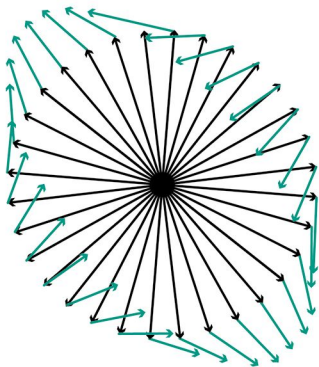# Linear Algebra
for the 21st Century

A. J. ROBERTS

LINEAR ALGEBRA FOR THE 21ST CENTURY

Linear Algebra for the 21st Century

A. J. ROBERTS

University of Adelaide
South Australia, 5005

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

Links to third party websites are provided by Oxford in good faith and
for information only. Oxford disclaims any responsibility for the materials
contained in any third party website referenced in this work.

To Barbara for her patience, understanding, and support during this long project, and to our children and grand-children, with the aim that this approach to mathematics will serve them well.

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 the beauty and power of this jewel? Answer: the 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, especially in data science, and also develops traditional linear algebra concepts, theory, and methods. No rigour is lost in this new route: indeed, this book demonstrates that most theories are better proved with an SVD rather than with the traditional RREF. This new route through linear algebra becomes a preferred approach both because of the ready availability of ubiquitous computing and because of the importance of data science 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)

---

[1] http://au.mathworks.com/company/newsletters/articles/professor-svd.html [Oct 2019]

Acknowledgements

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 is reshaped herein to meet modern demands.

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

For those who study more mathematics, the development here also provides a great foundation of key concepts, relationships, and transformations necessary for higher mathematics.

Important for all is to develop facility in manipulating, interpreting, and transforming between visualizations, 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, this book explicitly integrates computer support for developing the 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 equations, for approximation, and for data science; and the function `eig()` for probing structures. This provides a framework to understand key computational tools to effectively utilize the so-called third arm of science: namely, computation.

Throughout the book, examples (many graphical) introduce and illustrate the concepts and relationships between them. Working through these helps form the mathematical relationships essential for application. Interspersed throughout the text are questions labelled "Activity": these help form and test your developing understanding of the concepts.

Also included are many varied applications, described to varying levels of detail. 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.
>
> *Arnold (2014)*

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 a teacher you can use this book in several ways:

- as a reasonably rigorous mathematical development of concepts and interconnections by invoking its definitions, theorems, and proofs, all interleaved with examples.
- 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 organize 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.

The book typically introduces concepts in two or three dimensional cases, and subsequently develops the general theory of the concept. This is to help focus the learning, empowered by visualization, 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 references are sometimes forward to material not yet 'covered').

> information that is recalled grows stronger with each retrieval … spaced practice is preferable to massed practice.   *Halpern and Hakel (2003)*

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, especially in data science, are opened up by the relatively early introduction of orthogonality.

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

On visualization

> 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)*

Throughout, this book also integrates visualization. This visualization reflects the fundamentally geometric nature of linear algebra. It also empowers learners to utilize different parts of their brain and integrate the knowledge together from the different perspectives. Visualisation also facilitates greater skills at interpretation and modelling—skills essential in applications. But as commented by Fara (2009) [p.249] "just like reading, deciphering graphs and maps only becomes automatic with practice." So, lastly, visual exercise questions develop understanding without a learner 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)*

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 that 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 burden 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 criterion.[1] Further, we are as ruthless as possible in invoking herein the smallest feasible set of commands and functions from MATLAB/Octave so that students have the minimum to learn. Most teachers will find that 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 even 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 and Stodden (2015)*

---

[1] 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 and Joyner, 1998, e.g.). (Of course, be wary of this metric: e.g., APL would surely be too concise!)

Linear algebra for statisticians

This book forms an ideal companion to modern statistics courses, especially data science. The recently published Curriculum Guidelines for Undergraduate Programs in Statistical Science, by Horton et al. (2014), emphasizes that linear algebra courses must provide "matrix manipulations, linear transformations, projections in Euclidean space, eigenvalues/eigenvectors, and matrix decompositions". These are all core topics in this book, especially the statistically important SVD factorization (Chapters 3 and 5). Furthermore, this book explicitly makes the recommended "connections between concepts in these mathematical foundations courses and their applications in statistics" (Horton et al., 2014, p.12).

Moreover, with the aid of some indicative statistical applications along with the sustained invocation of "visualization" and "basic programming concepts", this book helps to underpin the requirement to "Encourage synthesis of theory, methods, computation, and applications" (Horton et al., 2014, p.13).

Contents

1 Vectors

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

Mathematics started with counting. The natural numbers $1, 2, 3, \ldots$ quantify how many objects have been counted. Historically, there were many existential arguments over many centuries about whether negative numbers and zero are meaningful. Nonetheless, eventually negative numbers and the zero were included to form the **integer**s $\ldots, -2, -1, 0, 1, 2, \ldots$. In the meantime people needed to quantify fractions such as two and a half bags, or a third of a cup, which led to the rational numbers such as $\frac{1}{3}$ or $2\frac{1}{2} = \frac{5}{2}$. Now **rational number**s are defined as all numbers writeable in the form $\frac{p}{q}$ for integers p and q (q nonzero). Over two thousand years

ago, Pythagoras[1] was forced to recognize that for many triangles the length of a side could not be rational, and hence there must be more numbers in the world about us than rationals could provide. To cope with non-rational numbers such as $\sqrt{2} = 1.41421\cdots$ and (pi) $\pi = 3.14159\cdots$, mathematicians define the **real numbers** to be all numbers which in principle can be written as a decimal expansion such as $\sqrt{2}, \pi$,

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

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[2] developed a procedure to solve cubic polynomial equations. But the procedure involved manipulating $\sqrt{-1}$ which seemed a crazy figment of imagination. Nonetheless the procedure worked. Subsequently, many practical uses were found for $\sqrt{-1}$, now denoted by i (or j in some disciplines). Consequently, many areas of modern science and engineering use **complex number**s which are those of the form $a + b\mathrm{i}$ for real numbers a and b. The symbol $\mathbb{C}$ denotes the set of all possible complex numbers. This book mostly uses integers and real numbers, but eventually (Chapter 7) we need the marvellous complex numbers.

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

## 1.1 Vectors have magnitude and direction

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

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

---

[1] Pythagoras was born on Samos, Greece, in 569 BC, studied in Egypt, imprisoned in Babylon, then settled in Crotona to found a philosophical and religious school. Pythagoras and his school established Pythagoras' Theorem, the sum of the angles of a triangle are two right-angles, properties of polygons and polyhedra, and solved equations geometrically.

[2] 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 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, Westfall (2012)*)

Example 1.1.1 *(displacement vector)* An important class of vectors are the so-called **displacement vector**s. Given two points in space, say A and B, the displacement vector $\overrightarrow{AB}$ is the directed line segment from the point A to the point B—as illustrated by the two displacement vectors $\overrightarrow{AB}$ and $\overrightarrow{CD}$ to the right. For example, if your home is at position A and your school at position B, then travelling from home to school is to move by the amount of the displacement vector $\overrightarrow{AB}$.
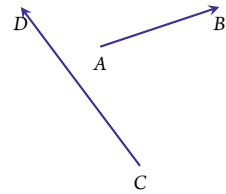
To be able to manipulate vectors we describe them with numbers. For such numbers to have meaning they must be set in the context of a coordinate system. So choose an origin for the coordinate system, usually denoted O, and draw coordinate axes in the plane (or space), as illustrated for the above two displacement vectors. Here the displacement vector $\overrightarrow{AB}$ goes three units to the right and one unit up, so we denote it by the ordered pair of numbers $\overrightarrow{AB} = (3, 1)$. Whereas the displacement vector $\overrightarrow{CD}$ goes three units to the left and four units up, so we denote it by the ordered pair of numbers $\overrightarrow{CD} = (-3, 4)$. Our choice of the origin O does not affect the number representation of these vectors. □
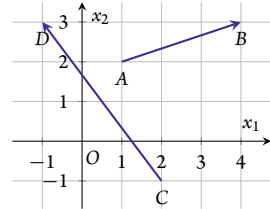
Example 1.1.2 *(position vector)* The next important class of vectors are the **position vector**s. Given some chosen fixed origin in space, usually denoted O, then $\overrightarrow{OA}$ is the position vector of the point A. This picture illustrates the position vectors of four points in the plane (A, B, C, D), from the given origin O.

Again, to be able to manipulate such vectors we describe them with numbers, and such numbers have meaning via a coordinate system. So draw coordinate axes in the plane (or space), as illustrated for the above four position vectors. Here the position vector $\overrightarrow{OA}$ goes one unit to the right and two units up so we denote it by $\overrightarrow{OA} = (1, 2)$. Similarly, the position vectors $\overrightarrow{OB} = (4, 3)$, $\overrightarrow{OC} = (2, -1)$, and $\overrightarrow{OD} = (-1, 3)$. Recognize that the ordered pairs of numbers in the position vectors are exactly the coordinates of each of the specified end-points. □
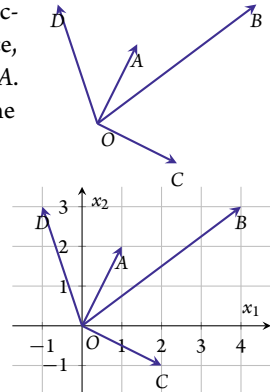
Example 1.1.3 *(velocity vector)* Consider an airplane in level flight at 900 km/hr (kilometres per hour) to the east-north-east. Choosing coordinate axes oriented to the east and the north, the direction of the airplane is at an angle $22.5°$ from the east, as illustrated on the right. Trigonometry then tells us that the eastward part of the speed of the airplane is $900\cos(22.5°) = 831.5$ km/hr, whereas the northward part of the speed is $900\sin(22.5°) = 344.4$ km/hr (as indicated quantitatively). Further, the airplane is in level flight, not going up or down, so in the third direction of space (vertically) its speed component is zero.
Putting these together forms the velocity vector $(831.5, 344.4, 0)$ in km/hr in space.
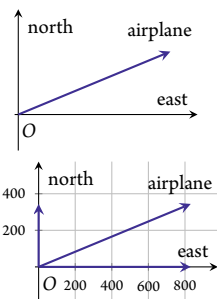
Another airplane takes off from an airport at 360 km/hr to the northwest and climbs at 2 m/s. The direction northwest is 45° to the east-west lines and 45° to the north-south lines. Trigonometry then tells us that the westward speed of the airplane is $360\cos(45°) = 360\cos(\frac{\pi}{4}) = 254.6$ km/hr, whereas the northward speed is $360\sin(45°) = 360\sin(\frac{\pi}{4}) = 254.6$ km/hr as illustrated.
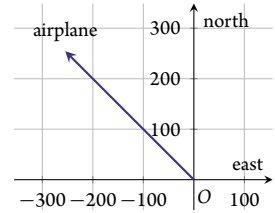
But west is the opposite direction to east, so if the coordinate system treats east as positive, then west must be negative. Consequently, together with the climb in the vertical, the velocity vector is $(-254.6\,\text{km/hr}, 254.6\,\text{km/hr}, 2\,\text{m/s})$. But we should avoid mixing units within a vector, so here convert all speeds to m/s: here 360 km/hr upon dividing by 3600 secs/hr and multiplying by 1000 m/km gives 360 km/hr $= 100$ m/s. Then the north and west speeds are both $100\cos(\frac{\pi}{4}) = 70.71$ m/s. Consequently, the velocity vector of the climbing airplane should be described as $(-70.71, 70.71, 2)$ in m/s. $\qquad\square$

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

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

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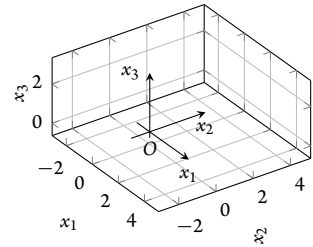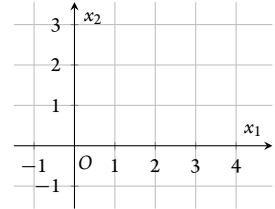
- Correspondingly, in so-called 'n dimensions' the standard coordinate system has n coordinate axes, all at right-angles to each other, and often labelled x_1, x_2, $\ldots$, x_n, respectively.

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

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

(they mean the same; it is just more convenient to usually use a row in parentheses in text, and a column in brackets in displayed mathematics). The real numbers x_1, x_2, ..., 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, ..., $x_n)$ then the position vector $\overrightarrow{OX}$ has the same magnitude and direction as the vector denoted $(x_1$, x_2, ..., $x_n)$.*

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

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

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

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

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

The *ordering* of the components is very important. For example, as illustrated to the right, the vector $(3,1)$ is very different from the vector $(1,3)$; similarly, the vector $(2,-1)$ is very different from the vector $(-1,2)$.
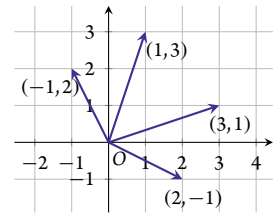
Definition 1.1.5 *The set of all vectors with n components is denoted $\mathbb{R}^n$. The vector with all components zero, $(0,0,\dots,0)$, is called the* **zero vector** *and denoted by* **0**.

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 **0**, this book generally denotes vectors by a bold letter (except for displacement vectors). The other common notation you may see elsewhere is to denote vectors by a small over-arrow such as in the "zero vector $\vec{0}$". Less commonly, some books and

articles use an over- or under-tilde ($\sim$) to denote vectors. Be aware of this different notation in reading other books.

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

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

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

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

(a) "The dog sat on the mat" is summarized by the vector $\boldsymbol{w} = (0,1,1,1,0)$.
(b) 'The cat scratched the dog" is summarized by the vector $\boldsymbol{w} = (1,1,0,0,1)$.
(c) "The cat and dog sat on the mat" is summarized by the vector $\boldsymbol{w} = (1,1,1,1,0)$.
(d) "The dog scratched" is summarized by the vector $\boldsymbol{w} = (0,1,0,0,1)$.
(e) An empty sentence is the zero vector $\boldsymbol{w} = (0,0,0,0,0)$.
(f) Together, the two sentences "The dog sat on the mat. The cat scratched the dog." are summarized by the vector $\boldsymbol{w} = (1,2,1,1,1)$.

Using such crude summary representations of some text, even of entire documents, empowers us to use powerful mathematical techniques to relate documents together, compare and contrast, express similarities, look for type clusters, and so on. In application we would not just count words for objects (nouns) and actions (verbs), but also qualifications (adjectives and adverbs).[3]

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. □

---

[3] Look up Latent Semantic Analysis, such as at https://en.wikipedia.org/wiki/Latent_semantic_analysis [July 2019]

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 2019]

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

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

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

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

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

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

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

Solution:

- $|a| = \sqrt{(-3)^2 + 4^2} = \sqrt{25} = 5$.
- $|b| = \sqrt{3^2 + 3^2} = \sqrt{18} = 3\sqrt{2}$.
- $|c| = \sqrt{1^2 + (-2)^2 + 3^2} = \sqrt{14}$.
- $|d| = \sqrt{1^2 + (-1)^2 + (-1)^2 + 1^2} = \sqrt{4} = 2$. $\square$

Example 1.1.11 Write down three different vectors, all three with the same number of components, for each of the following cases: (a) of length 5, (b) of length 3, and (c) of length -2.

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 perhaps $(1,2,2)$, $(2,-1,2)$, 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 case. □

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

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

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

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

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

Second establish that if a vector has length zero then it must be the zero vector. Let vector $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)$ in $\mathbb{R}^n$ have zero length. By squaring both sides of the Definition 1.1.9 for length we then know that
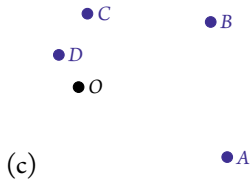
$$\underbrace{v_1^2}_{\geq 0} + \underbrace{v_2^2}_{\geq 0} + \cdots + \underbrace{v_n^2}_{\geq 0} = 0.$$

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 $\boldsymbol{0}$.
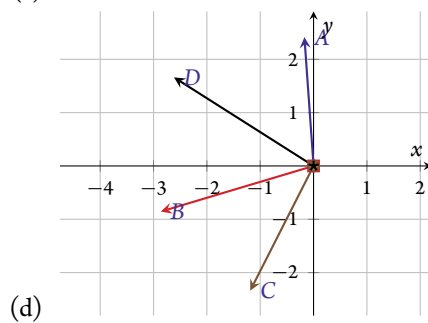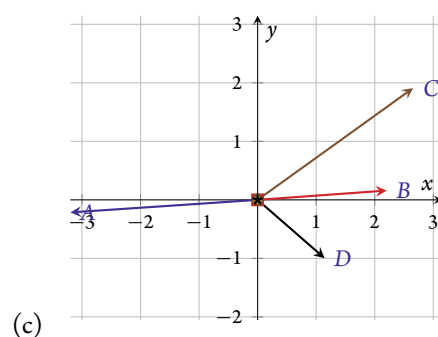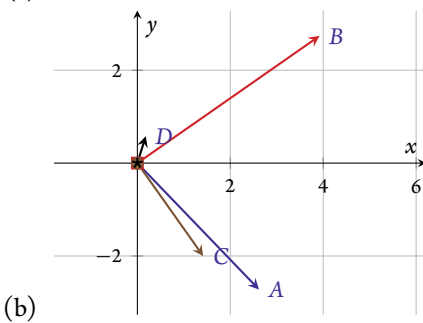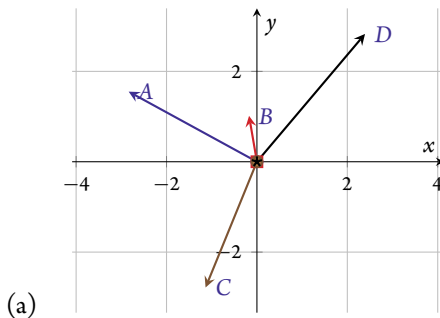
1.1.1 Exercises

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.

(c)

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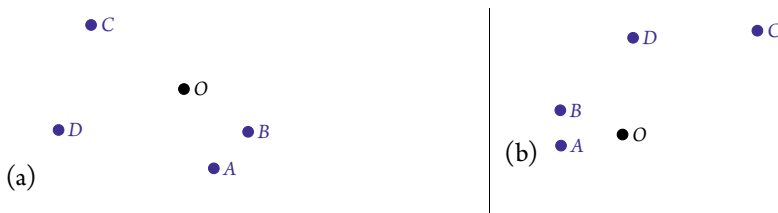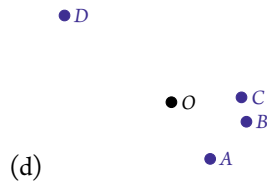
Exercise 1.1.2 For each case: roughly estimate (to say ± 0.2) each of the two components of the four position vectors of the points A, B, C, and D.

(a)

(c)

(b)

(d)

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 that your calculated lengths reasonably approximate your measured lengths.

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 $\mathbb{R}^5$.

Exercise 1.1.5 In a few sentences, answer/discuss each of the following.

(a) Why is a coordinate system important for a vector?

(b) Describe the distinction between a displacement vector and a position vector.

(c) Why do two vectors have to be the same size in order to be equal?

(d) What is the connection between the length of a vector and Pythagoras' theorem for triangles?

(e) Describe a problem that would occur if the ordering of the components in a vector was not significant?

(f) Recall that a vector has both a magnitude and a direction. Comment on why the zero vector is the only vector with zero magnitude.

(g) In what other courses have you seen vectors? What was the same and what was different?

1.2 Adding and stretching vectors

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

1.2.1 Basic operations

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

(a) Let's add the two vectors shown below-left: $(1,3) + (2,-1) = (1+2, 3+(-1)) = (3,2)$ as illustrated below-middle, where the vector $(2,-1)$ is drawn from the end of $(1,3)$, and the end-point of the result determines the vector addition $(3,2)$.
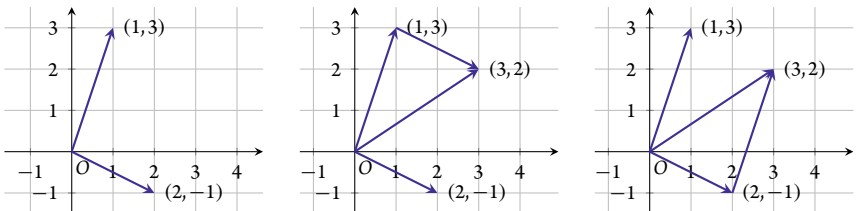
This result $(3,2)$ is the same if the vector $(1,3)$ is drawn from the end of $(2,-1)$ as shown above-right. That is, $(2,-1)+(1,3)=(1,3)+(2,-1)$. That the order of addition is immaterial is the commutative law of vector addition. Theorem 1.2.19(a) establishes this law in general.

(b) $(3,2,0)+(-1,3,2)=(3+(-1),2+3,0+$
   $2)=(2,5,2)$ as illustrated below where (given the two
   vectors as plotted to the right) the vector $(-1,3,2)$ is
   drawn from the end of $(3,2,0)$, and the end-point of the
   result determines the vector addition $(2,5,2)$. As below,
   find the same result by drawing the vector $(3,2,0)$
   from the end of $(-1,3,2)$.
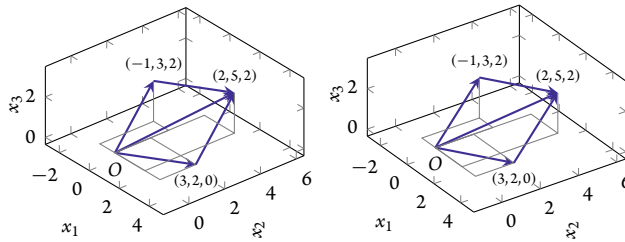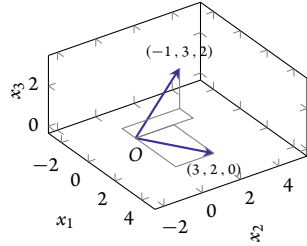
As drawn above, many of the three-D plots in this book are **stereo pair**s, drawing the plot from two slightly different viewpoints:[4] cross your eyes to merge two of the images, and then focus on the pair of plots to see the 3D effect. With practice viewing such 3D stereo pairs becomes less difficult![5]

(c) The addition $(1,3)+(3,2,0)$ is not defined and cannot be done because the two vectors have a different number of components; they have different sizes.
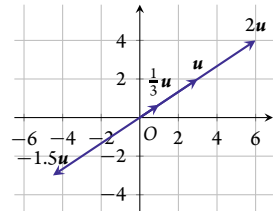
Example 1.2.2 To multiply a vector by a scalar, a number, multiply each component by the scalar. Equivalently, visualize the result through stretching the vector by a factor of the scalar.

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

$$2\boldsymbol{u}=2(3,2)=(2\cdot3,2\cdot2)=(6,4),$$
$$\tfrac{1}{3}\boldsymbol{u}=\tfrac{1}{3}(3,2)=(\tfrac{1}{3}\cdot3,\tfrac{1}{3}\cdot2)=(1,\tfrac{2}{3}),$$
$$(-1.5)\boldsymbol{u}=(-1.5\cdot3,-1.5\cdot2)=(-4.5,-3).$$

---

[4] I implement such cross-eyed stereo so that these stereo images are useful both printed and when projected on a screen.

[5] To help keeping your eyes crossed while you try to focus on the plots, hold a finger or pen in front of your nose so that your right-eye sight across the top of the finger/pen aligns with the centre of the left-picture, and simultaneously your left-eye sight across the top of the finger/pen aligns with the centre of the right-picture.

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

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

$$(-\tfrac{1}{2})v = -\frac{1}{2} \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \cdot 2 \\ -\frac{1}{2} \cdot 3 \\ -\frac{1}{2} \cdot 1 \end{bmatrix} = \begin{bmatrix} -1 \\ -\frac{3}{2} \\ -\frac{1}{2} \end{bmatrix}.$$
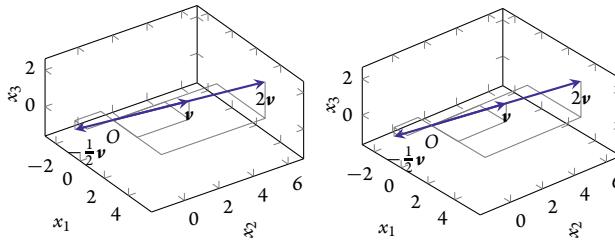
Activity 1.2.3 Combining multiplication and addition, what is $u + 2v$ for vectors $u = (4, 1)$ and $v = (-1, -3)$?

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

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

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

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

$$cu := (cu_1, cu_2, \ldots, cu_n).$$

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

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

Solution: Drawn below.

(a) **u + v**

(d) **v − u**

(b) **v + u**

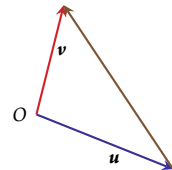(e) $\frac{1}{2}$**u**

(c) **u − v**

(f) **−v**

□
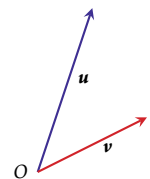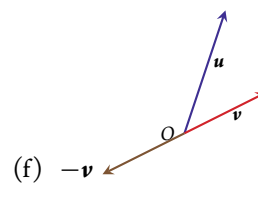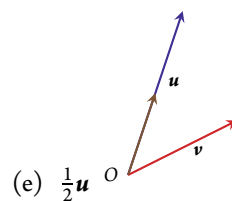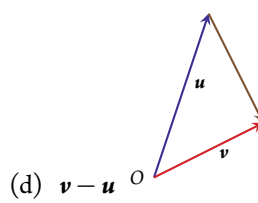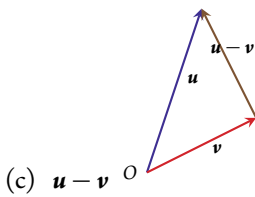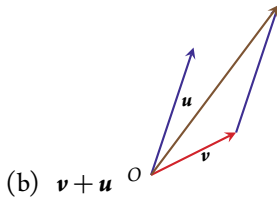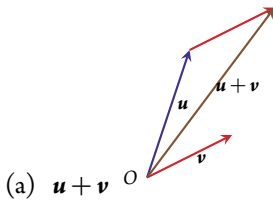
Activity 1.2.6 For the vectors **u** and **v** shown to the right, what is the result vector, which is also shown?

(a) **v − u**   (b) **u − v**   (c) **v + u**   (d) **u + v**

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

$$(3,2) = (3,0) + (0,2) \quad \text{(by addition)}$$
$$= 3(1,0) + 2(0,1) \quad \text{(by scalar mult)}$$
$$= 3\boldsymbol{i} + 2\boldsymbol{j} \quad \text{(by definition of } \boldsymbol{i} \text{ and } \boldsymbol{j}).$$
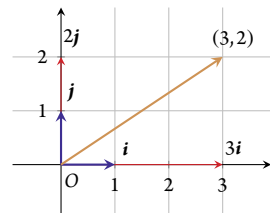
Similarly, in three-dimensional space, we often write vectors in terms of the three vectors **i**, **j**, and **k**, defined to be each of length one, aligned along the three coordinate axes. Hence **i** = (1,0,0), **j** = (0,1,0), and **k**=(0,0,1). For example,

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

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

Definition 1.2.7 *Given a standard coordinate system with n coordinate axes, all at right-angles to each other, the **standard unit vectors** $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, ... , $\boldsymbol{e}_n$ are the vectors of length one in the direction of the corresponding coordinate axis (as illustrated to the right for $\mathbb{R}^2$ and below for $\mathbb{R}^3$). That is,*
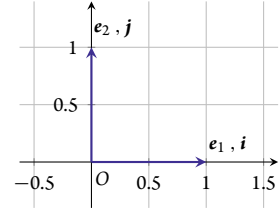
$$\boldsymbol{e}_1 = \begin{bmatrix} 1\\0\\\vdots\\0 \end{bmatrix}, \quad \boldsymbol{e}_2 = \begin{bmatrix} 0\\1\\\vdots\\0 \end{bmatrix}, \quad \dots, \quad \boldsymbol{e}_n = \begin{bmatrix} 0\\0\\\vdots\\1 \end{bmatrix}.$$

In $\mathbb{R}^2$ and $\mathbb{R}^3$, the symbols $\boldsymbol{i}$, $\boldsymbol{j}$, and $\boldsymbol{k}$ are often used as synonyms for $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, and $\boldsymbol{e}_3$, respectively (as illustrated below).
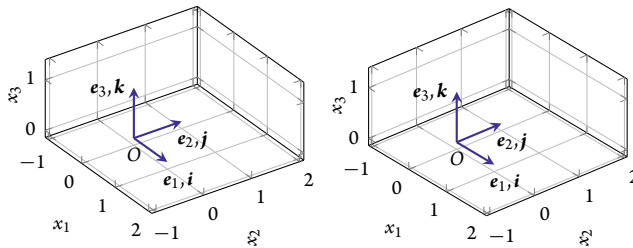
That is, for three examples, the following are equivalent ways of writing the same vector:

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

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

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

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

(a) $(0,3,0,0,1)$ (b) $(5,0,2)$ (c) $(3,1)$ (d) $(0,3,0,1)$

1.2.2 Distance

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

Example 1.2.9 We would like to say that $(1.2, 3.4) \approx (1.5, 3)$ to an error 0.5 (as illustrated to the right). Why is the error 0.5? Because the difference between the vectors $(1.5, 3) - (1.2, 3.4) = (0.3, -0.4)$ has length $\sqrt{0.3^2 + (-0.4)^2} = 0.5$.
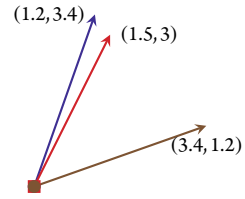
Conversely, we would like to recognize that vectors $(1.2, 3.4)$ and $(3.4, 1.2)$ are very different (as also illustrated)—there is a large 'distance' between them. Why is there a large 'distance'? Because the difference between the vectors $(1.2, 3.4) - (3.4, 1.2) = (-2.2, 2.2)$ has length $\sqrt{(-2.2)^2 + 2.2^2} = 2.2\sqrt{2} = 3.1113$, which is relatively large.

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

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

Example 1.2.11 Given three vectors $a = 3i + 2j - 2k$, $b = 5i + 5j + 4k$, and $c = 7i - 2j + 5k$ (shown right in stereo) use the concept of distance between vectors to answer the following: which pair are the closest to each other? And which pair are furthest from each other?
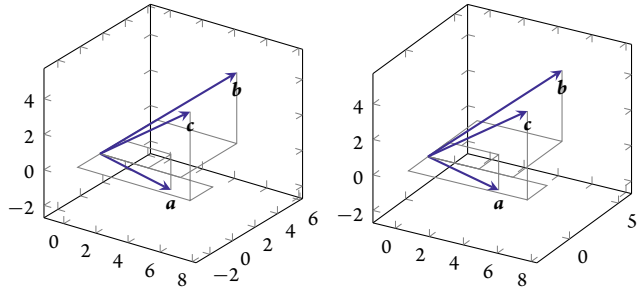
Solution: Compute the distances between each pair.

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

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

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

(a) b, c

(b) a, b

(c) two of the pairs

(d) a, c
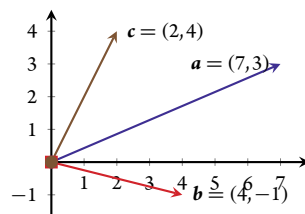
1.2.3 Parametric equation of a line

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

Example 1.2.13 Consider the line drawn to the right in some chosen coordinate system. Recall that one way to find an equation of the line is to find the intercepts with the axes, here at $x = 4$ and at $y = 2$, then write down $\frac{x}{4} + \frac{y}{2} = 1$ as an equation of the line. Algebraic rearrangement gives various other forms, such as $x + 2y = 4$ or $y = 2 - x/2$.

The alternative is to describe the line with vectors. Choose any point P on the line, such as $(2,1)$ as drawn to the right. Then view every other point on the line as having position vector that is the vector sum of $\overrightarrow{OP}$ and a vector aligned along the line. Denote $\overrightarrow{OP}$ by p as drawn. Then, for example, the point $(0,2)$ on the line has position vector $p + d$ for vector $d = (-2,1)$ because $p + d = (2,1) + (-2,1) = (0,2)$. Other points on the line are also given using the same vectors, p and d: for example, the point $(3,\frac{1}{2})$ has position vector $p - \frac{1}{2}d$ (as drawn) because $p - \frac{1}{2}d = (2,1) - \frac{1}{2}(-2,1) = (3,\frac{1}{2})$; and the point $(-2,3)$ has position vector $p + 2d = (2,1) + 2(-2,1)$. In general, every point on the line may be expressed as $p + td$ for some scalar t.

For every given line, there are many possible choices of p and 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 to the right. Let position vector $p = \overrightarrow{OP} = (0,2)$ and the vector $d = \overrightarrow{PQ} = (3,-\frac{3}{2})$, then every point on the line has position vector $p + td$ for some scalar t:
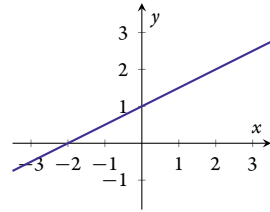
- $(2,1) = (0,2) + (2,-1) = (0,2) + \frac{2}{3}(3,-\frac{3}{2}) = p + \frac{2}{3}d$;
- $(6,-1) = (0,2) + (6,-3) = (0,2) + 2(3,-\frac{3}{2}) = p + 2d$;
- $(-1,\frac{5}{2}) = (0,2) + (-1,\frac{1}{2}) = (0,2) - \frac{1}{3}(3,-\frac{3}{2}) = p - \frac{1}{3}d$.
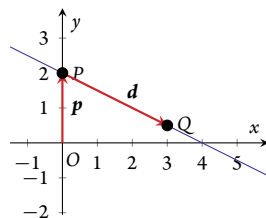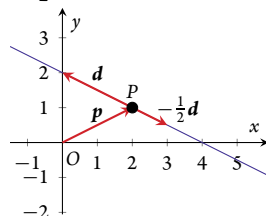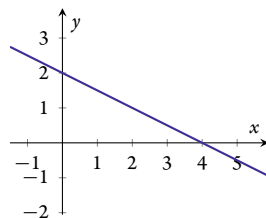
Other choices of points P and Q give other valid vector equations for a given line.

Activity 1.2.14 Which one of the following is *not* a valid vector equation for the line plotted to the right?

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

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

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

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

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

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

Example 1.2.16 Given that the line drawn to the right in space goes through points $(-4,-3,3)$ and $(3,2,1)$, find a parametric equation of the line.
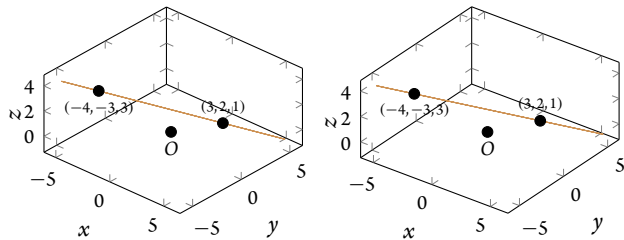
Solution: Let's call the points $(-4,-3,3)$ and $(3,2,1)$ as P and Q, respectively, and as shown below. First, choose a point on the line, say P, and set its position vector $p = \overrightarrow{OP} = (-4,-3,3) = -4i - 3j + 3k$, as drawn. Second, choose a direction vector to be, say, $d = \overrightarrow{PQ} = (3,2,1) - (-4,-3,3) = 7i + 5j - 2k$, also drawn. A parametric equation of the line is then $x = p + td$, specifically

$$x = (-4i - 3j + 3k) + t(7i + 5j - 2k)$$
$$= (-4 + 7t)i + (-3 + 5t)j + (3 - 2t)k.$$
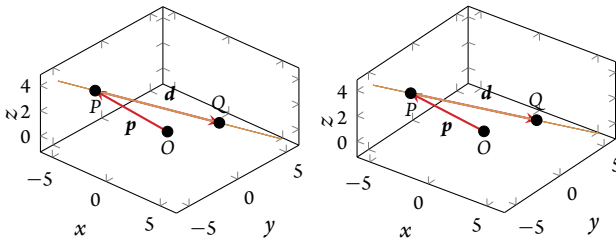
Example 1.2.17 Given the parametric equation of a line in space is $x = (-4 + 2t, 3 - t, -1 - 4t)$, find the value of the parameter t that gives each of the following points on the line: $(-1.6, 1.8, -5.8)$, $(-3, 2.5, -3)$, and $(-6, 4, 4)$.

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.

- 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.[6]

$\square$

1.2.4 Manipulation requires algebraic properties

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

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

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

(a) $u + v = v + u$ (commutative law);

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

(b) $(u + v) + w = u + (v + w)$ (associative law);

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

---

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

(c) $u + 0 = u$;

Solution: $u + 0 = (1,2) + (0,0) = (1+0, 2+0) = (1,2) = u.$ □

(d) $u + (-u) = 0$;

Solution: Recall $-u = (-1)u = (-1)(1,2) = (-1,-2)$, and so $u + (-u) = (1,2) + (-1,-2) = (1-1, 2-2) = (0,0) = 0.$ □

(e) $a(u + v) = au + av$ (a distributive law);

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

(f) $(a + b)u = au + bu$ (a distributive law);

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

(g) $(ab)u = a(bu)$;

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

(h) $1u = u$;

Solution: $1u = 1(1,2) = (1 \cdot 1, 1 \cdot 2) = (1,2) = u.$ □

(i) $0u = 0$;

Solution: $0u = 0(1,2) = (0 \cdot 1, 0 \cdot 2) = (0,0) = 0.$ □

(j) $|au| = |a| \cdot |u|$.

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

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

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

(a) $u + v = v + u$ (commutative law);

(b) $(u + v) + w = u + (v + w)$ (associative law);

(c) $u + 0 = 0 + u = u$;

(d) $u + (-u) = (-u) + u = 0$;

(e) $a(u + v) = au + av$ (a distributive law);

(f) $(a + b)u = au + bu$ (a distributive law);

(g) $(ab)u = a(bu)$;

(h) $1u = u$;

(i) $0u = 0$;

(j) $|au| = |a| \cdot |u|$.

Proof. We prove 1.2.19(a), 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 1.2.19(a) is the commutativity of vector addition. Example 1.2.1(a) shows graphically how the equality $u + v = v + u$ in just one case, and the diagram here shows another case. In general, let vectors $u = (u_1, u_2, \ldots, u_n)$ and $v = (v_1, v_2, \ldots, v_n)$ then
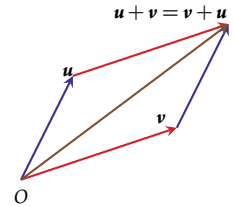
$u + v$

$= (u_1, u_2, \ldots, u_n) + (v_1, v_2, \ldots, v_n)$

$= (u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n)$ (by Definition 1.2.4)

$= (v_1 + u_1, v_2 + u_2, \ldots, v_n + u_n)$ (commutative scalar add)

$= (v_1, v_2, \ldots, v_n) + (u_1, u_2, \ldots, u_n)$ (by Definition 1.2.4)

$= v + u$.

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

Solution: The left diagram.
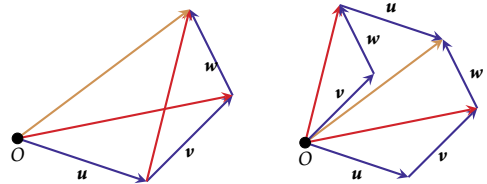
- In the left diagram, the two red vectors represent $u + v$ (left) and $v + w$ (right). Thus the left-red followed by the blue w represents $(u + v) + w$, whereas the u followed by the right-red represents $u + (v + w)$. The brown vector shows that they are equal: $(u + v) + w = u + (v + w)$.

- The right-hand of the two diagrams invokes the commutative law as well. The top-left part of the diagram shows $(v + w) + u$, whereas the bottom-right part shows $(u + v) + w$. That these are equal, the brown vector, requires both the commutative and associative laws.

$\square$

We frequently use the algebraic properties of Theorem 1.2.19 in rearranging and solving vector equations.

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

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

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

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

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

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

$\square$

1.2.5 Exercises

Exercise 1.2.1 For each of the pairs of vectors u and v shown below, draw the vectors $u + v$, $v + u$, $u - v$, $v - u$, $\tfrac{1}{2}u$, and $-v$.
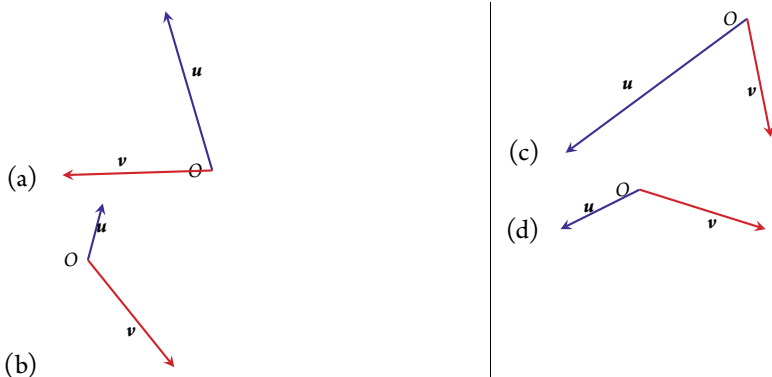
Exercise 1.2.2 For each of the following pairs of vectors shown below, use a ruler (or other measuring stick) to directly measure the distance between the pair of vectors.

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) $u = (-5, 0, 3)$, $v = (1, -6, 10)$, $w = (-4, 4, 11)$
(b) $u = 3i$, $v = 4i - 2j + 2k$, $w = 4i + 2j + 2k$
(c) $u = (-5, 3, 5, 6)$, $v = (-6, 1, 3, 10)$, $w = (-4, 6, 2, 15)$
(d) $u = 2e_1 + 4e_2 - e_3 + 5e_4$, $v = -2e_1 + 8e_2 - 6e_3 - 3e_4$, $w = -6e_3 + 11e_4$
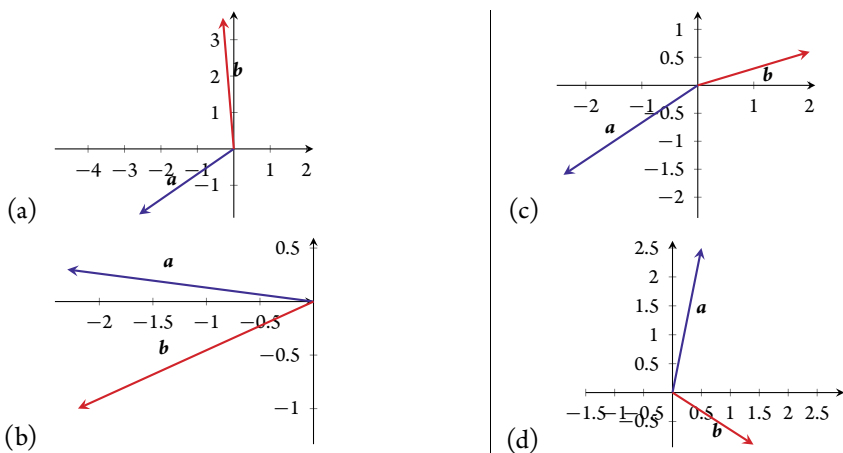
Exercise 1.2.4 Find a parametric equation of the line through the given pairs of points.

(a) $(-11, 0, 3)$, $(-3, -2, 2)$
(b) $(-4, 1, -2)$, $(3, -5, 5)$

(c) $(2.2, 5.8, 4, 3, 2)$,
$(-1.1, 2.2, -2.4, -3.2, 0.9)$
(d) $(1.8, -3.1, -1, -1.3, -3.3)$,
$(-1.4, 0.8, -2.6, 3.1, -0.8)$

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

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

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

(c) $u = (2, 1, 4, -2)$, $v = (-3, -2, 0, -1)$, $w = (-6, 5, 4, 2)$, $a = -4$, and $b = 3$.

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

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

(a) $x + a = 0$.

(b) $2x - b = 3b$.

(c) $3(x + a) = x + (a - 2x)$.

(d) $-4b = x + 3(a - x)$.

Exercise 1.2.8 In a few sentences, answer/discuss each of the following.

(a) What empowers us to write every vector in terms of the standard unit vectors?

(b) We use the distance $|u - v|$ to measure how close the two vectors are to each other. Invent an alternative way to measure closeness of two vectors, and comment on why your invented alternative measures closeness.

(c) What is it about the parametric equation of a line that means it does indeed describe a line in space?

(d) Comment on why many of the properties of vector operations (Theorem 1.2.19) appear the same as those for operations with real numbers.

1.3 The dot product determines angles and lengths

The previous Section 1.2 discussed how to add, subtract, and stretch vectors. Question: can we multiply two vectors? The answer is that 'vector multiplication' has major differences to the multiplication of scalar numbers. There are at least four ways of multiplying vectors together: each way useful in appropriate circumstances. Often the angle between vectors is denoted by the Greek letter theta, θ. This section introduces one such multiplication, the so-called dot product of two vectors that, among other attributes, gives a valuable way to determine the angle between two vectors.

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

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 plotted second. 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

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

(Recall that multiplication by $180/\pi$ converts an angle from radians to degrees: $1.3322 \cdot 180/\pi = 76.33°$.) Dividing by the product of the lengths then gives $\cos\theta = 9/(5\sqrt{58}) = 0.2364$, so the angle $\theta = \arccos(0.2364) = 1.3322 = 76.33°$ as is reasonable from the plots. $\square$
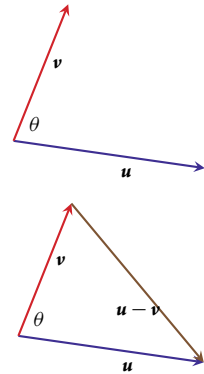
The interest in this Example 1.3.1 is the number 9 on the right-hand side of $|\boldsymbol{u}||\boldsymbol{v}|\cos\theta = 9$. The reason is that 9 just happens to be $14 - 5$, which in turn just happens to be $7 \cdot 2 + (-1) \cdot 5$, and it is no coincidence that this expression is the same as $u_1 v_1 + u_2 v_2$ in terms of vector components $\boldsymbol{u} = (u_1, u_2) = (7, -1)$ and $\boldsymbol{v} = (v_1, v_2) = (2, 5)$. Repeat this example for many pairs of vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ to find that always $|\boldsymbol{u}||\boldsymbol{v}|\cos\theta = u_1 v_1 + u_2 v_2$ (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.

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

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

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

When writing the vector dot product, the dot between the two vectors is essential. We sometimes also denote the scalar product by such a dot (to clarify a product) and sometimes omit the dot between the scalars, for example $a \cdot b = ab$ for scalars. But for the vector dot product, the dot must not be omitted: '$\boldsymbol{uv}$' is meaningless.

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

(a) $u = (-2, 5, -2), v = (3, 3, -2)$

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

(b) $u = (1, -3, 0), v = (1, 2)$

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

(c) $a = (-7, 3, 0, 2, 2), b = (-3, 4, -4, 2, 0)$

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

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

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

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

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

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

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

This picture illustrates the range of angles between two vectors: when they point in the same direction the angle is zero; when they are at right-angles to each other the angle is $\pi/2$, or equivalently 90°; when they point in opposite directions the angle is π, or equivalently 180°. Let's prove this theorem after some examples.
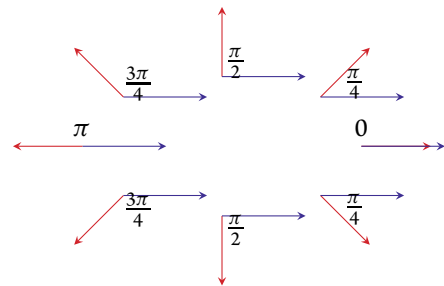
Example 1.3.6 Determine the angle between the following pairs of vectors.

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

Solution: These vectors (shown to the right) 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 $\cos\theta = 56/(5\cdot13)=0.8615$ and so angle $\theta=\arccos(0.8615)=0.5325=30.51°$. $\square$

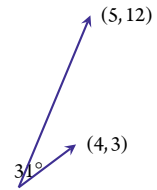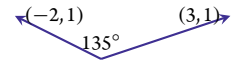(b) $(3,1)$ and $(-2,1)$

Solution: These vectors (shown to the right) 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). $\square$

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

Solution: These vectors (shown to the right) 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). $\square$
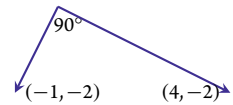
Activity 1.3.7 What is the angle between the two vectors $(1,\sqrt{3})$ and $(\sqrt{3},1)$?

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

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.

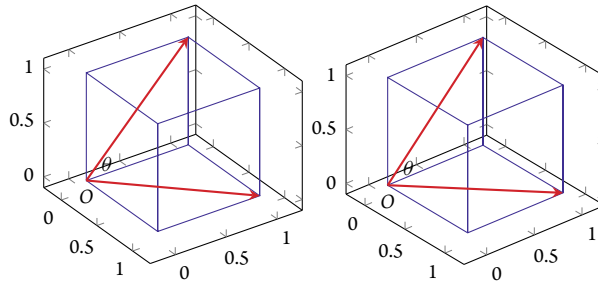**Table 1.1** When a cosine is one of these tabulated special values, then we know the corresponding angle exactly. In other cases, we usually use a calculator ($\texttt{arccos}$ or $\cos^{-1}$) or computer ($\texttt{acos()}$) to compute the angle numerically.

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 $\theta = \frac{\pi}{3} = 60°$. $\square$

(b) A body-centred cubic lattice (such as that formed by caesium chloride crystals) has one lattice point in the centre of the unit cell as well as the eight corner points. Consider the body-centred cube of atoms drawn in stereo below with the centre of the cube at the origin: what is the angle between the centre atom and any two adjacent corner atoms?
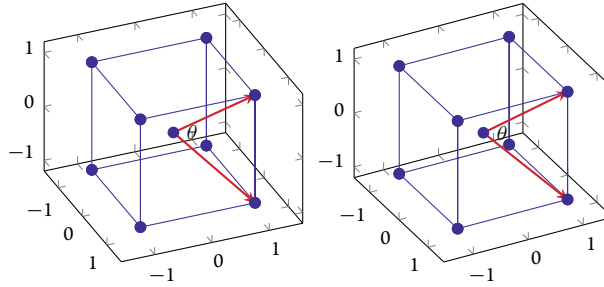
Solution: Draw two corresponding vectors from the centre atom: the above pair of vectors are $(1, 1, 1)$ and $(1, 1, -1)$. These have the same length $|(1, 1, 1)| = \sqrt{1^2 + 1^2 + 1^2} = \sqrt{3}$ and $|(1, 1, -1)| = \sqrt{1^2 + 1^2 + (-1)^2} = \sqrt{3}$. The dot product is $(1, 1, 1) \cdot (1, 1, -1) = 1 + 1 - 1 = 1$. 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 $\theta = \arccos(1/3) = 1.2310 = 70.53°$.

$\square$

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, $(N_{cat}, N_{dog}, N_{mat}, N_{sat}, N_{scratched})$ where the various N are the counts of each word in any sentence or document. For example,

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

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

Solution: First, these word vectors have lengths $|\boldsymbol{a}| = |\boldsymbol{b}| = \sqrt{3}$ and $|\boldsymbol{c}| = \sqrt{4} = 2$. Second, the 'angles' between these sentences are the following.

• 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 $\theta_{ab} = \arccos(1/3) = 1.2310 = 70.53°$ 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 $\theta_{ac} = \frac{\pi}{6} = 30°$ 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 $\theta_{bc} = \arccos(1/\sqrt{3}) = 0.9553 = 54.74°$ so the sentences are moderately dissimilar.

This 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).
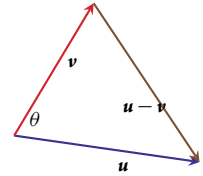

$\square$
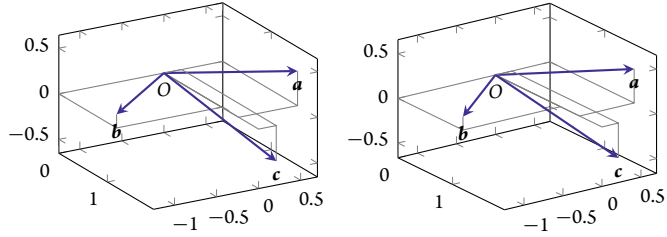
Proof. To prove the Angle Theorem 1.3.5, form a triangle from vectors $\boldsymbol{u}$, $\boldsymbol{v}$, and $\boldsymbol{u}-\boldsymbol{v}$ as illustrated to the right. Recall and apply the cosine rule for triangles

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

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

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

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

1.3.1 Work done involves the dot product

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

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

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

In general, the applied force and the displacement are not in the same direction (as illustrated to the right). Consider the general case when a vector force F acts on a body which moves a displacement vector d. Then the work done by the force on the body is the length of the displacement times the component of the force in the direction of the displacement—the component of the force at right-angles to the displacement does no work.

As illustrated to the right, draw a right-angled triangle to decompose the force 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 $W = F_0|d|$. As drawn, the force F makes an angle θ to the displacement d: the dot product determines this angle via $\cos\theta = (F \cdot d)/(|F||d|)$ (Theorem 1.3.5). By basic trigonometry, the adjacent side of the force triangle has length $F_0 = |F|\cos\theta = |F|\frac{F \cdot d}{|F||d|} = \frac{F \cdot d}{|d|}$. Finally, the work done $W = F_0|d| = \frac{F \cdot d}{|d|}|d| = F \cdot d$: that is, the work done is the dot product of the vector force and vector displacement.
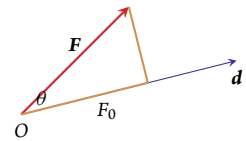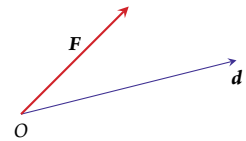
Example 1.3.10 A sailing boat travels a distance of 40 m east and 10 m north, as drawn to the right. The wind from abeam, of strength and direction $(1, -4)$ m/s, generates a force $F = (20, -10)$ (newtons) on the sail, as drawn. What is the work done by the wind?

Solution: The direction of the wind is immaterial except for the force it generates. The displacement vector $d = (40, 10)$ m. Then the work done is $W = F \cdot d = (40, 10) \cdot (20, -10) = 800 - 100 = 700$ joules. □
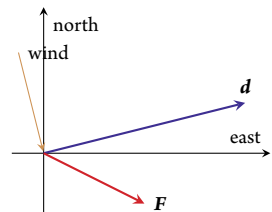
Activity 1.3.11 Recall the force of gravity on an object is the mass of the object times the acceleration of gravity, $9.8 \, \text{m/s}^2$. A 3 kg ball is thrown horizontally from a height of 2 m and lands 10 m away on the ground: what is the total work done by gravity on the ball?

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

Finding components of vectors in various directions is called projection. Such projection is surprisingly common in applications and is developed much further by Section 3.5.3.

1.3.2 Algebraic properties of the dot product

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

Example 1.3.12 Given vectors $\boldsymbol{u} = (-2, 5, -2)$, $\boldsymbol{v} = (3, 3, -2)$ and $\boldsymbol{w} = (2, 0, -5)$, and scalar $a = 2$, verify that (Theorems 1.3.13(c) and 1.3.13(d))

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

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

Solution:

- First, the following three evaluate to being equal:

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

- Second, the following two evaluate to being equal:

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

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

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

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

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

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

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

Proof. Here prove only the commutative law 1.3.13(a) and the inequality 1.3.13(e). Exercise 1.3.5 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 1.3.13(a) consider

$$
\begin{aligned}
\boldsymbol{u} \cdot \boldsymbol{v} &= u_1 v_1 + u_2 v_2 + \cdots + u_n v_n \quad \text{(by Definition 1.3.2)} \\
&= v_1 u_1 + v_2 u_2 + \cdots + v_n u_n \quad \text{(as each scalar multiplication commutes)} \\
&= \boldsymbol{v} \cdot \boldsymbol{u} \quad \text{(by Definition 1.3.2).}
\end{aligned}
$$

- To prove the inequality 1.3.13(e) consider

$$
\begin{aligned}
\boldsymbol{u} \cdot \boldsymbol{u} &= u_1 u_1 + u_2 u_2 + \cdots + u_n u_n \quad \text{(by Definition 1.3.2)} \\
&= u_1^2 + u_2^2 + \cdots + u_n^2 \\
&\geq 0 + 0 + \cdots + 0 \quad \text{(as each scalar term is } \geq 0) \\
&= 0.
\end{aligned}
$$

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

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

Second, suppose vector $\boldsymbol{u}$ in $\mathbb{R}^n$ satisfies $\boldsymbol{u} \cdot \boldsymbol{u} = 0$. Expanding the left-hand side in the components, $\boldsymbol{u} = (u_1, u_2, \ldots, u_n)$, gives 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 $\boldsymbol{0}$.

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

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

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

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

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

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

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

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

(b) $|u \cdot v| \le |u||v|$ (Cauchy–Schwarz inequality);

(c) $|u + v| \le |u| + |v|$ (triangle inequality).

Solution:

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

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

(c) Here $|u + v| = |(5, 5)| = \sqrt{5^2 + 5^2} = \sqrt{50} = 7.071$, whereas $|u| + |v| = 5 + \sqrt{5} = 7.236$. Hence $|u + v| = 7.071 \le 7.236 = |u| + |v|$. This is called the triangle inequality because the vectors u, v, and $u + v$ may be viewed as forming a triangle, as illustrated to the right, and this inequality follows because the length of a side of a triangle must be less than the sum of the lengths of the other two sides.
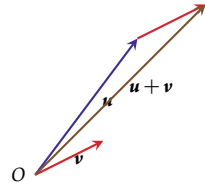
The Cauchy–Schwarz inequality is one point of distinction between this 'vector multiplication' and scalar multiplication: for scalars $|ab| = |a||b|$, whereas the dot product of vectors is typically less, $|u \cdot v| \le |u||v|$.

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 $u = (3, 4)$ and $v = (2, 1)$. Then consider the line given parametrically (Definition 1.2.15) as the position vectors $x = u + tv = (3 + 2t, 4 + t)$ for scalar parameter t— illustrated to the right. The position vector x of any point on the line has length ℓ (Definition 1.1.9) where
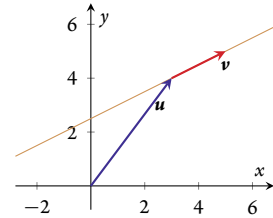
$$
\begin{aligned}
\ell^2 &= (3 + 2t)^2 + (4 + t)^2 \\
&= 9 + 12t + 4t^2 + 16 + 8t + t^2 \\
&= \underbrace{25}_{c} + \underbrace{20}_{b} t + \underbrace{5}_{a} t^2,
\end{aligned}
$$

a quadratic polynomial in t. We know that the length $\ell > 0$ (the line does not pass through the origin so no 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 $a = 5 = |v|^2$, $c = 25 = |u|^2$, and $b = 20 = 2 \cdot 10 = 2(u \cdot v)$. So $b^2 - 4ac < 0$, rewritten as $\frac{1}{4}b^2 < ac$, becomes the statement that $\frac{1}{4}[2(u \cdot v)]^2 = (u \cdot v)^2 < |v|^2|u|^2$. 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 u and v in $\mathbb{R}^n$ the following properties hold:*

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

(b) $|u \cdot v| \le |u||v|$ (***Cauchy–Schwarz inequality***);

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

Proof. Except for the first, each property depends upon the previous.

1.3.17(a) Expand and rearrange

$$\sqrt{u \cdot u} = \sqrt{u_1 u_1 + u_2 u_2 + \cdots + u_n u_n} \quad \text{(by Definition 1.3.2)}$$
$$= \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2} = |u| \quad \text{(by Definition 1.1.9).}$$

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

Second, for the case when $v \ne 0$, consider the line given parametrically by $x = u + tv$ for (real) scalar parameter t (Definition 1.2.15), as illustrated to the right. The distance ℓ of a point on the line from the origin is the length of its position vector, and by property 1.3.17(a)
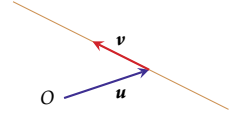
$$\begin{aligned}
\ell^2 &= x \cdot x \\
&= (u + tv) \cdot (u + tv) \\
&= u \cdot (u + tv) + (tv) \cdot (u + tv) \quad \text{(using distributivity 1.3.13(d))} \\
&= u \cdot u + u \cdot (tv) + (tv) \cdot u + (tv) \cdot (tv) \quad \text{(again using distributivity 1.3.13(d))} \\
&= u \cdot u + t(u \cdot v) + t(v \cdot u) + t^2(v \cdot v) \quad \text{(using scalar mult. property 1.3.13(c))} \\
&= |u|^2 + 2(u \cdot v)t + |v|^2 t^2 \quad \text{(using 1.3.17(a) and commutativity 1.3.13(a))} \\
&= at^2 + bt + c,
\end{aligned}$$

a quadratic in t, with coefficients $a = |v|^2 > 0$, $b = 2(u \cdot v)$, and $c = |u|^2$. 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 $\frac{1}{4}[2(u \cdot v)]^2 = (u \cdot v)^2 \le |v|^2 |u|^2$. Taking the square-root of both sides then establishes the Cauchy–Schwarz inequality $|u \cdot v| \le |u||v|$.

1.3.17(c) To prove the triangle inequality between vectors u and v first observe the Cauchy–Schwarz inequality implies $(u \cdot v) \le |u||v|$ (since the left-hand side has magnitude $\le$ the right-hand side). Then consider (analogous to the $t = 1$ case of the above)
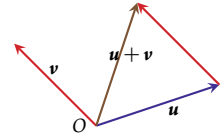
$$|u+v|^2 = (u+v) \cdot (u+v)$$
$$= u \cdot (u+v) + v \cdot (u+v) \quad \text{(using distributivity 1.3.13(d))}$$
$$= u \cdot u + u \cdot v + v \cdot u + v \cdot v \quad \text{(again using distributivity 1.3.13(d))}$$
$$= |u|^2 + 2(u \cdot v) + |v|^2 \quad \text{(using 1.3.17(a) and commutativity 1.3.13(a))}$$
$$\leq |u|^2 + 2|u||v| + |v|^2 \quad \text{(using Cauchy–Schwarz inequality)}$$
$$= (|u| + |v|)^2.$$

Take the square-root of both sides to establish the triangle inequality $|u+v| \leq |u| + |v|$. The minus case follows (illustrated) because $|u - v| = |u + (-v)| \leq |u| + |-v| = |u| + |v|$.
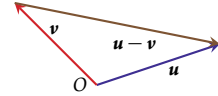
Example 1.3.18 Verify the Cauchy–Schwarz inequality and the triangle inequality ($+$ case) for the vectors $a = (-1, -2, 1, 3, -2)$ and $b = (-3, -2, 10, 2, 2)$.

Solution: We need the length of the vectors, and also the dot product:

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

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

Now, the length of the vector sum

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

Here $|a+b| = 13.342$ whereas $|a| + |b| = 11 + \sqrt{19} = 15.359$. Hence these indeed satisfy the triangle inequality $|a+b| \leq |a| + |b|$. $\quad \square$

1.3.3 Orthogonal vectors are at right-angles

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

Definition 1.3.19 *Two vectors **u** and **v** in $\mathbb{R}^n$ are termed **orthogonal** (or **perpendicular**) if and only if their dot product **u** · **v** = 0.*[7]

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

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

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

By Definition 1.3.19 these are orthogonal to each other.

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

Solution: Is the dot product zero? or not?

- $\mathbf{u} \cdot \mathbf{v} = (-1,1,-3,0) \cdot (2,4,2,-6) = -2 + 4 - 6 + 0 = -4 \neq 0$ so this pair are not perpendicular.

- $\mathbf{u} \cdot \mathbf{w} = (-1,1,-3,0) \cdot (-1,6,-2,3) = 1 + 6 + 6 + 0 = 13 \neq 0$ so this pair are not perpendicular.

- $\mathbf{v} \cdot \mathbf{w} = (2,4,2,-6) \cdot (-1,6,-2,3) = -2 + 24 - 4 - 18 = 0$ so this pair of vectors are perpendicular to each other. □

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

(a) no pair   (b) **x**, **y**   (c) **y**, **z**   (d) **x**, **z**

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

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

---

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

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

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 $\boldsymbol{i} + 4\boldsymbol{j} + 2\boldsymbol{k}$. $\qquad\square$

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

To introduce the first theorem, consider the 2D plane and try to draw a nonzero vector at right-angles to both the two standard unit vectors $\boldsymbol{i}$ and $\boldsymbol{j}$. The red vectors to the right illustrate failed attempts to draw a nonzero vector at right-angles to both $\boldsymbol{i}$ and $\boldsymbol{j}$. It cannot be done. No vector in the plane can be at right-angles to both the standard unit vectors in the plane.
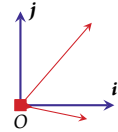
Theorem 1.3.24 *There is no nonzero vector orthogonal to all n standard unit vectors in $\mathbb{R}^n$.*

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

- $0 = \boldsymbol{u} \cdot \boldsymbol{e}_1 = (u_1, u_2, \ldots, u_n) \cdot (1, 0, \ldots, 0) = u_1 + 0 + \cdots + 0 = u_1$, and so the first component must be zero;
- $0 = \boldsymbol{u} \cdot \boldsymbol{e}_2 = (u_1, u_2, \ldots, u_n) \cdot (0, 1, \ldots, 0) = 0 + u_2 + 0 + \cdots + 0 = u_2$, and so the second component must be zero;
- and so on to
- $0 = \boldsymbol{u} \cdot \boldsymbol{e}_n = (u_1, u_2, \ldots, u_n) \cdot (0, 0, \ldots, 1) = 0 + 0 + \cdots + u_n = u_n$, 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 picture to the right illustrates one attempt. It cannot be done. There are at most two vectors in 2D that are all at right-angles to each other.
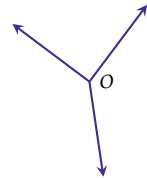
Theorem 1.3.25 (*orthogonal completeness*) *In a set of orthogonal unit vectors in $\mathbb{R}^n$, there can be no more than n vectors in the set.*[8]

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 $\mathbb{R}^2$ to the right). Theorem 1.3.24 then says there cannot be any more nonzero 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 $\mathbb{R}^n$.

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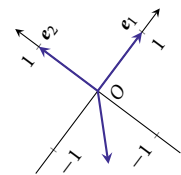
[8] For the pure at heart, this property is part of the definition of what we mean by $\mathbb{R}^n$. The representation of a vector in $\mathbb{R}^n$ by n components (here Definition 1.1.4) then follows as a consequence, instead of vice versa as here.

### 1.3.4 Normal vectors and equations of a plane

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

Example 1.3.26 First find the equation of the line that is perpendicular to the vector $(2,3)$ and that passes through the origin. Second, find the equation of the line that passes through the point $(4,1)$ (instead of the origin).
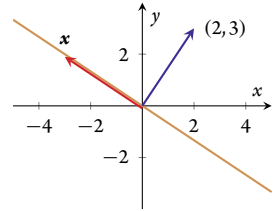
Solution: Recall that vectors at right-angles have a zero dot product (Section 1.3.3). Thus the position vector $\boldsymbol{x}$ of every point in the line satisfies the dot product $\boldsymbol{x} \cdot (2,3) = 0$. For $\boldsymbol{x} = (x,y)$, as illustrated to the above right, $\boldsymbol{x} \cdot (2,3) = 2x + 3y$ 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 vector $\boldsymbol{x} - (4,1)$ 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. $\square$
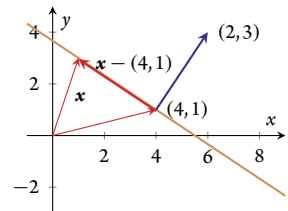
Activity 1.3.27 What is an equation of the line that is both through the point $(4,2)$, and at right-angles to the vector $(1,3)$?

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

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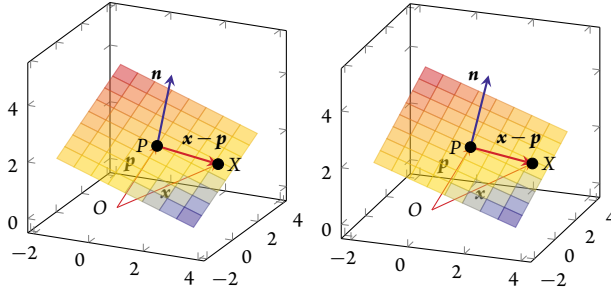rrow{PX}$ is orthogonal to $\boldsymbol{n}$. Denote the position vector of P by $\boldsymbol{p} = (x_0, y_0, z_0)$, the position vector of X by $\boldsymbol{x} = (x, y, z)$, and let the normal vector be $\boldsymbol{n} = (a, b, c)$. Then, as drawn below, the displacement vector $\overrightarrow{PX} =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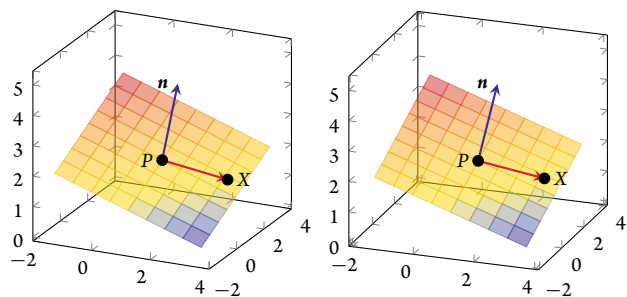
$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0,$$

equivalently, an equation of the plane is

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

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

Solution: Letting $x = (x,y,z)$ be the coordinates of a point in the plane, the above argument asserts an equation of the plane is $n \cdot (x - \overrightarrow{OP}) = 0$ 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) $3x - 6y + 2z = 4$;

(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, $n = (3,-6,2)$ (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 $n = (-0.2, 3.3, 1)$ (or any scalar multiple).

□

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

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

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

The basic idea, as illustrated to the right, is that given any plane (through the origin for the moment), then choosing almost any two vectors in the plane allows us to write all points in the plane as a sum of multiples of the two vectors. With the given vectors u and v shown to the right, illustrated are the points $u + 2v$, $\frac{1}{2}u - 2v$, and $-2u + 3v$. Similarly, every point in the plane has a position vector in the form $su + tv$ for some scalar parameters s and t. The grid shown to the right illustrates the sum of integral and half-integral multiples. The formula $x = su + tv$ for parameters s and t is called a parametric equation of the plane.
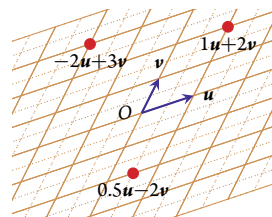
Example 1.3.31 Find a parametric equation of the plane that passes through the three points $P=(-1,2,3)$, $Q=(2,3,2)$, and $R=(0,4,5)$, drawn to the right in stereo.
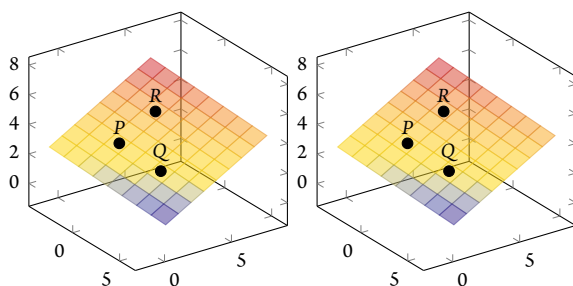
Solution: This plane does not pass through the origin, so we first choose a point and make the description relative to that point: say we choose the point P with position vector $p = \overrightarrow{OP} = -i + 2j + 3k$. Then, as illustrated below, two vectors parallel to the required plane are

$$
\begin{aligned}
u &= \overrightarrow{PQ} = \overrightarrow{OQ} - \overrightarrow{OP} \\
&= (2i + 3j + 2k) - (-i + 2j + 3k) \\
&= 3i + j - k, \\
v &= \overrightarrow{PR} = \overrightarrow{OR} - \overrightarrow{OP} \\
&= (4j + 5k) - (-i + 2j + 3k) \\
&= i + 2j + 2k.
\end{aligned}
$$
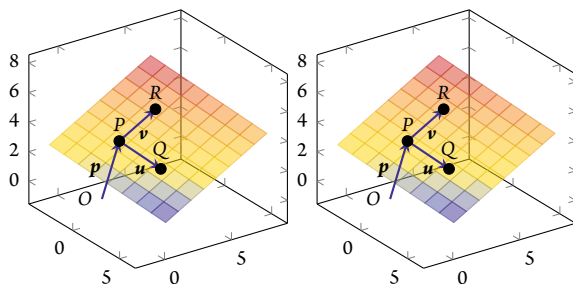
Lastly, every point in the plane is the sum of the displacement vector p and arbitrary multiples of the parallel vectors u and v. That is, a parametric equation of the plane is $x = p + su + tv$ which here is

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

□

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

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

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

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

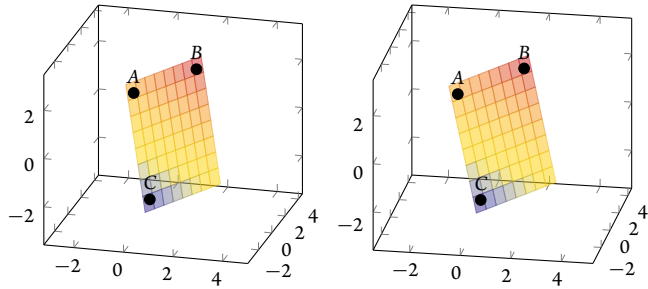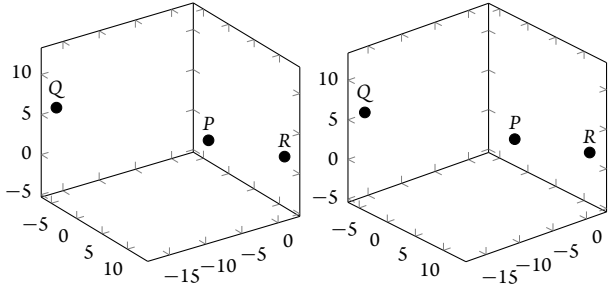$$u = \overrightarrow{PQ} = \overrightarrow{OQ} - \overrightarrow{OP} = (-4i - 18j + 7k) - (6i - 4j + 3k) = -10i - 14j + 4k,$$
$$v = \overrightarrow{PR} = \overrightarrow{OR} - \overrightarrow{OP} = (11i + 3j + k) - (6i - 4j + 3k) = 5i + 7j - 2k.$$

Oops: notice that $u = -2v$ so the vectors u and 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 to the right in stereo.

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

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

$$= (0.2i - 0.4j - 2.5k) - (-1.2i + 2.4j + 0.8k)$$
$$= 1.4i - 2.8j - 3.3k.$$
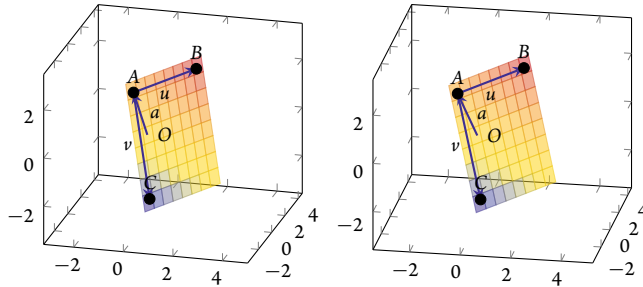
Lastly, every point in the plane is the sum of the displacement vector a, and some multiples of the 'parallel' vectors u and v. That is, a parametric equation of the plane is $x = a + su + tv$ which here is

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

□

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

(a) $i + sj + tk$

(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 $|u||v| \cos\theta = u \cdot v$ 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)
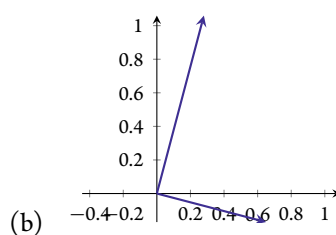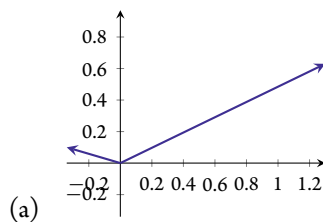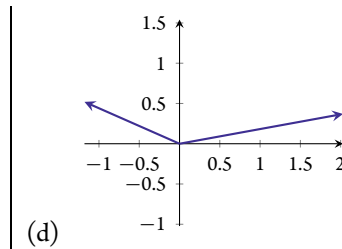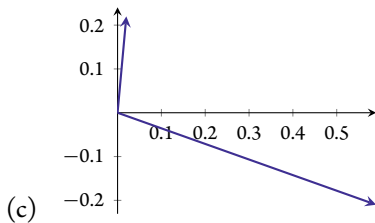
Exercise 1.3.3 Recall that Example 1.1.7 represented the following sentences by word vectors $w = (N_{cat}, N_{dog}, N_{mat}, N_{sat}, N_{scratched})$.

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

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 Suppose two nonzero word vectors are orthogonal. Explain what such orthogonality means in terms of the words of the original sentences.

Exercise 1.3.5 For the properties of the dot product, Theorem 1.3.13, prove some properties chosen from 1.3.13(b)–1.3.13(d).

Exercise 1.3.6 Verify the Cauchy–Schwarz inequality and also the triangle inequality ($+$ case) for the following pairs of vectors.

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

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

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

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

Exercise 1.3.7 Find an equation of the plane with the given normal vector n and through the given point P.

(a) $P = (1,2,-3), n = (2,-5,-2)$.

(b) $P = (5,-4,-13), n = (-1,0,-1)$.

(c) $P = (-7.3,-1.6,5.8),$
$n = (-2.8,-0.8,4.4)$.

(d) $P = (0,-1.2,2.2),$
$n = (-1.4,-8.1,-1.5)$.

Exercise 1.3.8 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) $0.1x = 1.5y + 1.1z + 0.7$

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

Exercise 1.3.9 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) $(-5.6, -2.2, -6.8), (-1.8, 4.3, -3.9), (2.5, -3.5, -1.7)$.

(d) $(1.8, -0.2, -0.7), (-1.6, 2, -3.7), (1.4, -0.5, 0.5)$.

Exercise 1.3.10 For each case of Exercise 1.3.9 that you have done, find two other parametric equations of the plane.

Exercise 1.3.11 In a few sentences, answer/discuss each of the following.

(a) When using the dot product to determine the angle between a pair of vectors we only discuss angles between $0°$ and $180°$ (0 and π radians). Why do we not discuss larger angles, such as $246°$ or $315°$? nor negative angles?

(b) What properties of the dot product differ from that of the multiplication of scalar numbers?

(c) Describe a geometric reason for the Cauchy–Schwarz inequality.

(d) Why do we phrase an equation for a plane in terms of its perpendicular vector?

(e) Given that $x = p + td$ parametrizes a line, and that $x = p + sc + td$ parametrizes a plane, what would $x = p + rb + sc + td$ describe? why? are there any provisos?

1.4 The cross product

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

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

Area of a parallelogram

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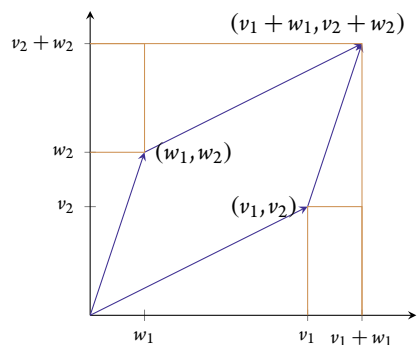
$$
\begin{aligned}
\text{area} &= (v_1 + w_1)(v_2 + w_2) - 2 w_1 v_2 - 2 \cdot \frac{1}{2} w_1 w_2 - 2 \cdot \frac{1}{2} v_1 v_2 \\
&= v_1 v_2 + v_1 w_2 + w_1 v_2 + w_1 w_2 - 2 w_1 v_2 - w_1 w_2 - v_1 v_2 \\
&= v_1 w_2 - v_2 w_1.
\end{aligned}
$$

In application, sometimes this right-hand side expression is negative because vectors v and w are the 'wrong way' around. Thus in general the parallelogram area $= |v_1 w_2 - v_2 w_1|$.

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

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 the parts about to form a rectangle roughly 3 by 5, and hence the area should be roughly 15.

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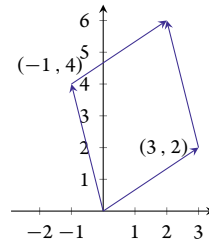
Activity 1.4.2 What is the area of the parallelogram (illustrated to the right) whose edges are formed by the vectors $(5,3)$ and $(2,-2)$?

(a) 16   (b) 11   (c) 4   (d) 19
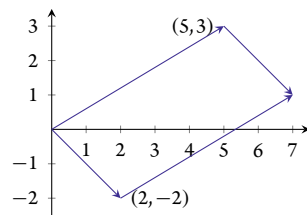
Interestingly, we meet this expression for area, $v_1 w_2 - v_2 w_1$, in another context: that of equations for a plane and its normal vector.

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 $x = (1,1,2) + (1,1,0)s + (0,3,1)t$. These determine the same plane; they are just different algebraic descriptions. One converts between these two descriptions using the cross product.

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

Solution: The 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 $(0,3,1)$. The following mysterious looking procedure may be a convenient way for you to remember an otherwise involved formula: if you prefer to remember the formula of Definition 1.4.5, then use that instead. (Those who have computed 3×3 determinants may recognize that 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 i, j, and k, in

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

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

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

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

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

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

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

$$= i - j + 3k.$$

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 $x = (4, -1, -2) + (1, -2, 1)s + (2, -3, -2)t$. Which of the following is your computed normal vector?

(a) $(7, 4, 1)$ (b) $(-4, 4, -10)$ (c) $(2, -2, 5)$ (d) $(5, 6, 7)$

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

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

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

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

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

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

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

$$= i(v_2 w_3 - v_3 w_2) - j(v_1 w_3 - v_3 w_1) + k(v_1 w_2 - v_2 w_1).$$

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

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

$$v \times w := i(v_2 w_3 - v_3 w_2) + j(v_3 w_1 - v_1 w_3) + k(v_1 w_2 - v_2 w_1).$$

Example 1.4.6 Among the standard unit vectors, derive that

(a) $i \times j = k$,

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

(c) $j \times k = i$,

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

(e) $k \times i = j$,

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

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

Solution: Using Definition 1.4.5:

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

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

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

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

(a) $(-3, -6, 6)$ (b) $(3, -6, -6)$ (c) $(-3, -6, 6)$ (d) $(3, 6, -6)$
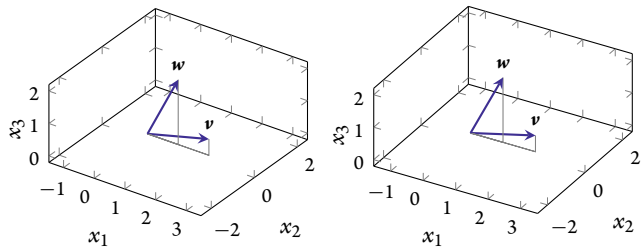
Geometry of a cross product

Example 1.4.8 *(parallelogram area)* Let's revisit the introduction to this section. Consider the parallelogram in the $x_1 x_2$-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 $= |v_1 w_2 - v_2 w_1|$. Compare this area with the cross product

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

Consequently, the length of this cross product equals the area of the parallelogram formed by $\boldsymbol{v}$ and $\boldsymbol{w}$ (Theorem 1.4.10(d)). (Also the direction of the cross product, $\pm \boldsymbol{k}$, is orthogonal to the $x_1 x_2$-plane containing the two vectors—Theorem 1.4.10(a)).

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

(a) $+\boldsymbol{i}$ (c) $+\boldsymbol{j}$

(b) $-\boldsymbol{j}$ (d) $-\boldsymbol{i}$

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

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

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

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

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