude, obtained by setting $y_2 = 0$ (Theorem 3.5.12); that is, $\boldsymbol{x}_{\text{smallest}} = (1.32, 0.83)$.

■

5.2.1 The SVD illuminates regularisation

> I think it is much more interesting to live with uncertainty than to live with answers that might be wrong.
>
> *Richard Feynman*

Procedure 5.2.2 (approximate linear equations). *Suppose the system of linear equation $A\boldsymbol{x} = \boldsymbol{b}$ arises from 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 entries are of size e (here "e" denotes "error", not the exponential e)*

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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 size; and*
   - *similarly for $\boldsymbol{b}$.*

   Determine the error e 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 e; that is, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k > e > \sigma_{k+1} \geq \cdots \geq 0$. Then the rank k approximation to A is*

$$\begin{aligned} A_k &:= US_kV^{\mathrm{T}} \\ &= \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^{\mathrm{T}} + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^{\mathrm{T}} + \cdots + \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^{\mathrm{T}} \\ &= \texttt{U(:,1:k)*S(1:k,1:k)*V(:,1:k)'}\,. \end{aligned}$$

   We usually do not construct A_k as we only need its SVD *to solve the system.*
4. *Solve the approximating linear equation $A_k\boldsymbol{x} = \boldsymbol{b}$ as in Theorems 3.5.7–3.5.12 (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 the needs of the application: often the smallest solution overall; or equally often a solution with the most zero components.*

That is, treat as zero any singular values smaller than the expected error in the matrix entries. For example, in computation on modern computers (with nearly sixteen significant decimal digits accuracy) any singular value smaller than $10^{-8}\sigma_1$ should be treated as zero. Certainly, any smaller than $10^{-16}\sigma_1$ must be treated as zero.

The final step in Procedure 5.2.2 arises because in many cases an infinite number of possible solutions are derived. The linear algebra cannot presume which is best for your application. Consequently, you will have to be aware of the freedom, and make a choice based on extra information from your particular application.

- For example, in a CTscan such as Example 3.5.16 one would usually prefer the greyest result in order to avoid diagnosing artifices.
- For another example, in the data mining task of fitting curves or surfaces to data, one would instead usually prefer a curve or surface with fewest non-zero coefficients.

Such extra information from the application is essential.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Example 5.2.3. For the following matrices A and right-hand side vectors $\boldsymbol{b}$, 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 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) $A = \begin{bmatrix} -0.2 & -0.6 & 1.8 \\ 0.0 & 0.2 & -0.4 \\ -0.3 & 0.7 & 0.3 \end{bmatrix}$, $\boldsymbol{b} = \begin{bmatrix} -0.5 \\ 0.1 \\ -0.2 \end{bmatrix}$

Solution: Enter the matrix and vector into MATLAB/Octave and solve with `x=A\b` to determine $\boldsymbol{x} = (0.06, -0.13, -0.31)$. To within the experimental error of ± 0.05 the following matrices approximate A: that is, they might have been what was measured for A. Then MATLAB/Octave `x=A\b` gives the corresponding equally valid solutions.

- $A' = \begin{bmatrix} -0.16 & -0.58 & 1.83 \\ 0.01 & 0.16 & -0.45 \\ -0.28 & 0.74 & 0.30 \end{bmatrix}$ gives $\boldsymbol{x}' = \begin{bmatrix} 0.85 \\ 0.12 \\ -0.16 \end{bmatrix}$
- $A'' = \begin{bmatrix} -0.22 & -0.62 & 1.77 \\ 0.01 & 0.17 & -0.42 \\ -0.26 & 0.66 & 0.26 \end{bmatrix}$ gives $\boldsymbol{x}'' = \begin{bmatrix} 0.42 \\ -0.04 \\ -0.24 \end{bmatrix}$

There are major differences between these equally valid solutions $\boldsymbol{x}$, $\boldsymbol{x}'$ and $\boldsymbol{x}''$. The problem is that, relative to the experimental error, there is a small singular value in the matrix A. We must use an SVD to find all solutions consistent with the experimental error. Consequently, compute `[U,S,V]=svd(A)` to find (2 d.p.)

```
U =
  -0.97   0.03  -0.23
   0.22  -0.10  -0.97
  -0.05  -0.99   0.09
S =
   1.96      0      0
      0   0.82      0
      0      0   0.02
V =
   0.11   0.36   0.93
   0.30  -0.90   0.31
  -0.95  -0.25   0.20
```

The singular value 0.02 is less than the error ± 0.05 so is effectively zero. Hence we solve the system as if this singular value is zero; that is, as if matrix A has rank two. Compute the smallest consistent solution with the three steps `z=U(:,1:2)'*b`, `y=z./diag(S(1:2,1:2))`, and `x=V(:,1:2)*y`. Then add an arbitrary multiple of the last column of `V`

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

to determine a general solution $\boldsymbol{x} = (0.10, -0.11, -0.30) + t(0.93, 0.31, 0.20)$.

(b) $A = \begin{bmatrix} -1.1 & 0.1 & 0.7 & -0.1 \\ 0.1 & -0.1 & 1.2 & -0.6 \\ 0.8 & -0.2 & 0.4 & -0.8 \\ 0.8 & 0.1 & -2.0 & 1.0 \end{bmatrix}$, $\boldsymbol{b} = \begin{bmatrix} -1.1 \\ -0.1 \\ 1.1 \\ 0.8 \end{bmatrix}$

Solution: Enter the matrix and vector into MATLAB/Octave and solve with `x=A\b` to determine $\boldsymbol{x} = (0.61, -0.64, -0.65, -0.93)$. To within experimental error the following matrices approximate A, and MATLAB/Octave `x=A\b` gives the corresponding solutions.

- $A' = \begin{bmatrix} -1.10 & 0.11 & 0.67 & -0.08 \\ 0.08 & -0.10 & 1.17 & -0.59 \\ 0.75 & -0.21 & 0.39 & -0.83 \\ 0.79 & 0.08 & -1.96 & 0.98 \end{bmatrix}$, $\boldsymbol{x}' = \begin{bmatrix} 0.64 \\ -0.40 \\ -0.64 \\ -0.95 \end{bmatrix}$
- $A'' = \begin{bmatrix} -1.10 & 0.08 & 0.66 & -0.14 \\ 0.08 & -0.09 & 1.22 & -0.58 \\ 0.77 & -0.18 & 0.39 & -0.78 \\ 0.75 & 0.11 & -2.01 & 1.04 \end{bmatrix}$, $\boldsymbol{x}'' = \begin{bmatrix} 0.87 \\ 1.09 \\ -0.58 \\ -1.09 \end{bmatrix}$

There are significant differences, mainly in the second component, between these equally valid solutions $\boldsymbol{x}$, $\boldsymbol{x}'$ and $\boldsymbol{x}''$. The problem is that, relative to the experimental error, there is a small singular value in the matrix A. We must use an SVD to find all solutions consistent with the experimental error: compute `[U,S,V]=svd(A)` to find (2 d.p.)

```
U =
  -0.33   0.59  -0.36   0.64
  -0.43  -0.31   0.71   0.46
  -0.16  -0.74  -0.59   0.27
   0.82  -0.07   0.12   0.55
S =
   2.89      0      0      0
      0   1.50      0      0
      0      0   0.26      0
      0      0      0   0.02
V =
   0.30  -0.88   0.35   0.09
   0.04   0.15   0.09   0.98
  -0.85  -0.08   0.52   0.00
   0.43   0.43   0.78  -0.16
```

The singular value 0.02 is less than the error ± 0.05 so is effectively zero. Hence solve the system as if this singular value is zero; that is, as if matrix A has rank three. Compute the smallest consistent solution with `z=U(:,1:3)'*b`, `y=z./diag(S(1:3,1:3))`, and `x=V(:,1:3)*y`. Then add an

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

arbitrary multiple of the last column of `V` to determine a general solution $\boldsymbol{x} = (0.65, -0.22, -0.65, -1.00) + t(0.09, 0.98, 0, -0.16)$.

That the second component of $(0.09, 0.98, 0, -0.16)$ is the largest corresponds to the second component in each of $\boldsymbol{x}$, $\boldsymbol{x}'$ and $\boldsymbol{x}''$ being the most sensitive, as seen in the above three cases.

Both of these examples gave an infinite number of solutions which are equally valid as far as the linear algebra is concerned. In each example, more information from an application would be needed to choose which of the infinity of solutions is preferred. ■

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 5.2.4 (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 $\mathbf{1} := (1, 1, 1, 1, 1)$.

Solution: Enter the matrix A into MATLAB/Octave with `A=hilb(5)` for the above 5×5 Hilbert matrix, and enter the right-hand side with `b=ones(5,1)`.

- First assume there is insignificant error in A (there is always the base error of 10^{-15} in computation). Then Procedure 2.2.5 finds that although the reciprocal of the condition number `rcond(A)` $\approx 10^{-6}$ is bad, the unique solution to $A\boldsymbol{x} = \mathbf{1}$, obtained via `x=A\b`, is

$$\boldsymbol{x} = (5, -120, 630, -1120, 630).$$

- Second suppose the errors in A are roughly 10^{-5}. This level of error is a concern as `rcond` $\approx 10^{-6}$ so errors would be magnified by 10^6 in a direct solution of $A\boldsymbol{x} = \mathbf{1}$. Here we explore when all errors are in A and none in the right-hand side vector $\mathbf{1}$. To explore, adopt Procedure 5.2.2.

  (a) Find an SVD $A = USV^{\mathrm{T}}$ via `[U,S,V]=svd(A)` (2 d.p.)

v0.2

```
U =
  -0.77   0.60  -0.21   0.05   0.01
  -0.45  -0.28   0.72  -0.43  -0.12
  -0.32  -0.42   0.12   0.67   0.51
  -0.25  -0.44  -0.31   0.23  -0.77
  -0.21  -0.43  -0.57  -0.56   0.38
S =
   1.57      0      0      0      0
      0   0.21      0      0      0
      0      0   0.01      0      0
      0      0      0   0.00      0
      0      0      0      0   0.00
V =
  -0.77   0.60  -0.21   0.05   0.01
  -0.45  -0.28   0.72  -0.43  -0.12
  -0.32  -0.42   0.12   0.67   0.51
  -0.25  -0.44  -0.31   0.23  -0.77
  -0.21  -0.43  -0.57  -0.56   0.38
```

More informatively, the singular values have the following wide range of magnitudes,

$$\sigma_1 = 1.57, \quad \sigma_2 = 0.21, \quad \sigma_3 = 1.14 \cdot 10^{-2},$$
$$\sigma_4 = 3.06 \cdot 10^{-4}, \quad \sigma_5 = 3.29 \cdot 10^{-6}.$$

(b) Because the assumed error 10^{-5} satisfies $\sigma_4 > 10^{-5} > \sigma_5$ the matrix A is effectively of rank four, $k = 4$.

(c) Solving the system $A\boldsymbol{x} = USV^{\mathrm{T}}\boldsymbol{x} = \mathbf{1}$ as rank four, in the least square sense, Procedure 3.5.4 gives (2 d.p.)

i. $\boldsymbol{z} = U^{\mathrm{T}}\mathbf{1} = (-2.00, -0.97, -0.24, -0.04, 0.00)$,

ii. neglect the fifth component of $\boldsymbol{z}$ as an error and obtain the first four components of $\boldsymbol{y}$ via `y=z(1:4)./diag(S(1:4,1:4)` so that

$$\boldsymbol{y} = (-1.28, -4.66, -21.43, -139.69, y_5),$$

iii. then the smallest, least square, solution determined with `x=V(:,1:4)*y` is

$$\boldsymbol{x} = (-3.82, 46.78, -93.41, -23.53, 92.27),$$

and a general solution includes the arbitrary multiple y_5 of the last column of V to be

$$\boldsymbol{x} = (-3.82, 46.78, -93.41, -23.53, 92.27) + y_5(0.01, -0.12, 0.51, -0.77, 0.38).$$

- Third suppose the errors in A are roughly 10^{-3}. Re-adopt Procedure 5.2.2.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(a) Use the same SVD, $A = USV^{\mathrm{T}}$.

(b) Because the assumed error 10^{-3} satisfies $\sigma_3 > 10^{-3} > \sigma_4$ the matrix A is effectively of rank three, $k = 3$.

(c) Solving the system $A\boldsymbol{x} = USV^{\mathrm{T}}\boldsymbol{x} = \mathbf{1}$ as rank three, in the least square sense, Procedure 3.5.4 gives (2 d.p.) the same $\boldsymbol{z}$, and the same first three components in

$$\boldsymbol{y} = (-1.28, -4.66, -21.43, y_4, y_5),$$

then the smallest, least square, solution determined with `x=V(:,1:3)*y` is

$$\boldsymbol{x} = (2.76, -13.66, -0.19, 9.03, 14.38),$$

and a general solution includes the arbitrary multiples of the last columns of V to be

$$\begin{aligned}\boldsymbol{x} = {} & (2.76, -13.66, -0.19, 9.03, 14.38) \\ & + y_4(0.05, -0.43, 0.67, 0.23, -0.56) \\ & + y_5(0.01, -0.12, 0.51, -0.77, 0.38).\end{aligned}$$

- Lastly suppose the errors in A are roughly 0.05. Re-adopt Procedure 5.2.2.

  (a) Use the same SVD, $A = USV^{\mathrm{T}}$.

  (b) Because the assumed error 0.05 satisfies $\sigma_2 > 0.05 > \sigma_3$ the matrix A is effectively of rank two, $k = 2$.

  (c) Solving the system $A\boldsymbol{x} = USV^{\mathrm{T}}\boldsymbol{x} = \mathbf{1}$ as rank two, in the least square sense, Procedure 3.5.4 gives (2 d.p.) the same $\boldsymbol{z}$, and the same first two components in

  $$\boldsymbol{y} = (-1.28, -4.66, y_3, y_4, y_5),$$

  then the smallest, least square, solution determined with `x=V(:,1:2)*y` is

  $$\boldsymbol{x} = (-1.83, 1.85, 2.39, 2.39, 2.27),$$

  and a general solution includes the arbitrary multiples of the last columns of V to be

  $$\begin{aligned}\boldsymbol{x} = {} & (-1.83, 1.85, 2.39, 2.39, 2.27) \\ & + y_3(-0.21, 0.72, 0.12, -0.31, -0.57) \\ & + y_4(0.05, -0.43, 0.67, 0.23, -0.56) \\ & + y_5(0.01, -0.12, 0.51, -0.77, 0.38).\end{aligned}$$

The level of error makes a major difference in the qualitative nature of allowable solutions: here from a unique solution through to a three parameter family of equally valid solutions. To appropriately solve systems of linear equations we must know the level of error.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

■

Example 5.2.5 (translating temperatures). Recall Example 2.2.12 attempts to fit a quartic polynomial to observations (plotted in the margin) of the relation between Celsius and Fahrenheit temperature. The attempt failed because `rcond` is too small. Let's try again now that we can cater for matrices with errors. Recall the data between temperatures reported by a European and an American are the following:

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 $\pm 0.5°$.

Solution: Now invoke Procedure 5.2.2 to approximate the system of linear equations, and solve the approximate problem.

(a) There is a problem in approximating the matrix: the columns are of wildly different sizes. In contrast, our mathematical analysis treats all columns the same. The problem is that each column comes from different powers of temperatures. To avoid the problem we must scale the temperature data for the matrix. The simplest is to divide by a typical temperature. That is, instead of seeking a fit in terms of powers of T_E, we seek a fit in powers of $T_E/20°$ as 20 degrees is a typical temperature in the data. Here we fit the data with the quartic polynomial

$$T_A = c_1 + c_2 \frac{T_E}{20} + c_3 \left(\frac{T_E}{20}\right)^2 + c_4 \left(\frac{T_E}{20}\right)^3 + c_5 \left(\frac{T_E}{20}\right)^4,$$

which gives the following system for the coefficients (2 d.p.)

$$\begin{bmatrix} 1 & 0.75 & 0.56 & 0.42 & 0.32 \\ 1 & 1.30 & 1.69 & 2.20 & 2.86 \\ 1 & 0.55 & 0.30 & 0.17 & 0.09 \\ 1 & 1.15 & 1.32 & 1.52 & 1.75 \\ 1 & 1.35 & 1.82 & 2.46 & 3.32 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_3 \\ c_4 \\ c_5 \end{bmatrix} = \begin{bmatrix} 60 \\ 80 \\ 51 \\ 74 \\ 81 \end{bmatrix}.$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Now all the components of the matrix are roughly the same size, as required.

There is no need to scale the right-hand side vector as all components are all roughly the same size, they are all simply 'American temperatures'.

In script construct the scaled matrix and right-hand side vector with

```
te=[15;26;11;23;27]
ta=[60;80;51;74;81]
tes=te/20
A=[ones(5,1) tes tes.^2 tes.^3 tes.^4]
```

(b) Compute an SVD, $A = USV^{\mathrm{T}}$, with `[U,S,V]=svd(A)` to get (2 d.p.)

```
U =
  -0.16   0.64   0.20   0.72  -0.12
  -0.59  -0.13  -0.00  -0.15  -0.78
  -0.10   0.67  -0.59  -0.45   0.05
  -0.42   0.23   0.68  -0.42   0.36
  -0.66  -0.28  -0.39   0.29   0.49
S =
   7.26      0      0      0      0
      0   1.44      0      0      0
      0      0   0.21      0      0
      0      0      0   0.02      0
      0      0      0      0   0.00
V =
  -0.27   0.78  -0.49  -0.27   0.09
  -0.32   0.39   0.36   0.66  -0.42
  -0.40   0.09   0.55  -0.09   0.72
  -0.50  -0.17   0.25  -0.62  -0.52
  -0.65  -0.44  -0.51   0.32   0.14
```

(c) Now choose the effective rank of the matrix to be the number of singular values bigger than the error. Here recall that the temperatures in the matrix have been divided by 20°. Hence the errors of roughly ±0.5° in each temperature becomes roughly $\pm 0.5/20 = \pm 0.025$ in the scaled components in the matrix. There are three singular values larger than the error 0.025, so the matrix effectively has rank three. The two singular values less than the error 0.025 are effectively zero. That is, although it is not necessary to construct, we

v0.2a

approximate the matrix A by (2 d.p.)

$$A_3 = US_3V^{\mathrm{T}} = \begin{bmatrix} 1 & 0.74 & 0.56 & 0.43 & 0.31 \\ 1 & 1.30 & 1.69 & 2.20 & 2.86 \\ 1 & 0.56 & 0.30 & 0.16 & 0.09 \\ 1 & 1.16 & 1.32 & 1.52 & 1.75 \\ 1 & 1.35 & 1.82 & 2.46 & 3.32 \end{bmatrix} :$$

the differences between this and A are only ± 0.01, so matrix A_3 is indeed close to A.

(d) Solve the equations as if matrix A has rank three.

i. Find $\boldsymbol{z} = U'\boldsymbol{b}$ via `z=U'*ta` to find

```
z =
  -146.53
    56.26
     0.41
     1.06
    -0.69
```

As matrix A has effective rank of three, we approximate the right-hand side data by neglecting the last two components in this $\boldsymbol{z}$. That the last two components in $\boldsymbol{z}$ are small compared to the others indicates this neglect is a reasonable approximation.

ii. Find $\boldsymbol{y}$ by solving $S\boldsymbol{y} = \boldsymbol{z}$ as a rank three system via `y=z(1:3)./diag(S(1:3,1:3))` to find (2 d.p.)

$$\boldsymbol{y} = (-20.19, 39.15, 1.95, y_4, y_5).$$

The smallest solution would be obtained by setting $y_4 = y_5 = 0$.

iii. Finally determine the coefficients $\boldsymbol{c} = V\boldsymbol{y}$ with command `c=V(:,1:3)*y` and then add arbitrary multiples of the remaining columns of V to obtain the general solution (2 d.p.)

$$\boldsymbol{c} = \begin{bmatrix} 35.09 \\ 22.50 \\ 12.77 \\ 3.99 \\ -5.28 \end{bmatrix} + y_4 \begin{bmatrix} -0.27 \\ 0.66 \\ -0.09 \\ -0.62 \\ 0.32 \end{bmatrix} + y_5 \begin{bmatrix} 0.09 \\ -0.42 \\ 0.72 \\ -0.52 \\ 0.14 \end{bmatrix}.$$

(e) Obtain the solution with smallest coefficients by setting $y_4 = y_5 = 0$. This would fit the data with the quartic polynomial

$$T_A = 35.09 + 22.50\frac{T_E}{20} + 12.77\left(\frac{T_E}{20}\right)^2 + 3.99\left(\frac{T_E}{20}\right)^3 - 5.28\left(\frac{T_E}{20}\right)^4.$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

But choosing the polynomial with smallest coefficients has little meaning in this application. Surely we prefer a polynomial with fewer terms, fewer non-zero coefficients. Surely we would prefer, say, the quadratic

$$T_A = 25.93 + 49.63\frac{T_E}{20} - 6.45\left(\frac{T_E}{20}\right)^2,$$

as plotted in the margin.

In this application, let's use the freedom in y_4 and y_5 to set two of the coefficients to zero in the quartic. Since $c_4 = 3.99$ and $c_5 = -5.28$ are the smallest coefficients, and because they correspond to the highest powers in the quartic, it is natural to choose to make them both zero. Let's redo the last step in the procedure. [7]

The last step in the procedure is to solve $V^{\mathrm{T}}\boldsymbol{c} = \boldsymbol{y}$ for $\boldsymbol{c}$. With the last two components of $\boldsymbol{c}$ set to zero, from the computed SVD this is the system of equations (2 d.p.)

$$\begin{bmatrix} -0.27 & -0.32 & -0.40 & -0.50 & -0.65 \\ 0.78 & 0.39 & 0.09 & -0.17 & -0.44 \\ -0.49 & 0.36 & 0.55 & 0.25 & -0.51 \\ -0.27 & 0.66 & -0.09 & -0.62 & 0.32 \\ 0.09 & -0.42 & 0.72 & -0.52 & 0.14 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -20.19 \\ 39.15 \\ 1.95 \\ y_4 \\ y_5 \end{bmatrix},$$

where y_4 and y_5 can be anything for equally good solutions. Considering only the first three rows of this system, and using the zeros in $\boldsymbol{c}$, this system becomes

$$\begin{bmatrix} -0.27 & -0.32 & -0.40 \\ 0.78 & 0.39 & 0.09 \\ -0.49 & 0.36 & 0.55 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} -20.19 \\ 39.15 \\ 1.95 \end{bmatrix}.$$

This is a basic system of three equations for three unknowns. Since the matrix is the first three rows and columns of V^{T} and the right-hand side is the three components of $\boldsymbol{y}$ already computed, we solve the equation by

- checking the condition number, `rcond(V(1:3,1:3))` is 0.05 which is good, and
- then `c=V(1:3,1:3)'\y` determines the coefficients $(c_1, c_2, c_3) = (25.93, 49.63, -6.45)$ (2 d.p.).

That is, a just as good polynomial fit, consistent with errors in the data, is the simpler quadratic polynomial

$$T_A = 25.93 + 49.63\frac{T_E}{20} - 6.45\left(\frac{T_E}{20}\right)^2,$$

[7] Alternatively, one could redo all the linear algebra to seek a quadratic from the outset rather than a quartic. The two alternative answers for a quadratic are not the same, but they are nearly the same. The small differences in the answers are because one modifies the matrix by recognising its errors, and the other does not.

© AJ Roberts, ORCID:0000-0001-6930-1992, September 1, 2010

as plotted previously in the margin.

■

Occam's razor: Non sunt multiplicanda entia sine necessitate [Entities must not be multiplied beyond necessity]
John Punch (1639)

Example 5.2.6. Recall that Exercise 3.5.16 introduced extra 'diagonal' measurements into a 2D CT-scan. As shown in the margin, the 2D region is divided into a 3×3 grid of nine blocks. Then measurements taken of the X-rays not absorbed along the shown nine paths: three horizontal, three vertical, and three diagonal. Suppose the measured fractions of X-ray energy are $\boldsymbol{f} = (0.048, 0.081, 0.042, 0.020, 0.106, 0.075, 0.177, 0.181, 0.105)$. Use an SVD to find the 'grayest' transmission factors consistent with the measurements and likely errors.

```
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
 0 1 0 1 0 0 0 0 0
 0 0 1 0 1 0 1 0 0
 0 0 0 0 0 1 0 1 0]
b=log([0.048
 0.081
 0.042
 0.020
 0.106
 0.075
 0.177
 0.181
 0.105])
[U,S,V]=svd(A)
z=U'*b
y=z(1:7)./diag(S(1:7,1:7))
x=V(:,1:7)*y
r=reshape(exp(A\b),3,3)
r7=reshape(exp(x),3,3)
```

Solution: Nine X-ray measurements are made through the body where $f_1, f_2, \dots, f_9$ denote the fraction of energy in the measurements relative to the power of the X-ray beam. Thus we need to solve nine equations for the nine unknown transmission factors:

$$r_1r_2r_3 = f_1\,, \quad r_4r_5r_6 = f_2\,, \quad r_7r_8r_9 = f_3\,,$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$r_1r_4r_7 = f_4\,, \quad r_2r_5r_8 = f_5\,, \quad r_3r_6r_9 = f_6\,,$$
$$r_2r_4 = f_7\,, \quad r_3r_5r_7 = f_8\,, \quad r_6r_8 = f_9\,.$$

Turn such nonlinear equations into linear equations by taking the logarithm (to any base, but here say the natural logarithm to base e) of both sides of all equations:

$$r_ir_jr_k = f_l \iff (\log r_i) + (\log r_j) + (\log r_k) = (\log f_l).$$

That is, letting new unknowns $x_i = \log r_i$ and new right-hand sides $b_i = \log f_i$, we aim to solve a system of nine linear equations for nine unknowns:

$$\begin{aligned}
x_1 + x_2 + x_3 &= b_1\,, & x_4 + x_5 + x_6 &= b_2\,, & x_7 + x_8 + x_9 &= b_3\,,\\
x_1 + x_4 + x_7 &= b_4\,, & x_2 + x_5 + x_8 &= b_5\,, & x_3 + x_6 + x_9 &= b_6\,,\\
x_2 + x_4 \qquad &= b_7\,, & x_3 + x_5 + x_7 &= b_8\,, & x_6 + x_8 \qquad &= b_9\,.
\end{aligned}$$

These forms the matrix-vector system $A\boldsymbol{x} = \boldsymbol{b}$ for 9×9 matrix

$$A = \begin{bmatrix}
1&1&1&0&0&0&0&0&0\\
0&0&0&1&1&1&0&0&0\\
0&0&0&0&0&0&1&1&1\\
1&0&0&1&0&0&1&0&0\\
0&1&0&0&1&0&0&1&0\\
0&0&1&0&0&1&0&0&1\\
0&1&0&1&0&0&0&0&0\\
0&0&1&0&1&0&1&0&0\\
0&0&0&0&0&1&0&1&0
\end{bmatrix}, \quad \boldsymbol{b} = \log\begin{bmatrix}.048\\.081\\.042\\.020\\.106\\.075\\.177\\.181\\.105\end{bmatrix} = \begin{bmatrix}-3.04\\-2.51\\-3.17\\-3.91\\-2.24\\-2.59\\-1.73\\-1.71\\-2.25\end{bmatrix}.$$

Implement Procedure 5.2.2.

(a) Here there is no need to scale the vector $\boldsymbol{b}$ as all entries are also the same. There is no need to scale the matrix A as all entries mean the same, namely simply zero or one depending upon whether beam passes through the a pixel square. However, the entries of A are in error in two ways.

- A diagonal beam has a path through a pixel that is up to 41% longer than horizontal or vertical beam–which is not accounted for. Further, a beam has finite width so it will also pass through part of some off-diagonal pixels—which is not represented.
- Similarly, a horizontal or vertical beam has finite width and may underrepresent the sides of the pixels it goes through, and/or involve parts of neighbouring pixels—neither effect is represented.

Consequently the entries in the matrix A could easily have error of 0.5.

(b) Compute an SVD, $A = USV^{\mathrm{T}}$, via `[U,S,V]=svd(A)` (2 d.p.)

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

```
U =
  0.33 -0.41  0.27  0.21  0.54  0.23 -0.29  0.13 -0.41
  0.38 -0.00 -0.47  0.36 -0.45 -0.19 -0.00  0.32 -0.41
  0.33  0.41  0.27 -0.57 -0.08  0.23  0.29  0.13 -0.41
  0.33 -0.41  0.27 -0.21 -0.54  0.23 -0.29  0.13  0.41
  0.38 -0.00 -0.47 -0.36  0.45 -0.19 -0.00  0.32  0.41
  0.33  0.41  0.27  0.57  0.08  0.23  0.29  0.13  0.41
  0.23 -0.41 -0.29 -0.00  0.00  0.30  0.58 -0.52 -0.00
  0.41  0.00  0.33  0.00  0.00 -0.74 -0.00 -0.43  0.00
  0.23  0.41 -0.29 -0.00 -0.00  0.30 -0.58 -0.52  0.00
S =
  2.84     0     0     0     0     0     0     0     0
     0  2.00     0     0     0     0     0     0     0
     0     0  1.84     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  1.51     0     0     0
     0     0     0     0     0     0  1.00     0     0
     0     0     0     0     0     0     0  0.51     0
     0     0     0     0     0     0     0     0  0.00
V =
  0.23 -0.41  0.29  0.00  0.00  0.30 -0.58  0.52 -0.00
  0.33 -0.41 -0.27 -0.08  0.57  0.23  0.29 -0.13 -0.41
  0.38  0.00  0.47  0.45  0.36 -0.19 -0.00 -0.32  0.41
  0.33 -0.41 -0.27  0.08 -0.57  0.23  0.29 -0.13  0.41
  0.41 -0.00 -0.33  0.00 -0.00 -0.74 -0.00  0.43 -0.00
  0.33  0.41 -0.27  0.54 -0.21  0.23 -0.29 -0.13 -0.41
  0.38  0.00  0.47 -0.45 -0.36 -0.19  0.00 -0.32 -0.41
  0.33  0.41 -0.27 -0.54  0.21  0.23 -0.29 -0.13  0.41
  0.23  0.41  0.29  0.00  0.00  0.30  0.58  0.52  0.00
```

(c) Here choose the rank of the matrix to be effectively seven as two of the nine singular values, namely 0.51 and 0.00, are less than about the size of the expected error, roughly 0.5.

(d) Use the rank seven SVD to solve the approximate system as in Procedure 3.5.4.

i. Find $\boldsymbol{z} = U^{\mathrm{T}}\boldsymbol{b}$ via `z=U'*b` to find

```
z =
  -7.63
   0.27
  -0.57
   0.42
   0.64
  -1.95
   0.64
  -0.40
  -0.01
```

ii. Neglect the last two rows in solving $S_7\boldsymbol{y} = \boldsymbol{z}$ to find via `y=z(1:7)./diag(S(1:7,1:7))` that the first seven components of $\boldsymbol{y}$ are

```
y =
  -2.69
   0.14
  -0.31
   0.24
   0.37
  -1.29
   0.64
```

The last two components of $\boldsymbol{y}$, y_8 and y_9, are free variables.

iii. Obtain a particular solution to $V^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{y}$, the one of smallest magnitude, by setting $y_8 = y_9 = 0$ and determining $\boldsymbol{x}$ from `x=V(:,1:7)*y` to get the smallest solution

```
x =
  -1.53
  -0.78
  -0.67
  -1.16
  -0.05
  -1.18
  -1.16
  -1.28
  -0.68
```

Obtain other equally valid solutions, in the context of the identified error in matrix A, by adding arbitrary multiples of the last two columns of V.

(e) Here we aim to make predictions from the CT-scan. The 'best' solution in this application is the one with least artificial features. The smallest magnitude $\boldsymbol{x}$ seems to reasonably implement this criteria. Thus use the above particular $\boldsymbol{x}$ to determine the transmission factors, $r_i = \exp(x_i)$. Here use `r=reshape(exp(x),3,3)` to compute and form into the 3×3 array of pixels

0.22	0.31	0.31
0.46	0.95	0.28
0.51	0.31	0.51

as illustrated with `colormap(gray),imagesc(r)`

The CT-scan identifies that there is a significant 'hole' in the middle of the body being scanned.

■

v0.2a

5.2.2 Tikhonov regularisation

This optional extension connects to much established practice that graduates may encounter.

Regularisation of poorly-posed linear equations is a widely used practical necessity. Many people have invented alternative techniques. Many have independently re-invented techniques. Perhaps the most common is the so-called Tikhonov regularisation. This section introduces and discusses Tikhonov regularisation.

> In statistics, the method is known as ridge regression, and with multiple independent discoveries, it is also variously known as the Tikhonov–Miller method, the Phillips–Twomey method, the constrained linear inversion method, and the method of linear regularization.
>
> *Wikipedia (2015)*

Definition 5.2.7. *In seeking to solve the poorly-posed system $A\boldsymbol{x} = \boldsymbol{b}$ for $m \times n$ matrix A, a* ***Tikhonov regularisation*** *is the system $(A^{\mathrm{T}}A + \alpha^2 I_n)\boldsymbol{x} = A^{\mathrm{T}}\boldsymbol{b}$ for some chosen regularisation parameter value $\alpha > 0$.* [8]

Example 5.2.8. Use Tikhonov regularisation to solve Example 5.2.1:

$$0.5x + 0.3y = 1 \quad \text{and} \quad 1.1x + 0.7y = 2\,,$$

Solution: Here the matrix and right-hand side vector are

$$A = \begin{bmatrix} 0.5 & 0.3 \\ 1.1 & 0.7 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

A Tikhonov regularisation, $(A^{\mathrm{T}}A + \alpha^2 I_n)\boldsymbol{x} = A^{\mathrm{T}}\boldsymbol{b}$, is then the system

$$\begin{bmatrix} 1.46 + \alpha^2 & 0.92 \\ 0.92 & 0.58 + \alpha^2 \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} 2.7 \\ 1.7 \end{bmatrix}.$$

Choose regularisation parameter α to be roughly the error: here the error is ± 0.05 so let's choose $\alpha = 0.1$ ($\alpha^2 = 0.01$). Enter into MATLAB/Octave with

```
A=[0.5 0.3;1.1 0.7]
b=[1.0;2.0]
At=(A'*A+0.01*eye(2))
bt=A'*b
rcondA=rcond(At)
x=At\bt
```

to find the Tikhonov regularised solution is $\boldsymbol{x} = (1.39, 0.72)$.[9] This solution is reasonably close to the smallest solution found by the

[8] Some will notice that a Tikhonov regularisation is closely connected to the so-called normal equation $A^{\mathrm{T}}A\boldsymbol{x} = A^{\mathrm{T}}\boldsymbol{b}$. Tikhonov regularisation shares some of the practical limitations of the normal equation.

[9] Interestingly, `rcond` $= 0.003$ for the Tikhonov system which is worse than `rcond`(A). The regularisation only works because pre-multiplying by A^{T} puts both sides in the row space of A (except for numerical error and the small $\alpha^2 I$ factor).

©A J Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

SVD which is $(1.32, 0.83)$. However, Tikhonov regularisation gives no hint of the reasonable general solutions found by the SVD approach of Example 5.2.1.

Change the regularisation parameter to $\alpha = 0.01$ and $\alpha = 1$ and see that both degrade the Tikhonov solution.

■

Do not apply Tikhonov regularisation blindly as it does introduce biases. The following example illustrates the bias.

Example 5.2.9. Recall the example at the start of Section 3.5.1 where my scales variously reported my weight in kg as 84.8, 84.1, 84.7 and 84.4. To best estimate my weight x we rewrote the problem in matrix-vector form

$$Ax = \boldsymbol{b}\,, \quad \text{namely} \quad \begin{bmatrix}1\\1\\1\\1\end{bmatrix} x = \begin{bmatrix}84.8\\84.1\\84.7\\84.4\end{bmatrix}.$$

A Tikhonov regularisation 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 $x = 84.5/(1+\alpha^2/4)\,\text{kg}$. This Tikhonov answer is biased because it is systematically below the average 84.5 kg. For small Tikhonov parameter α it is only a small bias, but even so such a bias is unpleasant. ■

Example 5.2.10. Use Tikhonov regularisation to solve $A\boldsymbol{x} = \boldsymbol{b}$ for the matrix and vector of Example 5.2.3a.

Solution: Here the matrix and right-hand side vector are

$$A = \begin{bmatrix}-0.2 & -0.6 & 1.8\\ 0.0 & 0.2 & -0.4\\ -0.3 & 0.7 & 0.3\end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix}-0.5\\0.1\\-0.2\end{bmatrix}.$$

A Tikhonov regularisation, $(A^{\mathrm{T}}A + \alpha^2 I_n)\boldsymbol{x} = A^{\mathrm{T}}\boldsymbol{b}$, is then the system

$$\begin{bmatrix}0.13+\alpha^2 & -0.09 & -0.45\\ -0.09 & 0.89+\alpha^2 & -0.95\\ -0.45 & -0.95 & 3.49+\alpha^2\end{bmatrix} \boldsymbol{x} = \begin{bmatrix}0.16\\0.18\\-1.00\end{bmatrix}.$$

Choose regularisation parameter α to be roughly the error: here the error is ± 0.05 so let's choose $\alpha = 0.1$ ($\alpha^2 = 0.01$). Enter into and solve with MATLAB/Octave via

```
A=[-0.2 -0.6 1.8
   0.0 0.2 -0.4
  -0.3 0.7 0.3 ]
b=[-0.5;0.1;-0.2]
At=(A'*A+0.01*eye(3))
bt=A'*b
rcondA=rcond(At)
x=At\bt
```

which then finds the Tikhonov regularised solution is $\boldsymbol{x} = (0.10, -0.11, -0.30)$. To two decimal places this is the same as the smallest solution found by an SVD. However, Tikhonov regularisation gives no hint of the reasonable general solutions found by the SVD approach of Example 5.2.3a.

■

Theorem 5.2.11 (Tikhonov regularisation)**.** *Solving the Tikhonov regularisation, with parameter α, of $A\boldsymbol{x} = \boldsymbol{b}$ is equivalent to finding the smallest, least square, solution of the system $\tilde{A}\boldsymbol{x} = \boldsymbol{b}$ where the matrix $\tilde{A}$ is obtained from A by replacing each of its non-zero singular values σ_i by $\tilde{\sigma}_i := \sigma_i + \alpha^2/\sigma_i$.*

Proof. Let $m \times n$ matrix A have SVD $A = USV^{\mathrm{T}}$. First, the left-hand side matrix in a Tikhonov regularisation is

$$\begin{aligned} A^{\mathrm{T}}A + \alpha^2 I_n &= (USV^{\mathrm{T}})^{\mathrm{T}}USV^{\mathrm{T}} + \alpha^2 I_n VV^{\mathrm{T}} \\ &= VS^{\mathrm{T}}U^{\mathrm{T}}USV^{\mathrm{T}} + \alpha^2 VI_nV^{\mathrm{T}} \\ &= VS^{\mathrm{T}}SV^{\mathrm{T}} + V(\alpha^2 I_n)V^{\mathrm{T}} \\ &= V(S^{\mathrm{T}}S + \alpha^2 I_n)V^{\mathrm{T}}, \end{aligned}$$

whereas the right-hand side is $A^{\mathrm{T}}\boldsymbol{b} = (USV^{\mathrm{T}})^{\mathrm{T}}\boldsymbol{b} = VS^{\mathrm{T}}U^{\mathrm{T}}\boldsymbol{b}$. Corresponding to the variables used in previous procedures, let $\boldsymbol{z} = U^{\mathrm{T}}\boldsymbol{b} \in \mathbb{R}^m$ and as yet unknown $\boldsymbol{y} = V^{\mathrm{T}}\boldsymbol{x} \in \mathbb{R}^n$. Then equating the above two sides, and premultiplying by the orthogonal V^{T}, means the Tikhonov regularisation is equivalent to solving $(S^{\mathrm{T}}S + \alpha^2 I_n)\boldsymbol{y} = S^{\mathrm{T}}\boldsymbol{z}$ for $\boldsymbol{y}$.

Second, suppose rank $A = r$ so that the singular value matrix

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

(where the bottom right zero block contains all the zero singular values). Consequently, the equivalent Tikhonov regularisation,

$(S^{\mathrm{T}}S + \alpha^2 I_n)\boldsymbol{y} = S^{\mathrm{T}}\boldsymbol{z}$, becomes

$$\begin{bmatrix} \sigma_1^2+\alpha^2 & \cdots & 0 & \\ \vdots & \ddots & \vdots & O_{r\times(n-r)} \\ 0 & \cdots & \sigma_r^2+\alpha^2 & \\ & & & \\ & O_{(n-r)\times r} & & \alpha^2 I_{n-r} \end{bmatrix} \boldsymbol{y} = \begin{bmatrix} \sigma_1 z_1 \\ \vdots \\ \sigma_r z_r \\ \\ \mathbf{0}_{n-r} \end{bmatrix} .$$

Dividing each of the first r rows by the corresponding non-zero singular value, $\sigma_1, \sigma_2, \dots, \sigma_r$, the equivalent system is

$$\begin{bmatrix} \sigma_1+\alpha^2/\sigma_1 & \cdots & 0 & \\ \vdots & \ddots & \vdots & O_{r\times(n-r)} \\ 0 & \cdots & \sigma_r+\alpha^2/\sigma_r & \\ & & & \\ & O_{(n-r)\times r} & & \alpha^2 I_{n-r} \end{bmatrix} \boldsymbol{y} = \begin{bmatrix} z_1 \\ \vdots \\ z_r \\ \\ \mathbf{0}_{n-r} \end{bmatrix} ,$$

with solution

- $y_i = z_i/(\sigma_i + \alpha^2/\sigma_i)$ for $i = 1, \dots, r$, and
- $y_i = 0$ for $i = r+1, \dots, n$ (since $\alpha^2 > 0$).

This establishes that solving the Tikhonov system is equivalent to performing the SVD Procedure 3.5.4 for the least square solution to $A\boldsymbol{x} = \boldsymbol{b}$ but with two changes in Step 3:

- for $i = 1, \dots, r$ divide by $\tilde\sigma_i := \sigma_i + \alpha^2/\sigma_i$ instead of the true singular value σ_i (the margin plots $\tilde\sigma$ versus σ), and
- for $i = r+1, \dots, n$ set $y_i = 0$ to obtain the smallest possible solution (Theorem 3.5.12).

Thus Tikhonov regularisation of $A\boldsymbol{x} = \boldsymbol{b}$ is equivalent to finding the smallest, least square, solution of the system $\tilde A\boldsymbol{x} = \boldsymbol{b}$. □

There is another reason to be careful when using Tikhonov regularisation. Yes, it gives a nice, neat, unique solution. However, it does not hint that there may be an infinite number of equally good nearby solutions (as found through Procedure 5.2.2). Among those equally good nearby solutions may be ones that you prefer in your application.

Choose a good regularisation parameter

- One strategy to choose the regularisation parameter α is that the effective change in the matrix, from A to $\tilde A$, should be about the size of errors expected in A. [10] Since changes in the matrix are largely measured by the singular values we need to consider the relation between $\tilde\sigma = \sigma + \alpha^2/\sigma$ and σ. From

[10] This strategic choice is sometimes called the discrepancy principle (Kress 2015, §7).

the marginal graph the small singular values are changed by a lot, but these are the ones for which we want $\tilde{\sigma}$ large in order give a 'least square' approximation. Significantly, the marginal graph also shows that singular values larger than 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 regularisation upon errors in the right-hand side vector. The condition number of A may be very bad. However, as the marginal graph shows the smallest $\tilde{\sigma} \geq 2\alpha$. Thus, in the regularised system the condition number of the effective matrix $\tilde{A}$ is approximately $\sigma_1/(2\alpha)$. We need to choose the regularisation parameter 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 3.3.29). It is only when the regularisation parameter α is big enough that the regularisation will be effective in finding a least square approximation.

5.2.3 Exercises

Exercise 5.2.1. For each of the following matrices, say A, and right-hand side vectors, say $\boldsymbol{b}_1$, solve $A\boldsymbol{x} = \boldsymbol{b}_1$. But suppose the matrix entries come from experiments and are only known to within errors ± 0.05. Thus within experimental error the given matrices A' and A'' may be the 'true' matrix A. Solve $A'\boldsymbol{x}' = \boldsymbol{b}_1$ and $A''\boldsymbol{x}'' = \boldsymbol{b}_1$ and comment on the results. Finally, use an SVD to find a general solution consistent with the error in the matrix.

(a) $A = \begin{bmatrix} -1.3 & -0.4 \\ 0.7 & 0.2 \end{bmatrix}$, $\boldsymbol{b}_1 = \begin{bmatrix} 2.4 \\ -1.3 \end{bmatrix}$, $A' = \begin{bmatrix} -1.27 & -0.43 \\ 0.71 & 0.19 \end{bmatrix}$, $A'' = \begin{bmatrix} -1.27 & -0.38 \\ 0.66 & 0.22 \end{bmatrix}$.

(b) $B = \begin{bmatrix} -1.8 & -1.1 \\ -0.2 & -0.1 \end{bmatrix}$, $\boldsymbol{b}_2 = \begin{bmatrix} -0.7 \\ -0.1 \end{bmatrix}$, $B' = \begin{bmatrix} -1.81 & -1.13 \\ -0.24 & -0.12 \end{bmatrix}$, $B'' = \begin{bmatrix} -1.81 & -1.13 \\ -0.18 & -0.1 \end{bmatrix}$.

(c) $C = \begin{bmatrix} 0.8 & -0.1 \\ -1.0 & 0.1 \end{bmatrix}$, $\boldsymbol{b}_3 = \begin{bmatrix} 0.2 \\ -0.3 \end{bmatrix}$, $C' = \begin{bmatrix} 0.81 & -0.07 \\ -1.01 & 0.06 \end{bmatrix}$, $C'' = \begin{bmatrix} 0.79 & -0.08 \\ -1.03 & 0.09 \end{bmatrix}$.

(d) $D = \begin{bmatrix} 0.0 & 0.5 & -0.5 \\ 0.6 & 0.5 & 0.9 \\ 0.6 & 1.3 & 0.0 \end{bmatrix}$, $\boldsymbol{b}_4 = \begin{bmatrix} -1.4 \\ 0.4 \\ -1.9 \end{bmatrix}$, $D' = \begin{bmatrix} -0.02 & 0.49 & -0.49 \\ 0.58 & 0.54 & 0.9 \\ 0.61 & 1.34 & -0.02 \end{bmatrix}$, $D'' = \begin{bmatrix} -0.04 & 0.52 & -0.48 \\ 0.64 & 0.52 & 0.87 \\ 0.57 & 1.33 & 0.04 \end{bmatrix}$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(e) $E = \begin{bmatrix} 0.6 & -0.8 & -0.2 \\ -0.9 & 1.0 & 1.2 \\ -0.9 & 0.9 & 1.4 \end{bmatrix}$, $\boldsymbol{b}_5 = \begin{bmatrix} 1.1 \\ -3.7 \\ -4.1 \end{bmatrix}$,

$E' = \begin{bmatrix} 0.57 & -0.78 & -0.23 \\ -0.91 & 0.99 & 1.22 \\ -0.93 & 0.9 & 1.39 \end{bmatrix}$,

$E'' = \begin{bmatrix} 0.56 & -0.77 & -0.21 \\ -0.87 & 1.01 & 1.22 \\ -0.87 & 0.9 & 1.39 \end{bmatrix}$.

(f) $F = \begin{bmatrix} 0.1 & -1.0 & 0.0 \\ 2.1 & -0.2 & -0.5 \\ 0.0 & -1.6 & 0.0 \end{bmatrix}$, $\boldsymbol{b}_6 = \begin{bmatrix} -0.2 \\ 1.6 \\ -0.5 \end{bmatrix}$,

$F' = \begin{bmatrix} 0.1 & -0.98 & -0.04 \\ 2.11 & -0.17 & -0.47 \\ -0.04 & -1.62 & -0.01 \end{bmatrix}$, $F'' = \begin{bmatrix} 0.14 & -0.96 & 0.01 \\ 2.13 & -0.23 & -0.47 \\ 0.0 & -1.57 & -0.02 \end{bmatrix}$.

(g) $G = \begin{bmatrix} 1.0 & -0.3 & 0.3 & -0.4 \\ 1.8 & 0.5 & 0.1 & 0.2 \\ 0.2 & -0.3 & 1.3 & -0.6 \\ 0.0 & 0.5 & 1.2 & 0.0 \end{bmatrix}$, $\boldsymbol{b}_7 = \begin{bmatrix} 2.0 \\ 1.6 \\ 1.4 \\ -0.2 \end{bmatrix}$,

$G' = \begin{bmatrix} 0.98 & -0.3 & 0.31 & -0.44 \\ 1.8 & 0.54 & 0.06 & 0.21 \\ 0.24 & -0.33 & 1.27 & -0.58 \\ 0.01 & 0.52 & 1.23 & -0.01 \end{bmatrix}$,

$G'' = \begin{bmatrix} 1.03 & -0.32 & 0.33 & -0.36 \\ 1.82 & 0.49 & 0.08 & 0.16 \\ 0.2 & -0.31 & 1.33 & -0.64 \\ 0.0 & 0.49 & 1.22 & 0.0 \end{bmatrix}$.

(h) $H = \begin{bmatrix} -0.9 & -0.5 & -0.3 & -0.4 \\ -0.1 & 0.1 & -0.2 & 0.8 \\ -1.0 & 0.4 & -1.1 & 0.6 \\ 1.0 & 2.2 & -1.0 & -0.1 \end{bmatrix}$, $\boldsymbol{b}_8 = \begin{bmatrix} 0.4 \\ 0.3 \\ 0.2 \\ -2.0 \end{bmatrix}$,

$H' = \begin{bmatrix} -0.88 & -0.52 & -0.33 & -0.41 \\ -0.11 & 0.13 & -0.17 & 0.78 \\ -0.96 & 0.44 & -1.12 & 0.61 \\ 0.98 & 2.19 & -0.99 & -0.13 \end{bmatrix}$,

$H'' = \begin{bmatrix} -0.86 & -0.49 & -0.29 & -0.37 \\ -0.06 & 0.14 & -0.18 & 0.83 \\ -0.96 & 0.38 & -1.11 & 0.58 \\ 1.01 & 2.21 & -1.04 & -0.13 \end{bmatrix}$.

Exercise 5.2.2. Recall Example 5.2.4 explores the effective rank of the 5×5 Hilbert matrix depending upon a supposed level of error. Similarly, explore the effective rank of the 7×7 Hilbert matrix (`hilb(7)` in MATLAB/Octave) depending upon supposed levels of error in the matrix. What levels of error in the components would give what effective rank of the matrix?

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Exercise 5.2.3. Recall Exercise 2.2.12 considered the inner four planets in the solar system. The exercise fitted fit a quadratic polynomial to the orbital period $T = c_1 + c_2 R + c_3 R^2$ as a function of distance R using the data of Table 2.4. In view of the bad condition number, $\texttt{rcond} = 6 \cdot 10^{-6}$, revisit the task with the more powerful techniques of this section. Use the data for Mercury, Venus and Earth to fit the quadratic and predict the period for Mars. Discuss how the bad condition number is due to the failure in Exercise 2.2.12 of scaling the data in the matrix.

Exercise 5.2.4. Recall Exercise 3.5.17 used a 4×4 grid of pixels in the computed tomography of a CT-scan. Redo this exercise recognising that the entries in matrix A have errors up to roughly 0.5. Discuss any change in the prediction.

Exercise 5.2.5. Reconsider each of the matrix-vector systems you explored in Exercise 5.2.1. Also solve each system using Tikhonov regularisation; for example, in the first system solve $A\boldsymbol{x} = \boldsymbol{b}_1$, $A'\boldsymbol{x}' = \boldsymbol{b}_1$ and $A''\boldsymbol{x}'' = \boldsymbol{b}_1$. Discuss why $\boldsymbol{x}$, $\boldsymbol{x}'$ and $\boldsymbol{x}''$ are all reasonably close to the smallest solution of those obtained via an SVD.

Exercise 5.2.6. Recall that Example 5.2.4 explores the effective rank of the 5×5 Hilbert matrix depending upon a supposed level of error. Here do the alternative and solve the system $A\boldsymbol{x} = \mathbf{1}$ via Tikhonov regularisation using a wide range of various regularisation parameters α. Comment on the relation between the solutions obtained for various α and those obtained in the example for the various presumed error—perhaps plot the components of $\boldsymbol{x}$ versus parameter α (on a log-log plot).

Exercise 5.2.7. Recall Example 5.2.6 used a 3×3 grid of pixels in the computed tomography of a CT-scan. Redo this example with Tikhonov regularisation recognising that the entries in matrix A have errors up to roughly 0.5. Discuss the relation between the solution of Example 5.2.6 that of Tikhonov regularisation.

Exercise 5.2.8. Recall Exercise 3.5.17 used a 4×4 grid of pixels in the computed tomography of a CT-scan. Redo this exercise with Tikhonov regularisation recognising that the entries in matrix A have errors up to roughly 0.5. Discuss any change in the prediction.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Answers to selected exercises

5.1.1b : 1.4

5.1.1d : 2.3

5.1.1f : 1.9

5.1.2b : 5

5.1.2d : 3

5.1.2f : 8

5.1.2h : 5

5.1.5 : Rank three.

5.1.8a : Use `A=[A A A;A 0*A A; A A A]`

5.1.9a : Use `A=[A 0*A A;0*A A 0*A; A 0*A A]`

5.1.10 : ranks 6, 24

5.1.13 : Book 3: angles 35°, 32°, 1°, 29°, 54°, 76°.

5.1.14c : Books 14, 4, 5, 13 and 2 are at angles 1°, 3°, 9°, 10° and 16°, respectively.

5.2.1b : $\boldsymbol{x} = (1.0\,,-1.0)$, $\boldsymbol{x}' = (0.54\,,-0.24)$, $\boldsymbol{x}'' = (1.92\,,-2.46)$, $\boldsymbol{x} = (0.28\,,0.17) + t(-0.52\,,0.85)$ (2 d.p.).

5.2.1d : $\boldsymbol{x} = (1.6\,,-2.2\,,0.6)$, $\boldsymbol{x}' = (1.31\,,-2.0\,,0.8)$, $\boldsymbol{x}'' = (-0.28\,,-1.35\,,1.47)$, $\boldsymbol{x} = (-0.09\,,-1.43\,,1.32) + t(0.85\,,-0.39\,,-0.36)$ (2 d.p.).

5.2.1f : $\boldsymbol{x} = (1.12\,,0.31\,,1.4)$, $\boldsymbol{x}' = (0.59\,,0.3\,,-0.84)$, $\boldsymbol{x}'' = (0.77\,,0.32\,,-0.08)$, $\boldsymbol{x} = (0.75\,,0.3\,,-0.18) + t(-0.23\,,-0.01\,,-0.97)$ (2 d.p.).

5.2.1h : $\boldsymbol{x} = (-0.76\,,-0.23\,,0.69\,,0.48)$, $\boldsymbol{x}' = (-1.36\,,0.33\,,1.35\,,0.43)$, $\boldsymbol{x}'' = (-0.09\,,-0.92\,,-0.17\,,0.47)$, $\boldsymbol{x} = (-0.38\,,-0.63\,,0.2\,,0.47) + t(0.52\,,-0.54\,,-0.67\,,-0.02)$ (2 d.p.).

5.2.4 : The matrix has effective rank of eleven. Pixel ten is still the most absorbing. The corner pixels are the most affected.

v0.2a

6 Determinants distinguish matrices

Chapter Contents

6.1 Geometry underlies determinants 539

6.1.1 Exercises 550

6.2 Laplace expansion theorem for determinants 556

6.2.1 Exercises 576

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 matrix A is invertible iff $\det A \neq 0$;
- 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 matrix A is invertible iff $\det A \neq 0$.

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 useful properties—especially those properties useful for understanding and developing the general eigenvalue problems and applications of Chapter 7.

6.1 Geometry underlies determinants

Section Contents

6.1.1 Exercises 550

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 areas and volumes.

Example 6.1.1. Consider the square matrix $A = \begin{bmatrix} \frac{1}{2} & 1 \\ 0 & 1 \end{bmatrix}$. Use matrix multiplication to find the image of the unit square under the transformation by A. How much is the area of the unit square scaled up/down? Compare with the determinant.

Solution: Consider the corner points of the unit square, under multiplication by A: $(0,0) \mapsto (0,0)$, $(1,0) \mapsto (\frac{1}{2},0)$, $(0,1) \mapsto (1,1)$, and $(1,1) \mapsto (\frac{3}{2},1)$, as shown in the marginal picture (the 'roof' is only plotted to uniquely identify the sides). The resultant parallelogram has area of $\frac{1}{2}$ as its base is $\frac{1}{2}$ and its height is 1. This parallelogram area is the same as the determinant since here (6.1) gives $\det A = \frac{1}{2} \cdot 1 - 0 \cdot 1 = \frac{1}{2}$.

■

Example 6.1.2. Consider the square matrix $B = \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$. Use matrix multiplication to find the image of the unit square under the transformation by B. How much is the unit area scaled up/down? Compare with the determinant.

Solution: Consider the corner points of the unit square, under multiplication by B: $(0,0) \mapsto (0,0)$, $(1,0) \mapsto (-1,1)$, $(0,1) \mapsto (1,1)$, and $(1,1) \mapsto (0,2)$, as shown in the marginal picture. Through multiplication by the matrix B, the unit square is expanded, rotated and reflected. The resultant square has area of 2 as its sides are all of length $\sqrt{2}$. This area has the same magnitude as the determinant since here (6.1) gives $\det B = (-1) \cdot 1 - 1 \cdot 1 = -2$.

■

Example 6.1.3. Consider the square matrix $A = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & \frac{3}{2} \end{bmatrix}$. Use matrix multiplication to find the image of the unit cube under the transformation by A. How much is the volume of the unit cube scaled up/down? Compare with the determinant.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Solution: Consider the corner points of the unit cube, under multiplication by A: $(0,0,0) \mapsto (0,0,0)$, $(1,0,0) \mapsto (2,1,0)$, $(0,1,0) \mapsto (0,1,0)$, $(0,0,1) \mapsto (0,0,\frac{3}{2})$, and so on, as shown below (in stereo):

Through multiplication by the matrix A, the unit cube is deformed to a parallelepiped. The resultant parallelepiped has volume of 3 as it has height $\frac{3}{2}$ and the parallelogram base has area $2 \cdot 1$. This volume is the same as the matrix determinant since (6.1) gives $\det A = 2 \cdot 1 \cdot \frac{3}{2} + 0 + 0 - 0 - 0 - 0 = 3$.

■

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 shown with corners at the origin, (a,c), (b,d) and $(a+b, c+d)$. Let's determine the area of the parallelogram by that of the containing rectangle less the two small rectangles and the four small triangles. The two small rectangles have the same area, namely bc. The two small triangles on the left and the right also have the same area, namely $\frac{1}{2}bd$. The two small triangles on the top and the bottom similarly 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

$$\begin{aligned} \text{area} &= (a+b)(c+d) - 2bc - 2 \cdot \frac{1}{2}bd - 2 \cdot \frac{1}{2}ac \\ &= ac + ad + bc + bd - 2bc - bd - ac \\ &= ad - bc = \det A\,. \end{aligned}$$

This picture is the case when the matrix does not also reflect the image: if the matrix also reflects, as in Example 6.1.2, then the determinant is the negative of the area. In either case, the

area of the unit square after transforming by the matrix A is the magnitude $|\det A|$.

Analogous geometric arguments relate determinants of 3×3 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 in the margin. By computing the volumes of various rectangular boxes, prisms and tetrahedra (or by other methods), the volume of such a parallelepiped can 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 generalise a square and cube to n dimensions ($\mathbb{R}^n$), n**D-volume** to generalise the notion of area and volume to n dimensions, and so on. When the dimension of the space is unspecified, then we may say **hyper-cube**, **hyper-volume**, and so on.

Definition 6.1.4. *Let A be an $n \times n$ square matrix, and let C be the unit nD-cube in $\mathbb{R}^n$. Consider the image C' in $\mathbb{R}^n$ of C by the transformation $\boldsymbol{x} \mapsto A\boldsymbol{x}$. Define the* ***determinant*** *of A, denoted either $\det A$ or $|A|$ 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.*

Example 6.1.5. Roughly estimate the determinant of the matrix that transforms the unit square to the parallelogram as shown in the margin.

Solution: The image is a parallelogram with a vertical base of length about 0.8 and a horizontal height of about 0.7 so the area of the image is about $0.8 \times 0.7 = 0.56 \approx 0.6$. But the image has been reflected as one cannot rotate and stretch to get the image (remember the origin is fixed under matrix multiplication): thus the determinant must be negative. Our estimate for the matrix determinant is -0.6.

■

Basic properties of a determinant follow direct from Definition 6.1.4.

Theorem 6.1.6. *(a) Let D be an $n \times n$ diagonal matrix. The determinant of D is the product of the diagonal entries: $\det D = d_{11}d_{22}\cdots d_{nn}$.*

(b) An orthogonal matrix Q has $\det Q = \pm 1$ (only one alternative, not both), and $\det Q = \det(Q^{\mathrm{T}})$.

(c) $\det(kA) = k^n \det A$ for any scalar k.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Proof. Use Definition 6.1.4.

6.1.6a. The unit nD-cube in $\mathbb{R}^n$ has edges $\boldsymbol{e}_1, \boldsymbol{e}_2, \dots, \boldsymbol{e}_n$ (the unit vectors). Multiplying each of these edges by the diagonal matrix $D = \operatorname{diag}(d_{11}, d_{22}, \dots, d_{nn})$ maps the unit nD-cube to a nD-rectangle with edges $d_{11}\boldsymbol{e}_1, d_{22}\boldsymbol{e}_2, \dots, d_{nn}\boldsymbol{e}_n$ (as illustrated in the margin). Being a nD-rectangle with all edges orthogonal, its nD-volume is the product of the length of the sides; that is, $|\det D| = |d_{11}| \cdot |d_{22}| \cdots |d_{nn}|$. The nD-cube is reflected only if there are an odd number of negative diagonal elements, hence the sign of the determinant is such that $\det D = d_{11}d_{22}\cdots d_{nn}$.

6.1.6b. Multiplication by an orthogonal matrix Q is a rotation and/or reflection as it preserves all lengths and angles (Theorem 3.2.48f). Hence it preserves nD-volumes. Consequently the image of the unit nD-cube under multiplication by Q has the same volume of one; that is, $|\det Q| = 1$. The sign of $\det Q$ characterises whether multiplication by Q has a reflection.

When Q is orthogonal then so is Q^{T} (Theorem 3.2.48d). Hence $\det Q^{\mathrm{T}} = \pm 1$. Multiplication by Q involves a reflection iff its inverse of multiplication by Q^{T} involves the reflection back again. Hence the signs of the two determinants must be the same: that is, $\det Q = \det Q^{\mathrm{T}}$.

6.1.6c. Let the matrix A transform the unit nD-cube to a nD-parallelepiped C' that has nD-volume $|\det A|$. Multiplication by the matrix (kA) then forms a nD-parallelepiped which is $|k|$-times bigger than C' in every direction. In $\mathbb{R}^n$ its nD-volume is then $|k|^n$-times bigger; that is, $|\det(kA)| = |k|^n |\det A| = |k^n \det A|$. If the scalar k is negative then the orientation of the image is reversed (reflected) only in odd n dimensions; that is, the sign of the determinant is multiplied by $(-1)^n$. (For example, the unit square shown in the margin is transformed through multiplication by $(-1)I_2$ and the effect is the same as rotation by 180°, without any reflection as $(-1)^2 = 1$.) Hence for all real k, the orientation is such that $\det(kA) = k^n \det A$.

□

Example 6.1.7. The determinant of the $n \times n$ identity matrix is one: that is, $\det I_n = 1$. We justify this result

- either because it is a diagonal matrix and hence its determinant is the product of the diagonal entries (Theorem 6.1.6a) here all ones,
- or because multiplication by the identity does not change the unit nD-cube and so does not change its nD-volume

(Definition 6.1.4).

The determinant of $-I_n$ is $(-1)^n$ either by the product of the diagonals (Theorem 6.1.6a), or because of the scalar multiplication Theorem 6.1.6c. ■

Example 6.1.8. 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 6.1.6 to deduce the determinants of the following matrices:

$$\begin{bmatrix} \frac{1}{3} & \frac{2}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\ \frac{2}{3} & \frac{1}{3} & \frac{2}{3} \end{bmatrix}, \quad \begin{bmatrix} -1 & 2 & -2 \\ -2 & -2 & -1 \\ 2 & -1 & -2 \end{bmatrix}, \quad \begin{bmatrix} \frac{1}{6} & -\frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \\ -\frac{1}{3} & \frac{1}{6} & \frac{1}{3} \end{bmatrix}.$$

Solution:

- Using (6.1) the determinant

$$\begin{aligned} \det Q &= \tfrac{1}{3}\tfrac{2}{3}\tfrac{2}{3} + (-\tfrac{2}{3})\tfrac{1}{3}(-\tfrac{2}{3}) + \tfrac{2}{3}\tfrac{2}{3}\tfrac{1}{3} \\ &\quad - \tfrac{2}{3}\tfrac{2}{3}(-\tfrac{2}{3}) - \tfrac{1}{3}\tfrac{1}{3}\tfrac{1}{3} - (-\tfrac{2}{3})\tfrac{2}{3}\tfrac{2}{3} \\ &= \tfrac{1}{27}(4+4+4+8-1+8) = 1\,. \end{aligned}$$

- The first matrix is Q^{T} for which $\det Q^{\mathrm{T}} = \det Q = 1$.
- The second matrix is minus three times Q so, being 3×3 matrices, its determinant is $(-3)^3 \det Q = -27$.
- The third matrix is half of Q so, being 3×3 matrices, its determinant is $(\frac{1}{2})^3 \det Q = \frac{1}{8}$.

■

A consequence of Theorem 6.1.6c is that a determinant characterises the transformation of any sized hyper-cube. Consider the transformation by a matrix A of an nD-cube of side length k ($k \geq 0$). The nD-cube has edges $k\boldsymbol{e}_1, k\boldsymbol{e}_2, \dots, k\boldsymbol{e}_n$. The transformation results in an nD-parallelepiped with edges $A(k\boldsymbol{e}_1), A(k\boldsymbol{e}_2), \dots, A(k\boldsymbol{e}_n)$, which by commutativity and associativity (Theorem 3.1.24d) are the same edges as $(kA)\boldsymbol{e}_1, (kA)\boldsymbol{e}_2, \dots, (kA)\boldsymbol{e}_n$. That is, the resulting nD-parallelepiped is the same as applying matrix kA to the unit nD-cube, and so must have nD-volume $k^n|\det A|$. Crucially, this property that matrix multiplication multiplies all sizes of hyper-cubes by the determinant holds for all other shapes and sizes, not just hyper-cubes. Let's see an specific example, before proving the general theorem.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Example 6.1.9. Multiplication by some specific matrix transforms the (blue) triangle C to the (red) triangle C' as shown in the margin. By finding the ratio of the areas, estimate the magnitude of the determinant of the matrix.

Solution: The (blue) triangle C has vertical height one and horizontal base one, so has area 0.5. The mapped (red) triangle C' has vertical base of 1.5 and horizontal height of 1.5 so its area is $\frac{1}{2} \times 1.5 \times 1.5 = 1.25$. The mapped area is thus $1.25/0.5 = 2.5$ times bigger than the initial area; hence the determinant of the transformation matrix has magnitude $|\det A| = 2.5$.

We cannot determine the sign of the determinant as we do not know about the orientation of C' relative to C.

■

Theorem 6.1.10. *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.

Proof. A proof is analogous to integration in calculus (Hannah 1996, p.402). In 2D, as drawn in the margin, we divide a given region C into many small squares of side length k, each of area k^2: each of these transforms to a small parallelogram of area $k^2|\det A|$ (by Theorem 6.1.6c), then the sum of the transformed areas is just $|\det A|$ times the original area of C.

In n-dimensions, divide a given region C into many small nD-cubes of side length k, each of nD-volume k^n: each of these transforms to a small nD-parallelepiped of nD-volume $k^n|\det A|$ (by Theorem 6.1.6c), then the sum of the transformed nD-volume is just $|\det A|$ times the nD-volume of C. □

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.

This property of transforming general areas and volumes also establishes the next crucial property of determinants, namely that the determinant of a product is the product of the determinants: $\det(AB) = \det(A)\det(B)$ for all matrices A and B (of the same size).

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Example 6.1.11. Recall the two 2×2 matrices of Examples 6.1.1 and 6.1.2:

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

Check that the determinant of their product is the product of their determinants.

Solution: First, Examples 6.1.1 and 6.1.2 computed $\det A = \frac{1}{2}$ and $\det B = -2$. Thus the product of their determinants is $\det(A)\det(B) = -1$.

Secondly, calculate the matrix product

$$AB = \begin{bmatrix} \frac{1}{2} & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{3}{2} \\ 1 & 1 \end{bmatrix},$$

whose action upon multiplication is illustrated in the margin. By (6.1), $\det(AB) = \frac{1}{2} \cdot 1 - \frac{3}{2} \cdot 1 = -1$, as required.

Thirdly, we should also check the other product (as the question does not specify the order of the product):

$$BA = \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} & 0 \\ \frac{1}{2} & 2 \end{bmatrix},$$

whose action upon multiplication is illustrated in the margin. By (6.1), $\det(BA) = -\frac{1}{2} \cdot 2 - 0 \cdot \frac{1}{2} = -1$, as required.

■

Theorem 6.1.12. *For any two $n \times n$ matrices A and B, $\det(AB) = \det(A)\det(B)$. Further, for $n \times n$ matrices $A_1, A_2, \ldots, A_\ell$, $\det(A_1 A_2 \cdots A_\ell) = \det(A_1)\det(A_2)\cdots\det(A_\ell)$.*

Proof. Consider the unit nD-cube C, its image C' upon transforming by B, and the image C'' after transforming C' by A. That is, each edge $\boldsymbol{e}_j$ of cube C is mapped to the edge $B\boldsymbol{e}_j$ of C', which is in turn mapped to edge $A(B\boldsymbol{e}_j)$ of C''. By Definition 6.1.4, C' has (signed) nD-volume $\det B$. Theorem 6.1.10 implies C'' has (signed) nD-volume $\det A$ times that of C'; that is, C'' has (signed) nD-volume $\det(A)\det(B)$.

By associativity, the jth edge $A(B\boldsymbol{e}_j)$ of C'' is the same as $(AB)\boldsymbol{e}_j$ and so C'' is the image of C under the transformation by matrix (AB). Consequently, the (signed) nD-volume of C'' is alternatively given by $\det(AB)$. These two expressions for the nD-volume of C'' must be equal: that is, $\det(AB) = \det(A)\det(B)$.

Exercise 6.1.13 uses induction to then prove the second statement in the theorem that $\det(A_1 A_2 \cdots A_\ell) = \det(A_1)\det(A_2)\cdots\det(A_\ell)$. □

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Example 6.1.13. (a) Confirm the product rule for determinants, Theorem 6.1.12, for the product

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

Solution: Although the determinant of the left-hand matrix is $(-3)(-3) - 3(-2) = 9 - (-6) = 15$, we cannot confirm the product rule because it does not apply: the matrices on the right-hand side are not square matrices and so do not have determinants.

(b) Given $\det A = 2$ and $\det B = \pi$, what is $\det(AB)$?

Solution: Strictly, there is no answer as we do not know that matrices A and B are of the same size. However, if we are *additionally* given that A and B are the same size, then Theorem 6.1.12 gives $\det(AB) = \det(A)\det(B) = 2\pi$.

■

Example 6.1.14. Use the product theorem to help find the determinant of matrix

$$C = \begin{bmatrix} 45 & -15 & 30 \\ -2\pi & \pi & 2\pi \\ \frac{1}{9} & \frac{2}{9} & -\frac{1}{3} \end{bmatrix}.$$

Solution: One route is to observe that there is a common factor in each row of the matrix so it may be factored as

$$C = \begin{bmatrix} 15 & 0 & 0 \\ 0 & \pi & 0 \\ 0 & 0 & \frac{1}{9} \end{bmatrix} \begin{bmatrix} 3 & -1 & 2 \\ -2 & 1 & 2 \\ 1 & 2 & -3 \end{bmatrix}.$$

The first matrix, being diagonal, has determinant that is the product of its diagonal elements (Theorem 6.1.6a) so its determinant $= 15\pi\frac{1}{9} = \frac{5}{3}\pi$. The second matrix, from (6.1), has determinant $= -9 - 2 - 8 - 2 - 12 + 6 = -27$. Theorem 6.1.12 then gives $\det C = \frac{5}{3}\pi \cdot (-27) = -45\pi$.

■

Theorem 6.1.15. *For any square matrix A, $\det A = \det(A^{\mathrm{T}})$.*

Proof. Use an SVD of the matrix, say $A = USV^{\mathrm{T}}$. Then

$$\begin{aligned} \det A &= \det(USV^{\mathrm{T}}) \\ &= \det(U)\det(S)\det(V^{\mathrm{T}}) \quad \text{(by Theorem 6.1.12 twice)} \end{aligned}$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$
\begin{aligned}
&= \det(U^{\mathrm{T}})(\sigma_1\sigma_2\cdots\sigma_n)\det(V) \quad \text{(by Theorem 6.1.6)}\\
&= \det(U^{\mathrm{T}})\det(S^{\mathrm{T}})\det(V)\\
&= \det(V)\det(S^{\mathrm{T}})\det(U^{\mathrm{T}}) \quad \text{(by scalar commutativity)}\\
&= \det(VS^{\mathrm{T}}U^{\mathrm{T}}) \quad \text{(by Theorem 6.1.12 twice)}\\
&= \det[(USV^{\mathrm{T}})^{\mathrm{T}}] = \det(A^{\mathrm{T}}).
\end{aligned}
$$

□

Example 6.1.16. Example 6.1.13a 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. ■

Example 6.1.17. A general 3×3 matrix $A = \begin{bmatrix}a & b & c\\ d & e & f\\ g & h & i\end{bmatrix}$ has determinant $\det A = |A| = aei + bfg + cdh - ceg - afh - bdi$. Its transpose, $A^{\mathrm{T}} = \begin{bmatrix}a & d & g\\ b & e & h\\ c & f & i\end{bmatrix}$, from the rule (6.1)

$$\begin{bmatrix}a & d & g\\ b & e & h\\ c & f & i\end{bmatrix}\begin{matrix}a & b\\ d & e\\ g & h\end{matrix}$$

has determinant

$$\det A^{\mathrm{T}} = aei + dhc + gbf - gec - ahf - dbi = \det A\,.$$

■

Example 6.1.18. Recall Example 3.3.4 showed that the following matrix has the given SVD:

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

Use this SVD to find the magnitude $|\det A|$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Solution: Given the SVD $A = USV^{\mathrm{T}}$, the Product Theorem 6.1.12 gives $\det A = \det(U)\det(S)\det(V^{\mathrm{T}})$.

- $\det U = \pm 1$ by Theorem 6.1.6b as U is an orthogonal matrix.
- Using the Transpose Theorem 6.1.15, $\det(V^{\mathrm{T}}) = \det V = \pm 1$ as V is orthogonal.
- Since $S = \operatorname{diag}(12, 6, 3)$ is diagonal, Theorem 6.1.6a asserts its determinant is the product of the diagonal elements; that is, $\det S = 12 \cdot 6 \cdot 3 = 216$.

Consequently $\det A = (\pm 1)216(\pm 1) = \pm 216$, so $|\det A| = 216$.

■

Theorem 6.1.19. *For any $n \times n$ square matrix A, the magnitude of its determinant $|\det A| = \sigma_1\sigma_2\cdots\sigma_n$, the product of all its singular values.*

Proof. Consider an SVD of the matrix $A = USV^{\mathrm{T}}$. Theorems 6.1.6 and 6.1.12 empowers the following identities:

$$\begin{aligned}
\det A &= \det(USV^{\mathrm{T}}) \\
&= \det(U)\det(S)\det(V^{\mathrm{T}}) \quad \text{(by Theorem 6.1.12)} \\
&= (\pm 1)\det(S)(\pm 1) \quad \text{(by Theorem 6.1.6b)} \\
&= \pm \det S \quad \text{(and since } S \text{ is diagonal)} \\
&= \pm\sigma_1\sigma_2\cdots\sigma_n . \quad \text{(by Theorem 6.1.6a)}
\end{aligned}$$

Hence $|\det A| = \sigma_1\sigma_2\cdots\sigma_n$. □

Example 6.1.20. Confirm Theorem 6.1.19 for the matrix $A = \begin{bmatrix} 10 & 2 \\ 5 & 11 \end{bmatrix}$ of Example 3.3.2.

Solution: Example 3.3.2 gave the SVD

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

so Theorem 6.1.19 asserts $|\det A| = 10\sqrt{2} \cdot 5\sqrt{2} = 100$. Using (6.1) directly, $\det A = 10 \cdot 11 - 2 \cdot 5 = 110 - 10 = 100$ which agrees with the product of the singular values.

■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Example 6.1.21. Use an SVD of the following matrix to find the magnitude of its determinant:

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

Solution: Although an SVD exists for this matrix and so we could form the product of its singular values, the concept of a determinant only applies to square matrices and so there is no such thing as a determinant for this 3×4 matrix A. The task is meaningless.

■

One of the main reasons for studying determinants is to establish when solutions to linear equations may exist or not (albeit only applicable 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 non-zero. Theorem 6.1.23 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 6.1.22. Consider multiplication by the matrix

$$A = \begin{bmatrix} 1 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

whose effect on the unit cube is illustrated below:

As illustrated, this matrix squashes the unit cube onto the x_1x_2-plane ($x_3 = 0$). Consequently the resultant volume is zero and so $\det A = 0$. Because many points in 3D space are squashed onto the same point in the $x_3 = 0$ plane, the action of the matrix cannot be undone. Hence the matrix is not invertible. That the matrix is not invertible and its determinant is zero is not a coincidence. ■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Theorem 6.1.23. *A square matrix A is invertible iff* $\det A \neq 0$ *. If matrix A is invertible, then* $\det(A^{-1}) = 1/(\det A)$.

Proof. First, Theorem 6.1.19 establishes that $\det A \neq 0$ iff all the singular values of square matrix A are non-zero, which by Theorem 3.4.43d is iff matrix A is invertible.

Second, as matrix A is invertible, an inverse A^{-1} exists such that $AA^{-1} = I_n$. Then the product of determinants

$$\begin{aligned}\det(A)\det(A^{-1}) &= \det(AA^{-1}) \quad \text{(by Theorem 6.1.12)}\\ &= \det I_n \quad (\text{from } AA^{-1} = I_n)\\ &= 1 \quad \text{(by Example 6.1.7)}\end{aligned}$$

For an invertible matrix A, $\det A \neq 0$; hence dividing by $\det A$ gives $\det(A^{-1}) = 1/\det A$. □

6.1.1 Exercises

Exercise 6.1.1. For each of the given illustrations of a linear transformation of the unit square, 'guesstimate' by eye the determinant of the matrix of the transformation (estimate to say 33% or so).

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Exercise 6.1.2. For each of the transformations illustrated in Exercise 6.1.1, estimate the matrix of the linear transformation (to within say 10%). Then use formula (6.1) to estimate the determinant of your matrix and confirm Exercise 6.1.1.

Exercise 6.1.3. For each of the following stereo illustrations of a linear transformation of the unit cube, estimate the matrix of the linear transformation (to within say 20%). Then use formula (6.1) to estimate the determinant of the matrix of the transformation.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(e)

(f)

Exercise 6.1.4. For each of the following matrices, use (6.1) to find all the values of k for which the matrix is *not* invertible.

(a) $A = \begin{bmatrix} 0 & 6-2k \\ -2k & -4 \end{bmatrix}$

(b) $B = \begin{bmatrix} 3k & 4-k \\ -4 & 0 \end{bmatrix}$

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

(d) $D = \begin{bmatrix} 2 & -2-4k & -1+k \\ -1-k & 0 & -5k \\ 0 & 0 & 4+2k \end{bmatrix}$

(e) $E = \begin{bmatrix} -1-2k & 5 & 1-k \\ 0 & -2 & 0 \\ 0 & 0 & -7+k \end{bmatrix}$

(f) $F = \begin{bmatrix} k & 0 & -3-3k \\ 3 & 7 & 3k \\ 0 & 2k & 2-k \end{bmatrix}$

Exercise 6.1.5. Find the determinants of each of the following matrices.

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

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

(c) $\begin{bmatrix} -1/2 & 0 & 0 & 0 \\ 0 & 3/2 & 0 & 0 \\ 0 & 0 & -5/2 & 0 \\ 0 & 0 & 0 & -1/2 \end{bmatrix}$

(d) $\begin{bmatrix} 5/6 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 7/6 & 0 \\ 0 & 0 & 0 & 2/3 \end{bmatrix}$

v0.2a

(e) $\begin{bmatrix} 1/3 & 0 & -4/3 & -1 \\ -1/3 & 1/3 & 5/3 & 4/3 \\ 0 & -1/3 & 5/3 & -1 \\ -1/3 & -1/3 & 8/3 & -1/3 \end{bmatrix}$

given $\det \begin{bmatrix} 1 & 0 & -4 & -3 \\ -1 & 1 & 5 & 4 \\ 0 & -1 & 5 & -3 \\ -1 & -1 & 8 & -1 \end{bmatrix} = -8$

(f) $\begin{bmatrix} -1 & 3/2 & -1/2 & 1/2 \\ -2 & -1/2 & -1 & -3/2 \\ -0 & -3/2 & -5/2 & 1/2 \\ -0 & -1/2 & 3 & 3/2 \end{bmatrix}$

given $\det \begin{bmatrix} 2 & -3 & 1 & -1 \\ 4 & 1 & 2 & 3 \\ 0 & 3 & 5 & -1 \\ 0 & 1 & -6 & -3 \end{bmatrix} = -524$

(g) $\begin{bmatrix} -0 & -2/3 & -2 & -4/3 \\ -0 & 1/3 & -1/3 & 1/3 \\ 1 & -1/3 & -5/3 & 2/3 \\ -7/3 & 1/3 & 4/3 & 2/3 \end{bmatrix}$

given $\det \begin{bmatrix} 0 & 2 & 6 & 4 \\ 0 & -1 & 1 & -1 \\ -3 & 1 & 5 & -2 \\ 7 & -1 & -4 & -2 \end{bmatrix} = 246$

(h) $\begin{bmatrix} -12 & -16 & -4 & 12 \\ -4 & 8 & -4 & -16 \\ -0 & -4 & -12 & 4 \\ 4 & -4 & -8 & 4 \end{bmatrix}$

given $\det \begin{bmatrix} 3 & 4 & 1 & -3 \\ 1 & -2 & 1 & 4 \\ 0 & 1 & 3 & -1 \\ -1 & 1 & 2 & -1 \end{bmatrix} = -34$

Exercise 6.1.6. Use Theorems 3.2.27 and 6.1.6a to prove that for any diagonal square matrix D, $\det(D^{-1}) = 1/\det D$ provided $\det D \neq 0$.

Exercise 6.1.7. For each pair of following matrices, by computing in full using (6.1) confirm $\det(AB) = \det(BA) = \det(A)\det(B)$. Show your working.

(a) $A = \begin{bmatrix} -2 & 3/2 \\ -1/2 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 0 & -1/2 \\ 1/2 & -1 \end{bmatrix}$

(b) $A = \begin{bmatrix} -1 & 1 \\ -7/2 & -3/2 \end{bmatrix}$, $B = \begin{bmatrix} 1/2 & -5/2 \\ -1/2 & 3/2 \end{bmatrix}$

(c) $A = \begin{bmatrix} 4 & -1 \\ 4 & -3 \end{bmatrix}$, $B = \begin{bmatrix} -3/2 & 0 \\ -1 & 3/2 \end{bmatrix}$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(d) $A = \begin{bmatrix} 0 & -1 \\ 1/2 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 2 & 0 \\ 1 & -2 \end{bmatrix}$

(e) $A = \begin{bmatrix} 0 & -1/2 & 1/2 \\ 2 & 0 & 0 \\ 0 & 1/2 & 0 \end{bmatrix}$, $B = \begin{bmatrix} -1 & -1 & 0 \\ 0 & 0 & -1 \\ -2 & 0 & 0 \end{bmatrix}$

(f) $A = \begin{bmatrix} 0 & 0 & -2 \\ 0 & 1 & 0 \\ -1/2 & 0 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$

(g) $A = \begin{bmatrix} -2 & -1/2 & 0 \\ 0 & 0 & 2 \\ 0 & -1 & 0 \end{bmatrix}$, $B = \begin{bmatrix} -1 & -1 & 0 \\ 0 & 0 & -1 \\ 0 & -2 & 0 \end{bmatrix}$

(h) $A = \begin{bmatrix} 1 & 2 & -3/2 \\ 0 & 3/2 & 0 \\ -1 & 0 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 0 & 1 & 0 \\ -4 & 2 & 0 \\ -1 & 0 & -5 \end{bmatrix}$

Exercise 6.1.8. Given that $\det(AB) = \det(A)\det(B)$ for any two square matrices of the same size, prove that $\det(AB) = \det(BA)$ (despite $AB \neq BA$ in general).

Exercise 6.1.9. Given that $n \times n$ matrices A and B have $\det A = 3$ and $\det B = -5$, determine the following determinants (if possible).

(a) $\det(AB)$

(b) $\det(B^2)$

(c) $\det(A^4)$

(d) $\det(A + B)$

(e) $\det(A^{-1}B)$

(f) $\det(2A)$

(g) $\det(B^{\mathrm{T}}/2)$

(h) $\det(B^{\mathrm{T}}B)$

Exercise 6.1.10. Let A and P be square matrices of the same size, and let matrix P be invertible. Prove that $\det(P^{-1}AP) = \det A$.

Exercise 6.1.11. Suppose square matrix A satisfies $A^2 = A$ (called idempotent). Determine all possible values of $\det A$. Invent and verify a nontrivial example of a idempotent matrix.

Exercise 6.1.12. Suppose a square matrix A satisfies $A^p = O$ for some integer exponent $p \geq 2$ (called nilpotent). Determine all possible values of $\det A$. Invent and verify a nontrivial example of a nilpotent matrix.

v0.2a

Exercise 6.1.13. Recall that $\det(AB) = \det(A)\det(B)$ for any two square matrices of the same size. For $n \times n$ matrices $A_1, A_2, \ldots, A_\ell$, use induction to prove $\det(A_1 A_2 \cdots A_\ell) = \det(A_1)\det(A_2)\cdots\det(A_\ell)$ for all integer $\ell \geq 2$.

Exercise 6.1.14. To complement the algebraic argument of Theorem 6.1.23, use a geometric argument based upon the transformation of nD-volumes to establish that $\det(A^{-1}) = 1/(\det A)$ for an invertible matrix A.

Exercise 6.1.15. Suppose square matrix $A = \begin{bmatrix} P & O \\ O & Q \end{bmatrix}$ for some square matrices P and Q, and appropriately sized zero matrices O. Give a geometric argument justifying that $\det A = (\det P)(\det Q)$.

v0.2a

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

6.2 Laplace expansion theorem for determinants

Section Contents

6.2.1 Exercises 576

This section develops a so-called row/column algebraic expansion for determinants. This expansion is useful for theoretical purposes. But there are vastly more efficient ways of computing determinants than using a row/column expansion. In MATLAB/Octave one may invoke `det(A)` to compute the determinant of a matrix. You may find this function useful for checking the results of some examples and exercises. However, just like computing an inverse, computing the determinant is expensive and error prone. In medium to large scale problems avoid computing the determinant, something else is almost always better.

Nonetheless, a row/column algebraic expansion for a determinant is useful for small matrix problems, as well as its beautiful theoretical uses. We start with examples of row properties that underpin a row/column algebraic expansion.

Example 6.2.1 (Theorem 6.2.5a)**.** Example 6.1.22 argued geometrically that the determinant is zero for the matrix

$$A = \begin{bmatrix} 1 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Confirm this determinant algebraically.

Solution: Using (6.1), $\det A = 1 \cdot \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 1 \cdot 0 + 0 \cdot 0 \cdot 0 - 0 \cdot \frac{1}{2} \cdot 0 - 1 \cdot 1 \cdot 0 - \frac{1}{2} \cdot 0 \cdot 0 = 0$. (There is a zero from the last row in every term.)

■

Example 6.2.2 (Theorem 6.2.5b)**.** Consider the matrix with two identical rows,

$$A = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{5} \\ 1 & \frac{1}{2} & \frac{1}{5} \\ 0 & \frac{1}{2} & 1 \end{bmatrix}.$$

Confirm algebraically that its determinant is zero. Give a geometric reason for why its determinant has to be zero.

Solution: Using (6.1), $\det A = 1 \cdot \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot \frac{1}{5} \cdot 0 + \frac{1}{5} \cdot 1 \cdot \frac{1}{2} - \frac{1}{5} \cdot \frac{1}{2} \cdot 0 - 1 \cdot \frac{1}{5} \cdot \frac{1}{2} - \frac{1}{2} \cdot 1 \cdot 1 = \frac{1}{2} + \frac{1}{10} - \frac{1}{10} - \frac{1}{2} = 0$.

Geometrically, consider the image of the unit cube under multiplication by A illustrated in stereo below.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Because the first two rows of A are identical the first two components of $A\boldsymbol{x}$ are always identical and hence all points are mapped onto the plane $x_1 = x_2$. The image of the cube thus has zero thickness and hence zero volume. By Definition 6.1.4, $\det A = 0$.

■

Example 6.2.3 (Theorem 6.2.5c). Consider the two matrices with two rows swapped:

$$A = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \\ \frac{1}{5} & \frac{1}{2} & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 1 & 1 \\ 1 & -1 & 0 \\ \frac{1}{5} & \frac{1}{2} & 1 \end{bmatrix}$$

Confirm algebraically that their determinants are the negative of each other. Give a geometric reason why this should be so.

Solution: Using (6.1), $\det A = 1 \cdot 1 \cdot 1 + (-1) \cdot 1 \cdot \frac{1}{5} + 0 \cdot 0 \cdot \frac{1}{2} - 0 \cdot 1 \cdot \frac{1}{5} - 1 \cdot 1 \cdot \frac{1}{2} - (-1) \cdot 0 \cdot 1 = 1 - \frac{1}{5} - \frac{1}{2} = \frac{3}{10}$. Using (6.1), $\det B = 0 \cdot (-1) \cdot 1 + 1 \cdot 0 \cdot \frac{1}{5} + 1 \cdot 1 \cdot \frac{1}{2} - 1 \cdot (-1) \cdot \frac{1}{5} - 0 \cdot 0 \cdot \frac{1}{2} - 1 \cdot 1 \cdot 1 = \frac{1}{2} + \frac{1}{5} - 1 = -\frac{3}{10} = -\det A$.

Geometrically, since the first two rows in A and B are swapped that means that multiplying by the matrix as in $A\boldsymbol{x}$ and $B\boldsymbol{x}$ has the first two components swapped. Hence $A\boldsymbol{x}$ and $B\boldsymbol{x}$ are always the reflection of each other in the plane $x_1 = x_2$. Consequently, the images of the unit cubes under multiplication by A and by B are the reflection of each other in the plane $x_1 = x_2$, as illustrated below, and so the determinants must be the negative of each other.

■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Example 6.2.4 (Theorem 6.2.5d). Compute the determinant of the matrix

$$B = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \\ 2 & 5 & 10 \end{bmatrix}.$$

Compare B with matrix A given in Example 6.2.3, and compare their determinants.

Solution: Using (6.1), $\det B = 1 \cdot 1 \cdot 10 + (-1) \cdot 1 \cdot 2 + 0 \cdot 0 \cdot 5 - 0 \cdot 1 \cdot 2 - 1 \cdot 1 \cdot 5 - (-1) \cdot 0 \cdot 10 = 10 - 2 - 5 = 3$. Matrix B is the same as matrix A except the third row is a factor of ten times bigger. Correspondingly, $\det B = 3 = 10 \times \frac{3}{10} = 10 \det A$.

■

The above four examples are specific cases of the four general properties established as the four parts of the following theorem.

Theorem 6.2.5 (row and column properties of determinants). *Let A be an $n \times n$ matrix.*

(a) If A has a zero row or column, then $\det A = 0$.

(b) If A has two identical rows or columns, then $\det A = 0$.

(c) Let B be obtained by interchanging two rows or columns of A, then $\det B = -\det A$.

(d) Let B be obtained by multiplying any one row or column of A by a scalar k, then $\det B = k \det A$.

Proof. We establish the properties for matrix rows. Then the same property holds for the columns because $\det(A^{\mathrm{T}}) = \det(A)$ (Theorem 6.1.15).

6.2.5d Suppose row i of matrix A is multiplied by k to give a new matrix B. Let the diagonal matrix

$$D := \operatorname{diag}(1, \dots, 1, k, 1, \dots, 1),$$

with the factor k being in the ith row and column. Then $\det D = 1 \times \cdots \times 1 \times k \times 1 \times \cdots \times 1 = k$ by Theorem 6.1.6a. Because multiplication by D multiplies the ith row by the factor k and leaves everything else the same, $B = DA$. Equate determinants of both sides and use the product Theorem 6.1.12: $\det B = \det(DA) = \det(D)\det(A) = k \det A$.

6.2.5a This arises from the case $k = 0$ of property 6.2.5d. Then $A = DA$ because multiplying the ith row of A by $k = 0$ maintains that the row is zero. Consequently, $\det A = \det(DA) = \det(D)\det(A) = 0 \det(A)$ and hence $\det A = 0$.

6.2.5c Suppose rows i and j of matrix A are swapped to form matrix B. Let the matrix E be the identity except with rows i and j swapped:

$$E := \begin{array}{c} \begin{bmatrix} \ddots & & & & \\ & 0 & & 1 & \\ & & \ddots & & \\ & 1 & & 0 & \\ & & & & \ddots \end{bmatrix} \begin{array}{l} \\ \text{row } i \\ \\ \text{row } j \\ \\ \end{array} \\ \begin{array}{ccccc} & i & & j & \end{array} \end{array}$$

where the diagonal dots $\ddots$ denote diagonals of ones, and all other entries of E are zero. Then $B = EA$ as multiplication by E copies row i into row j and vice versa. Equate determinants of both sides and use Theorem 6.1.12: $\det B = \det(EA) = \det(E)\det(A)$.

To find $\det E$ observe that $EE^{\mathrm{T}} = E^2 = I_n$ so E is orthogonal and hence $\det E = \pm 1$ by Theorem 6.1.6b. Geometrically, multiplication by E is a simple reflection in the $(n-1)$D-plane $x_i = x_j$ hence its determinant must be negative, so $\det E = -1$. Consequently, $\det B = \det(E)\det(A) = -\det A$.

6.2.5b Suppose rows i and j of matrix A are identical. Using the matrix E to swap these two identical rows results in the same matrix: that is, $A = EA$. Take determinants of both sides: $\det A = \det(EA) = \det(E)\det(A)$. Since $\det E = -1$ it follows that $\det A = -\det A$. Zero is the only number that equals its negative: thus $\det A = 0$.

□

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.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(a) $\begin{bmatrix} 0 & 2 & 3 & 0 & 4 \\ -2 & 2 & -2 & 0 & -3 \\ 4 & -2 & -4 & 0 & 0 \\ 2 & -1 & -4 & 0 & 2 \\ 5 & 4 & 3 & 0 & -5 \end{bmatrix}$

Solution: $\det = 0$ as the fourth column is all zeros.

(b) $\begin{bmatrix} 0 & 2 & 3 & 1 & 4 \\ -2 & 2 & -2 & 0 & -3 \\ 2 & -1 & -4 & 2 & 2 \\ 4 & -2 & -4 & 1 & 0 \\ 5 & 4 & 3 & -2 & -5 \end{bmatrix}$

Solution: $\det = -\det A = +9$ as the 3rd and 4th rows are swapped.

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

Solution: $\det = 0$ as the 2nd and 5th rows are identical.

(d) $\begin{bmatrix} 0 & 1 & 3 & 1 & 4 \\ -2 & 1 & -2 & 0 & -3 \\ 4 & -1 & -4 & 1 & 0 \\ 2 & -\frac{1}{2} & -4 & 2 & 2 \\ 5 & 2 & 3 & -2 & -5 \end{bmatrix}$

Solution: $\det = \frac{1}{2}\det A = -\frac{9}{2}$ as the 2nd column is half that of A

(e) $\begin{bmatrix} -2 & 2 & -6 & 0 & -3 \\ 4 & -2 & -12 & 1 & 0 \\ 2 & -1 & -12 & 2 & 2 \\ 0 & 2 & 9 & 1 & 4 \\ 5 & 4 & 9 & -2 & -5 \end{bmatrix}$

Solution: $\det = 3(-\det A) = 27$ as this matrix is A with 1st and 4th rows swapped, and the 3rd column multiplied by 3.

(f) $\begin{bmatrix} 0 & 3 & 3 & 1 & 4 \\ -2 & 0 & -2 & 0 & -5 \\ 5 & -1 & -4 & 1 & 0 \\ 2 & -1 & -4 & 2 & 2 \\ 5 & 4 & 6 & -2 & -5 \end{bmatrix}$

Solution: Cannot answer as none of these row and column operations on A appear to give this matrix.

■

Example 6.2.7. 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 the equation of the straight line in the xy-plane through the two points $(2, 3)$ and $(4, 5)$.

Solution: First, equation (6.2) is linear in variables x and y as at most one of x and y occur in each term of the determinant expansion:

$$\begin{bmatrix} 1 & x & y \\ 1 & 2 & 3 \\ 1 & 4 & 5 \end{bmatrix} \begin{matrix} 1 & x \\ 1 & 2 \\ 1 & 4 \end{matrix}$$

Since the equation is linear in x and y, any solution set of (6.2) must be a straight line. Second, equation (6.2) is satisfied when $(x, y) = (2, 3)$ or when $(x, y) = (4, 5)$ as then two rows in the determinant are identical, and so Theorem 6.2.5b assures us the determinant is zero. Thus the solution straight line passes through the two required points.

Let's evaluate the determinant to check: equation (6.2) becomes $1\cdot2\cdot5+x\cdot3\cdot1+y\cdot1\cdot4-y\cdot2\cdot1-1\cdot3\cdot4-x\cdot1\cdot5 = -2-2x+2y = 0$. That is, $y = x + 1$ which does indeed pass through $(2, 3)$ and $(4, 5)$.

■

Example 6.2.8. Without evaluating the determinant, use Theorem 6.2.5 to establish that the determinant equation

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

is, in xyz-space, the equation of the plane through the origin and the two points $(-1, -2, 2)$ and $(3, 5, 2)$.

Solution: As in the previous example, the determinant is linear in x, y and z, so the solutions must be those of one linear equation, namely a plane (Section 1.3.4).

- The solutions include the origin since when $x = y = z = 0$ the first row of the matrix is zero, hence the determinant is zero, and the equation satisfied.
- The solutions include the points $(-1, -2, 2)$ and $(3, 5, 2)$ since when (x, y, z) are either, then two rows in the determinant are identical, so the determinant is zero, and the equation satisfied.

Hence the solutions are the points in the plane passing through the origin and $(-1, -2, 2)$ and $(3, 5, 2)$.

■

The next step in developing a general 'formula' for a determinant is the special class of matrices for which one column or row is zero except for one element.

v0.2a

Example 6.2.9. 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.

Solution: Let's use two different arguments, both illustrating the next theorem and proof.

- Using (6.1),

$$\begin{aligned} \det A &= (-2)(-3)2 + (-1)(-2)\cdot 0 + (-1)1\cdot 0 \\ &\quad - (-1)(-3)0 - (-2)(-2)0 - (-1)1\cdot 2 \\ &= 2\big[(-2)(-3) - (-1)1\big] = 2\cdot 7 = 14\,. \end{aligned}$$

  Observe $\det A = 2 \times \det \begin{bmatrix} -2 & -1 \\ 1 & -3 \end{bmatrix}$ which is the expression to be generalised.

- Alternatively we use the product rule for determinants. Recognise that the matrix A may be factored to $A = FB$ where

$$F = \begin{bmatrix} 1 & 0 & -\frac{1}{2} \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} -2 & -1 & 0 \\ 1 & -3 & 0 \\ 0 & 0 & 2 \end{bmatrix};$$

  just multiply out and see that the last column of F 'fills in' the last column of A from that of B. Consider the geometry of the two transformations arising from multiplication by F and by B.

  - Multiplication by F shears the unit cube as illustrated below.

    Thus the volume of the unit cube after multiplication by F is the square base of area one, times the height of one, which is a volume of one. Consequently $\det F = 1$ by Definition 6.1.4.

  - As illustrated below, multiplication by B has two components.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Firstly, the 2×2 top-left sub-matrix, being bordered by zeros, maps the unit square in the x_1x_2-plane into the base parallelogram in the x_1x_2-plane. The shape of the parallelogram is determined by the top-left sub-matrix $\begin{bmatrix} -2 & -1 \\ 1 & -3 \end{bmatrix}$ (to be called the A_{33} minor) acting on the unit square. Thus the area of the parallelogram is $\det A_{33} = (-2)(-3) - (-1)1 = 7$. Secondly, the 2 in the bottom corner of B stretches objects vertically by a factor of 2 to form a parallelepiped of height 2. Thus the volume of the parallelepiped is $2 \cdot \det A_{33} = 2 \cdot 7 = 14$, which is $\det B$ by Definition 6.1.4.

By the product Theorem 6.1.12, $\det A = \det(FB) = \det(F)\det(B) = 1 \cdot 14 = 14$.

■

Theorem 6.2.10 (almost zero row/column). *Let A be an $n \times n$ matrix. 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) 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

$$\begin{matrix} + & - & + & - & + & \cdots \\ - & + & - & + & - & \cdots \\ + & - & + & - & + & \cdots \\ - & + & - & + & - & \cdots \\ + & - & + & - & + & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{matrix}$$

Proof. We establish the determinant formula (6.3) for matrix rows, then the same result holds for the columns because $\det(A^{\mathrm{T}}) = \det(A)$ (Theorem 6.1.15). First, if the entry $a_{ij} = 0$, then the whole ith row (or jth column) is zero and so $\det A = 0$ by Theorem 6.2.5a. Also, the expression $(-1)^{i+j} a_{ij} \det A_{ij} = 0$ as $a_{ij} = 0$. Consequently, the identity $\det A = (-1)^{i+j} a_{ij} \det A_{ij}$ holds. The rest of this proof addresses the case $a_{ij} \neq 0$.

Second, consider the special case when the last row *and* last column of matrix A is all zero except for $a_{nn} \neq 0$; that is,

$$A = \begin{bmatrix} A_{nn} & \mathbf{0} \\ \mathbf{0}^{\mathrm{T}} & a_{nn} \end{bmatrix}$$

for the minor A_{nn} and $\mathbf{0} \in \mathbb{R}^{n-1}$. Recall Definition 6.1.4: the image

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

of the nD-cube under multiplication by the matrix A is the image of the $(n-1)$D-cube under multiplication by A_{nn} extended orthogonally a length a_{nn} in the orthogonal direction $\boldsymbol{e}_n$ (as illustrated in the margin in 3D). The volume of the nD-image is thus $a_{nn}\times$(volume of the $(n-1)$D-image). Consequently, $\det A = a_{nn}\det A_{nn}$.

Third, consider the special case when the last row of matrix A is all zero except for $a_{nn}\neq 0$; that is,

$$A = \begin{bmatrix} A_{nn} & \boldsymbol{a}'_n \\ \mathbf{0}^{\mathrm{T}} & a_{nn} \end{bmatrix}$$

for the minor A_{nn}, and where $\boldsymbol{a}'_n = (a_{1n}, a_{2n}, \dots, a_{n-1,n})$. Define the two $n\times n$ matrices

$$F := \begin{bmatrix} I_{n-1} & \boldsymbol{a}'_n/a_{nn} \\ \mathbf{0}^{\mathrm{T}} & 1 \end{bmatrix} \quad \text{and } B := \begin{bmatrix} A_{nn} & \mathbf{0} \\ \mathbf{0}^{\mathrm{T}} & a_{nn} \end{bmatrix}.$$

Then $A = FB$ since

$$\begin{aligned} FB &= \begin{bmatrix} I_{n-1} & \boldsymbol{a}'_n/a_{nn} \\ \mathbf{0}^{\mathrm{T}} & 1 \end{bmatrix} \begin{bmatrix} A_{nn} & \mathbf{0} \\ \mathbf{0}^{\mathrm{T}} & a_{nn} \end{bmatrix} \\ &= \begin{bmatrix} I_{n-1}A_{nn} + \boldsymbol{a}'_n/a_{nn}\mathbf{0}^{\mathrm{T}} & I_{n-1}\mathbf{0} + \boldsymbol{a}'_n/a_{nn}\cdot a_{nn} \\ \mathbf{0}^{\mathrm{T}}A_{nn} + 1\cdot\mathbf{0}^{\mathrm{T}} & \mathbf{0}^{\mathrm{T}}\mathbf{0} + 1\cdot a_{nn} \end{bmatrix} \\ &= \begin{bmatrix} A_{nn} & \boldsymbol{a}'_n \\ \mathbf{0}^{\mathrm{T}} & a_{nn} \end{bmatrix} = A. \end{aligned}$$

By Theorem 6.1.12, $\det A = \det(FB) = \det(F)\det(B)$. From the previous part $\det B = a_{nn}\det A_{nn}$, so we just need to determine $\det F$. As illustrated for 3D in the margin, the action of matrix F on the unit nD-cube is that of a simple shear keeping the $(n-1)$D-cube base unchanged (due to the identity I_{n-1} in F). Since the height orthogonal to the $(n-1)$D-cube base is unchanged (due to the one in the bottom corner of F), the action of multiplying by F leaves the volume of the unit nD-cube unchanged at one. Hence $\det F = 1$. Thus $\det A = 1\det(B) = a_{nn}\det A_{nn}$ as required.

Fourth, suppose row i of matrix A is all zero except for entry a_{ij}. Swap rows i and $i+1$, then swap rows $i+1$ and $i+2$, and so on until the original row i is in the last row, and the order of all other rows are unchanged: this takes $(n-i)$ row swaps which changes the sign of the determinant $(n-i)$ times, that is, multiplies it by $(-1)^{n-i}$. Then swap columns j and $j+1$, then swap columns $j+1$ and $j+2$, and so on until the original column j is in the last column: this takes $(n-j)$ column swaps which change the determinant by a factor $(-1)^{n-j}$. The resulting matrix, say C, has the form

$$C = \begin{bmatrix} A_{ij} & \boldsymbol{a}'_j \\ \mathbf{0}^{\mathrm{T}} & a_{ij} \end{bmatrix}$$

for $\boldsymbol{a}'_j$ denoting the jth column of A with the ith entry omitted. Since matrix C has the form addressed in the first part, we know

$\det C = a_{ij} \det A_{ij}$. From the row and column swapping, $\det A = (-1)^{n-i}(-1)^{n-j} \det C = (-1)^{2n-i-j} \det C = (-1)^{-(i+j)} \det C = (-1)^{i+j} \det C = (-1)^{i+j} a_{ij} \det A_{ij}$. □

Example 6.2.11. Use Theorem 6.2.10 to evaluate the determinant of the following matrices.

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

Solution: There are two zeros in the bottom row so the determinant is $(-1)^6 2 \det \begin{bmatrix} -3 & -3 \\ -3 & 2 \end{bmatrix} = 2(-6-9) = -30$.

(b) $\begin{bmatrix} 2 & -1 & 7 \\ 0 & 3 & 0 \\ 2 & 2 & 5 \end{bmatrix}$

Solution: There are two zeros in the middle row so the determinant is $(-1)^4 3 \det \begin{bmatrix} 2 & 7 \\ 2 & 5 \end{bmatrix} = 3(10-14) = -12$.

(c) $\begin{bmatrix} 2 & 4 & 3 \\ 8 & 0 & -1 \\ -5 & 0 & -2 \end{bmatrix}$

Solution: There are two zeros in the middle column so the determinant is $(-1)^3 4 \det \begin{bmatrix} 8 & -1 \\ -5 & -2 \end{bmatrix} = -4(-16-5) = 84$.

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

Solution: There are two zeros in the first column so the determinant is $(-1)^2 2 \det \begin{bmatrix} -2 & -3 \\ 2 & 4 \end{bmatrix} = 2(-8+6) = -4$.

■

Example 6.2.12. Use Theorem 6.2.10 to evaluate the determinant of the so-called triangular matrix

$$A = \begin{bmatrix} 2 & -2 & 3 & 1 & 0 \\ 0 & 2 & -1 & -1 & -7 \\ 0 & 0 & 5 & -2 & -9 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 3 \end{bmatrix}$$

Solution: The last row is all zero except for the last element, so

$$\det A = (-1)^{10} 3 \det \begin{bmatrix} 2 & -2 & 3 & 1 \\ 0 & 2 & -1 & -1 \\ 0 & 0 & 5 & -2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

(as the last row is zero except the last)

$$= 3 \cdot (-1)^8 1 \det \begin{bmatrix} 2 & -2 & 3 \\ 0 & 2 & -1 \\ 0 & 0 & 5 \end{bmatrix}$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$
\begin{aligned}
&\quad \text{(as the last row is zero except the last)}\\
&= 3\cdot 1\cdot(-1)^6 5\det\begin{bmatrix}2 & -2\\ 0 & 2\end{bmatrix}\\
&\quad \text{(as the last row is zero except the last)}\\
&= 3\cdot 1\cdot 5(-1)^4 2\det\begin{bmatrix}2\end{bmatrix}\\
&= 3\cdot 1\cdot 5\cdot 2\cdot 2 = 60\,.
\end{aligned}
$$

■

The relative simplicity of finding the determinant in Example 6.2.12 indicates that there is something special and memorable about matrices with zeros in the lower-left 'triangle'. There is, as expressed by the following definition and theorem.

Definition 6.2.13. *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:* [1]

- *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)*

$$
\begin{bmatrix}
a_{11} & 0 & \cdots & 0 & 0\\
a_{21} & a_{22} & 0 & \cdots & 0\\
\vdots & \vdots & \ddots & \ddots & \vdots\\
a_{n-1\,1} & a_{n-1\,2} & \cdots & a_{n-1\,n-1} & 0\\
a_{n\,1} & a_{n\,2} & \cdots & a_{n\,n-1} & a_{n\,n}
\end{bmatrix}.
$$

Any square diagonal matrix is also an upper triangular matrix, and also a lower triangular matrix. Thus the following theorem encompasses square diagonal matrices and so generalises Theorem 6.1.6a.

[1] From time-to-time, some people may call an upper triangular matrix either a right triangular or an upper-right triangular matrix. Correspondingly, from time-to-time, some people may call a lower triangular matrix either a left triangular or a lower-left triangular matrix.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Theorem 6.2.14 (triangular matrix)**.** *Let A be an $n \times n$ triangular matrix. The determinant of A is the product of the diagonal entries,* $\det A = a_{11}a_{22}\cdots a_{nn}$.

Proof. A little induction proves the determinant of a triangular matrix is the product of its diagonal entries: only consider upper triangular matrices as transposing the matrix caters for lower triangular matrices.

First, for 1×1 matrices the result is trivial. The results is also straightforward for 2×2 matrices since the determinant

$$\begin{vmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{vmatrix} = a_{11}a_{22} - 0a_{12} = a_{11}a_{22}$$

which is the product of the diagonal entries as required.

Second, assume the property for $(n-1) \times (n-1)$ matrices. Also the upper triangular matrix A has the form

$$A = \begin{bmatrix} A_{nn} & \boldsymbol{a}'_n \\ \mathbf{0}^{\mathrm{T}} & a_{nn} \end{bmatrix}.$$

for $(n-1) \times (n-1)$ minor A_{nn}. Theorem 6.2.10 establishes $\det A = a_{nn} \det A_{nn}$. Since the minor A_{nn} is upper triangular and $(n-1) \times (n-1)$, by assumption $\det A_{nn} = a_{11}a_{22}\cdots a_{n-1,n-1}$. Consequently, $\det A = a_{nn} \det A_{nn} = a_{nn}a_{11}a_{22}\cdots a_{n-1,n-1}$. Induction then establishes the theorem. □

Example 6.2.15. Find the determinant of those of the following matrices which are triangular.

(a) $\begin{bmatrix} -1 & -1 & -1 & -5 \\ 0 & -4 & 1 & 4 \\ 0 & 0 & 7 & 0 \\ 0 & 0 & 0 & -3 \end{bmatrix}$

Solution: This is upper triangular, and its determinant is $(-1) \cdot (-4) \cdot 7 \cdot (-3) = -84$.

(b) $\begin{bmatrix} -3 & 0 & 0 & 0 \\ -4 & 2 & 0 & 0 \\ -1 & 1 & 1 & 0 \\ -2 & -3 & 7 & -1 \end{bmatrix}$

Solution: This is lower triangular, and its determinant is $(-3) \cdot 2 \cdot 1 \cdot) - 1) = 6$.

(c) $\begin{bmatrix} -4 & 0 & 0 & 0 \\ 2 & -2 & 0 & 0 \\ -5 & -3 & -2 & 0 \\ -2 & 5 & -2 & 0 \end{bmatrix}$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Solution: This is lower triangular, and its determinant is zero as it has a column of zeros.

(d) $\begin{bmatrix} 0.2 & 0 & 0 & 0 \\ 0 & 1.1 & 0 & 0 \\ 0 & 0 & -0.5 & 0 \\ 0 & 0 & 0 & 0.9 \end{bmatrix}$

Solution: This diagonal matrix is both upper and lower triangular, and its determinant is $0.2 \cdot 1.1 \cdot (-0.5) \cdot 0.9 = -0.099$.

(e) $\begin{bmatrix} 1 & -1 & 1 & -3 \\ 0 & 0 & 0 & -5 \\ 0 & 0 & -3 & -4 \\ 0 & -2 & 1 & -2 \end{bmatrix}$

Solution: This is not triangular, so we do not have to compute its determinant. Nonetheless, if we swap the 2nd and 4th rows, then the result is the upper triangular $\begin{bmatrix} 1 & -1 & 1 & -3 \\ 0 & -2 & 1 & -2 \\ 0 & 0 & -3 & -4 \\ 0 & 0 & 0 & -5 \end{bmatrix}$ and its determinant is $1 \cdot (-2) \cdot (-3) \cdot (-5) = -30$. But the row swap changes the sign so the determinant of the original matrix is $-(-30) = 30$.

(f) $\begin{bmatrix} 0 & 0 & 0 & -3 \\ 0 & 0 & 2 & -4 \\ 0 & -1 & 4 & -1 \\ -6 & 1 & 5 & 1 \end{bmatrix}$

Solution: This is *not* triangular, so we do not have to compute its determinant. Nonetheless, if we swap the 1st and 4th rows, and the 2nd and 3rd rows, then the result is the lower triangular $\begin{bmatrix} -3 & 0 & 0 & 0 \\ -4 & 2 & 0 & 0 \\ -1 & 4 & -1 & 0 \\ 1 & 5 & 1 & -6 \end{bmatrix}$ and its determinant is $(-3) \cdot 2 \cdot (-1) \cdot (-6) = -36$. But each row swap changes the sign so the determinant of the original matrix is $(-1)^2(-36) = 36$.

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

Solution: This is not triangular, so we do not have to compute its determinant. Nonetheless, if we swap the 2nd and 4th columns, the 1st and 2nd rows, and the 3rd and 4th rows, then the result is the lower triangular $\begin{bmatrix} -2 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -1 & 2 & 4 & 0 \\ 2 & -2 & -1 & -2 \end{bmatrix}$ and its determinant is $(-2) \cdot 1 \cdot 4 \cdot (-2) = 16$. But each row and column swap changes the sign so the determinant of the original matrix is $(-1)^3 16 = -16$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

■

The above case of triangular matrices is a short detour from the main development of this section which is to derive a formula for determinants in general. The following two examples introduce the next property we need before establishing a general formula for determinants.

Example 6.2.16. 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

$$\begin{aligned}\begin{vmatrix} a & b \\ c & d \end{vmatrix} &= ad - bc \\ &= (ad - 0c) + (0d - bc) \\ &= \begin{vmatrix} a & 0 \\ c & d \end{vmatrix} + \begin{vmatrix} 0 & b \\ c & d \end{vmatrix}.\end{aligned}$$

  That is, the original determinant is the same as the sum of two determinants, each with a zero in the first row and the other row unchanged. Equivalently, the first row is decomposed 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

$$\begin{aligned}\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} &= \quad aei + bfg + cdh - ceg - afh - bdi \\ &= + aei + 0fg + 0dh - 0eg - afh - 0di \\ &\quad + 0ei + bfg + 0dh - 0eg - 0fh - bdi \\ &\quad + 0ei + 0fg + cdh - ceg - 0fh - 0di \\ &= \begin{vmatrix} a & 0 & 0 \\ d & e & f \\ g & h & i \end{vmatrix} + \begin{vmatrix} 0 & b & 0 \\ d & e & f \\ g & h & i \end{vmatrix} + \begin{vmatrix} 0 & 0 & c \\ d & e & f \\ g & h & i \end{vmatrix}.\end{aligned}$$

  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. Equivalently, the first row is decomposed as $\begin{bmatrix} a & b & c \end{bmatrix} = \begin{bmatrix} a & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & b & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & c \end{bmatrix}$, while the other rows are unchanged.

This sort of rearrangement of a determinant makes progress because then Theorem 6.2.10 helps by finding the determinant of the resultant matrices that have almost all zero rows. ■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Example 6.2.17. 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 that of A. ■

Theorem 6.2.18 (sum formula)**.** *Let A, B and C be $n \times n$ matrices. If 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 property holds when "column" is replaced by "row" throughout.*

Proof. We establish the theorem for matrix columns. Then the same results holds for the rows because $\det(A^{\mathrm{T}}) = \det(A)$ (Theorem 6.1.15). As a prelude to the general geometric proof, consider the 2×2 case and the second column (as also established algebraically by Example 6.2.17). Write the matrices in terms of their column vectors, and draw the determinant parallelogram areas as shown below: let

$A = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 \end{bmatrix}$,

$B = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{b} \end{bmatrix}$,

$C = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{c} \end{bmatrix}$,

The matrices A, B and C all have the same first column $\boldsymbol{a}_1$, wheras the second columns satisfy $\boldsymbol{a}_2 = \boldsymbol{b} + \boldsymbol{c}$ by the condition of the

theorem. Because these parallelograms have common side $\boldsymbol{a}_1$ we can stack the area for $\det C$ on top of that for $\det B$, and because $\boldsymbol{a}_2 = \boldsymbol{b}+\boldsymbol{c}$ the top edge of the stack matches that for the area $\det A$, as shown below

The base of the stacked shape lies on the line $\mathbb{A}$, and let $\boldsymbol{n}$ denote the orthogonal/normal direction (as shown). Because the shape has the same cross-section in lines parallel to $\mathbb{A}$, its area is the area of the base times the height of the stacked shape in the direction $\boldsymbol{n}$. But this is precisely the same height and base as the area for $\det A$, hence $\det A = \det B + \det C$.

A general proof for the last column uses the same diagrams, albeit schematically. Let matrices

$$A = \begin{bmatrix} A' & \boldsymbol{a}_n \end{bmatrix}, \quad B = \begin{bmatrix} A' & \boldsymbol{b} \end{bmatrix}, \quad C = \begin{bmatrix} A' & \boldsymbol{c} \end{bmatrix},$$

where the $n \times (n-1)$ matrix $A' = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_{n-1} \end{bmatrix}$ is common to all three, and where the three last columns satisfy $\boldsymbol{a}_n = \boldsymbol{b} + \boldsymbol{c}$. Consider the nD-parallelepipeds whose nD-volumes are the three determinants, as before. Because these nD-parallelepipeds have common base of the $(n-1)$D-parallelepiped formed by the columns of A', we can and do stack the nD-volume for $\det C$ on top of that for $\det B$, and because $\boldsymbol{a}_n = \boldsymbol{b} + \boldsymbol{c}$ the top $(n-1)$D-parallelepiped of the stack matches that for the nD-volume $\det A$, as shown schematically before. The base of the stacked shape lies on the subspace $\mathbb{A} = \operatorname{span}\{\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_{n-1}\}$, and let $\boldsymbol{n}$ denote the orthogonal/normal direction to $\mathbb{A}$ (as shown schematically). Because the shape has the same cross-section parallel to $\mathbb{A}$, its nD-volume is the $(n-1)$D-volume of the base $(n-1)$D-parallelepiped times the height of the stacked shape in the direction $\boldsymbol{n}$. But this is precisely the same height and base as the nD-volume for $\det A$, hence $\det A = \det B + \det C$.

Lastly, when it is the jth column for which $\boldsymbol{a}_j = \boldsymbol{b}+\boldsymbol{c}$ and all others columns are identical, then swap column j with column n in all matrices. Theorem 6.2.5c asserts the signs of the three determinants are changed by this swapping. The above proof for the last column case then assures us $(-\det A) = (-\det B) + (-\det C)$; that is, $\det A = \det B + \det C$, as required. □

The sum formula Theorem 6.2.18 leads to the common way to compute determinants by hand for matrices larger than 3×3, albeit not generally practical for matrices significantly larger.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Example 6.2.19. Use Theorems 6.2.18 and 6.2.10 to evaluate the determinant of matrix

$$A = \begin{bmatrix} -2 & 1 & -1 \\ 1 & -6 & -1 \\ 2 & 1 & 0 \end{bmatrix}.$$

Solution: Write the first row of A as the sum

$$\begin{aligned} \begin{bmatrix} -2 & 1 & -1 \end{bmatrix} &= \begin{bmatrix} -2 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 1 & -1 \end{bmatrix} \\ &= \begin{bmatrix} -2 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & -1 \end{bmatrix}. \end{aligned}$$

Then using Theorem 6.2.18 twice, the determinant

$$\begin{aligned} &\begin{vmatrix} -2 & 1 & -1 \\ 1 & -6 & -1 \\ 2 & 1 & 0 \end{vmatrix} \\ &= \begin{vmatrix} -2 & 0 & 0 \\ 1 & -6 & -1 \\ 2 & 1 & 0 \end{vmatrix} + \begin{vmatrix} 0 & 1 & -1 \\ 1 & -6 & -1 \\ 2 & 1 & 0 \end{vmatrix} \\ &= \begin{vmatrix} -2 & 0 & 0 \\ 1 & -6 & -1 \\ 2 & 1 & 0 \end{vmatrix} + \begin{vmatrix} 0 & 1 & 0 \\ 1 & -6 & -1 \\ 2 & 1 & 0 \end{vmatrix} + \begin{vmatrix} 0 & 0 & -1 \\ 1 & -6 & -1 \\ 2 & 1 & 0 \end{vmatrix} \end{aligned}$$

Each of these last three matrices has the first row zero except for one element, so Theorem 6.2.10 applies to each of the three determinants to give

$$\begin{aligned} &\begin{vmatrix} -2 & 1 & -1 \\ 1 & -6 & -1 \\ 2 & 1 & 0 \end{vmatrix} \\ &= (-1)^2(-2)\begin{vmatrix} -6 & -1 \\ 1 & 0 \end{vmatrix} + (-1)^3(1)\begin{vmatrix} 1 & -1 \\ 2 & 0 \end{vmatrix} + (-1)^4(-1)\begin{vmatrix} 1 & -6 \\ 2 & 1 \end{vmatrix} \\ &= (-2)\cdot 1 - (1)\cdot 2 + (-1)\cdot 13 = -17 \end{aligned}$$

upon using the well-known formula (6.1) for the three 2×2 determinants.

Alternatively, we could have used any row or column instead of the first row. For example, let's use the last column as it usefully already has a zero entry: write the last column of matrix A as $(-1,-1,0) = (-1,0,0) + (0,-1,0)$, then by Theorem 6.2.18 the determinant

$$\begin{aligned} \begin{vmatrix} -2 & 1 & -1 \\ 1 & -6 & -1 \\ 2 & 1 & 0 \end{vmatrix} &= \begin{vmatrix} -2 & 1 & -1 \\ 1 & -6 & 0 \\ 2 & 1 & 0 \end{vmatrix} + \begin{vmatrix} -2 & 1 & 0 \\ 1 & -6 & -1 \\ 2 & 1 & 0 \end{vmatrix} \\ &\qquad \text{(so by Theorem 6.2.10)} \\ &= (-1)^4(-1)\begin{vmatrix} 1 & -6 \\ 2 & 1 \end{vmatrix} + (-1)^5(-1)\begin{vmatrix} -2 & 1 \\ 2 & 1 \end{vmatrix} \\ &= (-1)\cdot 13 - (-1)\cdot(-4) = -17, \end{aligned}$$

as before.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

■

Theorem 6.2.20 (Laplace expansion theorem). *Consider an $n \times n$ 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,*

$$\begin{aligned}\det A &= (-1)^{i+1} a_{i1} \det A_{i1} + (-1)^{i+2} a_{i2} \det A_{i2} \\ &\quad + \cdots + (-1)^{i+n} a_{in} \det A_{in} \\ &= (-1)^{j+1} a_{1j} \det A_{1j} + (-1)^{j+2} a_{2j} \det A_{2j} \\ &\quad + \cdots + (-1)^{j+n} a_{nj} \det A_{nj} \,. \end{aligned} \tag{6.4}$$

Proof. We establish the expansion for matrix rows: then the same property holds for the columns because $\det(A^{\mathrm{T}}) = \det(A)$ (Theorem 6.1.15). First prove the expansion for a first row expansion, and then second for any row. So first use the sum Theorem 6.2.18 $(n-1)$ times to deduce

$$\begin{aligned}&\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} \\ &= \begin{vmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} + \begin{vmatrix} 0 & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} \\ &\vdots \\ &= \begin{vmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} + \begin{vmatrix} 0 & a_{12} & \cdots & 0 \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} \\ &\quad + \cdots + \begin{vmatrix} 0 & 0 & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}\end{aligned}$$

As each of these n determinants has the first row zero except for one element, Theorem 6.2.10 applies to give

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

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$
\begin{aligned}
&= (-1)^2 a_{11} \begin{vmatrix} a_{22} & \cdots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{n2} & \cdots & a_{nn} \end{vmatrix} + (-1)^3 a_{12} \begin{vmatrix} a_{21} & \cdots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} \\
&\quad + \cdots + (-1)^{n+1} a_{1n} \begin{vmatrix} a_{21} & a_{22} & \cdots & a_{2,n-1} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{n,n-1} \end{vmatrix} \\
&= (-1)^2 a_{11} \det A_{11} + (-1)^3 a_{12} \det A_{12} \\
&\quad + \cdots + (-1)^{n+1} a_{1n} \det A_{1n} ,
\end{aligned}
$$

which is the case $i = 1$ of formula (6.4).

Second, for the general ith row expansion, let a new matrix B be obtained from A by swapping the ith row up $(i-1)$ times to form the first row of B and leaving the other rows from A in the same order. Then the elements $b_{1j} = a_{ij}$, and also the minors $B_{1j} = A_{ij}$. Apply formula (6.4) to the first row of B (just proved) to give

$$
\begin{aligned}
\det B &= (-1)^2 b_{11} \det B_{11} + (-1)^3 b_{12} \det B_{12} \\
&\quad + \cdots + (-1)^{n+1} b_{1n} \det B_{1n} \\
&= (-1)^2 a_{i1} \det A_{i1} + (-1)^3 a_{i2} \det A_{i2} \\
&\quad + \cdots + (-1)^{n+1} a_{in} \det A_{in} .
\end{aligned}
$$

But by Theorem 6.2.5c each of the $(i-1)$ row swaps in forming B changes the sign of the determinant: hence

$$
\begin{aligned}
\det A &= (-1)^{i-1} \det B \\
&= (-1)^{i-1+2} a_{i1} \det A_{i1} + (-1)^{i-1+3} a_{i2} \det A_{i2} \\
&\quad + \cdots + (-1)^{i-1+n+1} a_{in} \det A_{in} \\
&= (-1)^{i+1} a_{i1} \det A_{i1} + (-1)^{i+2} a_{i2} \det A_{i2} \\
&\quad + \cdots + (-1)^{i+n} a_{in} \det A_{in} ,
\end{aligned}
$$

as required. □

Example 6.2.21. Use the Laplace expansion (6.4) to find the determinant of the following matrices.

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

Solution: The first column has two zeros, so expand in the first column:

$$
\begin{aligned}
\det &= (-1)^3 (-1) \det \begin{bmatrix} 2 & 1 & 2 \\ 2 & -1 & -1 \\ -1 & -1 & 1 \end{bmatrix} \\
&\quad + (-1)^4 (1) \det \begin{bmatrix} 2 & 1 & 2 \\ 2 & -1 & -2 \\ -1 & -1 & 1 \end{bmatrix}
\end{aligned}
$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$
\begin{aligned}
&\quad \text{(using (6.1) for these } 3\times 3 \text{ matrices)}\\
&= (-2+1-4-2-2-2)\\
&\quad + (-2+2-4-2-4-2)\\
&= -23\,.
\end{aligned}
$$

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

Solution: The last column has three zeros, so expand in the last column:

$$
\begin{aligned}
\det &= (-1)^8(3)\det\begin{bmatrix} -3 & -1 & 1 \\ -2 & 0 & -2 \\ -3 & -2 & 0 \end{bmatrix}\\
&\quad \text{(expand in the middle row (say) due to its zero)}\\
&= 3\left\{(-1)^3(-2)\det\begin{bmatrix} -1 & 1 \\ -2 & 0 \end{bmatrix}\right.\\
&\qquad \left. + (-1)^5(-2)\det\begin{bmatrix} -3 & -1 \\ -3 & -2 \end{bmatrix}\right\}\\
&= 3\left\{2(0+2)+2(6-3)\right\}\\
&= 30\,.
\end{aligned}
$$

■

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 6.2.23), 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 6.2.14). In any case, remember that except for theoretical purposes there is rarely need to compute a medium to large determinant.

Example 6.2.22. 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 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. ■

v0.2a

Theorem 6.2.23. *The determinant of any $n \times n$ matrix expands to the sum $(\pm)$ of $n!$ terms, where each term is a product of n factors such that each factor comes from different rows and columns of the matrix.*

Proof. Use induction on the size of the matrix. First, the properties holds for 1×1 matrices as $\det \begin{bmatrix} a_{11} \end{bmatrix} = a_{11}$ is one term of one factor from the only row and column of the matrix.

Second, assume the determinant of any $(n-1) \times (n-1)$ matrix may be written the sum $(\pm)$ of $(n-1)!$ terms, where each term is a product of $(n-1)$ factors such that each factor comes from different rows and columns. Consider any $n \times n$ matrix A. By the Laplace Expansion Theorem 6.2.20, $\det A$ may be written as the sum $(\pm)$ of n terms of the form $a_{ij} \det A_{ij}$. By induction assumption, the $(n-1) \times (n-1)$ minors A_{ij} have determinants with $(n-1)!$ terms, each of $(n-1)$ factors and so the n terms in a Laplace Expansion of $\det A$ expands to $n(n-1)! = n!$ terms, each term being of n factors through the multiplication by the entry a_{ij}. Further, recall the minor A_{ij} is obtained from A by omitting row i and column j, and so the minor has no elements from the same row or column as a_{ij}. Consequently, each term in the determinant only has factors from different rows and columns, as required. By induction the theorem holds for all n. □

6.2.1 Exercises

Exercise 6.2.1. In each of the following, the determinant of a matrix is given. Use Theorem 6.2.5 on the row and column properties of a determinant to find the determinant of the other four listed matrices. Give reasons for your answers.

(a) $\det \begin{bmatrix} -2 & 1 & -4 \\ -2 & -1 & 2 \\ -2 & 5 & -1 \end{bmatrix} = 60$

i. $\begin{bmatrix} 1 & 1 & -4 \\ 1 & -1 & 2 \\ 1 & 5 & -1 \end{bmatrix}$ ii. $\begin{bmatrix} -2 & 1 & -4 \\ -2 & 1 & -4 \\ -2 & 5 & -1 \end{bmatrix}$

iii. $\begin{bmatrix} -2 & 1 & -4 \\ -0.2 & -0.1 & 0.2 \\ -2 & 5 & -1 \end{bmatrix}$ iv. $\begin{bmatrix} -2 & 1 & -4 \\ -2 & 5 & -1 \\ -2 & -1 & 2 \end{bmatrix}$

(b) $\det \begin{bmatrix} -1 & -1 & 4 & -6 \\ 4 & -2 & -2 & -1 \\ 0 & -3 & -1 & -4 \\ 3 & 2 & 1 & 1 \end{bmatrix} = 72$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

i. $\begin{bmatrix} 0 & -3 & -1 & -4 \\ 4 & -2 & -2 & -1 \\ -1 & -1 & 4 & -6 \\ 3 & 2 & 1 & 1 \end{bmatrix}$ ii. $\begin{bmatrix} -1 & -1 & 4 & -6 \\ 4 & -2 & -2 & -1 \\ 0 & 0 & 0 & 0 \\ 3 & 2 & 1 & 1 \end{bmatrix}$

iii. $\begin{bmatrix} -1 & -1 & 4 & -6 \\ 2 & -1 & -1 & -1/2 \\ 0 & -3 & -1 & -4 \\ 3 & 2 & 1 & 1 \end{bmatrix}$ iv. $\begin{bmatrix} -1 & -1 & 4 & -6 \\ -2 & 4 & -2 & -1 \\ -3 & 0 & -1 & -4 \\ 2 & 3 & 1 & 1 \end{bmatrix}$

(c) $\det \begin{bmatrix} 2 & -3 & 2 & -3 \\ 0 & -1 & -1 & -2 \\ 2 & 1 & -2 & -3 \\ -4 & -1 & -4 & 0 \end{bmatrix} = 16$

i. $\begin{bmatrix} 0 & -1 & -1 & -2 \\ 2 & -3 & 2 & -3 \\ -4 & -1 & -4 & 0 \\ 2 & 1 & -2 & -3 \end{bmatrix}$ ii. $\begin{bmatrix} 1 & 12 & 2 & -3 \\ 0 & 4 & -1 & -2 \\ 1 & -4 & -2 & -3 \\ -2 & 4 & -4 & 0 \end{bmatrix}$

iii. $\begin{bmatrix} 4 & 4 & -6 & -6 \\ 0 & -1 & -1 & -2 \\ 2 & -2 & 1 & -3 \\ -4 & -4 & -1 & 0 \end{bmatrix}$ iv. $\begin{bmatrix} 0 & -1 & -0.5 & -2 \\ 2 & -3 & 1 & -3 \\ 2 & 1 & -1 & -3 \\ -4 & -1 & -2 & 0 \end{bmatrix}$

(d) $\det \begin{bmatrix} 0.3 & -0.1 & -0.1 & 0.4 \\ 0.2 & 0.3 & 0 & 0.1 \\ 0.1 & -0.1 & -0.3 & -0.2 \\ -0.1 & -0.2 & 0.4 & 0.2 \end{bmatrix} = 0.01$

i. $\begin{bmatrix} 3 & -1 & -1 & 4 \\ 2 & 3 & 0 & 1 \\ 1 & -1 & -3 & -2 \\ -1 & -2 & 4 & 2 \end{bmatrix}$ ii. $\begin{bmatrix} 0.3 & 0.2 & 0 & 2 \\ 0.2 & -0.6 & 0 & 0.5 \\ 0.1 & 0.2 & 0 & -1 \\ -0.1 & 0.4 & 0 & 1 \end{bmatrix}$

iii. $\begin{bmatrix} 0.2 & 0.3 & 0 & 0.1 \\ 0.1 & -0.1 & -0.3 & -0.2 \\ -0.1 & -0.2 & 0.4 & 0.2 \\ 0.3 & -0.1 & -0.1 & 0.4 \end{bmatrix}$ iv. $\begin{bmatrix} 0.3 & 0.4 & -0.1 & 0.4 \\ 0.2 & 0.1 & 0 & 0.1 \\ 0.1 & -0.2 & -0.3 & -0.2 \\ -0.1 & 0.2 & 0.4 & 0.2 \end{bmatrix}$

Exercise 6.2.2. Recall Example 6.2.7. For each pair of given points, (x_1, y_1) and (x_2, y_2), evaluate the determinant in the equation

$$\det \begin{bmatrix} 1 & x & y \\ 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \end{bmatrix} = 0$$

to find an equation for the straight line through the two given points. Show your working.

(a) $(-3\,,-6)$, $(2\,,3)$ (b) $(3\,,-2)$, $(-3\,,0)$

v0.2a

(c) $(1\,,-4)$, $(-3\,,1)$ (d) $(-1\,,0)$, $(-2\,,1)$

(e) $(6\,,1)$, $(2\,,-1)$ (f) $(3\,,-8)$, $(7\,,-2)$

Exercise 6.2.3. Using mainly the properties of Theorem 6.2.5 detail an argument that the following determinant equations each give an equation for the line through two given points (x_1, y_1) and (x_2, y_2).

(a) $\det \begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x & y \\ 1 & x_2 & y_2 \end{bmatrix} = 0$ (b) $\det \begin{bmatrix} 1 & 1 & 1 \\ x_2 & x_1 & x \\ y_2 & y_1 & y \end{bmatrix} = 0$

Exercise 6.2.4. Recall Example 6.2.8. For each pair of given points, (x_1, y_1, z_1) and (x_2, y_2, z_2), evaluate the determinant in the equation

$$\det \begin{bmatrix} x & y & z \\ x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{bmatrix} = 0$$

to find an equation for the plane that passes through the two given points and the origin. Show your working.

(a) $(-1\,,-1\,,-3)$, $(3\,,-5\,,-1)$ (b) $(0\,,0\,,-2)$, $(4\,,-4\,,0)$

(c) $(-1\,,2\,,2)$, $(-1\,,-3\,,2)$ (d) $(4\,,-2\,,0)$, $(-3\,,-4\,,-1)$

(e) $(-4\,,-1\,,2)$, $(-3\,,-2\,,2)$ (f) $(2\,,2\,,3)$, $(2\,,1\,,4)$

Exercise 6.2.5. Using mainly the properties of Theorem 6.2.5 detail an argument that the following determinant equations each give an equation for the plane passing through the origin and the two given points (x_1, y_1, z_1) and (x_2, y_2, z_2).

(a) $\det \begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x & y & z \end{bmatrix} = 0$ (b) $\det \begin{bmatrix} x_2 & x & x_1 \\ y_2 & y & y_1 \\ z_2 & z & z_1 \end{bmatrix} = 0$

Exercise 6.2.6. Prove Theorems 6.2.5a, 6.2.5d and 6.2.5b using basic geometric arguments about the transformation of the unit nD-cube.

Exercise 6.2.7. Use Theorem 6.2.5 to prove that if a square matrix A has two non-zero rows proportional to each other, then $\det A = 0$. Why does it immediately follow that (instead of rows) if the matrix has two non-zero columns proportional to each other, then $\det A = 0$.

v0.2a

Exercise 6.2.8. Use Theorem 6.2.10, and then (6.1), to evaluate the following determinants. Show your working.

(a) $\det\begin{bmatrix} 6 & 1 & 1 \\ -1 & 3 & -8 \\ -6 & 0 & 0 \end{bmatrix}$ (b) $\det\begin{bmatrix} 4 & 8 & 0 \\ 3 & -2 & 0 \\ -1 & -1 & -3 \end{bmatrix}$

(c) $\det\begin{bmatrix} 0 & 0 & 3 \\ -1 & -3 & -3 \\ -3 & -5 & 2 \end{bmatrix}$ (d) $\det\begin{bmatrix} -4 & 0 & -5 \\ 1 & -7 & -1 \\ 4 & 0 & 4 \end{bmatrix}$

(e) $\det\begin{bmatrix} 2 & -4 & -2 & -2 \\ 0 & 1 & -3 & -2 \\ -2 & 0 & 0 & 0 \\ 5 & -8 & 1 & 7 \end{bmatrix}$ (f) $\det\begin{bmatrix} 0 & -5 & 0 & 0 \\ -7 & 2 & 2 & 1 \\ 1 & -2 & -2 & -5 \\ 6 & 8 & -2 & 0 \end{bmatrix}$

(g) $\det\begin{bmatrix} 0 & 2 & -4 & 3 \\ -6 & 6 & -2 & 0 \\ -1 & -8 & 4 & 0 \\ 2 & -2 & -1 & 0 \end{bmatrix}$ (h) $\det\begin{bmatrix} -3 & 4 & -1 & -1 \\ 4 & 8 & 1 & 6 \\ 0 & 7 & 0 & 0 \\ 2 & -6 & 1 & 2 \end{bmatrix}$

Exercise 6.2.9. Use the triangular matrix Theorem 6.2.14, as well as the row/column properties of Theorem 6.2.5, to find the determinants of each of the following matrices. Show your argument.

(a) $\begin{bmatrix} -6 & -4 & -7 & 2 \\ 0 & -2 & -1 & 1 \\ 0 & 0 & -4 & 1 \\ 0 & 0 & 0 & -2 \end{bmatrix}$ (b) $\begin{bmatrix} 2 & 0 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ -5 & -1 & 2 & 0 \\ 2 & 4 & -1 & 1 \end{bmatrix}$

(c) $\begin{bmatrix} 0 & 0 & -6 & -6 \\ -2 & -2 & -2 & 1 \\ 0 & 0 & 0 & -4 \\ 0 & 0 & -1 & 7 \end{bmatrix}$ (d) $\begin{bmatrix} 0 & 0 & -2 & 6 \\ 0 & 0 & 0 & 1 \\ -7 & -6 & 8 & 2 \\ 0 & -2 & -4 & 1 \end{bmatrix}$

(e) $\begin{bmatrix} 0 & 0 & 8 & 0 \\ -5 & -6 & 6 & -1 \\ 0 & 0 & -5 & 6 \\ 0 & -6 & -4 & 3 \end{bmatrix}$ (f) $\begin{bmatrix} 0 & 0 & 7 & 0 \\ 0 & -3 & 7 & -4 \\ 0 & 7 & -4 & 0 \\ 1 & 2 & -1 & -3 \end{bmatrix}$

(g) $\begin{bmatrix} 0 & 0 & 1 & 8 & 5 \\ 6 & 1 & -5 & -8 & -1 \\ 0 & 0 & 0 & 0 & 5 \\ 0 & -1 & -6 & -5 & 4 \\ 0 & 0 & 0 & -1 & -8 \end{bmatrix}$ (h) $\begin{bmatrix} -6 & 0 & 0 & 0 & 0 \\ -4 & -3 & 0 & 0 & 0 \\ 0 & -4 & 0 & 4 & 0 \\ 0 & 1 & -5 & 12 & 0 \\ -2 & -1 & -5 & -2 & 5 \end{bmatrix}$

Exercise 6.2.10. Given that the determinant

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = 6\,,$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

find the following determinants. Give reasons.

(a) $\begin{vmatrix} 3a & b & c \\ 3d & e & f \\ 3g & h & i \end{vmatrix}$ (b) $\begin{vmatrix} a & b & c/2 \\ -d & -e & -f/2 \\ g & h & i/2 \end{vmatrix}$

(c) $\begin{vmatrix} d & e & f \\ a & b & c \\ g & h & i \end{vmatrix}$ (d) $\begin{vmatrix} a+d & b+e & c+f \\ d & e & f \\ g & h & i \end{vmatrix}$

(e) $\begin{vmatrix} a & b & c-a \\ d & e & f-d \\ g & h & i-g \end{vmatrix}$ (f) $\begin{vmatrix} a & b & c \\ d & e & f \\ a+2g & b+2h & c+2i \end{vmatrix}$

(g) $\begin{vmatrix} d & e & f \\ g+a & h+b & i+c \\ a & b & c \end{vmatrix}$ (h) $\begin{vmatrix} a-3g & b & c \\ d/3-f & e/3 & f/3 \\ g-3i & h & i \end{vmatrix}$

Exercise 6.2.11. Consider a general 3×3 matrix $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$. Derive a first column Laplace expansion (Theorem 6.2.20) of the 3×3 determinant, and rearrange to show it is the same as the determinant formula (6.1).

Exercise 6.2.12. Use the Laplace expansion (Theorem 6.2.20) to find the determinant of the following matrices. Use rows or columns with many zeros. Show your working.

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

(c) $\begin{bmatrix} -4 & 0 & -2 & 2 \\ 0 & 1 & 5 & -2 \\ 4 & 0 & 0 & 4 \\ -2 & 4 & 4 & 5 \end{bmatrix}$ (d) $\begin{bmatrix} 3 & -1 & 5 & 0 \\ 1 & -2 & 0 & 2 \\ -1 & 3 & -6 & 2 \\ 0 & -2 & 0 & -3 \end{bmatrix}$

(e) $\begin{bmatrix} 0 & -1 & 0 & 0 & 6 \\ -4 & 0 & 1 & -1 & 0 \\ 0 & 0 & 3 & 0 & -4 \\ 5 & 0 & -4 & 0 & 0 \\ 3 & -1 & 0 & -3 & -6 \end{bmatrix}$ (f) $\begin{bmatrix} 0 & 3 & -7 & 2 & 0 \\ 0 & -4 & 0 & -6 & 0 \\ 0 & 1 & 3 & 0 & 0 \\ -3 & 0 & -4 & -3 & -3 \\ 0 & -6 & 0 & 0 & 5 \end{bmatrix}$

(g) $\begin{bmatrix} 0 & 0 & -2 & 0 & -6 \\ 0 & -3 & 0 & 1 & 0 \\ -4 & -5 & 2 & 0 & 2 \\ 0 & 2 & 0 & 3 & -3 \\ 2 & 0 & 0 & -1 & 0 \end{bmatrix}$ (h) $\begin{bmatrix} 0 & 0 & -2 & 7 & 4 \\ 0 & 4 & -4 & 1 & 0 \\ 2 & -2 & 0 & 1 & 0 \\ 3 & 2 & 0 & 2 & 0 \\ 0 & -2 & 0 & 0 & 1 \end{bmatrix}$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Exercise 6.2.13. For each of the following matrices, use the Laplace expansion (Theorem 6.2.20) to find all the values of k for which the matrix is *not* invertible. Show your working.

(a) $\begin{bmatrix} 3 & -2k & -1 & -1-2k \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & -2 & -5 & 3+2k \end{bmatrix}$

(b) $\begin{bmatrix} -1 & -2 & 0 & 2k \\ 0 & 0 & 5 & 0 \\ 0 & 2 & -1+k & -k \\ -2+k & 1+2k & 4 & 0 \end{bmatrix}$

(c) $\begin{bmatrix} -1+k & 0 & -2+3k & -1+k \\ -6 & 1+2k & -3 & k \\ 0 & 3k & -4 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$

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

(e) $\begin{bmatrix} 0 & 0 & -4 & 0 & 2+k \\ 0 & 0 & 5 & 0 & -1 \\ 0 & 0 & 0 & 1-2k & -3-2k \\ 1+2k & -1 & 3 & 1-4k & -4 \\ -1+2k & 0 & 0 & -1 & -2 \end{bmatrix}$

(f) $\begin{bmatrix} k & 1 & 0 & -5-k & -1-k \\ -4+6k & -1 & 1+3k & -3-5k & 0 \\ 0 & 2 & 0 & -2 & -5 \\ 0 & 0 & 0 & 0 & 2-k \\ -2k & 1+k & 0 & -2 & 3 \end{bmatrix}$

Exercise 6.2.14. Using Theorem 6.2.23 and the properties of Theorem 6.2.5, detail an argument that the following determinant equation generally forms an equation for the plane passing through the three given points (x_1, y_1, z_1), (x_2, y_2, z_2) and (x_3, y_3, z_3):

$$\det \begin{bmatrix} 1 & x & y & z \\ 1 & x_1 & y_1 & z_1 \\ 1 & x_2 & y_2 & z_2 \\ 1 & x_3 & y_3 & z_3 \end{bmatrix} = 0\,.$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Exercise 6.2.15. Using Theorem 6.2.23 and the properties of Theorem 6.2.5, detail an argument that the following determinant equation generally forms an equation for the parabola passing through the three given points (x_1, y_1), (x_2, y_2) and (x_3, y_3):

$$\det \begin{bmatrix} 1 & x & x^2 & y \\ 1 & x_1 & x_1^2 & y_1 \\ 1 & x_2 & x_2^2 & y_2 \\ 1 & x_3 & x_3^2 & y_3 \end{bmatrix} = 0\,.$$

Exercise 6.2.16. Using Theorem 6.2.23 and the properties of Theorem 6.2.5, detail an argument that the equation

$$\det \begin{bmatrix} 1 & x & x^2 & y & y^2 & xy \\ 1 & x_1 & x_1^2 & y_1 & y_1^2 & x_1y_1 \\ 1 & x_2 & x_2^2 & y_2 & y_2^2 & x_2y_2 \\ 1 & x_3 & x_3^2 & y_3 & y_3^2 & x_3y_3 \\ 1 & x_4 & x_4^2 & y_4 & y_4^2 & x_4y_4 \\ 1 & x_5 & x_5^2 & y_5 & y_5^2 & x_5y_5 \end{bmatrix} = 0$$

generally forms an equation for the conic section passing through the five given points (x_i, y_i), $i = 1, \dots, 5$.

v0.2a

Answers to selected exercises

6.1.1b : $\det \approx 1.26$

6.1.1d : $\det \approx -1.47$

6.1.1f : $\det \approx -0.9$

6.1.1h : $\det \approx -5.4$

6.1.1j : $\det \approx -1.32$

6.1.1l : $\det \approx -1.11$

6.1.3b : $\det \approx 1.6$

6.1.3d : $\det \approx 1.0$

6.1.3f : $\det \approx 2.3$

6.1.4b : 4

6.1.4d : $-\frac{1}{2}\,,-1\,,-2$

6.1.4f : $0\,,-\frac{1}{6}\,,-4$

6.1.5b : 3

6.1.5d : $-35/54$

6.1.5f : $-131/4$

6.1.5h : -8704

6.1.9b : 25

6.1.9d : Unknowable on the given information.

6.1.9f : $3\cdot 2^n$

6.1.9h : 25

6.2.1b : $-72\,,0\,,36\,,-72$

6.2.1d : $100\,,-0.1\,,-0.01\,,0$

6.2.2b : $-2(x+3y+3)=0$

6.2.2d : $-(x+y+1)=0$

6.2.2f : $2(-3x+2y+25)=0$

6.2.4b : $8(x+y)=0$

6.2.4d : $2(x+2y-11z)=0$

6.2.4f : $5x-2y-2z=0$

6.2.8b : 96

6.2.8d : -28

6.2.8f : 100

v0.2a

6.2.8h : -42

6.2.9b : 12

6.2.9d : -28

6.2.9f : 196

6.2.9h : 1800

6.2.10b : -3

6.2.10d : 6

6.2.10f : 12

6.2.10h : 2

6.2.12b : 140

6.2.12d : -137

6.2.12f : 1080

6.2.12h : 236

6.2.13b : $0\,,\,3/4$

6.2.13d : $0\,,\,-1$

6.2.13f : $-1/3\,,\,0\,,\,-9\,,\,2$

v0.2a

7 Eigenvalues and eigenvectors in general

Chapter Contents

7.1 Find eigenvalues and eigenvectors of matrices 593

7.1.1 A characteristic equation gives eigenvalues . . 593

7.1.2 Repeated eigenvalues are sensitive 608

7.1.3 Application: discrete dynamics of populations 612

7.1.4 Extension: Connect SVDs to eigen-problems 627

7.1.5 Exercises 629

7.2 Linear independent vectors may form a basis 639

7.2.1 Linearly (in)dependent sets 640

7.2.2 Form a basis for subspaces 651

7.2.3 Exercises 666

7.3 Diagonalisation identifies the transformation 672

7.3.1 Solve systems of differential equations 683

7.3.2 Exercises 694

Population modelling Suppose two species of animals interact: how do their populations evolve in time? Let $y(t)$ and $z(t)$ be the number of female animals in each of the species at time t in years (biologists usually just count females in population models as females usually determine reproduction). Modelling might deduce the populations interact according to the rule that the population one year later is $y(t+1) = 2y(t) - 4z(t)$ and $z(t+1) = -y(t) + 2z(t)$: that is, if it was not for the other species, then for each species the number of females would both double every year (since then $y(t+1) = 2y(t)$ and $z(t+1) = 2z(t)$); but the other species decreases each of these growths via the $-4z(t)$ and $-y(t)$ terms.

Question: can we find special solutions in the form $(y, z) = \boldsymbol{x}\lambda^t$ for some constant λ? Substitute $y = x_1\lambda^t$ 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 non-zero). Forming as a matrix-vector equation these last two equations are

$$\begin{bmatrix} 2 & -4 \\ -1 & 2 \end{bmatrix} \boldsymbol{x} = \lambda \boldsymbol{x} \,.$$

That is, this substitution $(y, z) = \boldsymbol{x}\lambda^t$ shows the question about solutions of the population equations reduces to solving $A\boldsymbol{x} = \lambda\boldsymbol{x}$, called an eigen-problem.

The linear algebra of the eigen-problem will empower us to predict that the general solution for the population is, in terms of two constants c_1 and c_2, that one species has female population $y(t) = 2c_1 4^t + 2c_2$ whereas the 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 can we find directions $\boldsymbol{x}$ such that matrix A acting on $\boldsymbol{x}$, resulting in $A\boldsymbol{x}$, is in the same direction as $\boldsymbol{x}$, namely $\lambda\boldsymbol{x}$ for some proportionality constant λ. Now $\boldsymbol{x} = \boldsymbol{0}$ is always a solution of the equation $A\boldsymbol{x} = \lambda\boldsymbol{x}$. Consequently, we are only interested in those values of the eigenvalue λ when non-zero 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 4.1.18 establishes that one way to find the eigenvalues λ 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 4.1.7).

Three general difficulties in eigen-problems Recall that Section 4.1 introduced one way to visually estimate eigenvectors and eigenvalues of a given matrix A (Schonefeld 1995). The 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 $\boldsymbol{x}$ which point in the same (or opposite) direction to $A\boldsymbol{x}$. Let's use this approach to identify three general difficulties.

The MATLAB function `eigshow(A)` provides an interactive alternative to this static view.

1. In this first picture, for matrix $A = \begin{bmatrix} 1 & 1 \\ \frac{1}{8} & 1 \end{bmatrix}$, the eigenvectors appear to be in directions $\boldsymbol{x}_1 \approx \pm(0.9, 0.3)$ 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, §4.1, so we always see ± pairs of eigenvectors in these pictures.) These pairs of eigenvectors are not orthogonal, not at right angles—as happens for symmetric matrices (Theorem 4.2.10). The lack of orthogonality in general means we soon generalise the concept of orthogonal sets of vectors to the new concept of linearly independent sets (Section 7.2).

2. In this second case, for $A = \begin{bmatrix} 0 & 1 \\ -1 & \frac{1}{2} \end{bmatrix}$, there appear to be no eigenvectors at all. No eigenvectors and eigenvalues is the answer if we require real answers. However, in most applications we find it sensible to have complex valued eigenvalues and eigenvectors (Section 7.1). 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 4.1.22). [1]

3. In this third case, for $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, there appears 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 7.3.12). 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 6.2.13). The reason is linked to Theorem 6.2.14 that the determinant of a triangular matrix is simply the product of its diagonal entries.

Example 7.0.1. Find algebraically the eigenvalues and eigenvectors of the triangular matrix $A = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}$.

Solution: Recall Procedure 4.1.18.

(a) Find all eigenvalues by solving the characteristic equation $\det(A-\lambda I) = 0$. Here $\det(A-\lambda I) = \det \begin{bmatrix} 2-\lambda & 1 \\ 0 & 3-\lambda \end{bmatrix}$ which being a triangular matrix has determinant that is the product of the diagonals, namely $\det(A - \lambda I) = (2 - \lambda)(3 - \lambda)$. This determinant is zero only for eigenvalues $\lambda = 2$ or 3. These are the diagonal entries in the triangular matrix.

(b) For each eigenvalue, find corresponding eigenvectors by solving the system $(A - \lambda I)\boldsymbol{x} = \boldsymbol{0}$.

[1] 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 complex valued eigenvalues and eigenvectors.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

- For $\lambda = 2$, the system is $\begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \boldsymbol{x} = \boldsymbol{0}$ which requires $x_2 = 0$. That is, all eigenvectors are $x_1(1,0)$.
- For $\lambda = 3$, the system is $\begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix} \boldsymbol{x} = \boldsymbol{0}$ which requires $x_1 = x_2$. That is, all eigenvectors are $x_2(1,1)$.

The eigenvectors corresponding to the different eigenvalues are not orthogonal as their dot product $(1,0)\cdot(1,1) = 1+0 = 1 \neq 0$: instead they are at 45° to each other.

■

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.*

Proof. We detail only the case of upper triangular matrices as the argument is similar for lower triangular matrices. First establish that the diagonal entries are eigenvalues, and second prove there are no others. Let λ be any value in the diagonal of the matrix A, and let k be the smallest index such that $a_{kk} = \lambda$ (this 'smallest' caters for duplicated diagonal values). Let's construct an eigenvector in the form $\boldsymbol{x} = (x_1, x_2, \dots, x_{k-1}, 1, 0, \dots, 0)$. Set $x_k = 1$ and $x_j = 0$ for $j > k$. Then set

$$\begin{aligned}
x_{k-1} &= -a_{k-1,k}x_k/(a_{k-1,k-1} - \lambda)\,, \\
x_{k-2} &= -(a_{k-2,k}x_k + a_{k-2,k-1}x_{k-1})/(a_{k-2,k-2} - \lambda)\,, \\
&\vdots \\
x_1 &= -(a_{1,k}x_k + a_{1,k-1}x_{k-1} + \cdots + a_{1,2}x_2)/(a_{1,1} - \lambda).
\end{aligned}$$

Since k is the smallest index for which $\lambda = a_{k,k}$ none of the above expressions involve divisions by zero, and so all are well defined. Rearranging the above equations shows that this vector $\boldsymbol{x}$ satisfies, for $\lambda = a_{k,k}$,

$$\begin{aligned}
(a_{1,1} - \lambda)x_1 + a_{1,2}x_2 + \cdots + a_{1,k-1}x_{k-1} + a_{1,k}x_k &= 0\,, \\
(a_{2,2} - \lambda)x_2 + \cdots + a_{2,k-1}x_{k-1} + a_{1,k}x_k &= 0\,, \\
\ddots \qquad\qquad \vdots \\
(a_{k-1,k-1} - \lambda)x_{k-1} + a_{k-1,k}x_k &= 0\,, \\
(a_{k,k} - \lambda)x_k &= 0\,;
\end{aligned}$$

that is, $(A - \lambda I)\boldsymbol{x} = \boldsymbol{0}$. Rearranging gives $A\boldsymbol{x} = \lambda\boldsymbol{x}$ for non-zero eigenvector $\boldsymbol{x}$ and corresponding eigenvalue $\lambda = a_{kk}$.

Second, there can be no other eigenvalues. Every eigenvalue has to have non-trivial solutions, eigenvectors $\boldsymbol{x}$, to $(A - \lambda I)\boldsymbol{x} = \boldsymbol{0}$, which by Theorem 6.1.23 requires $\det(A - \lambda I) = 0$. But matrix $(A - \lambda I)$

is upper triangular, as A is upper triangular, so Theorem 6.2.14 asserts the determinant

$$\det(A-\lambda I) = (a_{1,1}-\lambda)(a_{2,2}-\lambda)\cdots(a_{n,n}-\lambda) = 0$$

iff the eigenvalue λ is one of the diagonal elements of A.

As an example of the non-orthogonality of eigenvectors, consider the two eigenvalues $\lambda_1 = a_{1,1}$ and $\lambda_2 = a_{2,2}$ with corresponding eigenvectors $\boldsymbol{x}_1 = (1\,,0\,,0\,,\dots,0)$ and $\boldsymbol{x}_2 = (-a_{1,2}/(a_{1,1}-a_{2,2})\,,1\,,0\,,\dots,0)$. Then the dot product $\boldsymbol{x}_1\cdot\boldsymbol{x}_2 = -a_{1,2}/(a_{1,1}-a_{2,2}) \neq 0$ in general (the dot product is zero only when $a_{1,2}=0$). Since the dot product is generally non-zero, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are generally not orthogonal. Similarly for other pairs of eigenvectors corresponding to distinct eigenvalues. □

Example 7.0.3. Use Theorem 7.0.2 to find the eigenvalues, corresponding eigenvectors, and corresponding eigenspaces, of the following triangular matrices.

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

Solution: Matrix A is upper triangular so read off the eigenvalues from the diagonal to be -3 and ± 4.

- For $\lambda = -3$, and by inspection, all eigenvectors are proportional to $(1,0,0)$. Hence eigenspace $\mathbb{E}_{-3} = \operatorname{span}\{(1,0,0)\}$.
- For $\lambda = -4$ we need to solve

$$(A+4I)\boldsymbol{x} = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 8 \end{bmatrix}\boldsymbol{x} = \boldsymbol{0}\,.$$

  By inspection an eigenvector must be of the form $(x_1,1,0)$. And the first line of the system then asserts $x_1+2=0$. Hence eigenvectors are proportional to $(-2,1,0)$. That is, the eigenspace $\mathbb{E}_{-4} = \operatorname{span}\{(-2,1,0)\}$.
- For $\lambda = +4$ we need to solve

$$(A-4I)\boldsymbol{x} = \begin{bmatrix} -7 & 2 & 0 \\ 0 & -8 & 2 \\ 0 & 0 & 0 \end{bmatrix}\boldsymbol{x} = \boldsymbol{0}\,.$$

  Consider eigenvectors of the form $(x_1,x_2,1)$. The second line asserts $-8x_2+2=0$, that is $x_2=\frac{1}{4}$. The first line asserts $-7x_1+2x_2=0$, that is $x_1=\frac{2}{7}x_2=\frac{1}{14}$. Hence eigenvectors are proportional to $(\frac{1}{14},\frac{1}{4},1)$. That is, the eigenspace $\mathbb{E}_4 = \operatorname{span}\{(\frac{1}{14},\frac{1}{4},1)\}$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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

Solution: Matrix B is lower triangular so read the eigenvalues from the diagonal to be 3, -4, 0 and 1.

- For $\lambda = 1$, by inspection all eigenvectors are of the form $(0,0,0,1)$. Hence eigenspace $\mathbb{E}_1 = \operatorname{span}\{(0,0,0,1)\}$.
- For $\lambda = 0$, seek an eigenvector $(0,0,1,x_4)$ then the last line of the system

$$(B - 0I)\boldsymbol{x} = \begin{bmatrix} 3 & 0 & 0 & 0 \\ -2 & -4 & 0 & 0 \\ -3 & 1 & 0 & 0 \\ 0 & 0 & -3 & 1 \end{bmatrix} \boldsymbol{x} = \mathbf{0}$$

requires $-3 + x_4 = 0$. Hence eigenvectors are proportional to $(0,0,1,3)$. That is, the eigenspace $\mathbb{E}_0 = \operatorname{span}\{(0,0,1,3)\}$.

- For $\lambda = -4$, seek an eigenvector $(0,1,x_3,x_4)$ then the third line of the system

$$(B + 4I)\boldsymbol{x} = \begin{bmatrix} 7 & 0 & 0 & 0 \\ -2 & 0 & 0 & 0 \\ -3 & 1 & 4 & 0 \\ 0 & 0 & -3 & 5 \end{bmatrix} \boldsymbol{x} = \mathbf{0}$$

requires $1 + 4x_3 = 0$, that is $x_3 = -\frac{1}{4}$. Then the last line of the system requires $\frac{3}{4} + 5x_4 = 0$, that is $x_4 = -\frac{3}{20}$. Hence eigenvectors are proportional to $(0,1,-\frac{1}{4},-\frac{3}{20})$. That is, the eigenspace $\mathbb{E}_{-4} = \operatorname{span}\{(0,1,-\frac{1}{4},-\frac{3}{20})\}$.

- For $\lambda = 3$, seek an eigenvector $(1,x_2,x_3,x_4)$ then the second line of the system

$$(B - 3I)\boldsymbol{x} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ -2 & -7 & 0 & 0 \\ -3 & 1 & -3 & 0 \\ 0 & 0 & -3 & -2 \end{bmatrix} \boldsymbol{x} = \mathbf{0}$$

requires $-2 - 7x_2 = 0$, that is $x_2 = -\frac{2}{7}$. Then the third line of the system requires $-3 - \frac{2}{7} - 3x_3 = 0$, that is $x_3 = -\frac{23}{21}$. Lastly, the last line requires $\frac{23}{7} - 2x_4 = 0$, that is $x_4 = \frac{23}{14}$. Hence eigenvectors are proportional to $(1,-\frac{2}{7},-\frac{23}{21},\frac{23}{14})$. That is, the eigenspace $\mathbb{E}_3 = \operatorname{span}\{(1,-\frac{2}{7},-\frac{23}{21},\frac{23}{14})\}$.

(c) $C = \begin{bmatrix} -1 & 1 & -8 & -5 & 5 \\ -3 & 6 & 4 & -3 & 0 \\ 1 & -3 & 1 & 0 & 0 \\ -7 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \end{bmatrix}$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Solution: Matrix C is not a triangular matrix (Definition 6.2.13), so Theorem 7.0.2 does not apply. Row or column swaps could transform it to be triangular, but we have not investigated the effect of such swaps on eigenvalues and eigenvectors.

■

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 $n \times n$ diagonal matrix. Correspondingly, the next section establishes there are precisely n eigenvalues of general $n \times n$ matrices, provided we count the eigenvalues according to multiplicity and allow complex valued eigenvalues.

Exercise 7.0.1. Each of the following pictures applies to some specific real matrix, say called A. The pictures plot $A\boldsymbol{x}$ adjoined to the end of unit vectors $\boldsymbol{x}$. By inspection decide whether the matrix, in each case, has real eigenvalues or complex eigenvalues.

(a)

(b)

(c)

(d)

(e)

(f)

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(g)

(h)

Exercise 7.0.2. For each of the following triangular matrices, write down all eigenvalues and then find the corresponding eigenspaces. Show your working.

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

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

(c) $\begin{bmatrix} -1 & 0 & 3 \\ 0 & -1 & 2 \\ 0 & 0 & -5 \end{bmatrix}$

(d) $\begin{bmatrix} 0 & 0 & 0 \\ -3 & -4 & 0 \\ 1 & 5 & -2 \end{bmatrix}$

(e) $\begin{bmatrix} 1 & -5 & 0 \\ 0 & 0 & -4 \\ 0 & 0 & 0 \end{bmatrix}$

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

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

(h) $\begin{bmatrix} -2 & 4 & -2 & -2 \\ 0 & -2 & 1 & 7 \\ 0 & 0 & -3 & 1 \\ 0 & 0 & 0 & 2 \end{bmatrix}$

(i) $\begin{bmatrix} 8 & -2 & 3 & 2 \\ 0 & -6 & 1 & -2 \\ 0 & 0 & 3 & -2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

(j) $\begin{bmatrix} 0 & -2 & -5 & 2 \\ 0 & 7 & -1 & 2 \\ 0 & 0 & 3 & -4 \\ 0 & 0 & 0 & 3 \end{bmatrix}$

v0.2a

7.1 Find eigenvalues and eigenvectors of matrices

Section Contents

7.1.1 A characteristic equation gives eigenvalues . . 593
7.1.2 Repeated eigenvalues are sensitive 608
7.1.3 Application: discrete dynamics of populations 612
7.1.4 Extension: Connect SVDs to eigen-problems 627
7.1.5 Exercises 629

Given the additional determinant methods of Chapter 6, this section begins exploring the properties and some applications of the eigen-problem $A\boldsymbol{x} = \lambda\boldsymbol{x}$ for general matrices A. We establish that there are generally n eigenvalues of an $n \times n$ matrix, albeit possibly complex valued, and that repeated eigenvalues are sensitive to errors. Applications include population modelling, connecting to the computation of SVDs, and fitting exponentials to real data.

7.1.1 A characteristic equation gives eigenvalues

> The Fundamental Theorem of Algebra asserts that every polynomial equation over the complex field has a root. It is almost beneath the dignity of such a majestic theorem to mention that in fact it has precisely n roots.
> *J. H. Wilkinson, 1984 (Higham 1996, p.103)*

Recall that eigenvalues λ and non-zero eigenvectors $\boldsymbol{x}$ of a square matrix A must satisfy $(A-\lambda I)\boldsymbol{x} = \boldsymbol{0}$. Theorem 6.1.23 then implies the eigenvalues of a square matrix are precisely the solutions of the **characteristic equation** $\det(A-\lambda I) = 0$.

Theorem 7.1.1. *We call* $\det(A-\lambda I)$ *the* ***characteristic polynomial*** *of square matrix* A. [2] *The characteristic polynomial of an* $n \times n$ *matrix* A *is a polynomial of* n*th degree in* λ. *Consequently, there are at most* n *distinct eigenvalues of an* $n \times n$ *matrix* A.

Proof. Use induction on the size of the matrix. First, for 1×1 matrix $A = \begin{bmatrix} a_{11} \end{bmatrix}$, the determinant $\det(A-\lambda I) = \det\begin{bmatrix} a_{11}-\lambda \end{bmatrix} = a_{11}-\lambda$ which is of degree one in λ. Second, assume that for all $(n-1)\times(n-1)$ matrices A, $\det(A-\lambda I)$ is a polynomial of

[2] Alternatively, many call $\det(\lambda I - A)$ the characteristic polynomial, as does MATLAB/Octave. The distinction is immaterial as, for an $n \times n$ matrix A and by Theorem 6.1.6c with multiplicative factor $k = -1$, the only difference in the determinant is a factor of $(-1)^n$. In MATLAB/Octave, `poly(A)` computes the characteristic polynomial of the matrix, $\det(\lambda I - A)$, which might be useful for exercises, but is rarely useful in practice due to poor conditioning.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

degree $n-1$ in λ. Then use the Laplace Expansion Theorem 6.2.20 to give the first row expansion, in terms of minors of A and I,

$$\begin{aligned}\det(A-\lambda I) &= (a_{11}-\lambda)\det(A_{11}-\lambda I_{11})\\ &\quad - a_{12}\det(A_{12}-\lambda I_{12})\\ &\quad + \cdots - (-1)^n a_{1n}\det(A_{1n}-\lambda I_{1n}).\end{aligned}$$

Now the minor I_{11} is precisely the $(n-1)\times(n-1)$ identity, and hence by assumption $\det(A_{11}-\lambda I_{11})$ is a polynomial of degree $(n-1)$ in λ. But differently, the minors $I_{12}, \dots, I_{1n}$ have two of the ones removed from the $n\times n$ identity and hence $\lambda I_{12}, \dots, \lambda I_{1n}$ each have only $(n-2)$ λs: since each term in a determinant is a product of distinct entries of the matrix (Theorem 6.2.23) it follows that for $j\geq 2$ the determinant $\det(A_{1j}-\lambda I_{1j})$ is a polynomial in λ of degree $\leq n-2$. Consequently, the first row expansion of

$$\begin{aligned}\det(A-\lambda I) &= (a_{11}-\lambda)(\text{poly degree } n-1)\\ &\quad - a_{12}(\text{poly degree } \leq n-2)\\ &\quad + \cdots - (-1)^n a_{1n}(\text{poly degree } \leq n-2)\\ &= (a_{11}-\lambda)(\text{poly degree } n-1)\\ &\quad + (\text{poly degree } \leq n-2)\\ &= (\text{poly degree } n)\end{aligned}$$

as the highest power of λ cannot be cancelled by any term. Induction thus implies the characteristic polynomial of an $n\times n$ matrix A is a polynomial of nth degree in λ.

Lastly, because the characteristic polynomial of A is of nth degree in λ, the Fundamental Theorem of Algebra asserts that the polynomial has at most n roots (possibly complex). Hence there are at most n distinct eigenvalues. □

Example 7.1.2. 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) $A = \begin{bmatrix} 1 & -1 \\ -2 & 4 \end{bmatrix}$

Solution: The characteristic polynomial is $\det(A-\lambda I) = \det\begin{bmatrix} 1-\lambda & -1 \\ -2 & 4-\lambda \end{bmatrix} = (1-\lambda)(4-\lambda)-2 = \lambda^2-5\lambda+2$. Now $\det A = 4-2 = 2$ which is the coefficient of the constant term in this polynomial. Whereas the sum of the diagonal of A is $1+4=5$ which is the negative of the λ coefficient in the polynomial.

(b) $B = \begin{bmatrix} 4 & -2 & 1 \\ 1 & -2 & 0 \\ 8 & 2 & 6 \end{bmatrix}$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Solution: The characteristic polynomial is

$$\begin{aligned}\det(B-\lambda I) &= \det\begin{bmatrix}4-\lambda & -2 & 1\\ 1 & -2-\lambda & 0\\ 8 & 2 & 6-\lambda\end{bmatrix}\\ &= (4-\lambda)(-2-\lambda)(6-\lambda)+0+2\\ &\quad -(-2-\lambda)8-0-(-2)(6-\lambda)\\ &= -48+4\lambda+8\lambda^2-\lambda^3+2\\ &\quad +16+8\lambda+12+2\lambda\\ &= -\lambda^3+8\lambda^2+2\lambda-18\,.\end{aligned}$$

Now $\det B = 4(-2)6+0+2-(-2)8-0-(-2)6 = -48+2+16+12 = -18$ which is the coefficient of the constant term in the polynomial. Whereas the sum of the diagonal of B is $4-2+6=8$ which is the λ^2 coefficient in the polynomial.

■

These observations about the coefficients in the characteristic polynomials leads to the next theorem.

This optional theorem helps establish the nature of a characteristic polynomial.

Theorem 7.1.3. *For an $n\times n$ matrix A, the product of the eigenvalues equals $\det A$ and equals the constant term in the characteristic polynomial. The sum of the eigenvalues equals $(-1)^{n-1}$ times the coefficient of λ^{n-1} in the characteristic polynomial and equals the* ***trace*** *of the matrix, $a_{11}+a_{22}+\cdots+a_{nn}$.*

Proof. Theorem 7.1.1, and its proof, establishes that the characteristic polynomial has the form

$$\begin{aligned}\det(A-\lambda I) &= c_0+c_1\lambda+\cdots+c_{n-1}\lambda^{n-1}+(-1)^n\lambda^n\\ &= (\lambda_1-\lambda)(\lambda_2-\lambda)\cdots(\lambda_n-\lambda),\end{aligned}\tag{7.1}$$

where the second equality follows from the Fundamental Theorem of Algebra that an nth degree polynomial factors into n linear factors (albeit possibly complex). First, substitute $\lambda=0$ and this equation (7.1) gives $\det A = c_0 = \lambda_1\lambda_2\cdots\lambda_n$, as required.

Second, consider the term $c_{n-1}\lambda^{n-1}$ in equation (7.1). From the factorisation on the right-hand side, the λ^{n-1} term arises as

$$\begin{aligned}c_{n-1}\lambda^{n-1} &= \lambda_1(-\lambda)^{n-1}+(-\lambda)\lambda_2(-\lambda)^{n-2}+\cdots+(-\lambda)^{n-1}\lambda_n\\ &= (-1)^{n-1}(\lambda_1+\lambda_2+\cdots+\lambda_n)\lambda^{n-1},\end{aligned}$$

and hence the coefficient $c_{n-1}=(-1)^{n-1}(\lambda_1+\lambda_2+\cdots+\lambda_n)$, as required. Now use induction to prove the trace formula. Recall that the proof of Theorem 7.1.1 establishes that

$$\det(A-\lambda I) = (a_{11}-\lambda)\det(A_{11}-\lambda I_{11})$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$+ (\text{poly degree } \le n-2).$$

Assume the trace formula holds for $(n-1)\times(n-1)$ matrices such as the minor A_{11}. Then the previous identity gives

$$\begin{aligned}\det(A-\lambda I) &= (a_{11}-\lambda)\big[(-1)^{n-1}\lambda^{n-1} \\ &\qquad + (-1)^{n-2}(a_{22}+\cdots+a_{nn})\lambda^{n-2} \\ &\qquad + (\text{poly degree } \le n-3)\big] \\ &\quad + (\text{poly degree } \le n-2) \\ &= (-1)^n\lambda^n + (-1)^{n-1}(a_{11}+a_{22}+\cdots+a_{nn})\lambda^{n-1} \\ &\quad + (\text{poly degree } \le n-2).\end{aligned}$$

Hence the coefficient $c_{n-1} = (-1)^{n-1}(a_{11}+a_{22}+\cdots+a_{nn})$. Since the formula holds for the basic case $n=1$, namely $c_0 = +a_{11}$, induction implies the sum of the eigenvalues $\lambda_1+\lambda_2+\cdots+\lambda_n = a_{11}+a_{22}+\cdots+a_{nn}$, the trace of the matrix A. □

Example 7.1.4. (a) What are the two highest order terms and the constant term in the characteristic polynomial of the matrix

$$A = \begin{bmatrix} -2 & -1 & 3 & -2 \\ -1 & 3 & -2 & 2 \\ 2 & -3 & 0 & 1 \\ 0 & 1 & 0 & -3 \end{bmatrix}.$$

Solution: First compute the determinant using the Laplace expansion (Theorem 6.2.20). The two zeros in the last row suggest a last row expansion:

$$\begin{aligned}\det A &= (-1)^6 1 \det\begin{bmatrix} -2 & 3 & -2 \\ -1 & -2 & 2 \\ 2 & 0 & 1 \end{bmatrix} \\ &\quad + (-1)^8(-3)\det\begin{bmatrix} -2 & -1 & 3 \\ -1 & 3 & -2 \\ 2 & -3 & 0 \end{bmatrix} \\ &= (4+12+0-8-0+3) \\ &\quad - 3(0+4+9-18+12-0) = -10\,.\end{aligned}$$

This is the constant term in the characteristic polynomial. Second, the trace of A is $-2+3+0-3=-2$ so the cubic coefficient in the characteristic polynomial is $(-1)^3(-2)=2$. That is, the characteristic polynomial of A is of the form $\lambda^4+2\lambda^3+\cdots-10$.

(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?

Solution: No, because the trace of B is $-2-5-2+4+1=-4$ so the coefficient of the λ^4 term must be $(-1)^4(-4)=-4$ instead of the calculated 2.

(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?

Solution: No. By the column of zeros in C, $\det C$ must be zero instead of the calculated 668.

(d) What are the two highest order terms and the constant term in the characteristic polynomial of the matrix

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

Solution: First compute the determinant using the Laplace expansion (Theorem 6.2.20). The nearly all zeros in the last row suggests starting with a last row expansion (although others are just as good):

$$\det D=(-1)^8(-3)\det\begin{bmatrix}0&0&0&3&0\\-2&0&1&0&-2\\0&0&-1&0&0\\0&-5&0&-4&3\\0&-3&0&-4&0\end{bmatrix}$$

(using a 3rd row expansion)

$$=-3(-1)^6(-1)\det\begin{bmatrix}0&0&3&0\\-2&0&0&-2\\0&-5&-4&3\\0&-3&-4&0\end{bmatrix}$$

(using a 1st column expansion)

$$=3(-1)^3(-2)\det\begin{bmatrix}0&3&0\\-5&-4&3\\-3&-4&0\end{bmatrix}$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$
\begin{aligned}
&\text{(using a 3rd column expansion)}\\
&= 6(-1)^5 3 \det\begin{bmatrix} 0 & 3 \\ -3 & -4 \end{bmatrix}\\
&= -18(0+9) = -162\,.
\end{aligned}
$$

This is the constant term in the characteristic polynomial. Second, the trace of D is $0+0+0+0-4+0 = -4$ so the quintic coefficient in the characteristic polynomial is $(-1)^5(-4) = 4$. That is, the characteristic polynomial of D is of the form $\lambda^6 + 4\lambda^5 + \cdots - 162$.

■

Definition 7.1.5. *An eigenvalue λ_0 of a matrix A is said to have* ***multiplicity*** *m if the characteristic polynomial factorises to $\det(A-\lambda I) = (\lambda-\lambda_0)^m g(\lambda)$ where $g(\lambda_0) \neq 0$, and $g(\lambda)$ is a polynomial of degree $n-m$. Any eigenvalue of multiplicity $m \geq 2$ is also called a* ***repeated eigenvalue****.*

Example 7.1.6. Use the characteristic polynomial to find all eigenvalues and their multiplicity of the following matrices.

(a) $A = \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$

Solution: The characteristic equation is $\det(A-\lambda I) = \det\begin{bmatrix} 3-\lambda & 1 \\ 0 & 3-\lambda \end{bmatrix} = (3-\lambda)^2 - 0\cdot 1 = (\lambda-3)^2 = 0$. The only eigenvalue is $\lambda = 3$ with multiplicity two. (Since this is an upper triangular matrix, the eigenvalues are the diagonal elements, namely 3 twice.)

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

Solution: The characteristic equation is

$$
\begin{aligned}
\det(B-\lambda I) &= \det\begin{bmatrix} -1-\lambda & 1 & -2 \\ -1 & -\lambda & -1 \\ 0 & -3 & 1-\lambda \end{bmatrix}\\
&= (1+\lambda)\lambda(1-\lambda) + 0 - 6\\
&\quad - 0 + 3(1+\lambda) + (1-\lambda)\\
&= -\lambda^3 + 3\lambda - 2\\
&= -(\lambda-1)^2(\lambda+2) = 0\,.
\end{aligned}
$$

Eigenvalues are $\lambda = 1$ with multiplicity two, and $\lambda = -2$ with multiplicity one.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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

Solution: The characteristic equation is

$$\begin{aligned}
\det(C - \lambda I) &= \det \begin{bmatrix} -1-\lambda & 0 & -2 \\ 0 & -3-\lambda & 2 \\ 0 & -2 & 1-\lambda \end{bmatrix} \\
&= (-1-\lambda) \det \begin{bmatrix} -3-\lambda & 2 \\ -2 & 1-\lambda \end{bmatrix} \\
&= -(1+\lambda)\big[(-3-\lambda)(1-\lambda)+4\big] \\
&= -(\lambda+1)\big[\lambda^2+2\lambda+1\big] \\
&= -(\lambda+1)^3 = 0\,.
\end{aligned}$$

The only eigenvalue is $\lambda = -1$ with multiplicity three.

(d) $D = \begin{bmatrix} 2 & 0 & -1 \\ -5 & 3 & -5 \\ 5 & -2 & -2 \end{bmatrix}$

Solution: The characteristic equation is

$$\begin{aligned}
\det(D - \lambda I) &= \det \begin{bmatrix} 2-\lambda & 0 & -1 \\ -5 & 3-\lambda & -5 \\ 5 & -2 & -2-\lambda \end{bmatrix} \\
&= (2-\lambda)(3-\lambda)(-2-\lambda) + 0 - 10 \\
&\quad + 5(3-\lambda) - 10(2-\lambda) - 0 \\
&= -\lambda^3 + 3\lambda^2 + 9\lambda - 27 \\
&= -(\lambda-3)^2(\lambda+3) = 0\,.
\end{aligned}$$

Eigenvalues are $\lambda = 3$ with multiplicity two, and $\lambda = -3$ with multiplicity one.

(e) $E = \begin{bmatrix} 0 & 1 \\ -1 & 1 \end{bmatrix}$

Solution: The characteristic equation is

$$\begin{aligned}
\det(E - \lambda I) &= \det \begin{bmatrix} -\lambda & 1 \\ -1 & 1-\lambda \end{bmatrix} \\
&= -\lambda(1-\lambda) + 1 \\
&= \lambda^2 - \lambda + 1 = 0\,.
\end{aligned}$$

This quadratic equation does not factor easily so use the formula

$$\lambda = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{1 \pm \sqrt{1-4}}{2} = \frac{1}{2} \pm \frac{\sqrt{3}}{2} i\,,$$

for $i = \sqrt{-1}$, are two eigenvalues (complex valued) each of multiplicity one.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

■

Example 7.1.7. 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) $\begin{bmatrix} 2 & 2 & -1 \\ 0 & 1 & -2 \\ 0 & -1 & 0 \end{bmatrix}$

Solution: Execute `eig([2 2 -1;0 1 -2;0 -1 0])` to get

```
ans =
   2
   2
  -1
```

So the eigenvalue $\lambda = 2$ has multiplicity two, and the eigenvalue $\lambda = -1$ has multiplicity one.

(b) $\begin{bmatrix} -2 & -2 & -5 & 0 \\ 0 & -2 & 2 & 1 \\ -1 & 1 & 0 & -1 \\ -2 & 1 & 4 & 0 \end{bmatrix}$

Solution: In MATLAB/Octave execute

```
eig([-2 -2 -5 0
  0 -2 2 1
 -1 1 0 -1
 -2 1 4 0])
```

to get

```
ans =
  -3.0000 + 0.0000i
  -3.0000 + 0.0000i
   1.0000 + 1.4142i
   1.0000 - 1.4142i
```

There are two complex-valued eigenvalues, evidently $1 \pm \sqrt{2}i$, each of multiplicity one, and also the (real) eigenvalue $\lambda = -3$ which has multiplicity two.

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

Solution: In MATLAB/Octave execute

v0.2a

```
eig([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])
```

to get

```
ans =
   2.0000 + 2.8284i
   2.0000 - 2.8284i
   4.0000 + 0.0000i
  -0.0000 + 0.0000i
  -0.0000 - 0.0000i
```

There are three eigenvalues of multiplicity one, namely 4 and $2 \pm \sqrt{8}i$. The last two rows appear to be the eigenvalue $\lambda = 0$ with multiplicity two.

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

Solution: In MATLAB/Octave execute

```
eig([-1 0 0 0
 -1 2 -3 3
  3 1 -1 0
  0 3 -2 1])
```

to get

```
ans =
   4.0000 + 0.0000i
  -1.0000 + 0.0000i
  -1.0000 - 0.0000i
  -1.0000 + 0.0000i
```

There is one eigenvalue of multiplicity one, $\lambda = 4$. The last three rows appear to be the eigenvalue $\lambda = -1$ with multiplicity three.

■

To find eigenvalues and eigenvectors, the following restates Procedure 4.1.18 with a little more information, and now empowered to address larger matrices with the determinant tools from Chapter 6.

v0.2a

Procedure 7.1.8 (eigenvalues and eigenvectors). *To find by hand eigenvalues and eigenvectors of a (small) square matrix A:*

1. *find all eigenvalues (possibly complex) by solving the* ***characteristic equation*** *of A, $\det(A-\lambda I)=0$;*
2. *for each eigenvalue λ, solve $(A-\lambda I)\boldsymbol{x}=\mathbf{0}$ 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 7.2.15 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 complex eigenvalues).

Correspondingly, the following restates the computational procedure of section 4.1.1, but slightly more generally: the extra generality caters for non-symmetric matrices.

Compute in Matlab/Octave. For a given square matrix `A`, execute `[V,D]=eig(A)`, then the diagonal entries of `D`, `diag(D)`, are the eigenvalues of `A`. Corresponding to the eigenvalue `D(j,j)` is an eigenvector $\boldsymbol{v}_j=$ `V(:,j)`, the jth column of `V`. [3] If an eigenvalue is repeated (multiplicity more than one), then the corresponding columns of `V` span the eigenspace (and, as Section 7.2 discusses, the column vectors are a so-called basis for the eigenspace when they have a property called linear independence).

Example 7.1.9. Find the eigenspaces corresponding to the eigenvalues found for the first three matrices of Example 7.1.6.

7.1.6a. $A=\begin{bmatrix}3&1\\0&3\end{bmatrix}$

Solution: The only eigenvalue is $\lambda=3$ with multiplicity two. Its eigenvectors $\boldsymbol{x}$ satisfy

$$(A-3I)\boldsymbol{x}=\begin{bmatrix}0&1\\0&0\end{bmatrix}\boldsymbol{x}=\mathbf{0}\,.$$

The second component of this equation is the trivial $0=0$. The first component of the equation is $0x_1+1x_2=0$, hence $x_2=0$. All eigenvectors are of the form $\boldsymbol{x}=(1,0)x_1$. That is the eigenspace $\mathbb{E}_3=\operatorname{span}\{(1,0)\}$.

In MATLAB/Octave, executing `[V,D]=eig([3 1;0 3])` gives us

[3] Be aware that MATLAB/Octave does not use the determinant to find the eigenvalues, nor does it solve the linear equations to find eigenvectors. For any but the smallest matrices such a 'by hand' approach is far too expensive. Instead, to find eigenvalues and eigenvectors, just as for the SVD, MATLAB/Octave uses an intriguing iteration based upon what is called the QR factorisation.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

```
V =
   1.0000  -1.0000
   0.0000   0.0000
D =
   3   0
   0   3
```

Diagonal matrix `D` confirms the only eigenvalue is three, whereas the two columns of `V` confirm the eigenspace $\mathbb{E}_1 = \operatorname{span}\{(1,0),(-1,0)\} = \operatorname{span}\{(1,0)\}$.

7.1.6b. $B = \begin{bmatrix} -1 & 1 & -2 \\ -1 & 0 & -1 \\ 0 & -3 & 1 \end{bmatrix}$

Solution: The eigenvalues are $\lambda = -2$ (multiplicity one) and $\lambda = 1$ (multiplicity two).

- For $\lambda = -2$ solve

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

  The third component of this equation requires $-3x_2 + 3x_3 = 0$, that is, $x_2 = x_3$. The second component requires $-x_1 + 2x_2 - x_3 = 0$, that is, $x_1 = 2x_2 - x_3 = 2x_3 - x_3 = x_3$. The first component requires $x_1 + x_2 - 2x_3 = 0$ which is also satisfied by $x_1 = x_2 = x_3$. All eigenvectors are of the form $x_3(1,1,1)$. That is, the eigenspace $\mathbb{E}_{-2} = \operatorname{span}\{(1,1,1)\}$.

- For $\lambda = 1$ solve

$$(B - 1I)\boldsymbol{x} = \begin{bmatrix} -2 & 1 & -2 \\ -1 & -1 & -1 \\ 0 & -3 & 0 \end{bmatrix} \boldsymbol{x} = \boldsymbol{0}\,.$$

  The third component of this equation requires $-3x_2 = 0$, that is, $x_2 = 0$. The second component requires $-x_1 - x_2 - x_3 = 0$, that is, $x_1 = -x_2 - x_3 = 0 - x_3 = -x_3$. The first component requires $-2x_1 + x_2 - 2x_3 = 0$ which is also satisfied by $x_1 = -x_3$ and $x_2 = 0$. All eigenvectors are of the form $x_3(-1,0,1)$. That is, the eigenspace $\mathbb{E}_1 = \operatorname{span}\{(-1,0,1)\}$.

Alternatively, in MATLAB/Octave, executing

```
B=[-1 1 -2
 -1 0 -1
 0 -3 1]
[V,D]=eig(B)
```

gives us

```
V =
    -0.5774     0.7071    -0.7071
    -0.5774     0.0000     0.0000
    -0.5774    -0.7071     0.7071
D =
          -2          0          0
           0          1          0
           0          0          1
```

Diagonal matrix `D` confirms the eigenvalues. The first column of `V` confirms the eigenspace

$$\begin{aligned}\mathbb{E}_{-2} &= \operatorname{span}\{(-0.5774, -0.5774, -0.5774)\}\\ &= \operatorname{span}\{(1,1,1)\}.\end{aligned}$$

Whereas the last two columns of `V` confirm the eigenspace

$$\begin{aligned}\mathbb{E}_{1} &= \operatorname{span}\{(0.7071, 0, -0.7071), (-0.7071, 0, 0.7071)\}\\ &= \operatorname{span}\{(-1,0,1)\}.\end{aligned}$$

7.1.6c. $C = \begin{bmatrix} -1 & 0 & -2 \\ 0 & -3 & 2 \\ 0 & -2 & 1 \end{bmatrix}$

Solution: The only eigenvalue is $\lambda = -1$ with multiplicity three. Its eigenvectors $\boldsymbol{x}$ satisfy

$$(C + 1I)\boldsymbol{x} = \begin{bmatrix} 0 & 0 & -2 \\ 0 & -2 & 2 \\ 0 & -2 & 2 \end{bmatrix} \boldsymbol{x} = \boldsymbol{0}\,.$$

The first component of this equation requires $x_3 = 0$. The second and third components both requires $-2x_2 + 2x_3 = 0$, hence $x_2 = x_3 = 0$. Since x_1 is unconstrained, all eigenvectors are of the form $\boldsymbol{x} = x_1(1,0,0)$. That is the eigenspace $\mathbb{E}_{-1} = \operatorname{span}\{(1,0,0)\}$.

Alternatively, in MATLAB/Octave, executing

```
C=[-1 0 -2
 0 -3 2
 0 -2 1]
[V,D]=eig(C)
```

gives us

```
V =
           1          -1          -1
           0           0           0
           0           0           0
D =
          -1           0           0
           0          -1           0
           0           0          -1
```

Diagonal matrix `D` confirms the only eigenvalue is one with multiplicity three. The three columns of `V` confirm the eigenspace $\mathbb{E}_{-1} = \operatorname{span}\{(1,0,0)\}$.

■

The matrices in Example 7.1.9 all have repeated eigenvalues. For these repeated eigenvalues the corresponding eigenspaces were 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 4.1.10 and 4.1.15. Subsequent sections work towards Theorem 7.3.12 which establishes that for non-symmetric matrices an eigenspace has dimensionality between one and the multiplicity of the corresponding eigenvalue.

Example 7.1.10. By hand, find the eigenvalues and eigenspaces of the matrix

$$A = \begin{bmatrix} 0 & 3 & 0 & 0 & 0 \\ 1 & 0 & 3 & 0 & 0 \\ 0 & 1 & 0 & 3 & 0 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Confirm your answer with `eig()` in MATLAB/Octave.

Solution: Adopt Procedure 7.1.8.

(a) The characteristic equation is

$$\begin{aligned}
&\det(A - \lambda I) \\
&= \begin{vmatrix} -\lambda & 3 & 0 & 0 & 0 \\ 1 & -\lambda & 3 & 0 & 0 \\ 0 & 1 & -\lambda & 3 & 0 \\ 0 & 0 & 1 & -\lambda & 3 \\ 0 & 0 & 0 & 1 & -\lambda \end{vmatrix} \\
&\quad \text{(by first row expansion (6.4))} \\
&= (-\lambda) \begin{vmatrix} -\lambda & 3 & 0 & 0 \\ 1 & -\lambda & 3 & 0 \\ 0 & 1 & -\lambda & 3 \\ 0 & 0 & 1 & -\lambda \end{vmatrix} - 3 \begin{vmatrix} 1 & 3 & 0 & 0 \\ 0 & -\lambda & 3 & 0 \\ 0 & 1 & -\lambda & 3 \\ 0 & 0 & 1 & -\lambda \end{vmatrix} \\
&\quad \text{(by first row and first column expansion, respectively)} \\
&= (-\lambda)^2 \begin{vmatrix} -\lambda & 3 & 0 \\ 1 & -\lambda & 3 \\ 0 & 1 & -\lambda \end{vmatrix} - (-\lambda)3 \begin{vmatrix} 1 & 3 & 0 \\ 0 & -\lambda & 3 \\ 0 & 1 & -\lambda \end{vmatrix} \\
&\quad - 3 \begin{vmatrix} -\lambda & 3 & 0 \\ 1 & -\lambda & 3 \\ 0 & 1 & -\lambda \end{vmatrix} \\
&\quad \text{(by common factor, and first column expansion)}
\end{aligned}$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$
\begin{aligned}
&= (\lambda^2-3)\begin{vmatrix} -\lambda & 3 & 0 \\ 1 & -\lambda & 3 \\ 0 & 1 & -\lambda \end{vmatrix} + 3\lambda \begin{vmatrix} -\lambda & 3 \\ 1 & -\lambda \end{vmatrix} \\
&\quad \text{(using (6.1))} \\
&= (\lambda^2-3)\big[(-\lambda)^3+0+0-0+3\lambda+3\lambda\big] + 3\lambda(\lambda^2-3) \\
&= (\lambda^2-3)(-\lambda^3+9\lambda) \\
&= -\lambda(\lambda^2-3)(\lambda^2-9) = 0\,.
\end{aligned}
$$

The five eigenvalues are thus $\lambda = 0, \pm\sqrt{3}, \pm 3$, all of multiplicity one.

(b) Consider each eigenvalue in turn.

$\lambda = 0$ Solve

$$
(A-0I)\boldsymbol{x} = \begin{bmatrix} 0 & 3 & 0 & 0 & 0 \\ 1 & 0 & 3 & 0 & 0 \\ 0 & 1 & 0 & 3 & 0 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \boldsymbol{x} = \boldsymbol{0}\,.
$$

The last row requires $x_4 = 0$. The fourth row requires $x_3 + 3x_5 = 0$, that is, $x_3 = -3x_5$. The third row requires $x_2 + 3x_4 = 0$, that is, $x_2 = -3x_4 = -3\cdot 0 = 0$. The second row requires $x_1 + 3x_3 = 0$, that is, $x_1 = -3x_3 = 9x_5$. The first row requires $3x_2 = 0$, which is satisfied as $x_2 = 0$. Since all eigenvectors are of the form $(9x_5, 0, -3x_5, 0, x_5)$, the eigenspace $\mathbb{E}_0 = \operatorname{span}\{(9,0,-3,0,1)\}$.

$\lambda = \pm\sqrt{3}$ Being careful with the upper and lower signs, solve $(A \mp \sqrt{3}I)\boldsymbol{x} = \boldsymbol{0}$, that is,

$$
\begin{bmatrix} \mp\sqrt{3} & 3 & 0 & 0 & 0 \\ 1 & \mp\sqrt{3} & 3 & 0 & 0 \\ 0 & 1 & \mp\sqrt{3} & 3 & 0 \\ 0 & 0 & 1 & \mp\sqrt{3} & 3 \\ 0 & 0 & 0 & 1 & \mp\sqrt{3} \end{bmatrix} \boldsymbol{x} = \boldsymbol{0}\,.
$$

The last row requires $x_4 \mp \sqrt{3}x_5 = 0$, that is, $x_4 = \pm\sqrt{3}x_5$. The fourth row requires $x_3 \mp \sqrt{3}x_4 + 3x_5 = 0$, that is, $x_3 = \pm\sqrt{3}x_4 - 3x_5 = 3x_5 - 3x_5 = 0$. The third row requires $x_2 \mp \sqrt{3}x_3 + 3x_4 = 0$, that is, $x_2 = \pm\sqrt{3}x_3 - 3x_4 = \mp 3\sqrt{3}x_5$. The second row requires $x_1 \mp \sqrt{3}x_2 + 3x_3 = 0$, that is, $x_1 = \pm\sqrt{3}x_2 - 3x_3 = -9x_5$. The first row requires $\mp\sqrt{3}x_1 + 3x_2 = 0$, which is satisfied as $\mp\sqrt{3}(-9x_5) + 3(\mp 3\sqrt{3}x_5) = 0$. Since all eigenvectors are of the form $(-9x_5, \mp 3\sqrt{3}x_5, 0, \pm\sqrt{3}x_5, x_5)$, the eigenspaces $\mathbb{E}_{\pm\sqrt{3}} = \operatorname{span}\{(-9, \mp 3\sqrt{3}, 0, \pm\sqrt{3}, 1)\}$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$\lambda = \pm 3$ Being careful with the upper and lower signs, solve $(A \mp 3I)\boldsymbol{x} = \mathbf{0}$, that is,

$$\begin{bmatrix} \mp 3 & 3 & 0 & 0 & 0 \\ 1 & \mp 3 & 3 & 0 & 0 \\ 0 & 1 & \mp 3 & 3 & 0 \\ 0 & 0 & 1 & \mp 3 & 3 \\ 0 & 0 & 0 & 1 & \mp 3 \end{bmatrix} \boldsymbol{x} = \mathbf{0}\,.$$

The last row requires $x_4 \mp 3x_5 = 0$, that is, $x_4 = \pm 3x_5$. The fourth row requires $x_3 \mp 3x_4 + 3x_5 = 0$, that is, $x_3 = \pm 3x_4 - 3x_5 = 9x_5 - 3x_5 = 6x_5$. The third row requires $x_2 \mp 3x_3 + 3x_4 = 0$, that is, $x_2 = \pm 3x_3 - 3x_4 = \pm 9x_5$. The second row requires $x_1 \mp 3x_2 + 3x_3 = 0$, that is, $x_1 = \pm 3x_2 - 3x_3 = 9x_5$. The first row requires $\mp 3x_1 + 3x_2 = 0$, which is satisfied as $\mp 3(9x_5) + 3(\pm 9x_5) = 0$. Since all eigenvectors are of the form $(9x_5, \pm 9x_5, 6x_5, \pm 3x_5, x_5)$, the eigenspaces $\mathbb{E}_{\pm 3} = \operatorname{span}\{(9, \pm 9, 6, \pm 3, 1)\}$.

■

Example 7.1.11. Use MATLAB/Octave to confirm the eigenvalues and eigenvectors found for the matrix of Example 7.1.10.

Solution: 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 columns of eigenvectors are more easily seen to confirm the hand calculation of Example 7.1.10 when we divide each column by its last element via `V*diag(1./V(5,:))` which gives the more appealing (2 d.p.)

v0.2a

```
ans =
   9.00   9.00   9.00  -9.00  -9.00
   9.00  -9.00   0.00   5.20  -5.20
   6.00   6.00  -3.00   0.00   0.00
   3.00  -3.00   0.00  -1.73   1.73
   1.00   1.00   1.00   1.00   1.00
```

■

7.1.2 Repeated eigenvalues are sensitive

This optional subsection does not prove the sensitivity: it just uses examples to introduce and illustrate.

Albeit hidden in Example 7.1.7, repeated eigenvalues are exquisitely sensitive to errors in either the matrix or the computation. If the matrix or the computation has an error e, then expect a repeated eigenvalue of multiplicity m to appear as m eigenvalues all within about $e^{1/m}$ of each other. Thus when we find or compute m eigenvalues all within about $e^{1/m}$, then suspect it to actually be one eigenvalue of multiplicity m.

For example, 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. Repeat some of the previous Examples 7.1.7, preceded by the MATLAB/Octave command `format long`, to see that the repeated eigenvalues are sensitive to the computational errors.

Similarly, repeated eigenvalues are sensitive to errors in the matrix. The following examples show the sensitivity when we are uncertain about the components of a matrix.

Example 7.1.12. Compute eigenvalues of the following matrices and comment on the effect on repeated eigenvalues of errors in the matrix.

(a) $A = \begin{bmatrix} a & 1 \\ 0.0001 & a \end{bmatrix}$ for any parameter a.

Solution: By hand, the characteristic equation is

$$\det \begin{bmatrix} a-\lambda & 1 \\ 0.0001 & a-\lambda \end{bmatrix} = (a-\lambda)^2 - 0.0001 = 0\,.$$

Rearranging gives $(\lambda - a)^2 = 0.0001$. Taking square roots, $\lambda - a = \pm 0.01$; that is, the two eigenvalues are $\lambda = a \pm 0.01$. If we consider that the entry 0.0001 in the matrix is an error, that the entry should really be zero, then the eigenvalues should really be one repeated eigenvalue $\lambda = a$ of multiplicity two. However, the 'error' 0.0001 splits the repeated eigenvalue into two by an amount $\sqrt{0.0001} = 0.01$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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

Solution: In MATLAB/Octave execute

```
eig([-5 -3 1 0 1
 -2 -2 2 -1 -3
  1 1 -1 -4 -5
 -2 -1 0 -4 1
  0 0 0 2 -3])
```

to get

```
ans =
   2.7961e-08
  -2.7961e-08
  -6.4142
  -5.0000
  -3.5858
```

There are three eigenvalues of multiplicity one, namely -5 and $-5 \pm \sqrt{2}$. The two values $\pm$`2.7961e-08` at first sight appear to be two eigenvalues, $\pm 2.7961 \cdot 10^{-8}$, each of multiplicity one. However, computers usually work to about 15 digits accuracy, that is, the typical error is about 10^{-15}, so an eigenvalue of multiplicity two is generally computed as two eigenvalues separated by about $\sqrt{10^{-15}} \approx 3 \cdot 10^{-8}$. Thus we suspect that the two values $\pm$`2.7961e-08` actually represent one eigenvalue $\lambda = 0$ with multiplicity two.

(c) Suppose the previous matrix B is obtained from some experiment where there are experimental errors in the entries with error about 0.0001. Randomly perturb the entries in matrix B to see the effects of such errors on the eigenvalues (use `randn()`, Table 3.1).

Solution: In MATLAB/Octave execute

```
B=[-5 -3 1 0 1
 -2 -2 2 -1 -3
  1 1 -1 -4 -5
 -2 -1 0 -4 1
  0 0 0 2 -3]
eig(B+0.0001*randn(5))
```

to get something like

```
ans =
   0.0307
  -0.0308
  -6.4143
```

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

```
  -4.9996
  -3.5862
```

Observe the repeated eigenvalue $\lambda = 0$ splits into two eigenvalues, $\lambda = \pm 0.031$, of size roughly $\sqrt{0.0001} = 0.01$. The other eigenvalues are also perturbed by the errors but only by amounts of size roughly 0.0001.

Depending upon the random numbers, another possible answer is

```
ans =
  -0.0001 + 0.0112i
  -0.0001 - 0.0112i
  -6.4143 + 0.0000i
  -4.9999 + 0.0000i
  -3.5856 + 0.0000i
```

where the repeated eigenvalue of zero splits to be a pair of complex valued eigenvalues of roughly $\pm i\sqrt{0.0001} = \pm i0.01$.

(d) $C = \begin{bmatrix} -1 & 0 & 0 & 0 \\ -1 & 2 & -3 & 3 \\ 3 & 1 & -1 & 0 \\ 0 & 3 & -2 & 1 \end{bmatrix}$ perturbed by errors of size 10^{-6}

Solution: In MATLAB/Octave execute

```
C=[-1 0 0 0
 -1 2 -3 3
  3 1 -1 0
  0 3 -2 1]
eig(C+1e-6*randn(4))
```

to get

```
ans =
   4.0000 + 0.0000i
  -1.0156 + 0.0000i
  -0.9922 + 0.0139i
  -0.9922 - 0.0139i
```

The eigenvalue 4 of multiplicity one is not noticeably affected by the errors about 10^{-6}. However, the repeated eigenvalue of $\lambda = -1$ with multiplicity three is split into three close eigenvalues (two complex-valued) all differing by roughly 0.01 which is indeed the cube-root of the perturbation 10^{-6}.

■

But symmetric matrices are OK The eigenvalues of a symmetric matrix are not so sensitive. This is fortunate as many applications give rise to symmetric matrices (Chapter 4). Such

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

symmetry often reflects some symmetry in the natural world such as Newton's law of every action having an equal and opposite reaction. For symmetric matrices, the eigenvalues and eigenvectors are robust to computational perturbations and experimental errors.

Example 7.1.13. Compute the eigenvalues of the symmetric matrix

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

and see matrix A has an eigenvalue of multiplicity two. Explore the effects on the eigenvalues of errors in the matrix by perturbing the entries by random amounts of size 0.0001.

Solution: In MATLAB/Octave execute

```
A=[1 1 0 2
 1 0 2 -1
 0 2 1 4
 2 -1 4 1]
eig(A)
eig(A+0.0001*randn(4))
```

to get

```
ans =
  -4.6235
   1.0000
   1.0000
   5.6235
```

showing the eigenvalue $\lambda = 1$ has multiplicity two. Whereas the eigenvalues of the perturbed matrix depend upon the random numbers and so might be

```
ans =
  -4.6236
   5.6235
   0.9998
   0.9999
```

or perhaps

```
ans =
  -4.6236 + 0.0000i
   5.6234 + 0.0000i
   1.0001 + 0.0000i
   1.0001 - 0.0000i
```

In either case, the perturbation by amounts 0.0001 only change the eigenvalues, whether repeated or not, by an amount of about the same size.

v0.2a

■

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, let's define

- $y_1(t)$ to be the number of juveniles (including infants),
- $y_2(t)$ the number of adolescents, and
- $y_3(t)$ the number of adults.

Mostly, biologists only count females as females are the determining sex for reproduction. (Some bacteria/algae have seven sexes!) How do these numbers of females evolve over time? from generation to generation? First we need to choose a basic time interval (the unit of time): it could be one year, one month, one day, or maybe six months. Whatever we choose as convenient, we then quantify the number of events that happen to the females in each time interval as shown schematically in the diagram below:

Over any one time interval:

- 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:

$$\begin{aligned} y_1(t+1) &= \cdots , \\ y_2(t+1) &= \cdots , \\ y_3(t+1) &= \cdots . \end{aligned}$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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

$$\begin{aligned} y_1(t+1) &= (1-k_1)y_1(t) + \cdots , \\ y_2(t+1) &= +k_1 y_1(t) + (1-k_2)y_2(t), \\ y_3(t+1) &= +k_2 y_2(t) + \cdots . \end{aligned}$$

- A fraction k_3 of the adults die mean 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 $k_0 y_3$ 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.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Example 7.1.14 (orangutans). From the following extract of the Wikipedia entry on orangutans [20 Mar 2014] derive a mathematical model for the age structure of the orangutans from one year to the next.

> Gestation lasts for 9 months, with females giving birth to their first offspring between the ages of 14 and 15 years. Female orangutans have [seven to] eight-year intervals between births, the longest interbirth intervals among the great apes. ... Infant orangutans are completely dependent on their mothers for the first two years of their lives. The mother will carry the infant during travelling, as well as feed it and sleep with it in the same night nest. For the first four months, the infant is carried on its belly and never relieves physical contact. In the following months, the time an infant spends with its mother decreases. When an orangutan reaches the age of two, its climbing skills improve and it will travel through the canopy holding hands with other orangutans, a behaviour known as "buddy travel". Orangutans are juveniles from about two to five years of age and will start to temporarily move away from their mothers. Juveniles are usually weaned at about four years of age. Adolescent orangutans will socialize with their peers while still having contact with their mothers. Typically, orangutans live over 30 years in both the wild and captivity.

Suppose the initial population of orangutans in some area at year zero of a study is that of 30 adolescent females and 15 adult females. Use the mathematical model to predict the population for the next five years.

Solution: *Choose level:* we choose a time unit of one year, and choose to model three age categories.[4]

Define:

- $y_1(t)$ is the number of juvenile females (including infant) at time t (years), say age ≤ 5 years;
- $y_2(t)$ is the number of adolescent females, say $6 \leq$ age ≤ 14 years;
- $y_3(t)$ is the number of adult females, say age ≥ 15 years.

Quantify changes in a year: from Wikipedia information (with numbers slightly modified here to make the results numerically simpler):

- Orangutans are juvenile for say 5 years, so in any one year a

[4] A coarse model considers just the total number in a species; a fine model might count the number of each age in years (here 30 years); a 'micro' model might simulate each and every orangutan as individuals (thousands).

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

fraction 1/5 of them grow to be adolescent and a fraction 4/5 remain as juveniles. Say an adult female gives birth every 7–8 years, so on average it gives birth to a juvenile female every 15 years. Thus a fraction 1/15 of adults give birth to a juvenile female in any year. Consequently, model as $y_1(t+1) = \frac{4}{5}y_1(t) + \frac{1}{15}y_3(t)$.

- Adolescent orangutans become breeding adults after another 9–10 years, so in any one year a fraction 1/10 of them grow to adults and 9/10ths remain adolescents. Consequently, $y_2(t+1) = +\frac{1}{5}y_1(t) + \frac{9}{10}y_2(t)$.
- Orangutans live to 30 years, about 15 years of adulthood so in any one year a fraction 14/15 of adult females live to the next year. Consequently, $y_3(t+1) = \frac{1}{10}y_2(t) + \frac{9}{10}y_3(t)$.

The mathematical model of the age structure is to let vector $\boldsymbol{y} = (y_1, y_2, y_3)$ then

$$\boldsymbol{y}(t+1) = A\boldsymbol{y}(t) = \begin{bmatrix} \frac{4}{5} & 0 & \frac{1}{15} \\ \frac{1}{5} & \frac{9}{10} & 0 \\ 0 & \frac{1}{10} & \frac{14}{15} \end{bmatrix} \boldsymbol{y}(t)$$

To predict the population we need to know the current population. The given information is that there are initially 30 adolescent females and 15 adult females. This information specifies that at time zero the population vector $\boldsymbol{y}(0) = (0, 30, 15)$ in the study area.

(a) Then the rule $\boldsymbol{y}(t+1) = A\boldsymbol{y}(t)$ with time $t = 0$ gives

$$\begin{aligned} \boldsymbol{y}(1) &= A\boldsymbol{y}(0) \\ &= \begin{bmatrix} \frac{4}{5} & 0 & \frac{1}{15} \\ \frac{1}{5} & \frac{9}{10} & 0 \\ 0 & \frac{1}{10} & \frac{14}{15} \end{bmatrix} \begin{bmatrix} 0 \\ 30 \\ 15 \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ 27 \\ 17 \end{bmatrix}. \end{aligned}$$

That is, during the first year there is one birth of a female juvenile, three adolescents matured to adults, and one adult died.

(b) Then the rule $\boldsymbol{y}(t+1) = A\boldsymbol{y}(t)$ with time $t = 1$ year gives

$$\begin{aligned} \boldsymbol{y}(2) &= A\boldsymbol{y}(1) \\ &= \begin{bmatrix} \frac{4}{5} & 0 & \frac{1}{15} \\ \frac{1}{5} & \frac{9}{10} & 0 \\ 0 & \frac{1}{10} & \frac{14}{15} \end{bmatrix} \begin{bmatrix} 1 \\ 27 \\ 17 \end{bmatrix} \end{aligned}$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$= \begin{bmatrix} \frac{29}{15} \\ \frac{49}{2} \\ \frac{557}{30} \end{bmatrix} = \begin{bmatrix} 1.93 \\ 24.50 \\ 18.57 \end{bmatrix} \text{ (2 d.p.).}$$

In real life we cannot have such fractions of an orangutan. These predictions are averages, or expectations on average, and must be interpreted as such. Thus after two years, we expect on average nearly two juveniles, 24 or 25 adolescents, and 18 or 19 adults.

(c) Continuing on with the aid of MATLAB/Octave or calculator, the rule $\boldsymbol{y}(t+1) = A\boldsymbol{y}(t)$ with time $t = 2$ years gives

$$\begin{aligned} \boldsymbol{y}(3) &= A\boldsymbol{y}(2) \\ &= \begin{bmatrix} \frac{4}{5} & 0 & \frac{1}{15} \\ \frac{1}{5} & \frac{9}{10} & 0 \\ 0 & \frac{1}{10} & \frac{14}{15} \end{bmatrix} \begin{bmatrix} 1.93 \\ 24.50 \\ 18.57 \end{bmatrix} \\ &= \begin{bmatrix} 2.78 \\ 22.44 \\ 19.78 \end{bmatrix} \text{ (2 d.p.).} \end{aligned}$$

In MATLAB/Octave do this calculation via

```
A=[4/5 0 1/15;1/5 9/10 0;0 1/10 14/15]
y0=[0;30;15]
y1=A*y0
y2=A*y1
y3=A*y2
y4=A*y3
y5=A*y4
```

(d) Consequently, the rule $\boldsymbol{y}(t+1) = A\boldsymbol{y}(t)$ with time $t = 3$ years gives

$$\begin{aligned} \boldsymbol{y}(4) &= A\boldsymbol{y}(3) \\ &= \begin{bmatrix} \frac{4}{5} & 0 & \frac{1}{15} \\ \frac{1}{5} & \frac{9}{10} & 0 \\ 0 & \frac{1}{10} & \frac{14}{15} \end{bmatrix} \begin{bmatrix} 2.78 \\ 22.44 \\ 19.78 \end{bmatrix} \\ &= \begin{bmatrix} 3.55 \\ 20.75 \\ 20.70 \end{bmatrix} \text{ (2 d.p.).} \end{aligned}$$

(e) Lastly, the rule $\boldsymbol{y}(t+1) = A\boldsymbol{y}(t)$ with time $t = 4$ years gives

$$\begin{aligned} \boldsymbol{y}(5) &= A\boldsymbol{y}(4) \\ &= \begin{bmatrix} \frac{4}{5} & 0 & \frac{1}{15} \\ \frac{1}{5} & \frac{9}{10} & 0 \\ 0 & \frac{1}{10} & \frac{14}{15} \end{bmatrix} \begin{bmatrix} 3.55 \\ 20.75 \\ 20.70 \end{bmatrix} \end{aligned}$$

v0.2a

$$= \begin{bmatrix} 4.22 \\ 19.38 \\ 21.40 \end{bmatrix} \quad (2 \text{ d.p.}).$$

Thus after five years the mathematical model predicts about 4 juveniles, 19 adolescents, and 21 adults (on average).

Notice that the five-year population of 44 females (45 if you add all the fractions) is the same as the starting population. This nearly static total population is no accident, as we next see, and contributes to why orangutans are critically endangered.

■

The mathematical model $\boldsymbol{y}(t+1) = A\boldsymbol{y}(t)$ does forecast the future populations. However, to make predictions for many years and for general initial populations we prefer the formula solution given by the next Theorem 7.1.16 and introduced in the next example.

Example 7.1.15. A vector $\boldsymbol{y}(t) \in \mathbb{R}^2$ changes with time t according to

$$\boldsymbol{y}(t+1) = A\boldsymbol{y}(t) = \begin{bmatrix} 1 & -1 \\ -4 & 1 \end{bmatrix} \boldsymbol{y}(t).$$

First, what is $\boldsymbol{y}(3)$ if the initial $\boldsymbol{y}(0) = (0,1)$? Second, find a general formula for $\boldsymbol{y}(t)$ from any $\boldsymbol{y}(0)$.

Solution: First, given $\boldsymbol{y}(0) = (0,1)$ just compute (as drawn in the margin)

$$\begin{aligned} \boldsymbol{y}(1) &= A\boldsymbol{y}(0) = \begin{bmatrix} 1 & -1 \\ -4 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \\ \boldsymbol{y}(2) &= A\boldsymbol{y}(1) = \begin{bmatrix} 1 & -1 \\ -4 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ 5 \end{bmatrix}, \\ \boldsymbol{y}(3) &= A\boldsymbol{y}(2) = \begin{bmatrix} 1 & -1 \\ -4 & 1 \end{bmatrix} \begin{bmatrix} -2 \\ 5 \end{bmatrix} = \begin{bmatrix} -7 \\ 13 \end{bmatrix}. \end{aligned}$$

Second, let's suppose there may be solutions in the form $\boldsymbol{y} = \lambda^t \boldsymbol{v}$ for some non-zero scalar λ and some vector $\boldsymbol{v} \in \mathbb{R}^2$. Substitute into $\boldsymbol{y}(t+1) = A\boldsymbol{y}(t)$ to find $\lambda^{t+1}\boldsymbol{v} = A\lambda^t\boldsymbol{v}$. Divide by (non-zero) λ^t to find we require $A\boldsymbol{v} = \lambda\boldsymbol{v}$: this is an eigen-problem. That is, for every eigenvalue λ with corresponding eigenvector $\boldsymbol{v}$ there is a solution $\boldsymbol{y} = \lambda^t\boldsymbol{v}$.

To find the eigenvalues here solve the characteristic equation

$$\det(A - \lambda I) = \begin{vmatrix} 1-\lambda & -1 \\ -4 & 1-\lambda \end{vmatrix} = (1-\lambda)^2 - 4 = 0\,.$$

Rearrange to $(\lambda - 1)^2 = 4$ and take square roots to give $\lambda - 1 = \pm 2$, that is, eigenvalues $\lambda = 1 \pm 2 = -1, 3$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

- For eigenvalue $\lambda_1 = -1$ the corresponding eigenvector has to satisfy
$$(A+I)\boldsymbol{v}_1 = \begin{bmatrix} 2 & -1 \\ -4 & 2 \end{bmatrix} \boldsymbol{v}_1 = \boldsymbol{0}\,.$$
That is, $\boldsymbol{v}_1 \propto (1,2)$; let's take $\boldsymbol{v}_1 = (1,2)$.
- For eigenvalue $\lambda_2 = 3$ the corresponding eigenvector has to satisfy
$$(A-3I)\boldsymbol{v}_2 = \begin{bmatrix} -2 & -1 \\ -4 & -2 \end{bmatrix} \boldsymbol{v}_2 = \boldsymbol{0}\,.$$
That is, $\boldsymbol{v}_2 \propto (1,-2)$; let's take $\boldsymbol{v}_2 = (-1,2)$.

Summarising so far: $\boldsymbol{y}'(t) = (-1)^t(1,2)$ and $\boldsymbol{y}''(t) = 3^t(-1,2)$ are both solutions of $\boldsymbol{y}(t+1) = A\boldsymbol{y}(t)$.

Further, the equation $\boldsymbol{y}(t+1) = A\boldsymbol{y}(t)$ is linear in $\boldsymbol{y}$, so any linear combination of solutions is also a solution. Let's try $\boldsymbol{y}(t) = c_1\boldsymbol{y}'(t) + c_2\boldsymbol{y}''(t)$. Substituting

$$\begin{aligned} A\boldsymbol{y}(t) &= A\big(c_1\boldsymbol{y}'(t) + c_2\boldsymbol{y}''(t)\big) \\ &= c_1A\boldsymbol{y}'(t) + c_2A\boldsymbol{y}''(t) \\ &= c_1\boldsymbol{y}'(t+1) + c_2\boldsymbol{y}''(t+1) = \boldsymbol{y}(t+1), \end{aligned}$$

as required. So, $\boldsymbol{y}(t) = c_1(-1)^t(1,2) + c_2 3^t(-1,2)$ are solutions for all constants c_1 and c_2, and over all time t.

Now, what values of constants c_1 and c_2 should be chosen for any given initial $\boldsymbol{y}(0)$? Substitute time $t=0$ into $\boldsymbol{y}(t) = c_1(-1)^t(1,2) + c_2 3^t(-1,2)$ to require $\boldsymbol{y}(0) = c_1(-1)^0(1,2) + c_2 3^0(-1,2)$. Since the zero powers $(-1)^0 = 3^0 = 1$, this requires $\boldsymbol{y}(0) = c_1(1,2) + c_2(-1,2)$. Write as the matrix-vector system

$$\begin{bmatrix} 1 & -1 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \boldsymbol{y}(0) \iff \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} \\ -\frac{1}{2} & \frac{1}{4} \end{bmatrix} \boldsymbol{y}(0)$$

upon invoking the inverse (Theorem 3.2.7) of the matrix of eigenvectors, $P = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 \end{bmatrix}$. That is, because the matrix of eigenvectors is invertible, we find constants c_1 and c_2 to suit any initial $\boldsymbol{y}(0)$.

For example, with initial $\boldsymbol{y}(0) = (0,1)$ the above formula gives $(c_1,c_2) = (\frac{1}{4},\frac{1}{4})$ and so the corresponding formula solution is $\boldsymbol{y}(t) = \frac{1}{4}(-1)^t(1,2) + \frac{1}{4}3^t(-1,2)$. To check, evaluate at say $t=3$ to find $\boldsymbol{y}(3) = -\frac{1}{4}(1,2) + \frac{27}{4}(-1,2) = (-7,13)$, as before. In general, as here, as time t increases, the solution $\boldsymbol{y}(t)$ grows like 3^t with a little oscillation from the $(-1)^t$ term.

■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Theorem 7.1.16. *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 $\lambda_1, \lambda_2, \dots, \lambda_m$ be eigenvalues of A and $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_m$ be corresponding eigenvectors, then a solution of $\boldsymbol{y}(t+1) = A\boldsymbol{y}(t)$ is the linear combination

$$\boldsymbol{y}(t) = c_1\lambda_1^t\boldsymbol{v}_1 + c_2\lambda_2^t\boldsymbol{v}_2 + \cdots + c_m\lambda_m^t\boldsymbol{v}_m \tag{7.2}$$

for all constants $c_1, c_2, \dots, c_m$.

*(b) Further, if the number of eigenvectors $m = n$, the size of the matrix, and the matrix of eigenvectors $P = \begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n\end{bmatrix}$ is invertible, then the general linear combination (7.2) is a **general solution** in that constants $c_1, c_2, \dots, c_n$ can be found for every given initial $\boldsymbol{y}(0)$.*

Proof.

7.1.16a Just premultiply (7.2) by matrix A to find that

$$\begin{aligned}
A\boldsymbol{y}(t) &= A(c_1\lambda_1^t\boldsymbol{v}_1 + c_2\lambda_2^t\boldsymbol{v}_2 + \cdots + c_m\lambda_m^t\boldsymbol{v}_m) \\
&\qquad \text{(using distributivity Thm. 3.1.22)} \\
&= c_1\lambda_1^t A\boldsymbol{v}_1 + c_2\lambda_2^t A\boldsymbol{v}_2 + \cdots + c_m\lambda_m^t A\boldsymbol{v}_m \\
&\qquad \text{(as eigenvectors } A\boldsymbol{v}_j = \lambda_j\boldsymbol{v}_j) \\
&= c_1\lambda_1^t\lambda_1\boldsymbol{v}_1 + c_2\lambda_2^t\lambda_2\boldsymbol{v}_2 + \cdots + c_m\lambda_m^t\lambda_m\boldsymbol{v}_m \\
&= c_1\lambda_1^{t+1}\boldsymbol{v}_1 + c_2\lambda_2^{t+1}\boldsymbol{v}_2 + \cdots + c_m\lambda_m^{t+1}\boldsymbol{v}_m
\end{aligned}$$

which is the given formula (7.2) for $\boldsymbol{y}(t+1)$. Hence (7.2) is a solution of $\boldsymbol{y}(t+1) = A\boldsymbol{y}(t)$.

7.1.16b For any given initial value $\boldsymbol{y}(0)$, the solution (7.2) will hold if we can find constants $c_1, c_2, \dots, c_m$ such that the solution (7.2) evaluates to $\boldsymbol{y}(0)$ at time $t = 0$. Let's do this given the preconditions that the number of eigenvectors $m = n$ and the matrix $P = \begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n\end{bmatrix}$ is invertible. Evaluating the solution (7.2) at $t = 0$ we need to solve, since the zeroth power $\lambda_j^0 = 1$,

$$\boldsymbol{y}(0) = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n\,.$$

Writing as a matrix-vector system this equation requires $P\boldsymbol{c} = \boldsymbol{y}(0)$ for constant vector $\boldsymbol{c} = (c_1, c_2, \dots, c_n)$. Since P is invertible, $P\boldsymbol{c} = \boldsymbol{y}(0)$ always has the unique solution $\boldsymbol{c} = P^{-1}\boldsymbol{y}(0)$ (Theorem 3.4.43) which are the requisite constants.

□

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Example 7.1.17. 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 $\boldsymbol{y}(0) = (1,0)$? Second, find a general solution.

Solution: First, just compute (as illustrated)

$$\begin{aligned}
\boldsymbol{y}(1) &= A\boldsymbol{y}(0) = \begin{bmatrix} 1 & 3 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \\
\boldsymbol{y}(2) &= A\boldsymbol{y}(1) = \begin{bmatrix} 1 & 3 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \end{bmatrix}, \\
\boldsymbol{y}(3) &= A\boldsymbol{y}(2) = \begin{bmatrix} 1 & 3 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} -2 \\ -2 \end{bmatrix} = \begin{bmatrix} -8 \\ 0 \end{bmatrix}.
\end{aligned}$$

Interestingly, after three steps in time $\boldsymbol{y}(3)$ is (-8) times the initial $\boldsymbol{y}(0)$. This suggest after six steps in time $\boldsymbol{y}(6)$ will be $(-8)^2 = 64$ times the initial $\boldsymbol{y}(0)$, and so on. Perhaps the solution grows in size roughly like 2^t but in some irregular manner.

Second, find a general solution via the eigenvalues and eigenvectors of the matrix A. Its characteristic equation is

$$\det(A - \lambda I) = \begin{vmatrix} 1-\lambda & 3 \\ -1 & 1-\lambda \end{vmatrix} = (1-\lambda)^2 + 3 = 0\,.$$

That is, $(\lambda - 1)^2 = -3$ which upon taking square roots gives the complex conjugate pair of eigenvalues $\lambda = 1 \pm i\sqrt{3}$. Theorem 7.1.16 applies for complex eigenvalues and eigenvectors so we proceed.

- For eigenvalue $\lambda_1 = 1 + i\sqrt{3}$ the corresponding eigenvectors $\boldsymbol{v}_1$ satisfy

  $$(A - (1+i\sqrt{3})I)\boldsymbol{v}_1 = \begin{bmatrix} -i\sqrt{3} & 3 \\ -1 & -i\sqrt{3} \end{bmatrix} \boldsymbol{v}_1 = \boldsymbol{0}\,.$$

  Solutions are proportional to $\boldsymbol{v}_1 = (-i\sqrt{3}, 1)$.

- For eigenvalue $\lambda_2 = 1 - i\sqrt{3}$ the corresponding eigenvectors $\boldsymbol{v}_2$ satisfy

  $$(A - (1-i\sqrt{3})I)\boldsymbol{v}_2 = \begin{bmatrix} i\sqrt{3} & 3 \\ -1 & i\sqrt{3} \end{bmatrix} \boldsymbol{v}_2 = \boldsymbol{0}\,.$$

  Solutions are proportional to $\boldsymbol{v}_2 = (+i\sqrt{3}, 1)$.

Theorem 7.1.16 then establishes that a solution to $\boldsymbol{y}(t+1) = A\boldsymbol{y}(t)$ is

$$\boldsymbol{y}(t) = c_1(1+i\sqrt{3})^t \begin{bmatrix} -i\sqrt{3} \\ 1 \end{bmatrix} + c_2(1-i\sqrt{3})^t \begin{bmatrix} +i\sqrt{3} \\ 1 \end{bmatrix}.$$

This is a general solution since the matrix of the two eigenvectors (albeit complex valued) is invertible:

$$P = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 \end{bmatrix} = \begin{bmatrix} -i\sqrt{3} & i\sqrt{3} \\ 1 & 1 \end{bmatrix} \quad \text{has } P^{-1} = \begin{bmatrix} \frac{i}{2\sqrt{3}} & \frac{1}{2} \\ \frac{-i}{2\sqrt{3}} & \frac{1}{2} \end{bmatrix}$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

as its inverse (Theorem 3.2.7).

For example, if $\boldsymbol{y}(0) = (1,0)$, then the coefficient constants are $(c_1, c_2) = P^{-1}(1,0) = (i,-i)/(2\sqrt{3})$. Then the solution becomes

$$\begin{aligned}\boldsymbol{y}(t) &= \frac{i}{2\sqrt{3}}(1+i\sqrt{3})^t\begin{bmatrix}-i\sqrt{3}\\1\end{bmatrix} - \frac{i}{2\sqrt{3}}(1-i\sqrt{3})^t\begin{bmatrix}+i\sqrt{3}\\1\end{bmatrix}\\ &= \tfrac{1}{2}(1+i\sqrt{3})^t\begin{bmatrix}1\\\frac{i}{\sqrt{3}}\end{bmatrix} + \tfrac{1}{2}(1-i\sqrt{3})^t\begin{bmatrix}1\\-\frac{i}{\sqrt{3}}\end{bmatrix}.\end{aligned}$$

Through the magic of the complex conjugate form of the two terms in this expression, the complex parts cancel to always give a real result. For example, this complex formula predicts at time step $t = 1$

$$\begin{aligned}\boldsymbol{y}(1) &= \tfrac{1}{2}(1+i\sqrt{3})\begin{bmatrix}1\\\frac{i}{\sqrt{3}}\end{bmatrix} + \tfrac{1}{2}(1-i\sqrt{3})\begin{bmatrix}1\\-\frac{i}{\sqrt{3}}\end{bmatrix}\\ &= \frac{1}{2}\begin{bmatrix}1+i\sqrt{3}+1-i\sqrt{3}\\\frac{i}{\sqrt{3}}-1-\frac{i}{\sqrt{3}}-1\end{bmatrix}\\ &= \begin{bmatrix}1\\-1\end{bmatrix},\end{aligned}$$

as computed directly at the start of this example.

■

One crucial qualitative aspect we need to know is whether components in the solution (7.2) grow, decay, or stay the same size as time increases. This is determined by the eigenvalues as λ_j^t is the only place that the time appears in the formula (7.2).

- For example, in the general solution for Example 7.1.15, $\boldsymbol{y}(t) = c_1(-1)^t(1,2) + c_2 3^t(-1,2)$, 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 we know that the term involving the factor 3^t will dominate the solution as it grows.

- In Example 7.1.17 with complex conjugate eigenvalues the situation is more complicated. Let's write any given complex eigenvalue in polar form $\lambda = r(\cos\theta + i\sin\theta)$ where magnitude $r = |\lambda|$ and angle θ such that $\tan\theta = (\Im\lambda)/(\Re\lambda)$. For example, $1-i$ has magnitude $r = |1-1i| = \sqrt{1^2+(-1)^2} = \sqrt{2}$ and angle $\theta = -\frac{\pi}{4}$ since $\tan(-\frac{\pi}{4}) = -1/1$. Question: how does this help understand the solution which has λ_j^t in it? Answer: De Moivre's theorem says that if $\lambda = r[\cos\theta + i\sin\theta]$, then $\lambda^t = r^t\big[\cos(\theta t) + i\sin(\theta t)\big]$. Since the magnitude $|\cos(\theta t) + i\sin(\theta t)| = \sqrt{\cos^2(\theta t) + \sin^2(\theta t)} = \sqrt{1} = 1$, the magnitude $|\lambda^t| = r^t$. For example, the magnitude $|(1-i)^2| = (\sqrt{2})^2 = 2$

which we check by computing $(1-i)^2 = 1^2 - 2i + i^2 = -2i$ and $|-2i| = 2$.

In Example 7.1.17, the eigenvalue $\lambda_1 = 1 + i\sqrt{3}$ so its magnitude is $r_1 = |\lambda_1| = |1 + i\sqrt{3}| = \sqrt{1+3} = 2$. Hence the magnitude $|\lambda_1^t| = 2^t$ at any time step t. Similarly, the magnitude $|\lambda_2^t| = 2^t$ at any time step t. Consequently, the general solution

$$\boldsymbol{y}(t) = c_1\lambda_1^t \begin{bmatrix} -i\sqrt{3} \\ 1 \end{bmatrix} + c_2\lambda_2^t \begin{bmatrix} +i\sqrt{3} \\ 1 \end{bmatrix}$$

will grow in magnitude roughly like 2^t as both components grow like 2^t. It is a 'rough' growth because the components $\cos(\theta t)$ and $\sin(\theta t)$ cause 'oscillations' in time t. Nonetheless the overall growth like $|\lambda_1|^t = |\lambda_2|^t = 2^t$ is inexorable—and seen previously in the particular solution where we observe $\boldsymbol{y}(3)$ 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 will, as time t increases,

- grow to infinity if $|\lambda| > 1$,
- decay to zero if $|\lambda| < 1$, and
- remain the same magnitude if $|\lambda| = 1$.

Example 7.1.18 (orangutans over many years). Extend orangutan analysis of Example 7.1.14. Use Theorem 7.1.16 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.

Solution: Example 7.1.14 derived that the age structure population $\boldsymbol{y} = (y_1, y_2, y_3)$ satisfies $\boldsymbol{y}(t+1) = A\boldsymbol{y}(t)$ for matrix

$$A = \begin{bmatrix} \frac{4}{5} & 0 & \frac{1}{15} \\ \frac{1}{5} & \frac{9}{10} & 0 \\ 0 & \frac{1}{10} & \frac{14}{15} \end{bmatrix}.$$

Let's find the eigenvalues and eigenvectors of the matrix A using MATLAB/Octave via

```
A=[4/5 0 1/15;1/5 9/10 0;0 1/10 14/15]
[V,D]=eig(A)
```

to find

```
V =
 -0.3077+0.2952i -0.3077-0.2952i  0.2673+0.0000i
  0.7385+0.0000i  0.7385+0.0000i  0.5345+0.0000i
 -0.4308-0.2952i -0.4308+0.2952i  0.8018+0.0000i
D =
  0.8167+0.0799i  0.0000+0.0000i  0.0000+0.0000i
  0.0000+0.0000i  0.8167-0.0799i  0.0000+0.0000i
  0.0000+0.0000i  0.0000+0.0000i  1.0000+0.0000i
```

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Evidently there is one real eigenvalue of $\lambda_3 = 1$ and two complex conjugate eigenvalues $\lambda_{1,2} = 0.8167 \pm i0.0799$. Corresponding eigenvectors are the columns $\boldsymbol{v}_j$ of V. Thus a solution for the orangutan population is

$$\boldsymbol{y}(t) = c_1\lambda_1^t\boldsymbol{v}_1 + c_2\lambda_2^t\boldsymbol{v}_2 + c_3\lambda_3^t\boldsymbol{v}_3 .$$

- For the initial population $\boldsymbol{y}(0) = (0, 30, 15)$ we need to find constants $\boldsymbol{c} = (c_1, c_2, c_3)$ such that $V\boldsymbol{c} = \boldsymbol{y}(0)$. Solve this linear equation in MATLAB/Octave with

```
y0=[0;30;15]
rcond(V)
c=V\y0
```

which gives the answer

```
ans =
    0.1963
c =
  10.1550+2.1175i
  10.1550-2.1175i
  28.0624+0.0000i
```

The `rcond` value of 0.1963 indicates that matrix V is invertible. Then the backslash operator computes the above coefficients $\boldsymbol{c}$. Via the magic of complex conjugates cancelling, the real population of orangutans is for all times predicted to be (2 d.p.)

$$\begin{aligned}\boldsymbol{y}(t) &= (10.16 + 2.12i)(0.82 + 0.08i)^t \begin{bmatrix} -0.31 + 0.30i \\ 0.74 \\ -0.43 - 0.30i \end{bmatrix} \\ &\quad + (10.16 - 2.12i)(0.82 + 0.08i)^t \begin{bmatrix} -0.31 - 0.30i \\ 0.74 \\ -0.43 + 0.30i \end{bmatrix} \\ &\quad + 28.06 \begin{bmatrix} 0.27 \\ 0.53 \\ 0.80 \end{bmatrix}\end{aligned}$$

since $\lambda_3^t = 1^t = 1$.

Since the magnitude $|\lambda_1| = |\lambda_2| = 0.82$ (2 d.p.), the first two terms in this expression decay to zero as time t increases. For example, $|\lambda_1^{12}| = |\lambda_2^{12}| = 0.09$. Hence the model predicts that over long times the population

$$\boldsymbol{y}(t) \approx 28.06 \begin{bmatrix} 0.27 \\ 0.53 \\ 0.80 \end{bmatrix} = \begin{bmatrix} 7.5 \\ 15.0 \\ 22.5 \end{bmatrix}$$

Such a static population means that the orangutans are highly sensitive to disease, or deforestation, or chance events, and so on.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

- Such unfortunate sensitivity is typical for orangutans. It is not a quirk of the initial population. Recall the general prediction for the orangutans is

$$\boldsymbol{y}(t) = c_1\lambda_1^t\boldsymbol{v}_1 + c_2\lambda_2^t\boldsymbol{v}_2 + c_3\lambda_3^t\boldsymbol{v}_3\,.$$

The initial population determines the constants $\boldsymbol{c}$. However, the long term population is always predicted to be static. The reason is that the magnitude of the eigenvalues $|\lambda_1| = |\lambda_2| = 0.82$ and so the first two terms in this general solution will in time always decay to zero. Further, the remaining third eigenvalue has magnitude $|\lambda_3| = |1| = 1$ and so the third term in the population prediction is always constant in time t. That is, over long times the population is always

$$\boldsymbol{y}(t) \approx c_3\boldsymbol{v}_3\,.$$

Such a static population means that the orangutans are always sensitive to disease, or deforestation, or chance events, and so on.

■

Bliss et al. (2016) discuss mathematical modelling. On page 23 they comment "*Modelling (like real life) is open-ended and messy*": in our two examples here you have to extract the important factors from many unneeded details, and use them in the context of an imperfect model. Bliss et al. (2016) [p.23] also comment that 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, Bliss et al. (2016) recommend to "*focus on the process, not the product*": depending upon your choices and interpretations you will develop alternative plausible models in these scenarios—it is the process of forming plausible models and interpreting the results that are important here.

Example 7.1.19 (servals grow)**.** The serval is a member of the cat family that lives in Africa. Given next is an extract from Wikipedia of a serval's Reproduction and Life History.

> Kittens are born shortly before the peak breeding period of local rodent populations. A serval is able to give birth to multiple litters throughout the year, but commonly does so only if the earlier litters die shortly after birth. Gestation lasts from 66 to 77 days and commonly results in the birth of two kittens, although sometimes as few as one or as many as four have been recorded.
>
> The kittens are born in dense vegetation or sheltered locations such as abandoned aardvark burrows. If such

> an ideal location is not available, a place beneath a shrub may be sufficient. The kittens weigh around 250 gm at birth, and are initially blind and helpless, with a coat of greyish woolly hair. They open their eyes at 9 to 13 days of age, and begin to take solid food after around a month. At around six months, they acquire their permanent canine teeth and begin to hunt for themselves; they leave their mother at about 12 months of age. They may reach sexual maturity from 12 to 25 months of age.
>
> Life expectancy is about 10 years in the wild.

From the information in this extract, create a plausible, age structured, population model of the serval: give reasons for estimates of the coefficients in the model. Choose three age categories of kittens, juveniles, sexually mature adults. What does the model predict over long times?

Solution: Recall we only model the number of *female* servals as females are the limiting breeders. Define

- $y_1(t)$ is the number of female kittens, less than 0.5 years old from when they "begin to hunt for themselves";
- $y_2(t)$ is the number of female juveniles, between 0.5 years and 1.5 years which is when they "reach sexual maturity" on average;
- $y_3(t)$ is the number of female breeding adults, older than 1.5 years, and dying at about the "life expectancy" of 10 years;
- since servals transition from one age category to another in multiples of six months (0.5 years), let the unit of time be six months, equivalently a half-year. Consequently, time $t+1$ is the time a half-year later than time t.

Modelling of the servals leads to the following equations.

- Kittens are commonly born once a year to each female, and the common litter size is two, so *on average* one female kitten is born per year per adult, that is, on average $\frac{1}{2}$ female kitten is born per half-year per adult female: Also all kittens age to juveniles after 0.5 years, so none remain as kittens. Hence the kitten equation is $y_1(t+1) = 0y_1(t) + \frac{1}{2}y_3(t)$.
- Juveniles mature from the kittens, and age to an adult after about one year: that is, on average half of them become adults every half-year, and half remain juveniles. So the juvenile equation is $y_2(t+1) = y_1(t) + \frac{1}{2}y_2(t)$.
- Adults mature from the juveniles, and die after about 8.5 years which is about a rate 1/8.5 per year: that is, a rate of $\frac{1}{17}$ per half-year leaving $\frac{16}{17}$ of them to live into the next half-year. So the adult equation is $y_3(t+1) = \frac{1}{2}y_2(t) + \frac{16}{17}y_3(t)$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Bring these equations together, the age structure population $\boldsymbol{y} = (y_1, y_2, y_3)$ satisfies $\boldsymbol{y}(t+1) = A\boldsymbol{y}(t)$ for matrix

$$A = \begin{bmatrix} 0 & 0 & \frac{1}{2} \\ 1 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{16}{17} \end{bmatrix} .$$

Find the eigenvalues and eigenvectors of the matrix A using MATLAB/Octave via

```
A=[0 0 1/2;1 1/2 0;0 1/2 16/17]
[V,D]=eig(A)
```

to find

```
V =
 -0.3066+0.3439i -0.3066-0.3439i  0.3352+0.0000i
  0.7838+0.0000i  0.7838+0.0000i  0.4633+0.0000i
 -0.3684-0.1942i -0.3684+0.1942i  0.8203+0.0000i
D =
  0.1088+0.4387i  0.0000+0.0000i  0.0000+0.0000i
  0.0000+0.0000i  0.1088-0.4387i  0.0000+0.0000i
  0.0000+0.0000i  0.0000+0.0000i  1.2236+0.0000i
```

Evidently there is one real eigenvalue of $\lambda_3 = 1.2236$ and two complex conjugate eigenvalues $\lambda_{1,2} = 0.1088 \pm i0.4387$. Corresponding eigenvectors are the columns $\boldsymbol{v}_j$ of V. Thus a general solution for the serval population is (Theorem 7.1.16)

$$\boldsymbol{y}(t) = c_1\lambda_1^t\boldsymbol{v}_1 + c_2\lambda_2^t\boldsymbol{v}_2 + c_3\lambda_3^t\boldsymbol{v}_3 .$$

In this general solution, the first two terms will decay in time to zero. The reason is that the magnitudes $|\lambda_1| = |\lambda_2| = |0.1088 \pm i0.4387| = \sqrt{0.1088^2 + 0.4387^2} = 0.4520$, and since this magnitude is less than one, then λ_1^t and λ_2^t will decay to zero with increasing time t. However, the third term increases in time as $\lambda_3 = 1.2236 > 1$. The model predicts the serval population increases by about 22% per half-year (about 50% per year).

Predation, disease, and food shortages are just some processes not included in this model which act to limit the serval's population in ways not included in this model. ■

Crucial in this section—so that we find a solution for all initial states—is that the matrix of eigenvectors is invertible. The next Section 7.2 relates the invertibility of a matrix of eigenvectors to the concept of 'linear independence' defined in that section (Theorem 7.2.31).

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

7.1.4 Extension: Connect SVDs to eigen-problems

This optional section connects the SVD of a general matrix to a symmetric eigen-problem, in principle.

Recall that Chapter 4 starts by illustrating the close connection between the SVD of a symmetric matrix and the eigenvalues and eigenvectors of that symmetric matrix. This subsection establishes that an SVD of a general matrix is closely connected to the eigenvalues and eigenvectors of a specific matrix of double the size. The connection depends upon determinants and solving linear systems and so, in principle, is an approach to compute an SVD distinct from the inductive maximisation.

Example 7.1.20. Compute the eigenvalues and eigenvectors of the (symmetric) matrix

$$B = \begin{bmatrix} 0 & 0 & 10 & 2 \\ 0 & 0 & 5 & 11 \\ 10 & 5 & 0 & 0 \\ 2 & 11 & 0 & 0 \end{bmatrix}.$$

Solution: In MATLAB/Octave execute

```
B=[0 0 10 2
   0 0 5 11
  10 5 0 0
  2 11 0 0]
[V,D]=eig(B)
```

and obtain (2 d.p.)

```
V =
   0.42   0.57   0.57  -0.42
   0.57  -0.42  -0.42  -0.57
  -0.50  -0.50   0.50  -0.50
  -0.50   0.50  -0.50  -0.50
D =
  -14.14       0       0       0
       0   -7.07       0       0
       0       0    7.07       0
       0       0       0   14.14
```

The eigenvalues are the pairs ± 7.07 and ± 14.14, with corresponding eigenvector pairs $(0.57, -0.42, \pm 0.50, \mp 0.50)$ and $(\mp 0.42, \mp 0.57, -0.50, -0.50)$.

These eigenvalues/vectors occur in $\pm$ pairs because this matrix has the form

$$B = \begin{bmatrix} O_2 & A \\ A^{\mathrm{T}} & O_2 \end{bmatrix}, \quad \text{here for matrix } A = \begin{bmatrix} 10 & 2 \\ 5 & 11 \end{bmatrix}$$

from Example 3.3.2. Observe that not only are the eigenvectors are orthogonal, because B is symmetric, but also the two parts of the eigenvectors are orthogonal: $(0.57, -0.42)$ from the first pair is orthogonal to $(-0.42, -0.57)$ from the second pair; and $(0.50, -0.50)$ from the first pair is orthogonal to $(-0.50, -0.50)$ from the second

v0.2a

pair. The next Theorem 7.1.21 establishes how these properties relate to an SVD for the matrix A. ■

Procedure 7.1.8 computes eigenvalues and eigenvectors by hand (in principle no matter how large the matrix). The procedure is independent of the SVD. We now invoke this procedure to find another method to find an SVD distinct from the inductive maximisation of the proof in Section 3.3.3. The following theorem is a step towards an efficient numerical computation of an SVD (Trefethen & Bau 1997, p.234).

Theorem 7.1.21 (SVD as an eigenproblem)**.** *Given an $m \times n$ matrix A, the singular values of A are the non-negative eigenvalues of the $(m+n)\times(m+n)$ matrix $B = \begin{bmatrix} O_m & A \\ A^{\mathrm{T}} & O_n \end{bmatrix}$. A 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 $\boldsymbol{v} \in \mathbb{R}^n$.*

Proof. First prove the SVD of A gives eigenvalues/vectors of B, and second the converse. For simplicity this proof considers only the case $m = n$; the case $m \neq n$ is similar but the more intricate details are of little interest.

First, let $n \times n$ matrix $A = USV^{\mathrm{T}}$ be an SVD (Theorem 3.3.6) for $n \times n$ orthogonal $U = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_n \end{bmatrix}$, orthogonal $V = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{bmatrix}$, and diagonal $S = \operatorname{diag}(\sigma_1, \sigma_2, \dots, \sigma_n)$. Post-multiply the SVD by orthogonal V gives $AV = US$. Also, transpose the SVD to give $A^{\mathrm{T}} = (USV^{\mathrm{T}})^{\mathrm{T}} = VS^{\mathrm{T}}U^{\mathrm{T}} = VSU^{\mathrm{T}}$ and post-multiply by orthogonal U to give $A^{\mathrm{T}}U = VS$. Now consider each of the $\pm$ cases of

$$B \begin{bmatrix} U \\ \pm V \end{bmatrix} = \begin{bmatrix} O_n & A \\ A^{\mathrm{T}} & O_n \end{bmatrix} \begin{bmatrix} U \\ \pm V \end{bmatrix} = \begin{bmatrix} \pm AV \\ A^{\mathrm{T}}U \end{bmatrix} = \begin{bmatrix} \pm US \\ VS \end{bmatrix} = \begin{bmatrix} U \\ \pm V \end{bmatrix} (\pm S).$$

Letting $\boldsymbol{w} = (\boldsymbol{u}_j, \pm\boldsymbol{v}_j) \neq \boldsymbol{0}$, the jth column of the above equation is $B\boldsymbol{w} = \pm\sigma_j\boldsymbol{w}$ and hence $(\boldsymbol{u}_j, \pm\boldsymbol{v}_j)$ is an eigenvector of B corresponding to eigenvalue $\pm\sigma_j$, respectively, for $j = 1, \dots, n$ and each of the $\pm$ cases.

Second, let $\boldsymbol{w} \in \mathbb{R}^{2n}$ be an eigenvector of B corresponding to an eigenvalue λ (real as B is symmetric, Theorem 4.2.8) and normalised so that $|\boldsymbol{w}| = 1$. Partition $\boldsymbol{w} = (\boldsymbol{u}, \boldsymbol{v})$ for $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$. Then the fundamental eigen-problem $B\boldsymbol{w} = \lambda\boldsymbol{w}$ (Definition 4.1.1) partitions into

$$\begin{bmatrix} O & A \\ A^{\mathrm{T}} & O \end{bmatrix} \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{bmatrix} = \lambda \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{bmatrix} \iff A\boldsymbol{v} = \lambda\boldsymbol{u} \text{ and } A^{\mathrm{T}}\boldsymbol{u} = \lambda\boldsymbol{v}. \tag{7.3}$$

Now eigenvalues and eigenvectors of B come in pairs as the eigenvalue $\lambda' = -\lambda$ corresponds to eigenvector $\boldsymbol{w}' = (\boldsymbol{u}, -\boldsymbol{v})$: substitute into (7.3) to check, $A(-\boldsymbol{v}) = -A\boldsymbol{v} = -\lambda\boldsymbol{u} = \lambda'\boldsymbol{u}$ and

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$A^{\mathrm{T}}\boldsymbol{u} = \lambda\boldsymbol{v} = (-\lambda)(-\boldsymbol{v}) = \lambda'(-\boldsymbol{v})$. If the eigenvalue $\lambda \neq 0$, then corresponding to the distinct eigenvalues $\pm\lambda$ the eigenvectors $(\boldsymbol{u}\,,\pm\boldsymbol{v})$ of symmetric B are orthogonal (Theorem 4.2.10) and so $0 = (\boldsymbol{u}\,,\boldsymbol{v})\cdot(\boldsymbol{u}\,,-\boldsymbol{v}) = \boldsymbol{u}\cdot\boldsymbol{u} - \boldsymbol{v}\cdot\boldsymbol{v} = |\boldsymbol{u}|^2 - |\boldsymbol{v}|^2$ and hence $|\boldsymbol{u}| = |\boldsymbol{v}| = \frac{1}{\sqrt{2}}$ as $|\boldsymbol{w}|^2 = |\boldsymbol{u}|^2 + |\boldsymbol{v}|^2 = 1$ (see Example 7.1.20). Further, for any two distinct eigenvalues $|\lambda_i| \neq |\lambda_j| \neq 0$, the eigenvectors $(\boldsymbol{u}_i\,,\boldsymbol{v}_i)$ and $(\boldsymbol{u}_j\,,\pm\boldsymbol{u}_j)$ are also orthogonal, hence $0 = (\boldsymbol{u}_i\,,\boldsymbol{v}_i)\cdot(\boldsymbol{u}_j\,,\pm\boldsymbol{u}_j) = \boldsymbol{u}_i\cdot\boldsymbol{u}_j \pm \boldsymbol{v}_i\cdot\boldsymbol{v}_j$. Taking the sum and the difference of the $\pm$ cases of this equation gives $\boldsymbol{u}_i\cdot\boldsymbol{u}_j = \boldsymbol{v}_i\cdot\boldsymbol{v}_j = 0$; that is, (Definition 3.2.38) $\{\boldsymbol{u}_1\,,\boldsymbol{u}_2\,,\dots,\boldsymbol{u}_n\}$ and $\{\boldsymbol{v}_1\,,\boldsymbol{v}_2\,,\dots,\boldsymbol{v}_n\}$ are both orthogonal sets. In the cases when matrix B has $2n$ distinct non-zero eigenvalues, choose $(\boldsymbol{u}_j\,,\boldsymbol{v}_j)$ to be a normalised eigenvector corresponding to positive eigenvalue λ_j, $j = 1\,,\dots,n$. Upon setting $U = \sqrt{2}\begin{bmatrix}\boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_n\end{bmatrix}$, $V = \sqrt{2}\begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n\end{bmatrix}$, and $S = \operatorname{diag}(\lambda_1\,,\lambda_2\,,\dots,\lambda_n)$, equation (7.3) gives the columns of $AV = US$ which since the columns of U and V are orthonormal gives the SVD $A = USV^{\mathrm{T}}$ for singular vectors $\boldsymbol{u}_j\,,\boldsymbol{v}_j \in \mathbb{R}^n$ and singular values $\lambda_j (> 0)$.

Extensions of this proof cater for the case when zero is an eigenvalue and/or eigenvalues are repeated and/or the dimensions $m \neq n$. □

7.1.5 Exercises

Exercise 7.1.1. For each of the following list of numbers, could the numbers be all the eigenvalues of a 4×4 matrix? Justify your answer.

(a) $-1.2\,,-0.6\,,0.2\,,-1.4$

(b) $\pm 2\,,-3$

(c) $0\,,3\,,\pm 5\,,8$

(d) $0\,,3\pm 5i\,,8$

(e) $-1.4\pm\sqrt{7}i\,,-4\,,3\pm 2i$

Exercise 7.1.2. For each of the following matrices, determine the two highest order terms and the constant term in the characteristic polynomial of the matrix.

(a) $\begin{bmatrix}-7 & 1\\ -2 & 2\end{bmatrix}$

(b) $\begin{bmatrix}3 & -3\\ 6 & 2\end{bmatrix}$

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

(d) $\begin{bmatrix}3 & -3 & 6\\ -1 & 4 & 0\\ 0 & 0 & -4\end{bmatrix}$

(e) $\begin{bmatrix}-1 & 0 & -6\\ -7 & 0 & -4\\ 0 & -6 & 0\end{bmatrix}$

(f) $\begin{bmatrix}-1 & 0 & -2 & 0\\ 0 & 0 & 0 & 0\\ -3 & -1 & 0 & -2\\ -4 & 0 & -5 & 0\end{bmatrix}$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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

(h) $\begin{bmatrix} 0 & 4 & 0 & 1 \\ 4 & -5 & 0 & -3 \\ 0 & -4 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{bmatrix}$

Exercise 7.1.3. For each of the following characteristic polynomials, write down the size of the corresponding matrix, and the matrix's trace and determinant.

(a) $\lambda^2 + 5\lambda - 6$

(b) $\lambda^2 + 2\lambda - 10$

(c) $\lambda^2 - \lambda$

(d) $-\lambda^3 + 5\lambda^2 + 7\lambda - 20$

(e) $-\lambda^3 + 28\lambda - 199$

(f) $-\lambda^3 + 8\lambda^2 - 5\lambda$

(g) $\lambda^4 + 3\lambda^3 + 143\lambda - 56$

(h) $\lambda^4 + 5\lambda^2 - 41\lambda - 5$

Exercise 7.1.4. For each the following matrices, determine the characteristic polynomial by hand, and hence find all eigenvalues of the matrix and their multiplicity. Show your working.

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

(b) $\begin{bmatrix} 0 & 5 \\ -2 & 2 \end{bmatrix}$

(c) $\begin{bmatrix} 0 & -3 \\ 3 & 6 \end{bmatrix}$

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

(e) $\begin{bmatrix} 4.5 & 16 \\ -1 & -3.5 \end{bmatrix}$

(f) $\begin{bmatrix} -1 & 1 & -1 \\ -6 & -6 & 2 \\ -5 & -3 & -3 \end{bmatrix}$

(g) $\begin{bmatrix} -2 & -5 & -1 \\ 0 & 3 & 1 \\ 0 & -6 & -2 \end{bmatrix}$

(h) $\begin{bmatrix} 9 & 3 & 0 \\ -12 & -3 & 0 \\ 2 & -4 & -2 \end{bmatrix}$

(i) $\begin{bmatrix} -14 & 24 & 52 \\ -4 & 8 & 18 \\ -2 & 3 & 6 \end{bmatrix}$

(j) $\begin{bmatrix} -1 & -2 & 2 \\ 7 & 18 & -12 \\ 7 & 17 & -11 \end{bmatrix}$

(k) $\begin{bmatrix} -10 & -10 & -16 \\ 4 & 4 & 6 \\ 3 & 3 & 5 \end{bmatrix}$

(l) $\begin{bmatrix} 1 & -15 & 7 \\ -1 & -1 & -1 \\ -5 & -15 & -1 \end{bmatrix}$

v0.2a

Exercise 7.1.5. For each the following matrices, use MATLAB/Octave to find all eigenvalues of the matrix and their multiplicity.

(a) $\begin{bmatrix} -2.7 & 0 & 1.6 \\ 6.3 & -1 & -27.8 \\ -0.1 & 0 & -1.9 \end{bmatrix}$

(b) $\begin{bmatrix} 4.9 & -8.1 & 5.4 \\ 8 & -11.2 & 8 \\ 3.9 & -3.9 & 3.4 \end{bmatrix}$

(c) $\begin{bmatrix} -6.7 & -0.6 & -6.6 & 3.6 \\ 3 & 0.1 & 3 & -2 \\ 2.8 & 0.6 & 2.7 & -1.6 \\ -6 & 0 & -6 & 3.1 \end{bmatrix}$

(d) $\begin{bmatrix} 11 & 17.9 & -33.4 & 46.4 \\ 1.2 & 0.9 & 2.8 & 2.2 \\ -12.8 & -21 & 37.2 & -54.8 \\ -12.3 & -19.7 & 33.7 & -51.3 \end{bmatrix}$

(e) $\begin{bmatrix} 0.6 & 0 & 0 & 0 \\ 9.6 & -1 & 33.6 & 17.6 \\ -9.6 & 1.6 & -33 & -17.6 \\ 19.2 & -3.2 & 67.2 & 35.8 \end{bmatrix}$

(f) $\begin{bmatrix} 52.8 & -93.3 & 73.4 & 57.1 & 104 \\ 18.3 & -30.7 & 28.5 & 20.6 & 36.6 \\ -20 & 34.9 & -29.4 & -22.3 & -40 \\ 18.2 & -31.8 & 27.4 & 20.8 & 36.4 \\ -6.8 & 13.6 & -6.8 & -6.8 & -12.8 \end{bmatrix}$

(g) $\begin{bmatrix} -356.4 & 264.6 & -880.8 & -689.9 & 455.5 \\ -58.5 & 41.7 & -142.8 & -111.4 & 77 \\ 157 & -115.6 & 392.9 & 311.4 & -201 \\ -78.5 & 57.8 & -198.4 & -159.6 & 100.5 \\ -58.5 & 45.6 & -142.8 & -111.4 & 73.1 \end{bmatrix}$

(h) $\begin{bmatrix} 309.4 & -29.7 & 451.3 & 337.3 & 305.9 \\ -217.6 & 20.3 & -313.5 & -236.1 & -215.9 \\ 3 & 0 & 0.7 & 3 & 3 \\ -232.6 & 24.1 & -336 & -254.9 & -230.9 \\ -83.6 & 5.6 & -119.8 & -89.2 & -81.8 \end{bmatrix}$

Exercise 7.1.6. For each of the following matrices, find by hand the eigenspace of the nominated eigenvalue. Confirm your answer with MATLAB/Octave. Show your working.

(a) $\begin{bmatrix} -12 & 10 \\ -15 & 13 \end{bmatrix}$, $\lambda = 3$

(b) $\begin{bmatrix} -1 & 9 \\ -1 & 5 \end{bmatrix}$, $\lambda = 2$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(c) $\begin{bmatrix} -1 & 0 \\ -2 & 1 \end{bmatrix}$, $\lambda = -1$

(d) $\begin{bmatrix} 11 & -4 & -12 \\ -27 & 10 & 27 \\ 19 & -7 & -20 \end{bmatrix}$, $\lambda = 1$

(e) $\begin{bmatrix} -1 & -7 & -2 \\ 8 & 14 & 2 \\ 0 & 0 & 7 \end{bmatrix}$, $\lambda = 7$

(f) $\begin{bmatrix} -12 & -82 & -17 \\ 3 & 18 & 3 \\ -6 & -26 & -1 \end{bmatrix}$, $\lambda = 0$

(g) $\begin{bmatrix} -4 & 0 & -4 \\ -2 & -4 & 0 \\ 8 & 4 & 6 \end{bmatrix}$, $\lambda = -2$

(h) $\begin{bmatrix} -3 & 2 & -6 \\ -4 & 3 & -6 \\ 2 & -1 & 4 \end{bmatrix}$, $\lambda = 1$

Exercise 7.1.7. For each of the following matrices, Use MATLAB/Octave to find their eigenvalues, with multiplicity, and to find eigenvectors corresponding to each eigenvalue (2 d.p.). (The next Section 7.2 discusses that for repeated eigenvalues we generally want to record the so-called 'linearly independent' eigenvectors.)

(a) $\begin{bmatrix} 14 & 3 & -6 \\ 14 & -2 & -2 \\ 31 & 4 & -11 \end{bmatrix}$

(b) $\begin{bmatrix} -1 & 0 & 0 \\ 5 & 5 & 6 \\ -5 & -6 & -7 \end{bmatrix}$

(c) $\begin{bmatrix} -144 & -374 & 316 & 18 \\ 21 & 45 & -42 & 0 \\ -49 & -138 & 112 & 9 \\ 134 & 336 & -286 & -13 \end{bmatrix}$

(d) $\begin{bmatrix} -50 & 30 & 46 & -62 \\ 0 & 2 & 0 & 0 \\ -104 & 62 & 104 & -142 \\ -39 & 24 & 42 & -58 \end{bmatrix}$

(e) $\begin{bmatrix} 4 & -11 & -12 & 9 & -19 \\ 0 & 0 & 0 & 3 & -2 \\ 4 & 16 & 19 & -12 & 23 \\ 6 & 16 & 20 & -7 & 31 \\ -1 & -3 & -3 & 3 & 1 \end{bmatrix}$

v0.2a

(f) $\begin{bmatrix} 75 & 7 & -13 & -51 & -129 \\ 120 & 12 & -24 & -84 & -208 \\ 62 & 6 & -12 & -42 & -106 \\ -48 & -5 & 9 & 32 & 83 \\ 62 & 6 & -11 & -42 & -107 \end{bmatrix}$

Exercise 7.1.8. Consider the following three matrices which are perturbed versions of the matrix $\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$. Which perturbations show the high sensitivity of the repeated eigenvalue? Give reasons.

(a) $A = \begin{bmatrix} 2 & 1 \\ -0.0001 & 2 \end{bmatrix}$ (b) $B = \begin{bmatrix} 2 & 1.0001 \\ 0 & 2 \end{bmatrix}$

(c) $C = \begin{bmatrix} 2.0001 & 1 \\ 0 & 2 \end{bmatrix}$

Exercise 7.1.9. For each of the following matrices, use MATLAB/Octave to compute the eigenvalues, and to compute the eigenvalues of matrices obtained by adding random perturbations of size 0.0001 (use `randn`). Give reasons for which eigenvalues appear sensitive and which appear to be not sensitive.

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

(b) $B = \begin{bmatrix} 0 & -3 \\ 3 & 6 \end{bmatrix}$

(c) $C = \begin{bmatrix} -10 & -10 & -16 \\ 4 & 4 & 6 \\ 3 & 3 & 5 \end{bmatrix}$

(d) $D = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$

(e) $E = \begin{bmatrix} -1 & 1 & -1 \\ -6 & -6 & 2 \\ -5 & -3 & -3 \end{bmatrix}$

(f) $F = \begin{bmatrix} -6.7 & -0.6 & -6.6 & 3.6 \\ 3 & 0.1 & 3 & -2 \\ 2.8 & 0.6 & 2.7 & -1.6 \\ -6 & 0 & -6 & 3.1 \end{bmatrix}$

(g) $G = \begin{bmatrix} 1.4 & -7.1 & -0.7 & 6.2 \\ -7.1 & -1.0 & -2.2 & -2.5 \\ -0.7 & -2.2 & -3.4 & -4.1 \\ 6.2 & -2.5 & -4.1 & -1.0 \end{bmatrix}$

v0.2a

(h) $H = \begin{bmatrix} 0.6 & 0 & 0 & 0 \\ 9.6 & -1 & 33.6 & 17.6 \\ -9.6 & 1.6 & -33 & -17.6 \\ 19.2 & -3.2 & 67.2 & 35.8 \end{bmatrix}$

Exercise 7.1.10. For each of the following matrices, predict $\boldsymbol{y}(1)$, $\boldsymbol{y}(2)$ and $\boldsymbol{y}(3)$, for the given initial $\boldsymbol{y}(0)$, and given that $\boldsymbol{y}(t+1) = A\boldsymbol{y}(t)$.

(a) $\begin{bmatrix} 3 & 0 \\ 3 & 2 \end{bmatrix}$, $\boldsymbol{y}(0) = \begin{bmatrix} 6 \\ -1 \end{bmatrix}$

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

(c) $\begin{bmatrix} 26 & 21 \\ -28 & -23 \end{bmatrix}$, $\boldsymbol{y}(0) = \begin{bmatrix} 3 \\ -4 \end{bmatrix}$

(d) $\begin{bmatrix} -2 & 5 \\ -2 & 4 \end{bmatrix}$, $\boldsymbol{y}(0) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

(e) $\begin{bmatrix} 11 & -14 & -4 \\ 7 & -10 & -2 \\ 4 & -4 & -2 \end{bmatrix}$, $\boldsymbol{y}(0) = \begin{bmatrix} 0 \\ 2 \\ -2 \end{bmatrix}$

(f) $\begin{bmatrix} 9 & 7 & -5 \\ -16 & -8 & 8 \\ 10 & 10 & -6 \end{bmatrix}$, $\boldsymbol{y}(0) = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$

(g) $\begin{bmatrix} 2 & -2 & 0 \\ 4 & -2 & 0 \\ 0 & -5 & -1 \end{bmatrix}$, $\boldsymbol{y}(0) = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}$

(h) $\begin{bmatrix} -4 & 14 & 5 \\ 2 & -1 & 0 \\ -13 & 26 & 8 \end{bmatrix}$, $\boldsymbol{y}(0) = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix}$

Exercise 7.1.11. For each of the matrices of the previous Exercise 7.1.10, find a general solution of $\boldsymbol{y}(t+1) = A\boldsymbol{y}(t)$, if possible. Then use the corresponding given initial $\boldsymbol{y}(0)$ to find a formula for the specific $\boldsymbol{y}(t)$. Finally, check that the formula reproduces the values of $\boldsymbol{y}(1)$, $\boldsymbol{y}(2)$ and $\boldsymbol{y}(3)$ found in Exercise 7.1.10. Show your working.

Exercise 7.1.12. From the following partial description of the Tasmanian Devil, derive a mathematical model in the form $\boldsymbol{y}(t+1) = A\boldsymbol{y}(t)$ for the age structure of the Tasmanian Devil. By finding eigenvalues and an eigenvector, predict the long-term growth of the population, and predict the long-term relative numbers of Devils of various ages.

https://en.wikipedia.org/wiki/Tasmanian_devil

Devils are not monogamous, and their reproductive process is very robust and competitive. Males fight one another for the females, and then guard their partners to prevent female infidelity. Females can ovulate three

v02a

times in as many weeks during the mating season, and 80% of two-year-old females are seen to be pregnant during the annual mating season. Females average four breeding seasons in their life and give birth to 20–30 live young after three weeks' gestation. The newborn are pink, lack fur, have indistinct facial features and weigh around 0.20 g (0.0071 oz) at birth. As there are only four nipples in the pouch, competition is fierce and few newborns survive. The young grow rapidly and are ejected from the pouch after around 100 days, weighing roughly 200 g (7.1 oz). The young become independent after around nine months, so the female spends most of her year in activities related to birth and rearing.

Wikipedia, 2016

Exercise 7.1.13. From the following partial description of the elephant, derive a mathematical model in the form $\boldsymbol{y}(t+1) = A\boldsymbol{y}(t)$ for the age structure of the elephant. By finding eigenvalues and an eigenvector, predict the long-term growth of the population, and predict the long-term relative numbers of elephants of various ages.

https://en.wikipedia.org/wiki/Elephant

Gestation in elephants typically lasts around two years with interbirth intervals usually lasting four to five years. Births tend to take place during the wet season. Calves are born 85 cm (33 in) tall and weigh around 120 kg (260 lb). Typically, only a single young is born, but twins sometimes occur. The relatively long pregnancy is maintained by five corpus luteums (as opposed to one in most mammals) and gives the foetus more time to develop, particularly the brain and trunk. As such, newborn elephants are precocial and quickly stand and walk to follow their mother and family herd. A new calf is usually the centre of attention for herd members. Adults and most of the other young will gather around the newborn, touching and caressing it with their trunks. For the first few days, the mother is intolerant of other herd members near her young. Alloparenting—where a calf is cared for by someone other than its mother—takes place in some family groups. Allomothers are typically two to twelve years old. When a predator is near, the family group gathers together with the calves in the centre.

For the first few days, the newborn is unsteady on its feet, and needs the support of its mother. It relies on touch, smell and hearing, as its eyesight is poor. It has little precise control over its trunk, which wiggles around and may cause it to trip. By its second week of life, the calf can walk more firmly and has more control over its

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

trunk. After its first month, a calf can pick up, hold and put objects in its mouth, but cannot suck water through the trunk and must drink directly through the mouth. It is still dependent on its mother and keeps close to her.

For its first three months, a calf relies entirely on milk from its mother for nutrition after which it begins to forage for vegetation and can use its trunk to collect water. At the same time, improvements in lip and leg coordination occur. Calves continue to suckle at the same rate as before until their sixth month, after which they become more independent when feeding. By nine months, mouth, trunk and foot coordination is perfected. After a year, a calf's abilities to groom, drink, and feed itself are fully developed. It still needs its mother for nutrition and protection from predators for at least another year. Suckling bouts tend to last 2–4 min/hr for a calf younger than a year and it continues to suckle until it reaches three years of age or older. Suckling after two years may serve to maintain growth rate, body condition and reproductive ability. Play behaviour in calves differs between the sexes; females run or chase each other, while males play-fight. The former are sexually mature by the age of nine years while the latter become mature around 14–15 years. Adulthood starts at about 18 years of age in both sexes. Elephants have long lifespans, reaching 60–70 years of age.

Wikipedia, 2016

Exercise 7.1.14. From the following partial description of the giant mouse lemur, derive a mathematical model in the form $\boldsymbol{y}(t+1) = A\boldsymbol{y}(t)$ for the age structure of the giant mouse lemur. By finding eigenvalues and an eigenvector, predict the long-term growth of the population, and predict the long-term relative numbers of giant mouse lemurs of various ages.

https://en.wikipedia.org/wiki/Giant_mouse_lemur

Reproduction starts in November for Coquerel's giant mouse lemur at Kirindy Forest; the estrous cycle runs approximately 22 days, while estrus lasts only a day or less. ...

One to three offspring (typically two) are born after 90 days of gestation, weighing approximately 12 g (0.42 oz). Because they are poorly developed, they initially remain in their mother's nest for up to three weeks, being transported by mouth between nests. Once they have grown sufficiently, typically after three weeks, the mother will park her offspring in vegetation while she forages nearby.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

After a month, the young begin to participate in social play and grooming with their mother, and between the first and second month, young males begin to exhibit early sexual behaviors (including mounting, neck biting, and pelvic thrusting). By the third month, the young forage independently, though they maintain vocal contact with their mother and use a small part of her range.

Females start reproducing after ten months, while males develop functional testicles by their second mating season. Testicle size in the northern giant mouse lemur does not appear to fluctuate by season, and is so large relative to the animal's body mass that it is the highest among all primates. This emphasis on sperm production in males, as well as the use of copulatory plugs, suggests a mating system best described as polygynandrous where males use scramble competition (roaming widely to find many females). In contrast, male Coquerel's giant mouse lemurs appear to fight for access to females (contest competition) during their breeding season. Males disperse from their natal range, and the age at which they leave varies from two years to several. Females reproduce every year, although postpartum estrus has been observed in captivity. In the wild, the lifespan of giant mouse lemurs is thought to rarely exceed five or six years *Wikipedia, 2016*

Exercise 7.1.15. From the following partial description of the dolphin (Indo-Pacific bottlenose dolphin), derive a mathematical model in the form $\boldsymbol{y}(t+1) = A\boldsymbol{y}(t)$ for the age structure of the dolphin. (Assume only one calf is born at a time.) By finding eigenvalues and an eigenvector, predict the long-term growth of the population, and predict the long-term relative numbers of dolphins of various ages.

https://en.wikipedia.org/wiki/Indo-Pacific_bottlenose_dolphin

Indo-Pacific bottlenose dolphins live in groups that can number in the hundreds, but groups of five to 15 dolphins are most common. In some parts of their range, they associate with the common bottlenose dolphin and other dolphin species, such as the humpback dolphin.

The peak mating and calving seasons are in the spring and summer, although mating and calving occur throughout the year in some regions. Gestation period is about 12 months. Calves are between 0.84 and 1.5 metres (2.8 and 4.9 ft) long, and weigh between 9 and 21 kilograms (20 and 46 lb). The calves are weaned between 1.5 and two years, but can remain with their mothers for up to

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

five years. The interbirth interval for females is typically four to six years.

In some parts of its range, this dolphin is subject to predation by sharks; its life span is more than 40 years.

Wikipedia, 2016

Exercise 7.1.16. You are given that a mathematical model of the age structure of some animal population is

$$\begin{aligned} y_1(t+1) &= 0.5y_1(t) + y_3(t), \\ y_2(t+1) &= 0.5y_1(t) + 0.7y_2(t), \\ y_3(t+1) &= 0.3y_2(t) + 0.9y_3(t). \end{aligned}$$

Invent an animal species, and time scale, and create details of a plausible scenario for the breeding and life cycle of the species that could lead to this mathematical model. Write a coherent paragraph about the breeding and life cycle of the species with enough information that someone could deduce this mathematical model from your description. Be creative.

Exercise 7.1.17. For each of the following matrices, say A for instance, find by hand calculation the eigenvalues and eigenvectors of the larger matrix $\begin{bmatrix} O & A \\ A^{\mathrm{T}} & O \end{bmatrix}$. Show your working. Relate these to an SVD of the matrix A.

(a) $A = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$

(b) $B = \begin{bmatrix} -5 & 12 \end{bmatrix}$

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

(d) $D = \begin{bmatrix} 0 & 1 \\ -4 & 0 \end{bmatrix}$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

7.2 Linear independent vectors may form a basis

Section Contents

7.2.1 Linearly (in)dependent sets 640
7.2.2 Form a basis for subspaces 651
Revisit unique solutions 665
7.2.3 Exercises 666

In Chapter 4 on symmetric matrices, the eigenvectors from distinct eigenvalues are orthogonal. 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 a similarly useful analogue of orthogonality? Answer: yes; we now extend "orthogonal" with the more general concept of "linear independence" which for problems with general matrices replaces orthonormality.

One reason that orthogonal vectors are useful is that they can form an orthonormal basis and hence act as the unit vectors of an orthogonal coordinate systems. The concept of linear independence is closely connected to coordinate systems which are not orthogonal.

Subspace coordinate systems In any given problem we want two things from a general solution:

- firstly, the general solution must encompass every possibility (the solution must span the possibilities); and
- secondly, each possible solution should have a unique algebraic description in the general solution.

For an example of the need for a unique algebraic form, let's suppose we wanted to find solutions to the differential equation $d^2y/dt^2 - y = 0$. You might find $y = 3e^x + 2e^{-x}$, whereas I find $y = 5\cosh x + \sinh x$, and a friend finds $y = e^x + 4\cosh x$. By looking at these disparate algebraic forms it is apparent that we all disagree. Should we all go and search for errors in the solution process? No. The reason is that all these solutions are the same. The apparent differences arise only because you choose exponentials to represent the solution, whereas I choose hyperbolic functions: the solutions are the same, it is only the algebraic representation that appears different. In general, when we cannot immediately distinguish identical solutions, all algebraic manipulation becomes immensely more difficult due to algebraic ambiguity. To avoid such immense difficulties, in both calculus and linear algebra, we need to introduce the concept of linear independence.

Linear independence empowers us, often implicitly, to use a non-orthogonal coordinate system in a subspace. We replace the or-

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

thonormal standard unit vectors by any suitable set of basis vectors. For example, in the plane any two vectors at an angle to each other suffice to be able to describe uniquely every vector (point) in the plane. As illustrated in the margin, every point in the plane (end point of a vector) is a unique linear combination of the two basis vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$. Such a pair of basis vectors, termed linearly independent, avoid the immense difficulties of algebraic ambiguity.

7.2.1 Linearly (in)dependent sets

This section defines linear (in)dependence, 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 in the margin.

Solution: Let's start with some specific example vectors.

(a) The vector $(0\,,2)$ may be written as the linear combination $(0\,,2) = -\frac{2}{3}\boldsymbol{v}_1 + \frac{4}{3}\boldsymbol{v}_2$ as shown.

(b) The vector $(2\,,2)$ may be written as the linear combination $(2\,,2) = \frac{2}{3}\boldsymbol{v}_1 + \frac{2}{3}\boldsymbol{v}_2$ as shown.

(c) The vector $(1\,,-1)$ may be written as the linear combination $(1\,,-1) = \boldsymbol{v}_1 - \boldsymbol{v}_2$ as shown.

(d) The vector $(-3\,,-3)$ may be written as the linear combination $(-3\,,-3) = -\boldsymbol{v}_1 - \boldsymbol{v}_2\,.$

Now proceed to consider a general vector (x, y) and seek it as a linear combination of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$, namely $(x, y) = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2\,.$ That is, let's write each and every point in the plane as a linear combination of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ as illustrated in the margin. Rewrite the

v0.2a

equation in matrix-vector form as

$$\begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2\end{bmatrix}\begin{bmatrix}c_1\\c_2\end{bmatrix}=\begin{bmatrix}x\\y\end{bmatrix},\quad \text{that is, } V\boldsymbol{c}=\begin{bmatrix}x\\y\end{bmatrix}\quad \text{for } V=\begin{bmatrix}2&1\\1&2\end{bmatrix}.$$

For any given (x,y), $V\boldsymbol{c}=(x,y)$ is a system of linear equations for the coefficients $\boldsymbol{c}$. Theorem 3.4.43 asserts the system has a unique solution $\boldsymbol{c}$ if and only iff the matrix V is invertible. Here the unique solution is then that the vector of coefficients

$$\boldsymbol{c}=V^{-1}\begin{bmatrix}x\\y\end{bmatrix}=\begin{bmatrix}\frac{2}{3}&-\frac{1}{3}\\-\frac{1}{3}&\frac{2}{3}\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}.$$

Equivalently, Theorem 3.4.43c asserts the system has a unique solution $\boldsymbol{c}$—unique coefficients $\boldsymbol{c}$—if and only iff the homogeneous system $V\boldsymbol{c}=\boldsymbol{0}$ has only the zero solution $\boldsymbol{c}=\boldsymbol{0}$. It is this last statement that leads to the upcoming Definition 7.2.3 of linear independence.

■

Example 7.2.2 (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 below in stereo. The solution here looks at the different issue of unique representation.

Solution: As one example, consider the vector $(1,0,-1)$:

$$\begin{aligned}(1,0,-1)&=-1\boldsymbol{v}_1+0\boldsymbol{v}_2+1\boldsymbol{v}_3\,;\\(1,0,-1)&=1\boldsymbol{v}_1+2\boldsymbol{v}_2+3\boldsymbol{v}_3\,;\\(1,0,-1)&=-2\boldsymbol{v}_1-1\boldsymbol{v}_2+0\boldsymbol{v}_3\,;\\(1,0,-1)&=(-1+t)\boldsymbol{v}_1+t\boldsymbol{v}_2+(1+t)\boldsymbol{v}_3\,,\quad \text{for any } t.\end{aligned}$$

This last combination shows there an infinite number of ways to write $(1,0,-1)$ as a linear combination of $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$. Such an infinity of linear combinations means that $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$ do not form a useful 'coordinate system' because we cannot easily distinguish between the different combinations all giving the same vector. The reason for the infinity of combinations is that there

v0.2a

is a nontrivial linear combination of $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$ which is zero, namely $\boldsymbol{v}_1+\boldsymbol{v}_2+\boldsymbol{v}_3=\boldsymbol{0}$. It is this last statement that leads to the Definition 7.2.3 of linear dependence.

■

Definition 7.2.3. *A set of vectors $\{\boldsymbol{v}_1,\boldsymbol{v}_2,\dots,\boldsymbol{v}_k\}$ is **linearly dependent** if there are scalars $c_1,c_2,\dots,c_k$, at least one of which is nonzero, such that $c_1\boldsymbol{v}_1+c_2\boldsymbol{v}_2+\cdots+c_k\boldsymbol{v}_k=\boldsymbol{0}$. A set of vectors that is not linearly dependent is called **linearly independent** (characterised 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 7.2.4. Are the following sets of vectors linearly dependent or linearly independent. Give reasons.

(a) $\{(-1,1,0),\ (1,-2,1),\ (0,1,-1)\}$

Solution: The set is linearly dependent as the linear combination $(-1,1,0)+(1,-2,1)+(0,1,-1)=(0,0,0)$.

(b) $\{(2,1),\ (1,2)\}$

Solution: The set is linearly independent because the linear combination equation $c_1(2,1)+c_2(1,2)=(0,0)$ is equivalent to the homogeneous matrix-vector system $\begin{bmatrix}2&1\\1&2\end{bmatrix}\boldsymbol{c}=\boldsymbol{0}$ which has *only* the zero solution $\boldsymbol{c}=\boldsymbol{0}$.

(c) $\{(-2,4,1,-1,0)\}$

Solution: This set of one vector in $\mathbb{R}^5$ is linearly independent as $c_1(-2,4,1,-1,0)=\boldsymbol{0}$ can only be satisfied with $c_1=0$.

Indeed, any one non-zero vector $\boldsymbol{v}$ in $\mathbb{R}^n$ forms a linearly independent set, $\{\boldsymbol{v}\}$, for the same reason.

(d) $\{(2,1),\ (0,0)\}$

Solution: The set is linearly dependent because the linear combination $0(2,1)+c_2(0,0)=(0,0)$ for any non-zero c_2.

(e) $\{\boldsymbol{0},\boldsymbol{v}_2,\boldsymbol{v}_3,\dots,\boldsymbol{v}_k\}$

Solution: Any set that includes the zero vector is linearly dependent as $c_1\boldsymbol{0}+0\boldsymbol{v}_2+\cdots+0\boldsymbol{v}_k=\boldsymbol{0}$ for any non-zero c_1.

(f) $\{\boldsymbol{e}_1,\boldsymbol{e}_2,\boldsymbol{e}_3\}$, the set of standard unit vectors in $\mathbb{R}^3$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Solution: This set is linearly independent as $c_1\boldsymbol{e}_1 + c_2\boldsymbol{e}_2 + c_3\boldsymbol{e}_3 = (c_1, c_2, c_3) = \mathbf{0}$ *only* when all three components are zero, $c_1 = c_2 = c_3 = 0$.

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

Solution: This set is linearly independent. Seek some linear combination $c_1(\frac{1}{3}, \frac{2}{3}, \frac{2}{3}) + c_2(\frac{2}{3}, \frac{1}{3}, -\frac{2}{3}) = \mathbf{0}$. Take the dot product of both sides of this equation with $(1, 2, 2)$:

$$\begin{aligned} & c_1 \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} + c_2 \begin{bmatrix} \frac{2}{3} \\ \frac{1}{3} \\ -\frac{2}{3} \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = \mathbf{0} \cdot \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \\ \implies\ & c_1 3 + c_2 0 = 0 \\ \implies\ & c_1 = 0\,. \end{aligned}$$

Similarly, take the dot product with $(2, 1, -2)$:

$$\begin{aligned} & c_1 \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix} + c_2 \begin{bmatrix} \frac{2}{3} \\ \frac{1}{3} \\ -\frac{2}{3} \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix} = \mathbf{0} \cdot \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix} \\ \implies\ & c_1 0 + c_2 3 = 0 \\ \implies\ & c_2 = 0\,. \end{aligned}$$

Hence $c_1 = c_2 = 0$ and so the vectors are linearly independent.

These last two cases generalise to the next Theorem 7.2.6 about the linear independence of any orthonormal set of vectors. ■

Example 7.2.5 (calculus extension)**.** In calculus the notion of a function corresponds precisely to the notion of a vector in our linear algebra. For the purposes of this example, consider 'vector' and 'function' to be synonymous, and that 'all components' and 'all x' are synonymous. Show that the set $\{e^x, e^{-x}, \cosh x, \sinh x\}$ is linearly dependent. What is a subset that is linearly independent?

Solution: The definition of the hyperbolic functions, namely that $\cosh x = (e^x + e^{-x})/2$ and $\sinh x = (e^x - e^{-x})/2$, immediately give two nontrivial linear combinations that are zero for all x, namely $2\cosh x - e^x - e^{-x} = 0$ and $2\sinh x - e^x + e^{-x} = 0$ for all x. Either one of these implies the set $\{e^x, e^{-x}, \cosh x, \sinh x\}$ is linearly dependent.

Because e^x and e^{-x} are not proportional to each other, there is no linear combination which is zero for all x, and hence the set $\{e^x, e^{-x}\}$ is linearly independent (as are any other pairs of the four functions).

■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Theorem 7.2.6. *Any orthonormal set (Definition 3.2.38) of vectors is linearly independent.*

Proof. Let $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k\}$ be an orthonormal set of vectors in $\mathbb{R}^n$. Let's find all possible scalars $c_1, c_2, \dots, c_k$ such that $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k = \mathbf{0}$. Taking the dot product of this equation with $\boldsymbol{v}_1$ implies $c_1\boldsymbol{v}_1\cdot\boldsymbol{v}_1 + c_2\boldsymbol{v}_2\cdot\boldsymbol{v}_1 + \cdots + c_k\boldsymbol{v}_k\cdot\boldsymbol{v}_1 = \mathbf{0}\cdot\boldsymbol{v}_1$; by orthonormality this identity becomes $c_1 1 + c_2 0 + \cdots + c_k 0 = 0$; that is, $c_1 = 0$. Similarly, taking the dot product with $\boldsymbol{v}_2$ implies $c_1\boldsymbol{v}_1\cdot\boldsymbol{v}_2 + c_2\boldsymbol{v}_2\cdot\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k\cdot\boldsymbol{v}_2 = \mathbf{0}\cdot\boldsymbol{v}_2$; by orthonormality this identity becomes $c_1 0 + c_2 1 + \cdots + c_k 0 = 0$; that is, $c_2 = 0$. And so on for all vectors in the set, implying the coefficients $c_1 = c_2 = \cdots = c_k = 0$ is the only possibility. By Definition 7.2.3, the orthonormal set must be linearly independent. □

In contrast to orthonormal vectors which are always linearly independent, a set of two vectors proportional to each other is always linearly dependent as seen in the following examples. This linear dependence of proportional vectors then generalises in the next Theorem 7.2.8.

Example 7.2.7. Show the following sets are linearly dependent.

(a) $\{(1,2),\ (3,6)\}$

Solution: Since $(3,6) = 3(1,2)$ then the linear combination $1(3,6) - 3(1,2) = \mathbf{0}$ and the set is linearly dependent.

(b) $\{(2.2,-2.1,0,1.5),\ (-8.8,8.4,0,-6)\}$

Solution: Since $(-8.8,8.4,0,-6) = -4(2.2,-2.1,0,1.5)$ then the linear combination

$$(-8.8,8.4,0,-6) + 4(2.2,-2.1,0,1.5) = \mathbf{0}\,,$$

and so the set is linearly dependent.

■

Theorem 7.2.8. *A set of vectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_m\}$ 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 multiple of the other.*

Proof. Exercise 7.2.3 establishes the particular case of a set of two vectors.

In the general case of m vectors, first establish that if one of the vectors can be expressed as a linear combination of the others, then the set is linearly dependent. Suppose we have labelled the set

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

of vectors so that it is vector $\boldsymbol{v}_1$ which is a linear combination of the others; that is, $\boldsymbol{v}_1 = c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 + \cdots + c_m\boldsymbol{v}_m$. Rearranging, $(-1)\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 + \cdots + c_m\boldsymbol{v}_m = \mathbf{0}$; that is, there is a non-trivial (as at least $c_1 = -1 \neq 0$) linear combination of the set of vectors which is zero. Hence the set is linearly dependent.

Second, establish the converse. Given the set is linearly dependent, there exist coefficients, not all zero, such that $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_m\boldsymbol{v}_m = \mathbf{0}$. Suppose that we have labelled the vectors so that $c_1 \neq 0$. Then rearranging the equation gives $c_1\boldsymbol{v}_1 = -c_2\boldsymbol{v}_2 - c_3\boldsymbol{v}_3 - \cdots - c_m\boldsymbol{v}_m$. Divide by the non-zero c_1 to deduce $\boldsymbol{v}_1 = -(c_2/c_1)\boldsymbol{v}_2 - (c_3/c_1)\boldsymbol{v}_3 - \cdots - (c_m/c_1)\boldsymbol{v}_m$; that is, $\boldsymbol{v}_1$ is a linear combination of the other vectors. □

Example 7.2.9. Invoke Theorem 7.2.8 to deduce whether the following sets are linearly independent or linearly dependent.

(a) $\{(-1,1,0),\ (1,-2,1),\ (0,1,-1)\}$

Solution: Since $(1,-2,1) = -(-1,1,0) - (0,1,-1)$ the set must be linearly dependent.

(b) The set of two vectors shown in the margin.

Solution: Since they are not proportional to each other, we cannot write either as a multiple of the other, and so the pair are linearly independent.

(c) The set of two vectors shown in the margin.

Solution: Since they appear proportional to each other, $\boldsymbol{v}_2 \approx (-3)\boldsymbol{v}_1$, so the pair appear linearly dependent.

(d) $\{(1,3,0,-1),\ (1,0,-4,2),\ (-2,3,0,-3),\ (0,6,-4,-2)\}$

Solution: Notice that the last vector is the sum of the first three, $(0,6,-4,-2) = (1,3,0,-1) + (1,0,-4,2) + (-2,3,0,-3)$, and so the set is linearly dependent.

■

Recall that Theorem 4.2.10 established that for every two distinct eigenvalues of a symmetric matrix A, any corresponding two eigenvectors are orthogonal. Consequently, for a symmetric A, a set of eigenvectors from distinct eigenvalues forms an orthogonal set. The following Theorem 7.2.10 generalises this property to non-symmetric matrices using the concept of linear independence.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Theorem 7.2.10. *For an $n \times n$ matrix A, let $\lambda_1, \lambda_2, \dots, \lambda_m$ be distinct eigenvalues of A with corresponding eigenvectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_m$. The set $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_m\}$ is linearly independent.*

Proof. Use contradiction. Assume the set $\{\boldsymbol{v}_1 , \boldsymbol{v}_2 , \dots , \boldsymbol{v}_m\}$ is linearly dependent. Choose $k < m$ such that the set $\{\boldsymbol{v}_1 , \boldsymbol{v}_2 , \dots , \boldsymbol{v}_k\}$ is linearly independent, whereas the set $\{\boldsymbol{v}_1 , \boldsymbol{v}_2 , \dots , \boldsymbol{v}_{k+1}\}$ is linearly dependent. Hence there exists non-trivial coefficients such that

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k + c_{k+1}\boldsymbol{v}_{k+1} = \boldsymbol{0}\,;$$

further, $c_{k+1} \neq 0$ as $\{\boldsymbol{v}_1 , \boldsymbol{v}_2 , \dots , \boldsymbol{v}_k\}$ is linearly independent. Pre-multiply the linear combination by matrix A:

$$\begin{aligned} & c_1A\boldsymbol{v}_1 + c_2A\boldsymbol{v}_2 + \cdots + c_kA\boldsymbol{v}_k + c_{k+1}A\boldsymbol{v}_{k+1} = A\boldsymbol{0} \\ \implies\ & c_1\lambda_1\boldsymbol{v}_1 + c_2\lambda_2\boldsymbol{v}_2 + \cdots + c_k\lambda_k\boldsymbol{v}_k + c_{k+1}\lambda_{k+1}\boldsymbol{v}_{k+1} = \boldsymbol{0}\,. \end{aligned}$$

Now subtract $\lambda_{k+1}\times$ the original linear combination:

$$\begin{aligned} & c_1\lambda_1\boldsymbol{v}_1 + c_2\lambda_2\boldsymbol{v}_2 + \cdots + c_k\lambda_k\boldsymbol{v}_k + c_{k+1}\lambda_{k+1}\boldsymbol{v}_{k+1} \\ & \quad - (c_1\lambda_{k+1}\boldsymbol{v}_1 + c_2\lambda_{k+1}\boldsymbol{v}_2 + \cdots + c_k\lambda_{k+1}\boldsymbol{v}_k + c_{k+1}\lambda_{k+1}\boldsymbol{v}_{k+1}) = \boldsymbol{0} \\ \implies\ & c_1(\lambda_1 - \lambda_{k+1})\boldsymbol{v}_1 + c_2(\lambda_2 - \lambda_{k+1})\boldsymbol{v}_2 + \cdots + c_k(\lambda_k - \lambda_{k+1})\boldsymbol{v}_k = \boldsymbol{0} \\ \implies\ & c_1'\boldsymbol{v}_1 + c_2'\boldsymbol{v}_2 + \cdots + c_k'\boldsymbol{v}_k = \boldsymbol{0} \end{aligned}$$

for coefficients $c_j' = c_j(\lambda_j - \lambda_{k+1})$. Since all the eigenvalues are distinct, $\lambda_j - \lambda_{k+1} \neq 0$, and since the coefficients c_j are not all zero, hence c_j' are not all zero. Thus we have created a non-trivial linear combination of $\boldsymbol{v}_1 , \boldsymbol{v}_2 , \dots , \boldsymbol{v}_k$ which is zero, and so the set $\{\boldsymbol{v}_1 , \boldsymbol{v}_2 , \dots , \boldsymbol{v}_k\}$ is linearly dependent. This contradiction of the choice of k proves the assumption must be wrong. Hence the set $\{\boldsymbol{v}_1 , \boldsymbol{v}_2 , \dots , \boldsymbol{v}_m\}$ is linearly independent, as required. □

Example 7.2.11. For each of the following matrices, show the eigenvectors from distinct eigenvalues form linearly independent sets.

(a) Recall from Example 7.1.9 the matrix $B = \begin{bmatrix} -1 & 1 & -2 \\ -1 & 0 & -1 \\ 0 & -3 & 1 \end{bmatrix}$

Solution: In MATLAB/Octave, executing

```
B=[-1 1 -2
 -1 0 -1
 0 -3 1]
[V,D]=eig(B)
```

gives eigenvectors and corresponding eigenvalues in

```
V =
    -0.5774     0.7071    -0.7071
    -0.5774     0.0000     0.0000
```

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

```
    -0.5774    -0.7071     0.7071
D =
          -2          0          0
           0          1          0
           0          0          1
```

Recognising $0.7071 = 1/\sqrt{2}$, the last two eigenvectors, $(1/\sqrt{2}, 0, -1/\sqrt{2})$ and $(-1/\sqrt{2}, 0, 1/\sqrt{2})$, form a linearly dependent set because they are proportional to each other. This linear dependence does not confound Theorem 7.2.12 because the corresponding eigenvalues are the same, not distinct, namely $\lambda = 1$. The theorem only applies to eigenvectors of distinct eigenvalues.

Here the two distinct eigenvalues are $\lambda = -2$ and $\lambda = 1$. Recognising $0.5774 = 1/\sqrt{3}$, two corresponding eigenvectors are $(-1/\sqrt{3}, -1/\sqrt{3}, -1/\sqrt{3})$ and $(1/\sqrt{2}, 0, -1/\sqrt{2})$. Because of the zero component in the second, these cannot be proportional to each other, and so the pair form a linearly independent set.

(b) Example 7.1.10 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
```

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

The five eigenvalues are all distinct, so Theorem 7.2.12 asserts a set of corresponding eigenvectors will be linearly independent. The five columns of V, call them $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_5$, 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 $\boldsymbol{c} = (c_1, c_2, \dots, c_5)$ such that $V\boldsymbol{c} = \mathbf{0}$. Because the five singular values of square matrix V are all non-zero,[5] obtained from `svd(V)` as

```
ans =
   1.7703
   1.1268
   0.6542
   0.3625
   0.1922
```

consequently Theorem 3.4.43 asserts $V\boldsymbol{c} = \mathbf{0}$ has only the zero solution. Hence, by Definition 7.2.3 the set of eigenvectors in the columns of V are linearly independent.

■

This last case of Example 7.2.11b connects the concept of linear (in)dependence to the existence or otherwise of non-zero solutions to a homogeneous system of linear equations, $V\boldsymbol{c} = \mathbf{0}$. So does Example 7.2.4b. The great utility of this connection is that we understand a lot about homogeneous systems of linear equations. The next Theorem 7.2.12 establishes this connection in general.

Theorem 7.2.12. *Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_m$ be vectors in $\mathbb{R}^n$. Let $n \times m$ matrix $V = \begin{bmatrix}\boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_m\end{bmatrix}$. Then $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_m\}$ is linearly dependent if and only if $V\boldsymbol{c} = \mathbf{0}$ has a nonzero solution $\boldsymbol{c}$.*

Proof. Now $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_m\}$ is linearly dependent if and only if there are scalars, not all zero, such that the equation $c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_m\boldsymbol{v}_m = \mathbf{0}$ holds (Definition 7.2.3). Let the vector $\boldsymbol{c} = (c_1, c_2, \dots, c_m)$, then this equation is equivalent to the statement $V\boldsymbol{c} = \mathbf{0}$. That is, if and only if $V\boldsymbol{c} = \mathbf{0}$ has a nonzero solution. □

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 generalisation: in $\mathbb{R}^n$ there can be no more than n vectors in a linearly independent set of vectors.

[5] One could alternatively compute the determinant `det(V)` = 0.09090 and because it is non-zero Theorem 7.2.31 asserts that the equation has only the solution $\boldsymbol{c} = \mathbf{0}$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Theorem 7.2.13. *Any set of m vectors in $\mathbb{R}^n$ is linearly dependent when the number of vectors $m > n$.*

Proof. Form the m vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_m \in \mathbb{R}^n$ into the $n \times m$ matrix $V = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_m \end{bmatrix}$. Consider the homogeneous system $V\boldsymbol{c} = \boldsymbol{0}$: as $m > n$, Theorem 2.2.31 (with the meaning of m and n swapped) asserts $V\boldsymbol{c} = \boldsymbol{0}$ has infinitely many solutions. Thus $V\boldsymbol{c} = \boldsymbol{0}$ has nonzero solutions, so Theorem 7.2.12 implies the set of eigenvectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_m\}$ is linearly dependent. □

Example 7.2.14. Determine if the following sets of vectors are linearly dependent or independent. Give reasons.

(a) $\{(-1,-2),(-1,4),(0,5),(2,3)\}$

Solution: As there are four vectors in $\mathbb{R}^2$ so Theorem 7.2.13 asserts the set is linearly dependent.

(b) $\{(-6,-4,-1,-2),\ (2,0,1,-2),\ (2,-1,-1,1)\}$

Solution: In MATLAB/Octave form the matrix with these vectors as columns

```
V=[-6   2   2
  -4   0  -1
  -1   1  -1
  -2  -2   1]
svd(V)
```

and find the three singular values are all non-zero (namely 7.7568, 2.7474, and 2.2988). Hence there are no free variables when solving $V\boldsymbol{c} = \boldsymbol{0}$ (Procedure 3.3.15), and consequently there is only the unique solution $\boldsymbol{c} = \boldsymbol{0}$. By Theorem 7.2.12, the set of vectors is linearly independent.

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

Solution: By inspection, the third vector is twice the first. Hence the linear combination $2(-1,-2,2,-1)+0(1,3,1,-1)-(-2,-4,4,-2) = \boldsymbol{0}$ and so the set of vectors is linearly dependent.

(d) $\{(3,3,-1,-1),\ (0,-3,-1,-7),\ (1,2,0,2)\}$

Solution: In MATLAB/Octave form the matrix with these vectors as columns

```
V=[3   0   1
   3  -3   2
  -1  -1   0
  -1  -7   2]
svd(V)
```

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

and find the three singular values are 8.1393, 4.6638, and 0.0000. The zero singular value implies there is a free variables when solving $V\boldsymbol{c} = \mathbf{0}$ (Procedure 3.3.15), and consequently there are infinitely many non-zero $\boldsymbol{c}$ that solve $V\boldsymbol{c} = \mathbf{0}$. By Theorem 7.2.12, the set of vectors is linearly dependent.

(e) $\{(10, 3, 3, 1), (2, -3, 0, -1), (1, -1, 2, -1), (2, -1, -3, 0), (-2, 0, 2, -1)\}$

Solution: As there are five vectors in $\mathbb{R}^4$ so Theorem 7.2.13 asserts the set is linearly dependent.

(f) $\{(-0.4, -1.8, -0.2, 0.7, -0.2), (-1.1, 2.8, 2.7, -3.0, -2.6), (-2.3, -2.3, 4.1, 3.4,$
$(-2.6, -5.3, -3.3, -1.3, -4.1), (1.4, 5.2, -6.9, -0.7, 0.6)\}$

Solution: In MATLAB/Octave form the matrix V with these vectors as columns

```
V=[-0.4 -1.1 -2.3 -2.6 1.4
  -1.8  2.8 -2.3 -5.3  5.2
  -0.2  2.7  4.1 -3.3 -6.9
   0.7 -3.0  3.4 -1.3 -0.7
  -0.2 -2.6 -1.6 -4.1  0.6]
svd(V)
```

and find the five singular values are 10.6978, 8.0250, 5.5920, 3.0277 and 0.0024. As the singular values are all non-zero, the homogeneous system $V\boldsymbol{c} = \mathbf{0}$ has the unique solution $\boldsymbol{c} = \mathbf{0}$ (Procedure 3.3.15), and hence the set of five vectors are linearly independent.

However, the answer depends upon the context. In the strict mathematical context the vectors are unequivocally linearly independent. But in the context of practical problems, where errors in matrix entries are likely, there are 'shades of grey' . Here, one of the singular values is quite small, namely 0.0024. If the context informs us that the entries in the matrix had errors of say 0.01, then this singular value is effectively zero (Section 5.2). In the context of such errors, this set of five vectors would be effectively linearly dependent.

■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

7.2.2 Form a basis for subspaces

Recall from Sections 2.3 and 3.4 the definition of subspaces and the span: namely that a subspace is a set of vectors closed under addition and scalar multiplication; and a span is a set of vectors whose linear combinations give all vectors in a given subspace. Also, Definition 3.4.18 defined an "orthonormal basis" for a subspace to be a set of orthonormal vectors that span a subspace. This section generalises orthonormal basis by relaxing the requirement of orthonormality.

Definition 7.2.15. *A **basis** for a subspace $\mathbb{W}$ of $\mathbb{R}^n$ is a set of vectors that both span $\mathbb{W}$ and is linearly independent.*

Example 7.2.16. (a) Recall Examples 7.2.4b and 7.2.1 showed that the two vectors $(2,1)$ and $(1,2)$ are linearly independent and span $\mathbb{R}^2$. Hence the set $\{(2,1),(1,2)\}$ is a basis of $\mathbb{R}^2$.

(b) Recall that Example 7.2.4a showed the set $\{(-1,1,0),(1,-2,1),(0,1,-1)\}$ is linearly dependent so it cannot be a basis.

However, remove one vector, such as the middle one, and consider the set $\{(-1,1,0),(0,1,-1)\}$. As the two vectors are not proportional to each other, this set is linearly independent (Theorem 7.2.8). Also, the plane $x+y+z=0$ is a subspace, say $\mathbb{W}$. It is characterised by $y=-x-z$. So every vector in $\mathbb{W}$ can be written as $(x,-x-z,z)=(x,-x,0)+(0,-z,z)=(-x)(-1,1,0)+(-z)(0,1,-1)$. That is, $\operatorname{span}\{(-1,1,0),(0,1,-1)\}=\mathbb{W}$. 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 below in stereo).

Solution: The vectors in the line may be written as $\boldsymbol{x}=(x,y,z)=(2.1t,1.3t,-1.1t)=(2.1,1.3,-1.1)t$. Since the parameter t may vary over all values, vectors in the line form $\operatorname{span}\{(2.1,1.3,-1.1)\}$. Since $\{(2.1,1.3,-1.1)\}$ is a linearly independent set of vectors (Example 7.2.4c), it thus forms a basis for the vectors in the given line.

(d) Find a basis for the line given parametrically as $x=5.7t-0.6$ and $y=6.8t+2.4$.

Solution: The vectors in the line may be written as $\boldsymbol{x} = (5.7t - 0.6, 6.8t + 2.4)$. But this does not form a subspace as it does not include the zero vector $\mathbf{0}$ (as illustrated in the margin): the x-component is zero for positive t whereas the y-component is zero for negative t so they are never zero for the same value of parameter t. Since this line is not a subspace, it cannot have a basis.

(e) Find a basis for the plane $3x - 2y + z = 0$.

Solution: Writing the equation of the plane as $z = -3x + 2y$ we then write the plane parametrically (section 1.3.4) as the vectors $\boldsymbol{x} = (x, y, -3x + 2y) = (x, 0, -3x) + (0, y, 2y) = x(1, 0, -3) + y(0, 1, 2)$. Since x and y may vary over all values, the plane is the subspace $\operatorname{span}\{(1, 0, -3), (0, 1, 2)\}$ (as illustrated below in stereo). Since $(1, 0, -3)$ and $(0, 1, 2)$ are not proportional to each other, they form a linearly independent set. Hence $\{(1, 0, -3), (0, 1, 2)\}$ is a basis for the plane.

(f) Prove that every orthonormal basis of a subspace $\mathbb{W}$ is also a basis of $\mathbb{W}$.

Solution: Theorem 7.2.6 establishes that any orthonormal basis is linearly independent. By Definition 3.4.18, an orthonormal basis of $\mathbb{W}$ spans $\mathbb{W}$. Hence an orthonormal basis of $\mathbb{W}$ is also a basis of $\mathbb{W}$.

■

Recall that Theorem 3.4.28 establishes that an orthonormal basis of a given subspace always has the same number of vectors. The following theorem establishes the same is true for general bases. The proof is direct generalisation of that for Theorem 3.4.28.

Theorem 7.2.17. *Any two bases for a given subspace have the same number of vectors.*

Proof. Let $\mathcal{U} = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_r\}$ and $\mathcal{V} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_s\}$ be any two bases for a subspace in $\mathbb{R}^n$. Prove the number of vectors $r = s$ by contradiction. In the first case, assume $r < s$. Since $\mathcal{U}$ is a basis for the subspace every vector in the set $\mathcal{V}$ can be written as a linear combination of vectors in $\mathcal{U}$ with some coefficients a_{ij}:

$$\boldsymbol{v}_1 = \boldsymbol{u}_1 a_{11} + \boldsymbol{u}_2 a_{21} + \cdots + \boldsymbol{u}_r a_{r1} ,$$

$$\begin{aligned}
\boldsymbol{v}_2 &= \boldsymbol{u}_1 a_{12} + \boldsymbol{u}_2 a_{22} + \cdots + \boldsymbol{u}_r a_{r2} \,, \\
&\vdots \\
\boldsymbol{v}_s &= \boldsymbol{u}_1 a_{1s} + \boldsymbol{u}_2 a_{2s} + \cdots + \boldsymbol{u}_r a_{rs} \,.
\end{aligned}$$

Write each of these, such as the first one, in the form

$$\boldsymbol{v}_1 = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_r \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{r1} \end{bmatrix} = U\boldsymbol{a}_1 \,,$$

where $n \times r$ matrix $U = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_r \end{bmatrix}$. Similarly for the other equations $\boldsymbol{v}_2 = \cdots = U\boldsymbol{a}_2$ through to $\boldsymbol{v}_s = \cdots = U\boldsymbol{a}_s$. Then the $n \times s$ matrix

$$\begin{aligned}
V &= \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_s \end{bmatrix} \\
&= \begin{bmatrix} U\boldsymbol{a}_1 & U\boldsymbol{a}_2 & \cdots & U\boldsymbol{a}_s \end{bmatrix} \\
&= U \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_s \end{bmatrix} = UA
\end{aligned}$$

for the $r \times s$ matrix $A = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_s \end{bmatrix}$. By assumption $r < s$ and so Theorem 2.2.31 assures us that the homogeneous system $A\boldsymbol{x} = \boldsymbol{0}$ has infinitely many solutions, choose any non-trivial solution $\boldsymbol{x} \neq \boldsymbol{0}$. Consider

$$\begin{aligned}
V\boldsymbol{x} &= UA\boldsymbol{x} \quad \text{(from above)} \\
&= U\boldsymbol{0} \quad (\text{since } A\boldsymbol{x} = \boldsymbol{0}) \\
&= \boldsymbol{0} \,.
\end{aligned}$$

The identity $V\boldsymbol{x} = \boldsymbol{0}$ implies there is a linear combination of the columns $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_s$ of V which gives zero, hence the set $\mathcal{V}$ is linearly dependent (Theorem 7.2.12). But this is a contradiction, so we cannot have $r < s$.

Second, a corresponding argument establishes we cannot have $s < r$. Hence $s = r$: all bases of a given subspace must have the same number of vectors. □

Example 7.2.18. Consider the plane $x + y + z = 0$ in $\mathbb{R}^3$. Each of the following are a basis for the plane:

- $\{(-1, 1, 0),\ (1, -2, 1)\}$;
- $\{(1, -2, 1),\ (0, 1, -1)\}$;
- $\{(0, 1, -1),\ (-1, 1, 0)\}$.

The reasons are that all three vectors involved are in the plane, and that each pair are linearly independent (as, in each pair, one is not proportional to the other). However, although each of the three vectors in the set $\{(-1, 1, 0),\ (1, -2, 1),\ (0, 1, -1)\}$ is in the

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

plane, this set is not a basis because it is not linearly independent (Example 7.2.4a). Also, 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 $\big\{(1,0,-1)/\sqrt{2},\ (1,-2,1)/\sqrt{6}\big\}$ is another basis for the plane: both vectors satisfy $x+y+z=0$ and are orthogonal and so linearly independent (Theorem 7.2.6). All these bases possess two vectors. ■

That all bases for a given subspace, including orthonormal bases, have the same number of vectors leads to the following theorem about the dimensionality.

Theorem 7.2.19. *Let $\mathbb{W}$ be a subspace of $\mathbb{R}^n$. The **dimension** of $\mathbb{W}$, denoted $\dim\mathbb{W}$, is the number of vectors in any basis for $\mathbb{W}$.*

Proof. Recall Definition 3.4.30 defined $\dim\mathbb{W}$ to be the number of vectors in any orthonormal basis for $\mathbb{W}$. Theorem 7.2.6 certifies that all orthonormal bases are also bases (Definition 7.2.15), so Theorem 7.2.17 implies every basis of $\mathbb{W}$ has $\dim\mathbb{W}$ vectors. □

Procedure 7.2.20 (basis for a span)**.** *To find a basis for the subspace $\mathbb{A} = \operatorname{span}\{\boldsymbol{a}_1,\boldsymbol{a}_2,\dots,\boldsymbol{a}_n\}$ given $\{\boldsymbol{a}_1,\boldsymbol{a}_2,\dots,\boldsymbol{a}_n\}$ is a set of n vectors in $\mathbb{R}^m$. Recall Procedure 3.4.23 underpins finding an orthonormal basis by the following.*

1. *Form matrix $A := \begin{bmatrix}\boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_n\end{bmatrix}$.*
2. *Factorise 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 $\{\boldsymbol{u}_1,\boldsymbol{u}_2,\dots,\boldsymbol{u}_r\}$ (where $\boldsymbol{u}_j$ denotes the columns of U) is a basis, specifically an orthonormal basis, for the rD subspace $\mathbb{A}$.*

Alternatively, if $r = n$, then the set $\{\boldsymbol{a}_1,\boldsymbol{a}_2,\dots,\boldsymbol{a}_n\}$ is linearly independent and span the subspace $\mathbb{A}$, and so is also a basis for the nD subspace $\mathbb{A}$.

Example 7.2.21. Apply Procedure 7.2.20 to find a basis for the following sets.

(a) Recall Example 7.2.18 identified that any 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.

Solution: In MATLAB/Octave form the matrix with these vectors as columns:

```
A=[-1 1 0
   1 -2 1
   0 1 -1]
[U,S,V]=svd(A)
```

Then the SVD obtains (2 d.p.)

```
U =
  -0.41  -0.71   0.58
   0.82   0.00   0.58
  -0.41   0.71   0.58
S =
   3.00      0      0
      0   1.00      0
      0      0   0.00
V = ...
```

The two non-zero singular values determine $\operatorname{rank} A = 2$ and hence the first two columns of `U` form an (orthonormal) basis for $\operatorname{span}\{(-1,1,0),\ (1,-2,1),\ (0,1,-1)\}$. That is, $\{0.41(-1,2,-1),\ 0.71(-1,0,1)\}$ is an (orthonormal) basis.

(b) The span of the three vectors

$$(-2,0,-4,1,1),\ (7,1,2,-1,-5),\ (-5,-1,2,3,-2).$$

Solution: In MATLAB/Octave it is often easiest to enter these vectors as rows, and then transpose with the dash operator to form the matrix with these as columns:

```
A=[ -2 0 -4 1 1
 7 1 2 -1 -5
 -5 -1 2 3 -2]'
[U,S,V]=svd(A)
```

Then the SVD obtains (2 d.p.)

```
U =
   0.86  -0.32  -0.02   0.40   0.02
   0.12  -0.11  -0.06  -0.40   0.90
   0.22   0.65   0.72   0.07   0.13
  -0.23   0.35  -0.38   0.73   0.37
  -0.38  -0.59   0.58   0.37   0.19
S =
   10.07       0       0
       0    5.87       0
       0       0    3.01
       0       0       0
       0       0       0
V = ...
```

The three non-zero singular values determine $\operatorname{rank} A = 3$ and so the original three vectors are linearly independent.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Consequently the original three vectors form a basis for their span.

If you prefer an orthonormal basis, then use the first three columns of U.

(c) The span of the four vectors $(1, 0, 3, -4, 0)$, $(-1, -1, 1, 4, 2)$, $(-3, 2, 2, 2, 1)$, $(3, -3, 2, -2, 1)$.

Solution: In MATLAB/Octave, enter these vectors as rows, and then transpose with the dash operator to form the matrix with these as columns:

```
A=[1 0 3 -4 0
 -1 -1 1 4 2
 -3 2 2 2 1
 3 -3 2 -2 1]'
[U,S,V]=svd(A)
```

Then the SVD obtains (2 d.p.)

```
U =
  -0.52   0.11   0.46  -0.71   0.02
   0.28  -0.44  -0.55  -0.61   0.24
  -0.19   0.66  -0.60  -0.17  -0.37
   0.78   0.32   0.36  -0.30  -0.27
   0.10   0.51  -0.00   0.03   0.86
S =
   7.64      0      0      0
      0   4.59      0      0
      0      0   4.30      0
      0      0      0   0.00
      0      0      0      0
V = ...
```

The three non-zero singular values determine $\operatorname{rank} A = 3$. Consequently, the first three columns of U form an orthonormal basis for the span.

Alternatively, you might notice that the sum of the first two vectors is the sum of the last two vectors. Consequently, given the rank is three, we obtain three linearly independent vectors by omitting any one. That is, any three of the given vectors form a basis for the span of the four.

■

The procedure is different if the subspace of interest is defined by a system of equations instead of the span of some vectors.

v0.2a

Example 7.2.22. Find a basis for the solutions of the system in $\mathbb{R}^3$ of $3x + y = 0$ and $3x + 2y + 3z = 0$.

Solution: By hand manipulation, the first equation gives $y = -3x$; which when substituted into the second gives $3x - 6x + 3z = 0$, namely $z = x$. That is, all solutions are of the form $(x, -3x, x)$, namely $\operatorname{span}\{(1, -3, 1)\}$. Thus a basis for the subspace of solutions is $\{(1, -3, 1)\}$. (Infinitely many other bases are possible answers.)

■

Example 7.2.23. Find a basis for the solutions of $-2x - y + 3z = 0$ in $\mathbb{R}^3$.

Solution: Rearrange the equation so that one variable is a function of the others, say $y = -2x + 3z$. Then the vector form of of solutions are $(x, y, z) = (x, -2x + 3z, z) = (1, -2, 0)x + (0, 3, 1)z$ in terms of free variables x and z. Since $(1, -2, 0)$ and $(0, 3, 1)$ are not proportional to each other, they are linearly independent, and so a basis for the solutions is $\{(1, -2, 0), (0, 3, 1)\}$. (Infinitely many other bases are possible answers.)

■

Procedure 7.2.24 (basis from equations)**.** *Suppose we seek a basis for a subspace $\mathbb{W}$ defined as the solutions of a system of equations.*

1. *Rewrite the system of equations as the homogeneous system $A\boldsymbol{x} = \mathbf{0}$, hence 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 the* SVD *factorisation $A = USV^{\mathrm{T}}$ 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 the general solution of $S\boldsymbol{y} = \boldsymbol{z} = \mathbf{0}$ is $\boldsymbol{y} = (0, \dots, 0, y_{r+1}, \dots, y_n)$. 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 7.2.25. 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 7.2.22.

Solution: Form matrix $A = \begin{bmatrix} 3 & 1 & 0 \\ 3 & 2 & 3 \end{bmatrix}$ and compute an SVD with `[U,S,V]=svd(A)` to obtain (2 d.p.)

```
U = ...
S =
   5.34      0      0
      0   1.86      0
V =
   0.77   0.56   0.30
   0.42  -0.09  -0.90
   0.48  -0.82   0.30
```

The two non-zero singular values determine $\operatorname{rank} A = 2$. Hence the solutions of the system are spanned by the last one column of V. That is, a basis for the solutions is $\{(0.3, -0.9, 0.3)\}$.

(b) $7x = 6y + z + 3$ and $4x + 9y + 2z + 2 = 0$.

Solution: This system is not homogeneous (due to the constant terms, Definition 2.2.28), therefore $\boldsymbol{x} = \mathbf{0}$ is not a solution. Consequently, the solutions of the system cannot form a subspace (Definition 3.4.3). Thus the concept of a basis does not apply (Definition 7.2.15).

(c) $w + x = z$, $3w = x + y + 5z$, $4x + y + 2z = 0$.

Solution: Rearrange to the matrix-vector system $A\boldsymbol{x} = \mathbf{0}$ for vector $\boldsymbol{x} = (w, x, y, z) \in \mathbb{R}^4$ and matrix

```
A=[1  1  0 -1
   3 -1 -1 -5
   0  4  1  2]
```

Enter into MATLAB/Octave as above and then find an SVD with `[U,S,V]=svd(A)` to obtain (2 d.p.)

```
U = ...
S =
   6.77      0      0      0
      0   3.76      0      0
      0      0   0.00      0
V =
  -0.40   0.45   0.09   0.80
   0.41   0.86  -0.19  -0.25
   0.20   0.10   0.97  -0.07
   0.80  -0.24  -0.10   0.54
```

There are two non-zero singular values, so $\operatorname{rank} A = 2$. There are thus $4 - 2 = 2$ free variables in solving $A\boldsymbol{x} = \mathbf{0}$ leading to a 2D subspace with orthonormal basis of the last two columns of V. That is, an orthonormal basis for the subspace of all solutions is $\{(0.09, -0.19, 0.97, -0.10), (0.80, -0.25, -0.07, 0.54)\}$.

■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Recall this Section 7.2 started by discussing the need to have a unique representation of solutions to problems. If we do not have uniqueness, then the ambiguity in algebraic representation ruins basic algebra. The following theorem assures us that the linear independence of a basis ensures the unique representation that we need. In essence it says that any basis, whether orthogonal or not, can be used to form a coordinate system.

Example 7.2.26 (a tale of two coordinate systems)**.** In the margin are plotted three vectors and the origin. Take the view that these are fixed physically meaningful vectors: the issue of this example is how do we code such vectors in mathematics.

In the standard orthogonal coordinate system these three vectors and the origin have coordinates as plotted left by their end-points. 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 we cannot say both $\boldsymbol{a} = (7,4)$ and $\boldsymbol{a} = (2,1)$: it appears nonsense. The reason for the different 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 apparent contradiction: these statements recognise the underlying standard unit vectors in the first expression and the underlying non-orthogonal basis vectors in the second.

Consequently we invent a new better notation. We write $[\boldsymbol{a}]_{\mathcal{B}} = (2,1)$ 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 $\boldsymbol{a} = (2,1)$, $\boldsymbol{b} = (1,-3)$ and $\boldsymbol{c} = (-2,3)$ is rationalised in this new notation by the convention that if no basis is specified, then the standard basis $\mathcal{E}$ is assumed.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

[6] ■

Theorem 7.2.27. *Let $\mathbb{W}$ be a subspace of $\mathbb{R}^n$ and let $\mathcal{B} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k\}$ be a basis for $\mathbb{W}$. There is exactly one way to write each and every vector $\boldsymbol{w} \in \mathbb{W}$ as a linear combination of the basis vectors: $\boldsymbol{w} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k$. Then $c_1, c_2, \dots, c_k$ are called the* ***coordinates of $\boldsymbol{w}$ with respect to $\mathcal{B}$****, and the column vector $[\boldsymbol{w}]_{\mathcal{B}} = (c_1, c_2, \dots, c_k)$ is called the* ***coordinate vector of $\boldsymbol{w}$ with respect to $\mathcal{B}$****.*

Proof. Consider any vector $\boldsymbol{w} \in \mathbb{W}$. Since $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k\}$ is a basis for the subspace $\mathbb{W}$, $\boldsymbol{w}$ can be written as a linear combination of the basis vectors. Let two such linear combinations be

$$\begin{aligned} \boldsymbol{w} &= c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k\,, \\ \boldsymbol{w} &= d_1\boldsymbol{v}_1 + d_2\boldsymbol{v}_2 + \cdots + d_k\boldsymbol{v}_k\,. \end{aligned}$$

Subtract the second of these equations from the first, grouping common vectors:

$$\boldsymbol{0} = (c_1 - d_1)\boldsymbol{v}_1 + (c_2 - d_2)\boldsymbol{v}_2 + \cdots + (c_k - d_k)\boldsymbol{v}_k\,.$$

Since $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k\}$ is linearly independent, this equation implies all the coefficients in parentheses are zero:

$$0 = (c_1 - d_1) = (c_2 - d_2) = \cdots = (c_k - d_k).$$

That is, $c_1 = d_1$, $c_2 = d_2$, $\dots$, $c_k = d_k$, and the two linear combinations are identical. That is, there is exactly one way to write a vector $\boldsymbol{w} \in \mathbb{W}$ as a linear combination of the basis vectors. □

Example 7.2.28. (a) Consider the diagram of six labelled vectors drawn below.

Estimate the coordinates of the four shown vectors $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ and $\boldsymbol{d}$ in the shown basis $\mathcal{B} = \{\boldsymbol{v}_1, \boldsymbol{v}_2\}$.

Solution: Draw in a grid corresponding to multiples of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ in both directions, and parallel to $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$, as shown below. Then from the grid, estimate that $\boldsymbol{a} \approx 3\boldsymbol{v}_1 + 2\boldsymbol{v}_2$ hence the coordinates $[\boldsymbol{a}]_{\mathcal{B}} \approx (3, 2)$.

[6] Given that the numerical representation of a vector changes with the coordinate basis, some of you will wonder whether the same thing happens for matrices. The answer is yes.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Similarly, $\boldsymbol{b} \approx \boldsymbol{v}_1 - 2\boldsymbol{v}_2$ hence the coordinates $[\boldsymbol{b}]_{\mathcal{B}} \approx (1\,,\,-2)$. Also, $\boldsymbol{c} \approx -2\boldsymbol{v}_1 + 0.5\boldsymbol{v}_2$ hence the coordinates $[\boldsymbol{c}]_{\mathcal{B}} \approx (-2\,,\,0.5)$. And lastly, $\boldsymbol{d} \approx -\boldsymbol{v}_1 - \boldsymbol{v}_2$ hence the coordinates $[\boldsymbol{d}]_{\mathcal{B}} \approx (-1\,,\,-1)$.

(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 $\mathcal{W} = \{\boldsymbol{w}_1\,,\,\boldsymbol{w}_2\}$.

Solution: Draw in a grid corresponding to multiples of $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ in both directions, and parallel to $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$, as shown below. Then from the grid, estimate that $\boldsymbol{a} \approx \boldsymbol{w}_1 - 1.5\boldsymbol{w}_2$ hence the coordinates $[\boldsymbol{a}]_{\mathcal{W}} \approx (1, -1.5)$.

Similarly, $\boldsymbol{b} \approx 3\boldsymbol{w}_1 - 0.5\boldsymbol{w}_2$ hence the coordinates $[\boldsymbol{b}]_{\mathcal{W}} \approx (3\,,\,-0.5)$. Also, $\boldsymbol{c} \approx -2.5\boldsymbol{w}_1 + \boldsymbol{w}_2$ hence the coordinates $[\boldsymbol{c}]_{\mathcal{W}} \approx (-2.5\,,\,1)$. And lastly, $\boldsymbol{d} \approx 0.5\boldsymbol{w}_2$ hence the coordinates $[\boldsymbol{d}]_{\mathcal{W}} \approx (0\,,\,0.5)$.

■

Example 7.2.29. 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 $\boldsymbol{v}_3 = (0, 4, 5)$ (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 $[\boldsymbol{a}]_{\mathcal{B}} = (3, -2, 1)$?

Solution: $\boldsymbol{a} = 3\boldsymbol{v}_1 - 2\boldsymbol{v}_2 + \boldsymbol{v}_3$ which has standard coordinates $[\boldsymbol{a}]_{\mathcal{E}} = 3(-1, 1, -1) - 2(1, -2, 0) + (0, 4, 5) = (-5, 11, 2)$.

(b) What is the vector with coordinates $[\boldsymbol{b}]_{\mathcal{B}} = (-1, 1, 1)$?

Solution: $\boldsymbol{b} = -\boldsymbol{v}_1 + \boldsymbol{v}_2 + \boldsymbol{v}_3$ which has standard coordinates $[\boldsymbol{b}]_{\mathcal{E}} = -(-1, 1, -1) + (1, -2, 0) + (0, 4, 5) = (2, 1, 6)$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(c) What are the coordinates in the basis $\mathcal{B}$ of the vector $\boldsymbol{c} = (-1,3,3)$ in the standard basis $\mathcal{E}$?

Solution: We seek coordinate values c_1, c_2, c_3 such that $\boldsymbol{c} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3$. Expressed in the standard basis this equation is

$$\begin{bmatrix}-1\\3\\3\end{bmatrix} = \begin{bmatrix}-1\\1\\-1\end{bmatrix} c_1 + \begin{bmatrix}1\\-2\\0\end{bmatrix} c_2 + \begin{bmatrix}0\\4\\5\end{bmatrix} c_3 .$$

A small system like this we solve by hand (recall section 2.2.2): write in component form as

$$\begin{cases} -c_1 \quad +c_2 \qquad\quad = -1 \\ \quad c_1 \ -2c_2 \ +4c_3 = 3 \\ -c_1 \qquad\quad +5c_3 = 3 \end{cases}$$

(add 1st row to 2nd and take from 3rd)

$$\begin{cases} -c_1 \ +c_2 \qquad\quad = -1 \\ \qquad -c_2 \ +4c_3 = 2 \\ \qquad -c_2 \ +5c_3 = 4 \end{cases}$$

(subtract 2nd row from 3rd)

$$\begin{cases} -c_1 \ +c_2 \qquad\quad = -1 \\ \qquad -c_2 \ +4c_3 = 2 \\ \qquad\qquad\quad c_3 = 2 \end{cases}$$

Solving this triangular system gives $c_3 = 2$, $c_2 = 4c_3 - 2 = 6$, and $c_1 = c_2 + 1 = 7$. Thus the coordinates $[\boldsymbol{c}]_{\mathcal{B}} = (7,6,2)$ in the basis $\mathcal{B}$.

(d) What are the coordinates in the basis $\mathcal{B}$ of the vector $\boldsymbol{d} = (-3,2,0)$ in the standard basis $\mathcal{E}$?

Solution: We seek coordinate values d_1, d_2, d_3 such that $\boldsymbol{d} = d_1\boldsymbol{v}_1 + d_2\boldsymbol{v}_2 + d_3\boldsymbol{v}_3$. Expressed in the standard basis this equation is

$$\begin{bmatrix}-3\\2\\0\end{bmatrix} = \begin{bmatrix}-1\\1\\-1\end{bmatrix} d_1 + \begin{bmatrix}1\\-2\\0\end{bmatrix} d_2 + \begin{bmatrix}0\\4\\5\end{bmatrix} d_3 .$$

To solve this system with MATLAB/Octave (procedure 2.2.5), enter the matrix (easiest by transposing rows of $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$) and the standard coordinates of $\boldsymbol{d}$:

```
A=[-1 1 -1
 1 -2 0
 0 4 5]'
d=[-3;2;0]
```

Then compute the coordinates $[\boldsymbol{d}]_{\mathcal{B}}$ with `dB=A\d` to determine $[\boldsymbol{d}]_{\mathcal{B}} = (20,17,4)$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

But remember, before using `A\` always first check `rcond(A)` which here is the poor 0.0053 (Procedure 2.2.5). Interestingly, this poor small value of `rcond` indicates that although the basis vectors in $\mathcal{B}$ are linearly independent, they are 'only just' linearly independent. A small change or error might make them linearly dependent and thus $\mathcal{B}$ be ruined as a basis for $\mathbb{R}^3$. The poor `rcond` indicates that $\mathcal{B}$ is a poor basis in practice.

■

Example 7.2.30. 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 $\boldsymbol{a} = 2\boldsymbol{w}_1 + 3\boldsymbol{w}_2 + \boldsymbol{w}_3$?

Solution: In the standard basis

$$[\boldsymbol{a}]_{\mathcal{E}} = 2\begin{bmatrix}1\\3\\-4\\-3\\3\end{bmatrix} + 3\begin{bmatrix}-4\\1\\-2\\-4\\1\end{bmatrix} + \begin{bmatrix}-1\\1\\0\\2\\-3\end{bmatrix} = \begin{bmatrix}-11\\10\\-14\\-16\\6\end{bmatrix}.$$

(b) What are the coordinates in the basis $\mathcal{W}$ of the vector $\boldsymbol{b} = (-1, 2, -6, -11, 10)$ (in the standard coordinates $\mathcal{E}$).

Solution: We need to find coefficients c_1, c_2, c_3 such that $\boldsymbol{b} = c_1\boldsymbol{w}_1 + c_2\boldsymbol{w}_2 + c_3\boldsymbol{w}_3$. This forms the set of linear equations

$$\begin{bmatrix}-1\\2\\-6\\-11\\10\end{bmatrix} = \begin{bmatrix}1\\3\\-4\\-3\\3\end{bmatrix} c_1 + \begin{bmatrix}-4\\1\\-2\\-4\\1\end{bmatrix} c_2 + \begin{bmatrix}-1\\1\\0\\2\\-3\end{bmatrix} c_3 .$$

Form as the matrix-vector system

$$\begin{bmatrix}1 & -4 & -1\\3 & 1 & 1\\-4 & -2 & 0\\-3 & -4 & 2\\3 & 1 & -3\end{bmatrix}\begin{bmatrix}c_1\\c_2\\c_3\end{bmatrix} = \begin{bmatrix}-1\\2\\-6\\-11\\10\end{bmatrix},$$

and perhaps solve with MATLAB/Octave. Since there are more equations than unknowns, we should use an SVD in order to check the system is consistent, namely, to check that $\boldsymbol{b} \in \mathbb{W}$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

i. Code the matrix and the vector:

```
W=[1 -4 -1
 3 1 1
 -4 -2 0
 -3 -4 2
 3 1 -3]
b=[-1;2;-6;-11;10]
```

ii. Then obtain an SVD with `[U,S,V]=svd(W)` (2 d.p.)

```
U =
   0.18  -0.88   0.04  -0.24  -0.37
  -0.32  -0.09   0.65   0.62  -0.28
   0.51   0.11  -0.46   0.56  -0.44
   0.64  -0.18   0.34   0.22   0.63
  -0.44  -0.42  -0.49   0.44   0.45
S =
   8.18      0      0
      0   4.52      0
      0      0   3.11
      0      0      0
      0      0      0
V =
  -0.74  -0.51   0.43
  -0.62   0.77  -0.15
   0.26   0.38   0.89
```

The three non-zero singular values establish that the three vectors in the basis $\mathcal{W}$ are indeed linearly independent (and since no singular value is small, then the vectors are robustly linearly independent).

iii. Find $\boldsymbol{z} = U^{\mathrm{T}}\boldsymbol{b}$ with `z=U'*b` to get

```
z =
  -15.3413
   -2.2284
   -4.6562
   -0.0000
    0.0000
```

The last two values of $\boldsymbol{z}$ being zero confirm the system of equations is consistent and so vector $\boldsymbol{b}$ is in the range of $\mathcal{W}$, that is, $\boldsymbol{b}$ is in the subspace $\mathbb{W}$.

iv. Find $y_j = z_j/\sigma_j$ with `y=z(1:3)./diag(S)` to get

```
y =
  -1.8761
  -0.4929
  -1.4958
```

v0.2a

v. Lastly, find the coefficients $[\boldsymbol{b}]_{\mathcal{W}} = V\boldsymbol{y}$ with `bw=V*y` to get

```
bw =
   1.00000
   1.00000
  -2.00000
```

That is, $[\boldsymbol{b}]_{\mathcal{W}} = (1, 1, -2)$.

That $[\boldsymbol{b}]_{\mathcal{W}}$ has three components and $[\boldsymbol{b}]_{\mathcal{E}}$ has five components is not a contradiction. The difference in components occurs because the subspace $\mathbb{W}$ is 3D but lies in $\mathbb{R}^5$. Using the basis $\mathcal{W}$ implicitly builds in the information that the vector $\boldsymbol{b}$ is in a lower dimensional space, and so needs fewer components.

■

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 7.2.31 (Unique Solutions: version 3)**.** *Let A be an $n \times n$ square matrix. Extending Theorem 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) $\operatorname{rank} A = n$;

(f) $\operatorname{nullity} A = 0$;

(g) the column vectors of A span $\mathbb{R}^n$;

(h) the row vectors of A span $\mathbb{R}^n$.

(i) $\det A \neq 0$;

(j) 0 is not an eigenvalue of A;

(k) the n column vectors of A are linearly independent;

(l) the n row vectors of A are linearly independent.

Proof. Recall Theorems 3.3.26 and 3.4.43 established the equivalence of 7.2.31a–7.2.31h, and Theorem 6.1.23 proved the equivalence

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

of 7.2.31a and 7.2.31i. To establish that Property 7.2.31j is equivalent to 7.2.31i, recall Theorem 7.1.3 proved that $\det A$ equals the product of the eigenvalues of A. Hence $\det A$ is not zero if and only if all the eigenvalues are non-zero.

Lastly, Property 7.2.31g says the n column vectors span $\mathbb{R}^n$, so they must be a basis, and hence linearly independent. Conversely, if the n columns of A are linearly independent then they must span $\mathbb{R}^n$. Hence Property 7.2.31k is equivalent to 7.2.31g. Similarly for Property 7.2.31h and the row vectors of 7.2.31l. □

7.2.3 Exercises

Exercise 7.2.1. By inspection or basic arguments, decide wether the following sets of vectors are linearly dependent, or linearly independent. Give reasons.

(a) $\{(-2,3,3),(-1,2,-1)\}$

(b) $\{(0,2),(2,-2),(0,-1)\}$

(c) $\{(-3,0,-3),(3,2,-2)\}$

(d) $\{(0,2,2)\}$

(e) $\{(-2,0,-1),(0,-2,2),(2,0,1)\}$

(f) $\{(0,3,-2),(-2,-1,1),(1,-2,-4)\}$

(g) $\{(-2,1),,(0,0)\}$

(h) $\{(-2,-1,1),(-2,-2,2),(2,-1,-1),(2,-2,-2)\}$

(i) $\{(2,4),(1,2)\}$

(j) $\{(1,-2,3),(1,1,2)\}$

Exercise 7.2.2. Compute an SVD to decide whether the following sets of vectors are linearly dependent, or linearly independent. Give reasons.

(a) $\begin{bmatrix}5\\0\\-1\end{bmatrix},\begin{bmatrix}-2\\5\\-1\end{bmatrix}$

(b) $\begin{bmatrix}-2\\-1\\1\end{bmatrix},\begin{bmatrix}3\\-4\\-1\end{bmatrix},\begin{bmatrix}5\\-3\\-2\end{bmatrix}$

(c) $\begin{bmatrix}2\\-4\\1\\6\end{bmatrix},\begin{bmatrix}-1\\1\\10\\-5\end{bmatrix},\begin{bmatrix}1\\6\\-2\\4\end{bmatrix}$

(d) $\begin{bmatrix}1\\0\\-5\\2\end{bmatrix},\begin{bmatrix}-1\\0\\-2\\4\end{bmatrix},\begin{bmatrix}-1\\3\\1\\1\end{bmatrix},\begin{bmatrix}3\\4\\-4\\-4\end{bmatrix}$

v0.2a

(e) $\begin{bmatrix}0\\-2\\2\\0\end{bmatrix}, \begin{bmatrix}-1\\-4\\-6\\-1\end{bmatrix}, \begin{bmatrix}4\\3\\-5\\0\end{bmatrix}$ (f) $\begin{bmatrix}0\\3\\2\\0\end{bmatrix}, \begin{bmatrix}-3\\-3\\1\\-2\end{bmatrix}, \begin{bmatrix}-3\\0\\3\\-2\end{bmatrix}$

(g) $\begin{bmatrix}-3\\2\\6\\3\\-2\end{bmatrix}, \begin{bmatrix}-3\\-3\\1\\6\\-4\end{bmatrix}, \begin{bmatrix}3\\-2\\-2\\4\\-1\end{bmatrix}, \begin{bmatrix}1\\-2\\-2\\3\\2\end{bmatrix}$ (h) $\begin{bmatrix}3\\0\\4\\1\\2\end{bmatrix}, \begin{bmatrix}-2\\1\\2\\2\\2\end{bmatrix}, \begin{bmatrix}2\\1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}-9\\-1\\-2\\1\\0\end{bmatrix}$

Exercise 7.2.3. Prove the particular case of Theorem 7.2.8, namely that 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.

Exercise 7.2.4. Prove that every (non-empty) subset of a linearly independent set is also linearly independent. (Perhaps use contradiction.)

Exercise 7.2.5. Let $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_m\}$ be a linearly independent set of vectors in $\mathbb{R}^n$. Given that a vector $\boldsymbol{u} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_m\boldsymbol{v}_m$ with coefficient $c_1 \neq 0$, prove that the set $\{\boldsymbol{v}_2, \boldsymbol{v}_3, \dots, \boldsymbol{v}_m, \boldsymbol{u}\}$ is linearly independent.

Exercise 7.2.6. For each of the following systems of equations find by hand two different bases for their solution set (among the infinitely many bases that are possible). Show your working.

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

(b) $6x + 4y + 2z = 0$ and $-2x - y - 2z = 0$

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

(d) $-7x + y - z = 0$ and $-3x + 2 - 2z = 0$

(e) $x + 2y + 2z = 0$

(f) $2x + 0y - 4z = 0$

(g) $-2x + 3y - 6z = 0$

(h) $9x + 4y - 9z = 6$

Exercise 7.2.7. Use Procedure 7.2.24 to compute in MATLAB/Octave an orthonormal basis for all solutions to each of the following systems of equations.

(a) $2x - 6y - 9z = 0$, $2x - 2z = 0$, $-x + z = 0$

(b) $3x - 3y + 8z = 0$, $-2x - 4y + 2z = 0$, $-4x + y - 7z = 0$

(c) $-2x + 2y - 2z = 0$, $x + 3y - z = 0$, $3x + 3y = 0$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(d) $2w + x + 2y + z = 0$, $w + 4x - 4y - z = 0$, $3w - 2x + 5y = 0$, $2w - x + y - 2z = 0$

(e) $5w+y-3z = 0$, $-5x-5y = 0$, $-3x-y+4z = 0$, $3x+y-4z = 0$

(f) $-w - 2y + 4z = 0$, $2w + 2y + 2z = 0$, $-2w + 3x + y + z = 0$, $-w + x - y + 5z = 0$

(g) $-2w+x+2y-6z = 0$, $-2w+3x+4y = 0$, $-2w+2x+3x-3z = 0$

(h) $-w-2x-3y+2z = 0$, $2x+2y-2z = 0$, $-w-3x-4y+3z = 0$

Exercise 7.2.8. Recall that Theorem 4.2.14 establishes there are at most n eigenvalues of an $n \times n$ symmetric matrix. Adapt the proof of that theorem, using linear independence, to prove there are at most n eigenvalues of an $n \times n$ non-symmetric matrix. (This is an alternative to the given proof of Theorem 7.1.1.)

Exercise 7.2.9. For each diagram, estimate roughly the components of each of the four vectors $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ and $\boldsymbol{d}$ in the basis $\mathcal{P} = \{\boldsymbol{p}_1, \boldsymbol{p}_2\}$.

(a)

(b)

(c)

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(d)

(e)

(f)

(g)

Exercise 7.2.10. Let the three given vectors $\boldsymbol{b}_1 = (-1, 1, -1)$, $\boldsymbol{b}_2 = (1, -2, 0)$ and $\boldsymbol{b}_3 = (0, 4, 5)$ form a basis $\mathcal{B} = \{\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3\}$ (where these vectors are specified in the standard basis $\mathcal{E}$ of the standard unit vectors $\boldsymbol{e}_1$, $\boldsymbol{e}_2$ and $\boldsymbol{e}_3$). For each of the following vectors with specified coordinates in basis $\mathcal{B}$, what is the vector when written in the standard basis?

(a) $[\boldsymbol{p}]_{\mathcal{B}} = (1\,,-1\,,2)$

(b) $[\boldsymbol{q}]_{\mathcal{B}} = (0\,,2\,,3)$

(c) $[\boldsymbol{r}]_{\mathcal{B}} = (1\,,-3\,,-2)$

(d) $[\boldsymbol{s}]_{\mathcal{B}} = (1\,,2\,,1)$

(e) $[\boldsymbol{t}]_{\mathcal{B}} = (1/2\,,-1/2\,,1)$

(f) $[\boldsymbol{u}]_{\mathcal{B}} = (-1/2\,,1/2\,,-1/2)$

v0.2a

(g) $[\boldsymbol{v}]_{\mathcal{B}} = (0\,,1/2\,,-1/2)$

(h) $[\boldsymbol{w}]_{\mathcal{B}} = (-0.7\,,0.5\,,1.1)$

(i) $[\boldsymbol{x}]_{\mathcal{B}} = (0.2\,,-0.1\,,0.9)$

(j) $[\boldsymbol{y}]_{\mathcal{B}} = (2.1\,,-0.2\,,0.1)$

Exercise 7.2.11. Repeat Exercise 7.2.10 but with the three basis vectors $\boldsymbol{b}_1 = (6,2,1)$, $\boldsymbol{b}_2 = (-2,-1,-2)$ and $\boldsymbol{b}_3 = (-3,-1,5)$.

Exercise 7.2.12. Let the two given vectors $\boldsymbol{b}_1 = (1,-2,2)$ and $\boldsymbol{b}_2 = (1,-1,-1)$ form a basis $\mathcal{B} = \{\boldsymbol{b}_1, \boldsymbol{b}_2\}$ for the subspace $\mathbb{B}$ of $\mathbb{R}^3$ (specified in the standard basis $\mathcal{E}$ of the standard unit vectors $\boldsymbol{e}_1$, $\boldsymbol{e}_2$ and $\boldsymbol{e}_3$). For each of the following vectors, specified in the standard basis $\mathcal{E}$, what is the vector when written in the basis $\mathcal{B}$? if possible.

(a) $[\boldsymbol{p}]_{\mathcal{E}} = (0\,,1\,,3)$

(b) $[\boldsymbol{q}]_{\mathcal{E}} = (-4\,,9\,,-11)$

(c) $[\boldsymbol{r}]_{\mathcal{E}} = (-4\,,7\,,-5)$

(d) $[\boldsymbol{s}]_{\mathcal{E}} = (0\,,2\,,-4)$

(e) $[\boldsymbol{t}]_{\mathcal{E}} = (0\,,-2\,,5)$

(f) $[\boldsymbol{u}]_{\mathcal{E}} = (-2/3\,,0\,,8/3)$

(g) $[\boldsymbol{v}]_{\mathcal{E}} = (-8/3\,,19/3\,,-19/3)$

(h) $[\boldsymbol{w}]_{\mathcal{E}} = (-10/3\,,5\,,-5/3)$

(i) $[\boldsymbol{x}]_{\mathcal{E}} = (-0.3\,,0.4\,,0)$

(j) $[\boldsymbol{y}]_{\mathcal{E}} = (0.5\,,-0.7\,,0.1)$

Exercise 7.2.13. Repeat Exercise 7.2.12 but with the two basis vectors $\boldsymbol{b}_1 = (-2,3,-1)$ and $\boldsymbol{b}_2 = (0,-1,3)$.

Exercise 7.2.14. Let the three vectors $\boldsymbol{b}_1 = (-1,-2,5,3,-2)$, $\boldsymbol{b}_2 = (-2,-2,2,-1,-1)$ and $\boldsymbol{b}_3 = (-4,6,-4,2,-1)$ form a basis $\mathcal{B} = \{\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3\}$ for the subspace $\mathbb{B}$ in $\mathbb{R}^5$ (specified in the standard basis $\mathcal{E}$). For each of the following vectors, use MATLAB/Octave to find the requested coordinates, if possible.

(a) Find $[\boldsymbol{p}]_{\mathcal{E}}$ when $[\boldsymbol{p}]_{\mathcal{B}} = (2,2,-1)$.

(b) Find $[\boldsymbol{q}]_{\mathcal{E}}$ when $[\boldsymbol{q}]_{\mathcal{B}} = (-5,0,-2)$.

(c) Find $[\boldsymbol{r}]_{\mathcal{E}}$ when $[\boldsymbol{r}]_{\mathcal{B}} = (0,3,3)$.

(d) Find $[\boldsymbol{s}]_{\mathcal{E}}$ when $[\boldsymbol{s}]_{\mathcal{B}} = (-1,5,0)$.

(e) Find $[\boldsymbol{t}]_{\mathcal{B}}$ when $[\boldsymbol{t}]_{\mathcal{E}} = (-31,26,-5,19,-14)$.

(f) Find $[\boldsymbol{u}]_{\mathcal{B}}$ when $[\boldsymbol{u}]_{\mathcal{E}} = (-1,6,4,14,-5)$.

(g) Find $[\boldsymbol{v}]_{\mathcal{B}}$ when $[\boldsymbol{v}]_{\mathcal{E}} = (-21,18,-7,9,-8)$.

(h) Find $[\boldsymbol{w}]_{\mathcal{B}}$ when $[\boldsymbol{w}]_{\mathcal{E}} = (-0.2,-0.6,1.8,1.3,-0.7)$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(i) Find $[\boldsymbol{x}]_{\mathcal{B}}$ when $[\boldsymbol{x}]_{\mathcal{E}} = (0.7, -0.4, -0.3, -0.3, 1.2)$.

(j) Find $[\boldsymbol{y}]_{\mathcal{B}}$ when $[\boldsymbol{y}]_{\mathcal{E}} = (4.8, -3.8, 0.8, -2.6, 2.1)$.

Exercise 7.2.15. Repeat Exercise 7.2.14 but with basis vectors $\boldsymbol{b}_1 = (-3, 8, -9, -1, 1)$, $\boldsymbol{b}_2 = (10, -20, 14, -7, 2)$ and $\boldsymbol{b}_3 = (-1, -2, 5, 3, -2)$ (specified in the standard basis $\mathcal{E}$).

v0.2a

7.3 Diagonalisation identifies the transformation

Section Contents

7.3.1 Solve systems of differential equations 683

7.3.2 Exercises 694

Population modelling Recall that this Chapter 7 started by introducing the dynamics of two interacting species of animals. Recall we let $y(t)$ and $z(t)$ be the number of female animals in each of the species at time t (years). Modelling might deduce the populations interact according to the rule that the population one year later is $y(t+1) = 2y(t) - 4z(t)$ and $z(t+1) = -y(t) + 2z(t)$. Then seeking solutions proportional to 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.

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

$$\boldsymbol{y}(t+1) = \begin{bmatrix} y(t+1) \\ z(t+1) \end{bmatrix} = \begin{bmatrix} 2y - 4z \\ -y + 2z \end{bmatrix} = \begin{bmatrix} 2 & -4 \\ -1 & 2 \end{bmatrix} \boldsymbol{y} \,.$$

Now let's ask if there is a basis $\mathcal{P} = \{\boldsymbol{p}_1, \boldsymbol{p}_2\}$ for the yz-plane that simplifies the 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 components Y_1 and Y_2 (where $(Y_1, Y_2) = \boldsymbol{Y} = [\boldsymbol{y}]_{\mathcal{P}}$, but to simplify writing we use the symbol $\boldsymbol{Y}$ not $[\boldsymbol{y}]_{\mathcal{P}}$). Write this relation 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 $\boldsymbol{Y}$; that is, $\boldsymbol{y}(t) = P\boldsymbol{Y}(t)$. Substitute this into the system of equations:

$$\boldsymbol{y}(t+1) = P\boldsymbol{Y}(t+1) = \begin{bmatrix} 2 & -4 \\ -1 & 2 \end{bmatrix} P\boldsymbol{Y} \,.$$

Multiply both sides by P^{-1} (which exists by linear independence of the columns, Theorem 7.2.31k) to give

$$\boldsymbol{Y}(t+1) = \underbrace{P^{-1} \begin{bmatrix} 1 & -4 \\ -1 & 1 \end{bmatrix} P}_{P^{-1}AP} \boldsymbol{Y} \,.$$

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: using eigenvalues and eigenvectors, most of the time the product $P^{-1}AP$ can be made into a simple diagonal matrix.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Recall that (section 4.2.2) for a symmetric matrix A we could always factor $A = VDV^{\mathrm{T}} = VDV^{-1}$ for orthogonal matrix V and diagonal matrix D: thus a symmetric matrix is always orthogonally diagonalisable. For non-symmetric matrices, a diagonalisation mostly (although not always) can be done with the difference being we need an invertible matrix, typically called P, instead of the orthogonal matrix V. Such a matrix is termed 'diagonalisable' instead of 'orthogonally diagonalisable'.

Definition 7.3.1. *An $n \times n$ square matrix A is **diagonalisable** if there exists a diagonal matrix D and an invertible matrix P such that $A = PDP^{-1}$, equivalently $AP = PD$ or $P^{-1}AP = D$.*

Example 7.3.2. (a) Show that $A = \begin{bmatrix} 0 & 1 \\ 2 & -1 \end{bmatrix}$ is diagonalisable by matrix $P = \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix}$.

Solution: First find the 2×2 inverse (Theorem 3.2.7)

$$P^{-1} = \frac{1}{3}\begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix}.$$

Second, compute the product

$$P^{-1}AP = P^{-1}\begin{bmatrix} 1 & 2 \\ 1 & -4 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 3 & 0 \\ 0 & -6 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix}.$$

Since this product is diagonal, $\operatorname{diag}(1, -2)$, the matrix A is diagonalisable.

(b) $B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ is not diagonalisable.

Solution: Assume B is diagonalisable by the invertible matrix $P = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Being invertible, P has inverse $P^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$ (Theorem 3.2.7). Then the product

$$P^{-1}BP = P^{-1}\begin{bmatrix} c & d \\ 0 & 0 \end{bmatrix} = \frac{1}{ad-bc}\begin{bmatrix} cd & d^2 \\ -c^2 & -cd \end{bmatrix}.$$

For this matrix $(P^{-1}BP)$ to be diagonal requires the off-diagonal elements to be zero: $d^2 = -c^2 = 0$. This requires both $c = d = 0$, but then the determinant $ad - bc = 0 - 0 = 0$ and so matrix P is not invertible (Theorem 3.2.7). This contradiction means that matrix B is not diagonalisable.

(c) Is matrix $C = \begin{bmatrix} 1.2 & 3.2 & 2.3 \\ 2.2 & -0.5 & -2.2 \end{bmatrix}$ diagonalisable?

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Solution: No, as it is not a square matrix. (Perhaps an SVD could answer the needs of whatever problem led to this question.)

■

Example 7.3.3. Example 7.3.2a showed that matrix $P = \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix}$ diagonalises matrix $A = \begin{bmatrix} 0 & 1 \\ 2 & -1 \end{bmatrix}$ to matrix $D = \operatorname{diag}(1, -2)$. As a prelude to the next Theorem 7.3.4, show that the columns of P are eigenvectors of A.

Solution: Invoke the original Definition 4.1.1 of an eigenvector for a matrix.

- The first column of P is $\boldsymbol{p}_1 = (1, 1)$. Multiplying $A\boldsymbol{p}_1 = (0 + 1, 2 - 1) = (1, 1) = 1\boldsymbol{p}_1$ so vector $\boldsymbol{p}_1$ is an eigenvector of A corresponding to the eigenvalue 1. Correspondingly, this eigenvalue is the first entry in the diagonal D.
- The second column of P is $\boldsymbol{p}_2 = (-1, 2)$. Multiplying $A\boldsymbol{p}_2 = (0+2, -2-2) = (2, -4) = -2\boldsymbol{p}_2$ so vector $\boldsymbol{p}_2$ is an eigenvector of A corresponding to the eigenvalue -2. Correspondingly, this eigenvalue is the second entry in the diagonal D.

■

Theorem 7.3.4. *An $n \times n$ square matrix A is diagonalisable if and only if A has n linearly independent eigenvectors. If A is diagonalisable, with diagonal matrix $D = P^{-1}AP$, then the diagonal entries of D are eigenvalues, and the columns of P are corresponding eigenvectors.*

Proof. First, let matrix A be diagonalisable by invertible P and diagonal D. Write $P = \begin{bmatrix} \boldsymbol{p}_1 & \boldsymbol{p}_2 & \cdots & \boldsymbol{p}_n \end{bmatrix}$ in terms of its columns, and let $D = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ in terms of its diagonal entries. Then $AP = PD$ becomes

$$A \begin{bmatrix} \boldsymbol{p}_1 & \boldsymbol{p}_2 & \cdots & \boldsymbol{p}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{p}_1 & \boldsymbol{p}_2 & \cdots & \boldsymbol{p}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

Multiplying the matrix-column products on both sides gives

$$\begin{bmatrix} A\boldsymbol{p}_1 & A\boldsymbol{p}_2 & \cdots & A\boldsymbol{p}_n \end{bmatrix} = \begin{bmatrix} \lambda_1\boldsymbol{p}_1 & \lambda_2\boldsymbol{p}_2 & \cdots & \lambda_n\boldsymbol{p}_n \end{bmatrix}.$$

Equating columns implies $A\boldsymbol{p}_1 = \lambda_1\boldsymbol{p}_1$, $A\boldsymbol{p}_2 = \lambda_2\boldsymbol{p}_2$, $\ldots$, $A\boldsymbol{p}_n = \lambda_n\boldsymbol{p}_n$. As the matrix P is invertible, all its columns must be non-zero (Theorem 6.2.5a). Hence $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n$ are eigenvectors

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

of matrix A corresponding to eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$. Since matrix P is invertible, Theorem 7.2.31k implies the columns vectors $\boldsymbol{p}_1, \boldsymbol{p}_2, \dots, \boldsymbol{p}_n$ are linearly independent.

Second, suppose matrix A has n linearly independent eigenvectors $\boldsymbol{p}_1, \boldsymbol{p}_2, \dots, \boldsymbol{p}_n$ with corresponding eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$. Then follow the above argument backwards to deduce $AP = PD$ for invertible matrix $P = \begin{bmatrix}\boldsymbol{p}_1 & \boldsymbol{p}_2 & \cdots & \boldsymbol{p}_n\end{bmatrix}$, and hence A is diagonalisable. Lastly, in these arguments, P is the matrix of eigenvectors and the diagonal of D are the corresponding eigenvalues, as required. □

Example 7.3.5. Recall Example 7.0.3 found the triangular matrix

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

has eigenvalues -3, -4 and 4 (from its diagonal) and corresponding eigenvectors are proportional to $(1,0,0)$, $(-2,1,0)$ and $(\frac{1}{14}, \frac{1}{4}, 1)$. Is matrix A diagonalisable?

Solution: These three eigenvectors are linearly independent as they correspond to distinct eigenvalues (Theorem 7.2.10). Hence the matrix is diagonalisable.

The previous paragraph answer the question. But further, forming these eigenvectors into the columns of matrix

$$P = \begin{bmatrix} 1 & -2 & \frac{1}{14} \\ 0 & 1 & \frac{1}{4} \\ 0 & 0 & 1 \end{bmatrix},$$

we know (Theorem 7.3.4) $P^{-1}AP = \operatorname{diag}(-3,-4,4)$ where the eigenvalues appear in the same order as that of the eigenvectors in P.

One may check this by hand or with MATLAB/Octave. Enter the matrices with

```
A=[-3 2 0;0 -4 2;0 0 4]
P=[1 -2 1/14;0 1 1/4;0 0 1]
```

then compute $D = P^{-1}AP$ with `D=P\A*P` to find as required the following diagonal result

```
D =
  -3.0000   0.0000   0.0000
   0.0000  -4.0000   0.0000
   0.0000   0.0000   4.0000
```

■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Example 7.3.6. Recall the Sierpinski network of Example 4.1.15 (shown in the margin). Is the 9×9 matrix encoding the network diagonalisable?

$$A = \begin{bmatrix} -3 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & -2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & -3 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -3 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & -3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -3 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -2 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & -3 \end{bmatrix}.$$

Solution: In that example we used MATLAB/Octave command `[V,D]=eig(A)` to compute a matrix of eigenvectors `V` and the corresponding diagonal matrix of eigenvalues `D` where (2 d.p.)

```
V =
 -0.41  0.51 -0.16 -0.21 -0.45  0.18 -0.40  0.06  0.33
  0.00 -0.13  0.28  0.63  0.13 -0.18 -0.58 -0.08  0.33
  0.41 -0.20 -0.49 -0.42  0.32  0.01 -0.36 -0.17  0.33
 -0.41 -0.11  0.52 -0.42  0.32  0.01  0.14 -0.37  0.33
 -0.00 -0.18 -0.26  0.37 -0.22  0.51  0.36 -0.46  0.33
  0.41  0.53  0.07  0.05 -0.10 -0.51  0.33 -0.23  0.33
 -0.41 -0.39 -0.36  0.05 -0.10 -0.51  0.25  0.31  0.33
  0.00  0.31 -0.03  0.16  0.55  0.34  0.22  0.55  0.33
  0.41 -0.33  0.42 -0.21 -0.45  0.18  0.03  0.40  0.33
D =
 -5.00     0     0     0     0     0     0     0     0
     0 -4.30     0     0     0     0     0     0     0
     0     0 -4.30     0     0     0     0     0     0
     0     0     0 -3.00     0     0     0     0     0
     0     0     0     0 -3.00     0     0     0     0
     0     0     0     0     0 -3.00     0     0     0
     0     0     0     0     0     0 -0.70     0     0
     0     0     0     0     0     0     0 -0.70     0
     0     0     0     0     0     0     0     0 -0.00
```

Since matrix A is symmetric, MATLAB/Octave computes for us an orthogonal matrix V with columns eigenvectors. Since V is orthogonal its columns are orthonormal and hence its columns are linearly independent (Theorem 7.2.6). Since there exist nine linearly independent eigenvectors, the matrix A is diagonalisable. Further, the product $V^{-1}AV = D$ for the above diagonal matrix D of eigenvalues in the order of the eigenvectors in V. (Also, since V is orthogonal, $V^{\mathrm{T}}AV = D$.)

■

Example 7.3.7. Recall Example 7.1.9 found eigenvalues and corresponding eigenspaces for various matrices. Revisit these cases and show none of the matrices are diagonalisable.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(a) Matrix $A = \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$ had one eigenvalue $\lambda = 3$ with multiplicity two and corresponding eigenspace $\mathbb{E}_3 = \text{span}\{(1,0)\}$. This matrix is not diagonalisable as it has only one linearly independent eigenvector, such as $(1,0)$ or any non-zero multiple, and it needs two.

(b) Matrix $B = \begin{bmatrix} -1 & 1 & -2 \\ -1 & 0 & -1 \\ 0 & -3 & 1 \end{bmatrix}$ has eigenvalues $\lambda = -2$ (multiplicity one) and $\lambda = 1$ (multiplicity two). The corresponding eigenspaces are $\mathbb{E}_{-2} = \text{span}\{(1,1,1)\}$ and $\mathbb{E}_1 = \text{span}\{(-1,0,1)\}$. Thus the matrix has only two linearly independent eigenvectors, one from each eigenspace, and it needs three to be diagonalisable.

(c) Matrix $C = \begin{bmatrix} -1 & 0 & -2 \\ 0 & -3 & 2 \\ 0 & -2 & 1 \end{bmatrix}$ has only the eigenvalue $\lambda = -1$ with multiplicity three. The corresponding eigenspace $\mathbb{E}_{-1} = \text{span}\{(1,0,0)\}$. With only one linearly independent eigenvector, the matrix is not diagonalisable.

■

Example 7.3.8. Use the results of Example 7.1.10 to show the following matrix is diagonalisable:

$$A = \begin{bmatrix} 0 & 3 & 0 & 0 & 0 \\ 1 & 0 & 3 & 0 & 0 \\ 0 & 1 & 0 & 3 & 0 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Solution: Example 7.1.10 derived the five eigenvalues are $\lambda = 0, \pm\sqrt{3}, \pm 3$, all of multiplicity one. Further, the corresponding eigenspaces are

$$\begin{aligned} &\mathbb{E}_0 = \text{span}\{(9,0,-3,0,1)\}, \\ &\mathbb{E}_{\pm\sqrt{3}} = \text{span}\{(-9,\mp 3\sqrt{3},0,\pm\sqrt{3},1)\}, \\ &\mathbb{E}_{\pm 3} = \text{span}\{(9,\pm 9,6,\pm 3,1)\}. \end{aligned}$$

Here there are five linearly independent eigenvectors, one from each distinct eigenspace (Theorem 7.2.10). Since A is a 5×5 matrix it is thus diagonalisable. Further, Theorem 7.3.4 establishes that the matrix formed from the columns of the five eigenvectors will be a

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

possible matrix [7]

$$P = \begin{bmatrix} 9 & -9 & -9 & 9 & 9 \\ 0 & -3\sqrt{3} & 3\sqrt{3} & 9 & -9 \\ -3 & 0 & 0 & 6 & 6 \\ 0 & \sqrt{3} & -\sqrt{3} & 3 & -3 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

Lastly, Theorem 7.3.4 establishes the diagonal matrix $P^{-1}AP = D = \operatorname{diag}(0, \sqrt{3}, -\sqrt{3}, 3, -3)$ is that of the eigenvalues in the order corresponding to the eigenvectors in P.

■

These examples illustrate a widely useful property. The 5×5 matrix in Example 7.3.8 has five distinct eigenvalues whose corresponding eigenvectors are necessarily linearly independent (Theorem 7.2.10) and so diagonalise the matrix (Theorem 7.3.4). The 3×3 matrix in Example 7.3.5 has three distinct eigenvalues whose corresponding eigenvectors are necessarily linearly independent (Theorem 7.2.10) and so diagonalise the matrix (Theorem 7.3.4). However, the matrices of Examples 7.3.6 and 7.3.7 have repeated eigenvalues—eigenvalues of multiplicity two or more—and these matrices may (Example 7.3.6) or may not (Example 7.3.7) be diagonalisable. The following theorem confirms that matrices with as many *distinct* eigenvalues as the size of the matrix are always diagonalisable.

Theorem 7.3.9. *If an $n \times n$ square matrix A has n distinct eigenvalues, then A is diagonalisable. Consequently, and allowing complex eigenvalues, a non-diagonalisable matrix must be non-symmetric and must have at least one repeated eigenvalue (an eigenvalue with multiplicity two or more).*

Proof. First, let $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_n$ be eigenvectors corresponding to the n distinct eigenvalues of matrix A. (Recall that Theorem 7.1.1 establishes that there cannot be more than n eigenvalues.) As the corresponding eigenvalues are distinct, Theorem 7.2.10 establishes that these eigenvectors are linearly independent. Theorem 7.3.4 then establishes the matrix A is diagonalisable.

Second, the converse of the first statement in the theorem then also holds. Since an $n \times n$ matrix has n eigenvalues, when counted accordingly to multiplicity and allowing for complex eigenvalues (Procedure 7.1.8), a non-diagonalisable matrix must have at least one repeated eigenvalue. But further, by Theorem 4.2.18 a symmetric matrix is always diagonalisable: hence a non-diagonalisable matrix must also be non-symmetric. □

[7] One could also scale each column of P by a different arbitrary non-zero constant, and the diagonalisation still holds.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Example 7.3.10. From the given information, are the matrices diagonalisable?

(a) The only eigenvalues of a 4×4 matrix are 1.8, -3, 0.4 and 3.2.

Solution: Theorem 7.3.9 implies the matrix must be diagonalisable.

(b) The only eigenvalues of a 5×5 matrix are 1.8, -3, 0.4 and 3.2.

Solution: Here there are only four distinct eigenvalues of the 5×5 matrix. Theorem 7.3.9 does not apply as the precondition that there be five distinct eigenvalues is not met: the matrix may or may not be diagonalisable—it is unknowable on this information.

(c) The only eigenvalues of a 3×3 matrix are 1.8, -3, 0.4 and 3.2.

Solution: An error has been made in determining the eigenvalues as a 3×3 matrix has at most three distinct eigenvalues (Theorem 7.1.1). Because of the error, we cannot answer.

■

Example 7.3.11. MATLAB/Octave computes the eigenvalues of matrix

$$A = \begin{bmatrix} -1 & 2 & -2 & 1 & -2 \\ -3 & -1 & -2 & 5 & 6 \\ 3 & 1 & 6 & -2 & -1 \\ 1 & 1 & 2 & 1 & -1 \\ 7 & 5 & -3 & 0 & 0 \end{bmatrix}$$

via `eig(A)` and reports them to be (2 d.p.)

```
ans =
  -3.45 + 3.50i
  -3.45 - 3.50i
   5.00
   5.00
   1.91
```

Is the matrix diagonalisable?

Solution: The matrix appears to have only four distinct eigenvalues (two of them complex valued), and so on the given information Theorem 7.3.9 cannot determine whether the matrix is diagonalisable or not.

However, upon reporting the eigenvalues to four decimal places we find the two eigenvalues of `5.00` (2 d.p.) more precisely are two separate eigenvalues of `5.0000` and `4.9961`. Hence this matrix has

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

five distinct eigenvalues and so Theorem 7.3.9 implies the matrix is diagonalisable. [8]

■

Theorem 7.3.12. *The dimension of the eigenspace $\mathbb{E}_{\lambda_j}$ corresponding to a given eigenvalue λ_j of a matrix A is less than or equal to the multiplicity of λ_j; that is, $1 \le \dim \mathbb{E}_{\lambda_j} \le$ multiplicity of λ_j. (Recall that, from Definition 4.1.11, for a symmetric matrix, $\dim \mathbb{E}_{\lambda_j} =$ multiplicity of λ_j.)*

Proof. Suppose λ_j is an eigenvalue of matrix A and $\dim \mathbb{E}_{\lambda_j} = p < n$ (the case $p = n$ is proved by Exercise 7.3.7). Then choose p orthonormal vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_p$ to span $\mathbb{E}_{\lambda_j}$: vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_p$ are eigenvectors as they are in the eigenspace. Let

$$P = \begin{bmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_p & \boldsymbol{w}_{p+1} & \cdots & \boldsymbol{w}_n \end{bmatrix}$$

be any orthogonal matrix with $\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_p$ as its first p columns. Equivalently, write as the partitioned matrix $P = \begin{bmatrix} V & W \end{bmatrix}$ for corresponding $n \times p$ and $n \times (n-p)$ matrices. Since P is orthogonal, its inverse $P^{-1} = P^{\mathrm{T}} = \begin{bmatrix} V^{\mathrm{T}} \\ W^{\mathrm{T}} \end{bmatrix}$. Since the columns of V are eigenvectors corresponding to eigenvalue λ_j, $AV = A\begin{bmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_p \end{bmatrix} = \begin{bmatrix} A\boldsymbol{v}_1 & \cdots & A\boldsymbol{v}_p \end{bmatrix} = \begin{bmatrix} \lambda_j\boldsymbol{v}_1 & \cdots & \lambda_j\boldsymbol{v}_p \end{bmatrix} = \lambda_j \begin{bmatrix} \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_p \end{bmatrix} = \lambda_j V$.

Now consider

$$\begin{aligned} P^{-1}AP &= \begin{bmatrix} V^{\mathrm{T}} \\ W^{\mathrm{T}} \end{bmatrix} A \begin{bmatrix} V & W \end{bmatrix} = \begin{bmatrix} V^{\mathrm{T}}AV & V^{\mathrm{T}}AW \\ W^{\mathrm{T}}AV & W^{\mathrm{T}}AW \end{bmatrix} \\ &= \begin{bmatrix} \lambda_j V^{\mathrm{T}}V & V^{\mathrm{T}}AW \\ \lambda_j W^{\mathrm{T}}V & W^{\mathrm{T}}AW \end{bmatrix} = \begin{bmatrix} \lambda_j I_p & V^{\mathrm{T}}AW \\ O & W^{\mathrm{T}}AW \end{bmatrix} \end{aligned}$$

where the last equality follows from the orthonormality of columns of $P = \begin{bmatrix} V & W \end{bmatrix}$. Then the characteristic polynomial of matrix A becomes

$$\begin{aligned} &\det(A - \lambda I_n) \\ &= \det(PP^{-1}APP^{-1} - \lambda PP^{-1}) \quad (\text{as } PP^{-1} = I_n) \\ &= \det[P(P^{-1}AP - \lambda I_n)P^{-1}] \\ &= \det P \det(P^{-1}AP - \lambda I_n) \det(P^{-1}) \quad (\text{product Thm. 6.1.12}) \\ &= \det P \det(P^{-1}AP - \lambda I_n) \frac{1}{\det P} \quad (\text{inverse Thm. 6.1.23}) \\ &= \det(P^{-1}AP - \lambda I_n) \end{aligned}$$

[8] Nonetheless, in an application where errors are significant then the matrix may be effectively non-diagonalisable. Such effective non-diagonalisability is indicated by poor conditioning of the matrix of eigenvectors which here has the poor `rcond` of 0.0004 (Procedure 2.2.5).

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$= \det \begin{bmatrix} (\lambda_j - \lambda)I_p & V^{\mathrm{T}}AW \\ O & W^{\mathrm{T}}AW - \lambda I_{n-p} \end{bmatrix} \quad \text{(by above } P^{-1}AP\text{)}$$
$$= (\lambda_j - \lambda)^p \det(W^{\mathrm{T}}AW - \lambda I_{n-p})$$

by p successive first column expansions of the determinant (Theorem 6.2.20). Because of the factor $(\lambda_j - \lambda)^p$ in the characteristic polynomial of A, the eigenvalue λ_j must have multiplicity of at least $p = \dim \mathbb{E}_{\lambda_j}$ (there may be more factors of $(\lambda_j - \lambda)$ hidden within $\det(W^{\mathrm{T}}AW - \lambda I_{n-p})$). □

Example 7.3.13. Show the following matrix has one eigenvalue of multiplicity three, and the corresponding eigenspace has dimension two:

$$A = \begin{bmatrix} 0 & 5 & 6 \\ -8 & 22 & 24 \\ 6 & -15 & -16 \end{bmatrix}$$

Solution: Find eigenvalues via the characteristic polynomial

$$\begin{aligned} \det(A - \lambda I) &= \begin{vmatrix} -\lambda & 5 & 6 \\ -8 & 22-\lambda & 24 \\ 6 & -15 & -16-\lambda \end{vmatrix} \\ &= -\lambda(22-\lambda)(-16-\lambda) + 5\cdot 24\cdot 6 + 6(-8)(-15) \\ &\quad - 6(22-\lambda)6 + \lambda 24(-15) - 5(-8)(-16-\lambda) \\ &= \cdots \\ &= -\lambda^3 + 6\lambda^2 - 12\lambda + 8 \\ &= -(\lambda - 2)^3. \end{aligned}$$

This characteristic polynomial is zero only for eigenvalue $\lambda = 2$ which is of multiplicity three.

The corresponding eigenspace comes from solving $(A - \lambda I)\boldsymbol{x} = \mathbf{0}$ which here is

$$\begin{bmatrix} -2 & 5 & 6 \\ -8 & 20 & 24 \\ 6 & -15 & -18 \end{bmatrix} \boldsymbol{x} = \mathbf{0}.$$

Observe that the second row is just four times the first row, and the third row is $(-3) \times$ the first row, hence all three equations in this system are equivalent to just the one from the first row, namely $-2x_1 + 5x_2 + 6x_3 = 0$. A general solution of this equation is $x_1 = \frac{5}{2}x_2 - 3x_3$. That is all solutions are $\boldsymbol{x} = (\frac{5}{2}x_2 - 3x_3, x_2, x_3) = x_2(\frac{5}{2}, 1, 0) + x_3(-3, 0, 1)$. Hence all solutions form the two dimensional eigenspace $\mathbb{E}_2 = \operatorname{span}\{(\frac{5}{2}, 1, 0), (-3, 0, 1)\}$.

■

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Example 7.3.14. Use MATLAB/Octave to find the eigenvalues and the dimension of the eigenspaces of the matrix

$$B = \begin{bmatrix} 344 & -1165 & -149 & -1031 & 1065 & -2816 \\ 90 & -306 & -38 & -272 & 280 & -742 \\ -45 & 140 & 12 & 117 & -115 & 302 \\ 135 & -470 & -70 & -421 & 445 & -1175 \\ -165 & 555 & 67 & 493 & -506 & 1338 \\ -105 & 360 & 48 & 322 & -335 & 886 \end{bmatrix}.$$

Solution: In MATLAB/Octave enter the matrix with

```
B=[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]
```

Then `[V,D]=eig(B)` computes something like the following (2 d.p.)

```
V =
  -0.19   0.19  -0.45   0.75  -0.20   0.15
  -0.38   0.38   0.12   0.26  -0.03   0.00
  -0.58   0.58   0.08  -0.00  -0.54   0.56
  -0.00  -0.00  -0.83   0.28  -0.50   0.52
   0.58  -0.58  -0.29  -0.45  -0.64   0.63
   0.38  -0.38   0.08  -0.29  -0.04   0.03
D =
   4.00      0      0      0      0      0
      0   4.00      0      0      0      0
      0      0   4.00      0      0      0
      0      0      0  -1.00      0      0
      0      0      0      0  -1.00      0
      0      0      0      0      0  -1.00
```

Evidently, the matrix B has two eigenvalues, $\lambda = 4$ and $\lambda = -1$, both of multiplicity three (although due to round-off error MATLAB/Octave will report these with errors of about 10^{-5} (section 7.1.2)—possibly complex errors in which case ignore small complex parts). For each eigenvalue MATLAB/Octave reports three corresponds columns of `V` containing corresponding eigenvectors. These eigenvectors do span the eigenspace, but are not necessarily linearly independent.

- For eigenvalue $\lambda = 4$ the first two columns of `V` are clearly the negative of each other, and so are essentially the same eigenvector. The third column of `V` is clearly not proportional to the first two columns and so is linearly independent (Theorem 7.2.8). Thus MATLAB/Octave has computed only two linearly independent eigenvectors, either the first and third

column, or the second and third column. Consequently, the dimension $\dim \mathbb{E}_4 = 2$.

One can confirm this dimensionality by computing the singular values of the first three columns of `V` with `svd(V(:,1:3))` to find they are 1.4278, 0.9806 and 0.0000. The two non-zero singular values indicate the dimension of the span is two (Procedure 3.4.23).

- For eigenvalue $\lambda = -1$ the last three columns of `V` look linearly independent and so we suspect the eigenspace dimension $\dim \mathbb{E}_{-1} = 3$.

  To confirm the eigenspace dimensionality (Procedure 3.4.23) compute `svd(V(:,4:6))` to find the three singular values are 1.4136, 1.0001 and 0.0414. Since all three singular values are non-zero, $\dim \mathbb{E}_{-1} = 3$.

MATLAB/Octave 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 small variations in the hidden MATLAB/Octave algorithm may produce quite large changes in the matrix `V` for repeated eigenvalues. However, each eigenspace spanned by the appropriate columns of `V` is robust. ■

7.3.1 Solve systems of differential equations

Population modelling The population modelling seen so far (section 7.1.3) expressed the changes of the population over discrete intervals in time. 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.

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 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 $(y, z) = \boldsymbol{y} = P\boldsymbol{Y}$ where the differential equations for $\boldsymbol{Y}$ are simple?

- First, form the differential equations into a matrix-vector system:

$$\begin{bmatrix} dy/dt \\ dz/dt \end{bmatrix} = \begin{bmatrix} y - 4z \\ -y + z \end{bmatrix} = \begin{bmatrix} 1 & -4 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix}.$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

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, see what happens when we transform to some, as yet unknown, new variables $\boldsymbol{Y}(t)$ such that $\boldsymbol{y} = P\boldsymbol{Y}$ for some invertible matrix P. Under such a transform: $\frac{d\boldsymbol{y}}{dt} = \frac{d}{dt}P\boldsymbol{Y} = P\frac{d\boldsymbol{Y}}{dt}$; and $A\boldsymbol{y} = AP\boldsymbol{Y}$. Hence substituting such an assumed transformation into the differential equations 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 diagonalise the matrix on the right-hand side. The procedure is to choose the columns of P to be eigenvectors of the matrix (Theorem 7.3.4).

- Third, here the matrix $A = \begin{bmatrix} 1 & -4 \\ -1 & 1 \end{bmatrix}$ has characteristic polynomial $\det(A - \lambda I) = (1-\lambda)^2 - 4$. This is zero for $(1-\lambda)^2 = 4$, that is, $(1-\lambda) = \pm 2$. Hence the eigenvalues $\lambda = 1 \pm 2 = 3, -1$.
  - For eigenvalue $\lambda_1 = 3$ the corresponding eigenvectors satisfy
  $$(A - \lambda_1 I)\boldsymbol{p}_1 = \begin{bmatrix} -2 & -4 \\ -1 & -2 \end{bmatrix}\boldsymbol{p}_1 = \boldsymbol{0}\,,$$
  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 (any scalar multiple of the two columns would also work)

$$P = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} \implies \frac{d\boldsymbol{Y}}{dt} = \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix}\boldsymbol{Y}\,.$$

- Fourth, having diagonalised the matrix, expand this diagonalised set of differential equations to write this system in terms of components:

$$\frac{dY_1}{dt} = 3Y_1 \quad \text{and} \quad \frac{dY_2}{dt} = -Y_2\,.$$

  Each of these differential equations have well-known exponential solutions, respectively $Y_1 = c_1 e^{3t}$ and $Y_2 = c_2 e^{-t}$, for any constants c_1 and c_2.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

- Lastly, what does this mean for the original problem? From 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 diagonalisation of the matrix empowers us to solve complicated systems of differential equations as a set of simple systems.

Such a general solution makes predictions. For example, suppose at time zero ($t = 0$) the initial population of female y-animals is 22 and the population of female z-animals is 9. From the above general solution we then know that at time $t = 0$

$$\begin{bmatrix} 22 \\ 9 \end{bmatrix} = \begin{bmatrix} y(0) \\ z(0) \end{bmatrix} = \begin{bmatrix} 2c_1 e^{3\cdot 0} + 2c_2 e^{-0} \\ -c_1 e^{3\cdot 0} + c_2 e^{-0} \end{bmatrix} = \begin{bmatrix} 2c_1 + 2c_2 \\ -c_1 + c_2 \end{bmatrix}$$

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 above graph of this solution shows that the population of y-animals grows in time, whereas the population of z-animals crashes and becomes extinct at about time 0.6 years.[9]

The next theorem confirms that the same approach solves general systems of differential equations: it corresponds to Theorem 7.1.16 for discrete dynamics.

[9] After time 0.6 years the differential equation model and its predictions becomes meaningless as there is no biological meaning to a negative number of animals z.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Theorem 7.3.15. *Let $n \times n$ square matrix A be diagonalisable by matrix $P = \begin{bmatrix} \boldsymbol{p}_1 & \boldsymbol{p}_2 & \cdots & \boldsymbol{p}_n \end{bmatrix}$ whose columns are eigenvectors corresponding to eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Then a general solution $\boldsymbol{x}(t)$ to the differential equation system $d\boldsymbol{x}/dt = A\boldsymbol{x}$ is the linear combination*

$$\boldsymbol{x}(t) = c_1\boldsymbol{p}_1 e^{\lambda_1 t} + c_2\boldsymbol{p}_2 e^{\lambda_2 t} + \cdots + c_n\boldsymbol{p}_n e^{\lambda_n t} \tag{7.4}$$

for arbitrary constants $c_1, c_2, \ldots, c_n$.

Proof. First, instead of finding solutions for $\boldsymbol{x}(t)$ directly, let's write the differential equations in terms of the alternate basis for $\mathbb{R}^n$, basis $\mathcal{P} = \{\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n\}$ (as $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n$ are linearly independent). That is, solve for the coordinates $\boldsymbol{X}(t) = [\boldsymbol{x}(t)]_{\mathcal{P}}$ with respect to basis $\mathcal{P}$. Now recall (Theorem 7.2.27) that $\boldsymbol{X} = [\boldsymbol{x}]_{\mathcal{P}}$ means that $\boldsymbol{x} = X_1\boldsymbol{p}_1 + X_2\boldsymbol{p}_2 + \cdots + X_n\boldsymbol{p}_n = P\boldsymbol{X}$. Substitute this into the differential equation $d\boldsymbol{x}/dt = A\boldsymbol{x}$ requires $\frac{d}{dt}(P\boldsymbol{X}) = A(P\boldsymbol{X})$ which is the same as $P\frac{d\boldsymbol{X}}{dt} = AP\boldsymbol{X}$. Since matrix P is invertibble, this equation is the same as $\frac{d\boldsymbol{X}}{dt} = P^{-1}AP\boldsymbol{X}$ Because the columns of matrix P are eigenvectors, the product $P^{-1}AP$ is the diagonal matrix $D = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, hence the system becomes $\frac{d\boldsymbol{X}}{dt} = D\boldsymbol{X}$. Because matrix D is diagonal, this is a much simpler system of differential equations. All the rows of the system are

$$\frac{dX_1}{dt} = \lambda_1 X_1\,, \quad \frac{dX_2}{dt} = \lambda_2 X_2\,, \quad \ldots, \quad \frac{dX_n}{dt} = \lambda_n X_n\,.$$

Each of these have general solution

$$X_1 = c_1 e^{\lambda_1 t}\,, \quad X_2 = c_2 e^{\lambda_2 t}\,, \quad \ldots, \quad X_n = c_n e^{\lambda_n t},$$

where $c_1, c_2, \ldots, c_n$ are arbitrary constants. To rewrite this solution for the original coordinates $\boldsymbol{x}$ use

$$\begin{aligned} \boldsymbol{x} &= P\boldsymbol{X} \\ &= \begin{bmatrix} \boldsymbol{p}_1 & \boldsymbol{p}_2 & \cdots & \boldsymbol{p}_n \end{bmatrix} \begin{bmatrix} c_1 e^{\lambda_1 t} \\ c_2 e^{\lambda_2 t} \\ \vdots \\ c_n e^{\lambda_n t} \end{bmatrix} \\ &= c_1\boldsymbol{p}_1 e^{\lambda_1 t} + c_2\boldsymbol{p}_2 e^{\lambda_2 t} + \cdots + c_n\boldsymbol{p}_n e^{\lambda_n t} \end{aligned}$$

to derive the solution (7.4).

Second, being able to use the constants $\boldsymbol{c} = (c_1, c_2, \ldots, c_n)$ to match any given initial condition shows formula (7.4) is a general solution. Suppose the value of $\boldsymbol{x}(0)$ is given. Recalling $e^0 = 1$, formula (7.4) evaluated at $t = 0$ requires

$$\begin{aligned} \boldsymbol{x}(0) &= c_1\boldsymbol{p}_1 e^{\lambda_1 0} + c_2\boldsymbol{p}_2 e^{\lambda_2 0} + \cdots + c_n\boldsymbol{p}_n e^{\lambda_n 0} \\ &= c_1\boldsymbol{p}_1 + c_2\boldsymbol{p}_2 + \cdots + c_n\boldsymbol{p}_n \\ &= \begin{bmatrix} \boldsymbol{p}_1 & \boldsymbol{p}_2 & \cdots & \boldsymbol{p}_n \end{bmatrix} \boldsymbol{c} \end{aligned}$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$= P\boldsymbol{c}\,.$$

If matrix P is invertible, then choose constants $\boldsymbol{c} = P^{-1}\boldsymbol{x}(0)$ for any given $\boldsymbol{x}(0)$. Theorems 7.3.4 and 7.2.31 establish that such an invertible P exists for diagonalisable matrices A. □

Example 7.3.16. Find (by hand) a general solution to the system of differential equations $\frac{du}{dt} = -2u + 2v$, $\frac{dv}{dt} = u - 2v + w$, and $\frac{dw}{dt} = 2v - 2w$.

Solution: Let vector $\boldsymbol{u} = (u, v, w)$, and then form the differential equations into the matrix-vector system

$$\begin{aligned}\frac{d\boldsymbol{u}}{dt} &= \begin{bmatrix}\frac{du}{dt}\\ \frac{dv}{dt}\\ \frac{dw}{dt}\end{bmatrix} = \begin{bmatrix}-2u+2v\\ u-2v+w\\ 2v-2w\end{bmatrix}\\ &= \begin{bmatrix}-2 & 2 & 0\\ 1 & -2 & 1\\ 0 & 2 & -2\end{bmatrix}\begin{bmatrix}u\\ v\\ w\end{bmatrix} = \underbrace{\begin{bmatrix}-2 & 2 & 0\\ 1 & -2 & 1\\ 0 & 2 & -2\end{bmatrix}}_{A}\boldsymbol{u}\,.\end{aligned}$$

To use Theorem 7.3.15 we need eigenvalues and eigenvectors of the matrix A. Here the characteristic polynomial of A is, using (6.1),

$$\begin{aligned}\det(A-\lambda I) &= \det\begin{bmatrix}-2-\lambda & 2 & 0\\ 1 & -2-\lambda & 1\\ 0 & 2 & -2-\lambda\end{bmatrix}\\ &= -(2+\lambda)^3 + 0 + 0 - 0 + 2(2+\lambda) + 2(2+\lambda)\\ &= (2+\lambda)\left[-(2+\lambda)^2 + 4\right]\\ &= (2+\lambda)[-\lambda^2 - 4\lambda]\\ &= -\lambda(\lambda+2)(\lambda+4).\end{aligned}$$

This determinant is only zero for eigenvalues $\lambda = 0, -2, -4$.

- For eigenvalue $\lambda = 0$, corresponding eigenvectors $\boldsymbol{p}$ satisfy

$$(A - 0I)\boldsymbol{p} = \begin{bmatrix}-2 & 2 & 0\\ 1 & -2 & 1\\ 0 & 2 & -2\end{bmatrix}\boldsymbol{p} = \boldsymbol{0}\,.$$

  The last row of this equation requires $p_3 = p_2$, and the first row requires $p_1 = p_2$. Hence all solutions may be written as $\boldsymbol{p} = (p_2, p_2, p_2)$. Choose any one, say $\boldsymbol{p} = (1, 1, 1)$.

- For eigenvalue $\lambda = -2$, corresponding eigenvectors $\boldsymbol{p}$ satisfy

$$(A + 2I)\boldsymbol{p} = \begin{bmatrix}0 & 2 & 0\\ 1 & 0 & 1\\ 0 & 2 & 0\end{bmatrix}\boldsymbol{p} = \boldsymbol{0}\,.$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

The first and last rows of this equation require $p_2 = 0$, and the second row requires $p_3 = -p_1$. Hence all solutions may be written as $\boldsymbol{p} = (p_1, 0, -p_1)$. Choose any one, say $\boldsymbol{p} = (1, 0, -1)$.

- For eigenvalue $\lambda = -4$, corresponding eigenvectors $\boldsymbol{p}$ satisfy

$$(A + 4I)\boldsymbol{p} = \begin{bmatrix} 2 & 2 & 0 \\ 1 & 2 & 1 \\ 0 & 2 & 2 \end{bmatrix} \boldsymbol{p} = \boldsymbol{0}\,.$$

The last row of this equation requires $p_3 = -p_2$, and the first row requires $p_1 = -p_2$. Hence all solutions may be written as $\boldsymbol{p} = (-p_2, p_2, -p_2)$. Choose any one, say $\boldsymbol{p} = (-1, 1, -1)$.

With these three distinct eigenvalues, corresponding eigenvectors are linearly independent, and so Theorem 7.3.15 gives a general solution of the differential equations as

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = c_1 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} e^{0t} + c_2 \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} e^{-2t} + c_3 \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix} e^{-4t}.$$

That is, $u(t) = c_1 + c_2 e^{-2t} - c_3 e^{-4t}$, $v(t) = c_1 + c_3 e^{-4t}$, and $w(t) = c_1 - c_2 e^{2t} - c_3 e^{-4t}$ for any constants c_1, c_2 and c_3.

■

Example 7.3.17. Use the general solution derived in Example 7.3.16 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 $w(0) = 4$.

Solution: Evaluating the general solution

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = c_1 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} e^{-2t} + c_3 \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix} e^{-4t}$$

at time $t = 0$ gives, using the initial conditions and $e^0 = 1$,

$$\begin{bmatrix} 0 \\ 0 \\ 4 \end{bmatrix} = c_1 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + c_3 \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix} = \begin{bmatrix} c_1 + c_2 - c_3 \\ c_1 + c_3 \\ c_1 - c_2 - c_3 \end{bmatrix}.$$

Solving by hand, the second row requires $c_3 = -c_1$, so the first row then requires $c_1 + c_2 + c_1 = 0$, that is, $c_2 = -2c_1$. Putting both of these into the third row requires $c_1 + 2c_1 + c_1 = 4$, that is, $c_1 = 1$. Then $c_2 = -2$ and $c_3 = -1$. Consequently, as drawn in the margin, the particular solution is

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - 2 \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} e^{-2t} - \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix} e^{-4t}$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$= \begin{bmatrix} 1 - 2e^{-2t} + e^{-4t} \\ 1 - e^{-4t} \\ 1 + 2e^{-2t} + e^{-4t} \end{bmatrix} .$$

■

Example 7.3.18. Use MATLAB/Octave to find a general solution to the system of differential equations

$$\begin{aligned} dx_1/dt &= -\tfrac{1}{2}x_1 - \tfrac{1}{2}x_2 + x_3 + 2x_4 \,, \\ dx_2/dt &= -\tfrac{1}{2}x_1 - \tfrac{1}{2}x_2 + 2x_3 + x_4 \,, \\ dx_3/dt &= x_1 + 2x_2 - \tfrac{1}{2}x_3 - \tfrac{1}{2}x_4 \,, \\ dx_4/dt &= 2x_1 + x_2 - \tfrac{1}{2}x_3 - \tfrac{1}{2}x_4 \,. \end{aligned}$$

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.

Solution: Write the system in matrix-vector form $\frac{d\boldsymbol{x}}{dt} = A\boldsymbol{x}$ for vector

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \quad \text{and matrix } A = \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} & 1 & 2 \\ -\frac{1}{2} & -\frac{1}{2} & 2 & 1 \\ 1 & 2 & -\frac{1}{2} & -\frac{1}{2} \\ 2 & 1 & -\frac{1}{2} & -\frac{1}{2} \end{bmatrix} .$$

Enter the matrix into MATLAB/Octave and then find its eigenvalues and eigenvectors as follows

```
A=[-1/2 -1/2 1 2
 -1/2 -1/2 2 1
 1 2 -1/2 -1/2
 2 1 -1/2 -1/2]
[V,D]=eig(A)
```

MATLAB/Octave tells us the eigenvectors and eigenvalues:

```
V =
  -0.5000   0.5000  -0.5000  -0.5000
  -0.5000  -0.5000   0.5000  -0.5000
   0.5000   0.5000   0.5000  -0.5000
   0.5000  -0.5000  -0.5000  -0.5000
D =
  -4.0000        0        0        0
        0  -1.0000        0        0
        0        0   1.0000        0
        0        0        0   2.0000
```

Then Theorem 7.3.15 gives that a general solution of the differential

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

equations is

$$\boldsymbol{x} = c_1 \begin{bmatrix} -\frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} e^{-4t} + c_2 \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix} e^{-t} + c_3 \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix} e^{t} + c_4 \begin{bmatrix} -\frac{1}{2} \\ -\frac{1}{2} \\ -\frac{1}{2} \\ -\frac{1}{2} \end{bmatrix} e^{2t}.$$

Given the specified initial conditions at $t = 0$, when all the above exponentials reduce to $e^0 = 1$, we just need to find the linear combination of the eigenvectors that equals the initial vector $\boldsymbol{x}(0) = (-5, -1, 0, 0)$; that is, we solve $V\boldsymbol{c} = \boldsymbol{x}(0)$. In MATLAB/Octave compute `c=V\[-5;-1;0;0]` to find the vector of coefficients is $\boldsymbol{c} = (3, -2, 2, 3)$. Hence the particular solution is

$$\boldsymbol{x} = \begin{bmatrix} -\frac{3}{2} \\ -\frac{3}{2} \\ \frac{3}{2} \\ \frac{3}{2} \end{bmatrix} e^{-4t} + \begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix} e^{-t} + \begin{bmatrix} -1 \\ 1 \\ 1 \\ -1 \end{bmatrix} e^{t} + \begin{bmatrix} -\frac{3}{2} \\ -\frac{3}{2} \\ -\frac{3}{2} \\ -\frac{3}{2} \end{bmatrix} e^{2t}.$$

■

Example 7.3.19. Find (by hand) a general solution to the system of differential equations $\frac{dy}{dt} = z$ and $\frac{dz}{dt} = -4y$.

Solution: Let vector $\boldsymbol{y} = (y, z)$, and then form the differential equations into the matrix-vector system

$$\frac{d\boldsymbol{y}}{dt} = \begin{bmatrix} \frac{dy}{dt} \\ \frac{dz}{dt} \end{bmatrix} = \begin{bmatrix} z \\ -4y \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -4 & 0 \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 \\ -4 & 0 \end{bmatrix}}_{A} \boldsymbol{y}.$$

To use Theorem 7.3.15 we need eigenvalues and eigenvectors of the matrix A. Here the characteristic polynomial of A is

$$\det(A - \lambda I) = \det \begin{bmatrix} -\lambda & 1 \\ -4 & -\lambda \end{bmatrix} = \lambda^2 + 4.$$

This determinant is only zero for $\lambda^2 = -4$, that is, $\lambda = \pm 2i$ —a pair of complex conjugate eigenvalues.

- For eigenvalue $\lambda = +2i$ the corresponding eigenvectors $\boldsymbol{p}$ satisfy

  $$(A - \lambda I)\boldsymbol{p} = \begin{bmatrix} -2i & 1 \\ -4 & -2i \end{bmatrix} \boldsymbol{p} = \boldsymbol{0}.$$

  The second row of this matrix is $-2i$ times the first row so we just need to satisfy the first row equation $\begin{bmatrix} -2i & 1 \end{bmatrix} \boldsymbol{p} = 0$. This equation is $-2ip_1 + p_2 = 0$, that is, $p_2 = 2ip_1$. Hence all eigenvectors are of the form $\boldsymbol{p} = (1, 2i)p_1$. Choose any one, say $\boldsymbol{p} = (1, 2i)$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

- Similarly, for eigenvalue $\lambda = -2i$ the corresponding eigenvectors $\boldsymbol{p}$ satisfy

$$(A - \lambda I)\boldsymbol{p} = \begin{bmatrix} 2i & 1 \\ -4 & 2i \end{bmatrix} \boldsymbol{p} = \boldsymbol{0}.$$

  The second row of this matrix is $2i$ times the first row so we just need to satisfy the first row equation $\begin{bmatrix} 2i & 1 \end{bmatrix} \boldsymbol{p} = 0$. This equation is $2ip_1 + p_2 = 0$, that is, $p_2 = -2ip_1$. Hence all eigenvectors are of the form $\boldsymbol{p} = (1, -2i)p_1$. Choose any one, say $\boldsymbol{p} = (1, -2i)$.

With these two distinct eigenvalues, corresponding eigenvectors are linearly independent, and so Theorem 7.3.15 gives a general solution of the differential equations as

$$\begin{bmatrix} y \\ z \end{bmatrix} = c_1 \begin{bmatrix} 1 \\ 2i \end{bmatrix} e^{i2t} + c_2 \begin{bmatrix} 1 \\ -2i \end{bmatrix} e^{-i2t}.$$

That is, $y(t) = c_1 e^{i2t} + c_2 e^{-i2t}$ and $z(t) = 2ic_1 e^{i2t} - 2ic_2 e^{-i2t}$ for any constants c_1 and c_2. These formulas answer the exercise. The next examples show that because of the complex exponentials, this solution describes oscillations in time t.

■

Example 7.3.20. Further consider Example 7.3.19. Suppose we additionally know that $y(0) = 3$ and $z(0) = 0$. Find the particular solution that satisfies these two initial conditions.

Solution: Use the derived general solution that $y(t) = c_1 e^{i2t} + c_2 e^{-i2t}$ and $z(t) = 2ic_1 e^{i2t} - 2ic_2 e^{-i2t}$. We find the constants c_1 and c_2 so that these satisfy the conditions $y(0) = 3$ and $z(0) = 0$. Substitute $t = 0$ into the general solution to require, using $e^0 = 1$,

$$\begin{aligned} y(0) &= 3 = c_1 e^{i2\cdot 0} + c_2 e^{-i2\cdot 0} = c_1 + c_2\,, \\ z(0) &= 0 = 2ic_1 e^{i2\cdot 0} - 2ic_2 e^{-i2\cdot 0} = 2ic_1 - 2ic_2\,, \end{aligned}$$

The second of these equations requires $2ic_1 = 2ic_2$, that is, $c_1 = c_2$. The first, $c_1 + c_2 = 3$, then requires that $2c_1 = 3$, that is, $c_1 = 3/2$ and so $c_2 = 3/2$. Hence the particular solution is

$$y(t) = \frac{3}{2}e^{i2t} + \frac{3}{2}e^{-i2t} \quad \text{and} \quad z(t) = 3ie^{i2t} - 3ie^{-i2t}.$$

But recall Euler's formula that $e^{i\theta} = \cos\theta + i\sin\theta$ for any θ. Invoking Euler's formula the above particular solution simplifies:

$$\begin{aligned} y(t) &= \frac{3}{2}[\cos 2t + i\sin 2t] + \frac{3}{2}[\cos(-2t) + i\sin(-2t)] \\ &= \frac{3}{2}\cos 2t + \frac{3}{2}i\sin 2t + \frac{3}{2}\cos 2t - \frac{3}{2}i\sin 2t \end{aligned}$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$
\begin{aligned}
&= 3\cos 2t\,,\\
z(t) &= 3i[\cos 2t + i\sin 2t] - 3i[\cos(-2t) + i\sin(-2t)]\\
&= 3i\cos 2t - 3\sin 2t - 3i\cos 2t - 3\sin 2t\\
&= -6\sin 2t\,.
\end{aligned}
$$

Because $y(t)$ and $z(t)$ are just trigonometric functions of t, they oscillate in time t, as illustrated in the margin.

■

Example 7.3.21. In a real application the complex numbers of the general solution to Example 7.3.19 are usually inconvenient. Instead we often express the solution solely in terms of real quantities as just done in the previous Example 7.3.20. Use Euler's formula, that $e^{i\theta} = \cos\theta + i\sin\theta$ for any θ, to rewrite the general solution of Example 7.3.19 in terms of real functions.

Solution: Here use Euler's formula in the expression for

$$
\begin{aligned}
y(t) &= c_1 e^{i2t} + c_2 e^{-i2t}\\
&= c_1[\cos 2t + i\sin 2t] + c_2[\cos(-2t) + i\sin(-2t)]\\
&= c_1\cos 2t + ic_1\sin 2t + c_2\cos 2t - ic_2\sin 2t\\
&= (c_1 + c_2)\cos 2t + (ic_1 - ic_2)\sin 2t\\
&= C_1\cos 2t + C_2\sin 2t
\end{aligned}
$$

for constants $C_1 = c_1 + c_2$ and $C_2 = i(c_1 - c_2)$. Let's view the arbitrariness in c_1 and c_2 as being 'transferred' to C_1 and C_2, then $y(t) = C_1\cos 2t + C_2\sin 2t$ is a general solution to the differential equations, and is expressed purely in real factors. In such a real form we explicitly see the oscillations in time t through the trigonometric functions $\cos 2t$ and $\sin 2t$.

The function $z(t) = 2ic_1e^{i2t} - 2ic_2e^{-i2t}$ has a corresponding form found by replacing c_1 and c_2 in terms of C_1 and C_2. From above, the constants $C_1 = c_1 + c_2$ and $C_2 = i(c_1 - c_2)$. Adding the first to $\pm i$ times the second determines $C_1 - iC_2 = 2c_1$ and $C_1 + iC_2 = 2c_2$, so that $c_1 = (C_1 - iC_2)/2$ and $c_2 = (C_1 + iC_2)/2$. Then the expression for

$$
\begin{aligned}
z(t) &= 2ic_1e^{i2t} - 2ic_2e^{-i2t}\\
&= 2i\frac{C_1 - iC_2}{2}[\cos 2t + i\sin 2t]\\
&\quad - 2i\frac{C_1 + iC_2}{2}[\cos 2t - i\sin 2t]\\
&= (iC_1 + C_2)[\cos 2t + i\sin 2t]\\
&\quad + (-iC_1 + C_2)[\cos 2t - i\sin 2t]\\
&= iC_1\cos 2t - C_1\sin 2t + C_2\cos 2t + iC_2\sin 2t\\
&\quad - iC_1\cos 2t - C_1\sin 2t + C_2\cos 2t - iC_2\sin 2t
\end{aligned}
$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

$$= -2C_1 \sin 2t + 2C_2 \cos 2t\,.$$

That is, the corresponding general solution $z(t) = -2C_1 \sin 2t + 2C_2 \cos 2t$ is now also expressed in real factors.

■

Example 7.3.22 (oscillating applications). A huge variety of vibrating systems are analogous to the basic oscillations of a mass on a spring, illustrated schematically in the margin. The mass generally will oscillate to and fro. Describe such a system mathematically with two differential equations, and solve the differential equations to confirm it oscillates.

Solution: At any time t, let the position of the mass relative to its rest position be denoted by $x(t)$: that is, we put an x-axis on the picture with $x = 0$ where the mass would stay at rest, as illustrated. At any time t let the mass be moving with velocity $v(t)$ (positive to the right, negative to the left). Then we know one differential equation, that $\frac{dx}{dt} = v$.

Newton's law, that mass × acceleration = force, provides another differential equation. Here the mass is denoted by m, and the acceleration is $\frac{dv}{dt}$. The force on the mass come from the spring: typically the force by the spring is proportional to the stretching of the spring, namely to x. Different springs give different strength forces so let's denote the constant of proportionality by k—a constant that varies from spring to spring. Then the force will be $-kx$ as springs try to pull/push the mass back towards $x = 0$. Consequently Newton's law gives us the differential equation, $m\frac{dv}{dt} = -kx$. Divide by mass m and the differential equation is $\frac{dv}{dt} = -\frac{k}{m}x$.

Write these two differential equations together as a matrix-vector system:

$$\begin{bmatrix} \frac{dx}{dt} \\ \frac{dv}{dt} \end{bmatrix} = \begin{bmatrix} v \\ -\frac{k}{m}x \end{bmatrix}, \quad \text{that is,} \quad \frac{d}{dt}\begin{bmatrix} x \\ v \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\frac{k}{m} & 0 \end{bmatrix}\begin{bmatrix} x \\ v \end{bmatrix}.$$

Theorem 7.3.15 asserts a general solution comes from the eigenvalues and eigenvectors of the matrix.

- Here the characteristic polynomial is

$$\det\begin{bmatrix} -\lambda & 1 \\ -\frac{k}{m} & -\lambda \end{bmatrix} = \lambda^2 + \frac{k}{m}\,.$$

  This polynomial is zero only when the eigenvalues $\lambda = \pm\sqrt{-k/m} = \pm i\sqrt{k/m}$: these are pure imaginary eigenvalues.

- The corresponding eigenvectors $\boldsymbol{p}$ satisfy

$$\begin{bmatrix} \mp i\sqrt{k/m} & 1 \\ -k/m & \mp i\sqrt{k/m} \end{bmatrix}\boldsymbol{p} = \boldsymbol{0}.$$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

The two rows of this equation are satisfied by the corresponding eigenvectors $\boldsymbol{p} \propto (1, \pm i\sqrt{k/m})$.

Then a general solution to the system of differential equations is

$$\begin{bmatrix} x \\ v \end{bmatrix} = c_1 \begin{bmatrix} 1 \\ i\sqrt{k/m} \end{bmatrix} e^{i\sqrt{k/m}t} + c_2 \begin{bmatrix} 1 \\ -i\sqrt{k/m} \end{bmatrix} e^{-i\sqrt{k/m}t}.$$

This formula shows that the mass on the spring generally oscillates as the complex exponentials are oscillatory.

However, in real applications we usually prefer a real algebraic expression. Just as in Example 7.3.21, we make the above formula real by changing from (complex) arbitrary constants c_1 and c_2 to new (real) arbitrary constants C_1 and C_2 where $c_1 = (C_1 - iC_2)/2$ and $c_2 = (C_1 + iC_2)/2$. Substitute these relations into the above general solution, and using Euler's formula, gives the position

$$\begin{aligned}
x(t) &= c_1 e^{i\sqrt{k/m}t} + c_2 e^{-i\sqrt{k/m}t} \\
&= \frac{C_1 - iC_2}{2}\big[\cos(\sqrt{k/m}t) + i\sin(\sqrt{k/m}t)\big] \\
&\quad + \frac{C_1 + iC_2}{2}\big[\cos(\sqrt{k/m}t) - i\sin(\sqrt{k/m}t)\big] \\
&= \frac{C_1}{2}\cos(\sqrt{k/m}t) + i\frac{C_1}{2}\sin(\sqrt{k/m}t) \\
&\quad - i\frac{C_2}{2}\cos(\sqrt{k/m}t) + \frac{C_2}{2}\sin(\sqrt{k/m}t) \\
&\quad + \frac{C_1}{2}\cos(\sqrt{k/m}t) - i\frac{C_1}{2}\sin(\sqrt{k/m}t) \\
&\quad + i\frac{C_2}{2}\cos(\sqrt{k/m}t) + \frac{C_2}{2}\sin(\sqrt{k/m}t) \\
&= C_1\cos(\sqrt{k/m}t) + C_2\sin(\sqrt{k/m}t).
\end{aligned}$$

For all values of the arbitrary constants C_1 and C_2, this formula describes the position $x(t)$ of the mass as purely real oscillations in time. Similarly for the velocity $v(t)$ (Exercise 7.3.12).

■

7.3.2 Exercises

Exercise 7.3.1. Which of the following matrices diagonalise the matrix $Z = \begin{bmatrix} 7 & 12 \\ -2 & -3 \end{bmatrix}$? Show your working.

(a) $P_a = \begin{bmatrix} -2 & 3 \\ 1 & -1 \end{bmatrix}$ (b) $P_b = \begin{bmatrix} 3 & -2 \\ -1 & 1 \end{bmatrix}$

(c) $P_c = \begin{bmatrix} 1 & -1 \\ -2 & 3 \end{bmatrix}$ (d) $P_d = \begin{bmatrix} 1 & 3 \\ 1 & 2 \end{bmatrix}$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(e) $P_e = \begin{bmatrix} 4 & 3 \\ -2 & -1 \end{bmatrix}$ (f) $P_f = \begin{bmatrix} -2 & 3 \\ 2 & -2 \end{bmatrix}$

(g) $P_g = \begin{bmatrix} -1 & 1 \\ 3 & -2 \end{bmatrix}$ (h) $P_h = \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix}$

Exercise 7.3.2. Redo Exercise 7.3.1 by finding which matrices $P_a, \dots, P_h$ diagonalise each of the following matrices.

(a) $A = \begin{bmatrix} 5 & 12 \\ -2 & -5 \end{bmatrix}$ (b) $B = \begin{bmatrix} -3 & -12 \\ 2 & 7 \end{bmatrix}$

(c) $C = \begin{bmatrix} -1 & -1 \\ 6 & 4 \end{bmatrix}$ (d) $D = \begin{bmatrix} 4 & 6 \\ -1 & -1 \end{bmatrix}$

Exercise 7.3.3. Following Example 7.3.2b, prove that the matrix $\begin{bmatrix} k & 1 \\ 0 & k \end{bmatrix}$ is not diagonalisable for any scalar k.

Exercise 7.3.4. In each of the following cases, you are given three linearly independent eigenvectors and corresponding eigenvalues for some 3×3 matrix A. Write down three different matrices P that will diagonalise the matrix A, and for each write down the corresponding diagonal matrix $D = P^{-1}AP$.

(a) $\lambda_1 = -1$, $\boldsymbol{p}_1 = (3, 2, -1)$; $\lambda_2 = 1$, $\boldsymbol{p}_2 = (-4, -2, 2)$; $\lambda_3 = 3$, $\boldsymbol{p}_3 = (-1, 0, 2)$.

(b) $\lambda_1 = -1$, $\boldsymbol{p}_1 = (2, 1, 2)$; $\lambda_2 = -1$, $\boldsymbol{p}_2 = (0, 3, 1)$; $\lambda_3 = 2$, $\boldsymbol{p}_3 = (4, -2, -2)$.

(c) $\lambda_1 = 1$, $\boldsymbol{p}_1 = (3, -7, 2)$; $\lambda_2 = -2$, $\boldsymbol{p}_2 = (-4, -5, 1)$; $\lambda_3 = 4$, $\boldsymbol{p}_3 = (1, 2, 3)$.

(d) $\lambda_1 = 3$, $\boldsymbol{p}_1 = (2, -3, 0)$; $\lambda_2 = 1$, $\boldsymbol{p}_2 = (-1, 2, -6)$; $\lambda_3 = -4$, $\boldsymbol{p}_3 = (-2, -1, -3)$.

Exercise 7.3.5. From the given information, are each of the matrices diagonalisable? Give reasons.

(a) The only eigenvalues of a 2×2 matrix are 2.2 and 0.1.

(b) The only eigenvalues of a 4×4 matrix are 2.2, 1.9, -1.8 and -1.

(c) The only eigenvalue of a 2×2 matrix is 0.7.

(d) The only eigenvalues of a 6×6 matrix are -1.6, 0.3, 0.1 and -2.3.

(e) The only eigenvalues of a 5×5 matrix are -1.7, 1.4, 1.3, 2.4, 0.5 and -2.3.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(f) The only eigenvalues of a 6×6 matrix are 1.2, -0.9, -0.8, 2.2, 0.2 and -0.2.

(g) The MATLAB/Octave function `eig(A)` returns the result

```
ans =
   2.6816
  -0.1445
   0.0798
   0.3844
```

(h) The MATLAB/Octave function `eig(A)` returns the result

```
ans =
   3.0821 + 0.0000i
  -2.7996 + 0.0000i
  -0.7429 + 1.6123i
  -0.7429 - 1.6123i
```

(i) The MATLAB/Octave function `eig(A)` returns the result

```
ans =
  -1.0000
   1.0000
   2.0000
  -1.0000
```

Exercise 7.3.6. For each of the following 3×3 matrices, show with hand algebra that each matrix has one eigenvalue of multiplicity three, and then determine the dimension of the corresponding eigenspace. Also compute the eigenvalue and eigenvectors with MATLAB/Octave: comment on any limitations in the computed 'eigenvalues' and 'eigenvectors'.

(a) $A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$

(b) $B = \begin{bmatrix} 2 & 1 & -2 \\ -8 & -4 & 8 \\ -2 & -1 & 2 \end{bmatrix}$

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

(d) $D = \begin{bmatrix} 7 & 1 & 22 \\ -6 & -1 & -20 \\ -1 & 0 & -3 \end{bmatrix}$

(e) $E = \begin{bmatrix} 3 & 5 & -1 \\ -4 & -6 & 1 \\ -5 & -6 & 0 \end{bmatrix}$

(f) $F = \begin{bmatrix} -6 & -5 & 3 \\ -4 & -7 & 3 \\ -12 & -15 & 7 \end{bmatrix}$

(g) $G = \begin{bmatrix} -3 & 2 & 7 \\ 2 & -3 & -9 \\ -1 & 1 & 3 \end{bmatrix}$

(h) $H = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$

v0.2a

Exercise 7.3.7. For a given $n \times n$ square matrix A, suppose λ_1 is an eigenvalue with n corresponding linearly independent eigenvectors $\boldsymbol{p}_1, \boldsymbol{p}_2, \dots, \boldsymbol{p}_n$. Adapt parts of the proof of Theorem 7.3.12 to prove that the characteristic polynomial of matrix A is $\det(A - \lambda I) = (\lambda_1 - \lambda)^n$. Then deduce that λ_1 is the only eigenvalue of A, and is of multiplicity n.

Exercise 7.3.8. For each of the following matrices, use MATLAB/Octave to find the eigenvalues, their multiplicity, and the dimension of the eigenspaces. Give reasons (remember computational error).

(a) $A = \begin{bmatrix} -30.5 & 25.5 & 22 & -5 & -48.5 \\ -111 & 88 & 76 & -20 & -173 \\ 87.5 & -70.5 & -61 & 16 & 136.5 \\ 51 & -43 & -36 & 9 & 81 \\ -4.5 & 2.5 & 2 & -1 & -6.5 \end{bmatrix}$

(b) $B = \begin{bmatrix} 2929 & 1140 & -1359 & 2352 & 406 \\ -2441 & -950 & 1132 & -1960 & -338 \\ -1070 & -416 & 495 & -858 & -148 \\ -2929 & -1140 & 1359 & -2352 & -406 \\ -895 & -347 & 412 & -715 & -124 \end{bmatrix}$

(c) $C = \begin{bmatrix} -168 & 115 & 305 & -120 & 70 \\ -4710 & 3202 & 8510 & -3360 & 1990 \\ 450 & -305 & -813 & 320 & -190 \\ -4545 & 3090 & 8205 & -3243 & 1920 \\ -2415 & 1640 & 4355 & -1720 & 1017 \end{bmatrix}$

(d) $D = \begin{bmatrix} -8 & 744 & -564 & 270 & 321 \\ 1 & -43 & 33 & -15 & -19 \\ 1 & 29 & -21 & 12 & 12 \\ 5 & -431 & 327 & -156 & -186 \\ -5 & 535 & -405 & 195 & 231 \end{bmatrix}$

(e) $E = \begin{bmatrix} 15.6 & 31.4 & -11.6 & -14.6 & -10.6 \\ -16.6 & -32.4 & 11.6 & 14.6 & 10.6 \\ 33.6 & 56.4 & -19.6 & -27.6 & -15.6 \\ 38.6 & 75.4 & -28.6 & -35.6 & -26.6 \\ -122 & -226 & 82 & 106 & 73 \end{bmatrix}$

(f) $F = \begin{bmatrix} -208 & 420 & -518 & 82 & 264 \\ 655 & -1336 & 1642 & -260 & -838 \\ 229 & -467 & 574 & -91 & -294 \\ -171 & 348 & -428 & 66 & 218 \\ -703 & 1431 & -1760 & 279 & 896 \end{bmatrix}$

v0.2a

Exercise 7.3.9. For each of the following systems of differential equations, find eigenvalues and eigenvectors to derive a general solution of the system of differential equations. Show your working.

(a) $\frac{dx}{dt} = x - 1.5y$, $\frac{dy}{dt} = 4x - 4y$

(b) $\frac{dx}{dt} = x$, $\frac{dy}{dt} = -12x + 5y$

(c) $\frac{dx}{dt} = 7x - 3y$

(d) $\frac{du}{dt} = 2.8u - 3.6v$, $\frac{dv}{dt} = -0.6u + 2.2v$

(e) $\frac{dp}{dt} = 14p + 16q$, $\frac{dq}{dt} = -8p - 10q$

(f) $\frac{dx}{dt} = 6.5x - 0.6y - 5.7z$, $\frac{dy}{dt} = -3x + 4.4y + 7.8z$

(g) $\frac{dx}{dt} = -31x + 26y - 24z$, $\frac{dy}{dt} = -48x + 39y - 36z$, $\frac{dz}{dt} = -14x + 10y - 9z$

(h) $\frac{dx}{dt} = 0.2x + 1.2z$, $\frac{dy}{dt} = -x$, $\frac{dz}{dt} = 1.8x + 0.8z$

(i) $\frac{du}{dt} = 4.5u + 7.5v + 7.5w$, $\frac{dv}{dt} = 3u + 4v + 5w$, $\frac{dw}{dt} = -7.5u - 11.5v - 12.5w$

(j) $\frac{dp}{dt} = -13p + 30q + 6r$, $\frac{dq}{dt} = -32p + 69q + 14r$, $\frac{dr}{dt} = 125p - 265q - 54r$

Exercise 7.3.10. In each of the following, a general solution to a differential equation is given. Find the particular solution that satisfies the specified initial conditions. Show your working.

(a) $(x, y) = c_1(0, 1)e^{-t} + c_2(1, 3)e^{2t}$ where $x(0) = 2$ and $y(0) = 1$

(b) $(x, y) = c_1(0, 1)e^{-2t} + c_2(1, 3)e^{t}$ where $x(0) = 0$ and $y(0) = 2$

(c) $x = 3c_1e^t + c_2e^{-t}$, $y = 5c_1e^t + 2c_2e^{-t}$ where $x(0) = 0$ and $y(0) = 2$

(d) $x = 3c_1e^t + c_2e^{2t}$, $y = 5c_1e^t + 2c_2e^{2t}$ where $x(0) = 1$ and $y(0) = 3$

(e) $(x, y, z) = c_1(0, 0, 1) + c_2(1, -1, 1)e^{2t} + c_3(-2, 3, 1)e^{-2t}$ where $x(0) = 3$, $y(0) = -4$ and $z(0) = 0$

(f) $(x, y, z) = c_1(0, 0, 1)e^{-3t} + c_2(1, -1, 1) + c_3(-2, 3, 1)e^{-t}$ where $x(0) = 3$, $y(0) = -4$ and $z(0) = 2$

Exercise 7.3.11. For each of the following systems of differential equations, use MATLAB/Octave to find eigenvalues and eigenvectors and hence derive a general solution of the system of differential equations. Record your working.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

(a)
$$\begin{aligned}
\tfrac{dx_1}{dt} &= 15.8x_1 + 17.1x_2 - 119.7x_3 + 153.9x_4\,,\\
\tfrac{dx_2}{dt} &= 1.4x_1 + 0.1x_2 + 12.9x_3 - 17.1x_4\,,\\
\tfrac{dx_3}{dt} &= 6.2x_1 + 6.2x_2 - 43x_3 + 57.6x_4\,,\\
\tfrac{dx_4}{dt} &= 3.4x_1 + 3.4x_2 - 25x_3 + 34x_4\,.
\end{aligned}$$

(b)
$$\begin{aligned}
\tfrac{dx_1}{dt} &= -12.2x_1 + 53.7x_2 + 50.1x_3 - 22.8x_4\,,\\
\tfrac{dx_2}{dt} &= 0.6x_1 - 2.3x_2 - 2.7x_3 + 1.2x_4\,,\\
\tfrac{dx_3}{dt} &= -20.4x_1 + 93.8x_2 + 90.2x_3 - 40.8x_4\,,\\
\tfrac{dx_4}{dt} &= -38.2x_1 + 177.9x_2 + 170.7x_3 - 77.2x_4\,.
\end{aligned}$$

(c)
$$\begin{aligned}
\tfrac{dx_1}{dt} &= x_1 + 29.4x_2 - 3.2x_3 - 12.9x_4\,,\\
\tfrac{dx_2}{dt} &= 1.4x_1 - 38.4x_2 + 5.6x_3 + 18.2x_4\,,\\
\tfrac{dx_3}{dt} &= 2.3x_1 - 80.3x_2 + 12.3x_3 + 36.7x_4\,,\\
\tfrac{dx_4}{dt} &= 2.4x_1 - 65x_2 + 9.2x_3 + 31.1x_4\,.
\end{aligned}$$

(d)
$$\begin{aligned}
\tfrac{dx_1}{dt} &= -50.2x_1 - 39.5x_2 - 20.2x_3 + 68.9x_4\,,\\
\tfrac{dx_2}{dt} &= 62.8x_1 + 50.2x_2 + 28.4x_3 - 85.2x_4\,,\\
\tfrac{dx_3}{dt} &= -17.3x_1 - 13.7x_2 - 8.9x_3 + 22.7x_4\,,\\
\tfrac{dx_4}{dt} &= -6.4x_1 - 4.6x_2 - 1.4x_3 + 9x_4\,.
\end{aligned}$$

Exercise 7.3.12. Recall the general complex solution that Example 7.3.22 derives for the oscillations of a mass on a spring. Show that substituting $c_1 = (C_1 - iC_2)/2$ and $c_2 = (C_1 + iC_2)/2$ for real C_1 and C_2 results in the velocity $v(t)$ being expressed algebraically in purely real terms.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Answers to selected exercises

7.0.1b : real

7.0.1d : complex

7.0.1f : complex

7.0.1h : complex

7.0.2b : $\mathbb{E}_2 = \operatorname{span}\{(0\,,1)\}$.

7.0.2d : $\mathbb{E}_{-2} = \operatorname{span}\{(0\,,0\,,1)\}$, $\mathbb{E}_{-4} = \operatorname{span}\{(0\,,2\,,-5)\}$, $\mathbb{E}_0 = \operatorname{span}\{(8\,,-6\,,-11)\}$.

7.0.2f : This matrix is not triangular so we cannot answer (as yet).

7.0.2h : $\mathbb{E}_{-2} = \operatorname{span}\{(1\,,0\,,0\,,0)\}$, $\mathbb{E}_{-3} = \operatorname{span}\{(6\,,-1\,,1\,,0)\}$, $\mathbb{E}_2 = \operatorname{span}\{(6\,,9\,,1\,,5)\}$.

7.0.2j : $\mathbb{E}_0 = \operatorname{span}\{(1\,,0\,,0\,,0)\}$, $\mathbb{E}_7 = \operatorname{span}\{(-2\,,7\,,0\,,0)\}$, $\mathbb{E}_3 = \operatorname{span}\{(-22\,,3\,,12\,,0)\}$.

7.1.1b : yes

7.1.1d : yes

7.1.2a : $\lambda^2 + 5\lambda - 12$

7.1.2c : $-\lambda^3 + 0\lambda^2 + \cdots + 4$

7.1.2e : $-\lambda^3 - \lambda^2 + \cdots - 228$

7.1.2g : $\lambda^4 - \lambda^3 + \cdots + 24$

7.1.3a : 2×2, -5, -6

7.1.3c : 2×2, 1, 0

7.1.3e : 3×3, 0, -199

7.1.3g : 4×4, -3, -56

7.1.4a : 1 once, -3 once

7.1.4c : 3 twice

7.1.4e : $1/2$ twice

7.1.4g : 0 once, -2 once, 1 once

7.1.4i : 2 once, $-1 \pm i$ once

7.1.4k : -1 once, 0 twice

7.1.5a : -1 once, -2.3 twice

7.1.5c : -0.9 once, -0.1 once, 0.1 twice

7.1.5e : 0.6 four times

7.1.5g : $1.7 \pm 1.3i$ once, -3.9 thrice

v0.2a

7.1.6a : $\mathbb{E}_3 = \text{span}\{(2\,,3)\}$

7.1.6c : $\mathbb{E}_{-1} = \text{span}\{(1\,,1)\}$

7.1.6e : $\mathbb{E}_7 = \text{span}\{(7\,,-8\,,0)\,,(-2\,,2\,,1)\}$

7.1.6g : $\mathbb{E}_{-2} = \text{span}\{(-2\,,2\,,1)\}$

7.1.7a : $\lambda = -3$, once, $(0,-0.89,-0.45)$; $\lambda = 2$, twice, $(0.27,0.53,0.80)$.

7.1.7c : $\lambda = 5$, once, $(-0.82\,,0\,,-0.41\,,0.41)$; $\lambda = 3$, once, $(0.55\,,0.28\,,0.55\,,0.55)$; $\lambda = -4$, twice, $(0.67\,,0\,,0.33\,,-0.67)$.

7.1.7e : $\lambda = -3.94$, once, $(-0.33\,,-0.41\,,0.72\,,0.44\,,-0.15)$; $\lambda = 1$, once, $(0\,,0.8\,,-0.53\,,0.27\,,0)$; $\lambda = 7.49 \pm 0.39i$, once, $(-0.24 \pm 0.55i\,,0.23 \pm 0.02i\,,0.043 \pm 0.01i\,,0.72\,,0.24 \mp 0.11i)$; $\lambda = 4.95$, once, $(-0.14\,,0.45\,,0.62\,,0.46\,,-0.43)$.

7.1.8a : Sensitive.

7.1.8c : Not sensitive.

7.1.9b : $\lambda = 3$ sensitive.

7.1.9d : neither sensitive, symmetric.

7.1.9f : -0.9 and -0.1 not sensitive, 0.1 sensitive

7.1.9h : 0.6 mixed sensitivity!

7.1.10b : $(0\,,-12)$, $(12\,,-24)$ and $(24\,,-96)$

7.1.10d : $(3\,,2)$, $(4\,,2)$ and $(2\,,0)$

7.1.10f : $(-7\,,16\,,-6)$, $(79\,,-64\,,126)$ and $(-367\,,256\,,-606)$

7.1.10h : $(-12\,,6\,,-39)$, $(-63\,,-30\,,0)$ and $(-168\,,-96\,,39)$

7.1.17b : $\lambda = 0\,,\pm 13$, $(0\,,\frac{12}{13}\,,\frac{5}{13})$ and $(\pm 1\,,-\frac{5}{13}\,,\frac{12}{13})$

7.1.17d : $\lambda = \pm 1\,,\pm 4$, $(1\,,0\,,0\,,\pm 1)$ and $(0\,,1\,,\mp 1\,,0)$

7.2.1b : lin. dep.

7.2.1d : lin. indep.

7.2.1f : lin. indep.

7.2.1h : lin. dep.

7.2.1j : lin. indep.

7.2.2b : lin. dep.

7.2.2d : lin. dep.

7.2.2f : lin. dep.

7.2.2h : lin. dep.

7.2.6b : Two possibilities are $\{(-3/4\,,1\,,1/4)\}$ and $\{(3\,,-4\,,-1)\}$.

7.2.6d : Two possibilities are $\{(0\,,1\,,1)\}$ and $\{(0\,,-2\,,-2)\}$.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

7.2.6f : Two possibilities are $\{(0,1,0),(2,0,1)\}$ and $\{(2,-4,1),(4,1,2)\}$.

7.2.6h : Does not have a solution subspace.

7.2.7b : $\{(0.67\,,-0.57\,,-0.47)\}$ (2 d.p.)

7.2.7d : $\{(-0.50\,,0.50\,,0.50\,,-0.50)\}$ (2 d.p.)

7.2.7f : $\{(0.61\,,0.61\,,-0.51\,,-0.10)\}$ (2 d.p.)

7.2.7h : $\{(0.62\,,0.45\,,-0.62\,,-0.17)\,,(-0.13\,,0.63\,,0.13\,,0.76)\}$ (2 d.p.)

7.2.9b : $(-1.5\,,0)$, $(0.5\,,-0.5)$, $(-1\,,2.5)$, $(2\,,0.5)$

7.2.9d : $(-1\,,-0.5)$, $(1\,,1)$, $(-1\,,0.5)$, $(0.5\,,0)$

7.2.9f : $(-0.9\,,0.3)$, $(0.6\,,0.2)$, $(-0.2\,,-1.1)$, $(0.7\,,-0.6)$

7.2.10a : $[\boldsymbol{p}]_{\mathcal{E}} = (-2\,,11\,,9)$

7.2.10c : $[\boldsymbol{r}]_{\mathcal{E}} = (-4\,,-1\,,-11)$

7.2.10e : $[\boldsymbol{t}]_{\mathcal{E}} = (-1\,,11/2\,,9/2)$

7.2.10g : $[\boldsymbol{v}]_{\mathcal{E}} = (1/2\,,-3\,,-5/2)$

7.2.10i : $[\boldsymbol{x}]_{\mathcal{E}} = (-0.3\,,4.0\,,4.3)$

7.2.12a : $[\boldsymbol{p}]_{\mathcal{B}} = (-1\,,1)$

7.2.12c : $[\boldsymbol{r}]_{\mathcal{B}} = (-3\,,-1)$

7.2.12e : not in $\mathcal{B}$

7.2.12g : not in $\mathcal{B}$

7.2.12i : $[\boldsymbol{x}]_{\mathcal{B}} = (-0.1\,,-0.2)$

7.2.14a : $[\boldsymbol{p}]_{\mathcal{E}} = (-2\,,-14\,,18\,,2\,,-5)$

7.2.14c : $[\boldsymbol{r}]_{\mathcal{E}} = (-18\,,12\,,-6\,,3\,,-6)$

7.2.14e : $[\boldsymbol{t}]_{\mathcal{B}} = (3\,,2\,,6)$

7.2.14g : $[\boldsymbol{v}]_{\mathcal{B}} = (1\,,2\,,4)$

7.2.14i : not in $\mathbb{B}$

7.3.1b : yes

7.3.1d : no

7.3.1f : no

7.3.1h : no

7.3.5b : yes

7.3.5d : unknown

7.3.5f : yes

7.3.5h : yes

7.3.6a : $\lambda = 2$, three; all good

v0.2a

7.3.6c : $\lambda = -2$, two; not lin. indep.

7.3.6e : $\lambda = -1$, one; errors 10^{-6}, all three eigenvectors $\pm$same

7.3.6g : $\lambda = -1$, one; errors 10^{-6}, all three eigenvectors $\pm$same

7.3.8a : $\lambda = 1$ twice, $\dim \mathbb{E}_1 = 1$; $\lambda = -1$ thrice, $\dim \mathbb{E}_{-1} = 2$

7.3.8c : $\lambda = 2$ twice, $\dim \mathbb{E}_2 = 2$; $\lambda = -3$ thrice, $\dim \mathbb{E}_{-3} = 3$

7.3.8e : $\lambda = 2$ twice, $\dim \mathbb{E}_2 = 1$; $\lambda = -1$ thrice, $\dim \mathbb{E}_{-1} = 3$

7.3.9a : $(x\,,y) = c_1(3\,,4)e^{-t} + c_2(1\,,2)e^{-2t}$

7.3.9c : Not possible: only one equation for two unknowns.

7.3.9e : $(p\,,q) = c_1(2\,,-1)e^{6t} + c_2(-1\,,1)e^{-2t}$

7.3.9g : $(x\,,y\,,z) = c_1(3\,,3\,,-1)e^{3t} + c_2(1\,,2\,,1)e^{-3t} + c_3(1\,,3\,,2)e^{-t}$

7.3.9i : $(u\,,v\,,w) = c_1(-1\,,-1\,,2)e^{-3t} + c_2(-5\,,0\,,3) + c_3(0\,,1\,,-1)e^{-t}$

7.3.10a : $(x\,,y) = -5(0\,,1)e^{-t} + 2(1\,,3)e^{2t}$

7.3.10c : $x = -6e^{t} + 6e^{-t}$, $y = -10e^{t} + 12e^{-t}$

7.3.10e : $(x\,,y\,,z) = (1\,,-1\,,1)e^{2t} + (2\,,-3\,,-1)e^{-2t}$

7.3.11a : (2 d.p.) $\boldsymbol{x} = c_1(0.71,-0.71,0,0)e^{-1.3t} + c_2(-0.63,-0.63,-0.42,-0.21)e^{4.4t} + c_3(0,0.64,0.64,0.43)e^{1.6t} + c_4(-0.82,0.41,0.41)e^{2.2t}$

7.3.11c : (2 d.p.) $\boldsymbol{x} = c_1(-0.63\,,0.32\,,0.32\,,0.63)e^{0.8t} + c_2(0.8\,,0.27\,,0\,,0.53)e^{2.2t} + c_3(0.76\,,0\,,-0.49-0.4i\,,0.091+0.12i)e^{(1.5-0.4i)t} + c_4(0.76\,,0\,,-0.49+0.4i\,,0.091-0.12i)e^{(1.5+0.4i)t}$

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Bibliography

Alpers, B., Demlova, M., Fant, C.-H., Gustafsson, T., Lawson, D., Mustoe, L., Olsson-Lehtonen, B., Robinson, C. & Velichova, D. (2013), A framework for mathematics curricula in engineering education, Technical report, European Society for Engineering Education (SEFI).
http://sefi.htw-aalen.de/curriculum.htm

Anton, H. & Rorres, C. (1991), *Elementary linear algebra. Applications version*, 6th edn, Wiley.

Arnold, V. I. (2014), *Mathematical understanding of nature*, Amer. Math. Soc.

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=1.192190826.137003948.1386128871

Bressoud, D. M., Friedlander, E. M. & Levermore, C. D. (2014), 'Meeting the challenges of improved post-secondary education in the mathematical sciences', *Notices of the AMS* **61**(5), 502–3.

Chartier, T. (2015), *When life is linear: from computer graphics to bracketology*, Math Assoc Amer.
http://www.maa.org/press/books/when-life-is-linear-from-computer-graphics-to-bracketology

Cowen, C. C. (1997), On the centrality of linear algebra in the curriculum, Technical report, Mathematical Association of America.
http://www.maa.org/centrality-of-linear-algebra

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.

Davis, B. & Uhl, J. (1999), *Matrices, Geometry and Mathematica*, Wolfram Research.

Donoho, D. L. & Stodden, V. (2015), Reproducible research in the mathematical sciences, *in* N. J. Higham, M. R. Dennis, P. Glendinning, P. A. Martin, F. Santosa & J. Tanner, eds, 'Princeton Companion to Applied Mathematics', Princeton, chapter VIII.5, pp. 916–925.

Driscoll, T. A. & Maki, K. L. (2007), 'Searching for rare growth factors using multicanonical monte carlo methods', *SIAM Review* **49**(4), 673–692.
http://link.aip.org/link/?SIR/49/673/1

Gorodetski, V. I., Popyack, L. J. & Samoilov, V. (2001), SVD-based approach to transparent embedding data into digital images, *in* V. I. Gorodetski, V. A. Skormin & L. J. Popyack, eds, 'Information assurance in computer netwroks: methods, models, and architectures for network security', Vol. 2052 of *Lecture Notes in Computer Science*, Springer, pp. 263–274.

Halpern, D. F. & Hakel, M. D. (2003), 'Applying the science of learning to the university and beyond: Teaching for long-term retention and transfer', *Change: The Magazine of Higher Learning* **35**(4), 36–41.

Hannah, J. (1996), 'A geometric approach to determinants', *The American Mathematical Monthly* **103**(5), 401–409.
http://www.jstor.org/stable/2974931

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.

Holt, J. (2013), *Linear algebra with applications*, Freeman.

Hughes-Hallett, D., Gleason, A. M. & McCallum, et al., W. G. (2013), *Calculus: single and multivariable*, 6th edn, Wiley.

Kleiber, M. (1947), 'Body size and metabolic rate', *Physiological Reviews* **27**, 511–541.

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.

Larson, R. (2013), *Elementary linear algebra*, 7th edn, Brooks/Cole Cengage learning.

Lay, D. C. (2012), *Linear Algebra and its Applications*, 4th edn, Addison–Wesley.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Lichman, M. (2013), 'UCI machine learning repository', [online]. http://archive.ics.uci.edu/ml

Mandelbrot, B. B. (1982), *The fractal geometry of nature*, W. H. Freeman.

Moody, D. L. (2009), 'The "physics" of notations: Towards a scientific basis for constructing visual notations in software engineering', *IEEE Trans. Soft. Engrg.* **35**(5), 756–778.

Nakos, G. & Joyner, D. (1998), *Linear algebra with applications*, Brooks/Cole.

Pashler, H., Bain, P. M., Bottge, B. A., Graesser, A., Koedinger, K., McDaniel, M. & Metcalfe, J. (2007), Organizing instruction and study to improve student learning: IES practice guide, Technical report, National Center for Education Research, Institute of Education Sciences, U.S. Dept of Education. http://files.eric.ed.gov/fulltext/ED498555.pdf

Poole, D. (2015), *Linear algebra: A modern introduction*, 4th edn, Cengage Learning.

Quarteroni, A. & Saleri, F. (2006), *Scientific Computing with MATLAB and Octave*, Vol. 2 of *Texts in Computational Science and Engineering*, 2nd edition edn, Springer. (Available in pdf via the library of the University of Adelaide.).

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

Schumacher, C. S., Siegel, M. J. & Zorn, P. (2015), 2015 CUPM curriculum guide to majors in the mathematical sciences, Technical report, The Mathematical Association of America. http://www.maa.org/programs/faculty-and-departments/curriculum-department-guidelines-recommendations/cupm

Trefethen, L. N. & Bau, III, D. (1997), *Numerical linear algebra*, SIAM.

Turner, P. R., Crowley, J. M., Humpherys, J., Levy, R., Socha, K. & Wasserstein, R. (2015), Modeling across the curriculum II. report on the second SIAM-NSF workshop, Alexandria, VA, Technical report, [http://www.siam.org/reports/modeling_14.pdf].

Uhlig, F. (2002), A new unified, balanced, and conceptual approach to teaching linear algebra, Technical report, Department of Mathematics, Auburn University, http://www.auburn.edu/~uhligfd/TLA/download/tlateach.pdf.

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Will, T. (2004), Introduction to the singular value decomposition, Technical report, [http://www.uwlax.edu/faculty/will/svd].

v0.2a

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

Index

∈, 261
$|\cdot|$, **17**
`()`, 80
`+,-,*`, 80, 82, 160
`.*`, 190
`./`, 190
`.^`, 106
`/`, 80
`=`, 80, 106
`[...]`, 80, 106
`^`, 80
2-norm, 477
2 d.p., 108

`A\`, 106
`acos()`, 40
`acos()`, 80
addition, **24**, **148**, 161, 166
adjacency matrix, 434
adolescent, 612–614
adult, 612–614, 625
age structure, 612–626
age structured population, 152, 156
angle, **38**, 39–41, 56, 66, 197, 217
`ans`, 80
approximate solution, **302**, 307
arc-cos, 80
area, 66
Arrow, Kenneth, 306
artificial intelligence, 352
associative law, 30, 31, 166, 167
augmented matrix, **115**, 115, 116, 121
average, 476, 496
`axis`, 476

Babbage, Charles, 79, 87, 510
basin area, 352
basis, 602, **651**, 654, 657, 660, 663, 669, 670
best straight line, 307
body-centered cubic, 41
bond angles, 39
bounded, 544
brackets, 106, **145**
bulls eye, 470, 483

$\mathbb{C}$, **11**
canonical form, **455**, 463
Cardano, Gerolamo, 11
Cauchy–Schwarz inequality, 46, **47**, 47, 49, 60, 480
characteristic equation, **424**, 432, 586, **593**, **602**
characteristic polynomial, **593**, 594–598, 602, 629, 630, 697
chemistry, 39
closed, 261
coastlines, 353
coefficient, **95**, **132**, 595
`colormap()`, 317
column space, **266**, 267, 275, 283, 293, 295, 297, 301, 321, 329, 339
column vector, 105, 109, **146**, 159, 202, 266, 290, 665
commutative law, 22, 30, 31, 45, 70, 166
complex conjugate, 441
complex eigenvalue, 440, 587, 591, 602, 620, 621
complex eigenvector, 587
complex numbers, **11**, 428, 441
components, **14**, 19, **146**
composition, **384**, 395
computed tomography, 317, 355, 536
computer algebra, 80
cond, 234, 242, 256, 257, 313, 383
condition number, 106, **234**, 234–244, 255, 383, 525, 534
conic section, **451**, 462, 463, 582
consistent, 100, 121, 134, 307
constant term, **95**, 124, 595
contradiction, 238, 348, 441, 447
coordinate axes, 462, 463
coordinate system, 12
coordinate vector, **660**
coordinates, **660**
cosine, 40
cosine rule, 37, 42, 58
cross product, **65**, 62–78
cross product direction, 66

v0.2a

`csvread()`, 476, 495
CT scan, 317, 526
`cumprod()`, 511

data mining, 352
data reduction, 488
De Moivre's theorem, 621
decimal places, 108
Descartes, 30
`det()`, 556
determinant, 73, **180**, 423, 424, 538, **541**, 541, 547, 550, 556, 567, 573, 630
diag, 189–192, 194–196, 209, 213, 227, 235, 356, 386, 393, 399, 416, 422, 448, 467, 484, 674
`diag()`, 190
diagonal entries, **189**
diagonal matrix, **189**, 188–197, 209, 412, 448, 541, 566, 672–674
diagonalisable, **673**, 674, 677, 678, 686
diagonalisation, 672–699
difference, **24**, **149**, 161
differential equation, 683, 686
differential equations, 683–694
dim, 282, 343, 654, 680, 703
dimension, **282**, 282–290, 296, 415, 488, **654**, 680, 696
direction vector, **28**
discrepancy principle, 533
discriminant, 47, 48
displacement vector, **12**, 12, 19
distance, **26**, 469
distributive law, 31, 45, 70, 166, 167
dolphin, 637
dot product, **38**, 39, 80, 158, **197**, 197, 501, 506
`dot()`, 80
double subscript, **146**
Duhem, Pierre, 299

$\boldsymbol{e}_j$, **25**
`eig()`, 416, 600
eigen-problem, 586, 617
eigenspace, **413**, 415, 419, 424, 433, 586, 592, 602, 631, 680, 696, 697
eigenvalue, **409**, 412, 413, 415, 416, 419, 424, 431, 433, 435, 437, 441, 442, 447, 448, 455, 456, 586, 588, 592, 593, 595, 598, 600, 602, 610, 619, 628–631, 638, 639, 646, 674, 678, 680, 686
eigenvector, **409**, 413, 416, 424, 431, 437, 442, 447, 455, 586, 588, 602, 619, 628, 638, 639, 646, 674, 686
El Nino, 276, 282
elementary row operation, **116**, 121
elements, **146**
elephant, 635
ellipsis, 11
ensemble of simulations, 281
entries, **146**
equal, **14**, **147**
equation of the plane, **53**, 53
Error using, 161, 164
`Error using`, 84
`error:`, 84
error: operator, 162, 164
Euler, 434, 474
Euler's formula, 429, 691, 692
`exp()`, 317
exponential, 317
`eye()`, 160

factorial, 575, 576
factorisation, 221
female, 612, 614
Feynman, Richard, 515
floating point, **80**
`format long`, 608
fractal, 509, 510
free variable, **119**, 119, 121, 233, 315, 657

Galileo Galilei, 259
Gauss–Jordan elimination, 121, 122
general solution, **619**, 620, 634, 639, 685, 686, 698
giant mouse lemur, 636
global positioning system, 96, 102, 112, 113, 130
GPS, 96, 102, 112, 113, 130

Hack's exponent, 352
Hankel matrix, 278
`hankel()`, 277, 298
Hawking, Stephen, 193
`hilb()`, 519
Hilbert matrix, 519, 535, 536

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

homogeneous, **124**, 125, 238, 269, 289, 413, 657, 665
hyper-cube, **541**, 543
hyper-volume, **541**

i, **25**
idempotent, 554
identical columns, 558
identical rows, 558
identity matrix, **146**, 151, 152, 159, 160, 177
identity transformation, 386
image compression, 470–488
`imagesc()`, 317
Impossibility Theorem, 306
`imread()`, 476
in, $\in$, 261
inconsistent, 100, **233**, 301, 302, 516
inconsistent equations, 299–363
inconsistent system, 321–324, 330
induction, 246, 447, 448
`Inf`, 106
infant, 612
inference, 352
inferring, 91
infinitely many solutions, 100, 123, 125
infinity, 106
initial condition, 688, 689, 691, 698
inner product, **38**
integer, **10**
integral equations, 150
interchanging two columns, 558
interchanging two rows, 116, 558
intersection, 338
`inv()`, 208
inverse, **178**, 180, 182, 556
inverse cosine, 80
inverse matrix, 208
inverse transformation, **386**, 395
invertible, **178**, 180, 182, 184, 185, 191, 202, 218, 238, 256, 289, **386**, 389, 423, 424, 435, 437, 538, 550, 619, 665, 673
Iris, 491

j, **25**
jewel in the crown, 4, 221
jpeg, 470
juvenile, 612–614, 625

k, **25**
Kepler's law, 310
kitten, 625
Kleiber's power law, 352
knowledge discovery, 352

Lagrange multiplier, 356
latent semantic indexing, 498–507
leading one, **119**, 119
least square, 302, 329, 349, 375, 532, 534
left triangular, 566
Leibnitz, Gottfried, 103
length, **17**, 18, 20, 26, 46, 47, 66, 80, 197, 217, 469, 476
Leslie matrix, 152, 156, 175, 186, 439
life expectancy, 307
linear combination, **132**, 132–134, 136, 138, 139, 619, 640, 641, 644, 660, 686
linear dependence, 639–671
linear equation, **95**, 100, 104, 114–116, 119, 121, 124, 125, 134, 182, 217, 249, 315, 515, 516
linear independence, 602, 639–671
linear transformation, **365**, 364–392, 550, 551
linearly dependent, **642**, 639–671
linearly independent, 587, **642**, 639–671, 674, 678
linguistic vector, 15
`log()`, 190, 317
`log10()`, 190, 312
logarithm, 317
Lovelace, Ada, 510
lower triangular, 566
lower-left triangular, 566

magnitude, **17**, 20, 80
Markov chain, 152
Matlab, 79–87
MATLAB, 2, 8–10, 79, **80**, 80, 82–87, 103, 105, **106**, 106–108, 117, 121, 127, 131, 145, 150, 159, **160**, 160–164, 172, 174, 176, 182, 188, **190**, 190, 192, 207–209, **228**, 228, 231, 234–236, 239, 241, 252, 254–256, 275, 277, 278, 293, 295, 298, 307, 310, 313–316, **317**, 317, 319, 320, 331, 354, 358, 360, 394, 411, **416**, 416, 419, 423, 432, 434, 445, 447, 449, 460, 461, 463, **476**, 476, 478, 479,

v0.2a

494, 498, 501, 504, 505, 509–511, 535, 556, 586, 593, 600, 602, 605, 607, 608, 631–633, 647, 667, 670, 679, 682, 683, 689, 696–698
matrix, **104**, **145**
matrix multiplication, 167
matrix norm, 476, **477**, 477–480, 482–484, 488, 494
matrix power, **156**, 156, 167, 169
matrix product, **155**
matrix-vector form, **104**, 134
matrix-vector product, **150**
maximum, 247
mean, 496
`mean()`, 476, 496
minor, **563**, 573
Molar, Cleve, 314
Moore–Penrose inverse, **376**
multiplicity, **415**, 416, 419, 433, **598**, 598, 600, 602, 630–632, 678, 680, 697

`NaN`, 106
natural numbers, 10
nD-cube, **541**, 541, 543
nD-parallelepiped, 543
nD-volume, **541**, 541, 543, 544
negative, **24**
nilpotent, 554
no solution, 100, 123, 233
non-diagonalisable matrix, 678
non-homogeneous, **124**
non-orthogonal coordinates, 672
nonconformant arguments, 84
nonlinear equation, 94, **95**, 96
`norm()`, 79, 80, 476, 478, 479
normal equation, 313, 530
normal vector, **53**, 52–54, 60, 63, 64, 69, 72, 75
not a number, 106
null, 269, 339, 340
nullity, **284**, 284, 286, 290, 296, 665
nullspace, **269**, 284, 295–297, 339, 657

Occam's razor, 526
Octave, 2, 8–10, **80**, 79–87, 103, 105, **106**, 106–108, 117, 121, 127, 131, 145, 150, 159, **160**, 160–164, 172, 174, 176, 182, 188, **190**, 190, 192, 207–209, **228**, 228, 231, 234–236, 239, 241, 252, 254–256, 275, 277, 278, 293, 295, 298, 307, 310, 313–316, **317**, 317, 319–321, 331, 354, 358, 360, 394, **416**, 416, 419, 423, 432, 434, 445, 447, 449, 460, 461, 463, **476**, 476, 478, 479, 494, 498, 501, 504, 505, 509–511, 535, 556, 593, 600, 602, 605, 607, 608, 631–633, 647, 667, 670, 679, 682, 683, 689, 696–698
ones matrix, 160
`ones()`, 106, 160
orangutan, 614
orbital period, 310
orthogonal, **50**, 51, 52, 66, 197, 198, 442, 455, 588
orthogonal basis, 433
orthogonal complement, **336**, 336, 338, 339, 359, 360, 362, 363
orthogonal decomposition, **348**, 362, 363
orthogonal matrix, **199**, 199–206, 216, 218, 219, 226, 242, 416, 448, 480, 541, 673
orthogonal projection, **322**, **324**, 321–349, 357, 358
orthogonal set, **198**, 197–199, 214, 215
orthogonal vectors, 50–52
orthogonally diagonalisable, **448**, 449, 461, 673
orthonormal, 226, 455
orthonormal basis, **271**, 270–283, 295, 297, 324, 331, 358, 500, 504, 651, 652, 654, 657, 667
orthonormal set, **198**, 197–199, 202, 214, 215, 218, 271, 644

parallelepiped, 368
parallelepiped volume, 71–74
parallelogram, 368
parallelogram area, 62, 66, 74
parameter, **28**, **56**
parametric equation, **28**, 29, 47, **56**, 54–58, 133, 139
parentheses, **146**
partial fraction, 110
particular solution, 685, 689, 691, 698
PCA, 488
perp, 344, 346, 347, 362
perpendicular, **50**, 50, 52

v0.2a

perpendicular component, **344**, 346, 361, 362
pigeonhole principle, 447
pixel pattern, 485
planets, 310
player rating, 301, 303, 330
Poincaré, Henri, 243
polar form, 621
`poly()`, 593
population, 612–626
population model, 625
population modelling, 585, 672
position vector, **12**, 12, 14, 19, 28, 56
precision, 80, 174
principal axes, **456**, 463
principal component analysis, 488, 499, 511
principal components, **493**, 488–500
principal vector, **493**, 488–498, 500
proj, 322–326, 329–334, 344, 347, 349, 356, 357, 369, 403
pseudo-inverse, **376**, 374–383, 393, 394
Punch, John, 526
Pythagoras, 10

QR factorisation, 228, 602
qr-code, 81, 107
quadratic equation, 451
quadratic form, **455**, 455, 456
`quit`, 80
quote, 160

$\mathbb{R}$, **11**
`randn()`, 160, 609
random matrix, 160
random perturbation, 633
range, 266
rank, **235**, 235, 236, 238, 242, 255, 257, 272, 283, 286, 290, 298, 302, 381, 394, 484, 654, 657, 665
rank theorem, **286**, 296
rational function, 131
rational numbers, 10
rcond, 234
`rcond()`, 105, 106, 179
real numbers, **11**
Recorde, Robert, 14
reduced row echelon form, 4, **119**, 119, 121, 142
reflection, 197, **199**
regularisation parameter, 530, 533, 534
relative error, 534
repeated eigenvalue, **598**, 608–612, 632, 633, 678
`reshape()`, 317
Richardson, L. F., 353
right triangular, 566
right-angles, 14, 25, 39, 43, 50–52
river length, 352
$\mathbb{R}^n$, 15
rotation, 197, **199**
rotation and/or reflection, **199**, 202
row space, **266**, 275, 283, 293, 295, 297, 339, 340
row vector, **146**, 202, 266, 290, 493, 665

scalar, **11**, 24, 28, 31, 38, 45, 56
scalar multiplication, **24**, 166
scalar product, **150**
scalar triple product, **72**, 73, 77
`scatter()`, 476, 492
`scatter3()`, 476
searching, 502
`semilogy()`, 476
sensitivity, 608–612, 633
serval, 624, 625
Seven Bridges of Königsberg, 434
shear transformation, 388
Sierpinski carpet, 509
Sierpinski network, 419
Sierpinski triangle, 509
significant digits, 80
Singular Spectrum Analysis, 276, 297
singular value, **226**, 233–235, 238, 272, 289, 447, 472, 476, 477, 484, 514, 516, 532, 533, 548, 628, 654, 657, 665
singular value decomposition, 4, 144, 221, 222, 225, **226**, 226–228, 231, 233, 235–237, 244–253, 255, 257, 258, 272, 275, 276, 278, 283, 284, 286, 295, 297–302, 307, 310, 314, 316, 317, 320–322, 327, 330, 349, 350, 354–356, 358, 377, 378, 393, 407, 408, 447–449, 460, 464, 469, 470, 472, 479, 482, 484, 488, 493, 494, 498, 500–502, 504, 507–512, 516, 517, 526, 534, 536, 538, 547, 549, 593, 602, 627, 628, 638, 654, 657, 666
singular vector, **226**, 272, 628
size, **14**, 14, **145**, 159

© AJ Roberts, ORCID:0000-0001-8930-1552, September 1, 2016

`size()`, 80, 160
smallest change, 300–302, 305, **307**
smallest solution, 314, **315**, 532
Southern Oscillation Index, 276
span, **136**, 136–138, 140, 142, 264–267, 269, 272, 273, 282, 285, 292, 293, 322, 325, 329, 330, 334, 336, 337, 339, 340, 344, 345, 348, 357, 358, 360–362, 404, 405, 415, 424, 442, 465, 602, 651, 654, 677, 700, 701
species, 491
spectral norm, 477
square matrix, **104**, 105, 106, 145, 160, 178, 238, 289, 409, 413, 416, 424, 435, 437, 448, 449, 541, 545, 546, 548, 550, 566, 593, 602, 619, 665, 672–674, 678, 686
square-root, 461
standard basis, 659, 661–663, 669–671
standard coordinate system, **13**, 14
standard deviation, 476, 496, 498
standard matrix, **371**, 376, 384, 389, 391, 393
standard unit vector, **25**, 25, 50, 51, 65, 71, 74, 76, 137, 197, 198, 282, 371, 659, 661, 669, 670
`std()`, 476, 496
steganography, 510
stereo pair, **23**, 194, 200, 205, 214, 219, 289
subspace, **261**, 259–271, 279, 282, 290, 292, 293, 324, 331, 336, 338, 343, 344, 346, 348, 413, 651, 652, 654, 657, 660
subtraction, **149**, 161
sum, **24**, **148**, 161
SVD, 4, 144, 221, 222, 225–228, 231, 233, 235–237, 244–253, 255, 257, 258, 272, 275, 276, 278, 283, 284, 286, 295, 297–302, 307, 310, 314, 316, 317, 320–322, 327, 330, 349, 350, 354–356, 358, 377, 378, 393, 407, 408, 447–449, 460, 464, 469, 470, 472, 479, 482, 484, 488, 493, 494, 498, 500–502, 504, 507–512, 516, 517, 526, 534, 536, 538, 547, 549, 593, 602, 627, 628, 638, 654, 657, 666
`svd()`, 228
`svds()`, 476, 496, 498, 500, 504
symmetric matrix, **158**, 159, 169, 175, 415, 441, 442, 447, 610, 611, 673, 680
system, **95**, 104, 115, 116, 121

table tennis, 301
Tasmanian Devil, 634
Tasmanian Devils, 175
Tikhonov regularisation, **530**, 530–533, 536
tipping point, 281
tomography, 317
trace, **595**, 630
transformation, 364, 365
transpose, **157**, 157–160, 164, 169, 175, 176
triangle inequality, 46, **47**, 49, 60, 479
triangular matrix, 565, **566**, 567, 588, 592, 675
triple product, scalar, 72

unique solution, 100, 105, 123, 182, 234, 238, 289, 665
unit cube, 368
unit square, 367
unit vector, **17**, 52, 197, 198, 408, 455, 456
upper triangular, 566
upper-right triangular, 566

van de Snepscheut, Jan L. A., 105
variance, 498
vector, **14**, 17
vector product, **65**
vectors, **12**
velocity vector, 13
volume, parallelepiped, 72, 77

walking gait, 297
weather forecasts, 281
Wikipedia, 11, 130, 306, 310, 318, 530, 614, 624
Wiles, Andrew, 244
wine recognition, 495
word vector, 21, 41, 59, 499, 503
work, 43

zero, 124, 315, 435, 447, 516
zero column, 558
zero matrix, **146**, 159, 160
zero row, 558
zero vector, **15**, 16, 18, 338, 413
`zeros()`, 160

v0.2a